Andrei D. Polyanin
Alexander V. Manzhirov

HANDBOOK OF
MATHEMATICS FOR ENGINEERS AND SCIENTISTS

Chapman & Hall/CRC
Taylor & Francis Group

HANDBOOK OF

MATHEMATICS FOR ENGINEERS AND SCIENTISTS

HANDBOOK OF
MATHEMATICS FOR ENGINEERS AND SCIENTISTS

Andrei D. Polyanin
Alexander V. Manzhirov

Chapman & Hall/CRC
Taylor & Francis Group
Boca Raton London New York

Chapman & Hall/CRC is an imprint of the
Taylor & Francis Group, an informa business

Chapman & Hall/CRC
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2007 by Taylor & Francis Group, LLC
Chapman & Hall/CRC is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-10: 1-58488-502-5 (Hardcover)
International Standard Book Number-13: 978-1-58488-502-3 (Hardcover)

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

No part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC) 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

CONTENTS

Authors . . . . . **xxv**

Preface . . . . . **xxvii**

Main Notation . . . . . **xxix**

Part I. Definitions, Formulas, Methods, and Theorems **1**

1. Arithmetic and Elementary Algebra **3**

1.1. Real Numbers 3
1.1.1. Integer Numbers 3
1.1.2. Real, Rational, and Irrational Numbers 4

1.2. Equalities and Inequalities. Arithmetic Operations. Absolute Value 5
1.2.1. Equalities and Inequalities 5
1.2.2. Addition and Multiplication of Numbers 6
1.2.3. Ratios and Proportions 6
1.2.4. Percentage 7
1.2.5. Absolute Value of a Number (Modulus of a Number) 8

1.3. Powers and Logarithms 8
1.3.1. Powers and Roots 8
1.3.2. Logarithms 9

1.4. Binomial Theorem and Related Formulas 10
1.4.1. Factorials. Binomial Coefficients. Binomial Theorem 10
1.4.2. Related Formulas 10

1.5. Arithmetic and Geometric Progressions. Finite Sums and Products 11
1.5.1. Arithmetic and Geometric Progressions 11
1.5.2. Finite Series and Products 12

1.6. Mean Values and Inequalities of General Form 13
1.6.1. Arithmetic Mean, Geometric Mean, and Other Mean Values. Inequalities for Mean Values 13
1.6.2. Inequalities of General Form 14

1.7. Some Mathematical Methods 15
1.7.1. Proof by Contradiction 15
1.7.2. Mathematical Induction 16
1.7.3. Proof by Counterexample 17
1.7.4. Method of Undetermined Coefficients 17

References for Chapter 1 18

2. Elementary Functions **19**

2.1. Power, Exponential, and Logarithmic Functions 19
2.1.1. Power Function: $y = x^{\alpha}$ 19
2.1.2. Exponential Function: $y = a^{x}$ 21
2.1.3. Logarithmic Function: $y = \log_a x$ 22

2.2. Trigonometric Functions 24
2.2.1. Trigonometric Circle. Definition of Trigonometric Functions 24
2.2.2. Graphs of Trigonometric Functions 25
2.2.3. Properties of Trigonometric Functions 27

2.3. Inverse Trigonometric Functions 30
2.3.1. Definitions. Graphs of Inverse Trigonometric Functions 30
2.3.2. Properties of Inverse Trigonometric Functions 33
2.4. Hyperbolic Functions 34
2.4.1. Definitions. Graphs of Hyperbolic Functions 34
2.4.2. Properties of Hyperbolic Functions 36
2.5. Inverse Hyperbolic Functions 39
2.5.1. Definitions. Graphs of Inverse Hyperbolic Functions 39
2.5.2. Properties of Inverse Hyperbolic Functions 41
References for Chapter 2 42

3. Elementary Geometry 43

3.1. Plane Geometry 43
3.1.1. Triangles 43
3.1.2. Polygons 51
3.1.3. Circle 56
3.2. Solid Geometry 59
3.2.1. Straight Lines, Planes, and Angles in Space 59
3.2.2. Polyhedra 61
3.2.3. Solids Formed by Revolution of Lines 65
3.3. Spherical Trigonometry 70
3.3.1. Spherical Geometry 70
3.3.2. Spherical Triangles 71
References for Chapter 3 75

4. Analytic Geometry 77

4.1. Points, Segments, and Coordinates on Line and Plane 77
4.1.1. Coordinates on Line 77
4.1.2. Coordinates on Plane 78
4.1.3. Points and Segments on Plane 81
4.2. Curves on Plane 84
4.2.1. Curves and Their Equations 84
4.2.2. Main Problems of Analytic Geometry for Curves 88
4.3. Straight Lines and Points on Plane 89
4.3.1. Equations of Straight Lines on Plane 89
4.3.2. Mutual Arrangement of Points and Straight Lines 93
4.4. Second-Order Curves 97
4.4.1. Circle 97
4.4.2. Ellipse 98
4.4.3. Hyperbola 101
4.4.4. Parabola 104
4.4.5. Transformation of Second-Order Curves to Canonical Form 107
4.5. Coordinates, Vectors, Curves, and Surfaces in Space 113
4.5.1. Vectors. Cartesian Coordinate System 113
4.5.2. Coordinate Systems 114
4.5.3. Vectors. Products of Vectors 120
4.5.4. Curves and Surfaces in Space 123

4.6. Line and Plane in Space 124
4.6.1. Plane in Space 124
4.6.2. Line in Space 131
4.6.3. Mutual Arrangement of Points, Lines, and Planes 135
4.7. Quadric Surfaces (Quadrics) 143
4.7.1. Quadrics (Canonical Equations) 143
4.7.2. Quadrics (General Theory) 148
References for Chapter 4 153

5. Algebra 155

5.1. Polynomials and Algebraic Equations 155
5.1.1. Polynomials and Their Properties 155
5.1.2. Linear and Quadratic Equations 157
5.1.3. Cubic Equations 158
5.1.4. Fourth-Degree Equation 159
5.1.5. Algebraic Equations of Arbitrary Degree and Their Properties 161
5.2. Matrices and Determinants 167
5.2.1. Matrices 167
5.2.2. Determinants 175
5.2.3. Equivalent Matrices. Eigenvalues 180
5.3. Linear Spaces 187
5.3.1. Concept of a Linear Space. Its Basis and Dimension 187
5.3.2. Subspaces of Linear Spaces 190
5.3.3. Coordinate Transformations Corresponding to Basis Transformations in a Linear Space 191
5.4. Euclidean Spaces 192
5.4.1. Real Euclidean Space 192
5.4.2. Complex Euclidean Space (Unitary Space) 195
5.4.3. Banach Spaces and Hilbert Spaces 196
5.5. Systems of Linear Algebraic Equations 197
5.5.1. Consistency Condition for a Linear System 197
5.5.2. Finding Solutions of a System of Linear Equations 198
5.6. Linear Operators 204
5.6.1. Notion of a Linear Operator. Its Properties 204
5.6.2. Linear Operators in Matrix Form 208
5.6.3. Eigenvectors and Eigenvalues of Linear Operators 209
5.7. Bilinear and Quadratic Forms 213
5.7.1. Linear and Sesquilinear Forms 213
5.7.2. Bilinear Forms 214
5.7.3. Quadratic Forms 216
5.7.4. Bilinear and Quadratic Forms in Euclidean Space 219
5.7.5. Second-Order Hypersurfaces 220
5.8. Some Facts from Group Theory 225
5.8.1. Groups and Their Basic Properties 225
5.8.2. Transformation Groups 228
5.8.3. Group Representations 230
References for Chapter 5 233

6. Limits and Derivatives **235**

6.1. Basic Concepts of Mathematical Analysis 235
6.1.1. Number Sets. Functions of Real Variable 235
6.1.2. Limit of a Sequence 237
6.1.3. Limit of a Function. Asymptotes 240
6.1.4. Infinitely Small and Infinitely Large Functions 242
6.1.5. Continuous Functions. Discontinuities of the First and the Second Kind 243
6.1.6. Convex and Concave Functions 245
6.1.7. Functions of Bounded Variation 246
6.1.8. Convergence of Functions 249

6.2. Differential Calculus for Functions of a Single Variable 250
6.2.1. Derivative and Differential, Their Geometrical and Physical Meaning 250
6.2.2. Table of Derivatives and Differentiation Rules 252
6.2.3. Theorems about Differentiable Functions. L'Hospital Rule 254
6.2.4. Higher-Order Derivatives and Differentials. Taylor's Formula 255
6.2.5. Extremal Points. Points of Inflection 257
6.2.6. Qualitative Analysis of Functions and Construction of Graphs 259
6.2.7. Approximate Solution of Equations (Root-Finding Algorithms for Continuous Functions) 260

6.3. Functions of Several Variables. Partial Derivatives 263
6.3.1. Point Sets. Functions. Limits and Continuity 263
6.3.2. Differentiation of Functions of Several Variables 264
6.3.3. Directional Derivative. Gradient. Geometrical Applications 267
6.3.4. Extremal Points of Functions of Several Variables 269
6.3.5. Differential Operators of the Field Theory 272

References for Chapter 6 272

7. Integrals **273**

7.1. Indefinite Integral 273
7.1.1. Antiderivative. Indefinite Integral and Its Properties 273
7.1.2. Table of Basic Integrals. Properties of the Indefinite Integral. Integration Examples 274
7.1.3. Integration of Rational Functions 276
7.1.4. Integration of Irrational Functions 279
7.1.5. Integration of Exponential and Trigonometric Functions 281
7.1.6. Integration of Polynomials Multiplied by Elementary Functions 283

7.2. Definite Integral 286
7.2.1. Basic Definitions. Classes of Integrable Functions. Geometrical Meaning of the Definite Integral 286
7.2.2. Properties of Definite Integrals and Useful Formulas 287
7.2.3. General Reduction Formulas for the Evaluation of Integrals 289
7.2.4. General Asymptotic Formulas for the Calculation of Integrals 290
7.2.5. Mean Value Theorems. Properties of Integrals in Terms of Inequalities. Arithmetic Mean and Geometric Mean of Functions 295
7.2.6. Geometric and Physical Applications of the Definite Integral 299
7.2.7. Improper Integrals with Infinite Integration Limit 301
7.2.8. General Reduction Formulas for the Calculation of Improper Integrals 304
7.2.9. General Asymptotic Formulas for the Calculation of Improper Integrals 307
7.2.10. Improper Integrals of Unbounded Functions 308
7.2.11. Cauchy-Type Singular Integrals 310

7.2.12. Stieltjes Integral 312
7.2.13. Square Integrable Functions 314
7.2.14. Approximate (Numerical) Methods for Computation of Definite Integrals 315
7.3. Double and Triple Integrals 317
7.3.1. Definition and Properties of the Double Integral 317
7.3.2. Computation of the Double Integral 319
7.3.3. Geometric and Physical Applications of the Double Integral 323
7.3.4. Definition and Properties of the Triple Integral 324
7.3.5. Computation of the Triple Integral. Some Applications. Iterated Integrals and Asymptotic Formulas 325
7.4. Line and Surface Integrals 329
7.4.1. Line Integral of the First Kind 329
7.4.2. Line Integral of the Second Kind 330
7.4.3. Surface Integral of the First Kind 332
7.4.4. Surface Integral of the Second Kind 333
7.4.5. Integral Formulas of Vector Calculus 334
References for Chapter 7 335

8. Series 337

8.1. Numerical Series and Infinite Products 337
8.1.1. Convergent Numerical Series and Their Properties. Cauchy's Criterion 337
8.1.2. Convergence Criteria for Series with Positive (Nonnegative) Terms 338
8.1.3. Convergence Criteria for Arbitrary Numerical Series. Absolute and Conditional Convergence 341
8.1.4. Multiplication of Series. Some Inequalities 343
8.1.5. Summation Methods. Convergence Acceleration 344
8.1.6. Infinite Products 346
8.2. Functional Series 348
8.2.1. Pointwise and Uniform Convergence of Functional Series 348
8.2.2. Basic Criteria of Uniform Convergence. Properties of Uniformly Convergent Series 349
8.3. Power Series 350
8.3.1. Radius of Convergence of Power Series. Properties of Power Series 350
8.3.2. Taylor and Maclaurin Power Series 352
8.3.3. Operations with Power Series. Summation Formulas for Power Series 354
8.4. Fourier Series 357
8.4.1. Representation of 2π-Periodic Functions by Fourier Series. Main Results 357
8.4.2. Fourier Expansions of Periodic, Nonperiodic, Odd, and Even Functions 359
8.4.3. Criteria of Uniform and Mean-Square Convergence of Fourier Series 361
8.4.4. Summation Formulas for Trigonometric Series 362
8.5. Asymptotic Series 363
8.5.1. Asymptotic Series of Poincaré Type. Formulas for the Coefficients 363
8.5.2. Operations with Asymptotic Series 364
References for Chapter 8 366

9. Differential Geometry 367

9.1. Theory of Curves 367
9.1.1. Plane Curves 367
9.1.2. Space Curves 379

9.2. Theory of Surfaces 386
9.2.1. Elementary Notions in Theory of Surfaces 386
9.2.2. Curvature of Curves on Surface 392
9.2.3. Intrinsic Geometry of Surface 395
References for Chapter 9 397
10. Functions of Complex Variable 399
10.1. Basic Notions 399
10.1.1. Complex Numbers. Functions of Complex Variable 399
10.1.2. Functions of Complex Variable 401
10.2. Main Applications 419
10.2.1. Conformal Mappings 419
10.2.2. Boundary Value Problems 427
References for Chapter 10 433
11. Integral Transforms 435
11.1. General Form of Integral Transforms. Some Formulas 435
11.1.1. Integral Transforms and Inversion Formulas 435
11.1.2. Residues. Jordan Lemma 435
11.2. Laplace Transform 436
11.2.1. Laplace Transform and the Inverse Laplace Transform 436
11.2.2. Main Properties of the Laplace Transform. Inversion Formulas for Some Functions 437
11.2.3. Limit Theorems. Representation of Inverse Transforms as Convergent Series and Asymptotic Expansions 440
11.3. Mellin Transform 441
11.3.1. Mellin Transform and the Inversion Formula 441
11.3.2. Main Properties of the Mellin Transform. Relation Among the Mellin, Laplace, and Fourier Transforms 442
11.4. Various Forms of the Fourier Transform 443
11.4.1. Fourier Transform and the Inverse Fourier Transform 443
11.4.2. Fourier Cosine and Sine Transforms 445
11.5. Other Integral Transforms 446
11.5.1. Integral Transforms Whose Kernels Contain Bessel Functions and Modified Bessel Functions 446
11.5.2. Summary Table of Integral Transforms. Areas of Application of Integral Transforms 448
References for Chapter 11 451
12. Ordinary Differential Equations 453
12.1. First-Order Differential Equations 453
12.1.1. General Concepts. The Cauchy Problem. Uniqueness and Existence Theorems 453
12.1.2. Equations Solved for the Derivative. Simplest Techniques of Integration 456
12.1.3. Exact Differential Equations. Integrating Factor 458
12.1.4. Riccati Equation 460
12.1.5. Abel Equations of the First Kind 462
12.1.6. Abel Equations of the Second Kind 464
12.1.7. Equations Not Solved for the Derivative 465
12.1.8. Contact Transformations 468
12.1.9. Approximate Analytic Methods for Solution of Equations 469
12.1.10. Numerical Integration of Differential Equations 471

12.2. Second-Order Linear Differential Equations 472
12.2.1. Formulas for the General Solution. Some Transformations 472
12.2.2. Representation of Solutions as a Series in the Independent Variable 475
12.2.3. Asymptotic Solutions 477
12.2.4. Boundary Value Problems 480
12.2.5. Eigenvalue Problems 482
12.2.6. Theorems on Estimates and Zeros of Solutions 487
12.3. Second-Order Nonlinear Differential Equations 488
12.3.1. Form of the General Solution. Cauchy Problem 488
12.3.2. Equations Admitting Reduction of Order 489
12.3.3. Methods of Regular Series Expansions with Respect to the Independent Variable 492
12.3.4. Movable Singularities of Solutions of Ordinary Differential Equations. Painlevé Transcendents 494
12.3.5. Perturbation Methods of Mechanics and Physics 499
12.3.6. Galerkin Method and Its Modifications (Projection Methods) 508
12.3.7. Iteration and Numerical Methods 511
12.4. Linear Equations of Arbitrary Order 514
12.4.1. Linear Equations with Constant Coefficients 514
12.4.2. Linear Equations with Variable Coefficients 518
12.4.3. Asymptotic Solutions of Linear Equations 522
12.4.4. Collocation Method and Its Convergence 523
12.5. Nonlinear Equations of Arbitrary Order 524
12.5.1. Structure of the General Solution. Cauchy Problem 524
12.5.2. Equations Admitting Reduction of Order 525
12.6. Linear Systems of Ordinary Differential Equations 528
12.6.1. Systems of Linear Constant-Coefficient Equations 528
12.6.2. Systems of Linear Variable-Coefficient Equations 539
12.7. Nonlinear Systems of Ordinary Differential Equations 542
12.7.1. Solutions and First Integrals. Uniqueness and Existence Theorems 542
12.7.2. Integrable Combinations. Autonomous Systems of Equations 545
12.7.3. Elements of Stability Theory 546
References for Chapter 12 550

13. First-Order Partial Differential Equations 553

13.1. Linear and Quasilinear Equations 553
13.1.1. Characteristic System. General Solution 553
13.1.2. Cauchy Problem. Existence and Uniqueness Theorem 556
13.1.3. Qualitative Features and Discontinuous Solutions of Quasilinear Equations . . 558
13.1.4. Quasilinear Equations of General Form. Generalized Solution, Jump Condition, and Stability Condition 567
13.2. Nonlinear Equations 570
13.2.1. Solution Methods 570
13.2.2. Cauchy Problem. Existence and Uniqueness Theorem 576
13.2.3. Generalized Viscosity Solutions and Their Applications 579
References for Chapter 13 584

14. Linear Partial Differential Equations **585**

14.1. Classification of Second-Order Partial Differential Equations 585
14.1.1. Equations with Two Independent Variables 585
14.1.2. Equations with Many Independent Variables 589

14.2. Basic Problems of Mathematical Physics 590
14.2.1. Initial and Boundary Conditions. Cauchy Problem. Boundary Value Problems 590
14.2.2. First, Second, Third, and Mixed Boundary Value Problems 593

14.3. Properties and Exact Solutions of Linear Equations 594
14.3.1. Homogeneous Linear Equations and Their Particular Solutions 594
14.3.2. Nonhomogeneous Linear Equations and Their Particular Solutions 598
14.3.3. General Solutions of Some Hyperbolic Equations 600

14.4. Method of Separation of Variables (Fourier Method) 602
14.4.1. Description of the Method of Separation of Variables. General Stage of Solution 602
14.4.2. Problems for Parabolic Equations: Final Stage of Solution 605
14.4.3. Problems for Hyperbolic Equations: Final Stage of Solution 607
14.4.4. Solution of Boundary Value Problems for Elliptic Equations 609

14.5. Integral Transforms Method 611
14.5.1. Laplace Transform and Its Application in Mathematical Physics 611
14.5.2. Fourier Transform and Its Application in Mathematical Physics 614

14.6. Representation of the Solution of the Cauchy Problem via the Fundamental Solution . . 615
14.6.1. Cauchy Problem for Parabolic Equations 615
14.6.2. Cauchy Problem for Hyperbolic Equations 617

14.7. Boundary Value Problems for Parabolic Equations with One Space Variable. Green's Function 618
14.7.1. Representation of Solutions via the Green's Function 618
14.7.2. Problems for Equation $s(x)\frac{\partial w}{\partial t} = \frac{\partial}{\partial x}\big[p(x)\frac{\partial w}{\partial x}\big] - q(x)w + \Phi(x,t)$ 620

14.8. Boundary Value Problems for Hyperbolic Equations with One Space Variable. Green's Function. Goursat Problem 623
14.8.1. Representation of Solutions via the Green's Function 623
14.8.2. Problems for Equation $s(x)\frac{\partial^2 w}{\partial t^2} = \frac{\partial}{\partial x}\big[p(x)\frac{\partial w}{\partial x}\big] - q(x)w + \Phi(x,t)$ 624
14.8.3. Problems for Equation $\frac{\partial^2 w}{\partial t^2} + a(t)\frac{\partial w}{\partial t} = b(t)\big\{\frac{\partial}{\partial x}\big[p(x)\frac{\partial w}{\partial x}\big] - q(x)w\big\} + \Phi(x,t)$ 626
14.8.4. Generalized Cauchy Problem with Initial Conditions Set Along a Curve 627
14.8.5. Goursat Problem (a Problem with Initial Data of Characteristics) 629

14.9. Boundary Value Problems for Elliptic Equations with Two Space Variables 631
14.9.1. Problems and the Green's Functions for Equation $a(x)\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + b(x)\frac{\partial w}{\partial x} + c(x)w = -\Phi(x,y)$ 631
14.9.2. Representation of Solutions to Boundary Value Problems via the Green's Functions 633

14.10. Boundary Value Problems with Many Space Variables. Representation of Solutions via the Green's Function 634
14.10.1. Problems for Parabolic Equations 634
14.10.2. Problems for Hyperbolic Equations 636
14.10.3. Problems for Elliptic Equations 637
14.10.4. Comparison of the Solution Structures for Boundary Value Problems for Equations of Various Types 638

14.11. Construction of the Green's Functions. General Formulas and Relations 639
14.11.1. Green's Functions of Boundary Value Problems for Equations of Various Types in Bounded Domains 639
14.11.2. Green's Functions Admitting Incomplete Separation of Variables 640
14.11.3. Construction of Green's Functions via Fundamental Solutions 642
14.12. Duhamel's Principles in Nonstationary Problems 646
14.12.1. Problems for Homogeneous Linear Equations 646
14.12.2. Problems for Nonhomogeneous Linear Equations 648
14.13. Transformations Simplifying Initial and Boundary Conditions 649
14.13.1. Transformations That Lead to Homogeneous Boundary Conditions 649
14.13.2. Transformations That Lead to Homogeneous Initial and Boundary Conditions 650
References for Chapter 14 650

15. Nonlinear Partial Differential Equations 653

15.1. Classification of Second-Order Nonlinear Equations 653
15.1.1. Classification of Semilinear Equations in Two Independent Variables 653
15.1.2. Classification of Nonlinear Equations in Two Independent Variables 653
15.2. Transformations of Equations of Mathematical Physics 655
15.2.1. Point Transformations: Overview and Examples 655
15.2.2. Hodograph Transformations (Special Point Transformations) 657
15.2.3. Contact Transformations. Legendre and Euler Transformations 660
15.2.4. Bäcklund Transformations. Differential Substitutions 663
15.2.5. Differential Substitutions 666
15.3. Traveling-Wave Solutions, Self-Similar Solutions, and Some Other Simple Solutions. Similarity Method 667
15.3.1. Preliminary Remarks 667
15.3.2. Traveling-Wave Solutions. Invariance of Equations Under Translations 667
15.3.3. Self-Similar Solutions. Invariance of Equations Under Scaling Transformations 669
15.3.4. Equations Invariant Under Combinations of Translation and Scaling Transformations, and Their Solutions 674
15.3.5. Generalized Self-Similar Solutions 677
15.4. Exact Solutions with Simple Separation of Variables 678
15.4.1. Multiplicative and Additive Separable Solutions 678
15.4.2. Simple Separation of Variables in Nonlinear Partial Differential Equations . . . 678
15.4.3. Complex Separation of Variables in Nonlinear Partial Differential Equations . 679
15.5. Method of Generalized Separation of Variables 681
15.5.1. Structure of Generalized Separable Solutions 681
15.5.2. Simplified Scheme for Constructing Solutions Based on Presetting One System of Coordinate Functions 683
15.5.3. Solution of Functional Differential Equations by Differentiation 684
15.5.4. Solution of Functional-Differential Equations by Splitting 688
15.5.5. Titov–Galaktionov Method 693
15.6. Method of Functional Separation of Variables 697
15.6.1. Structure of Functional Separable Solutions. Solution by Reduction to Equations with Quadratic Nonlinearities 697
15.6.2. Special Functional Separable Solutions. Generalized Traveling-Wave Solutions 697

15.6.3. Differentiation Method 700
15.6.4. Splitting Method. Solutions of Some Nonlinear Functional Equations and Their Applications 704
15.7. Direct Method of Symmetry Reductions of Nonlinear Equations 708
15.7.1. Clarkson–Kruskal Direct Method 708
15.7.2. Some Modifications and Generalizations 712
15.8. Classical Method of Studying Symmetries of Differential Equations 716
15.8.1. One-Parameter Transformations and Their Local Properties 716
15.8.2. Symmetries of Nonlinear Second-Order Equations. Invariance Condition 719
15.8.3. Using Symmetries of Equations for Finding Exact Solutions. Invariant Solutions 724
15.8.4. Some Generalizations. Higher-Order Equations 730
15.9. Nonclassical Method of Symmetry Reductions 732
15.9.1. Description of the Method. Invariant Surface Condition 732
15.9.2. Examples: The Newell–Whitehead Equation and a Nonlinear Wave Equation 733
15.10. Differential Constraints Method 737
15.10.1. Description of the Method 737
15.10.2. First-Order Differential Constraints 739
15.10.3. Second- and Higher-Order Differential Constraints 744
15.10.4. Connection Between the Differential Constraints Method and Other Methods 746
15.11. Painlevé Test for Nonlinear Equations of Mathematical Physics 748
15.11.1. Solutions of Partial Differential Equations with a Movable Pole. Method Description 748
15.11.2. Examples of Performing the Painlevé Test and Truncated Expansions for Studying Nonlinear Equations 750
15.11.3. Construction of Solutions of Nonlinear Equations That Fail the Painlevé Test, Using Truncated Expansions 753
15.12. Methods of the Inverse Scattering Problem (Soliton Theory) 755
15.12.1. Method Based on Using Lax Pairs 755
15.12.2. Method Based on a Compatibility Condition for Systems of Linear Equations 757
15.12.3. Solution of the Cauchy Problem by the Inverse Scattering Problem Method 760
15.13. Conservation Laws and Integrals of Motion 766
15.13.1. Basic Definitions and Examples 766
15.13.2. Equations Admitting Variational Formulation. Noetherian Symmetries . . . 767
15.14. Nonlinear Systems of Partial Differential Equations 770
15.14.1. Overdetermined Systems of Two Equations 770
15.14.2. Pfaffian Equations and Their Solutions. Connection with Overdetermined Systems 772
15.14.3. Systems of First-Order Equations Describing Convective Mass Transfer with Volume Reaction 775
15.14.4. First-Order Hyperbolic Systems of Quasilinear Equations. Systems of Conservation Laws of Gas Dynamic Type 780
15.14.5. Systems of Second-Order Equations of Reaction-Diffusion Type 796
References for Chapter 15 798

16. Integral Equations **801**

16.1. Linear Integral Equations of the First Kind with Variable Integration Limit 801
16.1.1. Volterra Equations of the First Kind 801
16.1.2. Equations with Degenerate Kernel: $K(x,t) = g_1(x)h_1(t) + \cdots + g_n(x)h_n(t)$ . . 802
16.1.3. Equations with Difference Kernel: $K(x,t) = K(x-t)$ 804
16.1.4. Reduction of Volterra Equations of the First Kind to Volterra Equations of the Second Kind 807
16.1.5. Method of Quadratures 808

16.2. Linear Integral Equations of the Second Kind with Variable Integration Limit 810
16.2.1. Volterra Equations of the Second Kind 810
16.2.2. Equations with Degenerate Kernel: $K(x,t) = g_1(x)h_1(t) + \cdots + g_n(x)h_n(t)$ . . 811
16.2.3. Equations with Difference Kernel: $K(x,t) = K(x-t)$ 813
16.2.4. Construction of Solutions of Integral Equations with Special Right-Hand Side 815
16.2.5. Method of Model Solutions 818
16.2.6. Successive Approximation Method 822
16.2.7. Method of Quadratures 823

16.3. Linear Integral Equations of the First Kind with Constant Limits of Integration 824
16.3.1. Fredholm Integral Equations of the First Kind 824
16.3.2. Method of Integral Transforms 825
16.3.3. Regularization Methods 827

16.4. Linear Integral Equations of the Second Kind with Constant Limits of Integration 829
16.4.1. Fredholm Integral Equations of the Second Kind. Resolvent 829
16.4.2. Fredholm Equations of the Second Kind with Degenerate Kernel 830
16.4.3. Solution as a Power Series in the Parameter. Method of Successive Approximations 832
16.4.4. Fredholm Theorems and the Fredholm Alternative 834
16.4.5. Fredholm Integral Equations of the Second Kind with Symmetric Kernel 835
16.4.6. Methods of Integral Transforms 841
16.4.7. Method of Approximating a Kernel by a Degenerate One 844
16.4.8. Collocation Method 847
16.4.9. Method of Least Squares 849
16.4.10. Bubnov–Galerkin Method 850
16.4.11. Quadrature Method 852
16.4.12. Systems of Fredholm Integral Equations of the Second Kind 854

16.5. Nonlinear Integral Equations 856
16.5.1. Nonlinear Volterra and Urysohn Integral Equations 856
16.5.2. Nonlinear Volterra Integral Equations 856
16.5.3. Equations with Constant Integration Limits 863

References for Chapter 16 871

17. Difference Equations and Other Functional Equations **873**

17.1. Difference Equations of Integer Argument 873
17.1.1. First-Order Linear Difference Equations of Integer Argument 873
17.1.2. First-Order Nonlinear Difference Equations of Integer Argument 874
17.1.3. Second-Order Linear Difference Equations with Constant Coefficients 877
17.1.4. Second-Order Linear Difference Equations with Variable Coefficients 879
17.1.5. Linear Difference Equations of Arbitrary Order with Constant Coefficients . . 881
17.1.6. Linear Difference Equations of Arbitrary Order with Variable Coefficients . . . 882
17.1.7. Nonlinear Difference Equations of Arbitrary Order 884

17.2. Linear Difference Equations with a Single Continuous Variable 885
17.2.1. First-Order Linear Difference Equations 885
17.2.2. Second-Order Linear Difference Equations with Integer Differences 894
17.2.3. Linear mth-Order Difference Equations with Integer Differences 898
17.2.4. Linear mth-Order Difference Equations with Arbitrary Differences 904

17.3. Linear Functional Equations 907
17.3.1. Iterations of Functions and Their Properties 907
17.3.2. Linear Homogeneous Functional Equations 910
17.3.3. Linear Nonhomogeneous Functional Equations 912
17.3.4. Linear Functional Equations Reducible to Linear Difference Equations with Constant Coefficients 916

17.4. Nonlinear Difference and Functional Equations with a Single Variable 918
17.4.1. Nonlinear Difference Equations with a Single Variable 918
17.4.2. Reciprocal (Cyclic) Functional Equations 919
17.4.3. Nonlinear Functional Equations Reducible to Difference Equations 921
17.4.4. Power Series Solution of Nonlinear Functional Equations 922

17.5. Functional Equations with Several Variables 922
17.5.1. Method of Differentiation in a Parameter 922
17.5.2. Method of Differentiation in Independent Variables 925
17.5.3. Method of Substituting Particular Values of Independent Arguments 926
17.5.4. Method of Argument Elimination by Test Functions 928
17.5.5. Bilinear Functional Equations and Nonlinear Functional Equations Reducible to Bilinear Equations 930

References for Chapter 17 935

18. Special Functions and Their Properties 937

18.1. Some Coefficients, Symbols, and Numbers 937
18.1.1. Binomial Coefficients 937
18.1.2. Pochhammer Symbol 938
18.1.3. Bernoulli Numbers 938
18.1.4. Euler Numbers 939

18.2. Error Functions. Exponential and Logarithmic Integrals 939
18.2.1. Error Function and Complementary Error Function 939
18.2.2. Exponential Integral 940
18.2.3. Logarithmic Integral 941

18.3. Sine Integral and Cosine Integral. Fresnel Integrals 941
18.3.1. Sine Integral 941
18.3.2. Cosine Integral 942
18.3.3. Fresnel Integrals 942

18.4. Gamma Function, Psi Function, and Beta Function 943
18.4.1. Gamma Function 943
18.4.2. Psi Function (Digamma Function) 944
18.4.3. Beta Function 945

18.5. Incomplete Gamma and Beta Functions 946
18.5.1. Incomplete Gamma Function 946
18.5.2. Incomplete Beta Function 947

18.6. Bessel Functions (Cylindrical Functions) 947
18.6.1. Definitions and Basic Formulas 947
18.6.2. Integral Representations and Asymptotic Expansions 949
18.6.3. Zeros and Orthogonality Properties of Bessel Functions 951
18.6.4. Hankel Functions (Bessel Functions of the Third Kind) 952
18.7. Modified Bessel Functions 953
18.7.1. Definitions. Basic Formulas 953
18.7.2. Integral Representations and Asymptotic Expansions 954
18.8. Airy Functions 955
18.8.1. Definition and Basic Formulas 955
18.8.2. Power Series and Asymptotic Expansions 956
18.9. Degenerate Hypergeometric Functions (Kummer Functions) 956
18.9.1. Definitions and Basic Formulas 956
18.9.2. Integral Representations and Asymptotic Expansions 959
18.9.3. Whittaker Functions 960
18.10. Hypergeometric Functions 960
18.10.1. Various Representations of the Hypergeometric Function 960
18.10.2. Basic Properties 960
18.11. Legendre Polynomials, Legendre Functions, and Associated Legendre Functions . . . 962
18.11.1. Legendre Polynomials and Legendre Functions 962
18.11.2. Associated Legendre Functions with Integer Indices and Real Argument . . 964
18.11.3. Associated Legendre Functions. General Case 965
18.12. Parabolic Cylinder Functions 967
18.12.1. Definitions. Basic Formulas 967
18.12.2. Integral Representations, Asymptotic Expansions, and Linear Relations . . . 968
18.13. Elliptic Integrals 969
18.13.1. Complete Elliptic Integrals 969
18.13.2. Incomplete Elliptic Integrals (Elliptic Integrals) 970
18.14. Elliptic Functions 972
18.14.1. Jacobi Elliptic Functions 972
18.14.2. Weierstrass Elliptic Function 976
18.15. Jacobi Theta Functions 978
18.15.1. Series Representation of the Jacobi Theta Functions. Simplest Properties . . 978
18.15.2. Various Relations and Formulas. Connection with Jacobi Elliptic Functions 978
18.16. Mathieu Functions and Modified Mathieu Functions 980
18.16.1. Mathieu Functions 980
18.16.2. Modified Mathieu Functions 982
18.17. Orthogonal Polynomials 982
18.17.1. Laguerre Polynomials and Generalized Laguerre Polynomials 982
18.17.2. Chebyshev Polynomials and Functions 983
18.17.3. Hermite Polynomials 985
18.17.4. Jacobi Polynomials and Gegenbauer Polynomials 986
18.18. Nonorthogonal Polynomials 988
18.18.1. Bernoulli Polynomials 988
18.18.2. Euler Polynomials 989
References for Chapter 18 990

19. Calculus of Variations and Optimization **991**

19.1. Calculus of Variations and Optimal Control 991
19.1.1. Some Definitions and Formulas 991
19.1.2. Simplest Problem of Calculus of Variations 993
19.1.3. Isoperimetric Problem 1002
19.1.4. Problems with Higher Derivatives 1006
19.1.5. Lagrange Problem 1008
19.1.6. Pontryagin Maximum Principle 1010

19.2. Mathematical Programming 1012
19.2.1. Linear Programming 1012
19.2.2. Nonlinear Programming 1027

References for Chapter 19 1028

20. Probability Theory **1031**

20.1. Simplest Probabilistic Models 1031
20.1.1. Probabilities of Random Events 1031
20.1.2. Conditional Probability and Simplest Formulas 1035
20.1.3. Sequences of Trials 1037

20.2. Random Variables and Their Characteristics 1039
20.2.1. One-Dimensional Random Variables 1039
20.2.2. Characteristics of One-Dimensional Random Variables 1042
20.2.3. Main Discrete Distributions 1047
20.2.4. Continuous Distributions 1051
20.2.5. Multivariate Random Variables 1057

20.3. Limit Theorems 1068
20.3.1. Convergence of Random Variables 1068
20.3.2. Limit Theorems 1069

20.4. Stochastic Processes 1071
20.4.1. Theory of Stochastic Processes 1071
20.4.2. Models of Stochastic Processes 1074

References for Chapter 20 1079

21. Mathematical Statistics **1081**

21.1. Introduction to Mathematical Statistics 1081
21.1.1. Basic Notions and Problems of Mathematical Statistics 1081
21.1.2. Simplest Statistical Transformations 1082
21.1.3. Numerical Characteristics of Statistical Distribution 1087

21.2. Statistical Estimation 1088
21.2.1. Estimators and Their Properties 1088
21.2.2. Estimation Methods for Unknown Parameters 1091
21.2.3. Interval Estimators (Confidence Intervals) 1093

21.3. Statistical Hypothesis Testing 1094
21.3.1. Statistical Hypothesis. Test 1094
21.3.2. Goodness-of-Fit Tests 1098
21.3.3. Problems Related to Normal Samples 1101

References for Chapter 21 1109

Part II. Mathematical Tables **1111**

T1. Finite Sums and Infinite Series **1113**

T1.1. Finite Sums 1113
T1.1.1. Numerical Sum 1113
T1.1.2. Functional Sums 1116

T1.2. Infinite Series 1118
T1.2.1. Numerical Series 1118
T1.2.2. Functional Series 1120

References for Chapter T1 1127

T2. Integrals **1129**

T2.1. Indefinite Integrals 1129
T2.1.1. Integrals Involving Rational Functions 1129
T2.1.2. Integrals Involving Irrational Functions 1134
T2.1.3. Integrals Involving Exponential Functions 1137
T2.1.4. Integrals Involving Hyperbolic Functions 1137
T2.1.5. Integrals Involving Logarithmic Functions 1140
T2.1.6. Integrals Involving Trigonometric Functions 1142
T2.1.7. Integrals Involving Inverse Trigonometric Functions 1147

T2.2. Tables of Definite Integrals 1147
T2.2.1. Integrals Involving Power-Law Functions 1147
T2.2.2. Integrals Involving Exponential Functions 1150
T2.2.3. Integrals Involving Hyperbolic Functions 1152
T2.2.4. Integrals Involving Logarithmic Functions 1152
T2.2.5. Integrals Involving Trigonometric Functions 1153

References for Chapter T2 1155

T3. Integral Transforms **1157**

T3.1. Tables of Laplace Transforms 1157
T3.1.1. General Formulas 1157
T3.1.2. Expressions with Power-Law Functions 1159
T3.1.3. Expressions with Exponential Functions 1159
T3.1.4. Expressions with Hyperbolic Functions 1160
T3.1.5. Expressions with Logarithmic Functions 1161
T3.1.6. Expressions with Trigonometric Functions 1161
T3.1.7. Expressions with Special Functions 1163

T3.2. Tables of Inverse Laplace Transforms 1164
T3.2.1. General Formulas 1164
T3.2.2. Expressions with Rational Functions 1166
T3.2.3. Expressions with Square Roots 1170
T3.2.4. Expressions with Arbitrary Powers 1172
T3.2.5. Expressions with Exponential Functions 1172
T3.2.6. Expressions with Hyperbolic Functions 1174
T3.2.7. Expressions with Logarithmic Functions 1174
T3.2.8. Expressions with Trigonometric Functions 1175
T3.2.9. Expressions with Special Functions 1176

T3.3. Tables of Fourier Cosine Transforms 1177
T3.3.1. General Formulas 1177
T3.3.2. Expressions with Power-Law Functions 1177
T3.3.3. Expressions with Exponential Functions 1178
T3.3.4. Expressions with Hyperbolic Functions 1179
T3.3.5. Expressions with Logarithmic Functions 1179
T3.3.6. Expressions with Trigonometric Functions 1180
T3.3.7. Expressions with Special Functions 1181
T3.4. Tables of Fourier Sine Transforms 1182
T3.4.1. General Formulas 1182
T3.4.2. Expressions with Power-Law Functions 1182
T3.4.3. Expressions with Exponential Functions 1183
T3.4.4. Expressions with Hyperbolic Functions 1184
T3.4.5. Expressions with Logarithmic Functions 1184
T3.4.6. Expressions with Trigonometric Functions 1185
T3.4.7. Expressions with Special Functions 1186
T3.5. Tables of Mellin Transforms 1187
T3.5.1. General Formulas 1187
T3.5.2. Expressions with Power-Law Functions 1188
T3.5.3. Expressions with Exponential Functions 1188
T3.5.4. Expressions with Logarithmic Functions 1189
T3.5.5. Expressions with Trigonometric Functions 1189
T3.5.6. Expressions with Special Functions 1190
T3.6. Tables of Inverse Mellin Transforms 1190
T3.6.1. Expressions with Power-Law Functions 1190
T3.6.2. Expressions with Exponential and Logarithmic Functions 1191
T3.6.3. Expressions with Trigonometric Functions 1192
T3.6.4. Expressions with Special Functions 1193
References for Chapter T3 1194
T4. Orthogonal Curvilinear Systems of Coordinate 1195
T4.1. Arbitrary Curvilinear Coordinate Systems 1195
T4.1.1. General Nonorthogonal Curvilinear Coordinates 1195
T4.1.2. General Orthogonal Curvilinear Coordinates 1196
T4.2. Special Curvilinear Coordinate Systems 1198
T4.2.1. Cylindrical Coordinates 1198
T4.2.2. Spherical Coordinates 1199
T4.2.3. Coordinates of a Prolate Ellipsoid of Revolution 1200
T4.2.4. Coordinates of an Oblate Ellipsoid of Revolution 1201
T4.2.5. Coordinates of an Elliptic Cylinder 1202
T4.2.6. Conical Coordinates 1202
T4.2.7. Parabolic Cylinder Coordinates 1203
T4.2.8. Parabolic Coordinates 1203
T4.2.9. Bicylindrical Coordinates 1204
T4.2.10. Bipolar Coordinates (in Space) 1204
T4.2.11. Toroidal Coordinates 1205
References for Chapter T4 1205

T5. Ordinary Differential Equations 1207
T5.1. First-Order Equations 1207
T5.2. Second-Order Linear Equations 1212
T5.2.1. Equations Involving Power Functions 1213
T5.2.2. Equations Involving Exponential and Other Functions 1220
T5.2.3. Equations Involving Arbitrary Functions 1222
T5.3. Second-Order Nonlinear Equations 1223
T5.3.1. Equations of the Form $y''_{xx} = f(x, y)$ 1223
T5.3.2. Equations of the Form $f(x, y)y''_{xx} = g(x, y, y'_x)$ 1225
References for Chapter T5 1228

T6. Systems of Ordinary Differential Equations 1229
T6.1. Linear Systems of Two Equations 1229
T6.1.1. Systems of First-Order Equations 1229
T6.1.2. Systems of Second-Order Equations 1232
T6.2. Linear Systems of Three and More Equations 1237
T6.3. Nonlinear Systems of Two Equations 1239
T6.3.1. Systems of First-Order Equations 1239
T6.3.2. Systems of Second-Order Equations 1240
T6.4. Nonlinear Systems of Three or More Equations 1244
References for Chapter T6 1246

T7. First-Order Partial Differential Equations 1247
T7.1. Linear Equations 1247
T7.1.1. Equations of the Form $f(x, y)\frac{\partial w}{\partial x} + g(x, y)\frac{\partial w}{\partial y} = 0$ 1247
T7.1.2. Equations of the Form $f(x, y)\frac{\partial w}{\partial x} + g(x, y)\frac{\partial w}{\partial y} = h(x, y)$ 1248
T7.1.3. Equations of the Form $f(x, y)\frac{\partial w}{\partial x} + g(x, y)\frac{\partial w}{\partial y} = h(x, y)w + r(x, y)$ 1250
T7.2. Quasilinear Equations 1252
T7.2.1. Equations of the Form $f(x, y)\frac{\partial w}{\partial x} + g(x, y)\frac{\partial w}{\partial y} = h(x, y, w)$ 1252
T7.2.2. Equations of the Form $\frac{\partial w}{\partial x} + f(x, y, w)\frac{\partial w}{\partial y} = 0$ 1254
T7.2.3. Equations of the Form $\frac{\partial w}{\partial x} + f(x, y, w)\frac{\partial w}{\partial y} = g(x, y, w)$ 1256
T7.3. Nonlinear Equations 1258
T7.3.1. Equations Quadratic in One Derivative 1258
T7.3.2. Equations Quadratic in Two Derivatives 1259
T7.3.3. Equations with Arbitrary Nonlinearities in Derivatives 1261
References for Chapter T7 1265

T8. Linear Equations and Problems of Mathematical Physics 1267
T8.1. Parabolic Equations 1267
T8.1.1. Heat Equation $\frac{\partial w}{\partial t} = a\frac{\partial^2 w}{\partial x^2}$ 1267
T8.1.2. Nonhomogeneous Heat Equation $\frac{\partial w}{\partial t} = a\frac{\partial^2 w}{\partial x^2} + \Phi(x, t)$ 1268
T8.1.3. Equation of the Form $\frac{\partial w}{\partial t} = a\frac{\partial^2 w}{\partial x^2} + b\frac{\partial w}{\partial x} + cw + \Phi(x, t)$ 1270
T8.1.4. Heat Equation with Axial Symmetry $\frac{\partial w}{\partial t} = a\left(\frac{\partial^2 w}{\partial r^2} + \frac{1}{r}\frac{\partial w}{\partial r}\right)$ 1270
T8.1.5. Equation of the Form $\frac{\partial w}{\partial t} = a\left(\frac{\partial^2 w}{\partial r^2} + \frac{1}{r}\frac{\partial w}{\partial r}\right) + \Phi(r, t)$ 1271
T8.1.6. Heat Equation with Central Symmetry $\frac{\partial w}{\partial t} = a\left(\frac{\partial^2 w}{\partial r^2} + \frac{2}{r}\frac{\partial w}{\partial r}\right)$ 1272
T8.1.7. Equation of the Form $\frac{\partial w}{\partial t} = a\left(\frac{\partial^2 w}{\partial r^2} + \frac{2}{r}\frac{\partial w}{\partial r}\right) + \Phi(r, t)$ 1273
T8.1.8. Equation of the Form $\frac{\partial w}{\partial t} = \frac{\partial^2 w}{\partial x^2} + \frac{1-2\beta}{x}\frac{\partial w}{\partial x}$ 1274

T8.1.9. Equations of the Diffusion (Thermal) Boundary Layer 1276
T8.1.10. Schrödinger Equation $i\hbar\frac{\partial w}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2 w}{\partial x^2} + U(x)w$ 1276
T8.2. Hyperbolic Equations 1278
T8.2.1. Wave Equation $\frac{\partial^2 w}{\partial t^2} = a^2\frac{\partial^2 w}{\partial x^2}$ 1278
T8.2.2. Equation of the Form $\frac{\partial^2 w}{\partial t^2} = a^2\frac{\partial^2 w}{\partial x^2} + \Phi(x,t)$ 1279
T8.2.3. Klein–Gordon Equation $\frac{\partial^2 w}{\partial t^2} = a^2\frac{\partial^2 w}{\partial x^2} - bw$ 1280
T8.2.4. Equation of the Form $\frac{\partial^2 w}{\partial t^2} = a^2\frac{\partial^2 w}{\partial x^2} - bw + \Phi(x,t)$ 1281
T8.2.5. Equation of the Form $\frac{\partial^2 w}{\partial t^2} = a^2\left(\frac{\partial^2 w}{\partial r^2} + \frac{1}{r}\frac{\partial w}{\partial r}\right) + \Phi(r,t)$ 1282
T8.2.6. Equation of the Form $\frac{\partial^2 w}{\partial t^2} = a^2\left(\frac{\partial^2 w}{\partial r^2} + \frac{2}{r}\frac{\partial w}{\partial r}\right) + \Phi(r,t)$ 1283
T8.2.7. Equations of the Form $\frac{\partial^2 w}{\partial t^2} + k\frac{\partial w}{\partial t} = a^2\frac{\partial^2 w}{\partial x^2} + b\frac{\partial w}{\partial x} + cw + \Phi(x,t)$ 1284
T8.3. Elliptic Equations 1284
T8.3.1. Laplace Equation $\Delta w = 0$ 1284
T8.3.2. Poisson Equation $\Delta w + \Phi(x) = 0$ 1287
T8.3.3. Helmholtz Equation $\Delta w + \lambda w = -\Phi(x)$ 1289
T8.4. Fourth-Order Linear Equations 1294
T8.4.1. Equation of the Form $\frac{\partial^2 w}{\partial t^2} + a^2\frac{\partial^4 w}{\partial x^4} = 0$ 1294
T8.4.2. Equation of the Form $\frac{\partial^2 w}{\partial t^2} + a^2\frac{\partial^4 w}{\partial x^4} = \Phi(x,t)$ 1295
T8.4.3. Biharmonic Equation $\Delta\Delta w = 0$ 1297
T8.4.4. Nonhomogeneous Biharmonic Equation $\Delta\Delta w = \Phi(x,y)$ 1298
References for Chapter T8 1299

T9. Nonlinear Mathematical Physics Equations 1301

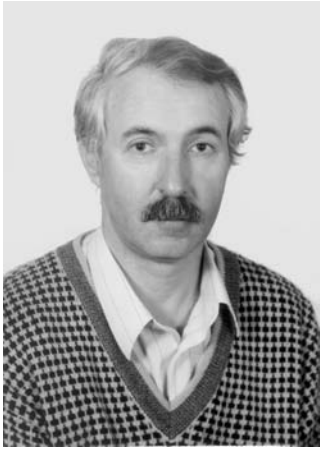
T9.1. Parabolic Equations 1301
T9.1.1. Nonlinear Heat Equations of the Form $\frac{\partial w}{\partial t} = \frac{\partial^2 w}{\partial x^2} + f(w)$ 1301
T9.1.2. Equations of the Form $\frac{\partial w}{\partial t} = \frac{\partial}{\partial x}\left[f(w)\frac{\partial w}{\partial x}\right] + g(w)$ 1303
T9.1.3. Burgers Equation and Nonlinear Heat Equation in Radial Symmetric Cases . . 1307
T9.1.4. Nonlinear Schrödinger Equations 1309
T9.2. Hyperbolic Equations 1312
T9.2.1. Nonlinear Wave Equations of the Form $\frac{\partial^2 w}{\partial t^2} = a\frac{\partial^2 w}{\partial x^2} + f(w)$ 1312
T9.2.2. Other Nonlinear Wave Equations 1316
T9.3. Elliptic Equations 1318
T9.3.1. Nonlinear Heat Equations of the Form $\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} = f(w)$ 1318
T9.3.2. Equations of the Form $\frac{\partial}{\partial x}\left[f(x)\frac{\partial w}{\partial x}\right] + \frac{\partial}{\partial y}\left[g(y)\frac{\partial w}{\partial y}\right] = f(w)$ 1321
T9.3.3. Equations of the Form $\frac{\partial}{\partial x}\left[f(w)\frac{\partial w}{\partial x}\right] + \frac{\partial}{\partial y}\left[g(w)\frac{\partial w}{\partial y}\right] = h(w)$ 1322
T9.4. Other Second-Order Equations 1324
T9.4.1. Equations of Transonic Gas Flow 1324
T9.4.2. Monge–Ampère Equations 1326
T9.5. Higher-Order Equations 1327
T9.5.1. Third-Order Equations 1327
T9.5.2. Fourth-Order Equations 1332
References for Chapter T9 1335

T10. Systems of Partial Differential Equations 1337

T10.1. Nonlinear Systems of Two First-Order Equations 1337
T10.2. Linear Systems of Two Second-Order Equations 1341

T10.3. Nonlinear Systems of Two Second-Order Equations 1343
T10.3.1. Systems of the Form $\frac{\partial u}{\partial t} = a\frac{\partial^2 u}{\partial x^2} + F(u,w)$, $\frac{\partial w}{\partial t} = b\frac{\partial^2 w}{\partial x^2} + G(u,w)$ 1343
T10.3.2. Systems of the Form $\frac{\partial u}{\partial t} = \frac{a}{x^n}\frac{\partial}{\partial x}\left(x^n\frac{\partial u}{\partial x}\right) + F(u,w)$, $\frac{\partial w}{\partial t} = \frac{b}{x^n}\frac{\partial}{\partial x}\left(x^n\frac{\partial w}{\partial x}\right) + G(u,w)$ 1357
T10.3.3. Systems of the Form $\Delta u = F(u,w)$, $\Delta w = G(u,w)$ 1364
T10.3.4. Systems of the Form $\frac{\partial^2 u}{\partial t^2} = \frac{a}{x^n}\frac{\partial}{\partial x}\left(x^n\frac{\partial u}{\partial x}\right) + F(u,w)$, $\frac{\partial^2 w}{\partial t^2} = \frac{b}{x^n}\frac{\partial}{\partial x}\left(x^n\frac{\partial w}{\partial x}\right) + G(u,w)$ 1368
T10.3.5. Other Systems 1373
T10.4. Systems of General Form 1374
T10.4.1. Linear Systems 1374
T10.4.2. Nonlinear Systems of Two Equations Involving the First Derivatives in t . . 1374
T10.4.3. Nonlinear Systems of Two Equations Involving the Second Derivatives in t 1378
T10.4.4. Nonlinear Systems of Many Equations Involving the First Derivatives in t . 1381
References for Chapter T10 1382
T11. Integral Equations 1385
T11.1. Linear Equations of the First Kind with Variable Limit of Integration 1385
T11.2. Linear Equations of the Second Kind with Variable Limit of Integration 1391
T11.3. Linear Equations of the First Kind with Constant Limits of Integration 1396
T11.4. Linear Equations of the Second Kind with Constant Limits of Integration 1401
References for Chapter T11 1406
T12. Functional Equations 1409
T12.1. Linear Functional Equations in One Independent Variable 1409
T12.1.1. Linear Difference and Functional Equations Involving Unknown Function with Two Different Arguments 1409
T12.1.2. Other Linear Functional Equations 1421
T12.2. Nonlinear Functional Equations in One Independent Variable 1428
T12.2.1. Functional Equations with Quadratic Nonlinearity 1428
T12.2.2. Functional Equations with Power Nonlinearity 1433
T12.2.3. Nonlinear Functional Equation of General Form 1434
T12.3. Functional Equations in Several Independent Variables 1438
T12.3.1. Linear Functional Equations 1438
T12.3.2. Nonlinear Functional Equations 1443
References for Chapter T12 1450
Supplement. Some Useful Electronic Mathematical Resources 1451
Index 1453

AUTHORS



Andrei D. Polyanin, D.Sc., Ph.D., is a well-known scientist of broad interests who is active in various areas of mathematics, mechanics, and chemical engineering sciences. He is one of the most prominent authors in the field of reference literature on mathematics and physics.

Professor Polyanin graduated with honors from the Department of Mechanics and Mathematics of Moscow State University in 1974. He received his Ph.D. in 1981 and his D.Sc. in 1986 from the Institute for Problems in Mechanics of the Russian (former USSR) Academy of Sciences. Since 1975, Professor Polyanin has been working at the Institute for Problems in Mechanics of the Russian Academy of Sciences; he is also Professor of Mathematics at Bauman Moscow State Technical University. He is a member of the Russian National Committee on Theoretical and Applied Mechanics and of the Mathematics and Mechanics Expert Council of the Higher Certification Committee of the Russian Federation.

Professor Polyanin has made important contributions to exact and approximate analytical methods in the theory of differential equations, mathematical physics, integral equations, engineering mathematics, theory of heat and mass transfer, and chemical hydrodynamics. He has obtained exact solutions for several thousand ordinary differential, partial differential, and integral equations.

Professor Polyanin is an author of more than 30 books in English, Russian, German, and Bulgarian as well as more than 120 research papers and three patents. He has written a number of fundamental handbooks, including A. D. Polyanin and V. F. Zaitsev, *Handbook of Exact Solutions for Ordinary Differential Equations*, CRC Press, 1995 and 2003; A. D. Polyanin and A. V. Manzhirov, *Handbook of Integral Equations*, CRC Press, 1998; A. D. Polyanin, *Handbook of Linear Partial Differential Equations for Engineers and Scientists*, Chapman & Hall/CRC Press, 2002; A. D. Polyanin, V. F. Zaitsev, and A. Moussiaux, *Handbook of First Order Partial Differential Equations*, Taylor & Francis, 2002; and A. D. Polyanin and V. F. Zaitsev, *Handbook of Nonlinear Partial Differential Equation*, Chapman & Hall/CRC Press, 2004.

Professor Polyanin is editor of the book series Differential and Integral Equations and Their Applications, Chapman & Hall/CRC Press, London/Boca Raton, and Physical and Mathematical Reference Literature, Fizmatlit, Moscow. He is also Editor-in-Chief of the international scientific-educational Website *EqWorld — The World of Mathematical Equations* (http://eqworld.ipmnet.ru), which is visited by over 1000 users a day worldwide. Professor Polyanin is a member of the Editorial Board of the journal *Theoretical Foundations of Chemical Engineering*.

In 1991, Professor Polyanin was awarded a Chaplygin Prize of the Russian Academy of Sciences for his research in mechanics. In 2001, he received an award from the Ministry of Education of the Russian Federation.

Address: Institute for Problems in Mechanics, Vernadsky Ave. 101 Bldg 1, 119526 Moscow, Russia
Home page: http://eqworld.ipmnet.ru/polyanin-ew.htm

Alexander V. Manzhirov, D.Sc., Ph.D., is a noted scientist in the fields of mechanics and applied mathematics, integral equations, and their applications.

After graduating with honors from the Department of Mechanics and Mathematics of Rostov State University in 1979, Professor Manzhirov attended postgraduate courses at Moscow Institute of Civil Engineering. He received his Ph.D. in 1983 from Moscow Institute of Electronic Engineering Industry and his D.Sc. in 1993 from the Institute for Problems in Mechanics of the Russian (former USSR) Academy of Sciences. Since 1983, Professor Manzhirov has been working at the Institute for Problems in Mechanics of the Russian Academy of Sciences, where he is currently head of the Laboratory for Modeling in Solid Mechanics.

Professor Manzhirov is also head of a branch of the Department of Applied Mathematics at Bauman Moscow State Technical University, professor of mathematics at Moscow State University of Engineering and Computer Science, vice-chairman of Mathematics and Mechanics Expert Council of the Higher Certification Committee of the Russian Federation, executive secretary of Solid Mechanics Scientific Council of the Russian Academy of Sciences, and an expert in mathematics, mechanics, and computer science of the Russian Foundation for Basic Research. He is a member of the Russian National Committee on Theoretical and Applied Mechanics and the European Mechanics Society (EUROMECH), and a member of the editorial board of the journal *Mechanics of Solids* and the international scientific-educational Website *EqWorld—The World of Mathematical Equations* (http://eqworld.ipmnet.ru).

Professor Manzhirov has made important contributions to new mathematical methods for solving problems in the fields of integral equations and their applications, mechanics of growing solids, contact mechanics, tribology, viscoelasticity, and creep theory. He is the author of ten books (including *Contact Problems in Mechanics of Growing Solids* [in Russian], Nauka, Moscow, 1991; *Handbook of Integral Equations*, CRC Press, Boca Raton, 1998; *Handbuch der Integralgleichungen: Exacte Lösungen*, Spektrum Akad. Verlag, Heidelberg, 1999; *Contact Problems in the Theory of Creep* [in Russian], National Academy of Sciences of Armenia, Erevan, 1999), more than 70 research papers, and two patents.

Professor Manzhirov is a winner of the First Competition of the Science Support Foundation 2001, Moscow.

Address: Institute for Problems in Mechanics, Vernadsky Ave. 101 Bldg 1, 119526 Moscow, Russia
Home page: http://eqworld.ipmnet.ru/en/board/manzhirov.htm

PREFACE

This book can be viewed as a reasonably comprehensive compendium of mathematical definitions, formulas, and theorems intended for researchers, university teachers, engineers, and students of various backgrounds in mathematics. The absence of proofs and a concise presentation has permitted combining a substantial amount of reference material in a single volume.

When selecting the material, the authors have given a pronounced preference to practical aspects, namely, to formulas, methods, equations, and solutions that are most frequently used in scientific and engineering applications. Hence some abstract concepts and their corollaries are not contained in this book.

• This book contains chapters on arithmetics, elementary geometry, analytic geometry, algebra, differential and integral calculus, differential geometry, elementary and special functions, functions of one complex variable, calculus of variations, probability theory, mathematical statistics, etc. Special attention is paid to formulas (exact, asymptotical, and approximate), functions, methods, equations, solutions, and transformations that are of frequent use in various areas of physics, mechanics, and engineering sciences.

• The main distinction of this reference book from other general (nonspecialized) mathematical reference books is a significantly wider and more detailed description of methods for solving equations and obtaining their exact solutions for various classes of mathematical equations (ordinary differential equations, partial differential equations, integral equations, difference equations, etc.) that underlie mathematical modeling of numerous phenomena and processes in science and technology. In addition to well-known methods, some new methods that have been developing intensively in recent years are described.

• For the convenience of a wider audience with different mathematical backgrounds, the authors tried to avoid special terminology whenever possible. Therefore, some of the methods and theorems are outlined in a schematic and somewhat simplified manner, which is sufficient for them to be used successfully in most cases. Many sections were written so that they could be read independently. The material within subsections is arranged in increasing order of complexity. This allows the reader to get to the heart of the matter quickly.

The material in the first part of the reference book can be roughly categorized into the following three groups according to meaning:

1. The main text containing a concise, coherent survey of the most important definitions, formulas, equations, methods, and theorems.

2. Numerous specific examples clarifying the essence of the topics and methods for solving problems and equations.

3. Discussion of additional issues of interest, given in the form of remarks in small print.

For the reader's convenience, several long mathematical tables — finite sums, series, indefinite and definite integrals, direct and inverse integral transforms (Laplace, Mellin, and Fourier transforms), and exact solutions of differential, integral, functional, and other mathematical equations — which contain a large amount of information, are presented in the second part of the book.

This handbook consists of chapters, sections, subsections, and paragraphs (the titles of the latter are not included in the table of contents). Figures and tables are numbered separately in each section, while formulas (equations) and examples are numbered separately in each subsection. When citing a formula, we use notation like (3.1.2.5), which means

formula 5 in Subsection 3.1.2. At the end of each chapter, we present a list of main and additional literature sources containing more detailed information about topics of interest to the reader.

Special font highlighting in the text, cross-references, an extensive table of contents, and an index help the reader to find the desired information.

We would like to express our deep gratitude to Alexei Zhurov for fruitful discussions and valuable remarks. We also appreciate the help of Vladimir Nazaikinskii and Grigorii Yosifian for translating several chapters of this book and are thankful to Kirill Kazakov and Mikhail Mikhin for their assistance in preparing the camera-ready copy of the book.

The authors hope that this book will be helpful for a wide range of scientists, university teachers, engineers, and students engaged in the fields of mathematics, physics, mechanics, control, chemistry, biology, engineering sciences, and social and economical sciences. Some sections and examples can be used in lectures and practical studies in basic and special mathematical courses.

Andrei D. Polyanin
Alexander V. Manzhirov

Main Notation

Special symbols

$=$ equal to

$\equiv$ identically equal to

$\neq$ not equal to

$\approx$ approximately equal to

$\sim$ of same order as (used in comparisons of infinitesimals or infinites)

$<$ less than; "a less than b" is written as $a < b$ (or, equivalently, $b > a$)

$\leq$ less than or equal to; a less than or equal to b is written as $a \leq b$

$\ll$ much less than; a much less than b is written as $a \ll b$

$>$ greater than; a greater than b is written as $a > b$ (or, equivalently, $b < a$)

$\geq$ greater than or equal to; a greater than or equal to b is written as $a \geq b$

$\gg$ much greater than; a much greater than b is written as $a \gg b$

$+$ plus sign; the sum of numbers a and b is denoted by $a + b$ and has the property $a + b = b + a$

$-$ minus sign; the difference of numbers a and b is denoted by $a - b$

$\cdot$ multiplication sign; the product of numbers a and b is denoted by either ab or $a \cdot b$ (sometimes $a \times b$) and has the property $ab = ba$; the inner product of vectors $\mathbf{a}$ and $\mathbf{b}$ is denoted by $\mathbf{a} \cdot \mathbf{b}$

$\times$ multiplication sign; the product of numbers a and b is sometimes denoted by $a \times b$; the cross-product of vectors $\mathbf{a}$ and $\mathbf{b}$ is denoted by $\mathbf{a} \times \mathbf{b}$

$:$ division sign; the ratio of numbers a and b is denoted by $a : b$ or a/b

$!$ factorial sign: $0! = 1! = 1,\ \ n! = 1 \cdot 2 \cdot 3 \ldots (n-1)n,\quad n = 2,\ 3,\ 4,\ \ldots$

$!!$ double factorial sign: $0!! = 1!! = 1,\ (2n)!! = 2 \cdot 4 \cdot 6 \ldots (2n),\ (2n+1)!! = 1 \cdot 3 \cdot 5 \ldots (2n+1)$, where $n = 1,\ 2,\ 3,\ \ldots$

$\%$ percent sign; 1% is one hundredth of the entire quantity

∞ infinity

$\rightarrow$ tends (infinitely approaches) to; $x \rightarrow a$ means that x tends to a

$\Longrightarrow$ implies; consequently

$\Longleftrightarrow$ is equivalent to (if and only if ...)

$\forall$ for all, for any

$\exists$ there exists

$\in$ belongs to; $a \in A$ means that a is an element of the set A

$\notin$ does not belong to; $a \notin A$ means that a is not an element of the set A

$\cup$ union (Boolean addition); $A \cup B$ stands for the union of sets A and B

$\cap$ intersection (Boolean multiplication); $A \cap B$ stands for the intersection (common part) of sets A and B

$\subset$ inclusion; $A \subset B$ means that the set A is part of the set B

$\subseteq$ nonstrict inclusion; $A \subseteq B$ means that the set A is part of the set B or coincides with B

$\varnothing$ empty set

$\sum$ sum, $\sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n$

$\prod$ product, $\prod_{k=1}^{n} a_k = a_1 \cdot a_2 \cdot \ldots \cdot a_n$

∂ symbol used to denote partial derivatives and differential operators; ∂_x is the operator of differentiation with respect to x

∇	vector differential operator "nabla"; ∇a is the gradient of a scalar a
$\int$	integral; $\int_a^b f(x)\,dx$ is the integral of a function $f(x)$ over the interval $[a,\ b]$
$\oint$	contour integral (denotes an integral over a closed contour)
$\perp$	perpendicular
$\parallel$	parallel

Roman alphabet

$\operatorname{Arg} z$	argument of a complex number $z = x + iy$; by definition, $\tan(\operatorname{Arg} z) = y/x$
$\arg z$	principal value of $\operatorname{Arg} z$; by definition, $\arg z = \operatorname{Arg} z$, where $-\pi < \operatorname{Arg} z \le \pi$
$\sqrt{a}$	square root of a number a, defined by the property $(\sqrt{a})^2 = a$
$\sqrt[n]{a}$	nth root of a number a ($n = 2,\ 3,\ \ldots,\ a \ge 0$), defined by the property $(\sqrt[n]{a})^n = a$
$\lvert a\rvert$	absolute value (modulus) of a real number a, $\lvert a\rvert = \begin{cases} a & \text{if } a \ge 0 \\ -a & \text{if } a < 0 \end{cases}$
$\mathbf{a}$	vector, $\mathbf{a} = \{a_1, a_2, a_3\}$, where a_1, a_2, a_3 are the vector components
$\lvert\mathbf{a}\rvert$	modulus of a vector $\mathbf{a}$, $\lvert\mathbf{a}\rvert = \sqrt{\mathbf{a}\cdot\mathbf{a}}$
$\mathbf{a}\cdot\mathbf{b}$	inner product of vectors $\mathbf{a}$ and $\mathbf{b}$, denoted also by $(\mathbf{a}\cdot\mathbf{b})$
$\mathbf{a}\times\mathbf{b}$	cross-product of vectors $\mathbf{a}$ and $\mathbf{b}$
$[\mathbf{abc}]$	triple product of vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$
(a, b)	interval (open interval) $a < x < b$
$(a, b]$	half-open interval $a < x \le b$
$[a, b)$	half-open interval $a \le x < b$
$[a, b]$	interval (closed interval) $a \le x \le b$
$\arccos x$	arccosine, the inverse function of cosine: $\cos(\arccos x) = x,\ \lvert x\rvert \le 1$
$\operatorname{arccot} x$	arccotangent, the inverse function of cotangent: $\operatorname{cot}(\operatorname{arccot} x) = x$
$\arcsin x$	arcsine, the inverse function of sine: $\sin(\arcsin x) = x,\ \lvert x\rvert \le 1$
$\arctan x$	arctangent, the inverse function of tangent: $\tan(\arctan x) = x$
$\operatorname{arccosh} x$	hyperbolic arccosine, the inverse function of hyperbolic cosine; also denoted by $\operatorname{arccosh} x = \cosh^{-1} x;\ \operatorname{arccosh} x = \ln\left(x + \sqrt{x^2 - 1}\right)\ (x \ge 1)$
$\operatorname{arccoth} x$	hyperbolic arccotangent, the inverse function of hyperbolic cotangent; also denoted by $\operatorname{arccoth} x = \coth^{-1} x;\ \operatorname{arccoth} x = \frac{1}{2}\ln\frac{x+1}{x-1}\ (\lvert x\rvert > 1)$
$\operatorname{arcsinh} x$	hyperbolic arcsine, the inverse function of hyperbolic sine; also denoted by $\operatorname{arcsinh} x = \sinh^{-1} x;\ \operatorname{arcsinh} x = \ln\left(x + \sqrt{x^2 + 1}\right)$
$\operatorname{arctanh} x$	hyperbolic arctangent, the inverse function of hyperbolic tangent; also denoted by $\operatorname{arctanh} x = \tanh^{-1} x;\ \operatorname{arctanh} x = \frac{1}{2}\ln\frac{1+x}{1-x}\ (\lvert x\rvert < 1)$
C_n^k	binomial coefficients, also denoted by $\binom{n}{k}$, $C_n^k = \frac{n!}{k!\,(n-k)!},\ k = 1,\ 2,\ \ldots,\ n$
$\mathcal{C}$	Euler constant, $\mathcal{C} = \lim\limits_{n\to\infty}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \ln n\right) = 0.5772156\ldots$
$\cos x$	cosine, even trigonometric function of period 2π
$\operatorname{cosec} x$	cosecant, odd trigonometric function of period 2π: $\operatorname{cosec} x = \frac{1}{\sin x}$
$\cosh x$	hyperbolic cosine, $\cosh x = \frac{1}{2}(e^x + e^{-x})$
$\cot x$	cotangent, odd trigonometric function of period π, $\cot x = \cos x/\sin x$
$\coth x$	hyperbolic cotangent, $\coth x = \cosh x/\sinh x$
$\det A$	determinant of a matrix $A = (a_{ij})$

$\operatorname{div}\mathbf{a}$	divergence of a vector $\mathbf{a}$
e	the number "e" (base of natural logarithms), $e = 2.718281\ldots$; definition: $e = \lim\limits_{n\to\infty}\left(1+\dfrac{1}{n}\right)^n$
$\operatorname{erf} x$	Gauss error function, $\operatorname{erf} x = \dfrac{2}{\sqrt{\pi}}\displaystyle\int_0^x \exp\left(-\xi^2\right)\,d\xi$
$\operatorname{erfc} x$	complementary error function, $\operatorname{erfc} x = \dfrac{2}{\sqrt{\pi}}\displaystyle\int_x^\infty \exp\left(-\xi^2\right)\,d\xi$
$\exp x$	exponential (exponential function), denoted also by $\exp x = e^x$
$\operatorname{grad} a$	gradient of a scalar a, denoted also by ∇a
$H_n(x)$	Hermite polynomial, $H_n(x) = (-1)^n e^{x^2}\dfrac{d^n}{dx^n}\left(e^{-x^2}\right)$
$I_\nu(x)$	modified Bessel function of the first kind, $I_\nu(x) = \displaystyle\sum_{n=0}^{\infty}\frac{(x/2)^{\nu+2n}}{n!\,\Gamma(\nu+n+1)}$
$\operatorname{Im} z$	imaginary part of a complex number; if $z = x + iy$, then $\operatorname{Im} z = y$
$\inf A$	infimum of a (numerical) set A; if $A = (a, b)$ or $A = [a, b)$, then $\inf A = a$
$J_\nu(x)$	Bessel function of the first kind, $J_\nu(x) = \displaystyle\sum_{n=0}^{\infty}\frac{(-1)^n(x/2)^{\nu+2n}}{n!\,\Gamma(\nu+n+1)}$
$K_\nu(x)$	modified Bessel function of the second kind, $K_\nu(x) = \dfrac{\pi}{2}\dfrac{I_{-\nu}(x) - I_\nu(x)}{\sin(\pi\nu)}$
$\lim\limits_{x\to a} f(x)$	limit of a function $f(x)$ as $x \to a$
$\ln x$	natural logarithm (logarithm to base e)
$\log_a x$	logarithm to base a
$\max\limits_{a\le x\le b} f(x)$	maximum of a function $f(x)$ on the interval $a \le x \le b$
$\min\limits_{a\le x\le b} f(x)$	minimum of a function $f(x)$ on the interval $a \le x \le b$
$n!$	factorial: $0! = 1! = 1,\ \ n! = 1\cdot 2\cdot 3\ldots(n-1)n,\quad n = 2,\ 3,\ 4,\ \ldots$
$P_n(x)$	Legendre polynomial, $P_n(x) = \dfrac{1}{n!\,2^n}\dfrac{d^n}{dx^n}(x^2-1)^n$
$\mathbb{R}$	set of real numbers, $\mathbb{R} = \{-\infty < x < \infty\}$
$\operatorname{Re} z$	real part of a complex number; if $z = x + iy$, then $\operatorname{Re} z = x$
$r,\ \varphi,\ z$	cylindrical coordinates, $r = \sqrt{x^2+y^2}$ and $x = r\cos\varphi$, $y = r\sin\varphi$
$r,\ \theta,\ \varphi$	spherical coordinates, $r = \sqrt{x^2+y^2+z^2}$ and $x = r\sin\theta\cos\varphi$, $y = \sin\theta\sin\varphi$, $z = r\cos\theta$
$\operatorname{rank} A$	rank of a matrix A
$\operatorname{curl}\mathbf{a}$	curl of a vector $\mathbf{a}$, also denoted by $\operatorname{rot}\mathbf{a}$
$\sec x$	secant, even trigonometric function of period 2π: $\sec x = \dfrac{1}{\cos x}$
$\operatorname{sign} x$	"sign" function: it is equal to 1 if $x > 0$, -1 if $x < 0$, and 0 if $x = 0$
$\sin x$	sine, odd trigonometric function of period 2π
$\sinh x$	hyperbolic sine, $\sinh x = \frac{1}{2}(e^x - e^{-x})$
$\sup A$	supremum of a (numerical) set A; if $A = (a, b)$ or $A = (a, b]$, then $\sup A = b$
$\tan x$	tangent, odd trigonometric function of period π, $\tan x = \sin x/\cos x$
$\tanh x$	hyperbolic tangent, $\tanh x = \sinh x/\cosh x$
x	independent variable, argument
x, y, z	spatial variables (Cartesian coordinates)

$Y_\nu(x)$ Bessel function of the second kind; $Y_\nu(x) = \dfrac{J_\nu(x)\cos(\pi\nu) - J_{-\nu}(x)}{\sin(\pi\nu)}$

y dependent variable, function; one often writes $y = y(x)$ or $y = f(x)$

y'_x first derivative of a function $y = f(x)$, also denoted by y', $\dfrac{dy}{dx}$, $f'(x)$

y''_{xx} second derivative of a function $y = f(x)$, also denoted by y'', $\dfrac{d^2y}{dx^2}$, $f''(x)$

$y_x^{(n)}$ nth derivative of a function $y = f(x)$, also denoted by $\dfrac{d^ny}{dx^n}$

$z = x + iy$ complex number; x is the real part of z, y is the imaginary part of z, $i^2 = -1$

$\bar{z} = x - iy$ complex conjugate number, $i^2 = -1$

$|z|$ modulus of a complex number; if $z = x + iy$, then $|z| = \sqrt{x^2 + y^2}$.

Greek alphabet

$\Gamma(\alpha)$ gamma function, $\Gamma(\alpha) = \displaystyle\int_0^\infty e^{-t}t^{\alpha-1}\,dt$

$\gamma(\alpha, x)$ incomplete gamma function, $\gamma(\alpha, x) = \displaystyle\int_0^x e^{-t}t^{\alpha-1}\,dt$

$\Phi(a, b; x)$ degenerate hypergeometric function, $\Phi(a,b;x) = 1 + \displaystyle\sum_{n=1}^{\infty} \frac{a(a+1)\ldots(a+n-1)}{b(b+1)\ldots(b+n-1)}\,\frac{x^n}{n!}$

Δ Laplace operator; in the two-dimensional case, $\Delta w = \frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2}$, where x and y are Cartesian coordinates

Δx increment of the argument

Δy increment of the function; if $y = f(x)$, then $\Delta y = f(x + \Delta x) - f(x)$

δ_{nm} Kronecker delta, $\delta_{nm} = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{if } n \neq m \end{cases}$

π the number "pi" (ratio of the circumference to the diameter), $\pi = 3.141592\ldots$

Remarks

1. If a formula or a solution contains an expression like $\dfrac{f(x)}{a-2}$, it is often not stated explicitly that the assumption $a \neq 2$ is implied.

2. If a formula or a solution contains derivatives of some functions, then the functions are assumed to be differentiable.

3. If a formula or a solution contains definite integrals, then the integrals are supposed to be convergent.

4. ODE and PDE are conventional abbreviations for ordinary differential equation and partial differential equation, respectively.

Part I

Definitions, Formulas, Methods, and Theorems

Chapter 1
Arithmetic and Elementary Algebra

1.1. Real Numbers

1.1.1. Integer Numbers

1.1.1-1. Natural, integer, even, and odd numbers.

Natural numbers: 1, 2, 3, . . . (all positive whole numbers).

Integer numbers (or simply *integers*): $0, \pm1, \pm2, \pm3, \ldots$

Even numbers: 0, 2, 4, . . . (all nonnegative integers that can be divided evenly by 2). An even number can generally be represented as $n = 2k$, where $k = 0, 1, 2, \ldots$

Remark 1. Sometimes all integers that are multiples of 2, such as $0, \pm2, \pm4, \ldots$, are considered to be even numbers.

Odd numbers: 1, 3, 5, . . . (all natural numbers that cannot be divided evenly by 2). An odd number can generally be represented as $n = 2k + 1$, where $k = 0, 1, 2, \ldots$

Remark 2. Sometimes all integers that are not multiples of 2, such as $\pm1, \pm3, \pm5, \ldots$, are considered to be odd numbers.

All integers as well as even numbers and odd numbers form *infinite countable sets*, which means that the elements of these sets can be enumerated using the natural numbers 1, 2, 3, . . .

1.1.1-2. Prime and composite numbers.

A *prime number* is a positive integer that is greater than 1 and has no positive integer divisors other than 1 and itself. The prime numbers form an infinite countable set. The first ten prime numbers are: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, . . .

A *composite number* is a positive integer that is greater than 1 and is not prime, i.e., has factors other than 1 and itself. Any composite number can be uniquely factored into a product of prime numbers. The following numbers are composite: $4 = 2 \times 2$, $6 = 2 \times 3$, $8 = 2^3$, $9 = 3^2$, $10 = 2 \times 5$, $12 = 2^2 \times 3$, . . .

The number 1 is a special case that is considered to be neither composite nor prime.

1.1.1-3. Divisibility tests.

Below are some simple rules helping to determine if an integer is divisible by another integer.

All integers are divisible by 1.

Divisibility by 2: last digit is divisible by 2.

Divisibility by 3: sum of digits is divisible by 3.

Divisibility by 4: two last digits form a number divisible by 4.

Divisibility by 5: last digit is either 0 or 5.

Divisibility by 6: divisible by both 2 and 3.

Divisibility by 9: sum of digits is divisible by 9.

Divisibility by 10: last digit is 0.

Divisibility by 11: the difference between the sum of the odd numbered digits (1st, 3rd, 5th, etc.) and the sum of the even numbered digits (2nd, 4th, etc.) is divisible by 11.

Example 1. Let us show that the number 80729 is divisible by 11.

The sum of the odd numbered digits is $\Sigma_1 = 8 + 7 + 9 = 24$. The sum of the even numbered digits is $\Sigma_2 = 0 + 2 = 2$. The difference between them is $\Sigma_1 - \Sigma_2 = 22$ and is divisible by 11. Consequently, the original number is also divisible by 11.

1.1.1-4. Greatest common divisor and least common multiple.

1°. The *greatest common divisor* of natural numbers $a_1, a_2, \ldots, a_n$ is the largest natural number, b, which is a common divisor to $a_1, \ldots, a_n$.

Suppose some positive numbers $a_1, a_2, \ldots, a_n$ are factored into products of primes so that

$$a_1 = p_1^{k_{11}} p_2^{k_{12}} \ldots p_m^{k_{1m}}, \quad a_2 = p_1^{k_{21}} p_2^{k_{22}} \ldots p_m^{k_{2m}}, \quad \ldots, \quad a_n = p_1^{k_{n1}} p_2^{k_{n2}} \ldots p_m^{k_{nm}},$$

where $p_1, p_2, \ldots, p_n$ are different prime numbers, the k_{ij} are positive integers ($i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$). Then the greatest common divisor b of $a_1, a_2, \ldots, a_n$ is calculated as

$$b = p_1^{\sigma_1} p_2^{\sigma_2} \ldots p_m^{\sigma_m}, \qquad \sigma_j = \min_{1 \le i \le n} k_{ij}.$$

Example 2. The greatest common divisor of 180 and 280 is $2^2 \times 5 = 20$ due to the following factorization:

$$180 = 2^2 \times 3^2 \times 5 = 2^2 \times 3^2 \times 5^1 \times 7^0,$$
$$280 = 2^3 \times 5 \times 7 = 2^3 \times 3^0 \times 5^1 \times 7^1.$$

2°. The *least common multiple* of n natural numbers $a_1, a_2, \ldots, a_n$ is the smallest natural number, A, that is a multiple of all the a_k.

Suppose some natural numbers $a_1, \ldots, a_n$ are factored into products of primes just as in Item 1°. Then the least common multiple of all the a_k is calculated as

$$A = p_1^{\nu_1} p_2^{\nu_2} \ldots p_m^{\nu_m}, \qquad \nu_j = \max_{1 \le i \le n} k_{ij}.$$

Example 3. The least common multiple of 180 and 280 is equal to $2^3 \times 3^2 \times 5^1 \times 7^1 = 2520$ due to their factorization given in Example 2.

1.1.2. Real, Rational, and Irrational Numbers

1.1.2-1. Real numbers.

The *real numbers* are all the positive numbers, negative numbers, and zero. Any real number can be represented by a *decimal fraction* (or simply *decimal*), finite or infinite. The set of all real numbers is denoted by $\mathbb{R}$.

All real numbers are categorized into two classes: the *rational* numbers and *irrational* numbers.

1.1.2-2. Rational numbers.

A *rational number* is a real number that can be written as a fraction (ratio) p/q with integer p and q ($q \neq 0$). It is only the rational numbers that can be written in the form of finite (terminating) or periodic (recurring) decimals (e.g., $1/8 = 0.125$ and $1/6 = 0.16666\ldots$). Any integer is a rational number.

The rational numbers form an infinite countable set. The set of all rational numbers is everywhere dense. This means that, for any two distinct rational numbers a and b such that $a < b$, there exists at least one more rational number c such that $a < c < b$, and hence there are infinitely many rational numbers between a and b. (Between any two rational numbers, there always exist irrational numbers.)

1.1.2-3. Irrational numbers.

An *irrational number* is a real number that is not rational; no irrational number can be written as a fraction p/q with integer p and q ($q \neq 0$). To the irrational numbers there correspond nonperiodic (nonrepeating) decimals. Examples of irrational numbers: $\sqrt{3} = 1.73205\ldots$, $\pi = 3.14159\ldots$

The set of irrational numbers is everywhere dense, which means that between any two distinct irrational numbers, there are both rational and irrational numbers. The set of irrational numbers is uncountable.

1.2. Equalities and Inequalities. Arithmetic Operations. Absolute Value

1.2.1. Equalities and Inequalities

1.2.1-1. Basic properties of equalities.

► *Throughout Paragraphs 1.2.1-1 and 1.2.1-2, it is assumed that a, b, c, d are real numbers.*

1. If $a = b$, then $b = a$.
2. If $a = b$, then $a + c = b + c$, where c is any real number; furthermore, if $a + c = b + c$, then $a = b$.
3. If $a = b$, then $ac = bc$, where c is any real number; furthermore, if $ac = bc$ and $c \neq 0$, then $a = b$.
4. If $a = b$ and $b = c$, then $a = c$.
5. If $ab = 0$, then either $a = 0$ or $b = 0$; furthermore, if $ab \neq 0$, then $a \neq 0$ and $b \neq 0$.

1.2.1-2. Basic properties of inequalities.

1. If $a < b$, then $b > a$.
2. If $a \leq b$ and $b \leq a$, then $a = b$.
3. If $a \leq b$ and $b \leq c$, then $a \leq c$.
4. If $a < b$ and $b \leq c$ (or $a \leq b$ and $b < c$), then $a < c$.
5. If $a < b$ and $c < d$ (or $c = d$), then $a + c < b + d$.
6. If $a \leq b$ and $c > 0$, then $ac \leq bc$.
7. If $a \leq b$ and $c < 0$, then $ac \geq bc$.
8. If $0 < a \leq b$ (or $a \leq b < 0$), then $1/a \geq 1/b$.

1.2.2. Addition and Multiplication of Numbers

1.2.2-1. Addition of real numbers.

The sum of real numbers is a real number.

Properties of addition:

$$a + 0 = a \qquad \text{(property of zero)},$$
$$a + b = b + a \qquad \text{(addition is commutative)},$$
$$a + (b + c) = (a + b) + c = a + b + c \qquad \text{(addition is associative)},$$

where a, b, c are arbitrary real numbers.

For any real number a, there exists its unique *additive inverse*, or its *opposite*, denoted by $-a$, such that

$$a + (-a) = a - a = 0.$$

1.2.2-2. Multiplication of real numbers.

The product of real numbers is a real number.

Properties of multiplication:

$$a \times 0 = 0 \qquad \text{(property of zero)},$$
$$ab = ba \qquad \text{(multiplication is commutative)},$$
$$a(bc) = (ab)c = abc \qquad \text{(multiplication is associative)},$$
$$a \times 1 = 1 \times a = a \qquad \text{(multiplication by unity)},$$
$$a(b + c) = ab + ac \qquad \text{(multiplication is distributive)},$$

where a, b, c are arbitrary real numbers.

For any nonzero real number a, there exists its unique *multiplicative inverse*, or its *reciprocal*, denoted by a^{-1} or $1/a$, such that

$$aa^{-1} = 1 \quad (a \neq 0).$$

1.2.3. Ratios and Proportions

1.2.3-1. Operations with fractions and properties of fractions.

Ratios are written as fractions: $a : b = a/b$. The number a is called the *numerator* and the number b ($b \neq 0$) is called the *denominator* of a fraction.

Properties of fractions and operations with fractions:

$$\frac{a}{1} = a, \qquad \frac{a}{b} = \frac{ab}{bc} = \frac{a:c}{b:c} \qquad \text{(simplest properties of fractions)};$$

$$\frac{a}{b} \pm \frac{c}{b} = \frac{a \pm c}{b}, \qquad \frac{a}{b} \pm \frac{c}{d} = \frac{ad \pm bc}{bd} \qquad \text{(addition and subtraction of fractions)};$$

$$\frac{a}{b} \times c = \frac{ac}{b}, \qquad \frac{a}{b} \times \frac{c}{d} = \frac{ac}{bc} \qquad \text{(multiplication by a number and by a fraction)};$$

$$\frac{a}{b} : c = \frac{a}{bc}, \qquad \frac{a}{b} : \frac{c}{d} = \frac{ad}{bc} \qquad \text{(division by a number and by a fraction)}.$$

1.2.3-2. Proportions. Simplest relations. Derivative proportions.

A proportion is an equation with a ratio on each side. A proportion is denoted by $a/b = c/d$ or $a : b = c : d$.

1°. The following simplest relations follow from $a/b = c/d$:

$$ad = bc, \quad \frac{a}{c} = \frac{b}{d}, \quad a = \frac{bc}{d}, \quad b = \frac{ad}{c}.$$

2°. The following derivative proportions follow from $a/b = c/d$:

$$\frac{ma + nb}{pa + qb} = \frac{mc + nd}{pc + qd},$$
$$\frac{ma + nc}{pa + qc} = \frac{mb + nd}{pb + qd},$$

where m, n, p, q are arbitrary real numbers.

Some special cases of the above formulas:

$$\frac{a \pm b}{b} = \frac{c \pm d}{d}, \qquad \frac{a - b}{a + b} = \frac{c - d}{c + d}.$$

1.2.4. Percentage

1.2.4-1. Definition. Main percentage problems.

A *percentage* is a way of expressing a ratio or a fraction as a whole number, by using 100 as the denominator. One *percent* is one per one hundred, or one hundredth of a whole number; notation: 1%.

Below are the statements of main percentage problems and their solutions.

1°. Find the number b that makes up $p\%$ of a number a. Answer: $b = \frac{ap}{100}$.

2°. Find the number a whose $p\%$ is equal to a number b. Answer: $a = \frac{100b}{p}$.

3°. What percentage does a number b make up of a number a? Answer: $p = \frac{100b}{a}\%$.

1.2.4-2. Simple and compound percentage.

1°. *Simple percentage*. Suppose a cash deposit is increased yearly by the same amount defined as a percentage, $p\%$, of the initial deposit, a. Then the amount accumulated after t years is calculated by the simple percentage formula

$$x = a\Big(1 + \frac{pt}{100}\Big).$$

2°. *Compound percentage*. Suppose a cash deposit is increased yearly by an amount defined as a percentage, $p\%$, of the deposit in the previous year. Then the amount accumulated after t years is calculated by the compound percentage formula

$$x = a\Big(1 + \frac{p}{100}\Big)^t,$$

where a is the initial deposit.

1.2.5. Absolute Value of a Number (Modulus of a Number)

1.2.5-1. Definition.

The absolute value of a real number a, denoted by $|a|$, is defined by the formula

$$|a| = \begin{cases} a & \text{if } a \ge 0, \\ -a & \text{if } a < 0. \end{cases}$$

An important property: $|a| \ge 0$.

1.2.5-2. Some formulas and inequalities.

1°. The following relations hold true:

$$|a| = |-a| = \sqrt{a^2}, \quad a \le |a|,$$
$$\big||a| - |b|\big| \le |a + b| \le |a| + |b|,$$
$$\big||a| - |b|\big| \le |a - b| \le |a| + |b|,$$
$$|ab| = |a|\,|b|, \quad |a/b| = |a|/|b|.$$

2°. From the inequalities $|a| \le A$ and $|b| \le B$ it follows that $|a + b| \le A + B$ and $|ab| \le AB$.

1.3. Powers and Logarithms

1.3.1. Powers and Roots

1.3.1-1. Powers and roots: the main definitions.

Given a positive real number a and a positive integer n, the nth power of a, written as a^n, is defined as the multiplication of a by itself repeated n times:

$$a^n = \underbrace{a \times a \times a \times \cdots \times a}_{n \text{ multipliers}}.$$

The number a is called the *base* and n is called the *exponent*.

Obvious properties: $0^n = 0$, $1^n = 1$, $a^1 = a$.

Raising to the zeroth power: $a^0 = 1$, where $a \ne 0$. Sometimes 0^0 is taken as undefined, but it is often sensibly defined as 1.

Raising to a negative power: $a^{-n} = \dfrac{1}{a^n}$, where n is a positive integer.

If a is a positive real number and n is a positive integer, then the *nth arithmetic root* or *radical* of a, written as $\sqrt[n]{a}$, is the unique positive real number b such that $b^n = a$. In the case of $n = 2$, the brief notation $\sqrt{a}$ is used to denote $\sqrt[2]{a}$.

The following relations hold:

$$\sqrt[n]{0} = 0, \qquad \sqrt[n]{1} = 1, \qquad \left(\sqrt[n]{a}\right)^n = a.$$

Raising to a fractional power $p = m/n$, where m and n are natural numbers:

$$a^p = a^{m/n} = \sqrt[n]{a^m}, \qquad a \ge 0.$$

1.3.1-2. Operations with powers and roots.

The properties given below are valid for any rational and real exponents p and q ($a > 0$, $b > 0$):

$$a^{-p} = \frac{1}{a^p}, \quad a^p a^q = a^{p+q}, \quad \frac{a^p}{a^q} = a^{p-q},$$
$$(ab)^p = a^p b^p, \quad \left(\frac{a}{b}\right)^q = \frac{a^q}{b^q}, \quad (a^p)^q = a^{pq}.$$

In operations with roots (radicals) the following properties are used:

$$\sqrt[n]{ab} = \sqrt[n]{a}\,\sqrt[n]{b}, \quad \sqrt[n]{\frac{a}{b}} = \frac{\sqrt[n]{a}}{\sqrt[n]{b}}, \quad \sqrt[n]{a^m} = \left(\sqrt[n]{a}\right)^m, \quad \sqrt[n]{\sqrt[m]{a}} = \sqrt[mn]{a}.$$

Remark. It often pays to represent roots as powers with rational exponents and apply the properties of operations with powers.

1.3.2. Logarithms

1.3.2-1. Definition. The main logarithmic identity.

The *logarithm of a positive number b to a given base a* is the exponent of the power c to which the base a must be raised to produce b. It is written as $\log_a b = c$.

Equivalent representations:

$$\log_a b = c \quad \Longleftrightarrow \quad a^c = b,$$

where $a > 0$, $a \neq 1$, and $b > 0$.

Main logarithmic identity:

$$a^{\log_a b} = b.$$

Simple properties:

$$\log_a 1 = 0, \quad \log_a a = 1.$$

1.3.2-2. Properties of logarithms. The common and natural logarithms.

Properties of logarithms:

$$\log_a(bc) = \log_a b + \log_a c, \qquad \log_a\left(\frac{b}{c}\right) = \log_a b - \log_a c,$$
$$\log_a(b^k) = k \log_a b, \qquad \log_{a^k} b = \frac{1}{k}\log_a b \quad (k \neq 0),$$
$$\log_a b = \frac{1}{\log_b a} \quad (b \neq 1), \qquad \log_a b = \frac{\log_c b}{\log_c a} \quad (c \neq 1),$$

where $a > 0$, $a \neq 1$, $b > 0$, $c > 0$, and k is any number.

The logarithm to the base 10 is called the *common* or *decadic logarithm* and written as

$$\log_{10} b = \log b \quad \text{or sometimes} \quad \log_{10} b = \lg b.$$

The logarithm to the base e (the *base of natural logarithms*) is called the *natural logarithm* and written as

$$\log_e b = \ln b,$$

where $e = \lim\limits_{n\to\infty}\left(1 + \frac{1}{n}\right)^n = 2.718281\ldots$

1.4. Binomial Theorem and Related Formulas

1.4.1. Factorials. Binomial Coefficients. Binomial Theorem

1.4.1-1. Factorials. Binomial coefficients.

Factorial:

$$0! = 1! = 1,$$
$$n! = 1 \times 2 \times 3 \times \ldots \times (n-1)n, \quad n = 2,\ 3,\ 4,\ \ldots$$

Double factorial:

$$0!! = 1!! = 1,$$
$$n!! = \begin{cases} (2k)!! & \text{if } n = 2k, \\ (2k+1)!! & \text{if } n = 2k+1, \end{cases}$$
$$(2k)!! = 2 \times 4 \times 6 \times \ldots \times (2k-2)(2k) = 2^k k!,$$
$$(2k+1)!! = 1 \times 3 \times 5 \times \ldots \times (2k-1)(2k+1),$$

where n and k are natural numbers.

Binomial coefficients:

$$C_n^k = \binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{n(n-1)\ldots(n-k+1)}{k!}, \quad k = 1,\ 2,\ 3,\ \ldots,\ n;$$
$$C_a^k = \frac{a(a-1)\ldots(a-k+1)}{k!}, \quad \text{where} \quad k = 1,\ 2,\ 3, \ldots,$$

where n is a natural number and a is any number.

1.4.1-2. Binomial theorem.

Let a, b, and c be real (or complex) numbers. The following formulas hold true:

$$(a \pm b)^2 = a^2 \pm 2ab + b^2,$$
$$(a \pm b)^3 = a^3 \pm 3a^2 b + 3ab^2 \pm b^3,$$
$$(a \pm b)^4 = a^4 \pm 4a^3 b + 6a^2 b^2 \pm 4ab^3 + b^4,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$(a+b)^n = \sum_{k=0}^{n} C_n^k a^{n-k} b^k, \quad n = 1,\ 2,\ \ldots$$

The last formula is known as the *binomial theorem*, where the C_n^k are binomial coefficients.

1.4.2. Related Formulas

1.4.2-1. Formulas involving powers ≤ 4.

$$a^2 - b^2 = (a-b)(a+b),$$
$$a^3 + b^3 = (a+b)(a^2 - ab + b^2),$$
$$a^3 - b^3 = (a-b)(a^2 + ab + b^2),$$
$$a^4 - b^4 = (a-b)(a+b)(a^2 + b^2),$$
$$(a+b+c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc,$$
$$a^4 + a^2 b^2 + b^4 = (a^2 + ab + b^2)(a^2 - ab + b^2).$$

1.4.2-2. Formulas involving arbitrary powers.

Let n be any positive integer. Then

$$a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + \cdots + ab^{n-2} + b^{n-1}).$$

If n is a positive even number, then

$$\begin{aligned} a^n - b^n &= (a+b)(a^{n-1} - a^{n-2}b + \cdots + ab^{n-2} - b^{n-1}) \\ &= (a-b)(a+b)(a^{n-2} + a^{n-4}b^2 + \cdots + a^2b^{n-4} + b^{n-2}). \end{aligned}$$

If n is a positive odd number, then

$$a^n + b^n = (a+b)(a^{n-1} - a^{n-2}b + \cdots - ab^{n-2} + b^{n-1}).$$

1.5. Arithmetic and Geometric Progressions. Finite Sums and Products

1.5.1. Arithmetic and Geometric Progressions

1.5.1-1. Arithmetic progression.

1°. An *arithmetic progression*, or *arithmetic sequence*, is a sequence of real numbers for which each term, starting from the second, is the previous term plus a constant d, called the *common difference*. In general, the terms of an arithmetic progression are expressed as

$$a_n = a_1 + (n-1)d, \qquad n = 1, 2, 3, \ldots,$$

where a_1 is the first term of the progression. An arithmetic progression is called *increasing* if $d > 0$ and *decreasing* if $d < 0$.

2°. An arithmetic progression has the property

$$a_n = \tfrac{1}{2}(a_{n-1} + a_{n+1}).$$

3°. The sum of n first terms of an arithmetic progression is called an *arithmetic series* and is calculated as

$$S_n = a_1 + \cdots + a_n = \tfrac{1}{2}(a_1 + a_n)n = \tfrac{1}{2}[2a_1 + (n-1)d]n.$$

1.5.1-2. Geometric progression.

1°. A *geometric progression*, or *geometric sequence*, is a sequence of real numbers for which each term, starting from the second, is the previous term multiplied by a constant q, called the *common ratio*. In general, the terms of a geometric progression are expressed as

$$a_n = a_1 q^{n-1}, \qquad n = 1, 2, 3, \ldots,$$

where a_1 is the first term of the progression.

2°. A geometric progression with positive terms has the property

$$a_n = \sqrt{a_{n-1}a_{n+1}}.$$

3°. The sum of n first terms of an arithmetic progression is called a *geometric series* and is calculated as ($q \neq 1$)

$$S_n = a_1 + \cdots + a_n = a_1 \frac{1-q^n}{1-q}.$$

1.5.2. Finite Series and Products

1.5.2-1. Notations for finite series and products.

A finite series is just the sum of a finite number of terms and a finite product is the product of a finite number of terms. These are written as

$$a_1 + a_2 + \cdots + a_n = \sum_{k=1}^{n} a_k, \qquad a_m + a_{m+1} + \cdots + a_n = \sum_{k=m}^{n} a_k;$$

$$a_1 a_2 \ldots a_n = \prod_{k=1}^{n} a_k, \qquad a_m a_{m+1} \ldots a_n = \prod_{k=m}^{n} a_k,$$

where m is a nonnegative integer ($m \le n$). The variable k appearing on the right-hand sides of the above formulas is called the *index of summation* (for series) or the *index of multiplication* (for products). The 1 and n (or the m and n) are the *upper* and *lower limits of summation* (*multiplication*).

The values of sums (products) are independent of the names used to denote the index of summation (multiplication): $\sum\limits_{k=1}^{n} a_k = \sum\limits_{j=1}^{n} a_j$, $\prod\limits_{k=1}^{n} a_k = \prod\limits_{i=1}^{n} a_i$. Such indices are called dummy indices.

1.5.2-2. Formulas for summation of some finite series.

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2};$$

$$\sum_{k=1}^{n} (-1)^k k = (-1)^n \left[\frac{n-1}{2}\right], \qquad [m] \text{ is the integer part of } m;$$

$$\sum_{k=0}^{n} (2k+1) = (n+1)^2;$$

$$\sum_{k=0}^{n} (-1)^k (2k+1) = (-1)^n (n+1);$$

$$\sum_{k=1}^{n} k^2 = \frac{1}{6} n(n+1)(2n+1);$$

$$\sum_{k=1}^{n} (-1)^k k^2 = (-1)^n \frac{n(n+1)}{2};$$

$$\sum_{k=0}^{n} (2k+1)^2 = \frac{1}{3}(n+1)(2n+1)(2n+3);$$

$$\sum_{k=0}^{n} (-1)^k (2k+1)^2 = 2(-1)^n (n+1)^2 - \frac{1}{2}\left[1 + (-1)^n\right];$$

$$\sum_{k=1}^{n} (k+a)(k+b) = \frac{1}{6} n(n+1)(2n+1+3a+3b) + nab.$$

► *A large number of formulas for the summation of various finite series can be found in Section T1.1.*

1.6. Mean Values and Inequalities of General Form

1.6.1. Arithmetic Mean, Geometric Mean, and Other Mean Values. Inequalities for Mean Values

1.6.1-1. Arithmetic mean, geometric mean, and other mean values.

The *arithmetic mean* of a set of n real numbers $a_1, a_2, \ldots, a_n$ is defined as

$$m_{\mathrm{a}} = \frac{a_1 + a_2 + \cdots + a_n}{n}. \tag{1.6.1.1}$$

Geometric mean of n positive numbers $a_1, a_2, \ldots, a_n$:

$$m_{\mathrm{g}} = (a_1 a_2 \ldots a_n)^{1/n}. \tag{1.6.1.2}$$

Harmonic mean of n real numbers $a_1, a_2, \ldots, a_n$:

$$m_{\mathrm{h}} = \frac{n}{(1/a_1) + (1/a_2) + \cdots + (1/a_n)}, \qquad a_k \neq 0. \tag{1.6.1.3}$$

Quadratic mean (or *root mean square*) of n real numbers $a_1, a_2, \ldots, a_n$:

$$m_{\mathrm{q}} = \sqrt{\frac{a_1^2 + a_2^2 + \cdots + a_n^2}{n}}. \tag{1.6.1.4}$$

1.6.1-2. Basic inequalities for mean values.

Given n positive numbers $a_1, a_2, \ldots, a_n$, the following inequalities hold true:

$$m_{\mathrm{h}} \le m_{\mathrm{g}} \le m_{\mathrm{a}} \le m_{\mathrm{q}}, \tag{1.6.1.5}$$

where the mean values are defined above by (1.6.1.1)–(1.6.1.4). The equalities in (1.6.1.5) are attained only if $a_1 = a_2 = \cdots = a_n$.

To make it easier to remember, let us rewrite inequalities (1.6.1.5) in words as

harmonic mean ≤ geometric mean ≤ arithmetic mean ≤ quadratic mean.

1.6.1-3. General approach to defining mean values.

Let $f(x)$ be a continuous monotonic function defined on the interval $0 \le x < \infty$.

The *functional mean* with respect to the function $f(x)$ for n positive real numbers $a_1, a_2, \ldots, a_n$ is introduced as follows:

$$m_f = f^{-1}\left(\frac{f(a_1) + f(a_2) + \cdots + f(a_n)}{n}\right), \tag{1.6.1.6}$$

where $f^{-1}(y)$ is the inverse of $f(x)$.

The mean values defined by (1.6.1.1)–(1.6.1.4) in Paragraph 1.6.1-1 are all special cases of the functional mean (1.6.1.6), provided the real numbers $a_1, a_2, \ldots, a_n$ are all positive. Specifically,

the arithmetic mean is the functional mean with respect to $f(x) = x$,
the geometric mean is the functional mean with respect to $f(x) = \ln x$,
the harmonic mean is the functional mean with respect to $f(x) = 1/x$,
the quadratic mean is the functional mean with respect to $f(x) = x^2$.

1.6.2. Inequalities of General Form

1.6.2-1. Triangle inequality, Cauchy inequality, and related inequalities.

Let a_k and b_k be real numbers with $k = 1, 2, \ldots, n$.

Generalized triangle inequality:

$$\left|\sum_{k=1}^{n} a_k\right| \le \sum_{k=1}^{n} |a_k|.$$

Cauchy's inequality (also known as the *Cauchy–Bunyakovski inequality* or *Cauchy–Schwarz–Bunyakovski inequality*):

$$\left(\sum_{k=1}^{n} a_k b_k\right)^2 \le \left(\sum_{k=1}^{n} a_k^2\right)\left(\sum_{k=1}^{n} b_k^2\right).$$

Minkowski's inequality:

$$\left(\sum_{k=1}^{n} |a_k + b_k|^p\right)^{\frac{1}{p}} \le \left(\sum_{k=1}^{n} |a_k|^p\right)^{\frac{1}{p}} + \left(\sum_{k=1}^{n} |b_k|^p\right)^{\frac{1}{p}}, \qquad p \ge 1.$$

Hölder's inequality (reduces to Cauchy's inequality at $p = 2$):

$$\left|\sum_{k=1}^{n} a_k b_k\right| \le \left(\sum_{k=1}^{n} |a_k|^p\right)^{\frac{1}{p}} \left(\sum_{k=1}^{n} |b_k|^{\frac{p}{p-1}}\right)^{\frac{p-1}{p}}, \qquad p > 1.$$

1.6.2-2. Chebyshev's inequalities.

Chebyshev's inequalities:

$$\left(\sum_{k=1}^{n} a_k\right)\left(\sum_{k=1}^{n} b_k\right) \le n\left(\sum_{k=1}^{n} a_k b_k\right) \quad \text{if} \quad 0 < a_1 \le a_2 \le \cdots < a_n,\ 0 < b_1 \le b_2 \le \cdots < b_n;$$

$$\left(\sum_{k=1}^{n} a_k\right)\left(\sum_{k=1}^{n} b_k\right) \ge n\left(\sum_{k=1}^{n} a_k b_k\right) \quad \text{if} \quad 0 < a_1 \le a_2 \le \cdots < a_n,\ b_1 \ge b_2 \ge \cdots \ge b_n > 0;$$

Generalized Chebyshev inequalities:

$$\left(\frac{1}{n}\sum_{k=1}^{n}a_k^p\right)^{1/p}\left(\frac{1}{n}\sum_{k=1}^{n}b_k^p\right)^{1/p}\le\left(\frac{1}{n}\sum_{k=1}^{n}a_k^pb_k^p\right)^{1/p}$$
$$\text{if}\quad 0<a_1\le a_2\le\cdots<a_n,\quad 0<b_1\le b_2\le\cdots<b_n;$$
$$\left(\frac{1}{n}\sum_{k=1}^{n}a_k^p\right)^{1/p}\left(\frac{1}{n}\sum_{k=1}^{n}b_k^p\right)^{1/p}\ge\left(\frac{1}{n}\sum_{k=1}^{n}a_k^pb_k^p\right)^{1/p}$$
$$\text{if}\quad 0<a_1\le a_2\le\cdots<a_n,\quad b_1\ge b_2\ge\cdots\ge b_n>0.$$

1.6.2-3. Generalizations of inequalities for means.

1°. The following inequality holds:

$$\left(a_1^{p_1}a_2^{p_2}\ldots a_n^{p_n}\right)^{\frac{1}{p_1+p_2+\cdots+p_n}}\le\frac{a_1p_1+a_2p_2+\cdots+a_np_n}{p_1+p_2+\cdots+p_n},$$

where the a_k and p_k are all positive. In the special case $p_1=p_2=\cdots=p_n=1$, we have the well-known inequality stating that the geometric mean of a series of positive numbers does not exceed their arithmetic mean (see Paragraph 1.6.1-2).

2°. The following inequality holds:

$$\frac{p_1+p_2+\cdots+p_n}{(p_1/a_1)+(p_2/a_2)+\cdots+(p_n/a_n)}\le\left(a_1^{p_1}a_2^{p_2}\ldots a_n^{p_n}\right)^{\frac{1}{p_1+p_2+\cdots+p_n}},$$

where the a_k and p_k are all positive. In the special case $p_1=p_2=\cdots=p_n=1$, we have the well-known inequality stating that the harmonic mean of a series of positive numbers does not exceed their geometric mean (see Paragraph 1.6.1-2).

1.6.2-4. Jensen's inequality.

If $f(x)$ is a convex function (in particular, with $f''>0$), then the following *Hölder–Jensen inequality* holds:

$$f\left(\frac{\sum p_kx_k}{\sum p_k}\right)\le\frac{\sum p_kf(x_k)}{\sum p_k},\tag{1.6.2.1}$$

where the x_k are any numbers and the p_k are any positive numbers; the summation is performed over all k (the limits are omitted for simplicity). The equality is attained if and only if either $x_1=x_2=\cdots=x_n$ or $f(x)$ is a linear function. If $f(x)$ is a concave function ($f''<0$), inequality (1.6.2.1) is the other way around.

The Hölder–Jensen inequality is often used to obtain various inequalities; in particular, the previous two inequalities as well as the Hölder inequality follow from it.

1.7. Some Mathematical Methods

1.7.1. Proof by Contradiction

Proof by contradiction (also known as *reductio ad absurdum*) is an indirect method of mathematical proof. It is based on the following reasoning:

1. Suppose one has to prove some statement S.

2. One assumes that the opposite of S is true.

3. Based on known axioms, definitions, theorems, formulas, and the assumption of Item 2, one arrives at a contradiction (deduces some obviously false statement).

4. One concludes that the assumption of Item 2 is false and hence the original statement S is true, which was to be proved.

Example. (Euclid's proof of the irrationality of the square root of 2 by contradiction.)

1. It is required to prove that $\sqrt{2}$ is an irrational number, that is, a real number that cannot be represented as a fraction p/q, where p and q are both integers.

2. Assume the opposite: $\sqrt{2}$ is a rational number. This means that $\sqrt{2}$ can be represented as a fraction

$$\sqrt{2} = p/q. \tag{1.7.1.1}$$

Without loss of generality the fraction p/q is assumed to be irreducible, implying that p and q are mutually prime (have no common factor other than 1).

3. Square both sides of (1.7.1.1) and then multiply by q^2 to obtain

$$2q^2 = p^2. \tag{1.7.1.2}$$

The left-hand side is divisible by 2. Then the right-hand side, p^2, and hence p is also divisible by 2. Consequently, p is an even number so that

$$p = 2n, \tag{1.7.1.3}$$

where n is an integer. Substituting (1.7.1.3) into (1.7.1.2) and then dividing by 2 yields

$$q^2 = 2p^2. \tag{1.7.1.4}$$

Now it can be concluded, just as above, that q^2 and hence q must be divisible by 2. Consequently, q is an even number so that

$$q = 2m, \tag{1.7.1.5}$$

where m is an integer.

It is now apparent from (1.7.1.3) and (1.7.1.5) that the fraction p/q is not simple, since p and q have a common factor 2. This contradicts the assumption made in Item 2.

4. It follows from the results of Item 3 that the representation of $\sqrt{2}$ in the form of a fraction (1.7.1.1) is false, which means that $\sqrt{2}$ is irrational.

1.7.2. Mathematical Induction

The method of proof by (complete) *mathematical induction* is based on the following reasoning:

1. Let $A(n)$ be a statement dependent on n with $n = 1, 2, \ldots$ (A is a hypothesis at this stage).

2. Base case. Suppose the initial statement $A(1)$ is true. This is usually established by direct substitution $n = 1$.

3. Induction step. Assume that $A(n)$ is true for any n and then, based on this assumption, prove that $A(n + 1)$ is also true.

4. Principle of mathematical induction. From the results of Items 2–3 it is concluded that the statement $A(n)$ is true for any n.

Example.

1. Prove the formula for the sum of odd numbers,

$$1 + 3 + 5 + \cdots + (2n - 1) = n^2, \tag{1.7.2.1}$$

for any natural n.

2. For $n = 1$, we have an obvious identity: $1 = 1$.

3. Let us assume that formula (1.7.2.1) holds for any n. To consider the case of $n + 1$, let us add the next term, $(2n + 1)$, to both sides of (1.7.2.1) to obtain

$$1 + 3 + 5 + \cdots + (2n - 1) + (2n + 1) = n^2 + (2n + 1) = (n + 1)^2.$$

Thus, from the assumption of the validity of formula (1.7.2.1) for any n it follows that (1.7.2.1) is also valid for $n + 1$.

4. According to the principle of mathematical induction, this proves formula (1.7.2.1).

Remark. The first step, the formulation of an original hypothesis, is the most difficult part of the method of mathematical induction. This step is often omitted from the method.

1.7.3. Proof by Counterexample

A *counterexample* is an example which is used to prove that a statement (proposition) is false. Counterexamples play an important role in mathematics. Whereas a complicated proof may be the only way to demonstrate the validity of a particular theorem, a single counter example is all that is need to refute the validity of a proposed theorem.

In general, the scheme of a proof by counterexample is as follows:

1. Given a proposition: all elements a that belong to a set A also belong to a set (possess a property) B.

2. Refutation of the proposition: one specifies an element a_* (counterexample) that belongs to A but does not belong to B.

Example. *Proposition:* Numbers in the form $2^{2^n}+1$, where n is a positive integer, were once thought to be prime.

These numbers are prime for $n = 1, 2, 3, 4$. But for $n = 5$, we have a counterexample, since

$$2^{2^5}+1 = 4294967297 = 641 \times 6700417;$$

it is a composite number.

Conclusion: When faced with a number in the form $2^{2^n}+1$, we are not allowed to assume it is either prime or composite, unless we know for sure for some other reason.

1.7.4. Method of Undetermined Coefficients

The *method of undetermined coefficients* is employed to find coefficients in expressions (such as formulas, series expansions, solutions to mathematical equations), the form of which is either known in advance or assigned based on intuitive judgment.

Example. The fractional function

$$\frac{x+2}{x(x^2-1)}, \tag{1.7.4.1}$$

whose denominator can be rewritten in the factored form $x(x+1)(x-1)$, can be represented as the sum of partial fractions

$$\frac{A}{x}+\frac{B}{x+1}+\frac{C}{x-1}, \tag{1.7.4.2}$$

where A, B, C are (undetermined) coefficients whose values are to be found. Equating (1.7.4.1) with (1.7.4.2), multiplying by $x(x^2-1)$, and rearranging, one obtains

$$(A+B+C)x^2+(-B+C-1)x-A-2=0.$$

For this equation to be valid for any x, the coefficients of the different powers of x must be set equal to zero. This results in the system of linear algebraic equations

$$A+B+C=0, \quad C-B-1=0, \quad -A-2=0.$$

On solving this system, one determines the coefficients in (1.7.4.2):

$$A=-2, \quad B=\tfrac{1}{2}, \quad C=\tfrac{3}{2}.$$

References for Chapter 1

Bronshtein, I. N. and Semendyayev, K. A., *Handbook of Mathematics, 4th Edition*, Springer-Verlag, Berlin, 2004.

Courant, R. and Robbins, H., *What Is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd Edition*, Oxford University Press, Oxford, 1996.

Franklin, J. and Daoud, A., *Introduction to Proofs in Mathematics*, Prentice Hall, New York, 1988.

Garnier, R. and Taylor, J., *100% Mathematical Proof*, John Wiley & Sons, New York, 1996.

Gelbaum, B. R. and Olmsted, J. M. H., *Theorems and Counterexamples in Mathematics*, Springer-Verlag, New York, 1990.

Jordan, B. E. and Palow, W. P., *Integrated Arithmetic & Algebra*, Addison-Wesley, Boston, 1999.

Krantz, S. G., *Dictionary of Algebra, Arithmetic, and Trigonometry*, CRC Press, Boca Raton, 2001.

Pólya, G., *Mathematics and Plausible Reasoning, Vol. 1: Induction and Analogy in Mathematics*, Princeton University Press, Princeton, 1990.

Rossi, R. J., *Theorems, Corollaries, Lemmas, and Methods of Proof*, Wiley-Interscience, Hoboken, N.J., 2006.

Thompson, J. E., *Arithmetic for the Practical Man*, Van Nostrand Reinhold, New York, 1973.

Weisstein, E. W., *CRC Concise Encyclopedia of Mathematics, 2nd Edition*, CRC Press, Boca Raton, 2003.

Zwillinger, D., *CRC Standard Mathematical Tables and Formulae, 31st Edition*, CRC Press, Boca Raton, 2002.

Chapter 2
Elementary Functions

Basic elementary functions: power, exponential, logarithmic, trigonometric, and inverse trigonometric (arc-trigonometric or antitrigonometric) functions. All other elementary functions are obtained from the basic elementary functions and constants by means of the four arithmetic operations (addition, subtraction, multiplication, and division) and the operation of composition (composite functions).

The graphs and the main properties of the basic as well as some other frequently occurring elementary functions of the real variable are described below.

2.1. Power, Exponential, and Logarithmic Functions

2.1.1. Power Function: $y = x^\alpha$ (α is an Arbitrary Real Number)

2.1.1-1. Graphs of the power function.

General properties of the graphs: the point $(1, 1)$ belongs to all the graphs, and $y > 0$ for $x > 0$. For $\alpha > 0$, the graphs pass through the origin $(0, 0)$; for $\alpha < 0$, the graphs have the vertical asymptote $x = 0$ ($y \to +\infty$ as $x \to 0$, $x > 0$). For $\alpha = 0$, the graph is a straight line parallel to the x-axis.

Consider more closely the following cases.

Case 1: $y = x^{2n}$, where n is a positive integer ($n = 1, 2, \ldots$). This function is defined for all real x and its range consists of all $y \geq 0$. This function is even, nonperiodic, and unbounded. It crosses the axis Oy and is tangential to the axis Ox at the origin $x = 0$, $y = 0$. On the interval $(-\infty, 0)$ this function decreases, and it increases on the interval $(0, +\infty)$. It attains its minimum $y = 0$ at $x = 0$. The graph of the function $y = x^2$ (parabola) is given in Fig. 2.1 *a*.

Case 2: $y = x^{2n+1}$, where n is a positive integer. This function is defined on the entire x-axis and its range coincides with the y-axis. This function is odd, nonperiodic, and unbounded. It crosses the x-axis and the y-axis at the origin $x = 0$, $y = 0$. It is an increasing function on the entire real axis with no points of extremum, the origin being its inflection point. The graph of the function $y = x^3$ (cubic parabola) is shown in Fig. 2.1 *a*.

Case 3: $y = x^{-2n}$, where n is a positive integer. This function is defined for all $x \neq 0$, and its range is the semiaxis $y > 0$. It is an even, nonperiodic, unbounded function having no intersection with the coordinate axes. It increases on the interval $(-\infty, 0)$, decreases on the interval $(0, +\infty)$, and has no points of extremum. This function has a vertical asymptote $x = 0$. The graph of the function $y = x^{-2}$ is given in Fig. 2.1 *b*.

Case 4: $y = x^{-2n+1}$, where n is a positive integer. This function is defined for all $x \neq 0$, and its range is the entire y-axis. It is an odd, nonperiodic, unbounded function with no intersections with the coordinate axes. This is a decreasing function on the entire real axis with no points of extremum. It has a vertical asymptote $x = 0$. The graph of the function $y = x^{-1}$ is given in Fig. 2.1 *b*.

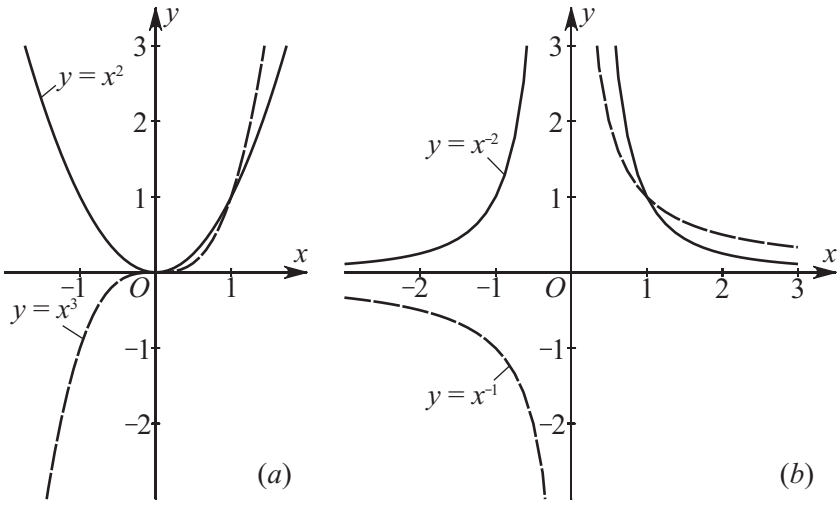


Figure 2.1. Graphs of the power function $y = x^n$, where n is an integer.

Case 5: $y = x^\alpha$ with a noninteger $\alpha > 0$. This function is defined for all* $x \geq 0$ and its range is the semiaxis $y \geq 0$. This function is neither odd nor even and it is nonperiodic and unbounded. It crosses the axes Ox and Oy at the origin $x = 0$, $y = 0$ and increases everywhere in its domain, taking its smallest value at the limit point $x = 0$, $y = 0$. The graph of the function $y = x^{1/2}$ is given in Fig. 2.2.

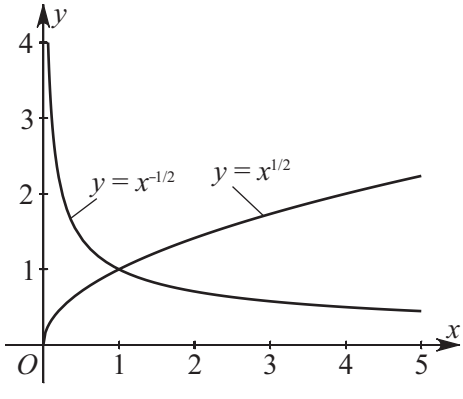


Figure 2.2. Graphs of the power function $y = x^\alpha$, where α is a noninteger.

Case 6: $y = x^\alpha$ with a noninteger $\alpha < 0$. This function is defined for all $x \geq 0$ and its range is the semiaxis $y \geq 0$. This function is neither odd nor even, it is nonperiodic and unbounded, and it has no intersections with the coordinate axes, which coincide with its horizontal and vertical asymptotes. This function is decreasing on its entire domain and has no points of extremum. The graph of the function $y = x^{-1/2}$ is given in Fig. 2.2.

* In fact, the power function $y = x^{1/n}$ with an odd integer n is defined and negative for all $x < 0$. Here, however, it is always assumed that $x \geq 0$. A similar assumption is made with regard to the functions of the form $y = x^{m/n}$, where m is a positive integer and m/n is an irreducible fraction.

2.1.1-2. Properties of the power function.

Basic properties of the power function:

$$x^\alpha x^\beta = x^{\alpha+\beta}, \quad (x_1x_2)^\alpha = x_1^\alpha x_2^\alpha, \quad (x^\alpha)^\beta = x^{\alpha\beta},$$

for any α and β, where $x > 0$, $x_1 > 0$, $x_2 > 0$.

Differentiation and integration formulas:

$$(x^\alpha)' = \alpha x^{\alpha-1}, \qquad \int x^\alpha\,dx = \begin{cases} \dfrac{x^{\alpha+1}}{\alpha+1} + C & \text{if } \alpha \neq -1, \\ \ln|x| + C & \text{if } \alpha = -1. \end{cases}$$

The Taylor series expansion in a neighborhood of an arbitrary point:

$$x^\alpha = \sum_{n=0}^{\infty} C_\alpha^n x_0^{\alpha-n}(x-x_0)^n \quad \text{for} \quad |x-x_0| < |x_0|,$$

where $C_\alpha^n = \dfrac{\alpha(\alpha-1)\ldots(\alpha-n+1)}{n!}$ are binomial coefficients.

2.1.2. Exponential Function: $y = a^x$ ($a > 0$, $a \neq 1$)

2.1.2-1. Graphs of the exponential function.

This function is defined for all x and its range is the semiaxis $y > 0$. This function is neither odd nor even, it is nonperiodic and unbounded, and it crosses the axis Oy at $y = 1$ and does not cross the axis Ox. For $a > 1$, it is an increasing function on the entire real axis; for $0 < a < 1$, it is a decreasing function. This function has no extremal points; the axis Ox is its horizontal asymptote. The graphs of these functions have the following common property: they pass through the point $(0, 1)$. The graph of $y = a^x$ is symmetrical to the graph of $y = (1/a)^x$ with respect to the y-axis. For $a > 1$, the function a^x grows faster than any power of x as $x \to +\infty$, and it decays faster than any power of $1/x$ as $x \to -\infty$. The graphs of the functions $y = 2^x$ and $y = (1/2)^x$ are given in Fig. 2.3.

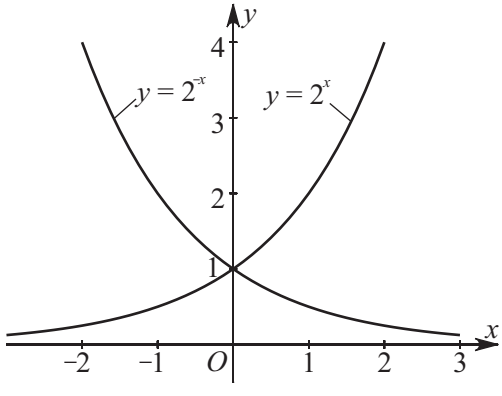


Figure 2.3. Graphs of the exponential function.

2.1.2-2. Properties of the exponential function.

Basic properties of the exponential function:

$$a^{x_1}a^{x_2} = a^{x_1+x_2}, \quad a^x b^x = (ab)^x, \quad (a^{x_1})^{x_2} = a^{x_1 x_2}.$$

Number e, *base of natural (Napierian) logarithms*, and the function e^x:

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = 2.718281\ldots, \quad e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n.$$

The formula for passing from an arbitrary base a to the base e of natural logarithms:

$$a^x = e^{x\ln a}.$$

The inequality

$$a^{x_1} > a^{x_2} \quad \Longleftrightarrow \quad \begin{cases} x_1 > x_2 & \text{if } a > 1, \\ x_1 < x_2 & \text{if } 0 < a < 1. \end{cases}$$

The limit relations for any $a > 1$ and $b > 0$:

$$\lim_{x\to+\infty}\frac{a^x}{|x|^b} = \infty, \quad \lim_{x\to-\infty} a^x|x|^b = 0.$$

Differentiation and integration formulas:

$$(e^x)' = e^x, \qquad \int e^x\,dx = e^x + C;$$
$$(a^x)' = a^x\ln a, \qquad \int a^x\,dx = \frac{a^x}{\ln a} + C.$$

The expansion in power series:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots = \sum_{k=0}^{\infty}\frac{x^k}{k!}.$$

2.1.3. Logarithmic Function: $y = \log_a x$ ($a > 0$, $a \neq 1$)

2.1.3-1. Graphs of the logarithmic function.

This function is defined for all $x > 0$ and its range is the entire y-axis. The function is neither odd nor even; it is nonperiodic and unbounded; it crosses the axis Ox at $x = 1$ and does not cross the axis Oy. For $a > 1$, this function is increasing, and for $0 < a < 1$, it is a decreasing function; it has no extremal points, and the axis Oy is its vertical asymptote. The common property of the graphs of such functions is that they all pass through the point $(1, 0)$. The graph of the function $y = \log_a x$ is symmetric to that of $y = \log_{1/a} x$ with respect to the x-axis. The modulus of the logarithmic function tends to infinity slower than any power of x as $x \to +\infty$; and it tends to zero slower than any power of $1/x$ as $x \to +0$. The graphs of the functions $y = \log_2 x$ and $y = \log_{1/2} x$ are shown in Fig. 2.4.

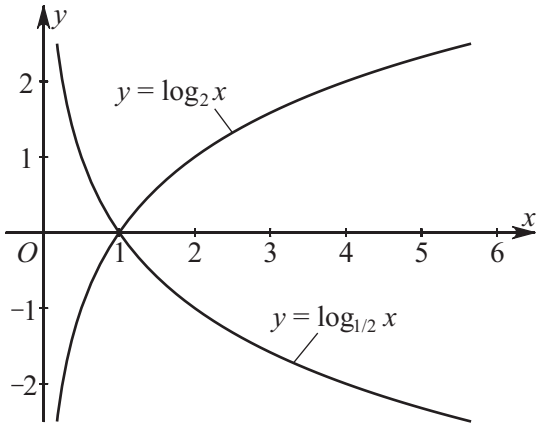


Figure 2.4. Graphs of the logarithmic function.

2.1.3-2. Properties of the logarithmic function.

By definition, the logarithmic function is the inverse of the exponential function. The following equivalence relation holds:

$$y = \log_a x \quad \Longleftrightarrow \quad x = a^y,$$

where $a > 0$, $a \neq 1$.

Basic properties of the logarithmic function:

$$a^{\log_a x} = x, \qquad \log_a(x_1 x_2) = \log_a x_1 + \log_a x_2,$$

$$\log_a(x^k) = k \log_a x, \quad \log_a x = \frac{\log_b x}{\log_b a},$$

where $x > 0$, $x_1 > 0$, $x_2 > 0$, $a > 0$, $a \neq 1$, $b > 0$, $b \neq 1$.

The simplest inequality:

$$\log_a x_1 > \log_a x_2 \quad \Longleftrightarrow \quad \begin{cases} x_1 > x_2 & \text{if } a > 1, \\ x_1 < x_2 & \text{if } 0 < a < 1. \end{cases}$$

For any $b > 0$, the following limit relations hold:

$$\lim_{x \to +\infty} \frac{\log_a x}{x^b} = 0, \qquad \lim_{x \to +0} x^b \log_a x = 0.$$

The logarithmic function with the base e (*base of natural logarithms, Napierian base*) is denoted by

$$\log_e x = \ln x,$$

where $e = \lim\limits_{n \to \infty} \left(1 + \dfrac{1}{n}\right)^n = 2.718281\ldots$

Formulas for passing from an arbitrary base a to the Napierian base e:

$$\log_a x = \frac{\ln x}{\ln a}.$$

Differentiation and integration formulas:

$$(\ln x)' = \frac{1}{x}, \qquad \int \ln x \, dx = x \ln x - x + C.$$

Expansion in power series:

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n-1} \frac{x^n}{n} + \cdots = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{x^k}{k}, \qquad -1 < x < 1.$$

2.2. Trigonometric Functions

2.2.1. Trigonometric Circle. Definition of Trigonometric Functions

2.2.1-1. Trigonometric circle. Degrees and radians.

Trigonometric circle is the circle of unit radius with center at the origin of an orthogonal coordinate system Oxy. The coordinate axes divide the circle into four quarters (*quadrants*); see Fig. 2.5. Consider rotation of the polar radius issuing from the origin O and ending at a point M of the trigonometric circle. Let α be the angle between the x-axis and the polar radius OM measured from the positive direction of the x-axis. This angle is assumed positive in the case of counterclockwise rotation and negative in the case of clockwise rotation.

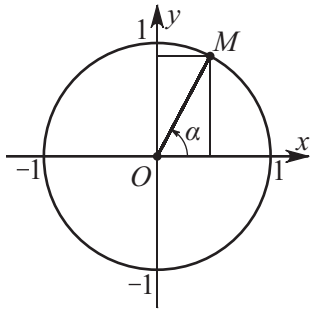


Figure 2.5. Trigonometric circle.

Angles are measured either in radians or in degrees. One radian is the angle at the vertex of the sector of the trigonometric circle supported by its arc of unit length. One degree is the angle at the vertex of the sector of the trigonometric circle supported by its arc of length $\pi/180$. The radians are related to the degrees by the formulas

$$1 \text{ radian} = \frac{180^\circ}{\pi}; \qquad 1^\circ = \frac{\pi}{180}.$$

2.2.1-2. Definition of trigonometric functions.

The *sine* of α is the ordinate (the projection to the axis Oy) of the point on the trigonometric circle corresponding to the angle of α radians. The *cosine* of α is the abscissa (projection to the axis Ox) of that point (see Fig. 2.5). The sine and the cosine are basic trigonometric functions and are denoted, respectively, by $\sin\alpha$ and $\cos\alpha$.

Other trigonometric functions are *tangent*, *cotangent*, *secant*, and *cosecant*. These are derived from the basic trigonometric functions, sine and cosine, as follows:

$$\tan\alpha = \frac{\sin\alpha}{\cos\alpha}, \qquad \cot\alpha = \frac{\cos\alpha}{\sin\alpha}, \qquad \sec\alpha = \frac{1}{\cos\alpha}, \qquad \operatorname{cosec}\alpha = \frac{1}{\sin\alpha}.$$

Table 2.1 gives the signs of the trigonometric functions in different quadrants. The signs and the values of $\sin\alpha$ and $\cos\alpha$ do not change if the argument α is incremented by $\pm 2\pi n$, where $n = 1, 2, \ldots$ The signs and the values of $\tan\alpha$ and $\cot\alpha$ do not change if the argument α is incremented by $\pm\pi n$, where $n = 1, 2, \ldots$

Table 2.2 gives the values of trigonometric functions for some values of their argument (the symbol ∞ means that the function is undefined for the corresponding value of its argument).

TABLE 2.1
Signs of trigonometric functions in different quarters

Quarter	Angle in radians	$\sin\alpha$	$\cos\alpha$	$\tan\alpha$	$\cot\alpha$	$\sec\alpha$	$\operatorname{cosec}\alpha$
I	$0<\alpha<\frac{\pi}{2}$	+	+	+	+	+	+
II	$\frac{\pi}{2}<\alpha<\pi$	+	−	−	−	−	+
III	$\pi<\alpha<\frac{3\pi}{2}$	−	−	+	+	−	−
IV	$\frac{3\pi}{2}<\alpha<2\pi$	−	+	−	−	+	−

TABLE 2.2
Numerical values of trigonometric functions for some angles α (in radians)

Angle α	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{3\pi}{4}$	$\frac{5\pi}{6}$	π
$\sin\alpha$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0
$\cos\alpha$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	$-\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{3}}{2}$	−1
$\tan\alpha$	0	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$	∞	$-\sqrt{3}$	−1	$-\frac{\sqrt{3}}{3}$	0
$\cot\alpha$	∞	$\sqrt{3}$	1	$\frac{\sqrt{3}}{3}$	0	$-\frac{\sqrt{3}}{3}$	−1	$-\sqrt{3}$	∞

2.2.2. Graphs of Trigonometric Functions

2.2.2-1. Sine: $y = \sin x$.

This function is defined for all x and its range is $y \in [-1, 1]$. The sine is an odd, bounded, periodic function (with period 2π). It crosses the axis Oy at the point $y = 0$ and crosses the axis Ox at the points $x = \pi n$, $n = 0, \pm1, \pm2, \ldots$ The sine is an increasing function on every segment $[-\frac{\pi}{2} + 2\pi n, \frac{\pi}{2} + 2\pi n]$ and is a decreasing function on every segment $[\frac{\pi}{2} + 2\pi n, \frac{3}{2}\pi + 2\pi n]$. For $x = \frac{\pi}{2} + 2\pi n$, it attains its maximal value ($y = 1$), and for $x = -\frac{\pi}{2} + 2\pi n$ it attains its minimal value ($y = -1$). The graph of the function $y = \sin x$ is called the *sinusoid* or *sine curve* and is shown in Fig. 2.6.

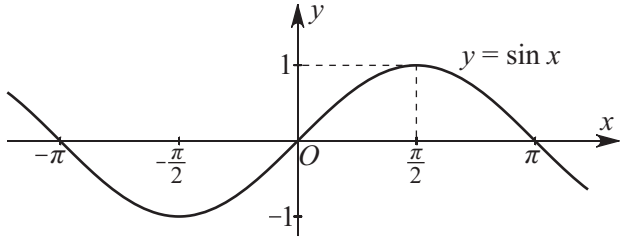


Figure 2.6. The graph of the function $y = \sin x$.

2.2.2-2. Cosine: $y = \cos x$.

This function is defined for all x and its range is $y \in [-1, 1]$. The cosine is a bounded, even, periodic function (with period 2π). It crosses the axis Oy at the point $y = 1$, and

crosses the axis Ox at the points $x = \frac{\pi}{2} + \pi n$. The cosine is an increasing function on every segment $[-\pi + 2\pi n, 2\pi n]$ and is a decreasing function on every segment $[2\pi n, \pi + 2\pi n]$, $n = 0, \pm 1, \pm 2, \ldots$ For $x = 2\pi n$ it attains its maximal value ($y = 1$), and for $x = \pi + 2\pi n$ it attains its minimal value ($y = -1$). The graph of the function $y = \cos x$ is a sinusoid obtained by shifting the graph of the function $y = \sin x$ by $\frac{\pi}{2}$ to the left along the axis Ox (see Fig. 2.7).

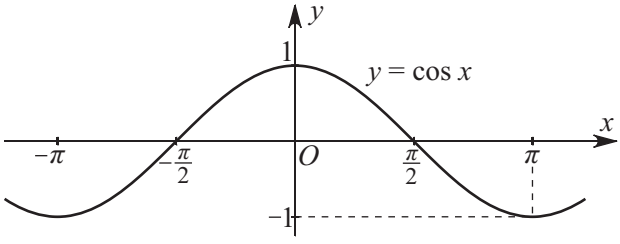


Figure 2.7. The graph of the function $y = \cos x$.

2.2.2-3. Tangent: $y = \tan x$.

This function is defined for all $x \neq \frac{\pi}{2} + \pi n$, $n = 0, \pm 1, \pm 2, \ldots$, and its range is the entire y-axis. The tangent is an unbounded, odd, periodic function (with period π). It crosses the axis Oy at the point $y = 0$ and crosses the axis Ox at the points $x = \pi n$. This is an increasing function on every interval $(-\frac{\pi}{2} + \pi n, \frac{\pi}{2} + \pi n)$. This function has no points of extremum and has vertical asymptotes at $x = \frac{\pi}{2} + \pi n$, $n = 0, \pm 1, \pm 2, \ldots$ The graph of the function $y = \tan x$ is given in Fig. 2.8.

2.2.2-4. Cotangent: $y = \cot x$.

This function is defined for all $x \neq \pi n$, $n = 0, \pm 1, \pm 2, \ldots$, and its range is the entire y-axis. The cotangent is an unbounded, odd, periodic function (with period π). It crosses the axis Ox at the points $x = \frac{\pi}{2} + \pi n$, and does not cross the axis Oy. This is a decreasing function on every interval $(\pi n, \pi + \pi n)$. This function has no extremal points and has vertical asymptotes at $x = \pi n$, $n = 0, \pm 1, \pm 2, \ldots$ The graph of the function $y = \cot x$ is given in Fig. 2.9.

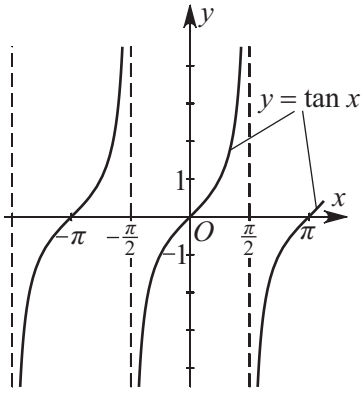


Figure 2.8. The graph of the function $y = \tan x$.

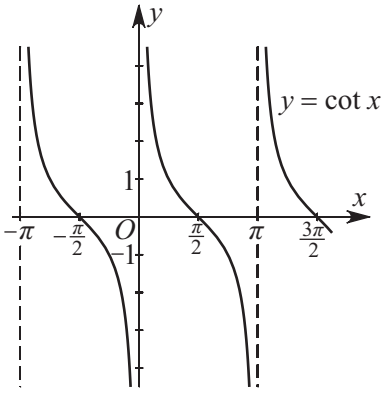


Figure 2.9. The graph of the function $y = \cot x$.

2.2.3. Properties of Trigonometric Functions

2.2.3-1. Simplest relations.

$$\sin^2 x + \cos^2 x = 1, \qquad \tan x \cot x = 1,$$
$$\sin(-x) = -\sin x, \qquad \cos(-x) = \cos x,$$
$$\tan x = \frac{\sin x}{\cos x}, \qquad \cot x = \frac{\cos x}{\sin x},$$
$$\tan(-x) = -\tan x, \qquad \cot(-x) = -\cot x,$$
$$1 + \tan^2 x = \frac{1}{\cos^2 x}, \qquad 1 + \cot^2 x = \frac{1}{\sin^2 x}.$$

2.2.3-2. Reduction formulas.

$$\sin(x \pm 2n\pi) = \sin x, \qquad \cos(x \pm 2n\pi) = \cos x,$$
$$\sin(x \pm n\pi) = (-1)^n \sin x, \qquad \cos(x \pm n\pi) = (-1)^n \cos x,$$
$$\sin\left(x \pm \frac{2n+1}{2}\pi\right) = \pm(-1)^n \cos x, \qquad \cos\left(x \pm \frac{2n+1}{2}\pi\right) = \mp(-1)^n \sin x,$$
$$\sin\left(x \pm \frac{\pi}{4}\right) = \frac{\sqrt{2}}{2}(\sin x \pm \cos x), \qquad \cos\left(x \pm \frac{\pi}{4}\right) = \frac{\sqrt{2}}{2}(\cos x \mp \sin x),$$
$$\tan(x \pm n\pi) = \tan x, \qquad \cot(x \pm n\pi) = \cot x,$$
$$\tan\left(x \pm \frac{2n+1}{2}\pi\right) = -\cot x, \qquad \cot\left(x \pm \frac{2n+1}{2}\pi\right) = -\tan x,$$
$$\tan\left(x \pm \frac{\pi}{4}\right) = \frac{\tan x \pm 1}{1 \mp \tan x}, \qquad \cot\left(x \pm \frac{\pi}{4}\right) = \frac{\cot x \mp 1}{1 \pm \cot x},$$

where $n = 1, 2, \ldots$

2.2.3-3. Relations between trigonometric functions of single argument.

$$\sin x = \pm\sqrt{1 - \cos^2 x} = \pm\frac{\tan x}{\sqrt{1 + \tan^2 x}} = \pm\frac{1}{\sqrt{1 + \cot^2 x}},$$
$$\cos x = \pm\sqrt{1 - \sin^2 x} = \pm\frac{1}{\sqrt{1 + \tan^2 x}} = \pm\frac{\cot x}{\sqrt{1 + \cot^2 x}},$$
$$\tan x = \pm\frac{\sin x}{\sqrt{1 - \sin^2 x}} = \pm\frac{\sqrt{1 - \cos^2 x}}{\cos x} = \frac{1}{\cot x},$$
$$\cot x = \pm\frac{\sqrt{1 - \sin^2 x}}{\sin x} = \pm\frac{\cos x}{\sqrt{1 - \cos^2 x}} = \frac{1}{\tan x}.$$

The sign before the radical is determined by the quarter in which the argument takes its values.

2.2.3-4. Addition and subtraction of trigonometric functions.

$$\sin x + \sin y = 2\sin\Big(\frac{x+y}{2}\Big)\cos\Big(\frac{x-y}{2}\Big),$$
$$\sin x - \sin y = 2\sin\Big(\frac{x-y}{2}\Big)\cos\Big(\frac{x+y}{2}\Big),$$
$$\cos x + \cos y = 2\cos\Big(\frac{x+y}{2}\Big)\cos\Big(\frac{x-y}{2}\Big),$$
$$\cos x - \cos y = -2\sin\Big(\frac{x+y}{2}\Big)\sin\Big(\frac{x-y}{2}\Big),$$
$$\sin^2 x - \sin^2 y = \cos^2 y - \cos^2 x = \sin(x+y)\sin(x-y),$$
$$\sin^2 x - \cos^2 y = -\cos(x+y)\cos(x-y),$$
$$\tan x \pm \tan y = \frac{\sin(x \pm y)}{\cos x \cos y}, \quad \cot x \pm \cot y = \frac{\sin(y \pm x)}{\sin x \sin y},$$
$$a\cos x + b\sin x = r\sin(x+\varphi) = r\cos(x-\psi).$$

Here, $r = \sqrt{a^2+b^2}$, $\sin\varphi = a/r$, $\cos\varphi = b/r$, $\sin\psi = b/r$, and $\cos\psi = a/r$.

2.2.3-5. Products of trigonometric functions.

$$\sin x \sin y = \tfrac{1}{2}[\cos(x-y) - \cos(x+y)],$$
$$\cos x \cos y = \tfrac{1}{2}[\cos(x-y) + \cos(x+y)],$$
$$\sin x \cos y = \tfrac{1}{2}[\sin(x-y) + \sin(x+y)].$$

2.2.3-6. Powers of trigonometric functions.

$$\cos^2 x = \tfrac{1}{2}\cos 2x + \tfrac{1}{2}, \qquad \sin^2 x = -\tfrac{1}{2}\cos 2x + \tfrac{1}{2},$$
$$\cos^3 x = \tfrac{1}{4}\cos 3x + \tfrac{3}{4}\cos x, \qquad \sin^3 x = -\tfrac{1}{4}\sin 3x + \tfrac{3}{4}\sin x,$$
$$\cos^4 x = \tfrac{1}{8}\cos 4x + \tfrac{1}{2}\cos 2x + \tfrac{3}{8}, \qquad \sin^4 x = \tfrac{1}{8}\cos 4x - \tfrac{1}{2}\cos 2x + \tfrac{3}{8},$$
$$\cos^5 x = \tfrac{1}{16}\cos 5x + \tfrac{5}{16}\cos 3x + \tfrac{5}{8}\cos x, \qquad \sin^5 x = \tfrac{1}{16}\sin 5x - \tfrac{5}{16}\sin 3x + \tfrac{5}{8}\sin x,$$

$$\cos^{2n} x = \frac{1}{2^{2n-1}}\sum_{k=0}^{n-1} C_{2n}^k \cos[2(n-k)x] + \frac{1}{2^{2n}}C_{2n}^n,$$

$$\cos^{2n+1} x = \frac{1}{2^{2n}}\sum_{k=0}^{n} C_{2n+1}^k \cos[(2n-2k+1)x],$$

$$\sin^{2n} x = \frac{1}{2^{2n-1}}\sum_{k=0}^{n-1}(-1)^{n-k} C_{2n}^k \cos[2(n-k)x] + \frac{1}{2^{2n}}C_{2n}^n,$$

$$\sin^{2n+1} x = \frac{1}{2^{2n}}\sum_{k=0}^{n}(-1)^{n-k} C_{2n+1}^k \sin[(2n-2k+1)x].$$

Here, $n = 1, 2, \ldots$ and $C_m^k = \dfrac{m!}{k!\,(m-k)!}$ are binomial coefficients $(0! = 1)$.

2.2.3-7. Addition formulas.

$$\sin(x \pm y) = \sin x \cos y \pm \cos x \sin y, \qquad \cos(x \pm y) = \cos x \cos y \mp \sin x \sin y,$$
$$\tan(x \pm y) = \frac{\tan x \pm \tan y}{1 \mp \tan x \tan y}, \qquad \cot(x \pm y) = \frac{1 \mp \tan x \tan y}{\tan x \pm \tan y}.$$

2.2.3-8. Trigonometric functions of multiple arguments.

$$\cos 2x = 2\cos^2 x - 1 = 1 - 2\sin^2 x, \qquad \sin 2x = 2 \sin x \cos x,$$
$$\cos 3x = -3\cos x + 4\cos^3 x, \qquad \sin 3x = 3\sin x - 4\sin^3 x,$$
$$\cos 4x = 1 - 8\cos^2 x + 8\cos^4 x, \qquad \sin 4x = 4\cos x\,(\sin x - 2\sin^3 x),$$
$$\cos 5x = 5\cos x - 20\cos^3 x + 16\cos^5 x, \qquad \sin 5x = 5\sin x - 20\sin^3 x + 16\sin^5 x,$$
$$\cos(2nx) = 1 + \sum_{k=1}^{n} (-1)^k \frac{n^2(n^2-1)\ldots[n^2-(k-1)^2]}{(2k)!} 4^k \sin^{2k} x,$$
$$\cos[(2n+1)x] = \cos x \left\{ 1 + \sum_{k=1}^{n} (-1)^k \frac{[(2n+1)^2-1][(2n+1)^2-3^2]\ldots[(2n+1)^2-(2k-1)^2]}{(2k)!} \sin^{2k} x \right\},$$
$$\sin(2nx) = 2n\cos x \left[\sin x + \sum_{k=1}^{n} (-4)^k \frac{(n^2-1)(n^2-2^2)\ldots(n^2-k^2)}{(2k-1)!} \sin^{2k-1} x \right],$$
$$\sin[(2n+1)x] = (2n+1) \left\{ \sin x + \sum_{k=1}^{n} (-1)^k \frac{[(2n+1)^2-1][(2n+1)^2-3^2]\ldots[(2n+1)^2-(2k-1)^2]}{(2k+1)!} \sin^{2k+1} x \right\},$$
$$\tan 2x = \frac{2\tan x}{1-\tan^2 x}, \qquad \tan 3x = \frac{3\tan x - \tan^3 x}{1 - 3\tan^2 x}, \qquad \tan 4x = \frac{4\tan x - 4\tan^3 x}{1 - 6\tan^2 x + \tan^4 x},$$

where $n = 1, 2, \ldots$

2.2.3-9. Trigonometric functions of half argument.

$$\sin^2 \frac{x}{2} = \frac{1-\cos x}{2}, \qquad \cos^2 \frac{x}{2} = \frac{1+\cos x}{2},$$
$$\tan \frac{x}{2} = \frac{\sin x}{1+\cos x} = \frac{1-\cos x}{\sin x}, \qquad \cot \frac{x}{2} = \frac{\sin x}{1-\cos x} = \frac{1+\cos x}{\sin x},$$
$$\sin x = \frac{2\tan \frac{x}{2}}{1+\tan^2 \frac{x}{2}}, \qquad \cos x = \frac{1-\tan^2 \frac{x}{2}}{1+\tan^2 \frac{x}{2}}, \qquad \tan x = \frac{2\tan \frac{x}{2}}{1-\tan^2 \frac{x}{2}}.$$

2.2.3-10. Differentiation formulas.

$$\frac{d\sin x}{dx} = \cos x, \qquad \frac{d\cos x}{dx} = -\sin x, \qquad \frac{d\tan x}{dx} = \frac{1}{\cos^2 x}, \qquad \frac{d\cot x}{dx} = -\frac{1}{\sin^2 x}.$$

2.2.3-11. Integration formulas.

$$\int \sin x\,dx = -\cos x + C, \qquad \int \cos x\,dx = \sin x + C,$$
$$\int \tan x\,dx = -\ln|\cos x| + C, \qquad \int \cot x\,dx = \ln|\sin x| + C,$$

where C is an arbitrary constant.

2.2.3-12. Expansion in power series.

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + \cdots \qquad (|x| < \infty),$$
$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + \cdots \qquad (|x| < \infty),$$
$$\tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} + \frac{17x^7}{315} + \cdots + \frac{2^{2n}(2^{2n}-1)|B_{2n}|}{(2n)!}x^{2n-1} + \cdots \qquad (|x| < \pi/2),$$
$$\cot x = \frac{1}{x} - \left(\frac{x}{3} + \frac{x^3}{45} + \frac{2x^5}{945} + \cdots + \frac{2^{2n}|B_{2n}|}{(2n)!}x^{2n-1} + \cdots\right) \qquad (0 < |x| < \pi),$$

where B_n are Bernoulli numbers (see Subsection 18.1.3).

2.2.3-13. Representation in the form of infinite products.

$$\sin x = x\left(1 - \frac{x^2}{\pi^2}\right)\left(1 - \frac{x^2}{4\pi^2}\right)\left(1 - \frac{x^2}{9\pi^2}\right)\ldots\left(1 - \frac{x^2}{n^2\pi^2}\right)\ldots$$
$$\cos x = \left(1 - \frac{4x^2}{\pi^2}\right)\left(1 - \frac{4x^2}{9\pi^2}\right)\left(1 - \frac{4x^2}{25\pi^2}\right)\ldots\left(1 - \frac{4x^2}{(2n+1)^2\pi^2}\right)\ldots$$

2.2.3-14. Euler and de Moivre formulas. Relationship with hyperbolic functions.

$$e^{y+ix} = e^y(\cos x + i\sin x), \quad (\cos x + i\sin x)^n = \cos(nx) + i\sin(nx), \quad i^2 = -1,$$
$$\sin(ix) = i\sinh x, \quad \cos(ix) = \cosh x, \quad \tan(ix) = i\tanh x, \quad \cot(ix) = -i\coth x.$$

2.3. Inverse Trigonometric Functions

2.3.1. Definitions. Graphs of Inverse Trigonometric Functions

2.3.1-1. Definitions of inverse trigonometric functions.

Inverse trigonometric functions (*arc functions*) are the functions that are inverse to the trigonometric functions. Since the trigonometric functions $\sin x$, $\cos x$, $\tan x$, $\cot x$ are periodic, the corresponding inverse functions, denoted by $\operatorname{Arcsin} x$, $\operatorname{Arccos} x$, $\operatorname{Arctan} x$,

Arccot x, are multi-valued. The following relations define the multi-valued inverse trigonometric functions:

$$\sin\bigl(\operatorname{Arcsin} x\bigr) = x, \quad \cos\bigl(\operatorname{Arccos} x\bigr) = x,$$
$$\tan\bigl(\operatorname{Arctan} x\bigr) = x, \quad \cot\bigl(\operatorname{Arccot} x\bigr) = x.$$

These functions admit the following verbal definitions: Arcsin x is the angle whose sine is equal to x; Arccos x is the angle whose cosine is equal to x; Arctan x is the angle whose tangent is equal to x; Arccot x is the angle whose cotangent is equal to x.

The principal (single-valued) branches of the inverse trigonometric functions are denoted by

$\arcsin x \equiv \sin^{-1} x$ (arcsine is the inverse of sine),

$\arccos x \equiv \cos^{-1} x$ (arccosine is the inverse of cosine),

$\arctan x \equiv \tan^{-1} x$ (arctangent is the inverse of tangent),

$\operatorname{arccot} x \equiv \cot^{-1} x$ (arccotangent is the inverse of cotangent)

and are determined by the inequalities

$$-\tfrac{\pi}{2} \le \arcsin x \le \tfrac{\pi}{2}, \quad 0 \le \arccos x \le \pi \qquad (-1 \le x \le 1);$$
$$-\tfrac{\pi}{2} < \arctan x < \tfrac{\pi}{2}, \quad 0 < \operatorname{arccot} x < \pi \qquad (-\infty < x < \infty).$$

The following equivalent relations can be taken as definitions of single-valued inverse trigonometric functions:

$$y = \arcsin x, \quad -1 \le x \le 1 \quad \Longleftrightarrow \quad x = \sin y, \quad -\frac{\pi}{2} \le y \le \frac{\pi}{2};$$
$$y = \arccos x, \quad -1 \le x \le 1 \quad \Longleftrightarrow \quad x = \cos y, \quad 0 \le y \le \pi;$$
$$y = \arctan x, \quad -\infty < x < +\infty \quad \Longleftrightarrow \quad x = \tan y, \quad -\frac{\pi}{2} < y < \frac{\pi}{2};$$
$$y = \operatorname{arccot} x, \quad -\infty < x < +\infty \quad \Longleftrightarrow \quad x = \cot y, \quad 0 < y < \pi.$$

The multi-valued and the single-valued inverse trigonometric functions are related by the formulas

$$\operatorname{Arcsin} x = (-1)^n \arcsin x + \pi n,$$
$$\operatorname{Arccos} x = \pm \arccos x + 2\pi n,$$
$$\operatorname{Arctan} x = \arctan x + \pi n,$$
$$\operatorname{Arccot} x = \operatorname{arccot} x + \pi n,$$

where $n = 0, \pm 1, \pm 2, \ldots$

The graphs of inverse trigonometric functions are obtained from the graphs of the corresponding trigonometric functions by mirror reflection with respect to the straight line $y = x$ (with the domain of each function being taken into account).

2.3.1-2. Arcsine: $y = \arcsin x$.

This function is defined for all $x \in [-1, 1]$ and its range is $y \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. The arcsine is an odd, nonperiodic, bounded function that crosses the axes Ox and Oy at the origin $x = 0$, $y = 0$. This is an increasing function in its domain, and it takes its smallest value $y = -\frac{\pi}{2}$ at the point $x = -1$; it takes its largest value $y = \frac{\pi}{2}$ at the point $x = 1$. The graph of the function $y = \arcsin x$ is given in Fig. 2.10.

2.3.1-3. Arccosine: $y = \arccos x$.

This function is defined for all $x \in [-1, 1]$, and its range is $y \in [0, \pi]$. It is neither odd nor even. It is a nonperiodic, bounded function that crosses the axis Oy at the point $y = \frac{\pi}{2}$ and crosses the axis Ox at the point $x = 1$. This is a decreasing function in its domain, and at the point $x = -1$ it takes its largest value $y = \pi$; at the point $x = 1$ it takes its smallest value $y = 0$. For all x in its domain, the following relation holds: $\arccos x = \frac{\pi}{2} - \arcsin x$. The graph of the function $y = \arccos x$ is given in Fig. 2.11.

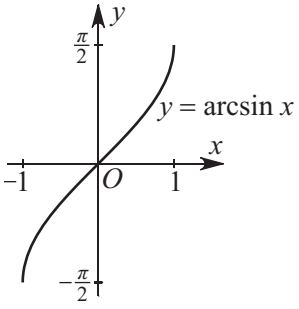


Figure 2.10. The graph of the function $y = \arcsin x$.

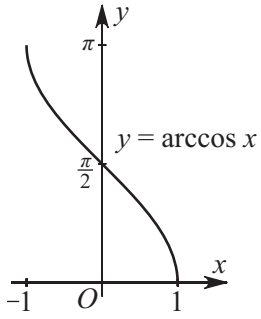


Figure 2.11. The graph of the function $y = \arccos x$.

2.3.1-4. Arctangent: $y = \arctan x$.

This function is defined for all x, and its range is $y \in (-\frac{\pi}{2}, \frac{\pi}{2})$. The arctangent is an odd, nonperiodic, bounded function that crosses the coordinate axes at the origin $x = 0$, $y = 0$. This is an increasing function on the real axis with no points of extremum. It has two horizontal asymptotes: $y = -\frac{\pi}{2}$ (as $x \to -\infty$) and $y = \frac{\pi}{2}$ (as $x \to +\infty$). The graph of the function $y = \arctan x$ is given in Fig. 2.12.

2.3.1-5. Arccotangent: $y = \operatorname{arccot} x$.

This function is defined for all x, and its range is $y \in (0, \pi)$. The arccotangent is neither odd nor even. It is a nonperiodic, bounded function that crosses the axis Oy at the point $y = \frac{\pi}{2}$ and does not cross the axis Ox. This is a decreasing function on the entire real axis with no points of extremum. It has two horizontal asymptotes $y = 0$ (as $x \to +\infty$) and $y = \pi$ (as $x \to -\infty$). For all x, the following relation holds: $\operatorname{arccot} x = \frac{\pi}{2} - \arctan x$. The graph of the function $y = \operatorname{arccot} x$ is given in Fig. 2.13.

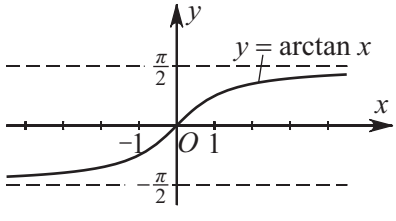


Figure 2.12. The graph of the function $y = \arctan x$.

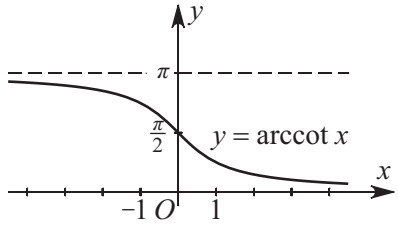


Figure 2.13. The graph of the function $y = \operatorname{arccot} x$.

2.3.2. Properties of Inverse Trigonometric Functions

2.3.2-1. Simplest formulas.

$$\sin(\arcsin x) = x, \quad \cos(\arccos x) = x,$$
$$\tan(\arctan x) = x, \quad \cot(\operatorname{arccot} x) = x.$$

2.3.2-2. Some properties.

$$\arcsin(-x) = -\arcsin x, \qquad \arccos(-x) = \pi - \arccos x,$$
$$\arctan(-x) = -\arctan x, \qquad \operatorname{arccot}(-x) = \pi - \operatorname{arccot} x,$$

$$\arcsin(\sin x) = \begin{cases} x - 2n\pi & \text{if } 2n\pi - \frac{\pi}{2} \le x \le 2n\pi + \frac{\pi}{2}, \\ -x + 2(n+1)\pi & \text{if } (2n+1)\pi - \frac{\pi}{2} \le x \le 2(n+1)\pi + \frac{\pi}{2}, \end{cases}$$

$$\arccos(\cos x) = \begin{cases} x - 2n\pi & \text{if } 2n\pi \le x \le (2n+1)\pi, \\ -x + 2(n+1)\pi & \text{if } (2n+1)\pi \le x \le 2(n+1)\pi, \end{cases}$$

$$\arctan(\tan x) = x - n\pi \quad \text{if} \quad n\pi - \tfrac{\pi}{2} < x < n\pi + \tfrac{\pi}{2},$$
$$\operatorname{arccot}(\cot x) = x - n\pi \quad \text{if} \quad n\pi < x < (n+1)\pi.$$

2.3.2-3. Relations between inverse trigonometric functions.

$$\arcsin x + \arccos x = \tfrac{\pi}{2}, \qquad \arctan x + \operatorname{arccot} x = \tfrac{\pi}{2};$$

$$\arcsin x = \begin{cases} \arccos\sqrt{1-x^2} & \text{if } 0 \le x \le 1, \\ -\arccos\sqrt{1-x^2} & \text{if } -1 \le x \le 0, \\ \arctan\dfrac{x}{\sqrt{1-x^2}} & \text{if } -1 < x < 1, \\ \operatorname{arccot}\dfrac{\sqrt{1-x^2}}{x} - \pi & \text{if } -1 \le x < 0; \end{cases} \qquad \arccos x = \begin{cases} \arcsin\sqrt{1-x^2} & \text{if } 0 \le x \le 1, \\ \pi - \arcsin\sqrt{1-x^2} & \text{if } -1 \le x \le 0, \\ \arctan\dfrac{\sqrt{1-x^2}}{x} & \text{if } 0 < x \le 1, \\ \operatorname{arccot}\dfrac{x}{\sqrt{1-x^2}} & \text{if } -1 < x < 1; \end{cases}$$

$$\arctan x = \begin{cases} \arcsin\dfrac{x}{\sqrt{1+x^2}} & \text{for any } x, \\ \arccos\dfrac{1}{\sqrt{1+x^2}} & \text{if } x \ge 0, \\ -\arccos\dfrac{1}{\sqrt{1+x^2}} & \text{if } x \le 0, \\ \operatorname{arccot}\dfrac{1}{x} & \text{if } x > 0; \end{cases} \qquad \operatorname{arccot} x = \begin{cases} \arcsin\dfrac{1}{\sqrt{1+x^2}} & \text{if } x > 0, \\ \pi - \arcsin\dfrac{1}{\sqrt{1+x^2}} & \text{if } x < 0, \\ \arctan\dfrac{1}{x} & \text{if } x > 0, \\ \pi + \arctan\dfrac{1}{x} & \text{if } x < 0. \end{cases}$$

2.3.2-4. Addition and subtraction of inverse trigonometric functions.

$$\arcsin x + \arcsin y = \arcsin\left(x\sqrt{1-y^2} + y\sqrt{1-x^2}\right) \qquad \text{for} \quad x^2 + y^2 \le 1,$$
$$\arccos x \pm \arccos y = \pm\arccos\left[xy \mp \sqrt{(1-x^2)(1-y^2)}\right] \quad \text{for} \quad x \pm y \ge 0,$$

$$\arctan x + \arctan y = \arctan\frac{x+y}{1-xy} \quad \text{for} \quad xy < 1,$$
$$\arctan x - \arctan y = \arctan\frac{x-y}{1+xy} \quad \text{for} \quad xy > -1.$$

2.3.2-5. Differentiation formulas.

$$\frac{d}{dx}\arcsin x = \frac{1}{\sqrt{1-x^2}}, \qquad \frac{d}{dx}\arccos x = -\frac{1}{\sqrt{1-x^2}},$$
$$\frac{d}{dx}\arctan x = \frac{1}{1+x^2}, \qquad \frac{d}{dx}\operatorname{arccot} x = -\frac{1}{1+x^2}.$$

2.3.2-6. Integration formulas.

$$\int \arcsin x\,dx = x\arcsin x + \sqrt{1-x^2} + C, \qquad \int \arccos x\,dx = x\arccos x - \sqrt{1-x^2} + C,$$
$$\int \arctan x\,dx = x\arctan x - \frac{1}{2}\ln(1+x^2) + C, \qquad \int \operatorname{arccot} x\,dx = x\operatorname{arccot} x + \frac{1}{2}\ln(1+x^2) + C,$$

where C is an arbitrary constant.

2.3.2-7. Expansion in power series.

$$\arcsin x = x + \frac{1}{2}\frac{x^3}{3} + \frac{1\times3}{2\times4}\frac{x^5}{5} + \frac{1\times3\times5}{2\times4\times6}\frac{x^7}{7} + \cdots + \frac{1\times3\times\cdots\times(2n-1)}{2\times4\times\cdots\times(2n)}\frac{x^{2n+1}}{2n+1} + \cdots \quad (|x| < 1),$$
$$\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots + (-1)^{n-1}\frac{x^{2n-1}}{2n-1} + \cdots \quad (|x| \le 1),$$
$$\arctan x = \frac{\pi}{2} - \frac{1}{x} + \frac{1}{3x^3} - \frac{1}{5x^5} + \cdots + (-1)^n\frac{1}{(2n-1)x^{2n-1}} + \cdots \quad (|x| > 1).$$

The expansions for $\arccos x$ and $\operatorname{arccot} x$ can be obtained from the relations $\arccos x = \frac{\pi}{2} - \arcsin x$ and $\operatorname{arccot} x = \frac{\pi}{2} - \arctan x$.

2.4. Hyperbolic Functions

2.4.1. Definitions. Graphs of Hyperbolic Functions

2.4.1-1. Definitions of hyperbolic functions.

Hyperbolic functions are defined in terms of the exponential functions as follows:

$$\sinh x = \frac{e^x - e^{-x}}{2}, \quad \cosh x = \frac{e^x + e^{-x}}{2}, \quad \tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad \coth x = \frac{e^x + e^{-x}}{e^x - e^{-x}}.$$

The graphs of hyperbolic functions are given below.

2.4.1-2. Hyperbolic sine: $y = \sinh x$.

This function is defined for all x and its range is the entire y-axis. The hyperbolic sine is an odd, nonperiodic, unbounded function that crosses the axes Ox and Oy at the origin $x = 0$, $y = 0$. This is an increasing function in its domain with no points of extremum. The graph of the function $y = \sinh x$ is given in Fig. 2.14.

2.4.1-3. Hyperbolic cosine: $y = \cosh x$.

This function is defined for all x, and its range is $y \in [1, +\infty)$. The hyperbolic cosine is a nonperiodic, unbounded function that crosses the axis Oy at the point 1 and does not cross the axis Ox. This function is decreasing on the interval $(-\infty, 0)$ and is increasing on the interval $(0, +\infty)$; it takes its smallest value $y = 1$ at $x = 0$. The graph of the function $y = \cosh x$ is given in Fig. 2.15.

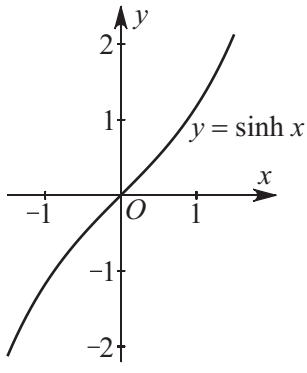


Figure 2.14. The graph of the function $y = \sinh x$.

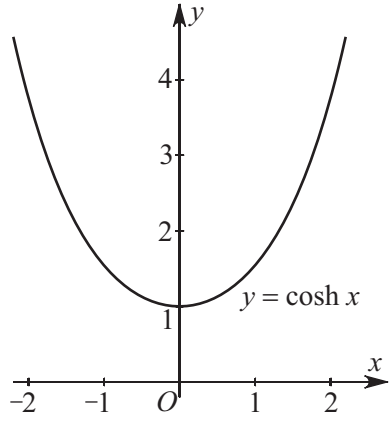


Figure 2.15. The graph of the function $y = \cosh x$.

2.4.1-4. Hyperbolic tangent: $y = \tanh x$.

This function is defined for all x, and its range is $y \in (-1, 1)$. The hyperbolic tangent is an odd, nonperiodic, bounded function that crosses the coordinate axes at the origin $x = 0$, $y = 0$. This is an increasing function on the entire real axis and has two horizontal asymptotes: $y = -1$ (as $x \to -\infty$) and $y = 1$ (as $x \to +\infty$). The graph of the function $y = \tanh x$ is given in Fig. 2.16.

2.4.1-5. Hyperbolic cotangent: $y = \coth x$.

This function is defined for all $x \neq 0$, and its range consists of all $y \in (-\infty, -1)$ and $y \in (1, +\infty)$. The hyperbolic cotangent is an odd, nonperiodic, unbounded function that does not cross the coordinate axes. This is a decreasing function on each of the semiaxes of its domain; it has no points of extremum and does not cross the coordinate axes. It has two horizontal asymptotes: $y = -1$ (as $x \to -\infty$) and $y = 1$ (as $x \to +\infty$). The graph of the function $y = \coth x$ is given in Fig. 2.17.

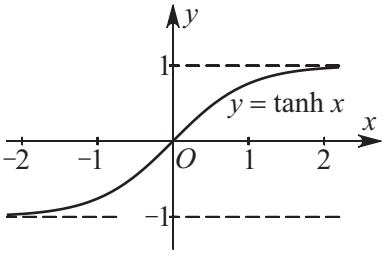


Figure 2.16. The graph of the function $y = \tanh x$.

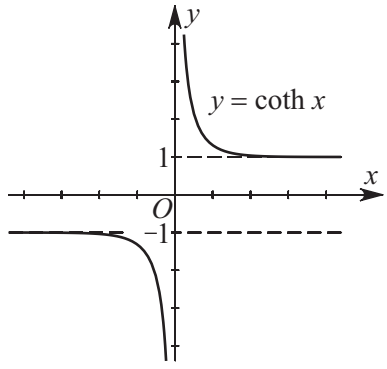


Figure 2.17. The graph of the function $y = \coth x$.

2.4.2. Properties of Hyperbolic Functions

2.4.2-1. Simplest relations.

$$\cosh^2 x - \sinh^2 x = 1, \qquad \tanh x \coth x = 1,$$
$$\sinh(-x) = -\sinh x, \qquad \cosh(-x) = \cosh x,$$
$$\tanh x = \frac{\sinh x}{\cosh x}, \qquad \coth x = \frac{\cosh x}{\sinh x},$$
$$\tanh(-x) = -\tanh x, \qquad \coth(-x) = -\coth x,$$
$$1 - \tanh^2 x = \frac{1}{\cosh^2 x}, \qquad \coth^2 x - 1 = \frac{1}{\sinh^2 x}.$$

2.4.2-2. Relations between hyperbolic functions of single argument $(x \geq 0)$.

$$\sinh x = \sqrt{\cosh^2 x - 1} = \frac{\tanh x}{\sqrt{1 - \tanh^2 x}} = \frac{1}{\sqrt{\coth^2 x - 1}},$$
$$\cosh x = \sqrt{\sinh^2 x + 1} = \frac{1}{\sqrt{1 - \tanh^2 x}} = \frac{\coth x}{\sqrt{\coth^2 x - 1}},$$
$$\tanh x = \frac{\sinh x}{\sqrt{\sinh^2 x + 1}} = \frac{\sqrt{\cosh^2 x - 1}}{\cosh x} = \frac{1}{\coth x},$$
$$\coth x = \frac{\sqrt{\sinh^2 x + 1}}{\sinh x} = \frac{\cosh x}{\sqrt{\cosh^2 x - 1}} = \frac{1}{\tanh x}.$$

2.4.2-3. Addition formulas.

$$\sinh(x \pm y) = \sinh x \cosh y \pm \sinh y \cosh x, \qquad \cosh(x \pm y) = \cosh x \cosh y \pm \sinh x \sinh y,$$
$$\tanh(x \pm y) = \frac{\tanh x \pm \tanh y}{1 \pm \tanh x \tanh y}, \qquad \coth(x \pm y) = \frac{\coth x \coth y \pm 1}{\coth y \pm \coth x}.$$

2.4.2-4. Addition and subtraction of hyperbolic functions.

$$\sinh x \pm \sinh y = 2\sinh\Big(\frac{x \pm y}{2}\Big)\cosh\Big(\frac{x \mp y}{2}\Big),$$
$$\cosh x + \cosh y = 2\cosh\Big(\frac{x + y}{2}\Big)\cosh\Big(\frac{x - y}{2}\Big),$$
$$\cosh x - \cosh y = 2\sinh\Big(\frac{x + y}{2}\Big)\sinh\Big(\frac{x - y}{2}\Big),$$
$$\sinh^2 x - \sinh^2 y = \cosh^2 x - \cosh^2 y = \sinh(x + y)\sinh(x - y),$$
$$\sinh^2 x + \cosh^2 y = \cosh(x + y)\cosh(x - y),$$
$$(\cosh x \pm \sinh x)^n = \cosh(nx) \pm \sinh(nx),$$
$$\tanh x \pm \tanh y = \frac{\sinh(x \pm y)}{\cosh x \cosh y}, \quad \coth x \pm \coth y = \pm\frac{\sinh(x \pm y)}{\sinh x \sinh y},$$

where $n = 0, \pm1, \pm2, \ldots$

2.4.2-5. Products of hyperbolic functions.

$$\sinh x \sinh y = \tfrac{1}{2}[\cosh(x + y) - \cosh(x - y)],$$
$$\cosh x \cosh y = \tfrac{1}{2}[\cosh(x + y) + \cosh(x - y)],$$
$$\sinh x \cosh y = \tfrac{1}{2}[\sinh(x + y) + \sinh(x - y)].$$

2.4.2-6. Powers of hyperbolic functions.

$$\cosh^2 x = \tfrac{1}{2}\cosh 2x + \tfrac{1}{2}, \qquad \sinh^2 x = \tfrac{1}{2}\cosh 2x - \tfrac{1}{2},$$
$$\cosh^3 x = \tfrac{1}{4}\cosh 3x + \tfrac{3}{4}\cosh x, \qquad \sinh^3 x = \tfrac{1}{4}\sinh 3x - \tfrac{3}{4}\sinh x,$$
$$\cosh^4 x = \tfrac{1}{8}\cosh 4x + \tfrac{1}{2}\cosh 2x + \tfrac{3}{8}, \qquad \sinh^4 x = \tfrac{1}{8}\cosh 4x - \tfrac{1}{2}\cosh 2x + \tfrac{3}{8},$$
$$\cosh^5 x = \tfrac{1}{16}\cosh 5x + \tfrac{5}{16}\cosh 3x + \tfrac{5}{8}\cosh x, \qquad \sinh^5 x = \tfrac{1}{16}\sinh 5x - \tfrac{5}{16}\sinh 3x + \tfrac{5}{8}\sinh x,$$

$$\cosh^{2n} x = \frac{1}{2^{2n-1}}\sum_{k=0}^{n-1} C_{2n}^k \cosh[2(n-k)x] + \frac{1}{2^{2n}}C_{2n}^n,$$

$$\cosh^{2n+1} x = \frac{1}{2^{2n}}\sum_{k=0}^{n} C_{2n+1}^k \cosh[(2n-2k+1)x],$$

$$\sinh^{2n} x = \frac{1}{2^{2n-1}}\sum_{k=0}^{n-1}(-1)^k C_{2n}^k \cosh[2(n-k)x] + \frac{(-1)^n}{2^{2n}}C_{2n}^n,$$

$$\sinh^{2n+1} x = \frac{1}{2^{2n}}\sum_{k=0}^{n}(-1)^k C_{2n+1}^k \sinh[(2n-2k+1)x].$$

Here, $n = 1, 2, \ldots$; and C_m^k are binomial coefficients.

2.4.2-7. Hyperbolic functions of multiple argument.

$$\cosh 2x = 2\cosh^2 x - 1, \qquad \sinh 2x = 2\sinh x \cosh x,$$

$$\cosh 3x = -3\cosh x + 4\cosh^3 x, \qquad \sinh 3x = 3\sinh x + 4\sinh^3 x,$$

$$\cosh 4x = 1 - 8\cosh^2 x + 8\cosh^4 x, \qquad \sinh 4x = 4\cosh x(\sinh x + 2\sinh^3 x),$$

$$\cosh 5x = 5\cosh x - 20\cosh^3 x + 16\cosh^5 x, \qquad \sinh 5x = 5\sinh x + 20\sinh^3 x + 16\sinh^5 x.$$

$$\cosh(nx) = 2^{n-1}\cosh^n x + \frac{n}{2}\sum_{k=0}^{[n/2]} \frac{(-1)^{k+1}}{k+1} C_{n-k-2}^{k-2} 2^{n-2k-2} (\cosh x)^{n-2k-2},$$

$$\sinh(nx) = \sinh x \sum_{k=0}^{[(n-1)/2]} 2^{n-k-1} C_{n-k-1}^{k} (\cosh x)^{n-2k-1}.$$

Here, C_m^k are binomial coefficients and $[A]$ stands for the integer part of the number A.

2.4.2-8. Hyperbolic functions of half argument.

$$\sinh\frac{x}{2} = \operatorname{sign} x\sqrt{\frac{\cosh x - 1}{2}}, \qquad \cosh\frac{x}{2} = \sqrt{\frac{\cosh x + 1}{2}},$$

$$\tanh\frac{x}{2} = \frac{\sinh x}{\cosh x + 1} = \frac{\cosh x - 1}{\sinh x}, \qquad \coth\frac{x}{2} = \frac{\sinh x}{\cosh x - 1} = \frac{\cosh x + 1}{\sinh x}.$$

2.4.2-9. Differentiation formulas.

$$\frac{d\sinh x}{dx} = \cosh x, \qquad \frac{d\cosh x}{dx} = \sinh x,$$

$$\frac{d\tanh x}{dx} = \frac{1}{\cosh^2 x}, \qquad \frac{d\coth x}{dx} = -\frac{1}{\sinh^2 x}.$$

2.4.2-10. Integration formulas.

$$\int \sinh x\,dx = \cosh x + C, \qquad \int \cosh x\,dx = \sinh x + C,$$

$$\int \tanh x\,dx = \ln\cosh x + C, \qquad \int \coth x\,dx = \ln|\sinh x| + C,$$

where C is an arbitrary constant.

2.4.2-11. Expansion in power series.

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \cdots + \frac{x^{2n}}{(2n)!} + \cdots \qquad (|x| < \infty),$$

$$\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} + \cdots + \frac{x^{2n+1}}{(2n+1)!} + \cdots \qquad (|x| < \infty),$$

$$\tanh x = x - \frac{x^3}{3} + \frac{2x^5}{15} - \frac{17x^7}{315} + \cdots + (-1)^{n-1}\frac{2^{2n}(2^{2n}-1)|B_{2n}|x^{2n-1}}{(2n)!} + \cdots \qquad (|x| < \pi/2),$$

$$\coth x = \frac{1}{x} + \frac{x}{3} - \frac{x^3}{45} + \frac{2x^5}{945} - \cdots + (-1)^{n-1}\frac{2^{2n}|B_{2n}|x^{2n-1}}{(2n)!} + \cdots \qquad (|x| < \pi),$$

where B_n are Bernoulli numbers (see Subsection 18.1.3).

2.4.2-12. Relationship with trigonometric functions.

$$\sinh(ix) = i\sin x, \quad \cosh(ix) = \cos x, \quad \tanh(ix) = i\tan x, \quad \coth(ix) = -i\cot x, \quad i^2 = -1.$$

2.5. Inverse Hyperbolic Functions

2.5.1. Definitions. Graphs of Inverse Hyperbolic Functions

2.5.1-1. Definitions of inverse hyperbolic functions.

Inverse hyperbolic functions are the functions that are inverse to hyperbolic functions. The following notation is used for inverse hyperbolic functions:

$\operatorname{arcsinh} x \equiv \sinh^{-1} x$ (inverse of hyperbolic sine),

$\operatorname{arccosh} x \equiv \cosh^{-1} x$ (inverse of hyperbolic cosine),

$\operatorname{arctanh} x \equiv \tanh^{-1} x$ (inverse of hyperbolic tangent),

$\operatorname{arccoth} x \equiv \coth^{-1} x$ (inverse of hyperbolic cotangent).

Inverse hyperbolic functions can be expressed in terms of logarithmic functions:

$$\operatorname{arcsinh} x = \ln\left(x + \sqrt{x^2+1}\,\right) \quad (x \text{ is any}); \qquad \operatorname{arccosh} x = \ln\left(x + \sqrt{x^2-1}\,\right) \quad (x \geq 1);$$

$$\operatorname{arctanh} x = \frac{1}{2}\ln\frac{1+x}{1-x} \quad (|x| < 1); \qquad \operatorname{arccoth} x = \frac{1}{2}\ln\frac{x+1}{x-1} \quad (|x| > 1).$$

Here, only one (principal) branch of the function $\operatorname{arccosh} x$ is listed, the function itself being double-valued. In order to write out both branches of $\operatorname{arccosh} x$, the symbol $\pm$ should be placed before the logarithm on the right-hand side of the formula.

Below, the graphs of the inverse hyperbolic functions are given. These are obtained from the graphs of the corresponding hyperbolic functions by mirror reflection with respect to the straight line $y = x$ (with the domain of each function being taken into account).

2.5.1-2. Inverse hyperbolic sine: $y = \operatorname{arcsinh} x$.

This function is defined for all x, and its range coincides with the y-axis. The $\operatorname{arcsinh} x$ is an odd, nonperiodic, unbounded function that crosses the axes Ox and Oy at the origin $x = 0$, $y = 0$. This is an increasing function on the entire real axis with no points of extremum. The graph of the function $y = \operatorname{arcsinh} x$ is given in Fig. 2.18.

2.5.1-3. Inverse hyperbolic cosine: $y = \operatorname{arccosh} x$.

This function is defined for all $x \in [1, +\infty)$, and its range consists of $y \in [0, +\infty)$. The $\operatorname{arccosh} x$ is neither odd nor even; it is nonperiodic and unbounded. It does not cross the axis Oy and crosses the axis Ox at the point $x = 1$. It is an increasing function in its domain with the minimal value $y = 0$ at $x = 1$. The graph of the function $y = \operatorname{arccosh} x$ is given in Fig. 2.19.

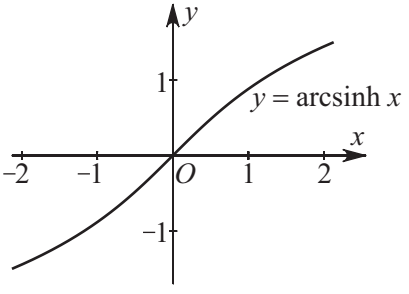


Figure 2.18. The graph of the function $y = \operatorname{arcsinh} x$.

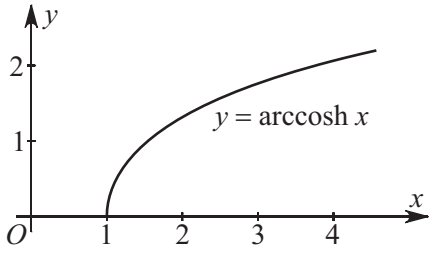


Figure 2.19. The graph of the function $y = \operatorname{arccosh} x$.

2.5.1-4. Inverse hyperbolic tangent: $y = \operatorname{arctanh} x$.

This function is defined for all $x \in (-1, 1)$, and its range consists of all y. The $\operatorname{arctanh} x$ is an odd, nonperiodic, unbounded function that crosses the coordinate axes at the origin $x = 0$, $y = 0$. This is an increasing function in its domain with no points of extremum and an inflection point at the origin. It has two vertical asymptotes: $x = \pm 1$. The graph of the function $y = \operatorname{arctanh} x$ is given in Fig. 2.20.

2.5.1-5. Inverse hyperbolic cotangent: $y = \operatorname{arccoth} x$.

This function is defined for $x \in (-\infty, -1)$ and $x \in (1, +\infty)$. Its range consists of all $y \neq 0$. The $\operatorname{arccoth} x$ is an odd, nonperiodic, unbounded function that does not cross the coordinate axes. It is a decreasing function on each of the semiaxes of its domain. This function has no points of extremum and has one horizontal asymptote $y = 0$ and two vertical asymptotes $x = \pm 1$. The graph of the function $y = \operatorname{arccoth} x$ is given in Fig. 2.21.

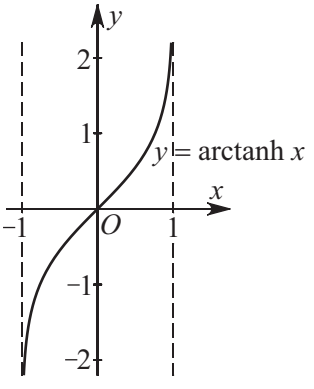


Figure 2.20. The graph of the function $y = \operatorname{arctanh} x$.

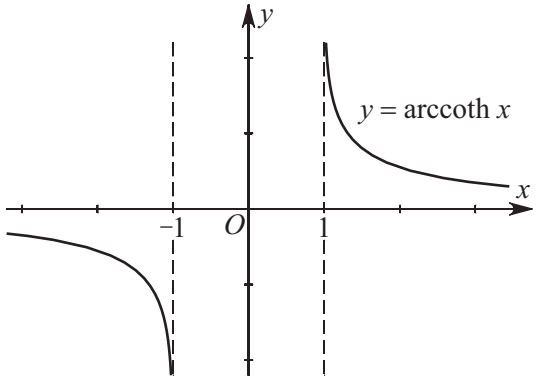


Figure 2.21. The graph of the function $y = \operatorname{arccoth} x$.

2.5.2. Properties of Inverse Hyperbolic Functions

2.5.2-1. Simplest relations.

$$\operatorname{arcsinh}(-x) = -\operatorname{arcsinh} x, \qquad \operatorname{arctanh}(-x) = -\operatorname{arctanh} x, \qquad \operatorname{arccoth}(-x) = -\operatorname{arccoth} x.$$

2.5.2-2. Relations between inverse hyperbolic functions.

$$\operatorname{arcsinh} x = \operatorname{arccosh}\sqrt{x^2+1} = \operatorname{arctanh}\frac{x}{\sqrt{x^2+1}},$$
$$\operatorname{arccosh} x = \operatorname{arcsinh}\sqrt{x^2-1} = \operatorname{arctanh}\frac{\sqrt{x^2-1}}{x},$$
$$\operatorname{arctanh} x = \operatorname{arcsinh}\frac{x}{\sqrt{1-x^2}} = \operatorname{arccosh}\frac{1}{\sqrt{1-x^2}} = \operatorname{arccoth}\frac{1}{x}.$$

2.5.2-3. Addition and subtraction of inverse hyperbolic functions.

$$\operatorname{arcsinh} x \pm \operatorname{arcsinh} y = \operatorname{arcsinh}\left(x\sqrt{1+y^2} \pm y\sqrt{1+x^2}\right),$$
$$\operatorname{arccosh} x \pm \operatorname{arccosh} y = \operatorname{arccosh}\left[xy \pm \sqrt{(x^2-1)(y^2-1)}\right],$$
$$\operatorname{arcsinh} x \pm \operatorname{arccosh} y = \operatorname{arcsinh}\left[xy \pm \sqrt{(x^2+1)(y^2-1)}\right],$$
$$\operatorname{arctanh} x \pm \operatorname{arctanh} y = \operatorname{arctanh}\frac{x \pm y}{1 \pm xy}, \qquad \operatorname{arctanh} x \pm \operatorname{arccoth} y = \operatorname{arctanh}\frac{xy \pm 1}{y \pm x}.$$

2.5.2-4. Differentiation formulas.

$$\frac{d}{dx}\operatorname{arcsinh} x = \frac{1}{\sqrt{x^2+1}}, \qquad \frac{d}{dx}\operatorname{arccosh} x = \frac{1}{\sqrt{x^2-1}},$$
$$\frac{d}{dx}\operatorname{arctanh} x = \frac{1}{1-x^2} \quad (x^2 < 1), \qquad \frac{d}{dx}\operatorname{arccoth} x = \frac{1}{1-x^2} \quad (x^2 > 1).$$

2.5.2-5. Integration formulas.

$$\int \operatorname{arcsinh} x\,dx = x\operatorname{arcsinh} x - \sqrt{1+x^2} + C,$$
$$\int \operatorname{arccosh} x\,dx = x\operatorname{arccosh} x - \sqrt{x^2-1} + C,$$
$$\int \operatorname{arctanh} x\,dx = x\operatorname{arctanh} x + \frac{1}{2}\ln(1-x^2) + C,$$
$$\int \operatorname{arccoth} x\,dx = x\operatorname{arccoth} x + \frac{1}{2}\ln(x^2-1) + C,$$

where C is an arbitrary constant.

2.5.2-6. Expansion in power series.

$$\operatorname{arcsinh} x = x - \frac{1}{2}\frac{x^3}{3} + \frac{1\times 3}{2\times 4}\frac{x^5}{5} - \cdots + (-1)^n \frac{1\times 3\times\cdots\times(2n-1)}{2\times 4\times\cdots\times(2n)}\frac{x^{2n+1}}{2n+1} + \cdots \qquad (|x|<1),$$

$$\operatorname{arcsinh} x = \ln(2x) + \frac{1}{2}\frac{1}{2x^2} + \frac{1\times 3}{2\times 4}\frac{1}{4x^4} + \cdots + \frac{1\times 3\times\cdots\times(2n-1)}{2\times 4\times\cdots\times(2n)}\frac{1}{2nx^{2n}} + \cdots \qquad (|x|>1),$$

$$\operatorname{arccosh} x = \ln(2x) - \frac{1}{2}\frac{1}{2x^2} - \frac{1\times 3}{2\times 4}\frac{1}{4x^4} - \cdots - \frac{1\times 3\times\cdots\times(2n-1)}{2\times 4\times\cdots\times(2n)}\frac{1}{2nx^{2n}} - \cdots \qquad (|x|>1),$$

$$\operatorname{arctanh} x = x + \frac{x^3}{3} + \frac{x^5}{5} + \frac{x^7}{7} + \cdots + \frac{x^{2n+1}}{2n+1} + \cdots \qquad (|x|<1),$$

$$\operatorname{arccoth} x = \frac{1}{x} + \frac{1}{3x^3} + \frac{1}{5x^5} + \frac{1}{7x^7} + \cdots + \frac{1}{(2n+1)x^{2n+1}} + \cdots \qquad (|x|>1).$$

References for Chapter 2

Abramowitz, M. and Stegun, I. A. (Editors), *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, National Bureau of Standards Applied Mathematics, Washington, D. C., 1964.

Adams, R., *Calculus: A Complete Course, 6th Edition*, Pearson Education, Toronto, 2006.

Anton, H., Bivens, I., and Davis, S., *Calculus: Early Transcendental Single Variable, 8th Edition*, John Wiley & Sons, New York, 2005.

Bronshtein, I. N. and Semendyayev, K. A., *Handbook of Mathematics, 4th Edition*, Springer-Verlag, Berlin, 2004.

Courant, R. and John, F., *Introduction to Calculus and Analysis, Vol. 1*, Springer-Verlag, New York, 1999.

Edwards, C. H., and Penney, D., *Calculus, 6th Edition*, Pearson Education, Toronto, 2002.

Gradshteyn, I. S. and Ryzhik, I. M., *Tables of Integrals, Series, and Products, 6th Edition*, Academic Press, New York, 2000.

Kline, M., *Calculus: An Intuitive and Physical Approach, 2nd Edition*, Dover Publications, New York, 1998.

Korn, G. A. and Korn, T. M., *Mathematical Handbook for Scientists and Engineers, 2nd Edition*, Dover Publications, New York, 2000.

Prudnikov, A. P., Brychkov, Yu. A., and Marichev, O. I., *Integrals and Series, Vol. 1, Elementary Functions*, Gordon & Breach, New York, 1986.

Sullivan, M., *Trigonometry, 7th Edition*, Prentice Hall, Englewood Cliffs, 2004.

Thomas, G. B. and Finney, R. L., *Calculus and Analytic Geometry, 9th Edition*, Addison Wesley, Reading, Massachusetts, 1996.

Weisstein, E. W., *CRC Concise Encyclopedia of Mathematics, 2nd Edition*, CRC Press, Boca Raton, 2003.

Zill, D. G. and Dewar, J. M., *Trigonometry, 2nd Edition*, McGraw-Hill, New York, 1990.

Zwillinger, D., *CRC Standard Mathematical Tables and Formulae, 31st Edition*, CRC Press, Boca Raton, 2002.

Chapter 3
Elementary Geometry

3.1. Plane Geometry

3.1.1. Triangles

3.1.1-1. Plane triangle and its properties.

1°. A *plane triangle*, or simply a *triangle*, is a plane figure bounded by three straight line segments (*sides*) connecting three noncollinear points (*vertices*) (Fig. 3.1*a*). The smaller angle between the two rays issuing from a vertex and passing through the other two vertices is called an (*interior*) *angle* of the triangle. The angle adjacent to an interior angle is called an *external angle* of the triangle. An external angle is equal to the sum of the two interior angles to which it is not adjacent.

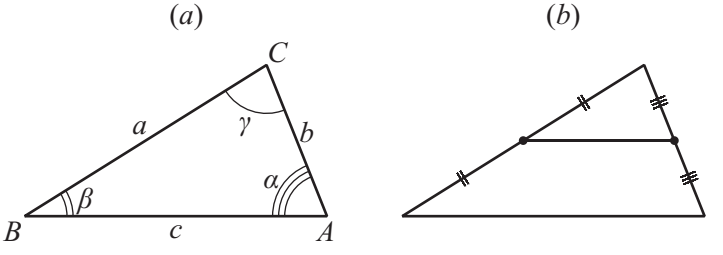


Figure 3.1. Plane triangle (*a*). Midline of a triangle (*b*).

A triangle is uniquely determined by any of the following sets of its parts:

1. Two angles and their included side.
2. Two sides and their included angle.
3. Three sides.

Depending on the angles, a triangle is said to be:

1. *Acute* if all three angles are acute.
2. *Right* (or *right-angled*) if one of the angles is right.
3. *Obtuse* if one of the angles is obtuse.

Depending on the relation between the side lengths, a triangle is said to be:

1. *Regular* (or *equilateral*) if all sides have the same length.
2. *Isosceles* if two of the sides are of equal length.
3. *Scalene* if all sides have different lengths.

2°. *Congruence tests for triangles:*

1. If two sides of a triangle and their included angle are congruent to the corresponding parts of another triangle, then the triangles are congruent.
2. If two angles of a triangle and their included side are congruent to the corresponding parts of another triangle, then the triangles are congruent.

3. If three sides of a triangle are congruent to the corresponding sides of another triangle, then the triangles are congruent.

3°. Triangles are said to be *similar* if their corresponding angles are equal and their corresponding sides are proportional.

Similarity tests for triangles:

1. If all three pairs of corresponding sides in a pair of triangles are in proportion, then the triangles are similar.
2. If two pairs of corresponding angles in a pair of triangles are congruent, then the triangles are similar.
3. If two pairs of corresponding sides in a pair of triangles are in proportion and the included angles are congruent, then the triangles are similar.

The areas of similar triangles are proportional to the squares of the corresponding linear parts (such as sides, altitudes, diagonals, etc.).

4°. The line connecting the midpoints of two sides of a triangle is called a *midline* of the triangle. The midline is parallel to and half as long as the third side (Fig. 3.1*b*).

Let a, b, and c be the lengths of the sides of a triangle; let α, β, and γ be the respective opposite angles (Fig. 3.1*a*); let R and r be the circumradius and the inradius, respectively; and let $p = \frac{1}{2}(a + b + c)$ be the semiperimeter.

Table 3.1 represents the basic properties and relations characterizing triangles.

TABLE 3.1
Basic properties and relations characterizing plane triangles

No.	The name of property	Properties and relations
1	Triangle inequality	The length of any side of a triangle does not exceed the sum of lengths of the other two sides
2	Sum of angles of a triangle	$\alpha + \beta + \gamma = 180^\circ$
3	Law of sines	$\dfrac{a}{\sin\alpha} = \dfrac{b}{\sin\beta} = \dfrac{c}{\sin\gamma} = 2R$
4	Law of cosines	$c^2 = a^2 + b^2 - 2ab\cos\gamma$
5	Law of tangents	$\dfrac{a+b}{a-b} = \dfrac{\tan\left[\frac{1}{2}(\alpha+\beta)\right]}{\tan\left[\frac{1}{2}(\alpha-\beta)\right]} = \dfrac{\cot\left(\frac{1}{2}\gamma\right)}{\tan\left[\frac{1}{2}(\alpha-\beta)\right]}$
6	Theorem on projections (law of cosines)	$c = a\cos\beta + b\cos\alpha$
7	Trigonometric angle formulas	$\sin\dfrac{\gamma}{2} = \sqrt{\dfrac{(p-a)(p-b)}{ab}}$, $\cos\dfrac{\gamma}{2} = \sqrt{\dfrac{p(p-c)}{ab}}$, $\tan\dfrac{\gamma}{2} = \sqrt{\dfrac{(p-a)(p-b)}{p(p-c)}}$, $\sin\gamma = \dfrac{2}{ab}\sqrt{p(p-a)(p-b)(p-c)}$
8	Law of tangents	$\tan\gamma = \dfrac{c\sin\alpha}{b - c\cos\alpha} = \dfrac{c\sin\beta}{a - c\cos\beta}$
9	Mollweide's formulas	$\dfrac{a+b}{c} = \dfrac{\cos\left[\frac{1}{2}(\alpha-\beta)\right]}{\sin\left(\frac{1}{2}\gamma\right)} = \dfrac{\cos\left[\frac{1}{2}(\alpha-\beta)\right]}{\cos\left[\frac{1}{2}(\alpha+\beta)\right]}$, $\dfrac{a-b}{c} = \dfrac{\sin\left[\frac{1}{2}(\alpha-\beta)\right]}{\cos\left(\frac{1}{2}\gamma\right)} = \dfrac{\sin\left[\frac{1}{2}(\alpha-\beta)\right]}{\sin\left[\frac{1}{2}(\alpha+\beta)\right]}$

Table 3.2 permits one to find the sides and angles of an arbitrary triangle if three appropriately chosen sides and/or angles are given. From the relations given in Tables 3.1 and 3.2, one can derive all missing relations by cyclic permutations of the sides a, b, and c and the angles α, β, and γ.

TABLE 3.2
Solution of plane triangles

No.	Three parts specified	Formulas for the remaining parts
1	Three sides a, b, c	*First method.* One of the angles is determined by the law of cosines, $\cos\alpha = \dfrac{b^2+c^2-a^2}{2bc}$. Then either the law of sines or the law of cosines is applied. *Second method.* One of the angles is determined by trigonometric angle formulas. Further proceed in a similar way. Remark. The sum of lengths of any two sides must be greater than the length of the third side.
2	Two sides a, b and the included angle γ	*First method.* The side c is determined by the law of cosines, $c = \sqrt{a^2+b^2-2ab\cos\gamma}$. The angle α is determined by either the law of cosines or the law of sines. The angle β is determined from the sum of angles in triangle, $\beta = 180^\circ - \alpha - \gamma$. *Second method.* $\alpha+\beta$ is found from the sum of angles in triangle, $\alpha+\beta = 180^\circ - \gamma$; $\alpha-\beta$ is found from the law of tangents, $\tan\dfrac{\alpha-\beta}{2} = \dfrac{a-b}{a+b}\cot\dfrac{\gamma}{2}$. Then α and β can be found. The third side c is determined by either the law of cosines or the law of sines.
3	A side c and the two angles α, β adjacent to it	The third angle γ is found from the sum of angles in triangle, $\gamma = 180^\circ - \alpha - \beta$. Sides a and b are determined by the law of sines.
4	Two sides a, b and the angle α opposite one of them	The second angle is determined by the law of sines, $\sin\beta = \dfrac{b}{a}\sin\alpha$. The third angle is $\gamma = 180^\circ - \alpha - \beta$. The third side is determined by the law of sines, $c = a\dfrac{\sin\gamma}{\sin\alpha}$. Remark. Five cases are possible: 1. $a > b$; i.e., the angle is opposite the greater side. Then $\alpha > \beta$, $\beta < 90^\circ$ (the larger angle is opposite the larger side), and the triangle is determined uniquely. 2. $a = b$; i.e., the triangle is isosceles and is determined uniquely. 3. $a < b$ and $b\sin\alpha < a$. Then there are two solutions, $\beta_1 + \beta_2 = 180^\circ$. 4. $a < b$ and $b\sin\alpha = a$. Then the solution is unique, $\beta = 90^\circ$. 5. $a < b$ and $b\sin\alpha > a$. Then there are no solutions.

3.1.1-2. Medians, angle bisectors, and altitudes of triangle.

A straight line through a vertex of a triangle and the midpoint of the opposite side is called a *median* of the triangle (Fig. 3.2*a*). The three medians of a triangle intersect in a single point lying strictly inside the triangle, which is called the *centroid* or *center of gravity* of the triangle. This point cuts the medians in the ratio 2:1 (counting from the corresponding vertices).

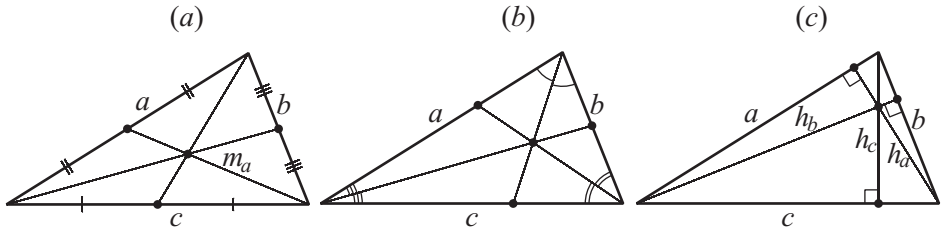


Figure 3.2. Medians (a), angle bisectors (b), and altitudes (c) of a triangle.

The length of the median m_a to the side a opposite the angle α is equal to

$$m_a = \frac{1}{2}\sqrt{2(b^2+c^2)-a^2} = \frac{1}{2}\sqrt{a^2+b^2+2ab\cos\gamma}. \tag{3.1.1.1}$$

An *angle bisector* of a triangle is a line segment between a vertex and a point of the opposite side and dividing the angle at that vertex into two equal parts (Fig. 3.2*b*). The three angle bisectors intersect in a single point lying strictly inside the triangle. This point is equidistant from all sides and is called the *incenter* (the center of the *incircle* of the triangle). Concerning the radius r of the incircle, see Paragraph 3.1.1-3. The angle bisector through a vertex cuts the opposite side in ratio proportional to the adjacent sides of the triangle.

The length of the angle bisector l_a drawn to the side a is given by the formulas

$$\begin{aligned} l_a &= \sqrt{bc-b_1c_1} = \frac{\sqrt{bc[(b+c)^2-a^2]}}{b+c} = \frac{\sqrt{4p(p-a)bc}}{b+c}, \\ l_a &= \frac{2cb\cos\left(\frac{1}{2}\alpha\right)}{b+c} = 2R\frac{\sin\beta\sin\gamma}{\cos\left[\frac{1}{2}(\beta-\gamma)\right]} = 2p\frac{\sin\left(\frac{1}{2}\beta\right)\sin\left(\frac{1}{2}\gamma\right)}{\sin\beta+\sin\gamma}, \end{aligned} \tag{3.1.1.2}$$

where b_1 and c_1 are the segments of the side a cut by bisector l_a and adjacent to the sides b and c, respectively, and R is the circumradius (see Paragraph 3.1.1-3).

An *altitude* of a triangle is a straight line passing through a vertex and perpendicular to the straight line containing the opposite side (Fig. 3.2*c*). The three altitudes of a triangle intersect in a single point, called the *orthocenter* of the triangle.

The length of the altitude h_a to the side a is given by the formulas

$$\begin{aligned} h_a &= b\sin\gamma = c\sin\beta = \frac{bc}{2R}, \\ h_a &= 2(p-a)\cos\frac{\alpha}{2}\cos\frac{\beta}{2}\cos\frac{\gamma}{2} = 2(p-b)\sin\frac{\alpha}{2}\sin\frac{\beta}{2}\cos\frac{\gamma}{2}. \end{aligned} \tag{3.1.1.3}$$

The lengths of the altitude, the angle bisector, and the median through the same vertex satisfy the inequality $h_a \le l_a \le m_a$. If $h_a = l_a = m_a$, then the triangle is isosceles; moreover, the first equality implies the second, and vice versa.

3.1.1-3. Circumcircle, incircle, and excircles.

A straight line passing through the midpoint of a segment and perpendicular to it is called the *perpendicular bisector* of the segment. The circle passing through the vertices of a triangle is called the *circumcircle* of the triangle. The center O_1 of the circumcircle, called the *circumcenter*, is the point where the perpendicular bisectors of the sides of the triangle

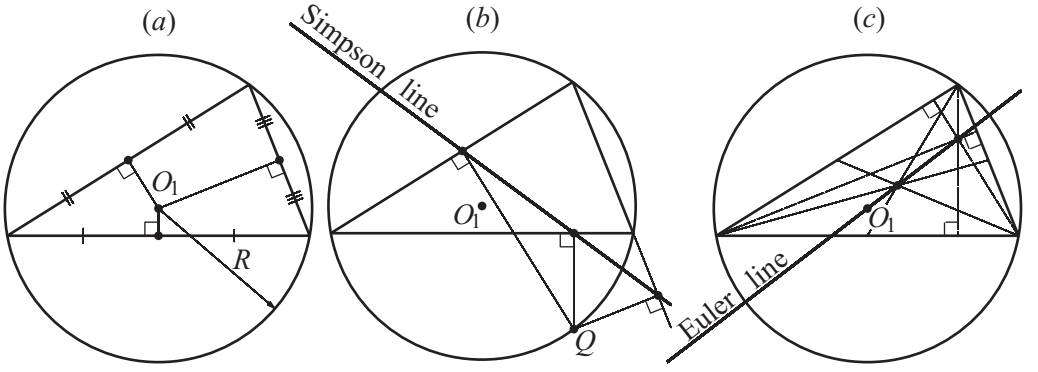


Figure 3.3. The circumcircle of a triangle. The circumcenter (*a*), the Simpson line (*b*), and the Euler line (*c*).

meet (Fig. 3.3*a*). The feet of the perpendiculars drawn from a point Q on the circumcircle to the three sides of the triangle lie on the same straight line called the *Simpson line* of Q with respect to the triangle (Fig. 3.3*b*). The circumcenter, the orthocenter, and the centroid lie on a single line, called the *Euler line* (Fig. 3.3*c*).

The circle tangent to the three sides of a triangle and lying inside the triangle is called the *incircle* of the triangle. The center O_2 of the incircle (the *incenter*) is the point where the angle bisectors meet (Fig. 3.4*a*). The straight lines connecting the vertices of a triangle with the points at which the incircle is tangent to the respective opposite sides intersect in a single point G called the *Gergonne point* (Fig. 3.4*b*).

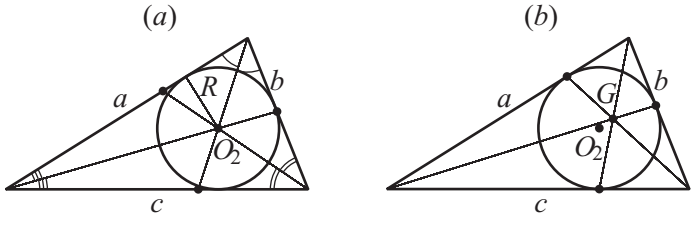


Figure 3.4. The incircle of a triangle (*a*). The incenter and the Gergonne point (*b*).

The circle tangent to one side of a triangle and to the extensions of the other two sides is called an *excircle* of the triangle. Each triangle has three excircles. The center of an excircle (an *excenter*) is the point of concurrency of two external angle bisectors and an interior angle bisector. The straight lines connecting the vertices of a triangle with the points at which the respective opposite sides are tangent to the excircles intersect in a single point N, called the *Nagel point* (Fig. 3.5).

The inradius r, the circumradius R, and the exradii ρ_a, ρ_b, and ρ_c satisfy the relations

$$r = \sqrt{\frac{(p-a)(p-b)(p-c)}{p}} = p\tan\frac{\alpha}{2}\tan\frac{\beta}{2}\tan\frac{\gamma}{2}$$

$$= 4R\sin\frac{\alpha}{2}\sin\frac{\beta}{2}\sin\frac{\gamma}{2} = (p-c)\tan\frac{\gamma}{2} = \frac{S}{p}, \qquad (3.1.1.4)$$

$$R = \frac{a}{2\sin\alpha} = \frac{b}{2\sin\beta} = \frac{c}{2\sin\gamma} = \frac{abc}{4S} = \frac{p}{4\cos\left(\frac{1}{2}\alpha\right)\cos\left(\frac{1}{2}\beta\right)\cos\left(\frac{1}{2}\gamma\right)}, \qquad (3.1.1.5)$$

$$\frac{1}{r} = \frac{1}{\rho_a} + \frac{1}{\rho_b} + \frac{1}{\rho_c}. \qquad (3.1.1.6)$$

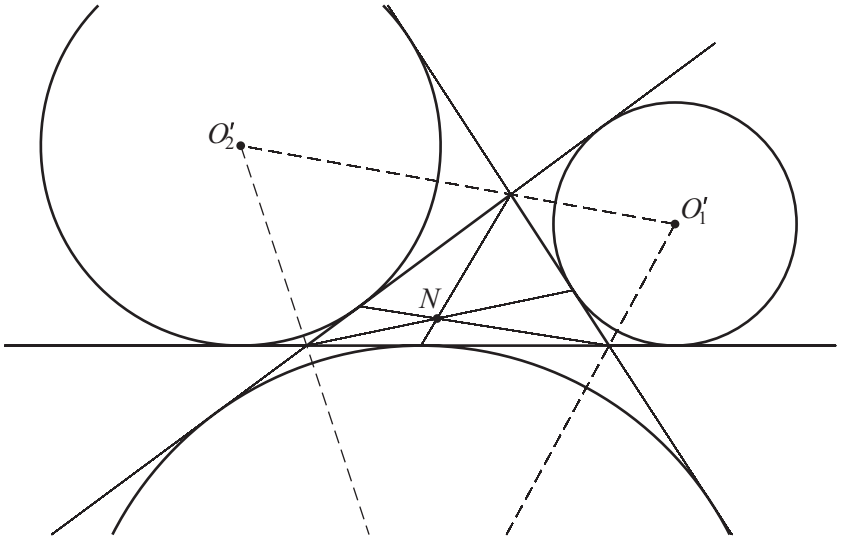


Figure 3.5. Excircles of a triangle. The Nagel point.

The distance d_1 between the circumcenter and the incenter and the distance d_2 between the circumcenter and the excenter are given by the expressions

$$d_1 = \sqrt{R^2 - 2Rr}, \tag{3.1.1.7}$$

$$d_2 = \sqrt{R^2 + 2R\rho_a}. \tag{3.1.1.8}$$

3.1.1-4. Area of a triangle.

The area S of a triangle is given by the formulas

$$\begin{aligned} S &= ah_a = \frac{1}{2}ab\sin\gamma = rp, \\ S &= \sqrt{p(p-a)(p-b)(p-c)} \quad (\textit{Heron's formula}), \\ S &= \frac{abc}{4R} = 2R^2\sin\alpha\sin\beta\sin\gamma, \\ S &= c^2\frac{\sin\alpha\sin\beta}{2\sin\gamma} = c^2\frac{\sin\alpha\sin\beta}{2\sin(\alpha+\beta)}. \end{aligned} \tag{3.1.1.9}$$

3.1.1-5. Theorems about points and lines related to triangle.

CEVA'S THEOREM. *Let points C_1, A_1, and B_1 lie on the sides AB, BC, and CA, respectively, of a triangle (Fig. 3.6). The straight lines AA_1, BB_1, and CC_1 are concurrent or parallel if and only if*

$$\frac{AC_1}{C_1B}\cdot\frac{BA_1}{A_1C}\cdot\frac{CB_1}{B_1A} = 1. \tag{3.1.1.10}$$

STEWART'S THEOREM. *If a straight line through a vertex of a triangle divides the opposite side into segments of lengths m and n (Fig. 3.7), then*

$$(m+n)(p^2+mn) = b^2m + c^2n. \tag{3.1.1.11}$$

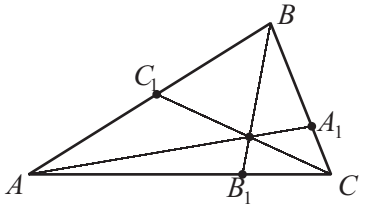


Figure 3.6. Ceva's theorem.

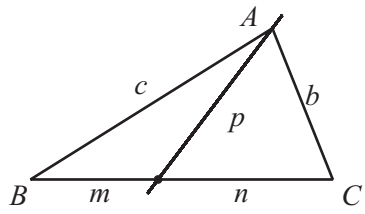


Figure 3.7. Stewart's theorem.

MENELAUS'S THEOREM. *If a straight line intersects sides AB, BC, and CA of a triangle (Fig. 3.8) or their extensions at points C_1, A_1, and B_1, respectively, then*

$$\frac{AC_1}{C_1B} \cdot \frac{BA_1}{A_1C} \cdot \frac{CB_1}{B_1A} = -1. \tag{3.1.1.12}$$

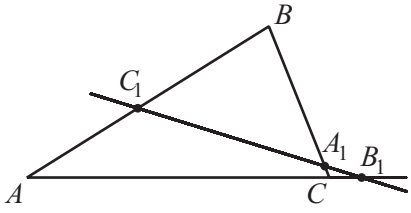


Figure 3.8. Menelaus's theorem.

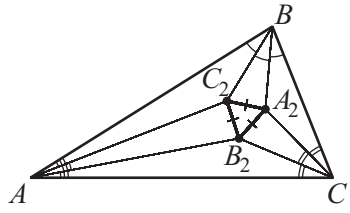


Figure 3.9. Morley's theorem.

Straight lines dividing the interior angles of a triangle into three equal parts are called *angle trisectors*.

MORLEY'S THEOREM. *The three points of intersection of adjacent angle trisectors of a triangle form an equilateral triangle (Fig. 3.9).*

In a triangle, the midpoints of the three sides, the feet of the three altitudes, and the midpoints of the segments of the altitudes between the orthocenter and the vertices all lie on a single circle, the *nine-point circle* (Fig. 3.10).

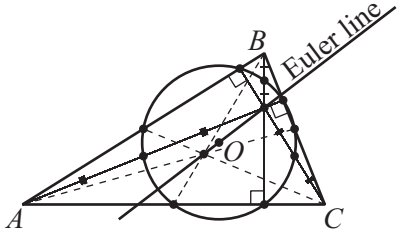


Figure 3.10. Nine-point circle.

FEUERBACH'S THEOREM. *The nine-point circle is tangent to the incircle and the three excircles. The points of tangency are called the Feuerbach points. The center of the nine-point circle lies on the Euler line (see Paragraph 3.1.1-3).*

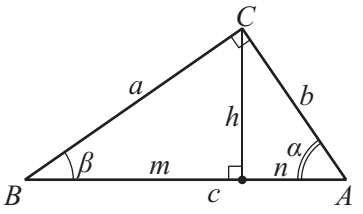


Figure 3.11. A right triangle.

3.1.1-6. Right (right-angled) triangles.

A *right* triangle is a triangle with a right angle. The side opposite the right angle is called the *hypotenuse*, and the other two sides are called the *legs* (Fig. 3.11).

The hypotenuse c, the legs a and b, and the angles α and β opposite the legs satisfy the following relations:

$$\begin{aligned} &\alpha + \beta = 90^\circ; \\ &\sin\alpha = \cos\beta = \frac{a}{c}, \quad \sin\beta = \cos\alpha = \frac{b}{c}, \\ &\tan\alpha = \cot\beta = \frac{a}{b}, \quad \tan\beta = \cot\alpha = \frac{b}{a}. \end{aligned} \tag{3.1.1.13}$$

One also has

$$a^2 + b^2 = c^2 \quad \text{(PYTHAGOREAN THEOREM)}, \tag{3.1.1.14}$$

$$h^2 = mn, \quad a^2 = mc, \quad b^2 = nc, \tag{3.1.1.15}$$

where h is the length of the altitude drawn to the hypotenuse; moreover, the altitude cuts the hypotenuse into segments of lengths m and n.

In a right triangle, the length of the median m_c drawn from the vertex of the right angle coincides with the circumradius R and is equal to half the length of the hypotenuse c, $m_c = R = \frac{1}{2}c$. The inradius is given by the formula $r = \frac{1}{2}(a + b - c)$. The area of the right triangle is $S = ah_a = \frac{1}{2}ab$ (see also Paragraphs 3.1.1-2 to 3.1.1-4).

3.1.1-7. Isosceles and equilateral triangles.

1°. An *isosceles* triangle is a triangle with two equal sides. These sides are called the *legs*, and the third side is called the *base* (Fig. 3.12*a*).

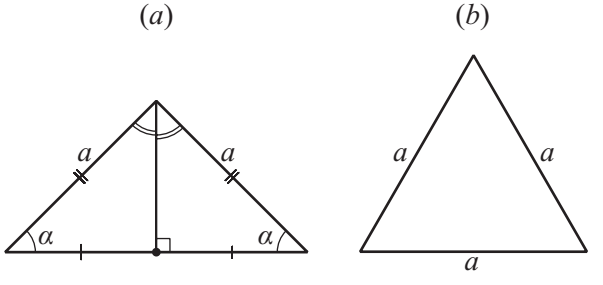


Figure 3.12. An isosceles triangle (*a*). An equilateral triangle (*b*).

Properties of isosceles triangles:

1. In an isosceles triangle, the angles adjacent to the base are equal.
2. In an isosceles triangle, the median drawn to the base is the angle bisector and the altitude.
3. In an isosceles triangle, the sum of distances from a point of the base to the legs is constant.

Criteria for a triangle to be isosceles:

1. If two angles in a triangle are equal, then the triangle is isosceles.
2. If a median in a triangle is also an altitude, then the triangle is isosceles.
3. If a bisector in a triangle is also an altitude, then the triangle is isosceles.

2°. An *equilateral* (or *regular*) triangle is a triangle with all three sides equal (Fig. 3.12*b*). All angles of an equilateral triangle are equal to 60°. In an equilateral triangle, the circumradius R and the inradius r satisfy the relation $R = 2r$.

For an equilateral triangle with side length a, the circumradius and the inradius are given by the formulas $R = \frac{\sqrt{3}}{3}a$ and $r = \frac{\sqrt{3}}{6}a$, and the area is equal to $S = \frac{\sqrt{3}}{4}a^2$.

3.1.2. Polygons

3.1.2-1. Polygons. Basic information.

A *polygon* is a plane figure bounded by a closed broken line, i.e., a line obtained if one takes n distinct points such that no three successive points are collinear and draws a straight line segment between each of these points and its successor as well as between the last point and the first point (Fig. 3.13*a*). The segments forming a polygon are called the *sides* (or *edges*), and the points themselves are called the *vertices* of the polygon. Two sides sharing a vertex, as well as two successive vertices (the endpoints of the same edge), are said to be adjacent. A polygon can be self-intersecting, but the points of self-intersection should not be vertices (Fig. 3.13*b*). A polygon is said to be *plane* if its vertices are coplanar. A polygon is said to be *simple* if its nonadjacent sides do not have common interior or endpoints. A polygon is said to be *convex* if it lies on one side of any straight line passing through two neighboring vertices (Fig. 3.13*c*). In what follows, we consider only plane simple convex polygons.

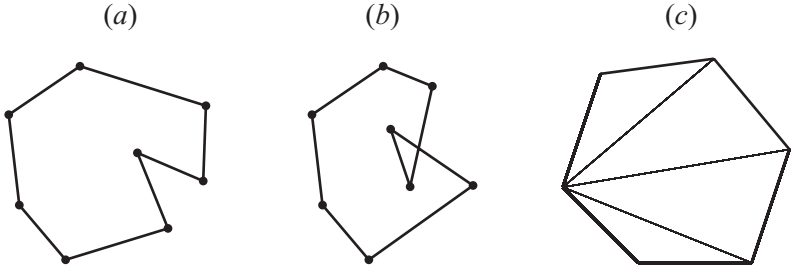


Figure 3.13. Polygons. Nonself-intersecting (*a*), self-intersecting (*b*), and convex (*c*) polygon.

An *(interior) angle* of a convex polygon is the angle between two sides meeting in a vertex. An angle adjacent to an interior angle is called an *external angle* of the convex polygon. A convex polygon is said to be *inscribed* in a circle if all of its vertices lie on the circle. A polygon is said to be *circumscribed* about a circle if all of its sides are tangent to the circle.

For a convex polygon with n sides, the sum of interior angles is equal to $180°(n-2)$, and the sum of external angles is equal to 360°.

One can find the area of an arbitrary polygon by dividing it into triangles.

3.1.2-2. Properties of quadrilaterals.

1. The diagonals of a convex quadrilateral meet.
2. The sum of interior angles of a convex quadrilateral is equal to 360° (Fig. 3.14*a* and *b*).
3. The lengths of the sides a, b, c, and d, the diagonals d_1 and d_2, and the segment m connecting the midpoints of the diagonals satisfy the relation $a^2 + b^2 + c^2 + d^2 = d_1^2 + d_2^2 + 4m^2$.
4. A convex quadrilateral is circumscribed if and only if $a + c = b + d$.
5. A convex quadrilateral is inscribed if and only if $\alpha + \gamma = \beta + \delta$.
6. The relation $ac + bd = d_1 d_2$ holds for inscribed quadrilaterals (PTOLEMY'S THEOREM).

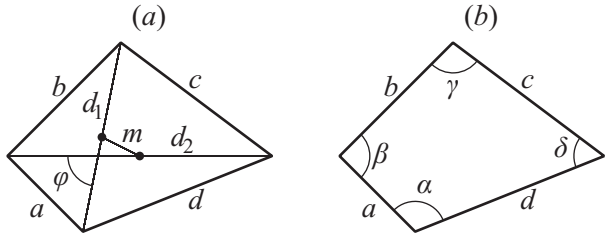


Figure 3.14. Quadrilaterals.

3.1.2-3. Areas of quadrilaterals.

The area of a convex quadrilateral is equal to

$$S = \frac{1}{2} d_1 d_2 \sin\varphi = \sqrt{p(p-a)(p-b)(p-c)(p-d) - abcd\cos^2\frac{\beta+\delta}{2}}, \qquad (3.1.2.1)$$

where φ is the angle between the diagonals d_1 and d_2 and $p = \frac{1}{2}(a + b + c + d)$.

The area of an inscribed quadrilateral is

$$S = \sqrt{p(p-a)(p-b)(p-c)(p-d)}. \qquad (3.1.2.2)$$

The area of a circumscribed quadrilateral is

$$S = \sqrt{abcd\sin^2\frac{\beta+\delta}{2}}. \qquad (3.1.2.3)$$

If a quadrilateral is simultaneously inscribed and circumscribed, then

$$S = \sqrt{abcd}. \qquad (3.1.2.4)$$

3.1.2-4. Basic quadrilaterals.

1°. A *parallelogram* is a quadrilateral such that both pairs of opposite sides are parallel (Fig. 3.15*a*).

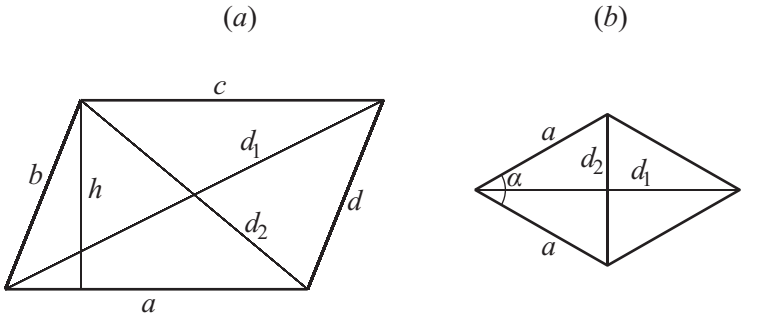


Figure 3.15. A parallelogram (*a*) and a rhombus (*b*).

Attributes of parallelograms (a quadrilateral is a parallelogram if):

1. Both pairs of opposite sides have equal length.
2. Both pairs of opposite angles are equal.
3. Two opposite sides are parallel and have equal length.
4. The diagonals meet and bisect each other.

Properties of parallelograms:

1. The diagonals meet and bisect each other.
2. Opposite sides have equal length, and opposite angles are equal.
3. The diagonals and the sides satisfy the relation $d_1^2 + d_2^2 = 2(a^2 + b^2)$.
4. The area of a parallelogram is $S = ah$, where h is the altitude (see also Paragraph 3.1.2-3).

2°. A *rhombus* is a parallelogram in which all sides are of equal length (Fig. 3.15*b*).

Properties of rhombi:

1. The diagonals are perpendicular.
2. The diagonals are angle bisectors.
3. The area of a rhombus is $S = ah = a^2 \sin\alpha = \frac{1}{2} d_1 d_2$.

3°. A *rectangle* is a parallelogram in which all angles are right angles (Fig. 3.16*a*).

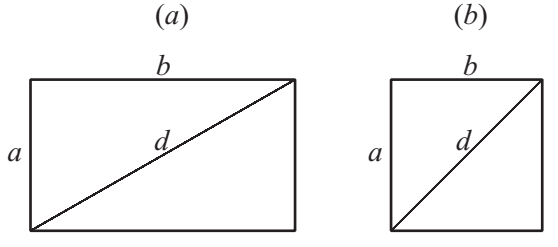


Figure 3.16. A rectangle (*a*) and a square (*b*).

Properties of rectangles:

1. The diagonals have equal lengths.
2. The area of a rectangle is $S = ab$.

4°. A *square* is a rectangle in which all sides have equal lengths (Fig. 3.16*b*). A square is also a special case of a rhombus (all angles are right angles).

Properties of squares:

1. All angles are right angles.
2. The diagonals are equal to $d = a\sqrt{2}$.
3. The diagonals meet at a right angle and are angle bisectors.
4. The area of a square is equal to $S = a^2 = \frac{1}{2}d^2$.

5°. A *trapezoid* is a quadrilateral in which two sides are parallel and the other two sides are nonparallel (Fig. 3.17). The parallel sides a and b are called the *bases* of the trapezoid, and the other two sides are called the *legs*. In an *isosceles* trapezoid, the legs are of equal length. The line segment connecting the midpoints of the legs is called the *median* of the trapezoid. The length of the median is equal to half the sum of the lengths of the bases, $m = \frac{1}{2}(a + b)$.

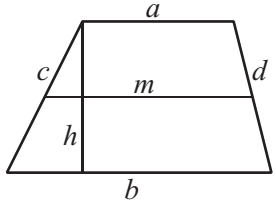


Figure 3.17. A trapezoid.

The perpendicular distance between the bases is called the *altitude* of a trapezoid.

Properties of trapezoids:

1. A trapezoid is circumscribed if and only if $a + b = c + d$.
2. A trapezoid is inscribed if and only if it is isosceles.
3. The area of a trapezoid is $S = \frac{1}{2}(a + b)h = mh = \frac{1}{2}d_1 d_2 \sin\varphi$, where φ is the angle between the diagonals d_1 and d_2.
4. The segment connecting the midpoints of the diagonals is parallel to the bases and has the length $\frac{1}{2}(b - a)$.

Example 1. Consider an application of plane geometry to measuring distances in geodesy. Suppose that the angles α, β, γ, and δ between a straight line AB and the directions to points D and C are known at points A and B (Fig. 3.18*a*). Suppose also that the distance $a = AB$ (or $b = DC$) is known and the task is to find the distance $b = DC$ (or $a = AB$).

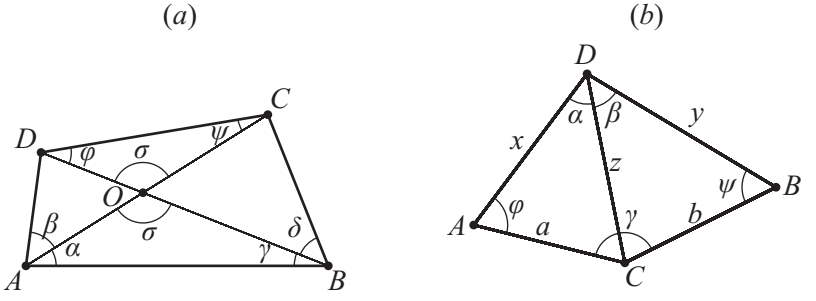


Figure 3.18. Applications of plane geometry in geodesy.

Let us find the angles φ and ψ. Since σ is the angle at the vertex O in both triangles AOB and DOC, it follows that $\alpha + \gamma = \varphi + \psi$. Let $\varepsilon_1 = \frac{1}{2}(\varphi + \psi)$. We twice apply the law of sines (Table 3.1) and find the half-difference of the desired angles. The main formulas read

$$\frac{AD}{a} = \frac{\sin\gamma}{\sin(\pi - \alpha - \beta - \gamma)} = \frac{\sin\gamma}{\sin(\alpha + \beta + \gamma)}, \quad \frac{BC}{a} = \frac{\sin\alpha}{\sin(\alpha + \gamma + \delta)},$$

$$\frac{b}{AD} = \frac{\sin\beta}{\sin\gamma}, \qquad \frac{b}{BC} = \frac{\sin\delta}{\sin\varphi}.$$

These relations imply that

$$\frac{b}{a} = \frac{\sin\beta\sin\gamma}{\sin\psi\sin(\alpha+\beta+\gamma)} = \frac{\sin\delta\sin\alpha}{\sin\varphi\sin(\alpha+\gamma+\delta)} \tag{3.1.2.5}$$

and hence

$$\frac{\sin\varphi}{\sin\psi} = \frac{\sin\delta\sin\alpha\sin(\alpha+\beta+\gamma)}{\sin\beta\sin\gamma\sin(\alpha+\gamma+\delta)} = \cot\eta,$$

where η is an auxiliary angle. By adding and subtracting, we obtain

$$\frac{\sin\varphi-\sin\psi}{\sin\varphi+\sin\psi} = \frac{\cot\eta-1}{\cot\eta+1}, \qquad \frac{2\cos\left[\frac{1}{2}(\varphi+\psi)\right]\sin\left[\frac{1}{2}(\varphi-\psi)\right]}{2\sin\left[\frac{1}{2}(\varphi+\psi)\right]\cos\left[\frac{1}{2}(\varphi-\psi)\right]} = \frac{\cot\left(\frac{1}{4}\pi\right)\cot\eta-1}{\cot\eta+\cot\left(\frac{1}{4}\pi\right)},$$

$$\tan\frac{\varphi-\psi}{2} = \tan\frac{\varphi+\psi}{2}\cot\left(\frac{\pi}{4}+\eta\right) = \tan\frac{\alpha+\gamma}{2}\cot\left(\frac{\pi}{4}+\eta\right).$$

From this we find $\varepsilon_2 = \frac{1}{2}(\varphi-\psi)$ and, substituting $\varphi = \varepsilon_1+\varepsilon_2$ and $\psi = \varepsilon_1-\varepsilon_2$ into (3.1.2.5), obtain the desired distance.

Example 2. Suppose that the mutual position of three points A, B, and C is determined by the segments $AC = a$ and $BC = b$, and the angle $\angle ACB = \gamma$. Suppose that the following angles have been measured at some point D: $\angle CDA = \alpha$ and $\angle CDB = \beta$.

In the general case, one can find the position of point D with respect to A, B, and C, i.e., uniquely determine the segments x, y, and z (Fig. 3.18*b*). For this to be possible, it is necessary that D does not lie on the circumcircle of the triangle ABC. We have

$$\varphi+\psi = 2\pi-(\alpha+\beta+\gamma) = 2\varepsilon_1. \tag{3.1.2.6}$$

By the law of sines (Table 3.1), we obtain

$$\sin\varphi = \frac{z}{a}\sin\alpha, \quad \sin\psi = \frac{z}{b}\sin\beta, \tag{3.1.2.7}$$

which implies that

$$\frac{\sin\varphi}{\sin\psi} = \frac{b\sin\alpha}{a\sin\beta} = \cot\eta, \tag{3.1.2.8}$$

where η is an auxiliary angle. We find the angles φ and ψ from (3.1.2.6) and (3.1.2.8), substitute them into (3.1.2.7) to determine z, and finally apply the law of sines to obtain x and y.

3.1.2-5. Regular polygons.

A convex polygon is said to be *regular* if all of its sides have the same length and all of its interior angles are equal. A convex n-gon is regular if and only if it is taken to itself by the rotation by an angle of $2\pi/n$ about some point O. The point O is called the *center* of the regular polygon. The angle between two rays issuing from the center and passing through two neighboring vertices is called the *central angle* (Fig. 3.19).

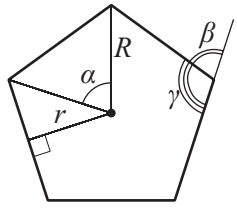


Figure 3.19. A regular polygon.

Properties of regular polygons:

1. The center is equidistant from all vertices as well as from all sides of a regular polygon.
2. A regular polygon is simultaneously inscribed and circumscribed; the centers of the circumcircle and the incircle coincide with the center of the polygon itself.
3. In a regular polygon, the central angle is $\alpha = 360^\circ/n$, the external angle is $\beta = 360^\circ/n$, and the interior angle is $\gamma = 180^\circ - \beta$.
4. The circumradius R, the inradius r, and the side length a of a regular polygon satisfy the relations
$$a = 2\sqrt{R^2 - r^2} = 2R \sin\frac{\alpha}{2} = 2r \tan\frac{\alpha}{2}. \tag{3.1.2.9}$$
5. The area S of a regular n-gon is given by the formula
$$S = \frac{arn}{2} = nr^2 \tan\frac{\alpha}{2} = nR^2 \sin\frac{\alpha}{2} = \frac{1}{4}na^2 \cot\frac{\alpha}{2}. \tag{3.1.2.10}$$

Table 3.3 presents several useful formulas for regular polygons.

TABLE 3.3
Regular polygons (a is the side length)

No.	Name	Inradius r	Circumradius R	Area S
1	Regular polygon	$\frac{a}{2\tan\frac{\pi}{n}}$	$\frac{a}{2\sin\frac{\pi}{n}}$	$\frac{1}{2}arn$
2	Triangle	$\frac{\sqrt{3}}{6}a$	$\frac{\sqrt{3}}{3}a$	$\frac{\sqrt{3}}{4}a^2$
3	Square	$\frac{1}{2}a$	$\frac{1}{\sqrt{2}}a$	a^2
4	Pentagon	$\sqrt{\frac{5+2\sqrt{5}}{20}}a$	$\sqrt{\frac{5+\sqrt{5}}{10}}a$	$\frac{\sqrt{25+10\sqrt{5}}}{4}a^2$
5	Hexagon	$\frac{\sqrt{3}}{2}a$	a	$\frac{3\sqrt{3}}{2}a^2$
6	Octagon	$\frac{1+\sqrt{2}}{2}a$	$\frac{\sqrt{2+\sqrt{2}}}{2}a$	$2(1+\sqrt{2})a^2$
7	Enneagon	$\frac{5+2\sqrt{5}}{2}a$	$\frac{1+\sqrt{5}}{2}a$	$\frac{\sqrt{5+2\sqrt{5}}}{2}a^2$
8	Dodecagon	$\frac{2+\sqrt{3}}{2}a$	$\frac{3+\sqrt{3}}{\sqrt{6}}a$	$3(2+\sqrt{3})a^2$

3.1.3. Circle

3.1.3-1. Some definitions and formulas.

The *circle of radius R centered at O* is the set of all points of the plane at a fixed distance R from a fixed point O (Fig. 3.20*a*). A plane figure bounded by a circle is called a *disk*. A segment connecting two points on a circle is called a *chord*. A chord passing through the center of a circle is called a *diameter* of the circle (Fig. 3.20*b*). The diameter length is $d = 2R$. A straight line that meets a circle at a single point is called a *tangent*, and the common point is called the *point of tangency* (Fig. 3.20*c*). An angle formed by two radii is called a *central angle*. An angle formed by two chords with a common endpoint is called an *inscribed angle*.

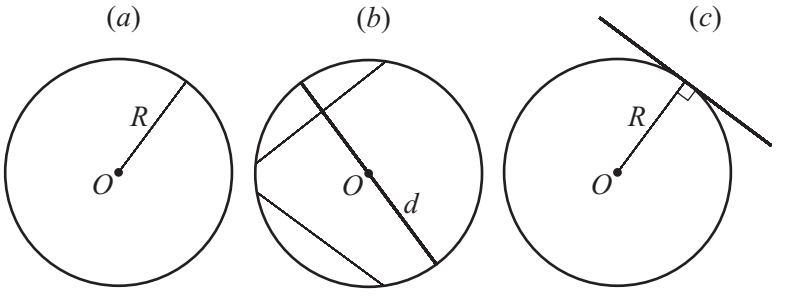


Figure 3.20. A circle (*a*). A diameter (*b*) and a tangent (*c*) of a circle.

Properties of circles and disks:

1. The circumference is $L = 2\pi R = \pi d = 2\sqrt{\pi S}$.
2. The area of a disk is $S = \pi R^2 = \frac{1}{4}\pi d^2 = \frac{1}{4}Ld$.
3. The diameter of a circle is a longest chord.
4. The diameter passing through the midpoint of the chord is perpendicular to the chord.
5. The radius drawn to the point of tangency is perpendicular to the tangent.
6. An inscribed angle is half the central angle subtended by the same chord, $\alpha = \frac{1}{2}\angle BOC$ (Fig. 3.21*a*).
7. The angle between a chord and the tangent to the circle at an endpoint of the chord is $\beta = \frac{1}{2}\angle AOC$ (Fig. 3.21*a*).
8. The angle between two chords is $\gamma = \frac{1}{2}(\overset{\smile}{BC} + \overset{\smile}{ED})$ (Fig. 3.21*b*).
9. The angle between two secants is $\alpha = \frac{1}{2}(\overset{\smile}{DE} - \overset{\smile}{BC})$ (Fig. 3.21*c*).

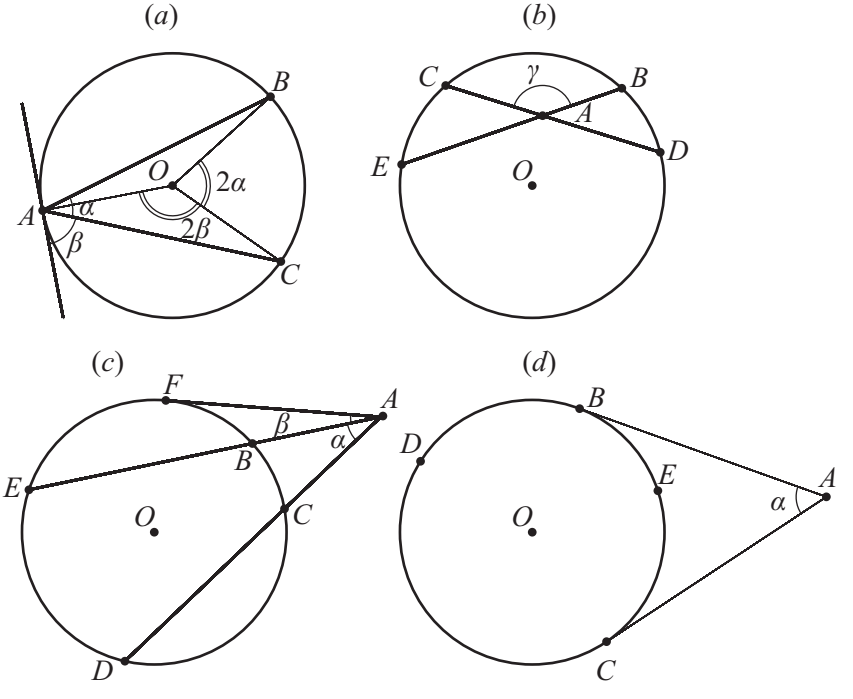


Figure 3.21. Properties of circles and disks.

10. The angle between a secant and the tangent to the circle at an endpoint of the secant is

$\beta = \frac{1}{2}(\overset{\smile}{FE} - \overset{\smile}{BF})$ (Fig. 3.21*c*).

11. The angle between two tangents is $\alpha = \frac{1}{2}(B\overset{\smile}{DC} - B\overset{\smile}{EC})$ (Fig. 3.21*d*).
12. If two chords meet, then $AC \cdot AD = AB \cdot AE = R^2 - m^2$ (Fig. 3.21*b*).
13. For secants, $AC \cdot AD = AB \cdot AE = m^2 - R^2$ (Fig. 3.21*c*).
14. For a tangent and a secant, $AF \cdot AF = AC \cdot AD$ (Fig. 3.21*c*).

3.1.3-2. Segment and sector.

A plane figure bounded by two radii and one of the subtending arcs is called a *(circular) sector*. A plane figure bounded by an arc and the corresponding chord is called a *segment* (Fig. 3.22*a*). If R is the radius of the circle, l is the arc length, a is the chord length, α is the central angle (in degrees), and h is the height of the segment, then the following formulas hold:

$$\begin{aligned} a &= 2\sqrt{2hR - h^2} = 2R \sin\frac{\alpha}{2}, \\ h &= R - \sqrt{R^2 - \frac{a^2}{4}} = R\left(1 - \cos\frac{\alpha}{2}\right) = \frac{a}{2}\tan\frac{\alpha}{4}, \\ l &= \frac{2\pi R\alpha}{360} \approx 0.01745\, R\alpha. \end{aligned} \tag{3.1.3.1}$$

The area of a circular sector is given by the formula

$$S = \frac{lR}{2} = \frac{\pi R^2 \alpha}{360} \approx 0.00873\, R^2\alpha, \tag{3.1.3.2}$$

and the area of a segment not equal to a half-disk is given by the expression

$$S_1 = \frac{\pi R^2\alpha}{360} \pm S_\Delta, \tag{3.1.3.3}$$

where S_Δ is the area of the triangle with vertices at the center of the disk and at the endpoints of the radii bounding the corresponding sector. One takes the minus sign for $\alpha < 180$ and the plus sign for $\alpha > 180$.

The arc length and the area of a segment can be found by the approximate formulas

$$\begin{aligned} l &\approx \frac{8b - a}{3}, \quad l \approx \sqrt{a^2 + \frac{16h^2}{3}}, \\ S_1 &\approx \frac{h(6a + 8b)}{15}, \end{aligned} \tag{3.1.3.4}$$

where b is the chord of the half-segment (see Fig. 3.22*a*).

3.1.3-3. Annulus.

An *annulus* is a plane figure bounded by two concentric circles of distinct radii (Fig. 3.22*b*). Let R be the outer radius of an annulus (the radius of the outer bounding circle), and let r

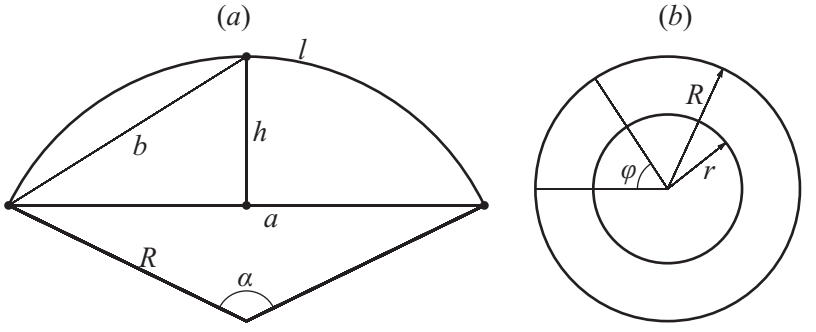


Figure 3.22. A segment (*a*) and an annulus (*b*).

be the inner radius (the radius of the inner bounding circle). Then the area of the annulus is given by the formula

$$S = \pi(R^2 - r^2) = \frac{\pi}{4}(D^2 - d^2) = 2\pi\rho\delta, \tag{3.1.3.5}$$

where $D = 2R$ and $d = 2r$ are the outer and inner diameters, $\rho = \frac{1}{2}(R + r)$ is the midradius, and $\delta = R - r$ is the width of the annulus.

The area of the part of the annulus contained in a sector of central angle φ, given in degrees (see Fig. 3.22*b*), is given by the formula

$$S = \frac{\pi\varphi}{360}(R^2 - r^2) = \frac{\pi\varphi}{1440}(D^2 - d^2) = \frac{\pi\varphi}{180}\rho\delta. \tag{3.1.3.6}$$

3.2. Solid Geometry

3.2.1. Straight Lines, Planes, and Angles in Space

3.2.1-1. Mutual arrangement of straight lines and planes.

1°. Two distinct straight lines lying in a single plane either have exactly one point of intersection or do not meet at all. In the latter case, they are said to be *parallel*. If two straight lines do not lie in a single plane, then they are called *skew lines*.

The angle between skew lines is determined as the angle between lines parallel to them and lying in a single plane (Fig. 3.23*a*). The distance between skew lines is the length of the straight line segment that meets both lines and is perpendicular to them.

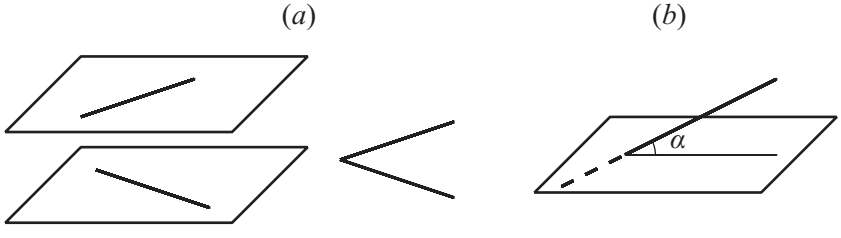


Figure 3.23. The angle between skew lines (*a*). The angle between a line and a plane (*b*).

2°. Two distinct planes either intersect in a straight line or do not have common points. In the latter case, they are said to be *parallel*. Coinciding planes are also assumed to be parallel. If two planes are perpendicular to a single straight line or each of them contains a pair of intersecting straight lines parallel to the corresponding lines in the other pair, then the planes are parallel.

3°. A straight line either entirely lies in the plane, meets the plane at a single point, or has no common points with the plane. In the last case, the line is said to be *parallel* to the plane.

The angle between a straight line and a plane is equal to the angle between the line and its projection onto the plane (Fig. 3.23*b*). If a straight line is perpendicular to two intersecting straight lines on a plane, then it is perpendicular to each line on the plane, i.e., *perpendicular to the plane*.

3.2.1-2. Polyhedral angles.

1°. A *dihedral angle* is a figure in space formed by two half-planes issuing from a single straight line as well as the part of space bounded by these half-planes. The half-planes are called the *faces* of the dihedral angle, and their common straight line is called the *edge*. A dihedral angle is measured by its linear angle ABC (Fig. 3.24*a*), i.e., by the angle between the perpendiculars raised to the edge DE of the dihedral angle in both planes (*faces*) at the same point.

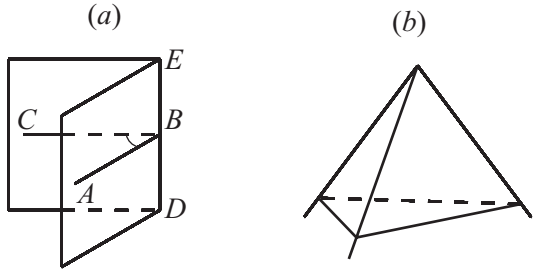


Figure 3.24. A dihedral (*a*) and a trihedral (*b*) angle.

2°. A part of space bounded by an infinite triangular pyramid is called a *trihedral angle* (Fig. 3.24*b*). The faces of this pyramid are called the *faces* of the trihedral angle, and the vertex of the pyramid is called the *vertex of a trihedral angle*. The rays in which the faces intersect are called the *edges* of a trihedral angle. The edges form *face angles*, and the faces form the dihedral angles of the trihedral angle. As a rule, one considers trihedral angles with dihedral angles less than π (or 180°), i.e., convex trihedral angles. Each face angle of a convex trihedral angle is less than the sum of the other two face angles and greater than their difference.

Two trihedral angles are equal if one of the following conditions is satisfied:

1. Two face angles, together with the included dihedral angle, of the first trihedral angle are equal to the respective parts (arranged in the same order) of the second trihedral angle.
2. Two dihedral angles, together with the included face angle, of the first trihedral angle are equal to the respective parts (arranged in the same order) of the second trihedral angle.
3. The three face angles of the first trihedral angle are equal to the respective face angles (arranged in the same order) of the second trihedral angle.
4. The three dihedral angles of the first trihedral angle are equal to the respective dihedral angles (arranged in the same order) of the second trihedral angle.

3°. A *polyhedral angle* $OABCDE$ (Fig. 3.25*a*) is formed by several planes (*faces*) having a single common point (the *vertex*) and successively intersecting along straight lines OA,

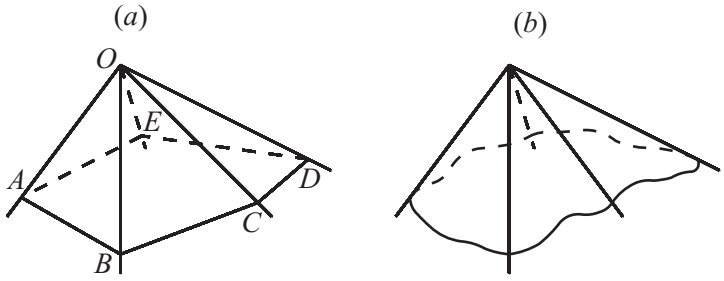


Figure 3.25. A polyhedral (*a*) and a solid (*b*) angle.

$OB, \ldots, OE$ (the *edges*). Two edges belonging to the same face form a *face angle* of the polyhedral angle, and two neighboring faces form a *dihedral angle*.

Polyhedral angles are equal (*congruent*) if one can be transformed into the other by translations and rotations. For polyhedral angles to be congruent, the corresponding parts (face and dihedral angles) must be equal. However, if the corresponding equal parts are arranged in reverse order, then the polyhedral angles cannot be transformed into each other by translations and rotations. In this case, they are said to be *symmetric*.

A convex polyhedral angle lies entirely on one side of each of its faces. The sum $\angle AOB + \angle BOC + \cdots + \angle EOA$ of face angles (Fig. 3.25*a*) of any convex polyhedral angle is less that 2π (or 360°).

4°. A *solid angle* is a part of space bounded by straight lines issuing from a single point (vertex) to all points of some closed curve (Fig. 3.25*b*). Trihedral and polyhedral angles are special cases of solid angles. A solid angle is measured by the area cut by the solid angle on the sphere of unit radius centered at the vertex. Solid angles are measured in steradians. The entire sphere forms a solid angle of 4π steradians.

3.2.2. Polyhedra

3.2.2-1. General concepts.

A *polyhedron* is a solid bounded by planes. In other words, a polyhedron is a set of finitely many plane polygons satisfying the following conditions:

1. Each side of each polygon is simultaneously a side of a unique other polygon, which is said to be adjacent to the first polygon (via this side).
2. From each of the polygons forming a polyhedron, one can reach any other polygon by successively passing to adjacent polygons.

These polygons are called the *faces*, their sides are called the *edges*, and their vertices are called the *vertices* of a polyhedron.

A polyhedron is said to be *convex* if it lies entirely on one side of the plane of any of its faces; if a polyhedron is convex, then so are its faces.

EULER'S THEOREM. *If the number of vertices in a convex polyhedron is e, the number of edges is f, and the number of faces is g, then $e + f - g = 2$.*

3.2.2-2. Prism. Parallelepiped.

1°. A *prism* is a polyhedron in which two faces are n-gons (the *base faces* of the prism) and the remaining n faces (*joining faces*) are parallelograms. The base faces of a prism are

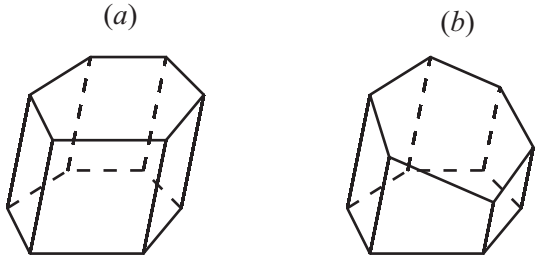


Figure 3.26. A prism (*a*) and a truncated prism (*b*).

equal (congruent) and lie in parallel planes (Fig. 3.26*a*). A *right* prism is a prism in which the joining faces are perpendicular to the base faces. A right prism is said to be *regular* if its base face is a regular polygon.

If l is the joining edge length, S is the area of the base face, H is the altitude of the prism, P_{sec} is the perimeter of a perpendicular section, and S_{sec} is the area of the perpendicular section, then the area of the lateral surface S_{lat} and the volume V of the prism can be determined by the formulas

$$\begin{aligned} S_{\text{lat}} &= P_{\text{sec}} l \\ V &= SH = S_{\text{sec}} l. \end{aligned} \tag{3.2.2.1}$$

The portion of a prism cut by a plane nonparallel to the base face is called a *truncated prism* (Fig. 3.26*b*). The volume of a truncated prism is

$$V = LP_1, \tag{3.2.2.2}$$

where L is the length of the segment connecting the centers of the base faces and P_1 is the area of the section of the prism by a plane perpendicular to this segment.

2°. A prism whose bases are parallelograms is called a *parallelepiped*. All four diagonals in a parallelepiped intersect at a single point and bisect each other (Fig. 3.27*a*). A parallelepiped is said to be *rectangular* if it is a right prism and its base faces are rectangles. In a rectangular parallelepiped, all diagonals are equal (Fig. 3.27*b*).

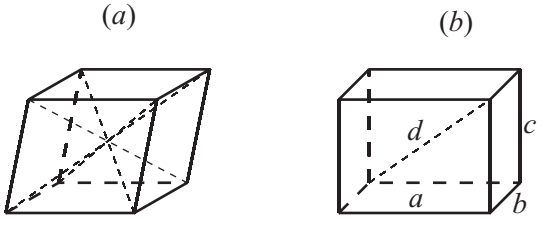


Figure 3.27. A parallelepiped (*a*) and a rectangular parallelepiped (*b*).

If a, b, and c are the lengths of the edges of a rectangular parallelepiped, then the diagonal d can be determined by the formula $d^2 = a^2 + b^2 + c^2$. The volume of a rectangular parallelepiped is given by the formula $V = abc$, and the lateral surface area is $S_{\text{lat}} = PH$, where P is the perimeter of the base face.

3°. A rectangular parallelepiped all of whose edges are equal ($a = b = c$) is called a *cube*. The diagonal of a cube is given by the formula $d^2 = 3a^2$. The volume of the cube is $V = a^3$, and the lateral surface area is $S_{lat} = 4a^2$.

3.2.2-3. Pyramid, obelisk, and wedge.

1°. A *pyramid* is a polyhedron in which one face (the *base* of the pyramid) is an arbitrary polygon and the other (*lateral*) faces are triangles with a common vertex, called the *apex* of the pyramid (Fig. 3.28*a*). The base of an n-*sided* pyramid is an n-gon. The perpendicular through the apex to the base of a pyramid is called the *altitude* of the pyramid.

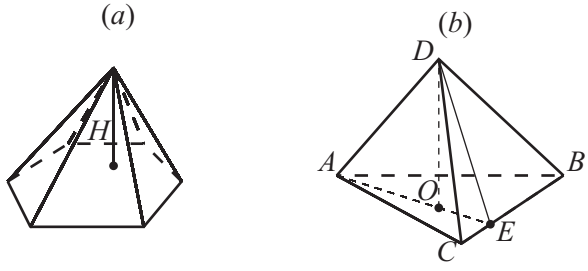


Figure 3.28. A pyramid (*a*). The attitude DO, the plane DAE, and the side BC in a triangular pyramid (*b*).

The volume of a pyramid is given by the formula

$$V = \frac{1}{3}SH, \tag{3.2.2.3}$$

where S is the area of the base and H is the altitude of the pyramid.

The apex of a pyramid is projected onto the circumcenter of the base if one of the following conditions is satisfied:

1. The lengths of all lateral edges are equal.
2. All lateral edges make equal angles with the base plane.

The apex of a pyramid is projected onto the incenter of the base if one of the following conditions is satisfied:

3. All lateral faces have equal apothems.
4. The angles between all lateral faces and the base are the same.

If DO is the altitude of the pyramid $ABCD$ and $DA \perp BC$, then the plane DAE is perpendicular to BC (Fig. 3.28*b*).

If the pyramid is cut by a plane (Fig. 3.29*a*) parallel to the base, then

$$\begin{gathered} \frac{SA_1}{A_1A} = \frac{SB_1}{B_1B} = \cdots = \frac{SO_1}{O_1O}, \\ \frac{S_{ABCDEF}}{S_{A_1B_1C_1D_1E_1F_1}} = \left(\frac{SO}{SO_1}\right)^2, \end{gathered} \tag{3.2.2.4}$$

where SO is the altitude of the pyramid, i.e., the segment of the perpendicular through the vertex to the base.

The altitude of a *triangular pyramid* passes through the orthocenter of the base if and only if all pairs of opposite edges of the pyramid are perpendicular. The volume of a triangular pyramid (Fig. 3.29*b*), where $DA = a$, $DB = b$, $DC = c$, $BC = p$, $AC = q$, and $AB = r$, is given by the formula

$$V^2 = \frac{1}{288}\begin{vmatrix} 0 & r^2 & q^2 & a^2 & 1 \\ r^2 & 0 & p^2 & b^2 & 1 \\ q^2 & p^2 & 0 & c^2 & 1 \\ a^2 & b^2 & c^2 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}, \tag{3.2.2.5}$$

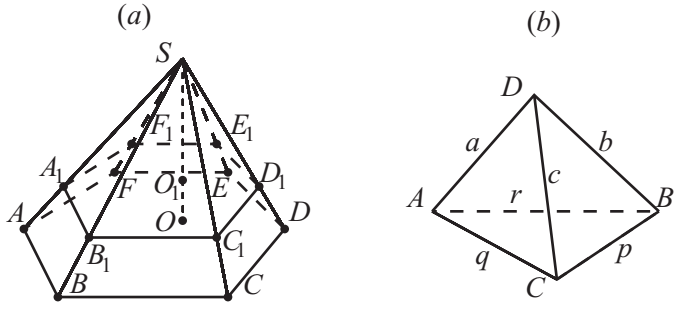


Figure 3.29. The pyramid cut by a plane and the original pyramid (a). A triangular pyramid (b).

where the right-hand side contains a determinant.

A pyramid is said to be *regular* if its base is a regular n-gon and the altitude passes through the center of the base. The altitude (issuing from the apex) of a lateral face is called the *apothem* of a regular pyramid. For a regular pyramid, the lateral surface area is

$$S_{\text{lat}} = \frac{1}{2}Pl, \tag{3.2.2.6}$$

where P is the perimeter of the base and l is the apothem.

2°. If a pyramid is cut by a plane parallel to the base, then it splits into two parts, a pyramid similar to the original pyramid and the *frustum* (Fig. 3.30a). The volume of the frustum is

$$V = \frac{1}{3}h(S_1 + S_2 + \sqrt{S_1 S_2}) = \frac{1}{3}hS_2\left[1 + \frac{a}{A} + \frac{a^2}{A^2}\right], \tag{3.2.2.7}$$

where S_1 and S_2 are the areas of the bases, a and A are two respective sides of the bases, and h is the altitude (the perpendicular distance between the bases).

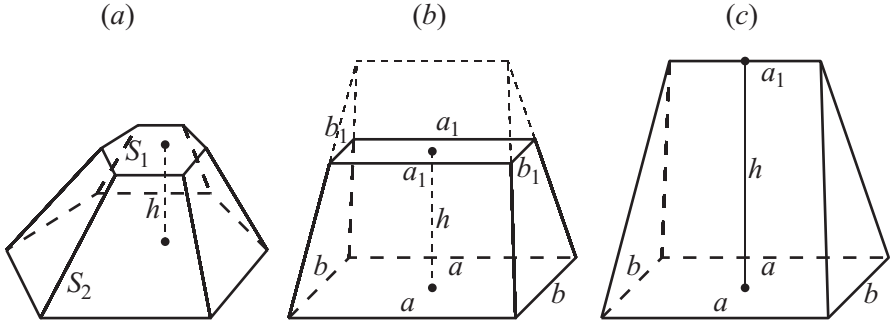


Figure 3.30. A frustum of a pyramid (a), an obelisk (b), and a wedge (c).

For a regular frustum, the lateral surface area is

$$S_{\text{lat}} = \frac{1}{2}(P_1 + P_2)l, \tag{3.2.2.8}$$

where P_1 and P_2 are the perimeters of the bases and l is the altitude of the lateral face.

3°. A hexahedron whose bases are rectangles lying in parallel planes and whose lateral faces form equal angles with the base, but do not meet at a single point, is called an *obelisk* (Fig. 3.30*b*). If a, b and a_1, b_1 are the sides of the bases and h is the altitude, then the volume of the hexahedron is

$$V = \frac{h}{6}[(2a + a_1)b + (2a_1 + a)b_1]. \tag{3.2.2.9}$$

4°. A pentahedron whose base is a rectangle and whose lateral faces are isosceles triangles and isosceles trapezoids is called a *wedge* (Fig. 3.30*c*). The volume of the wedge is

$$V = \frac{h}{6}(2a + a_1)b. \tag{3.2.2.10}$$

3.2.2-4. Regular polyhedra.

A polyhedron is said to be *regular* if all of its faces are equal regular polygons and all polyhedral angles are equal to each other. There exist five regular polyhedra (Fig. 3.31), whose properties are given in Table 3.4.

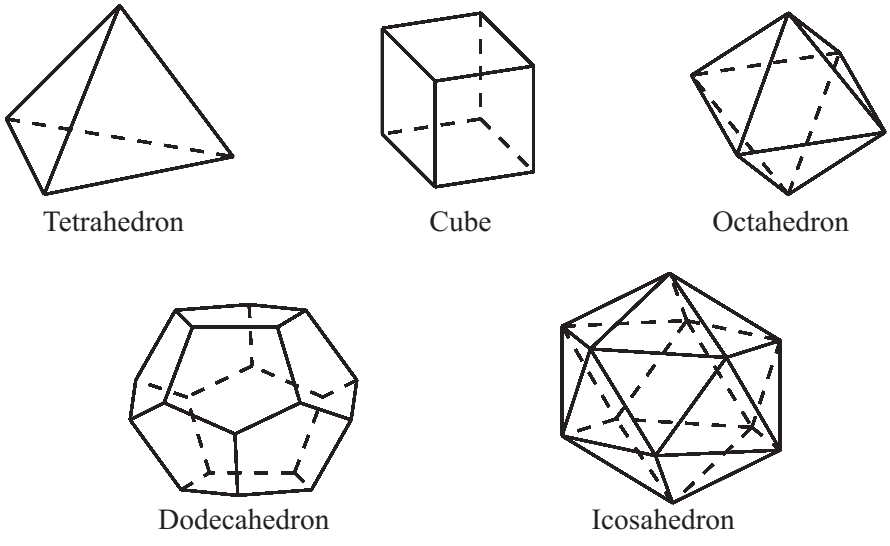


Figure 3.31. Five regular polyhedra.

3.2.3. Solids Formed by Revolution of Lines

3.2.3-1. Cylinder.

A *cylindrical surface* is a surface in space swept by a straight line (the *generator*) moving parallel to a given direction along some curve (the *directrix*) (Fig. 3.32*a*).

1°. A solid bounded by a closed cylindrical surface and two planes is called a *cylinder*; the planes are called the *bases* of the cylinder (Fig. 3.32*b*).

If P is the perimeter of the base, P_{sec} is the perimeter of the section perpendicular to the generator, S_{sec} is the area of this section, S_{bas} is the area of the base, and l is the length of

TABLE 3.4
Regular polyhedra (a is the edge length)

No.	Name	Number of faces and its shape	Number of vertices	Number of edges	Total surface area	Volume
1	Tetrahedron	4 triangles	4	6	$a^2\sqrt{3}$	$\frac{a^3\sqrt{2}}{12}$
2	Cube	6 squares	8	12	$6a^2$	a^3
3	Octahedron	8 triangles	6	12	$2a^2\sqrt{3}$	$\frac{a^3\sqrt{2}}{3}$
4	Dodecahedron	12 pentagons	20	30	$3a^2\sqrt{25+10\sqrt{5}}$	$\frac{a^3}{4}(15+7\sqrt{5})$
5	Icosahedron	20 triangles	12	30	$3a^2\sqrt{3}$	$\frac{5a^3}{12}(3+\sqrt{5})$

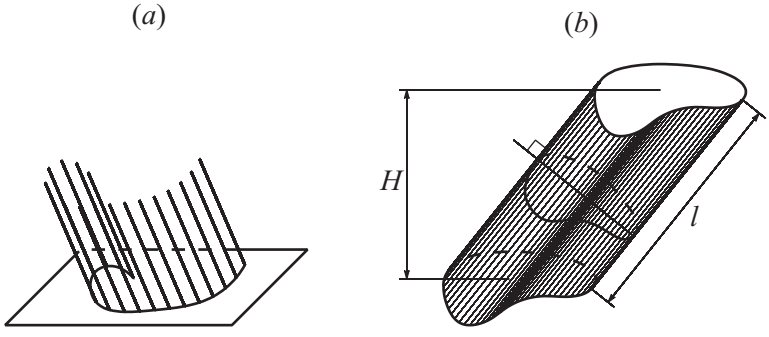


Figure 3.32. A cylindrical surface (a). A cylinder (b).

the generator, then the lateral surface area S_{lat} and the volume V of the cylinder are given by the formulas

$$\begin{aligned} S_{\text{lat}} &= PH = P_{\text{sec}}l, \\ V &= S_{\text{bas}}H = S_{\text{sec}}l. \end{aligned} \tag{3.2.3.1}$$

In a *right* cylinder, the bases are perpendicular to the generator. In particular, if the bases are disks, then one speaks of a *right circular cylinder*. The volume, the lateral surface area, and the total surface area of a right circular cylinder are given by the formulas

$$\begin{aligned} V &= \pi R^2 H, \\ S_{\text{lat}} &= 2\pi RH, \\ S &= 2\pi R(R+H), \end{aligned} \tag{3.2.3.2}$$

where R is the radius of the base.

A right circular cylinder is also called a *round cylinder*, or simply a *cylinder*.

2°. The part of a cylinder cut by a plane nonparallel to the base is called a *truncated cylinder* (Fig. 3.33a).

The volume, the lateral surface area, and the total surface area of a truncated cylinder

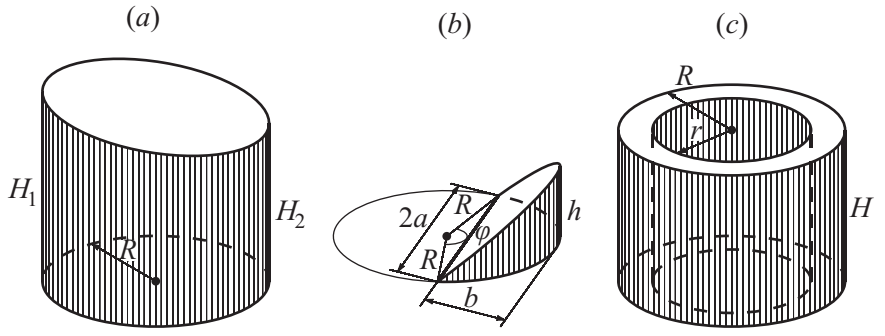


Figure 3.33. A truncated cylinder (*a*), a "hoof" (*b*), and a cylindrical tube (*c*).

are given by the formulas

$$\begin{aligned} V &= \pi R^2 \frac{H_1 + H_2}{2}, \\ S_{\text{lat}} &= \pi R(H_1 + H_2), \\ S &= \pi R\left[H_1 + H_2 + R + \sqrt{R^2 + \left(\frac{H_2 - H_1}{2}\right)^2}\right], \end{aligned} \tag{3.2.3.3}$$

where H_1 and H_2 are the maximal and minimal generators.

3°. A *segment of a round cylinder (a "hoof")* is a portion of the cylinder cut by a plane that is nonparallel to the base and intersects it. If R is the radius of the cylindrical segment, h is the height of the "hoof," and b is its width (for the other notation, see Fig. 3.33*b*), then the volume V and the lateral surface area S_{lat} of the "hoof" can be determined by the formulas

$$\begin{aligned} V &= \frac{h}{3b}\left[a(3R^2 - a^2) + 3R^2(b - R)\alpha\right] = \frac{hR^3}{b}\left(\sin\alpha - \frac{\sin^3\alpha}{3} - \alpha\cos\alpha\right), \\ S_{\text{lat}} &= \frac{2\pi R}{b}[(b - R)\alpha + a], \end{aligned} \tag{3.2.3.4}$$

where $\alpha = \frac{1}{2}\varphi$ is measured in radians.

4°. A solid bounded by two closed cylindrical surfaces and two planes is called a *cylindrical tube*; the planes are called the bases of the tube. The volume of a round cylindrical tube (Fig. 3.33*c*) is

$$V = \pi H(R^2 - r^2) = \pi H\delta(2R - r) = \pi H\delta(2r + \delta) = 2\pi H\delta\rho, \tag{3.2.3.5}$$

where R and r are the outer and inner radii, $\delta = R - r$ is the thickness, $\rho = \frac{1}{2}(R + r)$ is the midradius, and H is the height of the pipe.

3.2.3-2. Conical surface. Cone. Frustum of cone.

A *conical surface* is the union of straight lines (*generators*) passing through a fixed point (the *apex*) in space and any point of some space curve (the *directrix*) (Fig. 3.34*a*).

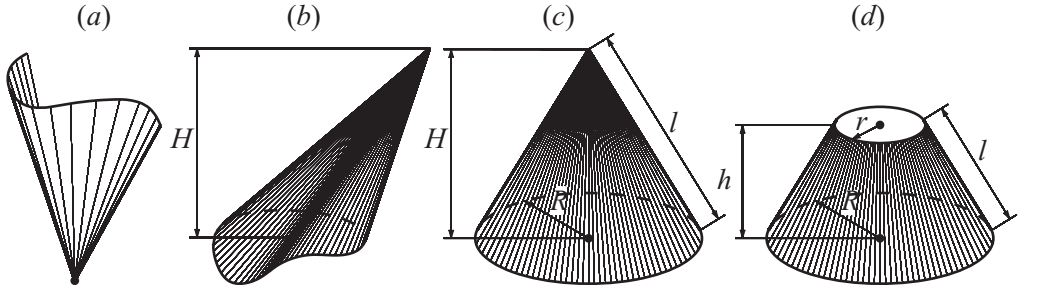


Figure 3.34. Conical surface (*a*). A cone (*b*), a right circular cone (*c*), and a frustum of a cone (*d*).

1°. A solid bounded by a conical surface with closed directrix and a plane is called a *cone*; the plane is the base of the cone (Fig. 3.34*b*). The volume of an arbitrary cone is given by the formula

$$V = \frac{1}{3} H S_{\text{bas}}, \tag{3.2.3.6}$$

where H is the altitude of the cone and S_{bas} is the area of the base.

A *right circular cone* (Fig. 3.34*c*) has a disk as the base, and its vertex is projected onto the center of the disk. If l is the length of the generator and R is the radius of the base, then the volume, the lateral surface area, and the total surface area of the right circular cone are given by the formulas

$$\begin{aligned} V &= \frac{1}{3}\pi R^2 H, \\ S_{\text{lat}} &= \pi R l = \pi R\sqrt{R^2 + H^2}, \\ S &= \pi R(R + l). \end{aligned} \tag{3.2.3.7}$$

2°. If a cone is cut by a plane parallel to the base, then we obtain a *frustum of a cone* (Fig. 3.34*d*). The length l of the generator, the volume V, the lateral surface area S_{lat}, and the total surface area S of the frustum of a right circular cone are given by the formulas

$$\begin{aligned} l &= \sqrt{h^2 + (R - r)^2}, \\ V &= \frac{\pi h}{3}(R^2 + r^2 + Rr), \\ S_{\text{lat}} &= \pi l(R + r), \\ S &= \pi[l(R + r) + R^2 + r^2], \end{aligned} \tag{3.2.3.8}$$

where r is the radius of the upper base and h is the altitude of the frustum of a cone.

3.2.3-3. Sphere. Spherical parts. Torus.

1°. The *sphere of radius R centered at O* is the set of points in space at the distance R from the point O (Fig. 3.35*a*). A solid bounded by a sphere is called a *ball*. Any section of the sphere by a plane is a circle. The section of the sphere by a plane passing through its center is called a *great circle* of radius R. There exists exactly one great circle passing through two arbitrary points on the sphere that are not antipodal (i.e., are not the opposite endpoints of a diameter). The smaller arc of this great circle is the shortest distance on the sphere between these points. Concerning the geometry of the sphere, see Section 3.3. The

surface area S of the sphere and the volume V of the ball bounded by the sphere are given by the formulas

$$S = 4\pi R = \pi D^2 = \sqrt[3]{36\pi V^2},$$
$$V = \frac{4\pi R^3}{3} = \frac{\pi D^3}{6} = \frac{1}{6}\sqrt{\frac{S^3}{\pi}}, \tag{3.2.3.9}$$

where $D = 2R$ is the *diameter* of the sphere.

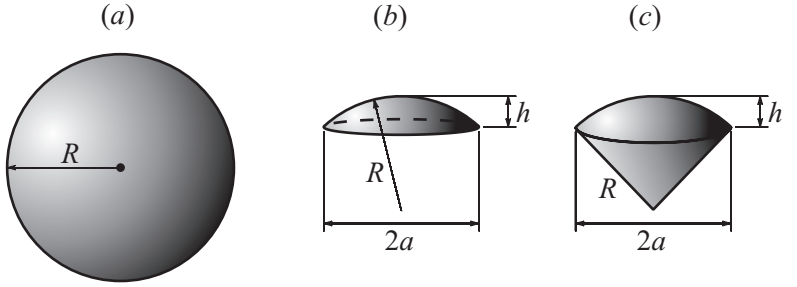


Figure 3.35. A sphere (a), a spherical cap (b), and a spherical sector (c).

2°. A portion of a ball cut from it by a plane is called a *spherical cap* (Fig. 3.35b). The width a, the area S_{lat} of the curved surface, the total surface area S, and the volume V of a spherical cap can be found from the formulas

$$\begin{aligned} &a^2 = h(2R - h),\\ &S_{\text{lat}} = 2\pi Rh = \pi(a^2 + h^2),\\ &S = S_{\text{lat}} + \pi a^2 = \pi(2Rh + a^2) = \pi(h^2 + 2a^2),\\ &V = \frac{\pi h}{6}(3a^2 + h^2) = \frac{\pi h^2}{3}(3R - h), \end{aligned} \tag{3.2.3.10}$$

where R and h are the radius and the height of the spherical cap.

3°. A portion of a ball bounded by the curved surface of a spherical cap and the conical surface whose base is the base of the cap and whose vertex is the center of the ball is called a *spherical sector* (Fig. 3.35c). The total surface area S and the volume V of a spherical sector are given by the formulas

$$\begin{aligned} &S = \pi R(2h + a),\\ &V = \frac{2}{3}\pi R^2 h, \end{aligned} \tag{3.2.3.11}$$

where a is the width, h is the height, and R is the radius of the sector.

4°. A portion of a ball contained between two parallel plane secants is called a *spherical segment* (Fig. 3.36a). The curved surface of a spherical segment is called a *spherical zone*, and the plane circular surfaces are the *bases* of a spherical segment. The radius R of the ball, the radii a and b of the bases, and the height h of a spherical segment satisfy the relation

$$R^2 = a^2 + \left(\frac{a^2 - b^2 - h^2}{2h}\right)^2. \tag{3.2.3.12}$$

The curved surface area S_{lat}, the total surface area S, and the volume V of a spherical segment are given by the formulas

$$\begin{aligned} S_{\text{lat}} &= 2\pi Rh, \\ S &= S_{\text{lat}} + \pi(a^2 + b^2) = \pi(2Rh + a^2 + b^2), \\ V &= \frac{\pi h}{6}(3a^2 + 3b^2 + h^2). \end{aligned} \tag{3.2.3.13}$$

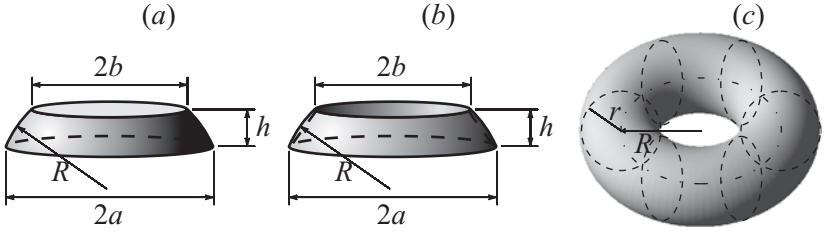


Figure 3.36. A spherical segment (*a*) and a spherical segment without the truncated cone inscribed in it (*b*). A torus (*c*).

If V_1 is the volume of the truncated cone inscribed in a spherical segment (Fig. 3.36*b*) and l is the length of its generator, then

$$V - V_1 = \frac{\pi h l^2}{6}. \tag{3.2.3.14}$$

4°. A *torus* is a surface generated by revolving a circle about an axis coplanar with the circle but not intersecting it. If the directrix is a circle (Fig. 3.36*c*), the radius R of the directrix is not less than the radius r of the generating circle ($R \geq r$), and the center of the generator moves along the directrix, then the surface area and the volume of the torus are given by the formulas

$$\begin{aligned} S &= 4\pi^2 Rr = \pi^2 Dd, \\ V &= 2\pi^2 Rr^2 = \frac{\pi^2 Dd^2}{4}, \end{aligned} \tag{3.2.3.15}$$

where $D = 2R$ and $d = 2r$ are the diameters of the generator and the directrix.

3.3. Spherical Trigonometry

3.3.1. Spherical Geometry

3.3.1-1. Great circle.

A *great circle* is a section of a sphere by a plane passing through the center.

Properties of great circles:

1. The radius of a great circle is equal to the radius of the sphere.
2. There is only one great circle through two arbitrary points that are not the opposite endpoints of a diameter.

The smaller arc of the great circle through two given points is called a *geodesic*, and the length of this arc is the *shortest distance* on the sphere between the two points. The great circles on the sphere play a role similar to the role of straight lines on the plane.

Any two points on the sphere determine a pencil of planes. The intersection of each plane in the pencil with the sphere is a circle. If two points are not the opposite endpoints of a diameter, then the plane passing through the center of the sphere determines the largest circle in the pencil, which is a great circle. The other circles are called *small circles*; the intersection with the sphere of the plane perpendicular to the plane containing the great circle is the smallest circle.

3.3.1-2. Measurement of arcs and angles on sphere. Spherical biangles.

The distances on the sphere are measured along great circle arcs. The great circle arc length between points A and B is given by the relation

$$\overset{\smile}{AB} = R\alpha, \tag{3.3.1.1}$$

where R is the radius of the sphere and α is the corresponding central angle (in radians). If only the unit sphere (the radius $R = 1$) is considered, then each great circle arc can be characterized by the corresponding central angle (in radians). The angle between two intersecting great circle arcs is measured by the linear angle between the tangents to the great circles at the point of intersection or, which is the same, by the dihedral angle between the planes of the great circles.

Two intersecting great circles on the sphere form four *spherical biangles*. The area of a spherical biangle with the angle α is given by the formula

$$S = 2R^2\alpha. \tag{3.3.1.2}$$

3.3.2. Spherical Triangles

3.3.2-1. Basic notions and properties.

A figure formed by three great circle arcs pairwise connecting three arbitrary points on the sphere is called a *spherical triangle* (Fig. 3.37*a*). The *vertices* of a spherical triangle are the points of intersection of three rays issuing from the center of the sphere with the sphere. The angles less than π between the rays are called the sides a, b, and c of a spherical triangle. Such spherical triangles are called *Euler triangles*. To each side of a triangle there corresponds a great circle arc on the sphere. The angles α, β, and γ opposite the sides a, b, and c of a spherical triangle are the angles between the great circle arcs corresponding to the sides of the triangle, or, equivalently, the angles between the planes determined by these rays.

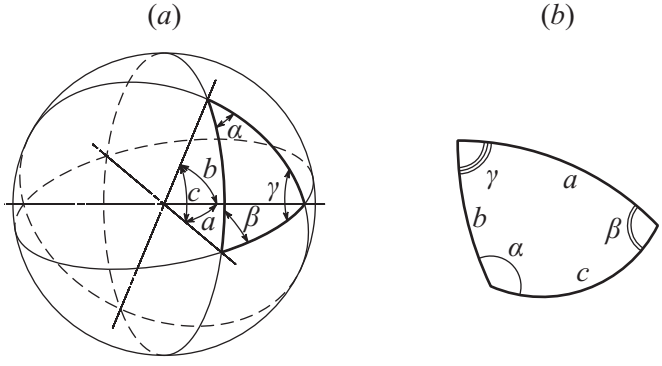


Figure 3.37. A spherical triangle.

By analogy with the circumcircle of a plane triangle, there is a "circumscribed cone of revolution" that contains the three straight lines determining the triangle; the axis of this cone is the intersection of the planes perpendicular to the sides at their midpoints. There also exists an "inscribed cone of revolution" that is tangent to the three planes corresponding to the spherical triangle; the axis of this cone is the intersection of the angle bisector planes. The "circumradius" R and the "inradius" r are defined as the angles equal to half the angles at the vertices of the first and the second cone, respectively.

If R is the radius of the sphere, then the area S of the spherical triangle is given by the formula

$$S = R^2\varepsilon, \tag{3.3.2.1}$$

where ε is the *spherical excess* defined as

$$\varepsilon = \alpha + \beta + \gamma - \pi \tag{3.3.2.2}$$

and measured in radians.

A spherical triangle is uniquely determined (up to a symmetry transformation) by:

1. Three sides.
2. Three angles.
3. Two sides and their included angle.
4. Two angles and their included side.

Let α, β, and γ be the angles and a, b, and c the sides opposite these angles in a spherical triangle (Fig. 3.37*b*). Table 3.5 presents the basic properties and relations characterizing spherical triangles (with the notation $2p = a + b + c$ and $2P = \alpha + \beta + \gamma - \pi$). From the relations given in Table 3.5, one can derive all missing relations by cyclically permuting the sides a, b, and c and the angles α, β, and γ.

LEGENDRE'S THEOREM. *The area of a spherical triangle with small sides (i.e., with sides that are small compared with the radius of the sphere) is approximately equal to the area of a plane triangle with the same sides; the difference between each angle of the plane triangle and the corresponding angle of the spherical triangle is approximately equal to one-third of the spherical excess.*

The law of sines, the law of cosines, and the half-angle theorem in spherical trigonometry for small sides become the corresponding theorems of the linear (plane) trigonometry.

Table 3.6 allows one to find the sides and angles of an arbitrary spherical triangle if three appropriately chosen sides and/or angles are given.

3.3.2-2. Rectangular spherical triangle.

A spherical triangle is said to be *rectangular* if at least one of its angles, for example, γ, is equal to $\frac{1}{2}\pi$ (Fig. 3.38*a*); the opposite side c is called the *hypotenuse*.

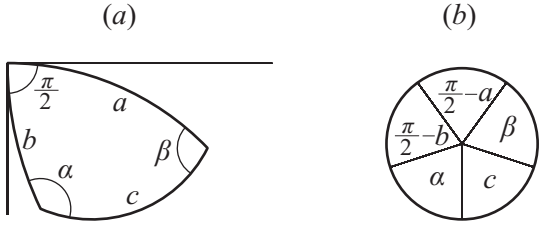


Figure 3.38. A rectangular spherical triangle (*a*). The Neper rules (*b*).

TABLE 3.5
Basic properties and relations characterizing spherical triangles

No.	The name of property	Properties and relations
1	Triangle inequality	The sum of lengths of two sides is greater than the length of the third side. The absolute value of the difference between the lengths of two sides is less than the length of the third side, $a+b>c, \quad \lvert a-b\rvert<c$
2	Sum of two angles of a triangle	The sum of two angles of a triangle is greater than the third angle increased by π, $\alpha+\beta<\pi+\gamma$
3	The greatest side and the greatest angle	The greatest side is opposite the greatest angle, $a<b \quad \text{if} \quad \alpha<\beta$; $a=b \quad \text{if} \quad \alpha=\beta$
4	Sum of angles of a triangle	The sum of the angles lies between π and 3π, $\pi<\alpha+\beta+\gamma<3\pi$
5	Sum of sides of a triangle	The sum of sides lies between 0 and 2π $0<a+b+c<2\pi$
6	The law of sines	$\dfrac{\sin a}{\sin\alpha}=\dfrac{\sin b}{\sin\beta}=\dfrac{\sin c}{\sin\gamma}$
7	The law of cosines of sides	$\cos c=\cos a\cos b+\sin a\sin b\cos\gamma$
8	The law of cosines of angles	$\cos\gamma=-\cos\alpha\cos\beta+\sin\alpha\sin\beta\cos c$
9	Half-angle formulas	$\sin\dfrac{\gamma}{2}=\sqrt{\dfrac{\sin(p-a)\sin(p-b)}{\sin a\sin b}}, \quad \cos\dfrac{\gamma}{2}=\sqrt{\dfrac{\sin p\sin(p-c)}{\sin a\sin b}}, \quad \tan\dfrac{\gamma}{2}=\sqrt{\dfrac{\sin(p-a)\sin(p-b)}{\sin p\sin(p-c)}}$
10	Half-side theorem	$\sin\dfrac{c}{2}=\sqrt{\dfrac{-\sin P\sin(P-\gamma)}{\sin\alpha\sin\beta}}, \quad \cos\dfrac{c}{2}=\sqrt{\dfrac{\sin(P-\alpha)\sin(P-\beta)}{\sin\alpha\sin\beta}}, \quad \tan\dfrac{c}{2}=\sqrt{\dfrac{-\sin P\sin(P-\gamma)}{\sin(P-\alpha)\sin(P-\beta)}}$
11	Neper's analogs	$\tan\dfrac{c}{2}\cos\dfrac{\alpha-\beta}{2}=\tan\dfrac{a+b}{2}\cos\dfrac{\alpha+\beta}{2}$, $\tan\dfrac{c}{2}\sin\dfrac{\alpha-\beta}{2}=\tan\dfrac{a-b}{2}\sin\dfrac{\alpha+\beta}{2}$, $\cot\dfrac{\gamma}{2}\cos\dfrac{a-b}{2}=\tan\dfrac{\alpha+\beta}{2}\cos\dfrac{a+b}{2}$, $\cot\dfrac{\gamma}{2}\sin\dfrac{a-b}{2}=\tan\dfrac{\alpha-\beta}{2}\sin\dfrac{a+b}{2}$
12	D'Alembert (Gauss) formulas	$\sin\dfrac{\gamma}{2}\sin\dfrac{a+b}{2}=\sin\dfrac{c}{2}\cos\dfrac{\alpha-\beta}{2}, \quad \sin\dfrac{\gamma}{2}\sin\dfrac{a+b}{2}=\cos\dfrac{c}{2}\cos\dfrac{\alpha+\beta}{2}$, $\cos\dfrac{\gamma}{2}\sin\dfrac{a-b}{2}=\sin\dfrac{c}{2}\sin\dfrac{\alpha-\beta}{2}, \quad \cos\dfrac{\gamma}{2}\cos\dfrac{a-b}{2}=\cos\dfrac{c}{2}\sin\dfrac{\alpha+\beta}{2}$
13	Product formulas	$\sin a\cos\beta=\cos b\sin c-\cos\alpha\sin b\cos c$, $\sin a\cos b=\cos\beta\sin c-\cos a\sin\beta\cos\gamma$
14	The "circumradius" $\overline{R}$	$\cot\overline{R}=\sqrt{\dfrac{\sin(P-\alpha)\sin(P-\beta)\sin(P-\gamma)}{\sin P}}=\cot\dfrac{\alpha}{2}\sin(\alpha-P)$
15	The "inradius" r	$\tan r=\sqrt{\dfrac{\sin(p-\alpha)\sin(p-\beta)\sin(p-\gamma)}{\sin p}}=\tan\dfrac{\alpha}{2}\sin(p-\alpha)$

TABLE 3.5 (*continued*)
Basic properties and relations characterizing spherical triangles

No.	The name of property	Properties and relations
16	Willier's formula for the spherical excess ε	$\tan\frac{P}{2} = \tan\frac{\varepsilon}{4} = \sqrt{\tan\frac{p}{2}\tan\frac{p-a}{2}\tan\frac{p-b}{2}\tan\frac{p-c}{2}}$
17	L'Huiller equation	$\tan\left(\frac{\gamma}{2} - \frac{\varepsilon}{4}\right) = \sqrt{\frac{\tan\frac{p-a}{2}\tan\frac{p-b}{2}}{\tan\frac{p}{2}\tan\frac{p-c}{2}}}$

TABLE 3.6.
Solution of spherical triangles

No.	Three parts specified	Formulas for the remaining parts
1	Three sides a, b, c	The angles α, β, and γ are determined by the half-angle formulas and the cyclic permutation. Remark. $0<a+b+c<2\pi$. The sum and difference of two sides are greater than the third.
2	Three angles α, β, γ	The sides a, b, and c are determined by the half-side theorems and the cyclic permutation. Remark. $\pi<\alpha+\beta+\gamma<3\pi$. The sum of two angles is less than π plus the third angle.
3	Two sides a, b and the included angle γ	*First method.* $\alpha+\beta$ and $\alpha-\beta$ are determined from Neper's analogs, then α and β can be found; side a is determined from the law of cosines, $\sin c = \sin\gamma\frac{\sin a}{\sin\alpha}$. *Second method.* The law of cosines of sides is applied, $\cos c = \cos a\cos b + \sin a\sin b\cos\gamma$, $\cos\beta = \frac{\cos b - \sin a\sin c}{\sin a\sin c}$, $\cos\alpha = \frac{\cos a - \sin b\sin c}{\sin b\sin c}$. Remark 1. If $\gamma>\beta$ ($\gamma<\beta$), then c must be chosen so that $c>b$ ($c<b$). Remark 2. The quantities c, α, and β are determined uniquely.
4	A side c and the two angles α, β adjacent to it	*First method.* $a+b$ and $a-b$ are determined from Neper's analogs, then a and b can be found; angle γ is determined from the law of sines, $\sin\gamma = \sin c\frac{\sin\alpha}{\sin a}$. *Second method.* The law of cosines of angles is applied, $\cos\gamma = -\cos\alpha\cos\beta + \sin\alpha\sin\beta\cos c$, $\cos a = \frac{\cos\alpha + \cos\beta\cos\gamma}{\sin\beta\sin\gamma}$, $\cos b = \frac{\cos\beta + \cos\alpha\cos\gamma}{\sin\alpha\sin\gamma}$. Remark 1. If $c>b$ ($c<b$), then γ must be chosen so that $\gamma>\beta$ ($\gamma<\beta$). Remark 2. The quantities γ, a, and b are determined uniquely.
5	Two sides a, b and the angle α opposite one of them	β is determined by the law of sines, $\sin\beta = \sin\alpha\frac{\sin b}{\sin a}$. The elements c and γ can be found from Neper's analogs. Remark 1.The problem has a solution for $\sin b\sin\alpha \le \sin a$. Remark 2. Different cases are possible: 1. If $\sin a \ge \sin b$, then the solution is determined uniquely. 2. If $\sin b\sin\alpha < \sin a$, then there are two solutions β_1 and β_2, $\beta_1+\beta_2=\pi$. 3. If $\sin b\sin\alpha = \sin a$, then the solution is unique: $\beta = \frac{1}{2}\pi$.
6	Two angles α, β and the side a opposite one of them	b is determined by the law of sines, $\sin b = \sin a\frac{\sin\beta}{\sin\alpha}$. The elements c and γ can be found from Neper's analogs. Remark 1. The problem has a solution for $\sin a\sin\beta \le \sin\alpha$. Remark 2. Different cases are possible: 1. If $\sin\alpha \ge \sin\beta$, then the solution is determined uniquely. 2. If $\sin\beta\sin\alpha < \sin a$, then there are two solutions b_1 and b_2, $b_1+b_2=\pi$. 3. If $\sin\beta\sin\alpha = \sin a$, then the solution is unique: $b = \frac{1}{2}\pi$.

The following basic relations hold for spherical triangles:

$$\begin{aligned}
&\sin a = \cos\left(\tfrac{\pi}{2} - a\right) = \sin\alpha \sin c = \cot\left(\tfrac{\pi}{2} - b\right) \cot\beta = \tan b \cot\beta,\\
&\sin b = \cos\left(\tfrac{\pi}{2} - b\right) = \sin\beta \sin c = \cot\left(\tfrac{\pi}{2} - a\right) \cot\alpha = \tan a \cot\alpha,\\
&\cos c = \sin\left(\tfrac{\pi}{2} - a\right) \sin\left(\tfrac{\pi}{2} - b\right) = \cos a \cos b = \cot\alpha \cot\beta,\\
&\cos\alpha = \sin\left(\tfrac{\pi}{2} - a\right) \sin\beta = \cos a \sin\beta = \cot\left(\tfrac{\pi}{2} - b\right) \cot c = \tan b \cot c,\\
&\cos\beta = \sin\left(\tfrac{\pi}{2} - b\right) \sin\alpha = \cos b \sin\alpha = \cot\left(\tfrac{\pi}{2} - a\right) \cot c = \tan a \cot c,
\end{aligned} \tag{3.3.2.3}$$

which can be obtained from the *Neper rules*: if the five parts of a spherical triangle (the right angle being omitted) are written in the form of a circle in the order in which they appear in the triangle and the legs a and b are replaced by their complements to $\frac{1}{2}\pi$ (Fig. 3.38*b*), then the cosine of each part is equal to the product of sines of the two parts not adjacent to it, as well as to the product of the cotangents of the two parts adjacent to it.

References for Chapter 3

Alexander, D. C. and Koeberlein, G. M., *Elementary Geometry for College Students, 3rd Edition*, Houghton Mifflin Company, Boston, 2002.

Alexandrov, A. D., Verner, A. L., and Ryzhik, B. I., *Solid Geometry [in Russian]*, Alpha, Moscow, 1998.

Chauvenet, W., *A Treatise on Elementary Geometry*, Adamant Media Corporation, Boston, 2001.

Fogiel, M. (Editor), *High School Geometry Tutor, 2nd Edition*, Research & Education Association, Englewood Cliffs, New Jersey, 2003.

Gustafson, R. D. and Frisk, P. D., *Elementary Geometry, 3rd Edition*, Wiley, New York, 1991.

Hadamard, J., *Leçons de géométrie élémentaire, Vols 1 and 2, Rep. Edition*, H.-C. Hege and K. Polthier (Editors), Editions J. Gabay, Paris, 1999.

Hartshorne, R., *Geometry: Euclid and Beyond*, Springer, New York, 2005.

Jacobs, H. R., *Geometry, 2nd Edition*, W. H. Freeman & Company, New York, 1987.

Jacobs, H. R., *Geometry: Seeing, Doing, Understanding, 3rd Edition*, W. H. Freeman & Company, New York, 2003.

Jurgensen, R. and Brown, R. G., *Geometry*, McDougal Littell/Houghton Mifflin, Boston, 2000.

Kay, D., *College Geometry: A Discovery Approach, 2nd Edition*, Addison Wesley, Boston, 2000.

Kiselev, A. P., *Plain and Solid Geometry [in Russian]*, Fizmatlit Publishers, Moscow, 2004.

Leff, L. S., *Geometry the Easy Way, 3rd Edition*, Barron's Educational Series, Hauppauge, New York, 1997.

Moise, E., *Elementary Geometry from an Advanced Standpoint, 3rd Edition*, Addison Wesley, Boston, 1990.

Musser, G. L., Burger, W. F., and Peterson, B. E., *Mathematics for Elementary Teachers: A Contemporary Approach, 6th Edition*, Wiley, New York, 2002.

Musser, G. L., Burger, W. F., and Peterson, B. E., *Essentials of Mathematics for Elementary Teachers: A Contemporary Approach, 6th Edition*, Wiley, New York, 2003.

Musser, G. L. and Trimpe, L. E., *College Geometry: A Problem Solving Approach with Applications*, Prentice Hall, Englewood Cliffs, New Jersey, 1994.

Pogorelov, A. V., *Elementary Geometry [in Russian]*, Nauka Publishers, Moscow, 1977.

Pogorelov, A., *Geometry*, Mir Publishers, Moscow, 1987.

Prasolov, V. V., *Problems in Plane Geometry [in Russian]*, MTsNMO, Moscow, 2001.

Prasolov, V. V. and Tikhomirov, V. M., *Geometry, Translations of Mathematical Monographs, Vol. 200*, American Mathematical Society, Providence, Rhode Island, 2001.

Roe, J., *Elementary Geometry*, Oxford University Press, Oxford, 1993.

Schultze, A. and Sevenoak, F. L., *Plane and Solid Geometry*, Adamant Media Corporation, Boston, 2004.

Tussy, A. S., *Basic Geometry for College Students: An Overview of the Fundamental Concepts of Geometry, 2nd Edition*, Brooks Cole, Stamford, 2002.

Vygodskii, M. Ya., *Mathematical Handbook: Elementary Mathematics, Rev. Edition*, Mir Publishers, Moscow, 1972.

Chapter 4
Analytic Geometry

4.1. Points, Segments, and Coordinates on Line and Plane

4.1.1. Coordinates on Line

4.1.1-1. Axis and segments on axis.

A straight line on which a sense is chosen is called an *axis*. If an axis is given and a *scale segment*, i.e., a linear unit used to measure any segment of the axis, is indicated, then the *segment length* is defined (see Fig. 4.1).

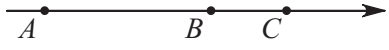


Figure 4.1. Axis.

A segment bounded by points A and B is called a *directed segment* if its initial point and endpoint are chosen. Such a segment with initial point A and endpoint B is denoted by $\overrightarrow{AB}$. Directed segments are usually called simply "segments" for brevity.

The *value of a segment* $\overrightarrow{AB}$ of some axis is defined as the number AB equal to its length taken with the plus sign if the senses of the interval and the axis coincide, and with the minus sign if the senses are opposite. Obviously, the length of a segment is its absolute value. The segment length is usually denoted by the symbol $|AB|$. It follows from the above that

$$AB = -BA, \qquad |AB| = |BA|. \tag{4.1.1.1}$$

Main identity. For any arbitrary arrangement of points A, B, and C on the axis, the values of the segments $\overrightarrow{AB}$, $\overrightarrow{BC}$, and $\overrightarrow{AC}$ satisfy the relation

$$AB + BC = AC. \tag{4.1.1.2}$$

4.1.1-2. Coordinates on line. Number axis.

One says that a *coordinate system* is introduced on an axis if there is a one-to-one correspondence between points of the axis and numbers.

Suppose that a sense, a scale segment, and a point O called the *origin* are chosen on a line. The value of a segment $\overrightarrow{OA}$ is called the *coordinate of the point* A on the axis. It is usually denoted by the letter x. The coordinates of different points are usually denoted by subscripts; for example, the coordinates of points $A_1, \ldots, A_n$ are $x_1, \ldots, x_n$. The point A_n with coordinate x_n is denoted by $A_n(x_n)$. An axis with a coordinate system on it is called a *number axis*.

4.1.1-3. Distance between points on axis.

The value A_1A_2 of the segment $\overrightarrow{A_1A_2}$ on an axis is equal to the difference between the coordinate x_2 of the endpoint and the coordinate x_1 of the initial point:

$$A_1A_2 = x_2 - x_1. \tag{4.1.1.3}$$

The distance d between two arbitrary points $A_1(x_1)$ and $A_2(x_2)$ on the line is given by the relation

$$d = |A_1A_2| = |x_2 - x_1|. \tag{4.1.1.4}$$

Remark. If segments do not lie on some axis but are treated as arbitrary segments on the plane or in space, then there is no reason to assign any sign to their lengths. In such cases, the symbol of absolute value is usually omitted in the notation of lengths of segments. We adopt this convention in the sequel.

4.1.2. Coordinates on Plane

4.1.2-1. Rectangular Cartesian coordinates on plane.

If a one-to-one correspondence between points on the plane and numbers (pairs of numbers) is specified, then one says that a *coordinate system* is introduced on the plane.

A *rectangular Cartesian coordinate system* is determined by a scale segment for measuring lengths and two mutually perpendicular axes. The point of intersection of the axes is usually denoted by the letter O and is called the *origin*, while the axes themselves are called the *coordinate axes*. As a rule, one of the coordinate axes is horizontal and the right sense is positive. This axis is called the *abscissa axis* and is denoted by the letter X or by OX. On the vertical axis, which is called the *ordinate axis* and is denoted by Y or OY, the upward sense is usually positive (see Fig. 4.2*a*). The coordinate system introduced above is often denoted by XY or OXY.

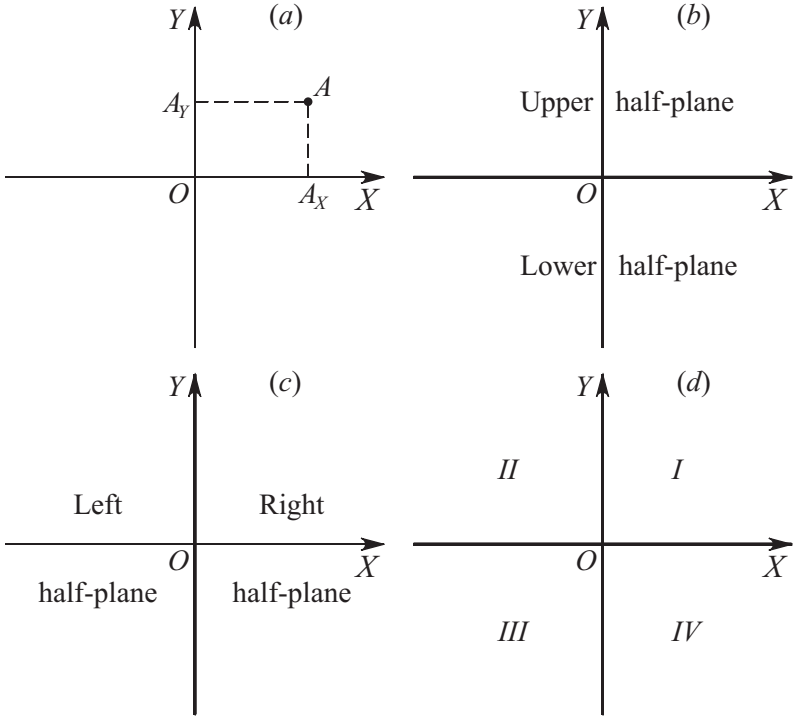


Figure 4.2. A rectangular Cartesian coordinate system.

The abscissa axis divides the plane into the *upper* and *lower* half-planes (see Fig. 4.2*b*), while the ordinate axis divides the plane into the *right* and *left* half-planes (see Fig. 4.2*c*). The two coordinate axes divide the plane into four parts, which are called *quadrants* and numbered as shown in Fig. 4.2*d*.

Take an arbitrary point A on the plane and project it onto the coordinate axes, i.e., draw perpendiculars to the axes OX and OY through A. The points of intersection of the perpendiculars with the axes are denoted by A_X and A_Y, respectively (see Fig. 4.2*a*). The numbers

$$x = OA_X, \qquad y = OA_Y, \tag{4.1.2.1}$$

where OA_X and OA_Y are the respective values of the segments $\overrightarrow{OA}_X$ and $\overrightarrow{OA}_Y$ on the abscissa and ordinate axes, are called the *coordinates of the point* A in the rectangular Cartesian coordinate system. The number x is the first coordinate, or the *abscissa*, of the point A, and y is the second coordinate, or the *ordinate*, of the point A. One says that the point A has the coordinates (x, y) and uses the notation $A(x, y)$.

Example 1. Let A be an arbitrary point in the right half-plane. Then the segment $\overrightarrow{OA}_X$ has the positive sense on the axis OX, and hence the abscissa $x = OA_X$ of A is positive. But if A lies in the left half-plane, then the segment A_X has the negative sense on the axis OX, and the number $x = OA_X$ is negative. If the point A lies on the axis OY, then its projection on the axis OX coincides with the point O and $x = OA_X = 0$.

Thus all points in the right half-plane have positive abscissas ($x > 0$), all points in the left half-plane have negative abscissas ($x < 0$), and the abscissas of points lying on the axis OY are zero ($x = 0$).

Similarly, all points in the upper half-plane have positive ordinates ($y > 0$), all points in the lower half-plane have negative ordinates ($y < 0$), and the ordinates of points lying on the axis OX are zero ($y = 0$).

Remark 1. Strictly speaking, the coordinate system introduced above is a *right rectangular Cartesian coordinate system*. A *left rectangular Cartesian coordinate system* can, for example, be obtained by changing the sense of one of the axes. There also exist right and left *oblique Cartesian coordinate systems*, where the coordinate axes intersect at an arbitrary angle.

Remark 2. A right rectangular Cartesian coordinate system is usually called simply a Cartesian coordinate system.

4.1.2-2. Transformation of Cartesian coordinates under parallel translation of axes.

Suppose that two rectangular Cartesian coordinate systems OXY and $\widehat{O}\widehat{X}\widehat{Y}$ are given and the first system is taken to the second by the translation of the origin O of the first system to the origin $\widehat{O}$ of the second system. Under this translation, the axes preserve their directions (the respective axes of the systems are parallel), and the origin moves by x_0 in the direction of the OX-axis and by y_0 in the direction of the OY-axis (see Fig. 4.3*a*). Obviously, the point $\widehat{O}$ has the coordinates (x_0, y_0) in the coordinate system OXY.

Let an arbitrary point A have coordinates (x, y) in the system OXY and coordinates $(\hat{x}, \hat{y})$ in the system $\widehat{O}\widehat{X}\widehat{Y}$. The transformation of rectangular Cartesian coordinates by the parallel translation of the axes is given by the formulas

$$\begin{aligned} x &= \hat{x} + x_0, \\ y &= \hat{y} + y_0 \end{aligned} \qquad \text{or} \qquad \begin{aligned} \hat{x} &= x - x_0, \\ \hat{y} &= y - y_0. \end{aligned} \tag{4.1.2.2}$$

4.1.2-3. Transformation of Cartesian coordinates under rotation of axes.

Suppose that two rectangular Cartesian coordinate systems OXY and $O\widehat{X}\widehat{Y}$ are given and the first system is taken to the second by the rotation around the point O by an angle α (see Fig. 4.3*b*).

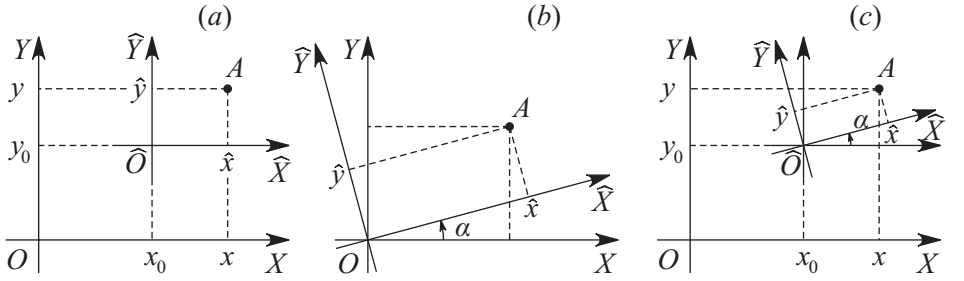


Figure 4.3. Transformation of Cartesian coordinates under parallel translation (*a*), under rotation (*b*), and under translation and rotation (*c*) of axes.

Let an arbitrary point A have coordinates (x, y) in the system OXY and coordinates $(\hat{x}, \hat{y})$ in the system $O\widehat{X}\widehat{Y}$. The transformation of rectangular Cartesian coordinates by the rotation of axes is given by the formulas

$$\begin{aligned} x &= \hat{x}\cos\alpha - \hat{y}\sin\alpha, \\ y &= \hat{x}\sin\alpha + \hat{y}\cos\alpha \end{aligned} \qquad \text{or} \qquad \begin{aligned} \hat{x} &= x\cos\alpha + y\sin\alpha, \\ \hat{y} &= -x\sin\alpha + y\cos\alpha. \end{aligned} \tag{4.1.2.3}$$

4.1.2-4. Transformation of coordinates under translation and rotation of axes.

Suppose that two rectangular Cartesian coordinate systems OXY and $\widehat{O}\widehat{X}\widehat{Y}$ are given and the first system is taken to the second by the translation of the origin $O(0, 0)$ of the first system to the origin $\widehat{O}(x_0, y_0)$ of the second system followed by the rotation of the system around the point $\widehat{O}$ by an angle α (see Fig. 4.3*c* and Paragraphs 4.1.2-2 and 4.1.2-3).

Let an arbitrary point A have coordinates (x, y) in the system OXY and coordinates $(\hat{x}, \hat{y})$ in the system $\widehat{O}\widehat{X}\widehat{Y}$. The transformation of rectangular Cartesian coordinates by the parallel translation and rotation of axes is given by the formulas

$$\begin{aligned} x &= \hat{x}\cos\alpha - \hat{y}\sin\alpha + x_0, \\ y &= \hat{x}\sin\alpha + \hat{y}\cos\alpha + y_0, \end{aligned} \qquad \text{or} \qquad \begin{aligned} \hat{x} &= (x - x_0)\cos\alpha + (y - y_0)\sin\alpha, \\ \hat{y} &= -(x - x_0)\sin\alpha + (y - y_0)\cos\alpha. \end{aligned} \tag{4.1.2.4}$$

4.1.2-5. Polar coordinates.

A *polar coordinate system* is determined by a point O called the *pole*, a ray OA issuing from this point, which is called the *polar axis*, a scale segment for measuring lengths, and the positive sense of rotation around the pole. Usually, the anticlockwise sense is assumed to be positive (see Fig. 4.4*a*).

The position of each point B on the plane is determined by two *polar coordinates*, the *polar radius* $\rho = |OB|$ and the *polar angle* $\theta = \angle AOB$ (the values of the angle θ are defined up to the addition of $\pm 2\pi n$, where n is an integer). To be definite, one usually assumes that $0 \le \theta \le 2\pi$ or $-\pi \le \theta \le \pi$. The polar radius of the pole is zero, and its polar angle does not have any definite value.

4.1.2-6. Relationship between Cartesian and polar coordinates.

Suppose that B is an arbitrary point on the plane, (x, y) are its rectangular Cartesian coordinates, and (ρ, θ) are its polar coordinates (see Fig. 4.4*b*). The formulas of transformation

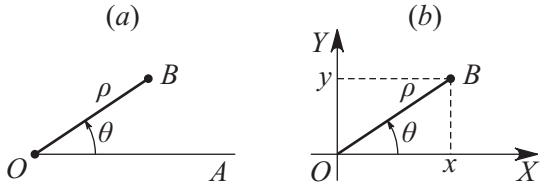


Figure 4.4. A polar coordinate system (*a*). Relationship between Cartesian and polar coordinates (*b*).

from one coordinate system to the other have the form

$$\begin{aligned} x &= \rho\cos\theta, \\ y &= \rho\sin\theta \end{aligned} \qquad \text{or} \qquad \begin{aligned} \rho &= \sqrt{x^2+y^2}, \\ \tan\theta &= y/x, \end{aligned} \tag{4.1.2.5}$$

where the polar angle θ is determined with regard to the quadrant where the point B lies.

Example 2. Let us find the polar coordinates ρ, θ $(0 \le \theta \le 2\pi)$ of the point B whose Cartesian coordinates are $x = -3$, $y = -3$.

From formulas (4.1.2.5), we obtain $\rho = \sqrt{(-3)^2 + (-3)^2} = 3\sqrt{2}$ and $\tan\theta = \frac{-3}{-3} = 1$. Since the point B lies in the third quadrant, we have $\theta = \arctan 1 + \pi = \frac{5}{4}\pi$.

4.1.3. Points and Segments on Plane

4.1.3-1. Distance between points on plane.

The distance d between two arbitrary points A_1 and A_2 on the plane is given by the formula

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \tag{4.1.3.1}$$

where x and y with the corresponding subscripts are the Cartesian coordinates of these points, and by the formula

$$d = \sqrt{\rho_1^2 + \rho_2^2 - 2\rho_1\rho_2\cos(\theta_2 - \theta_1)}, \tag{4.1.3.2}$$

where ρ and θ with the corresponding subscripts are the polar coordinates of these points.

4.1.3-2. Segment and its projections.

Suppose that an axis u and an arbitrary segment $\overrightarrow{A_1A_2}$ are given on the plane (see Fig. 4.5*a*). From the points A_1 and A_2, we draw the perpendiculars to u and denote the points of intersection of the perpendiculars with the axis by P_1 and P_2. The value P_1P_2 of the segment $\overrightarrow{P_1P_2}$ of the axis u is called the *projection of the segment* $\overrightarrow{A_1A_2}$ onto the axis u. Usually one writes $\mathrm{pr}_u\,\overrightarrow{A_1A_2} = P_1P_2$. If φ $(0 \le \varphi \le \pi)$ is the angle between the segment $\overrightarrow{A_1A_2}$ and the axis u, then

$$\mathrm{pr}_u\,\overrightarrow{A_1A_2} = d\cos\varphi. \tag{4.1.3.3}$$

For two arbitrary points $A_1(x_1, y_1)$ and $A_2(x_2, y_2)$, the projections x and y of the segment $\overrightarrow{A_1A_2}$ onto the coordinate X- and Y-axes are given by the formulas (see Fig. 4.5*b*)

$$\mathrm{pr}_X\,\overrightarrow{A_1A_2} = x_2 - x_1, \quad \mathrm{pr}_Y\,\overrightarrow{A_1A_2} = y_2 - y_1. \tag{4.1.3.4}$$

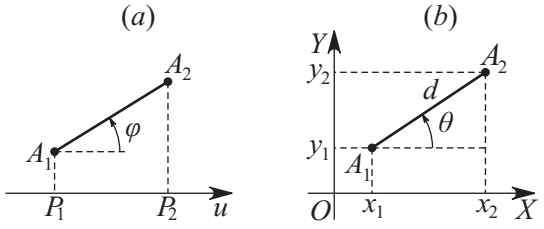


Figure 4.5. Projection of the segment onto the axis u (*a*) and onto the coordinate X- and Y-axes (*b*).

Thus, to obtain the projections of a segment onto the coordinate axes, one subtracts the coordinates of its initial point from the respective coordinates of its endpoint.

The projections of the segment $\overrightarrow{A_1A_2}$ onto the coordinate axes can be found if its length d (see (4.1.3.1)) and *polar angle* θ are known (see Fig. 4.5*b*). The corresponding formulas are

$$\mathrm{pr}_X \overrightarrow{A_1A_2} = d\cos\theta, \quad \mathrm{pr}_Y \overrightarrow{A_1A_2} = d\sin\theta, \quad \tan\theta = \frac{y_2 - y_1}{x_2 - x_1}. \tag{4.1.3.5}$$

4.1.3-3. Angles between coordinate axes and segments.

The angles $\alpha_x \equiv \theta$ and α_y between the segment $\overrightarrow{A_1A_2}$ and the coordinate x- and y-axes are determined by the expressions

$$\cos\alpha_x = \frac{x_2 - x_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}, \quad \cos\alpha_y = \frac{y_2 - y_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}, \tag{4.1.3.6}$$

and $\alpha_y = \pi - \alpha_x$.

The angle β between arbitrary segments $\overrightarrow{A_1A_2}$ and $\overrightarrow{A_3A_4}$ joining the points $A_1(x_1, y_1)$, $A_2(x_2, y_2)$ and $A_3(x_3, y_3)$, $A_4(x_4, y_4)$, respectively, can be found from the relation

$$\cos\beta = \frac{(x_2 - x_1)(x_4 - x_3) + (y_2 - y_1)(y_4 - y_3)}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}\sqrt{(x_4 - x_3)^2 + (y_4 - y_3)^2}}. \tag{4.1.3.7}$$

4.1.3-4. Division of segment in given ratio.

The number $\lambda = p/q$, where $p = A_1A$ and $q = AA_2$ are the values of the directed segments $\overrightarrow{A_1A}$ and $\overrightarrow{AA_2}$, is called the ratio in which point A divides the segment $\overrightarrow{A_1A_2}$. It is independent of the sense of the segment (i.e., one could use the segment $\overrightarrow{A_2A_1}$) and the scale segment. The coordinates of the point A dividing the segment $\overrightarrow{A_1A_2}$ in a ratio λ are given by the formulas

$$x = \frac{x_1 + \lambda x_2}{1 + \lambda} = \frac{qx_1 + px_2}{q + p}, \quad y = \frac{y_1 + \lambda y_2}{1 + \lambda} = \frac{qy_1 + py_2}{q + p}, \tag{4.1.3.8}$$

where $-\infty \le \lambda \le \infty$.

For the coordinates of the midpoint of the segment $\overrightarrow{A_1A_2}$, we have

$$x = \frac{x_1 + x_2}{2}, \quad y = \frac{y_1 + y_2}{2}; \tag{4.1.3.9}$$

i.e., each coordinate of the midpoint of a segment is equal to the half-sum of the respective coordinates of its endpoints.

4.1.3-5. Area of triangle area.

The area S_3 of the triangle with vertices A_1, A_2, and A_3 is given by the formula

$$\pm S_3 = \frac{1}{2}[(x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1)]$$
$$= \frac{1}{2}\begin{vmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1 \end{vmatrix} = \frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}, \qquad (4.1.3.10)$$

where x and y with respective subscripts are the Cartesian coordinates of the vertices, and by the formula

$$\pm S_3 = \frac{1}{2}[\rho_1\rho_2 \sin(\theta_2 - \theta_1) + \rho_2\rho_3 \sin(\theta_3 - \theta_2) + \rho_3\rho_1 \sin(\theta_1 - \theta_3)], \qquad (4.1.3.11)$$

where ρ and θ with respective subscripts are the polar coordinates of the vertices. In formulas (4.1.3.10) and (4.1.3.11), one takes the plus sign if the vertices are numbered anticlockwise (see Fig. 4.6*a*) and the minus sign otherwise.

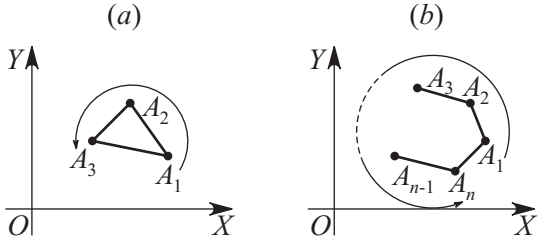


Figure 4.6. Area of triangle (*a*) and of a polygon (*b*).

4.1.3-6. Area of a polygon.

The area S_n of the polygon with vertices $A_1, \ldots, A_n$ is given by the formula

$$\pm S_n = \frac{1}{2}[(x_1 - x_2)(y_1 + y_2) + (x_2 - x_3)(y_2 + y_3) + \cdots + (x_n - x_1)(y_n + y_1)], \qquad (4.1.3.12)$$

where x and y with respective subscripts are the Cartesian coordinates of the vertices, and by the formula

$$\pm S_n = \frac{1}{2}[\rho_1\rho_2 \sin(\theta_2 - \theta_1) + \rho_2\rho_3 \sin(\theta_3 - \theta_2) + \cdots + \rho_n\rho_1 \sin(\theta_1 - \theta_n)], \qquad (4.1.3.13)$$

where ρ and θ with respective subscripts are the polar coordinates of the vertices. In formulas (4.1.3.12) and (4.1.3.13), one takes the plus sign if the vertices are numbered anticlockwise (see Fig. 4.6*b*) and the minus sign otherwise.

Remark. One often says that formulas (4.1.3.10)–(4.1.3.13) express the *oriented area* of the corresponding figures.

4.2. Curves on Plane

4.2.1. Curves and Their Equations

4.2.1-1. Basic definitions.

A *curve on the plane* determined by an equation in some coordinate system is the geometric locus of points of the plane whose coordinates satisfy this equation.

An *equation of a curve on the plane* in a given coordinate system is an equation with two variables such that the coordinates of the points lying on the curve satisfy the equation and the coordinates of the points that do not lie on the curve do not satisfy it.

The coordinates of an arbitrary point of a curve occurring in an equation of the curve are called *current coordinates*.

4.2.1-2. Equation of curve in Cartesian coordinate system.

An *equation of a curve in the Cartesian coordinate system* OXY can be written as

$$F(x, y) = 0. \tag{4.2.1.1}$$

The image of a curve determined by an equation of the form

$$y = f(x) \tag{4.2.1.2}$$

is called the *graph of the function* $f(x)$.

Example 1. Let us plot the curve determined by the equation $x^2 - y = 0$. We express one coordinate via the other (e.g., y via x) from this equation: $y = x^2$. Specifying various values of x, we find the corresponding values of y and thus construct a sequence of points of the desired curve. By joining these points, we obtain the curve itself (see Fig. 4.7*a*).

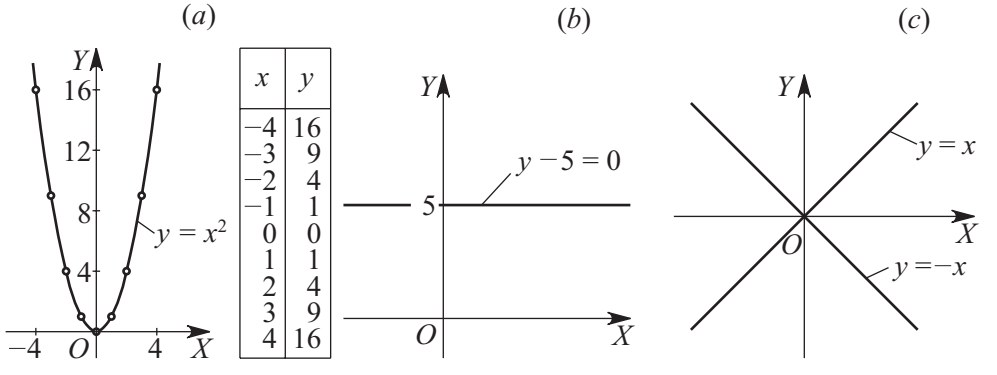


Figure 4.7. Cartesian coordinate system. Loci of points for equations $x^2 - y = 0$ (*a*), $y - 5 = 0$ (*b*), and $x^2 - y^2 = 0$ (*c*).

4.2.1-3. Special kinds of equations.

1. The equation of a curve on the plane may contain only one of the current coordinates but still determine a certain curve.

Example 2. Suppose that the equation $y-5=0$ (or $y=5$) is given. The locus of points whose coordinates are equal to five is the straight line parallel to the axis OX and passing through the point $y=5$ of the axis OY (see Fig. 4.7*b*).

Similarly, the equation $x+7=0$ determines a straight line parallel to the axis OY.

2. If the left-hand side of equation (4.2.1.1) can be factorized, then, equating each factor separately with zero, we obtain several new equations, each of which can determine a certain curve.

Example 3. Consider the equation $x^2-y^2=0$. Factorizing the left-hand side of this equation, we obtain $(x+y)(x-y)=0$. Obviously, the latter equation determines the pair of straight lines $x+y=0$ and $x-y=0$, which are the bisectors of the coordinate angles (see Fig. 4.7*c*).

3. Equation (4.2.1.1) may determine a locus consisting of one or several isolated points.

Example 4. The equation $x^2+y^2=0$ determines the single point with coordinates $(0,0)$.

Example 5. The equation $(x^2-9)^2+(y^2-25)^2=0$ defines the locus consisting of the four points $(3,5)$, $(3,-5)$, $(-3,5)$, and $(-3,-5)$.

4. There exist equations that do not determine any locus.

Example 6. The equation $x^2+y^2+5=0$ does not have solutions for any real x and y.

4.2.1-4. Equation of curve in polar coordinate system.

An *equation of a curve in a polar coordinate system* can be written as

$$\Phi(\rho,\theta)=0, \tag{4.2.1.3}$$

where ρ is the polar radius and θ is the polar angle. This equation is satisfied by the polar coordinates of any point lying on the curve and is not satisfied by the coordinates of the points that do not lie on the curve.

Example 7. Consider the equation $\rho-a\cos\theta=0$ (or $\rho=a\cos\theta$), where a is a positive number. By B we denote the point with polar coordinates (ρ,θ), and by A we denote the point with coordinates $(a,0)$. If $\rho=a\cos\theta$, then the angle OBA is a right angle, and vice versa. Therefore, the locus of points whose coordinates satisfy this equation is a circle with diameter a (see Fig. 4.8).

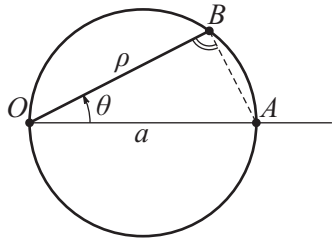


Figure 4.8. Polar coordinate system. Locus of points for equations $\rho-a\cos\theta=0$.

Example 8. Consider the equation $\rho-a\theta=0$ (or $\rho=a\theta$), where a is a positive constant. The curve determined by this equation is called a *spiral of Archimedes*.

As θ increases starting from zero, the point $B(\rho,\theta)$ issues from the pole and moves around it in the positive sense, simultaneously moving away from it. For each point of this curve with positive coordinates (ρ,θ), one has the corresponding point $(-\rho,-\theta)$ on the same curve. Figures 4.9*a* and *b* show the branches of the spiral of Archimedes corresponding to the positive and negative values of θ, respectively.

Note that the spiral of Archimedes divides each polar ray into equal segments (except for the segment nearest to the pole).

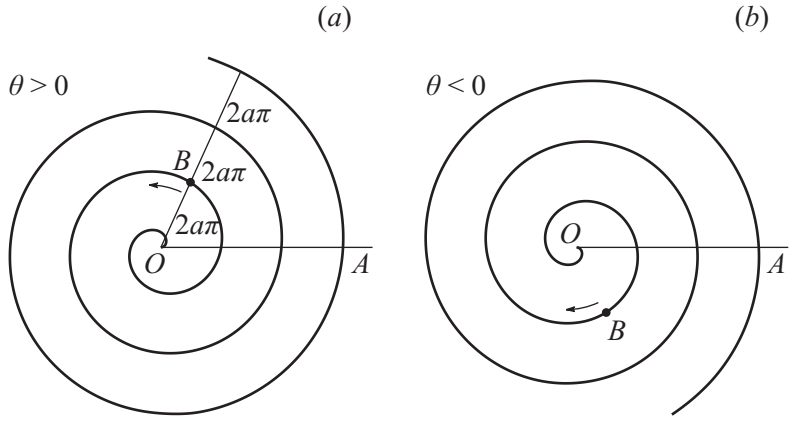


Figure 4.9. A spiral of Archimedes $\rho = a\theta$ corresponding to the positive (a) and negative (b) values of θ.

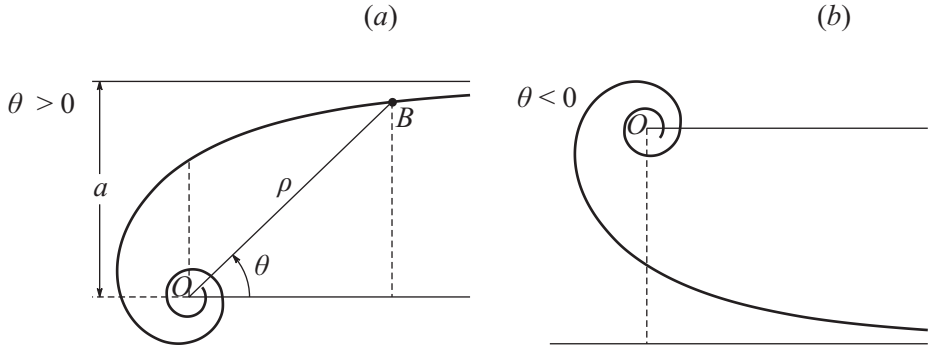


Figure 4.10. A hyperbolic spiral $\rho = a/\theta$ corresponding to the positive (a) and negative (b) values of θ.

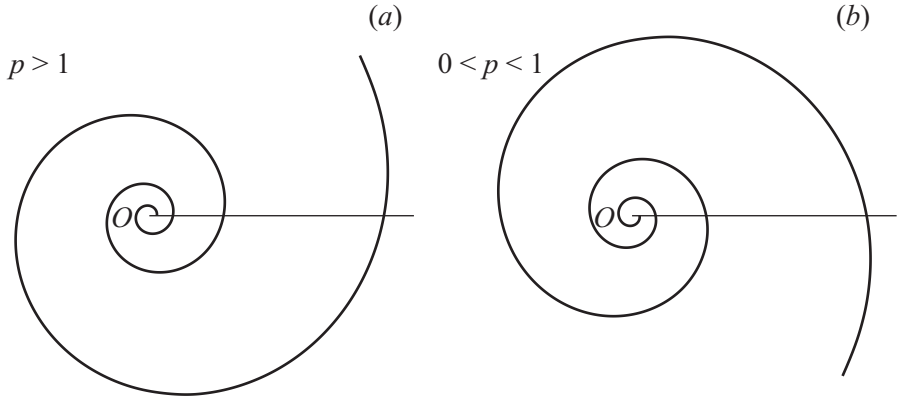


Figure 4.11. A logarithmic spiral $\rho = a^{\theta}$ corresponding to the values $a > 1$ (a) and $0 < a < 1$ (b).

Example 9. Consider the equation $\rho - a/\theta = 0$ (or $\rho = a/\theta$), where a is a positive number. The curve determined by this equation is called a *hyperbolic spiral*.

As θ increases, the point $B(\rho, \theta)$ moves around the pole in the positive sense while approaching it endlessly. As θ tends to zero, the point B approaches the line $y = a$ while moving to infinity. Figures 4.10*a* and *b* show the branches of the hyperbolic spiral corresponding to the positive and negative values of θ, respectively.

Example 10. Consider the equation $\rho - a^{\theta} = 0$ (or $\rho = a^{\theta}$), where a is a positive number. The curve determined by this equation is called the *logarithmic spiral*.

Figures 4.11*a* and *b* show the branches of the logarithmic spiral corresponding to the values $a > 1$ and $0 < a < 1$, respectively. For $a = 1$, we obtain the equation of a circle.

4.2.1-5. Parametric equations of a curve.

Parametric equations of a curve on the plane have the form

$$x = \varphi(t), \qquad y = \psi(t), \tag{4.2.1.4}$$

where x and y are treated as the coordinates of some point A for each value of the *variable parameter* t. In general, the variables x and y vary with t, and the point A moves on the plane.

Parametric equations play an important role in applied mathematics and mechanics, where they are called the *equations of motion of a mass point*. The parameter t has the meaning of time.

Remark 1. Eliminating the parameter t from equations (4.2.1.4), we obtain an equation of the curve in the form (4.2.1.1).

Remark 2. In different problems, the variable parameter in equations (4.2.1.4) may have different meanings.

Example 11. The circle of radius a centered at the origin is described by the following parametric equations for the Cartesian system:

$$x = a\cos t, \quad y = a\sin t.$$

By squaring these equations and by adding them, we obtain the equation of the circle in the form

$$x^2 + y^2 = a^2.$$

Example 12. The spiral of Archimedes (a), the hyperbolic spiral (b), and the logarithmic spiral (c) are described by the following equations in the polar coordinate system:

$$(a)\ \rho = a\theta; \qquad (b)\ \rho = \frac{a}{\theta}; \qquad (c)\ \rho = a^{\theta}.$$

The parametric equations for the Cartesian coordinates of these curves have the form

$$\begin{aligned}
&(a) \quad x = a\theta\cos\theta, \quad y = a\theta\sin\theta;\\
&(b) \quad x = \frac{a\cos\theta}{\theta}, \quad y = \frac{a\sin\theta}{\theta};\\
&(c) \quad x = a^{\theta}\cos\theta, \quad y = a^{\theta}\sin\theta.
\end{aligned}$$

In all three cases, the variable parameter is the polar angle θ.

4.2.1-6. Algebraic curves.

The curves given by algebraic equations of the form

$$\begin{aligned}
&Ax + By + C = 0,\\
&Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0,\\
&Ax^3 + Bx^2y + Cxy^2 + Dy^3 + Ex^2 + Fxy + Gy^2 + Hx + Iy + K = 0,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned} \tag{4.2.1.5}$$

in a rectangular Cartesian coordinate system are called *algebraic curves on the plane*.

A curve given by an algebraic equation of degree n in a rectangular Cartesian coordinate system is called an *nth-order algebraic curve*.

When passing from one rectangular Cartesian coordinate system to another, the degree of the equation of an algebraic curve does not change; i.e., any nth-order algebraic curve remains such in any rectangular Cartesian coordinate system.

4.2.2. Main Problems of Analytic Geometry for Curves

4.2.2-1. Construction of equation for given curve.

Suppose that a curve is defined on the plane as a locus of points and one needs to construct an equation of this curve. This is the first main problem of analytic geometry for curves.

Example 1. Suppose that a circle of radius R is given on the plane. In the rectangular Cartesian coordinate system OXY, its origin is point $A(x_0, y_0)$. It is required to derive an equation of this circle (see Fig. 4.12*a*).

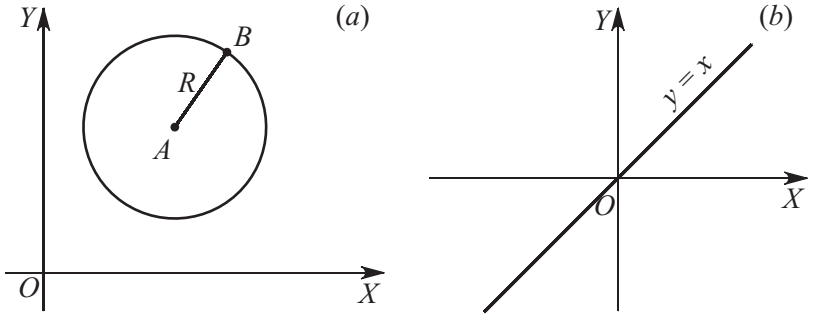


Figure 4.12. Construction of equation for given curve (a) and construction of curve with given equation (b).

By $B(x, y)$ we denote the variable point of the circle. Obviously, the circle is the locus of points lying at the distance R from the point $A(x_0, y_0)$; i.e., $AB = R$, or

$$\sqrt{(x - x_0)^2 + (y - y_0)^2} = R.$$

This is the desired equation of the circle of radius R centered at the point $A(x_0, y_0)$, which is usually written as

$$(x - x_0)^2 + (y - y_0)^2 = R^2.$$

4.2.2-2. Construction of curve with given equation.

Suppose that an equation with two variables is given and one needs to construct the curve determined by this equation on the plane. This is the second main problem of analytic geometry for curves.

Example 2. Consider the equation $x - y = 0$ or $y = x$. The points whose coordinates satisfy this equation lie in the first and third quadrants at equal distances from the coordinate axes. Thus the locus of points whose coordinates satisfy this equation is the bisector of the first and third coordinate angles (see Fig. 4.12*b*).

4.2.2-3. Intersection of two curves.

Consider two curves determined by the equations

$$F(x, y) = 0, \qquad G(x, y) = 0. \tag{4.2.2.1}$$

To find the points of intersection of these curves, one solves system (4.2.2.1). Each real solution of this system gives a point of intersection. If the system is inconsistent or does not have real solutions, then the curves do not meet.

Example 3. Let us find the points of intersection of two curves (a circle and a straight line) with the equations

$$x^2 + y^2 - 25 = 0, \qquad 2x - y + 5 = 0.$$

These equations form a system of two equations with two unknowns. For example, if we express y from the second equation and substitute the resulting expression into the first equation, then we obtain $x^2 + 4x = 0$, whence we find $x_1 = 0$, $x_2 = -4$ and $y_1 = 5$, $y_2 = -3$. This means that the curves have two points of intersection, $(0, 5)$ and $(-4, -3)$.

Remark. For equations (4.2.2.1) written, for example, in a polar coordinate system, it is natural to use ρ and θ instead of the variables x and y.

4.3. Straight Lines and Points on Plane

4.3.1. Equations of Straight Lines on Plane

4.3.1-1. Slope-intercept equation of straight line.

The tangent of the angle of inclination of a straight line to the axis OX is called the *slope* of the straight line. The slope characterizes the direction of the line. For straight lines perpendicular to the OX-axis, slope does not make sense, although one often says that the slope of such straight lines is equal to infinity.

The *slope-intercept equation of a straight line* in the rectangular Cartesian coordinate system OXY has the form

$$y = kx + b, \tag{4.3.1.1}$$

where $k = \tan\varphi = (y - b)/x$ is the slope of the line and b is the y-intercept of the line, i.e., the signed distance from the point of intersection of the line with the ordinate axis to the origin. Equation (4.3.1.1) is meaningful for any straight line that is not perpendicular to the abscissa axis (see Fig. 4.13*a*).

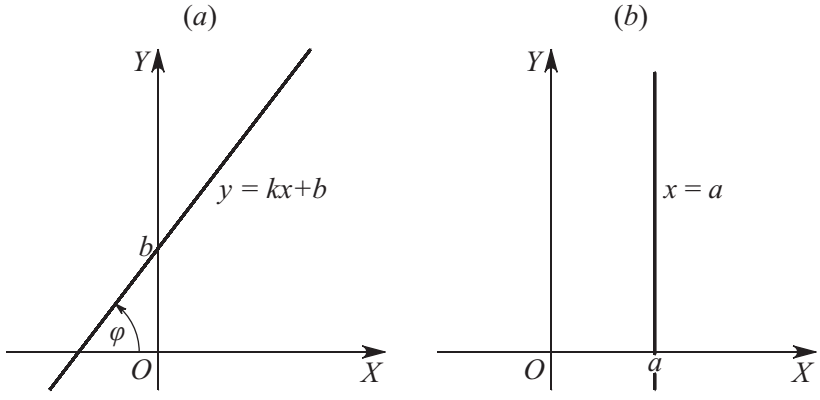


Figure 4.13. Straight lines on plane.

If a straight line is not perpendicular to the OX-axis, then its equation can be written as (4.3.1.1), but if a straight line is perpendicular to the OX-axis, then its equation can be written as

$$x = a, \tag{4.3.1.2}$$

where a is the abscissa of the point of intersection of this line with the OX-axis (see Fig. 4.13*b*).

For the slope of a straight line, we also have the formula

$$k = \frac{y_2 - y_1}{x_2 - x_1}, \tag{4.3.1.3}$$

where $A_1(x_1, y_1)$ and $A_2(x_2, y_2)$ are two arbitrary points of the line.

4.3.1-2. Point-slope equation of straight line.

In the rectangular Cartesian coordinate system OXY, the equation of a straight line with slope k passing through a point $A(x_1, y_1)$ has the form

$$y - y_1 = k(x - x_1). \tag{4.3.1.4}$$

Remark. If we set $x_1 = 0$ and $y_1 = b$ in equation (4.3.1.4), then we obtain equation (4.3.1.1).

4.3.1-3. Equation of straight line passing through two given points.

In the rectangular Cartesian coordinate system OXY, the equation of a straight line with slope k passing through points $A_1(x_1, y_1)$ and $A_2(x_2, y_2)$ has the form (4.3.1.4), where k is given by the expression (4.3.1.3):

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1). \tag{4.3.1.5}$$

This equation is usually written as

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1}. \tag{4.3.1.6}$$

Equation (4.3.1.6) is also called the *canonical equation of the straight line passing through two given points on the plane*.

Sometimes one writes this equation in terms of a third-order determinant as follows:

$$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0. \tag{4.3.1.7}$$

Example 1. Let us derive the equation of the straight line passing through the points $A_1(5, 1)$ and $A_2(7, 3)$. Substituting the coordinates of these points into formula (4.3.1.5), we obtain

$$\frac{x - 5}{2} = \frac{y - 1}{2} \qquad \text{or} \qquad y = x - 4.$$

4.3.1-4. General equation of straight line on plane.

An equation of the form

$$Ax + By + C = 0 \tag{4.3.1.8}$$

is called the *general equation of a straight line* in the rectangular Cartesian coordinate system OXY. In rectangular Cartesian coordinates, each straight line is determined by an equation of degree 1, and, conversely, each equation of degree 1 determines a straight line.

If $B \neq 0$, then equation (4.3.1.8) can be written as (4.3.1.1), where $k = -A/B$ and $b = -C/B$. If $B = 0$, then equation (4.3.1.8) can be written as (4.3.1.2), where $a = -C/A$.

If $C = 0$, then the equation of a straight line becomes $Ax + By = 0$ and determines a straight line passing through the origin.

If $B = 0$ and $A \neq 0$, then the equation of a straight line becomes $Ax + C = 0$ and determines a straight line parallel to the axis OY.

If $A = 0$ and $B \neq 0$, then the equation of a straight line becomes $By + C = 0$ and determines a straight line parallel to the axis OX.

4.3.1-5. General equation of straight line passing through given points on plane.

In the rectangular Cartesian coordinate system OXY, the general equation of a straight line passing through the point $A(x_1, y_1)$ on the plane has the form

$$A(x - x_1) + B(y - y_1) = 0. \tag{4.3.1.9}$$

If this equation is written in the form

$$\frac{x - x_1}{B} = \frac{y - y_1}{-A}, \tag{4.3.1.10}$$

then it is called the *canonical equation of a straight line passing through a given point on the plane*. If $B = 0$, then one sets $x - x_0 = 0$, and if $A = 0$, then one sets $y - y_0 = 0$.

Remark. The general equation of the straight line passing through two given points on the plane has the form (4.3.1.6).

4.3.1-6. Parametric equations of straight line on plane.

The *parametric equations of a straight line on the plane* through the point $A(x_1, y_1)$ in the rectangular Cartesian coordinate system OXY have the form

$$x = x_1 + Bt, \quad y = y_1 - At, \tag{4.3.1.11}$$

where A and B are the coefficients of the general equation (4.3.1.8) or (4.3.1.9) of a straight line and t is a variable parameter.

In the rectangular Cartesian coordinate system OXY, the parametric equations of the straight line passing through two points $A(x_1, y_1)$ and $A(x_2, y_2)$ on the plane can be written as

$$\begin{aligned} x &= x_1(1-t) + x_2 t, \\ y &= y_1(1-t) - y_2 t. \end{aligned} \tag{4.3.1.12}$$

Remark. Eliminating the parameter t from equations (4.3.1.11) and (4.3.1.12), we obtain equations (4.3.1.9) and (4.3.1.6), respectively.

4.3.1-7. Intercept-intercept equation of straight line.

The *intercept-intercept equation of a straight line* in the rectangular Cartesian coordinate system OXY has the form

$$\frac{x}{a} + \frac{y}{b} = 1, \tag{4.3.1.13}$$

where a and b are the x- and y-intercepts of the line, i.e., the signed distances from the points of intersection of the line with the coordinate axes to the origin (see Fig. 4.14).

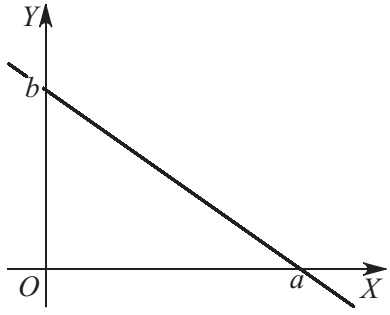


Figure 4.14. A straight line with intercept-intercept equation.

4.3.1-8. Normalized equation of straight line.

Suppose that a rectangular Cartesian coordinate system OXY and a straight line are given on the plane. We draw the perpendicular to the straight line through the origin. This perpendicular is called the *normal* to the line. By P we denote the point of intersection of the normal with the line.

The equation

$$x\cos\alpha + y\sin\alpha - p = 0, \tag{4.3.1.14}$$

where α is the polar angle of the normal and p is the length of the segment OP (the distance from the origin to the straight line) (see Fig. 4.15), is called the *normalized equation* of

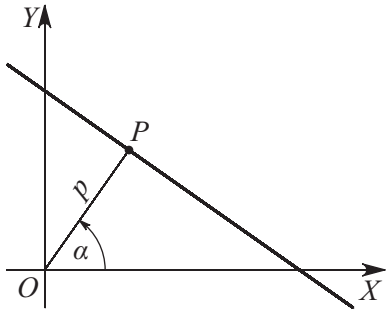


Figure 4.15. A straight line with normalized equation.

the straight line in the rectangular Cartesian coordinate system OXY. In the normalized equation of a straight line, $p \geq 0$ and $\cos^2 \alpha + \sin^2 \alpha = 1$.

For all positions of the straight line with respect to the coordinate axes, its equation can always be written in normalized form.

The general equation of a straight line (4.3.1.8) can be reduced to a normalized form (4.3.1.14) by setting

$$\cos \alpha = \pm \frac{A}{\sqrt{A^2 + B^2}}, \quad \sin \alpha = \pm \frac{B}{\sqrt{A^2 + B^2}}, \quad p = \mp \frac{C}{\sqrt{A^2 + B^2}}, \tag{4.3.1.15}$$

where the upper sign is taken for $C < 0$ and the lower sign for $C > 0$. For $C = 0$, either sign can be taken.

4.3.1-9. Equation of straight line in polar coordinates.

The *equation of a straight line in polar coordinates* can be written as

$$\rho \cos(\theta - \alpha) = p, \tag{4.3.1.16}$$

where ρ and θ are the polar radius and the polar angle of the current point B of the straight line, α is the polar angle of the normal, and p is the distance from the pole to the line (see Fig. 4.16).

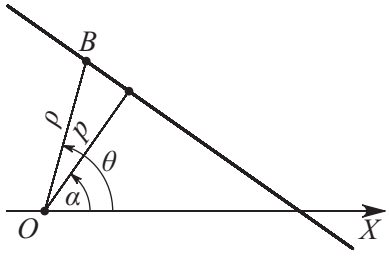


Figure 4.16. A straight line with equation in polar coordinates.

4.3.1-10. Equation of a pencil of straight lines.

The set of all straight lines passing through a point A on the plane is called a *pencil of straight lines*, and the point A itself is called the *center of the pencil*. The equations determining all straight lines in the pencil are called *equations of the pencil*.

1°. If the Cartesian coordinates of the pencil center $A(x_0, y_0)$ are known, then the equation of any straight line in the pencil that is not parallel to the OY-axis has the form

$$y - y_0 = k(x - x_0), \tag{4.3.1.17}$$

where k is the slope of the straight line. For the straight line in the pencil parallel to the OY-axis,

$$x = x_0. \tag{4.3.1.18}$$

2°. If the equations of two straight lines in the pencil are known, for example, $A_1x + B_1y + C_1 = 0$ and $A_2x + B_2y + C_2 = 0$, then the equation of the pencil can be written as

$$\alpha(A_1x + B_1y + C_1) + \beta(A_2x + B_2y + C_2) = 0, \tag{4.3.1.19}$$

where α and β are any numbers that are not simultaneously zero.

4.3.2. Mutual Arrangement of Points and Straight Lines

4.3.2-1. Condition for three points to be collinear.

Suppose that points $A_1(x_1, y_1)$, $A_2(x_2, y_2)$, and $A_3(x_3, y_3)$ are given in the Cartesian coordinate system OXY on the planes. They are collinear (lie on the same straight line) if and only if

$$\frac{x_3 - x_1}{x_2 - x_1} = \frac{y_3 - y_1}{y_2 - y_1}. \tag{4.3.2.1}$$

4.3.2-2. Deviation of point from straight line.

The number δ, equal to the length of the perpendicular to a straight line through a point taken with the plus sign if the point and the origin lie on opposite sides of the line and with the minus sign otherwise, is called the *deviation of the point from the line*. Obviously, the deviation is zero for points lying on the line.

To find the deviation of a point $A(x_0, y_0)$ from a straight line, one substitutes the coordinates of A into the left-hand side of the normalized equation of the line (see Paragraph 4.3.1-8) for the current Cartesian coordinates:

$$\delta = x_0 \cos\alpha + y_0 \sin\alpha - p. \tag{4.3.2.2}$$

4.3.2-3. Distance from point to straight line.

The *distance d from a point to a straight line* is the absolute value of the deviation. It can be calculated by the formula

$$d = |x_0 \cos\alpha + y_0 \sin\alpha - p|. \tag{4.3.2.3}$$

The distance from a point $A(x_0, y_0)$ to a straight line given by the general equation $Ax + By + C = 0$ can be calculated by the formula

$$d = \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}}. \tag{4.3.2.4}$$

Example 1. Let us find the distance from the point $A(2, 1)$ to the straight line $3x + 4y + 5 = 0$. We use formula (4.3.2.4) to obtain

$$d = \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}} = \frac{|3 \cdot 2 + 4 \cdot 1 + 5|}{\sqrt{3^2 + 4^2}} = \frac{15}{5} = 3.$$

4.3.2-4. Angle between two straight lines.

We consider two straight lines given by the equations

$$y = k_1 x + b_1 \quad \text{and} \quad y = k_2 x + b_2, \tag{4.3.2.5}$$

where $k_1 = \tan\varphi_1$ and $k_2 = \tan\varphi_2$ are the slopes of the respective lines (see Fig. 4.17). The angle α between these lines can be obtained by the formula

$$\tan\alpha = \frac{k_2 - k_1}{1 + k_1 k_2}, \tag{4.3.2.6}$$

where $k_1 k_2 \neq -1$. If $k_1 k_2 = -1$, then $\alpha = \frac{1}{2}\pi$.

Remark. If at least one of the lines is perpendicular to the axis OX, then formula (4.3.2.6) does not make sense. In this case, the angle between the lines can be calculated by the formula

$$\alpha = \varphi_2 - \varphi_1. \tag{4.3.2.7}$$

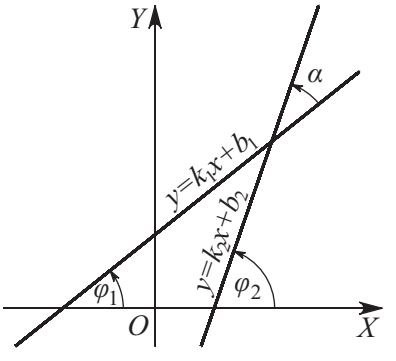


Figure 4.17. Angle between two straight lines.

The angle α between the two straight lines given by the general equations

$$A_1 x + B_1 y + C_1 = 0 \quad \text{and} \quad A_2 x + B_2 y + C_2 = 0 \tag{4.3.2.8}$$

can be calculated using the expression

$$\tan\alpha = \frac{A_1 B_2 - A_2 B_1}{A_1 A_2 + B_1 B_2}, \tag{4.3.2.9}$$

where $A_1 A_2 + B_1 B_2 \neq 0$. If $A_1 A_2 + B_1 B_2 = 0$, then $\alpha = \frac{1}{2}\pi$.

Remark. If one needs to find the angle between straight lines and the order in which they are considered is not defined, then this order can be chosen arbitrarily. Obviously, a change in the order results in a change in the sign of the tangent of the angle.

4.3.2-5. Point of intersection of straight lines.

Suppose that two straight lines are defined by general equations in the form (4.3.2.8). Consider the system of two first-order algebraic equations (4.3.2.8):

$$\begin{aligned} A_1x + B_1y + C_1 = 0,\\ A_2x + B_2y + C_2 = 0. \end{aligned} \tag{4.3.2.10}$$

Each common solution of equations (4.3.2.10) determines a common point of the tow lines.

If the determinant of system (4.3.2.10) is not zero, i.e.,

$$\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix} = A_1B_2 - A_2B_1 \neq 0, \tag{4.3.2.11}$$

then the system is consistent and has a unique solution; hence these straight lines are distinct and nonparallel and meet at the point $A(x_0, y_0)$, where

$$x_0 = \frac{B_1C_2 - B_2C_1}{A_1B_2 - A_2B_1}, \quad y_0 = \frac{C_1A_2 - C_2A_1}{A_1B_2 - A_2B_1}. \tag{4.3.2.12}$$

Condition (4.3.2.11) is often written as

$$\frac{A_1}{A_2} \neq \frac{B_1}{B_2}. \tag{4.3.2.13}$$

Example 2. To find the point of intersection of the straight lines $y = 2x - 1$ and $y = -4x + 5$, we solve system (4.3.2.10):

$$\begin{aligned} 2x - y - 1 = 0,\\ -4x - y + 5 = 0, \end{aligned}$$

and obtain $x = 1$, $y = 1$. Thus the intersection point has the coordinates $(1, 1)$.

4.3.2-6. Condition for straight lines to be perpendicular.

For two straight lines determined by slope-intercept equations (4.3.2.5) to be perpendicular, it is necessary and sufficient that

$$k_1k_2 = -1. \tag{4.3.2.14}$$

Relation (4.3.2.14) is usually written as

$$k_1 = -\frac{1}{k_2}, \tag{4.3.2.15}$$

and one also says that the slopes of perpendicular straight lines are inversely proportional in absolute value and opposite in sign.

If the straight lines are given by general equations (4.3.2.8), then a necessary and sufficient condition for them to be perpendicular can be written as (see Paragraph 4.3.2-4)

$$A_1A_2 + B_1B_2 = 0. \tag{4.3.2.16}$$

Example 3. The lines $3x+y-3=0$ and $x-3y+8=0$ are perpendicular since they satisfy condition (4.3.2.16):

$$A_1A_2 + B_1B_2 = 3 \cdot 1 + 1 \cdot (-3) = 0.$$

4.3.2-7. Condition for straight lines to be parallel.

For two straight lines defined by slope-intercept equations (4.3.2.5) to be parallel and not to coincide, it is necessary and sufficient that

$$k_1 = k_2, \quad b_1 \neq b_2. \tag{4.3.2.17}$$

If the straight lines are given by general equations (4.3.2.8), then a necessary and sufficient condition for them to be parallel can be written as

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} \neq \frac{C_1}{C_2}; \tag{4.3.2.18}$$

in this case, the straight lines do not coincide (see Fig. 4.18).

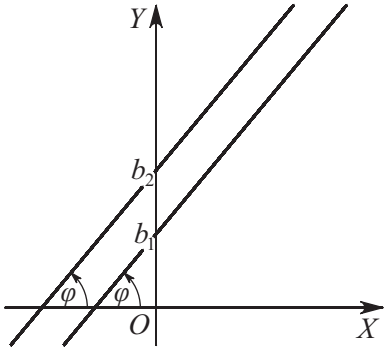


Figure 4.18. Parallel straight lines.

Example 4. The straight lines $3x + 4y + 5 = 0$ and $3/2\,x + 2y + 6 = 0$ are parallel since the following condition (4.3.2.18) is satisfied:

$$\frac{3}{3/2} = \frac{4}{2} \neq \frac{5}{6}.$$

4.3.2-8. Condition for straight lines to coincide.

For two straight lines given by slope-intercept equations (4.3.2.5) to coincide, it is necessary and sufficient that

$$k_1 = k_2, \quad b_1 = b_2. \tag{4.3.2.19}$$

If the straight lines are given by general equations (4.3.2.8), then a necessary and sufficient condition for them to coincide has the form

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2}. \tag{4.3.2.20}$$

Remark. Sometimes the case of coinciding straight lines is considered as a special case of parallel straight lines and it not distinguished as an exception.

4.3.2-9. Distance between parallel lines.

The distance between the parallel lines given by equations (see Paragraph 4.3.2-7)

$$A_1x + B_1y + C_1 = 0 \quad \text{and} \quad A_1x + B_1y + C_2 = 0 \tag{4.3.2.21}$$

can be found using the formula (see Paragraph 4.3.2-3)

$$d = \frac{|C_1 - C_2|}{\sqrt{A_1^2 + B_1^2}}. \tag{4.3.2.22}$$

4.3.2-10. Condition for a straight line to separate points of plane.

Suppose that a straight line in the Cartesian coordinate system OXY is given by an equation of the form

$$Ax + By + C = 0. \tag{4.3.2.23}$$

Obviously, this straight line divides the plane into two half-planes. We consider two arbitrary points $A_1(x_1, y_1)$ and $A_2(x_2, y_2)$ of the plane that do not lie on the line. The points are said to be *nonseparated by the straight line* if they belong to the same half-plane (lie on the same side of the straight line and possibly coincide). The points are said to be *separated by the straight line* if they belong to different half-planes (lie on opposite sides of the straight line).

Two points $A_1(x_1, y_1)$ and $A_2(x_2, y_2)$ that do not belong to the straight line (4.3.2.23) are separated by this line if and only if the numbers $Ax_1 + By_1 + C$ and $Ax_2 + By_2 + C$ have opposite signs.

4.4. Second-Order Curves

4.4.1. Circle

4.4.1-1. Definition and canonical equation of circle.

A curve on the plane is called a *circle* if there exists a rectangular Cartesian coordinate system OXY in which the equation of this curve has the form (see Fig. 4.19*a*)

$$x^2 + y^2 = a^2, \tag{4.4.1.1}$$

where the point $O(0, 0)$ is the center of the circle and $a > 0$ is its radius. Equation (4.4.1.1) is called the *canonical equation of a circle*.

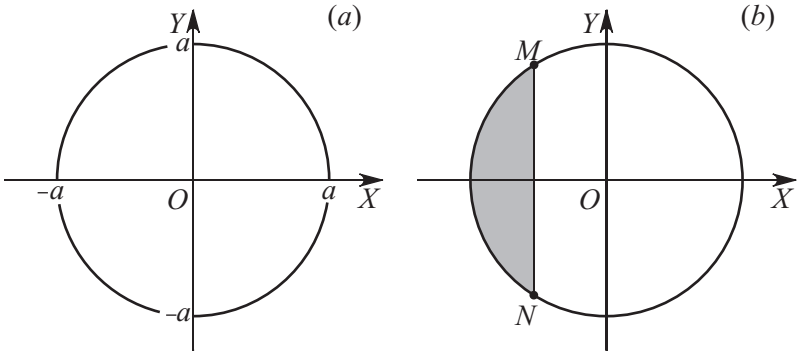


Figure 4.19. Circle.

The circle defined by equation (4.4.1.1) is the locus of points equidistant (lying at the distance a) from its center. If a circle of radius a is centered at a point $C(x_0, y_0)$, then its equation can be written as

$$(x - x_0)^2 + (y - y_0)^2 = a^2. \tag{4.4.1.2}$$

The area of the *disk* bounded by a circle of radius a is given by the formula $S = \pi a^2$. The length of this circle is $L = 2\pi a$. The area of the figure bounded by the circle and the chord passing through the points $M(x_0, y_0)$ and $N(x_0, -y_0)$ is (see Fig. 19*b*)

$$S = \frac{\pi a^2}{2} + x_0\sqrt{a^2 - x_0^2} + a^2 \arcsin\frac{x_0}{a}. \tag{4.4.1.3}$$

4.4.1-2. Parametric and other equations of circle.

The parametric equations of the circle (4.4.1.1) have the form

$$x = a\cos\theta, \quad y = a\sin\theta, \tag{4.4.1.4}$$

where the angle in the polar coordinate system plays the role of the variable parameter (see Paragraphs 4.2.1-4 and 4.2.1-5).

The equation of the circle (4.4.1.1) in the polar coordinate system has the form

$$\rho = a \tag{4.4.1.5}$$

and does not contain the polar angle θ.

Remark. The form of the equation of a circle in a polar coordinate system depends on the choice of the pole and the polar axis (see Example 7 in Subsection 4.2.1).

4.4.2. Ellipse

4.4.2-1. Definition and canonical equation of ellipse.

A curve on the plane is called an *ellipse* if there exists a rectangular Cartesian coordinate system OXY in which the equation of the curve has the form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \tag{4.4.2.1}$$

where $a \geq b > 0$ (see Fig. 4.20). The coordinates in which the equation of an ellipse has the form (4.4.2.1) are called the *canonical coordinates* for this ellipse, and equation (4.4.2.1) itself is called the *canonical equation of the ellipse*.

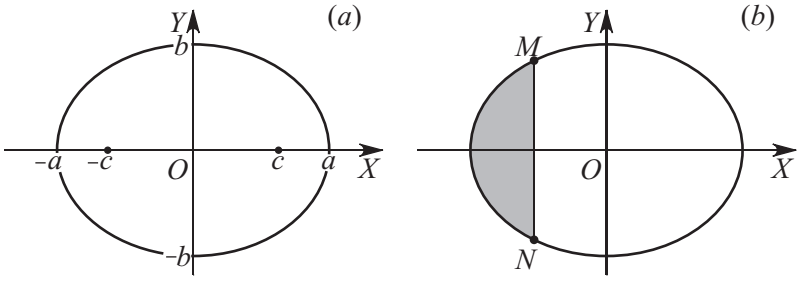


Figure 4.20. Ellipse.

The segments A_1A_2 and B_1B_2 joining the opposite vertices of an ellipse, as well as their lengths $2a$ and $2b$, are called the *major* and *minor axes*, respectively, *of the ellipse*. The axes of an ellipse are its axes of symmetry. In Fig. 4.20*a*, the axes of symmetry of the ellipse coincide with the axes of the rectangular Cartesian coordinate system OXY. The numbers a and b are called the *semimajor* and *semiminor axes of the ellipse*. The number $c = \sqrt{a^2 - b^2}$ is called the *linear eccentricity*, and the number $2c$ is called the *focal distance*. The number $e = c/a = \sqrt{1 - a^2/b^2}$, where, obviously, $0 \leq e < 1$, is called the *eccentricity* or the *numerical eccentricity*. The number $p = b^2/a$ is called the *focal parameter* or simply the *parameter of the ellipse*.

The point $O(0,0)$ is called the *center of the ellipse*. The points of intersection $A_1(-a,0)$, $A_2(a,0)$ and $B_1(0,-b)$, $B_2(0,b)$ of the ellipse with the axes of symmetry are called its *vertices*. The points $F_1(-c,0)$ and $F_2(c,0)$ are the *focus of the ellipse*. This explains why the major axis of an ellipse is sometimes called its *focal axis*. The straight lines $x = \pm a/e$ ($e \neq 0$) are called the *directrices*. The focus $F_2(c,0)$ and the directrix $x = a/e$ are said to be *right*, and the focus $F_1(-c,0)$ and the directrix $x = -a/e$ are said to be *left*. A focus and a directrix are said to be *like* if both of them are right or left simultaneously.

The segments joining a point $M(x,y)$ of an ellipse with the foci $F_1(-c,0)$ and $F_2(c,0)$ are called the *left* and *right focal radii* of this point. We denote the lengths of the left and right focal radii by $r_1 = |F_1M_0|$ and $r_2 = |F_2M_0|$, respectively.

Remark. For $a = b$ ($c = 0$), equation (4.4.2.1) becomes $x^2 + y^2 = a^2$ and determines a circle; hence a circle can be considered as an ellipse for which $b = a$, $c = 0$, $e = 0$, and $\rho = 0$, i.e., the semiaxes are equal to each other, the foci coincide with the center, the eccentricity is zero (the directrices are not defined), and the focal parameter is zero.

The area of the figure bounded by the ellipse is given by the formula $S = \pi ab$. The length of the ellipse can be calculated approximately by the formula $L \approx \pi\left[1.5\,(a+b) - \sqrt{ab}\,\right]$. The area of the figure bounded by the ellipse and the chord passing through the points $M(x_0, y_0)$ and $N(x_0, -y_0)$ is equal to (see Fig. 20*b*)

$$S = \frac{\pi ab}{2} + \frac{b}{a}\left(x_0\sqrt{a^2 - x_0^2} + a^2 \arcsin\frac{x_0}{a}\right). \tag{4.4.2.2}$$

4.4.2-2. Focal property of ellipse.

The ellipse defined by equation (4.4.2.1) is the locus of points on the plane for which the sum of distances to the foci F_1 and F_2 is equal to $2a$ (see Fig. 4.21). We write this property as

$$r_1 + r_2 = 2a, \tag{4.4.2.3}$$

where r_1 and r_2 satisfy the relations

$$\begin{aligned} r_1 &= \sqrt{(x+c)^2 + y^2} = a + ex, \\ r_2 &= \sqrt{(x-c)^2 + y^2} = a - ex. \end{aligned} \tag{4.4.2.4}$$

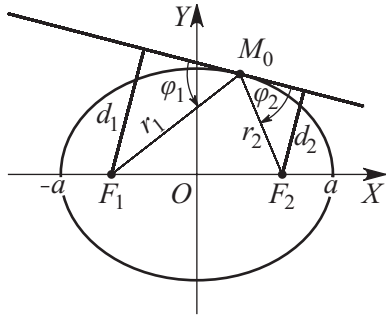


Figure 4.21. Focal property of ellipse.

Remark. One can show that equation (4.4.2.1) implies equation (4.4.2.3) and vice versa; hence the focal property of an ellipse is often used as its definition.

4.4.2-3. Focus-directrix property of ellipse.

The ellipse determined by equation (4.4.2.1) on the plane is the locus of points for which the ratio of distances to a focus and the like directrix is equal to e:

$$r_1\left|x+\frac{a}{e}\right|^{-1}=e, \qquad r_2\left|x-\frac{a}{e}\right|^{-1}=e. \tag{4.4.2.5}$$

4.4.2-4. Equation of tangent and optical property of ellipse.

The tangent to the ellipse (4.4.2.1) at an arbitrary point $M_0(x_0, y_0)$ is described by the equation

$$\frac{x_0x}{a^2}+\frac{y_0y}{b^2}=1. \tag{4.4.2.6}$$

The distances d_1 and d_2 from the foci $F_1(-c, 0)$ and $F_2(c, 0)$ to the tangent to the ellipse at the point $M_0(x_0, y_0)$ are given by the formulas (see Paragraph 4.3.2-4)

$$d_1=\frac{1}{Na}|x_0e+a|=\frac{r_1(M_0)}{Na}, \qquad d_2=\frac{1}{Na}|x_0e-a|=\frac{r_2(M_0)}{Na}, \qquad N=\sqrt{\left(\frac{x_0}{a^2}\right)^2+\left(\frac{y_0}{b^2}\right)^2}, \tag{4.4.2.7}$$

where $r_1(M_0)$ and $r_2(M_0)$ are the lengths of the focal radii of M_0.

The tangent at an arbitrary point $M_0(x_0, y_0)$ of an ellipse forms acute angles φ_1 and φ_2 with the focal radii of the point of tangency, and

$$\sin\varphi_1=\frac{d_1}{r_1}=\frac{1}{Na}, \qquad \sin\varphi_2=\frac{d_2}{r_2}=\frac{1}{Na}. \tag{4.4.2.8}$$

This implies the *optical property of the ellipse*:

$$\varphi_1=\varphi_2, \tag{4.4.2.9}$$

which means that all light rays issuing from one focus of the ellipse converge at the other focus after the reflection in the ellipse.

4.4.2-5. Diameters of ellipse.

A straight line passing through the midpoints of parallel chords of an ellipse is called a *diameter of the ellipse*. All diameters of an ellipse pass through its center. Two diameters of an ellipse are said to be *conjugate* if their slopes satisfy the relation

$$k_1k_2=-\frac{b^2}{a^2}. \tag{4.4.2.10}$$

Two perpendicular conjugate diameters are called the *principal diameters of the ellipse*.

Remark. If $a = b$, i.e., the ellipse is a circle, then condition (4.4.2.10) becomes the perpendicularity condition: $k_1k_2 = -1$. Thus any two conjugate diameters of a circle are perpendicular to each other, and each of the diameters is a principal diameter.

4.4.2-6. Ellipse in polar coordinate system.

In polar coordinates (ρ, φ), the equation of an ellipse becomes

$$\rho = \frac{p}{1 - e\cos\varphi}, \tag{4.4.2.11}$$

where $0 \le \varphi \le 2\pi$.

4.4.3. Hyperbola

4.4.3-1. Definition and canonical equation of hyperbola.

A curve on the plane is called a *hyperbola* if there exists a rectangular Cartesian coordinate system OXY in which the equation of this curve has the form

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \tag{4.4.3.1}$$

where $a > 0$ and $b > 0$ (see Fig. 4.22*a*). The coordinates in which the equation of a hyperbola has the form (4.4.3.1) are called the *canonical coordinates* for the hyperbola, and equation (4.4.3.1) itself is called the *canonical equation of the hyperbola*.

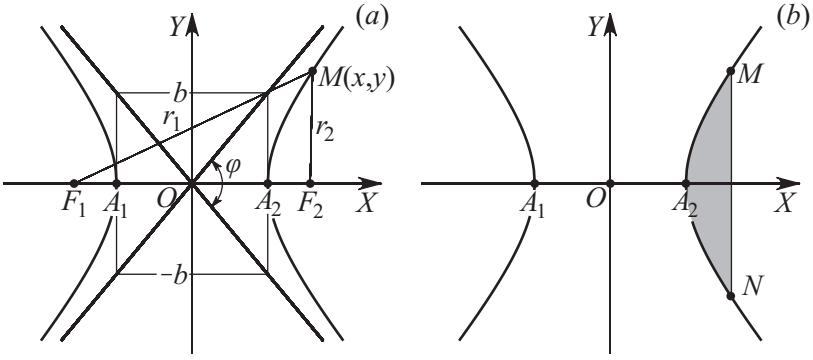


Figure 4.22. Hyperbola.

The hyperbola is a central curve of second order. It is described by equation (4.4.3.1) and consists of two connected parts (*arms*) lying in the domains $x > a$ and $x < -a$. The hyperbola has two *asymptotes* given by the equations

$$y = \frac{b}{a}x \quad \text{and} \quad y = -\frac{b}{a}x. \tag{4.4.3.2}$$

More precisely, its arms lie in the two vertical angles formed by the asymptotes and are called the *left* and *right arms* of the hyperbola. A hyperbola is symmetric about the axes OX and OY, which are called the *principal (real, or focal, and imaginary) axes*.

The angle between the asymptotes of a hyperbola is determined by the equation

$$\tan\frac{\varphi}{2} = \frac{b}{a}, \tag{4.4.3.3}$$

and if $a = b$, then $\varphi = \frac{1}{2}\pi$ (an *equilateral hyperbola*).

The number a is called the *real semiaxis*, and the number b is called the *imaginary semiaxis*. The number $c = \sqrt{a^2 + b^2}$ is called the *linear eccentricity*, and $2c$ is called the *focal distance*. The number $e = c/a = \sqrt{a^2 + b^2}/a$, where, obviously, $e > 1$, is called the *eccentricity*, or the *numerical eccentricity*. The number $p = b^2/a$ is called the *focal parameter* or simply the *parameter of the hyperbola*.

The point $O(0, 0)$ is called the *center of the hyperbola*. The points $A_1(-a, 0)$ and $A_2(a, 0)$ of intersection of the hyperbola with the real axis are called the *vertices of the hyperbola*. Points $F_1(-c, 0)$ and $F_2(c, 0)$ are called the *foci of the hyperbola*. This is why the real axis of a hyperbola is sometimes called the focal axis. The straight lines $x = \pm a/e$ ($y \neq 0$) are called the *directrices of the hyperbola* corresponding to the foci F_2 and F_1. The focus $F_2(c, 0)$ and the directrix $x = a/e$ are said to be *right*, and the focus $F_1(-c, 0)$ and the directrix $x = -a/e$ are said to be *left*. A focus and a directrix are said to be *like* if both of them are right or left simultaneously.

The segments joining a point $M(x, y)$ of the hyperbola with the foci $F_1(-c, 0)$ and $F_2(c, 0)$ are called the *left* and *right focal radii* of this point. We denote the lengths of the left and right focal radii by $r_1 = |F_1M|$ and $r_2 = |F_2M|$, respectively.

Remark. For $a = b$, the hyperbola is said to be *equilateral*, and its asymptotes are mutually perpendicular. The equation of an equilateral hyperbola has the form $x^2 - y^2 = a^2$. If we take the asymptotes to be the coordinate axes, then the equation of the hyperbola becomes $xy = a^2/2$; i.e., an equilateral hyperbola is the graph of inverse proportionality.

The curvature radius of a hyperbola at a point $M(x, y)$ is

$$R = a^2b^2\left(\frac{x^2}{a^4} + \frac{y^2}{b^4}\right)^{3/2} = \frac{\sqrt{(r_1r_2)^3}}{ab}. \tag{4.4.3.4}$$

The area of the figure bounded by the right arm of the hyperbola and the chord passing through the points $M(x_1, y_1)$ and $N(x_1, -y_1)$ is equal to (see Fig. 4.22*b*)

$$S = x_1y_1 - ab\ln\left(\frac{x_1}{a} + \frac{y_1}{b}\right). \tag{4.4.3.5}$$

4.4.3-2. Focal properties of hyperbola.

The hyperbola determined by equation (4.4.3.1) is the locus of points on the plane for which the difference of the distances to the foci F_1 and F_2 has the same absolute value $2a$ (see Fig. 4.22*a*). We write this property as

$$|r_1 - r_2| = 2a, \tag{4.4.3.6}$$

where r_1 and r_2 satisfy the relations

$$\begin{aligned} r_1 &= \sqrt{(x+c)^2 + y^2} = \begin{cases} a + ex & \text{for } x > 0, \\ -a - ex & \text{for } x < 0, \end{cases} \\ r_2 &= \sqrt{(x-c)^2 + y^2} = \begin{cases} -a + ex & \text{for } x > 0, \\ a - ex & \text{for } x < 0. \end{cases} \end{aligned} \tag{4.4.3.7}$$

Remark. One can show that equation (4.4.3.1) implies equation (4.4.3.6) and vice versa; hence the focal property of a hyperbola is often used as the definition.

4.4.3-3. Focus-directrix property of hyperbola.

The hyperbola defined by equation (4.4.3.1) on the plane is the locus of points for which the ratio of distances to a focus and the like directrix is equal to e:

$$r_1\left|x+\frac{a}{e}\right|^{-1}=e, \qquad r_2\left|x-\frac{a}{e}\right|^{-1}=e. \tag{4.4.3.8}$$

4.4.3-4. Equation of tangent and optical property of hyperbola.

The tangent to the hyperbola (4.4.3.1) at an arbitrary point $M_0(x_0, y_0)$ is described by the equation

$$\frac{x_0x}{a^2}-\frac{y_0y}{b^2}=1. \tag{4.4.3.9}$$

The distances d_1 and d_2 from the foci $F_1(-c, 0)$ and $F_2(0, c)$ to the tangent to the hyperbola at the point $M_0(x_0, y_0)$ are given by the formulas (see Paragraph 4.3.2-4)

$$d_1=\frac{Na}{|x_0e+a|}=\frac{r_1}{Na}, \qquad d_2=\frac{Na}{|x_0e-a|}=\frac{r_2}{Na}, \qquad N=\sqrt{\left(\frac{x_0}{a^2}\right)^2+\left(\frac{y_0}{b^2}\right)^2}, \tag{4.4.3.10}$$

where r_1 and r_2 are the lengths of the focal radii of the point M_0.

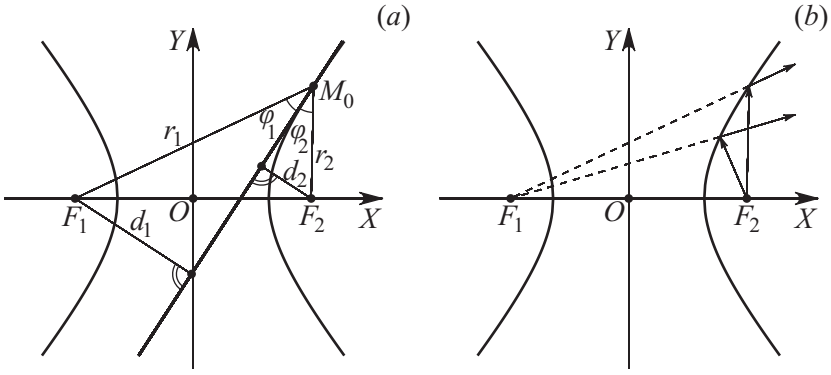


Figure 4.23. The tangent to the hyperbola (*a*). Optical property of a hyperbola (*b*).

The tangent at any point $M_0(x_0, y_0)$ of the hyperbola forms acute angles φ_1 and φ_2 with the focal radii of the point of tangency (see Fig. 4.23*a*), and

$$\sin\varphi_1=\frac{d_1}{r_1}=\frac{1}{Na}, \qquad \sin\varphi_2=\frac{d_2}{r_2}=\frac{1}{Na}. \tag{4.4.3.11}$$

This implies the *optical property of a hyperbola*:

$$\varphi_1=\varphi_2, \tag{4.4.3.12}$$

which means that all light rays issuing from a focus appear to be issuing from the other focus after the mirror reflection in the hyperbola (see Fig. 4.23*b*).

The tangent and normal to a hyperbola at any point bisect the angles between the straight lines joining this point with the foci. The tangent to a hyperbola at either of its vertices intersects the asymptotes at two points such that the distance between them is equal to $2b$.

4.4.3-5. Diameters of hyperbola.

A straight line passing through the midpoints of parallel chords of a hyperbola is called a *diameter of the hyperbola*. Two diameters of a hyperbola are said to be *conjugate* if their slopes satisfy the relation

$$k_1 k_2 = \frac{b^2}{a^2}. \tag{4.4.3.13}$$

A hyperbola meets the diameter $y = kx$ if and only if

$$k^2 < \frac{b^2}{a^2}. \tag{4.4.3.14}$$

The lengths l_1 and l_2 of the conjugate diameters with slopes k_1 and k_2 satisfy the relation

$$l_1 l_2 \sin(\arctan k_2 - \arctan k_1) = 4ab. \tag{4.4.3.15}$$

Two perpendicular conjugate diameters are called the *principal diameters of a hyperbola*; they are its principal axes.

4.4.3-6. Hyperbola in polar coordinate system.

In polar coordinates (ρ, φ), the equation for two connected parts of a hyperbola becomes

$$\rho = \frac{\pm p}{1 \mp e \cos\varphi}, \tag{4.4.3.16}$$

where upper and lower signs correspond to right and left parts of a hyperbola, respectively.

4.4.4. Parabola

4.4.4-1. Definition and canonical equation of parabola.

A curve on the plane is called a *parabola* if there exists a rectangular Cartesian coordinate system OXY, in which the equation of this curve has the form

$$y^2 = 2px, \tag{4.4.4.1}$$

where $p > 0$ (see Fig. 4.24*a*). The coordinates in which the equation of a parabola has the form (4.4.4.1) are called the *canonical coordinates* for the parabola, and equation (4.4.4.1) itself is called the *canonical equation of the parabola*.

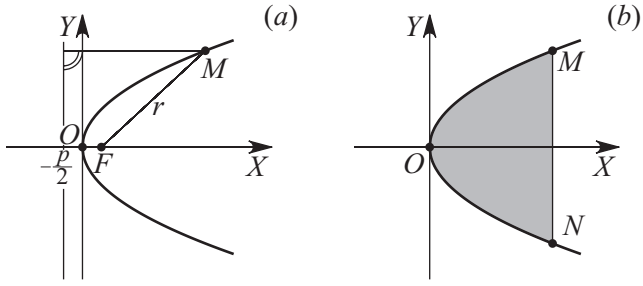


Figure 4.24. Parabola.

A parabola is a noncentral line of second order. It consists of an infinite branch symmetric about the OX-axis. The point $O(0,0)$ is called the *vertex of the parabola*. The point $F(p/2,0)$ is called the *focus of the parabola*. The straight line $x = -p/2$ is called the *directrix*. The focal parameter p is the distance from the focus to the directrix. The number $p/2$ is called the *focal distance*.

The segment joining a point $M(x,y)$ of the parabola with the focus $F(p/2,0)$ is called the *focal radius* of the point. The curvature radius of the parabola at a point $M(x,y)$ can be found from the formula

$$R = \frac{(p+2x)^{3/2}}{\sqrt{p}}. \tag{4.4.4.2}$$

The area of the figure bounded by the parabola and the chord passing through the points $M(x_1,y_1)$ and $N(x_1,-y_1)$ is equal to (see Fig. 4.24*b*)

$$S = \frac{4}{3}x_1y_1. \tag{4.4.4.3}$$

4.4.4-2. Focal properties of parabola.

The parabola defined by equation (4.4.4.1) on the plane is the locus of points equidistant from the focus $F(p/2,0)$ and the directrix $x = -p/2$ (see Fig. 4.24*a*).

We denote the length of the focal radius by r and write this property as

$$r = x + \frac{p}{2}, \tag{4.4.4.4}$$

where r satisfies the relation

$$r = \sqrt{\left(x - \frac{p}{2}\right)^2 + y^2}. \tag{4.4.4.5}$$

Remark. One can show that equation (4.4.4.1) implies equation (4.4.4.5) and vice versa; hence the focal property of a parabola is often used as the definition.

4.4.4-3. Focus-directrix property of parabola.

The parabola defined by equation (4.4.4.1) on the plane is the locus of points for which the ratio of distances to the focus and the directrix is equal to 1:

$$\frac{r}{|x + p/2|} = 1. \tag{4.4.4.6}$$

4.4.4-4. Equation of tangent and optical property of parabola.

The tangent to the parabola (4.4.4.1) at an arbitrary point $M_0(x_0,y_0)$ is described by the equation

$$yy_0 = p(x + x_0). \tag{4.4.4.7}$$

The direction vector of the tangent (4.4.4.7) has the coordinates (y_0,p), and the direction vector of the line passing through the points $M_0(x_0,y_0)$ and $F(p/2,0)$ has the coordinates

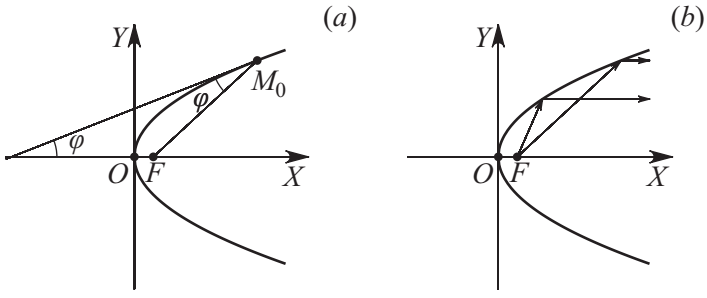


Figure 4.25. The tangent to the parabola (a). Optical property of a parabola (b).

$(x_0 - p/2, y_0)$ (see Fig. 4.25a). Thus, in view of the focus-directrix property, the angle φ between these lines satisfies the relation

$$\cos\varphi = \frac{y_0(x_0 - p/2) + py_0}{\sqrt{y_0^2 + p^2}\sqrt{(x_0 - p/2)^2 + y_0^2}} = \frac{y_0}{\sqrt{y_0^2 + p^2}}\,. \qquad (4.4.4.8)$$

But the same relation also holds for the angle between the tangent (4.4.4.7) and the OX-axis.

This property of a parabola is called the *optical property*: all light rays issuing from the focus of a parabola form a pencil parallel to the axis of the parabola after the mirror reflection in the parabola (see Fig. 4.25b).

The tangent and normal to a parabola at any point bisect the angles between the focal radius and the diameter.

4.4.4-5. Diameters of parabola.

A straight line passing through the midpoints of parallel chords of a parabola is called a *diameter of the parabola*. The diameter corresponding to the chords perpendicular to the axis of the parabola is the axis itself. The diameter of the parabola $y^2 = 2px$ corresponding to the chords with slope k ($k > 0$) is given by the equation

$$y = \frac{p}{k}\,. \qquad (4.4.4.9)$$

The OX-axis (the axis of symmetry of a parabola), in contrast to the other diameters of the parabola, is the diameter perpendicular to the chords conjugate to it. This diameter is called the *principal diameter of the parabola*. The slope of any diameter of a parabola is zero. A parabola does not have mutually conjugate diameters.

4.4.4-6. Parabola with vertical axis.

The equation of a parabola with vertical axis has the form

$$y = ax^2 + bx + c \quad (a \neq 0). \qquad (4.4.4.10)$$

For $a > 0$, the vertex of the parabola is directed downward, and for $a < 0$, the vertex is directed upward. The vertex of a parabola has the coordinates

$$x_0 = \frac{b}{2}\,, \quad y_0 = \frac{4ac - b^2}{4a}\,. \qquad (4.4.4.11)$$

4.4.4-7. Parabola in polar coordinates.

In the polar coordinates (ρ, φ) (the pole lies at the focus of the parabola, and the polar axis is directed along the parabola axis), the equation of the parabola has the form

$$\rho = \frac{p}{1 - \cos\varphi}, \tag{4.4.4.12}$$

where $-\frac{1}{2}\pi \le \varphi \le \frac{1}{2}\pi$.

4.4.5. Transformation of Second-Order Curves to Canonical Form

4.4.5-1. General equation of second-order curve.

The set of points on the plane whose coordinates in the rectangular Cartesian coordinate system satisfy the second-order algebraic equation

$$\begin{aligned} &a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0 \quad \text{or} \\ &(a_{11}x + a_{12}y + a_{13})x + (a_{21}x + a_{22}y + a_{23})y + a_{31}x + a_{32}y + a_{33} = 0, \\ &a_{ij} = a_{ji} \quad (i, j = 1, 2, 3) \end{aligned} \tag{4.4.5.1}$$

is called a *second-order curve*.

4.4.5-2. Nine canonical second-order curves.

There exists a rectangular Cartesian coordinate system in which equations (4.4.5.1) can be reduced to one of the following *nine canonical forms*:

1. $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$, an ellipse;
2. $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$, a hyperbola;
3. $y = 2px$, a parabola;
4. $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = -1$, an imaginary ellipse;
5. $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 0$, a pair of intersecting straight lines;
6. $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 0$, a pair of imaginary intersecting straight lines;
7. $x^2 - a^2 = 0$, a pair of parallel straight lines;
8. $x^2 + a^2 = 0$, a pair of imaginary parallel straight lines;
9. $x^2 = 0$, a pair of coinciding straight lines.

4.4.5-3. Invariants of second-order curves.

Second-order curves can be studied with the use of the three *invariants*

$$I = a_{11} + a_{22}, \quad \delta = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}, \quad \Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}, \tag{4.4.5.2}$$

whose values do not change under parallel translation and rotation of the coordinate axes, and the *semi-invariant*

$$\sigma = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}, \tag{4.4.5.3}$$

whose value does not change under rotation of the coordinate axes.

The invariant Δ is called the *large discriminant* of equation (4.4.2.1). The invariant δ is called the *small discriminant*.

Table 1 presents the classification of second-order curves based on invariants.

TABLE 4.1
Classification of second-order curves

<table>
<tr><th>Type</th><th colspan="2">$\Delta \neq 0$</th><th colspan="2">$\Delta = 0$</th></tr>
<tr><td rowspan="2">Elliptic
$\delta > 0$</td><td>$\frac{\Delta}{I} > 0$</td><td>Real ellipse (for $I^2 = 4\delta$ or $a_{11} = a_{22}$ and $a_{12} = 0$, this is a circle)</td><td colspan="2" rowspan="2">Pair of imaginary straight lines intersecting at a real point (ellipse degenerating into a point)</td></tr>
<tr><td>$\frac{\Delta}{I} < 0$</td><td>Imaginary ellipse (no real points)</td></tr>
<tr><td>Hyperbolic
$\delta < 0$</td><td colspan="2">Hyperbola</td><td colspan="2">Pair of real intersecting straight lines (degenerate hyperbola)</td></tr>
<tr><td rowspan="3">Parabolic
$\delta = 0$</td><td colspan="2" rowspan="3">Parabola</td><td>$\sigma > 0$</td><td>Pair of imaginary parallel straight lines</td></tr>
<tr><td>$\sigma < 0$</td><td>Pair of real parallel straight lines</td></tr>
<tr><td>$\sigma = 0$</td><td>Pair of coinciding straight lines (a single straight line)</td></tr>
</table>

4.4.5-4. Characteristic equation of second-order curves.

The properties of second-order curves can be studied with the use of the characteristic equation

$$\begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = 0 \quad \text{or} \quad \lambda^2 - I\lambda + \delta = 0. \tag{4.4.5.4}$$

The roots λ_1 and λ_2 of the characteristic equation (4.4.5.4) are the eigenvalues of the real symmetric matrix $[a_{ij}]$ and, as a consequence, are real.

Obviously, the invariants I and δ of second-order curves are expressed as follows in terms of the roots λ_1 and λ_2 of the characteristic equation (4.4.5.4):

$$I = \lambda_1 + \lambda_2, \quad \delta = \lambda_1\lambda_2. \tag{4.4.5.5}$$

4.4.5-5. Centers and diameters of second-order curves.

A straight line passing through the midpoints of parallel chords of a second-order curve is called a *diameter* of this curve. A diameter is said to be *conjugate* to the chords (or to the direction of chords) which it divides into two parts. The diameter conjugate to chords forms an angle φ with the positive direction of the OX-axis and is determined by the equation

$$(a_{11}x + a_{12}y + a_{13})\cos\varphi + (a_{21}x + a_{22}y + a_{23})\sin\varphi = 0. \tag{4.4.5.6}$$

All diameters of a second-order curve with $\delta \neq 0$ meet at a single point called the *center of the curve*, and in this case the curve is said to be *central*. The center coordinates (x_0, y_0) satisfy the system of equations

$$\begin{aligned} a_{11}x_0 + a_{12}y_0 + a_{13} &= 0, \\ a_{21}x_0 + a_{22}y_0 + a_{23} &= 0, \end{aligned} \tag{4.4.5.7}$$

which implies that

$$x_0 = -\frac{1}{\delta}\begin{vmatrix} a_{13} & a_{12} \\ a_{23} & a_{22} \end{vmatrix}, \quad y_0 = -\frac{1}{\delta}\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}. \tag{4.4.5.8}$$

All diameters of a second-order curve with $\delta = 0$ are parallel or coincide. A second-order curve does not have a center if and only if $\delta = 0$ and $\Delta \neq 0$. A second-order curve has a center line if and only if $\delta = 0$ and $\Delta = 0$.

Example 1. The centers of nine canonical second-order curves are as follows:
1. Ellipse, $\delta = 1$ and $\Delta = -1$: the single center $O(0, 0)$;
2. Hyperbola, $\delta = -1$ and $\Delta = 1$: the single center $O(0, 0)$;
3. Parabola, $\delta = 0$ and $\Delta = -1$: no centers;
4. Imaginary ellipse, $\delta = 1$ and $\Delta = -1$: the single center $O(0, 0)$;
5. Pair of intersecting straight lines, $\delta = -1$ and $\Delta = 0$: the single center $O(0, 0)$;
6. Pair of imaginary intersecting straight lines, $\delta = 1$ and $\Delta = 0$: the single center $O(0, 0)$;
7. Pair of imaginary parallel straight lines, $\delta = 0$ and $\Delta = 0$: the center line $y = 0$;
8. Pair of coinciding straight lines, $\delta = 0$ and $\Delta = 0$: the center line $y = 0$.

Each of the two conjugate diameters of a central second-order curve bisects the chords parallel to the other diameter.

4.4.5-6. Principal axes.

A diameter perpendicular to the chords conjugate to it is called a *principal axis*. A principal axis is a symmetry axis of a second-order curve. For each central second-order curve ($\delta \neq 0$), either there are two perpendicular principal axes or each of its diameters is a principal axis (the circle). A second-order curve with $\delta = 0$ has a unique principal axis. The points of intersection of a second-order curve with its principal axes are called its *vertices*.

The directions of principal axes coincide with the directions of the eigenvectors of the symmetric matrix $[a_{ij}]$ $(i, j = 1, 2)$; i.e., the direction cosines $\cos\theta$, $\sin\theta$ of the normals to the principal axes are determined from the system of equations

$$\begin{aligned} (a_{11} - \lambda)\cos\theta + a_{12}\sin\theta &= 0, \\ a_{21}\cos\theta + (a_{22} - \lambda)\sin\theta &= 0, \end{aligned} \tag{4.4.5.9}$$

where λ is a nonzero root of characteristic equation (4.4.5.4).

The directions of principal axes and of their conjugate chords are called the *principal directions* of a second-order curve. The angle between the positive direction of the axis OX and each of the two principal directions of a second-order curve is given by the formula

$$\tan 2\varphi = \tan 2\theta = \frac{2a_{12}}{a_{11} - a_{22}}. \tag{4.4.5.10}$$

Remark. The circle has undetermined principal directions.

4.4.5-7. Reduction of central second-order curves to canonical form.

A second-order curve with $\delta \neq 0$ has a center. By shifting the origin to the center $O_1(x_0, y_0)$ whose coordinates are determined by formula (4.4.5.8), we can reduce equation (4.4.5.1) to the form

$$a_{11}x_1^2 + 2a_{12}x_1y_1 + a_{22}y_1^2 + \frac{\Delta}{\delta} = 0, \tag{4.4.5.11}$$

where x_1 and y_1 are the new coordinates.

By rotating the axes O_1X_1 and O_1Y_1 by the angle θ determined by (4.4.5.10), we transform equation (4.4.5.11) as follows:

$$A_{11}\hat{x}^2 + A_{22}\hat{y}^2 + \frac{\Delta}{\delta} = 0. \tag{4.4.5.12}$$

The coefficients A_{11} and A_{22} are the roots of the characteristic equation (4.4.5.4).

We note the following formulas for the ellipse:

$$a^2 = -\frac{1}{\lambda_2}\frac{\Delta}{\delta} = -\frac{\Delta}{\lambda_1\lambda_2^2}, \quad b^2 = -\frac{1}{\lambda_1}\frac{\Delta}{\delta} = -\frac{\Delta}{\lambda_1^2\lambda_2}, \tag{4.4.5.13}$$

where a and b are the parameters of the canonical equation, δ and Δ are the invariants, and λ_1 and λ_2 are the roots of the characteristic equation (4.4.5.4).

Similarly, for the hyperbola one has

$$a^2 = -\frac{1}{\lambda_1}\frac{\Delta}{\delta} = -\frac{\Delta}{\lambda_1^2\lambda_2}, \quad b^2 = \frac{1}{\lambda_1}\frac{\Delta}{\delta} = \frac{\Delta}{\lambda_1^2\lambda_2}. \tag{4.4.5.14}$$

4.4.5-8. Reduction of noncentral second-order curves to canonical form.

If $\delta = 0$, then the curve does not have any center or does not have a definite center, and its equation can be written as

$$(\alpha x + \beta y)^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0. \tag{4.4.5.15}$$

If the coefficients a_{13} and a_{23} are proportional to the coefficients α and β, i.e., $a_{13} = k\alpha$ and $a_{23} = k\beta$, then equation (4.4.5.15) becomes $(\alpha x + \beta y)^2 + 2k(\alpha x + \beta y) + a_{33} = 0$, and hence

$$\alpha x + \beta y = -k \pm \sqrt{k^2 - a_{33}} \tag{4.4.5.16}$$

is a pair of real parallel straight lines.

If the coefficients a_{13} and a_{23} are not proportional to the coefficients α and β, then equation (4.4.5.15) can be written as

$$(\alpha x + \beta y + \gamma)^2 + 2k(\beta x - \alpha x + q) = 0. \tag{4.4.5.17}$$

The parameters k, γ, and q can be determined by comparing the coefficients in equations (4.4.5.15) and (4.4.5.17). For the axis O_1X one should take the line $\alpha x + \beta y + \gamma = 0$, and for the axis O_1Y, the line $\beta x - \alpha x + q = 0$. We denote

$$\hat{x} = \frac{\beta x - \alpha x + q}{\pm\sqrt{\alpha^2 + \beta^2}}, \quad \hat{y} = \frac{\alpha x + \beta y + \gamma}{\pm\sqrt{\alpha^2 + \beta^2}}; \tag{4.4.5.18}$$

then equation (4.4.5.17) acquires the form

$$\hat{y}^2 = 2p\hat{x}, \tag{4.4.5.19}$$

where $p = |k|/\sqrt{\alpha^2 + \beta^2}$. The axis O_1X points into the half-plane where the sign of $\beta x - \alpha x + q$ is opposite to that of k.

If one only needs to find the canonical equation of a parabola and it is not necessary to construct the graph of the parabola in the coordinate system OXY, then the parameter p is determined via the invariants I, δ, and Δ and the roots λ_1 and λ_2 ($\lambda_1 \geq \lambda_2$) of the characteristic equation (4.4.5.4) by the formulas

$$p = \frac{1}{I}\sqrt{-\frac{\Delta}{I}} = \frac{1}{\lambda_1}\sqrt{-\frac{\Delta}{\lambda_1}} > 0, \quad \lambda_2 = 0. \tag{4.4.5.20}$$

4.4.5-9. Geometric definition of nondegenerate second-order curve.

There exists a coordinate system in which equation (4.4.5.1) has the form

$$y^2 = 2px - (1 - e^2)x^2, \tag{4.4.5.21}$$

where $p > 0$ is a parameter and e is the eccentricity. Obviously, the curve (4.4.5.21) passes through the origin of the new coordinate system. The axis OX is a symmetry axis of the curve.

The equation of the directrix of the curve (4.4.5.21) is

$$x = -\frac{p}{(1 + e)e}. \tag{4.4.5.22}$$

The coordinates of the focus are

$$x = \frac{p}{1 + e}, \quad y = 0. \tag{4.4.5.23}$$

The distance from the focus to the directrix is equal to p/e. For a central second-order curve, the line

$$x = \frac{p}{1 - e^2} = a \tag{4.4.5.24}$$

is a symmetry axis.

Remark. All types of second-order curves can be obtained as plane sections of a right circular cone for various positions of the secant plane with respect to the cone.

4.4.5-10. Tangents and normals to second-order curves.

The *equation of the tangent* to a second-order curve at a point $M_0(x_0, y_0)$ has the form

$$(a_{11}x_0 + a_{12}y_0 + a_{13})x + (a_{21}x_0 + a_{22}y_0 + a_{23})y + a_{31}x_0 + a_{32}y_0 + a_{33} = 0. \tag{4.4.5.25}$$

The *equation of the normal* to a second-order curve at the point $M_0(x_0, y_0)$ has the form

$$\frac{x - x_0}{a_{11}x_0 + a_{12}y_0 + a_{13}} = \frac{y - y_0}{a_{21}x_0 + a_{22}y_0 + a_{23}}. \tag{4.4.5.26}$$

TABLE 4.2
Ellipse, hyperbola, parabola. Main formulas

	Ellipse	Hyperbola	Parabola
Canonical equation	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$	$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$	$y^2=2px$
Equation in polar coordinates	$\rho=\frac{p}{1-e\cos\varphi}$	$\rho=\frac{\pm p}{1\mp e\cos\varphi}$	$\rho=\frac{p}{1-\cos\varphi}$
Eccentricity	$e=\frac{c}{a}=\sqrt{1-\frac{b^2}{a^2}}<1$	$e=\frac{c}{a}=\sqrt{1+\frac{b^2}{a^2}}>1$	$e=\frac{c}{a}=1$
Foci	$(\pm ae,0)$	$(\pm ae,0)$	$\left(\frac{p}{2},0\right)$
Focal radii (distance from the foci to an arbitrary point (x,y) of curve)	$r_1=a+ex$ $r_2=a-ex$	$r_1=\begin{cases}a+ex & \text{for } x>0\\ -a-ex & \text{for } x<0\end{cases}$ $r_2=\begin{cases}-a+ex & \text{for } x>0\\ a-ex & \text{for } x<0\end{cases}$	$r=x+\frac{p}{2}$
Focal parameter	$p=\frac{b^2}{a}$	$p=\frac{b^2}{a}$	p
Equation of directrices	$x=\pm\frac{a}{e}$	$x=\pm\frac{a}{e}$	$x=-\frac{p}{2}$
Equation of diameter conjugate to chords with slope k	$y=-\frac{b^2}{a^2k}x$	$y=\frac{b^2}{a^2k}x$	$y=\frac{p}{k}$
Area of segment bounded by an arc convex to the left and the chord joining points (x_0,y_0) and $(x_0,-y_0)$	$\frac{\pi ab}{2}+\frac{b}{a}\left(x_0\sqrt{a^2-x_0^2}+a^2\arcsin\frac{x_0}{a}\right)$	$x_0y_0-ab\ln\left(\frac{x_0}{a}+\frac{y_0}{b}\right)$ $=x_0y_0-ab\operatorname{arccosh}\frac{x_0}{a}$	$\frac{4}{3}x_0y_0$
Curvature radius at point (x,y)	$a^2b^2\left(\frac{x^2}{a^4}+\frac{y^2}{b^4}\right)^{3/2}=\frac{\sqrt{(r_1r_2)^3}}{ab}$	$a^2b^2\left(\frac{x^2}{a^4}+\frac{y^2}{b^4}\right)^{3/2}=\frac{\sqrt{(r_1r_2)^3}}{ab}$	$\frac{(p+2x)^{3/2}}{\sqrt{p}}$
Equations of tangents to a curve which pass through an arbitrary point (x_0,y_0)	$\frac{y-y_0}{x-x_0}$ $=\frac{-x_0y_0\pm\sqrt{a^2y_0^2+b^2x_0^2-a^2b^2}}{a^2-x_0^2}$	$\frac{y-y_0}{x-x_0}$ $=\frac{-x_0y_0\pm\sqrt{a^2y_0^2-b^2x_0^2+a^2b^2}}{a^2-x_0^2}$	$\frac{y-y_0}{x-x_0}$ $=\frac{y_0\pm\sqrt{y_0^2-2px_0}}{2x_0}$
Equation of tangent at point (x_0,y_0)	$\frac{x_0x}{a^2}+\frac{y_0y}{b^2}=1$	$\frac{x_0x}{a^2}-\frac{y_0y}{b^2}=1$	$yy_0=p(x+x_0)$
Equation of tangent with slope k	$y=kx\pm\sqrt{k^2a^2+b^2}$	$y=kx\pm\sqrt{k^2a^2-b^2}$	$y=kx+\frac{p}{2k}$
Equation of normal at point (x_0,y_0)	$\frac{y-y_0}{x-x_0}=\frac{a^2y_0}{b^2x_0}$	$\frac{y-y_0}{x-x_0}=-\frac{a^2y_0}{b^2x_0}$	$\frac{y-y_0}{x-x_0}=-\frac{y_0}{p}$
Coordinates of pole (x_0,y_0) of straight line $Ax+By+C=0$ w.r.t. a curve	$x_0=-\frac{a^2A}{C},\quad y_0=-\frac{b^2B}{C}$	$x_0=-\frac{a^2A}{C},\quad y_0=\frac{b^2B}{C}$	$x_0=\frac{A}{C},\ y_0=-\frac{pB}{C}$

The *polar* is the set of points Q that are harmonically conjugate to a point P, called the pole of a second-order curve, with respect to points R_1 and R_2 of intersection of the second-order curve with secants passing through P; i.e.,

$$\frac{R_1P}{PR_2}=-\frac{R_1Q}{QR_2}.$$

If the point P lies outside the second-order curve (one can draw two tangents through P), then the polar line passes through the points at which this curve is tangent to straight lines drawn through point P. If the point lies on a second-order curve, then the polar line is a straight line tangent to this curve at this point. If the polar line of a point P passes through a point Q, then the polar line of Q passes through P.

4.5. Coordinates, Vectors, Curves, and Surfaces in Space

4.5.1. Vectors. Cartesian Coordinate System

4.5.1-1. Notion of vector.

A directed segment with initial point A and endpoint B (see Fig. 4.26) is called the *vector* $\overrightarrow{AB}$. A nonnegative number equal to the length of the segment AB joining the points A and B is called the *length* $|\overrightarrow{AB}|$ of the *vector* $\overrightarrow{AB}$. The vector $\overrightarrow{BA}$ is said to be *opposite to the vector* $\overrightarrow{AB}$.

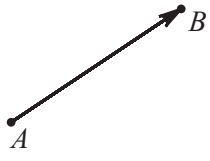


Figure 4.26. Vector.

Two directed segments $\overrightarrow{AB}$ and $\overrightarrow{CD}$ of the same length and the same direction determine the same vector **a**; i.e., $\mathbf{a} = \overrightarrow{AB} = \overrightarrow{CD}$.

Two vectors are said to be *collinear (parallel)* if they lie on the same straight line or on parallel lines. Three vectors are said to be *coplanar* if they lie in the same plane or in parallel planes. A vector 0 whose initial point and endpoint coincide is called the *zero vector*. The length of the zero vector is equal to zero ($|0| = 0$), and the direction of the zero vector is assumed to be arbitrary. A vector **e** of unit length is called a *unit vector*.

The *sum* $\mathbf{a} + \mathbf{b}$ *of vectors* **a** *and* **b** is defined as the vector directed from the initial point of **a** to the endpoint of **b** under the condition that **b** is applied at the endpoint of **a**. The rule for addition of vectors, which is contained in this definition, is called the *triangle rule* or the *rule of closing a chain of vectors* (see Fig. 4.27*a*). The sum $\mathbf{a} + \mathbf{b}$ can also be found using the *parallelogram rule* (see Fig. 4.27*b*). The *difference* $\mathbf{a} - \mathbf{b}$ *of vectors* **a** *and* **b** is defined as follows: $\mathbf{b} + (\mathbf{a} - \mathbf{b}) = \mathbf{a}$ (see Fig. 4.27*c*).

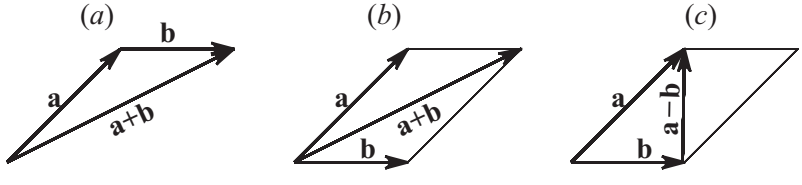


Figure 4.27. The sum of vectors: triangle rule (*a*) and parallelogram rule (*b*). The difference of vectors (*c*).

The *product* $\lambda\mathbf{a}$ *of a vector* **a** *by a number* λ is defined as the vector whose length is equal to $|\lambda\mathbf{a}| = |\lambda||\mathbf{a}|$ and whose direction coincides with that of the vector **a** if $\lambda > 0$ or is opposite to the direction of the vector **a** if $\lambda < 0$.

Remark. If $\mathbf{a} = 0$ or $\lambda = 0$, then the absolute value of the product is zero, i.e., it is the zero vector. In this case, the direction of the product $\lambda\mathbf{a}$ is undetermined.

Main properties of operations with vectors:

1. $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ (commutativity).
2. $\mathbf{a} + (\mathbf{b} + \mathbf{c}) = \mathbf{a} + (\mathbf{b} + \mathbf{c})$ (associativity of addition).
3. $\mathbf{a} + 0 = \mathbf{a}$ (existence of the zero vector).
4. $\mathbf{a} + (-\mathbf{a}) = 0$ (existence of the opposite vector).
5. $\lambda(\mathbf{a} + \mathbf{b}) = \lambda\mathbf{a} + \lambda\mathbf{b}$ (distributivity with respect to addition of vectors).
6. $(\lambda + \mu)\mathbf{a} = \lambda\mathbf{a} + \mu\mathbf{a}$ (distributivity with respect to addition of constants).
7. $\lambda(\mu\mathbf{a}) = (\lambda\mu)\mathbf{a}$ (associativity of product).
8. $1\mathbf{a} = \mathbf{a}$ (multiplication by unity).

4.5.1-2. Projection of vector onto axis.

A straight line with a unit vector $\mathbf{e}$ lying on it determining the positive sense of the line is called an *axis*. The *projection* $\mathrm{pr}_{\mathbf{e}}\,\mathbf{a}$ of a vector $\mathbf{a}$ onto the axis (see Fig. 4.28) is defined as the directed segment on the axis whose signed length is equal to the scalar product of $\mathbf{a}$ by the unit vector $\mathbf{e}$, i.e., is determined by the formula

$$\mathrm{pr}_{\mathbf{e}}\,\mathbf{a} = |\mathbf{a}|\cos\varphi, \tag{4.5.1.1}$$

where φ is the angle between the vectors $\mathbf{a}$ and $\mathbf{e}$.

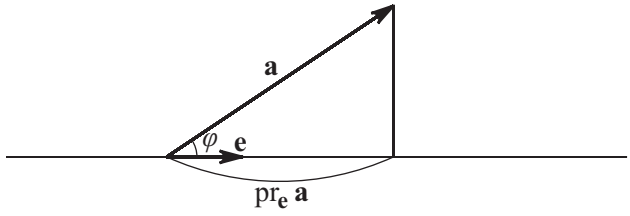


Figure 4.28. Projection of a vector onto the axes.

Properties of projections:

1. $\mathrm{pr}_{\mathbf{e}}(\mathbf{a} + \mathbf{b}) = \mathrm{pr}_{\mathbf{e}}\,\mathbf{a} + \mathrm{pr}_{\mathbf{e}}\,\mathbf{b}$ (additivity).
2. $\mathrm{pr}_{\mathbf{e}}(\lambda\mathbf{a}) = \lambda\,\mathrm{pr}_{\mathbf{e}}\,\mathbf{a}$ (homogeneity).

4.5.2. Coordinate Systems

4.5.2-1. Cartesian coordinate system.

If a one-to-one correspondence between points in space and numbers (triples of numbers) is given, then one says that a *coordinate system* is introduced in space.

A *rectangular Cartesian coordinate system* is determined by a scale segment for measuring lengths and three pairwise perpendicular directed straight lines OX, OY, and OZ (the *coordinate axes*) concurrent at a single point O (the *origin*). The three coordinate axes divide the space into eight parts called *octants*.

We choose an arbitrary point M in space and project it onto the coordinate axes, i.e., draw the perpendiculars to the axes OX, OY, and OZ through M. We denote the points

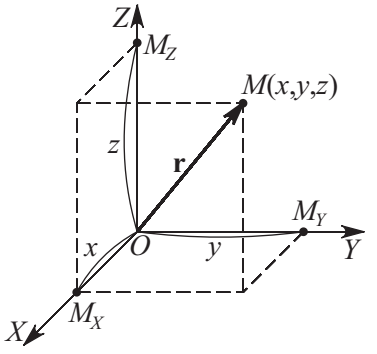


Figure 4.29. Point in rectangular Cartesian coordinate system.

of intersection of the perpendiculars with the axes by M_X, M_Y, and M_Z, respectively. The numbers (see Fig. 4.29)

$$x = OM_X, \quad y = OM_Y, \quad z = OM_Z, \tag{4.5.2.1}$$

where OM_X, OM_Y, and OM_Z are the signed lengths of the directed segments $\overrightarrow{OM}_X$, $\overrightarrow{OM}_Y$, and $\overrightarrow{OM}_Z$ of the axes OX, OY, and OZ, respectively, are called the *coordinates of the point* M in the rectangular Cartesian coordinate system. The number x is called the first coordinate or the *abscissa* of the point M, the number y is called the second coordinate or the *ordinate* of the point M, and the number z is called the third coordinate or the *applicate* of the point M. Usually one says that the point M has the coordinates (x, y, z), and the notation $M(x, y, z)$ is used.

To each point M of three-dimensional space one can assign its position vector. The directed segment $\overrightarrow{OM}$ is called the *position vector of the point* M. The position vector determines the vector $\mathbf{r}$ ($\mathbf{r} = \overrightarrow{OM}$) whose coordinates are its projections on the axes OX, OY, and OZ, respectively. Obviously, the triple (x, y, z) of numbers can be called the point M whose coordinates are these numbers, or the position vector $\overrightarrow{OM}$ whose projections are these numbers. An arbitrary vector (x, y, z) can be represented as

$$(x, y, z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}, \tag{4.5.2.2}$$

where $\mathbf{i} = (1, 0, 0)$, $\mathbf{j} = (0, 1, 0)$, and $\mathbf{k} = (0, 0, 1)$ are the unit vectors with the same directions as the coordinate axes OX, OY, and OZ. The distance between points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$ is determined by the formula

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} = |\mathbf{r}_2 - \mathbf{r}_1|, \tag{4.5.2.3}$$

where $\mathbf{r}_2 = \overrightarrow{OM}_2$ and $\mathbf{r}_1 = \overrightarrow{OM}_1$ are the position vectors of the points M_1 and M_2, respectively (see Fig. 4.30).

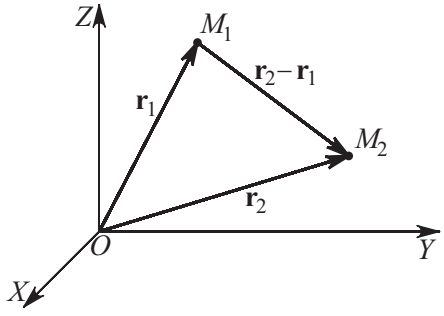


Figure 4.30. Distance between points.

Two vectors $\mathbf{a} = (x_1, y_1, z_1)$ and $\mathbf{b} = (x_2, y_2, z_2)$ are equal to each other if and only if the following relations hold simultaneously:

$$x_1 = x_2, \quad y_1 = y_2, \quad z_1 = z_2.$$

For arbitrary vectors, one has the relations

$$\begin{aligned} &(x_1, y_1, z_1) \pm (x_2, y_2, z_2) = (x_1 \pm x_2, y_1 \pm y_2, z_1 \pm z_2), \\ &\alpha(x, y, z) = (\alpha x, \alpha y, \alpha z). \end{aligned} \tag{4.5.2.4}$$

If a point M divides the directed segment $\overrightarrow{M_1M_2}$ in the ratio $p/q = \lambda$, then the coordinates of this point are given by the formulas

$$\begin{aligned} x &= \frac{x_1 + \lambda x_2}{1 + \lambda} = \frac{qx_1 + px_2}{q + p}, \\ y &= \frac{y_1 + \lambda y_2}{1 + \lambda} = \frac{qy_1 + py_2}{q + p}, \qquad \text{or} \qquad \mathbf{r} = \frac{\mathbf{r}_1 + \lambda \mathbf{r}_2}{1 + \lambda} = \frac{q\mathbf{r}_1 + p\mathbf{r}_2}{q + p}, \\ z &= \frac{z_1 + \lambda z_2}{1 + \lambda} = \frac{qz_1 + pz_2}{q + p}, \end{aligned} \tag{4.5.2.5}$$

where $-\infty \le \lambda \le \infty$. In the special case of the midpoint of the segment $\overrightarrow{M_1M_2}$ ($p = q$, $\lambda = 1$), the coordinates are

$$x = \frac{x_1 + x_2}{2}, \quad y = \frac{y_1 + y_2}{2}, \quad z = \frac{z_1 + z_2}{2} \qquad \text{or} \qquad \mathbf{r} = \frac{\mathbf{r}_1 + \mathbf{r}_2}{2};$$

i.e., each coordinate of the midpoint of a segment is equal to the half-sum of the corresponding coordinates of its endpoints.

The angles α, β, and γ between the segment $\overrightarrow{M_1M_2}$ and the coordinate axes OX, OY, and OZ are determined by the expressions

$$\cos\alpha = \frac{x_2 - x_1}{|\mathbf{r}_2 - \mathbf{r}_1|}, \quad \cos\beta = \frac{y_2 - y_1}{|\mathbf{r}_2 - \mathbf{r}_1|}, \quad \cos\gamma = \frac{z_2 - z_1}{|\mathbf{r}_2 - \mathbf{r}_1|}, \tag{4.5.2.6}$$

and

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1.$$

The numbers $\cos\alpha$, $\cos\beta$, and $\cos\gamma$ are called the *direction cosines of the segment* $\overrightarrow{M_1M_2}$.

The angle φ between arbitrary directed segments $\overrightarrow{M_1M_2}$ and $\overrightarrow{M_3M_4}$ joining the points $M_1(x_1, y_1, z_1)$, $M_2(x_2, y_2, z_2)$ and $M_3(x_3, y_3, z_3)$, $M_4(x_4, y_4, z_4)$, respectively, can be found from the relation

$$\cos\varphi = \frac{(x_2 - x_1)(x_4 - x_3) + (y_2 - y_1)(y_4 - y_3) + (z_2 - z_1)(z_4 - z_3)}{|\mathbf{r}_2 - \mathbf{r}_1|\,|\mathbf{r}_4 - \mathbf{r}_3|}. \tag{4.5.2.7}$$

The area of the triangle with vertices M_1, M_2, and M_3 is given by the formula

$$S = \frac{1}{4}\sqrt{\begin{vmatrix} y_1 & z_1 & 1 \\ y_2 & z_2 & 1 \\ y_3 & z_3 & 1 \end{vmatrix}^2 + \begin{vmatrix} z_1 & x_1 & 1 \\ z_2 & x_2 & 1 \\ z_3 & x_3 & 1 \end{vmatrix}^2 + \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}^2}. \tag{4.5.2.8}$$

The volume of the pyramid with vertices M_1, M_2, M_3, and M_4 is equal to

$$V = \pm\frac{1}{6}\begin{vmatrix} x_2-x_1 & y_2-y_1 & z_2-z_1 \\ x_3-x_1 & y_3-y_1 & z_3-z_1 \\ x_4-x_1 & y_4-y_1 & z_4-z_1 \end{vmatrix} = \pm\frac{1}{6}\begin{vmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \\ 1 & x_4 & y_4 & z_4 \end{vmatrix}, \tag{4.5.2.9}$$

and the volume of the parallelepiped spanned by vectors $\overrightarrow{M_1M_2}$, $\overrightarrow{M_1M_3}$, and $\overrightarrow{M_1M_4}$ is equal to

$$V = \pm\begin{vmatrix} x_2-x_1 & y_2-y_1 & z_2-z_1 \\ x_3-x_1 & y_3-y_1 & z_3-z_1 \\ x_4-x_1 & y_4-y_1 & z_4-z_1 \end{vmatrix} = \pm\begin{vmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \\ 1 & x_4 & y_4 & z_4 \end{vmatrix}. \tag{4.5.2.10}$$

The coordinate surfaces, on which one of the coordinates is constant, are planes parallel to the coordinate planes, and the coordinate lines, along which only one coordinate varies, are straight lines parallel to the coordinate axes. Coordinate surfaces meet in the coordinate lines.

4.5.2-2. Transformation of Cartesian coordinates under parallel translation of axes.

Suppose that two rectangular Cartesian coordinate systems $OXYZ$ and $\widehat{O}\widehat{X}\widehat{Y}\widehat{Z}$ are given and the first system can be made to coincide with the second system by translating the origin O of the first system to the origin $\widehat{O}$ of the second system. Under this translation the axes preserve their direction (the respective axes of the systems are parallel), and the origin moves by x_0 in the direction of the axis OX, by y_0 in the direction of the axis OY, and by z_0 in the direction of the axis OZ. Obviously, the point $\widehat{O}$ in the coordinate system $OXYZ$ has the coordinates (x_0, y_0, z_0).

An arbitrary point M has coordinates (x, y, z) in the system $OXYZ$ and coordinates $(\hat{x}, \hat{y}, \hat{z})$ in the system $\widehat{O}\widehat{X}\widehat{Y}\widehat{Z}$. The transformation of rectangular Cartesian coordinates by the parallel translation of axes is determined by the formulas

$$\begin{aligned} x &= \hat{x} + x_0, \\ y &= \hat{y} + y_0, \\ z &= \hat{z} + z_0 \end{aligned} \qquad \text{or} \qquad \begin{aligned} \hat{x} &= x - x_0, \\ \hat{y} &= y - y_0, \\ \hat{z} &= z - z_0. \end{aligned} \tag{4.5.2.11}$$

4.5.2-3. Transformation of Cartesian coordinates under rotation of axes.

Suppose that two rectangular Cartesian coordinate systems $OXYZ$ and $O\widehat{X}\widehat{Y}\widehat{Z}$ are given and the first system can be made to coincide with the second system by rotating the first system around the point O.

An arbitrary point M has coordinates (x, y, z) in the system $OXYZ$ and coordinates $(\hat{x}, \hat{y}, \hat{z})$ in the system $O\widehat{X}\widehat{Y}\widehat{Z}$. If the axis $O\widehat{X}$ has the direction cosines e_{11}, e_{21}, e_{31}, the axis $O\widehat{Y}$ has the direction cosines e_{12}, e_{22}, e_{32}, and the axis $O\widehat{Z}$ has the direction cosines e_{13}, e_{23}, e_{33} in the coordinate system $OXYZ$, then the axis OX has the direction cosines e_{11}, e_{12}, e_{13}, the axis OY has the direction cosines e_{21}, e_{22}, e_{23}, and the axis OZ has the

direction cosines e_{31}, e_{32}, e_{33} in the coordinate system $O\widehat{X}\widehat{Y}\widehat{Z}$. The transformation of rectangular Cartesian coordinates by the rotation of axes is determined by the formulas

$$\begin{aligned} x &= e_{11}\hat{x} + e_{12}\hat{y} + e_{13}\hat{z}, \\ y &= e_{21}\hat{x} + e_{22}\hat{y} + e_{23}\hat{z}, \\ z &= e_{31}\hat{x} + e_{32}\hat{y} + e_{33}\hat{z}, \end{aligned} \qquad \text{or} \qquad \begin{aligned} \hat{x} &= e_{11}x + e_{21}y + e_{31}z, \\ \hat{y} &= e_{12}x + e_{22}y + e_{32}z, \\ \hat{z} &= e_{13}x + e_{23}y + e_{33}z. \end{aligned} \tag{4.5.2.12}$$

4.5.2-4. Transformation of coordinates under translation and rotation of axes.

Suppose that two rectangular Cartesian coordinate systems $OXYZ$ and $\widehat{O}\widehat{X}\widehat{Y}\widehat{Z}$ are given and the first system can be made to coincide with the second system by translating the origin $O(0,0,0)$ of the first system to the origin $\widehat{O}(x_0, y_0, z_0)$ of the second system, and then by rotating the first system around the point $\widehat{O}$ (see Paragraphs 4.5.1-2 and 4.5.1-3).

An arbitrary point M has coordinates (x, y, z) in the system $OXYZ$ and coordinates $(\hat{x}, \hat{y}, \hat{z})$ in the system $\widehat{O}\widehat{X}\widehat{Y}\widehat{Z}$. The transformation of rectangular Cartesian coordinates by the parallel translation and the rotation of axes is determined by the formulas

$$\begin{aligned} x &= e_{11}\hat{x} + e_{12}\hat{y} + e_{13}\hat{z} + x_0, \\ y &= e_{21}\hat{x} + e_{22}\hat{y} + e_{23}\hat{z} + y_0, \\ z &= e_{31}\hat{x} + e_{32}\hat{y} + e_{33}\hat{z} + z_0, \end{aligned} \qquad \text{or} \qquad \begin{aligned} \hat{x} &= e_{11}(x - x_0) + e_{21}(y - y_0) + e_{31}(z - z_0), \\ \hat{y} &= e_{12}(x - x_0) + e_{22}(y - y_0) + e_{32}(z - z_0), \\ \hat{z} &= e_{13}(x - x_0) + e_{23}(y - y_0) + e_{33}(z - z_0). \end{aligned} \tag{4.5.2.13}$$

4.5.2-5. Cylindrical and spherical coordinates.

A more general curvilinear coordinate system is obtained if one introduces three families of coordinate surfaces such that exactly one surface of each family passes through each point of space. The position of a point in such a system is determined by the values of the parameters of the coordinate surfaces passing through this point. The most commonly used curvilinear coordinate systems (cylindrical and spherical) are described below.

The *cylindrical coordinates* of a point M are defined as the polar coordinates ρ and φ (see Paragraph 4.1.2-5) of the projection of M onto the base plane (usually OXY) and the distance (usually z) from M to the base plane, which is called the *applicate* (see Fig. 4.31*a*). To be definite, one usually assumes that $0 < \varphi \le 2\pi$ or $-\pi < \varphi \le \pi$. For cylindrical coordinates, the coordinate surfaces are the planes $z = \text{const}$ perpendicular to the axis OZ, the half-planes $\varphi = \text{const}$ bounded by the axis OZ, and the cylindrical surfaces $\rho = \text{const}$ with axis OZ. The coordinate surfaces intersect in the coordinate lines.

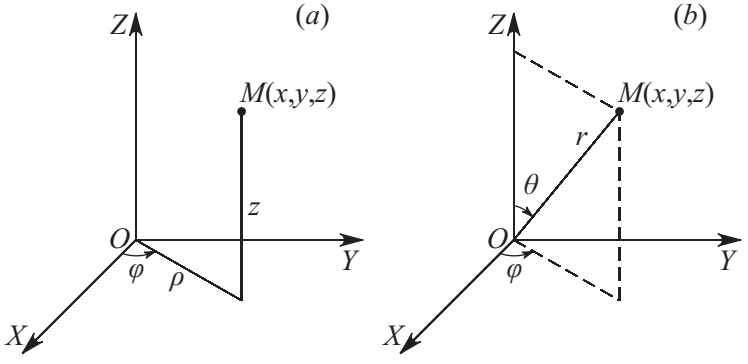


Figure 4.31. Point in cylindrical (*a*) and spherical (*b*) coordinates.

The *spherical (polar) coordinates* are defined as the length $r = |\overrightarrow{OM}|$ of the radius vector, the *longitude* φ, and the polar distance θ, called the *latitude* (see Fig. 4.31*b*). To be definite, one usually assumes that $0 < \varphi \le 2\pi$, $0 \le \theta \le \pi$ or $-\pi < \varphi \le \pi$, and $0 \le \theta \le \pi$. For spherical coordinates, the coordinate surfaces are the spheres $r = \text{const}$ centered at the origin, the half-planes $\varphi = \text{const}$ bounded by the axis OZ, and the cones $\theta = \text{const}$ with vertex O and axis OZ. The coordinate surfaces intersect in the coordinate lines.

4.5.2-6. Relationship between Cartesian, cylindrical, and spherical coordinates.

Let M be an arbitrary point in space with rectangular Cartesian coordinates (x, y, z), cylindrical coordinates (ρ, φ, z), and spherical coordinates (r, φ, θ). The formulas of transition from the cylindrical coordinate system to the Cartesian coordinate system and vice versa have the form

$$\begin{aligned} &x = \rho\cos\varphi, && \rho = \sqrt{x^2 + y^2}, \\ &y = \rho\sin\varphi, \qquad \text{or} && \tan\varphi = y/x, \\ &z = z, && z = z, \end{aligned} \tag{4.5.2.14}$$

where the polar angle φ is taken with regard to the quadrant in which the projection of the point M onto the base plane lies. The formulas of transition from the spherical coordinate system to the Cartesian coordinate system and vice versa have the form

$$\begin{aligned} &x = r\sin\theta\cos\varphi, && r = \sqrt{x^2 + y^2 + z^2}, \\ &y = r\sin\theta\sin\varphi, \qquad \text{or} && \tan\varphi = y/x, \\ &z = r\cos\theta, && \tan\theta = \sqrt{x^2 + y^2}/z, \end{aligned} \tag{4.5.2.15}$$

where the angle φ is determined from the same considerations as in the case of cylindrical coordinates.

4.5.2-7. Surfaces and curves in space.

A *surface in space* determined by an equation in some coordinate system is the locus of points in space whose coordinates satisfy this equation.

An *equation of a surface in space* in a given coordinate system is an equation with three variables satisfied by the coordinates of points lying on the surface and not satisfied by the coordinates of points that do not lie on the surface.

The coordinates of an arbitrary point of the surface occurring in the equation of the surface are called the *current coordinates*.

Example 1. The equation

$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2$$

defines the sphere of radius r centered at the point (x_0, y_0, z_0), i.e., the locus of points lying at the distance r from the point (x_0, y_0, z_0).

Example 2. The equation $x^2 + y^2 + (z - 1)^2 = 0$ determines the single point with coordinates $(0, 0, 1)$.

Example 3. The equation $x^2 + y^2 + z^2 + 1 = 0$ does not have solutions for any real x, y, and z.

In the general case, the *equation of a surface in the Cartesian coordinate system* $OXYZ$ can be written as

$$F(x, y, z) = 0. \tag{4.5.2.16}$$

On the other hand, a surface (continuous surface) can be defined parametrically; i.e., a *surface* is defined as the set of points whose coordinates satisfy the system of parametric equations

$$x = x(u, v), \quad y = y(u, v), \quad z = z(u, v) \tag{4.5.2.17}$$

for appropriate values of the parameters u and v.

In spatial analytic geometry, each curve is treated as the intersection of two surfaces and hence is defined by a system of two equations

$$F(x, y, z) = 0, \qquad G(x, y, z) = 0. \tag{4.5.2.18}$$

On the other hand, each curve (continuous curve) can be defined parametrically; i.e., a *curve* is defined as the set of points whose coordinates satisfy the system of parametric equations

$$x = x(t), \quad y = y(t), \quad z = z(t) \quad \text{or} \quad \mathbf{r} = \mathbf{r}(t) \qquad (-\infty \le t_1 \le t \le t_2 \le \infty). \tag{4.5.2.19}$$

4.5.3. Vectors. Products of Vectors

4.5.3-1. Scalar product of two vectors.

The *scalar product* of two vectors is defined as the product of their absolute values times the cosine of the angle between the vectors (see Fig. 4.32),

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos \varphi. \tag{4.5.3.1}$$

If the angle between vectors $\mathbf{a}$ and $\mathbf{b}$ is acute, then $\mathbf{a} \cdot \mathbf{b} > 0$; if the angle is obtuse, then $\mathbf{a} \cdot \mathbf{b} < 0$; if the angle is right, then $\mathbf{a} \cdot \mathbf{b} = 0$. Taking into account (4.5.1.1), we can write the scalar product as

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos \varphi = |\mathbf{a}| \operatorname{pr}_{\mathbf{a}} \mathbf{b} = |\mathbf{b}| \operatorname{pr}_{\mathbf{b}} \mathbf{a}. \tag{4.5.3.2}$$

Remark. The scalar product of a vector $\mathbf{a}$ by a vector $\mathbf{b}$ is also denoted by $(\mathbf{a}, \mathbf{b})$ or $\mathbf{ab}$.

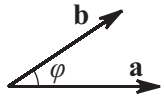


Figure 4.32. Scalar product of two vectors.

The *angle φ between vectors* is determined by the formula

$$\cos \varphi = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}||\mathbf{b}|} = \frac{a_x b_x + a_y b_y + a_z b_z}{\sqrt{a_x^2 + a_y^2 + a_z^2}\sqrt{b_x^2 + b_y^2 + b_z^2}}. \tag{4.5.3.3}$$

Properties of scalar product:

1. $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$ (commutativity).
2. $\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$ (distributivity with respect to addition of vectors). This property holds for any number of summands.
3. If vectors $\mathbf{a}$ and $\mathbf{b}$ are collinear, then $\mathbf{a} \cdot \mathbf{b} = \pm|\mathbf{a}||\mathbf{b}|$. (The sign + is taken if the vectors $\mathbf{a}$ and $\mathbf{b}$ have the same sense, and the sign – is taken if the senses are opposite.)
4. $(\lambda \mathbf{a}) \cdot \mathbf{b} = \lambda(\mathbf{a} \cdot \mathbf{b})$ (associativity with respect to a scalar factor).
5. $\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2$. The scalar product $\mathbf{a} \cdot \mathbf{a}$ is denoted by $\mathbf{a}^2$ (the *scalar square* of the vector $\mathbf{a}$).

6. The length of a vector is expressed via the scalar product by

$$|\mathbf{a}| = \sqrt{\mathbf{a} \cdot \mathbf{a}} = \sqrt{\mathbf{a}^2}.$$

7. Two nonzero vectors **a** and **b** are perpendicular if and only if **ab** = 0.
8. The scalar products of basis vectors are

$$\mathbf{i} \cdot \mathbf{j} = \mathbf{i} \cdot \mathbf{k} = \mathbf{j} \cdot \mathbf{k} = 0, \quad \mathbf{i} \cdot \mathbf{i} = \mathbf{j} \cdot \mathbf{j} = \mathbf{k} \cdot \mathbf{k} = 1.$$

9. If vectors are given by their coordinates, $\mathbf{a} = (a_x, a_y, a_z)$ and $\mathbf{b} = (b_x, b_y, b_z)$, then

$$\mathbf{a} \cdot \mathbf{b} = (a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{j})(b_x\mathbf{i} + b_y\mathbf{j} + b_z\mathbf{j}) = a_xb_x + a_yb_y + a_zb_z. \tag{4.5.3.4}$$

10. The *Cauchy–Schwarz inequality*

$$|\mathbf{a} \cdot \mathbf{b}| \le |\mathbf{a}||\mathbf{b}|.$$

11. The *Minkowski inequality*

$$|\mathbf{a} + \mathbf{b}| \le |\mathbf{a}| + |\mathbf{b}|.$$

4.5.3-2. Cross product of two vectors.

The *cross product* of a vector **a** by a vector **b** is defined as the vector **c** (see Fig. 4.33) satisfying the following three conditions:

1. Its absolute value is equal to the area of the parallelogram spanned by the vectors **a** and **b**; i.e.,

$$|\mathbf{c}| = |\mathbf{a} \times \mathbf{b}| = |\mathbf{a}||\mathbf{b}| \sin\varphi. \tag{4.5.3.5}$$

2. It is perpendicular to the plane of the parallelogram; i.e., $\mathbf{c} \perp \mathbf{a}$ and $\mathbf{c} \perp \mathbf{b}$.
3. The vectors **a**, **b**, and **c** form a *right-handed trihedral*; i.e., the vector **c** points to the side from which the sense of the shortest rotation from **a** to **b** is anticlockwise.

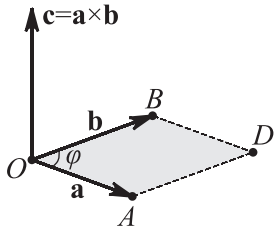


Figure 4.33. Cross product of two vectors.

Remark 1. The cross product of a vector **a** by a vector **b** is also denoted by $\mathbf{c} = [\mathbf{a}, \mathbf{b}]$.

Remark 2. If vectors **a** and **b** are collinear, then the parallelogram $OADB$ is degenerate and should be assigned the zero area. Hence the cross product of collinear vectors is defined to be the zero vector whose direction is arbitrary.

Properties of cross product:

1. $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$ (anticommutativity).
2. $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$ (distributivity with respect to the addition of vectors). This property holds for any number of summands.

3. Vectors $\mathbf{a}$ and $\mathbf{b}$ are collinear if and only if $\mathbf{a}\times\mathbf{b}=0$. In particular, $\mathbf{a}\times\mathbf{a}=0$ and $\mathbf{a}(\mathbf{a}\times\mathbf{b})=\mathbf{b}(\mathbf{a}\times\mathbf{b})=0$.
4. $(\lambda\mathbf{a})\times\mathbf{b}=\mathbf{a}\times(\lambda\mathbf{b})=\lambda(\mathbf{a}\times\mathbf{b})$ (associativity with respect to a scalar factor).
5. The cross product of basis vectors is

$$\mathbf{i}\times\mathbf{i}=\mathbf{j}\times\mathbf{j}=\mathbf{k}\times\mathbf{k}=0,\quad \mathbf{i}\times\mathbf{j}=\mathbf{k},\quad \mathbf{j}\times\mathbf{k}=\mathbf{i},\quad \mathbf{k}\times\mathbf{i}=\mathbf{j}.$$

6. If the vectors are given by their coordinates $\mathbf{a}=(a_x,a_y,a_z)$ and $\mathbf{b}=(b_x,b_y,b_z)$, then

$$\mathbf{a}\times\mathbf{b}=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\ a_x&a_y&a_z\\ b_x&b_y&b_z\end{vmatrix}=(a_yb_z-a_zb_y)\mathbf{i}+(a_zb_x-a_xb_z)\mathbf{j}+(a_xb_y-a_yb_x)\mathbf{k}. \quad (4.5.3.6)$$

7. The area of the parallelogram spanned by vectors $\mathbf{a}$ and $\mathbf{b}$ is equal to

$$S=|\mathbf{a}\times\mathbf{b}|=\sqrt{\begin{vmatrix}a_y&a_z\\ b_y&b_z\end{vmatrix}^2+\begin{vmatrix}a_x&a_z\\ b_x&b_z\end{vmatrix}^2+\begin{vmatrix}a_x&a_y\\ b_x&b_y\end{vmatrix}^2}. \quad (4.5.3.7)$$

8. The area of the triangle spanned by vectors $\mathbf{a}$ and $\mathbf{b}$ is equal to

$$S=\frac{1}{2}|\mathbf{a}\times\mathbf{b}|=\frac{1}{2}\sqrt{\begin{vmatrix}a_y&a_z\\ b_y&b_z\end{vmatrix}^2+\begin{vmatrix}a_x&a_z\\ b_x&b_z\end{vmatrix}^2+\begin{vmatrix}a_x&a_y\\ b_x&b_y\end{vmatrix}^2}. \quad (4.5.3.8)$$

Example 1. The moment with respect to the point O of a force $\mathbf{F}$ applied at a point M is the cross product of the position vector $\overrightarrow{OM}$ by the force $\mathbf{F}$; i.e., $\mathbf{M}=\overrightarrow{OM}\times\mathbf{F}$.

4.5.3-3. Conditions for vectors to be parallel or perpendicular.

A vector $\mathbf{a}$ is collinear to a vector $\mathbf{b}$ if

$$\mathbf{b}=\lambda\mathbf{a}\quad\text{or}\quad\mathbf{a}\times\mathbf{b}=0. \quad (4.5.3.9)$$

A vector $\mathbf{a}$ is perpendicular to a vector $\mathbf{b}$ if

$$\mathbf{a}\cdot\mathbf{b}=0. \quad (4.5.3.10)$$

Remark. In general, the condition $\mathbf{a}\cdot\mathbf{b}=0$ implies that the vectors $\mathbf{a}$ and $\mathbf{b}$ are perpendicular or one of them is the zero vector. The zero vector can be viewed to be perpendicular to any other vector.

4.5.3-4. Triple cross product.

The *triple cross product* of vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ is defined as the vector

$$\mathbf{d}=\mathbf{a}\times(\mathbf{b}\times\mathbf{c}). \quad (4.5.3.11)$$

The triple cross product is coplanar to the vectors $\mathbf{b}$ and $\mathbf{c}$; it can be expressed via $\mathbf{b}$ and $\mathbf{c}$ as follows:

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c})=\mathbf{b}\cdot(\mathbf{a}\cdot\mathbf{c})-\mathbf{c}\cdot(\mathbf{a}\cdot\mathbf{b}). \quad (4.5.3.12)$$

4.5.3-5. Scalar triple product of three vectors.

The *scalar triple product* of vectors **a**, **b**, and **c** is defined as the scalar product of **a** by the cross product of **b** and **c**:

$$[\mathbf{abc}] = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}). \tag{4.5.3.13}$$

Remark. The scalar triple product of three vectors **a**, **b**, and **c** is also denoted by **abc**.

Properties of scalar triple product:

1. $[\mathbf{abc}] = [\mathbf{bca}] = [\mathbf{cab}] = -[\mathbf{bac}] = -[\mathbf{cba}] = -[\mathbf{acb}]$.
2. $[(\mathbf{a} + \mathbf{b})\mathbf{cd}] = [\mathbf{acd}] + [\mathbf{bcd}]$ (distributivity with respect to addition of vectors). This property holds for any number of summands.
3. $[\lambda\mathbf{abc}] = \lambda[\mathbf{abc}]$ (associativity with respect to a scalar factor).
4. If the vectors are given by their coordinates $\mathbf{a} = (a_x, a_y, a_z)$, $\mathbf{b} = (b_x, b_y, b_z)$, and $\mathbf{c} = (c_x, c_y, c_z)$, then
$$[\mathbf{abc}] = \begin{vmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix}. \tag{4.5.3.14}$$
5. The scalar triple product $[\mathbf{abc}]$ is equal to the volume V of the parallelepiped spanned by the vectors **a**, **b**, and **c** taken with the sign + if the vectors **a**, **b**, and **c** form a right-handed trihedral and the sign – if the vectors form a left-handed trihedral,
$$[\mathbf{abc}] = \pm V. \tag{4.5.3.15}$$
6. Three nonzero vectors **a**, **b**, and **c** are coplanar if and only if $[\mathbf{abc}] = 0$. In this case, the vectors **a**, **b**, and **c** are linearly dependent; they satisfy a relation of the form $\mathbf{c} = \alpha\mathbf{a} + \beta\mathbf{b}$.

4.5.4. Curves and Surfaces in Space

4.5.4-1. Methods for defining curves.

A *continuous curve* in three-dimensional space is the set of points whose coordinates satisfy a system of parametric equations (4.5.2.19). This method for defining a curve is referred to as *parametric*. The curve can also be defined by an equivalent system of equations (4.5.2.18), i.e., described as the intersection of two surfaces (see Paragraph 4.5.4-2).

Remark 1. A curve may have more than one branch.

Remark 2. One can obtain the equation of the projection of the curve (4.5.2.19) onto the plane OXY by eliminating the variable x from equations (4.5.2.19)

4.5.4-2. Methods for defining surfaces.

A *continuous surface* in three-dimensional space is the set of points whose coordinates satisfy a system of parametric equations (4.5.2.17). This method for defining a surface is referred to as *parametric*. The surface can also be determined by an equation (4.5.2.16) or $z = f(x, y)$.

Remark 1. A surface may have more than one sheet.

Remark 2. The surfaces determined by the equations

$$F(x, y, z) = 0 \qquad \text{and} \qquad \lambda F(x, y, z) = 0$$

coincide provided that the constant λ is nonzero.

Remark 3. For any real number λ, the equation

$$F_1(x, y, z) + \lambda F_2(x, y, z) = 0$$

describes a surface passing through the line of intersection of the surfaces (4.5.2.18), provided that this line exists.

Remark 4. The equation

$$F_1(x, y, z) \cdot F_2(x, y, z) = 0$$

describes the surface that is formed by points of both surfaces in (4.5.2.18) and does not contain any other points.

4.6. Line and Plane in Space

4.6.1. Plane in Space

4.6.1-1. Equation of plane passing through point M_0 and perpendicular to vector **N**.

A *plane* is a first-order algebraic surface. In a Cartesian coordinate system, a plane is given by a first-order equation.

The equation of the plane passing through a point $M_0(x_0, y_0, z_0)$ and perpendicularly to a vector $\mathbf{N} = (A, B, C)$ has the form

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0, \qquad \text{or} \qquad (\mathbf{r} - \mathbf{r}_0) \cdot \mathbf{N} = 0, \tag{4.6.1.1}$$

where $\mathbf{r}$ and $\mathbf{r}_0$ are the position vectors of the point $M(x, y, z)$ and $M_0(x_0, y_0, z_0)$, respectively (see Fig. 4.34). The vector **N** is called a *normal vector*. Its direction cosines are

$$\cos\alpha = \frac{A}{\sqrt{A^2 + B^2 + C^2}}, \ \cos\beta = \frac{B}{\sqrt{A^2 + B^2 + C^2}}, \ \cos\gamma = \frac{C}{\sqrt{A^2 + B^2 + C^2}}. \tag{4.6.1.2}$$

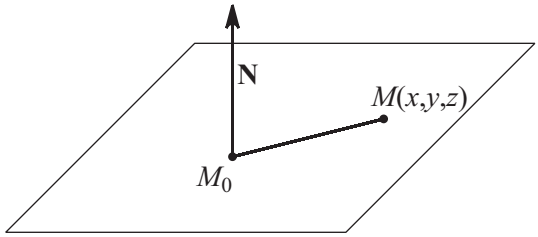


Figure 4.34. Plane passing through a point M_0 and perpendicularly to a vector **N**.

Example 1. Let us write out the equation of the plane that passes through the point $M_0(1, 2, 1)$ and is perpendicular to the vector $\mathbf{N} = (3, 2, 3)$.

According to (4.6.1.1), the desired equation is $3(x - 1) + 2(y - 2) + 3(z - 1) = 0$ or $3x + 2y + 3z - 10 = 0$.

4.6.1-2. General equation of plane.

The *general (complete) equation of a plane* has the form

$$Ax + By + Cz + D = 0, \qquad \text{or} \qquad \mathbf{r} \cdot \mathbf{N} + D = 0. \tag{4.6.1.3}$$

It follows from (4.6.1.1) that $D = -Ax_0 - By_0 - Cz_0$. If one of the coefficients in the equation of a plane is zero, then the equation is said to be *incomplete*:

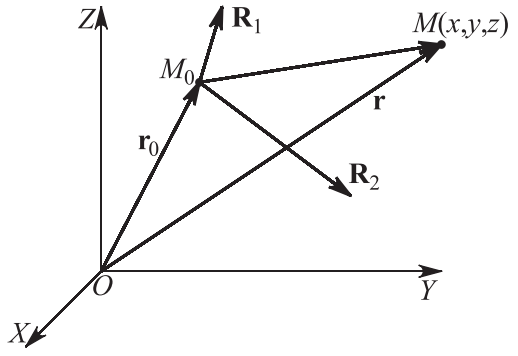


Figure 4.35. Basis in plane.

1. For $D = 0$, the equation has the form $Ax + By + Cz = 0$ and defines a plane passing through the origin.
2. For $A = 0$ (respectively, $B = 0$ or $C = 0$), the equation has the form $By + Cz + D = 0$ and defines a plane parallel to the axis OX (respectively, OY or OZ).
3. For $A=D=0$ (respectively, $B=D=0$ or $B=D=0$), the equation has the form $By+Cz=0$ and defines a plane passing through the axis OX (respectively, OY or OZ).
4. For $A=B=0$ (respectively, $A=C=0$ or $B=C=0$), the equation has the form $Cz+D=0$ and defines a plane parallel to the plane OXY (respectively, OXZ or OYZ).

4.6.1-3. Parametric equation of plane.

Each vector $\overrightarrow{M_0M} = \mathbf{r} - \mathbf{r}_0$ lying in a plane (where $\mathbf{r}$ and $\mathbf{r}_0$ are the position vectors of the points M and M_0, respectively) can be represented as (see Fig. 4.35)

$$\overrightarrow{M_0M} = t\mathbf{R}_1 + s\mathbf{R}_2, \tag{4.6.1.4}$$

where $\mathbf{R}_1 = (l_1, m_1, n_1)$ and $\mathbf{R}_2 = (l_2, m_2, n_2)$ are two arbitrary noncollinear vectors lying in the plane. Obviously, these two vectors form a basis in this plane. The *parametric equation of a plane* passing through the point $M_0(x_0, y_0, z_0)$ has the form

$$\mathbf{r} = \mathbf{r}_0 + t\mathbf{R}_1 + s\mathbf{R}_2, \qquad \text{or} \qquad \begin{aligned} x &= x_0 + tl_1 + sl_2, \\ y &= y_0 + tm_1 + sm_2, \\ z &= z_0 + tn_1 + sn_2. \end{aligned} \tag{4.6.1.5}$$

4.6.1-4. Intercept equation of plane.

A plane $Ax + By + Cz + D = 0$ that is not parallel to the axis Ox (i.e., $A \neq 0$) meets this axis at a (signed) distance $a = -D/A$ from the origin (see Fig. 4.36). The number a is called the x-intercept of the plane. Similarly, one defines the y-intercepts $b = -D/B$ (for $B \neq 0$) and the z-intercept $c = -D/C$ (for $C \neq 0$). Then such a plane can be defined by the equation

$$\frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 1, \tag{4.6.1.6}$$

which is called the *intercept equation of the plane*.

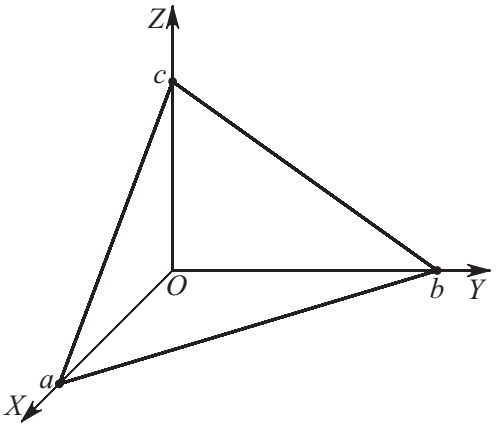


Figure 4.36. A plane with intercept equation.

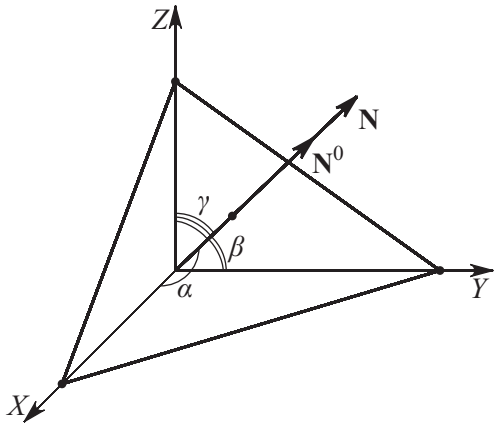


Figure 4.37. A plane with normalized equation.

Remark 1. Equation (4.6.1.6) can be obtained as the equation of the plane passing through three given points.

Remark 2. A plane parallel to the axis OX but nonparallel to the other two axes is defined by the equation $y/b + z/c = 1$, where b and c are the y- and z-intercepts of the plane. A plane simultaneously parallel to the axes OY and OZ can be represented in the form $z/c = 1$.

Example 2. Consider the plane given by the general equation $2x + 3y - z + 6 = 0$. Let us rewrite it in intercept form.

The x-, y-, and z-intercepts of this plane are

$$a = -\frac{D}{A} = -\frac{6}{2} = -3, \quad b = -\frac{D}{B} = -\frac{6}{3} = -2, \quad \text{and} \quad c = -\frac{D}{C} = -\frac{6}{-1} = 6.$$

Thus the intercept equation of the plane reads

$$\frac{x}{-3} + \frac{y}{-2} + \frac{z}{6} = 0.$$

4.6.1-5. Normalized equation of plane.

The *normalized equation of a plane* has the form

$$\mathbf{r} \cdot \mathbf{N}^0 - p = 0, \qquad \text{or} \qquad x\cos\alpha + y\cos\beta + z\cos\gamma - p = 0, \tag{4.6.1.7}$$

where $\mathbf{N}^0 = (\cos\alpha, \cos\beta, \cos\gamma)$ is a unit vector and p is the distance from the plane to the origin; here $\cos\alpha$, $\cos\beta$, and $\cos\gamma$ are the direction cosines of the normal to the plane (see Fig. 4.37). The numbers $\cos\alpha$, $\cos\beta$, $\cos\gamma$, and p can be expressed via the coefficients A, B, C as follows:

$$\begin{aligned} \cos\alpha &= \pm\frac{A}{\sqrt{A^2+B^2+C^2}}, \quad \cos\beta = \pm\frac{B}{\sqrt{A^2+B^2+C^2}}, \\ \cos\gamma &= \pm\frac{C}{\sqrt{A^2+B^2+C^2}}, \quad p = \mp\frac{D}{\sqrt{A^2+B^2+C^2}}, \end{aligned} \tag{4.6.1.8}$$

where the upper sign is taken if $D < 0$ and the lower sign is taken if $D > 0$. For $D = 0$, either sign can be taken.

The normalized equation (4.6.1.7) can be obtained from a general equation (4.6.1.3) by multiplication by the normalizing factor

$$\mu = \pm\frac{1}{\sqrt{A^2+B^2+C^2}}, \tag{4.6.1.9}$$

where the sign of μ must be opposite to that of D.

Example 3. Let us reduce the equation of the plane $-2x + 2y - z - 6 = 0$ to normalized form.
Since $D = -6 < 0$, we see that the normalizing factor is

$$\mu = \frac{1}{\sqrt{(-2)^2 + 2^2 + (-1)^2}} = \frac{1}{3}.$$

We multiply the equation by this factor and obtain

$$-\frac{2}{3}x + \frac{2}{3}y - \frac{1}{3}z - 2 = 0.$$

Hence for this plane we have

$$\cos\alpha = -\frac{2}{3}, \quad \cos\beta = \frac{2}{3}, \quad \cos\gamma = -\frac{1}{3}, \quad p = 2.$$

Remark. The numbers $\cos\alpha$, $\cos\beta$, $\cos\gamma$, and p are also called the *polar parameters of a plane*.

4.6.1-6. Equation of plane passing through point and parallel to another plane.

The plane that passes through a point $M_1(x_1, y_1, z_1)$ and is parallel to a plane $Ax + By + Cz + D = 0$ is given by the equation

$$A(x - x_1) + B(y - y_1) + C(z - z_1) + D = 0. \tag{4.6.1.10}$$

Example 4. Let us derive the equation of the plane that passes through the point $M_1(1, 2, -1)$ and is parallel to the plane $x + 2y + z + 2 = 0$.

According to (4.6.1.1), the desired equation is $(x - 1) + 2(y - 2) + (z + 1) + 2 = 0$ or

$$x + 2y + z - 2 = 0.$$

4.6.1-7. Equation of plane passing through three points.

The plane passing through three points $M_1(x_1, y_1, z_1)$, $M_2(x_2, y_2, z_2)$, and $M_3(x_3, y_3, z_3)$ (see Fig. 4.38) is described by the equation

$$\begin{vmatrix} x - x_1 & y - y_1 & z - z_1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 \end{vmatrix} = 0, \qquad \text{or} \qquad \big[(\mathbf{r} - \mathbf{r}_1)(\mathbf{r}_2 - \mathbf{r}_1)(\mathbf{r}_3 - \mathbf{r}_1)\big] = 0, \tag{4.6.1.11}$$

where $\mathbf{r}$, $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{r}_3$ are the position vectors of the points $M(x, y, z)$, $M_1(x_1, y_1, z_1)$, $M_2(x_2, y_2, z_2)$, and $M_3(x_3, y_3, z_3)$, respectively.

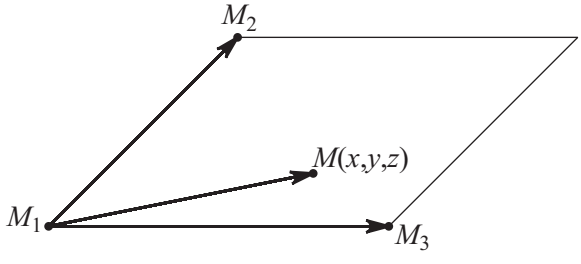


Figure 4.38. Plane passing through three points.

Remark 1. Equation (4.6.1.11) means that the vectors $\overrightarrow{M_1M}$, $\overrightarrow{M_1M_2}$, and $\overrightarrow{M_1M_3}$ are coplanar.

Remark 2. Equation (4.6.1.11) of the plane passing through three given points can be represented via a fourth-order determinant as follows:

$$\begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} = 0. \qquad (4.6.1.11a)$$

Remark 3. If the three points $M_1(x_1, y_1, z_1)$, $M_2(x_2, y_2, z_2)$, and $M_3(x_3, y_3, z_3)$ are collinear, then equations (4.6.1.11) and (4.6.1.11a) become identities.

Example 5. Let us construct an equation of the plane passing through the three points $M_1(1, 1, 1)$, $M_2(2, 2, 1)$, and $M_3(1, 2, 2)$.

Obviously, the points M_1, M_2, and M_3 are not collinear, since the vectors $\overrightarrow{M_1M_2} = (1, 1, 0)$ and $\overrightarrow{M_1M_3} = (0, 1, 1)$ are not collinear. According to (4.6.1.11), the desired equation is

$$\begin{vmatrix} x-1 & y-1 & z-1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{vmatrix} = 0,$$

whence

$$x - y + z - 1 = 0.$$

4.6.1-8. Equation of plane passing through two points and parallel to line.

The equation of the plane passing through two points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$ and parallel to a straight line with direction vector $\mathbf{R} = (l, m, n)$ (see Fig. 4.39) is

$$\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ x_2-x_1 & y_2-y_1 & z_2-z_1 \\ l & m & n \end{vmatrix} = 0, \qquad \text{or} \qquad \big[(\mathbf{r}-\mathbf{r}_1)(\mathbf{r}_2-\mathbf{r}_1)\mathbf{R}\big] = 0, \qquad (4.6.1.12)$$

where $\mathbf{r}$, $\mathbf{r}_1$, and $\mathbf{r}_2$ are the position vectors of the points $M(x, y, z)$, $M_1(x_1, y_1, z_1)$, and $M_2(x_2, y_2, z_2)$, respectively.

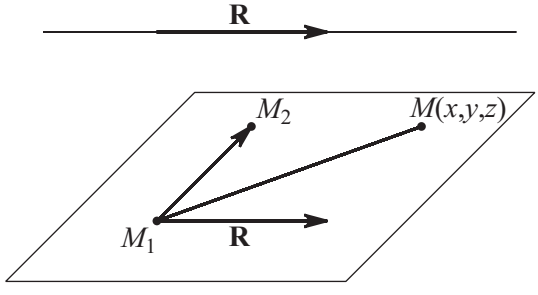


Figure 4.39. Plane passing through two points and parallel to line.

Remark. If the vectors $\overrightarrow{M_1M_2}$ and $\mathbf{R}$ are collinear, then equations (4.6.1.12) become identities.

Example 6. Let us construct an equation of the plane passing through the points $M_1(0, 1, 0)$ and $M_2(1, 1, 1)$ and parallel to the straight line with direction vector $\mathbf{R} = (0, 1, 1)$.

According to (4.6.1.12), the desired equation is

$$\begin{vmatrix} x-0 & y-1 & z-0 \\ 1-0 & 1-1 & 1-0 \\ 0 & 1 & 1 \end{vmatrix} = 0,$$

whence

$$-x - y + z + 1 = 0.$$

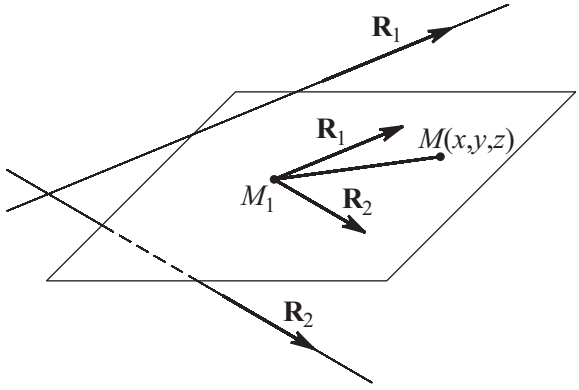


Figure 4.40. Plane passing through a point and parallel to two straight lines.

4.6.1-9. Equation of plane passing through point and parallel to two straight lines.

The equation of the plane passing through a point $M_1(x_1, y_1, z_1)$ and parallel to two straight lines with direction vectors $\mathbf{R}_1 = (l_1, m_1, n_1)$ and $\mathbf{R}_2 = (l_2, m_2, n_2)$ (see Fig. 4.40) is

$$\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{vmatrix} = 0, \qquad \text{or} \qquad \big[(\mathbf{r}-\mathbf{r}_1)\mathbf{R}_1\mathbf{R}_2\big] = 0, \tag{4.6.1.13}$$

where $\mathbf{r}$ and $\mathbf{r}_1$ are the position vectors of the points $M(x, y, z)$ and $M_1(x_1, y_1, z_1)$, respectively.

Example 7. Let us find the equation of the plane passing through the point $M_1(0, 1, 0)$ and parallel to the straight lines with direction vectors $\mathbf{R}_1 = (1, 0, 1)$ and $\mathbf{R}_2 = (0, 1, 2)$.

According to (4.6.1.13), the desired equation is

$$\begin{vmatrix} x-0 & y-1 & z-0 \\ 1 & 0 & 1 \\ 0 & 1 & 2 \end{vmatrix} = 0,$$

whence

$$-x - 2y + z + 2 = 0.$$

4.6.1-10. Plane passing through two points and perpendicular to given plane.

The plane (see Fig. 4.41) passing through two points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$ and perpendicular to the plane given by the equation $Ax + By + Cz + D = 0$ is determined by the equation

$$\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ x_2-x_1 & y_2-y_1 & z_2-z_1 \\ A & B & C \end{vmatrix} = 0, \qquad \text{or} \qquad \big[(\mathbf{r}-\mathbf{r}_1)(\mathbf{r}_2-\mathbf{r}_1)\mathbf{N}\big] = 0, \tag{4.6.1.14}$$

where $\mathbf{r}$, $\mathbf{r}_1$, and $\mathbf{r}_2$ are the position vectors of the points $M(x, y, z)$, $M_1(x_1, y_1, z_1)$, and $M_2(x_2, y_2, z_2)$, respectively.

Remark. If the straight line passing through points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$ is perpendicular to the original plane, then the desired plane is undetermined and equations (4.6.1.14) become identities.

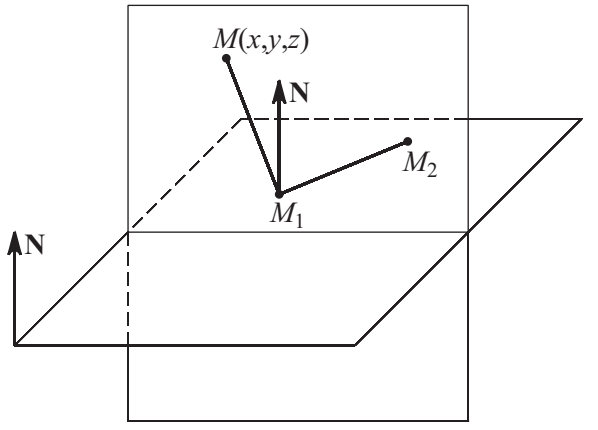


Figure 4.41. Plane passing through two points and perpendicular to given plane.

Example 8. Let us find an equation of the plane passing through the points $M_1(0, 1, 2)$ and $M_2(2, 2, 3)$ and perpendicular to the plane $x - y + z + 5 = 0$.

According to (4.6.1.14), the desired equation is

$$\begin{vmatrix} x-0 & y-1 & z-2 \\ 2-0 & 2-1 & 3-2 \\ 1 & -1 & 1 \end{vmatrix} = 0,$$

whence

$$2x - y - 3z + 7 = 0.$$

4.6.1-11. Plane passing through point and perpendicular to two planes.

The plane (see Fig. 4.42) passing through a point $M_1(x_1, y_1, z_1)$ and perpendicular to two (nonparallel) planes $A_1x + B_1y + C_1z + D_1 = 0$ and $A_2x + B_2y + C_2z + D_2 = 0$ is given by the equation

$$\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{vmatrix} = 0, \qquad \text{or} \qquad \big[(\mathbf{r}-\mathbf{r}_1)\mathbf{N}_1\mathbf{N}_2\big] = 0, \tag{4.6.1.15}$$

where $\mathbf{N}_1 = (A_1, B_1, C_1)$ and $\mathbf{N}_2 = (A_2, B_2, C_2)$ are the normals to the given planes and $\mathbf{r}$ and $\mathbf{r}_1$ are the position vectors of the points $M(x, y, z)$ and $M_1(x_1, y_1, z_1)$, respectively.

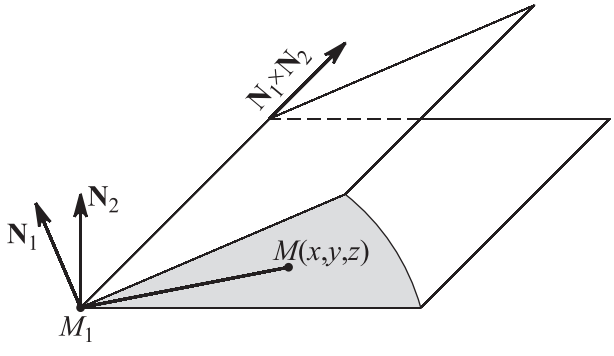


Figure 4.42. Plane passing through a point and perpendicular to two planes.

Remark 1. Equations (4.6.1.15) mean that the vectors $\overrightarrow{M_1M}$, $\mathbf{N}_1$, and $\mathbf{N}_2$ are coplanar.

Remark 2. If the original planes are parallel, then the desired plane is undetermined. In this case, equations (4.6.1.15) become identities.

Example 9. Let us find an equation of the plane passing through the point $M_1(0, 1, 2)$ and perpendicular to the planes $x - y + z + 5 = 0$ and $-x + y + z - 1 = 0$.

According to (4.6.1.15), the desired equation is

$$\begin{vmatrix} x-0 & y-1 & z-2 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \end{vmatrix} = 0,$$

whence

$$x + y - 1 = 0.$$

4.6.1-12. Equation of plane passing through line of intersection of planes.

The planes passing through the line of intersection of the planes $A_1x + B_1y + C_1z + D_1 = 0$ and $A_2x + B_2y + C_2z + D_2 = 0$ are given by the equation

$$\alpha(A_1x + B_1y + C_1z + D_1) + \beta(A_2x + B_2y + C_2z + D_2) = 0, \qquad (4.6.1.16)$$

which is called the *equation of a pencil of planes*. Here α and β are arbitrary parameters.

Let $\alpha \neq 0$. Set $\beta/\alpha = \lambda$; then equation (4.6.1.16) becomes

$$A_1x + B_1y + C_1z + D_1 + \lambda(A_2x + B_2y + C_2z + D_2) = 0. \qquad (4.6.1.17)$$

By varying the parameter λ from $-\infty$ to $+\infty$, we obtain all the planes in the pencil. For $\lambda = \pm 1$, we obtain equations of the planes that bisect the angles between the given planes provided that the equations of the latter are given in normalized form.

Remark. The passage from equation (4.6.1.16) to equation (4.6.1.17) excludes the case $\alpha = 0$. Equation (4.6.1.17) does not define the plane $A_2x + B_2y + C_2z + D_2 = 0$; i.e., equation (4.6.1.17) for various λ defines all the planes in the pencil but one (the second of the two given planes).

4.6.2. Line in Space

4.6.2-1. Parametric equation of straight line.

The *parametric equation of the line* that passes through a point $M_1(x_1, y_1, z_1)$ and is parallel to a direction vector $\mathbf{R} = (l, m, n)$ (see Fig. 4.43) is

$$x = x_1 + lt, \quad y = y_1 + mt, \quad z = z_1 + nt, \qquad \text{or} \qquad \mathbf{r} = \mathbf{r}_1 + t\mathbf{R}, \qquad (4.6.2.1)$$

where $\mathbf{r} = \overrightarrow{OM}$ and $\mathbf{r}_1 = \overrightarrow{M_1M}$. As the parameter t varies from $-\infty$ to $+\infty$, the point M with position vector $\mathbf{r} = (x, y, z)$ determined by formula (4.6.2.1) runs over the entire straight line in question. It is convenient to use the parametric equation (4.6.2.1) if one needs to find the point of intersection of a straight line with a plane.

The numbers l, m, and n characterize the direction of the straight line in space; they are called the *direction coefficients* of the straight line. For a unit vector $\mathbf{R} = \mathbf{R}^0$, the coefficients l, m, n are the cosines of the angles α, β, and γ formed by this straight line (the direction of the vector $\mathbf{R}^0$) with the coordinate axes OX, OY, and OZ. These cosines can be expressed via the coordinates of the direction vector $\mathbf{R}$ as

$$\cos\alpha = \frac{l}{\sqrt{l^2 + m^2 + n^2}}, \quad \cos\beta = \frac{m}{\sqrt{l^2 + m^2 + n^2}}, \quad \cos\gamma = \frac{n}{\sqrt{l^2 + m^2 + n^2}}. \qquad (4.6.2.2)$$

Example 1. Let us find the equation of the straight line that passes through the point $M_1(2, -3, 1)$ and is parallel to the direction vector $\mathbf{R} = (1, 2, -3)$.

According to (4.6.2.1), the desired equation is

$$x = 2 + t, \quad y = -3 + 2t, \quad z = 1 - 3t.$$

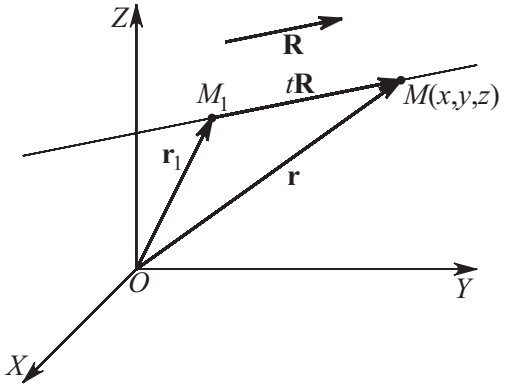


Figure 4.43. Straight line passing through a point and parallel to direction vector.

4.6.2-2. Canonical equation of straight line.

The equation

$$\frac{x-x_1}{l}=\frac{y-y_1}{m}=\frac{z-z_1}{n}, \qquad \text{or} \qquad (\mathbf{r}-\mathbf{r}_1)\times\mathbf{R}=0, \tag{4.6.2.3}$$

is called the *canonical equation of the straight line* passing through the point $M_1(x_1, y_1, z_1)$ with position vector $\mathbf{r}_1 = (x_1, y_1, z_1)$ and parallel to the direction vector $\mathbf{R} = (l, m, n)$.

Remark 1. One can obtain the canonical equation (4.6.2.3) from the parametric equations (4.6.2.1) by eliminating the parameter t.

Remark 2. In the canonical equation, all coefficients l, m, and n cannot be zero simultaneously, since $|\mathbf{R}| \neq 0$. But some of them may be zero. If one of the denominators in equations (4.6.2.3) is zero, this means that the corresponding numerator is also zero.

Example 2. The equations $(x-1)/1 = (y-3)/4 = (z-3)/0$ determine the straight line passing through the point $M_1(1, 3, 3)$ and perpendicular to the axis OZ. This means that the line lies in the plane $z = 3$, and hence $z - 3 = 0$ for all points of the line.

Example 3. Let us find the equation of the straight line passing through the point $M_1(2, -3, 1)$ and parallel to the direction vector $\mathbf{R} = (1, 2, -3)$.

According to (4.6.2.3), the desired equation is

$$\frac{x-2}{1}=\frac{y+3}{2}=\frac{z-1}{-3}.$$

4.6.2-3. General equation of straight line.

The *general equation of a straight line in space* defines it as the line of intersection of two planes (see Fig. 4.44) and is given analytically by a system of two linear equations

$$\begin{aligned} A_1x+B_1y+C_1z+D_1&=0,\\ A_2x+B_2y+C_2z+D_2&=0,\end{aligned} \qquad \text{or} \qquad \begin{aligned} \mathbf{r}\cdot\mathbf{N}_1+D_1&=0,\\ \mathbf{r}\cdot\mathbf{N}_2+D_2&=0,\end{aligned} \tag{4.6.2.4}$$

where $\mathbf{N}_1 = (A_1, B_1, C_1)$ and $\mathbf{N}_2 = (A_2, B_2, C_2)$ are the normals to the planes and $\mathbf{r}$ is the position vector of the point (x, y, z).

The direction vector $\mathbf{R}$ is equal to the cross product of the normals $\mathbf{N}_1$ and $\mathbf{N}_2$; i.e.,

$$\mathbf{R}=\mathbf{N}_1\times\mathbf{N}_2, \tag{4.6.2.5}$$

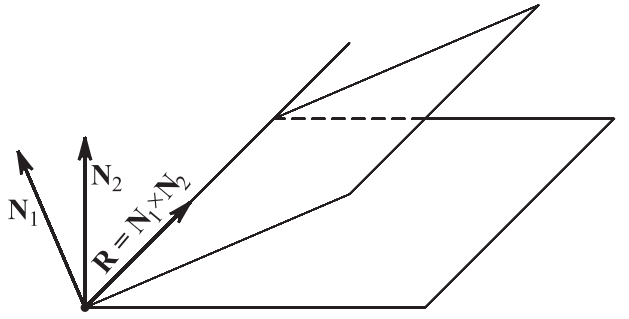


Figure 4.44. Straight line as intersection of two planes.

and its coordinates l, m, and n can be obtained by the formulas

$$l = \begin{vmatrix} B_1 & C_1 \\ B_2 & C_2 \end{vmatrix}, \quad m = \begin{vmatrix} C_1 & A_1 \\ C_2 & A_2 \end{vmatrix}, \quad n = \begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix}. \tag{4.6.2.6}$$

Remark 1. Simultaneous equations of the form (4.6.2.4) define a straight line if and only if the coefficients A_1, B_1, and C_1 in one of them are not proportional to the respective coefficients A_2, B_2, and C_2 in the other.

Remark 2. For $D_1 = D_2 = 0$ (and only in this case), the line passes through the origin.

Example 4. Let us reduce the equation of the straight line

$$x + 2y - z + 1 = 0, \quad x - y + z + 3 = 0$$

to canonical form.

We choose one of the coordinates arbitrarily; say, $x = 0$. Then

$$2y - z + 1 = 0, \quad -y + z + 3 = 0,$$

and hence $y = -4$, $z = -7$. Thus the desired line contains the point $M(0, -4, -7)$. We find the cross product of the vectors $\mathbf{N}_1 = (1, 2, -1)$ and $\mathbf{N}_2 = (1, -1, 1)$ and, according to (4.6.2.5), obtain the direction vector $\mathbf{R} = (1, -2, -3)$ of the desired line. Therefore, with (4.6.2.3) taken into account, the equation of the line becomes

$$\frac{x}{1} = \frac{y+4}{-2} = \frac{z+7}{-3}.$$

4.6.2-4. Equation of line in projections.

The *equation of a line in projections* can be obtained by eliminating first z and then y from the general equations (4.6.2.4):

$$y = kx + a, \quad z = hx + b. \tag{4.6.2.7}$$

Each of two equations (4.6.2.7) defines a plane projecting the straight line onto the planes OXY and OXZ (see Fig. 4.45).

Remark 1. For straight lines parallel to the plane OYZ, this form of the equations cannot be used; one should take the projections onto some other pair of coordinate planes.

Remark 2. Equations (4.6.2.7) can be represented in the canonical form

$$\frac{x-0}{1} = \frac{y-a}{k} = \frac{z-b}{h}. \tag{4.6.2.8}$$

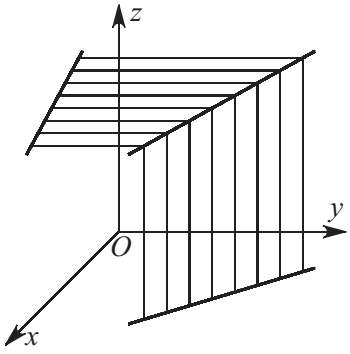


Figure 4.45. Straight line with equation in projections.

4.6.2-5. Equation of straight line passing through two points.

The canonical equation of the straight line (see Fig. 4.46) passing through two points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$ is

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} = \frac{z - z_1}{z_2 - z_1}, \qquad \text{or} \qquad (\mathbf{r} - \mathbf{r}_1) \times (\mathbf{r}_2 - \mathbf{r}_1) = 0, \tag{4.6.2.9}$$

where $\mathbf{r}$, $\mathbf{r}_1$, and $\mathbf{r}_2$ are the position vectors of the points $M(x, y, z)$, $M_1(x_1, y_1, z_1)$, and $M_2(x_2, y_2, z_2)$, respectively.

The parametric equations of the straight line passing through two points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$ in the rectangular Cartesian coordinate system $OXYZ$ can be written as

$$\begin{aligned} x &= x_1(1 - t) + x_2 t, \\ y &= y_1(1 - t) + y_2 t, \\ z &= z_1(1 - t) + z_2 t, \end{aligned} \qquad \text{or} \qquad \mathbf{r} = (1 - t)\mathbf{r}_1 + t\mathbf{r}_2. \tag{4.6.2.10}$$

Remark. Eliminating the parameter t from equations (4.6.2.10), we obtain equations (4.6.2.9).

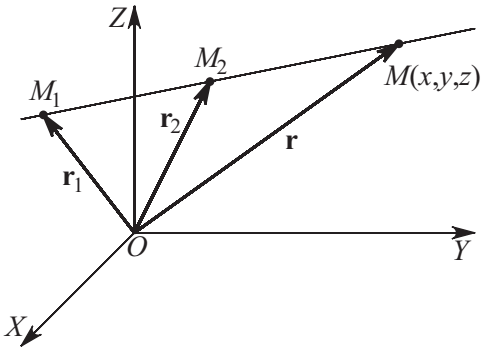


Figure 4.46. Straight line passing through two points.

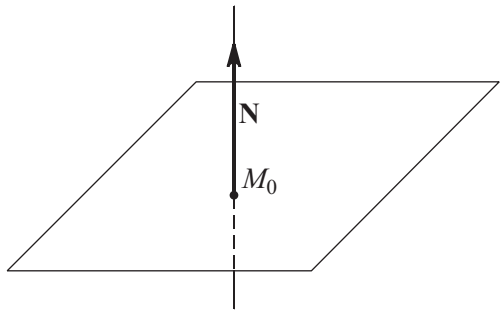


Figure 4.47. Straight line passing through point and perpendicular to plane.

4.6.2-6. Equation of straight line passing through point and perpendicular to plane.

The equation of the straight line passing through a point $M_0(x_0, y_0, z_0)$ and perpendicular to the plane given by the equation $Ax + By + Cz + D = 0$, or $\mathbf{r} \cdot \mathbf{N} + D = 0$ (see Fig. 4.47), is

$$\frac{x - x_0}{A} = \frac{y - y_0}{B} = \frac{z - z_0}{C}, \tag{4.6.2.11}$$

where $\mathbf{N} = (A, B, C)$ is the normal to the plane.

4.6.3. Mutual Arrangement of Points, Lines, and Planes

4.6.3-1. Angles between lines in space.

Consider two straight lines determined by vector parametric equations $\mathbf{r} = \mathbf{r}_1 + t\mathbf{R}_1$ and $\mathbf{r} = \mathbf{r}_2 + t\mathbf{R}_2$. The angle φ between these lines (see Fig. 4.48) can be obtained from the formulas

$$\cos\varphi = \frac{\mathbf{R}_1 \cdot \mathbf{R}_2}{|\mathbf{R}_1|\,|\mathbf{R}_2|}, \quad \sin\varphi = \frac{|\mathbf{R}_1 \times \mathbf{R}_2|}{|\mathbf{R}_1|\,|\mathbf{R}_2|}.$$

If the lines are given by the canonical equations

$$\frac{x-x_1}{l_1} = \frac{y-y_1}{m_1} = \frac{z-z_1}{n_1} \quad \text{and} \quad \frac{x-x_2}{l_2} = \frac{y-y_2}{m_2} = \frac{z-z_2}{n_2},$$

then the angle φ between the lines can be found from the formulas

$$\begin{aligned} \cos\varphi &= \frac{l_1 l_2 + m_1 m_2 + n_1 n_2}{\sqrt{l_1^2 + m_1^2 + n_1^2}\sqrt{l_2^2 + m_2^2 + n_2^2}}, \\ \sin\varphi &= \frac{\sqrt{\begin{vmatrix} m_1 & n_1 \\ m_2 & n_2 \end{vmatrix}^2 + \begin{vmatrix} n_1 & l_1 \\ n_2 & l_2 \end{vmatrix}^2 + \begin{vmatrix} l_1 & m_1 \\ l_2 & m_2 \end{vmatrix}^2}}{\sqrt{l_1^2 + m_1^2 + n_1^2}\sqrt{l_2^2 + m_2^2 + n_2^2}}, \end{aligned} \tag{4.6.3.1}$$

which coincide with formulas (4.6.3.1) written in coordinate form.

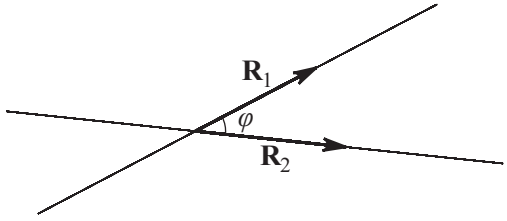


Figure 4.48. Angles between lines in space.

Example 1. Let us find the angle between the lines

$$\frac{x}{1} = \frac{y-2}{2} = \frac{z+1}{2} \qquad \text{and} \qquad \frac{x}{0} = \frac{y-2}{3} = \frac{z+1}{4}.$$

Using the first formula in (4.6.3.1), we obtain

$$\cos\varphi = \frac{1\cdot 0 + 3\cdot 2 + 4\cdot 2}{\sqrt{0^2+3^2+4^2}\sqrt{1^2+2^2+2^2}} = \frac{14}{15},$$

and hence $\varphi \approx 0.3672$ rad.

4.6.3-2. Conditions for two lines to be parallel.

Two straight lines given by vector parametric equations $\mathbf{r} = \mathbf{r}_1 + t\mathbf{R}_1$ and $\mathbf{r} = \mathbf{r}_2 + t\mathbf{R}_2$ are parallel if

$$\mathbf{R}_2 = \lambda\mathbf{R}_1 \qquad \text{or} \qquad \mathbf{R}_2 \times \mathbf{R}_1 = 0,$$

i.e., if their direction vectors $\mathbf{R}_1$ and $\mathbf{R}_2$ are collinear. If the straight lines are given by canonical equations, then the condition that they are parallel can be written as

$$\frac{l_1}{l_2} = \frac{m_1}{m_2} = \frac{n_1}{n_2}. \tag{4.6.3.2}$$

Remark. If parallel lines have a common point (i.e., $\mathbf{r}_1 = \mathbf{r}_2$ in parametric equations), then they coincide.

Example 2. Let us show that the lines

$$\frac{x-1}{2} = \frac{y-3}{1} = \frac{z}{2} \qquad \text{and} \qquad \frac{x-3}{4} = \frac{y+1}{2} = \frac{z}{4}$$

are parallel to each other.

Indeed, condition (4.6.3.2) is satisfied,

$$\frac{2}{4} = \frac{1}{2} = \frac{2}{4},$$

and hence the lines are parallel.

4.6.3-3. Conditions for two lines to be perpendicular.

Two straight lines given by vector parametric equation $\mathbf{r} = \mathbf{r}_1 + t\mathbf{R}_1$ and $\mathbf{r} = \mathbf{r}_2 + t\mathbf{R}_2$ are perpendicular if

$$\mathbf{R}_1 \cdot \mathbf{R}_2 = 0. \tag{4.6.3.3}$$

If the lines are given by canonical equations, then the condition that they are perpendicular can be written as

$$l_1 l_2 + m_1 m_2 + n_1 n_2 = 0, \tag{4.6.3.3a}$$

which coincides with formula (4.6.3.3) written in coordinate form.

Example 3. Let us show that the lines

$$\frac{x-1}{2} = \frac{y-3}{1} = \frac{z}{2} \qquad \text{and} \qquad \frac{x-2}{1} = \frac{y+1}{2} = \frac{z}{-2}$$

are perpendicular.

Indeed, condition (4.6.3.3a) is satisfied,

$$2 \cdot 1 + 1 \cdot 2 + 2 \cdot (-2) = 0,$$

and hence the lines are perpendicular.

4.6.3-4. Theorem on the arrangement of two lines in space.

THEOREM ON THE ARRANGEMENT OF TWO LINES IN SPACE. *Two lines in space can:*

a) *be skew;*
b) *lie in the same plane and not meet each other, i.e., be parallel;*
c) *meet at a point;*
d) *coincide.*

A general characteristic of all four cases is the determinant of the matrix

$$\begin{pmatrix} x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{pmatrix}, \tag{4.6.3.4}$$

whose entries are taken from the canonical equations

$$\frac{x-x_1}{l_1}=\frac{y-y_1}{m_1}=\frac{z-z_1}{n_1} \quad \text{and} \quad \frac{x-x_2}{l_2}=\frac{y-y_2}{m_2}=\frac{z-z_2}{n_2}$$

of the lines.

In cases *a*–*d* of the theorem, for the matrix (4.6.3.4) we have, respectively:

a) the determinant is nonzero;
b) the last two rows are proportional to each other but are not proportional to the first row;
c) the last two rows are not proportional, and the first row is their linear combination;
d) all rows are proportional.

4.6.3-5. Angles between planes.

Consider two planes given by the general equations

$$\begin{aligned} &A_1x+B_1y+C_1z+D_1=0,\\ &A_2x+B_2y+C_2z+D_2=0, \end{aligned} \quad \text{or} \quad \begin{aligned} &\mathbf{r}\cdot\mathbf{N}_1+D_1=0,\\ &\mathbf{r}\cdot\mathbf{N}_2+D_2=0, \end{aligned} \tag{4.6.3.5}$$

where $\mathbf{N}_1=(A_1,B_1,C_1)$ and $\mathbf{N}_2=(A_2,B_2,C_2)$ are the normals to the planes and $\mathbf{r}$ is the position vector of the point (x,y,z).

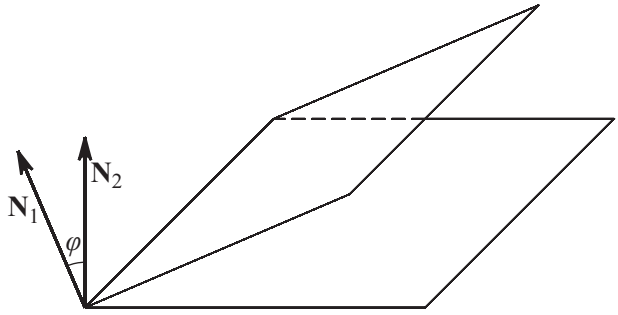


Figure 4.49. Angles between planes.

The angle between two planes (see Fig. 4.49) is defined as any of the two adjacent dihedral angles formed by the planes (if the planes are parallel, then the angle between them is by definition equal to 0 or π). One of these dihedral angles is equal to the angle φ between the normal vectors $\mathbf{N}_1=(A_1,B_1,C_1)$ and $\mathbf{N}_2=(A_2,B_2,C_2)$ to the planes, which can be determined by the formula

$$\cos\varphi=\frac{A_1A_2+B_1B_2+C_1C_2}{\sqrt{A_1^2+B_1^2+C_1^2}\sqrt{A_2^2+B_2^2+C_2^2}}=\frac{\mathbf{N}_1\cdot\mathbf{N}_2}{|\mathbf{N}_1|\,|\mathbf{N}_2|}. \tag{4.6.3.6}$$

If the planes are given by vector parametric equations

$$\mathbf{r}=\mathbf{r}_1+\mathbf{R}_1t+\mathbf{R}_2s \quad \text{or} \quad \mathbf{r}'=\mathbf{r}'_1+\mathbf{R}'_1t+\mathbf{R}'_2s, \tag{4.6.3.7}$$

then the angle between the planes is given by the formula

$$\cos\varphi=\frac{(\mathbf{R}_1\times\mathbf{R}_2)\cdot(\mathbf{R}'_1\times\mathbf{R}'_2)}{|\mathbf{R}_1\times\mathbf{R}_2|\,|\mathbf{R}'_1\times\mathbf{R}'_2|}. \tag{4.6.3.8}$$

4.6.3-6. Conditions for two planes to be parallel.

Two planes given by the general equations (4.6.3.5) in coordinate form are parallel if and only if the following *condition for the planes to be parallel* is satisfied:

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2} \neq \frac{D_1}{D_2}\,; \tag{4.6.3.9}$$

in this case, the planes do not coincide. For planes given by the general equations (4.6.3.5) in vector form, the condition becomes

$$\mathbf{N}_2 = \lambda \mathbf{N}_1 \quad \text{or} \quad \mathbf{N}_2 \times \mathbf{N}_1 = 0; \tag{4.6.3.10}$$

i.e., the planes are parallel if their normals are parallel.

Example 4. Let us show that the planes $x - y + z = 0$ and $2x - 2y + 2z + 5 = 0$ are parallel.
Since condition (4.6.3.9) is satisfied,

$$\frac{1}{2} = \frac{-1}{-2} = \frac{1}{2},$$

we see that the planes are parallel to each other.

4.6.3-7. Conditions for planes to coincide.

Two planes coincide if they are parallel and have a common point.

Two planes given by the general equations (4.6.3.5) coincide if and only if the following *condition for the planes to coincide* is satisfied:

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2} = \frac{D_1}{D_2}\,. \tag{4.6.3.11}$$

Remark. Sometimes the case in which the planes coincide is treated as a special case of parallel straight lines and is not distinguished as an exceptional case.

4.6.3-8. Conditions for two planes to be perpendicular.

Planes are perpendicular if their normals are perpendicular.

Two planes determined by the general equations (4.6.3.5) are perpendicular if and only if the following *condition for the planes to be perpendicular* is satisfied:

$$A_1A_2 + B_1B_2 + C_1C_2 = 0 \qquad \text{or} \qquad \mathbf{N}_1 \cdot \mathbf{N}_2 = 0, \tag{4.6.3.12}$$

where $\mathbf{N}_1 = (A_1, B_1, C_1)$ and $\mathbf{N}_2 = (A_2, B_2, C_2)$ are the normals to the planes.

Example 5. Let us show that the planes $x - y + z = 0$ and $x - y - 2z + 5 = 0$ are perpendicular.
Since condition (4.6.3.12) is satisfied,

$$1 \cdot 1 + (-1) \cdot (-1) + 1 \cdot (-2) = 0,$$

we see that the planes are perpendicular to each other.

4.6.3-9. Intersection of three planes.

The point of intersection of three planes given by the equations

$$A_1x + B_1y + C_1z + D_1 = 0, \quad A_2x + B_2y + C_2z + D_2 = 0, \quad A_3x + B_3y + C_3z + D_3 = 0,$$

has the following coordinates:

$$x_0 = -\frac{1}{\Delta}\begin{vmatrix} D_1 & B_1 & C_1 \\ D_2 & B_2 & C_2 \\ D_3 & B_3 & C_3 \end{vmatrix}, \; y_0 = -\frac{1}{\Delta}\begin{vmatrix} A_1 & D_1 & C_1 \\ A_2 & D_2 & C_2 \\ A_3 & D_3 & C_3 \end{vmatrix}, \; z_0 = -\frac{1}{\Delta}\begin{vmatrix} A_1 & B_1 & D_1 \\ A_2 & B_2 & D_2 \\ A_3 & B_3 & D_3 \end{vmatrix}, \tag{4.6.3.13}$$

where Δ is given by the formula

$$\Delta = \begin{vmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{vmatrix}. \tag{4.6.3.14}$$

Remark. Three planes are concurrent at a single point if $\Delta \neq 0$. If $\Delta = 0$ and at least one of the second-order minors is nonzero, then all planes are parallel to a single line. If all minors are zero, then the planes are concurrent in a single line.

4.6.3-10. Intersection of four planes.

If four planes given by the equations

$$A_1x + B_1y + C_1z + D_1 = 0, \qquad A_2x + B_2y + C_2z + D_2 = 0,$$
$$A_3x + B_3y + C_3z + D_3 = 0, \qquad A_4x + B_4y + C_4z + D_4 = 0,$$

are concurrent at a single point, then

$$\begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} = 0. \tag{4.6.3.15}$$

To find the points of intersection, it suffices to find the point of intersection of any three of them (see Paragraph 4.6.3-9). The remaining equation follows from the three other equations.

4.6.3-11. Angle between straight line and plane.

Consider a plane given by the general equation

$$Ax + By + Cz + D = 0, \qquad \text{or} \qquad \mathbf{r} \cdot \mathbf{N} + D = 0 \tag{4.6.3.16}$$

and a line given by the canonical equation

$$\frac{x - x_1}{l} = \frac{y - y_1}{m} = \frac{z - z_1}{n}, \qquad \text{or} \qquad (\mathbf{r} - \mathbf{r}_1) \times \mathbf{R} = 0, \tag{4.6.3.17}$$

where $\mathbf{N} = (A, B, C)$ is the normal to the plane, $\mathbf{r}$ and $\mathbf{r}_1$ are the respective position vectors of the points (x, y, z) and (x_1, y_1, z_1), and $\mathbf{R} = (l, m, n)$ is the direction vector of the line.

The angle between the line and the plane (see Fig. 4.50) is defined as the complementary angle θ of the angle φ between the direction vector $\mathbf{R}$ of the line and the normal $\mathbf{N}$ to the plane. For this angle, one has the formula

$$\sin\theta = \cos\varphi = \frac{Al + Bm + Cn}{\sqrt{A^2 + B^2 + C^2}\sqrt{l^2 + m^2 + n^2}} = \frac{\mathbf{N} \cdot \mathbf{R}}{|\mathbf{N}|\,|\mathbf{R}|}. \tag{4.6.3.18}$$

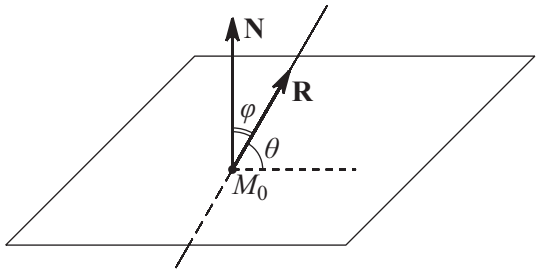


Figure 4.50. Angle between straight line and plane.

4.6.3-12. Conditions for line and plane to be parallel.

A plane given by the general equation (4.6.3.16) and a line given by the canonical equation (4.6.3.17) are parallel if

$$\begin{aligned} &Al + Bm + Cn = 0, \\ &Ax_1 + By_1 + Cz_1 + D \neq 0, \end{aligned} \qquad \text{or} \qquad \begin{aligned} &\mathbf{N} \cdot \mathbf{R} = 0, \\ &\mathbf{N} \cdot \mathbf{r}_1 + D \neq 0; \end{aligned} \tag{4.6.3.19}$$

i.e., a line is parallel to a plane if the direction vector of the line is perpendicular to the normal to the plane. Conditions (4.6.3.19) include the condition under which the line is not contained in the plane.

4.6.3-13. Condition for line to be entirely contained in plane.

A straight line given by the canonical equation (4.6.3.17) is entirely contained in a plane given by the general equation (4.6.3.16) if

$$\begin{aligned} &Al + Bm + Cn = 0, \\ &Ax_1 + By_1 + Cz_1 + D = 0, \end{aligned} \qquad \text{or} \qquad \begin{aligned} &\mathbf{N} \cdot \mathbf{R} = 0, \\ &\mathbf{N} \cdot \mathbf{r}_1 + D = 0. \end{aligned} \tag{4.6.3.20}$$

Remark. Sometimes the case in which a line is entirely contained in a plane is treated as a special case of parallel straight lines and is not distinguished as an exception.

4.6.3-14. Condition for line and plane to be perpendicular.

A line given by the canonical equation (4.6.3.17) and a plane given by the general equation (4.6.3.16) are perpendicular if the line is collinear to the normal to the plane (is a normal itself), i.e., if

$$\frac{A}{l} = \frac{B}{m} = \frac{C}{n}, \qquad \text{or} \qquad \mathbf{N} = \lambda \mathbf{R}, \qquad \text{or} \qquad \mathbf{N} \times \mathbf{R} = 0. \tag{4.6.3.21}$$

4.6.3-15. Intersection of line and plane.

Consider a plane given by the general equation (4.6.3.16) and a straight line given by the parametric equation

$$x = x_1 + lt, \quad y = y_1 + mt, \quad z = z_1 + nt, \qquad \text{or} \qquad \mathbf{r} = \mathbf{r}_1 + t\mathbf{R}.$$

The coordinates of the point $M_0(x_0, y_0, z_0)$ of intersection of the line with the plane (see Fig. 4.50), if the point exists at all, are determined by the formulas

$$x_0 = x_1 + lt_0, \quad y_0 = y_1 + mt_0, \quad z_0 = z_1 + nt_0, \qquad \text{or} \qquad \mathbf{r} = \mathbf{r}_1 + t_0\mathbf{R}, \tag{4.6.3.22}$$

where the parameter t_0 is determined from the relation

$$t_0 = -\frac{Ax_1 + By_1 + Cz_1 + D}{Al + Bm + Cn} = -\frac{\mathbf{N}\cdot\mathbf{r}_1 + D}{\mathbf{N}\cdot\mathbf{R}}. \tag{4.6.3.23}$$

Remark. To obtain formulas (4.6.3.22) and (4.6.3.23), one should rewrite the equation of the straight line in parametric form and replace x, y, and z in equation (4.6.3.16) of the plane by their expressions via t. From the resulting expression, one finds the parameter t_0 and then the coordinates x_0, y_0, and z_0 themselves.

Example 6. Let us find the point of intersection of the line $x/2 = (y-1)/1 = (z+1)/2$ with the plane $x + 2y + 3z - 29 = 0$.

We use formula (4.6.3.23) to find the value of the parameter t_0:

$$t_0 = -\frac{1\cdot 0 + 2\cdot 1 + 3\cdot(-1) - 29}{1\cdot 2 + 2\cdot 1 + 3\cdot 2} = -\frac{-30}{10} = 3.$$

Then, according to (4.6.3.22), we finally obtain the coordinates of the point of intersection in the form

$$x_0 = 0 - 2\cdot 3 = 6, \quad y_0 = 1 - 1\cdot 3 = 4, \quad z_0 = -1 - 2\cdot 3 = 5.$$

4.6.3-16. Distance from point to plane.

The *deviation of a point from a plane* is defined as the number δ equal to the length of the perpendicular drawn from this point to the plane and taken with sign + if the point and the origin lie on opposite sides of the plane and with sign – if they lie on the same side of the plane. Obviously, the deviation is zero for the points lying on the plane.

To obtain the deviation of a point $M_1(x_1, y_1, z_1)$ from a given plane, one should replace the current Cartesian coordinates (x, y, z) on the left-hand side in the normal equation (4.6.1.7) of this plane by the coordinates of the point M_1:

$$\delta = x_1\cos\alpha + y_1\cos\beta + z_1\cos\gamma - p = \mathbf{r}_1\cdot\mathbf{N}^0 - p, \tag{4.6.3.24}$$

where $\mathbf{N}^0 = (\cos\alpha, \cos\beta, \cos\gamma)$ is a unit vector and $\mathbf{r}_1$ is the position vector of the point $M_1(x_1, y_1, z_1)$. If the plane is given by the parametric equation (4.6.1.5), then the deviation of the point M_1 from the plane is equal to

$$\delta = \frac{[(\mathbf{r}_1 - \mathbf{r}_0)\mathbf{R}_1\mathbf{R}_2]}{|\mathbf{R}_1\times\mathbf{R}_2|}. \tag{4.6.3.25}$$

The *distance from a point to a plane* is defined as the nonnegative number d equal to the absolute value of the deviation; i.e.,

$$d = |\delta| = |x_0\cos\alpha + y_0\cos\beta + z_0\cos\gamma - p\,|. \tag{4.6.3.26}$$

Let us write out some more representations of the distance for the cases in which the plane is given by the general equation (4.6.1.3) and the parametric equation (4.6.1.5):

$$d = \frac{|Ax_0 + By_0 + Cz_0 + D|}{\sqrt{A^2 + B^2 + C^2}} = \frac{\big|[(\mathbf{r}_1 - \mathbf{r}_0)\mathbf{R}_1\mathbf{R}_2]\big|}{|\mathbf{R}_1\times\mathbf{R}_2|}. \tag{4.6.3.27}$$

Example 7. Let us find the distance from the point $M(5, 1, -1)$ to the plane $x - 2y - 2z + 1 = 0$.

Using formula (4.6.3.27), we obtain the desired distance

$$d = \frac{|1\cdot 5 + (-2)\cdot 1 + (-2)\cdot(-1) + 1|}{\sqrt{1^2 + (-2)^2 + (-2)^2}} = \frac{6}{3} = 2.$$

4.6.3-17. Distance between two parallel planes.

We consider two parallel planes given by the general equations $Ax + By + Cz + D_1 = 0$ and $Ax + By + Cz + D_2 = 0$. The distance between them is

$$d = \frac{|D_1 - D_2|}{\sqrt{A^2 + B^2 + C^2}}. \tag{4.6.3.28}$$

4.6.3-18. Distance from point to line.

The distance from a point $M_0(x_0, y_0, z_0)$ to a line given by the canonical equation (4.6.2.3) is determined by the formula

$$d = \frac{|\mathbf{R} \times (\mathbf{r}_0 - \mathbf{r}_1)|}{|\mathbf{R}|} = \frac{\sqrt{\begin{vmatrix} m & n \\ y_1 - y_0 & z_1 - z_0 \end{vmatrix}^2 + \begin{vmatrix} n & l \\ z_1 - z_0 & x_1 - x_0 \end{vmatrix}^2 + \begin{vmatrix} l & m \\ x_1 - x_0 & y_1 - y_0 \end{vmatrix}^2}}{\sqrt{l^2 + m^2 + n^2}}. \tag{4.6.3.29}$$

Note that the last formulas are significantly simplified if $\mathbf{R}$ is the unit vector ($l^2+m^2+n^2=1$).

Remark. The numerator of the fraction (4.6.3.29) is the area of the triangle spanned by the vectors $\mathbf{r}_0 - \mathbf{r}_1$ and $\mathbf{R}$, while the denominator of this fraction is the length of the base of the triangle. Hence the fraction itself is the altitude d of this triangle.

Example 8. Let us find the distance from the point $M_0(3, 0, 4)$ to the line $x/1 = (y-1)/2 = z/2$. We use formula (4.6.3.29) to obtain the desired distance

$$d = \frac{\sqrt{\begin{vmatrix} 2 & 2 \\ 1-0 & 0-4 \end{vmatrix}^2 + \begin{vmatrix} 2 & 1 \\ 0-4 & 0-3 \end{vmatrix}^2 + \begin{vmatrix} 1 & 2 \\ 0-3 & 1-0 \end{vmatrix}^2}}{\sqrt{1^2 + 2^2 + 2^2}} = \frac{\sqrt{153}}{3}.$$

4.6.3-19. Distance between lines.

Consider two nonparallel lines (see Fig. 4.51) given in the canonical form

$$\frac{x - x_1}{l_1} = \frac{y - y_1}{m_1} = \frac{z - z_1}{n_1}, \qquad \text{or} \qquad (\mathbf{r} - \mathbf{r}_1) \times \mathbf{R}_1 = 0$$

and

$$\frac{x - x_2}{l_2} = \frac{y - y_2}{m_2} = \frac{z - z_2}{n_2}, \qquad \text{or} \qquad (\mathbf{r} - \mathbf{r}_2) \times \mathbf{R}_2 = 0.$$

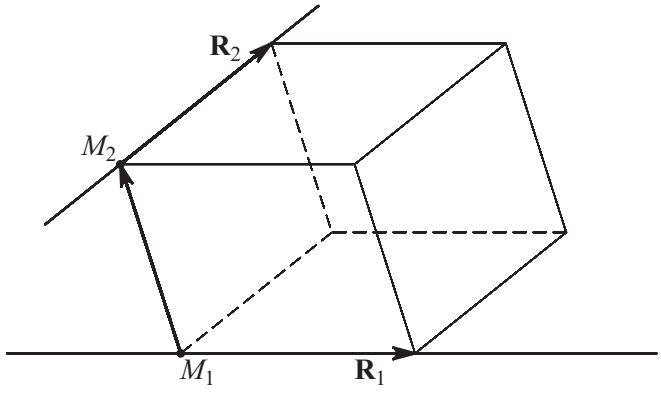


Figure 4.51. Distance between lines.

The distance between them can be calculated by the formula

$$d = \frac{\left|[(\mathbf{r}_1 - \mathbf{r}_2)\mathbf{R}_1\mathbf{R}_2]\right|}{|\mathbf{R}_1 \times \mathbf{R}_2|} = \frac{\pm \begin{vmatrix} x_1 - x_2 & y_1 - y_2 & z_1 - z_2 \\ l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{vmatrix}}{\sqrt{\begin{vmatrix} l_1 & m_1 \\ l_2 & m_2 \end{vmatrix}^2 + \begin{vmatrix} m_1 & n_1 \\ m_2 & n_2 \end{vmatrix}^2 + \begin{vmatrix} n_1 & l_1 \\ n_2 & l_2 \end{vmatrix}^2}}. \qquad (4.6.3.30)$$

The condition that the determinant in the numerator in (4.6.3.30) is zero is the *condition for the two lines in space to meet*.

Remark 1. The numerator of the fraction in (4.6.3.30) is the volume of the parallelepiped spanned by the vectors $\mathbf{r}_1 - \mathbf{r}_2$, $\mathbf{R}_1$, and $\mathbf{R}_2$, while the denominator of the fraction is the area of its base. Hence the fraction itself is the altitude d of this parallelepiped.

Remark 2. If the lines are parallel (i.e., $l_1 = l_2 = l$, $m_1 = m_2 = m$, and $n_1 = n_2 = n$, or $\mathbf{R}_1 = \mathbf{R}_2 = \mathbf{R}$), then the distance between them should be calculated by formula (4.6.3.29) with $\mathbf{r}_0$ replaced by $\mathbf{r}_2$.

4.7. Quadric Surfaces (Quadrics)

4.7.1. Quadrics (Canonical Equations)

4.7.1-1. Central surfaces.

A segment joining two points of a surface is called a *chord*. If there exists a point in space, not necessarily lying on the surface, that bisects all chords passing through it, then the surface is said to be *central* and the point is called the *center* of the surface.

The equations listed below in Paragraphs 4.7.1-2 to 4.7.1-4 for central surfaces are given in *canonical form*; i.e., the center of a surface is at the origin, and the surface symmetry axes are the coordinate axes. Moreover, the coordinate planes are symmetry planes.

4.7.1-2. Ellipsoid.

An *ellipsoid* is a surface defined by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1, \qquad (4.7.1.1)$$

where the numbers a, b, and c are the lengths of the segments called the *semiaxes* of the ellipsoid (see Fig 4.52*a*).

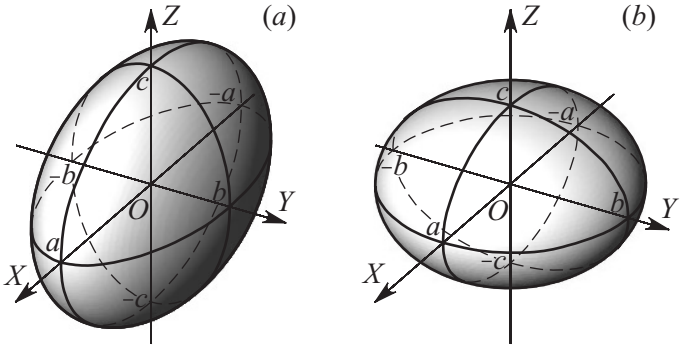


Figure 4.52. Triaxial ellipsoid (*a*) and spheroid (*b*).

If $a \neq b \neq c$, then the ellipsoid is said to be *triaxial*, or *scalene*. If $a = b \neq c$, then the ellipsoid is called a *spheroid*; it can be obtained by rotating the ellipse $x^2/a^2 + z^2/c^2 = 1$, $y = 0$ lying in the plane OXZ about the axis OZ (see Fig. 4.52*b*). If $a = b > c$, then the ellipsoid is an *oblate spheroid*, and if $a = b < c$, then the ellipsoid is a *prolate spheroid*. If $a = b = c$, then the ellipsoid is the sphere of radius a given by the equation $x^2 + y^2 + z^2 = a^2$.

An arbitrary plane section of an ellipsoid is an ellipse (in a special case, a circle). The volume of an ellipsoid is equal to $V = \frac{4}{3}\pi abc$.

Remark. About the sphere, see also Paragraph 3.2.3-3.

4.7.1-3. Hyperboloids.

A *one-sheeted hyperboloid* is a surface defined by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1, \tag{4.7.1.2}$$

where a and b are the *real semiaxes* and c is the *imaginary semiaxis* (see Fig. 4.53*a*).

A *two-sheeted hyperboloid* is a surface defined by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1, \tag{4.7.1.3}$$

where c is the *real semiaxis* and a and b are the *imaginary semiaxes* (see Fig 4.53*b*).

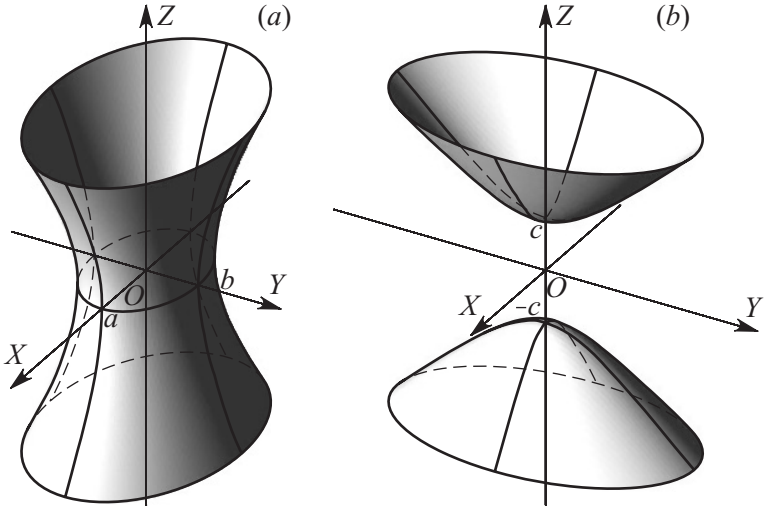


Figure 4.53. One-sheeted (*a*) and two-sheeted (*b*) hyperboloids.

A hyperboloid approaches the surface

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0,$$

which is called an *asymptotic cone*, infinitely closely.

A plane passing through the axis OZ intersects each of the hyperboloids (4.7.1.2) and (4.7.1.3) in two hyperbolas and the asymptotic cone in two straight lines, which are the

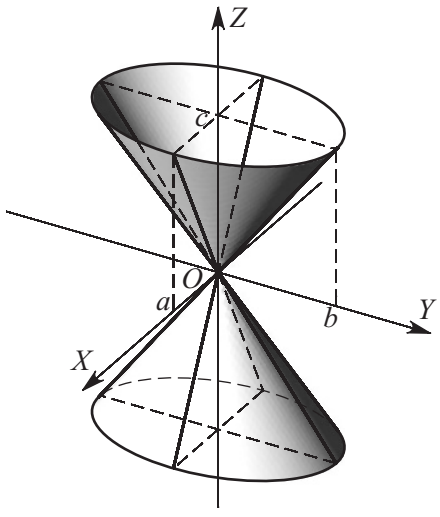


Figure 4.54. A cone.

asymptotes of these hyperbolas. The section of a hyperboloid by a plane parallel to OXY is an ellipse. The section of a one-sheeted hyperboloid by the plane $z = 0$ is an ellipse, which is called the *gorge* or *throat* ellipse.

For $a = b$, we deal with the *hyperboloid of revolution* obtained by rotating a hyperbola with semiaxes a and c about its focal axis $2c$ (which is an imaginary axis for a one-sheeted hyperboloid and a real axis for a two-sheeted hyperboloid). If $a = b = c$, then the hyperboloid of revolution is said to be *right*, and its sections by the planes OXZ and OYZ are equilateral hyperbolas.

A one-sheeted hyperboloid is a ruled surface (see Paragraph 4.7.1-6).

4.7.1-4. Cone.

A *cone* is a surface defined by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0. \tag{4.7.1.4}$$

The cone (see Fig. 4.54) defined by (4.7.1.4) has vertex at the origin, and for its base we can take the ellipse with semiaxes a and b in the plane perpendicular to the axis OZ at the distance c from the origin. This cone is the asymptotic cone for the hyperboloids (4.7.1.2) and (4.7.1.3). For $a = b$, we obtain a *right circular cone*.

A cone is a ruled surface (see Paragraph 4.7.1-6).

Remark. About the cone, see also Paragraph 3.2.3-2.

4.7.1-5. Paraboloids.

In contrast to the surfaces considered above, paraboloids are not central surfaces. For the equations listed below, the vertex of a paraboloid lies at the origin, the axis OZ is the symmetry axis, and the planes OXZ and OYZ are symmetry planes.

An *elliptic paraboloid* (see Fig 4.55*a*) is a surface defined by the equation

$$\frac{x^2}{p} + \frac{y^2}{q} = 2z, \tag{4.7.1.5}$$

where $p > 0$ and $q > 0$ are parameters.

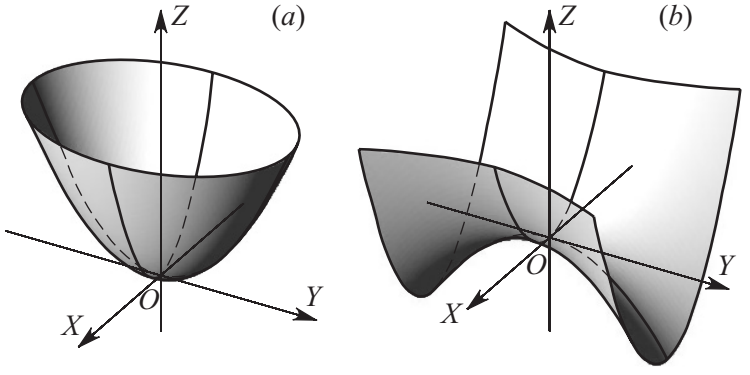


Figure 4.55. Elliptic (a) and hyperbolic (b) paraboloids.

The sections of an elliptic paraboloid by planes parallel to the axis OZ are parabolas, and the sections by planes parallel to the plane OXY are ellipses. For example, let the parabola $x^2 = 2pz$, $y = 0$, obtained by the section of an elliptic paraboloid by the plane OXZ be fixed and used as the directrix, and let the parabola $x^2 = 2qz$, $x = 0$, obtained by the section of the elliptic paraboloid by the plane OYZ be movable and used as the generator. Then the paraboloid can be obtained by parallel translation of the movable parabola (the generator) in a given direction along the fixed parabola (the directrix).

If $p = q$, then we have a *paraboloid of revolution*, which is obtained by rotating the parabola $2pz = x^2$ lying in the plane OXZ about its axis.

The volume of the part of an elliptic paraboloid cut by the plane perpendicular to its axis at a height h is equal to $V = \frac{1}{2}\pi abh$, i.e., half the volume of the elliptic cylinder with the same base and altitude.

A *hyperbolic paraboloid* (see Fig 4.55b) is a surface defined by the equation

$$\frac{x^2}{p} - \frac{y^2}{q} = 2z, \tag{4.7.1.6}$$

where $p > 0$ and $q > 0$ are parameters.

The sections of a hyperbolic paraboloid by planes parallel to the axis OZ are parabolas, and the sections by planes parallel to the plane OXY are hyperbolas. For example, let the parabola $x^2 = 2pz$, $y = 0$, obtained by the section of the hyperbolic paraboloid by the plane OXZ be fixed and used as the directrix, and let the parabola $x^2 = -2qz$, $x = 0$, obtained by the section of the hyperbolic paraboloid by the plane OYZ be movable and used as the generator. Then the paraboloid can be obtained by parallel translation of the movable parabola (the generator) in a given direction along the fixed parabola (the directrix).

A hyperbolic paraboloid is a ruled surface (see Paragraph 4.7.1-6).

4.7.1-6. Rulings of ruled surfaces.

A *ruled surface* is a surface swept out by a moving line in space. The straight lines forming a ruled surface are called *rulings*. Examples of ruled surfaces include the cone (see Paragraph 3.2.3-2 and 4.7.1-4), the cylinder (see Paragraph 3.2.3-1), the one-sheeted hyperboloid (see Paragraph 4.7.1-3), and the hyperbolic paraboloid (see Paragraph 4.7.1-5).

The cone (4.7.1.4) has one family of rulings,

$$\alpha x = \beta y, \quad \frac{\sqrt{\alpha^2 a^2 + \beta^2 b^2}}{ab} x = \frac{\beta}{c} z.$$

Properties of rulings of the cone:

1. There is a unique ruling through each point of the cone.
2. Two arbitrary distinct rulings of the cone meet at the point $O(0, 0, 0)$.
3. Three pairwise distinct rulings of the cone are not parallel to any plane.

The one-sheeted hyperboloid (4.7.1.2) has two families of rulings:

$$\begin{aligned} &\alpha\Big(\frac{x}{a} + \frac{z}{c}\Big) = \beta\Big(1 + \frac{y}{b}\Big), \quad \beta\Big(\frac{x}{a} - \frac{z}{c}\Big) = \alpha\Big(1 - \frac{y}{b}\Big); \\ &\gamma\Big(\frac{x}{a} + \frac{z}{c}\Big) = \delta\Big(1 - \frac{y}{b}\Big), \quad \delta\Big(\frac{x}{a} - \frac{z}{c}\Big) = \gamma\Big(1 + \frac{y}{b}\Big). \end{aligned} \tag{4.7.1.7}$$

One of these families is shown in Fig. 4.56*a*.

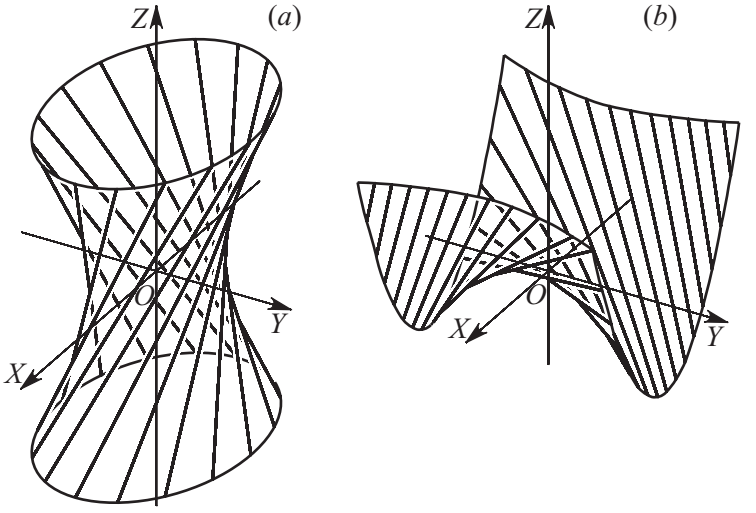


Figure 4.56. Families of rulings for one-sheeted hyperboloid (*a*) and for hyperbolic paraboloid (*b*).

Properties of rulings of the one-sheeted hyperboloid:

1. In either family, there is a unique ruling through each point of the one-sheeted hyperboloid.
2. Any two rulings in different families lie in a single plane.
3. Any two distinct rulings in the same family are skew.
4. Three distinct rulings in the same family are not parallel to any plane.

The hyperbolic paraboloid (4.7.1.6) has two families of rulings:

$$\begin{aligned} &\alpha\left(\frac{x}{\sqrt{p}} + \frac{y}{\sqrt{q}}\right) = 2\beta, \quad \beta\left(\frac{x}{\sqrt{p}} - \frac{y}{\sqrt{q}}\right) = \alpha z; \\ &\gamma\left(\frac{x}{\sqrt{p}} + \frac{y}{\sqrt{q}}\right) = \delta z, \quad \delta\left(\frac{x}{\sqrt{p}} - \frac{y}{\sqrt{q}}\right) = 2\gamma. \end{aligned} \tag{4.7.1.8}$$

One of these families is shown in Fig. 4.56*b*.

Properties of rulings of a hyperbolic paraboloid:

1. In either family, there is a unique ruling through each point of the hyperbolic paraboloid.
2. Any two rulings in different families lie in a single plane and meet.
3. Any two distinct rulings in the same family are skew.
4. All rulings in either family are parallel to a single plane.

4.7.2. Quadrics (General Theory)

4.7.2-1. General equation of quadric.

A *quadric* is a set of points in three-dimensional space whose coordinates in the rectangular Cartesian coordinate system satisfy a second-order algebraic equation

$$a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{13}xz + 2a_{23}yz + 2a_{14}x + 2a_{24}y + 2a_{34}z + a_{44} = 0, \qquad (4.7.2.1)$$

or

$$(a_{11}x + a_{12}y + a_{13}z + a_{14})x + (a_{21}x + a_{22}y + a_{23}z + a_{24})y + (a_{31}x + a_{32}y + a_{33}z + a_{34})z + a_{41}x + a_{42}y + a_{43}z + a_{44} = 0,$$

where $a_{ij} = a_{ji}$ $(i, j = 1, 2, 3, 4)$. If equation (4.7.2.1) does not define a real geometric object, then one says that this equation defines an *imaginary quadric*. Equation (4.7.2.1) in vector form reads

$$(\mathbf{Ar}) \cdot \mathbf{r} + 2\mathbf{a} \cdot \mathbf{r} + a_{44} = 0, \qquad (4.7.2.2)$$

where $\mathbf{A}$ is the affinor with coordinates $A^i_j = a_{ij}$ and $\mathbf{a}$ is the vector with coordinates $a_i = a_{i4}$.

4.7.2-2. Classification of quadrics.

There exists a rectangular Cartesian coordinate system in which equation (4.7.2.1), depending on the coefficients, has 1 of 17 *canonical forms*, each of which is associated with a certain class of quadrics (see Table 4.3).

4.7.2-3. Invariants of quadrics.

The shape of a quadric can be determined by using four *invariants* and two *semi-invariants* without reducing equation (4.7.2.1) to canonical form.

The main invariants are the quantities

$$S = a_{11} + a_{22} + a_{33}, \qquad (4.7.2.3)$$

$$T = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{23} \\ a_{23} & a_{33} \end{vmatrix}, \qquad (4.7.2.4)$$

$$\delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix}, \qquad (4.7.2.5)$$

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12} & a_{22} & a_{23} & a_{24} \\ a_{13} & a_{23} & a_{33} & a_{34} \\ a_{14} & a_{24} & a_{34} & a_{44} \end{vmatrix}, \qquad (4.7.2.6)$$

whose values are preserved under parallel translations and rotations of the coordinate axes.

TABLE 4.3
Canonical equations and classes of quadrics

No.	Surface	Canonical equation	Type	Class
Irreducible surfaces				
1	Ellipsoid	$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1$	Elliptic	Nondegenerate
2	Imaginary ellipsoid	$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=-1$		
3	One-sheeted hyperboloid	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1$	Hyperbolic	
4	Two-sheeted hyperboloid	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1$		
5	Elliptic paraboloid	$\frac{x^2}{p}+\frac{y^2}{q}=2z$	Parabolic ($p>0,\ q>0$)	
6	Hyperbolic paraboloid	$\frac{x^2}{p}-\frac{y^2}{q}=2z$		
7	Elliptic cylinder	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$	Cylindrical	Degenerate
8	Imaginary elliptic cylinder	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=-1$		
9	Hyperbolic cylinder	$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$		
10	Parabolic cylinder	$y^2=2px$		
11	Real cone	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0$	Conic	
12	Imaginary cone with real vertex	$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=0$		
Reducible surfaces				
13	Pair of real intersecting planes	$\frac{x^2}{a^2}-\frac{y^2}{b^2}=0$	Pairs of planes	Degenerate
14	Pair of imaginary planes intersecting in a real straight line	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=0$		
15	Pair of real parallel planes	$x^2=a^2$		
16	Pair of imaginary parallel planes	$x^2=-a^2$		
17	Pair of real coinciding planes	$x^2=0$		

The semi-invariants are the quantities

$$\sigma = \Delta_{11} + \Delta_{22} + \Delta_{33}, \tag{4.7.2.7}$$

$$\Sigma = \begin{vmatrix} a_{11} & a_{14} \\ a_{41} & a_{44} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{24} \\ a_{42} & a_{44} \end{vmatrix} + \begin{vmatrix} a_{33} & a_{34} \\ a_{44} & a_{44} \end{vmatrix}, \tag{4.7.2.8}$$

whose values are preserved only under rotations of the coordinate axes. Here Δ_{ij} is the cofactor of the entry a_{ij} in Δ.

The classification of quadrics based on the invariants S, T, δ, and Δ and the semi-invariants σ and Σ is given in Tables 4.4 and 4.5.

TABLE 4.4
Classification of quadrics (central surfaces $\delta \neq 0$)

Class		$S\delta > 0$ or $T > 0$	$S\delta > 0$ and $T > 0$ but not both
Nondegenerate surfaces ($\Delta \neq 0$)	$\Delta < 0$	Ellipsoid $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$	Two-sheeted hyperboloid $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1$
	$\Delta > 0$	Imaginary ellipsoid $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = -1$	One-sheeted hyperboloid $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$
Degenerate surfaces ($\Delta = 0$)		Imaginary cone with real vertex $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 0$	Real cone $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0$

TABLE 4.5
Classification of quadrics (central surfaces $\delta = 0$)

<table>
<tr><th colspan="2">Class and type</th><th>$T > 0$</th><th>$T < 0$</th><th>$T = 0$</th></tr>
<tr><td rowspan="2">Nondegenerate surfaces ($\Delta \neq 0$)</td><td>$\Delta > 0$</td><td></td><td>Hyperbolic paraboloid $\frac{x^2}{p} - \frac{y^2}{q} = 2z$</td><td></td></tr>
<tr><td>$\Delta < 0$</td><td>Elliptic paraboloid $\frac{x^2}{p} + \frac{y^2}{q} = 2z$</td><td></td><td></td></tr>
<tr><td rowspan="3">Degenerate surfaces ($\Delta = 0$)</td><td rowspan="2">Cylindrical surfaces ($\sigma \neq 0$)</td><td colspan="2">Elliptic cylinder</td><td rowspan="2">Hyperbolic cylinder $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$</td><td rowspan="2">Parabolic cylinder $y^2 = 2px$</td></tr>
<tr><td>Imaginary ($\sigma S > 0$) $\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1$</td><td>Real ($\sigma S < 0$) $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$</td></tr>
<tr><td>Reducible surfaces ($\sigma = 0$)</td><td colspan="2">Pair of imaginary planes, intersecting in a real straight line $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0$</td><td>Pair of real intersecting planes $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$</td><td colspan="2">Pair of real reducible planes ($\Sigma = 0$) $x^2 = 0$
Pair of imaginary parallel planes ($\Sigma > 0$) $x^2 = -a^2$ | Pair of real parallel planes ($\Sigma > 0$) $x^2 = a^2$</td></tr>
</table>

4.7.2-4. Characteristic quadratic form of quadric.

The *characteristic quadratic form*

$$F(x, y, z) = a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{13}xz + a_{23}yz \qquad (4.7.2.9)$$

corresponding to equation (4.7.2.1) and its *characteristic equation*

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{12} & a_{22} - \lambda & a_{23} \\ a_{13} & a_{23} & a_{33} - \lambda \end{vmatrix} = 0, \qquad \text{or} \qquad \lambda^3 - S\lambda^2 + T\lambda - \delta = 0 \qquad (4.7.2.10)$$

permit studying the main properties of quadrics.

The roots λ_1, λ_2, and λ_3 of the characteristic equation (4.7.2.10) are the eigenvalues of the real symmetric matrix $[a_{ij}]$ and hence are always real. The invariants S, T, and δ can be expressed in terms of the roots λ_1, λ_2, and λ_3 as follows:

$$S = \lambda_1 + \lambda_2 + \lambda_3, \quad T = \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3, \quad \delta = \lambda_1\lambda_2\lambda_3. \tag{4.7.2.11}$$

4.7.2-5. Diameters and diameter plane.

The locus of midpoints of parallel chords of a quadric is the *diameter plane* conjugate to these chords (or the direction of these chords). The diameter plane conjugate to the chords with direction cosines $\cos\alpha$, $\cos\beta$, and $\cos\gamma$ is determined by the equation

$$\begin{aligned}(a_{11}x + a_{12}y + a_{13}z + a_{14})\cos\alpha + (a_{21}x + a_{22}y + a_{23}z + a_{24})\cos\beta \\ + (a_{31}x + a_{32}y + a_{33}z + a_{34})\cos\gamma = 0.\end{aligned} \tag{4.7.2.12}$$

The line in which two diameter planes meet is called the *diameter* conjugate to the family of planes parallel to the conjugate chords of these diameter planes. The equation of the diameter conjugate to the family of planes with given direction cosines $\cos l$, $\cos m$, and $\cos n$ of the normals has the form

$$\frac{a_{11}x+a_{12}y+a_{13}z+a_{14}}{\cos l} = \frac{a_{21}x+a_{22}y+a_{23}z+a_{24}}{\cos m} = \frac{a_{31}x+a_{32}y+a_{33}z+a_{34}}{\cos n}. \tag{4.7.2.13}$$

If a surface is central (for $\delta \neq 0$), then all diameters are concurrent at a single point, called the *center* of the surface. For $\delta = 0$, all diameters are parallel or lie in a single plane.

For central quadrics, the coordinates x_0, y_0, z_0 of the center are determined by the system of equations

$$\begin{aligned}a_{11}x_0 + a_{12}y_0 + a_{13}z_0 + a_{14} = 0,\\ a_{21}x_0 + a_{22}y_0 + a_{23}z_0 + a_{24} = 0,\\ a_{31}x_0 + a_{32}y_0 + a_{33}z_0 + a_{34} = 0,\end{aligned}$$

whence it follows that

$$x_0 = -\frac{1}{\delta}\begin{vmatrix} a_{14} & a_{12} & a_{13} \\ a_{24} & a_{22} & a_{23} \\ a_{34} & a_{32} & a_{33} \end{vmatrix}, \quad y_0 = -\frac{1}{\delta}\begin{vmatrix} a_{11} & a_{14} & a_{13} \\ a_{21} & a_{24} & a_{23} \\ a_{31} & a_{34} & a_{33} \end{vmatrix}, \quad z_0 = -\frac{1}{\delta}\begin{vmatrix} a_{11} & a_{12} & a_{14} \\ a_{21} & a_{22} & a_{24} \\ a_{31} & a_{32} & a_{34} \end{vmatrix}. \tag{4.7.2.14}$$

For a central surface, the translation of the origin to its center (4.7.2.14) transforms the equation of the surface to the form

$$a_{11}x_1^2 + a_{22}y_1^2 + a_{33}z_1^2 + 2a_{12}x_1y_1 + 2a_{13}x_1z_1 + a_{23}y_1z_1 + \frac{\Delta}{\delta} = 0, \tag{4.7.2.15}$$

where x_1, y_1, and z_1 are the coordinates in the new system.

Three diameters of a central quadric are said to be *conjugate* if each of them is conjugate to the plane of the other two diameters.

4.7.2-6. Principal planes and principal axes.

A *principal plane* of a quadric is a diameter plane perpendicular to the chords conjugate to it. A principal plane is a plane of symmetry of the quadric. Any quadric has at least one principal plane and at least two symmetry planes. Cylinders have symmetry planes

perpendicular to the generators; these are not principal planes. Each nondegenerate quadric has at least two principal planes perpendicular to each other. Any central quadric has at least three principal planes, three of which are pairwise perpendicular.

A *principal axis* is a diameter that is the line of intersection of two principal planes. A principal axis is a symmetry axis of the quadric. If a quadric has two principal axes, then it also has the third axis perpendicular to the first two principal axes. Any nondegenerate quadric has at least one principal axis. A central quadric has at least three pairwise perpendicular principal axes, which are normal to its principal diameter planes; three of the principal axes of a central surface are pairwise perpendicular.

4.7.2-7. Transformation of equation of quadric to canonical form.

Translating the origin and rotating the coordinate axes so that the normal to each of the pairwise perpendicular symmetry planes of a surface becomes parallel to one of the new coordinate axes, we can transform equation (4.7.2.1) of any nondegenerate quadric to one of the forms listed in Table 4.6.

TABLE 4.6
Expression of the parameters of the main quadrics via the invariants T, δ, and Δ and the roots λ_1, λ_2, and λ_3 of the characteristic equation

Surface	Canonical equation	The parameters a, b, c, p, q expressed via the invariants T, δ, Δ and the roots $\lambda_1, \lambda_2, \lambda_3$ of the characteristic equation	Remarks
Ellipsoid	$\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}+\dfrac{z^2}{c^2}=1$	$a^2=-\dfrac{1}{\lambda_3}\dfrac{\Delta}{\delta}$, $b^2=-\dfrac{1}{\lambda_2}\dfrac{\Delta}{\delta}$, $c^2=-\dfrac{1}{\lambda_1}\dfrac{\Delta}{\delta}$, $\delta=\lambda_1\lambda_2\lambda_3$	$a\ge b\ge c$, $\lambda_1\ge\lambda_2\ge\lambda_3>0$
One-sheeted hyperboloid	$\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}-\dfrac{z^2}{c^2}=1$	$a^2=-\dfrac{1}{\lambda_2}\dfrac{\Delta}{\delta}$, $b^2=-\dfrac{1}{\lambda_1}\dfrac{\Delta}{\delta}$, $c^2=\dfrac{1}{\lambda_3}\dfrac{\Delta}{\delta}$, $\delta=\lambda_1\lambda_2\lambda_3$	$a\ge b$, $\lambda_1\ge\lambda_2>0>\lambda_3$
Two-sheeted hyperboloid	$\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}-\dfrac{z^2}{c^2}=-1$	$a^2=\dfrac{1}{\lambda_3}\dfrac{\Delta}{\delta}$, $b^2=\dfrac{1}{\lambda_2}\dfrac{\Delta}{\delta}$, $c^2=-\dfrac{1}{\lambda_1}\dfrac{\Delta}{\delta}$, $\delta=\lambda_1\lambda_2\lambda_3$	$a\ge b$, $\lambda_1>0>\lambda_2\ge\lambda_3$
Elliptic paraboloid	$\dfrac{x^2}{p}+\dfrac{y^2}{q}=2z$	$p=\dfrac{1}{\lambda_2}\sqrt{-\dfrac{\Delta}{T}}$, $q=\dfrac{1}{\lambda_1}\sqrt{-\dfrac{\Delta}{T}}$, $T=\lambda_1\lambda_2$	$p>0$, $q>0$, $\lambda_1\ge\lambda_2>\lambda_3=0$
Hyperbolic paraboloid	$\dfrac{x^2}{p}-\dfrac{y^2}{q}=2z$	$p=\dfrac{1}{\lambda_1}\sqrt{-\dfrac{\Delta}{T}}$, $q=-\dfrac{1}{\lambda_3}\sqrt{-\dfrac{\Delta}{T}}$, $T=\lambda_1\lambda_3$	$p>0$, $q>0$, $\lambda_1>\lambda_2=0>\lambda_3$

4.7.2-8. Tangent planes and normals to quadric.

The *equation of the tangent plane to the surface* (4.7.2.1) at a point $M(x_1, y_1, z_1)$ has the form

$$a_{11}x_1x+a_{22}y_1y+a_{33}z_1z+a_{12}(y_1x+x_1y)+a_{13}(z_1x+x_1z)$$
$$+a_{23}(y_1z+z_1y)+a_{14}(x_1+x)+a_{24}(y_1+y)+a_{34}(z_1+z)+a_{44}=0.$$

or

$$(a_{11}x_1 + a_{12}y_1 + a_{13}z_1 + a_{14})x + (a_{21}x_1 + a_{22}y_1 + a_{23}z_1 + a_{24})y + (a_{31}x_1 + a_{32}y_1 + a_{33}z_1 + a_{34})z + a_{41}x_1 + a_{42}y_1 + a_{43}z_1 + a_{44} = 0.$$

The *equation of the normal to the surface* (4.7.2.1) at a point $M(x_1, y_1, z_1)$ has the form

$$\frac{x - x_1}{a_{11}x_1 + a_{12}y_1 + a_{13}z_1 + a_{14}} = \frac{y - y_1}{a_{21}x_1 + a_{22}y_1 + a_{23}z_1 + a_{24}} = \frac{z - z_1}{a_{31}x_1 + a_{32}y_1 + a_{33}z_1 + a_{34}}.$$

References for Chapter 4

Alexander, D. C. and Koeberlein, G. M., *Elementary Geometry for College Students, 3rd Edition*, Houghton Mifflin Company, Boston, 2002.

Blau, H. I., *Foundations of Plane Geometry*, Prentice Hall, Englewood Cliffs, New Jersey, 2002.

Chauvenet, W., *A Treatise on Elementary Geometry*, Adamant Media Corporation, Boston, 2001.

Efimov, N. V., *Higher Geometry*, Mir Publishers, Moscow, 1980.

Fogiel, M. (Editor),*High School Geometry Tutor, 2nd Edition*, Research & Education Association, Englewood Cliffs, New Jersey, 2003.

Fuller, G. and Tarwater, D., *Analytic Geometry, 7th Edition*, Addison Wesley, Boston, 1993.

Gustafson, R. D. and Frisk, P. D., *Elementary Geometry, 3rd Edition*, Wiley, New York, 1991.

Hartshorne, R., *Geometry: Euclid and Beyond*, Springer, New York, 2005.

Jacobs, H. R., *Geometry, 2nd Edition*, W. H. Freeman & Company, New York, 1987.

Jacobs, H. R., *Geometry: Seeing, Doing, Understanding, 3rd Edition*, W. H. Freeman & Company, New York, 2003.

Jurgensen, R. and Brown, R. G., *Geometry*, McDougal Littell/Houghton Mifflin, Boston, 2000.

Kay, D., *College Geometry: A Discovery Approach, 2nd Edition*, Addison Wesley, Boston, 2000.

Kletenik, D. V., *Problems in Analytic Geometry*, University Press of the Pacific, Honolulu, Hawaii, 2002.

Kostrikin, A. I. (Editor), *Exercises in Algebra: A Collection of Exercises in Algebra, Linear Algebra and Geometry*, Gordon & Breach, New York, 1996.

Kostrikin, A. I. and Shafarevich, I. R. (Editors), *Algebra I: Basic Notions of Algebra*, Springer-Verlag, Berlin, 1990.

Leff, L. S., *Geometry the Easy Way, 3rd Edition*, Barron's Educational Series, Hauppauge, New York, 1997.

Moise, E., *Elementary Geometry from an Advanced Standpoint, 3rd Edition*, Addison Wesley, Boston, 1990.

Musser, G. L., Burger, W. F., and Peterson, B. E., *Mathematics for Elementary Teachers: A Contemporary Approach, 6th Edition*, Wiley, New York, 2002.

Musser, G. L., Burger, W. F., and Peterson, B. E., *Essentials of Mathematics for Elementary Teachers: A Contemporary Approach, 6th Edition*, Wiley, New York, 2003.

Musser, G. L. and Trimpe, L. E., *College Geometry: A Problem Solving Approach with Applications*, Prentice Hall, Englewood Cliffs, New Jersey, 1994.

Postnikov, M. M., *Lectures in Geometry, Semester I, Analytic Geometry*, Mir Publishers, Moscow, 1982.

Privalov, I. I., *Analytic Geometry, 32nd Edition [in Russian]*, Lan, Moscow, 2003.

Riddle, D. R., *Analytic Geometry, 6th Edition*, Brooks Cole, Stamford, 1995.

Roe, J., *Elementary Geometry*, Oxford University Press, Oxford, 1993.

Schultze, A. and Sevenoak, F. L., *Plane and Solid Geometry*, Adamant Media Corporation, Boston, 2004.

Suetin, P. K., Kostrikin, A. I., and Manin, Yu. I., *Linear Algebra and Geometry*, Gordon & Breach, New York, 1997.

Tussy, A. S., *Basic Geometry for College Students: An Overview of the Fundamental Concepts of Geometry, 2nd Edition*, Brooks Cole, Stamford, 2002.

Vygodskii, M. Ya., *Mathematical Handbook: Higher Mathematics*, Mir Publishers, Moscow, 1971.

Weeks, A. W. and Adkins, J. B., *A Course in Geometry: Plane and Solid*, Bates Publishing, Sandwich, 1982.

Woods, F. S., *Higher Geometry: An Introduction to Advanced Methods in Analytic Geometry, Phoenix Edition*, Dover Publications, New York, 2005.

Chapter 5
Algebra

5.1. Polynomials and Algebraic Equations

5.1.1. Polynomials and Their Properties

5.1.1-1. Definition of polynomial.

A *polynomial of degree* n of a scalar variable x is an expression of the form

$$f(x) \equiv a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \quad (a_n \neq 0), \tag{5.1.1.1}$$

where $a_0, \ldots, a_n$ are real or complex numbers ($n = 0, 1, 2, \ldots$). Polynomials of degree zero are nonzero numbers.

Two polynomials are *equal* if they have the same coefficients of like powers of the variable.

5.1.1-2. Main operations over polynomials.

1°. The *sum* (*difference*) of two polynomials $f(x)$ of degree n and $g(x)$ of degree m is the polynomial of degree $l \le \max\{n, m\}$ whose coefficient of each power of x is equal to the sum (difference) of the coefficients of the same power of x in $f(x)$ and $g(x)$, i.e. if

$$g(x) \equiv b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0, \tag{5.1.1.2}$$

then the sum (difference) of polynomials (5.1.1.1) and (5.1.1.2) is

$$f(x) \pm g(x) = c_l x^l + c_{l-1}x^{l-1} + \cdots + c_1 x + c_0, \quad \text{where} \quad c_k = a_k \pm b_k \quad (k = 0, 1, \ldots, l).$$

If $n > m$ then $b_{m+1} = \ldots = b_n = 0$; if $n < m$ then $a_{n+1} = \ldots = a_m = 0$.

2°. To *multiply* a polynomial $f(x)$ of degree n by a polynomial $g(x)$ of degree m, one should multiply each term in $f(x)$ by each term in $g(x)$, add the products, and collect similar terms. The degree of the resulting polynomial is $n+m$. The product of polynomials (5.1.1.1) and (5.1.1.2) is

$$f(x)g(x) = c_{n+m}x^{n+m} + c_{n+m-1}x^{n+m-1} + \cdots + c_1 x + c_0, \qquad c_k = \sum_{i,j=0}^{i+j=k} a_i b_j,$$

where $k = 0, 1, \ldots, n + m$.

3°. Each polynomial $f(x)$ of degree n can be *divided* by any other polynomial $p(x)$ of degree m ($p(x) \neq 0$) *with remainder*, i.e., uniquely represented in the form $f(x) = p(x)q(x) + r(x)$, where $q(x)$ is a polynomial of degree $n - m$ (for $m \le n$) or $q(x) = 0$ (for $m > n$), referred to as the *quotient*, and $r(x)$ is a polynomial of degree $l < m$ or $r(x) = 0$, referred to as the *remainder*.

If $r(x) = 0$, then $f(x)$ is said to be *divisible* by $p(x)$ (without remainder).

If $m > n$, then $q(x) = 0$ and $r(x) \equiv f(x)$.

5.1.1-3. Methods for finding quotient and remainder.

1°. *Horner's scheme.* To divide a polynomial $f(x)$ of degree n (see (5.1.1.1)) by the polynomial $p(x) = x - b$, one uses Horner's scheme: the coefficients of $f(x)$ are written out in a row, starting from a_n; b is written on the left; then one writes the number a_n under a_n, the number $a_n b + a_{n-1} = b_{n-1}$ under a_{n-1}, the number $b_{n-1}b + a_{n-2} = b_{n-2}$ under a_{n-2}, ..., the number $b_1 b + a_0 = b_0$ under a_0. The number b_0 is the remainder in the division of $f(x)$ by $p(x)$, and $a_n, b_{n-1}, \ldots, b_1$ are the coefficients of the quotient.

Remark. To divide $f(x)$ by $p(x) = ax + b$ ($a \neq 0$) with remainder, one first uses Horner's scheme to divide by $p_1(x) = x - (-\frac{b}{a})$; now if $q_1(x)$ and r_1 are the quotient and remainder in the division of $f(x)$ by $p_1(x)$, then $q(x) = \frac{1}{a}q_1(x)$ and $r = r_1$ are the quotient and remainder in the division of $f(x)$ by $p(x)$.

Example 1. Let us divide $f(x) = x^3 - 2x^2 - 10x + 3$ by $p(x) = 2x + 5$.
We use Horner's scheme to divide $f(x)$ by $p_1(x) = x + 5/2$:

	1	–2	–10	3
$-\frac{5}{2}$	1	$-\frac{9}{2}$	$\frac{5}{4}$	$-\frac{1}{8}$

Thus $f(x) = p(x)q(x) + r(x)$, where

$$q(x) = \frac{1}{2}\left(x^2 - \frac{9}{2}x + \frac{5}{4}\right) = \frac{1}{2}x^2 - \frac{9}{4}x + \frac{5}{8}, \quad r = -\frac{1}{8}.$$

POLYNOMIAL REMAINDER THEOREM. *The remainder in the division of a polynomial $f(x)$ by the polynomial $p(x) = x - b$ is the number equal to the value of the polynomial $f(x)$ at $x = b$.*

2°. *Long division.* To divide a polynomial $f(x)$ of degree n by a polynomial $p(x)$ of degree $m \le n$, one can use *long division*.

Example 2. Let us divide $f(x) = x^3 + 8x^2 + 14x - 5$ by $p(x) = x^2 + 3x - 1$.
We use long division:

$$\begin{array}{r|l} x^3 + 8x^2 + 14x - 5 & x^2 + 3x - 1 \\ \hline -\ x^3 + 3x^2 \quad - x & x + 5 \\ \hline -\quad 5x^2 + 15x - 5 & \\ \quad 5x^2 + 15x - 5 & \\ \hline 0 & \end{array}$$

Thus $f(x) = p(x)q(x) + r(x)$, where $q(x) = x + 5$ and $r(x) = 0$; i.e., $f(x)$ is divisible by $p(x)$.

Example 3. Let us divide $f(x) = x^3 - 4x^2 + x + 1$ by $p(x) = x^2 + 1$.
We use long division:

$$\begin{array}{r|l} x^3 - 4x^2 + x + 1 & x^2 + 1 \\ \hline -\ x^3 \quad + x & x - 4 \\ \hline -\ -4x^2 \quad + 1 & \\ \quad -4x^2 \quad - 4 & \\ \hline 5 & \end{array}$$

Thus $f(x) = p(x)q(x) + r(x)$, where $q(x) = x - 4$ and $r(x) = 5$.

5.1.1-4. Expansion of polynomials in powers of linear binomial.

For each polynomial $f(x)$ given by equation (5.1.1.1) and any number c, one can write out the *expansion of $f(x)$ in powers of $x - c$*:

$$f(x) = b_n(x-c)^n + b_{n-1}(x-c)^{n-1} + \cdots + b_1(x-c) + b_0.$$

To find the coefficients $b_0, \ldots, b_n$ of this expansion, one first divides $f(x)$ by $x - c$ with remainder. The remainder is b_0, and the quotient is some polynomial $g_0(x)$. Then one divides $g_0(x)$ by $x - c$ with remainder. The remainder is b_1, and the quotient is some polynomial $g_1(x)$. Then one divides $g_1(x)$ by $x - c$, obtaining the coefficient b_2 as the remainder, etc. It is convenient to perform the computations by Horner's scheme (see Paragraph 5.1.1-3).

Example 4. Expand the polynomial $f(x) = x^4 - 5x^3 - 3x^2 + 9$ in powers of the difference $x - 3$ ($c = 3$).

We write out Horner's scheme, where the first row contains the coefficients of the polynomial $f(x)$, the second row contains the coefficients of the quotient $g_0(x)$ and the remainder b_0 obtained when dividing $f(x)$ by $x - 3$, the third row contains the coefficients of the quotient $g_1(x)$ and the remainder b_1 obtained when dividing $g_0(x)$ by $x - 3$, etc.:

	1	–5	–3	0	9
3	1	–2	–9	–27	–72
	1	1	–6	–45	
	1	4	6		
	1	7			
	1				

Thus the expansion of $f(x)$ in powers of $x - 3$ has the form

$$f(x) = (x-3)^4 + 7(x-3)^3 + 6(x-3)^2 - 45(x-3) - 72.$$

The coefficients in the expansion of a polynomial $f(x)$ in powers of the difference $x - c$ are related to the values of the polynomial and its derivatives at $x = c$ by the formulas

$$b_0 = f(c), \quad b_1 = \frac{f'_x(c)}{1!}, \quad b_2 = \frac{f''_{xx}(c)}{2!}, \quad \ldots, \quad b_n = \frac{f^{(n)}_x(c)}{n!},$$

where the *derivative of a polynomial* $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + x_0$ with real or complex coefficients $a_0, \ldots, a_n$ is the polynomial $f'_x(x) = na_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + a_1$, $f''_{xx}(x) = [f'_x(x)]'_x$, etc. Thus Horner's scheme permits one to find the values of the derivatives of the polynomial $f(x)$ at $x = c$.

Example 5. In Example 4, the values of the derivatives of the polynomial $f(x)$ at $x = 3$ are

$$f(3) = -72, \quad f'(3) = -45 \times 1! = -45, \quad f''(3) = 6 \times 2! = 12, \quad f'''(3) = 7 \times 3! = 42, \quad f^{\mathrm{IV}}(3) = 1 \times 4! = 24.$$

The expansion of a polynomial in powers of $x - c$ can be used to compute the partial fraction decomposition of a rational function whose denominator is a power of a linear binomial.

Example 6. Find the partial fraction decomposition of the rational function $\Phi(x) = \dfrac{x^2 + x + 1}{(x-2)^4}$.

First, we expand the polynomial $f(x) = x^2 + x + 1$ in powers of the binomial $x - (-2) = x + 2$:

	1	1	1
–2	1	–1	3
	1	–3	
	1		

Thus $f(x) = (x+2)^2 - 3(x+2) + 3$. As a result, we obtain $\Phi(x) = \dfrac{f(x)}{(x+2)^4} = \dfrac{1}{(x+2)^2} - \dfrac{3}{(x+2)^3} + \dfrac{3}{(x+2)^4}$.

5.1.2. Linear and Quadratic Equations

5.1.2-1. Linear equations.

The linear equation

$$ax + b = 0 \quad (a \neq 0)$$

has the solution

$$x = -\frac{b}{a}.$$

5.1.2-2. Quadratic equations.

The quadratic equation

$$ax^2 + bx + c = 0 \quad (a \neq 0) \tag{5.1.2.1}$$

has the roots

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

The existence of real or complex roots is determined by the sign of the discriminant $D = b^2 - 4ac$:

Case $D > 0$. There are two distinct real roots.

Case $D < 0$. There are two distinct complex conjugate roots.

Case $D = 0$. There are two equal real roots.

VIÈTE THEOREM. *The roots of a quadratic equation (5.1.2.1) satisfy the following relations:*

$$x_1 + x_2 = -\frac{b}{a},$$
$$x_1 x_2 = \frac{c}{a}.$$

5.1.3. Cubic Equations

5.1.3-1. Incomplete cubic equation.

1°. *Cardano's solution.* The roots of the incomplete cubic equation

$$y^3 + py + q = 0 \tag{5.1.3.1}$$

have the form

$$y_1 = A + B, \quad y_{2,3} = -\frac{1}{2}(A + B) \pm i\frac{\sqrt{3}}{2}(A - B),$$

where

$$A = \left(-\frac{q}{2} + \sqrt{D}\right)^{1/3}, \quad B = \left(-\frac{q}{2} - \sqrt{D}\right)^{1/3}, \quad D = \left(\frac{p}{3}\right)^3 + \left(\frac{q}{2}\right)^2, \quad i^2 = -1,$$

and A, B are arbitrary values of the cubic roots such that $AB = -\frac{1}{3}p$.

The number of real roots of a cubic equation depends on the sign of the discriminant D:

Case $D > 0$. There is one real and two complex conjugate roots.

Case $D < 0$. There are three real roots.

Case $D = 0$. There is one real root and another real root of double multiplicity (this case is realized for $p = q = 0$).

2°. *Trigonometric solution.* If an incomplete cubic equation (5.1.3.1) has real coefficients p and q, then its solutions can be found with the help of the trigonometric formulas given below.

(*a*) Let $p < 0$ and $D < 0$. Then

$$y_1 = 2\sqrt{-\frac{p}{3}}\cos\frac{\alpha}{3}, \quad y_{2,3} = -2\sqrt{-\frac{p}{3}}\cos\left(\frac{\alpha}{3} \pm \frac{\pi}{3}\right),$$

where the values of the trigonometric functions are calculated from the relation

$$\cos\alpha = -\frac{q}{2\sqrt{-(p/3)^3}}.$$

(*b*) Let $p > 0$ and $D \ge 0$. Then

$$y_1 = 2\sqrt{\frac{p}{3}}\,\operatorname{cot}(2\alpha), \quad y_{2,3} = \sqrt{\frac{p}{3}}\left[\operatorname{cot}(2\alpha) \pm i\frac{\sqrt{3}}{\sin(2\alpha)}\right],$$

where the values of the trigonometric functions are calculated from the relations

$$\tan\alpha = \left(\tan\frac{\beta}{2}\right)^{1/3}, \quad \tan\beta = \frac{2}{q}\left(\frac{p}{3}\right)^{3/2}, \quad |\alpha| \le \frac{\pi}{4}, \quad |\beta| \le \frac{\pi}{2}.$$

(*c*) Let $p < 0$ and $D \ge 0$. Then

$$y_1 = -2\sqrt{-\frac{p}{3}}\,\frac{1}{\sin(2\alpha)}, \quad y_{2,3} = \sqrt{-\frac{p}{3}}\left[\frac{1}{\sin(2\alpha)} \pm i\sqrt{3}\,\operatorname{cot}(2\alpha)\right],$$

where the values of the trigonometric functions are calculated from the relations

$$\tan\alpha = \left(\tan\frac{\beta}{2}\right)^{1/3}, \quad \sin\beta = \frac{2}{q}\left(-\frac{p}{3}\right)^{3/2}, \quad |\alpha| \le \frac{\pi}{4}, \quad |\beta| \le \frac{\pi}{2}.$$

In the above three cases, the real value of the cubic root should be taken.

5.1.3-2. Complete cubic equation.

The roots of a complete cubic equation

$$ax^3 + bx^2 + cx + d = 0 \quad (a \ne 0) \tag{5.1.3.2}$$

are calculated by the formulas

$$x_k = y_k - \frac{b}{3a}, \quad k = 1,\ 2,\ 3,$$

where y_k are the roots of the incomplete cubic equation (5.1.3.1) with the coefficients

$$p = -\frac{1}{3}\left(\frac{b}{a}\right)^2 + \frac{c}{a}, \quad q = \frac{2}{27}\left(\frac{b}{a}\right)^3 - \frac{bc}{3a^2} + \frac{d}{a}.$$

VIÈTE THEOREM. *The roots of a complete cubic equation (5.1.3.2) satisfy the following relations:*

$$x_1 + x_2 + x_3 = -\frac{b}{a},$$
$$x_1x_2 + x_1x_3 + x_2x_3 = \frac{c}{a},$$
$$x_1x_2x_3 = -\frac{d}{a}.$$

5.1.4. Fourth-Degree Equation

5.1.4-1. Special cases of fourth-degree equations.

1°. The *biquadratic equation*

$$ax^4 + bx^2 + c = 0$$

can be reduced to a quadratic equation (5.1.2.1) by the substitution $\xi = x^2$. Therefore, the roots of the biquadratic equations are given by

$$x_{1,2} = \pm\sqrt{\frac{-b + \sqrt{b^2 - 4ac}}{2a}}, \quad x_{3,4} = \pm\sqrt{\frac{-b - \sqrt{b^2 - 4ac}}{2a}}.$$

2°. The *reciprocal (algebraic) equation*

$$ax^4 + bx^3 + cx^2 + bx + a = 0$$

can be reduced to a quadratic equation by the substitution

$$y = x + \frac{1}{x}.$$

The resulting quadratic equation has the form

$$ay^2 + by + c - 2a = 0.$$

3°. The *modified reciprocal equation*

$$ax^4 + bx^3 + cx^2 - bx + a = 0$$

can be reduced to a quadratic equation by the substitution

$$y = x - \frac{1}{x}.$$

The resulting quadratic equation has the form

$$ay^2 + by + 2a + c = 0.$$

4°. The *generalized reciprocal equation*

$$ab^2x^4 + bx^3 + cx^2 + dx + ad^2 = 0$$

can be reduced to a quadratic equation by the substitution

$$y = bx + \frac{d}{x}.$$

The resulting quadratic equation has the form

$$ay^2 + y + c - 2abd = 0.$$

5.1.4-2. General fourth-degree equation.

1°. *Reduction of a general equation of fourth-degree to an incomplete equation.* The general equation of fourth-degree

$$ax^4 + bx^3 + cx^2 + dx + e = 0 \quad (a \neq 0)$$

can be reduced to an incomplete equation of the form

$$y^4 + py^2 + qy + r = 0 \tag{5.1.4.1}$$

by the substitution

$$x = y - \frac{b}{4a}.$$

2°. *Descartes–Euler solution.* The roots of the incomplete equation (5.1.4.1) are given by the formulas

$$\begin{aligned} y_1 &= \tfrac{1}{2}\bigl(\sqrt{z_1} + \sqrt{z_2} + \sqrt{z_3}\bigr), \quad & y_2 &= \tfrac{1}{2}\bigl(\sqrt{z_1} - \sqrt{z_2} - \sqrt{z_3}\bigr), \\ y_3 &= \tfrac{1}{2}\bigl(-\sqrt{z_1} + \sqrt{z_2} - \sqrt{z_3}\bigr), \quad & y_4 &= \tfrac{1}{2}\bigl(-\sqrt{z_1} - \sqrt{z_2} + \sqrt{z_3}\bigr), \end{aligned} \tag{5.1.4.2}$$

where z_1, z_2, z_3 are the roots of the cubic equation (cubic resolvent of equation (5.1.4.1))

$$z^3 + 2pz^2 + (p^2 - 4r)z - q^2 = 0. \tag{5.1.4.3}$$

The signs of the roots in (5.1.4.2) are chosen from the condition

$$\sqrt{z_1}\sqrt{z_2}\sqrt{z_3} = -q.$$

The roots of the fourth-degree equation (5.1.4.1) are determined by the roots of the cubic resolvent (5.1.4.3); see Table 5.1.

3°. *Ferrari solution.* Let z_0 be any of the roots of the auxiliary cubic equation (5.1.4.3). Then the four roots of the incomplete equation (5.1.4.1) are found by solving the following two quadratic equations:

$$y^2 - \sqrt{z_0}\, y + \frac{p + z_0}{2} + \frac{q}{2\sqrt{z_0}} = 0,$$
$$y^2 + \sqrt{z_0}\, y + \frac{p + z_0}{2} - \frac{q}{2\sqrt{z_0}} = 0.$$

TABLE 5.1
Relations between the roots of an incomplete equation of fourth-degree and the roots of its cubic resolvent

Cubic resolvent (5.1.4.3)	Fourth-degree equation (5.1.4.1)
All roots are real and positive*	Four real roots
All roots are real: one is positive and two are negative*	Two pairs of complex conjugate roots
One real root and two complex conjugate roots	Two real roots and two complex conjugate roots

5.1.5. Algebraic Equations of Arbitrary Degree and Their Properties

5.1.5-1. Simplest equations of degree n and their solutions.

1°. The *binomial algebraic equation*

$$x^n - a = 0 \quad (a \neq 0)$$

has the solutions

$$x_{k+1} = \begin{cases} a^{1/n}\left(\cos\dfrac{2k\pi}{n} + i\sin\dfrac{2k\pi}{n}\right) & \text{for } a > 0, \\ |a|^{1/n}\left(\cos\dfrac{(2k+1)\pi}{n} + i\sin\dfrac{(2k+1)\pi}{n}\right) & \text{for } a < 0, \end{cases}$$

where $k = 0, 1, \ldots, n-1$ and $i^2 = -1$.

2°. Equations of the form

$$x^{2n} + ax^n + b = 0,$$
$$x^{3n} + ax^{2n} + bx^n + c = 0,$$
$$x^{4n} + ax^{3n} + bx^{2n} + cx^n + d = 0$$

are reduced by the substitution $y = x^n$ to a quadratic, cubic, and fourth-degree equation, respectively, whose solution can be expressed by radicals.

Remark. In the above equations, n can be noninteger.

3°. The *reciprocal (algebraic) equation*

$$a_0x^{2n} + a_1x^{2n-1} + a_2x^{2n-2} + \cdots + a_2x^2 + a_1x + a_0 = 0 \quad (a_0 \neq 0)$$

can be reduced to an equation of degree n by the substitution

$$y = x + \frac{1}{x}.$$

* By the Viète theorem, the product of the roots z_1, z_2, z_3 is equal to $q^2 \geq 0$.

Example 1. The equation

$$ax^6 + bx^5 + cx^4 + dx^3 + cx^2 + bx + a = 0$$

can be reduced to the cubic equation

$$ay^3 + by^2 + (c - 3a)y + d - 2b = 0$$

by the substitution $y = x + 1/x$.

5.1.5-2. Equations of general form and their properties.

An *algebraic equation of degree* n has the form

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0 \quad (a_n \neq 0), \tag{5.1.5.1}$$

where a_k are real or complex coefficients. Denote the polynomial of degree n on the right-hand side in equation (5.1.5.1) by

$$P_n(x) \equiv a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \quad (a_n \neq 0). \tag{5.1.5.2}$$

A value $x = x_1$ such that $P_n(x_1) = 0$ is called a *root* of equation (5.1.5.1) (and also a root of the polynomial $P_n(x)$). A value $x = x_1$ is called a *root of multiplicity* m if $P_n(x) = (x - x_1)^m Q_{n-m}(x)$, where m is an integer ($1 \le m \le n$), and $Q_{n-m}(x)$ is a polynomial of degree $n - m$ such that $Q_{n-m}(x_1) \neq 0$.

THEOREM 1 (FUNDAMENTAL THEOREM OF ALGEBRA). *Any algebraic equation of degree n has exactly n roots (real or complex), each root counted according to its multiplicity.*

Thus, the left-hand side of equation (5.1.5.1) with roots $x_1, x_2, \ldots, x_s$ of the respective multiplicities $k_1, k_2, \ldots, k_s$ ($k_1 + k_2 + \cdots + k_s = n$) can be factorized as follows:

$$P_n(x) = a_n(x - x_1)^{k_1}(x - x_2)^{k_2} \ldots (x - x_s)^{k_s}.$$

THEOREM 2. *Any algebraic equation of an odd degree with real coefficients has at least one real root.*

THEOREM 3. *Suppose that equation (5.1.5.1) with real coefficients has a complex root $x_1 = \alpha + i\beta$. Then this equation has the complex conjugate root $x_2 = \alpha - i\beta$, and the roots x_1, x_2 have the same multiplicity.*

THEOREM 4. *Any rational root of equation (5.1.5.1) with integer coefficients a_k is an irreducible fraction of the form p/q, where p is a divisor of a_0 and q is a divisor of a_n. If $a_n = 1$, then all rational roots of equation (5.1.5.1) (if they exist) are integer divisors of the free term.*

THEOREM 5 (ABEL–RUFFINI THEOREM). *Any equation (5.1.5.1) of degree $n \le 4$ is solvable by radicals, i.e., its roots can be expressed via its coefficients by the operations of addition, subtraction, multiplication, division, and taking roots (see Subsections 5.1.2–5.1.4). In general, equation (5.1.5.1) of degree $n > 4$ cannot be solved by radicals.*

5.1.5-3. Relations between roots and coefficients. Discriminant of an equation.

VIÈTE THEOREM. *The roots of equation (5.1.5.1) (counted according to their multiplicity) and its coefficients satisfy the following relations:*

$$(-1)^k \frac{a_{n-k}}{a_n} = S_k \quad (k = 1, 2, \ldots, n),$$

where S_k *are elementary symmetric functions of* $x_1, x_2, \ldots, x_n$*:*

$$S_1 = \sum_{i=1}^{n} x_i, \quad S_2 = \sum_{1\le i<j}^{n} x_i x_j, \quad S_3 = \sum_{1\le i<j<k}^{n} x_i x_j x_k, \quad \ldots, \quad S_n = x_1 x_2 \ldots x_n.$$

Note also the following relations:

$$(n-k)a_{n-k} + \sum_{j=1}^{k} a_{n-(k-j)} s_j = 0 \quad (k = 1, 2, \ldots, n)$$

with symmetric functions $s_j = \sum\limits_{i=1}^{n} x_i^j$.

The *discriminant* D *of an algebraic equation* is the product of a_n^{2n-2} and the squared Vandermonde determinant $\Delta(x_1, x_2, \ldots, x_n)$ of its roots:

$$D = a_n^{2n-2}[\Delta(x_1, x_2, \ldots, x_n)]^2 = a_n^{2n-2} \prod_{1\le j<i\le n} (x_i - x_j)^2.$$

The discriminant D is a symmetric function of the roots $x_1, x_2, \ldots, x_n$, and is equal to zero if and only if the polynomial $P_n(x)$ has at least one multiple root.

5.1.5-4. Bounds for the roots of algebraic equations with real coefficients.

1°. All roots of equation (5.1.5.1) in absolute value do not exceed

$$N = 1 + \frac{A}{|a_n|}, \tag{5.1.5.3}$$

where A is the largest of $|a_0|, |a_1|, \ldots, |a_{n-1}|$.

The last result admits the following generalization: all roots of equation (5.1.5.1) in absolute value do not exceed

$$N_1 = \rho + \frac{A_1}{|a_n|}, \tag{5.1.5.4}$$

where $\rho > 0$ is arbitrary and A_1 is the largest of

$$|a_{n-1}|, \quad \frac{|a_{n-2}|}{\rho}, \quad \frac{|a_{n-3}|}{\rho^2}, \quad \ldots, \quad \frac{|a_0|}{\rho^{n-1}}.$$

For $\rho = 1$, formula (5.1.5.4) turns into (5.1.5.3).

Remark. Formulas (5.1.5.3) and (5.1.5.4) can also be used for equations with complex coefficients.

Example 2. Consider the following equation of degree 4:

$$P_4(x) = 9x^4 - 9x^2 - 36x + 1.$$

Formula (5.1.5.3) for $n = 4$, $|a_n| = 9$, $A = 36$ yields a fairly rough estimate $N = 5$, i.e., the roots of the equation belong to the interval $[-5, 5]$. Formula (5.1.5.4) for $\rho = 2$, $n = 4$, $|a_n| = 9$, $A_1 = 9$ yields a better estimate for the bounds of the roots of this polynomial, $N_1 = 3$.

2°. A constant K is called an upper bound for the real roots of equation (5.1.5.1) or the polynomial $P_n(x)$ if equation (5.1.5.1) has no real roots greater than or equal to K; in a similar way, one defines a lower and an upper bound for positive and negative roots of an equation or the corresponding polynomial.

Let

K_1 be an upper bound for the positive roots of the polynomial $P_n(x)$,

K_2 be an upper bound for the positive roots of the polynomial $P_n(-x)$,

$K_3 > 0$ be an upper bound for the positive roots of the polynomial $x^n P_n(1/x)$,

$K_4 > 0$ be an upper bound for the positive roots of the polynomial $x^n P_n(-1/x)$.

Then all nonzero real roots of the polynomial $P_n(x)$ (if they exist) belong to the intervals $(-K_2, -1/K_4)$ and $(1/K_3, K_1)$.

Next, we describe three methods for finding upper bounds for positive roots of a polynomial.

Maclaurin method. Suppose that the first m leading coefficients of the polynomial (5.1.5.2) are nonnegative, i.e., $a_n > 0$, $a_{n-1} \geq 0, \ldots, a_{n-m+1} \geq 0$, and the next coefficient is negative, $a_{n-m} < 0$. Then

$$K = 1 + \left(\frac{B}{a_n}\right)^{1/m} \tag{5.1.5.5}$$

is an upper bound for the positive roots of this polynomial, where B is the largest of the absolute values of negative coefficients of $P_n(x)$.

Example 3. Consider the fourth-degree equation from Example 2. In this case, $m = 2$, $B = 36$ and formula (5.1.5.5) yields $K = K_1 = 1 + (36/9)^{1/2} = 3$. Now, consider the polynomial $P_4(-x) = 9x^4 - 9x^2 + 36x + 1$. Its positive roots has the upper bound $K_2 = 1 + (9/9)^{1/2} = 2$. For the polynomial $x^4 P_4(1/x) = x^4 - 36x^3 - 9x^2 + 9$, we have $m = 1$, $K_3 = 1 + 36 = 37$. Finally, for the polynomial $x^4 P_4(-1/x) = x^4 + 36x^3 - 9x^2 + 9$, we have $m = 2$, $k_4 = 1 + 9^{1/2} = 4$. Thus if $P_4(x)$ has real roots, they must belong to the intervals $(-2, -1/4)$ and $(1/37, 3)$.

Newton method. Suppose that for $x = c$, the polynomial $P_n(x)$ and all its derivatives $P_n'(x), \ldots, P_n^{(n)}(x)$ take positive values. Then c is an upper bound for the positive roots of $P_n(x)$.

Example 4. Consider the polynomial from Example 2 and calculate the derivatives

$$\begin{aligned} P_4(x) &= 9x^4 - 9x^2 - 36x + 1, \\ P_4'(x) &= 36x^3 - 18x - 36, \\ P_4''(x) &= 108x^2 - 18, \\ P_4'''(x) &= 216x, \\ P_4''''(x) &= 216. \end{aligned}$$

It is easy to check that for $x = 2$ this polynomial and all its derivatives take positive values, and therefore $c = 2$ is an upper bound for its positive roots.

A *method based on the representation of a polynomial as a sum of polynomials.* Assuming $a_n > 0$, let us represent the polynomial (5.1.5.4) (without rearranging its terms) as the sum $P_n(x) = f_1(x) + \ldots + f_m(x)$, where each polynomial $f_k(x)$ ($k = 1, 2, \ldots, m$) has a positive leading coefficient and the sequence of its coefficients does not change sign more than once. Suppose that for $c > 0$ all these polynomials are positive, $f_1(c) > 0, \ldots, f_m(c) > 0$. Then c is an upper bound for the positive roots of $P_n(x)$.

Example 5. The polynomial

$$P_7(x) = x^7 + 2x^6 - 4x^5 - 7x^4 + 2x^3 - 3x^2 + ax + b \quad (a > 0,\ b > 0)$$

can be represented as a sum of three polynomials

$$f_1(x) = x^7 + 2x^6 - 4x^5 - 7x^4 = x^4(x^3 + 2x^2 - 4x - 7), \quad f_2(x) = 2x^3 - 3x^2 = x^2(2x - 3), \quad f_3(x) = ax + b$$

(in the first two polynomials the sign of the sequence of coefficients changes once, and in the last polynomial the coefficients do not change sign). It is easy to see that all these polynomials are positive for $x = 2$. Therefore, $c = 2$ is an upper bound for the positive roots of the given polynomial.

5.1.5-5. Theorems on the number of real roots of polynomials.

The number all negative roots of a polynomial $P_n(x)$ is equal to the number of all positive roots of the polynomial $P_n(-x)$.

1°. The exact number of positive roots of a polynomial whose coefficients form a sequence that does not change sign or changes sign only once can be found with the help of the Descartes theorem.

DESCARTES THEOREM. *The number of positive roots (counted according to their multiplicity) of a polynomial $P_n(x)$ with real coefficients is either equal to the number of sign alterations in the sequence of its coefficients or is by an even number less.*

Applying the Descartes theorem to $P_n(-x)$, we obtain a similar theorem for the negative roots of the polynomial $P_n(x)$.

Example 6. Consider the cubic polynomial

$$P_3(x) = x^3 - 3x + 4.$$

Its coefficients have the signs + – +, and therefore we have two alterations of sign. Therefore, the number of positive roots of $P_3(x)$ is equal either to 2 or to 0. Now, consider the polynomial $P_3(-x) = -x^3 + 2x + 1$. The sequence of its coefficients changes sign only once. Therefore, the original equation has one negative root.

2°. *A stronger version of the Descartes theorem.* Suppose that all roots of a polynomial $P_n(x)$ are real*; then the number of positive roots of $P_n(x)$ is equal to the number of sign alterations in the sequence of its coefficients, and the number of its negative roots is equal to the number of sign alterations in the sequence of coefficients of the polynomial $P_n(-x)$.

Example 7. Consider the characteristic polynomial of the symmetric matrix

$$P_3(x) = \begin{vmatrix} -2-x & 1 & 1 \\ 1 & 1-x & 3 \\ 1 & 3 & 1-x \end{vmatrix} = -x^3 + 14x + 20,$$

which has only real roots. The sequence of its coefficients changes sign only once, and therefore it has a single positive root. The number of its negative roots is equal to two, since this polynomial has three nonzero real roots and only one of them can be positive.

3°. If two neighboring coefficients of a polynomial $P_n(x)$ are equal to zero, then the roots of the polynomial cannot be all real (in this case, the stronger version of the Descartes theorem cannot be used).

4°. The number of real roots of a polynomial $P_n(x)$ greater than a fixed c is either equal to the number of sign alterations in the sequence $P_n(c), \ldots, P_n^{(n)}(c)$ or is by an even number less. If all roots of $P_n(x)$ are real, then the number of its roots greater than c coincides with the number of sign alterations in the sequence $P_n(c), \ldots, P_n^{(n)}(c)$.

Example 8. Consider the polynomial

$$P_4(x) = x^4 - 3x^3 + 2x^2 - 2a^2x + a^2.$$

For $x = 1$, we have $P_4(1) = -a^2$, $P_4'(1) = -1 - 2a^2$, $P_4''(1) = -2$, $P_4'''(1) = 6$, $P_4''''(1) = 24$. Thus, there is a single sign alteration, and therefore the polynomial has a single real root greater than unity.

* This is the case, for instance, if we are dealing with the characteristic polynomial of a symmetric matrix.

5°. *Budan–Fourier method.* Let $N(x)$ be the number of sign alterations in the sequence $P_n(x), \dots, P_n^{(n)}(x)$ consisting of the values of the polynomial (5.1.5.2) and its derivatives. Then the number of real roots of equation (5.1.5.1) on the interval $[a, b]$ with $P_n(a) \neq 0$, $P_n(b) \neq 0$ is either equal to $N(a) - N(b)$ or is less by an even number. When calculating $N(a)$, zero terms of the sequence are dropped. When calculating $N(b)$, it may happen that $P_n^{(i)}(b) = 0$ for $k \leq i \leq m$ and $P_n^{(k-1)}(b) \neq 0$, $P_n^{(m+1)}(b) \neq 0$; then $P_n^{(i)}(b)$ should be replaced by $(-1)^{m+1-i}$ sign $P_n^{(m+1)}(b)$.

6°. *Sturm method for finding the number of real roots.* Consider a polynomial $P_n(x)$ with no multiple roots and denote by $N(x)$ the number of sign alterations in the sequence of values of the polynomials (zero terms of the sequence are not taken into account):

$$\begin{aligned} f_0(x) &= g_0(x)f_1(x) - f_2(x), \\ f_1(x) &= g_1(x)f_2(x) - f_3(x), \\ &\dots\dots\dots\dots\dots\dots\dots, \end{aligned}$$

where $f_0(x) = P_n(x)$, $f_1(x) = P_n'(x)$; for $k > 1$, every polynomial $-f_k(x)$ is the residue after dividing the polynomial $f_{k-2}(x)$ by $f_{k-1}(x)$; the last polynomial $f_n(x)$ is a nonzero constant. Then the number of all real roots of equation (5.1.5.1) on the segment $[a, b]$ for $P_n(a) \neq 0$, $P_n(b) \neq 0$ is equal to $N(a) - N(b)$.

Remark 1. Taking $a = -L$ and $b = L$ and passing to the limit as $L \to \infty$, we obtain the overall number of real roots of the algebraic equation.

Example 9. Consider the following cubic equation with the parameter a:

$$P_3(x) = x^3 + 3x^2 - a = 0.$$

The Sturm system for this equation has the form

$$\begin{aligned} P_3(x) &= f_0(x) = x^3 + 3x^2 - a, \\ [P_3(x)]_x' &= f_1(x) = 3x^2 + 6x, \\ f_2(x) &= 2x + a, \\ f_3(x) &= \tfrac{3}{4}a(4 - a). \end{aligned}$$

Case $0 < a < 4$. Let us find the number of sign alterations in the Sturm system for $x = -\infty$ and $x = \infty$:

x	$f_0(x)$	$f_1(x)$	$f_2(x)$	$f_3(x)$	number of sign alterations
$-\infty$	$-$	$+$	$-$	$+$	3
∞	$+$	$+$	$+$	$+$	0

It follows that $N(-\infty) - N(\infty) = 3$. Therefore, for $0 < a < 4$, the given polynomial has three real roots.

Case $a < 0$ *or* $a > 4$. Let us find the number of sign alterations in the Sturm system:

x	$f_0(x)$	$f_1(x)$	$f_2(x)$	$f_3(x)$	number of sign alterations
$-\infty$	$-$	$+$	$-$	$-$	2
∞	$+$	$+$	$+$	$-$	1

It follows that $N(-\infty) - N(\infty) = 1$, and therefore for $a < 0$ or $a > 4$, the given polynomial has one real root.

Remark 2. If equation $P_n(x) = 0$ has multiple roots, then $P_n(x)$ and $P_n'(x)$ have a common divisor and the multiple roots are found by equating to zero this divisor. In this case, $f_n(x)$ is nonconstant and $N(a) - N(b)$ is the number of roots between a and b, each multiple root counted only once.

5.1.5-6. Bounds for complex roots of polynomials with real coefficients.

1°. *Routh–Hurwitz criterion.* For an algebraic equation (5.1.5.1) with real coefficients, the number of roots with positive real parts is equal to the number of sign alterations in any of the two sequences

$$T_0,\quad T_1,\quad T_2/T_1,\quad \ldots,\quad T_n/T_{n-1};$$
$$T_0,\quad T_1,\quad T_1T_2,\quad \ldots,\quad T_{n-2}T_{n-1},\quad a_0;$$

where T_m (it is assumed that $T_m \neq 0$ for all m) are defined by

$$T_0 = a_n > 0, \qquad T_1 = a_{n-1}, \qquad T_2 = \begin{vmatrix} a_{n-1} & a_n \\ a_{n-3} & a_{n-2} \end{vmatrix}, \qquad T_3 = \begin{vmatrix} a_{n-1} & a_n & 0 \\ a_{n-3} & a_{n-2} & a_{n-1} \\ a_{n-5} & a_{n-4} & a_{n-3} \end{vmatrix},$$

$$T_4 = \begin{vmatrix} a_{n-1} & a_n & 0 & 0 \\ a_{n-3} & a_{n-2} & a_{n-1} & a_n \\ a_{n-5} & a_{n-4} & a_{n-3} & a_{n-2} \\ a_{n-7} & a_{n-6} & a_{n-5} & a_{n-4} \end{vmatrix}, \qquad T_5 = \begin{vmatrix} a_{n-1} & a_n & 0 & 0 & 0 \\ a_{n-3} & a_{n-2} & a_{n-1} & a_n & 0 \\ a_{n-5} & a_{n-4} & a_{n-3} & a_{n-2} & a_{n-1} \\ a_{n-7} & a_{n-6} & a_{n-5} & a_{n-4} & a_{n-3} \\ a_{n-9} & a_{n-8} & a_{n-7} & a_{n-6} & a_{n-5} \end{vmatrix}, \quad \ldots$$

2°. All roots of equation (5.1.5.1) have negative real parts if and only if all $T_0, T_1, \ldots, T_n$ are positive.

3°. All roots of an nth-degree equation (5.1.5.1) have negative real parts if and only if this is true for the following $(n-1)$st-degree equation:

$$a_{n-1}x^{n-1} + \left(a_{n-2} - \frac{a_n}{a_{n-1}}\,a_{n-3}\right)x^{n-2} + a_{n-3}x^{n-3} + \left(a_{n-4} - \frac{a_n}{a_{n-1}}\,a_{n-5}\right)x^{n-2} + \cdots = 0.$$

5.2. Matrices and Determinants

5.2.1. Matrices

5.2.1-1. Definition of a matrix. Types of matrices.

A *matrix* of *size* (or *dimension*) $m \times n$ is a rectangular table with *entries* a_{ij} ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$) arranged in m rows and n columns:

$$A \equiv \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Note that, for each entry a_{ij}, the index i refers to the ith row and the index j to the jth column. Matrices are briefly denoted by uppercase letters (for instance, A, as here), or by the symbol $[a_{ij}]$, sometimes with more details: $A \equiv [a_{ij}]$ ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$). The numbers m and n are called the *dimensions* of the matrix. A matrix is said to be *finite* if it has finitely many rows and columns; otherwise, the matrix is said to be *infinite*. In what follows, only finite matrices are considered.

The *null* or *zero matrix* is a matrix whose entries are all equal to zero: $a_{ij} = 0$ ($i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$).

A *column vector* or *column* is a matrix of size $m \times 1$. A *row vector* or *row* is a matrix of size $1 \times n$. Both column and row vectors are often simply called *vectors*.

TABLE 5.2
Types of square matrices ($\bar{a}_{ij}$ is the complex conjugate of a number a_{ij})

Type of square matrix $[a_{ij}]$	Entries
Unit (identity) $I = [\delta_{ij}]$	$a_{ij} = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$ (δ_{ij} is the Kronecker delta)
Diagonal	$a_{ij} = \begin{cases} \text{any}, & i = j, \\ 0, & i \neq j \end{cases}$
Upper triangular (superdiagonal)	$a_{ij} = \begin{cases} \text{any}, & i \leq j, \\ 0, & i > j \end{cases}$
Strictly upper triangular	$a_{ij} = \begin{cases} \text{any}, & i < j, \\ 0, & i \geq j \end{cases}$
Lower triangular (subdiagonal)	$a_{ij} = \begin{cases} \text{any}, & i \geq j, \\ 0, & i < j \end{cases}$
Strictly lower triangular	$a_{ij} = \begin{cases} \text{any}, & i > j, \\ 0, & i \leq j \end{cases}$
Symmetric	$a_{ij} = a_{ji}$ (see also Paragraph 5.2.1-3)
Skew-symmetric (antisymmetric)	$a_{ij} = -a_{ji}$ (see also Paragraph 5.2.1-3)
Hermitian (self-adjoint)	$a_{ij} = \bar{a}_{ji}$ (see also Paragraph 5.2.1-3)
Skew-Hermitian (antihermitian)	$a_{ij} = -\bar{a}_{ji}$ (see also Paragraph 5.2.1-3)
Monomial (generalized permutation)	Each column and each row contain exactly one nonzero entry

A *square matrix* is a matrix of size $n \times n$, and n is called the dimension of this square matrix. The *main diagonal* of a square matrix is its diagonal from the top left corner to the bottom right corner with the entries $a_{11}\ a_{22}\ \ldots\ a_{nn}$. The *secondary diagonal* of a square matrix is the diagonal from the bottom left corner to the top right corner with the entries $a_{n1}\ a_{(n-1)2}\ \ldots\ a_{1n}$. Table 5.2 lists the main types of square matrices (see also Paragraph 5.2.1-3).

5.2.1-2. Basic operations with matrices.

Two matrices are *equal* if they are of the same size and their respective entries are equal.

The *sum* of two matrices $A \equiv [a_{ij}]$ and $B \equiv [b_{ij}]$ of the same size $m \times n$ is the matrix $C \equiv [c_{ij}]$ of size $m \times n$ with the entries

$$c_{ij} = a_{ij} + b_{ij} \quad (i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n).$$

The sum of two matrices is denoted by $C = A + B$, and the operation is called *addition of matrices*.

Properties of addition of matrices:

$$A + O = A \qquad \text{(property of zero)},$$
$$A + B = B + A \qquad \text{(commutativity)},$$
$$(A + B) + C = A + (B + C) \qquad \text{(associativity)},$$

where matrices A, B, C, and zero matrix O have the same size.

The *difference* of two matrices $A \equiv [a_{ij}]$ and $B \equiv [b_{ij}]$ of the same size $m \times n$ is the matrix $C \equiv [c_{ij}]$ of size $m \times n$ with entries

$$c_{ij} = a_{ij} - b_{ij} \quad (i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n).$$

The difference of two matrices is denoted by $C = A - B$, and the operation is called *subtraction of matrices*.

The *product* of a matrix $A \equiv [a_{ij}]$ of size $m \times n$ by a scalar λ is the matrix $C \equiv [c_{ij}]$ of size $m \times n$ with entries

$$c_{ij} = \lambda a_{ij} \quad (i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n).$$

The product of a matrix by a scalar is denoted by $C = \lambda A$, and the operation is called *multiplication of a matrix by a scalar*.

Properties of multiplication of a matrix by a scalar:

$0A = O$	(property of zero),
$(\lambda\mu)A = \lambda(\mu A)$	(associativity with respect to a scalar factor),
$\lambda(A + B) = \lambda A + \lambda B$	(distributivity with respect to addition of matrices),
$(\lambda + \mu)A = \lambda A + \mu A$	(distributivity with respect to addition of scalars),

where λ and μ are scalars, matrices A, B, C, and zero matrix O have the same size.

The *additively inverse (opposite) matrix* for a matrix $A \equiv [a_{ij}]$ of size $m \times n$ is the matrix $C \equiv [c_{ij}]$ of size $m \times n$ with entries

$$c_{ij} = -a_{ij} \quad (i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n),$$

or, in matrix form,

$$C = (-1)A.$$

Remark. The difference C of two matrices A and B can be expressed as $C = A + (-1)B$.

The *product* of a matrix $A \equiv [a_{ij}]$ of size $m \times p$ and a matrix $B \equiv [b_{ij}]$ of size $p \times n$ is the matrix $C \equiv [c_{ij}]$ of size $m \times n$ with entries

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj} \quad (i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n);$$

i.e., the entry c_{ij} in the ith row and jth column of the matrix C is equal to the sum of products of the respective entries in the ith row of A and the jth column of B. Note that the product is defined for matrices of *compatible size*; i.e., the number of the columns in the first matrix should be equal to the number of rows in the second matrix. The product of two matrices A and B is denoted by $C = AB$, and the operation is called *multiplication of matrices*.

Example 1. Consider two matrices

$$A = \begin{pmatrix} 1 & 2 \\ 6 & -3 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 10 & 1 \\ -6 & -0.5 & 20 \end{pmatrix}.$$

The product of the matrix A and the matrix B is the matrix

$$C = AB = \begin{pmatrix} 1 & 2 \\ 6 & -3 \end{pmatrix} \begin{pmatrix} 0 & 10 & 1 \\ -6 & -0.5 & 20 \end{pmatrix}$$
$$= \begin{pmatrix} 1\times 0 + 2\times(-6) & 1\times 10 + 2\times(-0.5) & 1\times 1 + 2\times 20 \\ 6\times 0 + (-3)\times(-6) & 6\times 10 + (-3)\times(-0.5) & 6\times 1 + (-3)\times 20 \end{pmatrix} = \begin{pmatrix} -12 & 9 & 41 \\ 18 & 61.5 & -54 \end{pmatrix}.$$

Two square matrices A and B are said to *commute* if $AB = BA$, i.e., if their multiplication is subject to the commutative law.

Properties of multiplication of matrices:

$AO = O_1, \quad A + O = A$	(property of zero matrix),
$(AB)C = A(BC)$	(associativity of the product of three matrices),
$AI = A$	(multiplication by unit matrix),
$A(B + C) = AB + AC$	(distributivity with respect to a sum of two matrices),
$\lambda(AB) = (\lambda A)B = A(\lambda B)$	(associativity of the product of a scalar and two matrices),
$SD = DS$	(commutativity for any square and any diagonal matrices),

where λ is a scalar, matrices A, B, C, square matrix S, diagonal matrix D, zero matrices O and O_1, and unit matrix I have the compatible sizes.

5.2.1-3. Transpose, complex conjugate matrix, adjoint matrix.

The *transpose* of a matrix $A \equiv [a_{ij}]$ of size $m \times n$ is the matrix $C \equiv [c_{ij}]$ of size $n \times m$ with entries

$$c_{ij} = a_{ji} \quad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m).$$

The transpose is denoted by $C = A^T$.

Example 2. If $A = (a_1, a_2)$ then $A^T = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$.

Properties of transposes:

$$(A + B)^T = A^T + B^T, \quad (\lambda A)^T = \lambda A^T, \quad (A^T)^T = A,$$
$$(AC)^T = C^T A^T, \qquad O^T = O_1, \qquad I^T = I,$$

where λ is a scalar; matrices A, B, and zero matrix O have size $m \times n$; matrix C has size $n \times l$; zero matrix O_1 has size $n \times m$.

A square matrix A is said to be *orthogonal* if $A^T A = AA^T = I$, i.e., $A^T = A^{-1}$ (see Paragraph 5.2.1-6).

Properties of orthogonal matrices:

1. If A is an orthogonal matrix, then A^T is also orthogonal.
2. The product of two orthogonal matrices is an orthogonal matrix.
3. Any symmetric orthogonal matrix is involutive (see Paragraph 5.2.1-7).

The *complex conjugate* of a matrix $A \equiv [a_{ij}]$ of size $m \times n$ is the matrix $C \equiv [c_{ij}]$ of size $m \times n$ with entries

$$c_{ij} = \bar{a}_{ij} \quad (i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n),$$

where $\bar{a}_{ij}$ is the complex conjugate of a_{ij}. The complex conjugate matrix is denoted by $C = \overline{A}$.

The *adjoint matrix* of a matrix $A \equiv [a_{ij}]$ of size $m \times n$ is the matrix $C \equiv [c_{ij}]$ of size $n \times m$ with entries

$$c_{ij} = \bar{a}_{ji} \quad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m).$$

The adjoint matrix is denoted by $C = A^*$.

Properties of adjoint matrices:

$$(A + B)^* = A^* + B^*, \quad (\lambda A)^* = \bar{\lambda} A^*, \quad (A^*)^* = A,$$
$$(AC)^* = C^* A^*, \qquad O^* = O_1, \qquad I^* = I,$$

where λ is a scalar; matrices A, B, and zero matrix O have size $m \times n$; matrix C has size $n \times l$; zero matrix O_1 has a size $n \times m$.

Remark. If a matrix is *real* (i.e., all its entries are real), then the corresponding transpose and the adjoint matrix coincide.

A square matrix A is said to be *normal* if $A^*A = AA^*$. A normal matrix A is said to be *unitary* if $A^*A = AA^* = I$, i.e., $A^* = A^{-1}$ (see Paragraph 5.2.1-6).

5.2.1-4. Trace of a matrix.

The *trace* of a square matrix $A \equiv [a_{ij}]$ of size $n \times n$ is the sum S of its diagonal entries,

$$S = \mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

If λ is a scalar and square matrices A and B has the same size, then

$$\mathrm{Tr}(A + B) = \mathrm{Tr}(A) + \mathrm{Tr}(B), \quad \mathrm{Tr}(\lambda A) = \lambda \mathrm{Tr}(A), \quad \mathrm{Tr}(AB) = \mathrm{Tr}(BA),$$

5.2.1-5. Linear dependence of row vectors (column vectors).

A row vector (column vector) B is a *linear combination* of row vectors (column vectors) $A_1, \ldots, A_k$ if there exist scalars $\alpha_1, \ldots, \alpha_k$ such that

$$B = \alpha_1 A_1 + \cdots + \alpha_k A_k.$$

Row vectors (column vectors) $A_1, \ldots, A_k$ are said to be *linearly dependent* if there exist scalars $\alpha_1, \ldots, \alpha_k$ ($\alpha_1^2 + \cdots + \alpha_k^2 \neq 0$) such that

$$\alpha_1 A_1 + \cdots + \alpha_k A_k = O,$$

where O is the zero row vector (column vector).

Row vectors (column vectors) $A_1, \ldots, A_k$ are said to be *linearly independent* if, for any $\alpha_1, \ldots, \alpha_k$ ($\alpha_1^2 + \cdots + \alpha_k^2 \neq 0$) we have

$$\alpha_1 A_1 + \cdots + \alpha_k A_k \neq O.$$

THEOREM. *Row vectors (column vectors) $A_1, \ldots, A_k$ are linearly dependent if and only if one of them is a linear combination of the others.*

5.2.1-6. Inverse matrices.

Let A be a square matrix of size $n \times n$, and let I be the unit matrix of the same size.

A square matrix B of size $n \times n$ is called a *right inverse* of A if $AB = I$. A square matrix C of size $n \times n$ is called a *left inverse* of A if $CA = I$. If one of the matrices B or C exists, then the other exists, too, and these two matrices coincide. In such a case, the matrix A is said to be *nondegenerate (nonsingular)*.

THEOREM. *A square matrix is nondegenerate if and only if its rows (columns) are linearly independent.*

Remark. Generally, instead of the terms "left inverse matrix" and "right inverse matrix", the term "*inverse matrix*" is used with regard to the matrix $B = A^{-1}$ for a nondegenerate matrix A, since $AB = BA = I$.

UNIQUENESS THEOREM. *The matrix A^{-1} is the unique matrix satisfying the condition $AA^{-1} = A^{-1}A = I$ for a given nondegenerate matrix A.*

Remark. For the existence theorem, see Paragraph 5.2.2-7.

Properties of inverse matrices:

$$(AB)^{-1} = B^{-1}A^{-1}, \quad (\lambda A)^{-1} = \lambda^{-1}A^{-1},$$
$$(A^{-1})^{-1} = A, \quad (A^{-1})^T = (A^T)^{-1}, \quad (A^{-1})^* = (A^*)^{-1},$$

where square matrices A and B are assumed to be nondegenerate and scalar $\lambda \neq 0$.

The problem of finding the inverse matrix is considered in Paragraphs 5.2.2-7, 5.2.4-5, and 5.5.2-3.

5.2.1-7. Powers of matrices.

A product of several matrices equal to one and the same matrix A can be written as a *positive integer power* of the matrix A: $AA = A^2$, $AAA = A^2A = A^3$, etc. For a positive integer k, one defines $A^k = A^{k-1}A$ as the kth power of A. For a nondegenerate matrix A, one defines $A^0 = AA^{-1} = I$, $A^{-k} = (A^{-1})^k$. Powers of a matrix have the following properties:

$$A^pA^q = A^{p+q}, \quad (A^p)^q = A^{pq},$$

where p and q are arbitrary positive integers and A is an arbitrary square matrix; or p and q are arbitrary integers and A is an arbitrary nondegenerate matrix.

There exist matrices A^k whose positive integer power is equal to the zero matrix, even if $A \neq O$. If $A^k = O$ for some integer $k > 1$, then A is called a *nilpotent matrix*.

A matrix A is said to be *involutive* if it coincides with its inverse: $A = A^{-1}$ or $A^2 = I$.

5.2.1-8. Polynomials and matrices. Basic functions of matrices.

A *polynomial with matrix argument* is the expression obtained from a scalar polynomial $f(x)$ by replacing the scalar argument x with a square matrix X:

$$f(X) = a_0I + a_1X + a_2X^2 + \cdots,$$

where a_i ($i = 0, 1, 2, \ldots$) are real or complex coefficients. The polynomial $f(X)$ is a square matrix of the same size as X.

A *polynomial with matrix coefficients* is an expression obtained from a polynomial $f(x)$ by replacing its coefficients a_i ($i = 0, 1, 2, \ldots$) with matrices A_i ($i = 0, 1, 2, \ldots$) of the same size:

$$F(x) = A_0 + A_1x + A_2x^2 + \cdots.$$

Example 3. For the matrix

$$A = \begin{pmatrix} 4 & -8 & 1 \\ 5 & -9 & 1 \\ 4 & -6 & -1 \end{pmatrix},$$

the characteristic matrix (see Paragraph 5.2.3-2) is a polynomial with matrix coefficients and argument λ:

$$F(\lambda) \equiv A - \lambda I = A_0 + A_1\lambda = \begin{pmatrix} 4-\lambda & -8 & 1 \\ 5 & -9-\lambda & 1 \\ 4 & -6 & -1-\lambda \end{pmatrix},$$

where

$$A_0 = A = \begin{pmatrix} 4 & -8 & 1 \\ 5 & -9 & 1 \\ 4 & -6 & -1 \end{pmatrix}, \quad A_1 = -I = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

The corresponding adjugate matrix (see Paragraph 5.2.2-7) can also be represented as a polynomial with matrix coefficients:

$$G(\lambda) = \begin{pmatrix} \lambda^2 + 10\lambda + 15 & -8\lambda - 14 & \lambda + 1 \\ 5\lambda + 9 & \lambda^2 - 3\lambda - 8 & \lambda + 1 \\ 4\lambda + 6 & -6\lambda - 8 & \lambda^2 + 5\lambda + 4 \end{pmatrix} = A_0 + A_1\lambda + A_2\lambda^2,$$

where

$$A_0 = \begin{pmatrix} 15 & -14 & 1 \\ 9 & -8 & 1 \\ 6 & -8 & 4 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 10 & -8 & 1 \\ 5 & -3 & 1 \\ 4 & -6 & 5 \end{pmatrix}, \quad A_2 = I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The variable x in a polynomial with matrix coefficients can be replaced by a matrix X, which yields a polynomial of matrix argument with matrix coefficients. In this situation, one distinguishes between the "left" and the "right" values:

$$F(X) = A_0 + A_1X + A_2X^2 + \cdots,$$
$$\widehat{F}(X) = A_0 + XA_1 + X^2A_2 + \cdots.$$

The exponential function of a square matrix X can be represented as the following convergent series:

$$e^X = 1 + X + \frac{X^2}{2!} + \frac{X^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{X^k}{k!}.$$

The inverse matrix has the form

$$(e^X)^{-1} = e^{-X} = 1 - X + \frac{X^2}{2!} - \frac{X^3}{3!} + \cdots = \sum_{k=0}^{\infty} (-1)^k \frac{X^k}{k!}.$$

Remark. Note that $e^Xe^Y \neq e^Ye^X$, in general. The relation $e^Xe^Y = e^{X+Y}$ holds only for commuting matrices X and Y.

Some other functions of matrices can be expressed in terms of the exponential function:

$$\sin X = \frac{1}{2i}(e^{iX} - e^{-iX}), \quad \cos X = \frac{1}{2}(e^{iX} + e^{-iX}),$$
$$\sinh X = \frac{1}{2}(e^X - e^{-X}), \quad \cosh X = \frac{1}{2}(e^X + e^{-X}).$$

5.2.1-9. Decomposition of matrices.

THEOREM 1. *For any square matrix A, the matrix $S_1 = \frac{1}{2}(A + A^T)$ is symmetric and the matrix $S_2 = \frac{1}{2}(A - A^T)$ is skew-symmetric. The representation of A as the sum of symmetric and skew-symmetric matrices is unique: $A = S_1 + S_2$.*

THEOREM 2. *For any square matrix A, the matrices $H_1 = \frac{1}{2}(A+A^*)$ and $H_2 = \frac{1}{2i}(A-A^*)$ are Hermitian, and the matrix iH_2 is skew-Hermitian. The representation of A as the sum of Hermitian and skew-Hermitian matrices is unique: $A = H_1 + iH_2$.*

THEOREM 3. *For any square matrix A, the matrices AA^* and A^*A are nonnegative Hermitian matrices (see Paragraph 5.7.3-1).*

THEOREM 4. *Any square matrix A admits a polar decomposition*

$$A = QU \quad \text{and} \quad A = U_1Q_1,$$

where Q and Q_1 are nonnegative Hermitian matrices, $Q^2 = AA^*$ and $Q_1^2 = A^*A$, and U and U_1 are unitary matrices. The matrices Q and Q_1 are always unique, while the matrices U and U_1 are unique only in the case of a nondegenerate A.

5.2.1-10. Block matrices.

Let us split a given matrix $A \equiv [a_{ij}]$ ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$) of size $m \times n$ into separate rectangular cells with the help of $(M-1)$ horizontal and $(N-1)$ vertical lines. Each cell is a matrix $A_{\alpha\beta} \equiv [a_{ij}]$ ($i = i_\alpha, i_\alpha + 1, \ldots, i_\alpha + m_\alpha - 1$; $j = j_\beta, j_\beta + 1, \ldots, j_\beta + n_\beta - 1$) of size $m_\alpha \times n_\beta$ and is called a *block* of the matrix A. Here $i_\alpha = m_{\alpha-1} + i_{\alpha-1}$, $j_\beta = n_{\beta-1} + j_{\beta-1}$. Then the given matrix A can be regarded as a new matrix whose entries are the blocks: $A \equiv [A_{\alpha\beta}]$ ($\alpha = 1, 2, \ldots, M$; $\beta = 1, 2, \ldots, N$). This matrix is called a *block matrix*.

Example 4. The matrix

$$A \equiv \left(\begin{array}{ccc|cc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{array}\right)$$

can be regarded as the block matrix

$$A \equiv \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

of size 2×2 with the entries being the blocks

$$A_{11} \equiv \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}, \quad A_{12} \equiv \begin{pmatrix} a_{14} & a_{15} \\ a_{24} & a_{25} \end{pmatrix},$$
$$A_{21} \equiv \begin{pmatrix} a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \\ a_{51} & a_{52} & a_{53} \end{pmatrix}, \quad A_{22} \equiv \begin{pmatrix} a_{34} & a_{35} \\ a_{44} & a_{45} \\ a_{54} & a_{55} \end{pmatrix}$$

of size 2×3, 2×2, 3×3, 3×2, respectively.

Basic operations with block matrices are practically the same as those with common matrices, the role of the entries being played by blocks:

1. For matrices $A \equiv [a_{ij}] \equiv [A_{\alpha\beta}]$ and $B \equiv [b_{ij}] \equiv [B_{\alpha\beta}]$ of the same size and the same block structure, their sum $C \equiv [C_{\alpha\beta}] = [A_{\alpha\beta} + B_{\alpha\beta}]$ is a matrix of the same size and the same block structure.
2. For a matrix $A \equiv [a_{ij}]$ of size $m \times n$ regarded as a block matrix $A \equiv [A_{\alpha\beta}]$ of size $M \times N$, the multiplication by a scalar is defined by $\lambda A = [\lambda A_{\alpha\beta}] = [\lambda a_{ij}]$.
3. Let $A \equiv [a_{ik}] \equiv [A_{\alpha\gamma}]$ and $B \equiv [b_{kj}] \equiv [B_{\gamma\beta}]$ be two block matrices such that the number of columns of each block $A_{\alpha\gamma}$ is equal to the number of the rows of the block $B_{\gamma\beta}$. Then the product of the matrices A and B can be regarded as the block matrix $C \equiv [C_{\alpha\beta}] = [\sum_\gamma A_{\alpha\gamma} B_{\gamma\beta}]$.
4. For a matrix $A \equiv [a_{ij}]$ of size $m \times n$ regarded as a block matrix $A \equiv [A_{\alpha\beta}]$ of size $M \times N$, the transpose has the form $A^T = [A^T_{\beta\alpha}]$.
5. For a matrix $A \equiv [a_{ij}]$ of size $m \times n$ regarded as a block matrix $A \equiv [A_{\alpha\beta}]$ of size $M \times N$, the adjoint matrix has the form $A^* = [A^*_{\beta\alpha}]$.

Let A be a nondegenerate matrix of size $n \times n$ represented as the block matrix

$$A \equiv \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

where A_{11} and A_{22} are square matrices of size $p \times p$ and $q \times q$, respectively ($p + q = n$). Then the following relations, called the *Frobenius formulas*, hold:

$$A^{-1} = \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1} A_{12} N A_{21} A_{11}^{-1} & -A_{11}^{-1} A_{12} N \\ -N A_{21} A_{11}^{-1} & N \end{pmatrix},$$
$$A^{-1} = \begin{pmatrix} K & -K A_{12} A_{22}^{-1} \\ -A_{22}^{-1} A_{21} K & A_{22}^{-1} + A_{22}^{-1} A_{21} K A_{12} A_{22}^{-1} \end{pmatrix}.$$

Here, $N = (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1}$, $K = (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1}$; in the first formula, the matrix A_{11} is assumed nondegenerate, and in the second formula, A_{22} is assumed nondegenerate.

The *direct sum* of two square matrices A and B of size $m \times m$ and $n \times n$, respectively, is the block matrix $C = A \oplus B = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ of size $m + n$.

Properties of the direct sum of matrices:

1. For any square matrices A, B, and C the following relations hold:
$$(A \oplus B) \oplus C = A \oplus (B \oplus C) \quad \text{(associativity)},$$
$$\mathrm{Tr}(A \oplus B) = \mathrm{Tr}(A) + \mathrm{Tr}(B) \quad \text{(trace property)}.$$
2. For nondegenerate square matrices A and B, the following relation holds:
$$(A \oplus B)^{-1} = A^{-1} \oplus B^{-1}.$$
3. For square matrices A_m, B_m of size $m \times m$ and square matrices A_n, B_n of size $n \times n$, the following relations hold:
$$(A_m \oplus A_n) + (B_m \oplus B_n) = (A_m + B_m) \oplus (A_n + B_n);$$
$$(A_m \oplus A_n)(B_m \oplus B_n) = A_m B_m \oplus A_n B_n.$$

5.2.1-11. Kronecker product of matrices.

The *Kronecker product* of two matrices $A \equiv [a_{i_a j_a}]$ and $B \equiv [b_{i_b j_b}]$ of size $m_a \times n_a$ and $m_b \times n_b$, respectively, is the matrix $C \equiv [c_{kh}]$ of size $m_a m_b \times n_a n_b$ with entries

$$c_{kh} = a_{i_a j_a} b_{i_b j_b} \quad (k = 1, 2, \ldots, m_a m_b;\ h = 1, 2, \ldots, n_a n_b),$$

where the index k is the serial number of a pair (i_a, i_b) in the sequence $(1,1)$, $(1,2)$, $\ldots$, $(1, m_b)$, $(2,1)$, $(2,2)$, $\ldots (m_a, m_b)$, and the index h is the serial number of a pair (j_a, j_b) in a similar sequence. This Kronecker product can be represented as the block matrix $C \equiv [a_{i_a j_a} B]$.

Note that if A and B are square matrices and the number of rows in C is equal to the number of rows in A, and the number of rows in D is equal to the number of rows in B, then

$$(A \otimes B)(C \otimes D) = AC \otimes BD.$$

The following relations hold:

$$(A \otimes B)^T = A^T \otimes B^T, \qquad \mathrm{Tr}(A \otimes B) = \mathrm{Tr}(A)\mathrm{Tr}(B).$$

5.2.2. Determinants

5.2.2-1. Notion of determinant.

With each square matrix $A \equiv [a_{ij}]$ of size $n \times n$ one can associate a numerical characteristic, called its *determinant*. The determinant of such a matrix can be defined by induction with respect to the size n.

For a matrix of size 1×1 ($n = 1$), the *first-order determinant* is equal to its only entry,

$$\Delta \equiv \det A = a_{11}.$$

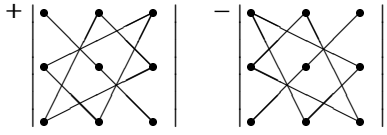
For a matrix of size 2×2 ($n = 2$), the *second-order determinant*, is equal to the difference of the product of its entries on the main diagonal and the product of its entries on the secondary diagonal:

$$\Delta \equiv \det A \equiv \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11} a_{22} - a_{12} a_{21}.$$

For a matrix of size 3×3 ($n = 3$), the *third-order determinant*,

$$\Delta \equiv \det A \equiv \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$
$$= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{21}a_{32}a_{13} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{23}a_{32}a_{11}.$$

This expression is obtained by the *triangle rule* (*Sarrus scheme*), illustrated by the following diagrams, where entries occurring in the same product with a given sign are joined by segments:



For a matrix of size $n \times n$ ($n > 2$), the *nth-order determinant* is defined as follows under the assumption that the $(n-1)$st-order determinant has already been defined for a matrix of size $(n-1) \times (n-1)$.

Consider a matrix $A = [a_{ij}]$ of size $n \times n$. The *minor* M_j^i corresponding to an entry a_{ij} is defined as the $(n-1)$st-order determinant of the matrix of size $(n-1) \times (n-1)$ obtained from the original matrix A by removing the ith row and the jth column (i.e., the row and the column whose intersection contains the entry a_{ij}). The *cofactor* A_j^i *of the entry* a_{ij} is defined by $A_j^i = (-1)^{i+j} M_j^i$ (i.e., it coincides with the corresponding minor if $i + j$ is even, and is the opposite of the minor if $i + j$ is odd).

The *nth-order determinant* of the matrix A is defined by

$$\Delta \equiv \det A \equiv \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \sum_{k=1}^{n} a_{ik} A_k^i = \sum_{k=1}^{n} a_{kj} A_j^k.$$

This formula is also called the *ith row expansion of the determinant of* A and also the *jth column expansion of the determinant of* A.

Example 1. Let us find the third-order determinant of the matrix

$$A = \begin{pmatrix} 1 & -1 & 2 \\ 6 & 1 & 5 \\ 2 & -1 & -4 \end{pmatrix}.$$

To this end, we use the second-column expansion of the determinant:

$$\det A = \sum_{i=1}^{3} (-1)^{i+2} a_{i2} M_2^i = (-1)^{1+2} \times (-1) \times \begin{vmatrix} 6 & 5 \\ 2 & -4 \end{vmatrix} + (-1)^{2+2} \times 1 \times \begin{vmatrix} 1 & 2 \\ 2 & -4 \end{vmatrix} + (-1)^{3+2} \times (-1) \times \begin{vmatrix} 1 & 2 \\ 6 & 5 \end{vmatrix}$$
$$= 1 \times [6 \times (-4) - 5 \times 2] + 1 \times [1 \times (-4) - 2 \times 2] + 1 \times [1 \times 5 - 2 \times 6] = -49.$$

5.2.2-2. Properties of determinants.

Basic properties:

1. Invariance with respect to transposition of matrices:
$$\det A = \det A^T.$$
2. Antisymmetry with respect to the permutation of two rows (or columns): if two rows (columns) of a matrix are interchanged, its determinant preserves its absolute value, but changes its sign.

3. Linearity with respect to a row (or column) of the corresponding matrix: suppose that the ith row of a matrix $A \equiv [a_{ij}]$ is a linear combination of two row vectors, $(a_{i1}, \ldots, a_{i3}) = \lambda(b_1, \ldots, b_n) + \mu(c_1, \ldots, c_n)$; then
$$\det A = \lambda \det A_b + \mu \det A_c,$$
where A_b and A_c are the matrices obtained from A by replacing its ith row with $(b_1, \ldots, b_n)$ and $(c_1, \ldots, c_n)$. This fact, together with the first property, implies that a similar linearity relation holds if a column of the matrix A is a linear combination of two column vectors.

Some useful corollaries from the basic properties:

1. The determinant of a matrix with two equal rows (columns) is equal to zero.
2. If all entries of a row are multiplied by λ, the determinant of the resulting matrix is multiplied by λ.
3. If a matrix contains a row (columns) consisting of zeroes, then its determinant is equal to zero.
4. If a matrix has two proportional rows (columns), its determinant is equal to zero.
5. If a matrix has a row (column) that is a linear combination of its other rows (columns), its determinant is equal to zero.
6. The determinant of a matrix does not change if a linear combination of some of its rows is added to another row of that matrix.

THEOREM (NECESSARY AND SUFFICIENT CONDITION FOR A MATRIX TO BE DEGENERATE). *The determinant of a square matrix is equal to zero if and only if its rows (columns) are linearly dependent.*

5.2.2-3. Minors. Basic minors. Rank and defect of a matrix.

Let $A \equiv [a_{ij}]$ be a matrix of size $n \times n$. Its *mth-order* ($m \le n$) *minor of the first kind, denoted by* $M^{i_1 i_2 \ldots i_m}_{j_1 j_2 \ldots j_m}$, is the mth-order determinant of a submatrix obtained from A by removing some of its $n - m$ rows and $n - m$ columns. Here, i_1, i_2, $\ldots$, i_m are the indices of the rows and j_1, j_2, $\ldots$, j_m are the indices of the columns involved in that submatrix. The *$(n-m)$th-order determinant of the second kind, denoted by* $\overline{M}^{i_1 i_2 \ldots i_m}_{j_1 j_2 \ldots j_m}$, is the $(n-m)$th-order determinant of the submatrix obtained from A by removing the rows and the columns involved in $M^{i_1 i_2 \ldots i_m}_{j_1 j_2 \ldots j_m}$. The *cofactor* of the minor $M^{i_1 i_2 \ldots i_m}_{j_1 j_2 \ldots j_m}$ is defined by $A^{i_1 i_2 \ldots i_m}_{j_1 j_2 \ldots j_m} = (-1)^{i_1+i_2+\cdots+i_m+j_1+j_2+\cdots+j_m} \overline{M}^{i_1 i_2 \ldots i_m}_{j_1 j_2 \ldots j_m}$.

Remark. minors of the first kind can be introduced for any rectangular matrix $A \equiv [a_{ij}]$ of size $m \times n$. Its kth-order ($k \le \min\{m, n\}$) *minor* $M^{i_1 i_2 \ldots i_k}_{j_1 j_2 \ldots j_k}$ is the determinant of the submatrix obtained from A by removing some of its $m - k$ rows and $n - k$ columns.

LAPLACE THEOREM. *Given m rows with indices $i_1, \ldots, i_m$ (or m columns with indices $j_1, \ldots, j_m$) of a square matrix A, its determinant Δ is equal to the sum of products of all mth-order minors $M^{i_1 i_2 \ldots i_m}_{j_1 j_2 \ldots j_m}$ in those rows (resp., columns) and their cofactors $A^{i_1 i_2 \ldots i_m}_{j_1 j_2 \ldots j_m}$, i.e.,*
$$\Delta \equiv \det A = \sum_{j_1, j_2, \ldots, j_m} M^{i_1 i_2 \ldots i_m}_{j_1 j_2 \ldots j_m} A^{i_1 i_2 \ldots i_m}_{j_1 j_2 \ldots j_m} = \sum_{i_1, i_2, \ldots, i_m} M^{i_1 i_2 \ldots i_m}_{j_1 j_2 \ldots j_m} A^{i_1 i_2 \ldots i_m}_{j_1 j_2 \ldots j_m}.$$
Here, in the first sum $i_1, \ldots, i_m$ are fixed, and in the second sum $j_1, \ldots, j_m$ are fixed.

Let $A \equiv [a_{ij}]$ be a matrix of size $m \times n$ with at least one nonzero entry. Then there is a positive integer $r \le n$ for which the following conditions hold:

i) the matrix A has an rth-order nonzero minor, and
ii) any minor of A of order $(r + 1)$ and higher (of it exists) is equal to zero.

The integer r satisfying these two conditions is called the *rank* of the matrix A and is denoted by $r = \operatorname{rank}(A)$. Any nonzero rth-order minor of the matrix A is called its *basic minor*. The rows and the columns whose intersection yields its basic minor are called *basic rows* and *basic columns* of the matrix. The rank of a matrix is equal to the maximal number of its linearly independent rows (columns). This implies that for any matrix, the number of its linearly independent rows is equal to the number of its linearly independent columns.

When calculating the rank of a matrix A, one should pass from submatrices of a smaller size to those of a larger size. If, at some step, one finds a submatrix A_k of size $k \times k$ such that it has a nonzero kth-order determinant and the $(k+1)$st-order determinants of all submatrices of size $(k+1)\times(k+1)$ containing A_k are equal to zero, then it can be concluded that k is the rank of the matrix A.

Properties of the rank of a matrix:

1. For any matrices A and B of the same size the following inequality holds:
$$\operatorname{rank}(A+B) \le \operatorname{rank}(A) + \operatorname{rank}(B).$$
2. For a matrix A of size $m\times n$ and a matrix B of size $n\times k$, the *Sylvester inequality* holds:
$$\operatorname{rank}(A) + \operatorname{rank}(B) - n \le \operatorname{rank}(AB) \le \min\{\operatorname{rank}(A), \operatorname{rank}(B)\}.$$

For a square matrix A of size $n\times n$, the value $d = n - \operatorname{rank}(A)$ is called the *defect* of the matrix A, and A is called a *d-fold degenerate matrix*. The rank of a nondegenerate square matrix $A \equiv [a_{ij}]$ of size $n\times n$ is equal to n.

THEOREM ON BASIC MINOR. *Basic rows (resp., basic columns) of a matrix are linearly independent. Any row (resp., any column) of a matrix is a linear combination of its basic rows (resp., columns).*

5.2.2-4. Expression of the determinant in terms of matrix entries.

1°. Consider a system of mutually distinct $\beta_1, \beta_2, \ldots, \beta_n$, with each β_i taking one of the values $1, 2, \ldots, n$. In this case, the system $\beta_1, \beta_2, \ldots, \beta_n$ is called a *permutation* of the set $1, 2, \ldots, n$. If we interchange two elements in a given permutation $\beta_1, \beta_2, \ldots, \beta_n$, leaving the remaining $n-2$ elements intact, we obtain another permutation, and this transformation of $\beta_1, \beta_2, \ldots, \beta_n$ is called *transposition*. All permutations can be arranged in such an order that the next is obtained from the previous by a single transposition, and one can start from an arbitrary permutation.

Example 2. Let us demonstrate this statement in the case of $n = 3$ (there are $n! = 6$ permutations).

If we start from the permutation 1 2 3, then we can order all permutations, for instance, like this (we underline the numbers to be interchanged):

$$\underline{1}\,\underline{2}\,3 \longrightarrow \underline{2}\,1\,\underline{3} \longrightarrow \underline{3}\,\underline{1}\,2 \longrightarrow \underline{1}\,3\,\underline{2} \longrightarrow \underline{2}\,\underline{3}\,1 \longrightarrow 321.$$

Thus, from any given permutation of n symbols, one can pass to any other permutation by finitely many transpositions.

One says that in a given permutation, the elements β_i and β_j form an *inversion* if $\beta_i > \beta_j$ for $i < j$. The total number of inversions in a permutation $\beta_1, \beta_2, \ldots, \beta_n$ is denoted by $N(\beta_1, \beta_2, \ldots, \beta_n)$. A permutation is said to be *even* if it contains an even number of inversions; otherwise, the permutation is said to be *odd*.

Example 3. The permutation 4 5 1 3 2 ($n = 5$) contains $N(4\ 5\ 1\ 3\ 2) = 7$ inversions and is, therefore, odd. Any of its transposition (for instance, that resulting in the permutation 4 3 1 5 2) yields an even permutation.

The nth-order determinant of a matrix $A \equiv [a_{ij}]$ of size $n\times n$ can be defined as follows:

$$\Delta \equiv \det A = \sum_{\beta_1,\beta_2,\ldots,\beta_n} (-1)^{N(\beta_1,\beta_2,\ldots,\beta_n)} a_{\beta_1 1} a_{\beta_2 2} \ldots a_{\beta_n n},$$

where the sum is over all possible permutations $\beta_1, \beta_2, \ldots, \beta_n$ of the set $1, 2, \ldots, n$.

Example 4. Using the last formula, let us calculate the third-order determinant of the matrix from Example 1. The numbers $\beta_1, \beta_2, \beta_3$ represent permutations of the set 1, 2, 3. We have

$$\begin{aligned}\Delta \equiv \det A &= (-1)^{N(1,2,3)} a_{11}a_{22}a_{33} + (-1)^{N(1,3,2)} a_{11}a_{32}a_{23} + (-1)^{N(2,1,3)} a_{21}a_{12}a_{33} \\ &+ (-1)^{N(2,3,1)} a_{21}a_{32}a_{13} + (-1)^{N(3,1,2)} a_{31}a_{12}a_{23} + (-1)^{N(3,2,1)} a_{31}a_{22}a_{13} \\ &= (-1)^0 \times 1 \times 1 \times (-4) + (-1)^1 \times 1 \times (-1) \times 5 + (-1)^1 \times 6 \times (-1) \times (-4) \\ &+ (-1)^2 \times 6 \times (-1) \times 2 + (-1)^2 \times 2 \times (-1) \times 5 + (-1)^3 \times 2 \times 1 \times 2 = -49,\end{aligned}$$

which coincides with the result of Example 1.

2°. The nth-order determinant can also be defined as follows:

$$\Delta \equiv \det A = \sum_{\beta_1=1}^{n} \sum_{\beta_2=1}^{n} \cdots \sum_{\beta_n=1}^{n} \delta_{\beta_1\beta_2\ldots\beta_n} a_{\beta_1 1} a_{\beta_2 2} \ldots a_{\beta_n n},$$

where $\delta_{\beta_1\beta_2\ldots\beta_n}$ is the *Levi-Civita symbol*:

$$\delta_{\beta_1\beta_2\ldots\beta_n} = \begin{cases} 0, & \text{if some of } \beta_1, \beta_1, \ldots, \beta_n \text{ coincide,} \\ 1, & \text{if } \beta_1, \beta_1, \ldots, \beta_n \text{ form an even permutation,} \\ -1, & \text{if } \beta_1, \beta_1, \ldots, \beta_n \text{ form an odd permutation.} \end{cases}$$

5.2.2-5. Calculation of determinants.

1°. Determinants of specific matrices are often calculated with the help of the formulas for row expansion or column expansion (see Paragraph 5.2.2-1). For this purpose, its is convenient to take rows or columns containing many zero entries.

2°. The determinant of a triangular (upper or lower) and a diagonal matrices is equal to the product of its entries on the main diagonal. In particular, the determinant of the unit matrix is equal to 1.

3°. The determinant of a strictly triangular (upper or lower) matrix is equal to zero.

4°. For block matrices, the following formula can be used:

$$\begin{vmatrix} A & O \\ B & C \end{vmatrix} = \begin{vmatrix} A & B \\ O & C \end{vmatrix} = \det A \det C,$$

where A, B, C are square matrices of size $n \times n$ and O is the zero matrix of size $n \times n$.

5°. The *Vandermonde determinant* is the determinant of the *Vandermonde matrix*:

$$\Delta(x_1, x_2, \ldots, x_n) \equiv \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{vmatrix} = \prod_{1 \le j < i \le n} (x_i - x_j).$$

5.2.2-6. Determinant of a sum and a product of matrices.

The determinant of the product of two matrices A and B of the same size is equal to the product of their determinants,

$$\det(AB) = \det A \det B.$$

The determinant of the direct sum of a matrix A of size $m \times m$ and B of size $n \times n$ is equal to the product of their determinants,

$$\det(A \oplus B) = \det A \det B.$$

The determinant of the direct product of a matrix A of size $m \times m$ and B of size $n \times n$ is calculated by the formula

$$\det(A \otimes B) = (\det A)^n (\det B)^m.$$

5.2.2-7. Relation between the determinant and the inverse matrix.

EXISTENCE THEOREM. *A square matrix A is invertible if and only if its determinant is different from zero.*

Remark. A square matrix A is nondegenerate if its determinant is different from zero.

The *adjugate (classical adjoint)* for a matrix $A \equiv [a_{ij}]$ of size $n \times n$ is a matrix $C \equiv [c_{ij}]$ of size $n \times n$ whose entries coincide with the cofactors of the entries of the transpose A^T, i.e.,

$$c_{ij} = A_{ji} \quad (i, j = 1, 2, \ldots, n). \tag{5.2.2.9}$$

The inverse matrix of a square matrix $A \equiv [a_{ij}]$ of size $n \times n$ is the matrix of size $n \times n$ obtained from the adjugate matrix by dividing all its entries by $\det A$, i.e.,

$$A^{-1} = \begin{pmatrix} \frac{A_{11}}{\det A} & \frac{A_{21}}{\det A} & \cdots & \frac{A_{n1}}{\det A} \\ \frac{A_{12}}{\det A} & \frac{A_{22}}{\det A} & \cdots & \frac{A_{n2}}{\det A} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{A_{1n}}{\det A} & \frac{A_{2n}}{\det A} & \cdots & \frac{A_{nn}}{\det A} \end{pmatrix}. \tag{5.2.2.10}$$

JACOBI THEOREM. *For minors of the matrix of cofactors of a matrix A, the following relations hold:*

$$\begin{vmatrix} A^{i_1}_{j_1} & A^{i_1}_{j_2} & \cdots & A^{i_1}_{j_k} \\ A^{i_2}_{j_1} & A^{i_2}_{j_2} & \cdots & A^{i_2}_{j_k} \\ \vdots & \vdots & \ddots & \vdots \\ A^{i_k}_{j_1} & A^{i_k}_{j_2} & \cdots & A^{i_k}_{j_k} \end{vmatrix} = (\det A)^{k-1} A^{i_1 i_2 \ldots i_k}_{j_1 j_2 \ldots j_k}.$$

5.2.3. Equivalent Matrices. Eigenvalues

5.2.3-1. Equivalence transformation.

Matrices A and $\widetilde{A}$ of size $m \times n$ are said to be *equivalent* if there exist nondegenerate matrices S and T of size $m \times m$ and $n \times n$, respectively, such that A and $\widetilde{A}$ are related by the *equivalence transformation*

$$\widetilde{A} = SAT.$$

THEOREM. *Two matrices of the same size are equivalent if and only if they are of the same rank.*

Remark 1. One of the square matrices S and T may coincide with the unit matrix. Thus, we have equivalent matrices A and B if there is a nondegenerate square matrix S such that $\widetilde{A} = SA$ or $\widetilde{A} = AS$.

Remark 2. *Triangular decomposition* of a matrix A corresponds to its equivalence transformation with $\widetilde{A} \equiv I$, so that $A = S^{-1}T^{-1} = LU$, where $L = S^{-1}$ and $P = T^{-1}$ are an upper and lower triangular matrix. This representation is also called the *LU-decomposition.*

Any equivalence transformation can be reduced to a sequence of *elementary transformations* of the following types:

1. Interchange of two rows (columns).
2. Multiplication of a row (column) by a nonzero scalar.
3. Addition to some row (column) of another row (column) multiplied by a scalar.

These elementary transformations are accomplished with the help of *elementary matrices* obtained from the unit matrix by the corresponding operations with its rows (columns). With the help of elementary transformations, an arbitrary matrix A of rank $r > 0$ can be reduced to *normal (canonical) form*, which has a block structure with the unit matrix I of size $r \times r$ in the top left corner.

Example 1. The LU-decomposition of a matrix

$$A = \begin{pmatrix} 2 & 1 & 4 \\ 3 & 2 & 1 \\ 1 & 3 & 3 \end{pmatrix}$$

can be obtained with the help of the following sequence of elementary transformations:

$$\overbrace{\begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}^{S_1}\begin{pmatrix} 2 & 1 & 4 \\ 3 & 2 & 1 \\ 1 & 3 & 3 \end{pmatrix} \to \overbrace{\begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}^{S_2}\begin{pmatrix} 1 & 1/2 & 2 \\ 3 & 2 & 1 \\ 1 & 3 & 3 \end{pmatrix} \to \overbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}}^{S_3}\begin{pmatrix} 1 & 1/2 & 2 \\ 0 & 1/2 & -5 \\ 1 & 3 & 3 \end{pmatrix} \to$$

$$\to \begin{pmatrix} 1 & 1/2 & 2 \\ 0 & 1/2 & -5 \\ 0 & 5/2 & 1 \end{pmatrix}\overbrace{\begin{pmatrix} 1 & -1/2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}^{T_1} \to \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1/2 & -5 \\ 0 & 5/2 & 1 \end{pmatrix}\overbrace{\begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}^{T_2} \to$$

$$\to \overbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}}^{S_4}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & -5 \\ 0 & 5/2 & 1 \end{pmatrix} \to \overbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -5/2 & 1 \end{pmatrix}}^{S_5}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -10 \\ 0 & 5/2 & 1 \end{pmatrix} \to$$

$$\to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -10 \\ 0 & 0 & 26 \end{pmatrix}\overbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 10 \\ 0 & 0 & 1 \end{pmatrix}}^{T_3} \to \overbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/26 \end{pmatrix}}^{S_6}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 26 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

These transformations amount to the equivalence transformation $I = SAT$, where $T = T_1T_2T_3$:

$$S = S_6S_5S_4S_3S_2S_1 = \begin{pmatrix} 1/2 & 0 & 0 \\ -3 & 2 & 0 \\ 7/26 & -5/26 & 1/26 \end{pmatrix} \quad \text{and} \quad T = T_1T_2T_3 = \begin{pmatrix} 1 & -1/2 & -7 \\ 0 & 1 & 10 \\ 0 & 0 & 1 \end{pmatrix}.$$

Hence, we obtain

$$L = S^{-1} = \begin{pmatrix} 2 & 0 & 0 \\ 3 & 1/2 & 0 \\ 1 & 5/2 & 26 \end{pmatrix} \quad \text{and} \quad U = T^{-1} = \begin{pmatrix} 1 & 1/2 & 2 \\ 0 & 1 & -10 \\ 0 & 0 & 1 \end{pmatrix}.$$

5.2.3-2. Similarity transformation.

Two square matrices A and $\widetilde{A}$ of the same size are said to be *similar* if there exists a square nondegenerate matrix S of the same size, the so-called *transforming matrix*, such that A and $\widetilde{A}$ are related by the *similarity transformation*

$$\widetilde{A} = S^{-1}AS \quad \text{or} \quad A = S\widetilde{A}S^{-1}.$$

Properties of similar matrices:

1. If A and B are square matrices of the same size and $C = A + B$, then

$$\widetilde{C} = \widetilde{A} + \widetilde{B} \quad \text{or} \quad S^{-1}(A + B)S = S^{-1}AS + S^{-1}BS.$$

2. If A and B are square matrices of the same size and $C = AB$, then
$$\widetilde{C} = \widetilde{A}\widetilde{B} \quad \text{or} \quad S^{-1}(AB)S = (S^{-1}AS)(S^{-1}BS).$$
3. If A is a square matrix and $C = \lambda A$ where λ is a scalar, then
$$\widetilde{C} = \lambda\widetilde{B} \quad \text{or} \quad S^{-1}(\lambda B)T = \lambda S^{-1}BS.$$
4. Two similar matrices have the same rank, the same trace, and the same determinant.

Under some additional conditions, there exists a similarity transformation that turns a square matrix A into a diagonal matrix with the eigenvalues of A (see Paragraph 5.2.3-5) on the main diagonal. There are three cases in which a matrix can be reduced to diagonal form:
1. All eigenvalues of A are mutually distinct (see Paragraph 5.2.3-5).
2. The defects of the matrices $A-\lambda_i I$ are equal to the multiplicities m_i' of the corresponding eigenvalues λ_i (see Paragraph 5.2.3-6). In this case, one says that the matrix has a *simple structure*.
3. Symmetric matrices.

For a matrix of general structure, one can only find a similarity transformation that reduces the matrix to the so-called *quasidiagonal canonical form* or the *canonical Jordan form* with a quasidiagonal structure. The main diagonal of the latter matrix consists of the eigenvalues of A, each repeated according to its multiplicity. The entries just above the main diagonal are equal either to 1 or 0. The other entries of the matrix are all equal to zero. The matrix in canonical Jordan form is a diagonal block matrix whose blocks form its main diagonal, each block being either a diagonal matrix or a so-called *Jordan cell* of the form
$$\Lambda_k \equiv \begin{pmatrix} \lambda_k & 1 & 0 & \cdots & 0 \\ 0 & \lambda_k & 1 & \cdots & 0 \\ 0 & 0 & \lambda_k & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_k \end{pmatrix}.$$

5.2.3-3. Congruent and orthogonal transformations.

Square matrices A and $\widetilde{A}$ of the same size are said to be *congruent* if there is a nondegenerate square matrix S such that A and $\widetilde{A}$ are related by the so-called *congruent* or *congruence transformation*
$$\widetilde{A} = S^T AS \quad \text{or} \quad A = S\widetilde{A}S^T.$$

This transformation is characterized by the fact that it preserves the symmetry of the original matrix.

For any symmetric matrix A of rank r there is a congruent transformation that reduces A to a canonical form which is a diagonal matrix of the form,
$$\widetilde{A} = S^T AS = \begin{pmatrix} I_p & & \\ & -I_{r-p} & \\ & & O \end{pmatrix},$$
where I_p and I_{r-p} are unit matrices of size $p\times p$ and $(r-p)\times(r-p)$. The number p is called the *index* of the matrix A, and $s = p-(r-p) = 2p-r$ is called its *signature*.

THEOREM. *Two symmetric matrices are congruent if they are of the same rank and have the same index (or signature).*

A similarity transformation defined by an orthogonal matrix S (i.e., $S^T = S^{-1}$) is said to be *orthogonal*. In this case
$$\widetilde{A} = S^{-1}AS = S^T AS.$$

Example 2. Consider a three-dimensional orthogonal coordinate system with the axes OX_1, OX_2, OX_3 and a new coordinate system obtained from this one by its rotation by the angle φ around the axis OX_3, i.e.,

$$\widetilde{x}_1 = x_1\cos\varphi - x_2\sin\varphi, \quad \widetilde{x}_2 = x_1\sin\varphi + x_2\cos\varphi, \quad \widetilde{x}_3 = x_3.$$

The matrix of this coordinate transformation has the form

$$S_3 = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Rotations of the given coordinate system by the angles ψ and θ around the axes OX_1 and OX_2, respectively, correspond to the matrices

$$S_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & -\sin\psi \\ 0 & \sin\psi & \cos\psi \end{pmatrix}, \quad S_2 = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}.$$

The matrices S_1, S_2, S_3 are orthogonal ($S_j^{-1} = S_j^T$).

The transformation that consists of simultaneous rotations around of the coordinate axes by the angles ψ, θ, φ is defined by the matrix

$$S = S_3S_2S_1.$$

5.2.3-4. Conjunctive and unitary transformations.

1°. Square matrices A and $\widetilde{A}$ of the same size are said to be *conjunctive* if there is a nondegenerate matrix S such that A and $\widetilde{A}$ are related by the *conjunctive transformation*

$$\widetilde{A} = S^*AS \quad \text{or} \quad A = S\widetilde{A}S^*,$$

where S^* is the adjoint of S.

2°. A similarity transformation of a matrix A is said to be *unitary* if it is defined by a unitary matrix S (i.e., $S^* = S^{-1}$). In this case,

$$\widetilde{A} = S^{-1}AS = S^*AS.$$

Some basic properties of the above matrix transformations are listed in Table 5.3.

TABLE 5.3
Matrix transformations

Transformation	$\widetilde{A}$	Invariants
Equivalence	SAT	Rank
Similarity	$S^{-1}AS$	Rank, determinant, eigenvalues
Congruent	S^TAS	Rank and symmetry
Orthogonal	$S^{-1}AS = S^TAS$	Rank, determinant, eigenvalues, and symmetry
Conjunctive	S^*AS	Rank and self-adjointness
Unitary	$S^{-1}AS = S^*AS$	Rank, determinant, eigenvalues, and self-adjointness

5.2.3-5. Eigenvalues and spectra of square matrices.

An *eigenvalue* of a square matrix A is any real or complex λ for which the matrix $F(\lambda) \equiv A - \lambda I$ is degenerate. The set of all eigenvalues of a matrix A is called its *spectrum*, and $F(\lambda)$ is called its *characteristic matrix*. The inverse of an eigenvalue, $\mu = 1/\lambda$, is called a *characteristic value*.

Properties of spectrum of a matrices:

1. Similar matrices have the same spectrum.
2. If λ is an eigenvalue of a normal matrix A (see Paragraph 5.2.1-3), then $\bar{\lambda}$ is an eigenvalue of the matrix A^*; $\operatorname{Re}\lambda$ is an eigenvalue of the matrix $H_1 = \frac{1}{2}(A + A^*)$; and $\operatorname{Im}\lambda$ is an eigenvalue of the matrix $H_2 = \frac{1}{2i}(A - A^*)$.
3. All eigenvalues of a normal matrix are real if and only if this matrix is similar to a Hermitian matrix.
4. All eigenvalues of a unitary matrix have absolute values equal to 1.
5. A square matrix is nondegenerate if and only if all its eigenvalues are different from zero.

A nonzero (column) vector X (see Paragraphs 5.2.1-1 and 5.2.1-2) satisfying the condition

$$AX = \lambda X$$

is called an *eigenvector* of the matrix A corresponding to the eigenvalue λ. Eigenvectors corresponding to distinct eigenvalues of A are linearly independent.

5.2.3-6. Reduction of a square matrix to triangular form.

THEOREM. *For any square matrix A there exists a similarity transformation $\widetilde{A} = S^{-1}AS$ such that $\widetilde{A}$ is a triangular matrix.*

The diagonal entries of any triangular matrix similar to a square matrix A of size $n \times n$ coincide with the eigenvalues of A; each eigenvalue λ_i of A occurs $m_i' \geq 1$ times on the diagonal. The positive integer m_i' is called the *algebraic multiplicity* of the eigenvalue λ_i. Note that $\sum\limits_i m_i' = n$.

The trace $\operatorname{Tr}(A)$ is equal to the sum of all eigenvalues of A, each eigenvalue counted according to its multiplicity, i.e.,

$$\operatorname{Tr}(A) = \sum_i m_i' \lambda_i.$$

The determinant $\det A$ is equal to the product of all eigenvalues of A, each eigenvalue counted according to its multiplicity, i.e.,

$$\det A = \prod_i \lambda_i^{m_i'}.$$

5.2.3-7. Reduction of a square matrix to diagonal form.

THEOREM 1. *If A is a square matrix similar to some normal matrix, then there is a similarity transformation $\widetilde{A} = S^{-1}AS$ such that the matrix $\widetilde{A}$ is diagonal.*

THEOREM 2. *Two Hermitian matrices A and B can be reduced to diagonal form by the same similarity transformation if and only if $AB = BA$.*

THEOREM 3. *For any Hermitian matrix A, there is a nondegenerate matrix S such that $\widetilde{A} = S^*AS$ is a diagonal matrix. The entries of $\widetilde{A}$ are real.*

THEOREM 4. *For any real symmetric matrix A, there is a real nondegenerate matrix T such that $\widetilde{A} = S^TAS$ is a diagonal matrix.*

Example 3. Consider the real symmetric matrix

$$A = \begin{pmatrix} 11 & -6 & 2 \\ -6 & 10 & -4 \\ 2 & -4 & 6 \end{pmatrix}.$$

Its eigenvalues are $\lambda_1 = 18$, $\lambda_2 = 6$, $\lambda_3 = 3$ and the respective eigenvectors are

$$X_1 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}.$$

Consider the matrix S with the columns X_1, X_2, and X_3:

$$S = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{pmatrix}.$$

Taking $\widetilde{A}_1 = S^T AS$, we obtain a diagonal matrix:

$$\widetilde{A}_1 = S^T AS = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{pmatrix} \begin{pmatrix} 11 & -6 & 2 \\ -6 & 10 & -4 \\ 2 & -4 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{pmatrix} = \begin{pmatrix} 27 & 0 & 0 \\ 0 & 54 & 0 \\ 0 & 0 & 162 \end{pmatrix}.$$

Taking $\widetilde{A}_2 = S^{-1} AS$, we obtain a diagonal matrix with the eigenvalues on the main diagonal:

$$\widetilde{A}_2 = S^{-1} AS = -\frac{1}{27} \begin{pmatrix} -3 & -6 & -6 \\ -6 & -3 & 6 \\ -6 & 6 & -3 \end{pmatrix} \begin{pmatrix} 11 & -6 & 2 \\ -6 & 10 & -4 \\ 2 & -4 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 18 \end{pmatrix}.$$

We note that $\widetilde{A}_1 = 9\widetilde{A}_2$.

5.2.3-8. Characteristic equation of a matrix.

The algebraic equation of degree n

$$f_A(\lambda) \equiv \det(A - \lambda I) \equiv \det\,[a_{ij} - \lambda\delta_{ij}] \equiv \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} = 0$$

is called the *characteristic equation* of the matrix A of size $n \times n$, and $f_A(\lambda)$ is called its *characteristic polynomial*. The spectrum of the matrix A (i.e., the set of all its eigenvalues) coincides with the set of all roots of its characteristic equation. The multiplicity of every root λ_i of the characteristic equation is equal to the multiplicity m_i' of the eigenvalue λ_i.

Example 4. The characteristic equation of the matrix

$$A = \begin{pmatrix} 4 & -8 & 1 \\ 5 & -9 & 1 \\ 4 & -6 & -1 \end{pmatrix}$$

has the form

$$f_A(\lambda) \equiv \det \begin{pmatrix} 4 - \lambda & -8 & 1 \\ 5 & -9 - \lambda & 1 \\ 4 & -6 & -1 - \lambda \end{pmatrix} = -\lambda^3 - 6\lambda^2 - 11\lambda - 6 = -(\lambda + 1)(\lambda + 2)(\lambda + 3).$$

Similar matrices have the same characteristic equation.

Let λ_j be an eigenvalue of a square matrix A. Then

1) $\alpha\lambda_j$ is an eigenvalue of the matrix αA for any scalar α;
2) λ_j^p is an eigenvalue of the matrix A^p ($p = 0, \pm 1, \ldots, \pm N$ for a nondegenerate A; otherwise, $p = 0, 1, \ldots, N$), where N is a natural number;
3) a polynomial $f(A)$ of the matrix A has the eigenvalue $f(\lambda)$.

Suppose that the spectra of matrices A and B consist of eigenvalues λ_j and μ_k, respectively. Then the spectrum of the Kronecker product $A \otimes B$ is the set of all products $\lambda_j \mu_k$. The spectrum of the direct sum of matrices $A = A_1 \oplus \ldots \oplus A_n$ is the union of the spectra of the matrices $A_1, \ldots, A_n$. The algebraic multiplicities of the same eigenvalues of matrices $A_1, \ldots, A_n$ are summed.

Regarding bounds for eigenvalues see Paragraph 5.6.3-4.

5.2.3-9. Cayley–Hamilton theorem. Sylvester theorem.

CAYLEY–HAMILTON THEOREM. *Each square matrix A satisfies its own characteristic equation; i.e., $f_A(A) = 0$.*

Example 5. Let us illustrate the Cayley–Hamilton theorem by the matrix in Example 4:

$$f_A(A) = -A^3 - 6A^2 - 11A - 6I$$
$$= -\begin{pmatrix} 70 & -116 & 19 \\ 71 & -117 & 19 \\ 64 & -102 & 11 \end{pmatrix} - 6\begin{pmatrix} -20 & 34 & -5 \\ -21 & 35 & -5 \\ -18 & 28 & -1 \end{pmatrix} - 11\begin{pmatrix} 4 & -8 & 1 \\ 5 & -9 & 1 \\ 4 & -6 & -1 \end{pmatrix} - 6\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = 0.$$

A scalar polynomial $p(\lambda)$ is called an *annihilating polynomial* of a square matrix A if $p(A) = 0$. For example, the characteristic polynomial $f_A(\lambda)$ is an annihilating polynomial of A. The unique monic annihilating polynomial of least degree is called the *minimal polynomial* of A and is denoted by $\psi(\lambda)$. The minimal polynomial is a divisor of every annihilating polynomial.

By dividing an arbitrary polynomial $f(\lambda)$ of degree n by an annihilating polynomial $p(\lambda)$ of degree m ($p(\lambda) \neq 0$), one obtains the representation

$$f(\lambda) = p(\lambda)q(\lambda) + r(\lambda),$$

where $q(\lambda)$ is a polynomial of degree $n - m$ (if $m \leq n$) or $q(\lambda) = 0$ (if $m > n$) and $r(\lambda)$ is a polynomial of degree $l < m$ or $r(\lambda) = 0$. Hence

$$f(A) = p(A)q(A) + r(A),$$

where $p(A) = 0$ and $f(A) = r(A)$. The polynomial $r(\lambda)$ in this representation is called the *interpolation polynomial* of A.

Example 6. Let

$$f(A) = A^4 + 4A^3 + 2A^2 - 12A - 10I,$$

where the matrix A is defined in Example 4. Dividing $f(\lambda)$ by the characteristic polynomial $f_A(\lambda) = -\lambda^3 - 6\lambda^2 - 11\lambda - 6$, we obtain the remainder $r(\lambda) = 3\lambda^2 + 4\lambda + 2$. Consequently,

$$f(A) = r(A) = 3A^2 + 4A + 2I.$$

THEOREM. *Every analytic function of a square $n \times n$ matrix A can be represented as a polynomial of the same matrix,*

$$f(A) = \frac{1}{\Delta(\lambda_1, \lambda_2, \ldots, \lambda_n)} \sum_{k=1}^{n} \Delta_{n-k} A^{n-k},$$

where $\Delta(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is the Vandermonde determinant and Δ_i is obtained from Δ by replacing the $(i+1)$st row by $(f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_n))$.

Example 7. Let us find $r(A)$ by this formula for the polynomial in Example 6.

We find the eigenvalues of A from the characteristic equation $f_A(\lambda) = 0$: $\lambda_1 = -1$, $\lambda_2 = -2$, and $\lambda_3 = -3$. Then the Vandermonde determinant is equal to $\Delta(\lambda_1, \lambda_2, \lambda_3) = -2$, and the other determinants are $\Delta_1 = -4$, $\Delta_2 = -8$, and $\Delta_3 = -6$. It follows that

$$f(A) = \frac{1}{-2}[(-6)A^2 + (-8)A + (-4)I] = 3A^2 + 4A + 2I.$$

The Cayley–Hamilton theorem can also be used to find the powers and the inverse of a matrix A (since if $f_A(A) = 0$, then $A^k f_A(A) = 0$ for any positive integer k).

Example 8. For the matrix in Examples 4–7, one has

$$f_A(A) = -A^3 - 6A^2 - 11A - 6I = 0.$$

Hence we obtain

$$A^3 = -6A^2 - 11A - 6I.$$

By multiplying this expression by A, we obtain

$$A^4 = -6A^3 - 11A^2 - 6A.$$

Now we use the representation of the cube of A via lower powers of A and eventually arrive at the formula

$$A^4 = 25A^2 + 60A + 36I.$$

For the inverse matrix, by analogy with the preceding, we obtain

$$A^{-1} f_A(A) = A^{-1}(-A^3 - 6A^2 - 11A - 6I) = -A^2 - 6A - 11I - 6A^{-1} = 0.$$

The definitive result is

$$A^{-1} = -\frac{1}{6}(A^2 + 6A + 11I).$$

In some cases, an analytic function of a matrix A can be computed by a formula in the following theorem.

SYLVESTER'S THEOREM. *If all eigenvalues of a matrix A are distinct, then*

$$f(A) = \sum_{k=1}^{n} f(\lambda_k) Z_k, \quad Z_k = \frac{\prod_{i \neq k}(A - \lambda_i I)}{\prod_{i \neq k}(\lambda_k - \lambda_i)},$$

and, moreover, $Z_k = Z_k^m$ ($m = 1, 2, 3, \ldots$).

5.3. Linear Spaces

5.3.1. Concept of a Linear Space. Its Basis and Dimension

5.3.1-1. Definition of a linear space.

A *linear space* or a *vector space* over a field of *scalars* (usually, the field of real numbers or the field of complex numbers) is a set $\mathcal{V}$ of elements $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, ... (also called *vectors*) of any nature for which the following conditions hold:

I. There is a rule that establishes correspondence between any pair of elements $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ and a third element $\mathbf{z} \in \mathcal{V}$, called the *sum* of the elements $\mathbf{x}$, $\mathbf{y}$ and denoted by $\mathbf{z} = \mathbf{x} + \mathbf{y}$.

II. There is a rule that establishes correspondence between any pair $\mathbf{x}$, λ, where $\mathbf{x}$ is an element of $\mathcal{V}$ and λ is a scalar, and an element $\mathbf{u} \in \mathcal{V}$, called the *product of a scalar λ and a vector* $\mathbf{x}$ and denoted by $\mathbf{u} = \lambda\mathbf{x}$.

III. The following eight axioms are assumed for the above two operations:

1. Commutativity of the sum: $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
2. Associativity of the sum: $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.
3. There is a zero element 0 such that $\mathbf{x} + 0 = \mathbf{x}$ for any $\mathbf{x}$.
4. For any element $\mathbf{x}$ there is an opposite element $\mathbf{x}'$ such that $\mathbf{x} + \mathbf{x}' = 0$.
5. A special role of the unit scalar 1: $1 \cdot \mathbf{x} = \mathbf{x}$ for any element $\mathbf{x}$.
6. Associativity of the multiplication by scalars: $\lambda(\mu\mathbf{x}) = (\lambda\mu)\mathbf{x}$.
7. Distributivity with respect to the addition of scalars: $(\lambda + \mu)\mathbf{x} = \lambda\mathbf{x} + \mu\mathbf{x}$.
8. Distributivity with respect to a sum of vectors: $\lambda(\mathbf{x} + \mathbf{y}) = \lambda\mathbf{x} + \lambda\mathbf{y}$.

This is the definition of an abstract linear space. We obtain a *specific linear space* if the nature of the elements and the operations of addition and multiplication by scalars are concretized.

Example 1. Consider the set of all free vectors in three-dimensional space. If addition of these vectors and their multiplication by scalars are defined as in analytic geometry (see Paragraph 4.5.1-1), this set becomes a linear space denoted by B_3.

Example 2. Consider the set $\{x\}$ whose elements are all positive real numbers. Let us define the sum of two elements x and y as the product of x and y, and define the product of a real scalar λ and an element x as the λth power of the positive real x. The number 1 is taken as the zero element of the space $\{x\}$, and the opposite of x is taken equal to $1/x$. It is easy to see that the set $\{x\}$ with these operations of addition and multiplication by scalars is a linear space.

Example 3. Consider the *n-dimensional coordinate space* $\mathbb{R}^n$, whose elements are ordered sets of n arbitrary real numbers $(x_1, \ldots, x_n)$. The generic element of this space is denoted by $\mathbf{x}$, i.e., $\mathbf{x} = (x_1, \ldots, x_n)$, and the reals $x_1, \ldots, x_n$ are called the *coordinates* of the element $\mathbf{x}$. From the algebraic standpoint, the set $\mathbb{R}^n$ may be regarded as the set of all row vectors with n real components.

The operations of addition of element of $\mathbb{R}^n$ and their multiplication by scalars are defined by the following rules:

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n),$$
$$\lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n).$$

Remark. If the field of scalars λ, μ, ... in the above definition is the field of all real numbers, the corresponding linear spaces are called *real linear spaces*. If the field of scalars is that of all complex numbers, the corresponding space is called a *complex linear space*. In many situations, it is clear from the context which field of scalars is meant.

The above axioms imply the following properties of an arbitrary linear space:

1. The zero vector is unique, and for any element $\mathbf{x}$ the opposite element is unique.
2. The zero vector 0 is equal to the product of any element $\mathbf{x}$ by the scalar 0.
3. For any element $\mathbf{x}$, the opposite element is equal to the product of $\mathbf{x}$ by the scalar –1.
4. The *difference* of two elements $\mathbf{x}$ and $\mathbf{y}$, i.e., the element $\mathbf{z}$ such that $\mathbf{z} + \mathbf{y} = \mathbf{x}$, is unique.

5.3.1-2. Basis and dimension of a linear space. Isomorphisms of linear spaces.

An element $\mathbf{y}$ is called a *linear combination* of elements $\mathbf{x}_1, \ldots, \mathbf{x}_k$ of a linear space $\mathcal{V}$ if there exist scalars $\alpha_1, \ldots, \alpha_k$ such that

$$\mathbf{y} = \alpha_1 \mathbf{x}_1 + \cdots + \alpha_k \mathbf{x}_k.$$

Elements $\mathbf{x}_1, \ldots, \mathbf{x}_k$ of the space $\mathcal{V}$ are said to be *linearly dependent* if there exist scalars $\alpha_1, \ldots, \alpha_k$ such that $|\alpha_1|^2 + \cdots + |\alpha_k|^2 \neq 0$ and

$$\alpha_1 \mathbf{x}_1 + \cdots + \alpha_k \mathbf{x}_k = 0,$$

where 0 is the zero element of $\mathcal{V}$.

Elements $\mathbf{x}_1, \ldots, \mathbf{x}_k$ of the space $\mathcal{V}$ are said to be *linearly independent* if for any scalars $\alpha_1, \ldots, \alpha_k$ such that $|\alpha_1|^2 + \cdots + |\alpha_k|^2 \neq 0$, we have

$$\alpha_1 \mathbf{x}_1 + \cdots + \alpha_k \mathbf{x}_k \neq 0.$$

THEOREM. *Elements $\mathbf{x}_1, \ldots, \mathbf{x}_k$ of a linear space $\mathcal{V}$ are linearly dependent if and only if one of them is a linear combination of the others.*

Remark. If at least one of the elements $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is equal to zero, then these elements are linearly dependent. If some of the elements $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly dependent, then all these elements are linearly dependent.

Example 4. The elements $\mathbf{i}_1 = (1, 0, \ldots, 0)$, $\mathbf{i}_2 = (0, 1, \ldots, 0)$, ..., $\mathbf{i}_n = (0, 0, \ldots, 1)$ of the space $\mathbb{R}^n$ (see Example 3) are linearly independent. For any $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, the vectors $\mathbf{x}, \mathbf{i}_1, \ldots, \mathbf{i}_n$ are linearly dependent.

A *basis* of a linear space $\mathcal{V}$ is defined as any system of linearly independent vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ such that for any element $\mathbf{x}$ of the space $\mathcal{V}$ there exist scalars $x_1, \ldots, x_n$ such that

$$\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n.$$

This relation is called the *representation of an element* $\mathbf{x}$ *in terms of the basis* $\mathbf{e}_1, \ldots, \mathbf{e}_n$, and the scalars $x_1, \ldots, x_n$ are called the *coordinates* of the element $\mathbf{x}$ in that basis.

UNIQUENESS THEOREM. *The representation of any element* $\mathbf{x} \in \mathcal{V}$ *in terms of a given basis* $\mathbf{e}_1, \ldots, \mathbf{e}_n$ *is unique.*

Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be any basis in $\mathcal{V}$ and vectors $\mathbf{x}$ and $\mathbf{y}$ have the coordinates $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ in that basis. Then the coordinates of the vector $\mathbf{x} + \mathbf{y}$ in that basis are $x_1 + y_1, \ldots, x_n + y_n$, and the coordinates of the vector $\lambda\mathbf{x}$ are $\lambda x_1, \ldots, \lambda x_n$ for any scalar λ.

Example 5. Any three noncoplanar vectors form a basis in the linear space B_3 of all free vectors. The n elements $\mathbf{i}_1 = (1, 0, \ldots, 0)$, $\mathbf{i}_2 = (0, 1, \ldots, 0)$, $\ldots$, $\mathbf{i}_n = (0, 0, \ldots, 1)$ form a basis in the linear space $\mathbb{R}^n$. Any basis of the linear space $\{x\}$ from Example 2 consists of a single element. This element can be arbitrarily chosen of nonzero elements of this space.

A linear space $\mathcal{V}$ is said to be n*-dimensional* if it contains n linearly independent elements and any $n + 1$ elements are linearly dependent. The number n is called the *dimension* of that space, $n = \dim \mathcal{V}$.

A linear space $\mathcal{V}$ is said to be *infinite-dimensional* ($\dim \mathcal{V} = \infty$) if for any positive integer N it contains N linearly independent elements.

THEOREM 1. *If* $\mathcal{V}$ *is a linear space of dimension* n*, then any* n *linearly independent elements of that space form its basis.*

THEOREM 2. *If a linear space* $\mathcal{V}$ *has a basis consisting of* n *elements, then* $\dim \mathcal{V} = n$.

Example 6. The dimension of the space B_3 of all vectors is equal to 3. The dimension of the space $\mathbb{R}^n$ is equal to n. The dimension of the space $\{x\}$ is equal to 1.

Two linear spaces $\mathcal{V}$ and $\mathcal{V}'$ over the same field of scalars are said to be *isomorphic* if there is a one-to-one correspondence between the elements of these spaces such that if elements $\mathbf{x}$ and $\mathbf{y}$ from $\mathcal{V}$ correspond to elements $\mathbf{x}'$ and $\mathbf{y}'$ from $\mathcal{V}'$, then the element $\mathbf{x} + \mathbf{y}$ corresponds to $\mathbf{x}' + \mathbf{y}'$ and the element $\lambda\mathbf{x}$ corresponds to $\lambda\mathbf{x}'$ for any scalar λ.

Remark. If linear spaces $\mathcal{V}$ and $\mathcal{V}'$ are isomorphic, then the zero element of one space corresponds to the zero element of the other.

THEOREM. *Any two* n*-dimensional real (or complex) spaces* $\mathcal{V}$ *and* $\mathcal{V}'$ *are isomorphic.*

5.3.1-3. Affine space.

An *affine space* is a nonempty set $\mathcal{A}$ that consists of elements of any nature, called *points*, for which the following conditions hold:

I. There is a given linear (vector) space $\mathcal{V}$, called the *associated linear space*.

II. There is a rule by which any ordered pair of points $A, B \in \mathcal{A}$ is associated with an element (vector) from $\mathcal{V}$; this vector is denoted by $\overrightarrow{AB}$ and is called the *vector issuing from the point* A *with endpoint at* B.

III. The following conditions (called *axioms of affine space*) hold:

1. For any point $A \in \mathcal{A}$ and any vector $\mathbf{a} \in \mathcal{V}$, there is a unique point $B \in \mathcal{A}$ such that $\overrightarrow{AB} = \mathbf{a}$.

2. $\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$ for any three points $A, B, C \in \mathcal{A}$.

By definition, the *dimension of an affine space* $\mathcal{A}$ is the dimension of the associated linear space $\mathcal{V}$, $\dim \mathcal{A} = \dim \mathcal{V}$.

Any linear space may be regarded as an affine space.

In particular, the space $\mathbb{R}^n$ can be naturally considered as an affine space. Thus if $A = (a_1, \ldots, a_n)$ and $B = (b_1, \ldots, b_n)$ are points of the affine space $\mathbb{R}^n$, then the corresponding vector $\overrightarrow{AB}$ from the linear space $\mathbb{R}^n$ is defined by $\overrightarrow{AB} = (b_1 - a_1, \ldots, b_n - a_n)$.

Let $\mathcal{A}$ be an n-dimensional affine space with the associated linear space $\mathcal{V}$. A *coordinate system* in the affine space $\mathcal{A}$ is a fixed point $O \in \mathcal{A}$, together with a fixed basis $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathcal{V}$. The point O is called the *origin* of this coordinate system.

Let M be a point of an affine space $\mathcal{A}$ with a coordinate system $O\mathbf{e}_1 \ldots \mathbf{e}_n$. One says that the point M has *affine coordinates* (or simply coordinates) $x_1, \ldots, x_n$ in this coordinate system, and one writes $M = (x_1, \ldots, x_n)$ if $x_1, \ldots x_n$ are the coordinates of the radius-vector $\overrightarrow{OM}$ in the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$, i.e., $\overrightarrow{OM} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$.

5.3.2. Subspaces of Linear Spaces

5.3.2-1. Concept of a linear subspace and a linear span.

A subset $\mathcal{L}$ of a linear space $\mathcal{V}$ is called a *linear subspace* of $\mathcal{V}$ if the following conditions hold:

1. If $\mathbf{x}$ and $\mathbf{y}$ belong to $\mathcal{L}$, then the sum $\mathbf{x} + \mathbf{y}$ belongs to $\mathcal{L}$.
2. If $\mathbf{x}$ belongs to $\mathcal{L}$ and λ is an arbitrary scalar, then the element $\lambda\mathbf{x}$ belongs to $\mathcal{L}$.

The *null subspace* in a linear space $\mathcal{V}$ is its subset consisting of the single element zero. The space $\mathcal{V}$ itself can be regarded as its own subspace. These two subspaces are called *improper subspaces*. All other subspaces are called *proper subspaces*.

Example 1. A subset B_2 consisting of all free vectors parallel to a given plane is a subspace in the linear space B_3 of all free vectors.

The *linear span* $L(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_m$ in a linear space $\mathcal{V}$ is, by definition, the set of all linear combinations of these vectors, i.e., the set of all vectors of the form

$$\alpha_1\mathbf{x}_1 + \cdots + \alpha_m\mathbf{x}_m,$$

where $\alpha_1, \ldots, \alpha_m$ are arbitrary scalars. The linear span $L(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ is the least subspace of $\mathcal{V}$ containing the elements $\mathbf{x}_1, \ldots, \mathbf{x}_m$.

If a subspace $\mathcal{L}$ of an n-dimensional space $\mathcal{V}$ does not coincide with $\mathcal{V}$, then $\dim \mathcal{L} < n = \dim \mathcal{V}$.

Let elements $\mathbf{e}_1, \ldots, \mathbf{e}_k$ form a basis in a k-dimensional subspace of an n-dimensional linear space $\mathcal{V}$. Then this basis can be supplemented by elements $\mathbf{e}_{k+1}, \ldots, \mathbf{e}_n$ of the space $\mathcal{V}$, so that the system $\mathbf{e}_1, \ldots, \mathbf{e}_k, \mathbf{e}_{k+1}, \ldots, \mathbf{e}_n$ forms a basis in the space $\mathcal{V}$.

THEOREM OF THE DIMENSION OF A LINEAR SPAN. *The dimension of a linear span $L(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ of elements $\mathbf{x}_m, \ldots, \mathbf{x}_m$ is equal to the maximal number of linearly independent vectors in the system $\mathbf{x}_1, \ldots, \mathbf{x}_m$.*

5.3.2-2. Sum and intersection of subspaces.

The *intersection* of subspaces $\mathcal{L}_1$ and $\mathcal{L}_2$ of one and the same linear space $\mathcal{V}$ is, by definition, the set of all elements $\mathbf{x}$ of $\mathcal{V}$ that belong simultaneously to both spaces $\mathcal{L}_1$ and $\mathcal{L}_2$. Such elements form a subspace of $\mathcal{V}$.

The *sum* of subspaces $\mathcal{L}_1$ and $\mathcal{L}_2$ of one and the same linear space $\mathcal{V}$ is, by definition, the set of all elements of $\mathcal{V}$ that can be represented in the form $\mathbf{y} + \mathbf{z}$, where $\mathbf{y}$ is an element of $\mathcal{V}_1$ and $\mathbf{z}$ is an element of $\mathcal{L}_2$. The sum of subspaces is also a subspace of $\mathcal{V}$.

THEOREM. *The sum of dimensions of arbitrary subspaces $\mathcal{L}_1$ and $\mathcal{L}_2$ of a finite-dimensional space $\mathcal{V}$ is equal to the sum of the dimension of their intersection and the dimension of their sum.*

Example 2. Let $\mathcal{V}$ be the linear space of all free vectors (in three-dimensional space). Denote by $\mathcal{L}_1$ the subspace of all free vectors parallel to the plane OXY, and by $\mathcal{L}_2$ the subspace of all free vectors parallel to the plane OXZ. Then the sum of the subspaces $\mathcal{L}_1$ and $\mathcal{L}_2$ coincides with $\mathcal{V}$, and their intersection consists of all free vectors parallel to the axis OX.

The dimension of each space $\mathcal{L}_1$ and $\mathcal{L}_2$ is equal to two, the dimension of their sum is equal to three, and the dimension of their intersection is equal to unity.

5.3.2-3. Representation of a linear space as a direct sum of its subspaces.

A linear space $\mathcal{V}$ can be represented as a *direct sum* of its subspaces, $\mathcal{V}_1$ and $\mathcal{V}_2$ if each element $\mathbf{x} \in \mathcal{V}$ admits the unique representation $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$, where $\mathbf{x}_1 \in \mathcal{V}_1$ and $\mathbf{x}_2 \in \mathcal{V}_2$. In this case, one writes $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2$.

Example 3. The space $\mathcal{V}$ of all free vectors (in three-dimensional space) can be represented as the direct sum of the subspace $\mathcal{V}_1$ formed by all free vectors parallel to the plane OXY and the subspace $\mathcal{V}_2$ formed by all free vectors parallel to the axis OZ.

THEOREM. *An n-dimensional space $\mathcal{V}$ is a direct sum of its subspaces $\mathcal{V}_1$ and $\mathcal{V}_2$ if and only if the intersection of $\mathcal{V}_1$ and $\mathcal{V}_2$ is the null subspace and* $\dim \mathcal{V} = \dim \mathcal{V}_1 + \dim \mathcal{V}_2$.

Remark. If R is the sum of its subspaces R_1 and R_2, but not the direct sum, then the representation $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ is nonunique, in general.

5.3.3. Coordinate Transformations Corresponding to Basis Transformations in a Linear Space

5.3.3-1. Basis transformation and its inverse.

Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_n$ be two arbitrary bases of an n-dimensional linear space $\mathcal{V}$. Suppose that the elements $\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_n$ are expressed via $\mathbf{e}_1, \ldots, \mathbf{e}_n$ by the formulas

$$\begin{aligned}
\widetilde{\mathbf{e}}_1 &= a_{11}\mathbf{e}_1 + a_{12}\mathbf{e}_2 + \cdots + a_{1n}\mathbf{e}_n,\\
\widetilde{\mathbf{e}}_2 &= a_{21}\mathbf{e}_1 + a_{22}\mathbf{e}_2 + \cdots + a_{2n}\mathbf{e}_n,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
\widetilde{\mathbf{e}}_n &= a_{n1}\mathbf{e}_1 + a_{n2}\mathbf{e}_2 + \cdots + a_{nn}\mathbf{e}_n.
\end{aligned}$$

Thus, the transition from the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ to the basis $\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_n$ is determined by the matrix

$$A \equiv \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

Note that $\det A \neq 0$, i.e., the matrix A is nondegenerate.

The transition from the basis $\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_n$ to the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is determined by the matrix $B \equiv [b_{ij}] = A^{-1}$. Thus, we can write

$$\widetilde{\mathbf{e}}_i = \sum_{j=1}^{n} a_{ij}\mathbf{e}_j, \qquad \mathbf{e}_k = \sum_{j=1}^{n} b_{kj}\widetilde{\mathbf{e}}_j \qquad (i, k = 1, 2, \ldots, n). \tag{5.3.3.1}$$

5.3.3-2. Relations between coordinate transformations and basis transformations.

Suppose that in a linear n-dimensional space $\mathcal{V}$, the transition from its basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ to another basis $\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_n$ is determined by the matrix A (see Paragraph 5.3.3-1). Let $\mathbf{x}$ be any element of the space $\mathcal{V}$ with the coordinates $(x_1, \ldots, x_n)$ in the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and the coordinates $(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ in the basis $\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_n$, i.e.,

$$\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n = \widetilde{x}_1\widetilde{\mathbf{e}}_1 + \cdots + \widetilde{x}_n\widetilde{\mathbf{e}}_n.$$

Then using formulas (5.3.3.1), we obtain the following relations between these coordinates:

$$x_j = \sum_{i=1}^{n} \widetilde{x}_i a_{ij}, \qquad \widetilde{x}_k = \sum_{l=1}^{n} x_l b_{lk}, \qquad j, k = 1, \ldots, n.$$

In terms of matrices and row vectors, these relations can be written as follows:

$$(x_1, \ldots, x_n) = (\widetilde{x}_1, \ldots, \widetilde{x}_n)A, \qquad (\widetilde{x}_1, \ldots, \widetilde{x}_n) = (x_1, \ldots, x_n)A^{-1}$$

or, in terms of column vectors,

$$(x_1, \ldots, x_n)^T = A^T(\widetilde{x}_1, \ldots, \widetilde{x}_n)^T, \qquad (\widetilde{x}_1, \ldots, \widetilde{x}_n)^T = (A^{-1})^T(x_1, \ldots, x_n)^T,$$

where the superscript T indicates the transpose of a matrix.

5.4. Euclidean Spaces

5.4.1. Real Euclidean Space

5.4.1-1. Definition and properties of a real Euclidean space.

A *real Euclidean space* (or simply, *Euclidean space*) is a real linear space $\mathcal{V}$ endowed with a *scalar product* (also called *inner product*), which is a real-valued function of two arguments $\mathbf{x} \in \mathcal{V}$, $\mathbf{y} \in \mathcal{V}$ called the scalar product of these elements, denoted by $\mathbf{x} \cdot \mathbf{y}$, and satisfying the following conditions (axioms of the scalar product):

1. Symmetry: $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$.
2. Distributivity: $(\mathbf{x}_1 + \mathbf{x}_2) \cdot \mathbf{y} = \mathbf{x}_1 \cdot \mathbf{y} + \mathbf{x}_2 \cdot \mathbf{y}$.
3. Homogeneity: $(\lambda\mathbf{x}) \cdot \mathbf{y} = \lambda(\mathbf{x} \cdot \mathbf{y})$ for any real λ.
4. Positive definiteness: $\mathbf{x} \cdot \mathbf{x} \geq 0$ for any $\mathbf{x}$, and $\mathbf{x} \cdot \mathbf{x} = 0$ if and only if $\mathbf{x} = 0$.

If the nature of the elements and the scalar product is concretized, one obtains a *specific Euclidean space*.

Example 1. Consider the linear space B_3 of all free vectors in three-dimensional space. The space B_3 becomes a Euclidean space if the scalar product is introduced as in analytic geometry (see Paragraph 4.5.3-1):

$$\mathbf{x} \cdot \mathbf{y} = |\mathbf{x}|\,|\mathbf{y}| \cos\varphi,$$

where φ is the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$.

Example 2. Consider the n-dimensional coordinate space $\mathbb{R}^n$ whose elements are ordered systems of n arbitrary real numbers, $\mathbf{x} = (x_1, \ldots, x_n)$. Endowing this space with the scalar product

$$\mathbf{x} \cdot \mathbf{y} = x_1y_1 + \cdots + x_ny_n,$$

we obtain a Euclidean space.

THEOREM. *For any two elements* $\mathbf{x}$ *and* $\mathbf{y}$ *of a Euclidean space, the Cauchy–Schwarz inequality holds:*

$$(\mathbf{x}\cdot\mathbf{y})^2 \le (\mathbf{x}\cdot\mathbf{x})(\mathbf{y}\cdot\mathbf{y}).$$

A linear space $\mathcal{V}$ is called a *normed space* if it is endowed with a *norm*, which is a real-valued function of $\mathbf{x}\in\mathcal{V}$, denoted by $\|\mathbf{x}\|$ and satisfying the following conditions:

1. Homogeneity: $\|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\|$ for any real λ.
2. Positive definiteness: $\|\mathbf{x}\| \ge 0$ and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = 0$.
3. The *triangle inequality* (also called the *Minkowski inequality*) holds for all elements $\mathbf{x}$ and $\mathbf{y}$:

$$\|\mathbf{x}+\mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|. \tag{5.4.1.1}$$

The value $\|\mathbf{x}\|$ is called the *norm of an element* $\mathbf{x}$ or its *length*.

THEOREM. *Any Euclidean space becomes a normed space if the norm is introduced by*

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}\cdot\mathbf{x}}. \tag{5.4.1.2}$$

COROLLARY. *In any Euclidean space with the norm (5.4.1.2), the triangle inequality (5.4.1.1) holds for all its elements* $\mathbf{x}$ *and* $\mathbf{y}$.

The *distance between elements* $\mathbf{x}$ and $\mathbf{y}$ of a Euclidean space is defined by

$$d(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|.$$

One says that $\varphi\in[0,2\pi]$ is the *angle* between two elements $\mathbf{x}$ and $\mathbf{y}$ of a Euclidean space if

$$\cos\varphi = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}.$$

Two elements $\mathbf{x}$ and $\mathbf{y}$ of a Euclidean space are said to be *orthogonal* if their scalar product is equal to zero, $\mathbf{x}\cdot\mathbf{y} = 0$.

PYTHAGOREAN THEOREM. *Let* $\mathbf{x}_1, \dots \mathbf{x}_m$ *be mutually orthogonal elements of a Euclidean space, i.e.,* $\mathbf{x}_j\cdot\mathbf{x}_j = 0$ *for* $i\ne j$. *Then*

$$\|\mathbf{x}_1+\cdots+\mathbf{x}_m\|^2 = \|\mathbf{x}_1\|^2+\cdots+\|\mathbf{x}_m\|^2.$$

Example 3. In the Euclidean space B_3 of free vectors with the usual scalar product (see Example 1), the following relations hold:

$$\|\mathbf{a}\| = |\mathbf{a}|, \quad (\mathbf{a}\cdot\mathbf{b})^2 \le |\mathbf{a}|^2|\mathbf{b}|^2, \quad |\mathbf{a}+\mathbf{b}| \le |\mathbf{a}|+|\mathbf{b}|.$$

In the Euclidean space $\mathbb{R}^n$ of ordered systems of n numbers with the scalar product defined in Example 2, the following relations hold:

$$\|\mathbf{x}\| = \sqrt{x_1^2+\cdots+x_n^2},$$
$$(x_1y_1+\cdots+x_ny_n)^2 \le (x_1^2+\cdots+x_n^2)(y_1^2+\cdots+y_n^2),$$
$$\sqrt{(x_1+y_1)^2+\cdots+(x_n+y_n)^2} \le \sqrt{x_1^2+\cdots+x_n^2}\sqrt{y_1^2+\cdots+y_n^2}.$$

5.4.1-2. Orthonormal basis in a finite-dimensional Euclidean space.

For elements $\mathbf{x}_1, \dots, \mathbf{x}_m$ of a Euclidean space, the mth-order determinant $\det[\mathbf{x}_i\cdot\mathbf{x}_j]$ is called their *Gram determinant*. These elements are linearly independent if and only if their Gram determinant is different from zero.

One says that n elements $\mathbf{i}_1, \ldots, \mathbf{i}_n$ of an n-dimensional Euclidean space $\mathcal{V}$ form its *orthonormal basis* if these elements have unit norm and are mutually orthogonal, i.e.,

$$\mathbf{i}_i \cdot \mathbf{i}_j = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases}$$

THEOREM. *In any n-dimensional Euclidean space $\mathcal{V}$, there exists an orthonormal basis.*

Orthogonalization of linearly independent elements:

Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be n linearly independent vectors of an n-dimensional Euclidean space $\mathcal{V}$. From these vectors, one can construct an orthonormal basis of $\mathcal{V}$ using the following algorithm (called *Gram–Schmidt orthogonalization*):

$$\mathbf{i}_i = \frac{\mathbf{g}_i}{\sqrt{\mathbf{g}_i \cdot \mathbf{g}_i}}, \quad \text{where} \quad \mathbf{g}_i = \mathbf{e}_i - \sum_{j=1}^{i} (\mathbf{e}_i \cdot \mathbf{i}_j)\mathbf{i}_j \quad (i = 1, 2, \ldots, n). \tag{5.4.1.3}$$

Remark. In any n-dimensional ($n > 1$) Euclidean space $\mathcal{V}$, there exist infinitely many orthonormal bases.

Properties of an orthonormal basis of a Euclidean space:

1. Let $\mathbf{i}_1, \ldots, \mathbf{i}_n$ be an orthonormal basis of a Euclidean space $\mathcal{V}$. Then the scalar product of two elements $\mathbf{x} = x_1\mathbf{i}_1 + \cdots + x_n\mathbf{i}_n$ and $\mathbf{y} = y_1\mathbf{i}_1 + \cdots + y_n\mathbf{i}_n$ is equal to the sum of products of their respective coordinates:

$$\mathbf{x} \cdot \mathbf{y} = x_1y_1 + \cdots + x_ny_n.$$

2. The coordinates of any vector $\mathbf{x}$ in an orthonormal basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ are equal to the scalar product of $\mathbf{x}$ and the corresponding vector of the basis (or the projection of the element $\mathbf{x}$ on the axis in the direction of the corresponding vector of the basis):

$$x_k = \mathbf{x} \cdot \mathbf{i}_k \quad (k = 1, 2, \ldots, n).$$

Remark. In an arbitrary basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of a Euclidean space, the scalar product of two elements $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$ and $\mathbf{y} = y_1\mathbf{e}_1 + \cdots + y_n\mathbf{e}_n$ has the form

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}x_iy_j,$$

where $a_{ij} = \mathbf{e}_i \cdot \mathbf{e}_j$ $(i, j = 1, 2, \ldots, n)$.

Let $\mathcal{X}, \mathcal{Y}$ be subspaces of a Euclidean space $\mathcal{V}$. The subspace $\mathcal{X}$ is called the *orthogonal complement* of the subspace $\mathcal{Y}$ in $\mathcal{V}$ if any element $\mathbf{x}$ of $\mathcal{X}$ is orthogonal to any element $\mathbf{y}$ of $\mathcal{Y}$ and $\mathcal{X} \oplus \mathcal{Y} = \mathcal{V}$.

THEOREM. *Any n-dimensional Euclidean space $\mathcal{V}$ can be represented as the direct sum of its arbitrary subspace $\mathcal{Y}$ and its orthogonal complement $\mathcal{X}$.*

Two Euclidean spaces $\mathcal{V}$ and $\widetilde{\mathcal{V}}$ are said to be *isomorphic* if one can establish a one-to-one correspondence between the elements of these spaces satisfying the following conditions: if elements $\mathbf{x}$ and $\mathbf{y}$ of $\mathcal{V}$ correspond to elements $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$ of $\widetilde{\mathcal{V}}$, then the element $\mathbf{x} + \mathbf{y}$ corresponds to $\widetilde{\mathbf{x}} + \widetilde{\mathbf{y}}$; the element $\lambda\mathbf{x}$ corresponds to $\lambda\widetilde{\mathbf{x}}$ for any λ; the scalar product $(\mathbf{x} \cdot \mathbf{y})_{\mathcal{V}}$ is equal to the scalar product $(\widetilde{\mathbf{x}} \cdot \widetilde{\mathbf{y}})_{\widetilde{\mathcal{V}}}$.

THEOREM. *Any two n-dimensional Euclidean spaces $\mathcal{V}$ and $\widetilde{\mathcal{V}}$ are isomorphic.*

5.4.2. Complex Euclidean Space (Unitary Space)

5.4.2-1. Definition and properties of complex Euclidean space (unitary space).

A *complex Euclidean space* (or *unitary space*) is a complex linear space $\mathcal{V}$ endowed with a *scalar product* (also called *inner product*), which is a complex-valued function of two arguments $\mathbf{x} \in \mathcal{V}$ and $\mathbf{y} \in \mathcal{V}$ called their scalar product, denoted by $\mathbf{x} \cdot \mathbf{y}$, satisfying the following conditions (called axioms of the scalar product):

1. Commutativity: $\mathbf{x} \cdot \mathbf{y} = \overline{\mathbf{y} \cdot \mathbf{x}}$.
2. Distributivity: $(\mathbf{x}_1 + \mathbf{x}_2) \cdot \mathbf{y} = \mathbf{x}_1 \cdot \mathbf{y} + \mathbf{x}_2 \cdot \mathbf{y}$.
3. Homogeneity: $(\lambda \mathbf{x}) \cdot \mathbf{y} = \lambda(\mathbf{x} \cdot \mathbf{y})$ for any complex λ.
4. Positive definiteness: $\mathbf{x} \cdot \mathbf{x} \geq 0$; and $\mathbf{x} \cdot \mathbf{x} = 0$ if and only if $\mathbf{x} = 0$.

Here $\overline{\mathbf{y} \cdot \mathbf{x}}$ is the complex conjugate of a number $\mathbf{y} \cdot \mathbf{x}$.

Example 1. Consider the n-dimensional complex linear space $\mathbb{R}^n_*$ whose elements are ordered systems of n complex numbers, $\mathbf{x} = (x_1, \ldots, x_n)$. We obtain a unitary space if the scalar product of two elements $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ is introduced by

$$\mathbf{x} \cdot \mathbf{y} = x_1 \bar{y}_1 + \cdots + x_n \bar{y}_n,$$

where $\bar{y}_j$ is the complex conjugate of y_j.

THEOREM. *For any two elements* $\mathbf{x}$ *and* $\mathbf{y}$ *of an arbitrary unitary space, the Cauchy–Schwarz inequality holds:*

$$|\mathbf{x} \cdot \mathbf{y}|^2 \leq (\mathbf{x} \cdot \mathbf{x})(\mathbf{y} \cdot \mathbf{y}).$$

THEOREM. *Any unitary space becomes a normed space if the norm of its element* $\mathbf{x}$ *is introduced by*

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}. \tag{5.4.2.1}$$

COROLLARY. *For any two elements* $\mathbf{x}$ *and* $\mathbf{y}$ *of a normed Euclidean space with the norm (5.4.2.1), the triangle inequality (5.4.1.1) holds.*

The *distance between elements* $\mathbf{x}$ and $\mathbf{y}$ of a unitary space is defined by

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|. \tag{5.4.2.2}$$

Two elements $\mathbf{x}$ and $\mathbf{y}$ of a unitary space are said to be *orthogonal* if their scalar product is equal to zero, $\mathbf{x} \cdot \mathbf{y} = 0$.

5.4.2-2. Orthonormal basis in a finite-dimensional unitary space.

Elements $\mathbf{x}_1, \ldots, \mathbf{x}_m$ of a unitary space $\mathcal{V}$ are linearly independent if and only if their *Gram determinant* is different from zero, $\det[\mathbf{x}_i \cdot \mathbf{x}_j] \neq 0$.

One says that elements $\mathbf{i}_1, \ldots, \mathbf{i}_n$ of an n-dimensional unitary space $\mathcal{V}$ form an *orthonormal basis* of that space if these elements are mutually orthogonal and have unit norm, i.e.,

$$\mathbf{i}_i \cdot \mathbf{i}_j = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases}$$

Given any n linearly independent elements of a unitary space, one can construct an orthonormal basis of that space using the procedure described in Paragraph 5.4.1-2 (see formulas (5.4.1.3)).

Properties of an orthonormal basis of a unitary space:

1. Let $\mathbf{i}_1, \ldots, \mathbf{i}_n$ be an orthonormal basis in a unitary space. Then the scalar product of two elements $\mathbf{x} = x_1\mathbf{i}_1 + \cdots + x_n\mathbf{i}_n$ and $\mathbf{y} = y_1\mathbf{i}_1 + \cdots + y_n\mathbf{i}_n$ is equal to the sum

$$\mathbf{x} \cdot \mathbf{y} = x_1\bar{y}_1 + \cdots + x_n\bar{y}_n.$$

2. The coordinates of any vector in an orthonormal basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ are equal to the scalar products of this vector and the vectors of the bases (or the projections of this element on the axes in the direction of the corresponding basis vectors):

$$x_k = \mathbf{x} \cdot \mathbf{i}_k \qquad (k = 1, 2, \ldots, n).$$

Two unitary spaces $\mathcal{V}$ and $\widetilde{\mathcal{V}}$ are said to be *isomorphic* if there is a one-to-one correspondence between their elements satisfying the following conditions: if elements $\mathbf{x}$ and $\mathbf{y}$ of $\mathcal{V}$ correspond to elements $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$ of $\widetilde{\mathcal{V}}$, then $\mathbf{x} + \mathbf{y}$ corresponds to $\widetilde{\mathbf{x}} + \widetilde{\mathbf{y}}$; the element $\lambda\mathbf{x}$ corresponds to $\lambda\widetilde{\mathbf{x}}$ for any complex λ; the scalar product $(\mathbf{x} \cdot \mathbf{y})_{\mathcal{V}}$ is equal to the scalar product $(\widetilde{\mathbf{x}} \cdot \widetilde{\mathbf{y}})_{\widetilde{\mathcal{V}}}$.

THEOREM. *Any two n-dimensional unitary spaces $\mathcal{V}$ and $\widetilde{\mathcal{V}}$ are isomorphic.*

5.4.3. Banach Spaces and Hilbert Spaces

5.4.3-1. Convergence in unitary spaces. Banach space.

Any normed linear space is a *metric space* with the metric (5.4.2.2).

A sequence $\{\mathbf{b}_s\}$ of elements of a normed space $\mathcal{V}$ is said to be convergent to an element $\mathbf{b} \in \mathcal{V}$ as $s \to \infty$ if $\lim\limits_{s\to\infty} \|\mathbf{b}_s - \mathbf{b}\| = 0$.

A series $\mathbf{x}_0 + \mathbf{x}_1 + \cdots$ with terms in a normed space is said to be convergent to an element $\mathbf{x}$ (called its sum; one writes $\mathbf{x} = \lim\limits_{n\to\infty} \sum\limits_{k=0}^{n} \mathbf{x}_k = \sum\limits_{k=0}^{\infty} \mathbf{x}_k$) if the sequence of its partial sums forms a sequence convergent to $\mathbf{x}$, i.e., $\lim\limits_{n\to\infty} \left\| \mathbf{x} - \sum\limits_{k=0}^{n} \mathbf{x}_k \right\| = 0$.

A normed linear space $\mathcal{V}$ is said to be *complete* if any sequence of its elements $\mathbf{s}_0$, $\mathbf{s}_1, \ldots$ satisfying the condition

$$\lim_{n,m\to\infty} \|\mathbf{s}_n - \mathbf{s}_m\| = 0$$

is convergent to some element $\mathbf{s}$ of the space $\mathcal{V}$.

A complete normed linear space is called a *Banach space*.

Remark. Any finite-dimensional normed linear space is complete.

5.4.3-2. Hilbert space.

A complete unitary space is called a *Hilbert space*.

Any complete subspace of a Hilbert space is itself a Hilbert space.

PROJECTION THEOREM. *Let $\mathcal{V}_1$ be a complete subspace of a unitary space $\mathcal{V}$. Then for any $\mathbf{x} \in \mathcal{V}$, there is a unique vector $\mathbf{x}_p \in \mathcal{V}_1$ such that*

$$\min_{\mathbf{y}\in\mathcal{V}_1} \|\mathbf{x} - \mathbf{y}\| = \|\mathbf{x} - \mathbf{x}_p\|.$$

Moreover, the vector $\mathbf{x}_p$ *is the unique element of* $\mathcal{V}_1$ *for which the difference* $\mathbf{x} - \mathbf{x}_p$ *is orthogonal to any element* $\mathbf{x}_1$ *of* $\mathcal{V}_1$, *i.e.,*

$$(\mathbf{x} - \mathbf{x}_p) \cdot \mathbf{x}_1 = 0 \quad \textit{for all} \quad \mathbf{x}_1 \in \mathcal{V}_1.$$

The mapping $\mathbf{x} \to \mathbf{x}_p$ is a bounded linear operator from $\mathcal{V}$ to $\mathcal{V}_1$ called the *orthogonal projection of the space* $\mathcal{V}$ *to its subspace* $\mathcal{V}_1$.

5.5. Systems of Linear Algebraic Equations

5.5.1. Consistency Condition for a Linear System

5.5.1-1. Notion of a system of linear algebraic equations.

A *system of* m *linear equations with* n *unknown quantities* has the form

$$\begin{aligned} &a_{11}x_1 + a_{12}x_2 + \cdots + a_{1k}x_k + \cdots + a_{1n}x_n = b_1, \\ &a_{21}x_1 + a_{22}x_2 + \cdots + a_{2k}x_k + \cdots + a_{2n}x_n = b_2, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mk}x_k + \cdots + a_{mn}x_n = b_m, \end{aligned} \tag{5.5.1.1}$$

where $a_{11}, a_{12}, \ldots, a_{mn}$ are the *coefficients of the system*; $b_1, b_2, \ldots, b_m$ are its *free terms*; and $x_1, x_2, \ldots, x_n$ are the unknown quantities.

System (5.5.1.1) is said to be *homogeneous* if all its free terms are equal to zero. Otherwise (i.e., if there is at least one nonzero free term) the system is called *nonhomogeneous*.

If the number of equations is equal to that of the unknown quantities ($m = n$), system (5.5.1.1) is called a *square system*.

A *solution* of system (5.5.1.1) is a set of n numbers $x_1, x_2, \ldots, x_n$ satisfying the equations of the system. A system is said to be *consistent* if it admits at least one solution. If a system has no solutions, it is said to be *inconsistent*. A consistent system of the form (5.5.1.1) is called a *determined system* — it has a unique solution. A consistent system with more than one solution is said to be *underdetermined*.

It is convenient to use matrix notation for systems of the form (5.5.1.1),

$$AX = B, \tag{5.5.1.2}$$

where $A \equiv [a_{ij}]$ is a matrix of size $m \times n$ called the *basic matrix* of the system; $X \equiv [x_i]$ is a column vector of size n; $B \equiv [b_i]$ is a column vector of size m.

5.5.1-2. Existence of nontrivial solutions of a homogeneous system.

Consider a homogeneous system

$$AX = O_m, \tag{5.5.1.3}$$

where $A \equiv [a_{ij}]$ is its basic matrix of size $m \times n$, $X \equiv [x_i]$ is a column vector of size n, and $O_m \equiv [0]$ is a column vector of size m. System (5.5.1.3) is always consistent since it always has the so-called *trivial solution* $X \equiv O_n$.

THEOREM. *A homogeneous system (5.5.1.3) has a nontrivial solution if and only if the rank of the matrix* A *is less than the number of the unknown quantities* n.

It follows that a square homogeneous system has a nontrivial solution if and only if the determinant of its matrix of coefficients is equal to zero, $\det A = 0$.

Properties of the set of all solutions of a homogeneous system:

1. All solutions of a homogeneous system (5.5.1.3) form a linear space.
2. The linear space of all solutions of a homogeneous system (5.5.1.3) with n unknown quantities and a basic matrix of rank r is isomorphic to the space A^{n-r} of all ordered systems of $(n-r)$ numbers. The dimension of the space of solutions is equal to $n-r$.
3. Any system of $(n-r)$ linearly independent solutions of the homogeneous system (5.5.1.3) forms a basis in the space of all its solutions and is called a *fundamental system of solutions* of that system. The fundamental system of solutions corresponding to the basis $\mathbf{i}_1 = (1, 0, \ldots, 0)$, $\mathbf{i}_2 = (0, 1, \ldots, 0)$, $\ldots$, $\mathbf{i}_{n-r} = (0, 0, \ldots, 1)$ of the space A^{n-r} is said to be *normal*.

5.5.1-3. Consistency condition for a general linear system.

System (5.5.1.1) or (5.5.1.2) is associated with two matrices: the basic matrix A of size $m \times n$ and the *augmented matrix* A_1 of size $m \times (n+1)$ formed by the matrix A supplemented with the column of the free terms, i.e.,

$$A_1 \equiv \left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right). \tag{5.5.1.4}$$

KRONECKER–CAPELLI THEOREM. *A linear system (5.5.1.1) [or (5.5.1.2)] is consistent if and only if its basic matrix and its augmented matrix (5.5.1.4) have the same rank, i.e.* $\operatorname{rank}(A_1) = \operatorname{rank}(A)$.

5.5.2. Finding Solutions of a System of Linear Equations

5.5.2-1. System of two equations with two unknown quantities.

A system of two equations with two unknown quantities has the form

$$\begin{aligned} a_1 x + b_1 y &= c_1, \\ a_2 x + b_2 y &= c_2. \end{aligned} \tag{5.5.2.1}$$

Depending on the coefficients a_k, b_k, c_k, the following three cases are possible:

1°. If $\Delta = a_1 b_2 - a_2 b_1 \neq 0$, then system (5.5.2.1) has a unique solution,

$$x = \frac{c_1 b_2 - c_2 b_1}{a_1 b_2 - a_2 b_1}, \quad y = \frac{a_1 c_2 - a_2 c_1}{a_1 b_2 - a_2 b_1}.$$

2°. If $\Delta = a_1 b_2 - a_2 b_1 = 0$ and $a_1 c_2 - a_2 c_1 = 0$ (the case of proportional coefficients), then system (5.5.2.1) has infinitely many solutions described by the formulas

$$x = t, \quad y = \frac{c_1 - a_1 t}{b_1} \quad (b_1 \neq 0),$$

where t is arbitrary.

3°. If $\Delta = a_1 b_2 - a_2 b_1 = 0$ and $a_1 c_2 - a_2 c_1 \neq 0$, then system (5.5.2.1) has no solutions.

5.5.2-2. General square system of linear equations.

A square system of linear equations has the form

$$AX = B, \tag{5.5.2.2}$$

where A is a square matrix.

1°. If the determinant of system (5.5.2.2) is different from zero, i.e. $\det A \neq 0$, then the system has a unique solution,

$$X = A^{-1}B.$$

2°. *Cramer rule*. If the determinant of the matrix of system (5.5.2.2) is different from zero, i.e. $\Delta = \det A \neq 0$, then the system admits a unique solution, which is expressed by

$$x_1 = \frac{\Delta_1}{\Delta}, \quad x_2 = \frac{\Delta_2}{\Delta}, \quad \ldots, \quad x_n = \frac{\Delta_n}{\Delta}, \tag{5.5.2.3}$$

where Δ_k ($k = 1, 2, \ldots, n$) is the determinant of the matrix obtained from A by replacing its kth column with the column of free terms:

$$\Delta_k = \begin{vmatrix} a_{11} & a_{12} & \ldots & b_1 & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & b_2 & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & b_n & \ldots & a_{nn} \end{vmatrix}.$$

Example 1. Using the Cramer rule, let us find the solution of the system of linear equations

$$\begin{aligned} 2x_1 + x_2 + 4x_3 &= 16, \\ 3x_1 + 2x_2 + x_3 &= 10, \\ x_1 + 3x_2 + 3x_3 &= 16. \end{aligned}$$

The determinant of its basic matrix is different from zero,

$$\Delta = \begin{vmatrix} 2 & 1 & 4 \\ 3 & 2 & 1 \\ 1 & 3 & 3 \end{vmatrix} = 26 \neq 0,$$

and we have

$$\Delta_1 = \begin{vmatrix} 16 & 1 & 4 \\ 10 & 2 & 1 \\ 16 & 3 & 3 \end{vmatrix} = 26, \quad \Delta_2 = \begin{vmatrix} 2 & 16 & 4 \\ 3 & 10 & 1 \\ 1 & 16 & 3 \end{vmatrix} = 52, \quad \Delta_3 = \begin{vmatrix} 2 & 1 & 16 \\ 3 & 2 & 10 \\ 1 & 3 & 16 \end{vmatrix} = 78.$$

Therefore, by the Cramer rule (5.5.2.3), the only solution of the system has the form

$$x_1 = \frac{\Delta_1}{\Delta} = \frac{26}{26} = 1, \quad x_2 = \frac{\Delta_2}{\Delta} = \frac{52}{26} = 2, \quad x_3 = \frac{\Delta_3}{\Delta} = \frac{78}{26} = 3.$$

3°. *Gaussian elimination of unknown quantities*.

Two systems are said to be *equivalent* if their sets of solutions coincide.

The method of Gaussian elimination consists in the reduction of a given system to an equivalent system with an upper triangular basic matrix. The latter system can be easily solved. This reduction is carried out in finitely many steps. On every step, one performs an *elementary transformation* of the system (or the corresponding augmented matrix) and obtains an equivalent system. The elementary transformations are of the following three types:

1. Interchange of two equations (or the corresponding rows of the augmented matrix).
2. Multiplication of both sides of one equation (or the corresponding row of the augmented matrix) by a nonzero constant.
3. Adding to both sides of one equation both sides of another equation multiplied by a nonzero constant (adding to some row of the augmented matrix its another row multiplied by a nonzero constant).

Suppose that $\det A \neq 0$. Then by consecutive elementary transformations, the augmented matrix of the system A_1 [see (5.5.1.4)] of size $n \times (n+1)$ can be reduced to the form

$$U_1 \equiv \left(\begin{array}{cccc|c} 1 & u_{12} & \cdots & u_{1n} & y_1 \\ 0 & 1 & \cdots & u_{2n} & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & y_n \end{array}\right)$$

and one obtains an equivalent system with an upper triangular basic matrix,

$$\begin{aligned} x_1 + u_{12}x_2 + u_{13}x_3 + \cdots + u_{1n}x_n &= y_1, \\ x_2 + u_{23}x_3 + \cdots + u_{2n}x_n &= y_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ x_n &= y_n. \end{aligned}$$

This system is solved by the so-called "backward substitution": inserting $x_n = y_n$ (obtained from the last equation) into the preceding $(n-1)$st equation, one finds x_{n-1}. Then inserting the values obtained for x_n, x_{n-1} into the $(n-2)$nd equation, one finds x_{n-2}. Proceeding in this way, one finally finds x_1. This back substitution process is described by the formulas

$$x_k = y_k - \sum_{s=k+1}^{n} u_{ks}x_s \quad (k = n-1, n-2, \ldots, 1).$$

Suppose that $\det A = 0$ and $\operatorname{rank}(A) = r$, $0 < r < n$. In this case, the system is either inconsistent (i.e., has no solutions) or has infinitely many solutions. By elementary transformations and, possibly, reindexing the unknown quantities (i.e., introducing new unknown quantities $y_1 = x_{\sigma(1)}, \ldots, y_n = x_{\sigma(n)}$, where $\sigma(1), \ldots, \sigma(n)$ is a permutation of the indices $1, 2, \ldots, n$), one obtains a system of the form (for the sake of brevity, we retain the notation x_j for the reindexed unknown quantities)

$$\begin{aligned} c_{11}x_1 + \cdots + c_{1r}x_r + c_{1,r+1}x_{r+1} + \cdots + c_{1n}x_n &= d_1, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ c_{rr}x_r + c_{r,r+1}x_{r+1} + \cdots + c_{rn}x_n &= d_r, \\ 0 &= d_{r+1}, \\ \cdots & \\ 0 &= d_n, \end{aligned}$$

where the matrix $[c_{ij}]$ $(i, j = 1, 2, \ldots, r)$ of size $r \times r$ is nondegenerate. If at least one of the right-hand sides $d_{r+1}, \ldots, d_n$ is different from zero, then the system is inconsistent. If $d_{r+1} = \ldots = d_n = 0$, then the last $n-r$ equations can be dropped and it remains to find all solutions of the first r equations. Transposing all terms containing the variables $x_{r+1}, \ldots, x_n$ to the right-hand sides and regarding these variables as arbitrary free parameters, we obtain a linear system for the unknown quantities $x_1, \ldots, x_r$ with the nondegenerate basic matrix $[c_{ij}]$ $(j, j = 1, 2, \ldots, r)$.

Example 2. Let us find a solution of the system from Example 1 by the Gaussian elimination method. By elementary transformations of the augmented matrix, we obtain

$$\left(\begin{array}{ccc|c} 2 & 1 & 4 & 16 \\ 3 & 2 & 1 & 10 \\ 1 & 3 & 3 & 16 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 1/2 & 2 & 8 \\ 0 & 1/2 & -5 & -14 \\ 0 & 5/2 & 1 & 8 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 1/2 & 2 & 8 \\ 0 & 1 & -10 & -28 \\ 0 & 0 & 26 & 78 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 1/2 & 2 & 8 \\ 0 & 1 & -10 & -28 \\ 0 & 0 & 1 & 3 \end{array}\right).$$

The transformed system has the form

$$x_1 + \tfrac{1}{2}x_2 + 2x_3 = 8,$$
$$x_2 - 10x_3 = -28,$$
$$x_3 = 3.$$

Hence, we find that

$$x_3 = 3, \quad x_2 = -28 + 10x_3 = 2, \quad x_1 = 8 - \tfrac{1}{2}x_2 - 2x_3 = 1.$$

4°. *Gauss-Jordan elimination of unknown quantities.*

This method consists of applying elementary transformations for reducing a system with a nondegenerate basic matrix to an equivalent system with the identity matrix. On the kth step ($k = 1, 2, \ldots, n$) the rows of the augmented matrix A_1' obtained on the preceding step can be transformed as follows:

$$a''_{kj} = \frac{a'_{kj}}{a'_{kk}}, \qquad b''_k = \frac{b'_k}{a'_{kk}} \qquad (j = k, k+1, \ldots, n),$$

$$a''_{ij} = a'_{ij} - a'_{ik}\frac{a'_{kj}}{a'_{kk}}, \quad b''_i = b'_i - a_{ik}\frac{b'_k}{a'_{kk}} \quad (i = 1, 2, \ldots, n,\ i \neq k,\ j = k, k+1, \ldots, n),$$

provided that the diagonal element obtained on each step is not equal to zero. After n steps, the basic matrix is transformed to the identity matrix and the right-hand side turns into the desired solution.

Example 3. For the linear system from Examples 1 and 2 we have

$$\left(\begin{array}{ccc|c} 2 & 1 & 4 & 16 \\ 3 & 2 & 1 & 10 \\ 1 & 3 & 3 & 16 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 1/2 & 2 & 8 \\ 0 & 1/2 & -5 & -14 \\ 0 & 5/2 & 1 & 8 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 0 & 7 & 22 \\ 0 & 1 & -10 & -28 \\ 0 & 0 & 26 & 78 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right),$$

and therefore $x_1 = 1$, $x_2 = 2$, $x_3 = 3$.

The diagonal element obtained on some step of the above elimination procedure may happen to be equal to zero. In this case, the formulas become more complicated and reindexing of the unknown quantities may be required.

5°. *Method of LU-decomposition.*

This method is based on the representation of the basic matrix A as the product of a lower triangular matrix L and an upper triangular matrix U, i.e., in the form $A = LU$. This factorization is called a *triangular representation* or the *LU-representation* of a matrix (see also Paragraph 5.2.3-1).

Given such an LU-representation of the matrix A, the system $AX = B$ can be represented in the form $LUX = B$, and its solution can be obtained by solving the following two systems:

$$LY = B, \quad UX = Y.$$

Due to the triangular structure of the matrices $L \equiv [l_{ij}]$ and $U \equiv [u_{ij}]$, these systems can be solved with the help of the formulas

$$y_i = \frac{1}{l_{ii}}\Big(b_i - \sum_{j=i}^{i-n} l_{ij}y_j\Big) \quad (i = 1, 2, \ldots, n),$$

$$x_k = y_k - \sum_{s=k+1}^{n} u_{ks}x_s \quad (k = n, n-1, \ldots, 1),$$

provided that $l_{ii} \neq 0$.

There exist various methods for the construction of LU-decompositions. In particular if the following conditions hold:

$$a_{11} \neq 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0, \quad \ldots, \quad \det A \neq 0,$$

then the elements of the desired matrices L and U can be calculated by the formulas

$$l_{ij} = \begin{cases} a_{ij} - \sum\limits_{s=1}^{j-1} l_{is} u_{sj} & \text{for } i \geq j, \\ 0 & \text{for } i < j, \end{cases}$$

$$u_{ij} = \begin{cases} \dfrac{1}{l_{ii}} \Big(a_{ij} - \sum\limits_{s=1}^{i-1} l_{is} u_{sj} \Big) & \text{for } i < j \\ 1 & \text{for } i = j, \\ 0 & \text{for } i > j. \end{cases}$$

5.5.2-3. Solutions of a square system with different right-hand sides.

1°. One often has to solve a system of linear equations with a given basic matrix A and different right-hand sides. For instance, consider the systems $AX^{(1)} = B^{(1)}$, …, $AX^{(m)} = B^{(m)}$. These m systems can be regarded as a single matrix equation $AX = B$, where X and B are matrices of size $n \times m$ whose columns coincide with $X^{(j)}$ and $B^{(j)}$ ($j = 1, 2, \ldots, m$).

Example 5. Suppose that we have to solve the equation $AX = B$ with the given basic matrix A and different right-hand sides:

$$A = \begin{pmatrix} 1 & 2 & -3 \\ 3 & -2 & 1 \\ -2 & 1 & 3 \end{pmatrix}, \quad B^{(1)} = \begin{pmatrix} 7 \\ 1 \\ 5 \end{pmatrix}, \quad B^{(2)} = \begin{pmatrix} 10 \\ 6 \\ -5 \end{pmatrix}.$$

Using the Gauss-Jordan procedure, we obtain

$$\left(\begin{array}{ccc|cc} 1 & 2 & -3 & 7 & 10 \\ 3 & -2 & 1 & 1 & 6 \\ -2 & 1 & 3 & 5 & -5 \end{array}\right) \to \left(\begin{array}{ccc|cc} 1 & 2 & -3 & 7 & 10 \\ 0 & -8 & 10 & -20 & -24 \\ 0 & 5 & -3 & 19 & 15 \end{array}\right) \to \left(\begin{array}{ccc|cc} 1 & 0 & -1/2 & 2 & 4 \\ 0 & 1 & -5/4 & 5/2 & 3 \\ 0 & 0 & 13/4 & 1/32 & 0 \end{array}\right) \to \left(\begin{array}{ccc|cc} 1 & 0 & 0 & 3 & 4 \\ 0 & 1 & 0 & 5 & 3 \\ 0 & 0 & 1 & 2 & 0 \end{array}\right).$$

Therefore,

$$X^{(1)} = \begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix}, \quad X^{(2)} = \begin{pmatrix} 4 \\ 3 \\ 0 \end{pmatrix}.$$

2°. If $B = I$, where I is the identity matrix of size $n \times n$, then the solution of the matrix equation $AX = I$ coincides with the matrix $X = A^{-1}$.

Example 6. Find the inverse of the matrix

$$A = \begin{pmatrix} 2 & 1 & 0 \\ -3 & 0 & 7 \\ -5 & 4 & 1 \end{pmatrix}.$$

Let us transform the augmented matrix of the system, using the Gauss-Jordan method. We get

$$\left(\begin{array}{ccc|ccc} 2 & 1 & 0 & 1 & 0 & 0 \\ -3 & 0 & 7 & 0 & 1 & 0 \\ -5 & 4 & -1 & 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 3/2 & 7 & 3/2 & 1 & 0 \\ 0 & 13/2 & -1 & 5/2 & 0 & 1 \end{array}\right) \to$$

$$\to \left(\begin{array}{ccc|ccc} 1 & 0 & -7/3 & 0 & -1/3 & 0 \\ 0 & 1 & 14/3 & 1 & 2/3 & 0 \\ 0 & 0 & -94/3 & -4 & -13/3 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 14/47 & -1/94 & -7/94 \\ 0 & 1 & 0 & 19/47 & 1/47 & 7/47 \\ 0 & 0 & 1 & 6/47 & 13/94 & -3/94 \end{array}\right).$$

5.5.2-4. General system of m linear equations with n unknown quantities.

Suppose that system (5.5.1.1) is consistent and its basic matrix A has rank r. First, in the matrix A, one finds a submatrix of size $r \times r$ with nonzero rth-order determinant and drops the $m - r$ equations whose coefficients do not belong to this submatrix (the dropped equations follow from the remaining ones and can, therefore, be neglected). In the remaining equations, the $n - r$ unknown quantities (free unknown quantities) that are not involved in the said submatrix should be transferred to the right-hand sides. Thus, one obtains a system of r equations with r unknown quantities, which can be solved by any of the methods described in Paragraph 5.5.2-2.

Remark. If the rank r of the basic matrix and the rank of the augmented matrix of system (5.5.1.1) are equal to the number of the unknown quantities n, then the system has a unique solution.

5.5.2-5. Solutions of homogeneous and corresponding nonhomogeneous systems.

1°. Suppose that the basic matrix A of the homogeneous system (5.5.1.3) has rank r and its submatrix in the left top corner, $B = [a_{ij}]$ $(i, j = 1, \ldots, r)$, is nondegenerate. Let $M = \det B \neq 0$ be the determinant of that submatrix. Any solution $x_1, \ldots, x_n$ has $n - r$ free components $x_{r+1}, \ldots, x_n$ and its first components $x_1, \ldots, x_r$ are expressed via the free components as follows:

$$
\begin{aligned}
x_1 &= -\frac{1}{M}[x_{r+1}M_1(a_{i(r+1)}) + x_{r+2}M_1(a_{i(r+2)}) + \cdots + x_nM_1(a_{in})],\\
x_2 &= -\frac{1}{M}[x_{r+1}M_2(a_{i(r+1)}) + x_{r+2}M_2(a_{i(r+2)}) + \cdots + x_nM_2(a_{in})],\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
x_r &= -\frac{1}{M}[x_{r+1}M_r(a_{i(r+1)}) + x_{r+2}M_r(a_{i(r+2)}) + \cdots + x_nM_r(a_{in})],
\end{aligned}
\qquad (5.5.2.4)
$$

where $M_j(a_{ik})$ is the determinant of the matrix obtained from B by replacing its jth column with the column whose components are $a_{1k}, a_{2k}, \ldots, a_{rk}$:

$$
M_j(a_{ik}) = \begin{vmatrix}
a_{11} & a_{12} & \ldots & a_{1k} & \ldots & a_{1r}\\
a_{21} & a_{22} & \ldots & a_{2k} & \ldots & a_{2r}\\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots\\
a_{r1} & a_{r2} & \ldots & a_{rk} & \ldots & a_{rr}
\end{vmatrix}.
$$

2°. Using (5.5.2.4), we obtain the following $n - r$ linearly independent solutions of the original system (5.5.1.3):

$$
\begin{aligned}
X_1 &= \left(-\frac{M_1(a_{i(r+1)})}{M} \quad -\frac{M_2(a_{i(r+1)})}{M} \quad \cdots \quad -\frac{M_r(a_{i(r+1)})}{M} \quad 1 \quad 0 \quad \cdots \quad 0\right),\\
X_2 &= \left(-\frac{M_1(a_{i(r+2)})}{M} \quad -\frac{M_2(a_{i(r+2)})}{M} \quad \cdots \quad -\frac{M_r(a_{i(r+2)})}{M} \quad 0 \quad 1 \quad \cdots \quad 0\right),\\
X_{n-r} &= \left(-\frac{M_1(a_{in})}{M} \quad -\frac{M_2(a_{in})}{M} \quad \cdots \quad -\frac{M_r(a_{in})}{M} \quad 0 \quad 0 \quad \cdots \quad 1\right).
\end{aligned}
$$

Any solution of system (5.5.1.3) can be represented as their linear combination

$$
X = C_1X_1 + C_2X_2 + \cdots + C_{n-r}X_{n-r}, \qquad (5.5.2.5)
$$

where $C_1, C_2, \ldots, C_{n-r}$ are arbitrary constants. This formula gives the *general solution* of the homogeneous system.

3°. Relations between solutions of the nonhomogeneous system (5.5.1.1) and solutions of the corresponding homogeneous system (5.5.1.3).

1. The sum of any solution of the nonhomogeneous system (5.5.1.1) and any solution of the corresponding homogeneous system (5.5.1.3) is a solution of system (5.5.1.1).
2. The difference of any two solutions of the nonhomogeneous system (5.5.1.1) is a solution of the homogeneous system (5.5.1.3).
3. The sum of a particular solution X_0 of the nonhomogeneous system (5.5.1.1) and the general solution (5.5.2.5) of the corresponding homogeneous system (5.5.1.3) yields the general solution X of the nonhomogeneous system (5.5.1.1).

5.6. Linear Operators

5.6.1. Notion of a Linear Operator. Its Properties

5.6.1-1. Definition of a linear operator.

An *operator* $\mathbf{A}$ acting from a linear space $\mathcal{V}$ of dimension n to a linear space $\mathcal{W}$ of dimension m is a mapping $\mathbf{A} : \mathcal{V} \to \mathcal{W}$ that establishes correspondence between each element $\mathbf{x}$ of the space $\mathcal{V}$ and some element $\mathbf{y}$ of the space $\mathcal{W}$. This fact is denoted by $\mathbf{y} = \mathbf{Ax}$ or $\mathbf{y} = \mathbf{A}(\mathbf{x})$.

An operator $\mathbf{A} : \mathcal{V} \to \mathcal{W}$ is said to be *linear* if for any elements $\mathbf{x}_1$ and $\mathbf{x}_2$ of the space $\mathcal{V}$ and any scalar λ, the following relations hold:

$$\mathbf{A}(\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{A}\mathbf{x}_1 + \mathbf{A}\mathbf{x}_2 \quad \text{(additivity of the operator)},$$
$$\mathbf{A}(\lambda\mathbf{x}) = \lambda\mathbf{A}\mathbf{x} \quad \text{(homogeneity of the operator)}.$$

A linear operator $\mathbf{A} : \mathcal{V} \to \mathcal{W}$ is said to be *bounded* if it has a finite *norm*, which is defined as follows:

$$\|\mathbf{A}\| = \sup_{\substack{\mathbf{x}\in V \\ \|\mathbf{x}\|\neq 0}} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} = \sup_{\|\mathbf{x}\|=1} \|\mathbf{Ax}\| \geq 0.$$

Remark. If $\mathbf{A}$ is a linear operator from a Hilbert space $\mathcal{V}$ into itself, then

$$\|\mathbf{A}\| = \sup_{\substack{\mathbf{x}\in V \\ \|\mathbf{x}\|\neq 0}} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} = \sup_{\|\mathbf{x}\|=1} \|\mathbf{Ax}\| = \sup_{\mathbf{x},\mathbf{y}\neq 0} \frac{|(\mathbf{x}, \mathbf{Ay})|}{\|\mathbf{x}\|\,\|\mathbf{y}\|} = \sup_{\|\mathbf{x}\|=\|\mathbf{y}\|=1} |(\mathbf{x}, \mathbf{Ay})|.$$

THEOREM. *Any linear operator in a finite-dimensional normed space is bounded.*

The set of all linear operators $\mathbf{A} : \mathcal{V} \to \mathcal{W}$ is denoted by $L(\mathcal{V}, \mathcal{W})$.

A linear operator $\mathbf{O}$ in $L(\mathcal{V}, \mathcal{W})$ is called the *zero operator* if it maps any element $\mathbf{x}$ of $\mathcal{V}$ to the zero element of the space $\mathcal{W}$: $\mathbf{Ox} = 0$.

A linear operator $\mathbf{A}$ in $L(\mathcal{V}, \mathcal{V})$ is also called a *linear transformation of the space* $\mathcal{V}$.

A linear operator $\mathbf{I}$ in $L(\mathcal{V}, \mathcal{V})$ is called the *identity operator* if it maps each element $\mathbf{x}$ of $\mathcal{V}$ into itself: $\mathbf{Ix} = \mathbf{x}$.

5.6.1-2. Basic operations with linear operators.

The *sum* of two linear operators $\mathbf{A}$ and $\mathbf{B}$ in $L(\mathcal{V}, \mathcal{W})$ is a linear operator denoted by $\mathbf{A} + \mathbf{B}$ and defined by

$$(\mathbf{A} + \mathbf{B})\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{x} \qquad \text{for any} \quad \mathbf{x} \in \mathcal{V}.$$

The *product* of a scalar λ and a linear operator $\mathbf{A}$ in $L(\mathcal{V}, \mathcal{W})$ is a linear operator denoted by $\lambda\mathbf{A}$ and defined by

$$(\lambda\mathbf{A})\mathbf{x} = \lambda\mathbf{A}\mathbf{x} \qquad \text{for any} \quad \mathbf{x} \in \mathcal{V}.$$

The *opposite* operator for an operator $\mathbf{A} \in L(\mathcal{V}, \mathcal{W})$ is an operator denoted by $-\mathbf{A}$ and defined by

$$-\mathbf{A} = (-1)\mathbf{A}.$$

The *product of two linear operators* $\mathbf{A}$ *and* $\mathbf{B}$ in $L(\mathcal{V}, \mathcal{V})$ is a linear operator denoted by $\mathbf{AB}$ and defined by

$$(\mathbf{AB})\mathbf{x} = \mathbf{A}(\mathbf{Bx}) \qquad \text{for any} \quad \mathbf{x} \in \mathcal{V}.$$

Properties of linear operators in $L(\mathcal{V}, \mathcal{V})$:

$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$ (associativity of the product of three operators),

$\lambda(\mathbf{AB}) = (\lambda\mathbf{A})\mathbf{B}$ (associativity of multiplication of a scalar and two operators),

$(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$ (distributivity with respect to the sum of operators),

where λ is a scalar; $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are linear operators in $L(\mathcal{V}, \mathcal{V})$.

Remark. Property 1 allows us to define the product $\mathbf{A}_1\mathbf{A}_2 \dots \mathbf{A}_k$ of finitely many operators in $L(\mathcal{V}, \mathcal{V})$ and the kth power of an operator $\mathbf{A}$,

$$\mathbf{A}^k = \underbrace{\mathbf{AA} \dots \mathbf{A}}_{k \text{ times}}.$$

The following relations hold:

$$\mathbf{A}^{p+q} = \mathbf{A}^p\mathbf{A}^q, \quad (\mathbf{A}^p)^q = \mathbf{A}^{pq}. \tag{5.6.1.1}$$

5.6.1-3. Inverse operators.

A linear operator $\mathbf{B}$ is called the inverse of an operator $\mathbf{A}$ in $L(\mathcal{V}, \mathcal{V})$ if $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$. The *inverse* operator is denoted by $\mathbf{B} = \mathbf{A}^{-1}$. If the inverse operator exists, the operator $\mathbf{A}$ is said to be *invertible* or *nondegenerate*.

Remark. If $\mathbf{A}$ is an invertible operator, then $\mathbf{A}^{-k} = (\mathbf{A}^{-1})^k = (\mathbf{A}^k)^{-1}$ and relations (5.6.1.1) still hold.

A linear operator $\mathbf{A}$ from $\mathcal{V}$ to $\mathcal{W}$ is said to be *injective* if it maps any two different elements of $\mathcal{V}$ into different elements of $\mathcal{W}$, i.e., for $\mathbf{x}_1 \neq \mathbf{x}_2$, we have $\mathbf{Ax}_1 \neq \mathbf{Ax}_2$.

If $\mathbf{A}$ is an injective linear operator from $\mathcal{V}$ to $\mathcal{V}$, then each element $\mathbf{y} \in \mathcal{V}$ is an image of some element $\mathbf{x} \in \mathcal{V}$: $\mathbf{y} = \mathbf{Ax}$.

THEOREM. *A linear operator $\mathbf{A} : \mathcal{V} \to \mathcal{V}$ is invertible if and only if it is injective.*

5.6.1-4. Kernel, range, and rank of a linear operator.

The *kernel* of a linear operator $\mathbf{A} : \mathcal{V} \to \mathcal{V}$ is the set of all $\mathbf{x}$ in $\mathcal{V}$ such that $\mathbf{Ax} = 0$. The kernel of an operator $\mathbf{A}$ is denoted by $\ker \mathbf{A}$ and is a linear subspace of $\mathcal{V}$.

The *range* of a linear operator $\mathbf{A} : \mathcal{V} \to \mathcal{V}$ is the set of all $\mathbf{y}$ in $\mathcal{V}$ such that $\mathbf{y} = \mathbf{Ax}$. The range of a linear operator $\mathbf{A}$ is denoted by $\operatorname{im} \mathbf{A}$ and is a subspace of $\mathcal{V}$.

Properties of the kernel, the range, and their dimensions:

1. For a linear operator $\mathbf{A} : \mathcal{V} \to \mathcal{V}$ in n-dimensional space $\mathcal{V}$, the following relation holds:
$$\dim(\operatorname{im} \mathbf{A}) + \dim(\ker \mathbf{A}) = n.$$
2. Let $\mathcal{V}_1$ and $\mathcal{V}_2$ be two subspaces of a linear space $\mathcal{V}$ and $\dim \mathcal{V}_1 + \dim \mathcal{V}_2 = \dim \mathcal{V}$. Then there exists a linear operator $\mathbf{A} : \mathcal{V} \to \mathcal{V}$ such that $\mathcal{V}_1 = \operatorname{im} \mathbf{A}$ and $\mathcal{V}_2 = \ker \mathbf{A}$.

A subspace $\mathcal{V}_1$ of the space $\mathcal{V}$ is called an *invariant subspace* of a linear operator $\mathbf{A} : \mathcal{V} \to \mathcal{V}$ if for any $\mathbf{x}$ in $\mathcal{V}_1$, the element $\mathbf{Ax}$ also belongs to $\mathcal{V}_1$. A linear operator $\mathbf{A} : \mathcal{V} \to \mathcal{V}$ is said to be *reducible* if $\mathcal{V}$ can be represented as a direct sum $\mathcal{V} = \mathcal{V}_1 \oplus \dots \oplus \mathcal{V}_N$ of two or more invariant subspaces $\mathcal{V}_1, \dots, \mathcal{V}_N$ of the operator $\mathbf{A}$, where N is a natural number.

Example 1. $\ker \mathbf{A}$ and $\operatorname{im} \mathbf{A}$ are invariant subspaces of any linear operator $\mathbf{A} : \mathcal{V} \to \mathcal{V}$.

The *rank* of a linear operator $\mathbf{A}$ is the dimension of its range: $\operatorname{rank}(\mathbf{A}) = \dim(\operatorname{im}\mathbf{A})$.
Properties of the rank of a linear operator:

$$\operatorname{rank}(\mathbf{AB}) \le \min\{\operatorname{rank}(\mathbf{A}), \operatorname{rank}(\mathbf{B})\},$$
$$\operatorname{rank}(\mathbf{A}) + \operatorname{rank}(\mathbf{B}) - n \le \operatorname{rank}(\mathbf{AB}),$$

where $\mathbf{A}$ and $\mathbf{B}$ are linear operators in $L(\mathcal{V}, \mathcal{V})$ and $n = \dim \mathcal{V}$.

Remark. If $\operatorname{rank}(\mathbf{A}) = n$ then $\operatorname{rank}(\mathbf{AB}) = \operatorname{rank}(\mathbf{BA}) = \operatorname{rank}(\mathbf{B})$.

THEOREM. *Let $\mathbf{A} : \mathcal{V} \to \mathcal{V}$ be a linear operator. Then the following statements are equivalent:*

1. $\mathbf{A}$ *is invertible (i.e., there exists* $\mathbf{A}^{-1}$*).*
2. $\ker \mathbf{A} = 0$.
3. $\operatorname{im}\mathbf{A} = \mathcal{V}$.
4. $\operatorname{rank}(\mathbf{A}) = \dim \mathcal{V}$.

5.6.1-5. Notion of a adjoint operator. Hermitian operators.

Let $\mathbf{A} \in L(\mathcal{V}, \mathcal{V})$ be a bounded linear operator in a Hilbert space $\mathcal{V}$. The operator $\mathbf{A}^*$ in $L(\mathcal{V}, \mathcal{V})$ is called its *adjoint operator* if

$$(\mathbf{Ax}) \cdot \mathbf{y} = \mathbf{x} \cdot (\mathbf{A}^*\mathbf{y})$$

for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathcal{V}$.

THEOREM. *Any bounded linear operator $\mathbf{A}$ in a Hilbert space has a unique adjoint operator.*

Properties of adjoint operators:

$$(\mathbf{A}+\mathbf{B})^* = \mathbf{A}^* + \mathbf{B}^*, \quad (\lambda\mathbf{A})^* = \bar{\lambda}\mathbf{A}^*, \quad (\mathbf{A}^*)^* = \mathbf{A},$$
$$(\mathbf{AB})^* = \mathbf{B}^*\mathbf{A}^*, \quad \mathbf{O}^* = \mathbf{O}, \quad \mathbf{I}^* = \mathbf{I},$$
$$(\mathbf{A}^{-1})^* = (\mathbf{A}^*)^{-1}, \quad \|\mathbf{A}^*\| = \|\mathbf{A}\|, \quad \|\mathbf{A}^*\mathbf{A}\| = \|\mathbf{A}\|^2,$$
$$(\mathbf{Ax}) \cdot (\mathbf{By}) \equiv \mathbf{x} \cdot (\mathbf{A}^*\mathbf{By}) \equiv (\mathbf{B}^*\mathbf{Ax}) \cdot \mathbf{y} \quad \text{for all } \mathbf{x} \text{ and } \mathbf{y} \text{ in } \mathcal{V},$$

where $\mathbf{A}$ and $\mathbf{B}$ are bounded linear operators in a Hilbert space $\mathcal{V}$, $\bar{\lambda}$ is the complex conjugate of a number λ.

A linear operator $\mathbf{A} \in L(\mathcal{V}, \mathcal{V})$ in a Hilbert space $\mathcal{V}$ is said to be *Hermitian (self-adjoint)* if

$$\mathbf{A}^* = \mathbf{A} \quad \text{or} \quad (\mathbf{Ax}) \cdot \mathbf{y} = \mathbf{x} \cdot (\mathbf{Ay}).$$

A linear operator $\mathbf{A} \in (\mathcal{V}, \mathcal{V})$ in a Hilbert space $\mathcal{V}$ is said to be *skew-Hermitian* if

$$\mathbf{A}^* = -\mathbf{A} \quad \text{or} \quad (\mathbf{Ax}) \cdot \mathbf{y} = -\mathbf{x} \cdot (\mathbf{Ay}).$$

5.6.1-6. Unitary and normal operators.

A linear operator $\mathbf{U} \in L(\mathcal{V}, \mathcal{V})$ in a Hilbert space $\mathcal{V}$ is called a *unitary operator* if for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathcal{V}$, the following relation holds:

$$(\mathbf{Ux}) \cdot (\mathbf{Uy}) = \mathbf{x} \cdot \mathbf{y}.$$

This relation is called the *unitarity condition.*

Properties of a unitary operator $\mathbf{U}$:

$$\mathbf{U}^* = \mathbf{U}^{-1} \quad \text{or} \quad \mathbf{U}^*\mathbf{U} = \mathbf{U}\mathbf{U}^* = \mathbf{I},$$
$$\|\mathbf{U}\mathbf{x}\| = \|\mathbf{x}\| \quad \text{for all } \mathbf{x} \text{ in } \mathcal{V}.$$

A linear operator $\mathbf{A}$ in $L(\mathcal{V}, \mathcal{V})$ is said to be *normal* if

$$\mathbf{A}^*\mathbf{A} = \mathbf{A}\mathbf{A}^*.$$

THEOREM. *A bounded linear operator* $\mathbf{A}$ *is normal if and only if* $\|\mathbf{A}\mathbf{x}\| = \|\mathbf{A}\|\,\|\mathbf{x}\|$.

Remark. Any unitary or Hermitian operator is normal.

5.6.1-7. Transpose, symmetric, and orthogonal operators.

The *transpose operator* of a bounded linear operator $\mathbf{A} \in L(\mathcal{V}, \mathcal{V})$ in a real Hilbert space $\mathcal{V}$ is the operator $\mathbf{A}^T \in L(\mathcal{V}, \mathcal{V})$ such that for all $\mathbf{x}$, $\mathbf{y}$ in $\mathcal{V}$, the following relation holds:

$$(\mathbf{A}\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (\mathbf{A}^T\mathbf{y}).$$

THEOREM. *Any bounded linear operator* $\mathbf{A}$ *in a real Hilbert space has a unique transpose operator.*

The properties of transpose operators in a real Hilbert space are similar to the properties of adjoint operators considered in Paragraph 5.6.1-5 if one takes $\mathbf{A}^T$ instead of $\mathbf{A}^*$.

A linear operator $\mathbf{A} \in L(\mathcal{V}, \mathcal{V})$ in a real Hilbert space $\mathcal{V}$ is said to be *symmetric* if

$$\mathbf{A}^T = \mathbf{A} \quad \text{or} \quad (\mathbf{A}\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (\mathbf{A}\mathbf{y}).$$

A linear operator $\mathbf{A} \in L(\mathcal{V}, \mathcal{V})$ in a real Hilbert space $\mathcal{V}$ is said to be *skew-symmetric* if

$$\mathbf{A}^T = -\mathbf{A} \quad \text{or} \quad (\mathbf{A}\mathbf{x}) \cdot \mathbf{y} = -\mathbf{x} \cdot (\mathbf{A}\mathbf{y}).$$

The properties of symmetric linear operators in a real Hilbert space are similar to the properties of Hermitian operators considered in Paragraph 5.6.1-5 if one takes $\mathbf{A}^T$ instead of $\mathbf{A}^*$.

A linear operator $\mathbf{P} \in L(\mathcal{V}, \mathcal{V})$ in a real Hilbert space $\mathcal{V}$ is said to be *orthogonal* if for any $\mathbf{x}$ and $\mathbf{y}$ in $\mathcal{V}$, the following relations hold:

$$(\mathbf{P}\mathbf{x}) \cdot (\mathbf{P}\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}.$$

This relation is called the *orthogonality condition*.

Properties of orthogonal operator $\mathbf{P}$:

$$\mathbf{P}^T = \mathbf{P}^{-1} \quad \text{or} \quad \mathbf{P}^T\mathbf{P} = \mathbf{P}\mathbf{P}^T = \mathbf{I},$$
$$\|\mathbf{P}\mathbf{x}\| = \|\mathbf{x}\| \quad \text{for all } \mathbf{x} \text{ in } \mathcal{V}.$$

5.6.1-8. Positive operators. Roots of an operator.

A Hermitian (symmetric, in the case of a real space) operator $\mathbf{A}$ is said to be

a) *nonnegative* (resp., *nonpositive*), and one writes $\mathbf{A} \geq 0$ (resp., $\mathbf{A} \leq 0$) if $(\mathbf{A}\mathbf{x}) \cdot \mathbf{x} \geq 0$ (resp., $(\mathbf{A}\mathbf{x}) \cdot \mathbf{x} \leq 0$) for any $\mathbf{x}$ in $\mathcal{V}$.

b) *positive* or *positive definite* (resp., *negative* or *negative definite*) and one writes $\mathbf{A} > 0$ ($\mathbf{A} < 0$) if $(\mathbf{A}\mathbf{x}) \cdot \mathbf{x} > 0$ (resp., $(\mathbf{A}\mathbf{x}) \cdot \mathbf{x} < 0$) for any $\mathbf{x} \neq 0$.

An *mth root* of an operator $\mathbf{A}$ is an operator $\mathbf{B}$ such that $\mathbf{B}^m = \mathbf{A}$.

THEOREM. *If* $\mathbf{A}$ *is a nonnegative Hermitian (symmetric) operator, then for any positive integer* m *there exists a unique nonnegative Hermitian (symmetric) operator* $\mathbf{A}^{1/m}$.

5.6.1-9. Decomposition theorems.

THEOREM 1. *For any bounded linear operator* $\mathbf{A}$ *in a Hilbert space* $\mathcal{V}$*, the operator* $\mathbf{H}_1 = \frac{1}{2}(\mathbf{A} + \mathbf{A}^*)$ *is Hermitian and the operator* $\mathbf{H}_2 = \frac{1}{2}(\mathbf{A} - \mathbf{A}^*)$ *is skew-Hermitian. The representation of* $\mathbf{A}$ *as a sum of Hermitian and skew-Hermitian operators is unique:* $\mathbf{A} = \mathbf{H}_1 + \mathbf{H}_2$.

THEOREM 2. *For any bounded linear operator* $\mathbf{A}$ *in a real Hilbert space, the operator* $\mathbf{S}_1 = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T)$ *is symmetric and the operator* $\mathbf{S}_2 = \frac{1}{2}(\mathbf{A} - \mathbf{A}^T)$ *is skew-symmetric. The representation of* $\mathbf{A}$ *as a sum of symmetric and skew-symmetric operators is unique:* $\mathbf{A} = \mathbf{S}_1 + \mathbf{S}_2$.

THEOREM 3. *For any bounded linear operator* $\mathbf{A}$ *in a Hilbert space,* $\mathbf{A}\mathbf{A}^*$ *and* $\mathbf{A}^*\mathbf{A}$ *are nonnegative Hermitian operators.*

THEOREM 4. *For any linear operator* $\mathbf{A}$ *in a Hilbert space* $\mathcal{V}$*, there exist polar decompositions*

$$\mathbf{A} = \mathbf{Q}\mathbf{U} \quad \text{and} \quad \mathbf{A} = \mathbf{U}_1\mathbf{Q}_1,$$

where $\mathbf{Q}$ *and* $\mathbf{Q}_1$ *are nonnegative Hermitian operators,* $\mathbf{Q}^2 = \mathbf{A}\mathbf{A}^*$, $\mathbf{Q}_1^2 = \mathbf{A}^*\mathbf{A}$, *and* $\mathbf{U}$, $\mathbf{U}_1$ *are unitary operators. The operators* $\mathbf{Q}$ *and* $\mathbf{Q}_1$ *are always unique, while the operators* $\mathbf{U}$ *and* $\mathbf{U}_1$ *are unique only if* $\mathbf{A}$ *is nondegenerate.*

5.6.2. Linear Operators in Matrix Form

5.6.2-1. Matrices associated with linear operators.

Let $\mathbf{A}$ be a linear operator in an n-dimensional linear space $\mathcal{V}$ with a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Then there is a matrix $[a_j^i]$ such that

$$\mathbf{A}\mathbf{e}_j = \sum_{i=1}^{n} a_j^i \mathbf{e}_i.$$

The coordinates y^j of the vector $\mathbf{y} = \mathbf{A}\mathbf{x}$ in that basis can be represented in the form

$$y^i = \sum_{j=1}^{n} a_j^i x^j \quad (i = 1, 2, \ldots, n), \tag{5.6.2.1}$$

where x^j are the coordinates of $\mathbf{x}$ in the same basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. The matrix $A \equiv [a_j^i]$ of size $n \times n$ is called the *matrix of the linear operator* $\mathbf{A}$ in a given basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$.

Thus, given a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$, any linear operator $\mathbf{y} = \mathbf{A}\mathbf{x}$ can be associated with its matrix in that basis with the help of (5.6.2.1).

If $\mathbf{A}$ is the zero operator, then its matrix is the zero matrix in any basis. If $\mathbf{A}$ is the unit operator, then its matrix is the unit matrix in any basis.

THEOREM 1. *Let* $\mathbf{e}_1, \ldots, \mathbf{e}_n$ *be a given basis in a linear space* $\mathcal{V}$ *and let* $A \equiv [a_j^i]$ *be a given square matrix of size* $n \times n$*. Then there exists a unique linear operator* $\mathbf{A} : \mathcal{V} \to \mathcal{V}$ *whose matrix in that basis coincides with the matrix* A.

THEOREM 2. *The rank of a linear operator* $\mathbf{A}$ *is equal to the rank of its matrix* A *in any basis:* $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(A)$.

THEOREM 3. *A linear operator* $\mathbf{A} : \mathcal{V} \to \mathcal{V}$ *is invertible if and only if* $\operatorname{rank}(A) = \dim V$. *In this case, the matrix of the operator* $\mathbf{A}$ *is invertible.*

5.6.2-2. Transformation of the matrix of a linear operator.

Suppose that the transition from the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ to another basis $\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_n$ is determined by a matrix $U \equiv [u_{ij}]$ of size $n \times n$, i.e.

$$\widetilde{\mathbf{e}}_i = \sum_{j=1}^{n} u_{ij}\mathbf{e}_j \qquad (i = 1, 2, \ldots, n).$$

THEOREM. *Let A and $\widetilde{A}$ be the matrices of a linear operator $\mathbf{A}$ in the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and the basis $\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_n$, respectively. Then*

$$A = U^{-1}\widetilde{A}U \quad \text{or} \quad \widetilde{A} = UAU^{-1}.$$

Note that the determinant of the matrix of a linear operator does not depend on the basis: $\det A = \det \widetilde{A}$. Therefore, one can correctly define the *determinant* $\det \mathbf{A}$ *of a linear operator* as the determinant of its matrix in any basis:

$$\det \mathbf{A} = \det A.$$

The trace of the matrix of a linear operator, $\mathrm{Tr}(A)$, is also independent of the basis. Therefore, one can correctly define the *trace* $\mathrm{Tr}(\mathbf{A})$ *of a linear operator* as the trace of its matrix in any basis:

$$\mathrm{Tr}(\mathbf{A}) = \mathrm{Tr}(A).$$

In the case of an orthonormal basis, a Hermitian, skew-Hermitian, normal, or unitary operator in a Hilbert space corresponds to a Hermitian, skew-Hermitian, normal, or unitary matrix; and a symmetric, skew-symmetric, or transpose operator in a real Hilbert space corresponds to a symmetric, skew-symmetric, or transpose matrix.

5.6.3. Eigenvectors and Eigenvalues of Linear Operators

5.6.3-1. Basic definitions.

1°. A scalar λ is called an *eigenvalue* of a linear operator $\mathbf{A}$ in a vector space $\mathcal{V}$ if there is a nonzero element $\mathbf{x}$ in $\mathcal{V}$ such that

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}. \tag{5.6.3.1}$$

A nonzero element $\mathbf{x}$ for which (5.6.3.1) holds is called an *eigenvector* of the operator $\mathbf{A}$ corresponding to the eigenvalue λ. Eigenvectors corresponding to distinct eigenvalues are linearly independent. For an eigenvalue $\lambda \neq 0$, the inverse $\mu = 1/\lambda$ is called a *characteristic value* of the operator A.

THEOREM. *If $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are eigenvectors of an operator $\mathbf{A}$ corresponding to its eigenvalue λ, then $\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k$ ($\alpha_1^2 + \cdots + \alpha_k^2 \neq 0$) is also an eigenvector of the operator $\mathbf{A}$ corresponding to the eigenvalue λ.*

The *geometric multiplicity* m_i of an eigenvalue λ_i is the maximal number of linearly independent eigenvectors corresponding to the eigenvalue λ_i. Thus, the geometric multiplicity of λ_i is the dimension of the subspace formed by all eigenvectors corresponding to the eigenvalue λ_i.

The *algebraic multiplicity* m_i' of an eigenvalue λ_i of an operator $\mathbf{A}$ is equal to the algebraic multiplicity of λ_i regarded as an eigenvalue of the corresponding matrix A.

The algebraic multiplicity m_i' of an eigenvalue λ_i is always not less than the geometric multiplicity m_i of this eigenvalue.

The trace $\mathrm{Tr}(\mathbf{A})$ is equal to the sum of all eigenvalues of the operator $\mathbf{A}$, each eigenvalue counted according to its multiplicity, i.e.,

$$\mathrm{Tr}(\mathbf{A}) = \sum_i m_i' \lambda_i.$$

The determinant $\det \mathbf{A}$ is equal to the product of all eigenvalues of the operator $\mathbf{A}$, each eigenvalue entering the product according to its multiplicity,

$$\det \mathbf{A} = \prod_i \lambda_i^{m_i'}.$$

5.6.3-2. Eigenvectors and eigenvalues of normal and Hermitian operators.

Properties of eigenvalues and eigenvectors of a normal operator:

1. A normal operator $\mathbf{A}$ in a Hilbert space $\mathcal{V}$ and its adjoint operator $\mathbf{A}^*$ have the same eigenvectors and their eigenvalues are complex conjugate.
2. For a normal operator $\mathbf{A}$ in a Hilbert space $\mathcal{V}$, there is a basis $\{\mathbf{e}_k\}$ formed by eigenvectors of the operators $\mathbf{A}$ and $\mathbf{A}^*$. Therefore, there is a basis in $\mathcal{V}$ in which the operator $\mathbf{A}$ has a diagonal matrix.
3. Eigenvectors corresponding to distinct eigenvalues of a normal operator are mutually orthogonal.
4. Any bounded normal operator $\mathbf{A}$ in a Hilbert space $\mathcal{V}$ is reducible. The space $\mathcal{V}$ can be represented as a direct sum of the subspace spanned by an orthonormal system of eigenvectors of $\mathbf{A}$ and the subspace consisting of vectors orthogonal to all eigenvectors of $\mathbf{A}$. In the finite-dimensional case, an orthonormal system of eigenvectors of $\mathbf{A}$ is a basis of $\mathcal{V}$.
5. The algebraic multiplicity of any eigenvalue λ of a normal operator is equal to its geometric multiplicity.

Properties of eigenvalues and eigenvectors of a Hermitian operator:

1. Since any Hermitian operator is normal, all properties of normal operators hold for Hermitian operators.
2. All eigenvalues of a Hermitian operator are real.
3. Any Hermitian operator $\mathbf{A}$ in an n-dimensional unitary space has n mutually orthogonal eigenvectors of unit length.
4. Any eigenvalue of a nonnegative (positive) operator is nonnegative (positive).
5. *Minimax property.* Let $\mathbf{A}$ be a Hermitian operator in an n-dimensional unitary space $\mathcal{V}$, and let $\mathcal{E}_m$ be the set of all m-dimensional subspaces of $\mathcal{V}$ $(m < n)$. Then the eigenvalues $\lambda_1, \ldots, \lambda_n$ of the operator $\mathbf{A}$ $(\lambda_1 \geq \ldots \geq \lambda_n)$ can be defined by the formulas

$$\lambda_{m+1} = \min_{\mathcal{Y} \in \mathcal{E}_m} \max_{\mathbf{x} \perp \mathcal{Y}} \frac{(\mathbf{A}\mathbf{x}) \cdot \mathbf{x}}{\mathbf{x} \cdot \mathbf{x}}.$$

6. Let $\mathbf{i}_1, \ldots, \mathbf{i}_n$ be an orthonormal basis in an n-dimensional space $\mathcal{V}$, and let all $\mathbf{i}_k$ are eigenvectors of a Hermitian operator $\mathbf{A}$, i.e., $\mathbf{A}\mathbf{i}_k = \lambda_k \mathbf{i}_k$. Then the matrix of the operator $\mathbf{A}$ in the basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ is diagonal and its diagonal elements have the form $a_k^k = \lambda_k$.

7. Let $\mathbf{i}_1, \ldots, \mathbf{i}_n$ be an arbitrary orthonormal basis in an n-dimensional Euclidean space $\mathcal{V}$. Then the matrix of an operator $\mathbf{A}$ in the basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ is symmetric if and only if the operator $\mathbf{A}$ is Hermitian.
8. In an orthonormal basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ formed by eigenvectors of a nonnegative Hermitian operator $\mathbf{A}$, the matrix of the operator $\mathbf{A}^{1/m}$ has the form

$$\begin{pmatrix} \lambda_1^{1/m} & 0 & \cdots & 0 \\ 0 & \lambda_2^{1/m} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^{1/m} \end{pmatrix}.$$

5.6.3-3. Characteristic polynomial of a linear operator.

Consider the finite-dimensional case. The algebraic equation

$$f_{\mathbf{A}}(\lambda) \equiv \det(\mathbf{A} - \lambda\mathbf{I}) = 0 \tag{5.6.3.2}$$

of degree n is called the *characteristic equation* of the operator $\mathbf{A}$ and $f_{\mathbf{A}}(\lambda)$ is called the *characteristic polynomial* of the operator $\mathbf{A}$.

Since the value of the determinant $\det(\mathbf{A} - \lambda\mathbf{I})$ does not depend on the basis, the coefficients of λ^k ($k = 0, 1, \ldots, n$) in the characteristic polynomial $f_{\mathbf{A}}(\lambda)$ are *invariants* (i.e., quantities whose values do not depend on the basis). In particular, the coefficient of λ^{k-1} is equal to the trace of the operator $\mathbf{A}$.

In the finite-dimensional case, λ is an eigenvalue of a linear operator $\mathbf{A}$ if and only if λ is a root of the characteristic equation (5.6.3.2) of the operator $\mathbf{A}$. Therefore, a linear operator always has eigenvalues.

In the case of a real space, a root of the characteristic equation can be an eigenvalue of a linear operator only if this root is real. In this connection, it would be natural to find a class of linear operators in a real Euclidean space for which all roots of the corresponding characteristic equations are real.

THEOREM. *The matrix A of a linear operator $\mathbf{A}$ in a given basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ is diagonal if and only if all $\mathbf{i}_i$ are eigenvectors of this operator.*

5.6.3-4. Bounds for eigenvalues of linear operators.

The modulus of any eigenvalue λ of a linear operator $\mathbf{A}$ in an n-dimensional unitary space satisfies the estimate:

$$|\lambda| \le \min(M_1, M_2), \qquad M_1 = \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}|, \quad M_2 = \max_{1\le j\le n} \sum_{i=1}^{n} |a_{ij}|,$$

where $A \equiv [a_{ij}]$ is the matrix of the operator $\mathbf{A}$. The real and the imaginary parts of eigenvalues satisfy the estimates:

$$\min_{1\le i\le n} (\operatorname{Re} a_{ii} - P_i) \le \operatorname{Re}\lambda \le \max_{1\le i\le n} (\operatorname{Re} a_{ii} + P_i),$$
$$\min_{1\le i\le n} (\operatorname{Im} a_{ii} - P_i) \le \operatorname{Im}\lambda \le \max_{1\le i\le n} (\operatorname{Im} a_{ii} + P_i),$$

where $P_i = \sum_{j=1,\ j\neq i}^{n} |a_{ij}|$, and P_i can be replaced by $Q_i = \sum_{j=1,\ i\neq i}^{n} |a_{ji}|$.

The modulus of any eigenvalue λ of a Hermitian operator $\mathbf{A}$ in an n-dimensional unitary space satisfies the inequalities

$$|\lambda|^2 \le \sum_i \sum_j |a_{ij}|^2, \quad |\lambda| \le \|\mathbf{A}\| = \sup_{\|\mathbf{x}\|=1} [(\mathbf{A}\mathbf{x})\cdot\mathbf{x}],$$

and its smallest and its largest eigenvalues, denoted, respectively, by m and M, can be found from the relations

$$m = \inf_{\|\mathbf{x}\|=1} [(\mathbf{A}\mathbf{x})\cdot\mathbf{x}], \quad M = \sup_{\|\mathbf{x}\|=1} [(\mathbf{A}\mathbf{x})\cdot\mathbf{x}].$$

5.6.3-5. Spectral decomposition of Hermitian operators.

Let $\mathbf{i}_1, \ldots, \mathbf{i}_n$ be a fixed orthonormal basis in an n-dimensional unitary space $\mathcal{V}$. Then any element of $\mathcal{V}$ can be represented in the form (see Paragraph 5.4.2-2)

$$\mathbf{x} = \sum_{j=1}^{n} (\mathbf{x}\cdot\mathbf{i}_j)\mathbf{i}_j.$$

The operator $\mathbf{P}_k$ ($k = 1, 2, \ldots, n$) defined by

$$\mathbf{P}_k\mathbf{x} = (\mathbf{x}\cdot\mathbf{i}_k)\mathbf{i}_k$$

is called the *projection* onto the one-dimensional subspace generated by the vector $\mathbf{i}_k$. The projection $\mathbf{P}_k$ is a Hermitian operator.

Properties of the projection $\mathbf{P}_k$:

$$\mathbf{P}_k\mathbf{P}_l = \begin{cases} \mathbf{P}_k & \text{for } k = l, \\ \mathbf{O} & \text{for } k \neq l, \end{cases} \qquad \mathbf{P}_k^m = \mathbf{P}_k \quad (m = 1, 2, 3, \ldots),$$

$$\sum_{j=1}^{n} \mathbf{P}_j = \mathbf{I}, \quad \text{where } \mathbf{I} \text{ is the identity operator.}$$

For a normal operator $\mathbf{A}$, there is an orthonormal basis consisting of its eigenvectors, $\mathbf{A}\mathbf{i}_k = \lambda\mathbf{i}_k$. Then one obtains the *spectral decomposition of a normal operator*:

$$\mathbf{A}^k = \sum_{j=1}^{n} \lambda_j^k \mathbf{P}_j \quad (k = 1, 2, 3, \ldots). \tag{5.6.3.3}$$

Consider an arbitrary polynomial $p(\lambda) = \sum_{j=1}^{m} c_j\lambda^j$. By definition, $p(\mathbf{A}) = \sum_{j=1}^{m} c_j\mathbf{A}^j$. Then, using (5.6.3.3), we get

$$p(\mathbf{A}) = \sum_{i=1}^{m} p(\lambda_i)\mathbf{P}_i.$$

CAYLEY-HAMILTON THEOREM. *Every normal operator satisfies its own characteristic equation, i.e.,* $f_{\mathbf{A}}(\mathbf{A}) = \mathbf{O}$.

5.6.3-6. Canonical form of linear operators.

An element $\mathbf{x}$ is called an *associated vector* of an operator $\mathbf{A}$ corresponding to its eigenvalue λ if for some $m \geq 1$, we have

$$(\mathbf{A} - \lambda\mathbf{I})^m \mathbf{x} \neq 0, \quad (\mathbf{A} - \lambda\mathbf{I})^{m+1} \mathbf{x} = 0.$$

The number m is called the *order of the associated vector* $\mathbf{x}$.

THEOREM. *Let $\mathbf{A}$ be a linear operator in an n-dimensional unitary space $\mathcal{V}$. Then there is a basis $\{\mathbf{i}_k^m\}$ ($k = 1, 2, \ldots, l$, $m = 1, 2, \ldots, n_k$, $n_1 + n_2 + \cdots + n_l = n$) in $\mathcal{V}$ consisting of eigenvectors and associated vectors of the operator $\mathbf{A}$ such that the action of the operator $\mathbf{A}$ is determined by the relations*

$$A\mathbf{i}_k^1 = \lambda_k \mathbf{i}_k^1 \quad (k = 1, 2, \ldots, l),$$
$$A\mathbf{i}_k^m = \lambda_k \mathbf{i}_k^m + \mathbf{i}_k^{m-1} \quad (k = 1, 2, \ldots, l,\ m = 2, 3, \ldots, n_k).$$

Remark 1. The vectors $\mathbf{i}_k^1$ ($k = 1, 2, \ldots, l$) are eigenvectors of the operator $\mathbf{A}$ corresponding to the eigenvalues λ_k.

Remark 2. The matrix A of the linear operator $\mathbf{A}$ in the basis $\{\mathbf{i}_k^m\}$ has canonical Jordan form, and the above theorem is also called the theorem on the reduction of a matrix to canonical Jordan form.

5.7. Bilinear and Quadratic Forms

5.7.1. Linear and Sesquilinear Forms

5.7.1-1. Linear forms in a unitary space.

A *linear form* or *linear functional* on $\mathcal{V}$ is a linear operator $\mathbf{A}$ in $L(\mathcal{V}, \mathcal{C})$, where $\mathcal{C}$ is the complex plane.

THEOREM. *For any linear form f in a finite-dimensional unitary space $\mathcal{V}$, there is a unique element $\mathbf{h}$ in $\mathcal{V}$ such that*

$$f(\mathbf{x}) = \mathbf{x} \cdot \mathbf{h} \quad \textit{for all} \quad \mathbf{x} \in \mathcal{V}.$$

Remark. This statement is true also for a Euclidean space $\mathcal{V}$ and a real-valued linear functional.

5.7.1-2. Sesquilinear forms in unitary space.

A *sesquilinear form* on a unitary space $\mathcal{V}$ is a complex-valued function $B(\mathbf{x}, \mathbf{y})$ of two arguments $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ such that for any $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ in $\mathcal{V}$ and any complex scalar λ, the following relations hold:

1. $B(\mathbf{x} + \mathbf{y}, \mathbf{z}) = B(\mathbf{x}, \mathbf{z}) + B(\mathbf{y}, \mathbf{z})$.
2. $B(\mathbf{x}, \mathbf{y} + \mathbf{z}) = B(\mathbf{x}, \mathbf{y}) + B(\mathbf{x}, \mathbf{z})$.
3. $B(\lambda\mathbf{x}, \mathbf{y}) = \lambda B(\mathbf{x}, \mathbf{y})$.
4. $B(\mathbf{x}, \lambda\mathbf{y}) = \bar{\lambda} B(\mathbf{x}, \mathbf{y})$.

Remark. Thus, $B(\mathbf{x}, \mathbf{y})$ is a scalar function that is linear with respect to its first argument and antilinear with respect to its second argument. For a real space $\mathcal{V}$, sesquilinear forms turn into bilinear forms (see Paragraph 5.7.2).

THEOREM. *Let $B(\mathbf{x}, \mathbf{y})$ be a sesquilinear form in a unitary space $\mathcal{V}$. Then there is a unique linear operator $\mathbf{A}$ in $L(\mathcal{V}, \mathcal{V})$ such that*

$$B(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot (\mathbf{A}\mathbf{y}).$$

COROLLARY. *If $B(\mathbf{x}, \mathbf{y})$ is a sesquilinear form in a unitary space V, then there is a unique linear operator $\mathbf{A}$ in $L(\mathcal{V}, \mathcal{V})$ such that*

$$B(\mathbf{x}, \mathbf{y}) = (\mathbf{A}\mathbf{x}) \cdot \mathbf{y}.$$

5.7.1-3. Matrix of a sesquilinear form.

Any sesquilinear form $B(\mathbf{x}, \mathbf{y})$ on an n-dimensional linear space with a given basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ can be uniquely represented as

$$B(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^{n} b_{ij} \xi_i \bar{\eta}_j, \qquad b_{ij} = B(\mathbf{e}_i, \mathbf{e}_j),$$

and ξ_i, η_j are the coordinates of $\mathbf{x}$ and $\mathbf{y}$ in the given basis. The matrix $B \equiv [b_{ij}]$ of size $n \times n$ is called the *matrix of the sesquilinear form* $B(\mathbf{x}, \mathbf{y})$ in the given basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. This sesquilinear form can also be represented as

$$B(\mathbf{x}, \mathbf{y}) = X^T B Y, \qquad X^T \equiv (\xi_1, \ldots, \xi_n), \quad Y^T \equiv (\bar{\eta}_1, \ldots, \bar{\eta}_n).$$

5.7.2. Bilinear Forms

5.7.2-1. Definition of a bilinear form.

A *bilinear form* on a real linear space $\mathcal{V}$ is a real-valued function $B(\mathbf{x}, \mathbf{y})$ of two arguments $\mathbf{x} \in L$, $\mathbf{y} \in \mathcal{V}$ satisfying the following conditions for any vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ in $\mathcal{V}$ and any real λ:

1. $B(\mathbf{x} + \mathbf{y}, \mathbf{z}) = B(\mathbf{x}, \mathbf{z}) + B(\mathbf{y}, \mathbf{z})$.
2. $B(\mathbf{x}, \mathbf{y} + \mathbf{z}) = B(\mathbf{x}, \mathbf{y}) + B(\mathbf{x}, \mathbf{z})$.
3. $B(\lambda\mathbf{x}, \mathbf{y}) = B(\mathbf{x}, \lambda\mathbf{y}) = \lambda B(\mathbf{x}, \mathbf{y})$.

THEOREM. *Let $B(\mathbf{x}, \mathbf{y})$ be a bilinear form in a Euclidean space $\mathcal{V}$. Then there is a unique linear operator $\mathbf{A}$ in $L(\mathcal{V}, \mathcal{V})$ such that*

$$B(\mathbf{x}, \mathbf{y}) = (\mathbf{A}\mathbf{x}) \cdot \mathbf{y}.$$

A bilinear form $B(\mathbf{x}, \mathbf{y})$ is said to be *symmetric* if for any $\mathbf{x}$ and $\mathbf{y}$, we have

$$B(\mathbf{x}, \mathbf{y}) = B(\mathbf{y}, \mathbf{x}).$$

A bilinear form $B(\mathbf{x}, \mathbf{y})$ is said to be *skew-symmetric* if for any $\mathbf{x}$ and $\mathbf{y}$, we have

$$B(\mathbf{x}, \mathbf{y}) = -B(\mathbf{y}, \mathbf{x}).$$

Any bilinear form can be represented as the sum of symmetric and skew-symmetric bilinear forms.

THEOREM. *A bilinear form $B(\mathbf{x}, \mathbf{y})$ on a Euclidean space $\mathcal{V}$ is symmetric if and only if the linear operator $\mathbf{A}$ in the representation (5.6.6.1) is Hermitian ($\mathbf{A} = \mathbf{A}^*$).*

5.7.2-2. Bilinear forms in finite-dimensional spaces.

Any bilinear form $B(\mathbf{x}, \mathbf{y})$ on an n-dimensional linear space with a given basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ can be uniquely represented as

$$B(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^{n} b_{ij}\xi_i\eta_j, \qquad b_{ij} = B(\mathbf{e}_i, \mathbf{e}_j),$$

and ξ_i, η_j are the coordinates of the vectors $\mathbf{x}$ and $\mathbf{y}$ in the given basis. The matrix $B \equiv [b_{ij}]$ of size $n \times n$ is called the *matrix of the bilinear form* in the given basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. The bilinear form can also be represented as

$$B(\mathbf{x}, \mathbf{y}) = X^T BY, \qquad X^T \equiv (\xi_1, \ldots, \xi_n), \quad Y^T \equiv (\eta_1, \ldots, \eta_n).$$

Remark. Any square matrix $B \equiv [b_{ij}]$ can be regarded as a matrix of some bilinear form in a given basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. If this matrix is symmetric (skew-symmetric), then the bilinear form is symmetric (skew-symmetric).

The *rank of a bilinear form* $B(\mathbf{x}, \mathbf{y})$ on a finite-dimensional linear space L is defined as the rank of the matrix B of this form in any basis: $\operatorname{rank} B(\mathbf{x}, \mathbf{y}) = \operatorname{rank}(B)$.

A bilinear form on a finite dimensional space $\mathcal{V}$ is said to be *nondegenerate* (*degenerate*) if its rank is equal to (is less than) the dimension of the space $\mathcal{V}$, i.e., $\operatorname{rank} B(\mathbf{x}, \mathbf{y}) = \dim \mathcal{V}$ ($\operatorname{rank} B(\mathbf{x}, \mathbf{y}) < \dim \mathcal{V}$).

5.7.2-3. Transformation of the matrix of a bilinear form in another basis.

Suppose that the transition from a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ to a basis $\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_n$ is determined by the matrix $U \equiv [u_{ij}]$ of size $n \times n$, i.e.

$$\widetilde{\mathbf{e}}_i = \sum_{j=1}^{n} u_{ij}\mathbf{e}_j \qquad (i = 1, 2, \ldots, n).$$

THEOREM. *The matrices B and $\widetilde{B}$ of a bilinear form $B(\mathbf{x}, \mathbf{y})$ in the bases $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_n$, respectively, are related by*

$$\widetilde{B} = U^T BU.$$

5.7.2-4. Multilinear forms.

A *multilinear form* on a linear space $\mathcal{V}$ is a scalar function $B(\mathbf{x}_1, \ldots, \mathbf{x}_p)$ of p arguments $\mathbf{x}_1, \ldots, \mathbf{x}_p \in \mathcal{V}$, which is linear in each argument for fixed values of the other arguments.

A multilinear form $B(\mathbf{x}, \mathbf{y})$ is said to be *symmetric* if for any two arguments $\mathbf{x}_l$ and $\mathbf{x}_l$, we have

$$B(\mathbf{x}_1, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_l, \ldots, \mathbf{x}_p) = B(\mathbf{x}_1, \ldots, \mathbf{x}_l, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_p).$$

A multilinear form $B(\mathbf{x}, \mathbf{y})$ is said to be *skew-symmetric* if for any two arguments $\mathbf{x}_l$ and $\mathbf{x}_l$, we have

$$B(\mathbf{x}_1, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_l, \ldots, \mathbf{x}_p) = -B(\mathbf{x}_1, \ldots, \mathbf{x}_l, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_p).$$

5.7.3. Quadratic Forms

5.7.3-1. Definition of a quadratic form.

A *quadratic form* on a real linear space is a scalar function $B(\mathbf{x}, \mathbf{x})$ obtained from a bilinear form $B(\mathbf{x}, \mathbf{y})$ for $\mathbf{x} = \mathbf{y}$.

Any symmetric bilinear form $B(\mathbf{x}, \mathbf{y})$ is *polar* with respect to the quadratic form $B(\mathbf{x}, \mathbf{x})$. These forms are related by

$$B(\mathbf{x}, \mathbf{y}) = \tfrac{1}{2}[B(\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) - B(\mathbf{x}, \mathbf{x}) - B(\mathbf{y}, \mathbf{y})].$$

5.7.3-2. Quadratic forms in a finite-dimensional linear space.

Any quadratic form $B(\mathbf{x}, \mathbf{x})$ in an n-dimensional linear space with a given basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ can be uniquely represented in the form

$$B(\mathbf{x}, \mathbf{x}) = \sum_{i,j=1}^{n} b_{ij}\xi_i\xi_j, \tag{5.7.3.1}$$

where ξ_i are the coordinates of the vector $\mathbf{x}$ in the given basis, and $B \equiv [b_{ij}]$ is a symmetric matrix of size $n \times n$, called the *matrix of the bilinear form* $B(\mathbf{x}, \mathbf{x})$ in the given basis. This quadratic form can also be represented as

$$B(\mathbf{x}, \mathbf{x}) = X^T BX, \qquad X^T \equiv (\xi_1, \ldots, \xi_n).$$

Remark. Any quadratic form can be represented in the form (5.7.3.1) with infinitely many matrices B such that $B(\mathbf{x}, \mathbf{x}) = X^T BX$. In what follows, we consider only one of such matrices, namely, the symmetric matrix. A quadratic form is real-valued if its symmetric matrix is real.

A real-valued quadratic form $B(\mathbf{x}, \mathbf{x})$ is said to be:

a) *positive definite* (*negative definite*) if $B(\mathbf{x}, \mathbf{x}) > 0$ ($B(\mathbf{x}, \mathbf{x}) < 0$) for any $\mathbf{x} \neq 0$;

b) *alternating* if there exist $\mathbf{x}$ and $\mathbf{y}$ such that $B(\mathbf{x}, \mathbf{x}) > 0$ and $B(\mathbf{y}, \mathbf{y}) < 0$;

c) *nonnegative* (*nonpositive*) if $B(\mathbf{x}, \mathbf{x}) \geq 0$ ($B(\mathbf{x}, \mathbf{x}) \leq 0$) for all $\mathbf{x}$.

If $B(\mathbf{x}, \mathbf{y})$ is a polar bilinear form with respect to some positive definite quadratic form $B(\mathbf{x}, \mathbf{x})$, then $B(\mathbf{x}, \mathbf{y})$ satisfies all axioms of the scalar product in a Euclidean space.

Remark. The axioms of the scalar product can be regarded as the conditions that determine a bilinear form that is polar to some positive definite quadratic form.

The *rank of a quadratic form* on a finite-dimensional linear space $\mathcal{V}$ is, by definition, the rank of the matrix of that form in any basis of $\mathcal{V}$, rank $B(\mathbf{x}, \mathbf{x})$ = rank (B).

A quadratic form on a finite-dimensional linear space $\mathcal{V}$ is said to be *nondegenerate* (*degenerate*) if its rank is equal to (is less than) the dimension of $\mathcal{V}$, i.e., rank $B(\mathbf{x}, \mathbf{x}) = \dim \mathcal{V}$ (rank $B(\mathbf{x}, \mathbf{x}) < \dim \mathcal{V}$).

5.7.3-3. Transformation of a bilinear form in another basis.

Suppose that the transition from the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ to the basis $\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_n$ is given by the matrix $U \equiv [u_{ij}]$ of size $n \times n$, i.e.

$$\widetilde{\mathbf{e}}_i = \sum_{j=1}^{n} u_{ij}\mathbf{e}_j \qquad (i = 1, 2, \ldots, n).$$

Then the matrices B and $\widetilde{B}$ of the quadratic form $B(\mathbf{x}, \mathbf{x})$ in the bases $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\widetilde{\mathbf{e}}_1, \ldots, \widetilde{\mathbf{e}}_n$, respectively, are related by

$$\widetilde{B} = U^T BU.$$

5.7.3-4. Canonical representation of a real quadratic form.

Let $\mathbf{g}_1, \ldots, \mathbf{g}_n$ be a basis in which the real quadratic form $B(\mathbf{x}, \mathbf{x})$ in a linear space $\mathcal{V}$ admits the representation

$$B(\mathbf{x}, \mathbf{x}) = \sum_{i=1}^{n} \lambda_i \eta_i^2, \tag{5.7.3.2}$$

where $\eta_1, \ldots, \eta_n$ are the coordinates of $\mathbf{x}$ in that basis. This representation is called a *canonical representation* of the quadratic form, the real coefficients $\lambda_1, \ldots, \lambda_n$ are called the *canonical coefficients*, and the basis $\mathbf{g}_1, \ldots, \mathbf{g}_n$ is called the *canonical basis*.

The number of nonzero canonical coefficients is equal to the rank of the quadratic form.

THEOREM. *Any real quadratic form on an n-dimensional real linear space $\mathcal{V}$ admits a canonical representation (5.7.3.2).*

1°. *Lagrange method*. The basic idea of the method consists of consecutive transformations of the quadratic form: on every step, one should single out the perfect square of some linear form.

Consider a quadratic form

$$B(\mathbf{x}, \mathbf{x}) = \sum_{i,j=1}^{n} b_{ij} \xi_i \xi_j.$$

Case 1. Suppose that for some m $(1 \le m \le n)$, we have $b_{mm} \ne 0$. Then, letting

$$B(\mathbf{x}, \mathbf{x}) = \frac{1}{b_{mm}} \left(\sum_{k=1}^{n} b_{mk} \xi_k \right)^2 + B_2(\mathbf{x}, \mathbf{x}),$$

one can easily verify that the quadratic form $B_2(\mathbf{x}, \mathbf{x})$ does not contain the variable ξ_m. This method of separating a perfect square in a quadratic form can always be applied if the matrix $[b_{ij}]$ $(i, j = 1, 2, \ldots, n)$ contains nonzero diagonal elements.

Case 2. Suppose that $b_{mm} = 0$, $b_{ss} = 0$, but $b_{ms} \ne 0$. In this case, the quadratic form can be represented as

$$B(\mathbf{x}, \mathbf{x}) = \frac{1}{2b_{sm}} \left[\sum_{k=1}^{n} (b_{mk} + b_{sk}) \xi_k \right]^2 - \frac{1}{2b_{sm}} \left[\sum_{k=1}^{n} (b_{mk} - b_{sk}) \xi_k \right]^2 + B_2(\mathbf{x}, \mathbf{x}),$$

where $B_2(\mathbf{x}, \mathbf{x})$ does not contain the variables ξ_m, ξ_s, and the linear forms in square brackets are linearly independent (and therefore can be taken as new independent variables or coordinates).

By consecutive combination of the above two procedures, the quadratic form $B(\mathbf{x}, \mathbf{x})$ can always be represented in terms of squared linear forms; these forms are linearly independent, since each contains a variable which is absent in the other linear forms.

2°. *Jacobi method*. Suppose that

$$\Delta_1 \equiv b_{11} \ne 0, \quad \Delta_2 \equiv \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} \ne 0, \quad \ldots, \quad \Delta_n \equiv \det B \ne 0,$$

where $B \equiv [b_{ij}]$ is the matrix of the quadratic form $B(\mathbf{x}, \mathbf{x})$ in some basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. One can obtain a canonical representation of this form using the formulas

$$\lambda_1 = \Delta_1, \qquad \lambda_i = \frac{\Delta_i}{\Delta_{i-1}} \qquad (i = 2, 3, \ldots, n).$$

The basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is transformed to the canonical basis $\mathbf{g}_1, \ldots, \mathbf{g}_n$ by the formulas

$$\mathbf{g}_i = \sum_{j=1}^{n} \alpha_{ij}\mathbf{e}_j \quad (i = 1, 2, \ldots, n),$$

$$\alpha_{ij} = (-1)^{i+j}\frac{\Delta_{i-1,j}}{\Delta_{i-1}},$$

where $\Delta_{i-1,j}$ is the minor of the submatrix of $B \equiv [b_{ij}]$ formed by the elements on the intersection of its rows with indices $1, 2, \ldots, i-1$ and columns with indices $1, 2, \ldots, j-1$, $j+1, i$.

5.7.3-5. Normal representation of a real quadratic form.

Let $\mathbf{g}_1, \ldots, \mathbf{g}_n$ be a basis of a linear space $\mathcal{V}$ in which the quadratic form $B(\mathbf{x}, \mathbf{x})$ is written as

$$B(\mathbf{x}, \mathbf{x}) = \sum_{i=1}^{n} \varepsilon_i \eta_i^2, \tag{5.7.3.3}$$

where $\eta_1, \ldots, \eta_n$ are the coordinates of $\mathbf{x}$ in that basis, and $\varepsilon_1, \ldots, \varepsilon_n$ are coefficients taking the values -1, 0, or 1. Such a representation of a quadratic form is called its *normal representation*.

Any real quadratic form $B(\mathbf{x}, \mathbf{x})$ in an n-dimensional real linear space $\mathcal{V}$ admits a normal representation (5.7.3.3). Such a representation can be obtained by the following transformations:

1. One obtains its canonical representation (see Paragraph 5.7.3-4):

$$B(\mathbf{x}, \mathbf{x}) = \sum_{i=1}^{n} \lambda_i \mu_i^2.$$

2. By the nondegenerate coordinate transformation

$$\eta_i = \begin{cases} \frac{1}{\sqrt{\lambda_i}}\mu_i & \text{for } \lambda_i > 0, \\ \frac{1}{\sqrt{-\lambda_i}}\mu_i & \text{for } \lambda_i < 0, \\ \mu_i & \text{for } \lambda_i = 0, \end{cases}$$

the canonical representation turns into a normal representation.

LAW OF INERTIA OF QUADRATIC FORMS. *The number of terms with positive coefficients and the number of terms with negative coefficients in any normal representation of a real quadratic form does not depend on the method used to obtain such a representation.*

The *index of inertia* of a real quadratic form is the integer k equal to the number of nonzero coefficients in its canonical representation (this number coincides with the rank of the quadratic form). Its *positive index of inertia* is the integer p equal to the number of positive coefficients in the canonical representation of the form, and its *negative index of inertia* is the integer q equal to the number of its negative canonical coefficients. The integer $s = p - q$ is called the *signature* of the quadratic form.

A real quadratic form $B(\mathbf{x}, \mathbf{x})$ on an n-dimensional real linear space $\mathcal{V}$ is

a) positive definite (resp., negative definite) if $p = n$ (resp., $q = n$);
b) alternating if $p \neq 0$, $q \neq 0$;
c) nonnegative (resp., nonpositive) if $q = 0$, $p < n$ (resp., $p = 0$, $q < n$).

5.7.3-6. Criteria of positive and negative definiteness of a quadratic form.

1°. A real quadratic form $B(\mathbf{x},\mathbf{x})$ is positive definite, negative definite, alternating, nonnegative, nonpositive if the eigenvalues λ_i of its matrix $B \equiv [b_{ij}]$ are all positive, are all negative, some are positive and some negative, are all nonnegative, are all nonpositive, respectively.

2°. *Sylvester criterion.* A real quadratic form $B(\mathbf{x},\mathbf{x})$ is positive definite if and only if the matrix of $B(\mathbf{x},\mathbf{x})$ in some basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ satisfies the conditions

$$\Delta_1 \equiv b_{11} > 0, \quad \Delta_2 \equiv \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} > 0, \ldots, \quad \Delta_n \equiv \det B > 0.$$

If the signs of the minor determinants alternate,

$$\Delta_1 < 0, \quad \Delta_2 > 0, \quad \Delta_3 < 0, \ldots,$$

then the quadratic form is negative definite.

3°. A real matrix B is nonnegative and symmetric if and only if there is a real matrix C such that $B = C^T C$.

5.7.4. Bilinear and Quadratic Forms in Euclidean Space

5.7.4-1. Reduction of a quadratic form to a sum of squares.

THEOREM 1. *Let $B(\mathbf{x},\mathbf{y})$ be a symmetric bilinear form on a n-dimensional Euclidean space $\mathcal{V}$. Then there is an orthonormal basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ in $\mathcal{V}$ and there are real numbers λ_k such that for any $\mathbf{x} \in \mathcal{V}$ the real quadratic form $B(\mathbf{x},\mathbf{x})$ can be represented as the sum of squares of the coordinates ξ_k of $\mathbf{x}$ in the basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$:*

$$B(\mathbf{x},\mathbf{x}) = \sum_{k=1}^{n} \lambda_k \xi_k^2.$$

THEOREM 2. *Let $A(\mathbf{x},\mathbf{y})$ and $B(\mathbf{x},\mathbf{y})$ be symmetric bilinear forms in a n-dimensional real linear space $\mathcal{V}$, and suppose that the quadratic form $A(\mathbf{x},\mathbf{x})$ is positive definite. Then there is a basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ of $\mathcal{V}$ such that the quadratic forms $A(\mathbf{x},\mathbf{x})$ and $B(\mathbf{x},\mathbf{x})$ can be represented in the form*

$$A(\mathbf{x},\mathbf{x}) = \sum_{k=1}^{n} \lambda_k \xi_k^2, \quad B(\mathbf{x},\mathbf{x}) = \sum_{k=1}^{n} \xi_k^2,$$

where ξ_k are the coordinates of $\mathbf{x}$ in the basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$. The set of real $\lambda_1, \ldots, \lambda_n$ coincides with the spectrum of eigenvalues of the matrix $B^{-1}A$ (the matrices A and B can be taken in any basis), and this set consists of the roots of the algebraic equation

$$\det(A - \lambda B) = 0.$$

5.7.4-2. Extremal properties of quadratic forms.

A point $\mathbf{x}_0$ on a smooth surface S is called a *stationary point* of a differentiable function f defined on S if the derivative of f at the point $\mathbf{x}_0$ in any direction on S is equal to zero. The value $f(\mathbf{x}_0)$ of the function f at a stationary point $\mathbf{x}_0$ is called its *stationary value*.

The *unit sphere* in a Euclidean space $\mathcal{V}$ is the set of all $\mathbf{x} \in \mathcal{V}$ such that

$$\mathbf{x} \cdot \mathbf{x} = 1 \quad (\|\mathbf{x}\| = 1). \tag{5.7.4.1}$$

THEOREM. *Let $B(\mathbf{x}, \mathbf{x})$ be a real quadratic form and let $B(\mathbf{x}, \mathbf{y}) = (\mathbf{A}\mathbf{x}) \cdot \mathbf{y}$ be the corresponding polar bilinear form, where $\mathbf{A}$ is a Hermitian operator. The stationary values of the quadratic form $B(\mathbf{x}, \mathbf{x})$ on the unit sphere (5.7.4.1) coincide with eigenvalues of the operator $\mathbf{A}$. These stationary values are attained, in particular, on the unit eigenvectors $\mathbf{e}_k$ of the operator $\mathbf{A}$.*

Remark. If the eigenvalues of the operator $\mathbf{A}$ satisfy the inequalities $\lambda_1 \geq \ldots \geq \lambda_n$, then λ_1 and λ_n are the largest and the smallest values of $B(\mathbf{x}, \mathbf{x})$ on the sphere $\mathbf{x} \cdot \mathbf{x} = 1$.

5.7.5. Second-Order Hypersurfaces

5.7.5-1. Definition of a second-order hypersurface.

A *second-order hypersurface* in an n-dimensional Euclidean space $\mathcal{V}$ is the set of all points $\mathbf{x} \in \mathcal{V}$ satisfying an equation of the form

$$A(\mathbf{x}, \mathbf{x}) + 2B(\mathbf{x}) + c = 0, \tag{5.7.5.1}$$

where $A(\mathbf{x}, \mathbf{x})$ is a real quadratic form different from identical zero, $B(\mathbf{x})$ is a linear form, and c is a real constant. Equation (5.7.5.1) is called the *general equation of a second-order hypersurface*.

Suppose that in some orthonormal basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$, we have

$$A(\mathbf{x}, \mathbf{x}) = X^T A X = \sum_{i,j=1}^{n} a_{ij} x_i x_k, \quad B(\mathbf{x}) = BX = \sum_{i=1}^{n} b_i x_i,$$
$$X^T = (x_1, \ldots, x_n), \quad A \equiv [a_{ij}], \quad B = (b_1, \ldots, b_n).$$

Then the general equation (5.7.5.1) of a second-order hypersurface in the Euclidean space $\mathcal{V}$ with the given orthonormal basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ can be written as

$$X^T A X + 2BX + c = 0.$$

The term $A(\mathbf{x}, \mathbf{x}) = X^T A X$ is called the *group of the leading terms* of equation (5.7.5.1), and the terms $B(\mathbf{x}) + c = BX + c$ are called the *linear part of the equation*.

5.7.5-2. Parallel translation.

A *parallel translation* in a Euclidean space $\mathcal{V}$ is a transformation defined by the formulas

$$X = X' + \overset{\circ}{X}, \tag{5.7.5.2}$$

where $\overset{\circ}{X}$ is a fixed point, called the *new origin*.

In terms of coordinates, (5.7.5.2) takes the form

$$x_k = x'_k + \overset{\circ}{x}_k \quad (k = 1, 2, \ldots, n),$$

where $X^T = (x_1, \ldots, x_n)$, $X'^T = (x'_1, \ldots, x'_n)$, $\overset{\circ}{X}{}^T = (\overset{\circ}{x}_1, \ldots, \overset{\circ}{x}_n)$.

Under parallel translations any basis remains unchanged.

The transformation of the space $\mathcal{V}$ defined by (5.7.5.2) reduces the hypersurface equation (5.7.5.1) to

$$A(\mathbf{x}', \mathbf{x}') + 2B'(\mathbf{x}') + c' = 0,$$

where the linear form $B'(\mathbf{x}')$ and the constant c' are defined by

$$B'(\mathbf{x}') = A(\mathbf{x}', \overset{\circ}{\mathbf{x}}) + B(\mathbf{x}'), \quad c' = A(\overset{\circ}{\mathbf{x}}, \overset{\circ}{\mathbf{x}}) + 2B(\overset{\circ}{\mathbf{x}}) + c,$$

or, in coordinate notation,

$$B'(\mathbf{x}') \equiv B'X' = \sum_{i=1}^{n} b'_i x'_i, \quad c' = \sum_{i=1}^{n} (b'_i + b_i)\overset{\circ}{x}_i + c, \qquad b'_i = \sum_{j=1}^{n} a_{ij}\overset{\circ}{x}_j + b_i.$$

Under parallel translation the group of the leading terms preserves its form.

5.7.5-3. Transformation of one orthonormal basis into another.

The transition from one orthonormal basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ to another orthonormal basis $\mathbf{i}'_1, \ldots, \mathbf{i}'_n$ is defined by an orthogonal matrix $P \equiv [p_{ij}]$ of size $n \times n$, i.e.,

$$\mathbf{i}'_i = \sum_{j=1}^{n} p_{ij}\mathbf{i}_j \qquad (i = 1, 2, \ldots, n).$$

Under this orthogonal transformation, the coordinates of points are transformed as follows:

$$X' = PX,$$

or, in coordinate notation,

$$x'_k = \sum_{i=1}^{n} p_{ki} x_i \quad (k = 1, 2, \ldots, n), \tag{5.7.5.3}$$

where $X^T = (x_1, \ldots, x_n)$, $X'^T = (x'_1, \ldots, x'_n)$.

If the transition from the orthonormal basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ to the orthonormal basis $\mathbf{i}'_1, \ldots, \mathbf{i}'_n$ is defined by an orthogonal matrix P, then the hypersurface equation (5.7.5.1) in the new basis takes the form

$$A'(\mathbf{x}', \mathbf{x}') + 2B'(\mathbf{x}') + c' = 0.$$

The matrix $A' \equiv [a'_{ij}]$ ($A'(\mathbf{x}', \mathbf{x}') = X'^T A' X'$) is found from the relation

$$A' = P^{-1}AP.$$

Thus, when passing from one orthonormal basis to another orthonormal basis, the matrix of a quadratic form is transformed similarly to the matrix of some linear operator. Note that the operator $\mathbf{A}$ whose matrix in an orthonormal basis coincides with the matrix of the quadratic form $A(\mathbf{x}, \mathbf{x})$ is Hermitian.

The coefficients b'_i of the linear form $B'(\mathbf{x}') = \sum\limits_{i=1}^{n} b'_i x'_i$ are found from the relations [to this end, one should use (5.7.5.3)]

$$\sum_{i=1}^{n} b'_i x'_i = \sum_{i=1}^{n} b_i x_i,$$

and the constant is $c' = c$.

5.7.5-4. Invariants of the general equation of a second-order hypersurface.

An *invariant* of the general second-order hypersurface equation (5.7.5.1) with respect to parallel translations and orthogonal transformations of an orthogonal basis is, by definition, any function $f(a_{11}, \ldots, a_{nn}, b_1, \ldots, b_n, c)$ of the coefficients of this equation that does not change under such transformations of the space.

THEOREM. *The coefficients of the characteristic polynomial of the matrix A of the quadratic form $A(\mathbf{x}, \mathbf{x})$ and the determinant* $\det \widetilde{A}$ *of the block matrix* $\widetilde{A} = \begin{pmatrix} A & B \\ B^T & c \end{pmatrix}$ *are invariants of the general second-order hypersurface equation (5.7.5.1).*

Remark. The quantities $\det A$, $\mathrm{Tr}(A)$, $\mathrm{rank}\,(A)$, and $\mathrm{rank}\,(\widetilde{A})$ are invariants of equation (5.7.5.1).

5.7.5-5. Center of a second-order hypersurface.

The *center* of a second-order hypersurface is a point $\overset{\circ}{\mathbf{x}}$ such that the linear form $B'(\mathbf{x}')$ becomes identically equal to zero after the parallel translation that makes $\overset{\circ}{\mathbf{x}}$ the new origin. Thus, the coordinates of the center can be found from the system of *equations of the center of a second-order hypersurface*

$$\sum_{j=1}^{n} a_{ij} \overset{\circ}{x}_j + b_i = 0 \qquad (i = 1, 2, \ldots, n).$$

If the center equations for a hypersurface S have a unique solution, then S is called a *central hypersurface*. If a hypersurface S has a center, then S consists of pairs of points, each pair being symmetric with respect to the center.

Remark 1. For a second-order hypersurface S with a center, the invariants $\det A$, $\det \widetilde{A}$, and the free term c' are related by

$$\det \widetilde{A} = c' \det A.$$

Remark 2. If the origin is shifted to the center of a central hypersurface S, then the equation of that hypersurface in new coordinates has the form

$$A(\mathbf{x}, \mathbf{x}) + \frac{\det \widetilde{A}}{\det A} = 0.$$

5.7.5-6. Simplification of a second-order hypersurface equation.

Let $\mathbf{A}$ be the operator whose matrix in an orthonormal basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ coincides with the matrix of a quadratic form $A(\mathbf{x}, \mathbf{x})$. Suppose that the transition from the orthonormal basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ to the orthonormal basis $\mathbf{i}'_1, \ldots, \mathbf{i}'_n$ is defined by an orthogonal matrix P, and $A' = P^{-1}AP$ is a diagonal matrix with the eigenvalues of the operator $\mathbf{A}$ on the main diagonal. Then the equation of the hypersurface (5.7.5.1) in the new basis takes the form

$$\sum_{i=1}^{n} \lambda_i x_i'^2 + 2\sum_{i=1}^{n} b'_i x'_i + c = 0, \tag{5.7.5.4}$$

where the coefficients b'_i are determined by the relations

$$\sum_{i=1}^{n} b'_i x'_i = \sum_{i=1}^{n} b_i x_i.$$

The reduction of any equation of a second-order hypersurface S to the form (5.7.5.4) is called the *standard simplification of this equation (by an orthogonal transformation of the basis)*.

5.7.5-7. Classification of central second-order hypersurfaces.

1°. Let $\mathbf{i}_2, \ldots, \mathbf{i}_n$ be an orthonormal basis in which a second-order central hypersurface is defined by the equation (called its *canonical equation*)

$$\sum_{i=1}^{n} \varepsilon_i \frac{x_i^2}{a_i^2} + \operatorname{sign} \frac{\det \widetilde{A}}{\det A} = 0, \tag{5.7.5.5}$$

where $x_1, \ldots, x_n$ are the coordinates of $\mathbf{x}$ in that basis, and the coefficients $\varepsilon_1, \ldots, \varepsilon_n$ take the values –1, 0, or 1. The constants $a_k > 0$ are called the *semiaxes* of the hypersurface.

The equation of any central hypersurface S can be reduced to the canonical equation (5.7.5.5) by the following transformations:

1. By the parallel translation that shifts the origin to the center of the hypersurface, its equation is transformed to (see Paragraph 5.7.5-5):

$$A(\mathbf{x}, \mathbf{x}) + \frac{\det \widetilde{A}}{\det A} = 0.$$

2. By the standard simplification of the last equation, one obtains an equation of the hypersurface in the form

$$\sum_{i=1}^{n} \lambda_i x_i^2 + \frac{\det \widetilde{A}}{\det A} = 0.$$

3. Letting

$$\frac{1}{a_k^2} = \begin{cases} |\lambda_k| \dfrac{|\det A|}{|\det \widetilde{A}|} & \text{if } \det \widetilde{A} \neq 0, \\ |\lambda_k| & \text{if } \det \widetilde{A} = 0, \end{cases} \qquad \varepsilon_k = \operatorname{sign} \lambda_k \qquad (k = 1, 2, \ldots, n),$$

one passes to the canonical equation (5.7.5.5) of the central second-order hypersurface.

2°. Let p be the number of positive eigenvalues of the matrix A and q the number of negative ones. Central second-order hypersurfaces admit the following classification. A hypersurface is called:

a) an $(n-1)$-*dimensional ellipsoid* if $p = n$ and sign $\frac{\det \widetilde{A}}{\det A} = -1$, or $q = n$ and sign $\frac{\det \widetilde{A}}{\det A} = 1$*;

b) an *imaginary ellipsoid* if $p = n$ and sign $\frac{\det \widetilde{A}}{\det A} = 1$, or $q = n$ and sign $\frac{\det \widetilde{A}}{\det A} = -1$;

c) a *hyperboloid* if $0 < p < n$ $(0 < q < n)$ and sign $\frac{\det \widetilde{A}}{\det A} \neq 0$;

d) *degenerate* if sign $\frac{\det \widetilde{A}}{\det A} = 0$.

5.7.5-8. Classification of noncentral second-order hypersurfaces.

1°. Let $\mathbf{i}_2, \ldots, \mathbf{i}_n$ be an orthonormal basis in which a noncentral second-order hypersurface is defined by the equation (called its *canonical equation*)

$$\sum_{i=1}^{p} \lambda_i x_i^2 + 2\mu x_n + c' = 0, \tag{5.7.5.6}$$

where $x_1, \ldots, x_n$ are the coordinates of $\mathbf{x}$ in that basis: $p = \operatorname{rank}(A)$.

The equation of any noncentral second-order hypersurface S can be reduced to the canonical form (5.7.5.6) by the following transformations:

1. If $p = \operatorname{rank}(A)$, then after the standard simplification and renumbering the basis vectors, equation (5.7.5.1) turns into

$$\sum_{i=1}^{p} \lambda_i x_i'^2 + 2\sum_{i=1}^{p} b_i' x_i' + 2\sum_{i=p+1}^{n} b_i' x_i' + c = 0.$$

2. After the parallel translation

$$x_k'' = \begin{cases} x_k' + \dfrac{b_k'}{\lambda_k} & \text{for } k = 1, 2, \ldots, p, \\ x_k' & \text{for } k = p+1, p+2, \ldots, n, \end{cases}$$

the last equation can be represented in the form

$$\sum_{i=1}^{p} \lambda_i x_i''^2 + 2\sum_{i=p+1}^{n} b_i' x_i'' + c' = 0, \qquad c' = c - \sum_{i=1}^{p} \frac{b_i'}{\lambda_i}.$$

3. Leaving intact the first p basis vectors and transforming the last basis vectors so that the term $\sum\limits_{i=p+1}^{n} b_i' x_i''$ turns into $\mu x_n'''$, one reduces the hypersurface equation to the canonical form (5.7.5.6).

2°. Noncentral second-order hypersurfaces admit the following classification.

A hypersurface is called:

a) a *paraboloid* if $\mu \neq 0$ and $p = n - 1$; in this case, the parallel translation in the direction of the x_n-axis by $-\frac{c'}{2\mu}$ yields the canonical equation of a paraboloid

$$\sum_{i=1}^{n-1} \lambda_i x_i^2 + 2\mu x_n = 0;$$

* If $a_1 = \ldots = a_n = R$, then the hypersurface is a *sphere* of radius R in n-dimensional space.

b) a *central cylinder* if $\mu = 0$, $p < n$; its canonical equation has the form

$$\sum_{i=1}^{p} \lambda_i x_i^2 + c' = 0;$$

c) a *paraboloidal cylinder* if $\mu \neq 0$, $p < n-1$; in this case, the parallel translation along the x_n-axis by $-\frac{c'}{2\mu}$ yields the canonical equation of a paraboloidal cylinder

$$\sum_{i=1}^{p} \lambda_i x_i^2 + 2\mu x_n = 0.$$

5.8. Some Facts from Group Theory

5.8.1. Groups and Their Basic Properties

5.8.1-1. Composition laws.

Let **T** be a mapping defined on ordered pairs a, b of elements of a set A and mapping each pair a, b to an element c of A. In this case, one says that a *composition law* is defined on the set A. The element $c \in A$ is called the *composition* of the elements $a, b \in A$ and is denoted by $c = a\mathbf{T}b$.

A composition law is commonly expressed in one of the two forms:

1. *Additive form*: $c = a + b$; the corresponding composition law is called *addition* and c is called the *sum* of a and b.
2. *Multiplicative form*: $c = ab$; the corresponding composition law is called *multiplication* and c is called the *product* of a and b.

A composition law is said to be *associative* if

$$a\mathbf{T}(b\mathbf{T}c) = (a\mathbf{T}b)\mathbf{T}c \qquad \text{for all } a, b, c \in A.$$

In additive form, this relation reads $a + (b + c) = (a + b) + c$; and in multiplicative form, $a(bc) = (ab)c$.

A composition law is said to be *commutative* if

$$a\mathbf{T}b = b\mathbf{T}a \qquad \text{for all } a, b \in A.$$

In additive form, this relation reads $a + b = b + a$; and in multiplicative form, $ab = ba$.

An element e of the set A is said to be *neutral* with respect to the composition law **T** if $a\mathbf{T}e = a$ for any $a \in A$.

In the additive case, a neutral element is called a *zero element*, and in the multiplicative case, an *identity element*.

An element b is called an *inverse* of $a \in A$ if $a\mathbf{T}b = e$. The inverse element is denoted by $b = a^{-1}$.

In the additive case, the inverse element of a is called the *negative* of a and it is denoted by $-a$.

Example 1. Addition and multiplication of real numbers are composition laws on the set of real numbers. Both these laws are commutative. The neutral element for the addition is zero. The neutral element for the multiplication is unity.

5.8.1-2. Notion of a group.

A *group* is a set G with a composition law $\mathbf{T}$ satisfying the conditions:
1. The law $\mathbf{T}$ is associative.
2. There is a neutral element $e \in G$.
3. For any $a \in G$, there is an inverse element a^{-1}.

A group G is said to be *commutative* or *abelian* if its composition law $\mathbf{T}$ is commutative.

Example 2. The set Z of all integer numbers is an abelian group with respect to addition. The set of all positive real numbers is an abelian group with respect to multiplication. Any linear space is an abelian group with respect to the addition of its elements.

Example 3. *Permutation groups.* Let E be a set consisting of finitely many elements $a, b, c, \ldots$. A *permutation* of E is a one-to-one mapping of E onto itself. A permutation f of the set E can be expressed in the form

$$f = \begin{pmatrix} a & b & c & \ldots \\ f(a) & f(b) & f(c) & \ldots \end{pmatrix}.$$

On the set P of all permutations of E, the composition law is introduced as follows: if f_1 and f_2 are two permutations of E, then their composition $f_2 \circ f_1$ is the permutation obtained by consecutive application of f_1 and f_2. This composition law is associative. The set of all permutations of E with this composition law is a group.

Example 4. The group Z_2 that consists of two elements 0 and 1 with the multiplication defined by

$$0 \cdot 0 = 0, \quad 0 \cdot 1 = 1, \quad 1 \cdot 0 = 1, \quad 1 \cdot 1 = 1$$

and the neutral element 0 is called the *group of modulo* 2 *residues*.

Properties of groups:
1. If $a\mathbf{T}a^{-1} = e$, then $a^{-1}\mathbf{T}a = e$.
2. $e\mathbf{T}a = a$ for any a.
3. If $a\mathbf{T}x = e$ and $a\mathbf{T}y = e$, then $x = y$.
4. The neutral element e is unique.

5.8.1-3. Homomorphisms and isomorphisms.

Recall that a mapping $f : A \to B$ of a set A into a set B is a correspondence that associates each element of A with an element of B. The *range* of the mapping f is the set of all $b \in B$ such that $b = f(a)$. One says that f is a *one-to-one mapping* if it maps different elements of A into different elements of B, i.e., for any $a_1, a_2 \in A$ such that $a_1 \neq a_2$, we have $f(a_1) \neq f(a_2)$.

A mapping $f : A \to B$ is called a *mapping of the set A onto the set B* if each element of B is an image of some element of A, i.e., for any $b \in B$, there is $a \in A$ such that $b = f(a)$.

A mapping f of A onto B is said to be *invertible* if there is a mapping $g : B \to A$ such that $g(f(a)) = a$ for any $a \in A$. The mapping g is called the *inverse* of the mapping f and is denoted by $g = f^{-1}$.

For definiteness, we use the multiplicative notation for composition laws in what follows, unless indicated otherwise.

Let G be a group and let $\widetilde{G}$ be a set with a composition law. A mapping $f : G \to \widetilde{G}$ is called a *homomorphism* if

$$f(ab) = f(a)f(b) \qquad \text{for all} \quad a, b \in G;$$

and the subset of $\widetilde{G}$ consisting of all elements of the form $f(a)$, $a \in G$, is called a *homomorphic image* of the group G and is denoted by $f(G)$. Note that here the set $\widetilde{G}$ with a composition law is not necessarily a group. However, the following result holds.

THEOREM. *The homomorphic image $f(G)$ is a group. The image $f(e)$ of the identity element $e \in G$ is the identity element of the group $f(G)$. Mutually inverse elements of G correspond to mutually inverse images in $f(G)$.*

Two groups G_1 and G_2 are said to be *isomorphic* if there exists a one-to-one mapping f of G_1 onto G_2 such that $f(ab) = f(a)f(b)$ for all $a, b \in G_1$. Such a mapping is called *an isomorphism* or *isomorphic mapping* of the group G_1 onto the group G_2.

THEOREM. *Any isomorphism of groups is invertible, and the inverse mapping is also an isomorphism.*

An isomorphic mapping of a group G onto itself is called an *automorphism* of G. If $f_1 : G \to G$ and $f_2 : G \to G$ are two automorphisms of a group G, one can define another automorphism $f_1 \circ f_2 : G \to G$ by letting $(f_1 \circ f_2)(g) = f_1(f_2(g))$ for all $g \in G$. This automorphism is called the *composition* of f_1 and f_2, and with this composition law, the set of all automorphisms of G becomes a group called the *automorphism group* of G.

5.8.1-4. Subgroups. Cosets. Normal subgroups.

Let G be a group. A subset G_1 of the group G is called a *subgroup* if the following conditions hold:

1. For any a and b belonging to G_1, the product ab belongs to G_1.
2. For any a belonging to G_1, its inverse a^{-1} belongs to G_1.

These conditions ensure that any subgroup of a group is itself a group.

Example 5. The identity element of a group is a subgroup. The subset of all even numbers is a subgroup of the additive group of all integers.

The *product* of two subsets H_1 and H_2 of a group G is a set H_3 that consists of all elements of the form h_1h_2, where $h_1 \in H_1$, $h_2 \in H_2$. In this case, one writes $H_3 = H_1H_2$.

Let H be a subgroup of a group G and a some fixed element of G. The set aH is called a *left coset*, and the set Ha is called a *right coset* of the subgroup H in G.

Properties of left cosets (right cosets have similar properties):

1. If $a \in H$, then $aH \equiv H$.
2. Cosets aH and bH coincide if $a^{-1}b \in H$.
3. Two cosets of the same subgroup H either coincide or have no common elements.
4. If aH is a coset, then $a \in aH$.

A subgroup H of a group G is called a *normal subgroup* of G if $H = a^{-1}Ha$ for any $a \in G$. This is equivalent to the condition that $aH = Ha$ for any $a \in G$, i.e., every right coset is a left coset.

5.8.1-5. Factor groups.

Let H be a normal subgroup of a group G. Then the product of two cosets aH and bH (as subsets of G) is the coset abH. Consider the set Q whose elements are cosets of the subgroup H in G, and define the product of the elements of Q as the product of cosets. Endowed with this product, Q becomes a group, denoted by $Q = G/H$ and called the *quotient group* of G with respect to the normal subgroup H.

The mapping $f : G \to G/H$ that maps each $a \in G$ to the corresponding coset aH is a homomorphism of G onto G/H.

If $f : G \to \overline{G}$ is a homomorphism of groups, the set of all elements of G mapped into the identity element of $\overline{G}$ is called the *kernel* of f and is denoted by $\ker f = \{g \in G : f(g) = f(e)\}$.

THEOREM 1. *If f is a homomorphism of a group G onto a group $\overline{G}$ and H is the set of all elements of G that are mapped to $f(e)$ (e is the identity element of G), then H is a normal subgroup in G.*

THEOREM 2 (ON GROUP HOMOMORPHISMS). *If f is a homomorphism of a group G onto a group $\overline{G}$ and H is the normal subgroup of G consisting of the elements mapped to the identity element of $\overline{G}$, then the group $\overline{G}$ and the quotient group G/H are isomorphic.*

Thus, given a homomorphism f of a group G onto a group $\widetilde{G}$, the kernel H of the homomorphism is a normal subgroup of G, and conversely any normal subgroup H of G is the kernel of the homomorphism of G onto the quotient group G/H.

Remark. Given a homomorphism of a group G onto a set $\overline{G}$, all elements of the group G are divided into mutually disjoint classes, each class containing all elements of G that are mapped into the same element of $\overline{G}$.

Example 6. Let $\mathbb{R}^n$ be the n-dimensional linear coordinate space, which is an abelian group with respect to addition of its elements. This space is the direct product of one-dimensional spaces:

$$\mathbb{R}^n = \mathbb{R}^1_{(1)} \otimes \cdots \otimes \mathbb{R}^1_{(n)}.$$

Since $\mathbb{R}^1_{(n)}$ is an abelian subgroup, the set $\mathbb{R}^1_{(n)}$ is a normal subgroup of the group $\mathbb{R}^n$. The coset corresponding to an element $a \in \mathbb{R}^n$ is the straight line passing through a in the direction parallel to the straight line $\mathbb{R}^1_{(n)}$, and the quotient group $\mathbb{R}^n/\mathbb{R}^1_{(n)}$ is isomorphic to the $(n-1)$-dimensional space $\mathbb{R}^{n-1}$:

$$\mathbb{R}^{n-1} = \mathbb{R}^n/\mathbb{R}^1_{(n)} = \mathbb{R}^1_{(1)} \otimes \cdots \otimes \mathbb{R}^1_{(n-1)}.$$

5.8.2. Transformation Groups

5.8.2-1. Group of linear transformations. Its subgroups.

Let V be a real finite-dimensional linear space and let $\mathbf{A} : V \to V$ be a nondegenerate linear operator. This operator can be regarded as a nondegenerate linear transformation of the space V, since A maps different elements of V into different elements, and for any $\mathbf{y} \in V$ there is a unique $\mathbf{x} \in V$ such that $\mathbf{Ax} = \mathbf{y}$.

The set of all nondegenerate linear transformations $\mathbf{A}$ of the n-dimensional real linear space V is denoted by $GL(n)$.

The *product* $\mathbf{AB}$ of linear transformations $\mathbf{A}$ and $\mathbf{B}$ in $GL(n)$ is defined by the relation

$$(\mathbf{AB})\mathbf{x} = \mathbf{A}(\mathbf{Bx}) \qquad \text{for all} \quad \mathbf{x} \in V.$$

This product is a composition law on $GL(n)$.

THEOREM. *The set $GL(n)$ of nondegenerate linear transformations of an n-dimensional real linear space V with the above product is a group.*

The group $GL(n)$ is called the *general linear group of dimension* n.

A subset of $GL(n)$ consisting of all linear transformations $\mathbf{A}$ such that $\det \mathbf{A} = 1$ is a subgroup of $GL(n)$ called the *special linear group of dimension* n and denoted by $SL(n)$.

A sequence $\{\mathbf{A}_k\}$ of elements of $GL(n)$ is said to be *convergent* to an element $\mathbf{A} \in GL(n)$ as $k \to \infty$ if the sequence $\{\mathbf{A}_k\mathbf{x}\}$ converges to $\mathbf{Ax}$ for any $\mathbf{x} \in V$.

Types of subgroups of $GL(n)$:

1. *Finite subgroups* are subgroups with finitely many elements.
2. *Discrete subgroups* are subgroups with countably many elements.
3. *Continuous subgroups* are subgroups with uncountably many elements.

Example 1. The subgroup of reflections with respect to the origin is finite and consists of two elements: the identity transformation and the reflection $\mathbf{x} \to -\mathbf{x}$.

The subgroup of rotations of a plane with respect to the origin by the angles $k\varphi$ ($k = 0, \pm1, \pm2, \ldots$ and φ is a fixed angle incommensurable with π) is a discrete subgroup.

The subgroups of all rotations of a three-dimensional space about a fixed axis are a continuous subgroup.

A continuous subgroup of $GL(n)$ is said to be *compact* if from any infinite sequence of its elements one can extract a subsequence convergent to some element of the subgroup.

5.8.2-2. Group of orthogonal transformations. Its subgroups.

Consider the set $O(n)$ that consists of all orthogonal transformations $\mathbf{P}$ of the n-dimensional Euclidean space V, i.e., $\mathbf{P}^T\mathbf{P} = \mathbf{P}\mathbf{P}^T = \mathbf{I}$ (see Paragraph 5.2.3-3 and Section 5.4). This set is a subgroup of $GL(n)$ called the *orthogonal group* of dimension n.

All orthogonal transformations are divided into two classes:

1. *Proper orthogonal transformations*, for which $\det \mathbf{P} = +1$.
2. *Improper orthogonal transformations*, for which $\det \mathbf{P} = -1$.

The set of proper orthogonal transformations forms a group called the *special orthogonal group* of dimension n and denoted by $SO(n)$.

In the two-dimensional orthogonal group $O(2)$ there is a subgroup of rotations by the angles $k\varphi$, where $k = 0, \pm1, \pm2, \ldots$ and φ is fixed. If a_k is its element corresponding to k and $a = a_1$, then the element a_k ($k > 0$) has the form

$$a_k = \underbrace{a \cdot a \cdot \ldots \cdot a}_{k \text{ times}} = a^k \qquad (k = 1, 2, 3, \ldots).$$

Denoting by a^{-1} the inverse of $a = a_1$, and the identity element by a^0, we see that each element of this group has the form

$$a_k = a^k \qquad (k = 0, \pm1, \pm2, \ldots).$$

Groups whose elements admit such a representation in terms of a single element are said to be *cyclic*. Such groups are discrete.

There are two cyclic groups of rotations (p and q are coprime numbers):

1. If $\varphi \neq 2\pi p/q$ (i.e., the angle φ is incommensurable with π), then all elements are distinct.
2. If $\varphi = 2\pi p/q$, then $a_{k+q} = a_k$ ($a^q = a^0$). Such groups are called *cyclic groups of order* q.

Consider *groups of mirror symmetry*. Each of them consists of two elements: the identity element and a reflection with respect to the origin.

Let $\{\mathbf{I}, \mathbf{P}\}$ be a subgroup of $O(3)$ consisting of the identity $\mathbf{I}$ and the reflection $\mathbf{P}$ of the three-dimensional space with respect to the origin, $\mathbf{Px} = -\mathbf{x}$. This is an improper subgroup. It is isomorphic to the group Z_2 of residues modulo 2. The subgroup $\{\mathbf{I}, \mathbf{P}\}$ is a normal subgroup in $O(3)$, and the subgroup $SO(3)$ (consisting of proper orthogonal transformations) is isomorphic to the quotient group $O(3)/\{\mathbf{I}, \mathbf{P}\}$.

5.8.2-3. Unitary groups.

By analogy with Paragraph 5.8.2-2, one can consider groups of linear transformations of a complex linear space.

In the general linear group of transformations of a unitary space, one considers *unitary groups* $U(n)$, which are analogues of orthogonal groups. In the group $U(n)$ of unitary transformations, one considers the subgroup $SU(n)$ that consists of unitary transformations whose determinant is equal to 1.

5.8.3. Group Representations

5.8.3-1. Linear representations of groups. Terminology.

A *linear representation* of a group G in the finite-dimensional Euclidean space $\mathcal{V}^n$ is a homomorphism of G to the group of nondegenerate linear transformations of $\mathcal{V}^n$; in other words, a linear representation of G is a mapping D that associates each element $a \in G$ with a nondegenerate linear transformation $D(a)$ of the space $\mathcal{V}^n$, so that for any a_1 and a_2 in G, we have $D(a_1a_2) = D(a_1)D(a_2)$.

Thus, for any $g \in G$, its image $D(g)$ is an element of the group $GL(n)$, and the set $D(G)$ consisting of all transformations $D(g)$, $g \in G$, is a subgroup of $GL(n)$ isomorphic to the quotient group $G/\ker D$, where $\ker D$ is the kernel of the homomorphism D, i.e., the set of all g such that $D(g)$ is the identity element of the group $GL(n)$.

The subgroup $D(G)$ is often also called a *representation* of the group G.

The space $\mathcal{V}^n$ is called the *representation space*; n is called the *dimension of the representation*; and the basis in $\mathcal{V}^n$ is called the *representation basis*.

The *trivial representation* of a group is its homomorphic mapping onto the identity element of the group $GL(n)$.

A *faithful representation* of a group G is an isomorphism of G onto a subgroup of $GL(n)$.

5.8.3-2. Matrices of linear representations. Equivalent representations.

If $D^{(\mu)}(G)$ is a representation of a group G, each $g \in G$ corresponds to a linear transformation $D^{(\mu)}(g)$, whose matrix in the basis of the representation $D^{(\mu)}(G)$ is denoted by $[D^{(\mu)}_{ij}(g)]$.

Two representations $D^{(\mu_1)}(G)$ and $D^{(\mu_2)}(G)$ of a group G in the same space E^n are said to be *equivalent* if there exists a nondegenerate linear transformation C of the space E^n such that $D^{(\mu_1)}(g) = C^{-1}D^{(\mu_2)}(g)C$ for each $g \in G$.

The choice of a basis in the representation space is important, since the matrices corresponding to the group elements may have some standard fairly simple form in that basis, and this allows one to make important conclusions with regards to a given representation.

5.8.3-3. Reducible and irreducible representations.

A subspace $\mathcal{V}'$ of $\mathcal{V}^n$ is called *invariant* for a representation $D(G)$ if it is invariant with respect to each linear operator in $D(G)$.

Suppose that all matrices of some three-dimensional representation $D(G)$ have the form

$$\begin{pmatrix} A_1 & A_2 \\ O & A_3 \end{pmatrix}, \qquad A_1 \equiv \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad A_2 \equiv \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix}, \quad A_3 \equiv (\,a_{33}\,), \quad O \equiv (\,0 \quad 0\,).$$

The product of such matrices has the form (see Paragraph 5.2.1-10)

$$\begin{pmatrix} A'_1 & A'_2 \\ O & A'_3 \end{pmatrix} \begin{pmatrix} A''_1 & A''_2 \\ O & A''_3 \end{pmatrix} = \begin{pmatrix} A'_1A''_1 & A'''_2 \\ O & A'_3A''_3 \end{pmatrix},$$

and therefore the structure of the matrices is preserved. Thus, the matrices A_1 form a two-dimensional representation of the given group G and the matrices A_3 form its one-dimensional representation. In such cases, one says that $D(G)$ is a *reducible representation*.

If all matrices of a representation have the form of size $n \times n$

$$\begin{pmatrix} A_1 & O \\ O & A_2 \end{pmatrix},$$

then the square matrices A_1 and A_2 form representations, the sum of their dimensions being equal to n. In this case, the representation is said to be *completely reducible*. The representation induced on an invariant space by a given representation $D(G)$ is called a *part* of the representation $D(G)$.

A representation $D(G)$ of a group G is said to be *irreducible* if it has only two invariant subspaces, $\mathcal{V}^n$ and O. Otherwise, it is said to be *reducible*. Any representation can be expressed in terms of irreducible representations.

5.8.3-4. Characters.

Let $D(G)$ be an n-dimensional representation of a group G, and let $[D_{ij}(g)]$ be the matrix of the operator corresponding to the element $g \in G$. The *character* of an element $g \in G$ in the representation $D(G)$ is defined by

$$\chi(g) = \sum_{i=1}^{n} D_{ii}(g) = \mathrm{Tr}([D_{ij}(g)]).$$

Thus, the character of an element does not depend on the representation basis and is, therefore, an invariant quantity.

An element $b \in G$ is said to be *conjugate* to the element $a \in G$ if there exists $u \in G$ such that

$$uau^{-1} = b.$$

Properties of conjugate elements:

1. Any element is conjugate to itself.
2. If b is conjugate to a, then a is conjugate to b.
3. If b is conjugate to a and c is conjugate to b, then c is conjugate to a.

The characters of all elements belonging to one and the same class of conjugate elements coincide. The characters of elements for equivalent representations coincide.

5.8.3-5. Examples of group representations.

1°. Let G be a group of symmetry of three-dimensional space consisting of two elements: the identity transformation **I** and the reflection **P** with respect to the origin, $G = \{\mathbf{I}, \mathbf{P}\}$.

The multiplication of elements of this group is described by the table

	I	**P**
I	**I**	**P**
P	**P**	**I**

1. *One-dimensional representation of the group G.*

In the space E^1, we chose a basis $\mathbf{e}_1$ and consider the matrix $A^{(1)}$ of the nondegenerate transformation $\mathbf{A}^1$ of this space: $A^{(1)} = (1)$. The transformation $\mathbf{A}^1$ forms a subgroup in the

group $GL(1)$ of all linear transformations of E^1, and the multiplication in this subgroup is described by the table

	$\mathbf{A}^{(1)}$
$\mathbf{A}^{(1)}$	$\mathbf{A}^{(1)}$

We obtain a one-dimensional representation $D^{(1)}(G)$ of the group G by letting $D^{(1)}(\mathbf{I}) = \mathbf{A}^{(1)}$, $D^{(1)}(\mathbf{P}) = \mathbf{A}^{(1)}$. These relations define a homomorphism of the group G to $GL(1)$ and thus define its representation.

2. *A two-dimensional representation of the group G.*

In E^2, we choose a basis $\mathbf{e}_1$, $\mathbf{e}_2$ and consider the matrices $A^{(2)}$, $B^{(2)}$ of linear transformations $\mathbf{A}^2$, $\mathbf{B}^2$ of this space: $A^{(2)} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, $B^{(2)} = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. The transformations $\mathbf{A}^2$, $\mathbf{B}^2$ form a subgroup in the group $GL(2)$ of linear transformations of E^2. The multiplication in this subgroup is defined by the table

	$\mathbf{A}^{(2)}$	$\mathbf{B}^{(2)}$
$\mathbf{A}^{(2)}$	$\mathbf{A}^{(2)}$	$\mathbf{B}^{(2)}$
$\mathbf{B}^{(2)}$	$\mathbf{B}^{(2)}$	$\mathbf{A}^{(2)}$

We obtain a two-dimensional representation $D^{(2)}(G)$ of the group G by letting $D^{(2)}(\mathbf{I}) = \mathbf{A}^{(2)}$, $D^{(2)}(\mathbf{P}) = \mathbf{B}^{(2)}$. These relations define an isomorphism of G onto the subgroup $\{\mathbf{A}^{(2)}, \mathbf{B}^{(2)}\}$ of $GL(2)$, and therefore define its representation.

3. *A three-dimensional representation of the group G.*

Consider the linear transformation $\mathbf{A}^{(3)}$ of E^3 defined by the matrix

$$A^{(3)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

This transformation forms a subgroup in $GL(3)$ with the multiplication law $\mathbf{A}^{(3)}\mathbf{A}^{(3)} = \mathbf{A}^{(3)}$. One obtains a three-dimensional representation $D^{(3)}(G)$ of the group G by letting $D^{(3)}(\mathbf{I}) = \mathbf{A}^{(3)}$, $D^{(3)}(\mathbf{P}) = \mathbf{A}^{(3)}$.

4. *A four-dimensional representation of the group G.*

Consider linear transformations $\mathbf{A}^{(4)}$ and $\mathbf{B}^{(4)}$ of E^4 defined by the matrices

$$A^{(4)} = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right), \quad B^{(4)} = \left(\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right).$$

The transformations $\mathbf{A}^{(4)}$ and $\mathbf{B}^{(4)}$ form a subgroup in $GL(4)$ with the multiplication defined by a table similar to that in the two-dimensional case. One obtains a four-dimensional representation $D^{(4)}(G)$ of the group G by letting $D^{(4)}(\mathbf{I}) = \mathbf{A}^{(4)}$, $D^{(4)}(\mathbf{P}) = \mathbf{A}^{(B)}$.

Remark. The matrices $A^{(4)}$ and $B^{(4)}$ may be written in the form $A^{(4)} = \left(\begin{smallmatrix} A^{(2)} & 0 \\ 0 & A^{(2)} \end{smallmatrix}\right)$, $B^{(4)} = \left(\begin{smallmatrix} B^{(2)} & 0 \\ 0 & B^{(2)} \end{smallmatrix}\right)$, and therefore the representation $D^{(4)}(G)$ is sometimes denoted by $D^{(4)}(G) = D^{(2)}(G) + D^{(2)}(G) = 2D^{(2)}(G)$. In a similar way, one may use the notation $D^{(3)}(G) = 3D^{(1)}(G)$. In this way, one can construct representations of the group G of arbitrary dimension.

2°. The symmetry group $G = \{\mathbf{I}, \mathbf{P}\}$ for the three-dimensional space is a normal subgroup of the group $O(3)$. The subgroup $SO(3) \subset O(3)$ formed by proper orthogonal transformations is isomorphic to the quotient group $O(3)/\{\mathbf{I}, \mathbf{P}\}$.

Since any group admits a homomorphic mapping onto its quotient group, there is a homomorphism of the group $O(3)$ onto $SO(3)$. This homomorphism is defined as follows: if a is a proper orthogonal transformation in $O(3)$, its image in $SO(3)$ coincides with a; and if a' is an improper orthogonal transformation, its image is the proper orthogonal transformation Pa'.

In this way, one obtains a three-dimensional representation $DO(3)$ of the group of orthogonal transformations $O(3)$ in terms of the group $SO(3)$ of proper orthogonal transformations.

References for Chapter 5

Anton, H., *Elementary Linear Algebra, 8th Edition*, Wiley, New York, 2000.

Barnett, S., *Matrices in Control Theory, Rev. Edition*, Krieger Publishing, Malabar, 1985.

Beecher, J. A., Penna, J. A., and Bittinger, M. L., *College Algebra, 2nd Edition*, Addison Wesley, Boston, 2004.

Bellman, R. E., *Introduction to Matrix Analysis* (McGraw-Hill Series in Matrix Theory), McGraw-Hill, New York, 1960.

Bernstein, D. S., *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*, Princeton University Press, Princeton, 2005.

Blitzer, R. F., *College Algebra, 3rd Edition*, Prentice Hall, Englewood Cliffs, New Jersey, 2003.

Bronson, R., *Schaum's Outline of Matrix Operations*, McGraw-Hill, New York, 1988.

Cullen, C. G., *Matrices and Linear Transformations, 2nd Edition*, Dover Publications, New York, 1990.

Davis, P. J. (Editor), *The Mathematics of Matrices: A First Book of Matrix Theory and Linear Algebra, 2nd Edition*, Xerox College Publications, Lexington, 1998.

Demmel, J. W., *Applied Numerical Linear Algebra*, Society for Industrial & Applied Mathematics, University City Science Center, Philadelphia, 1997.

Dixon, J. D., *Problems in Group Theory*, Dover Publications, New York, 1973.

Dugopolski, M., *College Algebra, 3rd Edition*, Addison Wesley, Boston, 2002.

Eves, H., *Elementary Matrix Theory*, Dover Publications, New York, 1980.

Franklin, J. N., *Matrix Theory*, Dover Publications, New York, 2000.

Gantmacher, F. R., *Matrix Theory, Vol. 1, 2nd Edition*, American Mathematical Society, Providence, Rhode Island, 1990.

Gantmacher, F. R., *Applications of the Theory of Matrices*, Dover Publications, New York, 2005.

Gelfand, I. M., *Lectures on Linear Algebra*, Dover Publications, New York, 1989.

Gelfand, I. M. and Shen, A., *Algebra*, Birkhauser, Boston, 2003.

Gilbert, J. and Gilbert, L., *Linear Algebra and Matrix Theory, 2nd Edition*, Brooks Cole, Stamford, 2004.

Golub, G. H. and Van Loan, C. F., *The Matrix Computations, 3nd Edition* (Johns Hopkins Studies in Mathematical Sciences), The Johns Hopkins University Press, Baltimore, Maryland, 1996.

Hazewinkel, M. (Editor), *Handbook of Algebra, Vol. 1*, North Holland, Amsterdam, 1996.

Hazewinkel, M. (Editor), *Handbook of Algebra, Vol. 2*, North Holland, Amsterdam, 2000.

Hazewinkel, M. (Editor), *Handbook of Algebra, Vol. 3*, North Holland, Amsterdam, 2003.

Horn, R. A. and Johnson, C. R., *Matrix Analysis, Rep. Edition*, Cambridge University Press, Cambridge, 1990.

Horn, R. A. and Johnson, C. R., *Topics in Matrix Analysis, New Ed. Edition*, Cambridge University Press, Cambridge, 1994.

Householder, A. S., *The Theory of Matrices in Numerical Analysis* (Dover Books on Mathematics), Dover Publications, New York, 2006.

Jacob, B., *Linear Functions and Matrix Theory* (Textbooks in Mathematical Sciences), Springer, New York, 1995.

Kostrikin, A. I. and Artamonov, V. A. (Editors), *Exercises in Algebra: A Collection of Exercises in Algebra, Linear Algebra and Geometry (Expanded)*, Gordon & Breach, New York, 1996.

Kostrikin, A. I. and Manin, Yu. I., *Linear Algebra and Geometry*, Gordon & Breach, New York, 1997.

Kostrikin, A. I. and Shafarevich, I. R. (Editor), *Algebra I: Basic Notions of Algebra*, Springer-Verlag, Berlin, 1990.

Kurosh, A. G., *Lectures on General Algebra*, Chelsea Publishing, New York, 1965.

Kurosh, A. G., *Algebraic Equations of Arbitrary Degrees*, Firebird Publishing, New York, 1977.

Kurosh, A. G., *Theory of Groups*, Chelsea Publishing, New York, 1979.

Lancaster, P. and Tismenetsky, M., *The Theory of Matrices, Second Edition: With Applications* (Computer Science and Scientific Computing), Academic Press, Boston, 1985.

Lay, D. C., *Linear Algebra and Its Applications, 3rd Edition*, Addison Wesley, Boston, 2002.
Lial, M. L., *Student's Solutions Manual for College Algebra, 3rd Sol. Mn Edition*, Addison Wesley, Boston, 2004.
Lial, M. L., Hornsby, J., and Schneider, D. I., *College Algebra, 9th Edition*, Addison Wesley, Boston, 2004.
Lipschutz, S., *3,000 Solved Problems in Linear Algebra*, McGraw-Hill, New York, 1989.
Lipschutz, S. and Lipson, M., *Schaum's Outline of Linear Algebra, 3rd Edition*, McGraw-Hill, New York, 2000.
MacDuffee, C. C., *The Theory of Matrices, Phoenix Edition*, Dover Publications, New York, 2004.
McWeeny, R., *Symmetry: An Introduction to Group Theory and Its Applications, Unabridged edition*, Dover Publications, New York, 2002.
Meyer, C. D., *Matrix Analysis and Applied Linear Algebra, Package Edition*, Society for Industrial & Applied Mathematics, University City Science Center, Philadelphia, 2001.
Mikhalev, A. V. and Pilz, G., *The Concise Handbook of Algebra*, Kluwer Academic, Dordrecht, Boston, 2002.
Perlis, S., *Theory of Matrices, Dover Ed. Edition*, Dover Publications, New York, 1991.
Poole, D., *Linear Algebra: A Modern Introduction*, Brooks Cole, Stamford, 2002.
Poole, D., *Student Solutions Manual for Poole's Linear Algebra: A Modern Introduction, 2nd Edition*, Brooks Cole, Stamford, 2005.
Rose, J. S., *A Course on Group Theory* (Dover Books on Advanced Mathematics), Dover Publications, New York, 1994.
Schneider, H. and Barker, G. Ph., *Matrices and Linear Algebra, 2nd Edition* (Dover Books on Advanced Mathematics), Dover Publications, New York, 1989.
Scott, W. R., *Group Theory*, Dover Publications, New York, 1987.
Shilov, G. E., *Linear Algebra, Rev. English Ed. Edition*, Dover Publications, New York, 1977.
Strang, G., *Introduction to Linear Algebra, 3rd Edition*, Wellesley Cambridge Pr., Wellesley, 2003.
Strang, G., *Linear Algebra and Its Applications, 4th Edition*, Brooks Cole, Stamford, 2005.
Sullivan, M., *College Algebra, 7th Edition*, Prentice Hall, Englewood Cliffs, New Jersey, 2004.
Tobey, J. and Slater, J., *Beginning Algebra, 6th Edition*, Prentice Hall, Englewood Cliffs, New Jersey, 2004.
Tobey, J. and Slater, J., *Intermediate Algebra, 5th Edition*, Prentice Hall, Englewood Cliffs, New Jersey, 2005.
Trefethen, L. N. and Bau, D., *Numerical Linear Algebra*, Society for Industrial & Applied Mathematics, University City Science Center, Philadelphia, 1997.
Turnbull, H. W. and Aitken, A. C., *An Introduction to the Theory of Canonical Matrices, Phoenix Edition*, Dover Publications, New York, 2004.
Vygodskii, M. Ya., *Mathematical Handbook: Higher Mathematics*, Mir Publishers, Moscow, 1971.
Zassenhaus, H. J., *The Theory of Groups, 2nd Ed. Edition*, Dover Publications, New York, 1999.
Zhang, F., *Matrix Theory*, Springer, New York, 1999.

Chapter 6
Limits and Derivatives

6.1. Basic Concepts of Mathematical Analysis

6.1.1. Number Sets. Functions of Real Variable

6.1.1-1. Real axis, intervals, and segments.

The *real axis* is a straight line with a point O chosen as the origin, a positive direction, and a scale unit.

There is a one-to-one correspondence between the set of all real numbers $\mathbb{R}$ and the set of all points of the real axis, with each real x being represented by a point on the real axis separated from O by the distance $|x|$ and lying to the right of O for $x > 0$, or to the left of O for $x < 0$.

One often has to deal with the following number sets (sets of real numbers or sets on the real axis).

1. Sets of the form (a, b), $(-\infty, b)$, $(a, +\infty)$, and $(-\infty, +\infty)$ consisting, respectively, of all $x \in \mathbb{R}$ such that $a < x < b$, $x < b$, $x > a$, and x is arbitrary are called *open intervals* (sometimes simply *intervals*).

2. Sets of the form $[a, b]$ consisting of all $x \in \mathbb{R}$ such that $a \le x \le b$ are called *closed intervals* or *segments*.

3. Sets of the form $(a, b]$, $[a, b)$, $(-\infty, b]$, $[a, +\infty)$ consisting of all x such that $a < x \le b$, $a \le x < b$, $x \le b$, $x \ge a$ are called *half-open intervals*.

A *neighborhood of a point* $x_\circ \in \mathbb{R}$ is defined as any open interval (a, b) containing $x_\circ$ ($a < x_\circ < b$). A neighborhood of the "point" $+\infty$, $-\infty$, or ∞ is defined, respectively, as any set of the form $(b, +\infty)$, $(-\infty, c)$ or $(-\infty, -a) \cup (a, +\infty)$ (here, $a \ge 0$).

6.1.1-2. Lower and upper bound of a set on a straight line.

The *upper bound* of a set of real numbers is the least number that bounds the set from above. The *lower bound* of a set of real numbers is the largest number that bounds the set from below.

In more details: let a set of real numbers $X \in \mathbb{R}$ be given. A number β is called its upper bound and denoted $\sup X$ if for any $x \in X$ the inequality $x \le \beta$ holds and for any $\beta_1 < \beta$ there exists an $x_1 \in X$ such that $x_1 > \beta_1$. A number α is called the lower bound of X and denoted $\inf X$ if for any $x \in X$ the inequality $x \ge \alpha$ holds and for any $\alpha_1 > \alpha$ there exists an $x_1 \in X$ such that $x_1 < \alpha_1$.

Example 1. For a set X consisting of two numbers a and b ($a < b$), we have

$$\inf X = a, \quad \sup X = b.$$

Example 2. For intervals (open, closed, and half-open), we have

$$\inf(a, b) = \inf[a, b] = \inf(a, b] = \inf[a, b) = a,$$
$$\sup(a, b) = \sup[a, b] = \sup(a, b] = \sup[a, b) = b.$$

One can see that the upper and lower bounds may belong to a given set (e.g., for closed intervals) and may not (e.g., for open intervals).

The symbol $+\infty$ (resp., $-\infty$) is called the upper (resp., lower) bound of a set unbounded from above (resp., from below).

6.1.1-3. Real-valued functions of real variable. Methods of defining a function.

1°. Let D and E be two sets of real numbers. Suppose that there is a relation between the points of D and E such that to each $x \in D$ there corresponds some $y \in E$, denoted by $y = f(x)$. In this case, one speaks of a *function* f defined on the set D and taking its values in the set E. The set D is called the *domain of the function* f, and the subset of E consisting of all elements $f(x)$ is called the *range of the function* f. This functional relation is often denoted by $y = f(x)$, $f : D \to E$, $f : x \mapsto y$.

The following terms are also used: x is the *independent variable* or the *argument*; y is the *dependent variable*.

2°. The most common and convenient way to define a function is the *analytic method*: the function is defined explicitly by means of a formula (or several formulas) depending on the argument x; for instance, $y = 2\sin x + 1$.

Implicit definition of a function consists of using an equation of the form $F(x, y) = 0$, from which one calculates the value y for any fixed value of the argument x.

Parametric definition of a function consists of defining the values of the independent variable x and the dependent variable y by a pair of formulas depending on an auxiliary variable t (parameter): $x = p(t)$, $y = q(t)$.

Quite often functions are defined in terms of convergent series or by means of tables or graphs. There are some other methods of defining functions.

3°. The *graph of a function* is the representation of a function $y = f(x)$ as a line on the plane with orthogonal coordinates x, y, the points of the line having the coordinates x, $y = f(x)$, where x is an arbitrary point from the domain of the function.

6.1.1-4. Single-valued, periodic, odd and even functions.

1°. A function is *single-valued* if each value of its argument corresponds to a unique value of the function. A function is *multi-valued* if there is at least one value of its argument corresponding to two or more values of the function. In what follows, we consider only single-valued functions, unless indicated otherwise.

2°. A function $f(x)$ is called *periodic* with period T (or T-*periodic*) if $f(x + T) = f(x)$ for any x.

3°. A function $f(x)$ is called *even* if it satisfies the condition $f(x) = f(-x)$ for any x. A function $f(x)$ is called *odd* if it satisfies the condition $f(x) = -f(-x)$ for any x.

6.1.1-5. Decreasing, increasing, monotone, and bounded functions.

1°. A function $f(x)$ is called *increasing or strictly increasing* (resp., *nondecreasing*) on a set $D \subset \mathbb{R}$ if for any $x_1, x_2 \in D$ such that $x_1 > x_2$, we have $f(x_1) > f(x_2)$ (resp., $f(x_1) \geq f(x_2)$). A function $f(x)$ is called *decreasing or strictly decreasing* (resp., *nonincreasing*) on a set D if for all $x_1, x_2 \in D$ such that $x_1 > x_2$, we have $f(x_1) < f(x_2)$ (resp., $f(x_1) \leq f(x_2)$). All such functions are called *monotone functions*. Strictly increasing or decreasing functions are called *strictly monotone*.

2°. A function $f(x)$ is called *bounded* on a set D if $|f(x)| < M$ for all $x \in D$, where M is a finite constant. A function $f(x)$ is called *bounded from above* (*bounded from below*) on a set D if $f(x) < M$ $(M < f(x))$ for all $x \in D$, where M is a real constant.

6.1.1-6. Composite and inverse functions.

1°. Consider a function $u = u(x)$, $x \in D$, with values $u \in E$, and let $y = f(u)$ be a function defined on E. Then the function $y = f\big(u(x)\big)$, $x \in D$, is called a *composite function* or the *superposition* of the functions f and u.

2°. Consider a function $y = f(x)$ that maps $x \in D$ into $y \in E$. The *inverse function* of $y = f(x)$ is a function $x = g(y)$ defined on E and such that $x = g(f(x))$ for all $x \in D$. The inverse function is often denoted by $g = f^{-1}$.

For strictly monotone functions $f(x)$, the inverse function always exists. In order to construct the inverse function $g(y)$, one should use the relation $y = f(x)$ to express x through y. The function $g(y)$ is monotonically increasing or decreasing together with $f(x)$.

6.1.2. Limit of a Sequence

6.1.2-1. Some definitions.

Suppose that there is a correspondence between each positive integer n and some (real or complex) number denoted, for instance, by x_n. In this case, one says that a *numerical sequence* (or, simply, a *sequence*) $x_1, x_2, \ldots, x_n, \ldots$ is defined. Such a sequence is often denoted by $\{x_n\}$; x_n is called the *generic term* of the sequence.

Example 1. For the sequence $\{n^2 - 2\}$, we have $x_1 = -1$, $x_2 = 2$, $x_3 = 7$, $x_4 = 14$, etc.

A sequence is called *bounded* (bounded from above, bounded from below) if there is a constant M such that $|x_n| < M$ (respectively, $x_n < M$, $x_n > M$) for all $n = 1, 2, \ldots$

6.1.2-2. Limit of a sequence.

A number b is called the *limit of a sequence* $x_1, x_2, \ldots, x_n, \ldots$ if for any $\varepsilon > 0$ there is $N = N(\varepsilon)$ such that $|x_n - b| < \varepsilon$ for all $n > N$.

If b is the limit of the sequence $\{x_n\}$, one writes $\lim\limits_{n\to\infty} x_n = b$ or $x_n \to b$ as $n \to \infty$.

The limit of a constant sequence $\{x_n = c\}$ exists and is equal to c, i.e., $\lim\limits_{n\to\infty} = c$. In this case, the inequality $|x_n - c| < \varepsilon$ takes the form $0 < \varepsilon$ and holds for all n.

Example 2. Let us show that $\lim\limits_{n\to\infty} \dfrac{n}{n+1} = 1$.

Consider the difference $\left|\dfrac{n}{n+1} - 1\right| = \dfrac{1}{n+1}$. The inequality $\dfrac{1}{n+1} < \varepsilon$ holds for all $n > \dfrac{1}{\varepsilon} - 1 = N(\varepsilon)$. Therefore, for any positive ε there is $N = \dfrac{1}{\varepsilon} - 1$ such that for $n > N$ we have $\left|\dfrac{n}{n+1} - 1\right| < \varepsilon$.

It may happen that a sequence $\{x_n\}$ has no limit at all, for instance, the sequence $\{x_n\} = \{(-1)^n\}$. A sequence that has a finite limit is called *convergent*.

THEOREM (BOLZANO–CAUCHY). *A sequence x_n has a finite limit if and only if for any $\varepsilon > 0$, there is N such that the inequality*

$$|x_n - x_m| < \varepsilon$$

holds for all $n > N$ and $m > N$.

6.1.2-3. Properties of convergent sequences.

1. Any convergent sequence can have only one limit.

2. Any convergent sequence is bounded. From any bounded sequence one can extract a convergent subsequence.*

3. If a sequence converges to b, then any of its subsequence also converges to b.

4. If $\{x_n\}$, $\{y_n\}$ are two convergent sequences, then the sequences $\{x_n \pm y_n\}$, $\{x_n \cdot y_n\}$, and $\{x_n/y_n\}$ (in this ratio, it is assumed that $y_n \neq 0$ and $\lim\limits_{n\to\infty} y_n \neq 0$) are also convergent and

$$\lim_{n\to\infty}(x_n \pm y_n) = \lim_{n\to\infty} x_n \pm \lim_{n\to\infty} y_n;$$

$$\lim_{n\to\infty}(cx_n) = c \lim_{n\to\infty} x_n \qquad (c = \text{const});$$

$$\lim_{n\to\infty}(x_n \cdot y_n) = \lim_{n\to\infty} x_n \cdot \lim_{n\to\infty} y_n;$$

$$\lim_{n\to\infty} \frac{x_n}{y_n} = \frac{\lim\limits_{n\to\infty} x_n}{\lim\limits_{n\to\infty} y_n}.$$

5. If $\{x_n\}$, $\{y_n\}$ are convergent sequences and the inequality $x_n \le y_n$ holds for all n, then $\lim\limits_{n\to\infty} x_n \le \lim\limits_{n\to\infty} y_n$.

6. If the inequalities $x_n \le y_n \le z_n$ hold for all n and $\lim\limits_{n\to\infty} x_n = \lim\limits_{n\to\infty} z_n = b$, then $\lim\limits_{n\to\infty} y_n = b$.

6.1.2-4. Increasing, decreasing, and monotone sequences.

A sequence $\{x_n\}$ is called *increasing or strictly increasing* (resp., *nondecreasing*) if the inequality $x_{n+1} > x_n$ (resp., $x_{n+1} \ge x_n$) holds for all n. A sequence $\{x_n\}$ is called *decreasing or strictly decreasing* (resp., *nonincreasing*) if the inequality $x_{n+1} < x_n$ (resp., $x_{n+1} \le x_n$) holds for all n. All such sequences are called *monotone sequences*. Strictly increasing or decreasing sequences are called *strictly monotone*.

THEOREM. *Any monotone bounded sequence has a finite limit.*

Example 3. It can be shown that the sequence $\left\{\left(1+\frac{1}{n}\right)^n\right\}$ is bounded and increasing. Therefore, it is convergent. Its limit is denoted by the letter e:

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \qquad (e \approx 2.71828).$$

Logarithms with the base e are called *natural* or *Napierian*, and $\log_e x$ is denoted by $\ln x$.

6.1.2-5. Properties of positive sequences.

1°. If a sequence x_n ($x_n > 0$) has a limit (finite or infinite), then the sequence

$$y_n = \sqrt[n]{x_1 \cdot x_2 \ldots x_n}$$

has the same limit.

* Let $\{x_n\}$ be a given sequence and let $\{n_k\}$ be a strictly increasing sequence with k and n_k being natural numbers. The sequence $\{x_{n_k}\}$ is called a *subsequence* of the sequence $\{x_n\}$.

2°. From property 1° for the sequence

$$x_1,\ \frac{x_2}{x_1},\ \frac{x_3}{x_2},\ \ldots,\ \frac{x_n}{x_{n-1}},\ \frac{x_{n+1}}{x_n},\ \ldots,$$

we obtain a useful corollary

$$\lim_{n\to\infty}\sqrt[n]{x_n}=\lim_{n\to\infty}\frac{x_{n+1}}{x_n},$$

under the assumption that the second limit exists.

Example 4. Let us show that $\lim\limits_{n\to\infty}\dfrac{n}{\sqrt[n]{n!}}=e$.

Taking $x_n=\dfrac{n^n}{n!}$ and using property 2°, we get

$$\lim_{n\to\infty}\frac{n}{\sqrt[n]{n!}}=\lim_{n\to\infty}\frac{x_{n+1}}{x_n}=\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n=e.$$

6.1.2-6. Infinitely small and infinitely large quantities.

A sequence x_n converging to zero as $n\to\infty$ is called *infinitely small or infinitesimal*.

A sequence x_n whose terms infinitely grow in absolute values with the growth of n is called *infinitely large* or "*tending to infinity*." In this case, the following notation is used: $\lim\limits_{n\to\infty}x_n=\infty$. If, in addition, all terms of the sequence starting from some number are positive (negative), then one says that the sequence x_n converges to "plus (minus) infinity," and one writes $\lim\limits_{n\to\infty}x_n=+\infty$ $\left(\lim\limits_{n\to\infty}x_n=-\infty\right)$. For instance, $\lim\limits_{n\to\infty}(-1)^n n^2=\infty$, $\lim\limits_{n\to\infty}\sqrt{n}=+\infty$, $\lim\limits_{n\to\infty}(-n)=-\infty$.

THEOREM (STOLZ). *Let x_n and y_n be two infinitely large sequences, $y_n\to+\infty$, and y_n increases with the growth of n (at least for sufficiently large n): $y_{n+1}>y_n$. Then*

$$\lim_{n\to\infty}\frac{x_n}{y_n}=\lim_{n\to\infty}\frac{x_n-x_{n-1}}{y_n-y_{n-1}},$$

provided that the right limit exists (finite or infinite).

Example 5. Let us find the limit of the sequence

$$z_n=\frac{1^k+2^k+\cdots+n^k}{n^{k+1}}.$$

Taking $x_n=1^k+2^k+\cdots+n^k$ and $y_n=n^{k+1}$ in the Stolz theorem, we get

$$\lim_{n\to\infty}z_n=\lim_{n\to\infty}\frac{n^k}{n^{k+1}-(n-1)^{k+1}}.$$

Since $(n-1)^{k+1}=n^{k+1}-(k+1)n^k+\cdots$, we have $n^{k+1}-(n-1)^{k+1}=(k+1)n^k+\cdots$, and therefore

$$\lim_{n\to\infty}z_n=\lim_{n\to\infty}\frac{n^k}{(k+1)n^k+\cdots}=\frac{1}{k+1}.$$

6.1.2-7. Upper and lower limits of a sequence.

The limit (finite or infinite) of a subsequence of a given sequence x_n is called a partial limit of x_n. In the set of all partial limits of any sequence of real numbers, there always exists the largest and the least (finite or infinite). The largest (resp., least) partial limit of a sequence is called its *upper* (resp., *lower*) *limit*. The upper and lower limits of a sequence x_n are denoted, respectively,

$$\overline{\lim_{n\to\infty}}\, x_n, \qquad \underline{\lim_{n\to\infty}}\, x_n.$$

Example 6. The upper and lower limits of the sequence $x_n = (-1)^n$ are, respectively,

$$\overline{\lim_{n\to\infty}}\, x_n = 1, \qquad \underline{\lim_{n\to\infty}}\, x_n = -1.$$

A sequence x_n has a limit (finite or infinite) if and only if its upper limit coincides with its lower limit:

$$\lim_{n\to\infty} x_n = \overline{\lim_{n\to\infty}}\, x_n = \underline{\lim_{n\to\infty}}\, x_n.$$

6.1.3. Limit of a Function. Asymptotes

6.1.3-1. Definition of the limit of a function. One-sided limits.

1°. One says that b is the *limit of a function* $f(x)$ as x tends to a if for any $\varepsilon > 0$ there is $\delta = \delta(\varepsilon) > 0$ such that $|f(x) - b| < \varepsilon$ for all x such that $0 < |x - a| < \delta$.

Notation: $\lim\limits_{x\to a} f(x) = b$ or $f(x) \to b$ as $x \to a$.

One says that b is the limit of a function $f(x)$ as x tends to $+\infty$ if for any $\varepsilon > 0$ there is $N = N(\varepsilon) > 0$ such that $|f(x) - b| < \varepsilon$ for all $x > N$.

Notation: $\lim\limits_{x\to+\infty} f(x) = b$ or $f(x) \to b$ as $x \to +\infty$.

In a similar way, one defines the limits for $x \to -\infty$ or $x \to \infty$.

THEOREM (BOLZANO–CAUCHY 1). *A function $f(x)$ has a finite limit as x tends to a (a is assumed finite) if and only if for any $\varepsilon > 0$ there is $\delta > 0$ such that the inequality*

$$|f(x_1) - f(x_2)| < \varepsilon \tag{6.1.3.1}$$

holds for all x_1, x_2 such that $|x_1 - a| < \delta$ and $|x_2 - a| < \delta$.

THEOREM (BOLZANO–CAUCHY 2). *A function $f(x)$ has a finite limit as x tends to $+\infty$ if and only if for any $\varepsilon > 0$ there is $\Delta > 0$ such that the inequality (6.1.3.1) holds for all $x_1 > \Delta$ and $x_2 > \Delta$.*

2°. One says that b is the *left-hand limit* (resp., *right-hand limit*) of a function $f(x)$ as x tends to a if for any $\varepsilon > 0$ there is $\delta = \delta(\varepsilon) > 0$ such that $|f(x) - b| < \varepsilon$ for $a - \delta < x < a$ (resp., for $a < x < a + \delta$).

Notation: $\lim\limits_{x\to a-0} f(x) = b$ or $f(a - 0) = b$ (resp., $\lim\limits_{x\to a+0} f(x) = b$ or $f(a + 0) = b$).

6.1.3-2. Properties of limits.

Let a be a number or any of the symbols ∞, $+\infty$, $-\infty$.

1. If a function has a limit at some point, this limit is unique.
2. If c is a constant function of x, then $\lim\limits_{x\to a} c = c$.

3. If there exist $\lim\limits_{x\to a} f(x)$ and $\lim\limits_{x\to a} g(x)$, then

$$\lim_{x\to a}\big[f(x)\pm g(x)\big] = \lim_{x\to a} f(x) \pm \lim_{x\to a} g(x);$$
$$\lim_{x\to a} cf(x) = c\lim_{x\to a} f(x) \quad (c = \text{const});$$
$$\lim_{x\to a} f(x)\cdot g(x) = \lim_{x\to a} f(x)\cdot \lim_{x\to a} g(x);$$
$$\lim_{x\to a}\frac{f(x)}{g(x)} = \frac{\lim\limits_{x\to a} f(x)}{\lim\limits_{x\to a} g(x)} \quad \Big(\text{if } \ g(x)\neq 0,\ \lim_{x\to a} g(x)\neq 0\Big).$$

4. Let $f(x)\le g(x)$ in a neighborhood of a point a ($x\neq a$). Then $\lim\limits_{x\to a} f(x)\le \lim\limits_{x\to a} g(x)$, provided that these limits exist.

5. If $f(x)\le g(x)\le h(x)$ in a neighborhood of a point a and $\lim\limits_{x\to a} f(x) = \lim\limits_{x\to a} h(x) = b$, then $\lim\limits_{x\to a} g(x) = b$.

These properties hold also for one-sided limits.

6.1.3-3. Limits of some functions.

First noteworthy limit: $\displaystyle\lim_{x\to 0}\frac{\sin x}{x} = 1.$

Second noteworthy limit: $\displaystyle\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x = e.$

Some other frequently used limits:

$$\lim_{x\to 0}\frac{(1+x)^n-1}{x} = n, \quad \lim_{x\to\infty}\frac{a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0}{b_nx^n+b_{n-1}x^{n-1}+\cdots+b_1x+b_0} = \frac{a_n}{b_n},$$
$$\lim_{x\to 0}\frac{1-\cos x}{x^2} = \frac{1}{2}, \quad \lim_{x\to 0}\frac{\tan x}{x} = 1, \quad \lim_{x\to 0}\frac{\arcsin x}{x} = 1, \quad \lim_{x\to 0}\frac{\arctan x}{x} = 1,$$
$$\lim_{x\to 0}\frac{e^x-1}{x} = 1, \quad \lim_{x\to 0}\frac{a^x-1}{x} = \ln a, \quad \lim_{x\to 0}\frac{\ln(1+x)}{x} = 1, \quad \lim_{x\to 0}\frac{\log_a(1+x)}{x} = \log_a e,$$
$$\lim_{x\to 0}\frac{\sinh x}{x} = 1, \quad \lim_{x\to 0}\frac{\tanh x}{x} = 1, \quad \lim_{x\to 0}\frac{\operatorname{arcsinh} x}{x} = 1, \quad \lim_{x\to 0}\frac{\operatorname{arctanh} x}{x} = 1,$$
$$\lim_{x\to +0} x^a\ln x = 0, \quad \lim_{x\to+\infty} x^{-a}\ln x = 0, \quad \lim_{x\to+\infty} x^a e^{-x} = 0, \quad \lim_{x\to +0} x^x = 1,$$

where $a>0$ and $b_n\neq 0$.

► See Paragraph 6.2.3-2, where L'Hospital rules for calculating limits with the help of derivatives are given.

6.1.3-4. Asymptotes of the graph of a function.

An *asymptote* of the graph of a function $y = f(x)$ is a straight line whose distance from the point (x, y) on the graph of $y = f(x)$ tends to zero if at least one of the coordinates (x, y) tends to zero.

The line $x = a$ is a *vertical asymptote* of the graph of the function $y = f(x)$ if at least one of the one-sided limits of $f(x)$ as $x \to a \pm 0$ is equal to $+\infty$ or $-\infty$.

The line $y = kx + b$ is an *oblique asymptote* of the graph of $y = f(x)$ if at least one of the limit relations holds: $\lim\limits_{x\to+\infty}[f(x) - kx - b] = 0$ or $\lim\limits_{x\to-\infty}[f(x) - kx - b] = 0$.

If there exist finite limits

$$\lim_{x\to+\infty}\frac{f(x)}{x} = k, \qquad \lim_{x\to+\infty}[f(x) - kx] = b, \tag{6.1.3.2}$$

then the line $y = kx + b$ is an oblique asymptote of the graph for $x \to +\infty$ (in a similar way, one defines an asymptote for $x \to -\infty$).

Example. Let us find the asymptotes of the graph of the function $y = \dfrac{x^2}{x-1}$.

1°. The graph has a vertical asymptote $x = 1$, since $\lim\limits_{x\to1}\dfrac{x^2}{x-1} = \infty$.

2°. Moreover, for $x \to \pm\infty$, there is an oblique asymptote $y = kx + b$ whose coefficients are determined by the formulas (6.1.3.2):

$$k = \lim_{x\to\pm\infty}\frac{x}{x-1} = 1, \qquad b = \lim_{x\to\pm\infty}\left(\frac{x^2}{x-1} - x\right) = \lim_{x\to\pm\infty}\frac{x}{x-1} = 1.$$

Thus, the equation of the oblique asymptote has the form $y = x + 1$. Fig. 6.1 shows the graph of the function under consideration and its asymptotes.

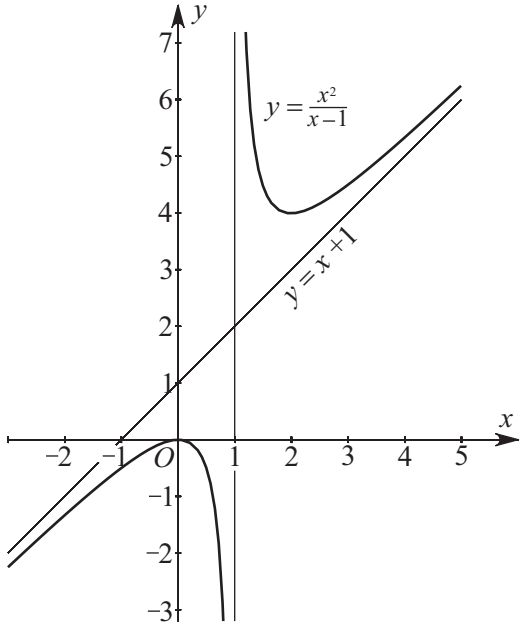


Figure 6.1. The graph and asymptotes of the function $y = \dfrac{x^2}{x-1}$.

6.1.4. Infinitely Small and Infinitely Large Functions

6.1.4-1. Definitions.

A function $f(x)$ is called *infinitely small* for $x \to a$ if $\lim\limits_{x\to a} f(x) = 0$.

A function $f(x)$ is said to be *infinitely large* for $x \to a$ if for any $K > 0$ the inequality $|f(x)| > K$ holds for all $x \neq a$ in a small neighborhood of the point a. In this case, one writes $f(x) \to \infty$ as $x \to a$ or $\lim\limits_{x \to a} f(x) = \infty$. (In these definitions, a is a finite number or any of the symbols ∞, $+\infty$, $-\infty$.) If $f(x)$ is infinitely large for $x \to a$ and $f(x) > 0$ ($f(x) < 0$) in a neighborhood of a (for $x \neq a$), one writes $\lim\limits_{x \to a} f(x) = +\infty$ (resp., $\lim\limits_{x \to a} f(x) = -\infty$).

6.1.4-2. Properties of infinitely small and infinitely large functions.

1. The sum and the product of finitely many infinitely small functions for $x \to a$ is an infinitely small function.
2. The product of an infinitely small function $f(x)$ for $x \to a$ and a function $g(x)$ which is bounded in a neighborhood U of the point a (i.e., $|g(x)| < M$ for all $x \in U$, where $M > 0$ is a constant) is an infinitely small function.
3. $\lim\limits_{x \to a} f(x) = b$ if and only if $f(x) = b + g(x)$, where $g(x)$ is infinitely small for $x \to a$.
4. A function $f(x)$ is infinitely large at some point if and only if the function $g(x) = \dfrac{1}{f(x)}$ is infinitely small at the same point.

6.1.4-3. Comparison of infinitely large quantities. Symbols of the order: O and o.

Functions $f(x)$ and $g(x)$ that are infinitely small for $x \to a$ are called *equivalent* near a if $\lim\limits_{x \to a} \dfrac{f(x)}{g(x)} = 1$. In this case one writes $f(x) \sim g(x)$.

Examples of equivalent infinitely small functions:

$$(1+\varepsilon)^n - 1 \sim n\varepsilon, \qquad a^\varepsilon - 1 \sim \varepsilon \ln a, \qquad \log_a(1+\varepsilon) \sim \varepsilon \log_a e,$$
$$\sin\varepsilon \sim \varepsilon, \quad \tan\varepsilon \sim \varepsilon, \quad 1 - \cos\varepsilon \sim \tfrac{1}{2}\varepsilon^2, \quad \arcsin\varepsilon \sim \varepsilon, \quad \arctan\varepsilon \sim \varepsilon,$$

where $\varepsilon = \varepsilon(x)$ is infinitely small for $x \to a$.

Functions $f(x)$ and $g(x)$ are said to be of the *same order* for $x \to a$, and one writes $f(x) = O\big(g(x)\big)$ if $\lim\limits_{x \to a} \dfrac{f(x)}{g(x)} = K$, $0 < |K| < \infty$.*

A function $f(x)$ is of a *higher order of smallness* compared with $g(x)$ for $x \to a$ if $\lim\limits_{x \to a} \dfrac{f(x)}{g(x)} = 0$, and in this case, one writes $f(x) = o\big(g(x)\big)$.

6.1.5. Continuous Functions. Discontinuities of the First and the Second Kind

6.1.5-1. Continuous functions.

A function $f(x)$ is called *continuous* at a point $x = a$ if it is defined in that point and its neighborhood and $\lim\limits_{x \to a} f(x) = f(a)$.

For continuous functions, a small variation of their argument $\Delta x = x - a$ corresponds to a small variation of the function $\Delta y = f(x) - f(a)$, i.e., $\Delta y \to 0$ as $\Delta x \to 0$. (This property is often used as a definition of continuity.)

* There is another definition of the symbol O. Namely, $f(x) = O\big(g(x)\big)$ for $x \to a$ if the inequality $|f(x)| \le K|g(x)|$, $K = \text{const}$, holds in some neighborhood of the point a (for $x \neq a$).

A function $f(x)$ is called *right-continuous* at a point $x = a$ if it is defined in that point (and to its right) and $\lim\limits_{x \to a+0} f(x) = f(a)$. A function $f(x)$ is called *left-continuous* at a point $x = a$ if it is defined in that point (and to its left) and $\lim\limits_{x \to a-0} f(x) = f(a)$.

6.1.5-2. Properties of continuous functions.

1. Suppose that functions $f(x)$ and $g(x)$ are continuous at some point a. Then the functions $f(x) \pm g(x)$, $cf(x)$, $f(x)g(x)$, $\dfrac{f(x)}{g(x)}$ $(g(a) \neq 0)$ are also continuous at a.

2. Suppose that a function $f(x)$ is continuous on the segment $[a, b]$ and takes values of different signs at its endpoints, i.e., $f(a)f(b) < 0$. Then there is a point c between a and b at which $f(x)$ vanishes:

$$f(c) = 0 \quad (a < c < b).$$

3. If $f(x)$ is continuous at a point a and $f(a) > 0$ (resp., $f(a) < 0$), then there is $\delta > 0$ such that $f(x) > 0$ (resp., $f(x) < 0$) for all $x \in (a - \delta, a + \delta)$.

4. Any function $f(x)$ that is continuous at each point of a segment $[a, b]$ attains its largest and its smallest values, M and m, on that segment.

5. A function $f(x)$ that is continuous on a segment $[a, b]$ takes any value $c \in [m, M]$ on that segment, where m and M are, respectively, its smallest and its largest values on $[a, b]$.

6. If $f(x)$ is continuous and increasing (resp., decreasing) on a segment $[a, b]$, then on the segment $\big[f(a), f(b)\big]$ (resp., $\big[f(b), f(a)\big]$) the inverse function $x = g(y)$ exists, and is continuous and increasing (resp., decreasing).

7. If $u(x)$ is continuous at a point a and $f(u)$ is continuous at $b = u(a)$, then the composite function $f\big(u(x)\big)$ is continuous at a.

Remark. Any elementary function is continuous at each point of its domain.

6.1.5-3. Points of discontinuity of a function.

A point a is called a *point of discontinuity of the first kind* for a function $f(x)$ if there exist finite one-sided limits $f(a+0)$ and $f(a-0)$, but the relations $\lim\limits_{x \to a+0} f(x) = \lim\limits_{x \to a-0} f(x) = f(a)$ do not hold. The value $|f(a + 0) - f(a - 0)|$ is called the *jump* of the function at the point a. In particular, if $f(a+0) = f(a-0) \neq f(a)$, then a is called a *point of removable discontinuity*.

Examples of function with discontinuities of the first kind.

1. The function $f(x) = \begin{cases} 0 & \text{for } x < 0, \\ 1 & \text{for } x \geq 0 \end{cases}$ has a jump equal to 1 at the discontinuity point $x = 0$.

2. The function $f(x) = \begin{cases} 0 & \text{for } x \neq 0, \\ 1 & \text{for } x = 0 \end{cases}$ has a removable discontinuity at the point $x = 0$.

A point a is called a *point of discontinuity of the second kind* if at least one of the one-sided limits $f(a + 0)$ or $f(a - 0)$ does not exist or is equal to infinity.

Examples of functions with discontinuities of the second kind.

1. The function $f(x) = \sin \dfrac{1}{x}$ has a second-kind discontinuity at the point $x = 0$ (since this function has no one-sided limits as $x \to \pm 0$).

2. The function $f(x) = 1/x$ has an infinite jump at the point $x = 0$.

6.1.5-4. Properties of monotone functions at points of discontinuity.

Any monotone function $f(x)$ always has a left-hand limit and a right-hand limit at its discontinuity point $x = x_0$; moreover, if $f(x)$ is a nonincreasing function, then

$$f(x_0 - 0) \geq f(x_0) \geq f(x_0 + 0);$$

if $f(x)$ is a nondecreasing function, then

$$f(x_0 - 0) \leq f(x_0) \leq f(x_0 + 0).$$

6.1.6. Convex and Concave Functions

6.1.6-1. Definition of convex and concave functions.

1°. A function $f(x)$ defined and continuous on a segment $[a, b]$ is called *convex* (or *convex downward*) if for any x_1, x_2 in $[a, b]$, the *Jensen inequality* holds:

$$f\left(\frac{x_1 + x_2}{2}\right) \leq \frac{f(x_1) + f(x_2)}{2}. \tag{6.1.6.1}$$

The geometrical meaning of convexity is that all points of the graph curve between two graph points lie below or on the rectilinear segment joining the two graph points (see Fig. 6.2 *a*).

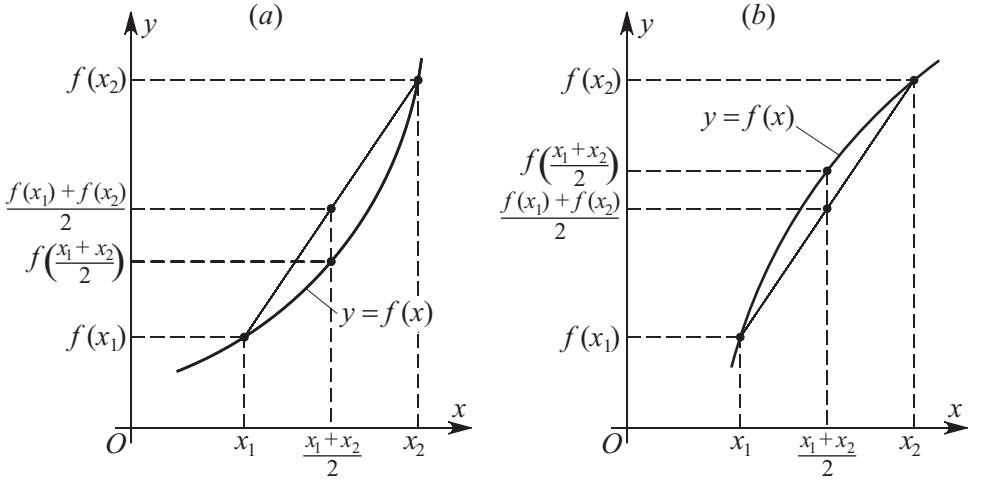


Figure 6.2. Graphs of convex (*a*) and concave (*b*) functions.

If for $x_1 \neq x_2$, condition (6.1.6.1) holds with $<$ instead of $\leq$, then the function $f(x)$ is called *strictly convex*.

2°. A function $f(x)$ defined and continuous on a segment $[a, b]$ is called *concave* (or *convex upward*) if for any x_1, x_2 in $[a, b]$ the following inequality holds:

$$f\left(\frac{x_1 + x_2}{2}\right) \geq \frac{f(x_1) + f(x_2)}{2}. \tag{6.1.6.2}$$

The geometrical meaning of concavity is that all points of the graph curve between two graph points lie above or on the rectilinear segment joining the two graph points (see Fig. 6.2 *b*).

If for $x_1 \neq x_2$, condition (6.1.6.2) holds with $>$ instead of $\geq$, then the function $f(x)$ is called *strictly concave*.

6.1.6-2. Generalized Jensen inequalities.

The inequalities (6.1.6.1) and (6.1.6.2) admit the following generalizations:

$$f(q_1x_1 + \cdots + q_nx_n) \le q_1f(x_1) + \cdots + q_nx_n \quad \text{for a convex function,}$$
$$f(q_1x_1 + \cdots + q_nx_n) \ge q_1f(x_1) + \cdots + q_nx_n \quad \text{for a concave function,}$$

where $q_1, \ldots, q_n$ are arbitrary positive numbers such that $q_1 + \cdots + q_n = 1$, and $x_1, \ldots, x_n$ are arbitrary points of the segment $[a, b]$.

6.1.6-3. Properties of convex and concave functions.

1. The product of a convex (concave) function and a positive constant is a convex (concave) function.

2. The sum of two or more convex (concave) functions is a convex (concave) function.

3. If $\varphi(u)$ is a convex increasing function and $u = f(x)$ is a convex function, then the composite function $\varphi(f(x))$ is convex. Some other properties of composite functions:

$$\varphi(u) \textit{ is convex and decreasing}, \quad u = f(x) \textit{ is concave} \quad \Longrightarrow \quad \varphi(f(x)) \textit{ is convex},$$
$$\varphi(u) \textit{ is concave and increasing}, \quad u = f(x) \textit{ is concave} \quad \Longrightarrow \quad \varphi(f(x)) \textit{ is concave},$$
$$\varphi(u) \textit{ is concave and decreasing}, \quad u = f(x) \textit{ is convex} \quad \Longrightarrow \quad \varphi(f(x)) \textit{ is concave}.$$

4. A non-constant convex (resp., concave) function $f(x)$ on a segment $[a, b]$ cannot attain its largest (resp., smallest) value inside the segment.

5. If $y = f(x)$ and $x = g(y)$ are single-valued mutually inverse functions (on the corresponding intervals), then the following properties hold:

$$f(x) \textit{ is convex and increasing} \quad \Longleftrightarrow \quad g(y) \textit{ is concave and increasing},$$
$$f(x) \textit{ is convex and decreasing} \quad \Longleftrightarrow \quad g(y) \textit{ is convex and decreasing},$$
$$f(x) \textit{ is concave and increasing} \quad \Longleftrightarrow \quad g(y) \textit{ is convex and increasing},$$
$$f(x) \textit{ is concave and decreasing} \quad \Longleftrightarrow \quad g(y) \textit{ is concave and decreasing}.$$

6. A function $f(x)$ that is continuous on a segment $[a, b]$ and twice differentiable on the interval (a, b) is convex downward (resp., convex upward) if and only if $f''(x) \ge 0$ (resp., $f''(x) \le 0$) on that interval.

7. Any convex function $f(x)$ satisfying the condition $f(x_0) = 0$ can be represented as the integral

$$f(x) = \int_{x_0}^{x} h(t)\,dt,$$

where $h(t)$ is a nondecreasing right-continuous function.

6.1.7. Functions of Bounded Variation

6.1.7-1. Definition of a function of bounded variation.

1°. Let $f(x)$ be a function defined on a finite segment $[a, b]$. Consider an arbitrary partition of the segment by the points

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$$

and construct the sum

$$v = \sum_{k=0}^{n-1} \left| f(x_{k+1}) - f(x_k) \right| \tag{6.1.7.1}$$

whose terms are absolute values of the increments of $f(x)$ on each segment of the partition.

If, for all partitions, the sums (6.1.7.1) are bounded by a constant independent of the partition, one says that the function $f(x)$ has *bounded variation* on the segment $[a, b]$. The supremum of all such sums over all partitions is called the *total variation* of the function $f(x)$ on the segment $[a,\ b]$. The total variation is denoted by

$$\bigvee_a^b f(x) = \sup\{v\}.$$

A function $f(x)$ is said to have bounded variation on the infinite interval $[a, \infty)$ if it is a function of bounded variation on any finite segment $[a, b]$ and its total variation on $[a, b]$ is bounded by a constant independent of b. By definition,

$$\bigvee_a^\infty f(x) = \sup_{b>a} \Big\{ \bigvee_a^b f(x) \Big\}.$$

2°. In the above definitions, the continuity of the function $f(x)$ is not mentioned. A continuous function (without additional conditions) may have bounded or unbounded variation.

Example. Consider the continuous function

$$f(x) = \begin{cases} x \cos \dfrac{\pi}{2x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0 \end{cases}$$

and the partition of the segment $[0, 1]$ by the points

$$0 < \frac{1}{2n} < \frac{1}{2n-1} < \cdots < \frac{1}{3} < \frac{1}{2} < 1.$$

Then the sums (6.1.7.1) corresponding to this partition have the form

$$v_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n} \to \infty \quad \text{as} \quad n \to \infty.$$

Therefore, $\bigvee_0^1 f(x) = \infty$.

6.1.7-2. Classes of functions of bounded variation.

Next, we list some common classes of functions of bounded variation.

1. Any bounded monotone function has bounded variation. Its total variation on the segment $[a, b]$ is defined by $\bigvee_a^b f(x) = |f(b) - f(a)|$.

Remark. The last statement is true for infinite intervals $(-\infty, a]$ and $[a, \infty)$; in the latter case, the total variation is equal to $\bigvee_a^\infty f(x) = |f(\infty) - f(a)|$.

2. Suppose that $f(x)$ is a bounded function on $[a, b]$ and this segment can be divided into finitely many parts

$$[a_k, a_{k+1}] \quad (k = 0,\ 1,\ \ldots\ , m-1; \quad a_0 = a, \quad a_m = b),$$

so that the function $f(x)$ is monotone on each part. Then $f(x)$ has bounded variation on $[a, b]$.

Remark. This statement is also true for infinite segments.

3. Let $f(x)$ be a function on a finite segment $[a, b]$ satisfying the *Lipschitz condition*

$$\big|f(x_1) - f(x_2)\big| \le L|x_1 - x_2|,$$

for any x_1 and x_2 in $[a, b]$, where L is a constant. Then $f(x)$ has bounded variation and $\bigvee_a^b f(x) \le L(b-a)$.

4. Let $f(x)$ be a function on a finite segment $[a, b]$ with a bounded derivative $|f'(x)| \le L$, where $L = \text{const}$. Then, $f(x)$ is of bounded variation and $\bigvee_a^b f(x) \le L(b-a)$.

5. Let $f(x)$ be a function on $[a, b]$ or $[a, \infty)$ and suppose that $f(x)$ can be represented as an integral with variable upper limit,

$$f(x) = c + \int_a^x \varphi(t)\,dt,$$

where $\varphi(t)$ is an absolutely continuous function on the interval under consideration. Then $f(x)$ has bounded variation and

$$\bigvee_a^b f(x) = \int_a^b |\varphi(x)|\,dx.$$

COROLLARY. *Suppose that $\varphi(t)$ on a finite segment $[a, b]$ or $[a, \infty)$ is integrable, but not absolutely integrable. Then the total variation of $f(x)$ is infinite.*

6.1.7-3. Properties of functions of bounded variation.

Here, all functions are considered on a finite segment $[a, b]$.

1. Any function of bounded variation is bounded.

2. The sum, difference, or product of finitely many functions of bounded variation is a function of bounded variation.

3. Let $f(x)$ and $g(x)$ be two functions of bounded variation and $|g(x)| \ge K > 0$. Then the ratio $f(x)/g(x)$ is a function of bounded variation.

4. Let $a < c < b$. If $f(x)$ has bounded variation on the segment $[a, b]$, then it has bounded variation on each segment $[a, c]$ and $[c, b]$; and the converse statement is true. In this case, the following additivity condition holds:

$$\bigvee_a^b f(x) = \bigvee_a^c f(x) + \bigvee_c^b f(x).$$

5. Let $f(x)$ be a function of bounded variation of the segment $[a, b]$. Then, for $a \le x \le b$, the variation of $f(x)$ with variable upper limit

$$F(x) = \bigvee_a^x f(x)$$

is a monotonically increasing bounded function of x.

6. Any function $f(x)$ of bounded variation on the segment $[a, b]$ has a left-hand limit $\lim\limits_{x \to x_0 - 0} f(x)$ and a right-hand limit $\lim\limits_{x \to x_0 + 0} f(x)$ at any point $x_0 \in [a, b]$.

6.1.7-4. Criteria for functions to have bounded variation.

1. A function $f(x)$ has bounded variation on a finite segment $[a, b]$ if and only if there is a monotonically increasing bounded function $\Phi(x)$ such that for all $x_1, x_2 \in [a, b]$ $(x_1 < x_2)$, the following inequality holds:

$$|f(x_2) - f(x_1)| \le \Phi(x_2) - \Phi(x_1).$$

2. A function $f(x)$ has bounded variation on a finite segment $[a, b]$ if and only if $f(x)$ can be represented as the difference of two monotonically increasing bounded functions on that segment: $f(x) = g_2(x) - g_1(x)$.

Remark. The above criteria are valid also for infinite intervals $(-\infty, a]$, $[a, \infty)$, and $(-\infty, \infty)$.

6.1.7-5. Properties of continuous functions of bounded variation.

1. Let $f(x)$ be a function of bounded variation on the segment $[a, b]$. If $f(x)$ is continuous at a point x_0 $(a < x_0 < b)$, then the function $F(x) = \bigvee_a^x f(x)$ is also continuous at that point.

2. A continuous function of bounded variation can be represented as the difference of two continuous increasing functions.

3. Let $f(x)$ be a continuous function on the segment $[a, b]$. Consider a partition of the segment

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$$

and the sum $v = \sum_{k=0}^{n-1} \big|f(x_{k+1}) - f(x_k)\big|$. Letting $\lambda = \max |x_{k+1} - x_k|$ and passing to the limit as $\lambda \to 0$, we get

$$\lim_{\lambda \to 0} v = \bigvee_a^b f(x).$$

6.1.8. Convergence of Functions

6.1.8-1. Pointwise, uniform, and nonuniform convergence of functions.

Let $\{f_n(x)\}$ be a sequence of functions defined on a set $X \subset \mathbb{R}$. The sequence $\{f_n(x)\}$ is said to be *pointwise convergent* to $f(x)$ as $n \to \infty$ if for any fixed $x \in X$, the numerical sequence $\{f_n(x)\}$ converges to $f(x)$. The sequence $\{f_n(x)\}$ is said to be *uniformly convergent* to a function $f(x)$ on X as $n \to \infty$ if for any $\varepsilon > 0$ there is an integer $N = N(\varepsilon)$ and such that for all $n > N$ and all $x \in X$, the following inequality holds:

$$|f_n(x) - f(x)| < \varepsilon. \tag{6.1.8.1}$$

Note that in this definition, N is independent of x. For a sequence $\{f_n(x)\}$ pointwise convergent to $f(x)$ as $n \to \infty$, by definition, for any $\varepsilon > 0$ and any $x \in X$, there is $N = N(\varepsilon, x)$ such that (6.1.8.1) holds for all $n > N(\varepsilon, x)$. If one cannot find such N independent of x and depending only on ε (i.e., one cannot ensure (6.1.8.1) uniformly; to be more precise, there is $\delta > 0$ such that for any $N > 0$ there is $k_N > N$ and $x_N \in X$ such that $|f_{k_N}(x_N) - f(x_N)| \ge \delta$), then one says that the sequence $\{f_n(x)\}$ *converges nonuniformly* to $f(x)$ on the set X.

6.1.8-2. Basic theorems.

Let X be an interval on the real axis.

THEOREM. *Let $f_n(x)$ be a sequence of continuous functions uniformly convergent to $f(x)$ on X. Then $f(x)$ is continuous on X.*

COROLLARY. *If the limit function $f(x)$ of a pointwise convergent sequence of continuous functions $\{f_n(x)\}$ is discontinuous, then the convergence of the sequence $\{f_n(x)\}$ is nonuniform.*

Example. The sequence $\{f_n(x)\} = \{x^n\}$ converges to $f(x) \equiv 0$ as $n \to \infty$ uniformly on each segment $[0, a]$, $0 < a < 1$. However, on the segment $[0, 1]$ this sequence converges nonuniformly to the discontinuous function $f(x) = \begin{cases} 0 & \text{for } 0 \le x < 1, \\ 1 & \text{for } x = 1. \end{cases}$

CAUCHY CRITERION. *A sequence of functions $\{f_n(x)\}$ defined on a set $X \in \mathbb{R}$ uniformly converges to $f(x)$ as $n \to \infty$ if and only if for any $\varepsilon > 0$ there is an integer $N = N(\varepsilon) > 0$ such that for all $n > N$ and $m > N$, the inequality $|f_n(x) - f_m(x)| < \varepsilon$ holds for all $x \in X$.*

6.1.8-3. Geometrical meaning of uniform convergence.

Let $f_n(x)$ be continuous functions on the segment $[a, b]$ and suppose that $\{f_n(x)\}$ uniformly converges to a continuous function $f(x)$ as $n \to \infty$. Then all curves $y = f_n(x)$, for sufficiently large $n > N$, belong to the strip between the two curves $y = f(x) - \varepsilon$ and $y = f(x) + \varepsilon$ (see Fig. 6.3).

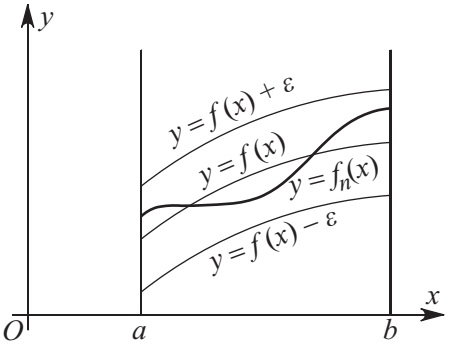


Figure 6.3. Geometrical meaning of uniform convergence of a sequence of functions $\{f_n(x)\}$ to a continuous function $f(x)$.

6.2. Differential Calculus for Functions of a Single Variable

6.2.1. Derivative and Differential, Their Geometrical and Physical Meaning

6.2.1-1. Definition of derivative and differential.

The *derivative* of a function $y = f(x)$ at a point x is the limit of the ratio

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x},$$

where $\Delta y = f(x+\Delta x) - f(x)$ is the increment of the function corresponding to the increment of the argument Δx. The derivative y' is also denoted by y'_x, $\dot{y}$, $\dfrac{dy}{dx}$, $f'(x)$, $\dfrac{df(x)}{dx}$.

Example 1. Let us calculate the derivative of the function $f(x) = x^2$.
By definition, we have

$$f'(x) = \lim_{\Delta x \to 0} \frac{(x + \Delta x)^2 - x^2}{\Delta x} = \lim_{\Delta x \to 0} (2x + \Delta x) = 2x.$$

The increment Δx is also called the differential of the independent variable x and is denoted by dx.

A function $f(x)$ that has a derivative at a point x is called *differentiable* at that point. The differentiability of $f(x)$ at a point x is equivalent to the condition that the increment of the function, $\Delta y = f(x + dx) - f(x)$, at that point can be represented in the form $\Delta y = f'(x)\,dx + o(dx)$ (the second term is an infinitely small quantity compared with dx as $dx \to 0$).

A function differentiable at some point x is continuous at that point. The converse is not true, in general; continuity does not always imply differentiability.

A function $f(x)$ is called *differentiable* on a set D (interval, segment, etc.) if for any $x \in D$ there exists the derivative $f'(x)$. A function $f(x)$ is called *continuously differentiable* on D if it has the derivative $f'(x)$ at each point $x \in D$ and $f'(x)$ is a continuous function on D.

The *differential* dy of a function $y = f(x)$ is the principal part of its increment Δy at the point x, so that $dy = f'(x)dx$, $\Delta y = dy + o(dx)$.

The approximate relation $\Delta y \approx dy$ or $f(x + \Delta x) \approx f(x) + f'(x)\Delta x$ (for small Δx) is often used in numerical analysis.

6.2.1-2. Physical and geometrical meaning of the derivative. Tangent line.

1°. Let $y = f(x)$ be the function describing the path y traversed by a body by the time x. Then the derivative $f'(x)$ is the velocity of the body at the instant x.

2°. The *tangent line* or simply the *tangent* to the graph of the function $y = f(x)$ at a point $M(x_0, y_0)$, where $y_0 = f(x_0)$, is defined as the straight line determined by the limit position of the secant MN as the point N tends to M along the graph. If α is the angle between the x-axis and the tangent line, then $f'(x_0) = \tan\alpha$ is the slope ratio of the tangential line (Fig. 6.4).

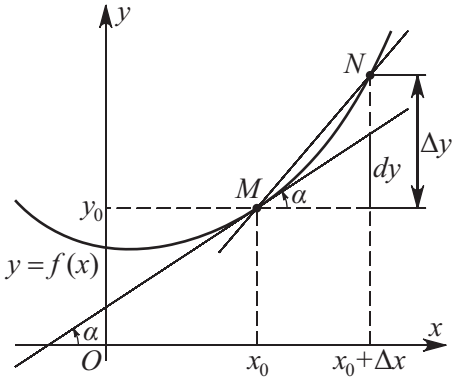


Figure 6.4. The tangent to the graph of a function $y = f(x)$ at a point (x_0, y_0).

Equation of the tangent line to the graph of a function $y = f(x)$ at a point (x_0, y_0):

$$y - y_0 = f'(x_0)(x - x_0).$$

Equation of the normal to the graph of a function $y = f(x)$ at a point (x_0, y_0):

$$y - y_0 = -\frac{1}{f'(x_0)}(x - x_0).$$

6.2.1-3. One-sided derivatives.

One-sided derivatives are defined as follows:

$$f'_+(x) = \lim_{\Delta x \to +0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to +0} \frac{f(x+\Delta x) - f(x)}{\Delta x} \quad \textit{right-hand derivative},$$
$$f'_-(x) = \lim_{\Delta x \to -0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to -0} \frac{f(x+\Delta x) - f(x)}{\Delta x} \quad \textit{left-hand derivative}.$$

Example 2. The function $y = |x|$ at the point $x = 0$ has different one-sided derivatives: $y'_+(0) = 1$, $y'_-(0) = -1$, but has no derivative at that point. Such points are called *angular points*.

Suppose that a function $y = f(x)$ is continuous at $x = x_0$ and has equal one-sided derivatives at that point, $y'_+(x_0) = y'_-(x_0) = a$. Then this function has a derivative at $x = x_0$ and $y'(x_0) = a$.

6.2.2. Table of Derivatives and Differentiation Rules

The derivative of any elementary function can be calculated with the help of derivatives of basic elementary functions and differentiation rules.

6.2.2-1. Table of derivatives of basic elementary functions (a = const).

$$(a)' = 0, \qquad (x^a)' = ax^{a-1},$$
$$(e^x)' = e^x, \qquad (a^x)' = a^x \ln a,$$
$$(\ln x)' = \frac{1}{x}, \qquad (\log_a x)' = \frac{1}{x \ln a},$$
$$(\sin x)' = \cos x, \qquad (\cos x)' = -\sin x,$$
$$(\tan x)' = \frac{1}{\cos^2 x}, \qquad (\cot x)' = -\frac{1}{\sin^2 x},$$
$$(\arcsin x)' = \frac{1}{\sqrt{1-x^2}}, \qquad (\arccos x)' = -\frac{1}{\sqrt{1-x^2}},$$
$$(\arctan x)' = \frac{1}{1+x^2}, \qquad (\operatorname{arccot} x)' = -\frac{1}{1+x^2},$$
$$(\sinh x)' = \cosh x, \qquad (\cosh x)' = \sinh x,$$
$$(\tanh x)' = \frac{1}{\cosh^2 x}, \qquad (\coth x)' = -\frac{1}{\sinh^2 x},$$
$$(\operatorname{arcsinh} x)' = \frac{1}{\sqrt{1+x^2}}, \qquad (\operatorname{arccosh} x)' = \frac{1}{\sqrt{x^2-1}},$$
$$(\operatorname{arctanh} x)' = \frac{1}{1-x^2}, \qquad (\operatorname{arccoth} x)' = \frac{1}{x^2-1}.$$

6.2.2-2. Differentiation rules.

1. *Derivative of a sum (difference) of functions*:
$$[u(x) \pm v(x)]' = u'(x) \pm v'(x).$$
2. *Derivative of the product of a function and a constant*:
$$[au(x)]' = au'(x) \qquad (a = \text{const}).$$
3. *Derivative of a product of functions*:
$$[u(x)v(x)]' = u'(x)v(x) + u(x)v'(x).$$
4. *Derivative of a ratio of functions*:
$$\left[\frac{u(x)}{v(x)}\right]' = \frac{u'(x)v(x) - u(x)v'(x)}{v^2(x)}.$$
5. *Derivative of a composite function*:
$$\big[f(u(x))\big]' = f'_u(u)u'(x).$$
6. *Derivative of a parametrically defined function* $x = x(t)$, $y = y(t)$:
$$y'_x = \frac{y'_t}{x'_t}.$$
7. *Derivative of an implicit function defined by the equation* $F(x, y) = 0$:
$$y'_x = -\frac{F_x}{F_y} \qquad (F_x \text{ and } F_y \text{ are partial derivatives}).$$
8. *Derivative of the inverse function* $x = x(y)$ (for details see footnote*):
$$x'_y = \frac{1}{y'_x}.$$
9. *Derivative of a composite exponential function*:
$$[u(x)^{v(x)}]' = u^v \ln u \cdot v' + vu^{v-1}u'.$$
10. *Derivative of a composite function of two arguments*:
$$[f(u(x), v(x))]' = f_u(u, v)u' + f_v(u, v)v' \qquad (f_u \text{ and } f_v \text{ are partial derivatives}).$$

Example 1. Let us calculate the derivative of the function $\dfrac{x^2}{2x+1}$.

Using the rule of differentiating the ratio of two functions, we obtain
$$\left(\frac{x^2}{2x+1}\right)' = \frac{(x^2)'(2x+1) - x^2(2x+1)'}{(2x+1)^2} = \frac{2x(2x+1) - 2x^2}{(2x+1)^2} = \frac{2x^2 + 2x}{(2x+1)^2}.$$

Example 2. Let us calculate the derivative of the function $\ln\cos x$.

Using the rule of differentiating composite functions and the formula for the logarithmic derivative from Paragraph 6.2.2-1, we get
$$(\ln\cos x)' = \frac{1}{\cos x}(\cos x)' = -\tan x.$$

Example 3. Let us calculate the derivative of the function x^x. Using the rule of differentiating the composite exponential function with $u(x) = v(x) = x$, we have
$$(x^x)' = x^x \ln x + xx^{x-1} = x^x(\ln x + 1).$$

* Let $y = f(x)$ be a differentiable monotone function on the interval (a, b) and $f'(x_0) \neq 0$, where $x_0 \in (a, b)$. Then the inverse function $x = g(y)$ is differentiable at the point $y_0 = f(x_0)$ and $g'(y_0) = \dfrac{1}{f'(x_0)}$.

6.2.3. Theorems about Differentiable Functions. L'Hospital Rule

6.2.3-1. Main theorems about differentiable functions.

ROLLE THEOREM. *If the function $y = f(x)$ is continuous on the segment $[a, b]$, differentiable on the interval (a, b), and $f(a) = f(b)$, then there is a point $c \in (a, b)$ such that $f'(c) = 0$.*

LAGRANGE THEOREM. *If the function $y = f(x)$ is continuous on the segment $[a, b]$ and differentiable on the interval (a, b), then there is a point $c \in (a, b)$ such that*

$$f(b) - f(a) = f'(c)(b - a).$$

This relation is called the *formula of finite increments*.

CAUCHY THEOREM. *Let $f(x)$ and $g(x)$ be two functions that are continuous on the segment $[a, b]$, differentiable on the interval (a, b), and $g'(x) \neq 0$ for all $x \in (a, b)$. Then there is a point $c \in (a, b)$ such that*

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}.$$

6.2.3-2. L'Hospital's rules on indeterminate expressions of the form $0/0$ and ∞/∞.

THEOREM 1. *Let $f(x)$ and $g(x)$ be two functions defined in a neighborhood of a point a, vanishing at this point, $f(a) = g(a) = 0$, and having the derivatives $f'(a)$ and $g'(a)$, with $g'(a) \neq 0$. Then*

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)}.$$

Example 1. Let us calculate the limit $\lim\limits_{x \to 0} \dfrac{\sin x}{1 - e^{-2x}}$.

Here, both the numerator and the denominator vanish for $x = 0$. Let us calculate the derivatives

$$f'(x) = (\sin x)' = \cos x \quad \Longrightarrow \quad f'(0) = 1,$$
$$g'(x) = (1 - e^{-2x})' = 2e^{-2x} \quad \Longrightarrow \quad g'(0) = 2 \neq 0.$$

By the L'Hospital rule, we find that

$$\lim_{x \to 0} \frac{\sin x}{1 - e^{-2x}} = \frac{f'(0)}{g'(0)} = \frac{1}{2}.$$

THEOREM 2. *Let $f(x)$ and $g(x)$ be two functions defined in a neighborhood of a point a, vanishing at a, together with their derivatives up to the order $n - 1$ inclusively. Suppose also that the derivatives $f^{(n)}(a)$ and $g^{(n)}(a)$ exist and are finite, $g^{(n)}(a) \neq 0$. Then*

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{f^{(n)}(a)}{g^{(n)}(a)}.$$

THEOREM 3. *Let $f(x)$ and $g(x)$ be differentiable functions and $g'(x) \neq 0$ in a neighborhood of a point a ($x \neq a$). If $f(x)$ and $g(x)$ are infinitely small or infinitely large functions for $x \to a$, i.e., the ratio $\dfrac{f(x)}{g(x)}$ at the point a is an indeterminate expression of the form $\dfrac{0}{0}$ or $\dfrac{\infty}{\infty}$, then*

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}$$

(provided that there exists a finite or infinite limit of the ratio of the derivatives).

Remark. The L'Hospital rule 3 is applicable also in the case of a being one of the symbols ∞, $+\infty$, $-\infty$.

6.2.3-3. Methods for interpreting other indeterminate expressions.

1°. Expressions of the form $0\cdot\infty$ and $\infty-\infty$ can be reduced to indeterminate expressions $\dfrac{0}{0}$ or $\dfrac{\infty}{\infty}$ by means of algebraic transformations, for instance:

$$u(x)\cdot v(x)=\frac{u(x)}{1/v(x)} \qquad \text{transformation rule} \quad 0\cdot\infty \Longrightarrow \frac{0}{0},$$

$$u(x)-v(x)=\left(\frac{1}{u(x)}-\frac{1}{v(x)}\right):\frac{1}{u(x)v(x)} \qquad \text{transformation rule} \quad \infty-\infty \Longrightarrow \frac{0}{0}.$$

2°. Indeterminate expressions of the form 1^{∞}, ∞^0, 0^0 can be reduced to expressions of the form $\dfrac{0}{0}$ or $\dfrac{\infty}{\infty}$ by taking logarithm and using the formulas $\ln u^v = v\ln u = \dfrac{\ln u}{1/v}$.

Example 2. Let us calculate the limit $\lim\limits_{x\to 0}(\cos x)^{1/x^2}$.

We have the indeterminate expression 1^{∞}. We find that

$$\ln\lim_{x\to 0}(\cos x)^{1/x^2}=\lim_{x\to 0}\ln(\cos x)^{1/x^2}=\lim_{x\to 0}\frac{\ln\cos x}{x^2}=\lim_{x\to 0}\frac{(\ln\cos x)'}{(x^2)'}=\lim_{x\to 0}\frac{(-\tan x)}{2x}=-\frac{1}{2}.$$

Therefore, $\lim\limits_{x\to 0}(\cos x)^{1/x^2}=e^{-1/2}=\dfrac{1}{\sqrt{e}}$.

6.2.4. Higher-Order Derivatives and Differentials. Taylor's Formula

6.2.4-1. Derivatives and differentials of higher orders.

The *second-order derivative* or the second derivative of a function $y=f(x)$ is the derivative of the derivative $f'(x)$. The second derivative is denoted by y'' and also by y''_{xx}, $\dfrac{d^2y}{dx^2}$, $f''(x)$.

The derivative of the second derivative of a function $y=f(x)$ is called the third-order derivative, $y'''=(y'')'$. The nth-order derivative of the function $y=f(x)$ is defined as the derivative of its $(n-1)$th derivative:

$$y^{(n)}=(y^{(n-1)})'.$$

The nth-order derivative is also denoted by $y_x^{(n)}$, $\dfrac{d^ny}{dx^n}$, $f^{(n)}(x)$.

The *second-order differential* is the differential of the first-order differential, $d^2y=d(dy)$. If x is the independent variable, then $d^2y=y''\cdot(dx)^2$. In a similar way, one defines differentials of higher orders.

6.2.4-2. Table of higher-order derivatives of some elementary functions.

$$(x^a)^{(n)}=a(a-1)\dots(a-n+1)x^{a-n}, \qquad (a^x)^{(n)}=(\ln a)^n a^x,$$

$$(\ln x)^{(n)}=(-1)^{n-1}(n-1)!\,\frac{1}{x^n}, \qquad (\log_a x)^{(n)}=(-1)^{n-1}\frac{(n-1)!}{\ln a}\,\frac{1}{x^n},$$

$$(\sin x)^{(n)}=\sin\left(x+\frac{\pi n}{2}\right), \qquad (\cos x)^{(n)}=\cos\left(x+\frac{\pi n}{2}\right),$$

$$(\sinh x)^{(n)}=\begin{cases}\cosh x & \text{if } n \text{ is odd},\\ \sinh x & \text{if } n \text{ is even},\end{cases} \qquad (\cosh x)^{(n)}=\begin{cases}\cosh x & \text{if } n \text{ is even},\\ \sinh x & \text{if } n \text{ is odd}.\end{cases}$$

6.2.4-3. Rules for calculating higher-order derivatives.

1. *Derivative of a sum (difference) of functions*:

$$[u(x) \pm v(x)]^{(n)} = u^{(n)}(x) \pm v^{(n)}(x).$$

2. *Derivatives of a function multiplied by a constant*:

$$[au(x)]^{(n)} = au^{(n)}(x) \qquad (a = \text{const}).$$

3. *Derivatives of a product*:

$$\begin{aligned}
&[u(x)v(x)]'' = u''(x)v(x) + 2u'(x)v'(x) + u(x)v''(x),\\
&[u(x)v(x)]''' = u'''(x)v(x) + 3u''(x)v'(x) + 3u'(x)v''(x) + u(x)v'''(x),\\
&[u(x)v(x)]^{(n)} = \sum_{k=0}^{n} C_n^k u^{(k)}(x)v^{(n-k)}(x) \qquad \text{(Leibnitz formula)},
\end{aligned}$$

where C_n^k are binomial coefficients, $u^{(0)}(x) = u(x)$, $v^{(0)}(x) = v(x)$.

4. *Derivatives of a composite function*:

$$\begin{aligned}
&\big[f(u(x))\big]'' = f''_{uu}(u'_x)^2 + f'_u u''_{xx},\\
&\big[f(u(x))\big]''' = f'''_{uuu}(u'_x)^3 + 3f''_{uu}u'_x u''_{xx} + f'_u u'''_{xxx}.
\end{aligned}$$

5. *Derivatives of a parametrically defined function* $x = x(t)$, $y = y(t)$:

$$y'' = \frac{x'_t y''_{tt} - y'_t x''_{tt}}{(x'_t)^3}, \quad y''' = \frac{(x'_t)^2 y'''_{ttt} - 3x'_t x''_{tt} y''_{tt} + 3y'_t (x''_{tt})^2 - x'_t y'_t x'''_{ttt}}{(x'_t)^5}, \quad y^{(n)} = \frac{(y^{(n-1)})'_t}{x'_t}.$$

6. *Derivatives of an implicit function defined by the equation* $F(x, y) = 0$:

$$\begin{aligned}
y'' =& \frac{1}{F_y^3}\big(-F_y^2 F_{xx} + 2F_x F_y F_{xy} - F_x^2 F_{yy}\big),\\
y''' =& \frac{1}{F_y^5}\big(-F_y^4 F_{xxx} + 3F_x F_y^3 F_{xxy} - 3F_x^2 F_y^3 F_{xyy} + F_x^3 F_y F_{yyy} + 3F_y^3 F_{xx} F_{xy}\\
&- 3F_x F_y^2 F_{xx} F_{yy} - 6F_x F_y^2 F_{xy}^2 - 3F_x^3 F_{yy}^2 + 9F_x^2 F_y F_{xy} F_{yy}\big),
\end{aligned}$$

where the subscripts denote the corresponding partial derivatives.

7. *Derivatives of the inverse function* $x = x(y)$:

$$x''_{yy} = -\frac{y''_{xx}}{(y'_x)^3}, \quad x'''_{yyy} = -\frac{y'''_{xxx}}{(y'_x)^4} + 3\frac{(y''_{xx})^2}{(y'_x)^5}, \quad x_y^{(n)} = \frac{1}{y'_x}[x_y^{(n-1)}]'_x.$$

6.2.4-4. Taylor's formula.

Suppose that in a neighborhood of a point $x = a$, the function $y = f(x)$ has derivatives up to the order $(n + 1)$ inclusively. Then for all x in that neighborhood, the following representation holds:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(x), \quad (6.2.4.1)$$

where $R_n(x)$ is the *remainder term* in the Taylor formula.

The remainder term can be represented in different forms (6.2.4.1):

$$R_n(x) = o[(x-a)^n] \qquad \text{(Peano)},$$

$$R_n(x) = \frac{f^{(n+1)}\big(a + k(x-a)\big)}{(n+1)!}(x-a)^{n+1} \qquad \text{(Lagrange)},$$

$$R_n(x) = \frac{f^{(n+1)}\big(a + k(x-a)\big)}{n!}(1-k)^n(x-a)^{n+1} \qquad \text{(Cauchy)},$$

$$R_n(x) = \frac{f^{(n+1)}\big(a + k(x-a)\big)}{n!p}(1-k)^{n+1-p}(x-a)^{n+1} \qquad \text{(Schlömilch and Roche)},$$

$$R_n(x) = \frac{1}{n!}\int_a^x f^{(n+1)}(t)(x-t)^n\,dt \qquad \text{(integral form)},$$

where $0 < k < 1$ and $p > 0$; k depends on x, n, and the structure of the remainder term. The remainders in the form of Lagrange and Cauchy can be obtained as special cases of the Schlömilch formula with $p = n + 1$ and $p = 1$, respectively.

For $a = 0$, the Taylor formula (6.2.4.1) turns into

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + R_n(x)$$

and is called the *Maclaurin formula*.

The Maclaurin formula for some functions:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + R_n(x),$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^n\frac{x^{2n+1}}{(2n+1)!} + R_{2n+1}(x),$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^n\frac{x^{2n}}{(2n)!} + R_{2n}(x).$$

6.2.5. Extremal Points. Points of Inflection

6.2.5-1. Maximum and minimum. Points of extremum.

Let $f(x)$ be a differentiable function on the interval (a, b) and $f'(x) > 0$ (resp., $f'(x) < 0$) on (a, b). Then $f(x)$ is an *increasing* (resp., *decreasing*) function on that interval*.

Suppose that there is a neighborhood of a point x_0 such that for all $x \neq x_0$ in that neighborhood we have $f(x) > f(x_0)$ (resp., $f(x) < f(x_0)$). Then x_0 is called a point of *local minimum* (resp., *local maximum*) of the function $f(x)$.

Points of local minimum or maximum are called points of *extremum*.

* At some points of the interval, the derivative may vanish.

6.2.5-2. Necessary and sufficient conditions for the existence of extremum.

NECESSARY CONDITION OF EXTREMUM. *A function $f(x)$ can have an extremum only at points in which its derivative either vanishes or does not exist (or is infinite).*

FIRST SUFFICIENT CONDITION OF EXTREMUM. *Suppose that $f(x)$ is continuous in some neighborhood $(x_0-\delta,\ x_0+\delta)$ of a point x_0 and differentiable at all points of the neighborhood except, possibly, x_0. If $f'(x) > 0$ for $x \in (x_0 - \delta,\ x_0)$ and $f'(x) < 0$ for $x \in (x_0,\ x_0 + \delta)$, then x_0 is a point of local maximum of this function. If $f'(x) < 0$ for $x \in (x_0 - \delta,\ x_0)$ and $f'(x) > 0$ for $x \in (x_0,\ x_0 + \delta)$, then x_0 is a point of local minimum of this function.*

If $f'(x)$ is of the same sign for all $x \neq x_0$, $x \in (x_0 - \delta,\ x_0 + \delta)$, then x_0 cannot be a point of extremum.

SECOND SUFFICIENT CONDITION OF EXTREMUM. *Let $f(x)$ be a twice differentiable function in a neighborhood of x_0. Then the following implications hold:*

(i) $f'(x_0) = 0$ *and* $f''(x_0) < 0 \quad \Longrightarrow \quad f(x)$ *has a local maximum at the point* x_0;

(ii) $f'(x_0) = 0$ *and* $f''(x_0) > 0 \quad \Longrightarrow \quad f(x)$ *has a local minimum at the point* x_0.

THIRD SUFFICIENT CONDITION OF EXTREMUM. *Let $f(x)$ be a function that is n times differentiable in a neighborhood of a point x_0 and $f'(x_0) = f''(x_0) = \cdots = f^{(n-1)}(x_0) = 0$, but $f^{(n)}(x_0) \neq 0$. Then the following implications hold:*

(i) n *is even and* $f^{(n)}(x_0) < 0 \quad \Longrightarrow \quad f(x)$ *has a local maximum at the point* x_0;

(ii) n *is even and* $f^{(n)}(x_0) > 0 \quad \Longrightarrow \quad f(x)$ *has a local minimum at the point* x_0.

If n is odd, then x_0 cannot be a point of extremum.

6.2.5-3. Largest and the smallest values of a function.

Let $y = f(x)$ be continuous on the segment $[a, b]$ and differentiable at all points of this segment except, possibly, finitely many points. Then the largest and the smallest values of $f(x)$ on $[a, b]$ belong to the set consisting of $f(a)$, $f(b)$, and the values $f(x_i)$, where $x_i \in (a, b)$ are the points at which $f'(x)$ is either equal to zero or does not exist (is infinite).

6.2.5-4. Direction of the convexity of the graph of a function.

The graph of a differentiable function $y = f(x)$ is said to be *convex upward* (resp., *convex downward*) on the interval (a, b) if for each point of this interval, the graph lies below (resp., above) the tangent line at that point.

If the function $y = f(x)$ is twice differentiable on the interval (a, b) and $f''(x) < 0$ (resp., $f''(x) > 0$), then its graph is convex upward (resp., downward) on that interval. (At some points of the interval, the second derivative may vanish.)

Thus, in order to find the intervals on which the graph of a twice differentiable function $f(x)$ is convex upward (resp., downward), one should solve the inequality $f''(x) < 0$ (resp., $f''(x) > 0$).

6.2.5-5. Inflection points.

An *inflection point* on the graph of a function $y = f(x)$ is defined as a point $(x_0,\ f(x_0))$ at which the graph passes from one side of its tangent line to another. At an inflection point, the graph changes the direction of its convexity.

Suppose that the function $y = f(x)$ has a continuous second derivative $f''(x)$ in some neighborhood of a point x_0. If $f''(x_0) = 0$ and $f''(x)$ changes sign as x passes through the point x_0, then $(x_0,\ f(x_0))$ is an inflection point.

6.2.6. Qualitative Analysis of Functions and Construction of Graphs

6.2.6-1. General scheme of analysis of a function and construction of its graph.

1. Determine the domain in which the function is defined.
2. Find the asymptotes of the graph.
3. Find extremal points and intervals of monotonicity.
4. Determine the directions of convexity of the graph and its inflection points.
5. Determine whether the function is odd or even and whether it is periodic.
6. Find the points at which the graph crosses the coordinate axes.
7. Draw the graph, using the properties 1 to 6.

Example. Let us examine the function $y = \dfrac{\ln x}{x}$ and construct its graph.

We use the above general scheme.

1. This function is defined for all x such that $0 < x < +\infty$.
2. The straight line $x = 0$ is a vertical asymptote, since $\lim\limits_{x\to+0} \dfrac{\ln x}{x} = -\infty$. We find the oblique asymptotes:

$$k = \lim_{x\to+\infty} \frac{y}{x} = 0, \quad b = \lim_{x\to+\infty} (y - kx) = 0.$$

Therefore, the line $y = 0$ is a horizontal asymptote of the graph.

3. The derivative $y' = \dfrac{1 - \ln x}{x^2}$ vanishes for $\ln x = 1$. Therefore, the function may have an extremum at $x = e$. For $x \in (0, e)$, we have $y' > 0$, i.e., the function is increasing on this interval. For $x \in (e, +\infty)$, we have $y' < 0$, and therefore the function is decreasing on this interval. At $x = e$ the function attains its maximal value $y_{\max} = \dfrac{1}{e}$.

One should also examine the points at which the derivative does not exist. There is only one such point, $x = 0$, and it corresponds to the vertical asymptote (see Item 1).

4. The second derivative $y'' = \dfrac{2\ln x - 3}{x^3}$ vanishes for $x = e^{3/2}$. On the interval $(0, e^{3/2})$, we have $y'' < 0$, and therefore the graph is convex upward on this interval. For $x \in (e^{3/2}, +\infty)$, we have $y'' > 0$, and therefore the graph is convex downward on this interval. The value $x = e^{3/2}$ corresponds to an inflection point of the graph, with the ordinate $y = \frac{3}{2}e^{-3/2}$.

5. This function is neither odd nor even, since it is defined only for $x > 0$ and the relations $f(-x) = f(x)$ or $f(-x) = -f(x)$ cannot hold. Obviously, this function is nonperiodic.

6. The graph of this function does not cross the y-axis, since for $x = 0$ the function is undefined. Further, $y = 0$ only if $x = 1$, i.e., the graph crosses the x-axis only at the point $(1, 0)$.

7. Using the above results, we construct the graph (Fig. 6.5).

6.2.6-2. Transformations of graphs of functions.

Let us describe some methods which in many cases allow us to construct the graph of a function if we have the graph of a simpler function.

1. The graph of the function $y = f(x) + a$ is obtained from that of $y = f(x)$ by shifting the latter along the axis Oy by the distance $|a|$. For $a > 0$ the shift is upward, and for $a < 0$ downward (see Fig. 6.6 *a*).

2. The graph of the function $y = f(x + a)$ is obtained from that of $y = f(x)$ by shifting the latter along the Ox by the distance $|a|$. For $a > 0$ the shift is to the left, and for $a < 0$ to the right (see Fig. 6.6 *b*).

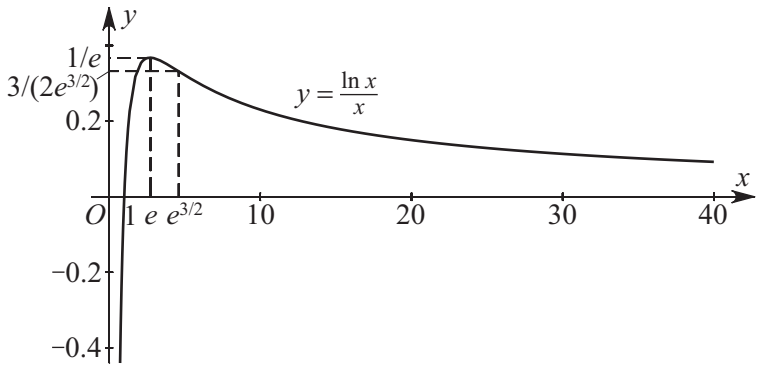


Figure 6.5. Graph of the function $y = \frac{\ln x}{x}$.

3. The graph of the function $y = -f(x)$ is obtained from that of $y = f(x)$ by symmetric reflection with respect to the axis Ox (see Fig. 6.6 *c*).

4. The graph of the function $y = f(-x)$ is obtained from that of $y = f(x)$ by symmetric reflection with respect to the axis Oy (see Fig. 6.6 *d*).

5. The graph of the function $y = kf(x)$ for $k > 1$ is obtained from that of $y = f(x)$ by extending the latter k times from the axis Ox, and for $0 < k < 1$ by contracting the latter $1/k$ times to the axis Ox. The points at which the graph crosses the axis Ox remain unchanged (see Fig. 6.6 *e*).

6. The graph of the function $y = f(kx)$ for $k > 1$ is obtained from that of $y = f(x)$ by contracting the latter k times to the axis Oy, and for $0 < k < 1$ by extending the latter $1/k$ times from the axis Oy. The points at which the graph crosses the axis Oy remain unchanged (see Fig. 6.6 *f*).

7. The graph of the function $y = |f(x)|$ is obtained from that of $y = f(x)$ by preserving the parts of the latter for which $f(x) \geq 0$ and symmetric reflection, with respect to the axis Ox, of the parts for which $f(x) < 0$ (see Fig. 6.6 *g*).

8. The graph of the inverse function $y = f^{-1}(x)$ is obtained from that of $y = f(x)$ by symmetric reflection with respect to the straight line $y = x$ (see Fig. 6.6 *h*).

6.2.7. Approximate Solution of Equations (Root-Finding Algorithms for Continuous Functions)

6.2.7-1. Preliminaries.

For a vast majority of algebraic (transcendental) equations of the form

$$f(x) = 0, \tag{6.2.7.1}$$

where $f(x)$ is a continuous function, there are no exact closed-form expressions for the roots.

When solving the equation approximately, the first step is to bracket the roots, i.e., find sufficiently small intervals containing exactly one root each. Such an interval $[a, b]$, where the numbers a and b satisfy the condition $f(a)f(b) < 0$ (which is assumed to hold in what follows), can be found, say, graphically.

The second step is to compute successive approximations $x_n \in [a, b]$ ($n = 1, 2, \ldots$) to the desired root $c = \lim\limits_{n\to\infty} x_n$, usually by one of the following methods.

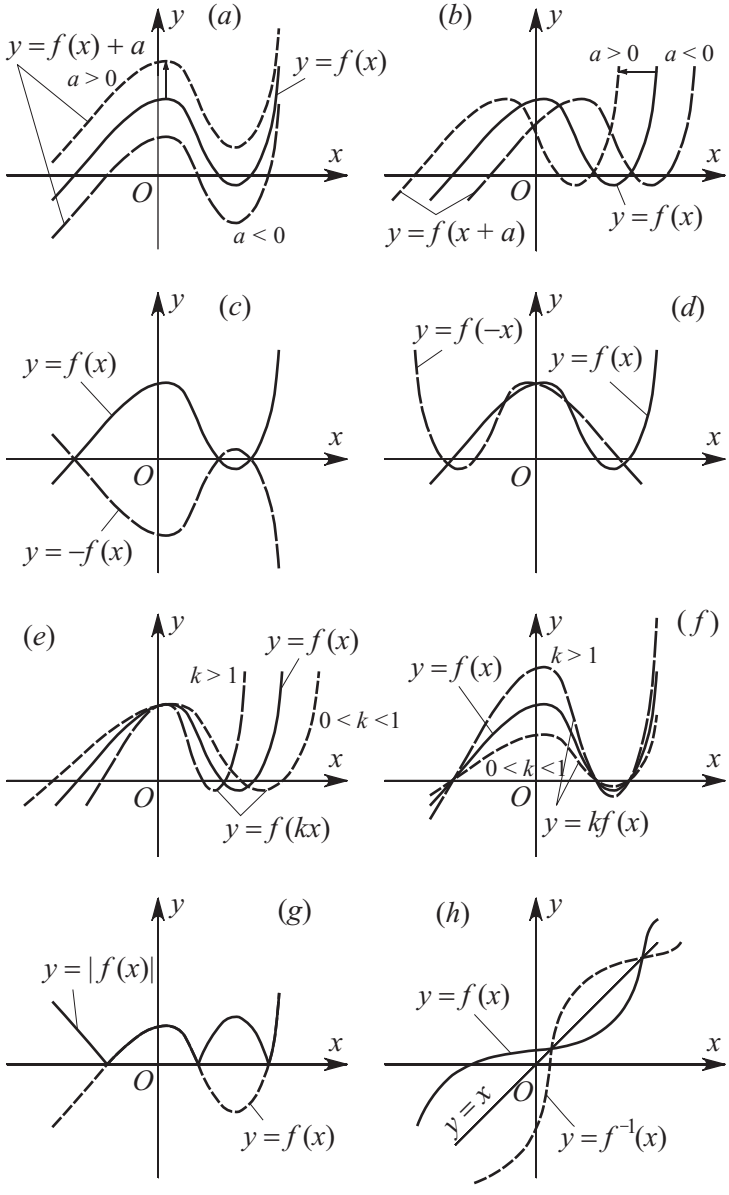


Figure 6.6. Transformations of graphs of functions.

6.2.7-2. Bisection method.

To find the root of equation (6.2.7.1) on the interval $[a, b]$, we bisect the interval. If $f\left(\dfrac{a+b}{2}\right) = 0$, then $c = \dfrac{a+b}{2}$ is the desired root. If $f\left(\dfrac{a+b}{2}\right) \neq 0$, then of the two intervals $\left[a, \dfrac{a+b}{2}\right]$ and $\left[\dfrac{a+b}{2}, b\right]$ we take the one at whose endpoints the function $f(x)$ has opposite signs. Now we bisect the new, smaller interval, etc. As a result, we obtain either an exact root of equation (6.2.7.1) at some step or an infinite sequence of nested intervals $[a_1, b_1]$, $[a_2, b_2]$, ... such that $f(a_n)f(b_n) < 0$. The root is given by the formula

$c = \lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} b_n$, and the estimate

$$0 \le c - a_n \le \frac{1}{2^n}(b-a)$$

is valid.

The following two methods are more efficient.

6.2.7-3. Regula falsi method (false position method).

Suppose that the derivatives $f'(x)$ and $f''(x)$ exist on the interval $[a, b]$ and the inequalities $f'(x) \ne 0$ and $f''(x) \ne 0$ hold for all $x \in [a, b]$.

If $f'(a)f''(a) > 0$, then we take $x_0 = a$ for the zero approximation; the subsequent approximations are given by the formulas

$$x_{n+1} = x_n - \frac{f(x_n)}{f(b) - f(x_n)}(b - x_n), \qquad n = 0,\ 1,\ \ldots$$

If $f'(a)f''(a) < 0$, then we take $x_0 = b$ for the zero approximation; the subsequent approximations are given by the formulas

$$x_{n+1} = x_n - \frac{f(x_n)}{f(a) - f(x_n)}(a - x_n), \qquad n = 0,\ 1,\ \ldots$$

The regula falsi method has the first order of local convergence as $n \to \infty$:

$$|x_{n+1} - c| \le k|x_n - c|,$$

where k is a constant depending on $f(x)$ and c is the root of equation (6.2.7.1).

The regula falsi method has a simple geometric interpretation. The straight line (secant) passing through the points $(a, f(a))$ and $(b, f(b))$ of the curve $y = f(x)$ meets the abscissa axis at the point x_1; the value x_{n+1} is the abscissa of the point where the line passing through the points $(x_0, f(x_0))$ and $(x_n, f(x_n))$ meets the x-axis (see Fig. 6.7 *a*).

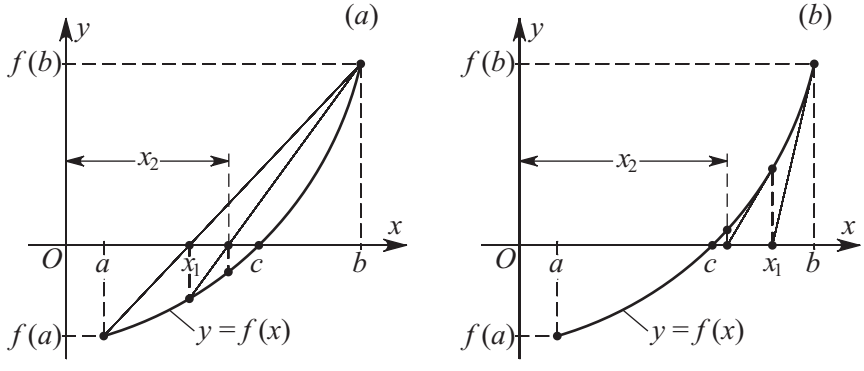


Figure 6.7. Graphical construction of successive approximations to the root of equation (6.2.7.1) by the regula falsi method (*a*) and the Newton–Raphson method (*b*).

6.2.7-4. Newton–Raphson method.

Suppose that the derivatives $f'(x)$ and $f''(x)$ exist on the interval $[a, b]$ and the inequalities $f'(x) \ne 0$ and $f''(x) \ne 0$ hold for all $x \in [a, b]$.

If $f(a)f''(a) > 0$, then we take $x_0 = a$ for the zero approximation; if $f(b)f''(b) > 0$, then $x_0 = b$. The subsequent approximations are computed by the formulas

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \qquad n = 0,\ 1,\ \ldots$$

If the initial approximation x_0 is sufficiently close to the desired root c, then the Newton–Raphson method exhibits quadratic convergence:

$$|x_{n+1} - c| \le \frac{M}{2m}\,|x_n - c|^2,$$

where $M = \max\limits_{a\le x\le b} |f''(x)|$ and $m = \min\limits_{a\le x\le b} |f'(x)|$.

The Newton–Raphson method has a simple geometric interpretation. The tangent to the curve $y = f(x)$ through the point $(x_n, f(x_n))$ meets the abscissa axis at the point x_{n+1} (see Fig. 6.7 *b*).

The Newton–Raphson method has a higher order of convergence than the regula falsi method. Hence the former is more often used in practice.

6.3. Functions of Several Variables. Partial Derivatives

6.3.1. Point Sets. Functions. Limits and Continuity

6.3.1-1. Sets on the plane and in space.

The distance between two points A and B on the plane and in space can be defined as follows:

$$\rho(A,B) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2} \qquad \text{(on the plane)},$$

$$\rho(A,B) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2} \qquad \text{(in three-dimensional space)},$$

$$\rho(A,B) = \sqrt{(x_{1A} - x_{1B})^2 + \cdots + (x_{nA} - x_{nB})^2} \qquad \text{(in } n\text{-dimensional space)}.$$

where x_A, y_A and x_B, y_B, and x_A, y_A, z_A and x_B, y_B, z_B, and x_{1A}, $\ldots$, x_{nA} and $x_{1B}, \ldots, x_{nB}$ are Cartesian coordinates of the corresponding points.

An *ε-neighborhood of a point* M_0 (on the plane or in space) is the set consisting of all points M (resp., on the plane or in space) such that $\rho(M, M_0) < \varepsilon$, where it is assumed that $\varepsilon > 0$. An *ε-neighborhood of a set* K (on the plane or in space) is the set consisting of all points M (resp., on the plane or in space) such that $\inf\limits_{M_0\in K} \rho(M, M_0) < \varepsilon$, where it is assumed that $\varepsilon > 0$.

An *interior point* of a set D is a point belonging to D, together with some neighborhood of that point. An *open set* is a set containing only interior points. A *boundary point* of a set D is a point such that any of its neighborhoods contains points outside D. A *closed set* is a set containing all its boundary points. A set D is called a *bounded set* if $\rho(A,B) < C$ for any points $A, B \in D$, where C is a constant independent of A, B. Otherwise (i.e., if there is no such constant), the set D is called *unbounded*.

6.3.1-2. Functions of two or three variables.

A (numerical) *function* on a set D is, by definition, a relation that sets up a correspondence between each point $M \in D$ and a unique numerical value. If D is a plane set, then each

point $M \in D$ is determined by two coordinates x, y, and a function $z = f(M) = f(x, y)$ is called a *function of two variables*. If D belongs to a three-dimensional space, then one speaks of a *function of three variables*. The set D on which the function is defined is called the *domain* of the function. For instance, the function $z = \sqrt{1 - x^2 - y^2}$ is defined on the closed circle $x^2 + y^2 \le 1$, which is its domain.

The *graph* of a function $z = f(x, y)$ is the surface formed by the points $(x, y, f(x, y))$ in three-dimensional space. For instance, the graph of the function $z = ax + by + c$ is a plane, and the graph of the function $z = \sqrt{1 - x^2 - y^2}$ is a half-sphere.

A *level line* of a function $z = f(x, y)$ is a line on the plane x, y with the following property: the function takes one and the same value $z = c$ at all points of that line. Thus, the equation of a level line has the form $f(x, y) = c$. A *level surface* of a function $u = f(x, y, z)$ is a surface on which the function takes a constant value, $u = c$; the equation of a level surface has the form $f(x, y, z) = c$.

A function $f(M)$ is called *bounded* on a set D if there is a constant C such that $|f(M)| \le C$ for all $M \in D$.

6.3.1-3. Limit of a function at a point and its continuity.

Let M be a point that comes infinitely close to some point M_0, i.e., $\rho = \rho(M_0, M) \to 0$. It is possible that the values $f(M)$ come close to some constant b.

One says that b is the *limit of the function* $f(M)$ at the point M_0 if for any (arbitrarily small) $\varepsilon > 0$, there is $\delta > 0$ such that for all points M belonging to the domain of the function and satisfying the inequality $0 < \rho(M_0, M) < \delta$, we have $|f(M) - b| < \varepsilon$. In this case, one writes $\lim\limits_{\rho(M, M_0) \to 0} f(M) = b$.

A function $f(M)$ is called *continuous* at a point M_0 if $\lim\limits_{\rho(M, M_0) \to 0} f(M) = f(M_0)$. A function is called *continuous on a set* D if it is continuous at each point of D. Any continuous function $f(M)$ on a closed bounded set is bounded on that set and attains its smallest and its largest values on that set.

6.3.2. Differentiation of Functions of Several Variables

For the sake of brevity, we consider the case of a function of two variables. However, all statements can be easily extended to the case of n variables.

6.3.2-1. Total and partial increments of a function. Partial derivatives.

A *total increment* of a function $z = f(x, y)$ at a point (x, y) is

$$\Delta z = f(x + \Delta x, y + \Delta y) - f(x, y),$$

where Δx, Δy are increments of the independent variables. *Partial increments* in x and in y are, respectively,

$$\Delta_x z = f(x + \Delta x, y) - f(x, y),$$
$$\Delta_y z = f(x, y + \Delta y) - f(x, y).$$

Partial derivatives of a function z in x and in y at a point (x, y) are defined as follows:

$$\frac{\partial z}{\partial x} = \lim_{\Delta x \to 0} \frac{\Delta_x z}{\Delta x}, \quad \frac{\partial z}{\partial y} = \lim_{\Delta y \to 0} \frac{\Delta_y z}{\Delta y}$$

(provided that these limits exist). Partial derivatives are also denoted by z_x and z_y, $\partial_x z$ and $\partial_y z$, or $f_x(x, y)$ and $f_y(x, y)$.

6.3.2-2. Differentiable functions. Differential.

A function $z = f(x, y)$ is called *differentiable* at a point (x, y) if its increment at that point can be represented in the form

$$\Delta z = A(x, y)\Delta x + B(x, y)\Delta y + o(\rho), \qquad \rho = \sqrt{(\Delta x)^2 + (\Delta y)^2},$$

where $o(\rho)$ is a quantity of a higher order of smallness compared with ρ as $\rho \to 0$ (i.e., $o(\rho)/\rho \to 0$ as $\rho \to 0$). In this case, there exist partial derivatives at the point (x, y), and $z'_x = A(x, y)$, $z'_y = B(x, y)$.

A function that has continuous partial derivatives at a point (x, y) is differentiable at that point.

The *differential of a function* $z = f(x, y)$ is defined as follows:

$$dz = f'_x(x, y)\Delta x + f'_y(x, y)\Delta y.$$

Taking the differentials dx and dy of the independent variables equal to Δx and Δy, respectively, one can also write $dz = f'_x(x, y)\,dx + f'_y(x, y)\,dy$.

The relation $\Delta z = dz + o(\rho)$ for small Δx and Δy is widely used for approximate calculations, in particular, for finding errors in numerical calculations of values of a function.

Example 1. Suppose that the values of the arguments of the function $z = x^2y^5$ are known with the error $x = 2 \pm 0.01$, $y = 1 \pm 0.01$. Let us calculate the approximate value of the function.

We find the increment of the function z at the point $x = 2$, $y = 1$ for $\Delta x = \Delta y = 0.01$, using the formula $\Delta z \approx dz = 2 \cdot 2 \cdot 1^5 \cdot 0.01 + 5 \cdot 2^2 \cdot 1^4 \cdot 0.01 = 0.24$. Therefore, we can accept the approximation $z = 4 \pm 0.24$.

If a function $z = f(x, y)$ is differentiable at a point (x_0, y_0), then

$$f(x, y) = f(x_0, y_0) + f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0) + o(\rho).$$

Hence, for small ρ (i.e., for $x \approx x_0$, $y \approx y_0$), we obtain the approximate formula

$$f(x, y) \approx f(x_0, y_0) + f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0).$$

The replacement of a function by this linear expression near a given point is called *linearization*.

6.3.2-3. Composite function.

Consider a function $z = f(x, y)$ and let $x = x(u, v)$, $y = y(u, v)$. Suppose that for $(u, v) \in D$, the functions $x(u, v)$, $y(u, v)$ take values for which the function $z = f(x, y)$ is defined. In this way, one defines a *composite function* on the set D, namely, $z(u, v) = f\big(x(u, v), y(u, v)\big)$. In this situation, $f(x, y)$ is called the outer function and $x(u, v)$, $y(u, v)$ are called the inner functions.

Partial derivatives of a composite function are expressed by

$$\frac{\partial z}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u},$$
$$\frac{\partial z}{\partial v} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v}.$$

For $z = z(t, x, y)$, let $x = x(t)$, $y = y(t)$. Thus, z is actually a function of only one variable t. The derivative $\frac{dz}{dt}$ is calculated by

$$\frac{dz}{dt} = \frac{\partial z}{\partial t} + \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

This derivative, in contrast to the partial derivative $\frac{\partial z}{\partial t}$, is called a *total derivative*.

6.3.2-4. Second partial derivatives and second differentials.

The *second partial derivatives* of a function $z = f(x, y)$ are defined as the derivatives of its first partial derivatives and are denoted as follows:

$$\frac{\partial^2 z}{\partial x^2} = z_{xx} \equiv (z_x)_x, \qquad \frac{\partial^2 z}{\partial x\,\partial y} = z_{xy} \equiv (z_x)_y,$$
$$\frac{\partial^2 z}{\partial y\,\partial x} = z_{yx} \equiv (z_y)_x, \qquad \frac{\partial^2 z}{\partial y^2} = z_{yy} \equiv (z_y)_y.$$

The derivatives z_{xy} and z_{yx} are called *mixed derivatives*. If the mixed derivatives are continuous at some point, then they coincide at that point, $z_{xy} = z_{yx}$.

In a similar way, one defines higher-order partial derivatives.

The *second differential* of a function $z = f(x, y)$ is the expression

$$d^2 z = d(dz) = (dz)_x \Delta x + (dz)_y \Delta y = z_{xx}(\Delta x)^2 + 2z_{xy}\Delta x \Delta y + z_{yy}(\Delta y)^2.$$

In a similar way, one defines $d^3 z$, $d^4 z$, etc.

6.3.2-5. Taylor's formula.

If at some point (x, y) the function $z = f(x, y)$ possesses partial derivatives up to the order n inclusively, then its increment Δz at that point can be expressed by

$$\Delta z = dz + \frac{d^2 z}{2!} + \frac{d^3 z}{3!} + \cdots + \frac{d^n z}{n!} + o(\rho^n),$$

where $\rho = \sqrt{(\Delta x)^2 + (\Delta y)^2}$.

6.3.2-6. Implicit functions and their differentiation.

Consider the equation $F(x, y) = 0$ with a solution (x_0, y_0). Suppose that the derivative $F_y(x, y)$ is continuous in a neighborhood of the point (x_0, y_0) and $F_y(x, y) \neq 0$ in that neighborhood. Then the equation $F(x, y) = 0$ defines a continuous function $y = y(x)$ (called an *implicit function*) of the variable x in a neighborhood of the point x_0. Moreover, if in a neighborhood of (x_0, y_0) there exists a continuous derivative F_x, then the implicit function $y = y(x)$ has a continuous derivative expressed by $\dfrac{dy}{dx} = -\dfrac{F_x}{F_y}$.

Consider the equation $F(x, y, z) = 0$ that establishes a relation between the variables x, y, z. If $F(x_0, y_0, z_0) = 0$ and in a neighborhood of the point (x_0, y_0, z_0) there exist continuous partial derivatives F_x, F_y, F_z such that $F_z(x_0, y_0, z_0) \neq 0$, then equation $F(x, y, z) = 0$, in a neighborhood of (x_0, y_0), has a unique solution $z = \varphi(x, y)$ such that $\varphi(x_0, y_0) = z_0$; moreover, the function $z = \varphi(x, y)$ is continuous and has continuous partial derivatives expressed by

$$\frac{\partial z}{\partial x} = -\frac{F_x}{F_z}, \qquad \frac{\partial z}{\partial y} = -\frac{F_y}{F_z}.$$

Example 2. For the equation $x \sin y + z + e^z = 0$ we have $F_z = 1 + e^z \neq 0$. Therefore, this equation defines a function $z = \varphi(x, y)$ on the entire plane, and its derivatives have the form $\dfrac{\partial z}{\partial x} = -\dfrac{\sin y}{1 + e^z}$, $\dfrac{\partial z}{\partial y} = -\dfrac{x \cos y}{1 + e^z}$.

6.3.2-7. Jacobian. Dependent and independent functions. Invertible transformations.

1°. Two functions $f(x,y)$ and $g(x,y)$ are called *dependent* if there is a function $\Phi(z)$ such that $g(x,y)=\Phi(f(x,y))$; otherwise, the functions $f(x,y)$ and $g(x,y)$ are called *independent*.

The *Jacobian* is the determinant of the matrix whose elements are the first partial derivatives of the functions $f(x,y)$ and $g(x,y)$:

$$\frac{\partial(f,g)}{\partial(x,y)}\equiv\begin{vmatrix}\frac{\partial f}{\partial x} & \frac{\partial f}{\partial y}\\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y}\end{vmatrix}. \tag{6.3.2.1}$$

1) If the Jacobian (6.3.2.1) in a domain D is identically equal to zero, then the functions $f(x,y)$ and $g(x,y)$ are dependent in D.

2) If the Jacobian (6.3.2.1) is separated from zero in D, then the functions $f(x,y)$ and $g(x,y)$ are independent in D.

2°. Functions $f_k(x_1,x_2,\dots,x_n)$, $k=1,2,\dots,n$, are called dependent in a domain D if there is a function $\Phi(z_1,z_2,\dots,z_n)$ such that

$$\Phi\big(f_1(x_1,x_2,\dots,x_n),f_2(x_1,x_2,\dots,x_n),\dots,f_n(x_1,x_2,\dots,x_n)\big)=0 \qquad (\text{in } D);$$

otherwise, these functions are called independent.

The *Jacobian* is the determinant of the matrix whose elements are the first partial derivatives:

$$\frac{\partial(f_1,f_2,\dots,f_n)}{\partial(x_1,x_2,\dots,x_n)}\equiv\det\left(\frac{\partial f_i}{\partial x_j}\right). \tag{6.3.2.2}$$

The functions $f_k(x_1,x_2,\dots,x_n)$ are dependent in a domain D if the Jacobian (6.3.2.2) is identically equal to zero in D. The functions $f_k(x_1,x_2,\dots,x_n)$ are independent in D if the Jacobian (6.3.2.2) does not vanish in D.

3°. Consider the transformation

$$y_k=f_k(x_1,x_2,\dots,x_n),\qquad k=1,2,\dots,n. \tag{6.3.2.3}$$

Suppose that the functions f_k are continuously differentiable and the Jacobian (6.3.2.2) differs from zero at the point $(x_1^\circ,x_2^\circ,\dots,x_n^\circ)$. Then, in a sufficiently small neighborhood of this point, equations (6.3.2.3) specify a one-to-one correspondence between the points of that neighborhood and the set of points $(y_1,y_2,\dots,y_n)$ consisting of the values of the functions (6.3.2.3) in the corresponding neighborhood of the point $(y_1^\circ,y_2^\circ,\dots,y_n^\circ)$. This means that the system (6.3.2.3) is locally solvable in a neighborhood of the point $(x_1^\circ,x_2^\circ,\dots,x_n^\circ)$, i.e., the following representation holds:

$$x_k=g_k(y_1,y_2,\dots,y_n),\qquad k=1,2,\dots,n,$$

where g_k are continuously differentiable functions in the corresponding neighborhood of the point $(y_1^\circ,y_2^\circ,\dots,y_n^\circ)$.

6.3.3. Directional Derivative. Gradient. Geometrical Applications

6.3.3-1. Directional derivative.

One says that a *scalar field* is defined in a domain D if any point $M(x,y)$ of that domain is associated with a certain value $z=f(M)=f(x,y)$. Thus, a thermal field and a pressure

field are examples of scalar fields. A *level line* of a scalar field is a level line of the function that specifies the field (see Subsection 6.3.1). Thus, isothermal and isobaric curves are, respectively, level lines of thermal and pressure fields.

In order to examine the behavior of a field $z = f(x, y)$ at a point $M_0(x_0, y_0)$ in the direction of a vector $\vec{a} = \{a_1, a_2\}$, one should construct a straight line passing through M_0 in the direction of the vector $\vec{a}$ (this line can be specified in terms of the parametric equations $x = x_0 + a_1 t$, $y = y_0 + a_2 t$) and study the function $z(t) = f(x_0 + a_1 t,\ y_0 + a_2 t)$. The derivative of the function $z(t)$ at the point M_0 (i.e., for $t = 0$) characterizes the change rate of the field at that point in the direction $\vec{a}$. Dividing $z'(0)$ by $|\vec{a}| = \sqrt{a_1^2 + a_2^2}$, we obtain the so-called *derivative in the direction* $\vec{a}$ of the given field at the given point:

$$\frac{\partial f}{\partial \vec{a}} = \frac{1}{|\vec{a}|}\left[\, a_1 f'_x(x_0, y_0) + a_2 f'_y(x_0, y_0) \,\right].$$

The *gradient* of the scalar field $z = f(x, y)$ is, by definition, the vector-valued function

$$\operatorname{grad} f = f'_x(x, y)\vec{i} + f'_y(x, y)\vec{j},$$

where $\vec{i}$ and $\vec{j}$ are unit vectors along the coordinate axes x and y. At each point, the gradient of a scalar field is orthogonal to the level line passing through that point. The gradient indicates the direction of maximal growth of the field. In terms of the gradient, the directional derivative can be expressed as follows:

$$\frac{\partial f}{\partial \vec{a}} = \frac{\vec{a}}{|\vec{a}|} \operatorname{grad} f.$$

The gradient is also denoted by $\nabla f = \operatorname{grad} f$.

Remark. The above facts for a plane scalar field obviously can be extended to the case of a spatial scalar field.

6.3.3-2. Geometrical applications of the theory of functions of several variables.

The *equation of the tangent plane* to the surface $z = f(x, y)$ at a point (x_0, y_0, z_0), where $z_0 = f(x_0, y_0)$, has the form

$$z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

The vector of the normal to the surface at that point is

$$\vec{n} = \left\{-f_x(x_0, y_0),\ -f_y(x_0, y_0),\ 1\right\}.$$

If a surface is defined implicitly by the equation $\Phi(x, y, z) = 0$, then the equation of its tangent plane at the point (x_0, y_0, z_0) has the form

$$\Phi_x(x_0, y_0, z_0)(x - x_0) + \Phi_y(x_0, y_0, z_0)(y - y_0) + \Phi_z(x_0, y_0, z_0)(z - z_0) = 0.$$

The vector of the normal to the surface at that point is

$$\vec{n} = \left\{\Phi_x(x_0, y_0, z_0),\ \Phi_y(x_0, y_0, z_0),\ \Phi_z(x_0, y_0, z_0)\right\}.$$

Consider a surface defined by the parametric equations

$$x = x(u, v), \quad y = y(u, v), \quad z = z(u, v)$$

or, in vector form, $\vec{r} = \vec{r}(u, v)$, where $\vec{r} = \{x, y, z\}$, and let $M_0\left(x(u_0, v_0),\ y(u_0, v_0),\ z(u_0, v_0)\right)$ be the point of the surface corresponding to the parameter values $u = u_0$, $v = v_0$. Then the vector of the normal to the surface at the point M_0 can be expressed by

$$\vec{n}(u, v) = \frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial v} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix},$$

where all partial derivatives are calculated at the point M_0.

6.3.4. Extremal Points of Functions of Several Variables

6.3.4-1. Conditions of extremum of a function of two variables.

1°. *Points of minimum, maximum, or extremum.* A point (x_0, y_0) is called a *point of local minimum* (resp., *maximum*) of a function $z = f(x, y)$ if there is a neighborhood of (x_0, y_0) in which the function is defined and satisfies the inequality $f(x, y) > f(x_0, y_0)$ (resp., $f(x, y) < f(x_0, y_0)$). Points of maximum or minimum are called *points of extremum*.

2°. A *necessary condition of extremum.* If a function has the first partial derivatives at a point of its extremum, these derivatives must vanish at that point. It follows that in order to find points of extremum of such a function $z = f(x, y)$, one should find solutions of the system of equations

$$f_x(x, y) = 0, \quad f_y(x, y) = 0.$$

The points whose coordinates satisfy this system are called *stationary points*. Any point of extremum of a differentiable function is its stationary point, but not every stationary point is a point of its extremum.

3°. *Sufficient conditions of extremum* are used for the identification of points of extremum among stationary points. Some conditions of this type are given below.

Suppose that the function $z = f(x, y)$ has continuous second derivatives at a stationary point. Let us calculate the value of the determinant at that point:

$$\Delta = f_{xx} f_{yy} - f_{xy}^2.$$

The following implications hold:

1) If $\Delta > 0$, $f_{xx} > 0$, then the stationary point is a point of local minimum;
2) If $\Delta > 0$, $f_{xx} < 0$, then the stationary point is a point of local maximum;
3) If $\Delta < 0$, then the stationary point cannot be a point of extremum.

In the degenerate case, $\Delta = 0$, a more delicate analysis of a stationary point is required. In this case, a stationary point may happen to be a point of extremum and maybe not.

Remark. In order to find points of extremum, one should check not only stationary points, but also points at which the first derivatives do not exist or are infinite.

4°. *The smallest and the largest values of a function.* Let $f(x, y)$ be a continuous function in a closed bounded domain D. Any such function takes its smallest and its largest values in D.

If the function has partial derivatives in D, except at some points, then the following method can be helpful for determining the coordinates of the points $(x_{\min}, y_{\min})$ and $(x_{\max}, y_{\max})$ at which the function attains its minimum and maximum, respectively. One should find all internal stationary points and all points at which the derivatives are infinite or do not exist. Then one should calculate the values of the function at these points and compare these with its values at the boundary points of the domain, and then choose the largest and the smallest values.

6.3.4-2. Extremal points of functions of three variables.

For functions of three variables, points of extremum are defined in exactly the same way as for functions of two variables. Let us briefly describe the scheme of finding extremal points of a function $u = \Phi(x, y, z)$. Finding solutions of the system of equations

$$\Phi_x(x, y, z) = 0, \quad \Phi_y(x, y, z) = 0, \quad \Phi_z(x, y, z) = 0,$$

we determine stationary points. For each stationary point, we calculate the values of

$$\Delta_1 = \Phi_{xx}, \quad \Delta_2 = \begin{vmatrix} \Phi_{xx} & \Phi_{xy} \\ \Phi_{xy} & \Phi_{yy} \end{vmatrix}, \quad \Delta_3 = \begin{vmatrix} \Phi_{xx} & \Phi_{xy} & \Phi_{xz} \\ \Phi_{xy} & \Phi_{yy} & \Phi_{yz} \\ \Phi_{xz} & \Phi_{yz} & \Phi_{zz} \end{vmatrix}.$$

The following implications hold:

1) If $\Delta_1 > 0,\ \Delta_2 > 0,\ \Delta_3 > 0,$ then the stationary point is a point of local minimum;
2) If $\Delta_1 < 0,\ \Delta_2 > 0,\ \Delta_3 < 0,$ then the stationary point is a point of local maximum.

6.3.4-3. Conditional extremum of a function of two variables. Lagrange function.

A point (x_0, y_0) is called a *point of conditional or constrained minimum* (resp., *maximum*) of a function

$$z = f(x, y) \tag{6.3.4.1}$$

under the additional condition*

$$\varphi(x, y) = 0 \tag{6.3.4.2}$$

if there is a neighborhood of the point (x_0, y_0) in which $f(x, y) > f(x_0, y_0)$ (resp., $f(x, y) < f(x_0, y_0)$) for all points (x, y) satisfying the condition (6.3.4.2).

For the determination of points of conditional extremum, it is common to use the *Lagrange function*

$$\Phi(x, y, \lambda) = f(x, y) + \lambda\varphi(x, y),$$

where λ is the so-called *Lagrange multiplier*. Solving the system of three equations (the last equation coincides with the condition (6.3.4.2))

$$\frac{\partial\Phi}{\partial x} = 0, \quad \frac{\partial\Phi}{\partial y} = 0, \quad \frac{\partial\Phi}{\partial\lambda} = 0,$$

one finds stationary points of the Lagrange function (and also the value of the coefficient λ). The stationary points may happen to be points of extremum. The above system yields only necessary conditions of extremum, but these conditions may be insufficient; it may happen that there is no extremum at some stationary points. However, with the help of other properties of the function under consideration, it is often possible to establish the character of a critical point.

Example 1. Let us find an extremum of the function

$$z = x^n y, \tag{6.3.4.3}$$

under the condition

$$x + y = a \qquad (a > 0,\ \ n > 0,\ \ x \geq 0,\ \ y \geq 0). \tag{6.3.4.4}$$

Taking $F(x, y) = x^n y$ and $\varphi(x, y) = x + y - a$, we construct the Lagrange function

$$\Phi(x, y, \lambda) = x^n y + \lambda(x + y - a).$$

Solving the system of equations

$$\begin{aligned} \Phi_x &\equiv nx^{n-1}y + \lambda = 0, \\ \Phi_y &\equiv x^n + \lambda = 0, \\ \Phi_\lambda &\equiv x + y - a = 0, \end{aligned}$$

we find the coordinates of a unique stationary point,

$$x_\circ = \frac{an}{n+1}, \quad y_\circ = \frac{a}{n+1}, \quad \lambda_\circ = -\left(\frac{an}{n+1}\right)^n,$$

which corresponds to the maximum of the given function, $z_{\max} = \dfrac{a^{n+1}n^n}{(n+1)^{n+1}}$.

* This condition is also called a *constraint*.

Remark. In order to find points of conditional extremum of functions of two variables, it is often convenient to express the variable y through x (or vice versa) from the additional equation (6.3.4.2) and substitute the resulting expression into the right-hand side of (6.3.4.1). In this way, the original problem is reduced to the problem of extremum for a function of a single variable.

Example 2. Consider again the extremum problem of Example 1 for the function of two variables (6.3.4.3) with the constraint (6.3.4.4). After the elimination of the variable y from (6.3.4.3)–(6.3.4.4), the original problem is reduced to the extremum problem for the function $z = x^n(a - x)$ of one variable.

6.3.4-4. Conditional extremum of functions of several variables.

Consider a function $u = f(x_1, \ldots, x_n)$ of n variables under the condition that $x_1, \ldots, x_n$ satisfy m equations ($m < n$):

$$\begin{cases} \varphi_1(x_1, \ldots, x_n) = 0, \\ \varphi_2(x_1, \ldots, x_n) = 0, \\ \ldots\ldots\ldots\ldots\ldots, \\ \varphi_m(x_1, \ldots, x_n) = 0. \end{cases}$$

In order to find the values of $x_1, \ldots, x_n$ for which f may have a conditional maximum or minimum, one should construct the Lagrange function

$$\Phi(x_1, \ldots, x_n; \lambda_1, \ldots, \lambda_m) = f + \lambda_1\varphi_1 + \lambda_2\varphi_2 + \cdots + \lambda_m\varphi_m$$

and equate to zero its first partial derivatives in the variables $x_1, \ldots, x_n$ and the parameters $\lambda_1, \ldots, \lambda_m$. From the resulting $n + m$ equations, one finds $x_1, \ldots, x_n$ (and also the values of the unknown Lagrange multipliers $\lambda_1, \ldots, \lambda_m$). As in the case of functions of two variables, the question whether the given function has points of conditional extremum can be answered on the basis of additional investigation.

Example 3. Consider the problem of finding the shortest distance from the point (x_0, y_0, z_0) to the plane

$$Ax + By + Cz + D = 0. \tag{6.3.4.5}$$

The squared distance between the points (x_0, y_0, z_0) and (x, y, z) is equal to

$$R^2 = (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2. \tag{6.3.4.6}$$

In our case, the coordinates (x, y, z) should satisfy equation (6.3.4.5) (this point should belong to the plane). Thus, our problem is to find the minimum of the expression (6.3.4.6) under the condition (6.3.4.5). The Lagrange function has the form

$$\Phi = (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 + \lambda(Ax + By + Cz + D).$$

Equating to zero the derivatives of Φ in x, y, z, λ, we obtain the following system of algebraic equations:

$$2(x - x_0) + A\lambda = 0, \quad 2(y - y_0) + B\lambda = 0, \quad 2(z - z_0) + C\lambda = 0, \quad Ax + By + Cz + D = 0.$$

Its solution has the form

$$x = x_0 - \frac{1}{2}A\lambda, \quad y = y_0 - \frac{1}{2}B\lambda, \quad z = z_0 - \frac{1}{2}C\lambda, \quad \lambda = \frac{2(Ax_0 + By_0 + Cz_0 + D)}{A^2 + B^2 + C^2}. \tag{6.3.4.7}$$

Thus we have a unique answer, and since the distance between a given point and the plane can be realized at a single point (x, y, z), the values obtained should correspond to that distance. Substituting the values (6.3.4.7) into (6.3.4.6), we find the squared distance

$$R^2 = \frac{(Ax_0 + By_0 + Cz_0 + D)^2}{A^2 + B^2 + C^2}.$$

6.3.5. Differential Operators of the Field Theory

6.3.5-1. Hamilton's operator and first-order differential operators.

The Hamilton's operator or the *nabla vector* is the symbolic vector

$$\nabla = \vec{i}\,\frac{\partial}{\partial x} + \vec{j}\,\frac{\partial}{\partial y} + \vec{k}\,\frac{\partial}{\partial z}.$$

This vector can be used for expressing the following differential operators:

1) gradient of a scalar function $u(x, y, z)$:

$$\operatorname{grad} u = \vec{i}\,\frac{\partial u}{\partial x} + \vec{j}\,\frac{\partial u}{\partial y} + \vec{k}\,\frac{\partial u}{\partial z} = \nabla u;$$

2) divergence of a vector field $\vec{a} = P\,\vec{i} + Q\,\vec{j} + R\,\vec{k}$:

$$\operatorname{div} \vec{a} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} = \nabla \cdot \vec{a}$$

(scalar product of the nabla vector and the vector $\vec{a}$);

3) rotation of a vector field $\vec{a} = P\,\vec{i} + Q\,\vec{j} + R\,\vec{k}$:

$$\operatorname{curl} \vec{a} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix} = \nabla \times \vec{a}$$

(vector product of the nabla vector and the vector $\vec{a}$).

Each scalar field $u(x, y, z)$ generates a vector field $\operatorname{grad} u$. A vector field $\vec{a}(x, y, z)$ generates two fields: the scalar field $\operatorname{div} \vec{a}$ and the vector field $\operatorname{curl} \vec{a}$.

6.3.5-2. Second-order differential operators.

The following differential identities hold:

1) $\operatorname{curl}\operatorname{grad} u = \vec{0}$ or $(\nabla \times \nabla)\, u = \vec{0}$,

2) $\operatorname{div}\operatorname{curl} \vec{a} = 0$ or $\nabla \cdot (\nabla \times \vec{a}) = 0$.

The following differential relations hold:

1) $\operatorname{div}\operatorname{grad} u = \Delta u = \dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2} + \dfrac{\partial^2 u}{\partial z^2}$,

2) $\operatorname{curl}\operatorname{curl} \vec{a} = \operatorname{grad}\operatorname{div} \vec{a} - \Delta \vec{a}$,

where Δ is the *Laplace operator*, $\Delta u = \nabla \cdot (\nabla u) = \nabla^2 u$.

References for Chapter 6

Adams, R., *Calculus: A Complete Course, 6th Edition*, Pearson Education, Toronto, 2006.

Boyer, C. B., *The History of the Calculus and Its Conceptual Development*, Dover Publications, New York, 1989.

Brannan, D., *A First Course in Mathematical Analysis*, Cambridge University Press, Cambridge, 2006.

Browder, A., *Mathematical Analysis: An Introduction*, Springer-Verlag, New York, 1996.

Courant, R. and John, F., *Introduction to Calculus and Analysis, Vol. 1*, Springer-Verlag, New York, 1999.

Edwards, C. H., and Penney, D., *Calculus, 6th Edition*, Pearson Education, Toronto, 2002.

Kline, M., *Calculus: An Intuitive and Physical Approach, 2nd Edition*, Dover Publications, New York, 1998.

Landau, E., *Differential and Integral Calculus*, American Mathematical Society, Providence, Rhode Island, 2001.

Silverman, R. A., *Essential Calculus with Applications*, Dover Publications, New York, 1989.

Zorich, V. A., *Mathematical Analysis*, Springer-Verlag, Berlin, 2004.

Chapter 7
Integrals

7.1. Indefinite Integral

7.1.1. Antiderivative. Indefinite Integral and Its Properties

7.1.1-1. Antiderivative.

An *antiderivative* (or *primitive function*) of a given function $f(x)$ on an interval (a, b) is a differentiable function $F(x)$ such that its derivative is equal to $f(x)$ for all $x \in (a, b)$:

$$F'(x) = f(x).$$

Example 1. Let $f(x) = 2x$. Then the functions $F(x) = x^2$ and $F_1(x) = x^2 - 1$ are antiderivatives of $f(x)$, since $(x^2)' = 2x$ and $(x^2 - 1)' = 2x$.

THEOREM. *Any function $f(x)$ continuous on an interval (a, b) has infinitely many continuous antiderivatives on (a, b). If $F(x)$ is one of them, then any other antiderivative has the form $F(x) + C$, where C is a constant.*

7.1.1-2. Indefinite integral.

The *indefinite integral* of a function $f(x)$ is the set, $F(x) + C$, of all its antiderivatives. This fact is conventionally written as

$$\int f(x)\,dx = F(x) + C.$$

Here, $f(x)$ is called the *integrand* (or the *integrand function*). The process of finding an integral is called integration. The differential dx indicates that the integration is carried out with respect to x.

Example 2. $\int 6x^2\,dx = 2x^3 + C$, since $(2x^3)' = 6x^2$.

7.1.1-3. Most important corollaries of the definition of the indefinite integral.

Differentiation is the inverse of integration:

$$\frac{d}{dx}\left(\int f(x)\,dx\right) = f(x).$$

Integration is the inverse of differentiation:*

$$\int f'(x)\,dx = f(x) + C.$$

The latter formula serves to make up tables of indefinite integrals. The procedure is often reverse here: an integral is first given in explicit form (i.e., the function $f(x)$ on the right-hand side is prescribed), and then the integrand is obtained by differentiation.

* Integration recovers the function from its derivative, to an additive constant.

7.1.2. Table of Basic Integrals. Properties of the Indefinite Integral. Integration Examples

7.1.2-1. Table of basic integrals.

Listed below are most common indefinite integrals, which are important for the integration of more complicated expressions:

$$\int x^a\,dx = \frac{x^{a+1}}{a+1} + C \quad (a \neq -1),$$

$$\int \frac{dx}{x} = \ln|x| + C,$$

$$\int \frac{dx}{x^2+a^2} = \frac{1}{a}\arctan\frac{x}{a} + C,$$

$$\int \frac{dx}{x^2-a^2} = \frac{1}{2a}\ln\left|\frac{x-a}{x+a}\right| + C,$$

$$\int \frac{dx}{\sqrt{a^2-x^2}} = \arcsin\frac{x}{a} + C,$$

$$\int \frac{dx}{\sqrt{x^2+a}} = \ln\left|x+\sqrt{x^2+a}\,\right| + C,$$

$$\int e^x\,dx = e^x + C,$$

$$\int a^x\,dx = \frac{a^x}{\ln a} + C,$$

$$\int \ln x\,dx = x\ln x - x + C,$$

$$\int \ln ax\,dx = x\ln ax - x + C,$$

$$\int \sin x\,dx = -\cos x + C,$$

$$\int \cos x\,dx = \sin x + C,$$

$$\int \tan x\,dx = -\ln|\cos x| + C,$$

$$\int \cot x\,dx = \ln|\sin x| + C,$$

$$\int \frac{dx}{\sin x} = \ln\left|\tan\frac{x}{2}\right| + C,$$

$$\int \frac{dx}{\cos x} = \ln\left|\tan\left(\frac{x}{2}+\frac{\pi}{4}\right)\right| + C,$$

$$\int \frac{dx}{\sin^2 x} = -\cot x + C,$$

$$\int \frac{dx}{\cos^2 x} = \tan x + C,$$

$$\int \arcsin x\,dx = x\arcsin x + \sqrt{1-x^2} + C,$$

$$\int \arccos x\,dx = x\arccos x - \sqrt{1-x^2} + C,$$

$$\int \arctan x\,dx = x\arctan x - \frac{1}{2}\ln(1+x^2) + C,$$

$$\int \operatorname{arccot} x\,dx = x\operatorname{arccot} x + \frac{1}{2}\ln(1+x^2) + C,$$

$$\int \sinh x\,dx = \cosh x + C,$$

$$\int \cosh x\,dx = \sinh x + C,$$

$$\int \tanh x\,dx = \ln\cosh x + C,$$

$$\int \coth x\,dx = \ln|\sinh x| + C,$$

$$\int \frac{dx}{\sinh x} = \ln\left|\tanh\frac{x}{2}\right| + C,$$

$$\int \frac{dx}{\cosh x} = 2\arctan e^x + C,$$

$$\int \frac{dx}{\sinh^2 x} = -\coth x + C,$$

$$\int \frac{dx}{\cosh^2 x} = \tanh x + C,$$

$$\int \operatorname{arcsinh} x\,dx = x\operatorname{arcsinh} x - \sqrt{1+x^2} + C,$$

$$\int \operatorname{arccosh} x\,dx = x\operatorname{arccosh} x - \sqrt{x^2-1} + C,$$

$$\int \operatorname{arctanh} x\,dx = x\operatorname{arctanh} x + \frac{1}{2}\ln(1-x^2) + C,$$

$$\int \operatorname{arccoth} x\,dx = x\operatorname{arccoth} x + \frac{1}{2}\ln(x^2-1) + C,$$

where C is an arbitrary constant.

► *A more extensive table of indefinite integrals can be found in Section T2.1.*

7.1.2-2. Properties of the indefinite integral.

1. *A constant factor can be taken outside the integral sign*:

$$\int af(x)\,dx = a\int f(x)\,dx \quad (a = \text{const}).$$

2. *Integral of the sum or difference of functions* (*additivity*):

$$\int [f(x) \pm g(x)]\,dx = \int f(x)\,dx \pm \int g(x)\,dx.$$

3. *Integration by parts*:

$$\int f(x)g'(x)\,dx = f(x)g(x) - \int f'(x)g(x)\,dx.$$

4. *Repeated integration by parts* (generalization of the previous formula):

$$\int f(x)g^{(n+1)}(x)\,dx = f(x)g^{(n)}(x) - f'(x)g^{(n-1)}(x) + \cdots + (-1)^n f^{(n)}(x)g(x) + (-1)^{n+1}\int f^{(n+1)}(x)g(x)\,dx, \qquad n = 0,\ 1,\ \ldots$$

5. *Change of variable* (*integration by substitution*):

$$\int f(x)\,dx = \int f(\varphi(t))\,\varphi'_t(t)\,dt, \qquad x = \varphi(t).$$

On computing the integral using the change of variable $x = \varphi(t)$, one should rewrite the resulting expression in terms of the original variable x using the back substitution $t = \varphi^{-1}(x)$.

7.1.2-3. Examples of direct integration of elementary functions.

1°. With simple algebraic manipulation and the properties listed in Paragraph 7.1.2-2, the integration may often be reduced to tabulated integrals.

Example 1. $\displaystyle\int \frac{2x-1}{\sqrt{x}}\,dx = \int\left(2\sqrt{x} - \frac{1}{\sqrt{x}}\right)dx = 2\int x^{1/2}\,dx - \int x^{-1/2}\,dx = \frac{4}{3}x^{3/2} - 2x^{1/2} + C.$

2°. Tabulated integrals can also be used where any function $\varphi(x)$ appears in place of x; for example,

$$\int e^x\,dx = e^x + C \quad\Longrightarrow\quad \int e^{\varphi(x)}\,d\varphi(x) = e^{\varphi(x)} + C;$$

$$\int \frac{dx}{x} = \ln|x| + C \quad\Longrightarrow\quad \int \frac{d\sin x}{\sin x} = \ln|\sin x| + C.$$

The reduction of an integral to a tabulated one may often be achieved by taking some function inside the differential sign.

Example 2. $\displaystyle\int \tan x\,dx = \int \frac{\sin x\,dx}{\cos x} = \int \frac{-d\cos x}{\cos x} = -\int \frac{d\cos x}{\cos x} = -\ln|\cos x| + C.$

3°. Integrals of the form $\displaystyle\int \frac{dx}{ax^2+bx+c}$, $\displaystyle\int \frac{dx}{\sqrt{ax^2+bx+c}}$ can be computed by making a perfect square:

$$ax^2 + bx + c = a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a} + c.$$

Then one should replace dx with the equal differential $d\left(x + \dfrac{b}{2a}\right)$ and use one of the four formulas in the second and third rows in the table of integrals given in Paragraph 7.1.2-1.

Example 3. $\displaystyle\int \frac{dx}{\sqrt{2x-x^2}} = \int \frac{dx}{\sqrt{1-(x-1)^2}} = \int \frac{d(x-1)}{\sqrt{1-(x-1)^2}} = \arcsin(x-1) + C.$

4°. The integration of a polynomial multiplied by an exponential function can be accomplished by using the formula of integration by parts (or repeated integration by parts) given in Paragraph 7.1.2-2.

Example 4. Compute the integral $\displaystyle\int (3x+1)\, e^{2x}\, dx$.

Taking $f(x) = 3x + 1$ and $g'(x) = e^{2x}$, one finds that $f'(x) = 3$ and $g(x) = \frac{1}{2}e^{2x}$. On substituting these expressions into the formula of integration by pars, one obtains

$$\int (3x+1)\, e^{2x}\, dx = \frac{1}{2}(3x+1)\, e^{2x} - \frac{3}{2}\int e^{2x}\, dx = \frac{1}{2}(3x+1)\, e^{2x} - \frac{3}{4}e^{2x} + C = \left(\frac{3}{2}x - \frac{1}{4}\right) e^{2x} + C.$$

Remark 1. More complex examples of the application of integration by parts or repeated integration by parts can be found in Subsection 7.1.6.

Remark 2. Examples of using a change of variables (see Item 5 in Paragraph 7.1.2-2) for the computation of integrals can be found in Subsections 7.1.4 and 7.1.5.

7.1.2-4. Remark on uncomputable integrals.

The differentiation of elementary functions is known to always result in elementary functions. However, this is not the case with integration, which is the reverse of differentiation. The integrals of elementary functions are often impossible to express in terms elementary functions using finitely many arithmetic operations and compositions.

Here are examples of integrals that cannot be expressed via elementary functions:

$$\int \frac{dx}{\sqrt{x^3+1}}, \quad \int \exp(-x^2)\, dx, \quad \int \frac{e^x}{x}\, dx, \quad \int \frac{dx}{\ln x}, \quad \int \frac{\cos x}{x}\, dx, \quad \int \sin(x^2)\, dx.$$

Such integrals are sometimes called intractable. It is significant that all these integrals exist; they generate nonelementary (special) functions.

7.1.3. Integration of Rational Functions

7.1.3-1. Partial fraction decomposition of a rational function.

A *rational function* (also know as a *rational polynomial function*) is a quotient of polynomials:

$$R(x) = \frac{P_n(x)}{Q_m(x)}, \tag{7.1.3.1}$$

where

$$P_n(x) = a_n x^n + \cdots + a_1 x + a_0,$$
$$Q_m(x) = b_m x^m + \cdots + b_1 x + b_0.$$

The fraction (7.1.3.1) is called *proper* if $m > n$ and *improper* if $m \le n$.

Every proper fraction (7.1.3.1) can be decomposed into a sum of partial fractions. To this end, one should factorize the denominator $Q_m(x)$ into irreducible multipliers of the form

$$(x-\alpha_i)^{p_i}, \qquad i = 1, 2, \ldots, k; \tag{7.1.3.2a}$$
$$(x^2+\beta_j x+\gamma_j)^{q_j}, \quad j = 1, 2, \ldots, s, \tag{7.1.3.2b}$$

where the p_i and q_j are positive integers satisfying the condition $p_1+\cdots+p_k+2(q_1+\cdots+q_s)=m$; $\beta_j^2-4\gamma_j<0$. The rational function (7.1.3.1) can be represented as a sum of irreducibles and to each irreducible of the form (7.1.3.2) there correspond as many terms as the power p_i or q_i:

$$\frac{A_{i,1}}{x-\alpha_i}+\frac{A_{i,2}}{(x-\alpha_i)^2}+\cdots+\frac{A_{i,p_i}}{(x-\alpha_i)^{p_i}}; \tag{7.1.3.3a}$$

$$\frac{B_{j,1}x+D_{j,1}}{x^2+\beta_j x+\gamma_j}+\frac{B_{j,2}x+D_{j,2}}{(x^2+\beta_j x+\gamma_j)^2}+\cdots+\frac{B_{j,q_j}x+D_{j,q_j}}{(x^2+\beta_j x+\gamma_j)^{q_j}}. \tag{7.1.3.3b}$$

The constants $A_{i,l}$, $B_{j,r}$, $D_{j,r}$ are found by the *method of undetermined coefficients*. To that end, one should equate the original rational fraction (7.1.3.1) with the sum of the above partial fractions (7.1.3.3) and reduce both sides of the resulting equation to a common denominator. Then, one collects the coefficients of like powers of x and equates them with zero, thus arriving at a system of linear algebraic equations for the $A_{i,l}$, $B_{j,r}$, and $D_{j,r}$.

Example 1. This is an illustration of how a proper fraction can be decomposed into partial fractions:

$$\frac{b_5x^5+b_4x^4+b_3x^3+b_2x^2+b_1x+b_0}{(x+a)(x+c)^3(x^2+k^2)}=\frac{A_{1,1}}{x+a}+\frac{A_{2,1}}{x+c}+\frac{A_{2,2}}{(x+c)^2}+\frac{A_{2,3}}{(x+c)^3}+\frac{Bx+D}{x^2+k^2}.$$

7.1.3-2. Integration of a proper fraction.

1°. To integrate a proper fraction, one should first rewrite the integrand (7.1.3.1) in the form of a sum of partial fractions. Below are the integrals of most common partial fractions (7.1.3.3*a*) and (7.1.3.3*b*) (with $q_j=1$):

$$\begin{aligned}&\int\frac{A}{x-\alpha}\,dx=A\ln|x-\alpha|, \qquad \int\frac{A}{(x-\alpha)^p}\,dx=-\frac{A}{(p-1)(x-\alpha)^{p-1}},\\ &\int\frac{Bx+D}{x^2+\beta x+\gamma}\,dx=\frac{B}{2}\ln(x^2+\beta x+\gamma)+\frac{2D-B\beta}{\sqrt{4\gamma-\beta^2}}\arctan\frac{2x+\beta}{\sqrt{4\gamma-\beta^2}}.\end{aligned} \tag{7.1.3.4}$$

The constant of integration C has been omitted here. More complex integrals of partial fractions (7.1.3.3*b*) with $q_j>1$ can be computed using the formula

$$\int\frac{Bx+D}{(x^2+\beta x+\gamma)^q}\,dx=\frac{P(x)}{(x^2+\beta x+\gamma)^{q-1}}+\lambda\int\frac{dx}{x^2+\beta x+\gamma}, \tag{7.1.3.5}$$

where $P(x)$ is a polynomial of degree $2q-3$. The coefficients of $P(x)$ and the constant λ can be found by the method of undetermined coefficients by differentiating formula (7.1.3.5).

Remark. The following recurrence relation may be used in order to compute the integrals on the left-hand side in (7.1.3.5):

$$\int\frac{Bx+D}{(x^2+\beta x+\gamma)^q}\,dx=\frac{(2D-B\beta)x+D\beta-2B\gamma}{(q-1)(4\gamma-\beta^2)(x^2+\beta x+\gamma)^{q-1}}+\frac{(2q-3)(2D-B\beta)}{(q-1)(4\gamma-\beta^2)}\int\frac{dx}{(x^2+\beta x+\gamma)^{q-1}}.$$

Example 2. Compute the integral $\displaystyle\int\frac{3x^2-x-2}{x^3+8}\,dx$.

Let us factor the denominator of the integrand, $x^3+8=(x+2)(x^2-2x+4)$, and perform the partial fraction decomposition:

$$\frac{3x^2-x-2}{(x+2)(x^2-2x+4)}=\frac{A}{x+2}+\frac{Bx+D}{x^2-2x+4}.$$

Multiplying both sides by the common denominator and collecting the coefficients of like powers of x, we obtain

$$(A + B - 3)x^2 + (-2A + 2B + D + 1)x + 4A + 2D + 2 = 0.$$

Now equating the coefficients of the different powers of x with zero, we arrive at a system of algebraic equations for A, B, and D:

$$A + B - 3 = 0, \quad -2A + 2B + D + 1 = 0, \quad 4A + 2D + 2 = 0.$$

Its solution is: $A = 1$, $B = 2$, $D = -3$. Hence, we have

$$\int \frac{3x^2 - x - 2}{x^3 + 8}\,dx = \int \frac{1}{x+2}\,dx + \int \frac{2x-3}{x^2 - 2x + 4}\,dx$$
$$= \ln|x+2| + \ln\left(x^2 - 2x + 4\right) - \frac{1}{\sqrt{3}} \arctan\frac{x-1}{\sqrt{3}} + C.$$

Here, the last integral of (7.1.3.4) has been used.

2°. The integrals of proper rational functions defined as the ratio of a polynomial to a power function $(x-\alpha)^m$ are given by the formulas

$$\int \frac{P_n(x)}{(x-\alpha)^m}\,dx = -\sum_{k=0}^{n} \frac{P_n^{(k)}(\alpha)}{k!\,(m-k-1)(x-\alpha)^{m-k-1}} + C, \qquad m > n+1;$$

$$\int \frac{P_n(x)}{(x-\alpha)^{n+1}}\,dx = -\sum_{k=0}^{n-1} \frac{P_n^{(k)}(\alpha)}{k!\,(n-k)(x-\alpha)^{n-k}} + \frac{P_n^{(n)}(\alpha)}{n!} \ln|x-\alpha| + C,$$

where $P_n(x)$ is a polynomial of degree n and $P_n^{(k)}(\alpha)$ is its kth derivative at $x = \alpha$.

3°. Suppose the roots in the factorization of the denominator of the fraction (7.1.3.1) are all real and distinct:

$$Q_m(x) = b_m x^m + \cdots + b_1 x + b_0 = b_m(x-\alpha_1)(x-\alpha_2)\ldots(x-\alpha_m), \qquad \alpha_i \neq \alpha_j.$$

Then the following formula holds:

$$\int \frac{P_n(x)}{Q_m(x)}\,dx = \sum_{k=1}^{m} \frac{P_n(\alpha_k)}{Q'_m(\alpha_k)} \ln|x-\alpha_k| + C,$$

where $m > n$ and the prime denotes a derivative.

7.1.3-3. Integration of improper fractions.

1°. In order to integrate an improper fraction, one should first isolate a proper fraction by division with remainder. As a result, the improper fraction is represented as the sum of a polynomial and a proper fraction,

$$\frac{a_n x^n + \cdots + a_1 x + a_0}{b_m x^m + \cdots + b_1 x + b_0} = c_m x^{n-m} + \cdots + c_1 x + c_0 + \frac{s_{m-1}x^{m-1} + \cdots + s_1 x + s_0}{b_m x^m + \cdots + b_1 x + b_0} \qquad (n \geq m),$$

which are then integrated separately.

Example 3. Evaluate the integral $I = \int \dfrac{x^2}{x-1}\,dx$.

Let us rewrite the integrand (improper fraction) as the sum of a polynomial and a proper fraction: $\dfrac{x^2}{x-1} = x + 1 + \dfrac{1}{x-1}$. Hence, $I = \int \left(x + 1 + \dfrac{1}{x-1}\right) dx = \frac{1}{2}x^2 + x + \ln|x-1| + C$.

2°. The integrals of improper rational functions defined as the ratio of a polynomial to a simple power function $(x-\alpha)^m$ are evaluated by the formula

$$\int \frac{P_n(x)}{(x-\alpha)^m}\,dx = \sum_{k=m}^{n} \frac{P_n^{(k)}(\alpha)}{k!\,(k-m+1)}(x-\alpha)^{k-m+1} + \frac{P_n^{(m-1)}(\alpha)}{(m-1)!}\ln|x-\alpha| - \sum_{k=0}^{m-2} \frac{P_n^{(k)}(\alpha)}{k!\,(m-k-1)(x-\alpha)^{m-k-1}} + C,$$

where $n \geq m$.

Remark 1. The indefinite integrals of rational functions are always expressed in terms of elementary functions.

Remark 2. Some of the integrals reducible to integrals of rational functions are considered in Subsections 7.1.5 and 7.1.6.

7.1.4. Integration of Irrational Functions

The integration of some irrational functions can be reduced to that of rational functions using a suitable change of variables. In what follows, the functions $R(x, y)$ and $R(x_1, \ldots, x_k)$ are assumed to be rational functions in each of the arguments.

7.1.4-1. Integration of expressions involving radicals of linear-fractional functions.

1°. The integrals with roots of linear functions

$$\int R\big(x,\ \sqrt[n]{ax+b}\big)\,dx$$

are reduced to integrals of rational functions by the change of variable $z = \sqrt[n]{ax+b}$.

Example 1. Evaluate the integral $I = \int x\sqrt{1-x}\,dx$.

With the change of variable $\sqrt{1-x} = z$, we have $x = 1-z^2$ and $dx = -2z\,dz$. Substituting these expressions into the integral yields

$$I = -2\int (1-z^2)z^2\,dz = -\frac{2}{3}z^3 + \frac{2}{5}z^5 + C = -\frac{2}{3}\sqrt{(1-x)^3} + \frac{2}{5}\sqrt{(1-x)^5} + C.$$

2°. The integrals with roots of linear-fractional functions

$$\int R\left(x,\ \sqrt[n]{\frac{ax+b}{cx+d}}\right) dx$$

are reduced to integrals of rational functions by the substitution $z = \sqrt[n]{\dfrac{ax+b}{cx+d}}$.

3°. The integrals of the more general form

$$\int R\left(x,\ \left(\frac{ax+b}{cx+d}\right)^{q_1},\ \ldots,\ \left(\frac{ax+b}{cx+d}\right)^{q_k}\right)dx,$$

where $q_1,\ldots,q_k$ are rational numbers, are reduced to integrals of rational functions using the change of variable $z^m=\dfrac{ax+b}{cx+d}$, where m is the common denominator of the fractions $q_1,\ldots,q_k$.

4°. Integrals containing the product of a polynomial by a simple power function of the form $(x-a)^\beta$ are evaluated by the formula

$$\int P_n(x)(x-a)^\beta\,dx=\sum_{k=0}^{n}\frac{P_n^{(k)}(a)}{k!\,(k+\beta+1)}(x-a)^{k+\beta+1},$$

where $P_n(x)$ is a polynomial of degree n, $P_n^{(k)}(a)$ is its kth derivative at $x=a$, and β is any positive or negative proper fraction (to be more precise, $\beta\neq -1,-2,\ldots,-n-1$).

7.1.4-2. Euler substitutions. Trigonometric substitutions.

We will be considering integrals involving the radical of a quadratic trinomial:

$$\int R\bigl(x,\sqrt{ax^2+bx+c}\,\bigr)\,dx,$$

where $b^2\neq 4ac$. Such integrals are expressible in terms of elementary functions.

1°. *Euler substitutions.* The given integral is reduced to the integral of a rational fraction by one of the following three Euler substitutions:

1) $\sqrt{ax^2+bx+c}=t\mp x\sqrt{a}$ if $a>0$;

2) $\sqrt{ax^2+bx+c}=xt\pm\sqrt{c}$ if $c>0$;

3) $\sqrt{ax^2+bx+c}=t(x-x_1)$ if $4ac-b^2<0$,

where x_1 is a root of the quadratic equation $ax^2+bx+c=0$. In all three cases, the variable x and the radical $\sqrt{ax^2+bx+c}$ are expressible in terms of the new variable t as (the formulas correspond to the upper signs in the substitutions):

1) $x=\dfrac{t^2-c}{2\sqrt{a}\,t+b}$, $\sqrt{ax^2+bx+c}=\dfrac{\sqrt{a}\,t^2+bt+c\sqrt{a}}{2\sqrt{a}\,t+b}$, $dx=2\dfrac{\sqrt{a}\,t^2+bt+c\sqrt{a}}{(2\sqrt{a}\,t+b)^2}\,dt$;

2) $x=\dfrac{2\sqrt{c}\,t-b}{a-t^2}$, $\sqrt{ax^2+bx+c}=\dfrac{\sqrt{c}\,t^2-bt+c\sqrt{a}}{a-t^2}$, $dx=2\dfrac{\sqrt{c}\,t^2-bt+c\sqrt{a}}{(a-t^2)^2}\,dt$;

3) $x=\dfrac{(t^2+a)x_1+b}{t^2-a}$, $\sqrt{ax^2+bx+c}=\dfrac{(2ax_1+b)t}{t^2-a}$, $dx=-2\dfrac{(2ax_1+b)t}{(t^2-a)^2}\,dt$.

2°. *Trigonometric substitutions.* The function $\sqrt{ax^2+bx+c}$ can be reduced, by making a perfect square in the radicand, to one of the three forms:

1) $\sqrt{a}\sqrt{(x-p)^2+q^2}$ if $a>0$;

2) $\sqrt{a}\sqrt{(x-p)^2-q^2}$ if $a>0$;

3) $\sqrt{-a}\sqrt{q^2-(x-p)^2}$ if $a<0$,

where $p = -\frac{1}{2}b/a$. Different trigonometric substitutions are further used in each case to evaluate the integral:

$$1)\quad x-p=q\tan t,\quad \sqrt{(x-p)^2+q^2}=\frac{q}{\cos t},\quad dx=\frac{q\,dt}{\cos^2 t};$$

$$2)\quad x-p=\frac{q}{\cos t},\quad \sqrt{(x-p)^2-q^2}=q\tan t,\quad dx=\frac{q\sin t\,dt}{\cos^2 t};$$

$$3)\quad x-p=q\sin t,\quad \sqrt{q^2-(x-p)^2}=q\cos t,\quad dx=q\cos t\,dt.$$

Example 2. Evaluate the integral $\int \sqrt{6+4x-2x^2}\,dx$.

This integral corresponds to case 3 with $a = -2$, $p = 1$, and $q = 2$. The integrand can be rewritten in the form:

$$\sqrt{6+4x-2x^2}=\sqrt{2}\sqrt{3+2x-x^2}=\sqrt{2}\sqrt{4-(x-1)^2}.$$

Using the trigonometric substitution $x-1=2\sin t$ and the formulas $\sqrt{3+2x-x^2}=2\cos t$ and $dx=2\cos t\,dt$, we obtain

$$\begin{aligned}\int \sqrt{6+4x-2x^2}\,dx &= 4\sqrt{2}\int \cos^2 t\,dt = 2\sqrt{2}\int (1+\cos 2t)\,dt\\ &= 2\sqrt{2}t+\sqrt{2}\sin 2t+C = 2\sqrt{2}\arcsin\frac{x-1}{2}+\sqrt{2}\sin\left(2\arcsin\frac{x-1}{2}\right)+C\\ &= 2\sqrt{2}\arcsin\frac{x-1}{2}+\frac{\sqrt{2}}{2}(x-1)\sqrt{4-(x-1)^2}+C.\end{aligned}$$

7.1.4-3. Integral of a differential binomial.

The *integral of a differential binomial*,

$$\int x^m(a+bx^n)^p\,dx,$$

where a and b are constants, and n, m, p are rational numbers, is expressible in terms of elementary functions in the following three cases only:

1) If p is an integer. For $p \geq 0$, removing the brackets gives the sum of power functions. For $p < 0$, the substitution $x = t^r$, where r is the common denominator of the fractions m and n, leads to the integral of a rational function.

2) If $\dfrac{m+1}{n}$ is an integer. One uses the substitution $a + bx^n = t^k$, where k is the denominator of the fraction p.

3) If $\dfrac{m+1}{n}+p$ is an integer. One uses the substitution $ax^{-n} + b = t^k$, where k is the denominator of the fraction p.

Remark. In cases 2 and 3, the substitution $z = x^n$ leads to integrals of the form 3° from Paragraph 7.1.4-1.

7.1.5. Integration of Exponential and Trigonometric Functions

7.1.5-1. Integration of exponential and hyperbolic functions.

1. Integrals of the form $\int R(e^{px}, e^{qx})\,dx$, where $R(x, y)$ is a rational function of its arguments and p, q are rational numbers, may be evaluated using the substitution $z^m = e^x$, where m is the common denominator of the fractions p and q. In the special case of integer p and q, we have $m = 1$, and the substitution becomes $z = e^x$.

Example 1. Evaluate the integral $\int \dfrac{e^{3x}\,dx}{e^x+2}$.

This integral corresponds to integer p and q: $p=1$ and $q=3$. So we use the substitution $z=e^x$. Then $x=\ln z$ and $dx=\dfrac{dz}{z}$. Therefore,

$$\int \frac{e^{3x}\,dx}{e^x+2}=\int \frac{z^2\,dz}{z+2}=\int\Big(z-2+\frac{4}{z+2}\Big)\,dz=\frac{1}{2}z^2-2z+4\ln|z+2|+C=\frac{1}{2}e^{2x}-2e^x+4\ln(e^x+2)+C.$$

2. Integrals of the form $\int R(\sinh ax,\cosh ax)\,dx$ are evaluated by converting the hyperbolic functions to exponentials, using the formulas $\sinh ax=\frac{1}{2}(e^{ax}-e^{-ax})$ and $\cosh ax=\frac{1}{2}(e^{ax}+e^{-ax})$, and performing the substitution $z=e^{ax}$. Then

$$\int R(\sinh ax,\cosh ax)\,dx=\frac{1}{a}\int R\Big(\frac{z^2-1}{2z},\frac{z^2+1}{2z}\Big)\frac{dz}{z}.$$

Alternatively, the substitution $t=\tanh\Big(\dfrac{ax}{2}\Big)$ can also be used to evaluate integrals of the above form. Then

$$\int R(\sinh ax,\cosh ax)\,dx=\frac{2}{a}\int R\Big(\frac{2t}{1-t^2},\frac{1+t^2}{1-t^2}\Big)\frac{dt}{1-t^2}.$$

7.1.5-2. Integration of trigonometric functions.

1. Integrals of the form $\int R(\sin ax,\cos ax)\,dx$ can be converted to integrals of rational functions using the versatile trigonometric substitution $t=\tan\Big(\dfrac{ax}{2}\Big)$:

$$\int R(\sin ax,\cos ax)\,dx=\frac{2}{a}\int R\Big(\frac{2t}{1+t^2},\frac{1-t^2}{1+t^2}\Big)\frac{dt}{1+t^2}.$$

Example 2. Evaluate the integral $\int \dfrac{dx}{2+\sin x}$.

Using the versatile trigonometric substitution $t=\tan\dfrac{x}{2}$, we have

$$\int\frac{dx}{2+\sin x}=2\int\frac{dt}{\Big(2+\dfrac{2t}{1+t^2}\Big)(1+t^2)}=\int\frac{dt}{t^2+t+1}=2\int\frac{d(2t+1)}{(2t+1)^2+3}$$
$$=\frac{2}{\sqrt{3}}\arctan\frac{2t+1}{\sqrt{3}}+C=\frac{2}{\sqrt{3}}\arctan\Big(\frac{2}{\sqrt{3}}\tan\frac{x}{2}+\frac{1}{\sqrt{3}}\Big)+C.$$

2. Integrals of the form $\int R(\sin^2 ax,\cos^2 ax,\tan ax)\,dx$ are converted to integrals of rational functions with the change of variable $z=\tan ax$:

$$\int R(\sin^2 ax,\cos^2 ax,\tan ax)\,dx=\frac{1}{a}\int R\Big(\frac{z^2}{1+z^2},\frac{1}{1+z^2},z\Big)\frac{dz}{1+z^2}.$$

3. Integrals of the form

$$\int \sin ax\cos bx\,dx,\quad \int\cos ax\cos bx\,dx,\quad \int\sin ax\sin bx\,dx$$

are evaluated using the formulas

$$\sin\alpha\cos\beta = \tfrac{1}{2}[\sin(\alpha+\beta)+\sin(\alpha-\beta)],$$
$$\cos\alpha\cos\beta = \tfrac{1}{2}[\cos(\alpha+\beta)+\cos(\alpha-\beta)],$$
$$\sin\alpha\sin\beta = \tfrac{1}{2}[\cos(\alpha-\beta)-\cos(\alpha+\beta)].$$

4. Integrals of the form $\int \sin^m x\cos^n x\,dx$, where m and n are integers, are evaluated as follows:

(a) if m is odd, one uses the change of variable $\cos x = z$, with $\sin x\,dx = -dz$;
(b) if n is odd, one uses the change of variable $\sin x = z$, with $\cos x\,dx = dz$;
(c) if m and n are both even nonnegative integers, one should use the degree reduction formulas

$$\sin^2 x = \tfrac{1}{2}(1-\cos 2x), \quad \cos^2 x = \tfrac{1}{2}(1+\cos 2x), \quad \sin x\cos x = \tfrac{1}{2}\sin 2x.$$

Example 3. Evaluate the integral $\int \sin^5 x\,dx$.

This integral corresponds to odd m: $m = 5$. With simple rearrangement and the change of variable $\cos x = z$, we have

$$\int \sin^5 x\,dx = \int (\sin^2 x)^2 \sin x\,dx = -\int (1-\cos^2 x)^2\,d\cos x = -\int (1-z^2)^2\,dz$$
$$= \tfrac{2}{3}z^3 - \tfrac{1}{5}z^5 - z + C = \tfrac{2}{3}\cos^3 x - \tfrac{1}{5}\cos^5 x - \cos x + C.$$

Remark. In general, the integrals $\int \sin^p x\cos^q x\,dx$ are reduced to the integral of a differential binomial by the substitution $y = \sin x$.

7.1.6. Integration of Polynomials Multiplied by Elementary Functions

► *Throughout this section, $P_n(x)$ designates a polynomial of degree n.*

7.1.6-1. Integration of the product of a polynomial by exponential functions.

General formulas:

$$\int P_n(x)e^{ax}\,dx = e^{ax}\left[\frac{P_n(x)}{a} - \frac{P_n'(x)}{a^2} + \cdots + (-1)^n\frac{P_n^{(n)}(x)}{a^{n+1}}\right] + C,$$

$$\int P_n(x)\cosh(ax)\,dx = \sinh(ax)\left[\frac{P_n(x)}{a} + \frac{P_n''(x)}{a^3} + \cdots\right] - \cosh(ax)\left[\frac{P_n'(x)}{a^2} + \frac{P_n'''(x)}{a^4} + \cdots\right] + C,$$

$$\int P_n(x)\sinh(ax)\,dx = \cosh(ax)\left[\frac{P_n(x)}{a} + \frac{P_n''(x)}{a^3} + \cdots\right] - \sinh(ax)\left[\frac{P_n'(x)}{a^2} + \frac{P_n'''(x)}{a^4} + \cdots\right] + C.$$

These formulas are obtained by repeated integration by parts; see formula 4 from Paragraph 7.1.2-2 with $f(x) = P_n(x)$ for $g^{(n+1)}(x) = e^{ax}$, $g^{(n+1)}(x) = \cosh(ax)$, and $g^{(n+1)}(x) = \sinh(ax)$, respectively.

In the special case $P_n(x) = x^n$, the first formula gives

$$\int x^n e^{ax}\,dx = e^{ax}\sum_{k=0}^{n}\frac{(-1)^{n-k}}{a^{n+1-k}}\frac{n!}{k!}x^k + C.$$

7.1.6-2. Integration of the product of a polynomial by a trigonometric function.

1°. General formulas:

$$\int P_n(x)\cos(ax)\,dx = \sin(ax)\left[\frac{P_n(x)}{a} - \frac{P_n''(x)}{a^3} + \cdots\right] + \cos(ax)\left[\frac{P_n'(x)}{a^2} - \frac{P_n'''(x)}{a^4} + \cdots\right] + C,$$

$$\int P_n(x)\sin(ax)\,dx = \sin(ax)\left[\frac{P_n'(x)}{a^2} - \frac{P_n'''(x)}{a^4} + \cdots\right] - \cos(ax)\left[\frac{P_n(x)}{a} - \frac{P_n''(x)}{a^3} + \cdots\right] + C.$$

These formulas are obtained by repeated integration by parts; see formula 4 from Paragraph 7.1.2-2 with $f(x) = P_n(x)$ for $g^{(n+1)}(x) = \cos(ax)$ and $g^{(n+1)}(x) = \sin(ax)$, respectively.

2°. To evaluate integrals of the form

$$\int P_n(x)\cos^m(ax)\,dx, \qquad \int P_n(x)\sin^m(ax)\,dx,$$

with $m = 2, 3, \ldots$, one should first use the trigonometric formulas

$$\cos^{2k}(ax) = \frac{1}{2^{2k-1}}\sum_{i=0}^{k-1} C_{2k}^i \cos[2(k-i)ax] + \frac{1}{2^{2k}}C_{2k}^k \qquad (m = 2k),$$

$$\cos^{2k+1}(ax) = \frac{1}{2^{2k}}\sum_{i=0}^{k} C_{2k+1}^i \cos[(2k-2i+1)ax] \qquad (m = 2k+1),$$

$$\sin^{2k}(ax) = \frac{1}{2^{2k-1}}\sum_{i=0}^{k-1}(-1)^{k-i} C_{2k}^i \cos[2(k-i)ax] + \frac{1}{2^{2k}}C_{2k}^k \qquad (m = 2k),$$

$$\sin^{2k+1}(ax) = \frac{1}{2^{2k}}\sum_{i=0}^{k}(-1)^{k-i} C_{2k+1}^i \sin[(2k-2i+1)ax] \qquad (m = 2k+1),$$

thus reducing the above integrals to those considered in Item 1°.

3°. Integrals of the form

$$\int P_n(x)\,e^{ax}\sin(bx)\,dx, \qquad \int P_n(x)\,e^{ax}\cos(bx)\,dx$$

can be evaluated by repeated integration by parts.

In particular,

$$\int x^n e^{ax}\sin(bx) = e^{ax}\sum_{k=1}^{n+1}\frac{(-1)^{k+1}n!}{(n-k+1)!\,(a^2+b^2)^{k/2}}x^{n-k+1}\sin(bx+k\theta) + C,$$

$$\int x^n e^{ax}\cos(bx) = e^{ax}\sum_{k=1}^{n+1}\frac{(-1)^{k+1}n!}{(n-k+1)!\,(a^2+b^2)^{k/2}}x^{n-k+1}\cos(bx+k\theta) + C,$$

where

$$\sin\theta = -\frac{b}{\sqrt{a^2+b^2}}, \qquad \cos\theta = \frac{a}{\sqrt{a^2+b^2}}.$$

7.1.6-3. Integrals involving power and logarithmic functions.

1°. The formula of integration by parts with $g'(x) = P_n(x)$ is effective in the evaluation of integrals of the form

$$\int P_n(x)\ln(ax)\,dx = Q_{n+1}(x)\ln(ax) - a\int \frac{Q_{n+1}(x)}{x}\,dx,$$

where $Q_{n+1}(x) = \int P_n(x)\,dx$ is a polynomial of degree $n+1$. The integral on the right-hand side is easy to take, since the integrand is the sum of power functions.

Example. Evaluate the integral $\int \ln x\,dx$.

Setting $f(x) = \ln x$ and $g'(x) = 1$, we find $f'(x) = \dfrac{1}{x}$ and $g(x) = x$. Substituting these expressions into the formula of integration by parts, we obtain $\int \ln x\,dx = x\ln x - \int dx = x\ln x - x + C$.

2°. The easiest way to evaluate integrals of the more general form

$$I = \int \sum_{i=0}^{n} \ln^i(ax)\left(\sum_{j=0}^{m} b_{ij}x^{\beta_{ij}}\right) dx,$$

where the β_{ij} are arbitrary numbers, is to use the substitution $z = \ln(ax)$, so that

$$I = \int \sum_{i=0}^{n} z^i\left(\sum_{j=0}^{m} \frac{b_{ij}}{a^{\beta_{ij}+1}}e^{(\beta_{ij}+1)z}\right) dz.$$

By removing the brackets, one obtains a sum of integrals like $\int x^n e^{ax}\,dx$, which are easy to evaluate by the last formula in Paragraph 7.1.6-1.

7.1.6-4. Integrals involving inverse trigonometric functions.

1°. The formula of integration by parts with $g'(x) = P_n(x)$ also allows the evaluation of the following integrals involving inverse trigonometric functions:

$$\int P_n(x)\arcsin(ax)\,dx = Q_{n+1}(x)\arcsin(ax) - a\int \frac{Q_{n+1}(x)}{\sqrt{1-a^2x^2}}\,dx,$$

$$\int P_n(x)\arccos(ax)\,dx = Q_{n+1}(x)\arccos(ax) + a\int \frac{Q_{n+1}(x)}{\sqrt{1-a^2x^2}}\,dx,$$

$$\int P_n(x)\arctan(ax)\,dx = Q_{n+1}(x)\arctan(ax) - a\int \frac{Q_{n+1}(x)}{a^2x^2+1}\,dx,$$

$$\int P_n(x)\operatorname{arccot}(ax)\,dx = Q_{n+1}(x)\operatorname{arccot}(ax) + a\int \frac{Q_{n+1}(x)}{a^2x^2+1}\,dx,$$

where $Q_{n+1}(x) = \int P_n(x)\,dx$ is a polynomial of degree $n+1$. The integrals with radicals on the right-hand side in the first two formulas can be evaluated using the techniques described in Paragraph 7.1.4-2. The integrals of rational functions on the right-hand side in the last two formulas can be evaluated using the techniques described in Subsection 7.1.3.

Remark. The above formulas can be generalized to contain any rational functions $R'(x)$ and $R(x)$ instead of the polynomials $P_n(x)$ and $Q_{n+1}(x)$, respectively.

2°. The following integrals are taken using a change of variable:

$$\int P_n(x)\arcsin^m(ax)\,dx = \frac{1}{a}\int \cos z\, P_n\left(\frac{\sin z}{a}\right) z^m\,dz, \qquad \text{substitution } z = \arcsin(ax);$$

$$\int P_n(x)\arccos^m(ax)\,dx = -\frac{1}{a}\int \sin z\, P_n\left(\frac{\cos z}{a}\right) z^m\,dz, \quad \text{substitution } z = \arccos(ax),$$

where $m = 2, 3, \ldots$ The expressions $\cos z \sin^k z$ and $\sin z \cos^k z$ $(k = 1, \ldots, n)$ in the integrals on the right-hand sides should be expressed as sums of sines and cosines with appropriate arguments. Then it remains to evaluate integrals considered in Paragraph 7.1.6-2.

7.2. Definite Integral

7.2.1. Basic Definitions. Classes of Integrable Functions. Geometrical Meaning of the Definite Integral

7.2.1-1. Basic definitions.

Let $y = f(x)$ be a bounded function defined on a finite closed interval $[a, b]$. Let us partition this interval into n elementary subintervals defined by a set of points $\{x_0, x_1, \ldots, x_n\}$ such that $a = x_0 < x_1 < \cdots < x_n = b$. Each subinterval $[x_{k-1}, x_k]$ will be characterized by its length $\Delta x_k = x_k - x_{k-1}$ and an arbitrarily chosen point $\xi_k \in [x_{k-1}, x_k]$. Let us make up an *integral sum* (a *Cauchy–Riemann sum*, also known as a *Riemann sum*)

$$s_n = \sum_{k=1}^{n} f(\xi_k)\Delta x_k \qquad (x_{k-1} \le \xi_k \le x_k).$$

If, as $n \to \infty$ and, accordingly, $\Delta x_k \to 0$ for all k, there exists a finite limit of the integral sums s_n and it depends on neither the way the interval $[a, b]$ was split up, nor the selection of the points ξ_k, then this limit is denoted $\int_a^b f(x)\,dx$ and is called the *definite integral* (also the *Riemann integral*) of the function $y = f(x)$ over the interval $[a, b]$:

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} s_n \qquad \Big(\max_{1\le k\le n} \Delta x_k \to 0 \ \text{ as } \ n \to \infty\Big).$$

In this case, the function $f(x)$ is called *integrable* on the interval $[a, b]$.

7.2.1-2. Classes of integrable functions.

1. If a function $f(x)$ is continuous on an interval $[a, b]$, then it is integrable on this interval.

2. If a bounded function $f(x)$ has finitely many jump discontinuities on $[a, b]$, then it is integrable on $[a, b]$.

3. A monotonic bounded function $f(x)$ is always integrable.

7.2.1-3. Geometric meaning of the definite integral.

If $f(x) \geq 0$ on $[a, b]$, then the integral $\int_a^b f(x)\,dx$ is equal to the area of the domain $D = \{a \leq x \leq b,\ 0 \leq y \leq f(x)\}$ (the area of the curvilinear trapezoid shown in Fig. 7.1).

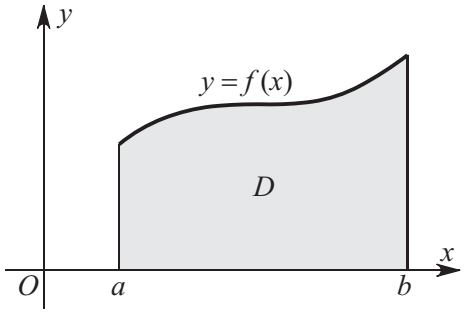


Figure 7.1. The integral of a nonnegative function $f(x)$ on an interval $[a, b]$ is equal to the area of the shaded region.

7.2.2. Properties of Definite Integrals and Useful Formulas

7.2.2-1. Qualitative properties of integrals.

1. If a function $f(x)$ is integrable on $[a, b]$, then the functions $cf(x)$, with $c = \text{const}$, and $|f(x)|$ are also integrable on $[a, b]$.

2. If two functions $f(x)$ and $g(x)$ are integrable on $[a, b]$, then their sum, difference, and product are also integrable on $[a, b]$.

3. If a function $f(x)$ is integrable on $[a, b]$ and its values lie within an interval $[c, d]$, where a function $g(y)$ is defined and continuous, then the composite function $g(f(x))$ is also integrable on $[a, b]$.

4. If a function $f(x)$ is integrable on $[a, b]$, then it is also integrable and on any subinterval $[\alpha, \beta] \subset [a, b]$. Conversely, if an interval $[a, b]$ is partitioned into a number of subintervals and $f(x)$ is integrable on each of the subintervals, then it is integrable on the whole interval $[a, b]$.

5. If the values of a function are changed at finitely many points, this will not affect the integrability of the function and will not change the value of the integral.

7.2.2-2. Properties of integrals in terms of identities.

1. *The integral over a zero-length interval is zero*:

$$\int_a^a f(x)\,dx = 0.$$

2. *Antisymmetry under the swap of the integration limits*:

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx.$$

This property can be taken as the definition of a definite integral with $a > b$.

3. *Linearity.* If functions $f(x)$ and $g(x)$ are integrable on an interval $[a, b]$, then

$$\int_a^b \big[Af(x) \pm Bg(x)\big]\,dx = A\int_a^b f(x)\,dx \pm B\int_a^b g(x)\,dx$$

for any numbers A and B.

4. *Additivity.* If $c \in [a, b]$ and $f(x)$ is integrable on $[a, b]$, then

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

5. *Differentiation with respect the variable upper limit.* If $f(x)$ is continuous on $[a, b]$, then the function $\Phi(x) = \int_a^x f(t)\,dt$ is differentiable on $[a, b]$, and $\Phi'(x) = f(x)$. This can be written in one relation:

$$\frac{d}{dx}\left(\int_a^x f(t)\,dt\right) = f(x).$$

6. *Newton–Leibniz formula*:

$$\int_a^b f(x)\,dx = F(x)\Big|_a^b = F(b) - F(a),$$

where $F(x)$ is an antiderivative of $f(x)$ on $[a, b]$.

7. *Integration by parts.* If functions $f(x)$ and $g(x)$ have continuous derivatives on $[a, b]$, then

$$\int_a^b f(x)g'(x)\,dx = \big[f(x)g(x)\big]\Big|_a^b - \int_a^b f'(x)g(x)\,dx.$$

8. *Repeated integration by parts*:

$$\int_a^b f(x)g^{(n+1)}(x)\,dx = \Big[f(x)g^{(n)}(x) - f'(x)g^{(n-1)}(x) + \cdots + (-1)^n f^{(n)}(x)g(x)\Big]_a^b + (-1)^{n+1}\int_a^b f^{(n+1)}(x)g(x)\,dx, \qquad n = 0,\ 1,\ \ldots$$

9. *Change of variable (substitution) in a definite integral.* Let $f(x)$ be a continuous function on $[a, b]$ and let $x(t)$ be a continuously differentiable function on $[\alpha, \beta]$. Suppose also that the range of values of $x(t)$ coincides with $[a, b]$, with $x(\alpha) = a$ and $x(\beta) = b$. Then

$$\int_a^b f(x)\,dx = \int_\alpha^\beta f\big(x(t)\big)\,x'(t)\,dt.$$

Example. Evaluate the integral $\displaystyle\int_0^3 \frac{dx}{(x-8)\sqrt{x+1}}$.

Perform the substitution $x + 1 = t^2$, with $dx = 2t\,dt$. We have $t = 1$ at $x = 0$ and $t = 2$ at $x = 3$. Therefore

$$\int_0^3 \frac{dx}{(x-8)\sqrt{x+1}} = \int_1^2 \frac{2t\,dt}{(t^2-9)t} = 2\int_1^2 \frac{dt}{t^2-9} = \frac{1}{3}\ln\left|\frac{t-3}{t+3}\right|\Bigg|_1^2 = \frac{1}{3}\ln\frac{2}{5}.$$

10. *Differentiation with respect to a parameter.* Let $f(x, \lambda)$ be a continuous function in a domain $a \le x \le b$, $\lambda_1 \le \lambda \le \lambda_2$ and let it has a continuous partial derivative $\frac{\partial}{\partial\lambda} f(x, \lambda)$ in the

same domain. Also let $u(\lambda)$ and $v(\lambda)$ be differentiable functions on the interval $\lambda_1 \le \lambda \le \lambda_2$ such that $a \le u(\lambda) \le b$ and $a \le v(\lambda) \le b$. Then

$$\frac{d}{d\lambda}\int_{v(\lambda)}^{u(\lambda)} f(x,\lambda)\,dx = f\big(u(\lambda),\lambda\big)\frac{du(\lambda)}{d\lambda} - f\big(v(\lambda),\lambda\big)\frac{dv(\lambda)}{d\lambda} + \int_{v(\lambda)}^{u(\lambda)} \frac{\partial}{\partial\lambda} f(x,\lambda)\,dx.$$

11. *Cauchy's formula for multiple integration*:

$$\int_a^x dx_1 \int_a^{x_1} dx_2 \ldots \int_a^{x_{n-1}} f(x_n)\,dx_n = \frac{1}{(n-1)!}\int_a^x (x-t)^{n-1} f(t)\,dt.$$

7.2.3. General Reduction Formulas for the Evaluation of Integrals

Below are some general formulas, involving arbitrary functions and parameters, that could facilitate the evaluation of integrals.

7.2.3-1. Integrals involving functions of a linear or rational argument.

$$\int_a^b f(a+b-x)\,dx = \int_a^b f(x)\,dx;$$

$$\int_0^a [f(x)+f(a-x)]\,dx = 2\int_0^a f(x)\,dx;$$

$$\int_0^a [f(x)-f(a-x)]\,dx = 0;$$

$$\int_{-a}^a f(x)\,dx = 0 \quad \text{if } f(x) \text{ is odd;}$$

$$\int_{-a}^a f(x)\,dx = 2\int_0^a f(x)\,dx \quad \text{if } f(x) \text{ is even;}$$

$$\int_a^b f(x, a+b-x)\,dx = 0 \quad \text{if } f(x,y) = -f(y,x);$$

$$\int_0^1 f\big(2x\sqrt{1-x^2}\big)\,dx = \int_0^1 f\big(1-x^2\big)\,dx.$$

7.2.3-2. Integrals involving functions with trigonometric argument.

$$\int_0^{\pi} f(\sin x)\,dx = 2\int_0^{\pi/2} f(\sin x)\,dx;$$

$$\int_0^{\pi/2} f(\sin x)\,dx = \int_0^{\pi/2} f(\cos x)\,dx;$$

$$\int_0^{\pi/2} f(\sin x, \cos x)\,dx = 0 \quad \text{if } f(x,y) = -f(y,x);$$

$$\int_0^{\pi/2} f(\sin 2x)\cos x\,dx = \int_0^{\pi/2} f(\cos^2 x)\cos x\,dx;$$

$$\int_0^{n\pi} xf(\sin x)\,dx = \pi n^2 \int_0^{\pi/2} f(\sin x)\,dx \quad \text{if } f(x) = f(-x);$$

$$\int_0^{n\pi} xf(\sin x)\,dx = (-1)^{n-1}\pi n \int_0^{\pi/2} f(\sin x)\,dx \quad \text{if } f(-x) = -f(x);$$

$$\int_0^{2\pi} f(a\sin x + b\cos x)\,dx = \int_0^{2\pi} f\bigl(\sqrt{a^2+b^2}\,\sin x\bigr)\,dx = 2\int_0^{\pi} f\bigl(\sqrt{a^2+b^2}\,\cos x\bigr)\,dx;$$

$$\int_0^{\pi} f\left(\frac{\sin^2 x}{1+2a\cos x+a^2}\right)dx = \int_0^{\pi} f(\sin^2 x)\,dx \quad \text{if } |a| \ge 1;$$

$$\int_0^{\pi} f\left(\frac{\sin^2 x}{1+2a\cos x+a^2}\right)dx = \int_0^{\pi} f\left(\frac{\sin^2 x}{a^2}\right)dx \quad \text{if } 0 < |a| < 1.$$

7.2.3-3. Integrals involving logarithmic functions.

$$\int_a^b f(x)\ln^n x\,dx = \left[\Bigl(\frac{d}{d\lambda}\Bigr)^n \int_a^b x^\lambda f(x)\,dx\right]_{\lambda=0},$$

$$\int_a^b f(x)\ln^n g(x)\,dx = \left[\Bigl(\frac{d}{d\lambda}\Bigr)^n \int_a^b f(x)[g(x)]^\lambda\,dx\right]_{\lambda=0},$$

$$\int_a^b f(x)[g(x)]^\lambda \ln^n g(x)\,dx = \Bigl(\frac{d}{d\lambda}\Bigr)^n \int_a^b f(x)[g(x)]^\lambda\,dx.$$

7.2.4. General Asymptotic Formulas for the Calculation of Integrals

Below are some general formulas, involving arbitrary functions and parameters, that may be helpful for obtaining asymptotics of integrals.

7.2.4-1. Asymptotic formulas for integrals with weak singularity as $\varepsilon \to 0$.

1°. We will consider integrals of the form

$$I(\varepsilon) = \int_0^a \frac{x^{\beta-1}f(x)\,dx}{(x+\varepsilon)^\alpha},$$

where $0 < a < \infty$, $\beta > 0$, $f(0) \ne 0$, and $\varepsilon > 0$ is a small parameter.

The integral diverges as $\varepsilon \to 0$ for $\alpha \ge \beta$, that is, $\lim\limits_{\varepsilon\to 0} I(\varepsilon) = \infty$. In this case, the leading term of the asymptotic expansion of the integral $I(\varepsilon)$ is given by

$$I(\varepsilon) = \frac{\Gamma(\beta)\Gamma(\alpha-\beta)}{\Gamma(\alpha)} f(0)\varepsilon^{\beta-\alpha} + O(\varepsilon^\sigma) \quad \text{if} \quad \alpha > \beta,$$

$$I(\varepsilon) = -f(0)\ln\varepsilon + O(1) \quad \text{if} \quad \alpha = \beta,$$

where $\Gamma(\beta)$ is the gamma function and $\sigma = \min[\beta - \alpha + 1, 0]$.

2°. The leading term of the asymptotic expansion, as $\varepsilon \to 0$, of the more general integral

$$I(\varepsilon) = \int_0^a \frac{x^{\beta-1} f(x)\,dx}{(x^k + \varepsilon^k)^\alpha}$$

with $0 < a < \infty$, $\beta > 0$, $k > 0$, $\varepsilon > 0$, and $f(0) \neq 0$ is expressed as

$$I(\varepsilon) = \frac{f(0)}{k\Gamma(\alpha)}\Gamma\left(\frac{\beta}{k}\right)\Gamma\left(\alpha - \frac{\beta}{k}\right)\varepsilon^{\beta-\alpha k} + O(\varepsilon^\sigma) \quad \text{if} \quad \alpha k > \beta,$$
$$I(\varepsilon) = -f(0)\ln\varepsilon + O(1) \quad \text{if} \quad \alpha k = \beta,$$

where $\sigma = \min[\beta - \alpha k + 1, 0]$.

3°. The leading terms of the asymptotic expansion, as $\varepsilon \to 0$, of the integral

$$I(\varepsilon) = \int_a^\infty x^\alpha \exp\bigl(-\varepsilon x^\beta\bigr) f(x)\,dx$$

with $a > 0$, $\beta > 0$, $\varepsilon > 0$, and $f(0) \neq 0$ has the form

$$I(\varepsilon) = \frac{1}{\beta} f(0)\Gamma\left(\frac{\alpha+1}{\beta}\right)\varepsilon^{-\frac{\alpha+1}{\beta}} \quad \text{if} \quad \alpha > -1,$$
$$I(\varepsilon) = -\frac{1}{\beta} f(0)\ln\varepsilon \quad \text{if} \quad \alpha = -1.$$

4°. Now consider *potential-type integrals*

$$\Pi(f) = \int_{-1}^1 \frac{f(\xi)\,d\xi}{\sqrt{(\xi - z)^2 + r^2}},$$

with z, r, φ being cylindrical coordinates in the three-dimensional space. The function $\Pi(f)$ is simple layer potential concentrated on the interval $z \in [-1, 1]$ with density $f(z)$. If the density is continuous, then $\Pi(f)$ satisfies the Laplace equation $\Delta\Pi = 0$ outside $z \in [-1, 1]$ and vanishes at infinity.

Asymptotics of the integral as $r \to 0$:

$$\Pi(f) = -2f(z)\ln r + O(1),$$

where $|z| \le 1 - \delta$ with $0 < \delta < 1$.

7.2.4-2. Asymptotic formulas for Laplace integrals of special form as $\lambda \to \infty$.

1°. Consider a *Laplace integral of the special form*

$$I(\lambda) = \int_0^a x^{\beta-1} \exp\bigl(-\lambda x^\alpha\bigr) f(x)\,dx,$$

where $0 < a < \infty$, $\alpha > 0$, and $\beta > 0$.

The following formula, called *Watson's asymptotic formula*, holds as $\lambda \to \infty$:

$$I(\lambda) = \frac{1}{\alpha}\sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!}\Gamma\Big(\frac{k+\beta}{\alpha}\Big)\lambda^{-(k+\beta)/\alpha} + O\big(\lambda^{-(n+\beta+1)/\alpha}\big).$$

Remark 1. Watson's formula also holds for improper integrals with $a = \infty$ if the original integral converges absolutely for some $\lambda_0 > 0$.

Remark 2. Watson's formula remains valid in the case of complex parameter λ as $|\lambda| \to \infty$, where $|\arg \lambda| \le \frac{\pi}{2} - \varepsilon < \frac{\pi}{2}$ ($\varepsilon > 0$ can be chosen arbitrarily small but independent of λ).

Remark 3. The Laplace transform corresponds to the above integral with $a = \infty$ and $\alpha = \beta = 1$.

2°. The leading term of the asymptotic expansion, as $\lambda \to \infty$, of the integral

$$I(\lambda) = \int_0^a x^{\beta-1}\,|\ln x|^{\gamma}\, e^{-\lambda x} f(x)\,dx$$

with $0 < a < \infty$, $\beta > 0$, and $f(0) \ne 0$ is expressed as

$$I(\lambda) = \Gamma(\beta) f(0) \lambda^{-\beta} (\ln \lambda)^{\gamma}.$$

7.2.4-3. Asymptotic formulas for Laplace integrals of general form as $\lambda \to \infty$.

Consider a *Laplace integral of the general form*

$$I(\lambda) = \int_a^b f(x)\exp[\lambda g(x)]\,dx, \tag{7.2.4.1}$$

where $[a, b]$ is a finite interval and $f(x)$, $g(x)$ are continuous functions.

1°. *Leading term of the asymptotic expansion of the integral (7.2.4.1) as $\lambda \to \infty$.* Suppose the function $g(x)$ attains a maximum on $[a, b]$ at only one point $x_0 \in [a, b]$ and is differentiable in a neighborhood of it, with $g'(x_0) = 0$, $g''(x_0) \ne 0$, and $f(x_0) \ne 0$. Then the leading term of the asymptotic expansion of the integral (7.2.4.1), as $\lambda \to \infty$, is expressed as

$$\begin{aligned} I(\lambda) &= f(x_0)\sqrt{-\frac{2\pi}{\lambda g''(x_0)}}\,\exp[\lambda g(x_0)] \qquad \text{if } \ a < x_0 < b, \\ I(\lambda) &= \frac{1}{2} f(x_0)\sqrt{-\frac{2\pi}{\lambda g''(x_0)}}\,\exp[\lambda g(x_0)] \quad \text{if } \ x_0 = a \ \text{ or } \ x_0 = b. \end{aligned} \tag{7.2.4.2}$$

Note that the latter formula differs from the former by the factor $1/2$ only.

Under the same conditions, if $g(x)$ attains a maximum at either endpoint, $x_0 = a$ or $x_0 = b$, but $g'(x_0) \ne 0$, then the leading asymptotic term of the integral, as $\lambda \to \infty$, is

$$I(\lambda) = \frac{f(x_0)}{|g'(x_0)|}\frac{1}{\lambda}\exp[\lambda g(x_0)], \qquad \text{where } \ x_0 = a \ \text{ or } \ x_0 = b. \tag{7.2.4.3}$$

For more accurate asymptotic estimates for the Laplace integral (7.2.4.1), see below.

2°. *Leading and subsequent asymptotic terms of the integral (7.2.4.1) as $\lambda \to \infty$.* Let $g(x)$ attain a maximum at only one internal point of the interval, $x_0 \in (a < x_0 < b)$, with $g'(x_0) = 0$ and $g''(x_0) \neq 0$, and let the functions $f(x)$ and $g(x)$ be, respectively, n and $n+1$ times differentiable in a neighborhood of $x = x_0$. Then the asymptotic formula

$$I(\lambda) = \exp[\lambda g(x_0)]\left(\sum_{k=0}^{n-1} c_k \lambda^{-k-1} + O(\lambda^{-n})\right) \tag{7.2.4.4}$$

holds as $\lambda \to \infty$, with

$$c_k = \frac{1}{(2k)!\,2^{k+1/2}}\Gamma\left(k+\frac{1}{2}\right)\lim_{x\to x_0}\left(\frac{d}{dx}\right)^k\left[f(x)\left(\frac{g(x_0)-g(x)}{(x-x_0)^2}\right)^{-k-1/2}\right].$$

Suppose $g(x)$ attains a maximum at the endpoint $x = a$ only, with $g'(a) \neq 0$. Suppose also that $f(x)$ and $g(x)$ are, respectively, n and $n+1$ times differentiable in a neighborhood of $x = a$. Then we have, as $\lambda \to \infty$,

$$I(\lambda) = \exp[\lambda g(a)]\left(\sum_{k=0}^{n-1} c_k \lambda^{-k-1} + O(\lambda^{-n})\right), \tag{7.2.4.5}$$

where

$$c_0 = -\frac{f(a)}{g'(a)}; \quad c_k = (-1)^{k+1}\left[\left(\frac{1}{g'(x)}\frac{d}{dx}\right)^k \frac{f(x)}{g'(x)}\right]_{x=a}, \quad k = 1, 2, \ldots$$

Remark 1. The asymptotic formulas (7.2.4.2)–(7.2.4.5) hold also for improper integrals with $b = \infty$ if the original integral (7.2.4.1) converges absolutely at some $\lambda_0 > 0$.

Remark 2. The asymptotic formulas (7.2.4.2)–(7.2.4.5) remain valid also in the case of complex λ as $|\lambda| \to \infty$, where $|\arg\lambda| \le \frac{\pi}{2} - \varepsilon < \frac{\pi}{2}$ ($\varepsilon > 0$ can be chosen arbitrarily small but independent of λ).

3°. *Some generalizations.* Let $g(x)$ attain a maximum at only one internal point of the interval, $x_0 \in (a < x_0 < b)$, with $g'(x_0) = \cdots = g^{(2m-1)}(x_0) = 0$ and $g^{(2m)}(x_0) \neq 0$, $m \ge 1$ and $f(x_0) \neq 0$. Then the leading asymptotic term of the integral (7.2.4.1), as $\lambda \to \infty$, is expressed as

$$I(\lambda) = \frac{1}{m}\Gamma\left(\frac{1}{2m}\right) f(x_0)\left[-\frac{(2m)!}{g^{(2m)}(x_0)}\right]^{\frac{1}{2m}} \lambda^{-\frac{1}{2m}} \exp[\lambda g(x_0)].$$

Let $g(x)$ attain a maximum at the endpoint $x = a$ only, with $g'(a) = \cdots = g^{(m-1)}(a) = 0$ and $g^{(m)}(a) \neq 0$, where $m \ge 1$ and $f(a) \neq 0$. Then the leading asymptotic term of the integral (7.2.4.1), as $\lambda \to \infty$, has the form

$$I(\lambda) = \frac{1}{m}\Gamma\left(\frac{1}{m}\right) f(a)\left[-\frac{m!}{g^{(m)}(a)}\right]^{\frac{1}{m}} \lambda^{-\frac{1}{m}} \exp[\lambda g(a)].$$

7.2.4-4. Asymptotic formulas for a power Laplace integral.

Consider the *power Laplace integral*, which is obtained from the exponential Laplace integral (7.2.4.1) by substituting $\ln g(x)$ for $g(x)$:

$$I(\lambda) = \int_a^b f(x)[g(x)]^\lambda \, dx, \tag{7.2.4.6}$$

where $[a, b]$ is a finite closed interval and $g(x) > 0$. It is assumed that the functions $f(x)$ and $g(x)$ appearing in the integral (7.2.4.6) are continuous; $g(x)$ is assumed to attain a maximum at only one point $x_0 = [a, b]$ and to be differentiable in a neighborhood of $x = x_0$, with $g'(x_0) = 0$, $g''(x_0) \neq 0$, and $f(x_0) \neq 0$. Then the leading asymptotic term of the integral, as $\lambda \to \infty$, is expressed as

$$I(\lambda) = f(x_0)\sqrt{-\frac{2\pi}{\lambda g''(x_0)}}\,[g(x_0)]^{\lambda+1/2} \qquad \text{if } \ a < x_0 < b,$$

$$I(\lambda) = \frac{1}{2} f(x_0)\sqrt{-\frac{2\pi}{\lambda g''(x_0)}}\,[g(x_0)]^{\lambda+1/2} \quad \text{if } \ x_0 = a \ \text{ or } \ x_0 = b.$$

Note that the latter formula differs from the former by the factor $1/2$ only.

Under the same conditions, if $g(x)$ attains a maximum at either endpoint, $x_0 = a$ or $x_0 = b$, but $g'(x_0) \neq 0$, then the leading asymptotic term of the integral, as $\lambda \to \infty$, is

$$I(\lambda) = \frac{f(x_0)}{|g'(x_0)|}\frac{1}{\lambda}[g(x_0)]^{\lambda+1/2}, \qquad \text{where } \ x_0 = a \ \text{ or } \ x_0 = b.$$

7.2.4-5. Asymptotic behavior of integrals with variable integration limit as $x \to \infty$.

Let $f(t)$ be a continuously differentiable function and let $g(t)$ be a twice continuously differentiable function. Also let the following conditions hold:

$$f(t) > 0, \quad g'(t) > 0; \quad g(t) \to \infty \ \text{ as } \ t \to \infty;$$
$$f'(t)/f(t) = o\big(g'(t)\big) \ \text{ as } \ t \to \infty; \quad g''(t) = o\big(g'^2(t)\big) \ \text{ as } \ t \to \infty.$$

Then the following asymptotic formula holds, as $x \to \infty$:

$$\int_0^x f(t)\exp[g(t)]\,dt \simeq \frac{f(x)}{g'(x)}\exp[g(x)].$$

7.2.4-6. Limiting properties of integrals involving periodic functions with parameter.

1°. *Riemann property of integrals involving periodic functions.* Let $f(x)$ be a continuous function on a finite interval $[a, b]$. Then the following limiting relations hold:

$$\lim_{\lambda\to\infty}\int_a^b f(x)\sin(\lambda x)\,dx = 0,$$
$$\lim_{\lambda\to\infty}\int_a^b f(x)\cos(\lambda x)\,dx = 0.$$

Remark. The condition of continuity of $f(x)$ can be replaced by the more general condition of absolute integrability of $f(x)$ on a finite interval $[a, b]$.

2°. *Dirichlet's formula.* Let $f(x)$ be a monotonically increasing and bounded function on a finite interval $[0, a]$, with $a > 0$. Then the following limiting formula holds:

$$\lim_{\lambda\to\infty} \int_0^a f(x)\frac{\sin(\lambda x)}{x}\,dx = \frac{\pi}{2} f(+0).$$

7.2.4-7. Limiting properties of other integrals with parameter.

Let $f(x)$ and $g(x)$ be continuous and positive functions on $[a, b]$. Then the following limiting relations hold:

$$\lim_{n\to\infty} \sqrt[n]{I_n} = \max_{x\in[a,b]} f(x),$$

$$\lim_{n\to\infty} \frac{I_{n+1}}{I_n} = \max_{x\in[a,b]} f(x),$$

where $I_n = \int_a^b g(x)[f(x)]^n\,dx$.

7.2.5. Mean Value Theorems. Properties of Integrals in Terms of Inequalities. Arithmetic Mean and Geometric Mean of Functions

7.2.5-1. Mean value theorems.

THEOREM 1. *If $f(x)$ is a continuous function on $[a, b]$, there exists at least one point $c \in (a, b)$ such that*

$$\int_a^b f(x)\,dx = f(c)(b-a).$$

The number $f(c)$ is called the mean value of the function $f(x)$ on $[a, b]$.

THEOREM 2. *If $f(x)$ is a continuous function on $[a, b]$, and $g(x)$ is integrable and of constant sign ($g(x) \ge 0$ or $g(x) \le 0$) on $[a, b]$, then there exists at least one point $c \in (a, b)$ such that*

$$\int_a^b f(x)g(x)\,dx = f(c)\int_a^b g(x)\,dx.$$

THEOREM 3. *If $f(x)$ is a monotonic and nonnegative function on an interval (a, b), with $a \ge b$, and $g(x)$ is integrable, then there exists at least one point $c \in (a, b)$ such that*

$$\int_a^b f(x)g(x)\,dx = f(a)\int_a^c g(x)\,dx \qquad \text{if } f(x) \text{ is nonincreasing;}$$

$$\int_a^b f(x)g(x)\,dx = f(b)\int_c^b g(x)\,dx \qquad \text{if } f(x) \text{ is nondecreasing.}$$

THEOREM 4. *If $f(x)$ and $g(x)$ are bounded and integrable functions on an interval $[a, b]$, with $a < b$, and $g(x)$ satisfies inequalities $A \le g(x) \le B$, then there exists a point $c \in [a, b]$*

such that

$$\int_a^b f(x)g(x)\,dx = A\int_a^c f(x)\,dx + B\int_c^b f(x)\,dx \qquad \text{if } g(x) \text{ is nondecreasing;}$$

$$\int_a^b f(x)g(x)\,dx = B\int_a^c f(x)\,dx + A\int_c^b f(x)\,dx \qquad \text{if } g(x) \text{ is nonincreasing;}$$

$$\int_a^b f(x)g(x)\,dx = g(a)\int_a^c f(x)\,dx + g(b)\int_c^b f(x)\,dx \qquad \text{if } g(x) \text{ is strictly monotonic.}$$

7.2.5-2. Properties of integrals in terms of inequalities.

1. *Estimation theorem.* If $m \le f(x) \le M$ on $[a, b]$, then

$$m(b-a) \le \int_a^b f(x)\,dx \le M(b-a).$$

2. *Inequality integration theorem.* If $\varphi(x) \le f(x) \le g(x)$ on $[a, b]$, then

$$\int_a^b \varphi(x)\,dx \le \int_a^b f(x)\,dx \le \int_a^b g(x)\,dx.$$

In particular, if $f(x) \ge 0$ on $[a, b]$, then $\int_a^b f(x)\,dx \ge 0$.

► *Further on, it is assumed that the integrals on the right-hand sides of the inequalities of Items 3–8 exist.*

3. *Absolute value theorem* (*integral analogue of the triangle inequality*):

$$\left|\int_a^b f(x)\,dx\right| \le \int_a^b |f(x)|\,dx.$$

4. *Bunyakovsky's inequality* (*Cauchy–Bunyakovsky inequality*):

$$\left(\int_a^b f(x)g(x)\,dx\right)^2 \le \int_a^b f^2(x)\,dx \int_a^b g^2(x)\,dx.$$

5. *Cauchy's inequality*:

$$\left(\int_a^b [f(x)+g(x)]^2\,dx\right)^{1/2} \le \left(\int_a^b f^2(x)\,dx\right)^{1/2} + \left(\int_a^b g^2(x)\,dx\right)^{1/2}.$$

6. *Minkowski's inequality* (generalization of Cauchy's inequality):

$$\left(\int_a^b |f(x)+g(x)|^p\,dx\right)^{\frac{1}{p}} \le \left(\int_a^b |f(x)|^p\,dx\right)^{\frac{1}{p}} + \left(\int_a^b |g(x)|^p\,dx\right)^{\frac{1}{p}}, \qquad p \ge 1.$$

7. *Hölder's inequality* (at $p = 2$, it translates into Bunyakovsky's inequality):

$$\left|\int_a^b f(x)g(x)\,dx\right| \le \left(\int_a^b |f(x)|^p\,dx\right)^{\frac{1}{p}} \left(\int_a^b |g(x)|^{\frac{p}{p-1}}\,dx\right)^{\frac{p-1}{p}}, \qquad p > 1.$$

8. *Chebyshev's inequality*:

$$\left(\int_a^b f(x)h(x)\,dx\right)\left(\int_a^b g(x)h(x)\,dx\right) \le \left(\int_a^b h(x)\,dx\right)\left(\int_a^b f(x)g(x)h(x)\,dx\right),$$

where $f(x)$ and $g(x)$ are monotonically increasing functions and $h(x)$ is a positive integrable function on $[a, b]$.

9. *Jensen's inequality*:

$$f\left(\frac{\int_a^b g(t)x(t)\,dt}{\int_a^b g(t)\,dt}\right) \le \frac{\int_a^b g(t)f(x(t))\,dt}{\int_a^b g(t)\,dt} \qquad \text{if } f(x) \text{ is convex } (f'' > 0);$$

$$f\left(\frac{\int_a^b g(t)x(t)\,dt}{\int_a^b g(t)\,dt}\right) \ge \frac{\int_a^b g(t)f(x(t))\,dt}{\int_a^b g(t)\,dt} \qquad \text{if } f(x) \text{ is concave } (f'' < 0),$$

where $x(t)$ is a continuous function ($a \le x \le b$) and $g(t) \ge 0$. The equality is attained if and only if either $x(t) = \text{const}$ or $f(x)$ is a linear function. Jensen's inequality serves as a general source for deriving various integral inequalities.

10. *Steklov's inequality*. Let $f(x)$ be a continuous function on $[0, \pi]$ and let it have everywhere on $[0, \pi]$, except maybe at finitely many points, a square integrable derivative $f'(x)$. If either of the conditions

$$\text{(a)} \quad f(0) = f(\pi) = 0,$$
$$\text{(b)} \quad \int_0^\pi f(x)\,dx = 0$$

is satisfied, then the following inequality holds:

$$\int_0^\pi [f'(x)]^2\,dx \ge \int_0^\pi [f(x)]^2\,dx.$$

The equality is only attained for functions $f(x) = A\sin x$ in case (a) and functions $f(x) = B\cos x$ in case (b).

11. A *π-related inequality*. If $a > 0$ and $f(x) \ge 0$ on $[0, a]$, then

$$\left(\int_0^a f(x)\,dx\right)^4 \le \pi^2\left(\int_0^a f^2(x)\,dx\right)\left(\int_0^a x^2 f^2(x)\,dx\right).$$

7.2.5-3. Arithmetic, geometric, harmonic, and quadratic means of functions.

Let $f(x)$ be a positive function integrable on $[a, b]$. Consider the values of $f(x)$ on a discrete set of points:

$$f_{kn} = f(a + k\delta_n), \qquad \delta_n = \frac{b-a}{n} \quad (k = 1, \ldots, n).$$

The arithmetic mean, geometric mean, harmonic mean, and quadratic mean of a function $f(x)$ on an interval $[a, b]$ are introduced using the definitions of the respective mean values for finitely many numbers (see Subsection 1.6.1) and going to the limit as $n \to \infty$.

1. *Arithmetic mean of a function* $f(x)$ *on* $[a, b]$:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f_{kn} = \frac{1}{b-a} \int_a^b f(x)\, dx.$$

This definition is in agreement with another definition of the mean value of a function $f(x)$ on $[a, b]$ given in Theorem 1 from Paragraph 7.2.5-1.

2. *Geometric mean of a function* $f(x)$ *on* $[a, b]$:

$$\lim_{n\to\infty} \left(\prod_{k=1}^{n} f_{kn} \right)^{1/n} = \exp\left(\frac{1}{b-a} \int_a^b \ln f(x)\, dx \right).$$

3. *Harmonic mean of a function* $f(x)$ *on* $[a, b]$:

$$\lim_{n\to\infty} n \left(\sum_{k=1}^{n} \frac{1}{f_{kn}} \right)^{-1} = (b-a) \left(\int_a^b \frac{dx}{f(x)} \right)^{-1}.$$

4. *Quadratic mean of a function* $f(x)$ *on* $[a, b]$:

$$\lim_{n\to\infty} \left(\frac{1}{n} \sum_{k=1}^{n} f_{kn}^2 \right)^{1/2} = \left(\frac{1}{b-a} \int_a^b f^2(x)\, dx \right)^{1/2}.$$

This definition differs from the common definition of the norm of a square integrable function given in Paragraph 7.2.13-2 by the constant factor $1/\sqrt{b-a}$.

The following inequalities hold:

$$(b-a) \left(\int_a^b \frac{dx}{f(x)} \right)^{-1} \le \exp\left(\frac{1}{b-a} \int_a^b \ln f(x)\, dx \right) \le \frac{1}{b-a} \int_a^b f(x)\, dx \le \left(\frac{1}{b-a} \int_a^b f^2(x)\, dx \right)^{1/2}.$$

To make it easier to remember, let us rewrite these inequalities in words as

$$\boxed{\text{harmonic mean}} \le \boxed{\text{geometric mean}} \le \boxed{\text{arithmetic mean}} \le \boxed{\text{quadratic mean}}.$$

The equality is attained for $f(x) = \text{const}$ only.

7.2.5-4. General approach to defining mean values.

Let $g(y)$ be a continuous monotonic function defined in the range $0 \le y < \infty$.

The *mean of a function* $f(x)$ *with respect to a function* $g(x)$ *on an interval* $[a, b]$ is defined as

$$\lim_{n\to\infty} g^{-1} \left(\frac{1}{n} \sum_{k=1}^{n} g(f_{kn}) \right) = g^{-1} \left(\frac{1}{b-a} \int_a^b g\big(f(x)\big)\, dx \right),$$

where $g^{-1}(z)$ is the inverse of $g(y)$.

The means presented in Paragraph 7.2.5-3 are special cases of the mean with respect to a function:

arithmetic mean of $f(x)$ = mean of $f(x)$ with respect to $g(y) = y$,

geometric mean of $f(x)$ = mean of $f(x)$ with respect to $g(y) = \ln y$,

harmonic mean of $f(x)$ = mean of $f(x)$ with respect to $g(y) = 1/y$,

quadratic mean of $f(x)$ = mean of $f(x)$ with respect to $g(y) = y^2$.

7.2.6. Geometric and Physical Applications of the Definite Integral

7.2.6-1. Geometric applications of the definite integral.

1. The *area of a domain* D bounded by curves $y = f(x)$ and $y = g(x)$ and straight lines $x = a$ and $x = b$ in the x, y plane (see Fig. 7.2 *a*) is calculated by the formula

$$S = \int_a^b \big[f(x) - g(x)\big]\, dx.$$

If $g(x) \equiv 0$, this formula gives the area of a curvilinear trapezoid bounded by the x-axis, the curve $y = f(x)$, and the straight lines $x = a$ and $x = b$.

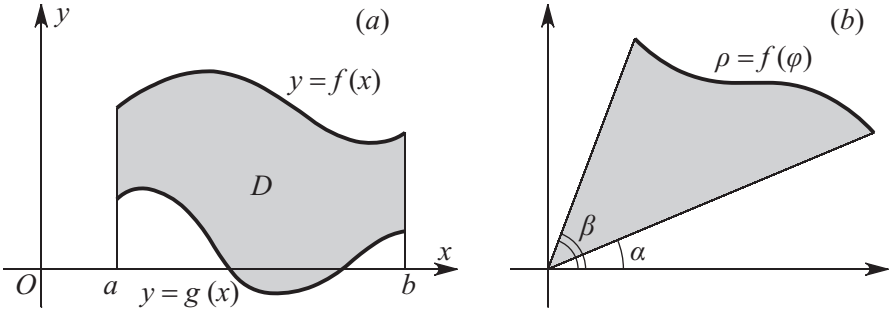


Figure 7.2. (*a*) A domain D bounded by two curves $y = f(x)$ and $y = g(x)$ on an interval $[a, b]$; (*b*) a curvilinear sector.

2. *Area of a domain* D. Let $x = x(t)$ and $y = y(t)$, with $t_1 \le t \le t_2$, be parametric equations of a piecewise-smooth simple closed curve bounding on its left (traced counterclockwise) a domain D with area S. Then

$$S = -\int_{t_1}^{t_2} y(t)x'(t)\, dt = \int_{t_1}^{t_2} x(t)y'(t)\, dt = \frac{1}{2}\int_{t_1}^{t_2} \big[x(t)y'(t) - y(t)x'(t)\big]\, dt.$$

3. *Area of a curvilinear sector*. Let a curve $\rho = f(\varphi)$, with $\varphi \in [\alpha, \beta]$, be defined in the polar coordinates ρ, φ. Then the area of the curvilinear sector $\{\alpha \le \varphi \le \beta;\ 0 \le \rho \le f(\varphi)\}$ (see Fig. 7.2 *b*) is calculated by the formula

$$S = \frac{1}{2}\int_{\alpha}^{\beta} [f(\varphi)]^2\, d\varphi.$$

4. *Area of a surface of revolution*. Let a surface of revolution be generated by rotating a curve $y = f(x) \ge 0$, $x \in [a, b]$, about the x-axis; see Fig. 7.3. The area of this surface is calculated as

$$S = 2\pi \int_a^b f(x)\sqrt{1 + [f'(x)]^2}\, dx.$$

5. *Volume of a body of revolution*. Let a body of revolution be obtained by rotating about the x-axis a curvilinear trapezoid bounded by a curve $y = f(x)$, the x-axis, and straight lines $x = a$ and $x = b$; see Fig. 7.3. Then the volume of this body is calculated as

$$V = \pi \int_a^b [f(x)]^2\, dx.$$

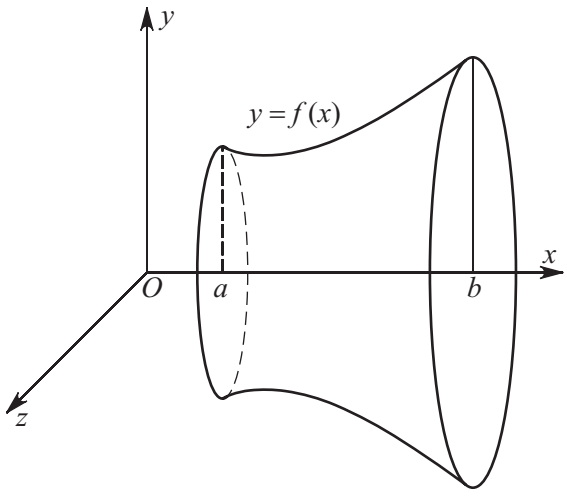


Figure 7.3. A surface of revolution generated by rotating a curve $y = f(x)$.

6. *Arc length of a plane curve defined in different ways.*

(*a*) If a curve is the graph of a continuously differentiable function $y = f(x)$, $x \in [a, b]$, then its length is determined as

$$L = \int_a^b \sqrt{1 + [f'(x)]^2}\, dx.$$

(*b*) If a plane curve is defined parametrically by equations $x = x(t)$ and $y = y(t)$, with $t \in [\alpha, \beta]$ and $x(t)$ and $y(t)$ being continuously differentiable functions, then its length is calculated by

$$L = \int_\alpha^\beta \sqrt{[x'(t)]^2 + [y'(t)]^2}\, dt.$$

(*c*) If a curve is defined in the polar coordinates ρ, φ by an equation $\rho = \rho(\varphi)$, with $\varphi \in [\alpha, \beta]$, then its length is found as

$$L = \int_\alpha^\beta \sqrt{\rho^2(\varphi) + [\rho'(\varphi)]^2}\, d\varphi.$$

7. The *arc length of a spatial curve* defined parametrically by equations $x = x(t)$, $y = y(t)$, and $z = z(t)$, with $t \in [\alpha, \beta]$ and $x(t)$, $y(t)$, and $z(t)$ being continuously differentiable functions, is calculated by

$$L = \int_\alpha^\beta \sqrt{[x'(t)]^2 + [y'(t)]^2 + [z'(t)]^2}\, dt.$$

7.2.6-2. Physical application of the integral.

1. *Work of a variable force.* Suppose a point mass moves along the x-axis from a point $x = a$ to a point $x = b$ under the action of a variable force $F(x)$ directed along the x-axis. The mechanical work of this force is equal to

$$A = \int_a^b F(x)\, dx.$$

2. *Mass of a rectilinear rod of variable density.* Suppose a rod with a constant cross-sectional area S occupies an interval $[0, l]$ on the x-axis and the density of the rod material is a function of x: $\rho = \rho(x)$. The mass of this rod is calculated as

$$m = S \int_0^l \rho(x)\, dx.$$

3. *Mass of a curvilinear rod of variable density.* Let the shape of a plane curvilinear rod with a constant cross-sectional area S be defined by an equation $y = f(x)$, with $a \le x \le b$, and let the density of the material be coordinate dependent: $\rho = \rho(x, y)$. The mass of this rod is calculated as

$$m = S \int_a^b \rho\big(x, f(x)\big) \sqrt{1 + [y'(x)]^2}\, dx.$$

If the shape of the rod is defined parametrically by $x = x(t)$ and $y = y(t)$, then its mass is found as

$$m = S \int_a^b \rho\big(x(t), y(t)\big) \sqrt{[x'(t)]^2 + [y'(t)]^2}\, dt.$$

4. The *coordinates of the center of mass of a plane homogeneous material curve* whose shape is defined by an equation $y = f(x)$, with $a \le x \le b$, are calculated by the formulas

$$x_c = \frac{1}{L} \int_a^b x \sqrt{1 + [y'(x)]^2}\, dx, \qquad y_c = \frac{1}{L} \int_a^b f(x) \sqrt{1 + [y'(x)]^2}\, dx,$$

where L is the length of the curve.

If the shape of a plane homogeneous material curve is defined parametrically by $x = x(t)$ and $y = y(t)$, then the coordinates of its center of mass are obtained as

$$x_c = \frac{1}{L} \int_a^b x(t) \sqrt{[x'(t)]^2 + [y'(t)]^2}\, dt, \qquad y_c = \frac{1}{L} \int_a^b y(t) \sqrt{[x'(t)]^2 + [y'(t)]^2}\, dt.$$

7.2.7. Improper Integrals with Infinite Integration Limit

An improper integral is an integral with an infinite limit (limits) of integration or an integral of an unbounded function.

7.2.7-1. Integrals with infinite limits.

1°. Let $y = f(x)$ be a function defined and continuous on an infinite interval $a \le x < \infty$. If there exists a finite limit $\lim\limits_{b\to\infty} \int_a^b f(x)\, dx$, then it is called a (convergent) *improper integral* of $f(x)$ on the interval $[a, \infty)$ and is denoted $\int_a^\infty f(x)\, dx$. Thus, by definition

$$\int_a^\infty f(x)\, dx = \lim_{b\to\infty} \int_a^b f(x)\, dx. \tag{7.2.7.1}$$

If the limit is infinite or does not exist, the improper integral is called *divergent*.

The *geometric meaning of an improper integral* is that the integral $\int_a^\infty f(x)\, dx$, with $f(x) \ge 0$, is equal to the area of the unbounded domain between the curve $y = f(x)$, its asymptote $y = 0$, and the straight line $x = a$ on the left.

2°. Suppose an antiderivative $F(x)$ of the integrand function $f(x)$ is known. Then the improper integral (7.2.7.1) is

(i) *convergent* if there exists a finite limit $\lim\limits_{x\to\infty} F(x) = F(\infty)$;

(ii) *divergent* if the limit is infinite or does not exist.

In case (i), we have

$$\int_a^\infty f(x)\,dx = F(x)\Big|_a^\infty = F(\infty) - F(a).$$

Example 1. Let us investigate the issue of convergence of the improper integral $I = \int_a^\infty \frac{dx}{x^\lambda}$, $a > 0$.

The integrand $f(x) = x^{-\lambda}$ has an antiderivative $F(x) = \frac{1}{1-\lambda}x^{1-\lambda}$. Depending on the value of the parameter λ, we have

$$\lim_{x\to\infty} F(x) = \frac{1}{1-\lambda}\lim_{x\to\infty} x^{1-\lambda} = \begin{cases} 0 & \text{if } \lambda > 1, \\ \infty & \text{if } \lambda \le 1. \end{cases}$$

Therefore, if $\lambda > 1$, the integral is convergent and is equal to $I = F(\infty) - F(a) = \frac{a^{1-\lambda}}{\lambda - 1}$, and if $\lambda \le 1$, the integral is divergent.

3°. Improper integrals for other infinite intervals are defined in a similar way:

$$\int_{-\infty}^b f(x)\,dx = \lim_{a\to-\infty}\int_a^b f(x)\,dx,$$
$$\int_{-\infty}^\infty f(x)\,dx = \int_{-\infty}^c f(x)\,dx + \int_c^\infty f(x)\,dx.$$

Note that if either improper integral on the right-hand side of the latter relation is convergent, then, by definition, the integral on the left-hand side is also convergent.

4°. Properties 2–4 and 6–9 from Paragraph 7.2.2-2, where a can be equal to $-\infty$ and b can be ∞, apply to improper integrals as well; it is assumed that all quantities on the right-hand sides exist (the integrals are convergent).

7.2.7-2. Sufficient conditions for convergence of improper integrals.

In many problems, it suffices to establish whether a given improper integral is convergent or not and, if yes, evaluate it. The theorems presented below can be useful in doing so.

THEOREM 1 (CAUCHY'S CONVERGENCE CRITERION). *For the integral (7.2.7.1) to be convergent it is necessary and sufficient that for any $\varepsilon > 0$ there exists a number R such that the inequality*

$$\left|\int_\alpha^\beta f(x)\,dx\right| < \varepsilon$$

holds for any $\beta > \alpha > R$.

THEOREM 2. *If $0 \le f(x) \le g(x)$ for $x \ge a$, then the convergence of the integral $\int_a^\infty g(x)\,dx$ implies the convergence of the integral $\int_a^\infty f(x)\,dx$; moreover, $\int_a^\infty f(x)\,dx \le \int_a^\infty g(x)\,dx$. If the integral $\int_a^\infty f(x)\,dx$ is divergent, then the integral $\int_a^\infty g(x)\,dx$ is also divergent.*

THEOREM 3. *If the integral* $\int_a^\infty |f(x)|\,dx$ *is convergent, then the integral* $\int_a^\infty f(x)\,dx$ *is also convergent; in this case, the latter integral is called absolutely convergent.*

Example 2. The improper integral $\int_1^\infty \frac{\sin x}{x^2}\,dx$ is absolutely convergent, since $\left|\frac{\sin x}{x^2}\right| \le \frac{1}{x^2}$ and the integral $\int_1^\infty \frac{1}{x^2}\,dx$ is convergent (see Example 1).

THEOREM 4. *Let* $f(x)$ *and* $g(x)$ *be integrable functions on any finite interval* $a \le x \le b$ *and let there exist a limit, finite or infinite,*

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = K.$$

Then the following assertions hold:

1. If $0 < K < \infty$, *both integrals*

$$\int_a^\infty f(x)\,dx, \qquad \int_a^\infty g(x)\,dx \tag{7.2.7.2}$$

are convergent or divergent simultaneously.

2. If $0 \le K < \infty$, *the convergence of the latter integral in (7.2.7.2) implies the convergence of the former integral.*

3. If $0 < K \le \infty$, *the divergence of the latter integral in (7.2.7.2) implies the divergence of the former integral.*

THEOREM 5 (COROLLARY OF THEOREM 4). *Given a function* $f(x)$, *let its asymptotics for sufficiently large* x *have the form*

$$f(x) = \frac{\varphi(x)}{x^\lambda} \qquad (\lambda > 0).$$

Then: (i) if $\lambda > 1$ *and* $\varphi(x) \le c < \infty$, *then the integral* $\int_a^\infty f(x)\,dx$ *is convergent; (ii) if* $\lambda \le 1$ *and* $\varphi(x) \ge c > 0$, *then the integral is divergent.*

THEOREM 6. *Let* $f(x)$ *be an absolutely integrable function on an interval* $[a, \infty)$ *and let* $g(x)$ *be a bounded function on* $[a, \infty)$. *Then the product* $f(x)g(x)$ *is an absolutely integrable function on* $[a, \infty)$.

THEOREM 7 (ANALOGUE OF ABEL'S TEST FOR CONVERGENCE OF INFINITE SERIES). *Let* $f(x)$ *be an integrable function on an interval* $[a, \infty)$ *such that the integral (7.2.7.1) is convergent (maybe not absolutely) and let* $g(x)$ *be a monotonic and bounded function on* $[a, \infty)$. *Then the integral*

$$\int_a^\infty f(x)g(x)\,dx \tag{7.2.7.3}$$

is convergent.

THEOREM 8 (ANALOGUE OF DIRICHLET'S TEST FOR CONVERGENCE OF INFINITE SERIES). *Let (i)* $f(x)$ *be an integrable function on any finite interval* $[a, A]$ *and*

$$\left|\int_a^A f(x)\,dx\right| \le K < \infty \qquad (a \le A < \infty);$$

(ii) $g(x)$ *be a function tending to zero monotonically as* $x \to \infty$: $\lim\limits_{x\to\infty} g(x) = 0$. *Then the integral (7.2.7.3) is convergent.*

Example 3. Let us show that the improper integral $\int_a^\infty \frac{\sin x}{x^\lambda}\,dx$ is convergent for $a > 0$ and $\lambda > 0$.

Set $f(x) = \sin x$ and $g(x) = x^{-\lambda}$ and verify conditions (i) and (ii) of Theorem 8. We have

(i) $\left|\int_a^A \sin x\,dx\right| = |\cos a - \cos A| \le 2$;

(ii) since $\lambda > 0$, the function $x^{-\lambda}$ is monotonically decreasing and goes to zero as $x \to \infty$.

So both conditions of Theorem 8 are met, and therefore the given improper integral is convergent.

7.2.7-3. Some remarks.

1°. If an improper integral is convergent and the integrand function tends to a limit as $x \to \infty$, then this limit can only be zero (it is such situations that were dealt with in Examples 1–3). However, the property $\lim\limits_{x\to\infty} f(x) = 0$ is not a *necessary condition for convergence of the integral* (7.2.7.1).

An integral can also be convergent if the integrand function does not have a limit as $x \to \infty$. For example, this is the case for *Fresnel's integrals*:

$$\int_0^\infty \sin(x^2)\,dx = \int_0^\infty \cos(x^2)\,dx = \frac{1}{2}\sqrt{\frac{\pi}{2}}.$$

Furthermore, it can be shown that the integral $\int \frac{x}{1 + x^6 \sin^2 x}\,dx$ is convergent regardless of the fact that the integrand function, being everywhere positive, is not even bounded ($f(\pi k) = \pi k,\ k = 1, 2, \ldots$). The graph of this function has infinitely many spikes with heights increasing indefinitely and base widths vanishing. At the points lying outside the spike bases, the function rapidly goes to zero.

2°. If $f(x)$ is a monotonic function for $x \ge 0$ and the improper integral $\int_0^\infty f(x)\,dx$ is convergent, then the following limiting relation holds:

$$\int_0^\infty f(x)\,dx = \lim_{\varepsilon\to 0} \varepsilon \sum_{n=1}^{\infty} f(\varepsilon n).$$

7.2.8. General Reduction Formulas for the Calculation of Improper Integrals

Below are some general formulas, involving arbitrary functions and parameters, that may facilitate the calculation of improper integrals.

7.2.8-1. Improper integrals involving power functions.

$$\int_0^\infty f\left(\frac{a+bx}{1+x}\right)\frac{dx}{(1+x)^2} = \frac{1}{b-a}\int_a^b f(x)\,dx;$$

$$\int_0^\infty \frac{f(ax) - f(bx)}{x}\,dx = \big[f(0) - f(\infty)\big]\ln\frac{b}{a} \quad \text{if } a > 0,\ b > 0,\ f(x) \text{ is continuous on } [0, \infty), \text{ and } f(\infty) = \lim_{x\to\infty} f(x) \text{ is a finite quantity;}$$

$$\int_0^\infty \frac{f(ax)-f(bx)}{x}\,dx = f(0)\ln\frac{b}{a} \quad \text{if } a>0,\ b>0,\ f(x) \text{ is continuous on } [0,\infty),$$

and the integral $\displaystyle\int_c^\infty \frac{f(x)}{x}\,dx$ exists; $c>0$;

$$\int_0^\infty f\left(\left|ax-\frac{b}{x}\right|\right)dx = \frac{1}{a}\int_0^\infty f(|x|)\,dx \quad \text{if } a>0,\ b>0;$$

$$\int_0^\infty x^2 f\left(\left|ax-\frac{b}{x}\right|\right)dx = \frac{1}{a^3}\int_0^\infty (x^2+ab)f(|x|)\,dx \quad \text{if } a>0,\ b>0;$$

$$\int_0^\infty f\left(\left|ax-\frac{b}{x}\right|\right)\frac{dx}{x^2} = \frac{1}{b}\int_0^\infty f(|x|)\,dx \quad \text{if } a>0,\ b>0;$$

$$\int_0^\infty f\left(x,\frac{1}{x}\right)\frac{dx}{x} = 2\int_0^1 f\left(x,\frac{1}{x}\right)\frac{dx}{x} \quad \text{if } f(x,y)=f(y,x);$$

$$\int_0^\infty f\left(x,\frac{a}{x}\right)\frac{dx}{x} = 0 \quad \text{if } f(x,y)=-f(y,x),\ a>0 \text{ (the integral is assumed to exist).}$$

7.2.8-2. Improper integrals involving logarithmic functions.

$$\int_0^\infty f\left(\frac{x}{a}+\frac{a}{x}\right)\frac{\ln x}{x}\,dx = \ln a\int_0^\infty f\left(\frac{x}{a}+\frac{a}{x}\right)\frac{dx}{x} \quad \text{if } a>0;$$

$$\int_0^\infty f\left(\frac{x^p}{a}+\frac{a}{x^p}\right)\frac{\ln x}{x}\,dx = \frac{\ln a}{p}\int_0^\infty f\left(\frac{x^p}{a}+\frac{a}{x^p}\right)\frac{dx}{x} \quad \text{if } a>0,\ p>0;$$

$$\int_0^\infty f(x^a+x^{-a})\frac{\ln x}{1+x^2}\,dx = 0 \quad \text{(a special case of the integral below);}$$

$$\int_0^\infty f\left(x,\frac{1}{x}\right)\frac{\ln x}{1+x^2}\,dx = 0 \quad \text{if } f(x,y)=f(y,x) \text{ (the integral is assumed to exist);}$$

$$\int_0^\infty f\left(x,\frac{1}{x}\right)\frac{\ln x}{x}\,dx = 0 \quad \text{if } f(x,y)=f(y,x) \text{ (the integral is assumed to exist).}$$

7.2.8-3. Improper integrals involving trigonometric functions.

$$\int_0^\infty f(x)\frac{\sin x}{x}\,dx = \int_0^{\pi/2} f(x)\,dx \quad \text{if } f(x)=f(-x) \text{ and } f(x+\pi)=f(x);$$

$$\int_0^\infty f(x)\frac{\sin x}{x}\,dx = \int_0^{\pi/2} f(x)\cos x\,dx \quad \text{if } f(x)=f(-x) \text{ and } f(x+\pi)=-f(x);$$

$$\int_0^\infty \frac{f(\sin x)}{x}\,dx = \int_0^{\pi/2}\frac{f(\sin x)}{\sin x}\,dx \quad \text{if } f(-x)=-f(x);$$

$$\int_0^\infty \frac{f(\sin x)}{x^2}\,dx = \int_0^{\pi/2}\frac{f(\sin x)}{\sin^2 x}\,dx \quad \text{if } f(x)=f(-x);$$

$$\int_0^\infty \frac{f(\sin x)}{x}\cos x\,dx = \int_0^{\pi/2}\frac{f(\sin x)}{\sin x}\cos^2 x\,dx \quad \text{if } f(-x)=-f(x);$$

$$\int_0^\infty \frac{f(\sin x)}{x}\tan x\,dx = \int_0^{\pi/2} f(\sin x)\,dx \quad \text{if } f(-x) = f(x);$$

$$\int_0^\infty \frac{f(\sin x)}{x^2+a^2}\,dx = \frac{\sinh(2a)}{2a}\int_0^{\pi/2} \frac{f(\sin x)\,dx}{\cosh^2 a - \cos^2 x} \quad \text{if } f(-x) = f(x);$$

$$\int_0^\infty f\left(x+\frac{1}{x}\right)\frac{\arctan x}{x}\,dx = \frac{\pi}{4}\int_0^\infty f\left(x+\frac{1}{x}\right)\frac{dx}{x}.$$

7.2.8-4. Calculation of improper integrals using analytic functions.

Suppose

$$F(z) = f(r,x) + ig(r,x), \quad z = r(\cos x + i\sin x), \quad i^2 = -1,$$

where $F(z)$ is a function analytic in a circle of radius r. Then the following formulas hold:

$$\int_0^\infty \frac{f(x,r)}{x^2+a^2}\,dx = \frac{\pi}{2a}F(re^{-a});$$

$$\int_0^\infty \frac{xg(x,r)}{x^2+a^2}\,dx = \frac{\pi}{2}[F(re^{-a}) - F(0)];$$

$$\int_0^\infty \frac{g(x,r)}{x}\,dx = \frac{\pi}{2}[F(r) - F(0)];$$

$$\int_0^\infty \frac{g(x,r)}{x(x^2+a^2)}\,dx = \frac{\pi}{2a^2}[F(r) - F(re^{-a})].$$

► *Paragraph 10.1.2-8 presents a method for the calculation of improper integrals using the theory of functions of a complex variable.*

7.2.8-5. Calculation of improper integrals using the Laplace transform.

The following classes of improper integrals may be evaluated using the Laplace transform:

$$\begin{aligned}
&\int_0^\infty \frac{f(x)}{x}\,dx = \int_0^\infty \widetilde{f}(p)\,dp,\\
&\int_0^\infty x^n f(x)\,dx = (-1)^{n+1}\left[\frac{d^n}{dp^n}\widetilde{f}(p)\right]_0^\infty, \qquad n = 1, 2, \ldots,
\end{aligned} \tag{7.2.8.1}$$

where $\widetilde{f}(p)$ is the *Laplace transform* of the function $f(x)$, which is defined as

$$\widetilde{f}(p) = \int_0^\infty e^{-px} f(x)\,dx.$$

Short notation for the Laplace transform: $\widetilde{f}(p) = \mathfrak{L}\{f(x)\}$.

Section 11.2 presents properties and methods for determining the *Laplace transform*, and Section T3.1 gives tables of the Laplace transforms of various functions.

Example 1. Evaluate the integral $\displaystyle\int_0^\infty \frac{\sin(ax)}{x}\,dx$.

Using Table 11.2 from Subsection 11.2.2 (or the table from Subsection T3.1.6), we find the Laplace transform of the function $\sin(ax)$: $\mathfrak{L}\{\sin(ax)\} = \dfrac{a}{a^2+p^2}$. Substitute this expression into the first formula in (7.2.8.1) and integrate to obtain

$$\int_0^\infty \frac{\sin(ax)}{x}\,dx = \int_0^\infty \frac{a\,dp}{a^2+p^2} = \arctan\frac{p}{a}\bigg|_0^\infty = \frac{\pi}{2}.$$

Example 2. Evaluate *Frullani's integral* $\displaystyle\int_0^\infty \frac{e^{-ax}-e^{-bx}}{x}\,dx$, where $a>0$ and $b>0$.

Using the first formula in (7.2.8.1) and Table 11.2 from Subsection 11.2.2 (or the table from Subsection T3.1.3), we obtain $\mathfrak{L}\{e^{-ax}\} = \dfrac{1}{p+a}$. Integrating yields

$$\int_0^\infty \frac{e^{-ax}-e^{-bx}}{x}\,dx = \int_0^\infty \left(\frac{1}{p+a}-\frac{1}{p+b}\right)dp = \ln\frac{p+a}{p+b}\bigg|_0^\infty = -\ln\frac{a}{b} = \ln\frac{b}{a}.$$

7.2.9. General Asymptotic Formulas for the Calculation of Improper Integrals

Below are some general formulas, involving arbitrary functions and parameters, that may be useful for determining the asymptotic behavior of improper integrals.

7.2.9-1. Asymptotic formulas for some improper integrals with parameter.

1°. For asymptotics of improper Laplace integrals

$$I(\lambda) = \int_a^\infty f(x)\exp[\lambda g(x)]\,dx$$

as $\lambda\to\infty$, see Remark 1 in Paragraph 7.2.4-3.

2°. For $\lambda\to\infty$, the following asymptotic expansions of improper integrals involving trigonometric functions and a Bessel function hold:

$$\int_0^\infty \cos(\lambda x)f(x)\,dx = \sum_{k=1}^{n}(-1)^k f^{(2k-1)}(0)\lambda^{-2k} + O(\lambda^{-2n-1}),$$

$$\int_0^\infty \sin(\lambda x)f(x)\,dx = \sum_{k=0}^{n-1}(-1)^k f^{(2k)}(0)\lambda^{-2k-1} + O(\lambda^{-2n-1}),$$

$$\int_0^\infty J_0(\lambda x)f(x)\,dx = \frac{1}{\sqrt{\pi}}\sum_{k=0}^{n-1}\frac{(-1)^k}{k!}\Gamma\left(k+\frac{1}{2}\right)f^{(2k)}(0)\lambda^{-2k-1} + O(\lambda^{-2n-1}).$$

The function $f(x)$ is assumed to have $2n+1$ partial derivatives with respect to x for $x\ge 0$ that monotonically go to zero as $x\to\infty$.

3°. For $\lambda \to \infty$, the following asymptotic expansions hold:

$$\int_0^\infty f(x)g\left(\frac{x}{\lambda}\right)dx = \sum_{k=0}^{n} \frac{(-1)^k}{k!} F^{(k)}(0)g^{(k)}(0)\lambda^{-k} + O(\lambda^{-n-1}),$$

$$\int_0^\infty f(\lambda x)g(x)\,dx = \sum_{k=0}^{n} \frac{(-1)^k}{k!} F^{(k)}(0)g^{(k)}(0)\lambda^{-k-1} + O(\lambda^{-n-2}),$$

where $F(t) = \int_0^\infty f(x)e^{-xt}\,dx$ is the Laplace transform of the function $f(x)$.

4°. For $\lambda \to \infty$, the following asymptotic expansions hold:

$$\int_0^\infty \frac{f(x)\,dx}{x+\lambda} = \sum_{k=0}^{n} \frac{F^{(k)}(0)}{\lambda^{k+1}} + O(\lambda^{-n-2}),$$

where $F(t) = \int_0^\infty f(x)e^{-xt}\,dx$ is the Laplace transform of the function $f(x)$.

7.2.9-2. Behavior of integrals with variable limit of integration as $x \to \infty$.

Let $f(t)$ be a continuously differentiable function, let $g(t)$ be a twice continuously differentiable function, and let the following conditions hold:

$$f(t) > 0, \quad g'(t) > 0; \quad g(t) \to \infty \text{ as } t \to \infty;$$
$$f'(t)/f(t) = o\big(g'(t)\big) \text{ as } t \to \infty; \quad g''(t) = o\big(g'^2(t)\big) \text{ as } t \to \infty.$$

Then the following asymptotic relation holds as $x \to \infty$:

$$\int_x^\infty f(t)\exp[-g(t)]\,dt \simeq \frac{f(x)}{g'(x)}\exp[-g(x)].$$

7.2.9-3. π-related inequality.

If $f(x) \ge 0$, the inequality

$$\left(\int_0^\infty f(x)\,dx\right)^4 \le \pi^2\left(\int_0^\infty f^2(x)\,dx\right)\left(\int_0^\infty x^2 f^2(x)\,dx\right)$$

holds, provided the integral on the left-hand side exists. The constant π^2 is best in the sense that there exist functions $f(x) \not\equiv 0$ for which the equality is attained.

7.2.10. Improper Integrals of Unbounded Functions

7.2.10-1. Basic definitions.

1°. Let a function $f(x)$ be defined and continuous for $a \le x < b$, but $\lim\limits_{x\to b-0} f(x) = \infty$. If there exists a finite limit $\lim\limits_{\lambda\to b-0}\int_a^\lambda f(x)\,dx$, it is called the (convergent) *improper integral of the unbounded function* $f(x)$ over the interval $[a, b]$. Thus, by definition

$$\int_a^b f(x)\,dx = \lim_{\lambda\to b-0}\int_a^\lambda f(x)\,dx. \tag{7.2.10.1}$$

If no finite limit exists, the integral is called *divergent*.

If $\lim\limits_{x\to a+0} f(x) = \infty$, then, by definition, it is assumed that

$$\int_a^b f(x)\,dx = \lim_{\gamma\to a+0} \int_\gamma^b f(x)\,dx.$$

Finally, if $f(x)$ is unbounded near a point $c \in (a, b)$ and both integrals $\int_a^c f(x)\,dx$ and $\int_c^b f(x)\,dx$ are convergent, then, by definition,

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

2°. The geometric meaning of an improper integral of an unbounded function and also sufficient conditions for convergence of such integrals are similar to those for improper integrals with infinite limit(s).

7.2.10-2. Convergence tests for improper integrals of unbounded functions.

Presented below are theorems for the case where the only singular point of the integrand function is the right endpoint of the interval $[a, b]$.

THEOREM 1 (CAUCHY'S CONVERGENCE CRITERION). *For the integral (7.2.10.1) to be convergent is it necessary and sufficient that for any $\varepsilon > 0$ there exists a number $\delta > 0$ such that for any δ_1 and δ_2 satisfying $0 < \delta_1 < \delta$ and $0 < \delta_2 < \delta$ the following inequality holds:*

$$\left|\int_{b-\delta_1}^{b-\delta_2} f(x)\,dx\right| < \varepsilon.$$

THEOREM 2. *If $0 \le f(x) \le g(x)$ for $a \le x < b$, then the convergence of the integral $\int_a^b g(x)\,dx$ implies the convergence of the integral $\int_a^b f(x)\,dx$, with $\int_a^b f(x)\,dx \le \int_a^b g(x)\,dx$. If the integral $\int_a^b f(x)\,dx$ is divergent, then the integral $\int_a^b g(x)\,dx$ is also divergent.*

Example. For any continuous function $\varphi(x)$ such that $\varphi(1) = 0$, the improper integral $\int_0^1 \dfrac{dx}{\varphi^2(x) + \sqrt{1-x}}$ is convergent and does not exceed 2, since $\dfrac{1}{\varphi^2(x) + \sqrt{1-x}} < \dfrac{1}{\sqrt{1-x}}$, while the integral $\int_0^1 \dfrac{dx}{\sqrt{1-x}}$ is convergent and is equal to 2.

THEOREM 3. *Let $f(x)$ and $g(x)$ be continuous functions on $[a, b)$ and let the following limit exist:*

$$\lim_{x\to b} \frac{f(x)}{g(x)} = K \qquad (0 < K < \infty).$$

Then both integrals

$$\int_a^b f(x)\,dx, \qquad \int_a^b g(x)\,dx$$

are either convergent or divergent simultaneously.

THEOREM 4. *Let a function $f(x)$ be representable in the form*

$$f(x) = \frac{\varphi(x)}{(b-x)^{\lambda}} \qquad (\lambda > 0),$$

where $\varphi(x)$ is continuous on $[a, b]$ and the condition $\varphi(b) \neq 0$ holds.

Then: (i) if $\lambda < 1$ and $\varphi(x) \le c < \infty$, then the integral $\int_a^b f(x)\,dx$ is convergent; (ii) if $\lambda \ge 1$ and $\varphi(x) \ge c > 0$, this integral is divergent.

Remark. The issue of convergence of the integral of an unbounded function (7.2.10.1) at $x = b$ can be reduced by a simple change of variable to the issue of convergence of an improper integral with an infinite limit:

$$\int_a^b f(x)\,dx = (b-a)\int_1^{\infty} f\Big(\frac{a-b}{z} + b\Big)\frac{dz}{z^2}, \qquad z = \frac{b-a}{b-x}.$$

7.2.10-3. Calculation of integrals using infinite sums of special form.

Let a function $f(x)$ be continuous and monotonic on the interval $(0, 1)$, whose endpoints can be singular. If the integral (proper or improper) $\int_0^1 f(x)\,dx$ exists, then the following limiting relations hold:

$$\int_0^1 f(x)\,dx = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n-1} f\left(\frac{k}{n}\right) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f\left(\frac{2k-1}{2n}\right).$$

7.2.11. Cauchy-Type Singular Integrals

7.2.11-1. Hölder and Lipschitz conditions.

We say that $f(x)$ satisfies the *Hölder condition* on $[a, b]$ if for any two points $x_1 \in [a, b]$ and $x_2 \in [a, b]$ we have

$$|f(x_2) - f(x_1)| < A|x_2 - x_1|^{\lambda}, \tag{7.2.11.1}$$

where A and λ are positive constants. The number A is called the *Hölder constant* and λ is called the *Hölder exponent*. If $\lambda > 1$, then by condition (7.2.11.1) the derivative $f'_x(x)$ vanishes everywhere, and $f(x)$ must be constant. Therefore, we assume that $0 < \lambda \le 1$. For $\lambda = 1$, the Hölder condition is often called the *Lipschitz condition*. Sometimes the Hölder condition is called the *Lipschitz condition of order* λ.

If x_1 and x_2 are sufficiently close to each other and if the Hölder condition holds for some exponent λ_1, then this condition certainly holds for each exponent $\lambda < \lambda_1$. In general, the converse assertion fails. The smaller λ, the broader is the class of Hölder continuous functions. The narrowest class is that of functions satisfying the Lipschitz condition.

It follows from the last property that if functions $f_1(x)$ and $f_2(x)$ satisfy the Hölder condition with exponents λ_1 and λ_2, respectively, then their sum and the product, as well as their ratio provided that the denominator is nonzero, satisfy the Hölder condition with exponent $\lambda = \min(\lambda_1, \lambda_2)$.

If $f(x)$ is differentiable and has a bounded derivative, then $f(x)$ satisfies the Lipschitz condition. In general, the converse assertion fails.

7.2.11-2. Principal value of a singular integral.

1°. Consider the integral

$$\int_a^b \frac{dx}{x-c}, \qquad a < c < b.$$

Evaluating this integral as an improper integral, we obtain

$$\int_a^b \frac{dx}{x-c} = \lim_{\substack{\varepsilon_1\to 0\\ \varepsilon_2\to 0}} \left(-\int_a^{c-\varepsilon_1} \frac{dx}{c-x} + \int_{c+\varepsilon_2}^b \frac{dx}{x-c} \right) = \ln\frac{b-c}{c-a} + \lim_{\substack{\varepsilon_1\to 0\\ \varepsilon_2\to 0}} \ln\frac{\varepsilon_1}{\varepsilon_2}. \qquad (7.2.11.2)$$

The limit of the last expression obviously depends on the way in which ε_1 and ε_2 tend to zero. Hence, the improper integral does not exist. This integral is called a *singular integral*. However, this integral can be assigned a meaning if we assume that there is some relationship between ε_1 and ε_2. For example, if the deleted interval is symmetric with respect to the point c, i.e.,

$$\varepsilon_1 = \varepsilon_2 = \varepsilon, \qquad (7.2.11.3)$$

we arrive at the notion of the Cauchy principal value of a singular integral.

The *Cauchy principal value* of the singular integral

$$\int_a^b \frac{dx}{x-c}, \qquad a < c < b$$

is the number

$$\lim_{\varepsilon\to 0}\left(\int_a^{c-\varepsilon} \frac{dx}{x-c} + \int_{c+\varepsilon}^b \frac{dx}{x-c} \right).$$

With regard to formula (7.2.11.2), we have

$$\int_a^b \frac{dx}{x-c} = \ln\frac{b-c}{c-a}. \qquad (7.2.11.4)$$

2°. Consider the more general integral

$$\int_a^b \frac{f(x)}{x-c}\,dx, \qquad (7.2.11.5)$$

where $f(x) \in [a, b]$ is a function satisfying the Hölder condition. Let us understand this integral in the sense of the Cauchy principal value, which we define as follows:

$$\int_a^b \frac{f(x)}{x-c}\,dx = \lim_{\varepsilon\to 0}\left(\int_a^{c-\varepsilon} \frac{f(x)}{x-c}\,dx + \int_{c+\varepsilon}^b \frac{f(x)}{x-c}\,dx \right).$$

We have the identity

$$\int_a^b \frac{f(x)}{x-c}\,dx = \int_a^b \frac{f(x)-f(c)}{x-c}\,dx + f(c)\int_a^b \frac{dx}{x-c};$$

moreover, the first integral on the right-hand side is convergent as an improper integral, because it follows from the Hölder condition that

$$\left| \frac{f(x)-f(c)}{x-c} \right| < \frac{A}{|x-c|^{1-\lambda}}, \qquad 0 < \lambda \le 1,$$

and the second integral coincides with (7.2.11.4).

Thus, we see that the singular integral (7.2.11.5), where $f(x)$ satisfies the Hölder condition, exists in the sense of the Cauchy principal value and is equal to

$$\int_a^b \frac{f(x)}{x-c}\,dx = \int_a^b \frac{f(x)-f(c)}{x-c}\,dx + f(c)\ln\frac{b-c}{c-a}.$$

Some authors denote singular integrals by special symbols like v.p. $\int$ (*valeur principale*). However, this is not necessary because, on the one hand, if an integral of the form (7.2.11.5) exists as a proper or an improper integral, then it exists in the sense of the Cauchy principal value, and their values coincide; on the other hand, we shall always understand a singular integral in the sense of the Cauchy principal value. For this reason, we denote a singular integral by the usual integral sign.

7.2.12. Stieltjes Integral

7.2.12-1. Basic definitions.

Let $f(x)$ and $\varphi(x)$ be functions defined on an interval $[a, b]$. Let us partition this interval into n elementary subintervals defined by a set of points $\{x_0, x_1, \ldots, x_n\}$ such that $a = x_0 < x_1 < \cdots < x_n = b$. Each subinterval $[x_{k-1}, x_k]$ will be characterized by its length $\Delta x_k = x_k - x_{k-1}$ and an arbitrarily chosen point $\xi_k \in [x_{k-1}, x_k]$. Let us make up a *Stieltjes integral sum*

$$s_n = \sum_{k=1}^{n} f(\xi_k)\Delta_k\varphi(x),$$

where $\Delta_k\varphi(x) = \varphi(x_k) - \varphi(x_{k-1})$ is the increment of the function $\varphi(x)$ on the kth elementary subinterval.

If there exists a limit of the integral sums s_n, as the number of subintervals n increases indefinitely so that the length of every subinterval Δx_k vanishes, and this limit depends on neither the way the interval $[a, b]$ was partitioned nor the way the points ξ_k were selected, then this limit is called the *Stieltjes integral* of the function $f(x)$ with respect to the function $\varphi(x)$ over the interval $[a, b]$:

$$\int_a^b f(x)\,d\varphi(x) = \lim_{\lambda\to 0} s_n \qquad \Big(\max_{1\le k\le n} \Delta x_k \to 0 \ \text{ as } \ n \to \infty\Big).$$

Then $f(x)$ is called an *integrable function* with respect to $\varphi(x)$, and $\varphi(x)$ is called an *integrating function*.

The Stieltjes integral is a generalization of the Riemann integral; the latter corresponds to the special case $\varphi(x) = x + \text{const}$.

7.2.12-2. Properties of the Stieltjes integral.

The Stieltjes integral has properties analogous to those of the definite Riemann integral:

1) $\displaystyle\int_a^b d\varphi(x) = \varphi(b) - \varphi(a);$

2) $\displaystyle\int_a^b \big[Af(x) \pm Bg(x)\big]\,d\varphi(x) = A\int_a^b f(x)\,d\varphi(x) \pm B\int_a^b g(x)\,d\varphi(x);$

3) $\displaystyle\int_a^b f(x)\,d[A\varphi(x)\pm B\psi(x)] = A\int_a^b f(x)\,d\varphi(x)\pm B\int_a^b f(x)\,d\psi(x);$

4) $\displaystyle\int_a^b f(x)\,d\varphi(x) = \int_a^c f(x)\,d\varphi(x) + \int_c^b f(x)\,d\varphi(x) \quad (a<c<b).$

It is assumed that all integrals on the left- and right-hand sides exist.

THEOREM (MEAN VALUE). *If a function $f(x)$ satisfies inequalities $m\le f(x)\le M$ on an interval $[a,b]$ and is integrable with respect to an increasing function $\varphi(x)$, then*

$$\int_a^b f(x)\,d\varphi(x) = \mu[\varphi(b)-\varphi(a)],$$

where $m<\mu<M$.

7.2.12-3. Existence theorems for the Stieltjes integral.

The existence of the Stieltjes integral and its reduction to the Riemann integral is established by the following theorem.

THEOREM 1. *If $f(x)$ is continuous on $[a,b]$ and $\varphi(x)$ has a bounded variation* on $[a,b]$, then the integral $\int_a^b f(x)\,d\varphi(x)$ exists.*

THEOREM 2. *Let $f(x)$ be integrable on $[a,b]$ in the sense of Riemann and let $\varphi(x)$ satisfy the Lipschitz condition*

$$|\varphi(x_2)-\varphi(x_1)|<K|x_2-x_1|,$$

where x_1 and x_2 are arbitrary points of the interval $[a,b]$ and K is a fixed positive constant. Then the function $f(x)$ is integrable with respect to the function $\varphi(x)$.

THEOREM 3. *Let $f(x)$ be integrable on $[a,b]$ in the sense of Riemann and let $\varphi(x)$ be differentiable and have an integrable derivative on $[a,b]$. Then the function $f(x)$ is integrable with respect to the function $\varphi(x)$ and, moreover,*

$$\int_a^b f(x)\,d\varphi(x) = \int_a^b f(x)\varphi'(x)\,dx,$$

where the integral on the right-hand side is understood in the sense of Riemann.

Remark. If a function $f(x)$ is integrable on an interval $[a,b]$ with respect to a function $\varphi(x)$, then, vice versa, the function $\varphi(x)$ is also integrable with respect to the function $f(x)$ on $[a,b]$. Owing to this property, the functions $f(x)$ and $\varphi(x)$ are interchangeable in Theorems 1 and 2.

THEOREM 4. *Let $f(x)$ be continuous on $[a,b]$ and let $\varphi(x)$ have an absolutely integrable derivative $\varphi'(x)$ everywhere on $[a,b]$, except, perhaps, finitely many points. Let, in addition, the function $\varphi(x)$ undergo a jump discontinuity at finitely many points*

$$a=c_0<c_1<\cdots<c_m=b.$$

* A function $\varphi(x)$ is said to have a *bounded variation* on an interval $[a,b]$ if there exists a number $M>0$ such that for any set of points $a=x_0<x_1<\cdots<x_n=b$ the inequality $\sum\limits_{k=1}^{n}|\varphi(x_{k+1})-\varphi(x_k)|<M$ holds (see also Subsection 6.1.7).

Then the Stieltjes integral exists and is calculated as

$$\int_a^b f(x)\,d\varphi(x) = \int_a^b f(x)\varphi'(x)\,dx + f(a)[\varphi(a+0)-\varphi(a)] + \sum_{k=1}^{m-1} f(c_k)[\varphi(c_k+0)-\varphi(c_k-0)] + f(b)[\varphi(b)-\varphi(b-0)],$$

where the right-hand side contains a Riemann integral. Note the presence of terms outside the integral on the right-hand side, where, apart from the ordinary jumps of the function $\varphi(x)$ at the internal points of discontinuity, there are terms with one-sided jumps at the endpoints (if there is no jump at either endpoint, the corresponding term vanishes).

The Stieltjes integral is useful for finding static moments, moments of inertia, and some other distributed quantities on an interval $[a, b]$, where, apart from continuous distributions, there are concentrated quantities like point masses that correspond to a discontinuous function $\varphi(x)$ with finite jumps.

7.2.13. Square Integrable Functions

7.2.13-1. Definitions.

A function $f(x)$ is said to be *square integrable* on an interval $[a, b]$ if $f^2(x)$ is integrable on $[a, b]$. The set of all square integrable functions is denoted by $L_2(a, b)$ or, briefly, L_2.* Likewise, the set of all integrable functions on $[a, b]$ is denoted by $L_1(a, b)$ or, briefly, L_1.

7.2.13-2. Basic properties of functions from L_2.

1°. The sum of two square integrable functions is a square integrable function.

2°. The product of a square integrable function by a constant is a square integrable function.

3°. The product of two square integrable functions is an integrable function.

4°. If $f(x) \in L_2$ and $g(x) \in L_2$, then the following Cauchy–Schwarz–Bunyakovsky inequality holds:

$$(f, g)^2 \le \|f\|^2 \|g\|^2,$$

$$(f, g) = \int_a^b f(x)g(x)\,dx, \quad \|f\|^2 = (f, f) = \int_a^b f^2(x)\,dx.$$

The number (f, g) is called the *inner product* of the functions $f(x)$ and $g(x)$ and the number $\|f\|$ is called the L_2*-norm* of $f(x)$.

5°. For $f(x) \in L_2$ and $g(x) \in L_2$, the following *triangle inequality* holds:

$$\|f + g\| \le \|f\| + \|g\|.$$

6°. Let functions $f(x)$ and $f_1(x), f_2(x), \ldots, f_n(x), \ldots$ be square integrable on an interval $[a, b]$. If

$$\lim_{n\to\infty} \int_a^b \big[f_n(x) - f(x)\big]^2\,dx = 0,$$

* In the most general case the integral is understood as the Lebesgue integral of measurable functions. As usual, two equivalent functions — i.e., equal everywhere, or distinct on a negligible set (of zero measure) — are regarded as one and the same element of L_2.

then the sequence $f_1(x)$, $f_2(x)$, ... is said to be *mean-square convergent* to $f(x)$.

Note that if a sequence of functions $\{f_n(x)\}$ from L_2 converges uniformly to $f(x)$, then $f(x) \in L_2$ and $\{f_n(x)\}$ is mean-square convergent to $f(x)$.

7.2.14. Approximate (Numerical) Methods for Computation of Definite Integrals

7.2.14-1. Rectangle, trapezoidal, and Simpson's rules.

For approximate computation of an integral like $\int_a^b f(x)\,dx$, let us break up the interval $[a, b]$ into n equal subintervals with length $h = \dfrac{b-a}{n}$. Introduce the notation: $x_0 = a$, x_1, ..., $x_n = b$ (the partition points), $y_i = f(x_i)$, $i = 0,\ 1,\ \ldots,\ n$.

1°. *Rectangle rules*:

$$\int_a^b f(x)\,dx \approx h(y_0 + y_1 + \cdots + y_{n-1}),$$

$$\int_a^b f(x)\,dx \approx h(y_1 + y_2 + \cdots + y_n).$$

The error of these formulas, R_n, is proportional to h and is estimated using the inequality

$$|R_n| \le \tfrac{1}{2}h(b-a)M_1, \qquad M_1 = \max_{a\le x\le b}\bigl|f'(x)\bigr|.$$

2°. *Trapezoidal rule*:

$$\int_a^b f(x)\,dx \approx h\Bigl(\frac{y_0 + y_n}{2} + y_1 + y_2 + \cdots + y_{n-1}\Bigr).$$

The error of this formula is proportional to h^2 and is estimated as

$$|R_n| \le \tfrac{1}{12}h^2(b-a)M_2, \qquad M_2 = \max_{a\le x\le b}\bigl|f''(x)\bigr|.$$

3°. *Simpson's rule*:

$$\int_a^b f(x)\,dx \approx \tfrac{1}{3}h[y_0 + y_n + 4(y_1 + y_3 + \cdots + y_{n-1}) + 2(y_2 + y_4 + \cdots + y_{n-2})],$$

where n is even. The error of approximation by Simpson's rule is proportional to h^4:

$$|R_n| \le \tfrac{1}{180}h^4(b-a)M_4, \qquad M_4 = \max_{a\le x\le b}\bigl|f^{(4)}(x)\bigr|.$$

Simpson's rule yields exact results for the case where the integrand function is a polynomial of degree two or three.

Remark. The above approximation formulas are often used for numerical computation of definite integrals; to achieve a higher accuracy, large n are normally taken.

7.2.14-2. Computation of integrals using uniformly distributed sequences.

1°. Consider a sequence of real numbers $\{x_n\}$ whose members all belong to the closed interval $[0, 1]$. Let $\nu_n(a, b)$ be the number of the members of the sequence with subscript $< n$ that belong to $[a, b] \in [0, 1]$. The sequence $\{x_n\}$ is called uniformly distributed on the interval $[0, 1]$ if

$$\lim_{n\to\infty} \frac{\nu_n(a, b)}{n} = b - a.$$

In the language of probability theory, this definition means that the probability of a randomly selected element of the sequence to fall into a subinterval $[a, b]$ is equal to the length of this subinterval.

A simple example of a uniformly distributed sequence is the sequence of all proper fractions m/n with $n = 2, 3, \ldots$ and $m = 1, 2, \ldots n-1$:

$$\frac{1}{2}, \quad \frac{1}{3}, \quad \frac{2}{3}, \quad \frac{1}{4}, \quad \frac{2}{4}, \quad \frac{3}{4}, \quad \ldots$$

The following theorem serves as an unlimited source for constructing uniformly distributed sequences.

THEOREM 1. *Let ϑ be an arbitrary irrational number. Then the sequence $x_n = n\vartheta - [n\vartheta]$, $n = 1, 2, \ldots$, is uniformly distributed on the interval $[0, 1]$; the symbol $[z]$ stands for the integer part of z (the maximum integer not exceeding z).*

2°. Uniformly distributed sequences can be used for the calculation of integrals on the basis of the following two theorems.

THEOREM 2. *If a number sequence $\{x_n\}$ is uniformly distributed on the interval $[0, 1]$, then the following limiting relation holds for any function $f(x)$ integrable on $[0, 1]$:*

$$\int_0^1 f(x)\,dx = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(x_k). \tag{7.2.14.1}$$

Conversely, if this relation holds, it follows that the sequence $\{x_n\}$ is uniformly distributed on $[0, 1]$.

Remark. Formula (7.2.14.1) can be used for approximate calculation of integrals by defining a uniformly distributed sequence $\{x_n\}$ and computing the sum with a large n. The integrals defined on an arbitrary interval $[a, b]$ are first reduced by the change of variable $z = \dfrac{x-a}{b-a}$ to an integral over the interval $[0, 1]$.

THEOREM 3. *If a number sequence $\{x_n\}$ is uniformly distributed on $[0, 1]$, then the following limiting relation holds for any function $f(x)$ integrable on $[0, 1]$ and any monotonically decreasing sequence of positive numbers $\{\alpha_n\}$ with a divergent sum:*

$$\int_0^1 f(x)\,dx = \lim_{n\to\infty} \frac{\alpha_1 f(x_1) + \alpha_2 f(x_2) + \cdots + \alpha_n f(x_n)}{\alpha_1 + \alpha_2 + \cdots + \alpha_n}.$$

7.2.14-3. Computation of integrals by series expansion of the integrand function.

THEOREM 4. *If a sequence of functions, $\{f_n(x)\}$, continuous on an interval $[a, b]$ uniformly converges on $[a, b]$ to a function $f(x)$, then the sequence of functions $\left\{\int_a^b f_n(x)\,dx\right\}$ uniformly converges on $[a, b]$ to the function $\left\{\int_a^b f(x)\,dx\right\}$.*

THEOREM 5. *If the functions of a sequence $\{u_n(x)\}$ are continuous on an interval $[a, b]$ and the series $\sum\limits_{n=1}^{\infty} u_n(x)$ uniformly converges on $[a, b]$ to a function $f(x)$, then the series $\sum\limits_{n=1}^{\infty} \int_a^x u_n(t)\,dt$ is also uniformly convergent on $[a, b]$ and*

$$\sum_{n=1}^{\infty} \int_a^b u_n(x)\,dx = \int_a^b \left(\sum_{n=1}^{\infty} u_n(x)\,dx\right) = \int_a^b f(x)\,dx;$$

$$\sum_{n=1}^{\infty} \int_a^x u_n(t)\,dt = \int_a^x \left(\sum_{n=1}^{\infty} u_n(t)\,dt\right) = \int_a^x f(t)\,dt \qquad (a \le x \le b).$$

This theorem is used for the calculation of integrals by expanding the integrand functions $f(x)$ into a uniformly convergent series.

Remark 1. Since a series of positive continuous functions with a continuous sum is always uniformly convergent, then this series can be integrated termwise.

Remark 2. In general, the convergence of a sequence of continuous integrand functions, $\{f_n(x)\}$, at each point of the integration interval does not guarantee that it is permitted to proceed to the limit under the sign of integral.

7.3. Double and Triple Integrals

7.3.1. Definition and Properties of the Double Integral

7.3.1-1. Definition and properties of the double integral.

Suppose there is a bounded set of points defined on the plane, so that it can be placed in a minimal enclosing circle. The diameter of this circle is called the *diameter of the set*. Consider a domain D in the x, y plane. Let us partition D into n nonintersecting subdomains (cells). The largest of the cell diameters is called the *partition diameter* and is denoted $\lambda = \lambda(\mathcal{D}_n)$, where $\mathcal{D}_n$ stands for the partition of the domain D into cells. Let a function $z = f(x, y)$ be defined in D. Select an arbitrary point in each cell (x_i, y_i), $i = 1, 2, \ldots, n$, and make up an *integral sum*,

$$s_n = \sum_{i=1}^{n} f(x_i, y_i)\,\Delta S_i,$$

where ΔS_i is the area of the ith subdomain.

If there exists a finite limit, $\mathcal{J}$, of the sums s_n as $\lambda \to 0$ and it depends on neither the partition $\mathcal{D}_n$ nor the selection of the points (x_i, y_i), this limit is denoted $\iint_D f(x, y)\,dx\,dy$ and is called the double integral of the function $f(x, y)$ over the domain D:

$$\iint_D f(x, y)\,dx\,dy = \lim_{\lambda \to 0} s_n.$$

This means that for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for all partitions $\mathcal{D}_n$ such that $\lambda(\mathcal{D}_n) < \delta$ and for any selection of the points (x_i, y_i), the inequality $|s_n - \mathcal{J}| < \varepsilon$ holds.

7.3.1-2. Classes of integrable functions.

Further on, it is assumed that D is a closed bounded domain.

1. If $f(x, y)$ is continuous in D, then the double integral $\iint_D f(x, y)\,dx\,dy$ exists.

2. If $f(x, y)$ is bounded and the set of points of discontinuity of $f(x, y)$ has a zero area (e.g., the points of discontinuity lie on finitely many continuous curves in the x, y plane), then the double integral of $f(x, y)$ over the domain D exists.

7.3.1-3. Properties of the double integral.

1. *Linearity.* If functions $f(x, y)$ and $g(x, y)$ are integrable in D, then

$$\iint_D \bigl[a f(x, y) \pm b g(x, y)\bigr]\,dx\,dy = a \iint_D f(x, y)\,dx\,dy \pm b \iint_D g(x, y)\,dx\,dy,$$

where a and b are any numbers.

2. *Additivity.* If the domain D is split into two subdomains D_1 and D_2 that do not have common internal points and if the function $f(x, y)$ is integrable in either subdomain, then

$$\iint_D f(x, y)\,dx\,dy = \iint_{D_1} f(x, y)\,dx\,dy + \iint_{D_2} f(x, y)\,dx\,dy.$$

3. *Estimation theorem.* If $m \le f(x, y) \le M$ in D, then

$$mS \le \iint_D f(x, y)\,dx\,dy \le MS,$$

where S is the area of the domain D.

4. *Mean value theorem.* If $f(x, y)$ is continuous in D, then there exists at least one internal point $(\bar{x}, \bar{y}) \in D$ such that

$$\iint_D f(x, y)\,dx\,dy = f(\bar{x}, \bar{y})\,S.$$

The number $f(\bar{x}, \bar{y})$ is called the *mean value of the function* $f(x, y)$ in D.

5. *Integration of inequalities.* If $\varphi(x, y) \le f(x, y) \le g(x, y)$ in D, then

$$\iint_D \varphi(x, y)\,dx\,dy \le \iint_D f(x, y)\,dx\,dy \le \iint_D g(x, y)\,dx\,dy.$$

In particular, if $f(x, y) \ge 0$ in D, then $\iint_D f(x, y)\,dx\,dy \ge 0$.

6. *Absolute value theorem*

$$\left|\iint_D f(x, y)\,dx\,dy\right| \le \iint_D \bigl|f(x, y)\bigr|\,dx\,dy.$$

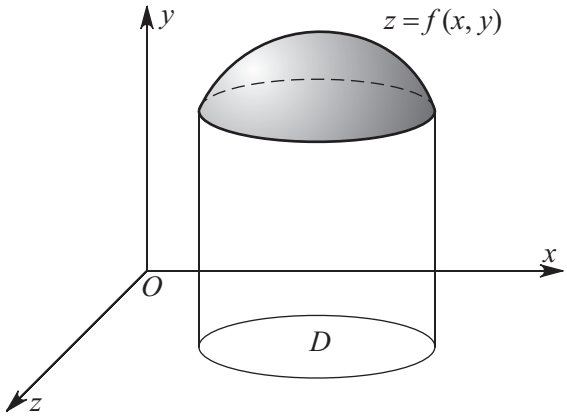


Figure 7.4. A double integral of a nonnegative function $f(x, y)$ over a domain D is equal to the volume of a cylindrical body with base D in the plane $z = 0$ and bounded from above by the surface $z = f(x, y)$.

7.3.1-4. Geometric meaning of the double integral.

Let a function $f(x, y)$ be nonnegative in D. Then the double integral $\iint_D f(x, y)\,dx\,dy$ is equal to the volume of a cylindrical body with base D in the plane $z = 0$ and bounded from above by the surface $z = f(x, y)$; see Fig. 7.4.

7.3.2. Computation of the Double Integral

7.3.2-1. Use of iterated integrals.

1°. If a domain D is defined in the x, y plane by the inequalities $a \le x \le b$ and $y_1(x) \le y \le y_2(x)$ (see Fig. 7.5 *a*), then*

$$\iint_D f(x, y)\,dx\,dy = \int_a^b dx \int_{y_1(x)}^{y_2(x)} f(x, y)\,dy. \tag{7.3.2.1}$$

The integral on the right-hand side is called an *iterated integral*.

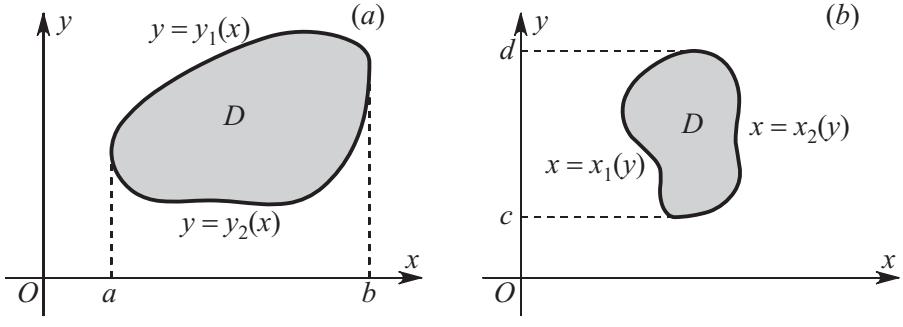


Figure 7.5. Computation of a double integral using iterated integrals: (*a*) illustration to formula (7.3.2.1), (*b*) illustration to formula (7.3.2.2).

* It is assumed that in (7.3.2.1) and (7.3.2.2) the double integral on the right-hand side and the inner integral on the right-hand side exist.

2°. If $D = \{c \le y \le d,\ x_1(y) \le x \le x_2(y)\}$ (see Fig. 7.5 *b*), then

$$\iint_D f(x,y)\,dx\,dy = \int_c^d dy \int_{x_1(y)}^{x_2(y)} f(x,y)\,dx. \tag{7.3.2.2}$$

Example 1. Compute the integral

$$I = \iint_D \frac{dx\,dy}{(ax+by)^2},$$

where $D = \{0 \le x \le 1,\ 1 \le y \le 3\}$ is a rectangle, $a > 0$, and $b > 0$.

Using formula (7.3.2.2), we get

$$\iint_D \frac{dx\,dy}{(ax+by)^2} = \int_1^3 dy \int_0^1 \frac{dx}{(ax+by)^2}.$$

Compute the inner integral:

$$\int_0^1 \frac{dx}{(ax+by)^2} = -\frac{1}{a(ax+by)}\bigg|_{x=0}^{x=1} = \frac{1}{a}\left(\frac{1}{by} - \frac{1}{by+a}\right).$$

It follows that

$$I = \frac{1}{a}\int_1^3 \left(\frac{1}{by} - \frac{1}{by+a}\right) dy = \frac{1}{ab}\ln\frac{3(a+b)}{a+3b}.$$

3°. Consider a domain D inscribed in a rectangle $\{a \le x \le b,\ c \le y \le d\}$. Let the boundary of D, within the rectangle, be intersected by straight lines parallel to the coordinate axes at two points only, as shown in Fig. 7.6 *a*. Then, by comparing formulas (7.3.2.1) and (7.3.2.2), we arrive at the relation

$$\int_a^b dx \int_{y_1(x)}^{y_2(x)} f(x,y)\,dy = \int_c^d dy \int_{x_1(y)}^{x_2(y)} f(x,y)\,dx,$$

which shows how the order of integration can be changed.

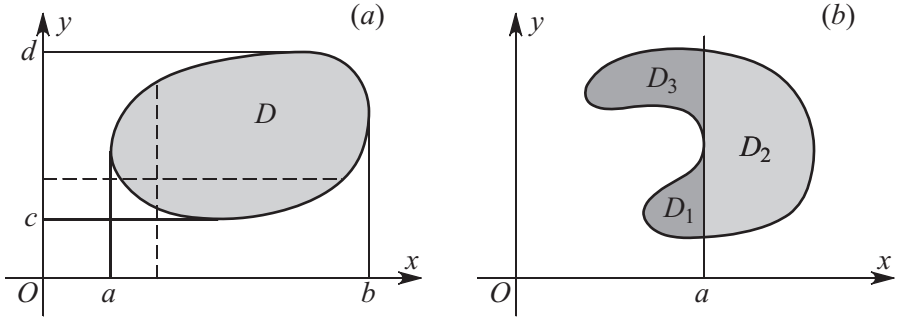


Figure 7.6. Illustrations to the computation of a double integral in a simple (*a*) and a complex (*b*) domain.

4°. In the general case, the domain D is first split into subdomains considered in Items 1° and 2°, and then the property of additivity of the double integral is used. For example, the domain D shown in Fig. 7.6 *b* is divided by the straight line $x = a$ into three subdomains D_1, D_2, and D_3. Then the integral over D is represented as the sum of three integrals over the resulting subdomains.

7.3.2-2. Change of variables in the double integral.

1°. Let $x = x(u, v)$ and $y = y(u, v)$ be continuously differentiable functions that map one-to-one a domain D_1 in the u, v plane onto a domain D in the x, y plane, and let $f(x, y)$ be a continuous function in D. Then

$$\iint_D f(x,y)\,dx\,dy = \iint_{D_1} f\big(x(u,v),\ y(u,v)\big)\,|J(u,v)|\,du\,dv,$$

where $J(u, v)$ is the *Jacobian* (or *Jacobian determinant*) of the mapping of D_1 onto D:

$$J(u,v) = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}.$$

The fraction before the determinant is a common notation for a Jacobian.

The absolute value of the Jacobian characterizes the extension (contraction) of an infinitesimal area element when passing from x, y to u, v.

2°. The Jacobian of the mapping defining the change from the Cartesian coordinates x, y to the polar coordinates ρ, φ,

$$x = \rho\cos\varphi, \qquad y = \rho\sin\varphi, \tag{7.3.2.3}$$

is equal to

$$J(\rho,\varphi) = \rho. \tag{7.3.2.4}$$

Example 2. Given a sphere of radius R and a right circular cylinder of radius $a < R$ whose axis passes through the sphere center, find the volume of the figure the cylinder cuts out of the sphere.

The volume of this figure is calculated as

$$V = \iint_{x^2+y^2\le a^2} \sqrt{R^2 - x^2 - y^2}\ dx\,dy.$$

Passing in the integral from x, y to the polar coordinates (7.3.2.3) and taking into account (7.3.2.4), we obtain

$$V = \int_0^{2\pi}\int_0^a \sqrt{R^2-\rho^2}\,\rho\,d\rho\,d\varphi = \frac{4\pi}{3}\big[R^3 - (R^2-a^2)^{3/2}\big].$$

3°. The Jacobians of some common transformations in the plane are listed in Table 7.1.

TABLE 7.1
Some common curvilinear coordinates in the plane and the respective Jacobians

Name of coordinates	Transformation	Jacobian, J
Polar coordinates ρ, φ	$x = \rho\cos\varphi,\ \ y = \rho\sin\varphi$	ρ
Generalized polar coordinates ρ, φ	$x = a\rho\cos\varphi,\ \ y = b\rho\sin\varphi$	$ab\rho$
Elliptic coordinates u, v (special system; $u \ge 0,\ 0 \le v \le \pi$)	$x = a\cosh u\cos v,\ \ y = a\sinh u\sin v$	$a^2(\sinh^2 u + \sin^2 v)$
Parabolic coordinates σ, τ	$x = a\sigma\tau,\ \ y = \frac{1}{2}a(\tau^2 - \sigma^2)$	$a^2(\sigma^2 + \tau^2)$
Bipolar coordinates σ, τ	$x = \dfrac{a\sinh\tau}{\cosh\tau - \cos\sigma},\ \ y = \dfrac{a\sin\sigma}{\cosh\tau - \cos\sigma}$	$\dfrac{a^2}{(\cosh\tau - \cos\sigma)^2}$

7.3.2-3. Differentiation of the double integral with respect to a parameter.

1°. In various applications, situations may arise where the integrand function and the integration domain depend on a parameter, t. The derivative of such a double integral with respect to t is expressed as

$$\frac{d}{dt}\iint_{D(t)} f(x,y,t)\,dx\,dy = \iint_{D(t)} \frac{\partial}{\partial t} f(x,y,t)\,dx\,dy + \int_{L(t)} (\vec{n}\cdot\vec{v}) f(x,y,t)\,dl, \quad (7.3.2.5)$$

where $L(t)$ is the boundary of the domain $D(t)$, $\vec{n}$ is outer unit normal to $L(t)$, and $\vec{v}$ is the velocity of motion of the points of $L(t)$.

2°. If the boundary $L(t)$ is specified by equations in parametric form

$$x = X(\lambda, t), \qquad y = Y(\lambda, t), \qquad \alpha \le \lambda \le \beta, \quad (7.3.2.6)$$

then

$$\vec{n} = \left\{ \frac{Y_\lambda}{\sqrt{X_\lambda^2 + Y_\lambda^2}}, -\frac{X_\lambda}{\sqrt{X_\lambda^2 + Y_\lambda^2}} \right\}, \quad \vec{v} = \{X_t, Y_t\}, \quad (\vec{n}\cdot\vec{v}) = \frac{Y_\lambda X_t - X_\lambda Y_t}{\sqrt{X_\lambda^2 + Y_\lambda^2}}, \quad dl = \sqrt{X_\lambda^2 + Y_\lambda^2};$$

the subscripts λ and t denote the respective partial derivatives. The last integral in (7.3.2.5) becomes

$$\int_{L(t)} (\vec{n}\cdot\vec{v}) f(x,y,t)\,dl = \int_\alpha^\beta f\big(X(\lambda,t), Y(\lambda,t), t\big)(Y_\lambda X_t - X_\lambda Y_t)\,d\lambda. \quad (7.3.2.7)$$

Example 3. Let $D(t)$ be a deformable plane domain bounded by an ellipse

$$\frac{x^2}{a^2(t)} + \frac{y^2}{b^2(t)} = 1. \quad (7.3.2.8)$$

Let us rewrite the equation of the ellipse (7.3.2.8) in parametric form as

$$x = a(t)\cos\lambda, \quad y = b(t)\sin\lambda, \quad 0 \le \lambda \le 2\pi,$$

which corresponds to the special case of (7.3.2.6) with $X(\lambda, t) = a(t)\cos\lambda$ and $Y(\lambda, t) = b(t)\sin\lambda$. Taking into account the aforesaid and using formula (7.3.2.7), we obtain

$$\int_{L(t)} (\vec{n}\cdot\vec{v}) f(x,y,t)\,dl = \int_0^{2\pi} f\big(a(t)\cos\lambda, b(t)\sin\lambda, t\big)(a_t' b\cos^2\lambda + a b_t' \sin^2\lambda)\,d\lambda. \quad (7.3.2.9)$$

Let us dwell on the simple special case of $f(x,y,t) = 1$, when the first integral on the right-hand side of (7.3.2.5) vanishes. Evaluate the second integral by formula (7.3.2.9) to obtain

$$\int_{L(t)} (\vec{n}\cdot\vec{v})\,dl = \int_0^{2\pi} (a_t' b\cos^2\lambda + a b_t' \sin^2\lambda)\,d\lambda = a_t' b\int_0^{2\pi} \cos^2\lambda\,d\lambda + a b_t' \int_0^{2\pi} \sin^2\lambda\,d\lambda$$
$$= \pi[a'(t)b(t) + a(t)b'(t)]. \quad (7.3.2.10)$$

Remark. Formula (7.3.2.10) is easy to derive directly from (7.3.2.5) noting that, for $f(x,y,t) = 1$, the integral on the left-hand side of (7.3.2.5) gives the area of the ellipse, $S = \pi a(t)b(t)$. Differentiating this formula yields (7.3.2.10).

3°. If the boundary $L(t)$ is specified by an equation in implicit form,

$$F(x,y,t) = 0,$$

then one should take into account in (7.3.2.5) that

$$\vec{n} = \left\{ \frac{F_x}{\sqrt{F_x^2 + F_y^2}}, \frac{F_y}{\sqrt{F_x^2 + F_y^2}} \right\}, \quad \vec{v} = \left\{ -\frac{F_t}{F_x}, -\frac{F_t}{F_y} \right\}, \quad (\vec{n}\cdot\vec{v}) = -\frac{2F_t}{\sqrt{F_x^2 + F_y^2}}.$$

For an elliptic domain, specified by equation (7.3.2.8), we have

$$F(x,y,t) = \frac{x^2}{a^2(t)} + \frac{y^2}{b^2(t)} - 1.$$

7.3.3. Geometric and Physical Applications of the Double Integral

7.3.3-1. Geometric applications of the double integral.

1. *Area of a domain D in the x, y plane*:

$$S = \iint_D dx\, dy.$$

2. *Area of a surface defined by an equation $z = f(x, y)$* with $(x, y) \in D$ (the surface is projected onto a domain D in the x, y plane):

$$S = \iint_D \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + 1}\; dx\, dy.$$

3. *Area of a surface defined parametrically by equations $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$*, with $(u, v) \in D_1$:

$$S = \iint_{D_1} \sqrt{EG - F^2}\, du\, dv.$$

Notation used:

$$E = \left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2 + \left(\frac{\partial z}{\partial u}\right)^2,$$
$$G = \left(\frac{\partial x}{\partial v}\right)^2 + \left(\frac{\partial y}{\partial v}\right)^2 + \left(\frac{\partial z}{\partial v}\right)^2,$$
$$F = \frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + \frac{\partial y}{\partial u}\frac{\partial y}{\partial v} + \frac{\partial z}{\partial u}\frac{\partial z}{\partial v}.$$

4. *Area of a surface defined by a vector equation $\vec{r} = \vec{r}(u, v) = x(u, v)\vec{i} + y(u, v)\vec{j} + z(u, v)\vec{k}$*, with $(u, v) \in D_1$:

$$S = \iint_{D_1} \left|\vec{n}(u, v)\right| du\, dv.$$

Here, the unit normal is calculated as $\vec{n}(u, v) = \vec{r}'_u \times \vec{r}'_v$.

Remark. The formulas from Items 3 and 4 are equivalent — they define one and the same surface in two forms, scalar and vector, respectively.

5. *Calculation of volumes.* If a domain U of the three-dimensional space is defined by $\big\{(x, y) \in D,\ f(x, y) \le z \le g(x, y)\big\}$, where D is a domain in the x, y plane, the volume of U is calculated as

$$V = \iint_D \big[g(x, y) - f(x, y)\big]\, dx\, dy.$$

The three-dimensional domain U is a cylinder with base D bounded by the surface $z = f(x, y)$ from below and the surface $z = g(x, y)$ from above. The lateral surface of this body consists of segments of straight lines parallel to the z-axis.

7.3.3-2. Physical applications of the double integral.

Consider a flat plate that occupies a domain D in the x, y plane. Let $\gamma(x, y)$ be the surface density of the plate material (the case $\gamma = \text{const}$ corresponds to a homogeneous plate).

1. *Mass of a flat plate*:

$$m = \iint_D \gamma(x, y)\, dx\, dy.$$

2. *Coordinates of the center of mass of a flat plate*:

$$x_c = \frac{1}{m} \iint_D x\gamma(x, y)\, dx\, dy, \quad y_c = \frac{1}{m} \iint_D y\gamma(x, y)\, dx\, dy,$$

where m is the mass of the plate.

3. *Moments of inertia of a flat plate* about the coordinate axes:

$$I_x = \iint_D y^2\gamma(x, y)\, dx\, dy, \quad I_y = \iint_D x^2\gamma(x, y)\, dx\, dy.$$

The moment of inertia of the plate about the origin of coordinates is calculated as $I_0 = I_x + I_y$.

7.3.4. Definition and Properties of the Triple Integral

7.3.4-1. Definition of the triple integral.

Let a function $f(x, y, z)$ be defined in a domain U of the three-dimensional space. Let us break up U into n subdomains (cells) that do not have common internal points. Denote by $\lambda = \lambda(\mathcal{U}_n)$ the *diameter* of the resulting partition $\mathcal{U}_n$, i.e., the maximum of the cell diameters (the diameter of a domain in space is the diameter of the minimal sphere enclosing the domain). Select an arbitrary point, (x_i, y_i, z_i), $i = 1, 2, \ldots, n$, in each cell and make up an *integral sum*

$$s_n = \sum_{i=1}^{n} f(x_i, y_i, z_i)\, \Delta V_i,$$

where ΔV_i is the volume of the ith cell. If there exists a finite limit of the sums s_n as $\lambda(\mathcal{U}_n) \to 0$ that depends on neither the partition $\mathcal{U}_n$ nor the selection of the points (x_i, y_i, z_i), then it is called the triple integral of the function $f(x, y, z)$ over the domain U and is denoted

$$\iiint_U f(x, y, z)\, dx\, dy\, dz = \lim_{\lambda \to 0} s_n.$$

7.3.4-2. Properties of the triple integral.

The properties of triple integrals are similar to those of double integrals.

1. *Linearity*. If functions $f(x, y, z)$ and $g(x, y, z)$ are integrable in a domain U, then

$$\iiint_U \big[a f(x, y, z) \pm b g(x, y, z)\big]\, dx\, dy\, dz = a \iiint_U f(x, y, z)\, dx\, dy\, dz \pm b \iiint_U g(x, y, z)\, dx\, dy\, dz,$$

where a and b are any numbers.

2. *Additivity*. If a domain U is split into two subdomains, U_1 and U_2, that do not have common internal points and if a function $f(x, y, z)$ is integrable in either subdomain, then

$$\iiint_U f(x,y,z)\,dx\,dy\,dz = \iiint_{U_1} f(x,y,z)\,dx\,dy\,dz + \iiint_{U_2} f(x,y,z)\,dx\,dy\,dz.$$

3. *Estimation theorem*. If $m \le f(x, y, z) \le M$ in a domain U, then

$$mV \le \iiint_U f(x,y,z)\,dx\,dy\,dz \le MV,$$

where V is the volume of U.

4. *Mean value theorem*. If $f(x, y, z)$ is continuous in U, then there exists at least one internal point $(\bar{x}, \bar{y}, \bar{z}) \in U$ such that

$$\iiint_U f(x,y,z)\,dx\,dy\,dz = f(\bar{x},\bar{y},\bar{z})\,V.$$

The number $f(\bar{x}, \bar{y}, \bar{z})$ is called the *mean value of the function* f in the domain U.

5. *Integration of inequalities*. If $\varphi(x, y, z) \le f(x, y, z) \le g(x, y, z)$ in a domain U, then

$$\iiint_U \varphi(x,y,z)\,dx\,dy\,dz \le \iiint_U f(x,y,z)\,dx\,dy\,dz \le \iiint_U g(x,y,z)\,dx\,dy\,dz.$$

6. *Absolute value theorem*:

$$\left|\iiint_U f(x,y,z)\,dx\,dy\,dz\right| \le \iiint_U \bigl|f(x,y,z)\bigr|\,dx\,dy\,dz.$$

7.3.5. Computation of the Triple Integral. Some Applications. Iterated Integrals and Asymptotic Formulas

7.3.5-1. Use of iterated integrals.

1°. Consider a three-dimensional body U bounded by a surface $z = g(x, y)$ from above and a surface $z = h(x, y)$ from below, with a domain D being the projection of it onto the x, y plane. In other words, the domain U is defined as $\{(x, y) \in D : h(x, y) \le z \le g(x, y)\}$. Then

$$\iiint_U f(x,y,z)\,dx\,dy\,dz = \iint_D dx\,dy \int_{h(x,y)}^{g(x,y)} f(x,y,z)\,dz.$$

2°. If, under the same conditions as in Item 1°, the domain D of the x, y plane is defined as $\{a \le x \le b,\ y_1(x) \le y \le y_2(x)\}$, then

$$\iiint_U f(x,y,z)\,dx\,dy\,dz = \int_a^b dx \int_{y_1(x)}^{y_2(x)} dy \int_{h(x,y)}^{g(x,y)} f(x,y,z)\,dz.$$

7.3.5-2. Change of variables in the triple integral.

1°. Let $x = x(u,v,w)$, $y = y(u,v,w)$, and $z = z(u,v,w)$ be continuously differentiable functions that map, one to one, a domain Ω of the u,v,w space into a domain U of the x,y,z space, and let a function $f(x,y,z)$ be continuous in U. Then

$$\iiint_U f(x,y,z)\,dx\,dy\,dz = \iiint_\Omega f\big(x(u,v,w),\ y(u,v,w),\ z(u,v,w)\big)\,|J(u,v,w)|\,du\,dv\,dw,$$

where $J(u,v,w)$ is the *Jacobian* of the mapping of Ω into U:

$$J(u,v,w) = \frac{\partial(x,y,z)}{\partial(u,v,w)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix}.$$

The expression in the middle is a very common notation for a Jacobian.

The absolute value of the Jacobian characterizes the expansion (or contraction) of an infinitesimal volume element when passing from x, y, z to u, v, w.

2°. The Jacobians of most common transformations in space are listed in Table 7.2.

TABLE 7.2
Common curvilinear coordinates in space and the respective Jacobians

Name of coordinates	Transformation	Jacobian, J
Cylindrical coordinates ρ,φ,z	$x = \rho\cos\varphi,\ \ y = \rho\sin\varphi,\ \ z = z$	ρ
Generalized cylindrical coordinates ρ,φ,z	$x = a\rho\cos\varphi,\ \ y = b\rho\sin\varphi,\ \ z = z$	$ab\rho$
Spherical coordinates r,φ,θ	$x = r\cos\varphi\,\sin\theta,\ y = r\sin\varphi\,\sin\theta,\ z = r\cos\theta$	$r^2\sin\theta$
Generalized spherical coordinates r,φ,θ	$x = ar\cos\varphi\,\sin\theta,\ y = br\sin\varphi\,\sin\theta,$ $z = cr\cos\theta$	$abcr^2\sin\theta$
Coordinates of prolate ellipsoid of revolution σ,τ,φ $(\sigma \ge 1 \ge \tau \ge -1)$	$x = a\sqrt{(\sigma^2-1)(1-\tau^2)}\,\cos\varphi,$ $y = a\sqrt{(\sigma^2-1)(1-\tau^2)}\,\sin\varphi,$ $z = a\sigma\tau$	$a^3(\sigma^2-\tau^2)$
Coordinates of oblate ellipsoid of revolution σ,τ,φ $(\sigma \ge 0,\ -1 \le \tau \le 1)$	$x = a\sqrt{(1+\sigma^2)(1-\tau^2)}\,\cos\varphi,$ $y = a\sqrt{(1+\sigma^2)(1-\tau^2)}\,\sin\varphi,$ $z = a\sigma\tau$	$a^3(\sigma^2+\tau^2)$
Parabolic coordinates σ,τ,φ	$x = \sigma\tau\cos\varphi,\ y = \sigma\tau\sin\varphi,\ z = \frac{1}{2}(\tau^2-\sigma^2)$	$\sigma\tau(\sigma^2+\tau^2)$
Parabolic cylinder coordinates σ,τ,z	$x = \sigma\tau,\ \ y = \frac{1}{2}(\tau^2-\sigma^2),\ \ z = z$	$\sigma^2+\tau^2$
Bicylindrical coordinates σ,τ,z	$x = \dfrac{a\sinh\tau}{\cosh\tau-\cos\sigma},\ y = \dfrac{a\sin\sigma}{\cosh\tau-\cos\sigma},\ \ z = z$	$\dfrac{a^2}{(\cosh\tau-\cos\sigma)^2}$
Toroidal coordinates σ,τ,φ $(-\pi \le \sigma \le \pi,\ 0 \le \tau < \infty,\ 0 \le \varphi < 2\pi)$	$x = \dfrac{a\sinh\tau\cos\varphi}{\cosh\tau-\cos\sigma},\ y = \dfrac{a\sinh\tau\sin\varphi}{\cosh\tau-\cos\sigma},$ $z = \dfrac{a\sin\sigma}{\cosh\tau-\cos\sigma}$	$\dfrac{a^3\sinh\tau}{(\cosh\tau-\cos\sigma)^2}$
Bipolar coordinates σ,τ,φ (σ is any, $0 \le \tau < \pi,\ 0 \le \varphi < 2\pi$)	$x = \dfrac{a\sin\tau\cos\varphi}{\cosh\sigma-\cos\tau},\ y = \dfrac{a\sin\tau\sin\varphi}{\cosh\sigma-\cos\tau},$ $z = \dfrac{a\sinh\sigma}{\cosh\sigma-\cos\tau}$	$\dfrac{a^3\sin\tau}{(\cosh\sigma-\cos\tau)^2}$

7.3.5-3. Differentiation of the triple integral with respect to a parameter.

Let the integrand function and the integration domain of a triple integral depend on a parameter, t. The derivative of this integral with respect to t is expressed as

$$\frac{d}{dt}\iiint_{U(t)} f(x,y,z,t)\,dx\,dy\,dz = \iiint_{U(t)} \frac{\partial}{\partial t} f(x,y,z,t)\,dx\,dy\,dz + \iint_{S(t)} (\vec{n}\cdot\vec{v}) f(x,y,z,t)\,ds,$$

where $S(t)$ is the boundary of the domain $U(t)$, $\vec{n}$ is the unit normal to $S(t)$, and $\vec{v}$ is the velocity of motion of the points of $S(t)$.

7.3.5-4. Some geometric and physical applications of the triple integral.

1. *Volume of a domain U*:

$$V = \iiint_U dx\,dy\,dz.$$

2. *Mass of a body of variable density $\gamma = \gamma(x,y,z)$* occupying a domain U:

$$m = \iiint_U \gamma\,dx\,dy\,dz.$$

3. *Coordinates of the center of mass*:

$$x_c = \frac{1}{m}\iiint_U x\gamma\,dx\,dy\,dz, \quad y_c = \frac{1}{m}\iiint_U y\gamma\,dx\,dy\,dz, \quad z_c = \frac{1}{m}\iiint_U z\gamma\,dx\,dy\,dz.$$

4. *Moments of inertia* about the coordinate axes:

$$I_x = \iiint_U \rho_{yz}^2\gamma\,dx\,dy\,dz, \quad I_y = \iiint_U \rho_{xz}^2\gamma\,dx\,dy\,dz, \quad I_z = \iiint_U \rho_{xy}^2\gamma\,dx\,dy\,dz,$$

where $\rho_{yz}^2 = y^2 + z^2$, $\rho_{xz}^2 = x^2 + z^2$, and $\rho_{xy}^2 = x^2 + y^2$.

If the body is homogeneous, then $\gamma = \text{const}$.

Example. Given a bounded homogeneous elliptic cylinder,

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad 0 \le z \le h,$$

find its moment of inertia about the z-axis.

Using the generalized cylindrical coordinates (see the second row in Table 7.2), we obtain

$$I_x = \gamma\iiint_U (x^2+y^2)\,dx\,dy\,dz = \gamma\int_0^h\int_0^{2\pi}\int_0^1 \rho^2(a^2\cos^2\varphi + b^2\sin^2\varphi)ab\rho\,d\rho\,d\varphi\,dz$$
$$= \frac{1}{4}ab\gamma\int_0^h\int_0^{2\pi}(a^2\cos^2\varphi + b^2\sin^2\varphi)\,d\varphi\,dz = \frac{1}{4}ab\gamma\int_0^{2\pi}\int_0^h (a^2\cos^2\varphi + b^2\sin^2\varphi)\,dz\,d\varphi$$
$$= \frac{1}{4}abh\gamma\int_0^{2\pi}(a^2\cos^2\varphi + b^2\sin^2\varphi)\,d\varphi = \frac{1}{4}ab(a^2+b^2)h\gamma.$$

5. *Potential of the gravitational field* of a body U at a point (x, y, z):

$$\Phi(x,y,z) = \iiint_U \gamma(\xi,\eta,\zeta)\frac{d\xi\,d\eta\,d\zeta}{r}, \qquad r = \sqrt{(x-\xi)^2+(y-\eta)^2+(z-\zeta)^2},$$

where $\gamma = \gamma(\xi,\eta,\zeta)$ is the body density. A material point of mass m is pulled by the gravitating body U with a force $\vec{F}$. The projections of $\vec{F}$ onto the x-, y-, and z-axes are given by

$$F_x = km\frac{\partial\Phi}{\partial x} = km\iiint_U \gamma(\xi,\eta,\zeta)\frac{\xi-x}{r^3}\,d\xi\,d\eta\,d\zeta,$$
$$F_y = km\frac{\partial\Phi}{\partial y} = km\iiint_U \gamma(\xi,\eta,\zeta)\frac{\eta-y}{r^3}\,d\xi\,d\eta\,d\zeta,$$
$$F_z = km\frac{\partial\Phi}{\partial z} = km\iiint_U \gamma(\xi,\eta,\zeta)\frac{\zeta-z}{r^3}\,d\xi\,d\eta\,d\zeta,$$

where k is the gravitational constant.

7.3.5-5. Multiple integrals. Asymptotic formulas.

Multiple integrals in n variables of integration are an obvious generalization of double and triple integrals.

1°. Consider the *Laplace-type multiple integral*

$$I(\lambda) = \int_\Omega f(\mathbf{x})\exp[\lambda g(\mathbf{x})]\,d\mathbf{x},$$

where $\mathbf{x} = \{x_1,\dots,x_n\}$, $d\mathbf{x} = dx_1\,\dots\,dx_n$, Ω is a bounded domain in $\mathbb{R}^n$, $f(\mathbf{x})$ and $g(\mathbf{x})$ are real-valued functions of n variable, and λ is a real or complex parameter.

Denote by

$$S_\varepsilon = \Big\{\lambda : \arg|\lambda| \le \frac{\pi}{2} - \varepsilon\Big\}, \qquad 0 < \varepsilon < \frac{\pi}{2},$$

a sector in the complex plane of λ.

THEOREM 1. *Let the following conditions hold:*

(1) the functions $f(\mathbf{x})$ and $g(\mathbf{x})$ are continuous in Ω,
(2) the maximum of $g(\mathbf{x})$ is attained at only one point $\mathbf{x}_0 \in \Omega$ ($\mathbf{x}_0$ is a nondegenerate maximum point), and
(3) the function $g(\mathbf{x})$ has continuous third derivatives in a neighborhood of $\mathbf{x}_0$.

Then the following asymptotic formula holds as $\lambda \to \infty$, $\lambda \in S_\varepsilon$:

$$I(\lambda) = (2\pi)^{n/2}\exp[\lambda g(\mathbf{x}_0)]\frac{f(\mathbf{x}_0)+O(\lambda^{-1})}{\sqrt{\lambda^n \det[g_{x_ix_j}(\mathbf{x}_0)]}},$$

where the $g_{x_ix_j}(\mathbf{x})$ are entries of the matrix of the second derivatives of $g(\mathbf{x})$.

2°. Consider the *power Laplace multiple integral*

$$I(\lambda) = \int_\Omega f(\mathbf{x})[g(\mathbf{x})]^\lambda\,d\mathbf{x}.$$

THEOREM 2. *Let $g(\mathbf{x}) > 0$ and let the conditions of Theorem 1 hold. Then the following asymptotic formula holds as $\lambda \to \infty$, $\lambda \in S_\varepsilon$:*

$$I(\lambda) = (2\pi)^{n/2}[g(\mathbf{x}_0)]^{(2\lambda+n)/2}\frac{f(\mathbf{x}_0)+o(1)}{\sqrt{\lambda^n \det[g_{x_ix_j}(\mathbf{x}_0)]}}.$$

7.4. Line and Surface Integrals

7.4.1. Line Integral of the First Kind

7.4.1-1. Definition of the line integral of the first kind.

Let a function $f(x, y, z)$ be defined on a piecewise smooth curve $\overset{\smile}{AB}$ in the three-dimensional space $\mathbb{R}^3$. Let the curve $\overset{\smile}{AB}$ be divided into n subcurves by points $A = M_0, M_1, M_2, \ldots, M_n = B$, thus defining a partition $\mathcal{L}_n$. The longest of the chords $M_0M_1, M_1M_2, \ldots, M_{n-1}M_n$ is called the *diameter of the partition* $\mathcal{L}_n$ and is denoted $\lambda = \lambda(\mathcal{L}_n)$. Let us select on each arc $\overset{\smile}{M_{i-1}M_i}$ an arbitrary point (x_i, y_i, z_i), $i = 1, 2, \ldots, n$, and make up an *integral sum*

$$s_n = \sum_{i=1}^{n} f(x_i, y_i, z_i)\,\Delta l_i,$$

where Δl_i is the length of $\overset{\smile}{M_{i-1}M_i}$.

If there exists a finite limit of the sums s_n as $\lambda(\mathcal{L}_n) \to 0$ that depends on neither the partition $\mathcal{L}_n$ nor the selection of the points (x_i, y_i, z_i), then it is called the *line integral of the first kind* of the function $f(x, y, z)$ over the curve $\overset{\smile}{AB}$ and is denoted

$$\int_{AB} f(x, y, z)\,dl = \lim_{\lambda \to 0} s_n.$$

A line integral is also called a *curvilinear integral* or a *path integral*.

If the function $f(x, y, z)$ is continuous, then the line integral exists. The line integral of the first kind does not depend of the direction the path $\overset{\smile}{AB}$ is traced; its properties are similar to those of the definite integral.

7.4.1-2. Computation of the line integral of the first kind.

1. If a plane curve is defined in the form $y = y(x)$, with $x \in [a, b]$, then

$$\int_{AB} f(x, y)\,dl = \int_a^b f\big(x, y(x)\big)\sqrt{1 + (y'_x)^2}\,dx.$$

2. If a curve $\overset{\smile}{AB}$ is defined in parametric form by equations $x = x(t)$, $y = y(t)$, and $z = z(t)$, with $t \in [\alpha, \beta]$, then

$$\int_{AB} f(x, y, z)\,dl = \int_\alpha^\beta f\big(x(t), y(t), z(t)\big)\sqrt{(x'_t)^2 + (y'_t)^2 + (z'_t)^2}\,dt. \qquad (7.4.1.1)$$

If a function $f(x, y)$ is defined on a plane curve $x = x(t)$, $y = y(t)$, with $t \in [\alpha, \beta]$, one should set $z'_t = 0$ in (7.4.1.1).

Example. Evaluate the integral $\int_{AB} xy\,dl$, where $\overset{\smile}{AB}$ is a quarter of an ellipse with semiaxes a and b.

Let us write out the equations of the ellipse for the first quadrant in parametric form:

$$x = a\cos t, \quad y = b\sin t \qquad (0 \le t \le \pi/2).$$

We have $\sqrt{(x_t')^2 + (y_t')^2} = \sqrt{a^2 \sin^2 t + b^2 \cos^2 t}$. To evaluate the integral, we use formula (7.4.1.1) with $z_t' = 0$:

$$\int_{AB} xy\, dl = \int_0^{\pi/2} (a\cos t)\,(b \sin t)\sqrt{a^2 \sin^2 t + b^2 \cos^2 t}\, dt$$
$$= \frac{ab}{2}\int_0^{\pi/2} \sin 2t \sqrt{\frac{a^2}{2}(1 - \cos 2t) + \frac{b^2}{2}(1 + \cos 2t)}\, dt = \frac{ab}{4}\int_{-1}^{1} \sqrt{\frac{a^2 + b^2}{2} + \frac{b^2 - a^2}{2} z}\, dz$$
$$= \frac{ab}{4}\,\frac{2}{b^2 - a^2}\,\frac{2}{3}\left(\frac{a^2 + b^2}{2} + \frac{b^2 - a^2}{2} z\right)^{3/2}\Bigg|_{-1}^{1} = \frac{ab}{3}\,\frac{a^2 + ab + b^2}{a + b}.$$

7.4.1-3. Applications of the line integral of the first kind.

1. *Length of a curve* $\overset{\smile}{AB}$:
$$L = \int_{AB} dl.$$

2. *Mass of a material curve* $\overset{\smile}{AB}$ with a given line density $\gamma = \gamma(x, y, z)$:
$$m = \int_{AB} \gamma\, dl.$$

3. *Coordinates of the center of mass* of a material curve $\overset{\smile}{AB}$:
$$x_c = \frac{1}{m}\int_{AB} x\gamma\, dl, \quad y_c = \frac{1}{m}\int_{AB} y\gamma\, dl, \quad z_c = \frac{1}{m}\int_{AB} z\gamma\, dl.$$

To a material line with uniform density there corresponds $\gamma = \text{const}$.

7.4.2. Line Integral of the Second Kind

7.4.2-1. Definition of the line integral of the second kind.

Let a vector field

$$\vec{a}(x, y, z) = P(x, y, z)\vec{i} + Q(x, y, z)\vec{j} + R(x, y, z)\vec{k}$$

and a piecewise smooth curve $\overset{\smile}{AB}$ be defined in some domain in $\mathbb{R}^3$. By dividing the curve by points $A = M_0, M_1, M_2, \ldots, M_n = B$ into n subcurves, we obtain a partition $\mathcal{L}_n$. Let us select on each arc $\overset{\smile}{M_{i-1}M_i}$ an arbitrary point (x_i, y_i, z_i), $i = 1, 2, \ldots, n$, and make up a sum of dot products

$$s_n = \sum_{i=1}^{n} \vec{a}(x_i, y_i, z_i) \cdot \overrightarrow{M_{i-1}M_i},$$

called an *integral sum*.

If there exists a finite limit of the sums s_n as $\lambda(\mathcal{L}_n) \to 0$ that depends on neither the partition $\mathcal{L}_n$ nor the selection of the points (x_i, y_i, z_i), then it is called the *line integral of the second kind* of the vector field $\vec{a}(x, y, z)$ along the curve $\overset{\smile}{AB}$ and is denoted

$$\int_{AB} \vec{a} \cdot d\vec{r}, \quad \text{or} \quad \int_{AB} P\, dx + Q\, dy + R\, dz.$$

The line integral of the second kind depends on the direction the path is traced, so that

$$\int_{AB} \vec{a} \cdot d\vec{r} = -\int_{BA} \vec{a} \cdot d\vec{r}.$$

A line integral over a closed contour $\mathcal{C}$ is called a *closed path integral* (or a *circulation*) of a vector field $\vec{a}$ around $\mathcal{C}$ and is denoted

$$\oint_{\mathcal{C}} \vec{a} \cdot d\vec{r}.$$

Physical meaning of the line integral of the second kind: $\int_{AB} \vec{a} \cdot d\vec{r}$ determines the work done by the vector field $\vec{a}(x, y, z)$ on a particle of unit mass when it travels along the arc $\overset{\smile}{AB}$.

7.4.2-2. Computation of the line integral of the second kind.

1°. For a plane curve $\overset{\smile}{AB}$ defined as $y = y(x)$, with $x \in [a, b]$, and a plane vector field $\vec{a}$, we have

$$\int_{AB} \vec{a} \cdot d\vec{r} = \int_a^b \big[P\big(x, y(x)\big) + Q\big(x, y(x)\big) y'_x(x)\big]\, dx.$$

2°. Let $\overset{\smile}{AB}$ be defined by a vector equation $\vec{r} = \vec{r}(t) = x(t)\vec{i} + y(t)\vec{j} + z(t)\vec{k}$, with $t \in [\alpha, \beta]$. Then

$$\int_{AB} \vec{a} \cdot d\vec{r} = \int_{AB} P\, dx + Q\, dy + R\, dz$$
$$= \int_\alpha^\beta \big[P\big(x(t), y(t), z(t)\big) x'_t(t) + Q\big(x(t), y(t), z(t)\big) y'_t(t) + R\big(x(t), y(t), z(t)\big) z'_t(t)\big]\, dt. \quad (7.4.2.1)$$

For a plane curve $\overset{\smile}{AB}$ and a plane vector field $\vec{a}$, one should set $z'(t) = 0$ in (7.4.2.1).

7.4.2-3. Potential and curl of a vector field.

1°. A vector field $\vec{a} = \vec{a}(x, y, z)$ is called *potential* if there exists a function $\Phi(x, y, z)$ such that

$$\vec{a} = \operatorname{grad} \Phi, \quad \text{or} \quad \vec{a} = \frac{\partial \Phi}{\partial x}\vec{i} + \frac{\partial \Phi}{\partial y}\vec{j} + \frac{\partial \Phi}{\partial z}\vec{k}.$$

The function $\Phi(x, y, z)$ is called a *potential* of the vector field $\vec{a}$. The line integral of the second kind of a potential vector field along a path $\overset{\smile}{AB}$ is equal to the increment of the potential along the path:

$$\int_{AB} \vec{a} \cdot d\vec{r} = \Phi\big|_B - \Phi\big|_A.$$

2°. The *curl* of a vector field $\vec{a}(x, y, z) = P\vec{i} + Q\vec{j} + R\vec{k}$ is the vector defined as

$$\operatorname{curl} \vec{a} = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\vec{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)\vec{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\vec{k} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix}.$$

The vector $\operatorname{curl} \vec{a}$ characterizes the rate of rotation of $\vec{a}$ and can also be described as the circulation density of $\vec{a}$. Alternative notations: $\operatorname{curl} \vec{a} \equiv \nabla \times \vec{a} \equiv \operatorname{rot} \vec{a}$.

7.4.2-4. Necessary and sufficient conditions for a vector field to be potential.

Let U be a simply connected domain in $\mathbb{R}^3$ (i.e., a domain in which any closed contour can be deformed to a point without leaving U) and let $\vec{a}(x, y, z)$ be a vector field in U. Then the following four assertions are equivalent to each other:

(1) the vector field $\vec{a}$ is potential;

(2) $\operatorname{curl} \vec{a} \equiv \vec{0}$;

(3) the circulation of $\vec{a}$ around any closed contour $\mathcal{C} \in U$ is zero, or, equivalently, $\oint_{\mathcal{C}} \vec{a} \cdot d\vec{r} = 0$;

(4) the integral $\int_{AB} \vec{a} \cdot d\vec{r}$ is independent of the shape of $\overset{\smile}{AB} \in U$ (it depends on the starting and the finishing point only).

7.4.3. Surface Integral of the First Kind

7.4.3-1. Definition of the surface integral of the first kind.

Let a function $f(x, y, z)$ be defined on a smooth surface D. Let us break up this surface into n elements (cells) that do not have common internal points and let us denote this partition by $\mathcal{D}_n$. The *diameter*, $\lambda(\mathcal{D}_n)$, of a partition $\mathcal{D}_n$ is the largest of the diameters of the cells (see Paragraph 7.3.4-1). Let us select in each cell an arbitrary point (x_i, y_i, z_i), $i = 1, 2, \ldots, n$, and make up an *integral sum*

$$s_n = \sum_{i=1}^{n} f(x_i, y_i, z_i)\, \Delta S_i,$$

where ΔS_i is the area of the ith element.

If there exists a finite limit of the sums s_n as $\lambda(\mathcal{D}_n) \to 0$ that depends on neither the partition $\mathcal{D}_n$ nor the selection of the points (x_i, y_i, z_i), then it is called the *surface integral of the first kind* of the function $f(x, y, z)$ and is denoted $\iint_D f(x, y, z)\, dS$.

7.4.3-2. Computation of the surface integral of the first kind.

1°. If a surface D is defined by an equation $z = z(x, y)$, with $(x, y) \in D_1$, then

$$\iint_D f(x, y, z)\, dS = \iint_{D_1} f\big(x, y, z(x, y)\big) \sqrt{1 + (z'_x)^2 + (z'_y)^2}\, dx\, dy.$$

2°. If a surface D is defined by a vector equation $\vec{r} = \vec{r}(x, y, z) = x(u, v)\, \vec{i} + y(u, v)\, \vec{j} + z(u, v)\, \vec{k}$, where $(u, v) \in D_2$, then

$$\iint_D f(x, y, z)\, dS = \iint_{D_2} f\big(x(u, v), y(u, v), z(u, v)\big)\, |\vec{n}(u, v)|\, du\, dv,$$

where $\vec{n}(u, v) = \vec{r}_u \times \vec{r}_v$ is the unit normal to the surface D; the subscripts u and v denote the respective partial derivatives.

7.4.3-3. Applications of the surface integral of the first kind.

1°. Area of a surface D:

$$S_D = \iint_D dS.$$

2°. Mass of a material surface D with a surface density $\gamma = \gamma(x, y, z)$:

$$m = \iint_D \gamma(x, y, z)\, dS.$$

3°. Coordinates of the center of mass of a material surface D:

$$x_c = \frac{1}{m} \iint_D x\gamma\, dS, \quad y_c = \frac{1}{m} \iint_D y\gamma\, dS, \quad z_c = \frac{1}{m} \iint_D z\gamma\, dS.$$

To the uniform surface density there corresponds $\gamma = \text{const}$.

7.4.4. Surface Integral of the Second Kind

7.4.4-1. Definition of the surface integral of the second kind.

Let D be an *oriented surface* defined by an equation

$$\vec{r} = \vec{r}(u, v) = x(u, v)\,\vec{i} + y(u, v)\,\vec{j} + z(u, v)\,\vec{k},$$

where u and v are parameters. The fact that D is oriented means that every point $M \in D$ has an associated unit normal $\vec{n}(M) = \vec{n}(u, v)$ continuously dependent on M. Two cases are possible: (i) the associated unit normal is $\vec{n}(u, v) = \vec{r}_u \times \vec{r}_v$ or (ii) the associated unit normal is opposite, $\vec{n}(u, v) = \vec{r}_v \times \vec{r}_u = -\vec{r}_u \times \vec{r}_v$.

Remark. If a surface is defined traditionally by an equation $z = z(x, y)$, its representation in vector form is as follows: $\vec{r} = \vec{r}(x, y) = x\,\vec{i} + y\,\vec{j} + z(x, y)\,\vec{k}$.

Let a vector field $\vec{a}(x, y, z) = P\vec{i} + Q\vec{j} + R\vec{k}$ be defined on a smooth oriented surface D. Let us perform a partition, $\mathcal{D}_n$, of the surface D into n elements (cells) that do not have common internal points. Also select an arbitrary point $M_i(x_i, y_i, z_i)$, $i = 1, 2, \ldots, n$, for each cell and make up an *integral sum* $s_n = \sum\limits_{i=1}^{n} \vec{a}(x_i, y_i, z_i) \cdot \vec{n}_i^{\circ}\, \Delta S_i$, where ΔS_i is area of the ith cell and $\vec{n}_i^{\circ}$ is the unit normal to the surface at the point M_i, the orientation of which coincides with that of the surface.

If there exists a finite limit of the sums s_n as $\lambda(\mathcal{D}_n) \to 0$ that depends on neither the partition $\mathcal{D}_n$ nor the selection of the points $M_i(x_i, y_i, z_i)$, then it is called the *surface integral of the second kind* (or the *flux of the vector field $\vec{a}$ across the oriented surface D*) and is denoted

$$\iint_D \vec{a}(x, y, z) \cdot \overrightarrow{dS}, \quad \text{or} \quad \iint_D P\, dy\, dz + Q\, dx\, dz + R\, dx\, dy.$$

Note that the surface integral of the second kind changes its sign when the orientation of the surface is reversed.

7.4.4-2. Computation of the surface integral of the second kind.

1°. If a surface D is defined by a vector equation $\vec{r} = \vec{r}(u, v)$, where $(u, v) \in D_1$, then

$$\iint_D \vec{a}(x, y, z) \cdot \overrightarrow{dS} = \pm \iint_{D_1} \vec{a}\big(x(u, v), y(u, v), z(u, v)\big) \cdot \vec{n}(u, v)\, du\, dv.$$

The plus sign is taken if the unit normal associated with the surface is $\vec{n}(u, v) = \vec{r}_u \times \vec{r}_v$, and the minus sign is taken in the opposite case.

2°. If a surface D is defined by an equation $z = z(x, y)$, with $(x, y) \in D_2$, then the normal $\vec{n}(x, y) = \vec{r}_x \times \vec{r}_y = -z_x\vec{i} - z_y\vec{j} + \vec{k}$ orients the surface D "upward," in the positive direction of the z-axis; the subscripts x and y denote the respective partial derivatives. Then

$$\iint_D \vec{a} \cdot \overrightarrow{dS} = \pm \iint_{D_2} \big(-z_x P - z_y Q + R\big)\, dx\, dy,$$

where $P = P\big(x, y, z(x, y)\big)$, $Q = Q\big(x, y, z(x, y)\big)$, and $R = R\big(x, y, z(x, y)\big)$. The plus sign is taken if the surface has the "upward" orientation, and the minus sign is chosen in the opposite case.

7.4.5. Integral Formulas of Vector Calculus

7.4.5-1. Ostrogradsky–Gauss theorem (divergence theorem).

Let a vector field $\vec{a}(x, y, z) = P(x, y, z)\,\vec{i} + Q(x, y, z)\,\vec{j} + R(x, y, z)\,\vec{k}$ be continuously differentiable in a finite simply connected domain $V \subset \mathbb{R}^3$ oriented by the outward normal and let S denote the boundary of V. Then the *Ostrogradsky–Gauss theorem* (or the *divergence theorem*) holds:

$$\iint_S \vec{a} \cdot \overrightarrow{dS} = \iiint_V \operatorname{div} \vec{a}\; dx\, dy\, dz,$$

where $\operatorname{div} \vec{a}$ is the *divergence* of the vector $\vec{a}$, which is defined as follows:

$$\operatorname{div} \vec{a} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}.$$

Thus, the flux of a vector field across a closed surface in the outward direction is equal to the triple integral of the divergence of the vector field over the volume bounded by the surface. In coordinate form, the Ostrogradsky–Gauss theorem reads

$$\iint_S P\, dy\, dz + Q\, dx\, dz + R\, dx\, dy = \iiint_V \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) dx\, dy\, dz.$$

7.4.5-2. Stokes's theorem (curl theorem).

1°. Let a vector field $\vec{a}(x, y, z)$ be continuously differentiable in a domain of the three-dimensional space $\mathbb{R}^3$ that contains an oriented surface D. The orientation of a surface uniquely defines the direction in which the boundary of the surface is traced; specifically, the boundary is traced counterclockwise when looked at from the direction of the normal to

the surface. Then the circulation of the vector field around the boundary $\mathcal{C}$ of the surface D is equal to the flux of the vector curl $\vec{a}$ across D:

$$\oint_{\mathcal{C}} \vec{a} \cdot d\vec{r} = \iint_D \operatorname{curl} \vec{a} \cdot \overrightarrow{dS}.$$

In coordinate notation, *Stokes's theorem* reads

$$\oint_{\mathcal{C}} P\,dx+Q\,dy+R\,dz = \iint_D \left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}\right) dy\,dz+\left(\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}\right) dx\,dz+\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right) dx\,dy.$$

2°. For a plane vector field $\vec{a}(x,y) = P(x,y)\,\vec{i} + Q(x,y)\,\vec{j}$, Stokes's theorem reduces to *Green's theorem*:

$$\oint_{\mathcal{C}} P\,dx + Q\,dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\,dy,$$

where the contour $\mathcal{C}$ of the domain D on the x, y plane is traced counterclockwise.

7.4.5-3. Green's first and second identities. Gauss's theorem.

1°. Let $\Phi = \Phi(x,y,z)$ and $\Psi = \Psi(x,y,z)$ be twice continuously differentiable functions defined in a finite simply connected domain $V \subset \mathbb{R}^3$ bounded by a piecewise smooth boundary S.

Then the following formulas hold:

$$\iiint_V \Psi\Delta\Phi\,dV + \iiint_V \nabla\Phi\cdot\nabla\Psi\,dV = \iint_S \Psi\frac{\partial\Phi}{\partial n}\,dS \quad (\textit{Green's first identity}),$$
$$\iiint_V (\Psi\Delta\Phi - \Phi\Delta\Psi)\,dV = \iint_S \left(\Psi\frac{\partial\Phi}{\partial n} - \Phi\frac{\partial\Psi}{\partial n}\right) dS \quad (\textit{Green's second identity}),$$

where $\frac{\partial}{\partial n}$ denotes a derivative along the (outward) normal to the surface S, and Δ is the Laplace operator.

2°. In applications, the following special cases of the above formulas are most common:

$$\iiint_V \Phi\Delta\Phi\,dV + \iiint_V |\nabla\Phi|^2\,dV = \iint_S \Phi\frac{\partial\Phi}{\partial n}\,dS \quad \text{(first identity with } \Psi = \Phi),$$
$$\iiint_V \Delta\Phi\,dV = \iint_S \frac{\partial\Phi}{\partial n}\,dS \quad \text{(second identity with } \Psi = 1).$$

References for Chapter 7

Adams, R., *Calculus: A Complete Course, 6th Edition*, Pearson Education, Toronto, 2006.

Anton, H., *Calculus: A New Horizon, 6th Edition*, Wiley, New York, 1999.

Anton, H., Bivens, I., and Davis, S., *Calculus: Early Transcendental Single Variable, 8th Edition*, John Wiley & Sons, New York, 2005.

Aramanovich, I. G., Guter, R. S., et al., *Mathematical Analysis (Differentiation and Integration)*, Fizmatlit Publishers, Moscow, 1961.

Borden, R. S., *A Course in Advanced Calculus*, Dover Publications, New York, 1998.

Brannan, D., *A First Course in Mathematical Analysis*, Cambridge University Press, Cambridge, 2006.

Bronshtein, I. N. and Semendyayev, K. A., *Handbook of Mathematics, 4th Edition*, Springer-Verlag, Berlin, 2004.

Browder, A., *Mathematical Analysis: An Introduction*, Springer-Verlag, New York, 1996.
Clark, D. N., *Dictionary of Analysis, Calculus, and Differential Equations*, CRC Press, Boca Raton, 2000.
Courant, R. and John, F., *Introduction to Calculus and Analysis, Vol. 1*, Springer-Verlag, New York, 1999.
Danilov, V. L., Ivanova, A. N., et al., *Mathematical Analysis (Functions, Limits, Series, Continued Fractions)* [in Russian], Fizmatlit Publishers, Moscow, 1961.
Dwight, H. B., *Tables of Integrals and Other Mathematical Data*, Macmillan, New York, 1961.
Edwards, C. H., and Penney, D., *Calculus, 6th Edition*, Pearson Education, Toronto, 2002.
Fedoryuk, M. V., *Asymptotics, Integrals and Series* [in Russian], Nauka Publishers, Moscow, 1987.
Fikhtengol'ts, G. M., *Fundamentals of Mathematical Analysis, Vol. 2*, Pergamon Press, London, 1965.
Fikhtengol'ts, G. M., *A Course of Differential and Integral Calculus, Vol. 2* [in Russian], Nauka Publishers, Moscow, 1969.
Gradshteyn, I. S. and Ryzhik, I. M., *Tables of Integrals, Series and Products, 6th Edition*, Academic Press, New York, 2000.
Kaplan, W., *Advanced Calculus, 5th Edition*, Addison Wesley, Reading, Massachusetts, 2002.
Kline, M., *Calculus: An Intuitive and Physical Approach, 2nd Edition*, Dover Publications, New York, 1998.
Landau, E., *Differential and Integral Calculus*, American Mathematical Society, Providence, 2001.
Marsden, J. E. and Weinstein, A., *Calculus, 2nd Edition*, Springer-Verlag, New York, 1985.
Mendelson, E., *3000 Solved Problems in Calculus*, McGraw-Hill, New York, 1988.
Polyanin, A. D., Polyanin, V. D., et al., *Handbook for Engineers and Students. Higher Mathematics. Physics. Theoretical Mechanics. Strength of Materials, 3rd Edition* [in Russian], AST/Astrel, Moscow, 2005.
Prudnikov, A. P., Brychkov, Yu. A., and Marichev, O. I., *Integrals and Series, Vol. 1, Elementary Functions*, Gordon & Breach, New York, 1986.
Silverman, R. A., *Essential Calculus with Applications*, Dover Publications, New York, 1989.
Strang, G., *Calculus*, Wellesley-Cambridge Press, Massachusetts, 1991.
Taylor, A. E. and Mann, W. R., *Advanced Calculus, 3rd Edition*, John Wiley, New York, 1983.
Thomas, G. B. and Finney, R. L., *Calculus and Analytic Geometry, 9th Edition*, Addison Wesley, Reading, Massachusetts, 1996.
Widder, D. V., *Advanced Calculus, 2nd Edition*, Dover Publications, New York, 1989.
Zorich, V. A., *Mathematical Analysis*, Springer-Verlag, Berlin, 2004.
Zwillinger, D., *CRC Standard Mathematical Tables and Formulae, 31st Edition*, CRC Press, Boca Raton, 2002.

Chapter 8
Series

8.1. Numerical Series and Infinite Products

8.1.1. Convergent Numerical Series and Their Properties. Cauchy's Criterion

8.1.1-1. Basic definitions.

Let $\{a_n\}$ be a numerical sequence. The expression

$$a_1 + a_2 + \cdots + a_n + \cdots = \sum_{n=1}^{\infty} a_n$$

is called a *numerical series* (*infinite sum*, *infinite numerical series*), a_n is the *generic term of the series*, and

$$s_n = a_1 + a_2 + \cdots + a_n = \sum_{k=1}^{n} a_k$$

is the n*th partial sum of the series*. If there exists a finite limit $\lim\limits_{n\to\infty} s_n = S$, the series is called *convergent*, and S is called the *sum of the series*. In this case, one writes $\sum\limits_{n=1}^{\infty} a_n = S$. If $\lim\limits_{n\to\infty} s_n$ does not exist (or is infinite), the series is called *divergent*. The series $a_{n+1} + a_{n+2} + a_{n+3} + \cdots$ is called the n*th remainder of the series*.

Example 1. Consider the series $\sum\limits_{n=1}^{\infty} aq^{n-1} = a + aq + aq^2 + \cdots$ whose terms form a *geometric progression* with ratio q. This series is convergent for $|q| < 1$ (its sum has the form $S = \frac{a}{1-q}$) and is divergent for $|q| \geq 1$.

8.1.1-2. Necessary condition for a series to be convergent. Cauchy's criterion.

1. *A necessary condition for a series to be convergent.* For a convergent series $\sum\limits_{n=1}^{\infty} a_n$, the generic term must tend to zero, $\lim\limits_{n\to\infty} a_n = 0$. If $\lim\limits_{n\to\infty} a_n \neq 0$, then the series is divergent.

Example 2. The series $\sum\limits_{n=1}^{\infty} \cos\dfrac{1}{n}$ is divergent, since its generic term $a_n = \cos\dfrac{1}{n}$ does not tend to zero as $n \to \infty$.

The above necessary condition is insufficient for the convergence of a series.

Example 3. Consider the series $\sum\limits_{n=1}^{\infty} \dfrac{1}{\sqrt{n}}$. Its generic term tends to zero, $\lim\limits_{n\to\infty} \dfrac{1}{\sqrt{n}} = 0$, but the series $\sum\limits_{n=1}^{\infty} \dfrac{1}{\sqrt{n}}$ is divergent because its partial sums are unbounded,

$$s_n = \frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \cdots + \frac{1}{\sqrt{n}} > n\frac{1}{\sqrt{n}} = \sqrt{n} \to \infty \quad \text{as} \quad n \to \infty.$$

2. *Cauchy's criterion of convergence of a series.* A series $\sum\limits_{n=1}^{\infty} a_n$ is convergent if and only if for any $\varepsilon > 0$ there is $N = N(\varepsilon)$ such that for all $n > N$ and any positive integer k, the following inequality holds: $|a_{n+1} + \cdots + a_{n+k}| < \varepsilon$.

8.1.1-3. Properties of convergent series.

1. If a series is convergent, then any of its remainders is convergent. Removal or addition of finitely many terms does not affect the convergence of a series.

2. If all terms of a series are multiplied by a nonzero constant, the resulting series preserves the property of convergence or divergence (its sum is multiplied by that constant).

3. If the series $\sum\limits_{n=1}^{\infty} a_n$ and $\sum\limits_{n=1}^{\infty} b_n$ are convergent and their sums are equal to S_1 and S_2, respectively, then the series $\sum\limits_{n=1}^{\infty}(a_n \pm b_n)$ are convergent and their sums are equal to $S_1 \pm S_2$.

4. Terms of a convergent series can be grouped in successive order; the resulting series has the same sum. In other words, one can insert brackets inside a series in an arbitrary order. The inverse operation of opening brackets is not always admissible. Thus, the series $(1-1)+(1-1)+\cdots$ is convergent (its sum is equal to zero), but, after removing the brackets, we obtain the divergent series $1 - 1 + 1 - 1 + \cdots$ (its generic term does not tend to zero).

8.1.2. Convergence Criteria for Series with Positive (Nonnegative) Terms

8.1.2-1. Basic convergence (divergence) criteria for series with positive terms.

1. *The first comparison criterion.* If $0 \le a_n \le b_n$ (starting from some n), then convergence of the series $\sum\limits_{n=1}^{\infty} b_n$ implies convergence of $\sum\limits_{n=1}^{\infty} a_n$; and divergence of the series $\sum\limits_{n=1}^{\infty} a_n$ implies divergence of $\sum\limits_{n=1}^{\infty} b_n$.

2. *The second convergence criterion.* Suppose that there is a finite limit

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \sigma,$$

where $0 < \sigma < \infty$. Then $\sum\limits_{n=1}^{\infty} a_n$ is convergent (resp., divergent) if and only if $\sum\limits_{n=1}^{\infty} b_n$ is convergent (resp., divergent).

Corollary. Suppose that $a_{n+1}/a_n \le b_{n+1}/b_n$ starting from some N (i.e., for $n > N$). Then convergence of the series $\sum\limits_{n=1}^{\infty} b_n$ implies convergence of $\sum\limits_{n=1}^{\infty} a_n$, and divergence of $\sum\limits_{n=1}^{\infty} a_n$ implies divergence of $\sum\limits_{n=1}^{\infty} b_n$.

3. *D'Alembert criterion.* Suppose that there exists the limit (finite or infinite)

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = D.$$

If $D < 1$, then the series $\sum\limits_{n=1}^{\infty} a_n$ is convergent. If $D > 1$, then the series is divergent. For $D = 1$, the D'Alembert criterion cannot be used for deciding whether the series is convergent or divergent.

Example 1. Let us examine convergence of the series $\sum\limits_{n=1}^{\infty} n^k x^n$ with $x > 0$, using the D'Alembert criterion. Taking $a_n = n^k x^n$, we get

$$\frac{a_{n+1}}{a_n} = \left(1 + \frac{1}{n}\right)^k x \to x \quad \text{as} \quad n \to \infty.$$

Therefore, $D = x$. It follows that the series is convergent for $x < 1$ and divergent for $x > 1$.

4. *Cauchy's criterion.* Suppose that there exists the limit (finite or infinite)

$$\lim_{n\to\infty} \sqrt[n]{a_n} = K.$$

For $K < 1$, the series $\sum\limits_{n=1}^{\infty} a_n$ is convergent; for $K > 1$, the series is divergent. For $K = 1$, the Cauchy criterion cannot be used to establish convergence of a series.

Remark. The Cauchy criterion is stronger than the D'Alembert criterion, but the latter is simpler than the former.

5. *Gauss' criterion.* Suppose that the ratio of two consecutive terms of a series can be represented in the form

$$\frac{a_n}{a_{n+1}} = \lambda + \frac{\mu}{n} + o\left(\frac{1}{n}\right) \quad \text{as} \quad n \to \infty.$$

The series $\sum\limits_{n=1}^{\infty} a_n$ is convergent if $\lambda > 1$ or $\lambda = 1$, $\mu > 1$. The series is divergent if $\lambda < 1$ or $\lambda = 1$, $\mu \le 1$.

6. *Maclaurin–Cauchy integral criterion.* Let $f(x)$ be a nonnegative nonincreasing continuous function on the interval $1 \le x < \infty$. Let $f(1) = a_1$, $f(2) = a_2, \ldots, f(n) = a_n, \ldots$ Then the series $\sum\limits_{n=1}^{\infty} a_n$ is convergent if and only if the improper integral $\int_1^{\infty} f(x)\,dx$ is convergent.

Example 2. The *harmonic series* $\sum\limits_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots$ is divergent, since the integral $\int_1^{\infty} \frac{1}{x}\,dx$ is divergent. In a similar way, one finds that the series $\sum\limits_{n=1}^{\infty} \frac{1}{n^{\alpha}}$ is convergent for $\alpha > 1$ and divergent for $\alpha \le 1$.

8.1.2-2. Other criteria of convergence (divergence) of series with positive terms.

1. *Raabe criterion.* Suppose that there exists the limit (finite or infinite)

$$\lim_{n\to\infty} n\left(\frac{a_n}{a_{n+1}} - 1\right) = \mathcal{R}.$$

Then, for $\mathcal{R} > 1$ the series $\sum\limits_{n=1}^{\infty} a_n$ is convergent, and for $\mathcal{R} < 1$ it is divergent. For $\mathcal{R} = 1$, the Raabe criterion is inapplicable.

2. *Bertrand criterion.* Suppose that there exists the limit (finite or infinite)

$$\lim_{n\to\infty} \ln n \left[n\left(\frac{a_n}{a_{n+1}} - 1 \right) - 1 \right] = \mathcal{B}.$$

Then, for $\mathcal{B} > 1$ the series $\sum\limits_{n=1}^{\infty} a_n$ is convergent, and for $\mathcal{B} < 1$ it is divergent. For $\mathcal{R} = 1$, the Bertrand criterion is inapplicable.

3. *Kummer criterion.* Let b_n be an arbitrary sequence with positive terms. Suppose that there exists the limit (finite or infinite)

$$\lim_{n\to\infty} \left(b_n \frac{a_n}{a_{n+1}} - b_{n+1} \right) = \mathcal{K}.$$

Then, for $\mathcal{K} > 0$, the series $\sum\limits_{n=1}^{\infty} a_n$ is convergent. If $\mathcal{K} < 0$ and the additional condition $\sum\limits_{n=1}^{\infty} \dfrac{1}{b_n} = \infty$ holds, then the series is divergent.

Remark. From the Kummer criterion, we obtain the D'Alembert criterion (taking $b_n = 1$), the Raabe criterion (taking $b_n = n$), and the Bertrand criterion (taking $b_n = n \ln n$).

4. *Ermakov criterion.* Let $f(x)$ be a positive monotonically decreasing continuous function on the interval $1 \le x < \infty$. Let $f(1) = a_1$, $f(2) = a_2, \ldots, f(n) = a_n, \ldots$ Then the following implications hold:

1) If $\dfrac{e^x f(e^x)}{f(x)} \le q < 1$, then the series $\sum\limits_{n=1}^{\infty} a_n$ is convergent.

2) If $\dfrac{e^x f(e^x)}{f(x)} \ge 1$, then the series $\sum\limits_{n=1}^{\infty} a_n$ is divergent.

Here, it suffices to have the inequalities on the left for sufficiently large $x \ge x_0$.

5. *Generalized Ermakov criterion.* Let $f(x)$ be the function involved in the Ermakov criterion, and let $\varphi(x)$ be an arbitrary positive monotonically increasing function that has a continuous derivative and satisfies the inequality $\varphi(x) > x$. Then the following implications hold:

1) If $\dfrac{\varphi(x) f\big(\varphi(x)\big)}{f(x)} \le q < 1$, then the series $\sum\limits_{n=1}^{\infty} a_n$ is convergent.

2) If $\dfrac{\varphi(x) f\big(\varphi(x)\big)}{f(x)} \ge 1$, then the series $\sum\limits_{n=1}^{\infty} a_n$ is divergent.

Here, it suffices to have the inequalities on the left for sufficiently large $x \ge x_0$.

6. *Sapogov criterion.* Let $a_1, a_2, \ldots$ be a monotonically increasing sequence. Then the series

$$\sum_{n=1}^{\infty} \left(1 - \frac{a_n}{a_{n+1}} \right) \qquad \left[\text{as well as } \sum_{n=1}^{\infty} \left(\frac{a_n}{a_{n+1}} - 1 \right) \right]$$

is convergent, provided that the sequence a_n is bounded ($a_n \le L$). Otherwise, this series is divergent.

7. *Special Cauchy criterion.* Suppose that $a_1, a_2, \dots$ is a monotonically decreasing sequence. Then the series $\sum\limits_{n=1}^{\infty} a_n$ is convergent (resp., divergent) if and only if the series $\sum\limits_{n=1}^{\infty} 2^n a_{2^n}$ is convergent (resp., divergent).

8. *Abel–Dini criterion.* If the series $\sum\limits_{n=1}^{\infty} a_n$ is divergent and s_n denotes its partial sum, then the series $\sum\limits_{n=1}^{\infty} \dfrac{a_n}{s_n}$ is also divergent, while the series $\sum\limits_{n=1}^{\infty} \dfrac{a_n}{s_n^{1+\sigma}}$ $(\sigma > 0)$ is convergent.

9. *Dini criterion.* If the series $\sum\limits_{n=1}^{\infty} a_n$ is convergent and γ_n denotes its remainder after the nth term, then the series $\sum\limits_{n=1}^{\infty} \dfrac{a_n}{\gamma_{n-1}}$ is divergent, while the series $\sum\limits_{n=1}^{\infty} \dfrac{a_n}{\gamma_{n-1}^{\sigma}}$ $(0 < \sigma < 1)$ is convergent.

10. *Bugaev criterion.* If the function $\varphi'(x)u\big(\varphi(x)\big)$ is monotone for large enough x, then the series $\sum\limits_{n=1}^{\infty} u(n)$ and $\sum\limits_{n=1}^{\infty} \varphi'(n)u\big(\varphi(n)\big)$ are convergent or divergent simultaneously.

11. *Lobachevsky criterion.* Let $u(x)$ be a monotonically decreasing function defined for all x. Then the series $\sum\limits_{n=1}^{\infty} u(n)$ is convergent or divergent if and only if the series $\sum\limits_{k=1}^{\infty} p_k 2^{-k}$ is convergent or divergent, where p_k is defined from the relation $u(p_k) = 2^{-k}$.

8.1.3. Convergence Criteria for Arbitrary Numerical Series. Absolute and Conditional Convergence

8.1.3-1. Arbitrary series. Leibnitz, Abel, and Dirichlet convergence criteria.

1. *Leibnitz criterion.* Suppose that the terms a_n of a series $\sum\limits_{n=1}^{\infty} a_n$ have alternating signs, their absolute values form a nonincreasing sequence, and $a_n \to 0$ as $n \to \infty$. Then this "alternating" series is convergent. If S is the sum of the series and s_n is its nth partial sum, then the following inequality holds for the error $|S - s_n| \le |a_{n+1}|$.

Example 1. The series $1 - \dfrac{1}{2^2} + \dfrac{1}{3^3} - \dfrac{1}{4^4} + \dfrac{1}{5^5} - \cdots$ is convergent by the Leibnitz criterion. Taking $S \approx s_4 = 1 - \dfrac{1}{2^2} + \dfrac{1}{3^3} - \dfrac{1}{4^4}$, we obtain the error less than $a_5 = \dfrac{1}{5^5} = 0.00032$.

2. *Abel criterion.* Consider the series

$$\sum_{n=1}^{\infty} a_n b_n = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n + \cdots, \tag{8.1.3.1}$$

where a_n and b_n are two sequences or real numbers.

Series (8.1.3.1) is convergent if the series

$$\sum_{n=1}^{\infty} b_n = b_1 + b_2 + \cdots + b_n + \cdots \tag{8.1.3.2}$$

is convergent and the a_n form a bounded monotone sequence ($|a_n| < K$).

3. *Dirichlet criterion.* Series (8.1.3.1) is convergent if partial sums of series (8.1.3.2) are bounded uniformly in n,

$$\left|\sum_{k=1}^{n} b_k\right| \le M \qquad (n = 1, 2, \ldots),$$

and the sequence $a_n \to 0$ is monotone.

Example 2. Consider the series $\sum\limits_{n=1}^{\infty} a_n \sin(nx)$, where $a_n \to 0$ is a monotonically decreasing sequence. Taking $b_n = \sin(nx)$ and using a well-known identity, we find the partial sum

$$s_n = \sum_{k=1}^{n} \sin(kx) = \frac{\cos\left(\frac{1}{2}x\right) - \cos\left[\left(n + \frac{1}{2}\right)x\right)}{2\sin\left(\frac{1}{2}x\right)} \qquad (x \ne 2m\pi;\ \ m = 0, \pm 1, \pm 2, \ldots).$$

This sum is bounded for $x \ne 2m\pi$:

$$|s_n| \le \frac{1}{\left|\sin\left(\frac{1}{2}x\right)\right|}.$$

Therefore, by the Dirichlet criterion, the series $\sum\limits_{n=1}^{\infty} a_n \sin(nx)$ is convergent for any $x \ne 2m\pi$. Direct verification shows that this series is also convergent for $x = 2m\pi$ (since all its terms at these points are equal to zero).

Remark. The Leibnitz and the Able criteria can be deduced from the Dirichlet criterion.

8.1.3-2. Absolute and conditional convergence.

1. *Absolutely convergent series.* A series $\sum\limits_{n=1}^{\infty} a_n$ (with terms of arbitrary sign) is called *absolutely convergent* if the series $\sum\limits_{n=1}^{\infty} |a_n|$ is convergent.

Any absolutely convergent series is convergent. In order to establish absolute convergence of a series, one can use all convergence criteria for series with nonnegative terms given in Subsection 8.1.2 (in these criteria, a_n should be replaced by $|a_n|$).

Example 3. The series $1 + \frac{1}{2^2} - \frac{1}{3^2} - \frac{1}{4^2} + \frac{1}{5^2} + \frac{1}{6^2} - \cdots$ is absolutely convergent, since the series with the absolute values of its terms, $\sum\limits_{n=1}^{\infty} \frac{1}{n^2}$, is convergent (see the second series in Example 2 of Subsection 8.1.2 for $\alpha = 2$).

2. *Conditionally convergent series.* A convergent series $\sum\limits_{n=1}^{\infty} a_n$ is called *conditionally convergent* if the series $\sum\limits_{n=1}^{\infty} |a_n|$ is divergent.

Example 4. The series $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$ is conditionally convergent, since it is convergent (by the Leibnitz criterion), but the series with absolute values of its terms is divergent (it is a harmonic series; see Example 2 in Subsection 8.1.2).

Any rearrangement of the terms of an absolutely convergent series (in particular, a convergent series with nonnegative terms) neither violates its absolute convergence nor changes its sum. Conditionally convergent series do not possess this property: the terms of a conditionally convergent series can be rearranged in such order that the sum of the new series becomes equal to any given value; its terms can also be rearranged so as to result in a divergent series.

8.1.4. Multiplication of Series. Some Inequalities

8.1.4-1. Multiplication of series. Cauchy, Mertens, and Abel theorems.

A product of two infinite series $\sum\limits_{n=0}^{\infty} a_n$ and $\sum\limits_{n=0}^{\infty} b_n$ is understood as a series whose terms have the form $a_n b_m$ (n, $m = 0, 1, \ldots$). The products $a_n b_m$ can be ordered to form a series in many different ways. The following theorems allow us to decide whether it is possible to multiply series.

CAUCHY THEOREM. *Suppose that the series* $\sum\limits_{n=0}^{\infty} a_n$ *and* $\sum\limits_{n=0}^{\infty} b_n$ *are absolutely convergent and their sums are equal to* A *and* B*, respectively. Then any product of these series is an absolutely convergent series and its sum is equal to* AB*. The following Cauchy multiplication formula holds*:

$$\Big(\sum_{n=0}^{\infty} a_n\Big)\Big(\sum_{n=0}^{\infty} b_n\Big) = \sum_{n=0}^{\infty}\Big(\sum_{m=0}^{n} a_m b_{n-m}\Big). \tag{8.1.4.1}$$

MERTENS THEOREM. *The Cauchy multiplication formula (8.1.4.1) is also valid if one of the series,* $\sum\limits_{n=0}^{\infty} a_n$ *or* $\sum\limits_{n=0}^{\infty} b_n$*, is absolutely convergent and the other is (conditionally) convergent. In this case, the product is a convergent series, possibly, not absolutely convergent.*

ABEL THEOREM. *Consider two convergent series with sums* A *and* B*. Suppose that the product of these series in the form of Cauchy (8.1.4.1) is a convergent series with sum* C*. Then* $C = AB$*.*

8.1.4-2. Inequalities.

1. *Generalized triangle inequality*:

$$\Big|\sum_{n=1}^{\infty} a_n\Big| \le \sum_{n=1}^{\infty} |a_n|.$$

2. *Cauchy inequality* (*Cauchy–Bunyakovsky inequality*):

$$\Big(\sum_{n=1}^{\infty} a_n b_n\Big)^2 \le \Big(\sum_{n=1}^{\infty} a_n^2\Big)\Big(\sum_{n=1}^{\infty} b_n^2\Big).$$

3. *Minkowski inequality*:

$$\Big(\sum_{n=1}^{\infty} |a_n + b_n|^p\Big)^{\frac{1}{p}} \le \Big(\sum_{n=1}^{\infty} |a_n|^p\Big)^{\frac{1}{p}} + \Big(\sum_{n=1}^{\infty} |b_n|^p\Big)^{\frac{1}{p}}, \qquad p \ge 1.$$

4. *Hölder inequality* (for $p = 2$ coincides with the Cauchy inequality):

$$\Big|\sum_{n=1}^{\infty} a_n b_n\Big| \le \Big(\sum_{n=1}^{\infty} |a_n|^p\Big)^{\frac{1}{p}} \Big(\sum_{n=1}^{\infty} |b_n|^{\frac{p}{p-1}}\Big)^{\frac{p-1}{p}}, \qquad p > 1.$$

5. *An inequality with* π:

$$\Big(\sum_{n=1}^{\infty} a_n\Big)^4 \le \pi^2\Big(\sum_{n=1}^{\infty} a_n^2\Big)\Big(\sum_{n=1}^{\infty} n^2 a_n^2\Big).$$

In all these inequalities it is assumed that the series in the right-hand sides are convergent.

8.1.5. Summation Methods. Convergence Acceleration

8.1.5-1. Some simple methods for calculating the sum of a series.

THEOREM 1. *Suppose that the terms of a series* $\sum_{n=1}^{\infty} a_n$ *can be represented in the form* $a_n = b_n - b_{n+1}$, *where* b_n *is a sequence with a finite limit* b_∞. *Then*

$$\sum_{n=1}^{\infty} a_n = b_1 - b_\infty.$$

Example 1. Let us find the sum of the infinite series $S = \sum_{n=1}^{\infty} \frac{1}{n(n+1)}$.

We have

$$a_n = \frac{1}{n(n+a)} = \frac{1}{n} - \frac{1}{n+1} \quad \Longrightarrow \quad b_n = \frac{1}{n}.$$

Since $b_1 = 1$ and $\lim\limits_{n\to\infty} b_n = 0$, we get $S = 1$.

THEOREM 2. *Suppose that the terms of a series* $\sum_{n=0}^{\infty} a_n$ *can be represented in the form*

$$a_n = b_n - b_{n+m},$$

where m *is a positive integer and the sequence* b_n *has a finite limit* b_∞. *Then*

$$\sum_{n=1}^{\infty} a_n = b_0 + b_1 + \cdots + b_{m-1} - mb_\infty.$$

THEOREM 3. *Suppose that the terms of a series* $\sum_{n=0}^{\infty} a_n$ *can be represented in the form*

$$a_n = \alpha_1 b_{n+1} + \alpha_2 b_{n+2} + \cdots + \alpha_m b_{n+m}, \tag{8.1.5.1}$$

where $m \ge 2$ *is a fixed positive integer, the sequence* b_n *has a finite limit* b_∞, *and* α_k *satisfy the condition*

$$\alpha_1 + \alpha_2 + \cdots + \alpha_m = 0. \tag{8.1.5.2}$$

Then the series is convergent and

$$\sum_{n=0}^{\infty} a_n = \alpha_1 b_1 + (\alpha_1+\alpha_2) b_2 + \cdots + (\alpha_1+\alpha_2+\cdots+\alpha_{m-1}) b_{m-1} + [\alpha_2 + 2\alpha_3 + \cdots + (m-1)\alpha_m] b_\infty. \tag{8.1.5.3}$$

Example 2. For the series $\sum_{n=0}^{\infty} \frac{4n+6}{(n+1)(n+2)(n+3)}$, we have

$$a_n = \frac{4n+6}{(n+1)(n+2)(n+3)} = \frac{1}{n+1} + \frac{2}{n+2} - \frac{3}{n+3},$$

which corresponds to the following values in (8.1.5.1):

$$\alpha_1 = 1, \quad \alpha_2 = 2, \quad \alpha_3 = -3, \quad b_n = \frac{1}{n}.$$

Thus, condition (8.1.5.2) holds: $1 + 2 - 3 = 0$, and the sequence b_n tends to zero: $b_\infty = 0$. Using (8.1.5.3), we find the sum of the series

$$\sum_{n=0}^{\infty} \frac{4n+6}{(n+1)(n+2)(n+3)} = 1 \cdot 1 + (1+2)\frac{1}{2} = \frac{5}{2}.$$

8.1.5-2. Summation of series with the help of Laplace transforms.

Let $a(k)$ be the Laplace transform of a given function $f(x)$,

$$a(k) = \int_0^\infty e^{-kx} f(x)\,dx.$$

Then the following summation formulas hold:

$$\begin{aligned} \sum_{k=1}^{\infty} a(k) &= \int_0^\infty \frac{f(x)\,dx}{e^x - 1}, \\ \sum_{k=1}^{\infty} (-1)^k a(k) &= -\int_0^\infty \frac{f(x)\,dx}{e^x + 1}, \end{aligned} \tag{8.1.5.4}$$

provided that the series are convergent.

Example 3. It is easy to check that

$$\frac{a}{k^2 + a^2} = \int_0^\infty e^{-kx} \sin(ax)\,dx.$$

Therefore, using the first formula in (8.1.5.4), we get

$$\sum_{k=1}^{\infty} \frac{1}{k^2 + a^2} = \frac{1}{a} \int_0^\infty \frac{\sin(ax)\,dx}{e^x - 1} = \frac{\pi}{2a} \coth(\pi a) - \frac{1}{2a^2}.$$

8.1.5-3. Kummer and Abel transformations. Acceleration of convergence of series.

1°. *Kummer transformation.* Consider a series with positive (nonnegative) terms

$$\sum_{n=1}^{\infty} a_n \tag{8.1.5.5}$$

and an auxiliary series with a finite sum

$$B = \sum_{n=1}^{\infty} b_n. \tag{8.1.5.6}$$

Suppose that there is a finite limit

$$K = \lim_{n\to\infty} \frac{b_n}{a_n} \neq 0.$$

Under these conditions, the series (8.1.5.5) is convergent and the following identity holds:

$$\sum_{n=1}^{\infty} a_n = \frac{B}{K} + \sum_{n=1}^{\infty} \left(1 - \frac{1}{K}\frac{b_n}{a_n}\right) a_n. \tag{8.1.5.7}$$

The right-hand side of (8.1.5.7) is called the Kummer transformation of the given series.

Remark. The Kummer transformation is used for accelerating convergence of series, since the generic term $a_n^{(1)} = \left(1 - \frac{1}{K}\frac{b_n}{a_n}\right)a_n$ of the transformed series (8.1.5.7) tends to zero faster that the generic term $\{a_n\}$ of the original series (8.1.5.5): $\lim\limits_{n\to\infty} a_n^{(1)}/a_n = 0$. The auxiliary sequence $\{b_n\}$ is chosen in such a way that the sum (8.1.5.6) is known beforehand.

2°. *Abel transformation*:

$$\sum_{n=1}^{\infty} a_n b_n = \sum_{n=1}^{\infty} (a_n - a_{n+1})B_n, \qquad B_n = \sum_{k=1}^{n} b_k.$$

Here, it is assumed that the sequence $\{B_n\}$ is bounded and $\lim\limits_{n\to\infty} a_n = 0$ (for example, these conditions are satisfied if the sequence $\{a_n\}$ has finitely many nonzero terms).

Remark. The Abel transformation is used for accelerating convergence of series whose convergence is slow (if $\lim\limits_{n\to\infty} a_{n+1}/a_n = 1$ and $b_n = O(1)$).

8.1.6. Infinite Products

8.1.6-1. Convergent and divergent infinite products.

An *infinite product* is an expression of the form

$$a_1 a_2 a_3 \cdots = \prod_{n=1}^{\infty} a_n \qquad (a_n \neq 0),$$

where a_n are real (in the general case, complex) numbers. The expression

$$p_n = a_1 a_2 \cdots a_n = \prod_{k=1}^{n} a_k \tag{8.1.6.1}$$

is called a *finite product*.

One says that an *infinite product is convergent to* p if there exists a finite nonzero limit of the partial products:

$$\lim_{n\to\infty} \prod_{k=1}^{n} a_k = p. \tag{8.1.6.2}$$

If there is no such limit, or $p = \infty$, or $p = 0$, one says that the *infinite product is divergent*. In the last case one says that the *infinite product diverges to zero* ($p = 0$).

The simplest examples of infinite products $\prod\limits_{n=1}^{\infty} a_n$:

for $a_n = 1$	the infinite product is convergent,	$p_n \to 1$;
for $a_n = (-1)^n$	the infinite product is divergent,	p_n has no limit;
for $a_n = n$	the infinite product is divergent,	$p_n \to \infty$;
for $a_n = 1/n$	the infinite product is divergent to zero,	$p_n \to 0$.

8.1.6-2. Infinite products with positive factors.

Taking logarithm of relation (8.1.6.1), one can reduce the problem of convergence of infinite products with positive factors a_n to the problem of convergence of partial sums of numerical series:

$$s_n \equiv \ln p_n = \sum_{k=1}^{n} \ln a_k.$$

Therefore, in order to examine convergence of infinite products, one can use the convergence criteria for infinite series considered in Subsections 8.1.2–8.1.3 (with a_n replaced by $\ln a_n$).

In a similar way, one can examine convergence of infinite products containing finitely many negative factors (infinite products with infinitely many negative factors are divergent; see Corollary in Paragraph 8.1.6-3). To that end, one should consider the part of the infinite product with positive factors.

8.1.6-3. Necessary condition for an infinite product to be convergent.

If an infinite product $\prod_{n=1}^{\infty} a_n$ is convergent, then $\lim_{n\to\infty} a_n = 1$; if $\lim_{n\to\infty} a_n \neq 1$, then the infinite product is divergent. (This necessary condition is insufficient to ensure convergence of a product.)

COROLLARY. *For a convergent infinite product, there is N such that $a_n > 0$ for all $n > N$ (i.e., a convergent infinite product can have only finitely many negative factors).*

8.1.6-4. Convergence criteria for infinite products.

In the criteria given below, Δ_n denotes the difference of the factor and its limit, which is equal to 1 (see the above necessary condition of convergence):

$$\Delta_n = a_n - 1.$$

THEOREM 1. *Suppose that for sufficiently large n, the difference Δ_n does not change sign. Then the infinite product $\prod_{n=1}^{\infty} a_n$ is convergent if and only if the series*

$$\sum_{n=1}^{\infty} \Delta_n \tag{8.1.6.3}$$

is convergent.

THEOREM 2. *If series (8.1.6.3) and the series*

$$\sum_{n=1}^{\infty} \Delta_n^2$$

are convergent, then the infinite product $\prod_{n=1}^{\infty} a_n$ is convergent. (Here, the difference Δ_n may change sign.)

THEOREM 3 (AN ANALOGUE OF THE D'ALEMBERT THEOREM). *Suppose that the terms of an infinite product satisfy the necessary condition of convergence, i.e.,* $a_n \to 1$ *as* $n \to \infty$, *and there exists the limit (finite or infinite)*

$$\lim_{n\to\infty} \frac{\Delta_{n+1}}{\Delta_n} = \lim_{n\to\infty} \frac{a_{n+1}-1}{a_n-1} = D.$$

Then, for $D < 1$ *the infinite product* $\prod_{n=1}^{\infty} a_n$ *is convergent, and for* $D > 1$ *it is divergent. (For* $D = 1$, *this criterion is inapplicable.)*

8.1.6-5. Absolute convergence of infinite products.

An infinite product $\prod_{n=1}^{\infty} a_n$ is said to be *absolutely convergent* if the product $\prod_{n=1}^{\infty} (1+|\Delta_n|)$ is convergent.

THEOREM 4. *An infinite product* $\prod_{n=1}^{\infty} a_n$ *is absolutely convergent if the series* $\sum_{n=1}^{\infty} |\Delta_n|$ *is convergent.*

An infinite product is *commutation convergent* (i.e., its value does not depend on the order of its factors) if and only if it is absolutely convergent.

8.2. Functional Series

8.2.1. Pointwise and Uniform Convergence of Functional Series

8.2.1-1. Convergence of a functional series at a point. Convergence domain.

A *functional series* is a series of the form

$$u_1(x) + u_2(x) + \cdots + u_n(x) + \cdots = \sum_{n=1}^{\infty} u_n(x),$$

where $u_n(x)$ are functions defined on a set X on the real axis. The series $\sum_{n=1}^{\infty} u_n(x)$ is called *convergent at a point* $x_0 \in X$ if the numerical series $\sum_{n=1}^{\infty} u_n(x_0)$ is convergent. The set of all $x \in X$ for which the functional series is convergent is called its *convergence domain*. The sum of the series is a function of x defined on its convergence domain.

In order to find the convergence domain for a functional series, one can use the convergence criteria for numerical series described in Subsections 8.1.2 and 8.1.3 (with the variable x regarded as a parameter).

A series $\sum_{n=1}^{\infty} u_n(x)$ is called *absolutely convergent on a set* X if the series $\sum_{n=1}^{\infty} |u_n(x)|$ is convergent on that set.

Example. The functional series $1 + x + x^2 + x^3 + \cdots$ is convergent for $-1 < x < 1$ (see Example 1 in Subsection 8.1.1). Its sum is defined on this interval, $S = \dfrac{1}{1-x}$.

The series $\sum\limits_{k=n+1}^{\infty} u_k(x)$ is called the *remainder of a functional series* $\sum\limits_{n=1}^{\infty} u_n(x)$. For a series convergent on a set X, the relation $S(x) = s_n(x) + r_n(x)$ ($s_n(x)$ is the partial sum of the series, $r_n(x)$ is the sum of its remainder) implies that $\lim\limits_{n\to\infty} r_n(x) = 0$ for $x \in X$.

8.2.1-2. Uniformly convergent series. Condition of uniform convergence.

A functional series is called *uniformly convergent* on a set X if for any $\varepsilon > 0$ there is N (depending on ε but not on x) such that for all $n > N$, the inequality $\left|\sum\limits_{k=n+1}^{\infty} u_k(x)\right| < \varepsilon$ holds for all $x \in X$.

A *necessary and sufficient condition of uniform convergence of a series.* A series $\sum\limits_{n=1}^{\infty} u_n(x)$ is uniformly convergent on a set X if and only if for any $\varepsilon > 0$ there is N (independent of x) such that for all $n > N$ and all $m = 1, 2, \ldots$, the inequality

$$\left|\sum_{k=n+1}^{n+m} u_k(x)\right| < \varepsilon$$

holds for all $x \in X$.

8.2.2. Basic Criteria of Uniform Convergence. Properties of Uniformly Convergent Series

8.2.2-1. Criteria of uniform convergence of series.

1. *Weierstrass criterion of uniform convergence.* A functional series $\sum\limits_{n=1}^{\infty} u_n(x)$ is uniformly convergent on a set X if there is a convergent series $\sum\limits_{n=1}^{\infty} a_n$ with nonnegative terms such that $|u_n(x)| \le a_n$ for all sufficiently large n and all $x \in X$. The series $\sum\limits_{n=1}^{\infty} a_n$ is called a *majorant series* for $\sum\limits_{n=1}^{\infty} u_n(x)$.

Example. The series $\sum\limits_{n=1}^{\infty}(-1)^n \dfrac{\sin nx}{n^2}$ is uniformly convergent for $-\infty < x < \infty$, since $\left|(-1)^n \dfrac{\sin nx}{n^2}\right| \le \dfrac{1}{n^2}$, and the numerical series $\sum\limits_{n=1}^{\infty} \dfrac{1}{n^2}$ is convergent (see the second series in Example 2 in Subsection 8.1.2).

2. *Abel criterion of uniform convergence of functional series.* Consider a functional series

$$\sum_{n=1}^{\infty} u_n(x)v_n(x) = u_1(x)v_1(x) + u_2(x)v_2(x) + \cdots + u_n(x)v_n(x) + \cdots, \qquad (8.2.2.1)$$

where $u_n(x)$ and $v_n(x)$ are sequences of functions of the real variable $x \in [a, b]$.

Series (8.2.2.1) is uniformly convergent on the interval $[a, b]$ if the series

$$\sum_{n=1}^{\infty} v_n(x) = v_1(x) + v_2(x) + \cdots + v_n(x) + \cdots \tag{8.2.2.2}$$

is uniformly convergent on $[a, b]$ and the functions $u_n(x)$ form a monotone sequence for each x and are uniformly bounded (i.e., $|u_n(x)| \le K$ with a constant K independent of n, x).

3. *Dirichlet criterion of uniform convergence of functional series.* Series (8.2.2.1) is uniformly convergent on the interval $[a, b]$ if the partial sums of the series (8.2.2.2) are uniformly bounded, i.e.,

$$\left|\sum_{k=1}^{n} v_k(x)\right| \le M = \text{const} \qquad (x \in [a, b], \quad n = 1, 2, \ldots),$$

and the functions $u_n(x)$ form a monotone sequence (for each x) that uniformly converges to zero on $[a, b]$ as $n \to \infty$.

8.2.2-2. Properties of uniformly convergent series.

Let $\sum\limits_{n=1}^{\infty} u_n(x)$ be a functional series that is uniformly convergent on a segment $[a, b]$, and let $S(x)$ be its sum. Then the following statements hold.

THEOREM 1. *If all terms $u_n(x)$ of the series are continuous at a point $x_0 \in [a, b]$, then the sum $S(x)$ is continuous at that point.*

THEOREM 2. *If the terms $u_n(x)$ are continuous on $[a, b]$, then the series admits term-by-term integration:*

$$\int_a^b S(x)\,dx = \int_a^b \left(\sum_{n=1}^{\infty} u_n(x)\right) dx = \sum_{n=1}^{\infty} \int_a^b u_n(x)\,dx.$$

Remark. The condition of continuity of the functions $u_n(x)$ on $[a, b]$ can be replaced by a weaker condition of their integrability on $[a, b]$.

THEOREM 3. *If all terms of the series have continuous derivatives and the functional series $\sum\limits_{n=1}^{\infty} u_n'(x)$ is uniformly convergent on $[a, b]$, then the sum $S(x)$ is continuously differentiable on $[a, b]$ and*

$$S'(x) = \left(\sum_{n=1}^{\infty} u_n(x)\right)' = \sum_{n=1}^{\infty} u_n'(x)$$

(i.e., the series admits term-by-term differentiation).

8.3. Power Series

8.3.1. Radius of Convergence of Power Series. Properties of Power Series

8.3.1-1. Abel theorem. Convergence radius of a power series.

A *power series* is a functional series of the form

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots \tag{8.3.1.1}$$

(the constants $a_0, a_1, \ldots$ are called the *coefficients* of the power series), and also a series of a more general form

$$\sum_{n=0}^{\infty} a_n(x-x_0)^n = a_0 + a_1(x-x_0) + a_2(x-x_0)^2 + a_3(x-x_0)^3 + \cdots,$$

where x_0 is a fixed point. Below, we consider power series of the first form, since the second series can be transformed into the first by the replacement $\bar{x} = x - x_0$.

ABEL THEOREM. *A power series* $\sum\limits_{n=0}^{\infty} a_n x^n$ *that is convergent for some* $x = x_1$ *is absolutely convergent for all* x *such that* $|x| < |x_1|$. *A power series that is divergent for some* $x = x_2$ *is divergent for all* x *such that* $|x| > |x_2|$.

Remark. There exist series convergent for all x, for instance, $\sum\limits_{n=1}^{\infty} \dfrac{x^n}{n!}$. There are series convergent only for $x = 0$, for instance, $\sum\limits_{n=1}^{\infty} n!\, x^n$.

For a given power series (8.3.1.1), let R be the least upper bound of all $|x|$ such that the series (8.3.1.1) is convergent at point x. Thus, by the Abel theorem, the series is (absolutely) convergent for all $|x| < R$, and the series is divergent for all $|x| > R$. The constant R is called the *radius of convergence* of the power series, and the interval $(-R, R)$ is called its *interval of convergence*. The problem of convergence of a power series at the endpoints of its convergence interval has to be studied separately in each specific case. If a series is convergent only for $x = 0$, the convergence interval degenerates into a point (and $R = 0$); if a series is convergent for all x, then, obviously, $R = \infty$.

8.3.1-2. Formulas for the radius of convergence of power series.

1°. The radius of convergence of a power series (8.3.1.1) with finitely many zero terms can be calculated by the formulas

$$R = \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right| \quad \text{(obtained from the D'Alembert criterion for numerical series)},$$

$$R = \lim_{n\to\infty} \frac{1}{\sqrt[n]{|a_n|}} \quad \text{(obtained from the Cauchy criterion for numerical series)}.$$

Example 1. For the power series $\sum\limits_{n=1}^{\infty} \dfrac{3^n}{n} x^n$, using the first formula for the radius of convergence, we get

$$R = \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right| = \lim_{n\to\infty} \left| \frac{n+1}{3n} \right| = \frac{1}{3}.$$

Therefore, the series is absolutely convergent on the interval $-\frac{1}{3} < x < \frac{1}{3}$ and is divergent outside that interval. At the left endpoint of the interval, for $x = -\frac{1}{3}$, we have the conditionally convergent series $\sum\limits_{n=1}^{\infty} \dfrac{(-1)^n}{n}$, and at the right endpoint, for $x = \frac{1}{3}$, we have the divergent numerical series $\sum\limits_{n=1}^{\infty} \dfrac{1}{n}$. Thus, the series under consideration is convergent on the semi-open interval $\left[-\frac{1}{3}, \frac{1}{3}\right)$.

2°. Suppose that a power series (8.3.1.1) is convergent at a boundary point of its convergence interval, say, for $x = R$. Then its sum is left-hand continuous at that point,

$$\lim_{x\to R-0} \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} a_n R^n.$$

Example 2. Having the expansion

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n+1}\frac{x^n}{n} + \cdots \qquad (R = 1)$$

in the domain $-1 < x < 1$ and knowing that the series

$$1 - \frac{1}{2} + \frac{1}{3} - \cdots + (-1)^{n+1}\frac{1}{n} + \cdots$$

is convergent (by the Leibnitz criterion for series with terms of alternating sign), we conclude that the sum of the last series is equal to $\ln 2$.

8.3.1-3. Properties of power series.

On any closed segment belonging to the (open) convergence interval of a power series, the series is uniformly convergent. Therefore, on any such segment, the series has all the properties of uniformly convergent series described in Subsection 8.2.2. Therefore, the following statements hold:

1. A power series (8.3.1.1) admits term-by-term integration on any segment $[0, x]$ for $|x| < R$,

$$\int_0^x \left(\sum_{n=0}^{\infty} a_n x^n\right) dx = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}$$
$$= a_0 x + \frac{a_1}{2}x^2 + \frac{a_2}{3}x^3 + \cdots + \frac{a_n}{n+1}x^{n+1} + \cdots .$$

Remark 1. The value of x in this formula may coincide with an endpoint of the convergence interval ($x = -R$ and/or $x = R$), provided that series (8.3.1.1) is convergent at that point.

Remark 2. The convergence radii of the original series and the series obtained by its term-by-term integration on the segment $[0, x]$ coincide.

2. Inside the convergence interval (for $|x| < R$), the series admits term-by-term differentiation of any order, in particular,

$$\frac{d}{dx}\left(\sum_{n=0}^{\infty} a_n x^n\right) = \sum_{n=1}^{\infty} n a_n x^{n-1}$$
$$= a_1 + 2a_2 x + 3a_3 x^2 + \cdots + n a_n x^{n-1} + \cdots .$$

Remark 1. This statement remains valid for an endpoint of the convergence interval if the series (8.3.1.1) is convergent at that point.

Remark 2. The convergence radii of the original series and the series obtained by its term-by-term differentiation coincide.

8.3.2. Taylor and Maclaurin Power Series

8.3.2-1. Basic definitions.

Let $f(x)$ is an infinitely differentiable function at a point x_0. The *Taylor series* for this function is the power series

$$\sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + \cdots ,$$

where $0! = 1$ and $f^{(0)}(x_0) = f(x_0)$.

A special case of the Taylor series (for $x_0 = 0$) is the *Maclaurin series*:

$$\sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(0) x^n = f(0) + f'(0)x + \frac{1}{2} f''(0)x^2 + \cdots .$$

A formal Taylor series (Maclaurin series) for a function $f(x)$ may be:

1) divergent for $x \neq x_0$,
2) convergent in a neighborhood of x_0 to a function different from $f(x)$,
3) convergent in a neighborhood of x_0 to the function $f(x)$.

In the last case, one says that $f(x)$ admits *expansion in Taylor series* in the said neighborhood, and one writes

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n.$$

8.3.2-2. Conditions of expansion in Taylor series.

A *necessary and sufficient condition for a function $f(x)$ to be represented by its Taylor series* in a neighborhood of a point x_0 is that the remainder term in the Taylor formula* should tend to zero as $n \to \infty$ in this neighborhood of x_0.

In order that $f(x)$ could be represented by its Taylor series in a neighborhood of x_0, it suffices that all its derivatives in that neighborhood be bounded by the same constant, $|f^{(n)}(x)| \leq M$ for all n.

Uniqueness of the Taylor series expansion. For a function $f(x)$ that can be represented as the sum of a power series, the coefficients of this series are determined uniquely (since this series is the Taylor series of $f(x)$ and its coefficients have the form $\dfrac{f^{(n)}(x_0)}{n!}$, where $n = 0,\ 1,\ 2, \ldots$). Therefore, in problems of representing a function by a power series, the answer does not depend on the method adopted for this purpose.

8.3.2-3. Representation of some functions by the Maclaurin series.

The following representations of elementary functions by Maclaurin series are often used in applications:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots ;$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^{n-1} \frac{x^{2n-1}}{(2n-1)!} + \cdots ;$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + \cdots ;$$

$$\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots + \frac{x^{2n-1}}{(2n-1)!} + \cdots ;$$

* Different representations of the remainder in the Taylor formula are given in Paragraph 6.2.4-4.

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots + \frac{x^{2n}}{(2n)!} + \cdots;$$

$$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!}x^2 + \cdots + \frac{\alpha(\alpha-1)\ldots(\alpha-n+1)}{n!}x^n + \cdots;$$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n+1}\frac{x^n}{n} + \cdots;$$

$$\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \cdots + (-1)^{n+1}\frac{x^{2n-1}}{2n-1} + \cdots.$$

The first five series are convergent for $-\infty < x < \infty$ ($R = \infty$), and the other series have unit radius of convergence, $R = 1$.

8.3.3. Operations with Power Series. Summation Formulas for Power Series

8.3.3-1. Addition, subtraction, multiplication, and division of power series.

1. *Addition and subtraction of power series.* Two series $\sum\limits_{n=0}^{\infty} a_n x^n$ and $\sum\limits_{n=0}^{\infty} b_n x^n$ with convergence radii R_a and R_b, respectively, admit term-by-term addition and subtraction on the intersection of their convergence intervals:

$$\sum_{n=0}^{\infty} a_n x^n \pm \sum_{n=0}^{\infty} b_n x^n = \sum_{n=0}^{\infty} c_n x^n, \qquad c_n = a_n \pm b_n.$$

The radius of convergence of the resulting series satisfies the inequality $R_c \geq \min[R_a, R_b]$.

2. *Multiplication of power series.* Two series $\sum\limits_{n=0}^{\infty} a_n x^n$ and $\sum\limits_{n=0}^{\infty} b_n x^n$, with the respective convergence radii R_a and R_b, can be multiplied on the intersection of their convergence intervals, and their product has the form

$$\left(\sum_{n=0}^{\infty} a_n x^n\right)\left(\sum_{n=0}^{\infty} b_n x^n\right) = \sum_{n=0}^{\infty} c_n x^n, \qquad c_n = \sum_{k=0}^{n} a_k b_{n-k}.$$

The convergence radius of the product satisfies the inequality $R_c \geq \min[R_a, R_b]$.

3. *Division of power series.* The ratio of two power series $\sum\limits_{n=0}^{\infty} a_n x^n$ and $\sum\limits_{n=0}^{\infty} b_n x^n$, $b_0 \neq 0$, with convergence radii R_a and R_b can be represented as a power series

$$\frac{\sum\limits_{n=0}^{\infty} a_n x^n}{\sum\limits_{n=0}^{\infty} b_n x^n} = c_0 + c_1 x + c_2 x^2 + \cdots = \sum_{n=0}^{\infty} c_n x^n, \tag{8.3.3.1}$$

whose coefficients can be found, by the method of indefinite coefficients, from the relation

$$(a_0 + a_1 x + a_2 x^2 + \cdots) = (b_0 + b_1 x + b_2 x^2 + \cdots)(c_0 + c_1 x + c_2 x^2 + \cdots).$$

Thus, for the unknown c_n, we obtain a triangular system of linear algebraic equations

$$a_n = \sum_{k=0}^{n} b_k c_{n-k}, \qquad n = 0,\ 1,\ \dots\ ,$$

which is solved consecutively, starting from the first equation:

$$c_0 = \frac{a_0}{b_0}, \quad c_1 = \frac{a_1 b_0 - a_0 b_1}{b_0^2}, \quad c_n = \frac{a_n}{b_0} - \frac{1}{b_0} \sum_{k=1}^{n} b_k c_{n-k}, \quad n = 2,\ 3,\ \dots$$

The convergence radius of the series (8.3.3.1) is determined by the formula

$$R_1 = \min\Bigl[R_a, \frac{\rho}{M+1}\Bigr],$$

where ρ is any constant such that $0 < \rho < R_b$; ρ can be chosen arbitrarily close to R_b; and M is the least upper bound of the quantities $|b_m/b_0|\rho^m$ ($m = 1,\ 2,\ \dots$), so that $|b_m/b_0|\rho^m \le M$ for all m.

8.3.3-2. Composition of functions representable by power series.

Consider a power series

$$z = f(y) = a_0 + a_1 y + a_2 y^2 + \cdots = \sum_{n=0}^{\infty} a_n y^n \tag{8.3.3.2}$$

with convergence radius R. Let the variable y be a function of x that can be represented by a power series

$$y = \varphi(x) = b_0 + b_1 x + a_2 x^2 + \cdots = \sum_{n=0}^{\infty} b_n x^n \tag{8.3.3.3}$$

with convergence radius r. It is required to represent z as a power series of x and find the convergence radius of this series.

Formal substitution of (8.3.3.3) into (8.3.3.2) yields

$$z = f\bigl(\varphi(x)\bigr) = \sum_{n=0}^{\infty} a_n \Bigl(\sum_{k=0}^{\infty} b_k x^k\Bigr)^n = A_0 + A_1 x + A_2 x^2 + \cdots = \sum_{n=0}^{\infty} A_n x^n, \tag{8.3.3.4}$$

where

$$\begin{aligned}
A_0 &= a_0 + a_1 b_0 + a_2 b_0^2 + \cdots\ ,\\
A_1 &= a_1 b_1 + 2a_2 b_0 b_1 + 3a_3 b_0^2 b_1 + \cdots\ ,\\
A_2 &= a_1 b_2 + a_2(b_1^2 + 2b_0 b_2) + 3a_3(b_0 b_1^2 + b_0^2 b_2) + \cdots\ ,\\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots
\end{aligned}$$

THEOREM ON CONVERGENCE OF SERIES (8.3.3.4).

(i) If series (8.3.3.2) is convergent for all y (i.e., $R = \infty$), then the convergence radius of series (8.3.3.4) coincides with the convergence radius r of series (8.3.3.3).

(ii) If $0 \le |b_0| < R$, then series (8.3.3.4) is convergent on the interval $(-R_1, R_1)$, where

$$R_1 = \frac{(R - |b_0|)\rho}{M + R - |b_0|},$$

and ρ is an arbitrary constant such that $0 < \rho < r$; ρ can be chosen arbitrarily close to r; and M is the least upper bound of the quantities $|b_m|\rho^m$ ($m = 1, 2, \ldots$), so that $|b_m|\rho^m \le M$ for all m.

(iii) If $|b_0| > R$, then series (8.3.3.4) is divergent.

Remark. Case (i) is realized, for instance if (8.3.3.2) has finitely many terms.

8.3.3-3. Local inversion of a function represented by power series.

1. Suppose that $y = y(x)$ is a function that can be represented, in a neighborhood of a point $x = x_0$, by the power series

$$y = y_0 + a(x - x_0) + b(x - x_0)^2 + c(x - x_0)^3 + d(x - x_0)^4 + \cdots, \qquad a \ne 0.$$

Then the inverse function $x = x(y)$, in a neighborhood of $y = y_0$, can be represented by the series

$$x = x_0 + \frac{1}{a}(y - y_0) - \frac{b}{a^3}(y - y_0)^2 + \frac{2b^2 - ac}{a^5}(y - y_0)^3 + \frac{5abc - 5b^3 - a^2 d}{a^7}(y - y_0)^4 + \cdots.$$

2. *Bürman–Lagrange formula.* Suppose that for a given function

$$y = f(x), \tag{8.3.3.5}$$

the auxiliary function

$$\varphi(x) = \frac{x}{f(x)}$$

is holomorphic in a neighborhood of the point $x = 0$ (i.e., it can be represented by a convergent power series in a neighborhood of that point). Then there is $\varepsilon > 0$ such that on the interval $|y| < \varepsilon$, the function (8.3.3.5) is invertible and its inverse $x = g(y)$ is holomorphic on that interval,

$$x = \sum_{n=1}^{\infty} b_n y^n, \qquad b_n = \frac{1}{n!}\left[\frac{d^{n-1}}{dx^{n-1}}\varphi^n(x)\right]_{x=0}.$$

The expression for the coefficients b_n is called the *Bürman–Lagrange formula.*

Example 1. Consider the function

$$y = x(x + b) \qquad (b \ne 0),$$

for which the auxiliary function has the form $\varphi(x) = (x + a)^{-1}$. Using the Bürman–Lagrange formula and the relation

$$\frac{d^{n-1}}{dx^{n-1}}\frac{1}{(x + a)^n} = \frac{(-1)^{n-1} n(n + 1)\cdots(2n - 2)}{(x + a)^{2n-1}},$$

we find the representation of the given function by power series:

$$x = \frac{y}{a} - \frac{y^2}{a^3} + \cdots + (-1)^{n-1}\frac{(2n - 2)!}{(n - 1)!\, n!}\frac{y^n}{a^{2n-1}} + \cdots.$$

8.3.3-4. Simplest summation formulas for power series.

Suppose that the sum of a power series is known,

$$\sum_{k=0}^{\infty} a_k x^k = S(x). \tag{8.3.3.6}$$

Then, using term-by-term integration (on the convergence interval), one can find the following sums:

$$\sum_{k=0}^{\infty} a_k k^m x^k = \left(x\frac{d}{dx}\right)^m S(x);$$

$$\sum_{k=0}^{\infty} a_k (nk+m) x^{nk+m-1} = \frac{d}{dx}\left[x^m S(x^n)\right];$$

$$\sum_{k=0}^{\infty} \frac{a_k}{nk+m} x^{nk+m} = \int_0^x x^{m-1} S(x^n)\,dx, \qquad n>0, \;\; m>0; \tag{8.3.3.7}$$

$$\sum_{k=0}^{\infty} a_k \frac{nk+s}{nk+m} x^{nk+s} = x\frac{d}{dx}\left[x^{s-m}\int_0^x x^{m-1} S(x^n)\,dx\right], \qquad n>0, \;\; m>0;$$

$$\sum_{k=0}^{\infty} a_k \frac{nk+m}{nk+s} x^{nk+s} = \int_0^x x^{s-m}\frac{d}{dx}\left[x^m S(x^n)\right]dx, \qquad n>0, \;\; s>0.$$

Example 2. Let us find the sum of the series $\sum\limits_{n=0}^{\infty} kx^{k-1}$.

We start with the well-known formula for the sum of an infinite geometrical progression:

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \qquad (|x|<1).$$

This series is a special case of (8.3.3.6) with $a_k = 1$, $S(x) = 1/(1-x)$. The series $\sum\limits_{n=0}^{\infty} kx^{k-1}$ can be obtained from the left-hand side of the second formula in (8.3.3.7) for $m = 0$ and $n = 1$. Substituting $S(x) = 1/(1-x)$ into the right-hand side of that formula, we get

$$\sum_{k=0}^{\infty} kx^{k-1} = \frac{d}{dx}\frac{1}{1-x} = \frac{1}{(1-x)^2} \qquad (|x|<1).$$

8.4. Fourier Series

8.4.1. Representation of 2π-Periodic Functions by Fourier Series. Main Results

8.4.1-1. Dirichlet theorem on representation of a function by Fourier series.

A function $f(x)$ is said to satisfy the *Dirichlet conditions* on an interval (a, b) if:

1) this interval can be divided into finitely many intervals on which $f(x)$ is monotone and continuous;

2) at any discontinuity point x_0 of the function, there exist finite one-sided limits $f(x_0+0)$ and $f(x_0-0)$.

DIRICHLET THEOREM. *Any 2π-periodic function that satisfies the Dirichlet conditions on the interval $(-\pi, \pi)$ can be represented by its Fourier series*

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \bigl(a_n \cos nx + b_n \sin nx\bigr) \tag{8.4.1.1}$$

whose coefficients are defined by the Euler–Fourier formulas

$$\begin{aligned} a_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, dx, \qquad n = 0,\ 1,\ 2, \ldots, \\ b_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, dx, \qquad n = 1,\ 2,\ 3, \ldots \end{aligned} \tag{8.4.1.2}$$

At the points of continuity of $f(x)$, the Fourier series converges to $f(x)$, and at any discontinuity point x_0, the series converges to $\frac{1}{2}[f(x_0 + 0) + f(x_0 - 0)]$.

The coefficients a_n and b_n of the series (8.4.1.1) are called the *Fourier coefficients*.

Remark. Instead of the integration limits $-\pi$ and π in (8.4.1.2), one can take c and $c + 2\pi$, where c is an arbitrary constant.

8.4.1-2. Lipschitz and Dirichlet–Jordan convergence criteria for Fourier series.

LIPSCHITZ CRITERION. *Suppose that $f(x)$ is continuous at a point x_0 and for sufficiently small $\varepsilon > 0$ satisfies the inequality $|f(x_0 \pm \varepsilon) - f(x_0)| \le K\varepsilon^{\sigma}$, where L and σ are constants, $0 < \sigma \le 1$. Then the representation (8.4.1.1)–(8.4.1.2) holds at $x = x_0$.*

In particular, the conditions of the Lipschitz criterion hold for continuous piecewise differentiable functions.

Remark. The Fourier series of a continuous periodic function with no additional conditions (for instance, of its regularity) may happen to be divergent at infinitely many (even uncountably many) points.

DIRICHLET–JORDAN CRITERION. *Suppose that $f(x)$ is a function of bounded variation on some interval $(x_0 - h, x_0 + h) \in (-\pi, \pi)$ (i.e., $f(x)$ can be represented as a difference of two monotonically increasing functions). Then the Fourier series (8.4.1.1)–(8.4.1.2) of the function $f(x)$ at the point x_0 converges to the value $\frac{1}{2}[f(x_0 + 0) + f(x_0 - 0)]$.*

8.4.1-3. Riemann localization principle.

RIEMANN LOCALIZATION PRINCIPLE. *The behavior of the Fourier series of a function $f(x)$ at a point x_0* depends only on its values near that point, i.e., values in an arbitrarily small neighborhood of that point.*

Thus, for two functions that coincide in a neighborhood of a point x_0, but differ outside that neighborhood, the corresponding Fourier series at x_0 are either both convergent or divergent and have the same sum in the case of convergence, although their Fourier coefficients may be different, being dependent on all values of the functions.

* What is meant here is the fact of convergence or divergence of the Fourier series at x_0, and also the numerical value of its sum in the case of convergence.

8.4.1-4. Asymptotic properties of Fourier coefficients.

1°. Fourier coefficients of an absolutely integrable function tend to zero as n goes to infinity: $a_n \to 0$ and $b_n \to 0$ as $n \to \infty$.

2°. Fourier coefficients of a continuous 2π-periodic function have the following limit properties:

$$\lim_{n\to\infty} (na_n) = 0, \qquad \lim_{n\to\infty} (nb_n) = 0,$$

i.e., $a_n = o(1/n)$, $b_n = o(1/n)$.

3°. If a continuous periodic function is continuously differentiable up to the order $m - 1$ inclusively, then its Fourier coefficients have the following limit properties:

$$\lim_{n\to\infty} (n^m a_n) = 0, \qquad \lim_{n\to\infty} (n^m b_n) = 0,$$

i.e., $a_n = o\big(n^{-m}\big)$, $b_n = o\big(n^{-m}\big)$.

8.4.1-5. Integration and differentiation of Fourier series.

1°. The Fourier series of a continuous periodic function of bounded variation admits term-by-term integration, and the resulting series is uniformly convergent.

2°. The Fourier series of a k times continuously differentiable function admits term-by-term differentiation $(k - 1)$ times, the resulting series still being uniformly convergent (the kth differentiation yields the kth derivative of the function, but the resulting series may have only mean-square convergence, not necessarily pointwise convergence).

8.4.2. Fourier Expansions of Periodic, Nonperiodic, Odd, and Even Functions

8.4.2-1. Expansion of $2l$-periodic and nonperiodic functions in Fourier series.

1°. The case of $2l$-periodic functions can be easily reduced to that of 2π-periodic functions by changing the variable x to $z = \dfrac{\pi x}{l}$. In this way, all the results described above for 2π-periodic functions can be easily extended to $2l$-periodic functions.

The Fourier expansion of a $2l$-periodic function $f(x)$ has the form

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{l} + b_n \sin \frac{n\pi x}{l} \right), \tag{8.4.2.1}$$

where

$$a_n = \frac{1}{l} \int_{-l}^{l} f(x) \cos \frac{n\pi x}{l}\, dx, \qquad b_n = \frac{1}{l} \int_{-l}^{l} f(x) \sin \frac{n\pi x}{l}\, dx. \tag{8.4.2.2}$$

2°. A nonperiodic (aperiodic) function $f(x)$ defined on the interval $(-l, l)$ can also be represented by a Fourier series (8.4.2.1)–(8.4.2.2); however, outside that interval, the sum of that series $S(x)$ may differ from $f(x)$*.

* The sum $S(x)$ is a $2l$-periodic function defined for all x, but $f(x)$ may happen to be nonperiodic, or even undefined outside the interval $(-l, l)$.

8.4.2-2. Fourier expansion of odd and even functions.

1°. Let $f(x)$ be an even function, i.e., $f(x) = f(-x)$. Then the Fourier expansion of $f(x)$ on the interval $(-l, l)$ has the form of the *cosine Fourier series*:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\frac{n\pi x}{l},$$

where the Fourier coefficients have the form

$$a_n = \frac{2}{l}\int_0^l f(x)\cos\frac{n\pi x}{l}\,dx \qquad (b_n = 0).$$

2°. Let $f(x)$ be an odd function, i.e., $f(x) = -f(-x)$. Then the Fourier expansion of $f(x)$ on the interval $(-l, l)$ has the form of the *sine Fourier series*:

$$f(x) = \sum_{n=1}^{\infty} b_n \sin\frac{n\pi x}{l},$$

where the Fourier coefficients have the form

$$b_n = \frac{2}{l}\int_0^l f(x)\sin\frac{n\pi x}{l}\,dx \qquad (a_n = 0).$$

Example. Let us find the Fourier expansion of the function $f(x) = x$ on the interval $(-\pi, \pi)$.
Taking $l = \pi$ and $f(x) = x$ in the formula for the Fourier coefficients and integrating by parts, we obtain

$$b_n = \frac{2}{\pi}\int_0^{\pi} x\sin(nx)\,dx = \frac{2}{\pi}\left(-\frac{1}{n}x\cos(nx)\Big|_0^{\pi} + \frac{1}{n}\int_0^{\pi}\cos(nx)\,dx\right) = -\frac{2}{n}\cos(n\pi) = (-1)^{n+1}\frac{2}{n}.$$

Therefore, the Fourier expansion of $f(x) = x$ has the form

$$f(x) = 2\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\sin(nx)}{n} \qquad (-\pi < x < \pi).$$

3°. If $f(x)$ is defined on the interval $(0, l)$ and satisfies the Dirichlet conditions, it can be represented by the cosine Fourier series, as well as the sine Fourier series (with the help of the above formulas).* Both series on the interval $(0, l)$ give the values of $f(x)$ at points of its continuity and the value $\frac{1}{2}[f(x_0 + 0) + f(x_0 - 0)]$ at points of its discontinuity; outside the interval $(0, l)$, these two series represent different functions.

8.4.2-3. Fourier series in complex form.

The complex Fourier expansion of a function $f(x)$ on an interval $(-l, l)$ has the form

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{i\omega_n x},$$

where

$$\omega_n = \frac{n\pi}{l}, \quad c_n = \frac{1}{2l}\int_{-l}^{l} f(x)e^{-i\omega_n x}\,dx; \quad n = 0, \pm1, \pm2, \ldots$$

The expressions $e^{i\omega_n x}$ are called *harmonics*, the coefficients c_n are *complex amplitudes*, ω_n are *wave numbers* of the function $f(x)$, and the set of all wave numbers $\{\omega_n\}$ is called the *discrete spectrum* of the function.

* The cosine Fourier expansion of $f(x)$ on the interval $(0, l)$ corresponds to the extension of $f(x)$ to the interval $(-l, 0)$ as an even function: $f(-x) = f(x)$. The sine Fourier expansion of $f(x)$ on $(0, l)$ corresponds to the extension of $f(x)$ to the interval $(-l, 0)$ as an odd function: $f(-x) = -f(x)$.

8.4.3. Criteria of Uniform and Mean-Square Convergence of Fourier Series

8.4.3-1. Criteria of uniform convergence of Fourier series.

LIPSCHITZ CRITERION. *The Fourier series of a function $f(x)$ converges uniformly to that function on an interval $[-l, l]$ if on a wider interval $[-L, L]$ $(-L < -l < l < L)$ the following inequality holds:*

$$|f(x_1) - f(x_2)| \le K|x_1 - x_2|^{\sigma} \quad \text{for all} \quad x_1, x_2 \in [L, L],$$

where K and σ are constants, $0 < \sigma \le 1$.

Corollary. The Fourier series of a continuous function $f(x)$ uniformly converges to that function on an interval $[-l, l]$ if on a wider interval the function $f(x)$ has a bounded derivative $f'(x)$.

DIRICHLET–JORDAN CRITERION. *Suppose that on an interval $(-l, l) \in [-L, L]$, a continuous function $f(x)$ has bounded variation (i.e., can be represented as the difference of two monotonically increasing functions). Then its Fourier series is uniformly convergent to that function on the interval $(-l, l)$.*

For any continuously differentiable $2l$-periodic function $f(x)$, its Fourier series [defined by formulas (8.4.2.1)–(8.4.2.2)] is uniformly convergent to $f(x)$.

8.4.3-2. Fourier series of square-integrable functions. Parseval identity.

1°. For a continuous 2π-periodic function $f(x)$, its Fourier series (8.4.1.1)–(8.4.1.2) converges to $f(x)$ in mean square, i.e.,

$$\int_{-\pi}^{\pi} \bigl|f(x) - f_n(x)\bigr|^2 \, dx \to 0 \quad \text{as} \quad n \to \infty,$$

where $f_n(x) = \frac{1}{2}a_0 + \sum\limits_{k=1}^{n}(a_k \cos kx + b_k \sin kx)$ is a partial sum of the Fourier series.

2°. If $f(x)$ is integrable on the segment $[-\pi, \pi]$ and the integral $\int_{-\pi}^{\pi} |f(x)|^2 \, dx$ exists as an improper integral with finitely many singularities, then the Fourier series (8.4.1.1)–(8.4.1.2) is mean-square convergent to $f(x)$.

3°. Let $f(x) \in L_2[-\pi, \pi]$ be a square-integrable function on the segment $[-\pi, \pi]$. Then its Fourier series (8.4.1.1)–(8.4.1.2) is mean-square convergent to $f(x)$, and the *Parseval identity* holds:

$$\frac{a_0^2}{2} + \sum_{n=1}^{\infty}(a_n^2 + b_n^2) = \frac{1}{\pi}\int_{-\pi}^{\pi} f^2(x)\, dx,$$

where a_n, b_n are defined by (8.4.1.2). Note that the functions considered in Items 1° and 2° belong to $L^2[-\pi, \pi]$.

8.4.4. Summation Formulas for Trigonometric Series

8.4.4-1. Summation of trigonometric series with the help of Laplace transforms.

When finding sums of trigonometric series, the following formulas may be useful:

$$\sum_{n=1}^{\infty} F(n)\sin(nx) = \frac{1}{2}\sin x \int_0^{\infty} \frac{f(t)\,dt}{\cosh t - \cos x}, \qquad 0 \le x \le \pi;$$

$$\sum_{n=1}^{\infty} F(n)\cos(nx) = \frac{1}{2}\int_0^{\infty} \frac{\cos x - e^{-t}}{\cosh t - \cos x} f(t)\,dt, \qquad 0 \le x \le \pi,$$

where

$$F(x) = \int_0^{\infty} f(t)e^{-xt}\,dt.$$

For specific $F(x)$, the corresponding functions $f(t)$ can be found in tables of inverse Laplace transforms (see Section T3.2).

8.4.4-2. Summation of series with the help of functions of complex variable.

Suppose that trigonometric series

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(nx) = f_1(x), \tag{8.4.4.1}$$

$$\sum_{n=1}^{\infty} a_n \sin(nx) = f_2(x) \tag{8.4.4.2}$$

have positive (nonnegative) coefficients a_n and the auxiliary series $\displaystyle\sum_{n=1}^{\infty} \frac{a_n}{n}$ is convergent. Then the series (8.4.4.1) and (8.4.4.2) are Fourier series representing continuous functions.

In order to find the sums $f_1(x)$ and $f_2(x)$ of the series (8.4.4.1) and (8.4.4.2), it is sometimes possible to use functions of complex variable. Let

$$\varphi(z) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n z^n$$

be the sum of a series that is convergent in the circle $|z| < 1$. If $\lim\limits_{|z|\to 1} \varphi(z) = \lim\limits_{r\to 1} \varphi(re^{ix}) = \varphi(e^{ix})$, then $f_1(x) + if_2(x) = \varphi(e^{ix})$. Thus, after separating the real and the imaginary parts of the function $\varphi(e^{ix})$, we get

$$f_1(x) = \mathrm{Re}[\varphi(e^{ix})], \qquad f_2(x) = \mathrm{Im}[\varphi(e^{ix})].$$

Example. Let us sum up the series

$$1 + \sum_{n=1}^{\infty} \frac{\cos(nx)}{n!} = f_1(x),$$

$$\sum_{n=1}^{\infty} \frac{\sin(nx)}{n!} = f_2(x).$$

Taking

$$\varphi(z) = 1 + \sum_{n=1}^{\infty} \frac{z^n}{n!} = e^z$$

and representing z in exponential form $z = re^{ix}$, let us pass to the limit as $r \to 1$. We find that

$$\exp(e^{ix}) = e^{\cos x + i \sin x} = e^{\cos x}[\cos(\sin x) + i \sin(\sin x)] = 1 + \sum_{n=1}^{\infty} \frac{(\cos x + i \sin x)^n}{n!}.$$

Using the Moivre formula $(\cos x + i \sin x)^n = \cos(nx) + i \sin(nx)$, we obtain

$$e^{\cos x} \cos(\sin x) + ie^{\cos x} \sin(\sin x) = 1 + \sum_{n=1}^{\infty} \frac{\cos(nx)}{n!} + i \sum_{n=1}^{\infty} \frac{\sin(nx)}{n!}.$$

It follows that

$$1 + \sum_{n=1}^{\infty} \frac{\cos(nx)}{n!} = e^{\cos x} \cos(\sin x),$$

$$\sum_{n=1}^{\infty} \frac{\sin(nx)}{n!} = e^{\cos x} \sin(\sin x).$$

8.5. Asymptotic Series

8.5.1. Asymptotic Series of Poincaré Type. Formulas for the Coefficients

8.5.1-1. Definition of asymptotic series. Illustrative example.

Suppose that for large x the function $f(x)$ can be represented in the form

$$f(x) = \sum_{k=0}^{n} \frac{a_k}{x^k} + o\big(x^{-n}\big) \quad \text{as} \quad x \to \infty. \tag{8.5.1.1}$$

Then one writes

$$f(x) \sim \sum_{k=0}^{\infty} \frac{a_k}{x^k}. \tag{8.5.1.2}$$

The infinite series on the right-hand side is called an *asymptotic series of Poincaré's type* (or an *asymptotic expansion*) of the function $f(x)$.

Asymptotic series may happen to be convergent or divergent.

Example. Consider the function defined by the integral

$$f(x) = \int_x^{\infty} e^{x-t} \frac{dt}{t}.$$

Repeated integration by parts yields

$$f(x) = \frac{1}{x} - \frac{1}{x^2} + \frac{2!}{x^3} - \cdots + (-1)^{n-1} \frac{(n-1)!}{x^n} + R_n(x), \tag{8.5.1.3}$$

where the remainder is defined by

$$R_n(x) = (-1)^n n! \int_x^{\infty} \frac{e^{x-t}}{t^{n+1}}\, dt.$$

The following estimate holds:

$$|R_n(x)| = n! \int_x^{\infty} \frac{e^{x-t}}{t^{n+1}}\, dt < n! \frac{1}{x^{n+1}} \int_x^{\infty} e^{x-t}\, dt = -\frac{n!}{x^{n+1}} e^{x-t}\Big|_{t=x}^{t=\infty} = \frac{n!}{x^{n+1}}.$$

Therefore, $R_n(x) = o\left(x^{-n}\right)$ and we can write

$$f(x) \sim \frac{1}{x} - \frac{1}{x^2} + \frac{2!}{x^3} - \cdots + (-1)^{n-1}\frac{(n-1)!}{x^n} + \cdots.$$

This asymptotic series is divergent, since for any fixed x we have

$$\lim_{n\to\infty} |a_n| = \lim_{n\to\infty} \frac{(n-1)!}{x^n} = \infty.$$

The motivation of using divergent asymptotic series of the form (8.5.1.2) is that a finite sum of the series (8.5.1.1) provides a good approximation for the given function for $x \to \infty$. For a fixed n, the accuracy of the approximations increases with the growth of x.

8.5.1-2. Uniqueness of an asymptotic series representing a function.

1. For a function $f(x)$ that admits the representation by asymptotic series (8.5.1.2), such a representation is unique and the coefficients a_n of the series are determined by

$$a_n = \lim_{x\to\infty} x^n \left[f(x) - \sum_{k=0}^{n-1} a_k x^{-k} \right].$$

2. There exist functions that are not identically equal to zero, but having all coefficients a_n in their asymptotic expansion (8.5.1.2) equal to zero. Such a function is called *asymptotic zero*. The class of asymptotic zeroes includes any function $f(x)$ such that for large x and any positive integer n the following estimate holds:

$$f(x) = o(x^{-n}).$$

For instance, the functions $f(x) = e^{-x}$, $f(x) = xe^{-2x}$, $f(x) = \exp(-x^2)$ belong to this class. Adding such functions to the left-hand side of the asymptotic expansion (8.5.1.2) does not change its right-hand side.

COROLLARY. *The coefficients of the asymptotic expansion of a function $f(x)$ do not uniquely determine the function.*

Remark. Addition of a suitable asymptotic zero can be used, in some cases, for the construction of approximations that not only correctly describe the behavior of a function for large x, but give fairly accurate numerical results for moderate values of x (i.e., $x = O(1)$; sometimes even for small x). This method is often used in applications, for instance in engineering, to obtain formulas applicable in a wide range of some characteristic parameter.

8.5.2. Operations with Asymptotic Series

8.5.2-1. Addition, subtraction, multiplication, and division of asymptotic series.

1. Asymptotic expansions of functions $f(x) \sim \sum\limits_{n=0}^{\infty} a_n x^{-n}$ and $g(x) \sim \sum\limits_{n=0}^{\infty} b_n x^{-n}$ admit term-by-term addition and subtraction:

$$f(x) \pm g(x) \sim \sum_{n=0}^{\infty} a_n x^{-n} \pm \sum_{n=0}^{\infty} b_n x^{-n} = \sum_{n=0}^{\infty} (a_n \pm b_n) x^{-n}.$$

2. Asymptotic expansions $f(x) \sim \sum\limits_{n=0}^{\infty} a_n x^{-n}$ and $g(x) \sim \sum\limits_{n=0}^{\infty} b_n x^{-n}$ can be formally multiplied, according to the Cauchy rule:

$$f(x)g(x) \sim \left(\sum_{n=0}^{\infty} a_n x^{-n}\right)\left(\sum_{n=0}^{\infty} b_n x^{-n}\right) = \sum_{n=0}^{\infty} c_n x^{-n}, \qquad c_n = \sum_{k=0}^{n} a_k b_{n-k}.$$

3. Asymptotic expansion $f(x) \sim \sum\limits_{n=0}^{\infty} a_n x^{-n}$ can be formally divided by $g(x) \sim \sum\limits_{n=0}^{\infty} b_n x^{-n}$, provided that $b_0 \neq 0$. The coefficients of the resulting asymptotic expansion $h(x) \sim \sum\limits_{n=0}^{\infty} b_n x^{-n}$ are found by the method of indefinite coefficients with the help of the relation $f(x) = g(x)h(x)$: the corresponding expansions should be inserted into this relation and then similar terms gathered.

8.5.2-2. Composition of series. Integration of asymptotic series.

1. Suppose that in a neighborhood of $y = 0$, the function $g(y)$ can be represented by a convergent power series, $g(y) = \sum\limits_{m=0}^{\infty} b_m y^m$, and for $x \to \infty$, the function $f(x)$ can be represented as an asymptotic series $f(x) \sim \sum\limits_{n=1}^{\infty} a_n x^{-n}$ with $a_0 = 0$. Then, for large enough x, the composite function

$$g\big(f(x)\big) = \sum_{m=0}^{\infty} b_m [f(x)]^m \tag{8.5.2.1}$$

makes sense and admits an asymptotic expansion, which can be obtained from (8.5.2.1) if $f(x)$ is replaced by its asymptotic expansion, after which similar terms should be gathered.

2. Suppose that the asymptotic expansion of a function $f(x)$ starts with $a_2 x^{-2}$ [i.e., $a_0 = a_1 = 0$ in (8.5.1.2)]. Then this asymptotic expansion admits term-by-term integration from x to ∞,

$$\int_x^{\infty} f(x)\,dx \sim \int_x^{\infty} \left(\sum_{n=2}^{\infty} \frac{a_n}{x^n}\right) dx = \sum_{n=2}^{\infty} \int_x^{\infty} \frac{a_n}{x^n}\,dx = \sum_{n=2}^{\infty} \frac{a_n}{(n-1)x^{n-1}}.$$

3. Term-by-term differentiation of asymptotic series is impossible, in general.

Example. To illustrate the last statement, consider the function $f(x) = e^{-x}\sin(e^x)$.

For all n, we have

$$\lim_{x\to\infty} x^n f(x) = 0,$$

and therefore the asymptotic expansion of this function has zero coefficients, $f(x) \sim 0$. Term-by-term differentiation of this relation also yields zero, $f'(x) \sim 0$.

On the other hand, the derivative of this function, $f'(x) = -e^{-x}\sin(e^x) + \cos(e^x)$, admits no asymptotic expansion at all, since there is no limit $\lim\limits_{x\to\infty} f'(x)$!

References for Chapter 8

Brannan, D., *A First Course in Mathematical Analysis*, Cambridge University Press, Cambridge, 2006.
Bromwich, T. J. I., *Introduction to the Theory of Infinite Series*, American Mathematical Society, Providence, Rhode Island, 2005.
Bronshtein, I. N. and Semendyayev, K. A., *Handbook of Mathematics, 4th Edition*, Springer-Verlag, Berlin, 2004.
Danilov, V. L., Ivanova, A. N., et al., *Mathematical Analysis (Functions, Limits, Series, Continued Fractions)* [in Russian], Fizmatlit Publishers, Moscow, 1961.
Davis, H. F., *Fourier Series and Orthogonal Functions*, Dover Publications, New York, 1989.
Fedoryuk, M. V. *Asymptotics, Integrals, Series* [in Russian], Nauka, Moscow, 1987.
Fichtenholz, G. M., *Functional Series*, Taylor & Francis, London, 1970.
Fikhtengol'ts (Fichtenholz), G. M., *A Course in Differential and Integral Calculus, Vols. 2 and 3* [in Russian], Nauka Publishers, Moscow, 1969.
Gradshteyn, I. S. and Ryzhik, I. M., *Tables of Integrals, Series, and Products, 6th Edition*, Academic Press, New York, 2000.
Hansen, E. R., *A Table of Series and Products*, Printice Hall, Englewood Cliffs, London, 1975.
Hirschman, I., *Infinite Series*, Greenwood Press, New York, 1978.
Hyslop, J. M., *Infinite Series*, Dover Publications, New York, 2006.
Jolley, L. B. W., *Summation of Series*, Dover Publications, New York, 1961.
Knopp, K., *Infinite Sequences and Series*, Dover Publications, New York, 1956.
Knopp, K., *Theory and Application of Infinite Series*, Dover Publications, New York, 1990.
Mangulis, V., *Handbook of Series for Scientists and Engineers*, Academic Press, New York, 1965.
Pinkus, A. and Zafrany, S., *Fourier Series and Integral Transforms*, Cambridge University Press, Cambridge, 1997.
Prudnikov, A. P., Brychkov, Yu. A., and Marichev, O. I., *Integrals and Series, Vol. 1, Elementary Functions*, Gordon & Breach, New York, 1986.
Zhizhiashvili, L.V., *Trigonometric Fourier Series and Their Conjugates*, Kluwer Academic, Dordrecht, 1996.
Zygmund, A., *Trigonometric Series, 3rd Edition*, Cambridge University Press, Cambridge, 2003.

Chapter 9
Differential Geometry

9.1. Theory of Curves

9.1.1. Plane Curves

9.1.1-1. Regular points of plane curve.

A plane curve Γ in a Cartesian coordinate system can be defined by equations in the following form:

Explicitly,

$$y = f(x). \tag{9.1.1.1}$$

Implicitly,

$$F(x, y) = 0.$$

Parametrically,

$$x = x(t), \quad y = y(t). \tag{9.1.1.2}$$

In vector form,

$$\mathbf{r} = \mathbf{r}(t), \quad \text{where} \quad \mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j}\,.$$

In a polar coordinate system, the curve is usually given by the equation

$$r = r(\varphi),$$

where the relationship between Cartesian and polar coordinates is given by formulas $x = r\cos\varphi$ and $y = r\sin\varphi$.

Remark. The explicit equation (9.1.1.1) can be obtained from the parametric equations (9.1.1.2) if the abscissa is taken for the parameter: $x = t$, $y = f(t)$.

A point $M(x(t), y(t))$ is said to be *regular* if the functions $x(t)$ and $y(t)$ have continuous first derivatives not simultaneously equal to zero in a sufficiently small neighborhood of this point. For implicitly defined functions, a point $M(x, y)$ is said to be *regular* if grad $F = \nabla F \neq 0$ at this point.

If a curve is given parametrically (9.1.1.2), then the *positive sense* is defined on this curve, i.e., the direction in which the point $M(x(t), y(t))$ of the curve moves as the parameter t increases. If the curve is given explicitly by (9.1.1.1), then the positive sense corresponds to the direction in which the abscissa increases (i.e., moves from left to right). In a polar coordinate system, the positive sense corresponds to the direction in which the angle φ increases (i.e., the positive sense is counterclockwise).

If s is the curve length from some constant point M_0 to M, then the infinitesimal length increment of the arc M_0M is approximately expressed by the formula for the arc length

differential ds; i.e., the following formulas hold:

$$\Delta s \approx ds = \sqrt{1 + (y'_x)^2}\, dx, \qquad \text{if the curve is given explicitly,}$$

$$\Delta s \approx ds = \sqrt{(x'_t)^2 + (y'_t)^2}\, dt, \qquad \text{if the curve is given parametrically,}$$

$$\Delta s \approx ds = \sqrt{r^2 + (r'_\varphi)^2}\, d\varphi, \qquad \text{for a curve in the polar coordinate system.}$$

Example 1. The arc length differential of the curve $y = \cos x$ has the form $ds = \sqrt{1 + \sin x}\, dx$.

Example 2. For the semicubical parabola $x = t^2$, $y = t^3$, the arc length differential is equal to $ds = t\sqrt{4 + 9t^2}\, dt$.

Example 3. For the *hyperbolic spiral* $r = a/\varphi$ for $r > 0$, the arc length differential is equal to $ds = a\sqrt{1 + \varphi^2}/\varphi^2\, d\varphi$.

9.1.1-2. Tangent and normal.

The *tangent* to a curve Γ at a regular point M_0 is defined to be the straight line that is the limit position of the secant M_0M_1 as the point M_1 approaches the point M_0; the *normal* is defined to be the straight line passing through M_1 and perpendicular to the tangent (Fig. 9.1).

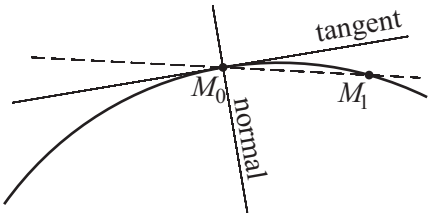


Figure 9.1. Tangent and normal.

At each regular point $M(x_0, y_0) = M(x(t_0), y(t_0))$, the curve Γ has a unique tangent given by one of the equations (depending on how the curve is defined)

$$y - y_0 = y'_x(x - x_0), \qquad \text{if the curve is given explicitly,}$$

$$F_x(x - x_0) + F_y(y - y_0) = 0, \qquad \text{if the curve is given implicitly,}$$

$$\frac{y - y_0}{y'_t} = \frac{x - x_0}{x'_t}, \qquad \text{if the curve is given parametrically,}$$

$$\mathbf{r} = \mathbf{r}_0 + \lambda \mathbf{r}_t, \qquad \text{if the curve is given in vector form,}$$

where $\mathbf{r}_0$ is the position vector of the point M_0, λ is an arbitrary parameter, and all derivatives are evaluated at $x = x_0$, $y = y_0$, and $t = t_0$.

The *slope of the tangent* is determined by the angle α between the positive direction of the OX-axis and the positive direction of the tangent (Fig. 9.2*a*). The slope of the tangent (and the angle α) is determined by the formulas

$$\tan\alpha = y'_x = -\frac{F_x}{F_y} = \frac{y'_t}{x'_t}.$$

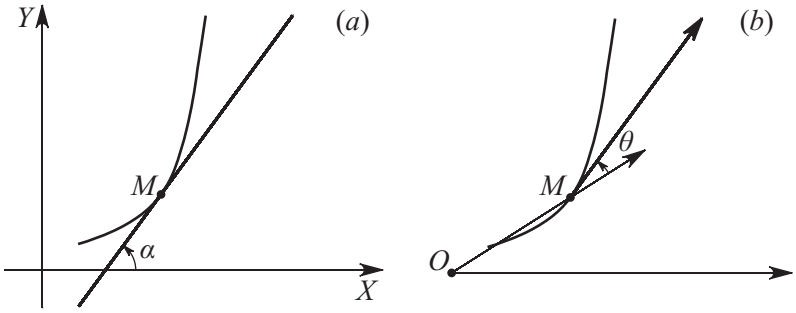


Figure 9.2. Slope of the tangent.

If a curve is given in the polar coordinate system, then the slope of the tangent is determined by the angle θ between the direction of the position vector $r = OM$ and the positive direction of the tangent (Fig. 9.2*b*). The angle θ is determined by the formula

$$\tan\theta = \frac{r}{\varphi'_r}.$$

The normal at each regular point $M(x_0, y_0) = M(x(t_0), y(t_0))$ is given, depending on the method for defining the curve, by the equations

$$y - y_0 = -\frac{x - x_0}{y'_x}, \qquad \text{if the curve is given explicitly,}$$

$$\frac{x - x_0}{F_x} = \frac{y - y_0}{F_y}, \qquad \text{if the curve is given implicitly,}$$

$$(y - y_0)y'_t + (x - x_0)x'_t = 0, \qquad \text{if the curve is given parametrically,}$$

where all the derivatives are evaluated at $x = x_0$, $y = y_0$, and $t = t_0$.

The *positive sense* of the tangent coincides with the positive sense of the curve at the point of tangency; and the *positive sense* of the normal can in some way be made consistent with the positive sense of the tangent; for example, it can be obtained from the positive sense of the tangent by counterclockwise rotation around M by an angle of 90° (Fig. 9.3). The point M divides the tangent and the normal into positive and negative half-lines.

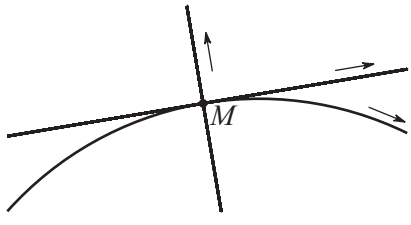


Figure 9.3. Positive sense of tangent.

Example 4. Let us find the equations of the tangent and the normal to the parametrically given semicubical parabola $x = t^2$, $y = t^3$ at the point $M_0(1, 1)$, $t = 1$.

The equation of the tangent,

$$\frac{y - t^3}{3t^2} = \frac{x - t^2}{2t} \quad \text{or} \quad y = \frac{3}{2}tx - \frac{1}{2}t^3,$$

at the point $M_0(1, 1)$ is

$$y = \frac{3}{2}x - \frac{1}{2}.$$

The equation of the normal,

$$2t(x - t^2) + 3t^2(y - t^3) = 0 \quad \text{or} \quad 2x + 3ty = t^2(2 + 3t^2),$$

at the point $M_0(1, 1)$ is (Fig. 9.4*a*)

$$2x + 3y = 5.$$

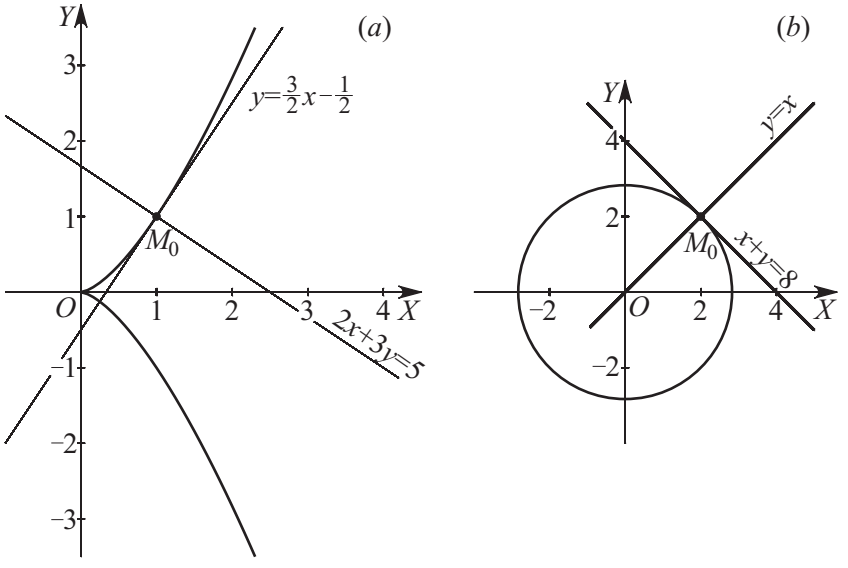


Figure 9.4. Tangents and normals to the semicubical parabola (*a*) and to the circle (*b*).

Example 5. Let us find the equation of the tangent and the normal to the circle $x^2 + y^2 = 8$ at the point $M_0(2, 2)$.

We write the equation of the circle as $F(x, y) = 0$:

$$x^2 + y^2 - 8 = 0,$$

i.e., $F(x, y) = x^2 + y^2 - 8$. Obviously, we obtain

$$F_x = 2x, \quad F_y = 2y.$$

The equation of the tangent is

$$2x_0(x - x_0) + 2y_0(y - y_0) = 0,$$

or, taking into account the original equation of the circle,

$$xx_0 + yy_0 = 8.$$

At the point $M_0(2, 2)$, we have

$$x + y = 4.$$

The equation of the normal is

$$\frac{x - x_0}{2x_0} = \frac{y - y_0}{2y_0},$$

or

$$y = \frac{y_0}{x_0}x.$$

At the point $M_0(2, 2)$ (Fig. 9.4*b*), we have

$$y = x.$$

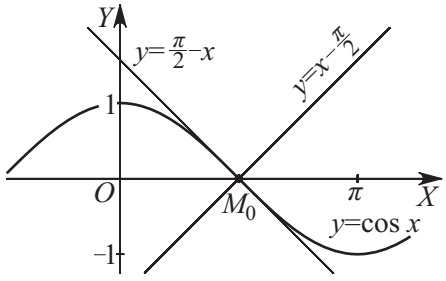


Figure 9.5. The tangent and the normal to the curve $y = \cos x$.

Example 6. Let us find the equations of the tangent and the normal to the curve $y = \cos x$ at the point $M_0(\pi/2, 0)$.

The equation of the tangent is

$$y - \cos x_0 = -\sin x_0(x - x_0) \quad \text{or} \quad y = \cos x_0 - \sin x_0(x - x_0).$$

At the point $M_0(\pi/2, 0)$, we have

$$y = \frac{\pi}{2} - x.$$

The equation of the normal is

$$y - \cos x_0 = -\frac{x - x_0}{-\sin x_0} \quad \text{or} \quad y = \cos x_0 + \frac{x - x_0}{\sin x_0}.$$

At the point $M_0(\pi/2, 0)$ (Fig. 9.5), we have

$$y = x - \frac{\pi}{2}.$$

9.1.1-3. Singular points.

A point is said to be *singular* if it is not regular.

Implicit equations of the form $F(x, y) = 0$ are used as a rule to find singular points of a curve and analyze their character. At any singular point $M_0(x_0, y_0)$, both partial derivatives of the function $F(x, y)$ are zero:

$$F_x(x_0, y_0) = 0 \quad \text{and} \quad F_y(x_0, y_0) = 0.$$

If both first partial derivatives are zero at M_0 and simultaneously at least one of the second derivatives F_{xx}, F_{xy}, and F_{yy} is nonzero, then M_0 is called a *double point*. This is the most widely known case of singular points. If both first partial derivatives and simultaneously all second partial derivatives are zero at M_0 but not all third partial derivatives are zero at M_0, then the point M_0 is said to be *triple*. In general, if all partial derivatives of $F(x, y)$ up to order $n - 1$ inclusive are zero at M_0 but at least one of the nth derivatives is nonzero at M_0, then the point M_0 is called an *n-fold singular point*. At an n-fold singular point M_0, the curve has n tangents, some of which may coincide or be imaginary. For example, for a double singular point M_0, the slopes $\lambda = y'_x$ of the two tangents at this point are the roots of the quadratic equation

$$F_{yy}(x_0, y_0)\lambda^2 + 2F_{xy}(x_0, y_0)\lambda + F_{xx}(x_0, y_0) = 0. \tag{9.1.1.3}$$

The roots of equation (9.1.1.3) depend on the sign of the expression

$$\Delta = \begin{vmatrix} F_{xx} & F_{xy} \\ F_{yx} & F_{yy} \end{vmatrix} = F_{xx}F_{yy} - F_{xy}^2,$$

where the second derivatives are evaluated at the point $M_0(x_0, y_0)$.

If $\Delta > 0$, then the roots of the quadratic equation (9.1.1.3) are complex conjugate. In this case, a sufficiently small neighborhood of $M_0(x_0, y_0)$ does not contain any other points of the curve except for M_0 itself. Such a point is called an *isolated point*.

Example 7. The curve

$$y^2 + 4x^2 - x^4 = 0$$

has the isolated point $(0, 0)$ (Fig. 9.6*a*).

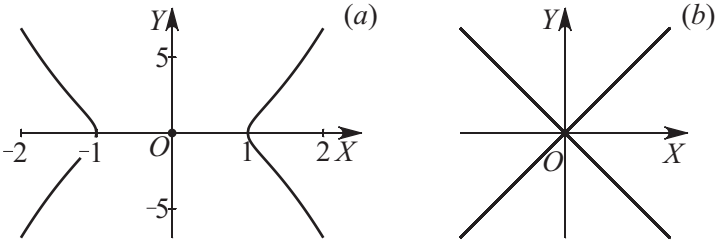


Figure 9.6. Examples of the isolated point (*a*) and the node (*b*).

If $\Delta < 0$, then the quadratic equation (9.1.1.3) has two distinct real roots. In this case, there are two branches of the curve passing through the point $M_0(x_0, y_0)$; these branches have distinct tangents whose directions are just determined by equation (9.1.1.3). Such a point is called a *node (a point of self-intersection)*.

Example 8. The point $(0, 0)$ of the curve

$$y^2 - x^2 = 0$$

is the node $(0, 0)$ (Fig. 9.6*b*).

If $\Delta = 0$, then the roots of the quadratic equation (9.1.1.3) coincide. In this case, the singular point of the curve is either isolated or characterized by the fact that all branches approaching the singular point M_0 have a common tangent at this point:

1. *Cusps of the first kind* are points approached by two branches of the curve that have a common tangent at this point and lie on the same side of the common normal and on opposite sides of the common tangent.
2. *Cusps of the second kind* are points approached by two branches of the curve that have a common tangent at this point and lie on the same side of the common normal and on the same side of the common tangent.
3. *Points of osculation* are points at which the curve is tangential to itself.

Example 9. For the curve (*cissoid of Diocles*)

$$(2a - x)y^2 - x^3 = 0,$$

shown in Fig. 9.7*a*, the origin is a cusp of the first kind, which is clear from the explicit equation

$$y = \pm\sqrt{\frac{x^3}{2a - x}}$$

of the curve. For $x < 0$, no y satisfy the equation, and for $x > 0$ the values $\pm y$ lie on opposite sides of the tangent $x = 0$ at the origin.

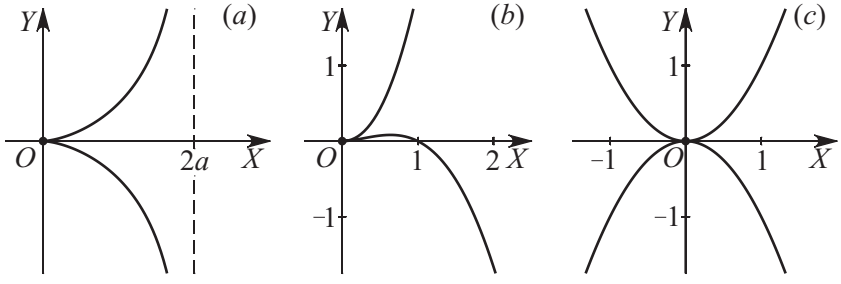


Figure 9.7. Examples of a cusp of the first kind (*a*), a cusp of the second kind (*b*), and an osculation point (*c*).

Example 10. For the curve (Fig. 9.7*b*)

$$(y^2 - x^2)^2 - x^5 = 0,$$

the origin is a cusp of the second kind, which easily follows from equation

$$y = x^2 \pm x^{5/2}$$

of the curve. It is also obvious that the curve consists of two branches tangent to the axis OX at the origin, and for $0 < x < 1$ the value of y is positive for both branches.

Example 11. The curve (Fig. 9.7*c*)

$$y^2 - x^4 = 0$$

has an osculation point at the origin.

Remark 1. If all the second partial derivatives are zero at the point M_0, i.e., $F_{xx} = F_{xy} = F_{yy} = 0$, then more than two branches of the curve can pass through this point. For example, for the *trefoil*

$$(x^2 + y^2)^2 - ax(x^2 - y^2) = 0,$$

three branches with tangents $x = 0$ and $x \pm y = 0$ pass through the origin (Fig. 9.8).

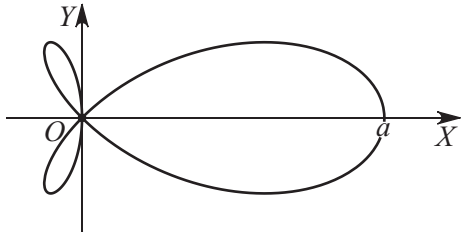


Figure 9.8. The trefoil.

Remark 2. If the equation $F(x, y) = 0$ does not contain constant terms and terms of degree 1, then the origin is a double point. The equation of the tangent at a double point can readily be obtained by equating all terms of degree 2 with zero. For example, for the cissoid of Diocles (Example 9), the equation of the tangent $x = 0$ follows from the equation $-xy^2 - x^3 = 0$. If the equation $F(x, y) = 0$ does not contain constant terms and terms of degrees 1 and 2, then the origin is a triple point, etc.

Along with the singular points listed above, there are many other singular points with specific names:

1. *Break points* are points at which the curve changes its direction by a "jump" and, in contrast to cusps, the tangents to both parts of the curve are distinct (Fig. 9.9*a*).
2. *Termination points* are points at which the curve terminates (Fig. 9.9*b*).
3. *Asymptotic points* are points around which the curve winds infinitely many times while infinitely approaching them (Fig. 9.9*c*).

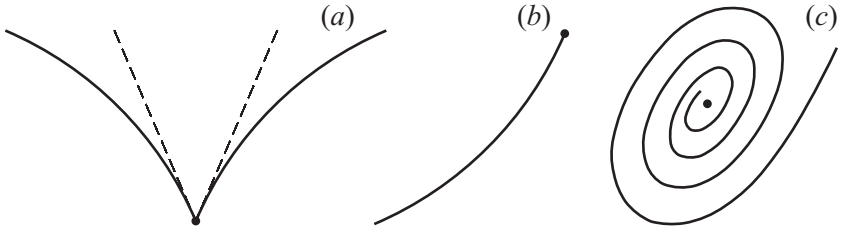


Figure 9.9. The break point (*a*), the termination point (*b*), and the asymptotic point (*c*).

Remark. A break point corresponds to a jump discontinuity of the derivative dy/dx. Termination points correspond to either a jump discontinuity or termination of the function $y = f(x)$. Asymptotic points can be found most easily in curves given in the polar coordinate system.

9.1.1-4. Asymptotes.

A straight line is called an *asymptote* of a curve Γ if the distance from a point $M(x, y)$ of the curve to this straight line tends to zero as $x^2 + y^2 \to \infty$. The limit position of the tangent to a regular point of the curve is an asymptote; the converse assertion is generally not true.

For a curve given explicitly as $y = f(x)$, vertical asymptotes are determined as points of discontinuity of the function $y = f(x)$, while horizontal and skew asymptotes have the form $y = kx + b$, where

$$k = \lim_{t\to\infty} \frac{f(x)}{x}, \quad b = \lim_{t\to\infty} [f(x) - kx];$$

both limits must be finite.

Example 12. Let us find the asymptotes of the curve $y = x^3/(x^2 + 1)$. Since the limits

$$k = \lim_{t\to\infty} \frac{f(x)}{x} = \lim_{t\to\infty} \frac{x^3}{x(x^2+1)} = 1,$$
$$b = \lim_{t\to\infty} [f(x) - kx] = \lim_{t\to\infty} \left(\frac{x^3}{x^2+1} - x \right) = \lim_{t\to\infty} \left(\frac{-x}{x^2+1} \right) = 0$$

exist, the asymptote is given by the equation $y = x$.

To find an asymptote of a parametrically defined curve $x = x(t)$, $y = y(t)$, one should find the values $t = t_i$ for which $x(t) \to \infty$ or $y(t) \to \infty$.

If

$$x(t_i) = \infty \quad \text{but} \quad y(t_i) = c \neq \infty,$$

then the straight line $y = c$ is a horizontal asymptote. If

$$y(t_i) = \infty \quad \text{but} \quad x(t_i) = a \neq \infty,$$

then the straight line $x = a$ is a vertical asymptote. If

$$x(t_i) = \infty \quad \text{and} \quad y(t_i) = \infty,$$

then one should calculate the following two limits:

$$k = \lim_{t\to t_i} \frac{y(t)}{x(t)} \quad \text{and} \quad b = \lim_{t\to t_i} [y(t) - kx(t)]. \tag{9.1.1.4}$$

If both limits exist, then the curve has the asymptote $y = kx + b$.

Example 13. Let us find the asymptote of the *Folium of Descartes*

$$x = \frac{3at}{t^3+1}, \quad y = \frac{3at^2}{t^3+1} \quad (-\infty \le t \le \infty).$$

Since $x(-1) = \infty$ and $y(-1) = \infty$, one should use formulas (9.1.1.4),

$$k = \lim_{t\to-1} \frac{y(t)}{x(t)} = \lim_{t\to-1} \frac{\dfrac{3at^2}{t^3+1}}{\dfrac{3at}{t^3+1}} = \lim_{t\to-1} t = -1,$$
$$b = \lim_{t\to-1} [y(t) - kx(t)] = \lim_{t\to-1} \left(\frac{3at^2}{t^3+1} + \frac{3at}{t^3+1} \right) = \lim_{t\to-1} \frac{3at(t+1)}{(t+1)(t^2-t+1)} = -a,$$

which imply that the asymptote is given by the equation $y = -x - a$.

Suppose that the function $F(x, y)$ in the equation $F(x, y) = 0$ is a polynomial in the variables x and y. We choose the terms of the highest order in $F(x, y)$. By $\Phi(x, y)$ we denote the set of highest-order terms and solve the equation for the variables x and y:

$$x = \varphi(y), \quad y = \psi(x).$$

The values $y_i = c$ for which $x = \infty$ give the horizontal asymptotes $y = c$; the values $x_i = a$ for which $y = \infty$ give the horizontal asymptotes $x = a$.

To find skew asymptotes, one should substitute the expression $y = kx + b$ into $F(x, y)$. We write the resulting polynomial $F(x, kx + b)$ as

$$F(x, kx + b) = f_1(k, b)x^n + f_2(k, b)x^{n-1} + \ldots$$

If the system of equations

$$f_1(k, b) = 0, \quad f_2(k, b) = 0$$

is consistent, then its solutions k, b are the parameters of the asymptotes $y = kx + b$.

9.1.1-5. Osculating circle.

The *osculating circle (circle of curvature)* of a curve Γ at a point M_0 is defined to be the limit position of the circle passing through M_0 and two neighboring points M_1 and M_2 of the curve as $M_1 \to M_0$ and $M_2 \to M_0$ (Fig. 9.10).

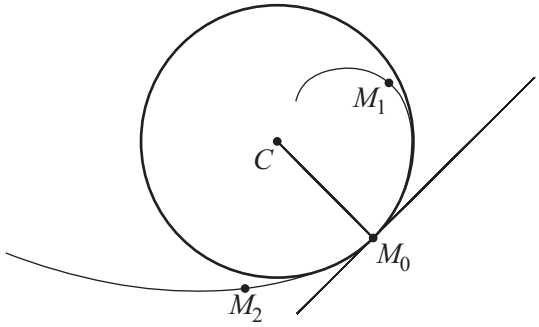


Figure 9.10. The osculating circle.

The center of this circle (the *center of curvature* of the curve Γ at the point M_1) is called the center of the osculating circle and lies on the normal to this curve (Fig. 9.10). The coordinates of the center of curvature can be found by the following formulas:

for a curve defined explicitly,

$$x_c = x_0 - \frac{y'_x\left[1 + (y'_x)^2\right]}{y''_{xx}}, \quad y_c = y_0 + \frac{1 + (y'_x)^2}{y''_{xx}};$$

for a curve defined implicitly,

$$x_c = x_0 - \frac{F_x\left(F_x{}^2 + F_y{}^2\right)}{2F_{xy}F_xF_y - F_x^2F_{yy} - F_y^2F_{xx}}, \quad y_c = y_0 + \frac{F_y\left(F_x{}^2 + F_y{}^2\right)}{2F_{xy}F_xF_y - F_x^2F_{yy} - F_y^2F_{xx}};$$

for a curve defined parametrically,

$$x_c = x_0 - \frac{y'_t\left[(x'_t)^2 + (y'_t)^2\right]}{x'_t y''_{tt} - y'_t x''_{tt}}, \quad y_c = y_0 + \frac{x'_t\left[(x'_t)^2 + (y'_t)^2\right]}{x'_t y''_{tt} - y'_t x''_{tt}};$$

for a curve in polar coordinates,

$$x_c = r_0\cos\varphi_0 - \frac{\left[r_0^2 + \left(r'_\varphi\right)^2\right]\left(r_0\cos\varphi_0 + r'_\varphi\sin\varphi_0\right)}{r_0^2 + 2\left(r'_\varphi\right)^2 - r_0 r''_{\varphi\varphi}},$$

$$y_c = r_0\sin\varphi_0 - \frac{\left[r_0^2 + \left(r'_\varphi\right)^2\right]\left(r_0\sin\varphi_0 - r'_\varphi\cos\varphi_0\right)}{r_0^2 + 2\left(r'_\varphi\right)^2 - r_0 r''_{\varphi\varphi}},$$

$$x_0 = r_0\cos\varphi_0, \quad y_0 = r_0\sin\varphi_0,$$

where all derivatives are evaluated at $x = x_0$, $y = y_0$, $t = t_0$, and $\varphi = \varphi_0$.

The radius of the osculating circle is called the *radius of curvature* of the curve at the point $M_0(x_0, y_0)$; its length in a Cartesian coordinate system is

$$\rho = \frac{\left[1 + (y'_x)^2\right]^{3/2}}{\left|y''_{xx}\right|} = \frac{\left(F_x{}^2 + F_y{}^2\right)^{3/2}}{\left|2F_{xy}F_xF_y - F_x^2F_{yy} - F_y^2F_{xx}\right|} = \frac{\left[\left(x'_t\right)^2 + \left(y'_t\right)^2\right]^{3/2}}{\left|x'_t y''_{tt} - y'_t x''_{tt}\right|},$$

and in the polar coordinate system it is

$$\rho = \frac{\left[r_0^2 + \left(r'_\varphi\right)^2\right]^{3/2}}{\left|r_0^2 + 2\left(r'_\varphi\right)^2 - r_0 r''_{\varphi\varphi}\right|},$$

where all derivatives are evaluated at $x = x_0$, $y = y_0$, $t = t_0$, and $\varphi = \varphi_0$.

9.1.1-6. Curvature of plane curves.

The limit ratio of the tangent rotation angle $\Delta\varphi$ to the corresponding arc length Δs of the curve Γ as $\Delta s \to 0$ (Fig. 9.11),

$$k = \lim_{\Delta s\to 0}\left|\frac{\Delta\varphi}{\Delta s}\right|,$$

is called the *curvature* of Γ at the point M_1.

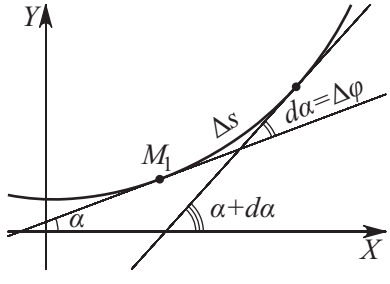


Figure 9.11. The curvature of the curve.

The curvature and the radius of curvature are reciprocal quantities,

$$k = \frac{1}{\rho}\,.$$

The more bent a curve is near a point, the larger k is and the smaller ρ is at this point. For a circle of radius a, the radius of curvature is $\rho = a$ and the curvature is $k = 1/a$ (they are constant at all points of the circle); for a straight line, $\rho = \infty$ and $k = 0$; for all other curves, the curvature varies from point to point.

Remark. All points of inflection are points of zero curvature.

9.1.1-7. Frénet formulas.

To each point M of a plane curve, one can naturally assign a local coordinate system. The role of the origin O is played by point M itself, and the role of the axes OX and OY are played by the tangent and normal at this point. The unit tangent and normal vectors to the curve are usually denoted by **t** and **n**, respectively (Fig. 9.12).

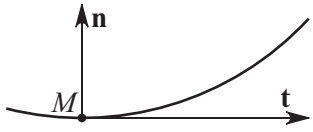


Figure 9.12. The unit tangent **t** and normal **n** vectors to the curve.

Suppose that the arc length is taken as a (natural) parameter on the curve:

$$\mathbf{r} = \mathbf{r}(s);$$

then the *Frénet formulas*

$$\mathbf{t}'_s = k\mathbf{n}, \quad \mathbf{n}'_s = -k\mathbf{t}$$

hold, where k is the curvature of the curve.

With first-order accuracy, the Frénet formulas determine the rotation of the vectors **t** and **n** when translated along the curve to a close point, $s \to s + \Delta s$.

9.1.1-8. Envelope of a family of curves.

A *one-parameter family of curves* is the set of curves defined by the equation

$$F(x, y, C) = 0, \tag{9.1.1.5}$$

which is called the *equation of the family*. Here C is a parameter varying in a certain range, $C_1 \le C \le C_2$; in particular, the range can be $-\infty \le C + \infty$.

A curve that is tangent at each point to some curve of a one-parameter family of curves (9.1.1.5) is called the *envelope* of the family. The point of tangency of the envelope to a curve of the family is called a *characteristic point* of the curve of the family (Fig. 9.13).

Figure 9.13. The envelope of the family of curves.

The equation of the envelope of a one-parameter family is obtained by elimination of the parameter C from the system of equations

$$\begin{aligned} &F(x, y, C) = 0, \\ &\frac{\partial F(x, y, C)}{\partial C} = 0, \end{aligned} \tag{9.1.1.6}$$

which determines the *discriminant curve* of this family.

The discriminant curve (9.1.1.6) of a one-parameter family is an envelope if it does not consist of singular points of the curves.

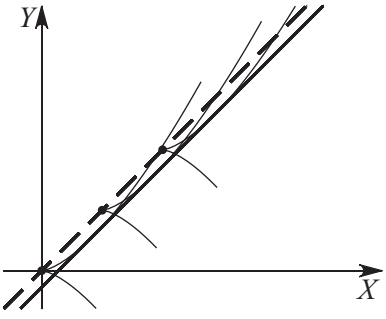


Figure 9.14. The envelope of the family of semicubical parabolas.

Example 14. Consider the family of semicubical parabolas (Fig. 9.14)

$$3(y - C)^2 - 2(x - C)^3 = 0.$$

Differentiating with respect to the parameter C, we obtain

$$y - C - (x - C)^2 = 0.$$

Solving the system

$$\begin{aligned} &3(y - C)^2 - 2(x - C)^3 = 0, \\ &y - C - (x - C)^2 = 0, \end{aligned}$$

we obtain

$$\begin{aligned} &x - C = 0, \quad y - C = 0; \\ &x - C = \frac{2}{3}, \quad y - C = \frac{4}{9}. \end{aligned}$$

Eliminating the parameter C, we see that the discriminant curve splits into the pair of straight lines $x = y$ and $x - y = 2/9$. Only the second of these two straight lines is an envelope, because the first straight line is the locus of singular points.

9.1.1-9. Evolute and evolvent.

The locus of centers of curvature of a curve is called its *evolute*. If a curve is defined via a natural parameter s, then the position vector of a point of the evolute of a plane curve $\mathbf{p}$ can be expressed in terms of the radius vector $\mathbf{r}$ of a point of this curve, the normal vector $\mathbf{n}$, and its radius of curvature ρ as follows:

$$\mathbf{p} = \mathbf{r} + \rho\mathbf{n}.$$

To obtain the equation of the evolute, it also suffices to treat x_c and y_c in the equations for the coordinates of the center of curvature as the current coordinates of the evolute.

Geometric properties of the evolute:

1. The normals to the original curve coincide with the tangents to the evolute at the corresponding points.
2. If the radius of curvature ρ varies monotonically on a given part of the curve, then its increment is equal to the distance passed by the center of curvature along the evolute.
3. At the points of extremum of the radius of curvature ρ, the evolute has a cusp of the first kind.
4. Since the radius of curvature ρ is always positive, any point of the evolute lies on a normal to the curve on the concave side.
5. The evolute is the envelope of the family of normals to the original curve.

A curve that intersects all curves of a family at the right angle is called an *orthogonal trajectory* of a one-parameter family of curves. A trajectory orthogonal to tangents to a given curve is called an *evolvent* of this curve.

If a curve is defined via its natural parameter s, then the vector equation of its evolvent has the form

$$\mathbf{p} = \mathbf{r} + (s_0 - s)\mathbf{t},$$

where s_0 is an arbitrary constant.

Basic properties of the evolvent:

1. The tangent to the original curve at each point is the normal to the evolvent at the corresponding point.
2. The distance between the corresponding points of two evolvents of a given curve is constant.
3. For $s = s_0$, the evolvent has cusps of the first kind.

The evolute and the evolvent are related to each other. The original curve is the evolvent of its evolute. The converse assertion is also true; i.e., the original curve is the evolute of its evolvent. The normal to the evolvent is tangent to the evolute.

9.1.2. Space Curves

9.1.2-1. Regular points of space curve.

A space curve Γ is in general determined parametrically or in vector form by the equations

$$x = x(t), \quad y = y(t), \quad z = z(t) \quad \text{or} \quad \mathbf{r} = \mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k},$$

where $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are the unit vectors (see (4.5.2.2)), t is an arbitrary parameter ($t \in [t_1, t_2]$), and t_1 and t_2 can be $-\infty$ and $+\infty$, respectively.

A point $M(x(t), y(t), z(t))$ is said to be *regular* if the functions $x(t)$, $y(t)$, and $z(t)$ have continuous first derivatives in a sufficiently small neighborhood of this point and these derivatives are not simultaneously zero, i.e., if $d\mathbf{r}/dt \neq 0$.

If the functions $x(t)$, $y(t)$, and $z(t)$ have continuous derivatives with respect to t and $d\mathbf{r}/dt \neq 0$ for all $t \in [t_1, t_2]$, then Γ is a *regular arc*.

For the parameter t it is convenient to take the arc length s, that is, the length of the arc from a point $M_0(x(t_0), y(t_0), z(t_0))$ to $M_1(x(t_1), y(t_1), z(t_1))$,

$$s = \int_\Gamma ds = \int_\Gamma \sqrt{d\mathbf{r} \cdot d\mathbf{r}} = \int_{t_0}^{t_1} \sqrt{(x_t')^2 + (y_t')^2 + (z_t')^2}\, dt. \qquad (9.1.2.1)$$

The sign of ds is chosen arbitrarily, and it determines the positive sense of the curve and the tangent.

A space curve can also be defined as the intersection of two surfaces

$$F_1(x, y, z) = 0, \quad F_2(x, y, z) = 0. \tag{9.1.2.2}$$

For a curve determined as the intersection of two planes, a point M_0 is regular if the vectors ∇F_1 and ∇F_2 are not linearly dependent at this point.

9.1.2-2. Tangents and normals.

A straight line is called the *tangent* to a curve Γ at a regular point M_0 if it is the limit position of the secant passing through M_0 and a point M_1 infinitely approaching the point M_0.

At a regular point M_0, the equation of the tangent has the form

$$\mathbf{r} = \mathbf{r}_0 + \lambda \mathbf{r}'_t(t_0), \tag{9.1.2.3}$$

where λ is a variable parameter.

Eliminating the parameter λ from (9.1.2.3), we obtain the canonical equation

$$\frac{x - x_0}{x'_t(t_0)} = \frac{y - y_0}{y'_t(t_0)} = \frac{z - z_0}{z'_t(t_0)}$$

of the tangent.

The equation of the tangent at a point M_0 of the curve Γ obtained as the intersection (9.1.2.2) of two planes is

$$\frac{x - x_0}{(F_1)_y(F_2)_z - (F_1)_z(F_2)_y} = \frac{y - y_0}{(F_1)_z(F_2)_x - (F_1)_x(F_2)_z} = \frac{z - z_0}{(F_1)_x(F_2)_y - (F_1)_y(F_2)_x},$$

where all derivatives are evaluated at $x = x_0$ and $y = y_0$.

A perpendicular to the tangent at the point of tangency is called a *normal* to a space curve. Obviously, at each point of the curve, there are infinitely many normals that form the plane perpendicular to the tangent.

The plane passing through the point of tangency and perpendicular to the tangent is called the *normal plane* (Fig. 9.15).

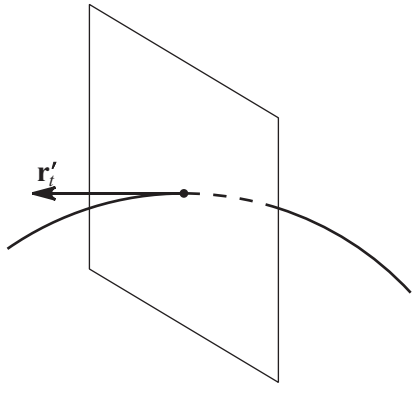


Figure 9.15. The normal plane.

At a regular point M_0, the equation of the normal plane has the form

$$(x-x_0)x_t'(t_0)+(y-y_0)y_t'(t_0)+(z-z_0)z_t'(t_0)=0 \quad \text{or} \quad (\mathbf{r}-\mathbf{r}_0)\cdot\mathbf{r}_t'(t_0)=0, \tag{9.1.2.4}$$

and for the curve Γ obtained as the intersection (9.1.2.2) of two planes, we have

$$\begin{vmatrix} x-x_0 & y-y_0 & z-z_0 \\ (F_1)_x & (F_1)_y & (F_1)_z \\ (F_2)_x & (F_2)_y & (F_2)_z \end{vmatrix} = 0, \tag{9.1.2.5}$$

where all derivatives are evaluated at $x=x_0$ and $y=y_0$.

1°. Consider a curve defined via its natural parameter s, $\mathbf{r}=\mathbf{r}(s)$. Unlike the derivatives with respect to an arbitrary parameter, the derivatives with respect to s will be denoted by primes. Consider the vectors $\mathbf{r}_s'$ and $\mathbf{r}_{ss}''$. It follows from (9.1.2.1) that $|\mathbf{r}_s'|=1$. Thus the first derivative with respect to the natural parameter s of the position vector of a point on a curve is the unit vector tangent to the curve.

2°. The second derivative with respect to the natural parameter s of the position vector of a point of a curve is equal to the first derivative of the unit vector $\mathbf{r}_s'$, i.e., of a vector of constant length, and hence it is perpendicular to this vector. But since the vector of the first derivative is tangent to the curve, the vector of the second derivative with respect to the natural parameter s is normal to the curve. This normal is called the *principal normal* to the curve.

At a regular point M_0, the equation of the principal normal has the form

$$\frac{x-x_0}{y_t'n-z_t'm}=\frac{y-y_0}{z_t'l-x_t'n}=\frac{z-z_0}{x_t'm-y_t'l} \quad \text{or} \quad \mathbf{r}=\mathbf{r}_0+\lambda\mathbf{r}_t'\times(\mathbf{r}_t'\times\mathbf{r}_{tt}''),$$
$$l=y_t'z_{tt}''-y_{tt}''z_t', \quad m=z_t'x_{tt}''-z_{tt}''x_t', \quad n=x_t'y_{tt}''-x_{tt}''y_t', \tag{9.1.2.6}$$

where all derivatives are evaluated at $t=t_0$.

9.1.2-3. Osculating plane.

Any plane passing through the tangent line to a curve is called a *tangent plane*. A tangent plane passing through a principal normal to the curve is called the *osculating plane*.

A curve has exactly one osculating plane at each of its points (assuming that the vectors $\mathbf{r}_t'$ and $\mathbf{r}_{tt}''$ are linearly independent). This plane passes through the vectors $\mathbf{r}_t'$ and $\mathbf{r}_{tt}''$ drawn from the point of tangency. The osculating plane is independent of the choice of the parameter t on the curve.

Remark. If t is time and $\mathbf{r}=\mathbf{r}(t)$ is an equation of motion, then the vector $\mathbf{r}_{tt}''$ is called the *acceleration vector* of a moving point. The acceleration vector always lies in the osculating plane of the trajectory of a moving point.

Since the family of all planes in space depends on three parameters and the position of a plane is determined by three noncollinear points, the osculating plane can be determined as the limit position of the plane passing through three points of the curve that infinitely approach one another (Fig. 9.16).

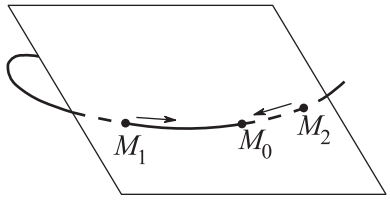


Figure 9.16. The osculating plane.

The equation of the osculating plane at the point M_0 has the form

$$\begin{vmatrix} x-x_0 & y-y_0 & z-z_0 \\ x'_t(t_0) & y'_t(t_0) & z'_t(t_0) \\ x''_{tt}(t_0) & y''_{tt}(t_0) & z''_{tt}(t_0) \end{vmatrix} = 0 \quad \text{or} \quad \left[(\mathbf{r}-\mathbf{r}_0)\mathbf{r}'_t(t_0)\mathbf{r}''_{tt}(t_0)\right] = 0.$$

This equation becomes meaningless for the points of the curve at which

$$\mathbf{r}'_t(t_0) \times \mathbf{r}''_{tt}(t_0) = 0\,. \tag{9.1.2.7}$$

The points satisfying (9.1.2.7) are called *points of rectification*. The osculating plane at these points is undefined and, in what follows, we exclude them from consideration together with singular points of the curve.

Remark. This is inadmissible for a curve entirely consisting of points of rectification, because a curve consisting of points of rectification is a straight line.

9.1.2-4. Moving trihedral of curve.

The normal perpendicular to the osculating plane is called the *binormal*.

At each point of a curve, the tangent, the principal normal, and the binormal determine a trihedral with three right angles at the vertex, which lies on the curve. This trihedral is called the *moving* or *natural trihedral* of a curve. The faces of the moving trihedral are three mutually perpendicular planes:

1. The *normal plane* is the plane containing the principal normal and the binormal.
2. The *osculating plane* is the plane containing the tangent and the principal normal.
3. The *rectifying plane* is the plane containing the tangent and the binormal.

The direction vector of the tangent is equal to the first derivative

$$\mathbf{T} = \mathbf{r}'_t.$$

The direction vector of the binormal is equal to the vector product of the vectors of the first and second derivatives (see Fig. 9.17):

$$\mathbf{B} = \mathbf{r}'_t \times \mathbf{r}''_{tt}.$$

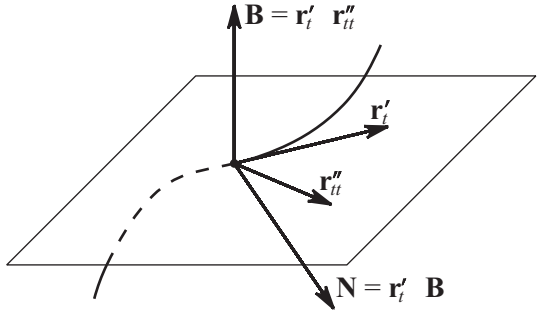


Figure 9.17. Direction vectors of the tangent, of the binormal, and of the principal normal.

The direction vector of the principal normal is perpendicular to the tangent and binormal vectors. Hence it can be set equal to their vector product

$$\mathbf{N} = \mathbf{r}'_t \times \mathbf{B}.$$

The moving trihedral permits one to assign a rectangular coordinate system to each point of the curve; the axes of this coordinate system coincide with the tangent, the principal normal, and the binormal. To determine the sense of these axes, one introduces positive unit vectors of these axes.

Consider a curve defined via the natural parameter s.

1. The unit tangent vector is defined to coincide with the first derivative of the position vector with respect to the natural parameter,
$$\mathbf{t} = \mathbf{r}'_s. \tag{9.1.2.8}$$
2. The unit principal normal vector $\mathbf{n}$ is defined in such a way that its sense coincides with that of the vector of the second derivative with respect to the parameter,
$$\mathbf{n} = \frac{\mathbf{r}''_{ss}}{|\mathbf{r}''_{ss}|} = \frac{1}{k}\mathbf{r}''_{ss}.$$
The vector $k\mathbf{n} = \mathbf{r}''_{ss}$ is called the *curvature vector*; here k is the curvature, studied in Paragraph 9.1.2-6.
3. The unit binormal vector $\mathbf{b}$ is chosen from the condition that $\mathbf{t}$, $\mathbf{n}$, and $\mathbf{b}$ is a right triple of vectors,
$$\mathbf{b} = \mathbf{t} \times \mathbf{n}.$$

If the direction of arc length increase is changed, then the tangent and binormal vectors also change their sense, but the sense of the principal normal vector remains the same.

9.1.2-5. Equations for elements of trihedral.

At a regular point M_0, the equation of the tangent has the form [see also (9.1.2.3)]

$$\mathbf{r} = \mathbf{r}_0 + \lambda\mathbf{t},$$

where λ is a variable parameter.

At a regular point M_0, the equation of the principal normal has the form [see also (9.1.2.5)]

$$\mathbf{r} = \mathbf{r}_0 + \lambda\mathbf{n},$$

where λ is a variable parameter.

At a regular point M_0, the equation of the binormal has the form

$$\mathbf{r} = \mathbf{r}_0 + \lambda\mathbf{b},$$

where λ is a variable parameter.

The vector equation of the osculating plane at a regular point M_0 is

$$(\mathbf{r} - \mathbf{r}_1) \cdot \mathbf{b} = 0$$

[see also (9.1.2.6)].

The vector equation of the normal plane at a regular point M_0 is

$$(\mathbf{r} - \mathbf{r}_0) \cdot \mathbf{t} = 0$$

[see also (9.1.2.4)].

The vector equation of the rectifying plane at a regular point M_0 is

$$(\mathbf{r} - \mathbf{r}_0) \cdot \mathbf{n} = 0.$$

9.1.2-6. Curvature of space curves.

The curvature of a space curve at a point M_0 is determined by analogy with the curvature of a plane curve (see Paragraph 9.1.1-7); i.e., the *curvature* of a curve at a point M_0 is the limit ratio of the tangent rotation angle along an arc shrinking to M_0 to the arc length (Fig. 9.18),

$$k = \lim_{\Delta s \to 0} \left| \frac{\Delta \mathbf{t}}{\Delta s} \right| = \left| \frac{d\mathbf{t}}{ds} \right|. \tag{9.1.2.9}$$

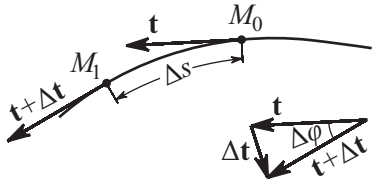


Figure 9.18. The curvature of a space curve.

Remark. The curvature k can be determined as the angular velocity of the vector $\mathbf{t}$ (or, which is the same, the angular velocity of the tangent) at a given point of the curve with respect to the distance s passed along the curve.

It follows from (9.1.2.8) and (9.1.2.9) that

$$k = |\mathbf{r}''_{ss}| = \sqrt{(x''_{ss})^2 + (y''_{ss})^2 + (z''_{ss})^2}. \tag{9.1.2.10}$$

For an arbitrary choice of the parameter, the curvature k is calculated by the formulas

$$k = \frac{|\mathbf{r}'_t \times \mathbf{r}''_{tt}|}{|\mathbf{r}'_t|^3},$$

$$k = \frac{\sqrt{\left[(x'_t)^2 + (y'_t)^2 + (z'_t)^2\right]\left[(x''_{tt})^2 + (y''_{tt})^2 + (z''_{tt})^2\right] - \left(x'_t x''_{tt} + y'_t y''_{tt} + z'_t z''_{tt}\right)^2}}{\left[(x'_t)^2 + (y'_t)^2 + (z'_t)^2\right]^{3/2}}.$$

The *radius of curvature* and the curvature are reciprocal quantities; i.e., $\rho = 1/k$. For space curves, ρ and k are always positive.

Example 1. Let us find the curvature of the helix $x = a\cos t$, $y = a\sin t$, $z = bt$.

Expressing the parameter t in terms of the arc length s as

$$s = \sqrt{a^2 + b^2}\,t,$$

we obtain

$$x = a\cos\frac{s}{\sqrt{a^2+b^2}}, \quad y = a\sin\frac{s}{\sqrt{a^2+b^2}}, \quad z = \frac{bs}{\sqrt{a^2+b^2}}.$$

Formula (9.1.2.10) implies

$$k = \frac{a}{a^2+b^2}, \quad \rho = \frac{a^2+b^2}{a}.$$

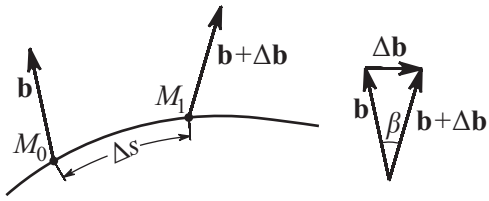


Figure 9.19. The torsion of a space curve.

9.1.2-7. Torsion of space curves.

In addition to curvature, space curves are also characterized by torsion. The *torsion* of a space curve is defined to be a quantity whose absolute value is equal to the limit ratio of the rotation angle of the binormal along an arc, shrinking to a given point to the arc length (Fig. 9.19):

$$|\tau| = \lim_{\Delta s\to 0}\frac{\beta}{\Delta s} = \lim_{\Delta s\to 0}\left|\frac{\Delta\mathbf{b}}{\Delta s}\right| = \left|\frac{d\mathbf{b}}{ds}\right|.$$

Remark 1. The torsion τ does not change (including the sign) if the direction of increase of the parameter s is changed to the opposite.

Remark 2. The torsion τ can be defined as the angular velocity of the binormal $\mathbf{b}$ at the corresponding point with respect to the distance s passed along the curve.

In the special case of a plane curve, the torsion is zero at all points; conversely, if the torsion is zero at all points of a curve, then the entire curve lies in a single plane, which is the common osculating plane for all of its points.

For an arbitrary choice of the parameter t, the torsion τ is calculated by the formulas

$$\tau = \rho^2\frac{\mathbf{r}'_t\mathbf{r}''_{tt}\mathbf{r}'''_{ttt}}{|\mathbf{r}'_t|^3} = \frac{\begin{vmatrix} x'_t & y'_t & z'_t \\ x''_{tt} & y''_{tt} & z''_{tt} \\ x'''_{ttt} & y'''_{ttt} & z'''_{ttt}\end{vmatrix}}{\left[(x'_t)^2+(y'_t)^2+(z'_t)^2\right]\left[(x''_{tt})^2+(y''_{tt})^2+(z''_{tt})^2\right]-(x'_tx''_{tt}+y'_ty''_{tt}+z'_tz''_{tt})^2}. \tag{9.1.2.11}$$

But if the arc length is taken as the parameter, then τ is calculated by the formulas

$$\tau = \rho^2\frac{(\mathbf{r}'_s\mathbf{r}''_{ss}\mathbf{r}'''_{sss})}{|\mathbf{r}'_s|^3} = \frac{\begin{vmatrix} x'_s & y'_s & z'_s \\ x''_{ss} & y''_{ss} & z''_{ss} \\ x'''_{sss} & y'''_{sss} & z'''_{sss}\end{vmatrix}}{\left[(x''_{ss})^2+(y''_{ss})^2+(z''_{ss})^2\right]^2}.$$

The quantity $\rho_\tau = 1/\tau$ inverse to the torsion is called the *radius of torsion*.

Example 2. Let us find the torsion of the helix $x = a\cos t$, $y = a\sin t$, $z = bt$. Using formula (9.1.2.11), we obtain

$$\tau = \frac{\begin{vmatrix} -a\sin t & a\cos t & b \\ -a\cos t & -a\sin t & 0 \\ a\sin t & -a\cos t & 0\end{vmatrix}}{\left(a^2\sin^2 t+a^2\cos^2 t+b^2\right)\left(a^2\cos^2 t+a^2\sin^2 t\right)-\left(a^2\sin t\cos t-a^2\cos t\sin t\right)^2} = \frac{b}{a^2+b^2}.$$

9.1.2-8. Serret–Frénet formulas.

The derivatives of the vectors **t**, **n**, and **b** with respect to the parameter s at each point of the curve satisfy the *Serret–Frénet formulas*

$$\mathbf{t}'_s = k\mathbf{n}, \quad \mathbf{n}'_s = -k\mathbf{t} + \tau\mathbf{b}, \quad \mathbf{b}'_s = -\tau\mathbf{n},$$

where k is the curvature of the curve and τ is the torsion of the curve.

As the parameter s increases, the point M moves along Γ; in this case

1. The tangent rotates around the instantaneous position of the binormal **b** at a positive angular velocity equal to the curvature k of Γ at M.
2. The binormal **b** rotates around the instantaneous position of the tangent at an angular velocity equal to the torsion τ of Γ at M; here the torsion is positive if the shape of the curve resembles a right helix.
3. The trihedral rotates as a solid around an instantaneous axis whose direction is determined by the *Darboux vector* $\Omega = \tau\mathbf{t} + k\mathbf{b}$ at a positive angular velocity equal to the total curvature of Γ M, i.e., to $|\Omega| = \sqrt{\tau^2 + k^2}$.

The Serret–Frénet formulas permit one to find the coefficients in the decomposition of the derivatives of the position vector in the vectors of the moving trihedral. The following recursion formulas hold for the decomposition of the derivatives of the position vector of a point on the curve with respect to the arc length:

$$\mathbf{r}_s^{(n)} = A_n\mathbf{t} + B_n\mathbf{n} + C_n\mathbf{b},$$

where

$$A_{n+1} = (A_n)'_s - B_nk, \quad B_{n+1} = (B_n)'_s + A_nk - C_n\tau, \quad C_{n+1} = (C_n)'_s + B_n\tau.$$

For example, using the relation $\mathbf{r}''_{ss} = k\mathbf{n}$, we can decompose $\mathbf{r}'''_{sss}$ as

$$\mathbf{r}'''_{sss} = -k^2\mathbf{t} + k'_s\mathbf{n} + k\tau\mathbf{b}.$$

9.2. Theory of Surfaces

9.2.1. Elementary Notions in Theory of Surfaces

9.2.1-1. Equation of surface.

Any surface can be determined by equations in one of the following forms:

Explicitly,

$$z = f(x, y).$$

Implicitly,

$$F(x, y, z) = 0. \tag{9.2.1.1}$$

Parametrically,

$$x = x(u, v), \quad y = y(u, v), \quad z = z(u, v).$$

In vector form,

$$\mathbf{r} = \mathbf{r}(u, v), \quad \text{or} \quad \mathbf{r} = x(u, v)\mathbf{i} + y(u, v)\mathbf{j} + z(u, v)\mathbf{k}.$$

Varying the parameters u and v arbitrarily, we obtain the position vector and the coordinates of various points on the surface. Eliminating the parameters u and v from a parametric equation of a surface, we obtain an implicit equation of the surface. The explicit equation is a special case ($u = x$ and $v = y$) of the parametric equation.

It is assumed that the vectors $\mathbf{r}_u \equiv \partial\mathbf{r}/\partial u$ and $\mathbf{r}_v \equiv \partial\mathbf{r}/\partial v$ are nonparallel, i.e.,

$$\mathbf{r}_u \times \mathbf{r}_v \neq 0. \tag{9.2.1.2}$$

The points at which condition (9.2.1.2) holds are said to be *regular*. This condition satisfied at a point M on the surface guarantees that the equation of the surface near this point can be solved for one of the coordinates. Condition (9.2.1.2) also guarantees a one-to-one correspondence between points on the surface and pairs of values of u and v in the corresponding range of u and v.

Condition (9.2.1.2) for a surface defined implicitly becomes

$$\operatorname{grad} F \neq 0\,.$$

For a surface defined parametrically, all functions have continuous first partial derivatives, and the rank of the matrix

$$\begin{pmatrix} x_u & y_u & z_u \\ x_v & y_v & z_v \end{pmatrix}$$

is equal to 2.

Example 3. The sphere defined by the implicit equation

$$x^2 + y^2 + z^2 - a^2 = 0$$

is described in parametric form as

$$x = a\cos u \sin v, \quad y = a \sin u \sin v, \quad z = a\cos v$$

and in vector form as

$$\mathbf{r} = a\cos u \sin v\mathbf{i} + a \sin u \sin v\mathbf{j} + a\cos v\mathbf{k}.$$

9.2.1-2. Curvilinear coordinates on surface.

We consider a surface or part of the surface such that it can be topologically (i.e., bijectively and continuously) mapped onto a plane domain and assume that a point M of this surface is taken to a point M_0 with rectangular coordinates u and v on the plane (see Fig. 9.20). If such a mapping is given, then the surface is said to be *parametrized*, and u and v are called *curvilinear (Gaussian) coordinates* of the point M on the surface. Since the mapping is continuous, each curve on the plane gives rise to a curve on the surface. In particular, the straight lines $u = \text{const}$ and $v = \text{const}$ are associated with the curves on the surface, which are called *parametric* or *curvilinear lines* of the surface. Since the mapping is one-to-one, there is a single curve of the family $u = \text{const}$ and a single curve of the family $v = \text{const}$ that pass through each point of the parametrized surface. Both families together form a regular net, which is called the *coordinate net*.

In the case of rectangular Cartesian coordinates, the coordinate net on the plane is formed by all possible straight lines parallel to coordinate axes; in the case of polar coordinates, the coordinate net is formed by circles centered at the pole and half-lines issuing from the pole.

Example 4. In the parametric equations

$$x = a\cos u \sin v, \quad y = a \sin u \sin v, \quad z = a\cos v$$

of the sphere, u is the longitude and v the polar distance of a point.

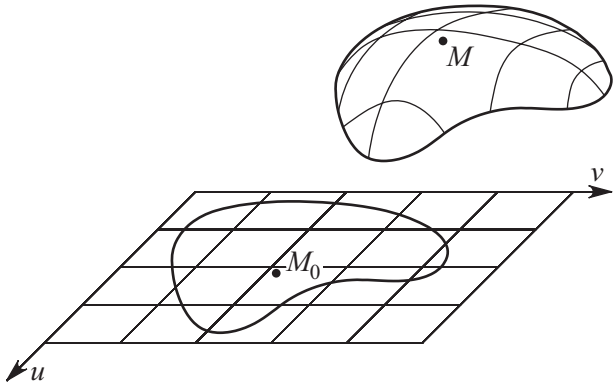


Figure 9.20. The parametrized surface.

9.2.1-3. Tangent line to surface.

A straight line is said to be *tangent* to a surface if it is tangent to a curve lying in this surface. Suppose that a surface is given in vector form (9.2.1.1) and a curve lying on it is parametrized by the parameter t. Then to each parameter value there corresponds a point of the curve, and the position of this point on the surface is specified by some values of the curvilinear coordinates u and v. Thus the curvilinear coordinates of points of a curve lying on a surface are functions of the parameter t.

The system of equations

$$u = u(t), \quad v = v(t) \tag{9.2.1.3}$$

is called the *intrinsic* equations of the curve on the surface.

The intrinsic equations completely characterize the curve if the vector equation of the surface is given, since the substitution of (9.2.1.3) into (9.2.1.1) results in the equation

$$\mathbf{r} = \mathbf{r}[u(t), v(t)], \tag{9.2.1.4}$$

which is called the *parametric equation* of the curve.

The differential of the position vector is equal to

$$d\mathbf{r} = \mathbf{r}_u\, du + \mathbf{r}_v\, dv, \tag{9.2.1.5}$$

where $du = u'_t(t)\, dt$ and $dv = v'_t(t)\, dt$. The vectors $\mathbf{r}_u$ and $\mathbf{r}_v$ are called the *coordinate vectors* corresponding to the point whose curvilinear coordinates have been used in the computations. The coordinate vectors are tangent vectors to coordinate curves (Fig. 9.21).

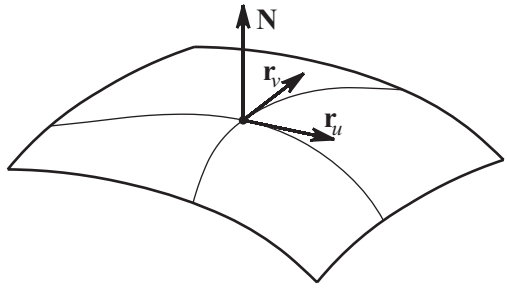


Figure 9.21. Coordinate vectors.

Formula (9.2.1.5) shows that the direction vector of any tangent line to a surface at a given point is a linear combination of the coordinate vectors corresponding to this point; i.e., the tangent to a curve lies in the plane spanned by the vectors $\mathbf{r}_u$ and $\mathbf{r}_v$ at this point.

The direction of the tangent to a curve on a surface at a point M is completely characterized by the ratio $dv:du$ of differentials taken along this curve.

9.2.1-4. Tangent plane and normal.

If all possible curves are drawn on a surface through a given regular point $M_0(r_0) = M_0(x_0, y_0, z_0) = M(u_0, v_0)$ of the surface, then their tangents at M_0 lie in the same plane, which is called the *tangent plane* to the surface at M_0. The tangent plane can be defined as the limit position of the plane passing through three distinct points M_0, M_1, and M_2 on the surface as $M_1 \to M_0$ and $M_2 \to M_0$; here M_1 and M_2 should move along curves with distinct tangents at M_0.

The tangent plane at M_0 can be viewed as the plane passing through M_0 and perpendicular to the vector

$$\mathbf{r}_u \times \mathbf{r}_v;$$

i.e., it passes through the vectors $\mathbf{r}_u$ and $\mathbf{r}_v$. Thus the tangent plane at M_0, depending on the method for defining the surface, is given by one of the equations

$$\begin{vmatrix} x-x_0 & y-y_0 & z-z_0 \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix} = 0, \quad [(\mathbf{r}-\mathbf{r}_0)\mathbf{r}_u\mathbf{r}_v] = 0,$$

$$F_x(x-x_0) + F_y(y-y_0) + F_z(z-z_0) = 0, \quad z_x(x-x_0) + z_y(y-y_0) = z - z_0,$$

where all the derivatives are evaluated at the point $M_0(r_0) = M_0(x_0, y_0, z_0) = M(u_0, v_0)$.

The straight line passing through M_0 and perpendicular to the tangent plane is called the *normal* to the surface at M_0. The vector $\mathbf{N} = \mathbf{r}_u \times \mathbf{r}_v / |\mathbf{r}_u \times \mathbf{r}_v|$ is called the *unit normal vector*. The sense of the vector $\mathbf{N}$ is called the *positive normal direction*; the vector $\mathbf{r}_u$, the vector $\mathbf{r}_v$, and the positive normal form a right triple.

The equation of the normal, depending on the method for defining the surface, has one of the forms

$$\frac{x-x_0}{\begin{vmatrix} y_u & z_u \\ y_v & z_v \end{vmatrix}} = \frac{y-y_0}{\begin{vmatrix} z_u & x_u \\ z_v & x_v \end{vmatrix}} = \frac{z-z_0}{\begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix}}, \quad \mathbf{r} = \mathbf{r}_0 + \lambda(\mathbf{r}_u \times \mathbf{r}_v) \quad \text{or} \quad \mathbf{r} = \mathbf{r}_0 + \lambda\mathbf{N},$$

$$\frac{x-x_0}{F_x} = \frac{y-y_0}{F_y} = \frac{z-z_0}{F_z}, \quad \frac{x-x_0}{z_x} = \frac{y-y_0}{z_y} = \frac{z-z_0}{-1},$$

where all derivatives are evaluated at the point $M_0(r_0) = M_0(x_0, y_0, z_0) = M(u_0, v_0)$.

Example 5. For the sphere given by the implicit equation

$$x^2 + y^2 + z^2 - a^2 = 0,$$

the tangent plane at the point $M_0(x_0, y_0, z_0)$ is given by the equation

$$2x_0(x-x_0) + 2y_0(y-y_0) + 2z_0(z-z_0) = 0 \quad \text{or} \quad xx_0 + yy_0 + zz_0 = 0,$$

and the normal is given by the equation

$$\frac{x-x_0}{2x_0} = \frac{y-y_0}{2y_0} = \frac{z-z_0}{2z_0} \quad \text{or} \quad \frac{x}{x_0} = \frac{y}{y_0} = \frac{z}{z_0}.$$

9.2.1-5. First quadratic form.

If a surface is given parametrically or in vector form, $M_1(u, v)$ is an arbitrary point, and $M_2(u + du, v + dv)$ is a nearby point on the surface, then the length of the arc M_1M_2 on the surface is approximately expressed in terms of the *arc length differential* or in terms of the *linear surface element* by the formula

$$ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2, \tag{9.2.1.6}$$

where the coefficients E, F, and G are given by the formulas

$$E = \mathbf{r}_u^2 = x_u^2 + y_u^2 + z_u^2,$$
$$F = \mathbf{r}_u \cdot \mathbf{r}_v = x_u x_v + y_u y_v + z_u z_v,$$
$$G = \mathbf{r}_v^2 = x_v^2 + y_v^2 + z_v^2.$$

The right-hand side of formula (9.2.1.6) is also called the *first quadratic form of the surface* given parametrically or in vector form; its coefficients E, F, and G depend on the point on the surface. At each regular point on the surface corresponding to (real) coordinates u and v, the first quadratic form (0.2.1.12) is positive definite; i.e.,

$$E > 0, \quad G > 0, \quad EG - F^2 > 0.$$

Example 6. For the sphere given by the equation

$$\mathbf{r} = a\cos u \sin v\mathbf{i} + a \sin u \sin v\mathbf{j} + a\cos v\mathbf{k},$$

the coefficients E, F, and G are equal to

$$E = a^2 \sin^2 v, \quad F = 0, \quad G = a^2,$$

and the first quadratic form is

$$ds^2 = a^2(\sin^2 v\,du^2 + dv^2).$$

For a surface given explicitly, the coefficients E, F, and G are given by the formulas

$$E = 1 + z_x^2, \quad F = z_x z_y, \quad G = 1 + z_y^2.$$

The *arc length* of the curve $u = u(t)$, $v = v(t)$, $t \in [t_0, t_1]$, on the surface can be calculated by the formula

$$L = \int_{t_0}^{t_1} ds = \int_{t_0}^{t_1} \sqrt{Eu_t^2 + 2Fu_t v_t + Gv_t^2}\,dt. \tag{9.2.1.7}$$

The angle γ between two curves (i.e., between their tangents) intersecting in a point M and having the direction vectors $d\mathbf{r} = (du, dv)$ and $\delta\mathbf{r} = (\delta u, \delta v)$ at this point (Fig. 9.22) can be calculated by the formula

$$\cos\gamma = \frac{d\mathbf{r}\,\delta\mathbf{r}}{|d\mathbf{r}||\delta\mathbf{r}|} = \frac{E\,du\,\delta u + F(du\,\delta v + dv\,\delta u) + G\,dv\,\delta v}{\sqrt{E\,du^2 + 2F\,du\,dv + G\,dv^2}\sqrt{E\,\delta u^2 + 2F\,\delta u\,\delta v + G\,\delta v^2}}. \tag{9.2.1.8}$$

(The coefficients E, F, and G are evaluated at point M.)

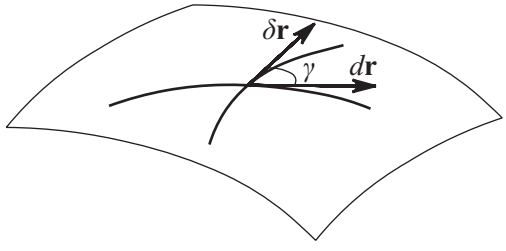


Figure 9.22. The angle between two space curves.

In particular, the angle γ_1 between the coordinate curves $u = \text{const}$ and $v = \text{const}$ passing through a point $M(u, v)$ is determined by the formulas

$$\cos\gamma_1 = \frac{F}{\sqrt{EG}}, \quad \sin\gamma_1 = \frac{\sqrt{EG - F^2}}{\sqrt{EG}}.$$

The coordinate lines are perpendicular if $F = 0$.

The area of a domain U bounded by some curve on the surface can be calculated as the double integral

$$S = \int_U dS = \int_U \sqrt{EG - F^2}\, du\, dv. \tag{9.2.1.9}$$

Thus if the coefficients E, F, and G of the first quadratic form are known, then one can measure lengths, angles, and areas on the surface according to formulas (9.2.1.7), (9.2.1.8), and (9.2.1.9); i.e., the first quadratic form completely determines the *intrinsic geometry of the surface* (see Subsection 9.2.3 for details).

To calculate surface areas in three-dimensional space, one can use the following theorems.

THEOREM 1. *If a surface is given in the explicit form $z = f(x, y)$ and a domain U on the surface is projected onto a domain V on the plane (x, y), then*

$$S = \int_V \sqrt{1 + f_x^2 + f_y^2}\, dx\, dy.$$

THEOREM 2. *If the surface is given implicitly ($F(x, y, z) = 0$) and a domain U on the surface is projected bijectively onto a domain V on the plane (x, y), then*

$$S = \int_V \frac{|\operatorname{grad} F|}{|F_z|}\, dx\, dy,$$

where $|F_z| = \partial F/\partial z \neq 0$ for (x, y, z) lying in the domain U.

THEOREM 3. *If a surface is the parametric form $\mathbf{r} = \mathbf{r}(u, v)$ or $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$, then*

$$S = \int_V |\mathbf{r}_u \times \mathbf{r}_v|\, du\, dv.$$

9.2.1-6. Singular (conic) points of surface.

A point $M_0(x_0, y_0, z_0)$ on a surface given implicitly, i.e., determined by the equation $F(x, y) = 0$, is said to be *singular (conic)* if its coordinates satisfy the system of equations

$$F_x(x_0, y_0, z_0) = 0, \quad F_y(x_0, y_0, z_0) = 0, \quad F_z(x_0, y_0, z_0) = 0, \quad F(x_0, y_0, z_0) = 0.$$

All tangents passing through a singular point $M_0(x_0, y_0, z_0)$ do not lie in the same plane but form a second-order cone defined by the equation

$$\begin{aligned} F_{xx}(x - x_0) + F_{yy}(y - y_0) + F_{zz}(z - z_0) + 2F_{xy}(x - x_0)(y - y_0) \\ + 2F_{yz}(y - y_0)(z - z_0) + 2F_{zx}(z - z_0)(x - x_0) = 0. \end{aligned}$$

The derivatives are evaluated at the point $M_0(x_0, y_0, z_0)$; if all six second partial derivatives are simultaneously zero, then the singular point is of a more complicated type (the tangents form a cone of third or higher order).

9.2.2. Curvature of Curves on Surface

9.2.2-1. Normal curvature. Meusnier's theorem.

Of the plane sections of a surface, the planes containing the normal to the surface at a given point are said to be *normal*. In this case, there exists a unique normal section Γ_0 containing a given tangent to the curve Γ.

MEUSNIER THEOREM. *The radius of curvature at a given point of a curve Γ lying on a surface is equal to the radius of curvature of the normal section Γ_0 taken at the same point with the same tangent, multiplied by the cosine of the angle α between the osculating plane of the curve at this point and the plane of the normal section Γ_0; i.e.,*

$$\rho = \rho_N \cos\alpha‡$$

The *normal curvature* of a curve Γ at a point $M(u, v)$ is defined as

$$k_N = k\mathbf{n} \cdot \mathbf{N} = \mathbf{r}''_{ss} \cdot \mathbf{N} = -\mathbf{r}'_s \cdot \mathbf{N}'_s.$$

The normal curvature is the curvature of the *normal section*.

The *geodesic curvature* of a curve Γ at a point $M(u, v)$ is defined as

$$k_G = k\mathbf{r}'_s\mathbf{n}\mathbf{N} = \mathbf{r}'_s\mathbf{r}''_{ss}\mathbf{N}.$$

The geodesic curvature is the angular velocity of the tangent to the curve around the normal. The geodesic curvature is the curvature of the projection of the curve Γ onto the tangent plane.

For any point u, v of the curve Γ given by equation (9.2.1.4) and lying on the surface, the curvature vector can be represented as a sum of two vectors,

$$\mathbf{r}''_{ss} = k\mathbf{n} = k_G\mathbf{N} \times \mathbf{r}'_s + k_N\mathbf{N}, \tag{9.2.2.1}$$

where $\mathbf{N}$ is the unit normal vector to the surface. The first term on the right-hand side in (9.2.2.1) is called the *geodesic (tangential) curvature vector*, and the second term is called the *normal curvature vector*. The geodesic curvature vector lies in the tangent plane, and the normal curvature vector is normal to the surface.

9.2.2-2. Second quadratic form. Curvature of curve on surface.

The quadratic differential form

$$-d\mathbf{r} \cdot d\mathbf{N} = L\,du^2 + 2M\,du\,dv + N\,dv^2,$$
$$L = -\mathbf{r}_u \cdot \mathbf{N}_u = \frac{\mathbf{r}_{uu}\mathbf{r}_u\mathbf{r}_v}{\sqrt{EG - F^2}}, \quad N = -\mathbf{r}_v \cdot \mathbf{N}_v = \frac{\mathbf{r}_{vv}\mathbf{r}_u\mathbf{r}_v}{\sqrt{EG - F^2}},$$
$$M = -\mathbf{r}_u \cdot \mathbf{N}_v = -\mathbf{r}_v \cdot \mathbf{N}_u = \frac{\mathbf{r}_{uv}\mathbf{r}_u\mathbf{r}_v}{\sqrt{EG - F^2}}$$

(all derivatives are evaluated at the point $M(u, v)$) is called the *second quadratic form of the surface*.

The coefficients L, N, and M for surfaces given parametrically or implicitly can be calculated by the formulas

$$L = \frac{1}{\sqrt{EG - F^2}} \begin{vmatrix} x_{uu} & y_{uu} & z_{uu} \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix} = z_{xx}\sqrt{1 + z_x^2 + z_y^2},$$

$$M = \frac{1}{\sqrt{EG - F^2}} \begin{vmatrix} x_{uv} & y_{uv} & z_{uv} \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix} = z_{xy}\sqrt{1 + z_x^2 + z_y^2},$$

$$N = \frac{1}{\sqrt{EG - F^2}} \begin{vmatrix} x_{vv} & y_{vv} & z_{vv} \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix} = z_{yy}\sqrt{1 + z_x^2 + z_y^2}.$$

The curvature k_N of a normal section can be calculated by the formula

$$k_N = -\frac{d\mathbf{r} \cdot d\mathbf{N}}{ds^2} = \frac{L\,du^2 + 2M\,du\,dv + N\,dv^2}{E\,du^2 + 2F\,du\,dv + G\,dv^2}.$$

A point on the surface at which the curvature ρ_N of a normal section takes the same value for any normal section ($L : M : N = E : F : G$) is said to be *umbilical (circular)*. At each nonumbilical point, there are two normal sections called the *principal normal sections*. They are characterized by the maximum and minimum values k_1 and k_2 of the curvature ρ_N, which are called the *principal curvatures* of the surface U at the point $M(u, v)$. The planes of principal normal sections are mutually perpendicular.

EULER THEOREM. *For a normal section at $M(u, v)$ whose plane forms an angle θ with the plane of one of the principal normal sections, one has*

$$k_N = k_1 \cos^2\theta + k_2 \sin^2\theta \quad \text{or} \quad k_N = k_1 + (k_2 - k_1)\sin^2\theta.$$

The quantities k_1 and k_2 are the roots of the characteristic equation

$$\begin{vmatrix} L - kE & M - kF \\ M - kF & N - kG \end{vmatrix} = 0.$$

The curves on a surface whose directions at each point coincide with the directions of the principal normal sections are called the *curvature lines*; their differential equation is

$$\begin{vmatrix} dv^2 & -dv\,du & du^2 \\ E & F & G \\ L & M & N \end{vmatrix} = 0.$$

Asymptotic lines are defined to be the curves for which $\rho_N = 0$ at each point. The asymptotic lines are determined by the differential equation

$$L\,du^2 + 2M\,du\,dv + N\,dv^2 = 0.$$

9.2.2-3. Mean and Gaussian curvatures.

The symmetric functions

$$H(u, v) = \frac{k_1 + k_2}{2} \quad \text{and} \quad K(u, v) = k_1 k_2$$

are called the *mean and Gaussian (extrinsic) curvature*, respectively, of the surface U at the point $M(u, v)$. They are given by the formulas

$$H(u,v) = \frac{EN - 2FM + LG}{2(EG - F^2)},$$
$$K(u,v) = \frac{LN - M^2}{EG - F^2}. \qquad (9.2.2.2)$$

The mean and Gaussian curvatures are related by the inequality

$$H^2 - K = \frac{(k_1 + k_2)^2}{4} - k_1 k_2 = \frac{(k_1 - k_2)^2}{4} \geq 0.$$

The mean and Gaussian curvatures can be used to characterize the deviation of the surface from a plane. In particular, if $H = 0$ and $K = 0$ at all points of the surface, then the surface is a plane.

Example 1. For a circular cylinder (of radius a),

$$H = \frac{a}{2}, \quad K = 0.$$

If a surface is represented by the equation $z = f(x, y)$, then the mean and Gaussian curvature can be determined by the formulas

$$H = \frac{r(1 + q^2) - 2pqs + t(1 + p^2)}{2\sqrt{(1 + p^2 + q^2)^3}}, \quad K = \frac{rt - s^2}{(1 + p^2 + q^2)^3},$$

where the following notation is used:

$$p = z_x, \quad q = z_y, \quad r = z_{xx}, \quad s = z_{xy}, \quad t = z_{yy}, \quad h = \sqrt{1 + p^2 + q^2}.$$

The surfaces for which the mean curvature H is zero at all points are said to be *minimal*. The surfaces for which the Gaussian curvature K is constant at all points are called *surfaces of constant curvature*.

9.2.2-4. Classification of points on surface.

The points of a surface can be classified according to the values of the Gaussian curvature:

1. A point M at which $K = k_1 k_2 > 0$ (the principal normal sections are convex in the same direction from the tangent plane; example: any point of an ellipsoid) is called an *elliptic point*; the analytic criterion for this case is $LN - M^2 > 0$. In the special case $k_1 = k_2$, the point is umbilical (circular): R = const for all normal sections at this point.
2. A point M at which $K = k_1 k_2 < 0$ (the principal normal sections are convex in opposite directions; the surface intersects the tangent plane and has a saddle character; example: any point of a one-sheeted hyperboloid) is called a *hyperbolic (saddle) point*; the analytic criterion for this case is $LN - M^2 < 0$.
3. A point M at which $K = k_1 k_2 = 0$ (one principal normal section has an inflection point or is a straight line; example: any point of a cylinder) is called a *parabolic point*; the analytic criterion for this case is $LN - M^2 = 0$.

Any umbilical point ($k_1 = k_2$) is either elliptic or parabolic.

9.2.3. Intrinsic Geometry of Surface

9.2.3-1. Intrinsic geometry and bending of surface.

Suppose that two surfaces U and U^* are given and there is one-to-one correspondence between their points such that the length of each curve on U is equal to the length of the corresponding curve on U^*. Such a one-to-one mapping of U into U^* is called a *bending* of the surface U into the surface U^*, and the surfaces U and U^* are said to be *applicable*.

The correspondence between U and U^* is said to be *isometric*.

The intrinsic geometry of a surface studies geometric constructions and quantities related to the surface that can be determined solely from the first quadratic form. The notions of length of a segment, angle between two curves, and area of part of a surface all belong in intrinsic geometry.

On the opposite, the curvature of a curve given on the surface by the equations

$$u = u(t), \quad v = v(t)$$

cannot be found using only the first quadratic form, and hence it does not belong in intrinsic geometry.

9.2.3-2. Index notation. Surface as Riemannian space.

From now on in this chapter, the following notation related to tensor analysis is used:

$$\begin{aligned}
&u^1 = u, \quad u^2 = v;\\
&g_{11} = E, \quad g_{12} = g_{21} = F, \quad g_{22} = G;\\
&b_{11} = L, \quad b_{12} = b_{21} = M, \quad b_{22} = N;\\
&\mathbf{r}_1 = \mathbf{r}_u, \quad \mathbf{r}_2 = \mathbf{r}_v, \quad \mathbf{r}_{12} = \mathbf{r}_{21} = \mathbf{r}_{vu} = \mathbf{r}_{uv}, \quad \mathbf{r}_{11} = \mathbf{r}_{uu}, \quad \mathbf{r}_{22} = \mathbf{r}_{vv}.
\end{aligned}$$

In the new notation, the first fundamental quadratic form becomes

$$E\,du^2 + 2F\,du\,dv + G\,dv^2 = \sum_{\alpha=1}^{2}\sum_{\beta=1}^{2} g_{\alpha\beta}\,du^\alpha\,du^\beta = g_{\alpha\beta}\,du^\alpha\,du^\beta,$$

and the second fundamental quadratic form is

$$L\,du^2 + 2M\,du\,dv + N\,dv^2 = \sum_{\alpha=1}^{2}\sum_{\beta=1}^{2} b_{\alpha\beta}\,du^\alpha\,du^\beta = b_{\alpha\beta}\,du^\alpha\,du^\beta.$$

The expressions for the coefficients of the first and second quadratic forms in the new notation become

$$g_{ij} = \mathbf{r}_i \cdot \mathbf{r}_j, \quad b_{ij} = \mathbf{r}_{ij} \cdot \mathbf{N},$$

where $\mathbf{N}$ is the unit normal vector to the surface; i and j are equal to either 1 or 2.

9.2.3-3. Derivation formula.

The *Christoffel symbols* $\Gamma_{k,ij}$ *of the first kind* are defined to be the scalar products of the vectors $\mathbf{r}_k$ and $\mathbf{r}_{ij}$; i.e.,

$$\mathbf{r}_k \cdot \mathbf{r}_{ij} = \Gamma_{k,ij}.$$

The Christoffel symbols of the first kind satisfy the formula

$$\Gamma_{k,ij} = \frac{1}{2}\Big(\frac{\partial g_{ki}}{\partial u^j} + \frac{\partial g_{kj}}{\partial u^i} - \frac{\partial g_{ij}}{\partial u^k}\Big).$$

This is one of the basic formulas in the theory of surfaces; this formula means that the scalar products of the second partial derivatives of the position vector $\mathbf{r}(u^i, u^j)$ by its partial derivatives can be expressed in terms of the coefficients of the first quadratic form (more precisely, in terms of their derivatives).

The *Christoffel symbols* Γ^k_{ij} *of the second kind* are defined by the relations

$$\Gamma^k_{ij} = \sum_{t=1}^{n} g^{kt}\Gamma_{k,ij},$$

where g^{kt} is given by

$$\sum_{k=1}^{n} g^{kt}g_{ks} = \begin{cases} 1, & \text{if } t = s, \\ 0, & \text{if } t \neq s. \end{cases}$$

The Christoffel symbols of the second kind are the coefficients in the decomposition of the vector $\mathbf{r}_{ij}$ in two noncollinear vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ and the unit normal vector $\mathbf{N}$:

$$\mathbf{r}_{ij} = \Gamma^1_{ij}\mathbf{r}_1 + \Gamma^2_{ij}\mathbf{r}_2 + b_{ij}\mathbf{N} \quad (i, j = 1, 2). \tag{9.2.3.1}$$

Formulas (9.2.3.1) are called the *first group of derivation formulas (the Gauss derivation formulas).*

The formulas

$$\mathbf{N}_1 = -b^1_1\mathbf{r}_1 - b^2_1\mathbf{r}_2, \quad \mathbf{N}_1 = -b^1_2\mathbf{r}_1 - b^2_2\mathbf{r}_2, \tag{9.2.3.2}$$

where $b^j_i = b_{i\alpha}g^{\alpha j}$, are called the *second group of derivation formulas (Weingarten formulas).* The formulas in the second group of derivation formulas express the partial derivatives of the unit normal vector $\mathbf{N}$ in terms of the variables u^1 and u^2 in the decomposition in the basis vectors $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{N}$.

Formulas (9.2.3.1) and (9.2.3.2) express the partial derivatives with respect to u^1 and u^2 of the basis vectors, i.e., of the two tangent vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ and the normal vector $\mathbf{N}$, at a given point on a surface. These partial derivatives of $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{N}$ are obtained as a decomposition in the vectors $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{N}$ themselves.

9.2.3-4. Gauss formulas. Peterson–Codazzi formulas.

If the first quadratic form of the surface is given, then the second quadratic form cannot be chosen arbitrarily, since its discriminant $(LN - M^2)$ is completely determined by the *Gauss formula*

$$b_{11}b_{22} - b^2_{12} = \frac{\partial^2 g_{12}}{\partial u^1 \partial u^2} - \frac{1}{2}\frac{\partial^2 g_{11}}{\partial u^2 \partial u^2} - \frac{1}{2}\frac{\partial^2 g_{22}}{\partial u^1 \partial u^1} + \Gamma^\gamma_{12}\Gamma^\delta_{12}g_{\gamma\delta} - \Gamma^\alpha_{11}\Gamma^\beta_{22}g_{\alpha\beta}, \tag{9.2.3.3}$$

where γ, δ, α, and β are independent summation indices equal to either 1 or 2.

We use (9.2.3.3) to reduce relation (9.2.2.2) to the form

$$K(u,v) = k_1 k_2 = \frac{LN - M^2}{EG - F^2} = \frac{b_{11}b_{22} - b_{12}^2}{g_{11}g_{22} - g_{12}^2}. \tag{9.2.3.4}$$

In view of (9.2.3.4), the Gaussian curvature of the surface is completely determined by the coefficients of the first quadratic form and by their first and second derivatives with respect to u^1 and u^2.

The coefficients b_{11}, b_{12}, b_{22} of the second quadratic form and their first derivatives are related to the coefficients g_{11}, g_{12}, g_{22} of the first quadratic form and their first derivatives (contained only in Γ_{ij}^k) by differential equations known as the *Peterson–Codazzi formulas*:

$$\frac{\partial b_{i1}}{\partial u^2} - \Gamma_{i2}^1 b_{11} - \Gamma_{i2}^2 b_{21} = \frac{\partial b_{i2}}{\partial u^1} - \Gamma_{i1}^1 b_{12} - \Gamma_{i1}^2 b_{22} \quad (i = 1, 2).$$

The Gauss formula and the Peterson–Codazzi formulas are necessary and sufficient conditions for two analytically determined quadratic differential forms to be the fist and second quadratic forms of some surface.

References for Chapter 9

Aubin, T., *A Course in Differential Geometry*, American Mathematical Society, Providendce, Rhoad Island, 2000.

Burke, W. L., *Applied Differential Geometry*, Cambridge University Press, Cambridge, 1985.

Byushgens, S. S., *Differential Geometry* [in Russian], Komkniga, Moscow, 2006.

Chern, S.-S., Chen, W.-H., and Lam, K. S., *Lectures on Differential Geometry*, World Scientific Publishing Co., Hackensack, New Jersey, 2000.

Danielson, D. A., *Vectors and Tensors in Engineering and Physics, 2nd Rep Edition*, Westview Press, Boulder, Colorado, 2003.

Dillen, F. J. E. and Verstraelen, L. C. A., *Handbook of Differential Geometry, Vol. 1*, North Holland, Amsterdam, 2000.

Dillen, F. J. E. and Verstraelen, L. C. A., *Handbook of Differential Geometry, Vol. 2*, North Holland, Amsterdam, 2006.

Guggenheimer, H. W., *Differential Geometry*, Dover Publications, New York, 1977.

Kay, D. C., *Schaum's Outline of Tensor Calculus*, McGraw-Hill, New York, 1988.

Kobayashi, S. and Nomizu, K., *Foundations of Differential Geometry, Vol. 1*, Wiley, New York, 1996.

Kreyszig, E., *Differential Geometry*, Dover Publications, New York, 1991.

Lang, S., *Fundamentals of Differential Geometry*, Springer, New York, 2001.

Lebedev, L. P. and Cloud, M. J., *Tensor Analysis*, World Scientific Publishing Co., Hackensack, New Jersey, 2003.

Lovelock, D. and Rund, H., *Tensors, Differential Forms, and Variational Principles*, Dover Publications, New York, 1989.

O'Neill, V., *Elementary Differential Geometry, Rev. 2nd Edition*, Academic Press, New York, 2006.

Oprea, J., *Differential Geometry and Its Applications, 2nd Edition*, Prentice Hall, Englewood Cliffs, New Jersey, 2003.

Pogorelov, A. V., *Differential Geometry*, P. Noordhoff, Groningen, 1967.

Postnikov, M. M., *Linear Algebra and Differential Geometry* (Lectures in Geometry), Mir Publishers, Moscow, 1982.

Pressley, A., *Elementary Differential Geometry*, Springer, New York, 2002.

Rashevsky, P. K., *A Course in Differential Geometry, 4th Edition* [in Russian], URSS, Moscow, 2003.

Simmonds, J. G., *A Brief on Tensor Analysis, 2nd Edition*, Springer, New York, 1997.

Somasundaram, D., *Differential Geometry: A First Course*, Alpha Science International, Oxford, 2004.

Spain, B., *Tensor Calculus: A Concise Course*, Dover Publications, New York, 2003.

Spivak, M., *A Comprehensive Introduction to Differential Geometry. Vols 1–5, 3rd Edition*, Publish or Perish, Houston, 1999.

Struik, D. J., *Lectures on Classical Differential Geometry, 2nd Edition*, Dover Publications, New York, 1988.

Temple, G., *Cartesian Tensors: An Introduction*, Dover Publications, New York, 2004.

Chapter 10
Functions of Complex Variable

10.1. Basic Notions

10.1.1. Complex Numbers. Functions of Complex Variable

10.1.1-1. Complex numbers.

The set of complex numbers is an extension of the set of real numbers. An expression of the form $z = x + iy$, where x and y are real numbers, is called a *complex number*, and the symbol i is called the *imaginary unit*: $i^2 = -1$. The numbers x and y are called, respectively, the *real* and *imaginary parts* of z and denoted by

$$x = \operatorname{Re} z \quad \text{and} \quad y = \operatorname{Im} z. \tag{10.1.1.1}$$

The complex number $x + i0$ is identified with real number x, and the number $0 + iy$ is denoted by iy and is said to be *pure imaginary*. Two complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ are assumed to be *equal* if $x_1 = x_2$ and $y_1 = y_2$.

The complex number $\bar{z} = x - iy$ is said to be *conjugate* to the number z.

The *sum or difference* of complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ is defined to be the number

$$z_1 \pm z_2 = x_1 \pm x_2 + i(y_1 \pm y_2). \tag{10.1.1.2}$$

Addition laws:

1. $z_1 + z_2 = z_2 + z_1$ (commutativity).
2. $z_1 + (z_2 + z_3) = (z_1 + z_2) + z_3$ (associativity).

The *product* $z_1 z_2$ of complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ is defined to be the number

$$z_1 z_2 = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 - x_2 y_1). \tag{10.1.1.3}$$

Multiplication laws:

1. $z_1 z_2 = z_2 z_1$ (commutativity).
2. $z_1(z_2 z_3) = (z_1 z_2) z_3$ (associativity).
3. $(z_1 + z_2) z_3 = z_1 z_3 + z_2 z_3$ (distributivity with respect to addition).

The product of a complex number $z = x + iy$ by its conjugate is always nonnegative:

$$z\bar{z} = x^2 + y^2. \tag{10.1.1.4}$$

For a positive integer n, the n-fold product of z by itself is called the nth *power* of the number z and is denoted by z^n. A number w is called an *nth root* of a number z and is denoted by $w = \sqrt[n]{z}$ if $w^n = z$.

If $z_2 \neq 0$, then the *quotient* of z_1 and z_2 is defined as

$$\frac{z_1}{z_2} = \frac{x_1 x_2 + y_1 y_2}{x_2^2 + y_2^2} + i\frac{x_2 y_1 - x_1 y_2}{x_2^2 + y_2^2}. \tag{10.1.1.5}$$

Relation (10.1.1.5) can be obtained by multiplying the numerator and the denominator of the fraction z_1/z_2 by $\bar{z}_2$.

10.1.1-2. Geometric interpretation of complex number.

There is a one-to-one correspondence between complex numbers $z = x + iy$ and points M with coordinates (x, y) on the plane with a Cartesian rectangular coordinate system OXY or with vectors $\overrightarrow{OM}$ connecting the origin O with M (Fig. 10.1). The length r of the vector $\overrightarrow{OM}$ is called the *modulus* of the number z and is denoted by $r = |z|$, and the angle φ formed by the vector $\overrightarrow{OM}$ and the positive direction of the OX-axis is called the *argument* of the number z and is denoted by $\varphi = \operatorname{Arg} z$.

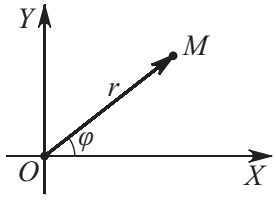


Figure 10.1. Geometric interpretation of complex number.

The modulus of a complex number is determined by the formula

$$|z| = \sqrt{z\bar{z}} = \sqrt{x^2 + y^2}. \tag{10.1.1.6}$$

The argument $\operatorname{Arg} z$ is determined up to a multiple of 2π, $\operatorname{Arg} z = \arg z + 2k\pi$, where k is an arbitrary integer and $\arg z$ is the *principal value* of $\operatorname{Arg} z$ determined by the condition $-\pi < \arg z \le \pi$. The principal value $\arg z$ is given by the formula

$$\arg z = \begin{cases} \arctan(y/x) & \text{for } x > 0, \\ \pi + \arctan(y/x) & \text{for } x < 0,\ y \ge 0, \\ -\pi + \arctan(y/x) & \text{for } x < 0,\ y < 0, \\ \pi/2 & \text{for } x = 0,\ y > 0, \\ -\pi/2 & \text{for } x = 0,\ y < 0. \end{cases} \tag{10.1.1.7}$$

For $z = 0$, $\operatorname{Arg} z$ is undefined.

Since $x = r\cos\varphi$ and $y = r\sin\varphi$, it follows that the complex number can be written in the *trigonometric form*

$$z = x + iy = r(\cos\varphi + i\sin\varphi). \tag{10.1.1.8}$$

For numbers $z_1 = r_1(\cos\varphi_1 + i\sin\varphi_1)$ and $z_2 = r_2(\cos\varphi_2 + i\sin\varphi_2)$, written in trigonometric form, the following rules of algebraic operations are valid:

$$\begin{aligned} z_1 z_2 &= r_1 r_2\big[\cos(\varphi_1 + \varphi_2) + i\sin(\varphi_1 + \varphi_2)\big], \\ \frac{z_1}{z_2} &= \frac{r_1}{r_2}\big[\cos(\varphi_1 - \varphi_2) + i\sin(\varphi_1 - \varphi_2)\big]. \end{aligned} \tag{10.1.1.9}$$

In the latter formula, it is assumed that $z \ne 0$. For any positive integer n, this implies the *de Moivre formula*

$$z^n = r^n(\cos n\varphi + i\sin n\varphi), \tag{10.1.1.10}$$

as well as the *formula for extracting the root of a complex number*. For $z \ne 0$, there are exactly n distinct values of the nth root of the number $z = r(\cos\varphi + i\sin\varphi)$. They are determined by the formulas

$$\sqrt[n]{z} = \sqrt[n]{r}\left(\cos\frac{\varphi + 2k\pi}{n} + i\sin\frac{\varphi + 2k\pi}{n}\right) \quad (k = 0, 1, 2, \ldots, n-1). \tag{10.1.1.11}$$

Example. Let us find all values of $\sqrt[3]{i}$.
We represent the complex number $z = i$ in trigonometric form. We have $r = 1$ and $\varphi = \arg z = \frac{1}{2}\pi$.
The distinct values of the third root are calculated by the formula

$$\omega_k = \sqrt[3]{i}\left(\cos\frac{\frac{\pi}{2}+2\pi k}{3} + i\sin\frac{\frac{\pi}{2}+2\pi k}{3}\right) \quad (k = 0, 1, 2),$$

so that

$$\omega_0 = \cos\frac{\pi}{6} + i\sin\frac{\pi}{6} = \frac{\sqrt{3}}{2} + i\frac{1}{2},$$
$$\omega_1 = \cos\frac{5\pi}{6} + i\sin\frac{5\pi}{6} = -\frac{\sqrt{3}}{2} + i\frac{1}{2},$$
$$\omega_2 = \cos\frac{3\pi}{2} + i\sin\frac{3\pi}{2} = -i.$$

The roots are shown in (Fig. 10.2).

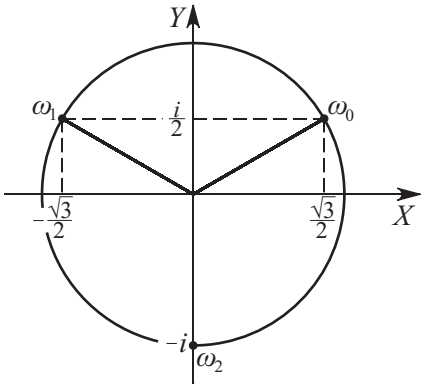


Figure 10.2. The roots of $\sqrt[3]{i}$.

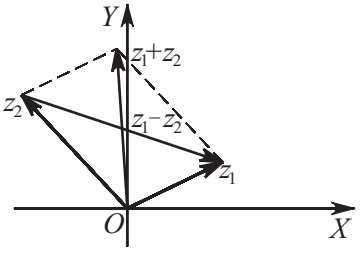


Figure 10.3. The sum and difference of complex numbers.

The plane OXY is called the *complex plane*, the axis OX is called the *real axis*, and the axis OY is called the *imaginary axis*. The notions of complex number and point on the complex plane are identical.

The *geometric meaning of the operations of addition and subtraction of complex numbers* is as follows: the sum and the difference of complex numbers z_1 and z_2 are the vectors equal to the directed diagonals of the parallelogram spanned by the vectors z_1 and z_2 (Fig. 10.3). The following inequalities hold (Fig. 10.3):

$$|z_1 + z_2| \le |z_1| + |z_2|, \quad |z_1 - z_2| \ge \big||z_1| - |z_2|\big|. \tag{10.1.1.12}$$

Inequalities (10.1.1.12) become equalities if and only if the arguments of the complex numbers z_1 and z_2 coincide (i.e., $\arg z_1 = \arg z_2$) or one of the numbers is zero.

10.1.2. Functions of Complex Variable

10.1.2-1. Notion of function of complex variable.

A subset D of the complex plane such that each point of D has a neighborhood contained in D (i.e., D is open) and two arbitrary points of D can be connected by a broken line lying in D (i.e., D is connected) is called a *domain* on the complex plane. A point that does not itself lie in D but whose arbitrary neighborhood contains points of D is called a *boundary point* of D. The set of all boundary points of D is called the *boundary* of D, and the union

of D with its boundary is called a *closed domain* and is denoted by $\overline{D}$. The *positive sense* of the boundary is defined to be the sense for which the domain lies to the left of the boundary. The boundary of a domain can consist of finitely many closed curves, segments, and points; the curves and cuts are assumed to be piecewise smooth.

The simplest examples of domains are neighborhoods of points on the complex plane. A *neighborhood* of a point a on the complex plane is understood as the set of points z such that $|z-a| < R$, i.e., the interior of the disk of radius $R > 0$ centered at the point a. The *extended complex plane* is obtained by augmenting the complex plane with the fictitious *point at infinity*. A *neighborhood of the point at infinity* is understood as the set of points z such that $|z| > R$ (including the point at infinity itself).

If to each point z of a domain D there corresponds a point w (resp., a set of points w), then one says that there is a *single-valued* (resp., *multivalued*) *function* $w = f(z)$ defined on the domain D. If we set $z = x + iy$ and $w = u + iv$, then defining a function $w = f(z)$ of the complex variable is equivalent to defining two functions $\operatorname{Re} f = u = u(x, y)$ and $\operatorname{Im} f = v = v(x, y)$ of two real variables. If the function $w = f(z)$ is single-valued on D and the images of distinct points of D are distinct, then the mapping determined by this function is said to be *schlicht*. The notions of boundedness, limit, and continuity for single-valued functions of the complex variable do not differ from the corresponding notions for functions of two real variables.

10.1.2-2. Differentiability and analyticity.

Let a single-valued function $w = f(z)$ be defined in a neighborhood of a point z. If there exists a limit

$$\lim_{h\to 0} \frac{f(z+h)-f(z)}{h} = f'_z(z), \tag{10.1.2.1}$$

then the function $w = f(z)$ is said to be *differentiable* at the point z and $f'_z(z)$ is called its *derivative* at the point z.

Cauchy–Riemann conditions. If the functions $u(x, y) = \operatorname{Re} f(z)$ and $v(x, y) = \operatorname{Im} f(z)$ are differentiable at a point (x, y), then the *Cauchy–Riemann conditions*

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \tag{10.1.2.2}$$

are necessary and sufficient for the function $w = f(z)$ to be differentiable at the point $z = x+iy$.

If the function $w = f(z)$ is differentiable, then

$$w'_z = u_x + iv_x = v_y - iu_y = u_x + iu_y = v_y + iv_x, \tag{10.1.2.3}$$

where the subscripts x and y indicate the corresponding partial derivatives.

Remark. The Cauchy–Riemann conditions are sometimes also called the d'Alembert–Euler conditions.

The rules for algebraic operations on the derivatives and for calculating the derivative of the composite function and the inverse function (if it exists) have exactly the same form as in the real case:

1. $\big[\alpha f_1(z) \pm \beta f_2(z)\big]'_z = \alpha[f_1(z)]'_z \pm \beta[f_2(z)]'_z$, where α and β are arbitrary complex constants.
2. $\big[f_1(z)f_2(z)\big]'_z = [f_1(z)]'_z f_2(z) + f_1(z)[f_2(z)]'_z$.
3. $\left[\dfrac{f_1(z)}{f_2(z)}\right]'_z = \dfrac{[f_1(z)]'_z f_2(z) - f_1(z)[f_2(z)]'_z}{f_2^2(z)}$ $\qquad (f_2(z) \not\equiv 0)$.
4. If a function $w = f(z)$ is differentiable at a point z and a function $W = F(w)$ is differentiable at the point $w = f(z)$, then the composite function $W = F(f(z))$ is differentiable at the point z and $W'_z = [F(f(z))]'_z = F'_f(f)f'_z(z)$.

5. If a function $w=f(z)$ is differentiable at a point z and the inverse function $z=g(w)\equiv f^{-1}(w)$ exists and is differentiable at the point w and satisfies $[f^{-1}(w)]'_w \neq 0$, then

$$f'_z(z) = \frac{1}{[f^{-1}(w)]'_w}.$$

A single-valued function differentiable in some neighborhood of a point z_0 is said to be *analytic (regular, holomorphic)* at this point.

A function $w = f(z)$ is analytic at a point z_0 if and only if it can be represented by a power series

$$f(z) = \sum_{k=0}^{\infty} c_k (z - z_0)^k \tag{10.1.2.4}$$

converging in some neighborhood of z_0.

A function analytic at each point of the domain D is said to be *analytic* in D.

A function $w = f(z)$ is said to be *analytic at the point at infinity* if the function $F(z) = f(1/z)$ is analytic at the point $z = 0$. In this case, $f'_z(\infty) = (-z^2 F'_z)\big|_{z=0}$ by definition.

A function $w = f(z)$ is analytic at the point at infinity if and only if this function can be represented by a power series

$$f(z) = \sum_{k=0}^{\infty} b_k z^{-k} \tag{10.1.2.5}$$

converging for sufficiently large $|z|$.

If a function $w = f(z)$ is analytic at a point z_0 and $f'_z(z_0) \neq 0$, then $f(z)$ has an analytic inverse function $z(w)$ defined in a neighborhood of the point $w_0 = f(z_0)$. If a function $w = f(z)$ is analytic at a point z_0 and the function $W = F(w)$ is analytic at the point $w_0 = f(z_0)$, then the composite function $W = F[f(z)]$ is analytic at the point z_0. If a function is analytic in a domain D and continuous in $\overline{D}$, then its value at any interior point of the domain is uniquely determined by its values on the boundary of the domain. The analyticity of a function at a point implies the existence and analyticity of its derivatives of arbitrary order at this point.

MAXIMUM MODULUS PRINCIPLE. *If a function $w = f(z)$ that is not identically constant is analytic in a domain D and continuous in $\overline{D}$, then its modulus cannot attain a maximum at an interior point of D.*

LIOUVILLE'S THEOREM. *If a function $w = f(z)$ is analytic and bounded in the entire complex plane, then it is constant.*

Remark. The Liouville theorem can be stated in the following form:
if a function $w = f(z)$ is analytic in the extended complex plane, then it is constant.

Geometric meaning of the absolute value of the derivative. Suppose that a function $w = f(z)$ is analytic at a point z_0 and $f'_z(z_0) \neq 0$. Then the value $|f'_z(z_0)|$ determines the dilatation (similarity) coefficient at the point z_0 under the mapping $w = f(z)$. The value $|f'_z(z_0)|$ is called the *dilatation ratio* if $|f'_z(z_0)| > 1$ and the *contraction ratio* if $|f'_z(z_0)| < 1$.

Geometric meaning of the argument of the derivative. The argument of the derivative $f'_z(z_0)$ is equal to the angle by which the tangent at the point z_0 to any curve passing through z_0 should be rotated to give the tangent to the image of the curve at the point $w_0 = f(z_0)$. For $\varphi = \arg f'_z(z) > 0$, the rotation is anticlockwise, and for $\varphi = \arg f'_z(z) < 0$, the rotation is clockwise.

Single-valued functions, as well as single-valued branches of multi-valued functions, are analytic everywhere on the domains where they are defined. It follows from (10.1.2.2)

that the real and imaginary parts $u(x, y)$ and $v(x, y)$ of a function analytic in a domain are *harmonic* in this domain, i.e., satisfy the Laplace equation

$$\Delta f = f_{xx} + f_{yy} = 0 \tag{10.1.2.6}$$

in this domain.

Remark. If $u(x, y)$ and $v(x, y)$ are two arbitrary harmonic functions, then the function $f(z) = u(x, y) + iv(x, y)$ is not necessarily analytic, since for the analyticity of $f(z)$ the functions $u(x, y)$ and $v(x, y)$ must satisfy the Cauchy–Riemann conditions.

Example 1. The function $w = z^2$ is analytic.

Indeed, since $z = x + iy$, we have $w = (x + iy)^2 = x^2 - y^2 + i2xy$, $u(x, y) = x^2 - y^2$, and $v(x, y) = 2xy$. The Cauchy–Riemann conditions

$$u_x = v_y = 2x, \quad u_y = v_x = -2y$$

are satisfied at all points of the complex plane, and the function $w = z^2$ is analytic.

Example 2. The function $w = \bar{z}$ is not analytic.

Indeed, since $z = x + iy$, we have $w = x - iy$, $u(x, y) = x$, $v(x, y) = -y$. The Cauchy–Riemann conditions are not satisfied,

$$u_x = 1 \neq -1 = v_y, \qquad u_y = v_x = 0,$$

and the function $w = \bar{z}$ is not analytic.

10.1.2-3. Elementary functions.

1°. The *functions* $w = z^n$ *and* $w = \sqrt[n]{z}$ for positive integer n are defined in Paragraph 10.1.1-2. The function

$$w = z^n \tag{10.1.2.7}$$

is single-valued. It is schlicht in the sectors $2\pi k/n < \varphi < 2\pi(k + 1)/n$, $k = 0, 1, 2, \ldots$, each of which is transformed by the mapping $w = z^n$ into the plane w with a cut on the positive semiaxis.

The function

$$w = \sqrt[n]{z} \tag{10.1.2.8}$$

is an n-valued function for $z \neq 0$, and its value is determined by the value of the argument chosen for the point z. If a closed curve C does not surround the point $z = 0$, then, as the point z goes around the entire curve C, the point $w = \sqrt[n]{z}$ for a chosen value of the root also moves along a closed curve and returns to the initial value of the argument. But if the curve C surrounds the origin, then, as the point z goes around the entire curve C in the positive sense, the argument of z increases by 2π and the corresponding point $w = \sqrt[n]{z}$ does not return to the initial position. It will return there only after the point z goes n times around the entire curve C. If a domain D does not contain a closed curve surrounding the point $z = 0$, then one can singe out n continuous single-valued functions, each of which takes only one of the values $w = \sqrt[n]{z}$; these functions are called the *branches* of the multi-valued function $w = \sqrt[n]{z}$. One cannot single n separate branches of the function $w = \sqrt[n]{z}$ in any neighborhood of the point $z = 0$; accordingly, the point $z = 0$ is called a *branch point* of this function.

2°. The *Zhukovskii function*

$$w = \frac{1}{2}\left(z + \frac{1}{z}\right) \tag{10.1.2.9}$$

is defined and single-valued for all $z \neq 0$; it is schlicht in any domain that does not simultaneously contain any points z_1 and z_2 such that $z_1 z_2 = 1$.

3°. The *exponential function* $w = e^z$ is defined by the formula

$$w = e^z = e^{x+iy} = e^x(\cos y + i \sin y). \tag{10.1.2.10}$$

The function $w = e^z$ is analytic everywhere. For the exponential function, the usual differentiation rule is preserved:

$$(e^z)'_z = e^z.$$

The basic property of the exponential function (*addition theorem*) is also preserved:

$$e^{z_1} e^{z_2} = e^{z_1+z_2}.$$

For $x = 0$ and $y = \varphi$, the definition of the exponential function implies the *Euler formula* $e^{i\varphi} = \cos\varphi + i \sin\varphi$, which permits one to write any complex number with modulus r and argument φ in the *exponential form*

$$z = r(\cos\varphi + i \sin\varphi) = re^{i\varphi}. \tag{10.1.2.11}$$

The exponential function is 2π-periodic, and the mapping $w = e^z$ is schlicht in the strip $0 \le y < 2\pi$.

4°. The *logarithm* is defined as the inverse of the exponential function: if $e^w = z$, then

$$w = \operatorname{Ln} z. \tag{10.1.2.12}$$

This function is defined for $z \ne 0$. The logarithm satisfies the following relations:

$$\operatorname{Ln} z_1 + \operatorname{Ln} z_2 = \operatorname{Ln}(z_1 z_2), \quad \operatorname{Ln} z_1 - \operatorname{Ln} z_2 = \operatorname{Ln} \frac{z_1}{z_2},$$
$$\operatorname{Ln}(z^n) = n \operatorname{Ln} z, \quad \operatorname{Ln} \sqrt[n]{z} = \frac{1}{n} \operatorname{Ln} z.$$

The exponential form of complex numbers readily shows that the logarithm is infinite-valued:

$$\operatorname{Ln} z = \ln|z| + i \operatorname{Arg} z = \ln|z| + i \arg z + 2\pi k i, \quad k = 0, \pm 1, \pm 2, \ldots \tag{10.1.2.13}$$

The value $\ln z = \ln|z| + i \arg z$ is taken to be the principal value of this function. Just as with the function $w = \sqrt[n]{z}$, we see that if the point $z = 0$ is surrounded by a closed curve C, then the point $w = \operatorname{Ln} z$ does not return to its initial position after z goes around C in the positive sense, since the argument of w increases by $2\pi i$. Thus if a domain D does not contain a closed curve surrounding the point $z = 0$, then in D one can single out infinitely many continuous and single-valued branches of the multi-valued function $w = \operatorname{Ln} z$; the differences between the values of these branches at each point of the domain have the form $2\pi k i$, where k is an integer. This cannot be done in an arbitrary neighborhood of the point $z = 0$, and this point is called a *branch point* of the logarithm.

5°. *Trigonometric functions* are defined in terms of the exponential function as follows:

$$\cos z = \frac{e^{iz} + e^{-iz}}{2}, \qquad \sin z = \frac{e^{iz} - e^{-iz}}{2i},$$
$$\tan z = \frac{\sin z}{\cos z} = -i \frac{e^{iz} - e^{-iz}}{e^{iz} + e^{-iz}}, \qquad \cot z = \frac{\cos z}{\sin z} = i \frac{e^{iz} + e^{-iz}}{e^{iz} - e^{-iz}}. \tag{10.1.2.14}$$

Properties of the functions $\cos z$ and $\sin z$:

1. They are analytic for any z.
2. The usual differentiation rules are valid:
$$(\sin z)'_z = \cos z, \quad (\cos z)'_z = \sin z.$$
3. They are periodic with real period $T = 2\pi$.
4. $\sin z$ is an odd function, and $\cos z$ is an even function.
5. In the complex domain, they are unbounded.
6. The usual trigonometric relations hold:
$$\cos^2 z + \sin^2 z = 1, \quad \cos 2z = \cos^2 z - \sin^2 z, \quad \text{etc.}$$

The function $\tan z$ is analytic everywhere except for the points

$$z_k = \frac{\pi}{2} + k\pi, \qquad k = 0, \pm 1, \pm 2, \ldots,$$

and the function $\cot z$ is analytic everywhere except for the points

$$z_k = k\pi, \qquad k = 0, \pm 1, \pm 2, \ldots$$

The functions $\tan z$ and $\cot z$ are periodic with real period $T = \pi$.

6°. *Hyperbolic functions* are defined by the formulas

$$\begin{aligned} &\cosh z = \frac{e^z + e^{-z}}{2}, && \sinh z = \frac{e^z - e^{-z}}{2}, \\ &\tanh z = \frac{\sinh z}{\cosh z} = \frac{e^z - e^{-z}}{e^z + e^{-z}}, && \coth z = \frac{\cosh z}{\sinh z} = \frac{e^z + e^{-z}}{e^z - e^{-z}}. \end{aligned} \tag{10.1.2.15}$$

For real values of the argument, each of these functions coincides with the corresponding real function. Hyperbolic and trigonometric functions are related by the formulas

$$\cosh z = \cos iz, \quad \sinh z = -i \sin iz, \quad \tanh z = -i \tan iz, \quad \coth z = i \cot iz.$$

7°. *Inverse trigonometric and hyperbolic functions* are expressed via the logarithm and hence are infinite-valued:

$$\begin{aligned} &\operatorname{Arccos} z = -i \operatorname{Ln}(z + \sqrt{z^2 - 1}), && \operatorname{Arcsin} z = -i \operatorname{Ln}(iz + \sqrt{1 - z^2}), \\ &\operatorname{Arctan} z = -\frac{i}{2} \operatorname{Ln} \frac{1 + iz}{1 - iz}, && \operatorname{Arccot} z = -\frac{i}{2} \operatorname{Ln} \frac{z + i}{z - i}, \\ &\operatorname{arccosh} z = \operatorname{Ln}(z + \sqrt{z^2 - 1}), && \operatorname{arcsinh} z = \operatorname{Ln}(z + \sqrt{z^2 - 1}), \\ &\operatorname{arctanh} z = \frac{1}{2} \operatorname{Ln} \frac{1 + z}{1 - z}, && \operatorname{arccoth} z = \frac{1}{2} \operatorname{Ln} \frac{z + 1}{z - 1}. \end{aligned} \tag{10.1.2.16}$$

The principal value of each of these functions is obtained by choosing the principal value of the corresponding logarithmic function.

8°. The *power* $w = z^\gamma$ is defined by the relation

$$z^\gamma = e^{\gamma \operatorname{Ln} z}, \tag{10.1.2.17}$$

where $\gamma = \alpha + i\beta$ is an arbitrary complex number. Substituting $z = re^{i\varphi}$ into (10.1.2.17), we obtain

$$z^\gamma = e^{\alpha \ln r - \beta(\varphi + 2k\pi)} e^{i\alpha(\varphi + 2k\pi) + i\beta \ln r}, \qquad k = 0, \pm 1, \pm 2, \ldots \tag{10.1.2.18}$$

It follows from relation (10.1.2.18) that the function $w = z^\gamma$ has infinitely many values for $\beta \neq 0$.

9°. The *general exponential function* is defined by the formula

$$w = \gamma^z = e^{z\,\mathrm{Ln}\,\gamma} = e^{z\ln|\gamma|}e^{zi\,\mathrm{Arg}\,\gamma}, \tag{10.1.2.19}$$

where $\gamma = \alpha + i\beta$ is an arbitrary nonzero complex number. The function (10.1.2.19) is a set of separate mutually independent single-valued functions that differ from one another by the factors $e^{2k\pi iz}$, $k = 0, \pm1, \pm2, \ldots$

Example 3. Let us calculate the values of some elementary functions at specific points:

1. $\cos 2i = \frac{1}{2}(e^{2ii} + e^{-2ii}) = \frac{1}{2}(e^2 + e^{-2}) = \cosh 2 \approx 3.7622$.
2. $\ln(-2) = \ln 2 + i\pi$, since $|-2| = 2$ and the principal value of the argument is equal to π.
3. $\mathrm{Ln}(-2)$ is calculated by formula (10.1.2.13):

$$\mathrm{Ln}(-2) = \ln 2 + i\pi + 2\pi ki = \ln 2 + (1 + 2k)i\pi \quad (k = 0, \pm1, \pm2, \ldots).$$

4. $i^i = e^{i\,\mathrm{Ln}\,i} = e^{i(i\pi/2+2\pi k)} = e^{-\pi/2-2\pi k} \quad (k = 0, \pm1, \pm2, \ldots)$.

The main elementary functions $w = f(z) = u(x, y) + iv(x, y)$ of the complex variable $z = x + iy$ are given in Table 10.1.

10.1.2-4. Integration of function of complex variable.

Suppose that an oriented curve C connecting points $z = a$ and $z = b$ is given on the complex plane and a function $w = f(z)$ of the complex variable is defined on the curve. We divide the curve C into n parts, $a = z_0, z_1, \ldots, z_{n-1}, z_n = b$, arbitrarily choose $\xi_k \in [z_k, z_{k+1}]$, and compose the integral sum

$$\sum_{k=0}^{n-1} f(\xi_k)(z_{k+1} - z_k).$$

If there exists a limit of this sum as $\max|z_{k+1} - z_k| \to 0$, independent of the construction of the partition and the choice of points ξ_k, then this limit is called the *integral* of the function $w = f(z)$ over the curve C and is denoted by

$$\int_C f(z)\,dz. \tag{10.1.2.20}$$

Properties of the integral of a function of a complex variable:

1. If α, β are arbitrary constants, then $\int_C [\alpha f(z) + \beta g(z)]\,dz = \alpha \int_C f(z)\,dz + \beta \int_C g(z)\,dz$.
2. If $\widetilde{C}$ is the same curve as C but with the opposite sense, then $\int_{\widetilde{C}} f(z)\,dz = -\int_C f(z)\,dz$.
3. If $C = C_1 \cup \cdots \cup C_n$, then $\int_C f(z)\,dz = \int_{C_1} f(z)\,dz + \ldots + \int_{C_n} f(z)\,dz$.
4. If $|f(z)| \le M$ at all points of the curve C, then the following *estimate of the absolute value of the integral* holds: $\left|\int_C f(z)\,dz\right| \le Ml$, where l is the length of the curve C.

If C is a piecewise smooth curve and $f(z)$ is bounded and piecewise continuous, then the integral (10.1.2.20) exists. If $z = x + iy$ and $w = u(x, y) + iv(x, y)$, then the computation of the integral (10.1.2.20) is reduced to finding two ordinary curvilinear integrals:

$$\int_C f(z)\,dz = \int_C u(x, y)\,dx - v(x, y)\,dy + i\int_C v(x, y)\,dx + u(x, y)\,dy. \tag{10.1.2.21}$$

Remark. Formula (10.1.2.21) can be rewritten in a form convenient for memorizing:

$$\int_C f(z)\,dz = \int_C (u + iv)(dx + i\,dy).$$

TABLE 10.1
Main elementary functions $w = f(z) = u(x, y) + iv(x, y)$ of complex variable $z = x + iy$

No.	Complex function $w = f(z)$	Algebraic form $f(z) = u(x, y) + iv(x, y)$	Zeros of nth order	Singularities
1	z	$x + iy$	$z = 0,\ n = 1$	$z = \infty$ is a first-order pole
2	z^2	$x^2 - y^2 + i\,2xy$	$z = 0,\ n = 2$	$z = \infty$ is a second-order pole
3	$\dfrac{1}{z-(x_0+iy_0)}$ (x_0, y_0 are real numbers)	$\dfrac{x-x_0}{(x-x_0)^2+(y-y_0)^2} + i\dfrac{-(y-y_0)}{(x-x_0)^2+(y-y_0)^2}$	$z = \infty,\ n = 1$	$z = x_0 + iy_0$ is a first-order pole
4	$\dfrac{1}{z^2}$	$\dfrac{x^2-y^2}{(x^2+y^2)^2} + i\dfrac{-2xy}{(x^2+y^2)^2}$	$z = \infty,\ n = 2$	$z = 0$ is a second-order pole
5	$\sqrt{z}$	$\pm\left[\left(\dfrac{x+\sqrt{x^2+y^2}}{2}\right)^{1/2} + i\left(\dfrac{-x+\sqrt{x^2+y^2}}{2}\right)^{1/2}\right]$	$z = 0$ is a branch point	$z = 0$ is a first-order branch point $z = \infty$ is a first-order branch point
6	e^z	$e^x \cos y + ie^x \sin y$	—	$z = \infty$ is an essential singular point
7	$\operatorname{Ln} z$	$\ln\lvert z\rvert + i(\arg z + 2k\pi)$, $k = 0, \pm1, \pm2, \ldots$	$z = 1,\ n = 1$ (for the branch corresponding to $k = 0$)	Logarithmic branch points for $z = 0$, $z = \infty$
8	$\sin z$	$\sin x \cosh y + i \cos x \sinh y$	$z = \pi k,\ n = 1$ ($k = 0, \pm1, \pm2, \ldots$)	$z = \infty$ is an essential singular point
9	$\cos z$	$\cos x \cosh y + i(-\sin x \sinh y)$	$z = \frac{1}{2}\pi + \pi k,\ n = 1$ ($k = 0, \pm1, \pm2, \ldots$)	$z = \infty$ is an essential singular point
10	$\tan z$	$\dfrac{\sin 2x}{\cos 2x + \cosh 2y} + i\dfrac{\sinh 2y}{\cos 2x + \cosh 2y}$	$z = \pi k,\ n = 1$ ($k = 0, \pm1, \pm2, \ldots$)	$z = \frac{1}{2}\pi + \pi k$ ($k = 0, \pm1, \pm2, \ldots$) are first-order poles

If the curve C is given by the parametric equations $x = x(t)$, $y = y(t)$ $(t_1 \le t \le t_2)$, then

$$\int_C f(z)\,dz = \int_{t_1}^{t_2} f(z(t))z'_t(t)\,dt, \tag{10.1.2.22}$$

where $z = z(t) = x(t) + iy(t)$ is the complex parametric equation of the curve C.

If $f(z)$ is an analytic function in a simply connected domain D containing the points $z = a$ and $z = b$, then the *Newton–Leibniz formula* holds:

$$\int_a^b f(z)\,dz = F(b) - F(a), \tag{10.1.2.23}$$

where $F(z)$ is a primitive of the function $f(z)$, i.e., $F'_z(z) = f(z)$ in the domain D.

If $f(z)$ and $g(z)$ are analytic functions in a simply connected domain D and $z = a$ and $z = b$ are arbitrary points of the domain D, then the *formula of integration by parts* holds:

$$\int_a^b f(z)\,dg(z) = f(b)g(b) - f(a)g(a) - \int_a^b g(z)\,df(z). \tag{10.1.2.24}$$

If an analytic function $z = g(w)$ determines a single-valued mapping of a curve $\widetilde{C}$ onto a curve C, then

$$\int_C f(z)\,dz = \int_{\widetilde{C}} f(g(w))g'_w(w)\,dw. \tag{10.1.2.25}$$

CAUCHY'S THEOREM FOR A SIMPLY CONNECTED DOMAIN. *If a function $f(z)$ is analytic in a simply connected domain D bounded by a contour C and is continuous in $\overline{D}$, then $\int_C f(z)\,dz = 0$.*

CAUCHY'S THEOREM FOR A MULTIPLY CONNECTED DOMAIN. *If a function $f(z)$ is analytic in a multiply connected domain D bounded by a contour Γ consisting of several closed curves and is continuous in $\overline{D}$, then $\int_\Gamma f(z)\,dz = 0$ provided that the sense of all curves forming Γ is chosen in such a way that the domain D lies to one side of the contour.*

If a function $f(z)$ is analytic in an n-connected domain D and continuous in $\overline{D}$, and C is the boundary of D, then for any interior point z of this domain the *Cauchy integral formula* holds:

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f(\xi)}{\xi - z}\,d\xi. \tag{10.1.2.26}$$

(Here integration is in the positive sense of C; i.e., the domain D lies to the left of C.) Under the same assumptions as above, formula (10.1.2.26) implies expressions for the value of the derivative of arbitrary order of the function $f(z)$ at any interior point z of the domain:

$$f_z^{(n)}(z) = \frac{n!}{2\pi i}\int_C \frac{f(\xi)}{(\xi - z)^{n+1}}\,d\xi \qquad (n = 1, 2, \ldots). \tag{10.1.2.27}$$

For an arbitrary smooth curve C, not necessarily closed, and for a function $f(\xi)$ everywhere continuous on C, possibly except for finitely many points at which this function has an integrable discontinuity, the right-hand side of formula (10.1.2.26) defines a *Cauchy-type integral*. The function $F(z)$ determined by a Cauchy-type integral is analytic at any point that does not belong to C. If C divides the plane into several domains, then the Cauchy-type integral generally determines different analytic functions in these domains.

Formulas (10.1.2.26) and (10.1.2.27) allow one to calculate the integrals

$$\int_C \frac{f(\xi)}{\xi - z}\,d\xi = 2\pi i f(z), \qquad \int_C \frac{f(\xi)}{(\xi - z)^n}\,d\xi = \frac{2\pi i}{n!} f_z^{(n)}(z). \tag{10.1.2.28}$$

Example 4. Let us calculate the integral

$$\int_C \operatorname{Im} z\,dz,$$

where C is the semicircle $|z| = 1$, $0 \le \arg z \le \pi$ (Fig. 10.4).

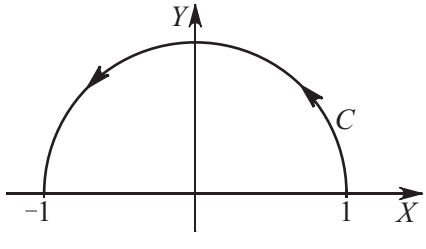


Figure 10.4. The semicircle $|z| = 1$, $0 \le \arg z \le \pi$.

Using formula (10.1.2.21), we obtain

$$\int_C \operatorname{Im} z\, dz = \int_C y(dx + i\, dy) = \int_C y\, dx + i \int_C y\, dy = \int_1^{-1} \sqrt{1-x^2}\, dx - i \int_1^{-1} x\, dx = -\frac{\pi}{2}.$$

Example 5. Let us calculate the integral

$$\int_C \frac{dz}{z - z_0},$$

where C is the circle of radius R centered at a point z_0 with anticlockwise sense.

Using the Cauchy integral formula (10.1.2.28), we obtain

$$\int_C \frac{1}{z - z_0}\, dz = 2\pi i.$$

Example 6. Let us calculate the integral

$$\int_C \frac{dz}{z^2 + 1},$$

where C is the circle of unit radius centered at the point i with anticlockwise sense.

To apply the Cauchy integral formula (10.1.2.26), we transform the integrand as follows:

$$\frac{1}{1+z^2} = \frac{1}{(z-i)(z+i)} = \frac{1}{z+i}\frac{1}{z-i} = \frac{f(z)}{z+i}, \quad f(z) = \frac{1}{z+i}.$$

The function $f(z) = 1/(z+i)$ is analytic in the interior of the domain under study and on its boundary; hence the Cauchy integral formula (10.1.2.26) and the first of formulas (10.1.2.28) hold. From the latter formula, we obtain

$$\int_C \frac{dz}{z^2+1} = \int_C \frac{f(z)}{z-i}\, dz = 2\pi i f(i) = 2\pi i \frac{1}{2i} = \pi.$$

Formulas (10.1.2.26) and (10.1.2.27) imply the *Cauchy inequalities*

$$|f_z^{(n)}(z)| \le \frac{n!}{2\pi}\left|\int_C \frac{f(\xi)}{(\xi - z)^{n+1}}\, d\xi\right| \le \frac{n! M l}{2\pi R^{n+1}}, \tag{10.1.2.29}$$

where $M = \max\limits_{z \in D} |f(z)|$ is the maximum modulus of the function $f(z)$ in the domain D, R is the distance from the point z to the boundary C, and l is the length of the boundary C.

If, in particular, $f(z)$ is analytic in the disk $D = |z - z_0| < R$, and bounded in $\bar{D}$, then we obtain the inequality

$$|f_z^{(n)}(z_0)| \le \frac{n! M}{R^n} \quad (n = 0, 1, 2, \ldots). \tag{10.1.2.30}$$

MORERA'S THEOREM. *If a function $f(z)$ is continuous in a simply connected domain D and $\int_C f(z)\, dz = 0$ for any closed curve C lying in D, then $f(z)$ is analytic in the domain D.*

10.1.2-5. Taylor and Laurent series.

If a series

$$\sum_{n=0}^{\infty} f_n(z) \tag{10.1.2.31}$$

of analytic functions in a simply connected domain D converges uniformly in this domain, then its sum is analytic in the domain D.

If a series (10.1.2.31) of functions analytic in a domain D and continuous in $\overline{D}$ converges uniformly in D, then it can be differentiated termwise any number of times and can be integrated termwise over any piecewise smooth curve C lying in D.

ABEL'S THEOREM. *If the power series*

$$\sum_{n=0}^{\infty} c_n (z-a)^n \tag{10.1.2.32}$$

converges at a point z_0, *then it also converges at any point* z *satisfying the condition* $|z-a| < |z_0-a|$. *Moreover, the series converges uniformly in any disk* $|z-a| < q|z_0-a|$, *where* $0 < q < 1$.

It follows from Abel's theorem that the domain of convergence of a power series is an open disk centered at the point a; moreover, this disk can fill the entire plane. The radius of this disk is called the *radius of convergence* of a power series. The sum of the power series inside the disk of convergence is an analytic function.

Remark. The radius of convergence R can be found by the *Cauchy–Hadamard formula*

$$\frac{1}{R} = \overline{\lim_{n\to\infty}} \sqrt[n]{|c_n|},$$

where $\overline{\lim}$ denotes the upper limit.

If a function $f(z)$ is analytic in the open disk D of radius R centered at a point $z = a$, then this function can be represented in this disk by its *Taylor series*

$$f(z) = \sum_{n=0}^{\infty} c_n (z-a)^n,$$

whose coefficients are determined by the formulas

$$c_n = \frac{f_z^{(n)}(a)}{n!} = \frac{1}{2\pi i} \int_C \frac{f(\xi)}{(\xi - z)^{n+1}}\, d\xi \qquad (n = 0, 1, 2, \ldots), \tag{10.1.2.33}$$

where C is the circle $|z-a| = qR$, $0 < q < 1$. In any closed domain belonging to the disk D, the Taylor series converges uniformly. Any power series expansion of an analytic function is its Taylor expansion. The Taylor series expansions of the functions given in Paragraph 10.1.2-3 in powers of z have the form

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \ldots \quad (|z| < \infty), \tag{10.1.2.34}$$

$$\cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \ldots, \quad \sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \ldots \quad (|z| < \infty), \tag{10.1.2.35}$$

$$\cosh z = 1 + \frac{z^2}{2!} + \frac{z^4}{4!} + \ldots, \quad \sinh z = z + \frac{z^3}{3!} + \frac{z^5}{5!} + \ldots \quad (|z| < \infty), \tag{10.1.2.36}$$

$$\ln(1+z) = z - \frac{z^2}{2!} + \frac{z^3}{3!} - \ldots \quad (|z| < 1), \tag{10.1.2.37}$$

$$(1+z)^a = 1 + az + \frac{a(a-1)}{2!} z^2 + \frac{a(a-1)(a-2)}{3!} z^3 + \ldots \quad (|z| < 1). \tag{10.1.2.38}$$

The last two expansions are valid for the single-valued branches for which the values of the functions for $z = 0$ are equal to 0 and 1, respectively.

Remark. Series expansions (10.1.2.34)–(10.1.2.38) coincide with analogous expansions of the corresponding elementary functions of the real variable (see Paragraph 8.3.2-3).

To obtain the Taylor series for other branches of the multi-valued function $\operatorname{Ln}(1+z)$, one has to add the numbers $2k\pi i$, $k = \pm 1, \pm 2, \ldots$ to the expression in the right-hand side:

$$\operatorname{Ln}(1+z) = z - \frac{z^2}{2!} + \frac{z^3}{3!} - \ldots + 2k\pi.$$

The domain of convergence of the function series $\sum\limits_{n=-\infty}^{\infty} c_n(z-a)^n$ is a circular annulus $K : r < |z-a| < R$, where $0 \le r \le \infty$ and $0 \le R \le \infty$. The sum of the series is an analytic function in the annulus of convergence. Conversely, in any annulus K where the function $f(z)$ is analytic, this function can be represented by the *Laurent series* expansion

$$f(z) = \sum_{n=-\infty}^{\infty} c_n(z-a)^n$$

with coefficients determined by the formulas

$$c_n = \frac{1}{2\pi i} \int_\gamma \frac{f(\xi)}{(\xi - z)^{n+1}}\, d\xi \quad (n = 0, \pm 1, \pm 2, \ldots), \tag{10.1.2.39}$$

where γ is the circle $|z-a| = \rho$, $r < \rho < R$. In any closed domain contained in the annulus K, the Laurent series converges uniformly.

The part of the Laurent series with negative numbers,

$$\sum_{n=-\infty}^{-1} c_n(z-a)^n = \sum_{n=1}^{\infty} \frac{c_{-n}}{(z-a)^n}, \tag{10.1.2.40}$$

is called its *principal part*, and the part with nonnegative numbers,

$$\sum_{n=0}^{\infty} c_n(z-a)^n, \tag{10.1.2.41}$$

is called the *regular part*. Any expansion of an analytic function in positive and negative powers of $z-a$ is its Laurent expansion.

Example 7. Let us consider Laurent series expansions of the function

$$f(z) = \frac{1}{z(1-z)}$$

in a Laurent series in the domain $0 < |z| < 1$. This function is analytic in the annulus $0 < |z| < 1$ and hence can be expanded in the corresponding Laurent series. We write this function as the sum of elementary fractions:

$$f(z) = \frac{1}{z(1-z)} = \frac{1}{z} + \frac{1}{1-z}.$$

Since $|z| < 1$, we can use formula (10.1.2.39) and obtain the expansion

$$\frac{1}{z(1-z)} = \frac{1}{z} + 1 + z + z^2 + \ldots$$

Example 8. Let us consider Laurent series expansions of the function

$$f(z) = e^{1/z}$$

in a Laurent series in a neighborhood of the point $z_0 = 0$. To this end, we use the well-known expansion (10.1.2.34), where we should replace z by $1/z$. Thus we obtain

$$e^{1/z} = 1 + \frac{1}{1!z} + \frac{1}{2!z^2} + \ldots + \frac{1}{n!z^n} + \ldots \qquad (z \ne 0).$$

10.1.2-6. Zeros and isolated singularities of analytic functions.

A point $z = a$ is called a *zero* of a function $f(z)$ if $f(a) = 0$. If $f(z)$ is analytic at the point a and is not zero identically, then the least order of nonzero coefficients in the Taylor expansion of $f(z)$ centered at a, in other words, the number n of the first nonzero derivative $f^{(n)}(a)$, is called the *order of zero* of this function. In a neighborhood of a zero a of order n, the Taylor expansion of $f(z)$ has the form

$$f(z) = c_n(z-a)^n + c_{n+1}(z-a)^{n+1} + \dots \qquad (c_n \neq 0,\ n \geq 1).$$

In this case, $f(z) = c_n(z-a)^n g(z)$, where the function $g(z)$ is analytic at the point a and $g(a) \neq 0$. A first-order zero is said to be *simple*. The point $z = \infty$ is a zero of order n for a function $f(z)$ if $z = 0$ is a zero of order n for $F(z) = f(1/z)$.

If a function $f(z)$ is analytic at a point a and is not identically zero in any neighborhood of a, then there exists a neighborhood of a in which $f(z)$ does not have any zeros other than a.

UNIQUENESS THEOREM. *If functions $f(z)$ and $g(z)$ are analytic in a domain D and their values coincide on some sequence a_k of points converging to an interior point a of the domain D, then $f(z) \equiv g(z)$ everywhere in D.*

ROUCHÉ'S THEOREM. *If functions $f(z)$ and $g(z)$ are analytic in a simply connected domain D bounded by a curve C, are continuous in $\overline{D}$, and satisfy the inequality $|f(z)| > |g(z)|$ on C, then the functions $f(z)$ and $f(z) + g(z)$ have the same number of zeros in D.*

A point a is called an *isolated singularity* of a single-valued analytic function $f(z)$ if there exists a neighborhood of this point in which $f(z)$ is analytic everywhere except for the point a itself. The point a is called

1. A *removable singularity* if $\lim\limits_{z\to a} f(z)$ exists and is finite.
2. A *pole* if $\lim\limits_{z\to a} f(z) = \infty$.
3. An *essential singularity* if $\lim\limits_{z\to a} f(z)$ does not exist.

A necessary and sufficient condition for a point a to be a removable singularity of a function $f(z)$ is that the Laurent expansion of $f(z)$ around a does not contain the principal part. If a function $f(z)$ is bounded in a neighborhood of an isolated singularity a, then a is a removable singularity of this function.

A necessary and sufficient condition for a point a to be a pole of a function $f(z)$ is that the principal part of the Laurent expansion of $f(z)$ around a contains finitely many terms:

$$f(z) = \frac{c_{-n}}{(z-a)^n} + \dots + \frac{c_{-1}}{(z-a)} + \sum_{k=0}^{\infty} c_k(z-a)^k. \qquad (10.1.2.42)$$

The *order of a pole* a of a function $f(z)$ is defined to be the order of the zero of the function $F(z) = f(1/z)$. If $c_{-n} \neq 0$ in expansion (10.1.2.42), then the order of the pole a of the function $f(z)$ is equal to n. For $n = 1$, we have a *simple pole*.

A necessary and sufficient condition for a point a to be an essential singularity of a function $f(z)$ is that the principal part of the Laurent expansion of $f(z)$ around a contains infinitely many nonzero terms.

SOKHOTSKII'S THEOREM. *If a is an essential singularity of a function $f(z)$, then for each complex number A there exists a sequence of points $z_k \to a$ such that $f(z_k) \to A$.*

Example 9. Let us consider some functions with different singular points.

1°. The function $f(z) = (1 - \cos z)/z^2$ has a removable singularity at the origin, since its Laurent expansion about the origin,

$$\frac{1-\cos z}{z^2} = \frac{1}{2} - \frac{z^2}{24} + \frac{z^4}{720} - \dots,$$

does not contain the principal part.

2°. The function $f(z) = 1/(1+e^{z^2})$ has infinitely many poles at the points $z = \pm\sqrt{(2k+1)\pi i}$ $(k = 0, \pm1, \pm2, \dots)$. All these poles are simple poles, since the function $1/f(z) = 1 + e^{z^2}$ has simple zeros at these points. (Its derivative is nonzero at these points.)

3°. The function $f(z) = \sin(1/z)$ has an essential singularity at the origin, since the principal part of its Laurent expansion

$$\sin\frac{1}{z} = \frac{1}{z} - \frac{1}{z^3 3!} + \dots$$

contains infinitely many terms.

The following two simplest classes of single-valued analytic functions are distinguished according to the character of singular points.

1. *Entire functions*. A function $f(z)$ is said to be *entire* if it does not have singular points in the finite part of the plane. An entire function can be represented by an everywhere convergent power series

$$f(z) = \sum_{n=0}^{\infty} c_n z^n.$$

An entire function can have only one singular point at $z = \infty$. If this singularity is a pole of order n, then $f(z)$ is a polynomial of degree n. If $z = \infty$ is an essential singularity, then $f(z)$ is called an *entire transcendental function*. If $z = \infty$ is a regular point (i.e., $f(z)$ is analytic for all z), then $f(z)$ is constant (*Liouville's theorem*). All polynomials, the exponential function, $\sin z$, $\cos z$, etc. are examples of entire functions. Sums, differences, and products of entire functions are themselves entire functions.

2. *Meromorphic functions*. A function $f(z)$ is said to be *meromorphic* if it does not have any singularities except for poles. The number of these poles in each finite closed domain D is always finite.

Suppose that a function $f(z)$ is analytic in a neighborhood of the point at infinity. The definition of singular points can be generalized to this function without any changes. But the criteria for the type of a singular point at infinity related to the Laurent expansion are different.

THEOREM. *In the case of a removable singularity at the point at infinity, the Laurent expansion of a function $f(z)$ in a neighborhood of this point does not contain positive powers of z. In the case of a pole, it contains finitely many positive powers of z. In the case of an essential singularity, it contains infinitely many powers of z.*

Let $f(z)$ be a multi-valued function defined in a neighborhood D of a point $z = a$ except possibly for the point a itself, and let $f_1(z)$, $f_2(z)$ . . . be its branches, which are single-valued continuous functions in the domain where they are defined. The point a is called a *branch point (ramification point)* of the function $f(z)$ if $f(z)$ passes from one branch to another as the point z goes along a closed curve around the point z in a neighborhood of D. If the original branch is reached again for the first time after going around this curve m times (in the same sense), then the number $m - 1$ is called the *order* of the branch point, and the point a itself is called a *branch point of order* $m - 1$.

If all branches $f_k(z)$ tend to the same finite or infinite limit as $z \to a$, then the point a is called an *algebraic* branch point. (For example, the point $z = 0$ is an algebraic branch point of the function $f(z) = \sqrt[m]{z}$.) In this case, the single-valued function

$$F(z) = f(z^m + a)$$

has a regular point or a pole for $z = 0$.

If the limit of $f_k(z)$ as $z \to a$ does not exist, then the point a is called a *transcendental* branch point. For example, the point $z = 0$ is a transcendental branch point of the function $f(z) = \exp(\sqrt[m]{1/z})$.

In a neighborhood of a branch point a of finite order, the function $f(z)$ can be expanded in a *fractional power series* (*Puiseux series*)

$$f(z) = \sum_{k=-\infty}^{\infty} c_k (z-a)^{k/m}. \tag{10.1.2.43}$$

If a new branch is obtained each time after going around this curve (in the same sense), then the point a is called a *branch point of infinite order (a logarithmic branch point)*. For example, the points $z = 0$ and $z = \infty$ are logarithmic branch points of the multivalued function $w = \operatorname{Ln} z$. A logarithmic branch point is classified as a transcendental branch point.

For $a \neq \infty$, the expansion (10.1.2.43) contains finitely many terms with negative k (infinitely many in the case of a transcendental point).

10.1.2-7. Residues.

The *residue* $\operatorname{res} f(a)$ of a function $f(z)$ at an isolated singularity a is defined as the number

$$\operatorname{res} f(a) = \frac{1}{2\pi i} \oint_C f(z)\, dz, \tag{10.1.2.44}$$

where the integral is taken in the positive sense over a contour C surrounding the point a and containing no other singularities of $f(z)$ in the interior.

Remark. Residues are sometimes denoted by $\operatorname{res}[f(z); a]$ or $\operatorname{res}_{z=a} f(z)$.

The residue $\operatorname{res} f(a)$ of a function $f(z)$ at a singularity a is equal to the coefficient of $(z-a)^{-1}$ in the Laurent expansion of $f(z)$ in a neighborhood of the point a,

$$\operatorname{res} f(a) = \frac{1}{2\pi i} \oint_C f(z)\, dz = c_{-1}. \tag{10.1.2.45}$$

Basic rules for finding the residues:

1. The residue of a function at a removable singularity is always zero.
2. If a is a pole of order n, then

$$\operatorname{res} f(a) = \frac{1}{(n-1)!} \lim_{z \to a} \frac{d^{n-1}}{dz^{n-1}} \big[f(z)(z-a)^n \big]. \tag{10.1.2.46}$$

3. For a simple pole ($n = 1$),

$$\operatorname{res} f(a) = \lim_{z \to a} \big[f(z)(z-a)^n \big]. \tag{10.1.2.47}$$

4. If $f(z)$ is the quotient of two analytic functions,

$$f(z) = \frac{\varphi(z)}{\psi(z)},$$

in a neighborhood of a point a and $\varphi(a) \neq 0$, $\psi(a) = 0$, but $\psi'_z(a) \neq 0$ (i.e., a is a simple pole of $f(z)$), then

$$\operatorname{res} f(a) = \frac{\varphi(a)}{\psi'_z(a)}. \tag{10.1.2.48}$$

5. If a is an essential singularity of $f(z)$, then to obtain $\operatorname{res} f(a)$, one has to find the coefficient c_{-1} in the Laurent expansion of $f(z)$ in a neighborhood of a.

A function $f(z)$ is said to be continuous on the boundary C of the domain D if for each boundary point z_0 there exists a limit $\lim\limits_{z \to z_0} f(z) = f(z_0)$ as $z \to z_0$, $z \in D$.

CAUCHY'S RESIDUE THEOREM. *Let $f(z)$ be a function continuous on the boundary C of a domain D and analytic in the interior of D everywhere except for finitely many points $a_1,\ldots,a_n$. Then*

$$\int_C f(z)\,dz = 2\pi i \sum_{k=1}^{n} \operatorname{res} f(a_k), \tag{10.1.2.49}$$

where the integral is taken in the positive sense of C.

The *logarithmic residue* of a function $f(z)$ at a point a is by definition the residue of its logarithmic derivative

$$\bigl[\ln f(z)\bigr]'_z = \frac{f'_z(z)}{f(z)}.$$

THEOREM. *The logarithmic derivative $f'_z(z)/f(z)$ has first-order poles at the zeros and poles of $f(z)$. Moreover, the logarithmic residue of $f(z)$ at a zero or a pole of $f(z)$ is equal to the order of the zero or minus the order of the pole, respectively.*

The residue of a function $f(z)$ at infinity is defined as

$$\operatorname{res} f(\infty) = \frac{1}{2\pi i} \oint_{\Gamma} f(z)\,dz, \tag{10.1.2.50}$$

where Γ is a circle of sufficiently large radius $|z| = \rho$ and the integral is taken in the clockwise sense (so that the neighborhood of the point $z = \infty$ remains to the left of the contour, just as in the case of a finite point).

The residue of $f(z)$ at infinity is equal to minus the coefficient of z^{-1} in the Laurent expansion of $f(z)$ in a neighborhood of the point $z = \infty$,

$$\operatorname{res} f(\infty) = -c_{-1}. \tag{10.1.2.51}$$

Note that

$$\operatorname{res} f(\infty) = \lim_{z \to \infty} [-z f(z)], \tag{10.1.2.52}$$

provided that this limit exists.

THEOREM. *If a function $f(z)$ has finitely many singular points $a_1,\ldots,a_n$ in the extended complex plane, then the sum of all its residues, including the residue at infinity, is zero:*

$$\operatorname{res} f(\infty) + \sum_{k=1}^{n} \operatorname{res} f(a_k) = 0. \tag{10.1.2.53}$$

Example 10. Let us calculate the integral

$$\oint_C \frac{\ln(z+2)}{z^2}\,dz,$$

where C is the circle $|z| = \frac{1}{2}$.

In the disk $|z| \le \frac{1}{2}$, there is only one singular point of the integrand, $z = 0$, which is a second-order pole. The residue of $f(z)$ at $z = 0$ is calculated by the formula (10.1.2.46)

$$\operatorname{res} f(0) = \lim_{z\to 0}\Big[z^2\frac{\ln(z+2)}{z^2}\Big]'_z = \lim_{z\to 0}[\ln(z+2)]'_z = \lim_{z\to 0}\frac{1}{z+2} = \frac{1}{2}.$$

Using formula (10.1.2.44), we obtain

$$\frac{1}{2} = \frac{1}{2\pi i}\oint_C \frac{\ln(z+2)}{z^2}\,dz, \quad \oint_C \frac{\ln(z+2)}{z^2}\,dz = \pi i.$$

10.1.2-8. Calculation of definite integrals.

Suppose that we need to calculate the integral of a real function $f(x)$ over a (finite or infinite) interval (a, b). Let us supplement the interval (a, b) with a curve Γ that, together with (a, b), bounds a domain D, and then analytically continue the function $f(x)$ into $\overline{D}$. Then the residue theorem can be applied to this analytic continuation of $f(z)$, and by this theorem,

$$\int_a^b f(x)\,dx + \int_\Gamma f(z)\,dz = 2\pi i\Lambda, \tag{10.1.2.54}$$

where Λ is the sum of residues of $f(z)$ in D. If $\int_\Gamma f(z)\,dz$ can be calculated or expressed in terms of the desired integral $\int_a^b f(x)\,dx$, then the problem will be solved.

When calculating integrals of the form $\int_{-\infty}^{\infty} f(x)\,dx$, one should apply (10.1.2.49) to the contour C that consists of the interval $(-R, R)$ of the real axis and the arc C_R of the circle $|z| = R$ in the upper half-plane. Sometimes it is only possible to find the limit as $R \to \infty$ of the integral over the contour C_R rather than to calculate it, and often it turns out that the limit of this integral is equal to zero.

The integral over the curve Γ can be estimated using the following lemmas.

JORDAN LEMMA. *If a function $g(z)$ tends to zero uniformly with respect to* $\arg z$ *along a sequence of circular arcs C_{R_n} : $|z| = R_n$, $\operatorname{Im} z > -a$ (where $R_n \to \infty$ and a is fixed), then*

$$\lim_{n\to\infty}\int_{C_{R_n}} g(z)e^{imz}\,dz = 0 \tag{10.1.2.55}$$

for each positive number m.

If a function $f(z)$ is analytic for $|z| > R_0$ and $zf(z) \to 0$ as $|z| \to \infty$ for $y \ge 0$, then

$$\lim_{R\to\infty}\int_{C_R} f(z)\,dz = 0, \tag{10.1.2.56}$$

where C_R is the arc of the circle $|z| = R$ in the upper half-plane.

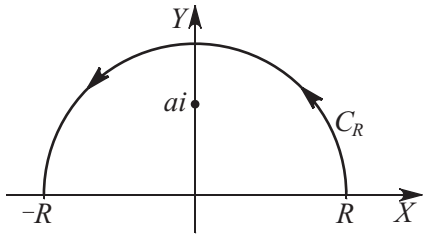


Figure 10.5. The contour to calculate the Laplace integral.

Example 11 (Laplace integral). To calculate the integral

$$\int_0^\infty \frac{\cos x}{x^2+a^2}\,dx,$$

one uses the auxiliary function

$$f(z)=\frac{e^{iz}}{z^2+a^2}=g(z)e^{iz},\quad g(z)=\frac{1}{z^2+a^2}$$

and the contour shown in Fig. 10.5. Since $g(z)$ satisfies the inequality $|g(z)|<(R^2-a^2)^{-1}$ on C_R, it follows that this function uniformly tends to zero as $R\to\infty$, and by the Jordan lemma we obtain

$$\int_{C_R} f(z)\,dz=\int_{C_R} g(z)e^{iz}\,dz\to 0$$

as $R\to\infty$.

By the residue theorem,

$$\int_{-R}^{R}\frac{e^{ix}}{x^2+a^2}\,dx+\int_{C_R} f(z)\,dz=2\pi i\frac{e^{-a}}{2ai}$$

for each $R>0$. (The residue at the singular point $z=ai$ of the function $f(z)$, which is a first-order pole and which is the only singular point of this function lying inside the contour, can be calculated by formula (10.1.2.48).) In the limit as $R\to\infty$, we obtain

$$\int_{-\infty}^{\infty}\frac{e^{ix}}{x^2+a^2}\,dx=\frac{\pi}{ae^a}.$$

Separating the real part and using the fact that the function is even, we obtain

$$\int_0^\infty \frac{\cos x}{x^2+a^2}\,dx=\frac{\pi}{2ae^a}.$$

10.1.2-9. Analytic continuation.

Let two domains D_1 and D_2 have a common part γ of the boundary, and let single-valued analytic functions $f_1(z)$ and $f_2(z)$, respectively, be given in these domains. The function $f_2(z)$ is called a *direct analytic continuation* of $f_1(z)$ into the domain D_2 if there exists a function $f(z)$ analytic in the domain $D_1\cup\gamma\cup D_2$ and satisfying the condition

$$f(z)=\begin{cases} f_1(z) & \text{for } z\in D_1,\\ f_2(z) & \text{for } z\in D_2.\end{cases}\qquad (10.1.2.57)$$

If such a continuation is possible, then the function $f(z)$ is uniquely determined. If the domains are simply connected and the functions $f_1(z)$ and $f_2(z)$ are continuous in $D_1\cup\gamma$ and $D_2\cup\gamma$, respectively, and coincide on γ, then $f_2(z)$ is the direct analytic continuation of $f_1(z)$ into the domain D_2. In addition, suppose that the domains D_1 and D_2 are allowed to have common interior points. A function $f_2(z)$ is called a direct analytic continuation of $f_1(z)$ through γ if $f_1(z)$ and $f_2(z)$ are continuous in $D_1\cup\gamma$ and $D_2\cup\gamma$, respectively, and their values on γ coincide. At the common interior points of D_1 and D_2, the function determined by relation (10.1.2.57) can be double-valued.

10.2. Main Applications

10.2.1. Conformal Mappings

10.2.1-1. Generalities.

A one-to-one mapping

$$w = f(z) = u(x, y) + iv(x, y) \tag{10.2.1.1}$$

of a domain D onto a domain D^* is said to be *conformal* if the principal linear part of this mapping at any point of D is an orthogonal orientation-preserving transformation.

Main properties of conformal mappings:

1. *Circular property.* A conformal mapping takes infinitesimal circles to infinitesimal circles (up to higher-order infinitesimals).
2. *Angle preservation property.* A conformal mapping preserves the angles between intersecting curves at points of intersection.

THEOREM. *A function $w = f(z)$ is a conformal mapping of a domain D if and only if it is analytic and schlicht in D and the derivative $f'_z(z)$ vanishes nowhere in D.*

The *main problem in the theory of conformal mappings* is as follows: given domains D and D^*, construct a function that gives a conformal mapping of one of the domains onto the other.

THE MAIN THEOREM OF THE THEORY OF CONFORMAL MAPPINGS (RIEMANN THEOREM). *For any simply connected domains D and D^* (with boundaries consisting of more than a single point), any points $z_0 \in D$ and $w_0 \in D^*$, and any real number α_0, there exists a unique conformal mapping*

$$w = f(z)$$

of D onto D^* such that

$$f(z_0) = w_0, \quad \arg f'_z(z_0) = \alpha_0.$$

10.2.1-2. Boundary correspondence.

On the boundary C of a domain D, let us introduce a real arc length parameter s reckoned from some point of C, so that $\zeta = \zeta(s)$ on C. If $f(z)$ is a continuous function in the closed domain $\overline{D}$, then on the boundary C one can set

$$f(\zeta) = f[\zeta(s)] = \varphi(s).$$

The function $\varphi(s)$ is called the *boundary function* for $f(z)$.

THEOREM ON THE BOUNDARY CORRESPONDENCE. *Suppose that a function $w = f(z)$ specifies a conformal mapping between domains D and D^*. Then the following assertions hold.*

1. *If the boundary of D^* does not have infinite branches, then $f(z)$ is continuous on the boundary of D and the boundary function $w = f(\zeta) = \varphi(s)$ is a continuous one-to-one correspondence between the boundaries of the domains D and D^*.*
2. *If the boundaries of D and D^* do not contain infinite branches and have a continuous curvature at each point, then the boundary function $\varphi(s)$ is continuously differentiable.*

BOUNDARY CORRESPONDENCE PRINCIPLE. *Let D and D^* be two simply connected domains with boundaries C and C^*, and let the domain D^* be bounded. Suppose that a function $w = f(z)$ satisfies the following conditions:*

1. *It is analytic in D and continuous in $\overline{D}$. If the point at infinity lies in the interior of the domain D^*, then the boundary correspondence principle remains valid provided that $w = f(z)$ is continuous in $\overline{D}$ and analytic in D everywhere except for an interior point z_0, at which this function has a simple pole.*
2. *It is a one-to-one sense-preserving mapping of C onto C^*.*

Then f(z) is a (schlicht) conformal mapping of D onto D^*.

Example 1. The exponential function $w = e^z$ maps

a) the strip between the straight lines $y = k(x - a_1)$ and $y = k(x - a_2)$ onto the strip lying between the logarithmic spirals (Fig. 10.6). (If $k(a_2 - a_1) = 2\pi$, then the spirals coincide, and we obtain a mapping onto the plane with the spiral cut; for $k(a_2 - a_1) > 2\pi$, the mapping is not schlicht);
b) the strip $0 < \operatorname{Im} z < \pi$ onto the upper half-plane (Fig. 10.7); here the point πi is taken to the point –1, and the point 0 is taken to the point 1;
c) the half-strip $0 < \operatorname{Im} z < \pi$, $\operatorname{Re} z < 0$, onto the half-disk $|w| < 1$, $\operatorname{Im} w > 0$ (Fig. 10.8);
d) a rectangle onto a half-annulus (Fig. 10.9).

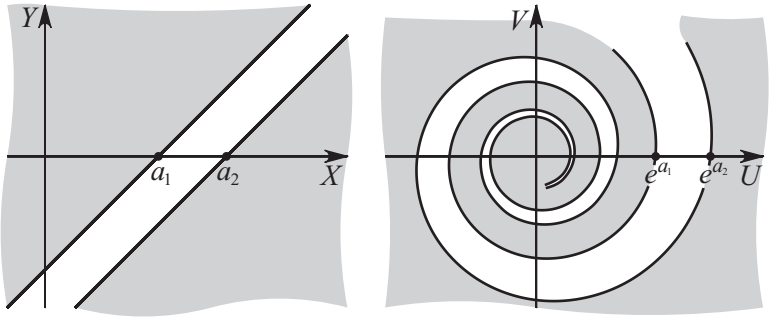


Figure 10.6. The exponential function $w = e^z$ maps the strip between the straight lines onto the strip lying between the logarithmic spirals.

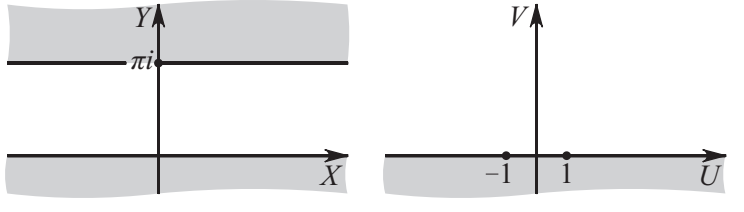


Figure 10.7. The exponential function $w = e^z$ maps the strip $0 < \operatorname{Im} z < \pi$ onto the upper half-plane.

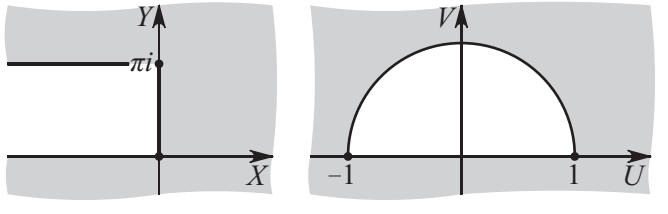


Figure 10.8. The exponential function $w = e^z$ maps the half-strip $0 < \operatorname{Im} z < \pi$, $\operatorname{Re} z < 0$, onto the half-disk.

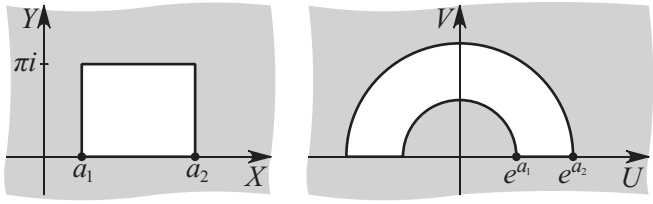


Figure 10.9. The exponential function $w = e^z$ maps a rectangle onto a half-annulus.

Example 2. The function $w = z^2$ maps the interior of a circle onto the interior of a cardioid (Fig. 10.10). The circle given in polar coordinates by the equation $r = \cos\varphi$ is taken to the cardioid $\rho = \frac{1}{2}(1 + \cos\theta)$, where $\theta = 2\varphi$.

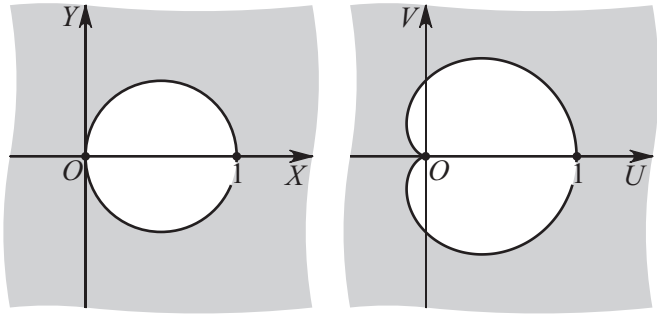


Figure 10.10. The function $w = z^2$ maps the interior of a circle onto the interior of a cardioid.

Example 3. The function $w = \sqrt{z}$ maps the interior of a circle onto the interior of the right branch of a lemniscate (Fig. 10.11). The circle $r = \cos\varphi$ is taken to the right branch of the lemniscate $\rho = \sqrt{\cos 2\theta}$, where $\theta = \frac{1}{2}\varphi$.

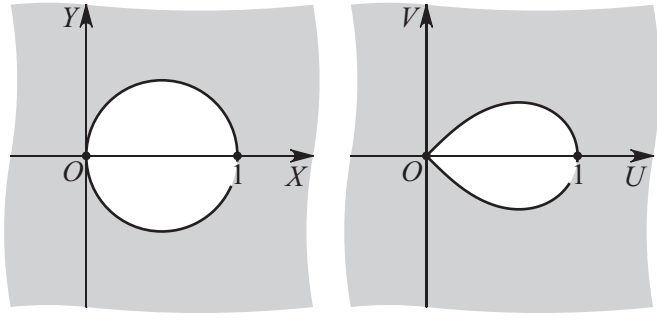


Figure 10.11. The function $w = \sqrt{z}$ maps the interior of a circle onto the interior of the right branch of a lemniscate.

Example 4. The function $w = -\ln(1 - z)$ maps the interior of the unit circle onto the interior of the curve $u = -\ln(2\cos v)$ (Fig. 10.12).

Example 5. The function $w = \ln\dfrac{z-1}{1+z}$ maps the upper half-plane onto the strip $0 < \operatorname{Im} z < \pi$ (Fig. 10.13). The function $z = -\coth\frac{1}{2}w$ specifies the inverse mapping of the strip $0 < \operatorname{Im} z < \pi$ onto the upper half-plane.

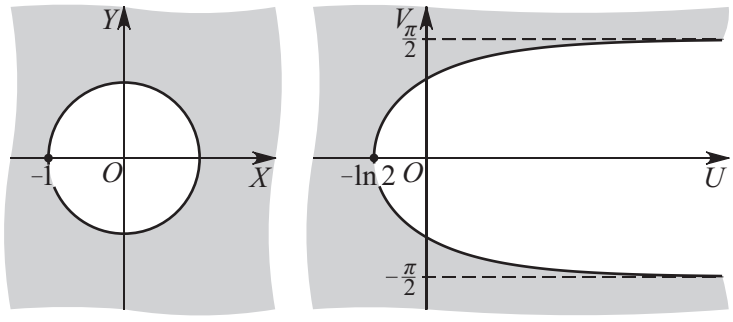


Figure 10.12. The function $w = -\ln(1 - z)$ maps the interior of the unit circle onto the interior of the curve $u = -\ln(2\cos v)$.

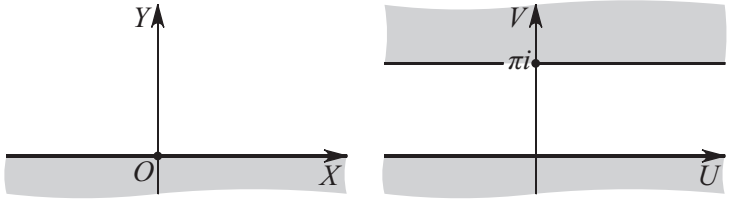


Figure 10.13. The function $w = \ln\dfrac{z-1}{1+z}$ maps the upper half-plane onto the strip $0 < \operatorname{Im} z < \pi$.

10.2.1-3. Linear-fractional mappings.

The mappings given by linear-fractional functions

$$w = \frac{az + b}{cz + d}, \tag{10.2.1.2}$$

where a, b, c, and d are complex constants and $ad - bc \neq 0$, are called *linear-fractional mappings*. The function (10.2.1.2) is defined on the extended complex plane. (Its value at the point $z = -d/c$ is defined to be ∞, and the value at the point $z = \infty$ is defined to be a/c.) A linear-fractional function defines a schlicht mapping of the extended z-plane onto the extended w-plane. Linear-fractional functions are the only functions with this property.

Points z and z^* are said to be symmetric about the circle $C_0 : |z - z_0| = R_0$ if they lie on the same ray passing through z_0 and $|z - z_0||z^* - z_0| = R_0^2$.

The transformation taking each point z to the point z^* symmetric to z about the circle C_0 is called the *symmetry*, or the *inversion*, about the circle.

Points z and z^* are symmetric about a circle C_0 if and only if they are the vertices of a pencil of circles orthogonal to the circle C_0.

THEOREM. *An arbitrary linear-fractional function*

$$w = \frac{az + b}{cz + d}, \quad ad - bc \neq 0,$$

defines a schlicht conformal mapping of the extended z-plane onto the extended w-plane. This mapping transforms any circle on the extended z-plane into a circle on the extended w-plane (the circular property) and transforms any pair of points symmetric about a circle C into a pair of points symmetric about the image of the circle C (preservation of symmetric points).

Let us present some formulas that allow one to find the images of straight lines and circles for an arbitrary linear-fractional mapping (10.2.1.2).

1. The straight lines $\operatorname{Re}(\lambda z) = \alpha$ that do not pass through the point $z = -d/c$ ($\alpha \neq -\operatorname{Re}(\lambda d/c)$) are taken to the circles $|w - w_0| = \rho$, where

$$w_0 = \frac{2a\alpha\bar{c} + a\bar{d}\bar{\lambda} + b\bar{c}\lambda}{2\alpha|c|^2 + 2\operatorname{Re}(c\bar{d}\bar{\lambda})}, \quad \rho = \left|\frac{a}{c} - w_0\right|.$$

2. The straight lines $\operatorname{Re}(\lambda z) = -\operatorname{Re}(\lambda d/c)$ passing through the point $z = -d/c$ are taken to the straight lines

$$\operatorname{Re}\Big(\frac{ad-bc}{c^2}\lambda\bar{w}\Big) = \operatorname{Re}\Big(\frac{ad-bc}{c^2}\frac{\lambda\bar{a}}{\bar{c}}\Big).$$

3. The circles $|z - z_0| = r$ that do not pass through the point $z = -d/c$ ($r \neq |z_0 + d/c|$) are taken to the circles $|w - w_0| = \rho$, where

$$w_0 = \frac{(az_0+b)(\bar{c}\bar{z}_0+\bar{d}) - a\bar{c}r^2}{|cz_0+d|^2 - |c|^2r^2}, \quad \rho = \frac{r|ad-bc|}{\big||cz_0+d|^2 - |c|^2r^2\big|}.$$

4. The circles $|z - z_0| = |z_0 + d/c|$ are taken to the straight lines

$$\operatorname{Re}\Big[\frac{ad-bc}{c(cz_0+d)}\bar{w}\Big] = \frac{|ad-bc|^2 + 2\operatorname{Re}[c(az_0+b)(\overline{ad}-\overline{bc})]}{2|c(cz_0+d)|^2}.$$

If a linear-fractional mapping takes four points z_1, z_2, z_3, and z to points w_1, w_2, w_3, and w, respectively, then the following relation holds:

$$\frac{w-w_1}{w-w_3}\frac{w_2-w_3}{w_2-w_1} = \frac{z-z_1}{z-z_3}\frac{z_2-z_3}{z_2-z_1}. \tag{10.2.1.3}$$

THEOREM. *There exists a unique linear-fractional mapping of the extended z-plane onto the extended w-plane taking three arbitrary distinct points z_1, z_2, and z_3 to three arbitrary distinct points w_1, w_2, and w_3, respectively.*

THEOREM. *Any disk of the extended z-plane can be transformed into any disk of the extended w-plane by a linear-fractional function.*

Example 6. A mapping of the upper half-plane onto the unit disk.

Let a be the point of the upper half-plane which should be taken to the center $w = 0$ of the disk (Fig. 10.14). Then the problem is solved by the linear-fractional function

$$w = e^{i\beta}\frac{z-a}{z-\bar{a}},$$

where β is an arbitrary real number. (Changing α means rotating the disk around the center $w = 0$.)

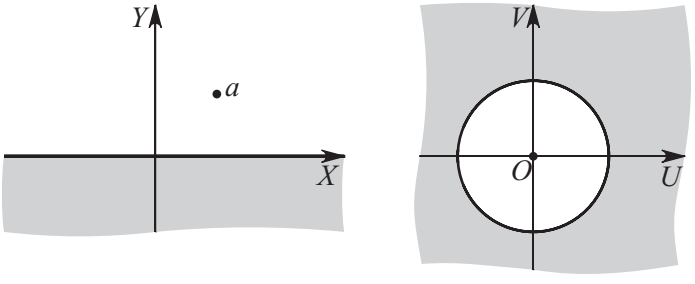


Figure 10.14. A mapping of the upper half-plane onto the unit disk.

Example 7. A mapping of the unit disk onto the upper half-plane.

Let $a = ih$ be the point of the upper half-plane to which the center $z = 0$ of the disk should be taken (Fig. 10.15). Then the problem is solved by the linear-fractional function

$$w = ih\frac{e^{i\beta} + z}{e^{i\beta} - z},$$

where β is an arbitrary real number. (Changing β means rotating the disk around the center $w = 0$.)

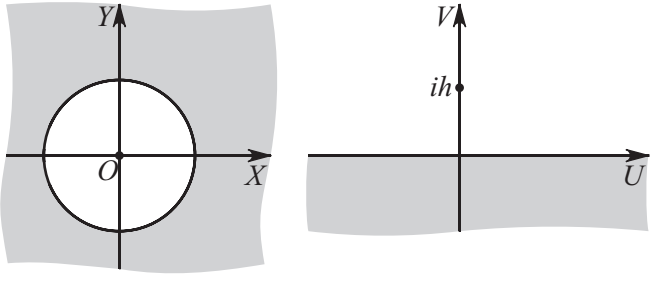


Figure 10.15. A mapping of the unit disk onto the upper half-plane.

Example 8. A mapping of the unit disk onto the unit disk.

Let a be the point of the disk $|z| < 1$ that should be taken to the center of the disk $|w| < 1$ (Fig. 10.16). Then the problem is solved by the linear-fractional function

$$w = e^{i\beta}\frac{z - a}{1 - z\bar{a}}, \tag{10.2.1.4}$$

where β is an arbitrary real number.

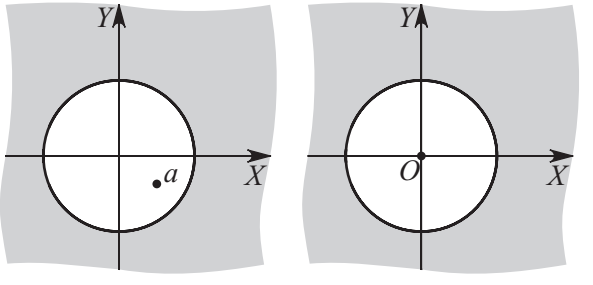


Figure 10.16. A mapping of the unit disk onto the unit disk.

Geometrically, β is the angle of rotation of the mapping (10.2.1.4) at the point a:

$$\beta = \arg\frac{dw}{dz}.$$

If the radius of the disk in the z-plane is equal to R, then the function $w = f(z)$ mapping this disk onto the disk $|w| < 1$ and satisfying $f(a) = 0$ and $\arg f'_z(a) = \beta$ has the form

$$w = e^{i\beta}\frac{R(z - a)}{R^2 - z\bar{a}}.$$

10.2.1-4. Mappings determined by the Zhukovskii function.

The mapping (see also Paragraph 10.1.2-3)

$$w = \frac{1}{2}\left(z + \frac{1}{z}\right) \tag{10.2.1.5}$$

is conformal except at the points $z = \pm 1$.

The mapping given by the Zhukovskii function is equivalent to the following mappings:

$$z = w + \sqrt{w^2 - 1} \quad \text{or} \quad \frac{w-1}{w+1} = \left(\frac{z-1}{z+1}\right)^2. \tag{10.2.1.6}$$

If we denote $z = re^{i\varphi}$ and $w = u + iv$, then the mapping (10.2.1.5) can be written as

$$u = \frac{1}{2}\left(r + \frac{1}{r}\right)\cos\varphi, \quad v = \frac{1}{2}\left(r - \frac{1}{r}\right)\sin\varphi. \tag{10.2.1.7}$$

The main properties of the mapping (10.2.1.5) are given in Table 10.2.

TABLE 10.2
Properties of mapping $w = \frac{1}{2}\left(z + \frac{1}{z}\right)$, where $z = x + iy = re^{i\varphi}$, $w = u + iv = \rho e^{i\theta}$

No.	z-plane	w-plane	Remarks
1	Circle $\lvert z\rvert = r_0 < 1$	Ellipse $\frac{u^2}{a^2} + \frac{v^2}{b^2} = 1$, where $a = \frac{1}{2}(1/r_0 + r_0)$, $b = \frac{1}{2}(1/r_0 - r_0)$	The ellipse has the negative sense. The foci are at the points ± 1.
2	Circle $\lvert z\rvert = r_0 > 1$	Ellipse $\frac{u^2}{a^2} + \frac{v^2}{b^2} = 1$, where $a = \frac{1}{2}(1/r_0 + r_0)$, $b = \frac{1}{2}(1/r_0 - r_0)$	The ellipse has the positive sense. The foci are at the points ± 1.
3	Radii $\arg z = \varphi_0 \quad (0 < r < 1)$	Hyperbolas $\frac{u^2}{\cos^2\varphi_0} - \frac{v^2}{\sin^2\varphi_0} = 1$	The foci are at the points ± 1.
4	Semicircle $\lvert z\rvert = 1, \quad \operatorname{Im} z \geq 0$	Segment $\lvert\operatorname{Re} w\rvert \leq 1, \quad \operatorname{Im} w = 0$	u decreases with increasing φ.
5	Semicircle $\lvert z\rvert = 1, \quad \operatorname{Im} z \leq 0$	Segment $\lvert\operatorname{Re} w\rvert \leq 1, \quad \operatorname{Im} w = 0$	u increases with increasing φ.
6	Segment $0 < \operatorname{Re} z < 1, \quad \operatorname{Im} z = 0$	Half-line $\operatorname{Re} w > 1, \quad \operatorname{Im} w = 0$	—
7	Segment $-1 < \operatorname{Re} z < 0, \quad \operatorname{Im} z = 0$	Half-line $\operatorname{Re} w < -1, \quad \operatorname{Im} w = 0$	—
8	Half-line $\operatorname{Re} z > 1, \quad \operatorname{Im} z = 0$	Half-line $\operatorname{Re} w > 1, \quad \operatorname{Im} w = 0$	—
9	Half-line $\operatorname{Re} z < -1, \quad \operatorname{Im} z = 0$	Half-line $\operatorname{Re} w < -1, \quad \operatorname{Im} w = 0$	—

Example 9. The Zhukovskii function defines the following conformal maps:

1. It maps the interior of the semicircle $|z| < 1$, $\operatorname{Im} z > 0$, onto the lower half-plane (Fig. 10.17). The point $z = 1$ is taken to the point $w = 1$, and the point $z = i$ is taken to the point $w = 0$.
2. It maps the upper half-plane with the disk $|z| < 1$ deleted onto the upper half-plane (Fig. 10.18).
3. It maps the half-annulus $1 < |z| < k$, $\operatorname{Im} z > 0$, onto the half-ellipse given by the equation

$$\left(\frac{2ku}{k^2+1}\right)^2 + \left(\frac{2kv}{k^2-1}\right)^2 = 1$$

on the w-plane (Fig. 10.19).

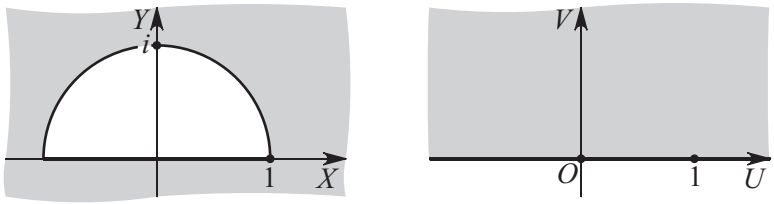


Figure 10.17. Zhukovskii function maps the interior of the semicircle $|z| < 1$, $\operatorname{Im} z > 0$, onto the lower half-plane.

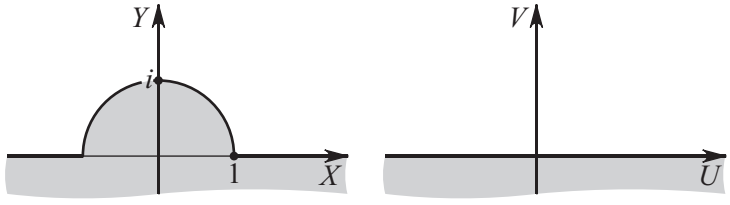


Figure 10.18. Zhukovskii function maps the upper half-plane with the disk $|z| < 1$ deleted onto the upper half-plane.

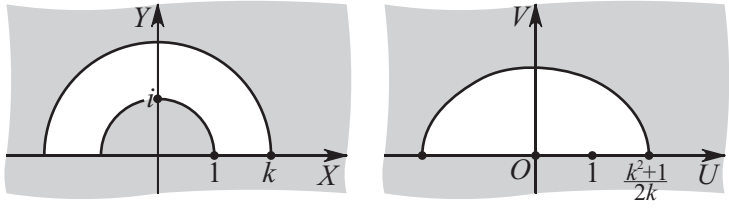


Figure 10.19. Zhukovskii function maps the half-annulus $1 < |z| < k$, $\operatorname{Im} z > 0$, onto the half-ellipse.

10.2.1-5. Symmetry principle and mapping of polygons.

In a special case, the symmetry principle gives a simple sufficient condition for the existence of an analytic continuation of a function realizing a conformal mapping.

THE RIEMANN–SCHWARZ THEOREM. *Suppose that a function $w = f_1(z)$ realizes a conformal mapping of a domain D_1 onto a domain D_1^* and takes a circular arc C of the boundary of D_1 to a circular arc C^* of the boundary of D_1^*. Then the function $f_1(z)$ admits an analytic continuation $f_2(z)$ through the arc C into a domain D_2 symmetric to D_1 about C, the function $w = f_2(z)$ realizes a conformal mapping of the domain D_2 onto the domain D_2^* symmetric to D_1^* about C^*, and the function*

$$w = f(z) = \begin{cases} f_1(z) & \text{in } D_1, \\ f_1(z) = f_2(z) & \text{on } C, \\ f_2(z) & \text{in } D_2 \end{cases}$$

realizes a conformal mapping of the domain $D_1 + C + D_2$ onto the domain $D_1^* + C^* + D_2^*$.

An arc C is said to be *analytic* if it can be described by parametric equations

$$x = x(t), \quad y = y(t) \quad (\alpha \le t \le \beta)$$

such that $x(t)$ and $y(t)$ are analytic functions of the real variable t on the interval (α, β).

SCHWARZ'S ANALYTIC CONTINUATION PRINCIPLE. *Suppose that a function $w = f(z)$ realizes a conformal mapping of a domain D onto a domain D^* and takes an analytic arc C of the boundary of D to an arc C^* of the boundary of D^*. Then the function $w = f(z)$ can be analytically continued through the arc C.*

THE SCHWARZ–CHRISTOFFEL THEOREM. *If a function $w = f(z)$ realizes a conformal mapping of the upper half-plane $\operatorname{Im} z > 0$ onto the interior of a bounded polygon Δ with angles $\pi\alpha_k$ ($0 < \alpha_k \le 2$, $k = 1, 2, \ldots, n$) at the vertices and if the points a_k of the real axis ($-\infty < a_1 < \ldots < a_n < \infty$) corresponding to the vertices of this polygon are known, then the function $f(z)$ can be represented by the Schwarz–Christoffel integral*

$$f(z) = C \int_{z_0}^{z} (z-a_1)^{\alpha_1-1}(z-a_2)^{\alpha_2-1} \cdots (z-a_n)^{\alpha_n-1}\, dz + C_1,$$

where z_0, C, and C_1 are some constants.

The Schwarz–Christoffel integral is obtained under the assumption that the points a_k corresponding to the vertices A_k of the polygon are known. In practice, only the vertices of the polygon are given, and the points a_k are unknown. Determining the points a_k is a very difficult task.

Table 10.3 presents some conformal mappings of given domains D onto the unit disk.

TABLE 10.3
Conformal mappings onto the unit disk $|w| \le 1$, where $z = x + iy = r(\cos\varphi + i\sin\varphi)$

No.	Domain	Mapping
1	Upper half-plane, $\operatorname{Im} z > 0$	$w = e^{i\beta}\dfrac{z-a}{z-\bar{a}}$ (β is a real number)
2	Right half-plane, $\operatorname{Re} z > 0$	$w = e^{i\beta}\dfrac{z-a}{z-\bar{a}}$ (β is a real number)
3	Disk of radius R, $\lvert z\rvert < R$	$w = e^{i\beta}\dfrac{R(z-a)}{R^2-\bar{a}z}$ (β is a real number)
4	Strip of width $\frac{1}{2}\pi$, $-\frac{1}{4}\pi < \operatorname{Re} z < \frac{1}{4}\pi$	$w = \tan z$
5	Sector of unit disk, $\lvert z\rvert < 1$, $0 < \arg z < \pi\alpha$	$w = \dfrac{(1+z^{1/\alpha})^2 - i(1-z^{1/\alpha})^2}{(1+z^{1/\alpha})^2 + i(1-z^{1/\alpha})^2}$
6	Plane with cut from $z = 0$ to $z = \infty$ along the positive real axis	$w = \dfrac{\sqrt{z}-i}{\sqrt{z}+i}$
7	Exterior of the ellipse, $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$	$z = R\Big(mw + \dfrac{1}{w}\Big)$, $R = \dfrac{a+b}{2}$, $m = \dfrac{a-b}{a+b}$
8	Exterior of the parabola, $r\cos^2\dfrac{\varphi}{2} = 1$	$z = \Big(\dfrac{2}{w+1}\Big)^2$
9	Interior of the parabola, $r\cos^2\dfrac{\varphi}{2} = 1$	$w = \tan^2\Big(\dfrac{\pi}{4}\sqrt{z}\Big)$
10	Half-disk, $\lvert z\rvert < R$, $\operatorname{Re} z > 0$	$w = i\dfrac{z^2 + 2Rz - R^2}{z^2 - 2Rz - R^2}$
Remark. In the items 1, 2, and 3 the point $z = a$ of the domain is taken to the center $w = 0$ of the disk; β determines the rotation of the disk about the center $w = 0$.		

10.2.2. Boundary Value Problems

10.2.2-1. Dirichlet problem.

Find a function $u(z)$ harmonic in the domain D, continuous in $\overline{D}$, and taking given continuous values $u(\xi)$ on the boundary of D.

Generalized Dirichlet problem. Given a function $u(\xi)$ defined on the boundary C of a domain D and continuous everywhere except for finitely many points $\xi_1, \ldots, \xi_n$, where it has jump discontinuities, find a function $u(z)$ harmonic and bounded in D and equal to $u(\xi)$ at all points of continuity of this function on C.

THEOREM ON THE UNIQUENESS OF A SOLUTION OF THE GENERALIZED DIRICHLET PROBLEM. *In a given domain for a given boundary function $u(\xi)$, there exists at most one solution of the generalized Dirichlet problem.*

THEOREM ON THE EXISTENCE OF A SOLUTION OF THE GENERALIZED DIRICHLET PROBLEM. *For any simply connected domain D and any piecewise continuous boundary function $u(\xi)$ with jump discontinuities, the generalized Dirichlet problem has a solution.*

POISSON'S THEOREM. *The solution of the generalized Dirichlet problem for the unit disk is given by the Poisson integral*

$$u(z) = \frac{1}{2\pi}\int_0^{2\pi} u(e^{it})\frac{1-r^2}{1-2r\cos(t-\varphi)+r^2}\,dt \qquad (z = re^{it}). \tag{10.2.2.1}$$

1°. Let z_0 be an arbitrary point of a domain D, and let

$$w = f(z; z_0), \quad f(z_0; z_0) = 0 \tag{10.2.2.2}$$

be a function mapping the domain D onto the unit disk $|w| < 1$. The function

$$g(z; z_0) = \ln\frac{1}{|f(z; z_0)|} \tag{10.2.2.3}$$

is called a *Green's function* of the domain D. A Green's function is harmonic everywhere in D except for the point z_0 at which it has a pole.

The solution of the generalized Dirichlet problem is given by *Green's formula*

$$u(z) = \frac{1}{2\pi}\int_0^{2\pi} u(\xi)g_{\mathbf{n}}(\xi, z)ds, \tag{10.2.2.4}$$

where $g_{\mathbf{n}}$ is the inward normal derivative.

Green's formula expresses the solution of the Dirichlet problem for some domain D in terms of the logarithm of the conformal mapping of the domain D onto the unit disk, i.e., reduces solving the Dirichlet problem to solving the conformal mapping problem. The converse statement is also true. If the solution of the Dirichlet problem is known for some domain D, then a conformal mapping of this domain onto the unit disk can be constructed.

2°. Suppose that we need to find a function $f(z)$ that is analytic in the disk $|z| < 1$ and whose real part on the circle takes given values $u(\xi)$ at each point of continuity of the function $u(\xi)$. The solution of this problem is given by the *Schwarz formula*

$$f(z) = \frac{1}{2\pi}\int_0^{2\pi} u(\xi)\frac{\xi+z}{\xi-z}\,dt + iC \quad (\xi = e^{it}), \tag{10.2.2.5}$$

where C is a real constant. The integral on the right-hand side in (10.2.2.5) is called the *Schwarz integral*.

3°. Suppose that a bounded function $u(t)$ with finitely many points of discontinuity is given on the real axis and there exist finite limits of $u(t)$ as $t \to \pm\infty$. The solution of the Dirichlet problem for the upper half-plane is given by the following *Poisson integral for the half-plane*:

$$u(z) = \frac{1}{\pi}\int_{-\infty}^{+\infty} u(t)\frac{y}{(t-x)^2+y^2}\,dt. \tag{10.2.2.6}$$

Since

$$\frac{y}{(t-x)^2+y^2} = \operatorname{Re}\frac{1}{i(t-z)},$$

we can also write out the *Schwarz integral for the half-plane* in the form

$$f(z) = \frac{1}{2\pi}\int_0^{2\pi} u(t)\frac{dt}{t-z} + iC, \tag{10.2.2.7}$$

where C is a real constant.

10.2.2-2. Neumann problem.

Find a function $u(z)$ harmonic in the domain D with given normal derivative

$$u_{\mathbf{n}} = u_x\cos\alpha + u_y\sin\alpha = g(\xi) \tag{10.2.2.8}$$

on the boundary C and given value $u(z_0)$ at a point z_0 of the domain $\overline{D}$.

It is assumed in (10.2.2.8) that the outward normal is considered and α is the angle between this normal and the axis OX. The function $g(\xi)$ is allowed to have only finitely many points of jump discontinuity on C; the function $u(z)$ and its first partial derivatives are assumed to be bounded.

Necessary condition for the solvability of Neumann problem:

$$\int_C g(\xi)\,ds = 0. \tag{10.2.2.9}$$

If, in addition, we assume that the partial derivatives are continuous in $\overline{D}$, then solving the Neumann problem can be reduced to solving the Dirichlet problem for the conjugate harmonic function. Suppose that $v(z)$ is a harmonic function conjugate to $u(z)$. By the Cauchy–Riemann conditions written for the curve C in the directions of $\mathbf{s}$ and $\mathbf{n}$, we have

$$v_{\mathbf{s}} = u_{\mathbf{n}} = g(\xi).$$

If $v_{\mathbf{s}}$ along the curve C is known, then straightforward integration gives

$$v(\xi) = \int_{\xi_0}^{\xi} v_{\mathbf{s}}\,ds = \int_{\xi_0}^{\xi} g(\xi)\,ds. \tag{10.2.2.10}$$

Now the problem of determining $v(z)$ in the domain D is reduced to the Dirichlet problem. If $v(z)$ is known, then the desired function $u(z)$ can also be obtained by integration.

Now suppose that the domain D is the unit disk. If we set $f(z) = u + iv$, then the function $f(z)$ satisfies the formula

$$f(z) = -\frac{1}{\pi}\int_0^{2\pi} g(e^{it})\ln(e^{it}-z)\,dt + \text{const}. \tag{10.2.2.11}$$

Separating the real part, we obtain the formula for the desired function:

$$u(z) = -\frac{1}{\pi}\int_0^{2\pi} g(e^{it})\ln|e^{it}-z|\,dt + \text{const}, \tag{10.2.2.12}$$

which is called the *Dini formula*.

10.2.2-3. Cauchy-type integral.

Suppose that C is an arbitrary curve without cusps, not necessarily closed. Let an arbitrary function $f(\xi)$, which is assumed to be finite and integrable, be given on this curve.

The integral

$$F(z) = \frac{1}{2\pi i}\int_C \frac{f(\xi)\,d\xi}{\xi - z} \tag{10.2.2.13}$$

is called a *Cauchy-type integral*.

The Cauchy-type integral is a function analytic at any point z that does not lie on C. If the curve C divides the plane into several domains, then, in general, the Cauchy-type integral defines different analytic functions in these domains.

One says that the function $f(\xi)$ satisfies the *Hölder condition* with exponent $\mu \le 1$ at a point $\xi = \xi_0$ of the contour C if there exists a constant M such that the inequality

$$|f(\xi) - f(\xi_0)| \le M|\xi - \xi_0|^\mu \quad (0 < \mu \le 1) \tag{10.2.2.14}$$

holds for all points $\xi \in C$ sufficiently close to ξ_0. The Hölder condition means that the increment of the function is an infinitesimal of order at least μ with respect to the increment of the argument.

The *principal value of the integral* is defined as the limit

$$\lim_{r\to 0}\int_{C-c} \frac{f(\xi)\,d\xi}{\xi - \xi_0} = \int_C \frac{f(\xi)\,d\xi}{\xi - \xi_0}, \tag{10.2.2.15}$$

where c is the segment of the curve C between the points of intersection of C with the circle $|z - \xi_0| = r$.

The *singular integral in the sense of the Cauchy principal value* is defined as the integral given by the formula

$$\int_{C-c} \frac{f(\xi)\,d\xi}{\xi - \xi_0} = \int_C \frac{f(\xi) - f(\xi_0)}{\xi - \xi_0}\,d\xi + f(\xi_0)\ln\frac{b - \xi_0}{a - \xi_0} + i\pi f(\xi_0) + O(r), \tag{10.2.2.16}$$

where a and b are the endpoints of C and $O(r) \to 0$ as $r \to 0$.

THEOREM. *If the function $f(\xi)$ satisfies the Hölder condition with exponent $\mu \le 1$ at a point ξ_0 which is a regular (nonsingular) point of the contour C and does not coincide with its endpoints, then the Cauchy-type integral exists at this point as a singular integral and its principal value can be expressed in terms of the usual integral by the formula*

$$F(\xi_0) = \frac{1}{2\pi i}\int_C \frac{f(\xi)\,d\xi}{\xi - \xi_0} = \frac{1}{2\pi i}\int_C \frac{f(\xi) - f(\xi_0)}{\xi - \xi_0}\,d\xi + \frac{f(\xi_0)}{2} + \frac{f(\xi_0)}{2\pi i}\ln\frac{b - \xi_0}{a - \xi_0}. \tag{10.2.2.17}$$

If the curve C is closed, then $a = b$ and formula (10.2.2.17) becomes

$$F(\xi_0) = \frac{1}{2\pi i}\int_C \frac{f(\xi)\,d\xi}{\xi - \xi_0} = \frac{1}{2\pi i}\int_C \frac{f(\xi) - f(\xi_0)}{\xi - \xi_0}\,d\xi + \frac{f(\xi_0)}{2}. \tag{10.2.2.18}$$

Suppose that the function $f(\xi)$ satisfies the Hölder condition with exponent $\mu \le 1$ at the point $\xi = \xi_0$ and the point z tends to ξ_0 so that the ratio of $h = |z - \xi_0|$ to d (dh is the shortest distance from z to the points of C) remains bounded. Then

$$\lim_{z\to\xi_0}\int_C \frac{f(\xi) - f(\xi_0)}{\xi - z}\,d\xi = \int_C \frac{f(\xi) - f(\xi_0)}{\xi - \xi_0}\,d\xi. \tag{10.2.2.19}$$

SOKHOTSKII'S THEOREM. *Suppose that ξ_0 is a regular (nonsingular) point of the contour C and does not coincide with its endpoints, the function $f(\xi)$ satisfies the Hölder condition with exponent $\mu \le 1$ at this point, and $z \to \xi_0$ so that the ratio h/d remains bounded. Then the Cauchy-type integral has limit values $F^+(\xi_0)$ and $F^-(\xi_0)$ to which this integral tends as $z \to \xi_0$ from the left and, respectively, from the right of C, and*

$$F^+(\xi_0) = F(\xi_0) + \frac{1}{2} f(\xi_0), \quad F^-(\xi_0) = F(\xi_0) - \frac{1}{2} f(\xi_0), \tag{10.2.2.20}$$

where $F(\xi_0)$ is the singular integral (10.2.2.18).

The Cauchy-type integral experiences a jump when passing through the integration contour C at the point ξ_0:

$$F^+(\xi_0) - F^-(\xi_0) = f(\xi_0). \tag{10.2.2.21}$$

The condition

$$F^-(\xi) = 0 \tag{10.2.2.22}$$

at each point of C is necessary and sufficient for a Cauchy-type integral to be the Cauchy integral.

THEOREM. *If a function $f(\xi)$ satisfies the Hölder condition with exponent $\mu \le 1$ at each point of a closed contour C, then, for its values to be the boundary values of a function analytic in the interior of C, it is necessary and sufficient that*

$$\int_C \xi^n f(\xi)\, d\xi = 0 \qquad (n = 0, 1, 2, \ldots). \tag{10.2.2.23}$$

THEOREM. *If a function $f(\xi)$ satisfies the Hölder condition with exponent $\mu \le 1$ at each point of a closed contour C, then, for the values of $f(\xi)$ to be the boundary values of a function analytic in the interior of C, it is necessary and sufficient that*

$$\frac{1}{2\pi i} \int_C \frac{f(\xi)\, d\xi}{\xi - z} = 0 \tag{10.2.2.24}$$

for all points z lying in the exterior of C.

THEOREM. *If a function $f(\xi)$ satisfies the Hölder condition with exponent $\mu \le 1$ at each point of a closed contour C, then, for the values of $f(\xi)$ to be the boundary values of a function analytic in the exterior of C, it is necessary and sufficient that*

$$\frac{1}{2\pi i} \int_C \frac{f(\xi)\, d\xi}{\xi - z} = f(\infty) \tag{10.2.2.25}$$

for all points z lying in the interior of C.

THEOREM. *For the values of a function $f(\xi)$ satisfying the Hölder condition with exponent $\mu \le 1$ to be the boundary values of a function analytic (a) in the interior of the disk $|z| < 1$ or (b) in the exterior of this disk, it is necessary and sufficient that the following respective conditions hold:*

for all z in the interior of C,

$$\frac{1}{2\pi i} \int_C \frac{\overline{f(\xi)}\, d\xi}{\xi - z} = \overline{f(0)}, \tag{10.2.2.26}$$

for all z in the exterior of C

$$\frac{1}{2\pi i} \int_C \frac{\overline{f(\xi)}\, d\xi}{\xi - z} = 0. \tag{10.2.2.27}$$

Example (the first main problem of elasticity).

Let D be the unit disk. Find the elastic equilibrium for given external stresses $F_n = X_n + iY_n$ on the unit circle C, where X_n and Y_n are the components of a surface force vector.

The problem is to find functions φ and ψ satisfying the boundary condition

$$\varphi(\xi) + \xi\overline{\varphi'_\xi(\xi)} + \overline{\psi(\xi)} = f(\xi),$$

where $f(\xi) = i\int_{\xi_0}^{\xi} F_n\,ds$ is a function given on C.

To be definite, we set

$$\psi(0) = \operatorname{Im}\varphi'_\xi(0) = 0.$$

By formula (10.2.2.26), the relation

$$\frac{1}{2\pi i}\int_C \frac{\overline{\psi(\xi)}\,d\xi}{\xi - z} = \frac{1}{2\pi i}\int_C \frac{f(\xi)\,d\xi}{\xi - z} - \frac{1}{2\pi i}\int_C \frac{\varphi(\xi)\,d\xi}{\xi - z} - \frac{1}{2\pi i}\int_C \frac{\xi\overline{\varphi'_\xi(\xi)}}{\xi - z}\,d\xi = 0$$

holds for all $|z| < 1$. Since the function $\varphi(z)$ is analytic in the disk $|z| < 1$, we can use the Cauchy formula and rewrite this relation as

$$\varphi(z) + \frac{1}{2\pi i}\int_C \frac{\xi\overline{\varphi'_\xi(\xi)}}{\xi - z}\,d\xi = \frac{1}{2\pi i}\int_C \frac{f(\xi)\,d\xi}{\xi - z}.$$

Thus we obtain an equation for the function $\varphi(z)$. Omitting the details, we write out the definitive result:

$$\varphi(z) = \frac{1}{2\pi i}\int_C \frac{f(\xi)\,d\xi}{\xi - z} - \frac{z}{4\pi i}\int_C \frac{f(\xi)\,d\xi}{\xi^2}.$$

To find the function $\psi(z)$, we pass from the boundary condition $\varphi(\xi) + \xi\overline{\varphi'_\xi(\xi)} + \overline{\psi(\xi)} = f(\xi)$ to the complex conjugate condition and solve it for $\psi(z)$. Thus we obtain

$$\psi(\xi) = \overline{f(\xi)} - \overline{\varphi(\xi)} - \overline{\xi}\varphi'_\xi(\xi).$$

We calculate the Cauchy-type integral of the expressions in both sides, which is reduced to the Cauchy integral in either case, and obtain

$$\psi(z) = \frac{1}{2\pi i}\int_C \frac{\overline{f(\xi)}\,d\xi}{\xi - z} + \frac{1}{4\pi i z}\int_C \frac{f(\xi)\,d\xi}{\xi^2} - \frac{\varphi'_\xi(\xi)}{z}.$$

10.2.2-4. Hilbert–Privalov boundary value problem.

Privalov boundary value problem. Given two complex functions $a(\xi) \neq 0$ and $b(\xi)$ satisfying the Hölder condition with exponent $\mu \leq 1$ on a closed curve C, find a function $f^-(z)$ analytic in the exterior of C including the point at infinity $z = \infty$ and a function $f^+(z)$ analytic in the interior of C such that the boundary values $f^-(\xi)$ and $f^+(\xi)$ of these functions on C exist and satisfy the relation

$$f^-(\xi) = a(\xi)f^+(\xi) + b(\xi). \tag{10.2.2.28}$$

If $b(\xi) = 0$, i.e., if the boundary relation has the form

$$f^-(\xi) = a(\xi)f^+(\xi), \tag{10.2.2.29}$$

then the Privalov boundary value problem is called the *Hilbert boundary value problem.*

The *index (winding number) of a function* $a(\xi)$ is defined to be the integer equal to the net increment of its argument along the closed curve C, divided by 2π:

$$\frac{1}{2\pi}\Delta_C \arg a(\xi) = \frac{1}{2\pi i}\int_C d\ln a(\xi). \tag{10.2.2.30}$$

GAKHOV'S FIRST THEOREM. *The Hilbert problem*

$$f^-(\xi) = a(\xi)f^+(\xi)$$

has a family of solutions depending on $n + 1$ arbitrary constants if the index n of the boundary function $a(\xi)$ is not positive. If the index n is positive, then the problem does not have solutions analytic in the corresponding domains.

The solutions of the Hilbert problem can be written as

$$\begin{aligned} f^-(\xi) &= \left(a_0 + \frac{a_1}{z} + \cdots + \frac{a_n}{z^n}\right) \exp[-F_1^-(z)], \\ f^+(\xi) &= (a_0 z^n + a_1 z^{n-1} + \cdots + a_n) \exp[-F_1^+(z)], \end{aligned} \tag{10.2.2.31}$$

where

$$F_1(z) = \frac{1}{2\pi i} \int_C \frac{\ln[\xi^n a(\xi)]}{\xi - z}\, d\xi.$$

The constants $a_0, \ldots, a_n$ in formula (10.2.2.31) are arbitrary, and a_0 is determined by the choice of the value $f^-(\infty)$.

GAKHOV'S SECOND THEOREM. *The Privalov problem*

$$f^-(\xi) = a(\xi) f^+(\xi) + b(\xi)$$

has a family of solutions depending on $n + 1$ arbitrary constants if the index n of the boundary function $a(\xi)$ is not positive. If the index n of the function $a(\xi)$ is positive, then the problem is solvable only if the function $b(\xi)$ satisfies the condition

$$\int_C \frac{b(\xi) \exp[-F_1^-(\xi)]}{\xi^{k+1}}\, d\xi = 0 \qquad (k = 1, 2, \ldots, n). \tag{10.2.2.32}$$

The solutions of the Privalov problem can be written as

$$\begin{aligned} f^-(\xi) &= \left[a_0 + \frac{a_1}{z} + \ldots + \frac{a_n}{z^n} + F_2^-(z)\right] \exp[-F_1^-(z)], \\ f^+(\xi) &= \left[a_0 z^n + a_1 z^{n-1} + \ldots + a_n + z^n F_2^+(z)\right] \exp[-F_1^+(z)], \end{aligned} \tag{10.2.2.33}$$

where $a_0, \ldots, a_n$ are arbitrary constants and $F_2(z)$ is determined by the formula

$$F_2(z) = -\frac{1}{2\pi i} \int_C \frac{b(\xi) \exp[-F_1(\xi)]}{\xi - z}\, d\xi.$$

References for Chapter 10

Ablowitz, M. J. and Fokas, A. S., *Complex Variables: Introduction and Applications* (Cambridge Texts in Applied Mathematics), *2nd Edition*, Cambridge University Press, Cambridge, 2003.

Berenstein, C. A. and Roger Gay, R., *Complex Variables: An Introduction* (Graduate Texts in Mathematics), Springer, New York, 1997.

Bieberbach, L., *Conformal Mapping*, American Mathematical Society, Providence, Rhode Island, 2000.

Bronshtein, I. N., Semendyayev, K. A., Musiol, G., and Mühlig, H., *Handbook of Mathematics, 4th Edition*, Springer, New York, 2004.

Brown, J. W. and Churchill, R. V., *Complex Variables and Applications, 7th Edition*, McGraw-Hill, New York, 2003.

Caratheodory, C., *Conformal Representation*, Dover Publications, New York, 1998.

Carrier, G. F., Krock, M., and Pearson, C. E., *Functions of a Complex Variable: Theory and Technique* (Classics in Applied Mathematics), Society for Industrial & Applied Mathematics, University City Science Center, Philadelphia, 2005.

Cartan, H., *Elementary Theory of Analytic Functions of One or Several Complex Variables*, Dover Publications, New York, 1995.

Conway, J. B., *Functions of One Complex Variable I* (Graduate Texts in Mathematics), *2nd Edition*, Springer, New York, 1995.

Conway, J. B., *Functions of One Complex Variable II* (Graduate Texts in Mathematics)*, 2nd Edition*, Springer, New York, 1996.
Dettman, J. W., Applied Complex Variables (Mathematics Series), Dover Publications, New York, 1984.
England, A. H., *Complex Variable Methods in Elasticity, Dover Edition*, Dover Publications, New York, 2003.
Fisher, S. D., *Complex Variables* (Dover Books on Mathematics), *2nd Edition*, Dover Publications, New York, 1999.
Flanigan, F. J., *Complex Variables, Dover Ed. Edition*, Dover Publications, New York, 1983.
Greene, R. E. and Krantz, S. G., *Function Theory of One Complex Variable* (Graduate Studies in Mathematics), *Vol. 40, 2nd Edition*, American Mathematical Society, Providence, Rhode Island, 2002.
Ivanov, V. I. and Trubetskov, M. K., *Handbook of Conformal Mapping with Computer-Aided Visualization*, CRC Press, Boca Raton, 1995.
Korn, G. A and Korn, T. M., *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review*, Dover Edition, Dover Publications, New York, 2000.
Krantz, S. G., *Handbook of Complex Variables*, Birkhäuser, Boston, 1999.
Lang, S., *Complex Analysis* (Graduate Texts in Mathematics), *4th Edition*, Springer, New York, 2003.
Lavrentiev, M. A. and Shabat, V. B., *Methods of the Theory of Functions of a Complex Variable, 5th Edition [in Russian]*, Nauka Publishers, Moscow, 1987.
LePage, W. R., *Complex Variables and the Laplace Transform for Engineers*, Dover Publications, New York, 1980.
Markushevich, A. I. and Silverman, R. A. (Editor), *Theory of Functions of a Complex Variable, 2nd Rev. Edition*, American Mathematical Society, Providence, Rhode Island, 2005.
Narasimhan, R. and Nievergelt, Y., *Complex Analysis in One Variable, 2nd Edition*, Birkhäuser, Boston, Basel, Stuttgard, 2000.
Needham, T., *Visual Complex Analysis, Rep. Edition*, Oxford University Press, Oxford, 1999.
Nehari, Z., *Conformal Mapping*, Dover Publications, New York, 1982.
Paliouras, J. D. and Meadows, D. S., *Complex Variables for Scientists and Engineers, Facsimile Edition*, Macmillan Coll. Div., New York, 1990.
Pierpont, J., *Functions of a Complex Variable (Phoenix Edition)*, Dover Publications, New York, 2005.
Schinzinger, R. and Laura, P. A. A., *Conformal Mapping: Methods and Applications*, Dover Publications, New York, 2003.
Silverman, R. A., *Introductory Complex Analysis*, Dover Publications, New York, 1984.
Spiegel, M. R., *Schaum's Outline of Complex Variables*, McGraw-Hill, New York, 1968.
Sveshnikov, A. G. and Tikhonov, A. N., The Theory of Functions of a Complex Variable, Mir Publishers, Moscow, 1982.
Wunsch, D. A., *Complex Variables with Applications, 2nd Edition*, Addison Wesley, Boston, 1993.

Chapter 11
Integral Transforms

11.1. General Form of Integral Transforms. Some Formulas

11.1.1. Integral Transforms and Inversion Formulas

Normally an integral transform has the form

$$\widetilde{f}(\lambda) = \int_a^b \varphi(x,\lambda) f(x)\,dx. \tag{11.1.1.1}$$

The function $\widetilde{f}(\lambda)$ is called the *transform* of the function $f(x)$ and $\varphi(x,\lambda)$ is called the *kernel* of the integral transform. The function $f(x)$ is called the *inverse transform* of $\widetilde{f}(\lambda)$. The limits of integration a and b are real numbers (usually, $a = 0$, $b = \infty$ or $a = -\infty$, $b = \infty$). For brevity, we rewrite formula (11.1.1.1) as follows: $\widetilde{f}(u) = \mathcal{L}\{f(x)\}$.

General properties of integral transforms (linearity):

$$\mathcal{L}\{kf(x)\} = k\mathcal{L}\{f(x)\},$$
$$\mathcal{L}\{f(x) \pm g(x)\} = \mathcal{L}\{f(x)\} \pm \mathcal{L}\{g(x)\}.$$

Here, k is an arbitrary constant; it is assumed that integral transforms of the functions $f(x)$ and $g(x)$ exist.

In Subsections 11.2–11.6, the most popular (Laplace, Mellin, Fourier, etc.) integral transforms are described. These subsections also describe the corresponding inversion formulas, which normally have the form

$$f(x) = \int_{\mathcal{C}} \psi(x,\lambda)\widetilde{f}(\lambda)\,d\lambda \tag{11.1.1.2}$$

and make it possible to recover $f(x)$ if $\widetilde{f}(\lambda)$ is given. The integration path $\mathcal{C}$ can lie either on the real axis or in the complex plane.

In many cases, to evaluate the integrals in the inversion formula (11.1.1.2) — in particular, to find the inverse Laplace, Mellin, and Fourier transforms — methods of the theory of functions of a complex variable can be applied, including the residue theorem and the Jordan lemma, which are briefly outlined below in Subsection 11.1.2.

11.1.2. Residues. Jordan Lemma

11.1.2-1. Residues. Calculation formulas.

The *residue* of a function $f(z)$ holomorphic in a deleted neighborhood of a point $z = a$ (thus, a is an isolated singularity of f) of the complex plane z is the number

$$\operatorname*{res}_{z=a} f(z) = \frac{1}{2\pi i}\int_{C_\varepsilon} f(z)\,dz, \quad i^2 = -1,$$

where C_ε is a circle of sufficiently small radius ε described by the equation $|z - a| = \varepsilon$.

If the point $z = a$ is a pole of order n* of the function $f(z)$, then we have

$$\operatorname*{res}_{z=a} f(z) = \frac{1}{(n-1)!} \lim_{z\to a} \frac{d^{n-1}}{dx^{n-1}} \bigl[(z-a)^n f(z)\bigr].$$

For a simple pole, which corresponds to $n = 1$, this implies

$$\operatorname*{res}_{z=a} f(z) = \lim_{z\to a} \bigl[(z-a) f(z)\bigr].$$

If $f(z) = \dfrac{\varphi(z)}{\psi(z)}$, where $\varphi(a) \neq 0$ and $\psi(z)$ has a simple zero at the point $z = a$, i.e., $\psi(a) = 0$ and $\psi'_z(a) \neq 0$, then

$$\operatorname*{res}_{z=a} f(z) = \frac{\varphi(a)}{\psi'_z(a)}.$$

11.1.2-2. Jordan lemma.

If a function $f(z)$ is continuous in the domain $|z| \geq R_0$, $\operatorname{Im} z \geq \alpha$, where α is a chosen real number, and if $\lim\limits_{z\to\infty} f(z) = 0$, then

$$\lim_{R\to\infty} \int_{C_R} e^{i\lambda z} f(z)\, dz = 0$$

for any $\lambda > 0$, where C_R is the arc of the circle $|z| = R$ that lies in this domain.

► *For more details about residues and the Jordan lemma, see Paragraphs 10.1.2-7 and 10.1.2-8.*

11.2. Laplace Transform

11.2.1. Laplace Transform and the Inverse Laplace Transform

11.2.1-1. Laplace transform.

The *Laplace transform* of an arbitrary (complex-valued) function $f(x)$ of a real variable x ($x \geq 0$) is defined by

$$\widetilde{f}(p) = \int_0^\infty e^{-px} f(x)\, dx, \tag{11.2.1.1}$$

where $p = s + i\sigma$ is a complex variable.

The Laplace transform exists for any continuous or piecewise-continuous function satisfying the condition $|f(x)| < Me^{\sigma_0 x}$ with some $M > 0$ and $\sigma_0 \geq 0$. In the following, σ_0 often means the greatest lower bound of the possible values of σ_0 in this estimate; this value is called the *growth exponent* of the function $f(x)$.

For any $f(x)$, the transform $\widetilde{f}(p)$ is defined in the half-plane $\operatorname{Re} p > \sigma_0$ and is analytic there.

For brevity, we shall write formula (11.2.1.1) as follows:

$$\widetilde{f}(p) = \mathfrak{L}\bigl\{f(x)\bigr\}, \qquad \text{or} \qquad \widetilde{f}(p) = \mathfrak{L}\bigl\{f(x),\, p\bigr\}.$$

* In a neighborhood of this point we have $f(z) \approx \text{const}\,(z-a)^{-n}$.

11.2.1-2. Inverse Laplace transform.

Given the transform $\widetilde{f}(p)$, the function $f(x)$ can be found by means of the inverse Laplace transform

$$f(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \widetilde{f}(p)e^{px}\,dp, \qquad i^2 = -1, \tag{11.2.1.2}$$

where the integration path is parallel to the imaginary axis and lies to the right of all singularities of $\widetilde{f}(p)$, which corresponds to $c > \sigma_0$.

The integral in inversion formula (11.2.1.2) is understood in the sense of the Cauchy principal value:

$$\int_{c-i\infty}^{c+i\infty} \widetilde{f}(p)e^{px}\,dp = \lim_{\omega\to\infty}\int_{c-i\omega}^{c+i\omega} \widetilde{f}(p)e^{px}\,dp.$$

In the domain $x < 0$, formula (11.2.1.2) gives $f(x) \equiv 0$.

Formula (11.2.1.2) holds for continuous functions. If $f(x)$ has a (finite) jump discontinuity at a point $x = x_0 > 0$, then the left-hand side of (11.2.1.2) is equal to $\frac{1}{2}[f(x_0-0)+f(x_0+0)]$ at this point (for $x_0 = 0$, the first term in the square brackets must be omitted).

For brevity, we write the Laplace inversion formula (11.2.1.2) as follows:

$$f(x) = \mathfrak{L}^{-1}\big\{\widetilde{f}(p)\big\} \qquad \text{or} \qquad f(x) = \mathfrak{L}^{-1}\big\{\widetilde{f}(p),\, x\big\}.$$

There are tables of direct and inverse Laplace transforms (see Sections T3.1 and T3.2), which are handy in solving linear differential and integral equations.

11.2.2. Main Properties of the Laplace Transform. Inversion Formulas for Some Functions

11.2.2-1. Convolution theorem. Main properties of the Laplace transform.

1°. The *convolution* of two functions $f(x)$ and $g(x)$ is defined as an integral of the form $\int_0^x f(t)g(x-t)\,dt$, and is usually denoted by $f(x) * g(x)$ or

$$f(x) * g(x) = \int_0^x f(t)\,g(x-t)\,dt.$$

By performing substitution $x - t = u$, we see that the convolution is symmetric with respect to the convolved functions: $f(x) * g(x) = g(x) * f(x)$.

The *convolution theorem* states that

$$\mathfrak{L}\big\{f(x) * g(x)\big\} = \mathfrak{L}\big\{f(x)\big\}\,\mathfrak{L}\big\{g(x)\big\}$$

and is frequently applied to solve Volterra equations with kernels depending on the difference of the arguments.

2°. The main properties of the correspondence between functions and their Laplace transforms are gathered in Table 11.1.

3°. The Laplace transforms of some functions are listed in Table 11.2; for more detailed tables, see Section T3.1 and the list of references at the end of this chapter.

TABLE 11.1
Main properties of the Laplace transform

No.	Function	Laplace transform	Operation
1	$af_1(x) + bf_2(x)$	$a\widetilde{f}_1(p) + b\widetilde{f}_2(p)$	Linearity
2	$f(x/a),\ a > 0$	$a\widetilde{f}(ap)$	Scaling
3	$f(x-a)$, $f(\xi) \equiv 0$ for $\xi < 0$	$e^{-ap}\widetilde{f}(p)$	Shift of the argument
4	$x^n f(x);\ n = 1, 2, \ldots$	$(-1)^n \widetilde{f}_p^{(n)}(p)$	Differentiation of the transform
5	$\dfrac{1}{x} f(x)$	$\displaystyle\int_p^\infty \widetilde{f}(q)\, dq$	Integration of the transform
6	$e^{ax} f(x)$	$\widetilde{f}(p-a)$	Shift in the complex plane
7	$f'_x(x)$	$p\widetilde{f}(p) - f(+0)$	Differentiation
8	$f_x^{(n)}(x)$	$p^n \widetilde{f}(p) - \displaystyle\sum_{k=1}^{n} p^{n-k} f_x^{(k-1)}(+0)$	Differentiation
9	$x^m f_x^{(n)}(x),\ m = 1, 2, \ldots$	$(-1)^m \dfrac{d^m}{dp^m}\left[p^n \widetilde{f}(p) - \displaystyle\sum_{k=1}^{n} p^{n-k} f_x^{(k-1)}(+0)\right]$	Differentiation
10	$\dfrac{d^n}{dx^n}\left[x^m f(x)\right],\ m \geq n$	$(-1)^m p^n \dfrac{d^m}{dp^m}\widetilde{f}(p)$	Differentiation
11	$\displaystyle\int_0^x f(t)\, dt$	$\dfrac{\widetilde{f}(p)}{p}$	Integration
12	$\displaystyle\int_0^x f_1(t) f_2(x-t)\, dt$	$\widetilde{f}_1(p)\widetilde{f}_2(p)$	Convolution

TABLE 11.2
The Laplace transforms of some functions

No.	Function, $f(x)$	Laplace transform, $\widetilde{f}(p)$	Remarks
1	1	$1/p$	
2	x^n	$\dfrac{n!}{p^{n+1}}$	$n = 1, 2, \ldots$
3	x^a	$\Gamma(a+1)p^{-a-1}$	$a > -1$
4	e^{-ax}	$(p+a)^{-1}$	
5	$x^a e^{-bx}$	$\Gamma(a+1)(p+b)^{-a-1}$	$a > -1$
6	$\sinh(ax)$	$\dfrac{a}{p^2 - a^2}$	
7	$\cosh(ax)$	$\dfrac{p}{p^2 - a^2}$	
8	$\ln x$	$-\dfrac{1}{p}(\ln p + \mathcal{C})$	$\mathcal{C} = 0.5772\ldots$ is the Euler constant
9	$\sin(ax)$	$\dfrac{a}{p^2 + a^2}$	
10	$\cos(ax)$	$\dfrac{p}{p^2 + a^2}$	
11	$\operatorname{erfc}\left(\dfrac{a}{2\sqrt{x}}\right)$	$\dfrac{1}{p}\exp\left(-a\sqrt{p}\right)$	$a \geq 0$
12	$J_0(ax)$	$\dfrac{1}{\sqrt{p^2 + a^2}}$	$J_0(x)$ is the Bessel function

11.2.2-2. Inverse transforms of rational functions.

Consider the important case in which the transform is a rational function of the form

$$\widetilde{f}(p) = \frac{R(p)}{Q(p)}, \tag{11.2.2.1}$$

where $Q(p)$ and $R(p)$ are polynomials in the variable p and the degree of $Q(p)$ exceeds that of $R(p)$.

Assume that the zeros of the denominator are simple, i.e.,

$$Q(p) \equiv \text{const}\,(p-\lambda_1)(p-\lambda_2)\ldots(p-\lambda_n).$$

Then the inverse transform can be determined by the formula

$$f(x) = \sum_{k=1}^{n} \frac{R(\lambda_k)}{Q'(\lambda_k)} \exp(\lambda_k x), \tag{11.2.2.2}$$

where the primes denote the derivatives.

If $Q(p)$ has multiple zeros, i.e.,

$$Q(p) \equiv \text{const}\,(p-\lambda_1)^{s_1}(p-\lambda_2)^{s_2}\ldots(p-\lambda_m)^{s_m},$$

then

$$f(x) = \sum_{k=1}^{m} \frac{1}{(s_k-1)!} \lim_{p\to s_k} \frac{d^{s_k-1}}{dp^{s_k-1}}\big[(p-\lambda_k)^{s_k}\widetilde{f}(p)e^{px}\big]. \tag{11.2.2.3}$$

Example 1. The transform

$$\widetilde{f}(p) = \frac{b}{p^2-a^2} \qquad (a, b \text{ real numbers})$$

can be represented as the fraction (11.2.2.1) with $R(p) = b$ and $Q(p) = (p-a)(p+a)$. The denominator $Q(p)$ has two simple roots, $\lambda_1 = a$ and $\lambda_2 = -a$. Using formula (11.2.2.2) with $n = 2$ and $Q'(p) = 2p$, we obtain the inverse transform in the form

$$f(x) = \frac{b}{2a}e^{ax} - \frac{b}{2a}e^{-ax} = \frac{b}{a}\sinh(ax).$$

Example 2. The transform

$$\widetilde{f}(p) = \frac{b}{p^2+a^2} \qquad (a, b \text{ real numbers})$$

can be written as the fraction (11.2.2.1) with $R(p) = b$ and $Q(p) = (p-ia)(p+ia)$, $i^2 = -1$. The denominator $Q(p)$ has two simple pure imaginary roots, $\lambda_1 = ia$ and $\lambda_2 = -ia$. Using formula (11.2.2.2) with $n = 2$, we find the inverse transform:

$$f(x) = \frac{b}{2ia}e^{iax} - \frac{b}{2ia}e^{-iax} = -\frac{bi}{2a}\big[\cos(ax) + i\sin(ax)\big] + \frac{bi}{2a}\big[\cos(ax) - i\sin(ax)\big] = \frac{b}{a}\sin(ax).$$

Example 3. The transform

$$\widetilde{f}(p) = ap^{-n},$$

where n is a positive integer, can be written as the fraction (11.2.2.1) with $R(p) = a$ and $Q(p) = p^n$. The denominator $Q(p)$ has one root of multiplicity n, $\lambda_1 = 0$. By formula (11.2.2.3) with $m = 1$ and $s_1 = n$, we find the inverse transform:

$$f(x) = \frac{a}{(n-1)!}x^{n-1}.$$

► *Fairly detailed tables of inverse Laplace transforms can be found in Section T3.2.*

11.2.2-3. Inversion of functions with finitely many singular points.

If the function $\widetilde{f}(p)$ has finitely many singular points, $p_1, p_2, \ldots, p_n$, and tends to zero as $p \to \infty$, then the integral in the Laplace inversion formula (11.2.1.2) may be evaluated using the residue theory by applying the Jordan lemma (see Subsection 11.1.2). In this case

$$f(x) = \sum_{k=1}^{n} \operatorname*{res}_{p=p_k} [\widetilde{f}(p)e^{px}]. \tag{11.2.2.4}$$

Formula (11.2.2.4) can be extended to the case where $\widetilde{f}(p)$ has infinitely many singular points. In this case, $f(x)$ is represented as an infinite series.

11.2.3. Limit Theorems. Representation of Inverse Transforms as Convergent Series and Asymptotic Expansions

11.2.3-1. Limit theorems.

THEOREM 1. *Let $0 \le x < \infty$ and $\widetilde{f}(p) = \mathfrak{L}\{f(x)\}$ be the Laplace transform of $f(x)$. If a limit of $f(x)$ as $x \to 0$ exists, then*

$$\lim_{x\to 0} f(x) = \lim_{p\to\infty} \bigl[p\widetilde{f}(p)\bigr].$$

THEOREM 2. *If a limit of $f(x)$ as $x \to \infty$ exists, then*

$$\lim_{x\to\infty} f(x) = \lim_{p\to 0} \bigl[p\widetilde{f}(p)\bigr].$$

11.2.3-2. Representation of inverse transforms as convergent series.

THEOREM 1. *Suppose the transform $\widetilde{f}(p)$ can be expanded into series in negative powers of p,*

$$\widetilde{f}(p) = \sum_{n=1}^{\infty} \frac{a_n}{p^n},$$

convergent for $|p| > R$, where R is an arbitrary positive number; note that the transform tends to zero as $|p| \to \infty$. Then the inverse transform can be obtained by the formula

$$f(x) = \sum_{n=1}^{\infty} \frac{a_n}{(n-1)!} x^{n-1},$$

where the series on the right-hand side is convergent for all x.

THEOREM 2. *Suppose the transform $\widetilde{f}(p)$, $|p| > R$, is represented by an absolutely convergent series,*

$$\widetilde{f}(p) = \sum_{n=0}^{\infty} \frac{a_n}{p^{\lambda_n}}, \tag{11.2.3.1}$$

where $\{\lambda_n\}$ is any positive increasing sequence, $0 < \lambda_0 < \lambda_1 < \cdots \to \infty$. Then it is possible to proceed termwise from series (11.2.3.1) to the following inverse transform series:

$$f(x) = \sum_{n=0}^{\infty} \frac{a_n}{\Gamma(\lambda_n)} x^{\lambda_n - 1}, \tag{11.2.3.2}$$

where $\Gamma(\lambda)$ is the Gamma function. Series (11.2.3.2) is convergent for all real and complex values of x other than zero (if $\lambda_0 \geq 1$, the series is convergent for all x).

11.2.3-3. Representation of inverse transforms as asymptotic expansions as $x \to \infty$.

1°. Let $p = p_0$ be a singular point of the Laplace transform $\widetilde{f}(p)$ with the greatest real part (it is assumed there is only one such point). If $\widetilde{f}(p)$ can be expanded near $p = p_0$ into an absolutely convergent series,

$$\widetilde{f}(p) = \sum_{n=0}^{\infty} c_n (p - p_0)^{\lambda_n} \qquad (\lambda_0 < \lambda_1 < \cdots \to \infty) \tag{11.2.3.3}$$

with arbitrary λ_n, then the inverse transform $f(x)$ can be expressed in the form of the asymptotic expansion

$$f(x) \sim e^{p_0 x} \sum_{n=0}^{\infty} \frac{c_n}{\Gamma(-\lambda_n)} x^{-\lambda_n - 1} \quad \text{as} \quad x \to \infty. \tag{11.2.3.4}$$

The terms corresponding to nonnegative integer λ_n must be omitted from the summation, since $\Gamma(0) = \Gamma(-1) = \Gamma(-2) = \cdots = \infty$.

2°. If the transform $\widetilde{f}(p)$ has several singular points, $p_1, \ldots, p_m$, with the same greatest real part, $\operatorname{Re} p_1 = \cdots = \operatorname{Re} p_m$, then expansions of the form (11.2.3.3) should be obtained for each of these points and the resulting expressions must be added together.

11.2.3-4. Post–Widder formula.

In applications, one can find $f(x)$ if the Laplace transform $\widetilde{f}(t)$ on the real semiaxis is known for $t = p \geq 0$. To this end, one uses the Post–Widder formula

$$f(x) = \lim_{n \to \infty} \left[\frac{(-1)^n}{n!} \left(\frac{n}{x} \right)^{n+1} \widetilde{f}_t^{(n)} \left(\frac{n}{x} \right) \right]. \tag{11.2.3.5}$$

Approximate inversion formulas are obtained by taking sufficiently large positive integer n in (11.2.3.5) instead of passing to the limit.

11.3. Mellin Transform

11.3.1. Mellin Transform and the Inversion Formula

11.3.1-1. Mellin transform.

Suppose that a function $f(x)$ is defined for positive x and satisfies the conditions

$$\int_0^1 |f(x)|\, x^{\sigma_1 - 1}\, dx < \infty, \qquad \int_1^{\infty} |f(x)|\, x^{\sigma_2 - 1}\, dx < \infty$$

for some real numbers σ_1 and σ_2, $\sigma_1 < \sigma_2$.

The Mellin transform of $f(x)$ is defined by

$$\hat{f}(s) = \int_0^\infty f(x)x^{s-1}\,dx, \tag{11.3.1.1}$$

where $s = \sigma + i\tau$ is a complex variable ($\sigma_1 < \sigma < \sigma_2$).

For brevity, we rewrite formula (11.3.1.1) as follows:

$$\hat{f}(s) = \mathfrak{M}\{f(x)\} \qquad \text{or} \qquad \hat{f}(s) = \mathfrak{M}\{f(x), s\}.$$

11.3.1-2. Inverse Mellin transform.

Given $\hat{f}(s)$, the function $f(x)$ can be found by means of the *inverse Mellin transform*

$$f(x) = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} \hat{f}(s)x^{-s}\,ds \qquad (\sigma_1 < \sigma < \sigma_2), \tag{11.3.1.2}$$

where the integration path is parallel to the imaginary axis of the complex plane s and the integral is understood in the sense of the Cauchy principal value.

Formula (11.3.1.2) holds for continuous functions. If $f(x)$ has a (finite) jump discontinuity at a point $x = x_0 > 0$, then the left-hand side of (11.3.1.2) is equal to $\frac{1}{2}\big[f(x_0-0)+f(x_0+0)\big]$ at this point (for $x_0 = 0$, the first term in the square brackets must be omitted).

For brevity, we rewrite formula (11.3.1.2) in the form

$$f(x) = \mathfrak{M}^{-1}\{\hat{f}(s)\} \qquad \text{or} \qquad f(x) = \mathfrak{M}^{-1}\{\hat{f}(s),\, x\}.$$

11.3.2. Main Properties of the Mellin Transform. Relation Among the Mellin, Laplace, and Fourier Transforms

11.3.2-1. Main properties of the Mellin transform.

1°. The main properties of the correspondence between the functions and their Mellin transforms are gathered in Table 11.3.

2°. The integral relations

$$\int_0^\infty f(x)g(x)\,dx = \mathfrak{M}^{-1}\{\hat{f}(s)\hat{g}(1-s)\},$$
$$\int_0^\infty f(x)g\Big(\frac{1}{x}\Big)\,dx = \mathfrak{M}^{-1}\{\hat{f}(s)\hat{g}(s)\}$$

hold for fairly general assumptions about the integrability of the functions involved (see Ditkin and Prudnikov, 1965).

11.3.2-2. Relation among the Mellin, Laplace, and Fourier transforms.

There are tables of direct and inverse Mellin transforms (see Sections T3.5 and T3.6 and the references listed at the end of the current chapter) that are useful in solving specific integral and differential equations. The Mellin transform is related to the Laplace and Fourier transforms as follows:

$$\mathfrak{M}\{f(x), s\} = \mathfrak{L}\{f(e^x), -s\} + \mathfrak{L}\{f(e^{-x}), s\} = \mathfrak{F}\{f(e^x), is\},$$

which makes it possible to apply much more common tables of direct and inverse Laplace and Fourier transforms.

TABLE 11.3
Main properties of the Mellin transform

No.	Function	Mellin transform	Operation
1	$af_1(x)+bf_2(x)$	$a\hat{f}_1(s)+b\hat{f}_2(s)$	Linearity
2	$f(ax),\ a>0$	$a^{-s}\hat{f}(s)$	Scaling
3	$x^a f(x)$	$\hat{f}(s+a)$	Shift of the argument of the transform
4	$f(x^2)$	$\frac{1}{2}\hat{f}\left(\frac{1}{2}s\right)$	Squared argument
5	$f(1/x)$	$\hat{f}(-s)$	Inversion of the argument of the transform
6	$x^\lambda f\left(ax^\beta\right),\ a>0, \beta\neq 0$	$\frac{1}{\beta}a^{-\frac{s+\lambda}{\beta}}\hat{f}\left(\frac{s+\lambda}{\beta}\right)$	Power law transform
7	$f'_x(x)$	$-(s-1)\hat{f}(s-1)$	Differentiation
8	$xf'_x(x)$	$-s\hat{f}(s)$	Differentiation
9	$f^{(n)}_x(x)$	$(-1)^n\dfrac{\Gamma(s)}{\Gamma(s-n)}\hat{f}(s-n)$	Multiple differentiation
10	$\left(x\dfrac{d}{dx}\right)^n f(x)$	$(-1)^n s^n\hat{f}(s)$	Multiple differentiation
11	$x^\alpha\displaystyle\int_0^\infty t^\beta f_1(xt)f_2(t)\,dt$	$\hat{f}_1(s+\alpha)\hat{f}_2(1-s-\alpha+\beta)$	Complicated integration
12	$x^\alpha\displaystyle\int_0^\infty t^\beta f_1\left(\frac{x}{t}\right)f_2(t)\,dt$	$\hat{f}_1(s+\alpha)\hat{f}_2(s+\alpha+\beta+1)$	Complicated integration

11.4. Various Forms of the Fourier Transform

11.4.1. Fourier Transform and the Inverse Fourier Transform

11.4.1-1. Standard form of the Fourier transform.

The *Fourier transform* is defined as follows:

$$\widetilde{f}(u)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}f(x)\,e^{-iux}\,dx. \tag{11.4.1.1}$$

For brevity, we rewrite formula (11.4.1.1) as follows:

$$\widetilde{f}(u)=\mathfrak{F}\{f(x)\}\qquad\text{or}\qquad\widetilde{f}(u)=\mathfrak{F}\{f(x),u\}.$$

Given $\widetilde{f}(u)$, the function $f(x)$ can be found by means of the *inverse Fourier transform*

$$f(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\widetilde{f}(u)\,e^{iux}\,du. \tag{11.4.1.2}$$

Formula (11.4.1.2) holds for continuous functions. If $f(x)$ has a (finite) jump discontinuity at a point $x=x_0$, then the left-hand side of (11.4.1.2) is equal to $\frac{1}{2}\left[f(x_0-0)+f(x_0+0)\right]$ at this point.

TABLE 11.4
Main properties of the Fourier transform

No.	Function	Fourier transform	Operation
1	$af_1(x)+bf_2(x)$	$a\widetilde{f}_1(u)+b\widetilde{f}_2(u)$	Linearity
2	$f(x/a),\ a>0$	$a\widetilde{f}(au)$	Scaling
3	$x^n f(x);\ n=1,2,\ldots$	$i^n\widetilde{f}_u^{(n)}(u)$	Differentiation of the transform
4	$f''_{xx}(x)$	$-u^2\widetilde{f}(u)$	Differentiation
5	$f_x^{(n)}(x)$	$(iu)^n\widetilde{f}(u)$	Differentiation
6	$\displaystyle\int_{-\infty}^{\infty} f_1(\xi)f_2(x-\xi)\,d\xi$	$\widetilde{f}_1(u)\widetilde{f}_2(u)$	Convolution

For brevity, we rewrite formula (11.4.1.2) as follows:

$$f(x)=\mathfrak{F}^{-1}\{\widetilde{f}(u)\}\qquad\text{or}\qquad f(x)=\mathfrak{F}^{-1}\{\widetilde{f}(u),\ x\}.$$

11.4.1-2. Asymmetric form of the Fourier transform. Alternative Fourier transform.

1°. Sometimes it is more convenient to define the Fourier transform by

$$\check{f}(u)=\int_{-\infty}^{\infty} f(x)e^{-iux}\,dx.$$

In this case, the *Fourier inversion formula* reads

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\check{f}(u)e^{iux}\,du.$$

2°. Sometimes the alternative Fourier transform is used (and called merely the *Fourier transform*), which corresponds to the renaming $e^{-iux}\rightleftarrows e^{iux}$ on the right-hand sides of (11.4.1.1) and (11.4.1.2).

11.4.1-3. Convolution theorem. Main properties of the Fourier transforms.

1°. The *convolution* of two functions $f(x)$ and $g(x)$ is defined as

$$f(x)*g(x)\equiv\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(x-t)g(t)\,dt.$$

By performing substitution $x-t=u$, we see that the convolution is symmetric with respect to the convolved functions: $f(x)*g(x)=g(x)*f(x)$.

The *convolution theorem* states that

$$\mathfrak{F}\big\{f(x)*g(x)\big\}=\mathfrak{F}\big\{f(x)\big\}\,\mathfrak{F}\big\{g(x)\big\}.$$

2°. The main properties of the correspondence between functions and their Fourier transforms are gathered in Table 11.4.

11.4.1-4. n-dimensional Fourier transform.

The Fourier transform admits n-dimensional generalization:

$$\widetilde{f}(\mathbf{u}) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(\mathbf{x}) e^{-i(\mathbf{u}\cdot\mathbf{x})}\, d\mathbf{x}, \qquad (\mathbf{u}\cdot\mathbf{x}) = u_1x_1 + \cdots + u_nx_n, \tag{11.4.1.3}$$

where $f(\mathbf{x}) = f(x_1, \ldots, x_n)$, $\widetilde{f}(\mathbf{u}) = f(u_1, \ldots, u_n)$, and $d\mathbf{x} = dx_1 \ldots dx_n$.

The corresponding inversion formula is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \widetilde{f}(\mathbf{u}) e^{i(\mathbf{u}\cdot\mathbf{x})}\, d\mathbf{u}, \qquad d\mathbf{u} = du_1 \ldots du_n.$$

The Fourier transform (11.4.1.3) is frequently used in the theory of linear partial differential equations with constant coefficients ($\mathbf{x} \in \mathbb{R}^n$).

11.4.2. Fourier Cosine and Sine Transforms

11.4.2-1. Fourier cosine transform.

1°. Let a function $f(x)$ be integrable on the semiaxis $0 \le x < \infty$. The *Fourier cosine transform* is defined by

$$\widetilde{f}_{\mathrm{c}}(u) = \sqrt{\frac{2}{\pi}} \int_0^\infty f(x) \cos(xu)\, dx, \qquad 0 < u < \infty. \tag{11.4.2.1}$$

For given $\widetilde{f}_{\mathrm{c}}(u)$, the function can be found by means of the *Fourier cosine inversion formula*

$$f(x) = \sqrt{\frac{2}{\pi}} \int_0^\infty \widetilde{f}_{\mathrm{c}}(u) \cos(xu)\, du, \qquad 0 < x < \infty. \tag{11.4.2.2}$$

The Fourier cosine transform (11.4.2.1) is denoted for brevity by $\widetilde{f}_{\mathrm{c}}(u) = \mathfrak{F}_{\mathrm{c}}\{f(x)\}$.

2°. It follows from formula (11.4.2.2) that the Fourier cosine transform has the property $\mathfrak{F}_{\mathrm{c}}^2 = 1$.

Some other properties of the Fourier cosine transform:

$$\mathfrak{F}_{\mathrm{c}}\{x^{2n} f(x)\} = (-1)^n \frac{d^{2n}}{du^{2n}} \mathfrak{F}_{\mathrm{c}}\{f(x)\}, \quad n = 1, 2, \ldots;$$
$$\mathfrak{F}_{\mathrm{c}}\{f''(x)\} = -u^2 \mathfrak{F}_{\mathrm{c}}\{f(x)\}.$$

Here, $f(x)$ is assumed to vanish sufficiently rapidly (exponentially) as $x \to \infty$. For the second formula, the condition $f'(0) = 0$ is assumed to hold.

Parseval's relation for the Fourier cosine transform:

$$\int_0^\infty \mathfrak{F}_{\mathrm{c}}\{f(x)\} \mathfrak{F}_{\mathrm{c}}\{g(x)\}\, du = \int_0^\infty f(x) g(x)\, dx.$$

There are tables of the Fourier cosine transform (see Section T3.3 and the references listed at the end of the current chapter).

3°. Sometimes the asymmetric form of the Fourier cosine transform is applied, which is given by the pair of formulas

$$\check{f}_{\mathrm{c}}(u) = \int_0^\infty f(x) \cos(xu)\, dx, \qquad f(x) = \frac{2}{\pi} \int_0^\infty \check{f}_{\mathrm{c}}(u) \cos(xu)\, du.$$

11.4.2-2. Fourier sine transform.

1°. Let a function $f(x)$ be integrable on the semiaxis $0 \le x < \infty$. The *Fourier sine transform* is defined by

$$\widetilde{f}_{\mathrm{s}}(u) = \sqrt{\frac{2}{\pi}} \int_0^\infty f(x) \sin(xu)\, dx, \qquad 0 < u < \infty. \tag{11.4.2.3}$$

For given $\widetilde{f}_{\mathrm{s}}(u)$, the function $f(x)$ can be found by means of the *inverse Fourier sine transform*

$$f(x) = \sqrt{\frac{2}{\pi}} \int_0^\infty \widetilde{f}_{\mathrm{s}}(u) \sin(xu)\, du, \qquad 0 < x < \infty. \tag{11.4.2.4}$$

The Fourier sine transform (11.4.2.3) is briefly denoted by $\widetilde{f}_{\mathrm{s}}(u) = \mathfrak{F}_{\mathrm{s}}\{f(x)\}$.

2°. It follows from formula (11.4.2.4) that the Fourier sine transform has the property $\mathfrak{F}_{\mathrm{s}}^2 = 1$.

Some other properties of the Fourier sine transform:

$$\mathfrak{F}_{\mathrm{s}}\{x^{2n} f(x)\} = (-1)^n \frac{d^{2n}}{du^{2n}} \mathfrak{F}_{\mathrm{s}}\{f(x)\}, \quad n = 1, 2, \ldots;$$
$$\mathfrak{F}_{\mathrm{s}}\{f''(x)\} = -u^2 \mathfrak{F}_{\mathrm{s}}\{f(x)\}.$$

Here, $f(x)$ is assumed to vanish sufficiently rapidly (exponentially) as $x \to \infty$. For the second formula, the condition $f(0) = 0$ is assumed to hold.

Parseval's relation for the Fourier sine transform:

$$\int_0^\infty \mathfrak{F}_{\mathrm{s}}\{f(x)\}\mathfrak{F}_{\mathrm{s}}\{g(x)\}\, du = \int_0^\infty f(x)g(x)\, dx.$$

There are tables of the Fourier cosine transform (see Section T3.4 and the references listed at the end of the current chapter).

3°. Sometimes it is more convenient to apply the asymmetric form of the Fourier sine transform defined by the following two formulas:

$$\check{f}_{\mathrm{s}}(u) = \int_0^\infty f(x) \sin(xu)\, dx, \qquad f(x) = \frac{2}{\pi} \int_0^\infty \check{f}_{\mathrm{s}}(u) \sin(xu)\, du.$$

11.5. Other Integral Transforms

11.5.1. Integral Transforms Whose Kernels Contain Bessel Functions and Modified Bessel Functions

11.5.1-1. Hankel transform.

1°. The *Hankel transform* is defined as follows:

$$\widetilde{f}_\nu(u) = \int_0^\infty x J_\nu(ux) f(x)\, dx, \qquad 0 < u < \infty, \tag{11.5.1.1}$$

where $\nu > -\frac{1}{2}$ and $J_\nu(x)$ is the Bessel function of the first kind of order ν (see Section SF.6).

For given $\widetilde{f}_\nu(u)$, the function $f(x)$ can be found by means of the *Hankel inversion formula*

$$f(x) = \int_0^\infty uJ_\nu(ux)\widetilde{f}_\nu(u)\,du, \qquad 0 < x < \infty. \tag{11.5.1.2}$$

Note that if $f(x) = O(x^\alpha)$ as $x \to 0$, where $\alpha + \nu + 2 > 0$, and $f(x) = O(x^\beta)$ as $x \to \infty$, where $\beta + \frac{3}{2} < 0$, then the integral (11.5.1.1) is convergent.

The inversion formula (11.5.1.2) holds for continuous functions. If $f(x)$ has a (finite) jump discontinuity at a point $x = x_0$, then the left-hand side of (11.5.1.2) is equal to $\frac{1}{2}[f(x_0 - 0) + f(x_0 + 0)]$ at this point.

For brevity, we denote the Hankel transform (11.5.1.1) by $\widetilde{f}_\nu(u) = \mathcal{H}_\nu\{f(x)\}$.

2°. It follows from formula (11.5.1.2) that the Hankel transform has the property $\mathcal{H}_\nu^2 = 1$.

Other properties of the Hankel transform:

$$\mathcal{H}_\nu\left\{\frac{1}{x}f(x)\right\} = \frac{u}{2\nu}\mathcal{H}_{\nu-1}\{f(x)\} + \frac{u}{2\nu}\mathcal{H}_{\nu+1}\{f(x)\},$$

$$\mathcal{H}_\nu\{f'(x)\} = \frac{(\nu-1)u}{2\nu}\mathcal{H}_{\nu+1}\{f(x)\} - \frac{(\nu+1)u}{2\nu}\mathcal{H}_{\nu-1}\{f(x)\},$$

$$\mathcal{H}_\nu\left\{f''(x) + \frac{1}{x}f'(x) - \frac{\nu^2}{x^2}f(x)\right\} = -u^2\mathcal{H}_\nu\{f(x)\}.$$

The conditions

$$\lim_{x\to 0}\left[x^\nu f(x)\right] = 0, \quad \lim_{x\to 0}\left[x^{\nu+1}f'(x)\right] = 0, \quad \lim_{x\to\infty}\left[x^{1/2}f(x)\right] = 0, \quad \lim_{x\to\infty}\left[x^{1/2}f'(x)\right] = 0$$

are assumed to hold for the last formula.

Parseval's relation for the Hankel transform:

$$\int_0^\infty u\mathcal{H}_\nu\{f(x)\}\mathcal{H}_\nu\{g(x)\}\,du = \int_0^\infty xf(x)g(x)\,dx, \qquad \nu > -\frac{1}{2}.$$

11.5.1-2. Meijer transform.

The *Meijer transform* is defined as follows:

$$\hat{f}_\mu(s) = \sqrt{\frac{2}{\pi}}\int_0^\infty \sqrt{sx}\,K_\mu(sx)f(x)\,dx, \qquad 0 < s < \infty,$$

where $K_\mu(x)$ is the modified Bessel function of the second kind (the Macdonald function) of order μ (see Section SF.7).

For given $\widetilde{f}_\mu(s)$, the function $f(x)$ can be found by means of the *Meijer inversion formula*

$$f(x) = \frac{1}{i\sqrt{2\pi}}\int_{c-i\infty}^{c+i\infty} \sqrt{sx}\,I_\mu(sx)\hat{f}_\mu(s)\,ds, \qquad 0 < x < \infty,$$

where $I_\mu(x)$ is the modified Bessel function of the first kind of order μ (see Section SF.7). For the Meijer transform, a convolution is defined and an operational calculus is developed.

11.5.1-3. Kontorovich–Lebedev transform.

The *Kontorovich–Lebedev transform* is introduced as follows:

$$F(\tau) = \int_0^\infty K_{i\tau}(x) f(x)\, dx, \qquad 0 < \tau < \infty,$$

where $K_\mu(x)$ is the modified Bessel function of the second kind (the Macdonald function) of order μ (see Section SF.7) and $i = \sqrt{-1}$.

For given $F(\tau)$, the function can be found by means of the *Kontorovich–Lebedev inversion formula*

$$f(x) = \frac{2}{\pi^2 x} \int_0^\infty \tau \sinh(\pi\tau) K_{i\tau}(x) F(\tau)\, d\tau, \qquad 0 < x < \infty.$$

11.5.1-4. Y-transform.

The *Y-transform* is defined by

$$F_\nu(u) = \int_0^\infty \sqrt{ux}\, Y_\nu(ux) f(x)\, dx,$$

where $Y_\nu(x)$ is the Bessel function of the second kind of order ν.

Given a transform $F_\nu(u)$, the inverse Y-transform $f(x)$ is found by the inversion formula

$$f(x) = \int_0^\infty \sqrt{ux}\, \mathbf{H}_\nu(ux) F_\nu(u)\, du,$$

where $\mathbf{H}_\nu(x)$ is the Struve function, which is defined as

$$\mathbf{H}_\nu(x) = \sum_{j=0}^\infty \frac{(-1)^j (x/2)^{\nu+2j+1}}{\Gamma\left(j+\frac{3}{2}\right)\Gamma\left(\nu+j+\frac{3}{2}\right)}.$$

11.5.2. Summary Table of Integral Transforms. Areas of Application of Integral Transforms

11.5.2-1. Summary table of integral transforms.

Table 11.5 summarizes the integral transforms considered above and also lists some other integral transforms; for the constraints imposed on the functions and parameters occurring in the integrand, see the references given at the end of this section.

11.5.2-2. Areas of application of integral transforms.

Integral transforms are widely used for the evaluation of integrals, summation of series, and solution of various mathematical equations and problems. In particular, the application of an appropriate integral transform to linear ordinary differential, integral, and difference equations reduces the problem to a linear algebraic equation for the transform; and linear partial differential equations are reduced to an ordinary differential equation.

TABLE 11.5
Summary table of integral transforms

Integral transform	Definition	Inversion formula
Laplace transform	$\widetilde{f}(p)=\int_0^\infty e^{-px}f(x)\,dx$	$f(x)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
Laplace–Carlson transform	$\widetilde{f}(p)=p\int_0^\infty e^{-px}f(x)\,dx$	$f(x)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{px}\frac{\widetilde{f}(p)}{p}\,dp$
Two-sided Laplace transform	$\widetilde{f}_*(p)=\int_{-\infty}^\infty e^{-px}f(x)\,dx$	$f(x)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}_*(p)\,dp$
Fourier transform	$\widetilde{f}(u)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-iux}f(x)\,dx$	$f(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{iux}\widetilde{f}(u)\,du$
Fourier sine transform	$\widetilde{f}_{\rm s}(u)=\sqrt{\frac{2}{\pi}}\int_0^\infty \sin(xu)f(x)\,dx$	$f(x)=\sqrt{\frac{2}{\pi}}\int_0^\infty \sin(xu)\widetilde{f}_{\rm s}(u)\,du$
Fourier cosine transform	$\widetilde{f}_{\rm c}(u)=\sqrt{\frac{2}{\pi}}\int_0^\infty \cos(xu)f(x)\,dx$	$f(x)=\sqrt{\frac{2}{\pi}}\int_0^\infty \cos(xu)\widetilde{f}_{\rm c}(u)\,du$
Hartley transform	$\widetilde{f}_{\rm h}(u)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty(\cos xu+\sin xu)f(x)\,dx$	$f(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty(\cos xu+\sin xu)\widetilde{f}_{\rm h}(u)\,du$
Mellin transform	$\widehat{f}(s)=\int_0^\infty x^{s-1}f(x)\,dx$	$f(x)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} x^{-s}\widehat{f}(s)\,ds$
Hankel transform	$\widehat{f}_\nu(w)=\int_0^\infty xJ_\nu(xw)f(x)\,dx$	$f(x)=\int_0^\infty wJ_\nu(xw)\widehat{f}_\nu(w)\,dw$
Y-transform	$F_\nu(u)=\int_0^\infty \sqrt{ux}\,Y_\nu(ux)f(x)\,dx$	$f(x)=\int_0^\infty \sqrt{ux}\,\mathbf{H}_\nu(ux)F_\nu(u)\,du$
Meijer transform (K-transform)	$\widehat{f}(s)=\sqrt{\frac{2}{\pi}}\int_0^\infty \sqrt{sx}\,K_\nu(sx)f(x)\,dx$	$f(x)=\frac{1}{i\sqrt{2\pi}}\int_{c-i\infty}^{c+i\infty}\sqrt{sx}\,I_\nu(sx)\widehat{f}(s)\,ds$
Bochner transform	$\widetilde{f}(r)=\int_0^\infty J_{n/2-1}(2\pi xr)G(x,r)f(x)\,dx,$ $G(x,r)=2\pi r(x/r)^{n/2},\quad n=1,2,\ldots$	$f(x)=\int_0^\infty J_{n/2-1}(2\pi rx)G(r,x)\widetilde{f}(r)\,dr$
Weber transform	$F_a(u)=\int_a^\infty W_\nu(xu,au)xf(x)\,dx,$ $W_\nu(\beta,\mu)\equiv J_\nu(\beta)Y_\nu(\mu)-J_\nu(\mu)Y_\nu(\beta)$	$f(x)=\int_0^\infty \frac{W_\nu(xu,au)}{J_\nu^2(au)+Y_\nu^2(au)}uF_a(u)\,du$
Hardy transform	$F(u)=\int_0^\infty C_\nu(xu)xf(x)\,dx,$ $C_\nu(z)\equiv\cos(\pi p)J_\nu(z)+\sin(\pi p)Y_\nu(z)$	$f(x)=\int_0^\infty \Phi(xu)uF(u)\,du$ $\Phi(z)=\sum_{n=0}^\infty \frac{(-1)^n(z/2)^{\nu+2p+2n}}{\Gamma(p+n+1)\Gamma(\nu+p+n+1)}$
Kontorovich–Lebedev transform	$F(\tau)=\int_0^\infty K_{i\tau}(x)f(x)\,dx$	$f(x)=\frac{2}{\pi^2x}\int_0^\infty \tau\sinh(\pi\tau)K_{i\tau}(x)F(\tau)\,d\tau$
Meler–Fock transform	$\widetilde{F}(\tau)=\int_1^\infty P_{-\frac12+i\tau}(x)f(x)\,dx$	$f(x)=\int_0^\infty \tau\tanh(\pi\tau)P_{-\frac12+i\tau}(x)\widetilde{F}(\tau)\,d\tau$
Euler transform of the 1st kind*	$F(x)=\frac{1}{\Gamma(\mu)}\int_a^x\frac{f(t)\,dt}{(x-t)^{1-\mu}}$ $0<\mu<1,\quad x>a$	$f(x)=\frac{1}{\Gamma(1-\mu)}\frac{d}{dx}\int_a^x\frac{F(t)\,dt}{(x-t)^\mu}$

TABLE 11.5 (*continued*)
Summary table of integral transforms

Integral transform	Definition	Inversion formula
Euler transform of the 2nd kind*	$F(x)=\dfrac{1}{\Gamma(\mu)}\displaystyle\int_x^a \frac{f(t)\,dt}{(t-x)^{1-\mu}}$ $0<\mu<1,\quad x<a$	$f(x)=-\dfrac{1}{\Gamma(1-\mu)}\dfrac{d}{dx}\displaystyle\int_x^a \frac{F(t)\,dt}{(t-x)^{\mu}}$
Gauss transform**	$F(x)=\dfrac{1}{\sqrt{\pi a}}\displaystyle\int_{-\infty}^{\infty}\exp\left[-\frac{(x-t)^2}{a}\right]f(t)\,dt$	$f(x)=\exp\left(-\dfrac{a}{4}\dfrac{d^2}{dx^2}\right)F(x)$
Hilbert transform***	$\widehat{F}(s)=\dfrac{1}{\pi}\displaystyle\int_{-\infty}^{\infty}\frac{f(x)}{x-s}\,dx$	$f(x)=-\dfrac{1}{\pi}\displaystyle\int_{-\infty}^{\infty}\frac{\widehat{F}(s)}{s-x}\,ds$

NOTATION: $i=\sqrt{-1}$, $J_\mu(x)$ and $Y_\mu(x)$ are the Bessel functions of the first and the second kind, respectively; $I_\mu(x)$ and $K_\mu(x)$ are the modified Bessel functions of the first and the second kind, respectively; $P_\mu(x)$ is the Legendre spherical function of the second kind; and $\mathbf{H}_\mu(x)$ is the Struve function (see Subsection 11.5.1-4).

REMARKS.

* The Euler transform of the first kind is also known as Riemann–Liouville integral (the left fractional integral of order μ or, for short, the fractional integral). The Euler transform of the second kind is also called the right fractional integral of order μ.

** If $a=4$, the Gauss transform is called the Weierstrass transform. In the inversion formula, the exponential is represented by an operator series: $\exp\left(k\frac{d^2}{dx^2}\right)\equiv 1+\sum\limits_{n=1}^{\infty}\frac{k^n}{n!}\frac{d^{2n}}{dx^{2n}}$.

*** In the direct and inverse Hilbert transforms, the integrals are understood in the sense of the Cauchy principal value.

Table 11.6 presents various areas of application of integral transforms with literature references.

Example.

Consider the Cauchy problem for the integro-differential equation

$$\frac{dy}{dx}+\int_0^x K(x-t)y(t)\,dt=f(x)\qquad(0\le t<\infty)\tag{11.5.2.1}$$

with the initial condition

$$y=a\quad\text{at}\quad t=0.\tag{11.5.2.2}$$

Multiply equation (11.5.2.1) by e^{-px} and then integrate with respect to x from zero to infinity. Using properties 7 and 12 of the Laplace transform (Table 11.1) and taking into account the initial condition (11.5.2.2), we obtain a linear algebraic equation for the transform $\widetilde{y}(p)$:

$$p\widetilde{y}(p)-a+\widetilde{K}(p)\widetilde{y}(p)=\widetilde{f}(p).$$

It follows that

$$\widetilde{y}(p)=\frac{\widetilde{f}(p)+a}{p+\widetilde{K}(p)}.$$

By the inversion formula (11.2.1.2), the solution to the original problem (11.5.2.1)–(11.5.2.2) is found in the form

$$y(x)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{\widetilde{f}(p)+a}{p+\widetilde{K}(p)}e^{px}\,dp,\qquad i^2=-1.\tag{11.5.2.3}$$

Consider the special case of $a=0$ and $K(x)=\cos(bx)$. From row 10 of Table 11.2 it follows that $\widetilde{K}(p)=\dfrac{p}{p^2+b^2}$. Rearrange the integrand of (11.5.2.3):

$$\frac{\widetilde{f}(p)}{p+\widetilde{K}(p)}=\frac{p^2+b^2}{p(p^2+b^2+1)}\widetilde{f}(p)=\left(\frac{1}{p}-\frac{1}{p(p^2+b^2+1)}\right)\widetilde{f}(p).$$

In order to invert this expression, let us use the convolution theorem (see formula 16 of Subsection T3.2.1) as well as formulas 1 and 28 for the inversion of rational functions, Subsection T3.2.1. As a result, we arrive at the solution in the form

$$y(x)=\int_0^x\frac{b^2+\cos\left(t\sqrt{b^2+1}\right)}{b^2+1}f(x-t)\,dt.$$

TABLE 11.6
Areas of application of integral transforms (first in the last column come references to appropriate sections of the current book)

Area of application	Integral transforms	References
Evaluation of improper integrals	Laplace, Mellin	Paragraph 7.2.8-5; Ditkin & Prudnikov (1965)
Summation of series	Laplace (direct and inverse)	Paragraphs 8.1.5-2 and 8.4.4-1
Computation of coefficients of asymptotic expansions	Laplace, Mellin	Paragraph 7.2.9-1
Linear constant- and variable-coefficient ordinary differential equations	Laplace, Mellin, Euler, and others	Paragraphs 12.4.1-3 and 12.4.2-6; Ditkin & Prudnikov (1965); Doetsch (1974); E. Kamke (1977); Sveshnikov & Tikhonov (1970); LePage (1980)
Systems of linear constant-coefficient ordinary differential equations	Laplace	Paragraph 12.6.1-4; Doetsch (1974); Ditkin & Prudnikov (1965)
Linear equations of mathematical physics	Laplace, Fourier, Fourier sine, Hankel, Kontorovich–Lebedev, and others	Section 14.5; Doetsch (1974); Ditkin & Prudnikov (1965); Antimirov (1993); Sneddon (1995); Zwillinger (1997); Bracewell (1999); Polyanin (2002); Duffy (2004)
Linear integral equations	Laplace, Mellin, Fourier, Meler–Fock, Euler, and others	Subsections 16.1.3, 16.2.3, 16.3.2, 16.4.6; Krasnov, Kiselev, & Makarenko (1971); Ditkin & Prudnikov (1965); Samko, Kilbas, & Marichev (1993); Polyanin & Manzhirov (1998)
Nonlinear integral equations	Laplace, Mellin, Fourier	Paragraphs 16.5.2-1 and 16.5.3-2; Krasnov, Kiselev, & Makarenko (1971); Polyanin & Manzhirov (1998)
Linear difference equations	Laplace	Ditkin & Prudnikov (1965)
Linear differential-difference equations	Laplace	Bellman & Roth (1984)
Linear integro-differential equations	Laplace, Fourier	Paragraph 11.5.2-2; LePage (1980)

References for Chapter 11

Antimirov, M. Ya., *Applied Integral Transforms*, American Mathematical Society, Providence, Rhode Island, 1993.

Bateman, H. and Erdélyi, A., *Tables of Integral Transforms. Vols. 1 and 2*, McGraw-Hill, New York, 1954.

Beerends, R. J., ter Morschem, H. G., and van den Berg, J. C., *Fourier and Laplace Transforms*, Cambridge University Press, Cambridge, 2003.

Bellman, R. and Roth, R., *The Laplace Transform*, World Scientific Publishing Co., Singapore, 1984.

Bracewell, R., *The Fourier Transform and Its Applications, 3rd Edition*, McGraw-Hill, New York, 1999.

Brychkov, Yu. A. and Prudnikov, A. P., *Integral Transforms of Generalized Functions*, Gordon & Breach, New York, 1989.

Davis, B., *Integral Transforms and Their Applications*, Springer-Verlag, New York, 1978.

Ditkin, V. A. and Prudnikov, A. P., *Integral Transforms and Operational Calculus*, Pergamon Press, New York, 1965.

Doetsch, G., *Handbuch der Laplace-Transformation. Theorie der Laplace-Transformation*, Birkhäuser, Basel–Stuttgart, 1950.

Doetsch, G., *Handbuch der Laplace-Transformation. Anwendungen der Laplace-Transformation*, Birkhäuser, Basel–Stuttgart, 1956.

Doetsch, G., *Introduction to the Theory and Application of the Laplace Transformation*, Springer-Verlag, Berlin, 1974.

Duffy, D. G., *Transform Methods for Solving Partial Differential Equations, 2nd Edition*, Chapman & Hall/CRC Press, Boca Raton, 2004.

Hirschman, I. I. and Widder, D. V., *The Convolution Transform*, Princeton University Press, Princeton, New Jersey, 1955.
Kamke, E., *Differentialgleichungen: Lösungsmethoden und Lösungen, I, Gewöhnliche Differentialgleichungen*, B. G. Teubner, Leipzig, 1977.
Krantz, S. G., *Handbook of Complex Variables*, Birkhäuser, Boston, 1999.
Krasnov, M. L., Kiselev, A. I., and Makarenko, G. I., *Problems and Exercises in Integral Equations*, Mir Publishers, Moscow, 1971.
LePage, W. R., *Complex Variables and the Laplace Transform for Engineers*, Dover Publications, New York, 1980.
Miles, J. W., *Integral Transforms in Applied Mathematics*, Cambridge University Press, Cambridge, 1971.
Oberhettinger, F., *Tables of Bessel Transforms*, Springer-Verlag, New York, 1972.
Oberhettinger, F., *Tables of Fourier Transforms and Fourier Transforms of Distributions*, Springer-Verlag, Berlin, 1980.
Oberhettinger, F., *Tables of Mellin Transforms*, Springer-Verlag, New York, 1974.
Oberhettinger, F. and Badii, L., *Tables of Laplace Transforms*, Springer-Verlag, New York, 1973.
Pinkus, A. and Zafrany, S., *Fourier Series and Integral Transforms*, Cambridge University Press, Cambridge, 1997.
Polyanin, A. D., *Handbook of Linear Partial Differential Equations for Engineers and Scientists*, Chapman & Hall/CRC Press, Boca Raton, 2002.
Polyanin, A. D. and Manzhirov, A. V., *Handbook of Integral Equations*, CRC Press, Boca Raton, 1998.
Poularikas, A. D., *The Transforms and Applications Handbook, 2nd Edition*, CRC Press, Boca Raton, 2000.
Prudnikov, A. P., Brychkov, Yu. A., and Marichev, O. I., *Integrals and Series, Vol. 4, Direct Laplace Transform*, Gordon & Breach, New York, 1992.
Prudnikov, A. P., Brychkov, Yu. A., and Marichev, O. I., *Integrals and Series, Vol. 5, Inverse Laplace Transform*, Gordon & Breach, New York, 1992.
Samko, S. G., Kilbas, A. A., and Marichev, O. I., *Fractional Integrals and Derivatives. Theory and Applications*, Gordon & Breach, New York, 1993.
Sneddon, I., *Fourier Transforms*, Dover Publications, New York, 1995.
Sneddon, I., *The Use of Integral Transforms*, McGraw-Hill, New York, 1972.
Sveshnikov, A. G. and Tikhonov, A. N., *Theory of Functions of a Complex Variable* [in Russian], Nauka Publishers, Moscow, 1970.
Titchmarsh, E. C., *Introduction to the Theory of Fourier Integrals, 3rd Edition*, Chelsea Publishing, New York, 1986.
Zwillinger, D., *CRC Standard Mathematical Tables and Formulae, 31st Edition*, CRC Press, Boca Raton, 2002.
Zwillinger, D., *Handbook of Differential Equations, 3rd Edition*, Academic Press, Boston, 1997.

Chapter 12
Ordinary Differential Equations

12.1. First-Order Differential Equations

12.1.1. General Concepts. The Cauchy Problem. Uniqueness and Existence Theorems

12.1.1-1. Equations solved for the derivative. General solution.

A *first-order ordinary differential equation** solved for the derivative has the form

$$y'_x = f(x, y). \tag{12.1.1.1}$$

Sometimes it is represented in terms of differentials as $dy = f(x, y)\,dx$.

A *solution of a differential equation* is a function $y(x)$ that, when substituted into the equation, turns it into an identity. The *general solution of a differential equation* is the set of all its solutions. In some cases, the general solution can be represented as a function $y = \varphi(x, C)$ that depends on one *arbitrary constant* C; specific values of C define specific solutions of the equation (*particular solutions*). In practice, the general solution more frequently appears in implicit form, $\Phi(x, y, C) = 0$, or parametric form, $x = x(t, C)$, $y = y(t, C)$.

Geometrically, the general solution (also called the general integral) of an equation is a family of curves in the xy-plane depending on a single parameter C; these curves are called *integral curves* of the equation. To each particular solution (particular integral) there corresponds a single curve that passes through a given point in the plane.

For each point (x, y), the equation $y'_x = f(x, y)$ defines a value of y'_x, i.e., the slope of the integral curve that passes through this point. In other words, the equation generates a field of directions in the xy-plane. From the geometrical point of view, the problem of solving a first-order differential equation involves finding the curves, the slopes of which at each point coincide with the direction of the field at this point.

Figure 12.1 depicts the tangent to an integral curve at a point (x_0, y_0); the slope of the integral curve at this point is determined by the right-hand side of equation (12.1.1.1): $\tan\alpha = f(x_0, y_0)$. The little lines show the field of tangents to the integral curves of the differential equation (12.1.1.1) at other points.

12.1.1-2. Equations integrable by quadrature.

To integrate a differential equation in closed form is to represent its solution in the form of formulas written using a predefined bounded set of allowed functions and mathematical operations. A solution is expressed as a quadrature if the set of allowed functions consists of the elementary functions and the functions appearing in the equation and the allowed

* In what follows, we often call an ordinary differential equation a "differential equation" or, even shorter, an "equation."

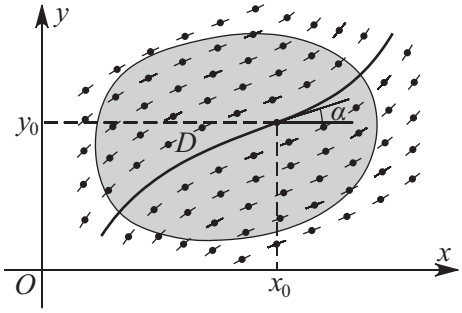


Figure 12.1. The direction field of a differential equation and the integral curve passing through a point (x_0, y_0).

mathematical operations are the arithmetic operations, a finite number of function compositions, and the indefinite integral. An equation is said to be integrable by quadrature if its general solution can be expressed in terms of quadratures.

12.1.1-3. Cauchy problem. The uniqueness and existence theorems.

The *Cauchy problem*: find a solution of equation (12.1.1.1) that satisfies the *initial condition*

$$y = y_0 \quad \text{at} \quad x = x_0, \tag{12.1.1.2}$$

where y_0 and x_0 are some numbers.

Geometrical meaning of the Cauchy problem: find an integral curve of equation (12.1.1.1) that passes through the point (x_0, y_0); see Fig. 12.1.

Condition (12.1.1.2) is alternatively written $y(x_0) = y_0$ or $y|_{x=x_0} = y_0$.

THEOREM (EXISTENCE, PEANO). *Let the function $f(x, y)$ be continuous in an open domain D of the xy-plane. Then there is at least one integral curve of equation (12.1.1.1) that passes through each point $(x_0, y_0) \in D$; each of these curves can be extended at both ends up to the boundary of any closed domain $D_0 \subset D$ such that (x_0, y_0) belongs to the interior of D_0.*

THEOREM (UNIQUENESS). *Let the function $f(x, y)$ be continuous in an open domain D and have in D a bounded partial derivative with respect to y (or the Lipschitz condition holds: $|f(x, y) - f(x, z)| \le M|y - z|$, where M is some positive number). Then there is a unique solution of equation (12.1.1.1) satisfying condition (12.1.1.2).*

12.1.1-4. Equations not solved for the derivative. The existence theorem.

A first-order differential equation not solved for the derivative can generally be written as

$$F(x, y, y'_x) = 0. \tag{12.1.1.3}$$

THEOREM (EXISTENCE AND UNIQUENESS). *There exists a unique solution $y = y(x)$ of equation (12.1.1.3) satisfying the conditions $y|_{x=x_0} = y_0$ and $y'_x|_{x=x_0} = t_0$, where t_0 is one of the real roots of the equation $F(x_0, y_0, t_0) = 0$ if the following conditions hold in a neighborhood of the point (x_0, y_0, t_0):*

1. *The function $F(x, y, t)$ is continuous in each of the three arguments.*
2. *The partial derivative F_t exists and is nonzero.*
3. *There is a bounded partial derivative with respect to y, $|F_y| \le M$.*

The solution exists for $|x - x_0| \le a$, where a is a (sufficiently small) positive number.

12.1.1-5. Singular solutions.

1°. A point (x, y) at which the uniqueness of the solution to equation (12.1.1.3) is violated is called a *singular* point. If conditions 1 and 3 of the existence and uniqueness theorem hold, then

$$F(x, y, t) = 0, \qquad F_t(x, y, t) = 0 \tag{12.1.1.4}$$

simultaneously at each singular point. Relations (12.1.1.4) define a *t-discriminant curve* in parametric form. In some cases, the parameter t can be eliminated from (12.1.1.4) to give an equation of this curve in implicit form, $\Psi(x, y) = 0$. If a branch $y = \psi(x)$ of the curve $\Psi(x, y) = 0$ consists of singular points and, at the same time, is an integral curve, then this branch is called a *singular integral curve* and the function $y = \psi(x)$ is a *singular solution* of equation (12.1.1.3).

2°. The singular solutions can be found by identifying the *envelope of the family of integral curves*, $\Phi(x, y, C) = 0$, of equation (12.1.1.3). The envelope is part of the *C-discriminant curve*, which is defined by the equations

$$\Phi(x, y, C) = 0, \qquad \Phi_C(x, y, C) = 0.$$

The branch of the C-discriminant curve at which

(a) there exist bounded partial derivatives, $|\Phi_x| < M_1$ and $|\Phi_y| < M_2$, and
(b) $|\Phi_x| + |\Phi_y| \neq 0$

is the envelope.

12.1.1-6. Point transformations.

In the general case, a point transformation is defined by

$$x = F(X, Y), \qquad y = G(X, Y), \tag{12.1.1.5}$$

where X is the new independent variable, $Y = Y(X)$ is the new dependent variable, and F and G are some (prescribed or unknown) functions.

The derivative y'_x under the point transformation (12.1.1.5) is calculated by

$$y'_x = \frac{G_X + G_Y Y'_X}{F_X + F_Y Y'_X},$$

where the subscripts X and Y denote the corresponding partial derivatives.

Transformation (12.1.1.5) is invertible if $F_X G_Y - F_Y G_X \neq 0$.

Point transformations are used to simplify equations and reduce them to known equations. Sometimes a point transformation allows the reduction of a nonlinear equation to a linear one.

Example. The *hodograph transformation is* an important example of a point transformation. It is defined by $x = Y$, $y = X$, which means that y is taken to be the independent variable and x the dependent one. In this case, the derivative is expressed as

$$y'_x = \frac{1}{X'_Y}.$$

Other examples of point transformations can be found in Subsections 12.1.2 and 12.1.4–12.1.6.

12.1.2. Equations Solved for the Derivative. Simplest Techniques of Integration

12.1.2-1. Equations with separated or separable variables.

1°. An *equation with separated variables* (a *separated equation*) has the form

$$f(y)y'_x = g(x).$$

Equivalently, the equation can be rewritten as $f(y)\,dy = g(x)\,dx$ (the right-hand side depends on x alone and the left-hand side on y alone). The general solution can be obtained by termwise integration:

$$\int f(y)\,dy = \int g(x)\,dx + C,$$

where C is an arbitrary constant.

2°. An *equation with separable variables* (a *separable equation*) is generally represented by

$$f_1(y)g_1(x)y'_x = f_2(y)g_2(x).$$

Dividing the equation by $f_2(y)g_1(x)$, one obtains a separated equation. Integrating yields

$$\int \frac{f_1(y)}{f_2(y)}\,dy = \int \frac{g_2(x)}{g_1(x)}\,dx + C.$$

Remark. In termwise division of the equation by $f_2(y)g_1(x)$, solutions corresponding to $f_2(y) = 0$ can be lost.

12.1.2-2. Equation of the form $y'_x = f(ax + by)$.

The substitution $z = ax + by$ brings the equation to a separable equation, $z'_x = bf(z) + a$; see Paragraph 12.1.2-1.

12.1.2-3. Homogeneous equations $y'_x = f(y/x)$.

1°. A *homogeneous equation* remains the same under simultaneous scaling (dilatation) of the independent and dependent variables in accordance with the rule $x \to \alpha x$, $y \to \alpha y$, where α is an arbitrary constant ($\alpha \neq 0$). Such equations can be represented in the form

$$y'_x = f\Big(\frac{y}{x}\Big).$$

The substitution $u = y/x$ brings a homogeneous equation to a separable one, $xu'_x = f(u) - u$; see Paragraph 12.1.2-1.

2°. The equations of the form

$$y'_x = f\Big(\frac{a_1x + b_1y + c_1}{a_2x + b_2y + c_2}\Big)$$

can be reduced to a homogeneous equation. To this end, for $a_1x + b_1y \neq k(a_2x + b_2y)$, one should use the change of variables $\xi = x - x_0$, $\eta = y - y_0$, where the constants x_0 and y_0 are determined by solving the linear algebraic system

$$\begin{aligned} a_1x_0 + b_1y_0 + c_1 &= 0,\\ a_2x_0 + b_2y_0 + c_2 &= 0. \end{aligned}$$

As a result, one arrives at the following equation for $\eta = \eta(\xi)$:

$$\eta'_\xi = f\left(\frac{a_1\xi + b_1\eta}{a_2\xi + b_2\eta}\right).$$

On dividing the numerator and denominator of the argument of f by ξ, one obtains a homogeneous equation whose right-hand side is dependent on the ratio η/ξ only:

$$\eta'_\xi = f\left(\frac{a_1 + b_1\eta/\xi}{a_2 + b_2\eta/\xi}\right).$$

For $a_1x + b_1y = k(a_2x + b_2y)$, see the equation of Paragraph 12.1.2-2.

12.1.2-4. Generalized homogeneous equations and equations reducible to them.

1°. A *generalized homogeneous equation* (a homogeneous equation in the generalized sense) remains the same under simultaneous scaling of the independent and dependent variables in accordance with the rule $x \to \alpha x$, $y \to \alpha^k y$, where $\alpha \neq 0$ is an arbitrary constant and k is some number. Such equations can be represented in the form

$$y'_x = x^{k-1} f(yx^{-k}).$$

The substitution $u = yx^{-k}$ brings a generalized homogeneous equation to a separable equation, $xu'_x = f(u) - ku$; see Paragraph 12.1.2-1.

Example. Consider the equation

$$y'_x = ax^2y^4 + by^2. \tag{12.1.2.1}$$

Let us perform the transformation $x = \alpha\bar{x}$, $y = \alpha^k\bar{y}$ and then multiply the resulting equation by α^{1-k} to obtain

$$\bar{y}'_{\bar{x}} = a\alpha^{3(k+1)}\bar{x}^2\bar{y}^4 + b\alpha^{k+1}\bar{y}^2. \tag{12.1.2.2}$$

It is apparent that if $k = -1$, the transformed equation (12.1.2.2) is the same as the original one, up to notation. This means that equation (12.1.2.1) is generalized homogeneous of degree $k = -1$. Therefore the substitution $u = xy$ brings it to a separable equation: $xu'_x = au^4 + bu^2 + u$.

2°. The equations of the form

$$y'_x = yf(e^{\lambda x}y)$$

can be reduced to a generalized homogeneous equation. To this end, one should use the change of variable $z = e^x$ and set $\lambda = -k$.

12.1.2-5. Linear equation $y'_x + f(x)y = g(x)$.

A first-order *linear equation* is written as

$$y'_x + f(x)y = g(x).$$

The solution is sought in the product form $y = uv$, where $v = v(x)$ is any function that satisfies the "truncated" equation $v'_x + f(x)v = 0$ [as $v(x)$ one takes the particular solution $v = e^{-F}$, where $F = \int f(x)\,dx$]. As a result, one obtains the following separable equation for $u = u(x)$: $v(x)u'_x = g(x)$. Integrating it yields the general solution:

$$y(x) = e^{-F}\left(\int e^F g(x)\,dx + C\right), \qquad F = \int f(x)\,dx,$$

where C is an arbitrary constant.

12.1.2-6. Bernoulli equation $y'_x + f(x)y = g(x)y^a$.

A *Bernoulli equation* has the form

$$y'_x + f(x)y = g(x)y^a, \qquad a \neq 0,\ 1.$$

(For $a = 0$ and $a = 1$, it is a linear equation; see Paragraph 12.1.2-5.) The substitution $z = y^{1-a}$ brings it to a linear equation, $z'_x + (1-a)f(x)z = (1-a)g(x)$, which is discussed in Paragraph 12.1.2-5. With this in view, one can obtain the general integral:

$$y^{1-a} = Ce^{-F} + (1-a)e^{-F}\int e^F g(x)\,dx, \qquad \text{where} \quad F = (1-a)\int f(x)\,dx.$$

12.1.2-7. Equation of the form $xy'_x = y + f(x)g(y/x)$.

The substitution $u = y/x$ brings the equation to a separable equation, $x^2u'_x = f(x)g(u)$; see Paragraph 12.1.2-1.

12.1.2-8. Darboux equation.

A *Darboux equation* can be represented as

$$\left[f\left(\frac{y}{x}\right) + x^a h\left(\frac{y}{x}\right)\right]y'_x = g\left(\frac{y}{x}\right) + yx^{a-1}h\left(\frac{y}{x}\right).$$

Using the substitution $y = xz(x)$ and taking z to be the independent variable, one obtains a Bernoulli equation, which is considered in Paragraph 12.1.2-6:

$$\big[g(z) - zf(z)\big]x'_z = xf(z) + x^{a+1}h(z).$$

► *Some other first-order equations integrable by quadrature are treated in Section T5.1.*

12.1.3. Exact Differential Equations. Integrating Factor

12.1.3-1. Exact differential equations.

An *exact differential equation* has the form

$$f(x,y)\,dx + g(x,y)\,dy = 0, \quad \text{where} \quad \frac{\partial f}{\partial y} = \frac{\partial g}{\partial x}.$$

The left-hand side of the equation is the total differential of a function of two variables $U(x,y)$.

The general integral, $U(x,y) = C$, where C is an arbitrary constant and the function U is determined from the system:

$$\frac{\partial U}{\partial x} = f, \quad \frac{\partial U}{\partial y} = g.$$

Integrating the first equation yields $U = \int f(x,y)\,dx + \Psi(y)$ (while integrating, the variable y is treated as a parameter). On substituting this expression into the second equation, one identifies the function Ψ (and hence, U). As a result, the general integral of an exact differential equation can be represented in the form

$$\int_{x_0}^{x} f(\xi,y)\,d\xi + \int_{y_0}^{y} g(x_0,\eta)\,d\eta = C,$$

where x_0 and y_0 are any numbers.

TABLE 12.1
An integrating factor $\mu = \mu(x, y)$ for some types of ordinary differential equations $f\,dx + g\,dy = 0$, where $f = f(x, y)$ and $g = g(x, y)$. The subscripts x and y indicate the corresponding partial derivatives

No.	Conditions for f and g	Integrating factor	Remarks
1	$f = y\varphi(xy),\ g = x\psi(xy)$	$\mu = \frac{1}{xf-yg}$	$xf - yg \neq 0$; $\varphi(z)$ and $\psi(z)$ are any functions
2	$f_x = g_y,\ f_y = -g_x$	$\mu = \frac{1}{f^2+g^2}$	$f + ig$ is an analytic function of the complex variable $x + iy$
3	$\frac{f_y-g_x}{g} = \varphi(x)$	$\mu = \exp\left[\int \varphi(x)\,dx\right]$	$\varphi(x)$ is any function
4	$\frac{f_y-g_x}{f} = \varphi(y)$	$\mu = \exp\left[-\int \varphi(y)\,dy\right]$	$\varphi(y)$ is any function
5	$\frac{f_y-g_x}{g-f} = \varphi(x+y)$	$\mu = \exp\left[\int \varphi(z)\,dz\right],\ z = x + y$	$\varphi(z)$ is any function
6	$\frac{f_y-g_x}{yg-xf} = \varphi(xy)$	$\mu = \exp\left[\int \varphi(z)\,dz\right],\ z = xy$	$\varphi(z)$ is any function
7	$\frac{x^2(f_y-g_x)}{yg+xf} = \varphi\left(\frac{y}{x}\right)$	$\mu = \exp\left[-\int \varphi(z)\,dz\right],\ z = \frac{y}{x}$	$\varphi(z)$ is any function
8	$\frac{f_y-g_x}{xg-yf} = \varphi(x^2+y^2)$	$\mu = \exp\left[\frac{1}{2}\int \varphi(z)\,dz\right],\ z = x^2+y^2$	$\varphi(z)$ is any function
9	$f_y - g_x = \varphi(x)g - \psi(y)f$	$\mu = \exp\left[\int \varphi(x)\,dx + \int \psi(y)\,dy\right]$	$\varphi(x)$ and $\psi(y)$ are any functions
10	$\frac{f_y-g_x}{g\omega_x-f\omega_y} = \varphi(\omega)$	$\mu = \exp\left[\int \varphi(\omega)\,d\omega\right]$	$\omega = \omega(x, y)$ is any function of two variables

Example. Consider the equation

$$(ay^n + bx)y'_x + by + cx^m = 0, \quad \text{or} \quad (by + cx^m)\,dx + (ay^n + bx)\,dy = 0,$$

defined by the functions $f(x, y) = by + cx^m$ and $g(x, y) = ay^n + bx$. Computing the derivatives, we have

$$\frac{\partial f}{\partial y} = b, \quad \frac{\partial g}{\partial x} = b \quad \Longrightarrow \quad \frac{\partial f}{\partial y} = \frac{\partial g}{\partial x}.$$

Hence the given equation is an exact differential equation. Its solution can be found using the last formula from Paragraph 12.1.3-1 with $x_0 = y_0 = 0$:

$$\frac{a}{n+1}y^{n+1} + bxy + \frac{c}{m+1}x^{m+1} = C.$$

12.1.3-2. Integrating factor.

An *integrating factor* for the equation

$$f(x, y)\,dx + g(x, y)\,dy = 0$$

is a function $\mu(x, y) \not\equiv 0$ such that the left-hand side of the equation, when multiplied by $\mu(x, y)$, becomes a total differential, and the equation itself becomes an exact differential equation.

An integrating factor satisfies the first-order partial differential equation,

$$g\frac{\partial \mu}{\partial x} - f\frac{\partial \mu}{\partial y} = \left(\frac{\partial f}{\partial y} - \frac{\partial g}{\partial x}\right)\mu,$$

which is not generally easier to solve than the original equation.

Table 12.1 lists some special cases where an integrating factor can be found in explicit form.

12.1.4. Riccati Equation

12.1.4-1. General Riccati equation. Simplest integrable cases.

A *Riccati equation* has the general form

$$y'_x = f_2(x)y^2 + f_1(x)y + f_0(x). \tag{12.1.4.1}$$

If $f_2 \equiv 0$, we have a linear equation (see Paragraph 12.1.2-5), and if $f_0 \equiv 0$, we have a Bernoulli equation (see Paragraph 12.1.2-6 for $a=2$), whose solutions were given previously. For arbitrary f_2, f_1, and f_0, the Riccati equation is not integrable by quadrature.

Listed below are some special cases where the Riccati equation (12.1.4.1) is integrable by quadrature.

1°. The functions f_2, f_1, and f_0 are proportional, i.e.,

$$y'_x = \varphi(x)(ay^2 + by + c),$$

where a, b, and c are constants. This equation is a separable equation; see Paragraph 12.1.2-1.

2°. The Riccati equation is homogeneous:

$$y'_x = a\frac{y^2}{x^2} + b\frac{y}{x} + c.$$

See Paragraph 12.1.2-3.

3°. The Riccati equation is generalized homogeneous:

$$y'_x = ax^n y^2 + \frac{b}{x}y + cx^{-n-2}.$$

See Paragraph 12.1.2-4 (with $k=-n-1$). The substitution $z = x^{n+1}y$ brings it to a separable equation: $xz'_x = az^2 + (b+n+1)z + c$.

4°. The Riccati equation has the form

$$y'_x = ax^{2n}y^2 + \frac{m-n}{x}y + cx^{2m}.$$

By the substitution $y = x^{m-n}z$, the equation is reduced to a separable equation: $x^{-n-m}z'_x = az^2 + c$.

► *Some other Riccati equations integrable by quadrature are treated in Section T5.1 (see equations T5.1.6 to T5.1.22).*

12.1.4-2. Polynomial solutions of the Riccati equation.

Let $f_2 = 1$, $f_1(x)$, and $f_0(x)$ be polynomials. If the degree of the polynomial

$$\Delta = f_1^2 - 2(f_1)'_x - 4f_0$$

is odd, the Riccati equation cannot possess a polynomial solution. If the degree of Δ is even, the equation involved may possess only the following polynomial solutions:

$$y = -\tfrac{1}{2}\big(f_1 \pm \big[\sqrt{\Delta}\,\big]\big),$$

where $\big[\sqrt{\Delta}\,\big]$ denotes an integer rational part of the expansion of $\sqrt{\Delta}$ in decreasing powers of x (for example, $\big[\sqrt{x^2-2x+3}\,\big] = x-1$).

12.1.4-3. Use of particular solutions to construct the general solution.

1°. Given a particular solution $y_0 = y_0(x)$ of the Riccati equation (12.1.4.1), the general solution can be written as

$$y = y_0(x) + \Phi(x)\Big[C - \int \Phi(x) f_2(x)\,dx\Big]^{-1}, \tag{12.1.4.2}$$

where C is an arbitrary constant and

$$\Phi(x) = \exp\Big\{\int \big[2f_2(x)y_0(x) + f_1(x)\big]\,dx\Big\}. \tag{12.1.4.3}$$

To the particular solution $y_0(x)$ there corresponds $C = \infty$.

2°. Let $y_1 = y_1(x)$ and $y_2 = y_2(x)$ be two different particular solutions of equation (12.1.4.1). Then the general solution can be calculated by

$$y = \frac{Cy_1 + U(x)y_2}{C + U(x)}, \quad \text{where} \quad U(x) = \exp\Big[\int f_2(y_1 - y_2)\,dx\Big].$$

To the particular solution $y_1(x)$, there corresponds $C = \infty$; and to $y_2(x)$, there corresponds $C = 0$.

3°. Let $y_1 = y_1(x)$, $y_2 = y_2(x)$, and $y_3 = y_3(x)$ be three distinct particular solutions of equation (12.1.4.1). Then the general solution can be found without quadrature:

$$\frac{y - y_2}{y - y_1}\,\frac{y_3 - y_1}{y_3 - y_2} = C.$$

This means that the Riccati equation has a fundamental system of solutions.

12.1.4-4. Some transformations.

1°. The transformation (φ, ψ_1, ψ_2, ψ_3, and ψ_4 are arbitrary functions)

$$x = \varphi(\xi), \quad y = \frac{\psi_4(\xi)u + \psi_3(\xi)}{\psi_2(\xi)u + \psi_1(\xi)}$$

reduces the Riccati equation (12.1.4.1) to a Riccati equation for $u = u(\xi)$.

2°. Let $y_0 = y_0(x)$ be a particular solution of equation (12.1.4.1). Then the substitution $y = y_0 + 1/w$ leads to a linear equation for $w = w(x)$:

$$w'_x + \big[2f_2(x)y_0(x) + f_1(x)\big]w + f_2(x) = 0.$$

For solution of linear equations, see Paragraph 12.1.2-5.

12.1.4-5. Reduction of the Riccati equation to a second-order linear equation.

The substitution

$$u(x) = \exp\Big(-\int f_2 y\,dx\Big)$$

reduces the general Riccati equation (12.1.4.1) to a second-order linear equation:

$$f_2 u''_{xx} - \big[(f_2)'_x + f_1 f_2\big]u'_x + f_0 f_2^2 u = 0,$$

which often may be easier to solve than the original Riccati equation.

12.1.4-6. Reduction of the Riccati equation to the canonical form.

The general Riccati equation (12.1.4.1) can be reduced with the aid of the transformation

$$x = \varphi(\xi), \quad y = \frac{1}{F_2}w - \frac{1}{2}\frac{F_1}{F_2} + \frac{1}{2}\left(\frac{1}{F_2}\right)'_\xi, \quad \text{where} \quad F_i(\xi) = f_i(\varphi)\varphi'_\xi, \qquad (12.1.4.4)$$

to the canonical form

$$w'_\xi = w^2 + \Psi(\xi). \qquad (12.1.4.5)$$

Here, the function Ψ is defined by the formula

$$\Psi(\xi) = F_0F_2 - \frac{1}{4}F_1^2 + \frac{1}{2}F_1' - \frac{1}{2}F_1\frac{F_2'}{F_2} - \frac{3}{4}\left(\frac{F_2'}{F_2}\right)^2 + \frac{1}{2}\frac{F_2''}{F_2};$$

the prime denotes differentiation with respect to ξ.

Transformation (12.1.4.4) depends on a function $\varphi = \varphi(\xi)$ that can be arbitrary. For a specific original Riccati equation, different functions φ in (12.1.4.4) will generate different functions Ψ in equation (12.1.4.5). In practice, transformation (12.1.4.4) is most frequently used with $\varphi(\xi) = \xi$.

12.1.5. Abel Equations of the First Kind

12.1.5-1. General form of Abel equations of the first kind. Some integrable cases.

An *Abel equation of the first kind* has the general form

$$y'_x = f_3(x)y^3 + f_2(x)y^2 + f_1(x)y + f_0(x), \qquad f_3(x) \not\equiv 0. \qquad (12.1.5.1)$$

In the degenerate case $f_2(x) = f_0(x) = 0$, we have a Bernoulli equation (see Paragraph 12.1.2-6 with $a = 3$). The Abel equation (12.1.5.1) is not integrable in closed form for arbitrary $f_n(x)$.

Listed below are some special cases where the Abel equation of the first kind is integrable by quadrature.

1°. If the functions $f_n(x)$ $(n = 0, 1, 2, 3)$ are proportional, i.e., $f_n(x) = a_n g(x)$, then (12.1.5.1) is a separable equation (see Paragraph 12.1.2-1).

2°. The Abel equation is homogeneous:

$$y'_x = a\frac{y^3}{x^3} + b\frac{y^2}{x^2} + c\frac{y}{x} + d.$$

See Paragraph 12.1.2-3.

3°. The Abel equation is generalized homogeneous:

$$y'_x = ax^{2n+1}y^3 + bx^ny^2 + \frac{c}{x}y + dx^{-n-2}.$$

See Paragraph 12.1.2-4 for $k = -n - 1$. The substitution $w = x^{n+1}y$ leads to a separable equation: $xw'_x = aw^3 + bw^2 + (c + n + 1)w + d$.

4°. The Abel equation

$$y'_x = ax^{3n-m}y^3 + bx^{2n}y^2 + \frac{m-n}{x}y + dx^{2m}$$

can be reduced with the substitution $y = x^{m-n}z$ to a separable equation: $x^{-n-m}z'_x = az^3 + bz^2 + c$.

5°. Let $f_0 \equiv 0$, $f_1 \equiv 0$, and $(f_3/f_2)'_x = af_2$ for some constant a. Then the substitution $y = f_2 f_3^{-1}u$ leads to a separable equation: $u'_x = f_2^2 f_3^{-1}(u^3 + u^2 + au)$.

6°. If

$$f_0 = \frac{f_1 f_2}{3f_3} - \frac{2f_2^3}{27f_3^2} - \frac{1}{3}\frac{d}{dx}\frac{f_2}{f_3}, \qquad f_n = f_n(x),$$

then the solution of equation (12.1.5.1) is given by

$$y(x) = E\Big(C - 2\int f_3E^2\,dx\Big)^{-1/2} - \frac{f_2}{3f_3}, \qquad \text{where} \quad E = \exp\left[\int\left(f_1 - \frac{f_2^2}{3f_3}\right)dx\right].$$

For other solvable Abel equations of the first kind, see the books by Kamke (1977) and Polyanin and Zaitsev (2003).

12.1.5-2. Reduction of the Abel equation of the first kind to the canonical form.

The transformation

$$y = U(x)\eta(\xi) - \frac{f_2}{3f_3}, \quad \xi = \int f_3U^2\,dx, \qquad \text{where} \quad U(x) = \exp\left[\int\left(f_1 - \frac{f_2^2}{3f_3}\right)dx\right],$$

brings equation (12.1.5.1) to the canonical (normal) form

$$\eta'_\xi = \eta^3 + \Phi(\xi).$$

Here, the function $\Phi(\xi)$ is defined parametrically (x is the parameter) by the relations

$$\Phi = \frac{1}{f_3U^3}\left(f_0 - \frac{f_1f_2}{3f_3} + \frac{2f_2^3}{27f_3^2} + \frac{1}{3}\frac{d}{dx}\frac{f_2}{f_3}\right), \quad \xi = \int f_3U^2\,dx.$$

12.1.5-3. Reduction to an Abel equation of the second kind.

Let $y_0 = y_0(x)$ be a particular solution of equation (12.1.5.1). Then the substitution

$$y = y_0 + \frac{E(x)}{z(x)}, \qquad \text{where} \quad E(x) = \exp\Big[\int(3f_3y_0^2 + 2f_2y_0 + f_1)\,dx\Big],$$

leads to an Abel equation of the second kind:

$$zz'_x = -(3f_3y_0 + f_2)Ez - f_3E^2.$$

For equations of this type, see Subsection 12.1.6.

12.1.6. Abel Equations of the Second Kind

12.1.6-1. General form of Abel equations of the second kind. Some integrable cases.

An *Abel equation of the second kind* has the general form

$$[y + g(x)]y'_x = f_2(x)y^2 + f_1(x)y + f_0(x), \qquad g(x) \not\equiv 0. \tag{12.1.6.1}$$

The Abel equation (12.1.6.1) is not integrable for arbitrary $f_n(x)$ and $g(x)$. Given below are some special cases where the Abel equation of the second kind is integrable by quadrature.

1°. If $g(x) = \text{const}$ and the functions $f_n(x)$ $(n = 0, 1, 2)$ are proportional, i.e., $f_n(x) = a_n g(x)$, then (12.1.6.1) is a separable equation (see Paragraph 12.1.2-1).

2°. The Abel equation is homogeneous:

$$(y + sx)y'_x = \frac{a}{x}y^2 + by + cx.$$

See Paragraph 12.1.2-3. The substitution $w = y/x$ leads to a separable equation.

3°. The Abel equation is generalized homogeneous:

$$(y + sx^n)y'_x = \frac{a}{x}y^2 + bx^{n-1}y + cx^{2n-1}.$$

See Paragraph 12.1.2-4 for $k = n$. The substitution $w = yx^{-n}$ leads to a separable equation: $x(w + s)w'_x = (a - n)w^2 + (b - ns)w + c$.

4°. The Abel equation

$$(y + a_2x + c_2)y'_x = b_1y + a_1x + c_1$$

is a special case of the equation treated in Paragraph 12.1.2-3 (see Item 2° with $f(w) = w$ and $b_2 = 1$).

5°. The unnormalized Abel equation

$$[(a_1x + a_2x^n)y + b_1x + b_2x^n]y'_x = c_2y^2 + c_1y + c_0$$

can be reduced to the form (12.1.6.1) by dividing it by $(a_1x + a_2x^n)$. Taking y to be the independent variable and $x = x(y)$ to be the dependent one, we obtain the Bernoulli equation

$$(c_2y^2 + c_1y + c_0)x'_y = (a_1y + b_1)x + (a_2y + b_2)x^n.$$

See Paragraph 12.1.2-6.

6°. The general solution of the Abel equation

$$(y + g)y'_x = f_2y^2 + f_1y + f_1g - f_2g^2, \qquad f_n = f_n(x), \quad g = g(x),$$

is given by

$$y = -g + CE + E\int (f_1 + g'_x - 2f_2g)E^{-1}\,dx, \qquad \text{where} \quad E = \exp\Big(\int f_2\,dx\Big).$$

7°. If $f_1 = 2f_2g - g'_x$, the general solution of the Abel equation (12.1.6.1) has the form

$$y = -g \pm E\Big[2\int (f_0 + gg'_x - f_2g^2)E^{-2}\,dx + C\Big]^{1/2}, \qquad \text{where} \quad E = \exp\Big(\int f_2\,dx\Big).$$

For other solvable Abel equations of the second kind, see the books by Kamke (1977) and Polyanin and Zaitsev (2003).

12.1.6-2. Reduction of the Abel equation of the second kind to the canonical form.

1°. The substitution

$$w = (y + g)E, \qquad \text{where} \quad E = \exp\Big(-\int f_2\,dx\Big), \tag{12.1.6.2}$$

brings equation (12.1.6.1) to the simpler form

$$ww'_x = F_1(x)w + F_0(x), \tag{12.1.6.3}$$

where

$$F_1 = (f_1 - 2f_2g + g'_x)E, \quad F_0 = (f_0 - f_1g + f_2g^2)E^2.$$

2°. In turn, equation (12.1.6.3) can be reduced, by the introduction of the new independent variable

$$z = \int F_1(x)\,dx, \tag{12.1.6.4}$$

to the *canonical form*

$$ww'_z - w = R(z). \tag{12.1.6.5}$$

Here, the function $R(z)$ is defined parametrically (x is the parameter) by the relations

$$R = \frac{F_0(x)}{F_1(x)}, \qquad z = \int F_1(x)\,dx.$$

Substitutions (12.1.6.2) and (12.1.6.4), which take the Abel equation to the canonical form, are called *canonical*.

Remark 1. The transformation $w = a\hat{w}$, $z = a\hat{z} + b$ brings (12.1.6.5) to a similar equation, $\hat{w}\hat{w}'_{\hat{z}} - \hat{w} = a^{-1}R(a\hat{z} + b)$. Therefore the function $R(z)$ in the right-hand side of the Abel equation (12.1.6.5) can be identified with the two-parameter family of functions $a^{-1}R(az + b)$.

Remark 2. Any Abel equations of the second kind related by linear (in y) transformations $\widetilde{x} = \varphi_1(x)$, $\widetilde{y} = \varphi_2(x)y + \varphi_3(x)$ have identical canonical forms (up to the two-parameter family of functions specified in Remark 1).

12.1.6-3. Reduction to an Abel equation of the first kind.

The substitution $y + g = 1/u$ leads to an Abel equation of the first kind:

$$u'_x + (f_0 - f_1g + f_2g^2)u^3 + (f_1 - 2f_2g + g'_x)u^2 + f_2u = 0.$$

For equations of this type, see Subsection 12.1.5.

12.1.7. Equations Not Solved for the Derivative

12.1.7-1. Method of "integration by differentiation."

In the general case, a first-order equation not solved for the derivative,

$$F(x, y, y'_x) = 0, \tag{12.1.7.1}$$

can be rewritten in the equivalent form

$$F(x, y, t) = 0, \qquad t = y'_x. \tag{12.1.7.2}$$

We look for a solution in parametric form: $x = x(t),\ y = y(t)$. In accordance with the first relation in (12.1.7.2), the differential of F is given by

$$F_x\,dx + F_y\,dy + F_t\,dt = 0. \tag{12.1.7.3}$$

Using the relation $dy = t\,dx$, we eliminate successively dy and dx from (12.1.7.3). As a result, we obtain the system of two first-order ordinary differential equations:

$$\frac{dx}{dt} = -\frac{F_t}{F_x + tF_y}, \qquad \frac{dy}{dt} = -\frac{tF_t}{F_x + tF_y}. \tag{12.1.7.4}$$

By finding a solution of this system, one thereby obtains a solution of the original equation (12.1.7.1) in parametric form, $x = x(t),\ y = y(t)$.

Remark 1. The application of the above method may lead to loss of individual solutions (satisfying the condition $F_x + tF_y = 0$); this issue should be additionally investigated.

Remark 2. One of the differential equations of system (12.1.7.4) can be replaced by the algebraic equation $F(x, y, t) = 0$; see equation (12.1.7.2). This technique is used subsequently in Paragraphs 12.1.7-2, 12.1.7-3, and 12.1.7-5.

12.1.7-2. Equations of the form $y = f(y'_x)$.

This equation is a special case of equation (12.1.7.1), with $F(x, y, t) = y - f(t)$. The procedure described in Paragraph 12.1.7-1 yields

$$\frac{dx}{dt} = \frac{f'(t)}{t}, \qquad y = f(t). \tag{12.1.7.5}$$

Here, the original equation is used instead of the second equation in system (12.1.7.4); this is valid because the first equation in (12.1.7.4) does not depend on y explicitly.

Integrating the first equation in (12.1.7.5) yields the solution in parametric form,

$$x = \int \frac{f'(t)}{t}\,dt + C, \qquad y = f(t).$$

12.1.7-3. Equations of the form $x = f(y'_x)$.

This equation is a special case of equation (12.1.7.1), with $F(x, y, t) = x - f(t)$. The procedure described in Paragraph 12.1.7-1 yields

$$x = f(t), \qquad \frac{dy}{dt} = tf'(t). \tag{12.1.7.6}$$

Here, the original equation is used instead of the first equation in system (12.1.7.4); this is valid because the second equation in (12.1.7.4) does not depend on x explicitly.

Integrating the second equation in (12.1.7.5) yields the solution in parametric form,

$$x = f(t), \qquad y = \int tf'(t)\,dt + C.$$

12.1.7-4. Clairaut's equation $y = xy'_x + f(y'_x)$.

Clairaut's equation is a special case of equation (12.1.7.1), with $F(x, y, t) = y - xt - f(t)$. It can be rewritten as

$$y = xt + f(t), \qquad t = y'_x. \tag{12.1.7.7}$$

This equation corresponds to the degenerate case $F_x + tF_y \equiv 0$, where system (12.1.7.4) cannot be obtained. One should proceed in the following way: the first relation in (12.1.7.7) gives $dy = x\,dt + t\,dx + f'(t)\,dt$; performing the substitution $dy = t\,dx$, which follows from the second relation in (12.1.7.7), one obtains

$$[x + f'(t)]\,dt = 0.$$

This equation splits into $dt = 0$ and $x + f'(t) = 0$. The solution of the first equation is obvious: $t = C$; it gives the general solution of Clairaut's equation,

$$y = Cx + f(C), \tag{12.1.7.8}$$

which is a family of straight lines. The second equation generates a solution in parametric form,

$$x = -f'(t), \qquad y = -tf'(t) + f(t), \tag{12.1.7.9}$$

which is a singular solution and is the envelope of the family of lines (12.1.7.8).

Remark. There are also "compound" solutions of Clairaut's equation; they consist of part of curve (12.1.7.9) joined with the tangents at finite points; these tangents are defined by formula (12.1.7.8).

12.1.7-5. Lagrange's equation $y = xf(y'_x) + g(y'_x)$.

Lagrange's equation is a special case of equation (12.1.7.1), with $F(x, y, t) = y - xf(t) - g(t)$. In the special case $f(t) \equiv t$, it coincides with Clairaut's equation; see Paragraph 12.1.7-4.

The procedure described in Paragraph 12.1.7-1 yields

$$\frac{dx}{dt} + \frac{f'(t)}{f(t) - t}x = \frac{g'(t)}{t - f(t)}, \qquad y = xf(t) + g(t). \tag{12.1.7.10}$$

Here, the original equation is used instead of the second equation in system (12.1.7.4); this is valid because the first equation in (12.1.7.4) does not depend on y explicitly.

The first equation of system (12.1.7.10) is linear. Its general solution has the form $x = \varphi(t)C + \psi(t)$; the functions φ and ψ are defined in Paragraph 12.1.2-5. Substituting this solution into the second equation in (12.1.7.10), we obtain the general solution of Lagrange's equation in parametric form,

$$x = \varphi(t)C + \psi(t), \qquad y = \big[\varphi(t)C + \psi(t)\big]f(t) + g(t).$$

Remark. With the above method, solutions of the form $y = t_k x + g(t_k)$, where the t_k are roots of the equation $f(t) - t = 0$, may be lost. These solutions can be particular or singular solutions of Lagrange's equation.

12.1.8. Contact Transformations

12.1.8-1. General form of contact transformations.

A contact transformation has the form

$$\begin{aligned} x &= F(X, Y, Y'_X), \\ y &= G(X, Y, Y'_X), \end{aligned} \tag{12.1.8.1}$$

where the functions $F(X, Y, U)$ and $G(X, Y, U)$ are chosen so that the derivative y'_x does not depend on Y''_{XX}:

$$y'_x = \frac{y'_X}{x'_X} = \frac{G_X + G_Y Y'_X + G_U Y''_{XX}}{F_X + F_Y Y'_X + F_U Y''_{XX}} = H(X, Y, Y'_X). \tag{12.1.8.2}$$

The subscripts X, Y, and U after F and G denote the respective partial derivatives (it is assumed that $F_U \not\equiv 0$ and $G_U \not\equiv 0$).

It follows from (12.1.8.2) that the relation

$$\frac{\partial G}{\partial U}\Big(\frac{\partial F}{\partial X} + U\frac{\partial F}{\partial Y}\Big) - \frac{\partial F}{\partial U}\Big(\frac{\partial G}{\partial X} + U\frac{\partial G}{\partial Y}\Big) = 0 \tag{12.1.8.3}$$

holds; the derivative is calculated by

$$y'_x = \frac{G_U}{F_U}, \tag{12.1.8.4}$$

where $G_U/F_U \not\equiv \text{const}$.

The application of contact transformations preserves the order of differential equations. The inverse of a contact transformation can be obtained by solving system (12.1.8.1) and (12.1.8.4) for X, Y, Y'_X.

12.1.8-2. Method for the construction of contact transformations.

Suppose the function $F = F(X, Y, U)$ in the contact transformation (12.1.8.1) is specified. Then relation (12.1.8.3) can be viewed as a linear partial differential equation for the second function G. The corresponding characteristic system of ordinary differential equations (see Subsection 13.1.1),

$$\frac{dX}{1} = \frac{dY}{U} = -\frac{F_U\, dU}{F_X + U F_Y},$$

admits the obvious first integral:

$$F(X, Y, U) = C_1, \tag{12.1.8.5}$$

where C_1 is an arbitrary constant. It follows that, to obtain the general representation of the function $G = G(X, Y, U)$, one has to deal with the ordinary differential equation

$$Y'_X = U, \tag{12.1.8.6}$$

whose right-hand side is defined in implicit form by (12.1.8.5). Let the first integral of equation (12.1.8.6) have the form

$$\Phi(X, Y, C_1) = C_2.$$

Then the general representation of $G = G(X, Y, U)$ in transformation (12.1.8.1) is given by

$$G = \Psi(F, \widetilde{\Phi}),$$

where $\Psi(F, \widetilde{\Phi})$ is an arbitrary function of two variables, $F = F(X, Y, U)$ and $\widetilde{\Phi} = \Phi(X, Y, F)$.

12.1.8-3. Examples of contact transformations.

Example 1. *Legendre transformation*:

$$x = Y'_X, \quad y = XY'_X - Y, \quad y'_x = X \quad \text{(direct transformation)};$$
$$X = y'_x, \quad Y = xy'_x - y, \quad Y'_X = x \quad \text{(inverse transformation)}.$$

This transformation is used for solving some equations. In particular, the nonlinear equation

$$(xy'_x - y)^a f(y'_x) + yg(y'_x) + xh(y'_x) = 0$$

can be reduced by the Legendre transformation to a Bernoulli equation: $[Xg(X)+h(X)]Y'_X = g(X)Y - f(X)Y^a$ (see Paragraph 12.1.2-6).

Example 2. Contact transformation ($a \neq 0$):

$$x = Y'_X + aY, \quad y = be^{aX}Y'_X, \quad y'_x = be^{aX} \quad \text{(direct transformation)};$$
$$X = \frac{1}{a}\ln\frac{y'_x}{b}, \quad Y = \frac{1}{a}\Big(x - \frac{y}{y'_x}\Big), \quad Y'_X = \frac{y}{y'_x} \quad \text{(inverse transformation)}.$$

Example 3. Contact transformation ($a \neq 0$):

$$x = Y'_X + aX, \quad y = \tfrac{1}{2}(Y'_X)^2 + aY, \quad y'_x = Y'_X \quad \text{(direct transformation)};$$
$$X = \frac{1}{a}\big(x - y'_x\big), \quad Y = \frac{1}{2a}\big[2y - (y'_x)^2\big], \quad Y'_X = y'_x \quad \text{(inverse transformation)}.$$

12.1.9. Approximate Analytic Methods for Solution of Equations

12.1.9-1. Method of successive approximations (Picard method).

The method of successive approximations consists of two stages. At the first stage, the Cauchy problem

$$y'_x = f(x, y) \qquad \text{(equation)}, \tag{12.1.9.1}$$
$$y(x_0) = y_0 \qquad \text{(initial condition)} \tag{12.1.9.2}$$

is reduced to the equivalent integral equation:

$$y(x) = y_0 + \int_{x_0}^{x} f(t, y(t))\,dt. \tag{12.1.9.3}$$

Then a solution of equation (12.1.9.3) is sought using the formula of successive approximations:

$$y_{n+1}(x) = y_0 + \int_{x_0}^{x} f(t, y_n(t))\,dt; \qquad n = 0,\ 1,\ 2,\ \ldots$$

The initial approximation $y_0(x)$ can be chosen arbitrarily; the simplest way is to take y_0 to be a number. The iterative process converges as $n \to \infty$, provided the conditions of the theorems in Paragraph 12.1.1-3 are satisfied.

12.1.9-2. Method of Taylor series expansion in the independent variable.

A solution of the Cauchy problem (12.1.9.1)–(12.1.9.2) can be sought in the form of the Taylor series in powers of $(x - x_0)$:

$$y(x) = y(x_0) + y'_x(x_0)(x - x_0) + \frac{y''_{xx}(x_0)}{2!}(x - x_0)^2 + \cdots . \tag{12.1.9.4}$$

The first coefficient $y(x_0)$ in solution (12.1.9.4) is prescribed by the initial condition (12.1.9.2). The values of the derivatives of $y(x)$ at $x = x_0$ are determined from equation (12.1.9.1) and its derivative equations (obtained by successive differentiation), taking into account the initial condition (12.1.9.2). In particular, setting $x = x_0$ in (12.1.9.1) and substituting (12.1.9.2), one obtains the value of the first derivative:

$$y'_x(x_0) = f(x_0, y_0). \tag{12.1.9.5}$$

Further, differentiating equation (12.1.9.1) yields

$$y''_{xx} = f_x(x, y) + f_y(x, y)y'_x. \tag{12.1.9.6}$$

On substituting $x = x_0$, as well as the initial condition (12.1.9.2) and the first derivative (12.1.9.5), into the right-hand side of this equation, one calculates the value of the second derivative:

$$y''_{xx}(x_0) = f_x(x_0, y_0) + f(x_0, y_0)f_y(x_0, y_0).$$

Likewise, one can determine the subsequent derivatives of y at $x = x_0$.

Solution (12.1.9.4) obtained by this method can normally be used in only some sufficiently small neighborhood of the point $x = x_0$.

Example. Consider the Cauchy problem for the equation

$$y' = e^y + \cos x$$

with the initial condition $y(0) = 0$.

Since $x_0 = 0$, we will be constructing a series in powers of x. If follows from the equation that $y'(0) = e^0 + \cos 0 = 2$. Differentiating the original equation yields $y'' = e^y y' - \sin x$. Using the initial condition and the condition $y'(0) = 2$ just obtained, we have $y''(0) = e^0 \times 2 - \sin 0 = 2$. Similarly, we find that $y''' = e^y y'' + e^y(y')^2 - \cos x$, whence $y'''(0) = e^0 \times 2 + e^0 \times 2^2 - \cos 0 = 5$.

Substituting the values of the derivatives at $x = 0$ into series (12.1.9.4), we obtain the desired series representation of the solution: $y = 2x + x^2 + \frac{5}{6}x^3 + \cdots$.

12.1.9-3. Method of regular expansion in the small parameter.

Consider a general first-order ordinary differential equation with a small parameter ε:

$$y'_x = f(x, y, \varepsilon). \tag{12.1.9.7}$$

Suppose the function f is representable as a series in powers of ε:

$$f(x, y, \varepsilon) = \sum_{n=0}^{\infty} \varepsilon^n f_n(x, y). \tag{12.1.9.8}$$

One looks for a solution of the Cauchy problem for equation (12.1.9.7) with the initial condition (12.1.9.2) as $\varepsilon \to 0$ in the form of a regular expansion in powers of the small parameter:

$$y = \sum_{n=0}^{\infty} \varepsilon^n Y_n(x). \tag{12.1.9.9}$$

Relation (12.1.9.9) is substituted in equation (12.1.9.7) taking into account (12.1.9.8). Then one expands the functions f_n into a power series in ε and matches the coefficients of like powers of ε to obtain a system of equations for $Y_n(x)$:

$$Y_0' = f_0(x, Y_0), \tag{12.1.9.10}$$

$$Y_1' = g(x, Y_0)Y_1 + f_1(x, Y_0), \qquad g(x, y) = \frac{\partial f_0}{\partial y}. \tag{12.1.9.11}$$

Only the first two equations are written out here. The prime denotes differentiation with respect to x. The initial conditions for Y_n can be obtained from (12.1.9.2) taking into account (12.1.9.9):

$$Y_0(x_0) = y_0, \qquad Y_1(x_0) = 0.$$

Success in the application of this method is primarily determined by the possibility of constructing a solution of equation (12.1.9.10) for the leading term in the expansion of Y_0. It is significant that the remaining terms of the expansion, Y_n with $n \geq 1$, are governed by linear equations with homogeneous initial conditions.

Remark 1. Paragraph 12.3.5-2 gives an example of solving a Cauchy problem by the method of regular expansion for a second-order equation and also discusses characteristic features of the method.

Remark 2. The methods of scaled coordinates, two-scale expansions, and matched asymptotic expansions are also used to solve problems defined by first-order differential equations with a small parameter. The basic ideas of these methods are given in Subsection 12.3.5.

12.1.10. Numerical Integration of Differential Equations

12.1.10-1. Method of Euler polygonal lines.

Consider the Cauchy problem for the first-order differential equation

$$y_x' = f(x, y)$$

with the initial condition $y(x_0) = y_0$. Our aim is to construct an approximate solution $y = y(x)$ of this equation on an interval $[x_0, x_*]$.

Let us split the interval $[x_0, x_*]$ into n equal segments of length $\Delta x = \dfrac{x_* - x_0}{n}$. We seek approximate values $y_1, y_2, \ldots, y_n$ of the function $y(x)$ at the partitioning points $x_1, x_2, \ldots, x_n = x_*$.

For a given initial value $y_0 = y(x_0)$ and a sufficiently small Δx, the values of the unknown function $y_k = y(x_k)$ at the other points $x_k = x_0 + k\Delta x$ are calculated successively by the formula

$$y_{k+1} = y_k + f(x_k, y_k)\Delta x \qquad (\textit{Euler polygonal line}),$$

where $k = 0, 1, \ldots, n-1$. The Euler method is a single-step method of the first-order approximation (with respect to the step Δx).

12.1.10-2. Single-step methods of the second-order approximation.

Two single-step methods for solving the Cauchy problem in the second-order approximation are specified by the recurrence formulas

$$y_{k+1} = y_k + f\big(x_k + \tfrac{1}{2}\Delta x,\ y_k + \tfrac{1}{2}f_k\Delta x\big)\Delta x,$$

$$y_{k+1} = y_k + \tfrac{1}{2}\big[f_k + f(x_{k+1},\ y_k + f_k\Delta x)\big]\Delta x,$$

where $f_k = f(x_k, y_k)$; $k = 0, 1, \ldots, n-1$.

12.1.10-3. Runge–Kutta method of the fourth-order approximation.

This is one of the widely used methods. The unknown values y_k are successively found by the formulas

$$y_{k+1} = y_k + \tfrac{1}{6}(f_1 + 2f_2 + 2f_3 + f_4)\Delta x,$$

where

$$f_1 = f(x_k,\ y_k), \qquad f_2 = f(x_k + \tfrac{1}{2}\Delta x,\ y_k + \tfrac{1}{2}f_1\Delta x),$$
$$f_3 = f(x_k + \tfrac{1}{2}\Delta x,\ y_k + \tfrac{1}{2}f_2\Delta x), \qquad f_4 = f(x_k + \Delta x,\ y_k + f_3\Delta x).$$

Remark 1. All methods described in Subsection 12.1.10 are special cases of the Runge–Kutta method (a detailed description of this method can be found in the monographs listed at the end of the current chapter).

Remark 2. In practice, calculations are performed on the basis of any of the above recurrence formulas with two different steps Δx, $\frac{1}{2}\Delta x$ and an arbitrarily chosen small Δx. Then one compares the results obtained at common points. If these results coincide within the given order of accuracy, one assumes that the chosen step Δx ensures the desired accuracy of calculations. Otherwise, the step is halved and the calculations are performed with the steps $\frac{1}{2}\Delta x$ and $\frac{1}{4}\Delta x$, after which the results are compared again, etc. (Quite often, one compares the results of calculations with steps varying by ten or more times.)

12.2. Second-Order Linear Differential Equations

12.2.1. Formulas for the General Solution. Some Transformations

12.2.1-1. Homogeneous linear equations. Formulas for the general solution.

1°. Consider a second-order homogeneous linear equation in the general form

$$f_2(x)y''_{xx} + f_1(x)y'_x + f_0(x)y = 0. \tag{12.2.1.1}$$

The *trivial solution*, $y = 0$, is a particular solution of the homogeneous linear equation.

Let $y_1(x)$, $y_2(x)$ be a fundamental system of solutions (nontrivial linearly independent particular solutions) of equation (12.2.1.1). Then the general solution is given by

$$y = C_1y_1(x) + C_2y_2(x), \tag{12.2.1.2}$$

where C_1 and C_2 are arbitrary constants.

2°. Let $y_1 = y_1(x)$ be any nontrivial particular solution of equation (12.2.1.1). Then its general solution can be represented as

$$y = y_1\left(C_1 + C_2\int \frac{e^{-F}}{y_1^2}\,dx\right), \qquad \text{where} \quad F = \int \frac{f_1}{f_2}\,dx. \tag{12.2.1.3}$$

3°. Consider the equation

$$y''_{xx} + f(x)y = 0,$$

which is written in the canonical form; see Paragraph 12.2.1-3 for the reduction of equations to this form. Let $y_1(x)$ be any nontrivial partial solution of this equation. The general solution can be constructed by formula (12.2.1.3) with $F = 0$ or formula (12.2.1.2) in which

$$y_2(x) = y_1\int \frac{[f(x)-1][y_1^2 - (y_1')^2]}{[y_1^2 + (y_1')^2]^2}\,dx + \frac{y_1'}{y_1^2 + (y_1')^2}.$$

Here, $y_1 = y_1(x)$ and the prime denotes differentiation with respect to x. The last formula is suitable where y_1 vanishes at some points.

4°. The second-order constant coefficient linear equation

$$y''_{xx} + ay'_x + by = 0$$

has the following fundamental system of solutions:

$$y_1(x) = \exp\left(-\tfrac{1}{2}ax\right)\sinh\left(\tfrac{1}{2}x\sqrt{a^2-4b}\right), \quad y_2(x) = \exp\left(-\tfrac{1}{2}ax\right)\cosh\left(\tfrac{1}{2}x\sqrt{a^2-4b}\right) \quad \text{if } a^2 > 4b;$$
$$y_1(x) = \exp\left(-\tfrac{1}{2}ax\right)\sin\left(\tfrac{1}{2}x\sqrt{4b-a^2}\right), \quad y_2(x) = \exp\left(-\tfrac{1}{2}ax\right)\cos\left(\tfrac{1}{2}x\sqrt{4b-a^2}\right) \quad \text{if } a^2 < 4b;$$
$$y_1(x) = \exp\left(-\tfrac{1}{2}ax\right), \quad y_2(x) = x\exp\left(-\tfrac{1}{2}ax\right) \quad \text{if } a^2 = 4b.$$

5°. The *Euler equation*

$$x^2y''_{xx} + axy'_x + by = 0$$

is reduced by the change of variable $x = ke^t$ ($k \neq 0$) to the second-order constant coefficient linear equation $y''_{tt} + (a-1)y'_t + by = 0$, which is treated in Item 4°.

► *Solutions to some other second-order linear equations can be found in Section T5.2.*

12.2.1-2. Wronskian determinant and Liouville's formula.

The *Wronskian determinant* (or *Wronskian*) is defined by

$$W(x) = \begin{vmatrix} y_1(x) & y_2(x) \\ y'_1(x) & y'_2(x) \end{vmatrix} \equiv y_1(y_2)'_x - y_2(y_1)'_x,$$

where $y_1(x)$, $y_2(x)$ is a fundamental system of solutions of equation (12.2.1.1).

Liouville's formula:

$$W(x) = W(x_0)\exp\left[-\int_{x_0}^{x}\frac{f_1(t)}{f_2(t)}\,dt\right].$$

12.2.1-3. Reduction to the canonical form.

1°. The substitution

$$y = u(x)\exp\left(-\frac{1}{2}\int\frac{f_1}{f_2}\,dx\right) \tag{12.2.1.4}$$

brings equation (12.2.1.1) to the canonical (or normal) form

$$u''_{xx} + f(x)u = 0, \qquad \text{where} \quad f = \frac{f_0}{f_2} - \frac{1}{4}\left(\frac{f_1}{f_2}\right)^2 - \frac{1}{2}\left(\frac{f_1}{f_2}\right)'_x. \tag{12.2.1.5}$$

2°. The substitution (12.2.1.4) is a special case of the more general transformation (φ is an arbitrary function)

$$x = \varphi(\xi), \quad y = u(\xi)\sqrt{|\varphi'_\xi(\xi)|}\,\exp\left(-\frac{1}{2}\int\frac{f_1(\varphi)}{f_2(\varphi)}\,d\varphi\right),$$

which also brings the original equation to the canonical form.

12.2.1-4. Reduction to the Riccati equation.

The substitution $u = y'_x/y$ brings the second-order homogeneous linear equation (12.2.1.1) to the Riccati equation:

$$f_2(x)u'_x + f_2(x)u^2 + f_1(x)u + f_0(x) = 0,$$

which is discussed in Subsection 12.1.4.

12.2.1-5. Nonhomogeneous linear equations. The existence theorem.

A second-order nonhomogeneous linear equation has the form

$$f_2(x)y''_{xx} + f_1(x)y'_x + f_0(x)y = g(x). \tag{12.2.1.6}$$

THEOREM (EXISTENCE AND UNIQUENESS). *On an open interval $a < x < b$, let the functions f_2, f_1, f_0, and g be continuous and $f_2 \neq 0$. Also let*

$$y(x_0) = A, \quad y'_x(x_0) = B$$

be arbitrary initial conditions, where x_0 is any point such that $a < x_0 < b$, and A and B are arbitrary prescribed numbers. Then a solution of equation (12.2.1.6) exists and is unique. This solutions is defined for all $x \in (a, b)$.

12.2.1-6. Nonhomogeneous linear equations. Formulas for the general solution.

1°. The general solution of the nonhomogeneous linear equation (12.2.1.6) is the sum of the general solution of the corresponding homogeneous equation (12.2.1.1) and any particular solution of the nonhomogeneous equation (12.2.1.6).

2°. Let $y_1 = y_1(x)$, $y_2 = y_2(x)$ be a fundamental system of solutions of the corresponding homogeneous equation, with $g \equiv 0$. Then the general solution of equation (12.2.1.6) can be represented as

$$y = C_1y_1 + C_2y_2 + y_2\int y_1\frac{g}{f_2}\frac{dx}{W} - y_1\int y_2\frac{g}{f_2}\frac{dx}{W}, \tag{12.2.1.7}$$

where $W = y_1(y_2)'_x - y_2(y_1)'_x$ is the Wronskian determinant.

3°. Given a nontrivial particular solution $y_1 = y_1(x)$ of the homogeneous equation (with $g \equiv 0$), a second particular solution $y_2 = y_2(x)$ can be calculated from the formula

$$y_2 = y_1\int \frac{e^{-F}}{y_1^2}\,dx, \qquad \text{where} \quad F = \int \frac{f_1}{f_2}\,dx, \quad W = e^{-F}. \tag{12.2.1.8}$$

Then the general solution of equation (12.2.1.6) can be constructed by (12.2.1.7).

4°. Let $\bar{y}_1$ and $\bar{y}_2$ be respective solutions of the nonhomogeneous differential equations $\mathrm{L}[\bar{y}_1] = g_1(x)$ and $\mathrm{L}[\bar{y}_2] = g_2(x)$, which have the same left-hand side but different right-hand sides, where $\mathrm{L}[y]$ is the left-hand side of equation (12.2.1.6). Then the function $\bar{y} = \bar{y}_1 + \bar{y}_2$ is a solution of the equation $\mathrm{L}[\bar{y}] = g_1(x) + g_2(x)$.

12.2.1-7. Reduction to a constant coefficient equation (a special case).

Let $f_2 = 1$, $f_0 \neq 0$, and the condition

$$\frac{1}{|f_0|}\frac{d}{dx}\sqrt{|f_0|} + \frac{f_1}{\sqrt{|f_0|}} = a = \text{const}$$

be satisfied. Then the substitution $\xi = \int \sqrt{|f_0|}\,dx$ leads to a constant coefficient linear equation,

$$y''_{\xi\xi} + ay'_\xi + y \operatorname{sign} f_0 = 0.$$

12.2.1-8. Kummer–Liouville transformation.

The transformation

$$x = \alpha(t), \quad y = \beta(t)z + \gamma(t), \tag{12.2.1.9}$$

where $\alpha(t)$, $\beta(t)$, and $\gamma(t)$ are arbitrary sufficiently smooth functions ($\beta \not\equiv 0$), takes any linear differential equation for $y(x)$ to a linear equation for $z = z(t)$. In the special case $\gamma \equiv 0$, a homogeneous equation is transformed to a homogeneous one.

Special cases of transformation (12.2.1.9) are widely used to simplify second- and higher-order linear differential equations.

12.2.2. Representation of Solutions as a Series in the Independent Variable

12.2.2-1. Equation coefficients are representable in the ordinary power series form.

Let us consider a homogeneous linear differential equation of the general form

$$y''_{xx} + f(x)y'_x + g(x)y = 0. \tag{12.2.2.1}$$

Assume that the functions $f(x)$ and $g(x)$ are representable, in the vicinity of a point $x = x_0$, in the power series form,

$$f(x) = \sum_{n=0}^{\infty} A_n(x - x_0)^n, \qquad g(x) = \sum_{n=0}^{\infty} B_n(x - x_0)^n, \tag{12.2.2.2}$$

on the interval $|x - x_0| < R$, where R stands for the minimum radius of convergence of the two series in (12.2.2.2). In this case, the point $x = x_0$ is referred to as an *ordinary point*, and equation (12.2.2.1) possesses two linearly independent solutions of the form

$$y_1(x) = \sum_{n=0}^{\infty} a_n(x - x_0)^n, \qquad y_2(x) = \sum_{n=0}^{\infty} b_n(x - x_0)^n. \tag{12.2.2.3}$$

The coefficients a_n and b_n are determined by substituting the series (12.2.2.2) into equation (12.2.2.1) followed by extracting the coefficients of like powers of $(x - x_0)$.*

* Prior to that, the terms containing the same powers $(x - x_0)^k$, $k = 0, 1, \ldots$, should be collected.

12.2.2-2. Equation coefficients have poles at some point.

Assume that the functions $f(x)$ and $g(x)$ are representable, in the vicinity of a point $x = x_0$, in the form

$$f(x) = \sum_{n=-1}^{\infty} A_n (x - x_0)^n, \qquad g(x) = \sum_{n=-2}^{\infty} B_n (x - x_0)^n, \tag{12.2.2.4}$$

on the interval $|x - x_0| < R$. In this case, the point $x = x_0$ is referred to as a *regular singular point*. Let λ_1 and λ_2 be roots of the quadratic equation

$$\lambda_1^2 + (A_{-1} - 1)\lambda + B_{-2} = 0.$$

There are three cases, depending on the values of the exponents of the singularity.

1. If $\lambda_1 \neq \lambda_2$ and $\lambda_1 - \lambda_2$ is not an integer, equation (12.2.2.1) has two linearly independent solutions of the form

$$\begin{aligned} y_1(x) &= |x - x_0|^{\lambda_1} \Big[1 + \sum_{n=1}^{\infty} a_n (x - x_0)^n \Big], \\ y_2(x) &= |x - x_0|^{\lambda_2} \Big[1 + \sum_{n=1}^{\infty} b_n (x - x_0)^n \Big]. \end{aligned} \tag{12.2.2.5}$$

2. If $\lambda_1 = \lambda_2 = \lambda$, equation (12.2.2.1) possesses two linearly independent solutions:

$$\begin{aligned} y_1(x) &= |x - x_0|^{\lambda} \Big[1 + \sum_{n=1}^{\infty} a_n (x - x_0)^n \Big], \\ y_2(x) &= y_1(x) \ln|x - x_0| + |x - x_0|^{\lambda} \sum_{n=0}^{\infty} b_n (x - x_0)^n. \end{aligned}$$

3. If $\lambda_1 = \lambda_2 + N$, where N is a positive integer, equation (12.2.2.1) has two linearly independent solutions of the form

$$\begin{aligned} y_1(x) &= |x - x_0|^{\lambda_1} \Big[1 + \sum_{n=1}^{\infty} a_n (x - x_0)^n \Big], \\ y_2(x) &= k y_1(x) \ln|x - x_0| + |x - x_0|^{\lambda_2} \sum_{n=0}^{\infty} b_n (x - x_0)^n, \end{aligned}$$

where k is a constant to be determined (it may be equal to zero).

To construct the solution in each of the three cases, the following procedure should be performed: substitute the above expressions of y_1 and y_2 into the original equation (12.2.2.1) and equate the coefficients of $(x - x_0)^n$ and $(x - x_0)^n \ln|x - x_0|$ for different values of n to obtain recurrence relations for the unknown coefficients. From these recurrence relations the solution sought can be found.

12.2.3. Asymptotic Solutions

This subsection presents asymptotic solutions, as $\varepsilon \to 0$ $(\varepsilon > 0)$, of some second-order linear ordinary differential equations containing arbitrary functions (sufficiently smooth), with the independent variable being real.

12.2.3-1. Equations not containing y'_x. Leading asymptotic terms.

1°. Consider the equation

$$\varepsilon^2 y''_{xx} - f(x)y = 0 \tag{12.2.3.1}$$

on a closed interval $a \le x \le b$.

Case 1. With the condition $f \ne 0$, the leading terms of the asymptotic expansions of the fundamental system of solutions, as $\varepsilon \to 0$, are given by the formulas

$$y_1 = f^{-1/4}\exp\Big(-\frac{1}{\varepsilon}\int \sqrt{f}\,dx\Big), \qquad y_2 = f^{-1/4}\exp\Big(\frac{1}{\varepsilon}\int \sqrt{f}\,dx\Big) \qquad \text{if } f > 0,$$

$$y_1 = (-f)^{-1/4}\cos\Big(\frac{1}{\varepsilon}\int \sqrt{-f}\,dx\Big), \quad y_2 = (-f)^{-1/4}\sin\Big(\frac{1}{\varepsilon}\int \sqrt{-f}\,dx\Big) \quad \text{if } f < 0.$$

Case 2. Discuss the asymptotic solution of equation (12.2.3.1) in the vicinity of the point $x = x_0$, where function $f(x)$ vanishes, $f(x_0) = 0$ (such a point is referred to as a *transition point*). We assume that the function f can be presented in the form

$$f(x) = (x_0 - x)\psi(x), \qquad \text{where} \quad \psi(x) > 0.$$

In this case, the fundamental solutions, as $\varepsilon \to 0$, are described by three different formulas:

$$y_1 = \begin{cases} \dfrac{1}{|f(x)|^{1/4}}\sin\Big[\dfrac{1}{\varepsilon}\displaystyle\int_{x_0}^{x}\sqrt{|f(x)|}\,dx + \dfrac{\pi}{4}\Big] & \text{if } x - x_0 \ge \delta, \\ \dfrac{\sqrt{\pi}}{[\varepsilon\psi(x_0)]^{1/6}}\,\mathrm{Ai}(z) & \text{if } |x - x_0| \le \delta, \\ \dfrac{1}{2[f(x)]^{1/4}}\exp\Big[-\dfrac{1}{\varepsilon}\displaystyle\int_{x}^{x_0}\sqrt{f(x)}\,dx\Big] & \text{if } x_0 - x \ge \delta, \end{cases}$$

$$y_2 = \begin{cases} \dfrac{1}{|f(x)|^{1/4}}\cos\Big[\dfrac{1}{\varepsilon}\displaystyle\int_{x_0}^{x}\sqrt{|f(x)|}\,dx + \dfrac{\pi}{4}\Big] & \text{if } x - x_0 \ge \delta, \\ \dfrac{\sqrt{\pi}}{[\varepsilon\psi(x_0)]^{1/6}}\,\mathrm{Bi}(z) & \text{if } |x - x_0| \le \delta, \\ \dfrac{1}{[f(x)]^{1/4}}\exp\Big[\dfrac{1}{\varepsilon}\displaystyle\int_{x}^{x_0}\sqrt{f(x)}\,dx\Big] & \text{if } x_0 - x \ge \delta, \end{cases}$$

where $\mathrm{Ai}(z)$ and $\mathrm{Bi}(z)$ are the Airy functions of the first and second kind, respectively (see Section 18.8), $z = \varepsilon^{-2/3}[\psi(x_0)]^{1/3}(x_0 - x)$, and $\delta = O(\varepsilon^{2/3})$.

12.2.3-2. Equations not containing y'_x. Two-term asymptotic expansions.

The two-term asymptotic expansions of the solution of equation (12.2.3.1) with $f > 0$, as $\varepsilon \to 0$, on a closed interval $a \le x \le b$, has the form

$$y_1 = f^{-1/4}\exp\Big(-\frac{1}{\varepsilon}\int_{x_0}^{x}\sqrt{f}\,dx\Big)\Big\{1-\varepsilon\int_{x_0}^{x}\Big[\frac{1}{8}\frac{f''_{xx}}{f^{3/2}}-\frac{5}{32}\frac{(f'_x)^2}{f^{5/2}}\Big]\,dx+O(\varepsilon^2)\Big\},$$

$$y_2 = f^{-1/4}\exp\Big(\frac{1}{\varepsilon}\int_{x_0}^{x}\sqrt{f}\,dx\Big)\Big\{1+\varepsilon\int_{x_0}^{x}\Big[\frac{1}{8}\frac{f''_{xx}}{f^{3/2}}-\frac{5}{32}\frac{(f'_x)^2}{f^{5/2}}\Big]\,dx+O(\varepsilon^2)\Big\}, \tag{12.2.3.2}$$

where x_0 is an arbitrary number satisfying the inequality $a \le x_0 \le b$.

The asymptotic expansions of the fundamental system of solutions of equation (12.2.3.1) with $f < 0$ are derived by separating the real and imaginary parts in either formula (12.2.3.2).

12.2.3-3. Equations of special form not containing y'_x.

Consider the equation

$$\varepsilon^2 y''_{xx} - x^{m-2}f(x)y = 0 \tag{12.2.3.3}$$

on a closed interval $a \le x \le b$, where $a < 0$ and $b > 0$, under the conditions that m is a positive integer and $f(x) \ne 0$. In this case, the leading term of the asymptotic solution, as $\varepsilon \to 0$, in the vicinity of the point $x = 0$ is expressed in terms of a simpler model equation, which results from substituting the function $f(x)$ in equation (12.2.3.3) by the constant $f(0)$ (the solution of the model equation is expressed in terms of the Bessel functions of order $1/m$).

We specify below formulas by which the leading terms of the asymptotic expansions of the fundamental system of solutions of equation (12.2.3.3) with $a < x < 0$ and $0 < x < b$ are related (excluding a small vicinity of the point $x = 0$). Three different cases can be extracted.

1°. Let m be an even integer and $f(x) > 0$. Then,

$$y_1 = \begin{cases} [f(x)]^{-1/4}\exp\Big[\dfrac{1}{\varepsilon}\displaystyle\int_0^x\sqrt{f(x)}\,dx\Big] & \text{if } x < 0,\\ k^{-1}[f(x)]^{-1/4}\exp\Big[\dfrac{1}{\varepsilon}\displaystyle\int_0^x\sqrt{f(x)}\,dx\Big] & \text{if } x > 0,\end{cases}$$

$$y_2 = \begin{cases} [f(x)]^{-1/4}\exp\Big[-\dfrac{1}{\varepsilon}\displaystyle\int_0^x\sqrt{f(x)}\,dx\Big] & \text{if } x < 0,\\ k[f(x)]^{-1/4}\exp\Big[-\dfrac{1}{\varepsilon}\displaystyle\int_0^x\sqrt{f(x)}\,dx\Big] & \text{if } x > 0,\end{cases}$$

where $f = f(x)$, $k = \sin\Big(\dfrac{\pi}{m}\Big)$.

2°. Let m be an even integer and $f(x) < 0$. Then,

$$y_1 = \begin{cases} |f(x)|^{-1/4}\cos\Big[-\dfrac{1}{\varepsilon}\displaystyle\int_0^x\sqrt{|f(x)|}\,dx+\dfrac{\pi}{4}\Big] & \text{if } x < 0,\\ k^{-1}|f(x)|^{-1/4}\cos\Big[\dfrac{1}{\varepsilon}\displaystyle\int_0^x\sqrt{|f(x)|}\,dx-\dfrac{\pi}{4}\Big] & \text{if } x > 0,\end{cases}$$

$$y_2 = \begin{cases} |f(x)|^{-1/4}\cos\Big[-\dfrac{1}{\varepsilon}\displaystyle\int_0^x\sqrt{|f(x)|}\,dx-\dfrac{\pi}{4}\Big] & \text{if } x < 0,\\ k|f(x)|^{-1/4}\cos\Big[\dfrac{1}{\varepsilon}\displaystyle\int_0^x\sqrt{|f(x)|}\,dx+\dfrac{\pi}{4}\Big] & \text{if } x > 0,\end{cases}$$

where $f = f(x)$, $k = \tan\Big(\dfrac{\pi}{2m}\Big)$.

3°. Let m be an odd integer. Then,

$$y_1 = \begin{cases} |f(x)|^{-1/4}\cos\Big[-\dfrac{1}{\varepsilon}\displaystyle\int_0^x \sqrt{|f(x)|}\,dx + \dfrac{\pi}{4}\Big] & \text{if } x < 0, \\ \dfrac{1}{2}k^{-1}[f(x)]^{-1/4}\exp\Big[\dfrac{1}{\varepsilon}\displaystyle\int_0^x \sqrt{f(x)}\,dx\Big] & \text{if } x > 0, \end{cases}$$

$$y_2 = \begin{cases} |f(x)|^{-1/4}\cos\Big[-\dfrac{1}{\varepsilon}\displaystyle\int_0^x \sqrt{|f(x)|}\,dx - \dfrac{\pi}{4}\Big] & \text{if } x < 0, \\ k[f(x)]^{-1/4}\exp\Big[-\dfrac{1}{\varepsilon}\displaystyle\int_0^x \sqrt{f(x)}\,dx\Big] & \text{if } x > 0, \end{cases}$$

where $f = f(x)$, $k = \sin\Big(\dfrac{\pi}{2m}\Big)$.

12.2.3-4. Equations not containing y'_x. Equation coefficients are dependent on ε.

Consider an equation of the form

$$\varepsilon^2 y''_{xx} - f(x,\varepsilon)y = 0 \tag{12.2.3.4}$$

on a closed interval $a \le x \le b$ under the condition that $f \ne 0$. Assume that the following asymptotic relation holds:

$$f(x,\varepsilon) = \sum_{k=0}^{\infty} f_k(x)\varepsilon^k, \qquad \varepsilon \to 0.$$

Then the leading terms of the asymptotic expansions of the fundamental system of solutions of equation (12.2.3.4) are given by the formulas

$$y_1 = f_0^{-1/4}(x)\exp\Big[-\frac{1}{\varepsilon}\int \sqrt{f_0(x)}\,dx + \frac{1}{2}\int \frac{f_1(x)}{\sqrt{f_0(x)}}\,dx\Big]\big[1 + O(\varepsilon)\big],$$
$$y_2 = f_0^{-1/4}(x)\exp\Big[\frac{1}{\varepsilon}\int \sqrt{f_0(x)}\,dx + \frac{1}{2}\int \frac{f_1(x)}{\sqrt{f_0(x)}}\,dx\Big]\big[1 + O(\varepsilon)\big].$$

12.2.3-5. Equations containing y'_x.

1°. Consider an equation of the form

$$\varepsilon y''_{xx} + g(x)y'_x + f(x)y = 0$$

on a closed interval $0 \le x \le 1$. With $g(x) > 0$, the asymptotic solution of this equation, satisfying the boundary conditions $y(0) = C_1$ and $y(1) = C_2$, can be represented in the form

$$y = (C_1 - kC_2)\exp\big[-\varepsilon^{-1}g(0)x\big] + C_2\exp\Big[\int_x^1 \frac{f(x)}{g(x)}\,dx\Big] + O(\varepsilon),$$

where $k = \exp\Big[\displaystyle\int_0^1 \frac{f(x)}{g(x)}\,dx\Big]$.

2°. Now let us take a look at an equation of the form

$$\varepsilon^2 y''_{xx} + \varepsilon g(x) y'_x + f(x) y = 0 \tag{12.2.3.5}$$

on a closed interval $a \le x \le b$. Assume

$$D(x) \equiv [g(x)]^2 - 4f(x) \ne 0.$$

Then the leading terms of the asymptotic expansions of the fundamental system of solutions of equation (12.2.3.5), as $\varepsilon \to 0$, are expressed by

$$y_1 = |D(x)|^{-1/4} \exp\Big[-\frac{1}{2\varepsilon}\int \sqrt{D(x)}\,dx - \frac{1}{2}\int \frac{g'_x(x)}{\sqrt{D(x)}}\,dx\Big]\big[1 + O(\varepsilon)\big],$$
$$y_2 = |D(x)|^{-1/4} \exp\Big[\frac{1}{2\varepsilon}\int \sqrt{D(x)}\,dx - \frac{1}{2}\int \frac{g'_x(x)}{\sqrt{D(x)}}\,dx\Big]\big[1 + O(\varepsilon)\big].$$

12.2.3-6. Equations of the general form.

The more general equation

$$\varepsilon^2 y''_{xx} + \varepsilon g(x,\varepsilon) y'_x + f(x,\varepsilon) y = 0$$

is reducible, with the aid of the substitution $y = w \exp\Big(-\dfrac{1}{2\varepsilon}\int g\,dx\Big)$, to an equation of the form (12.2.3.4),

$$\varepsilon^2 w''_{xx} + (f - \tfrac{1}{4}g^2 - \tfrac{1}{2}\varepsilon g'_x) w = 0,$$

to which the asymptotic formulas given above in Paragraph 12.2.3-4 are applicable.

12.2.4. Boundary Value Problems

12.2.4-1. First, second, third, and mixed boundary value problems ($x_1 \le x \le x_2$).

We consider the second-order nonhomogeneous linear differential equation

$$y''_{xx} + f(x) y'_x + g(x) y = h(x). \tag{12.2.4.1}$$

1°. *The first boundary value problem*: Find a solution of equation (12.2.4.1) satisfying the boundary conditions

$$y = a_1 \quad \text{at} \quad x = x_1, \qquad y = a_2 \quad \text{at} \quad x = x_2. \tag{12.2.4.2}$$

(The values of the unknown are prescribed at two distinct points x_1 and x_2.)

2°. *The second boundary value problem*: Find a solution of equation (12.2.4.1) satisfying the boundary conditions

$$y'_x = a_1 \quad \text{at} \quad x = x_1, \qquad y'_x = a_2 \quad \text{at} \quad x = x_2. \tag{12.2.4.3}$$

(The values of the derivative of the unknown are prescribed at two distinct points x_1 and x_2.)

3°. *The third boundary value problem*: Find a solution of equation (12.2.4.1) satisfying the boundary conditions

$$\begin{aligned} y'_x - k_1 y &= a_1 \quad \text{at} \quad x = x_1, \\ y'_x + k_2 y &= a_2 \quad \text{at} \quad x = x_2. \end{aligned} \tag{12.2.4.4}$$

4°. *The third boundary value problem*: Find a solution of equation (12.2.4.1) satisfying the boundary conditions

$$y = a_1 \quad \text{at} \quad x = x_1, \qquad y'_x = a_2 \quad \text{at} \quad x = x_2. \tag{12.2.4.5}$$

(The unknown itself is prescribed at one point, and its derivative at another point.)

Conditions (12.2.4.2), (12.2.4.3), (12.2.4.4), and (12.2.4.5) are called *homogeneous* if $a_1 = a_2 = 0$.

12.2.4-2. Simplification of boundary conditions. The self-adjoint form of equations.

1°. Nonhomogeneous boundary conditions can be reduced to homogeneous ones by the change of variable $z = A_2x^2 + A_1x + A_0 + y$ (the constants A_2, A_1, and A_0 are selected using the method of undetermined coefficients). In particular, the nonhomogeneous boundary conditions of the first kind (12.2.4.2) can be reduced to homogeneous boundary conditions by the linear change of variable

$$z = y - \frac{a_2 - a_1}{x_2 - x_1}(x - x_1) - a_1.$$

2°. On multiplying by $p(x) = \exp\left[\int f(x)\,dx\right]$, one reduces equation (12.2.4.1) to the self-adjoint form:

$$[p(x)y'_x]'_x + q(x)y = r(x). \tag{12.2.4.6}$$

Without loss of generality, we can further consider equation (12.2.4.6) instead of (12.2.4.1). We assume that the functions p, p'_x, q, and r are continuous on the interval $x_1 \le x \le x_2$, and p is positive.

12.2.4-3. Green's function. Linear problems for nonhomogeneous equations.

The *Green's function* of the first boundary value problem for equation (12.2.4.6) with homogeneous boundary conditions (12.2.4.2) is a function of two variables $G(x, s)$ that satisfies the following conditions:

1°. $G(x, s)$ is continuous in x for fixed s, with $x_1 \le x \le x_2$ and $x_1 \le s \le x_2$.

2°. $G(x, s)$ is a solution of the homogeneous equation (12.2.4.6), with $r = 0$, for all $x_1 < x < x_2$ exclusive of the point $x = s$.

3°. $G(x, s)$ satisfies the homogeneous boundary conditions $G(x_1, s) = G(x_2, s) = 0$.

4°. The derivative $G'_x(x, s)$ has a jump of $1/p(s)$ at the point $x = s$, that is,

$$G'_x(x, s)\big|_{x\to s,\, x>s} - G'_x(x, s)\big|_{x\to s,\, x<s} = \frac{1}{p(s)}.$$

For the second, third, and mixed boundary value problems, the Green's function is defined likewise except that in 3° the homogeneous boundary conditions (12.2.4.3), (12.2.4.4), and (12.2.4.5), with $a_1 = a_2 = 0$, are adopted, respectively.

The solution of the nonhomogeneous equation (12.2.4.6) subject to appropriate homogeneous boundary conditions is expressed in terms of the Green's function as follows:*

$$y(x) = \int_{x_1}^{x_2} G(x, s) r(s)\, ds.$$

12.2.4-4. Representation of the Green's function in terms of particular solutions.

We consider the first boundary value problem. Let $y_1 = y_1(x)$ and $y_2 = y_2(x)$ be linearly independent particular solutions of the homogeneous equation (12.2.4.6), with $r = 0$, that satisfy the conditions

$$y_1(x_1) = 0, \qquad y_2(x_2) = 0.$$

(Each of the solutions satisfies one of the homogeneous boundary conditions.)

The Green's function is expressed in terms of solutions of the homogeneous equation as follows:

$$G(x, s) = \begin{cases} \dfrac{y_1(x) y_2(s)}{p(s) W(s)} & \text{for } x_1 \le x \le s, \\ \dfrac{y_1(s) y_2(x)}{p(s) W(s)} & \text{for } s \le x \le x_2, \end{cases} \tag{12.2.4.7}$$

where $W(x) = y_1(x) y_2'(x) - y_1'(x) y_2(x)$ is the Wronskian determinant.

Remark. Formula (12.2.4.7) can also be used to construct the Green's functions for the second, third, and mixed boundary value problems. To this end, one should find two linearly independent solutions, $y_1 = y_1(x)$ and $y_2 = y_2(x)$, of the homogeneous equation; the former satisfies the corresponding homogeneous boundary condition at $x = x_1$ and the latter satisfies the one at $x = x_2$.

12.2.5. Eigenvalue Problems

12.2.5-1. Sturm–Liouville problem.

Consider the second-order homogeneous linear differential equation

$$[p(x) y_x']_x' + [\lambda s(x) - q(x)] y = 0 \tag{12.2.5.1}$$

subject to linear boundary conditions of the general form

$$\begin{aligned} \alpha_1 y_x' + \beta_1 y = 0 \quad &\text{at} \quad x = x_1, \\ \alpha_2 y_x' + \beta_2 y = 0 \quad &\text{at} \quad x = x_2. \end{aligned} \tag{12.2.5.2}$$

It is assumed that the functions p, p_x', s, and q are continuous, and p and s are positive on an interval $x_1 \le x \le x_2$. It is also assumed that $|\alpha_1| + |\beta_1| > 0$ and $|\alpha_2| + |\beta_2| > 0$.

The *Sturm–Liouville problem*: Find the values λ_n of the parameter λ at which problem (12.2.5.1), (12.2.5.2) has a nontrivial solution. Such λ_n are called *eigenvalues* and the corresponding solutions $y_n = y_n(x)$ are called *eigenfunctions* of the Sturm–Liouville problem (12.2.5.1), (12.2.5.2).

* The homogeneous boundary value problem — with $r(x) = 0$ and $a_1 = a_2 = 0$ — is assumed to have only the trivial solution.

12.2.5-2. General properties of the Sturm–Liouville problem (12.2.5.1), (12.2.5.2).

1°. There are infinitely (countably) many eigenvalues. All eigenvalues can be ordered so that $\lambda_1 < \lambda_2 < \lambda_3 < \cdots$. Moreover, $\lambda_n \to \infty$ as $n \to \infty$; hence, there can only be a finite number of negative eigenvalues. Each eigenvalue has multiplicity 1.

2°. The eigenfunctions are defined up to a constant factor. Each eigenfunction $y_n(x)$ has precisely $n-1$ zeros on the open interval (x_1, x_2).

3°. Any two eigenfunctions $y_n(x)$ and $y_m(x)$, $n \neq m$, are orthogonal with weight $s(x)$ on the interval $x_1 \le x \le x_2$:

$$\int_{x_1}^{x_2} s(x) y_n(x) y_m(x)\,dx = 0 \quad \text{if} \quad n \neq m.$$

4°. An arbitrary function $F(x)$ that has a continuous derivative and satisfies the boundary conditions of the Sturm–Liouville problem can be decomposed into an absolutely and uniformly convergent series in the eigenfunctions

$$F(x) = \sum_{n=1}^{\infty} F_n y_n(x),$$

where the Fourier coefficients F_n of $F(x)$ are calculated by

$$F_n = \frac{1}{\|y_n\|^2} \int_{x_1}^{x_2} s(x) F(x) y_n(x)\,dx, \qquad \|y_n\|^2 = \int_{x_1}^{x_2} s(x) y_n^2(x)\,dx.$$

5°. If the conditions

$$q(x) \ge 0, \quad \alpha_1\beta_1 \le 0, \quad \alpha_2\beta_2 \ge 0 \tag{12.2.5.3}$$

hold true, there are no negative eigenvalues. If $q \equiv 0$ and $\beta_1 = \beta_2 = 0$, the least eigenvalue is $\lambda_1 = 0$, to which there corresponds an eigenfunction $y_1 = \text{const}$. In the other cases where conditions (12.2.5.3) are satisfied, all eigenvalues are positive.

6°. The following asymptotic formula is valid for eigenvalues as $n \to \infty$:

$$\lambda_n = \frac{\pi^2 n^2}{\Delta^2} + O(1), \qquad \Delta = \int_{x_1}^{x_2} \sqrt{\frac{s(x)}{p(x)}}\,dx. \tag{12.2.5.4}$$

Paragraphs 12.2.5-3 through 12.2.5-6 will describe special properties of the Sturm–Liouville problem that depend on the specific form of the boundary conditions.

Remark 1. Equation (12.2.5.1) can be reduced to the case where $p(x) \equiv 1$ and $s(x) \equiv 1$ by the change of variables

$$\zeta = \int \sqrt{\frac{s(x)}{p(x)}}\,dx, \qquad u(\zeta) = \bigl[p(x)s(x)\bigr]^{1/4} y(x).$$

In this case, the boundary conditions are transformed to boundary conditions of similar form.

Remark 2. The second-order linear equation

$$\varphi_2(x) y''_{xx} + \varphi_1(x) y'_x + [\lambda + \varphi_0(x)] y = 0$$

can be represented in the form of equation (12.2.5.1) where $p(x)$, $s(x)$, and $q(x)$ are given by

$$p(x) = \exp\Bigl[\int \frac{\varphi_1(x)}{\varphi_2(x)}\,dx\Bigr], \quad s(x) = \frac{1}{\varphi_2(x)} \exp\Bigl[\int \frac{\varphi_1(x)}{\varphi_2(x)}\,dx\Bigr], \quad q(x) = -\frac{\varphi_0(x)}{\varphi_2(x)} \exp\Bigl[\int \frac{\varphi_1(x)}{\varphi_2(x)}\,dx\Bigr].$$

TABLE 12.2
Example estimates of the first eigenvalue λ_1 in Sturm–Liouville problems with boundary conditions of the first kind $y(0) = y(1) = 0$ obtained using the Rayleigh–Ritz principle [the right-hand side of relation (12.2.5.6)]

Equation	Test function	λ_1, approximate	λ_1, exact
$y''_{xx} + \lambda(1 + x^2)^{-2}y = 0$	$z = \sin \pi x$	15.337	15.0
$y''_{xx} + \lambda(4 - x^2)^{-2}y = 0$	$z = \sin \pi x$	135.317	134.837
$[(1 + x)^{-1}y'_x]'_x + \lambda y = 0$	$z = \sin \pi x$	7.003	6.772
$\left(\sqrt{1 + x}\, y'_x\right)'_x + \lambda y = 0$	$z = \sin \pi x$	11.9956	11.8985
$y''_{xx} + \lambda(1 + \sin \pi x)y = 0$	$z = \sin \pi x$ $z = x(1 - x)$	$0.54105\,\pi^2$ $0.55204\,\pi^2$	$0.54032\,\pi^2$ $0.54032\,\pi^2$

12.2.5-3. Problems with boundary conditions of the first kind.

Let us note some special properties of the Sturm–Liouville problem that is the first boundary value problem for equation (12.2.5.1) with the boundary conditions

$$y = 0 \quad \text{at} \quad x = x_1, \qquad y = 0 \quad \text{at} \quad x = x_2. \tag{12.2.5.5}$$

1°. For $n \to \infty$, the asymptotic relation (12.2.5.4) can be used to estimate the eigenvalues λ_n. In this case, the asymptotic formula

$$\frac{y_n(x)}{\|y_n\|} = \left[\frac{4}{\Delta^2 p(x)s(x)}\right]^{1/4} \sin\left[\frac{\pi n}{\Delta}\int_{x_1}^{x}\sqrt{\frac{s(x)}{p(x)}}\,dx\right] + O\Big(\frac{1}{n}\Big), \quad \Delta = \int_{x_1}^{x_2}\sqrt{\frac{s(x)}{p(x)}}\,dx$$

holds true for the eigenfunctions $y_n(x)$.

2°. If $q \ge 0$, the following upper estimate holds for the least eigenvalue (*Rayleigh–Ritz principle*):

$$\lambda_1 \le \frac{\displaystyle\int_{x_1}^{x_2}\left[p(x)(z'_x)^2 + q(x)z^2\right]dx}{\displaystyle\int_{x_1}^{x_2} s(x)z^2\,dx}, \tag{12.2.5.6}$$

where $z = z(x)$ is any twice differentiable function that satisfies the conditions $z(x_1) = z(x_2) = 0$. The equality in (12.2.5.6) is attained if $z = y_1(x)$, where $y_1(x)$ is the eigenfunction corresponding to the eigenvalue λ_1. One can take $z = (x - x_1)(x_2 - x)$ or $z = \sin\left[\dfrac{\pi(x - x_1)}{x_2 - x_1}\right]$ in (12.2.5.6) to obtain specific estimates.

It is significant to note that the left-hand side of (12.2.5.6) usually gives a fairly precise estimate of the first eigenvalue (see Table 12.2).

3°. The extension of the interval $[x_1, x_2]$ leads to decreasing in eigenvalues.

4°. Let the inequalities

$$0 < p_{\min} \le p(x) \le p_{\max}, \quad 0 < s_{\min} \le s(x) \le s_{\max}, \quad 0 < q_{\min} \le q(x) \le q_{\max}$$

be satisfied. Then the following bilateral estimates hold:

$$\frac{p_{\min}}{s_{\max}}\frac{\pi^2 n^2}{(x_2 - x_1)^2} + \frac{q_{\min}}{s_{\max}} \le \lambda_n \le \frac{p_{\max}}{s_{\min}}\frac{\pi^2 n^2}{(x_2 - x_1)^2} + \frac{q_{\max}}{s_{\min}}.$$

5°. In engineering calculations for eigenvalues, the approximate formula

$$\lambda_n = \frac{\pi^2 n^2}{\Delta^2} + \frac{1}{x_2 - x_1}\int_{x_1}^{x_2} \frac{q(x)}{s(x)}\,dx, \qquad \Delta = \int_{x_1}^{x_2} \sqrt{\frac{s(x)}{p(x)}}\,dx \tag{12.2.5.7}$$

may be quite useful. This formula provides an exact result if $p(x)s(x) = \text{const}$ and $q(x)/s(x) = \text{const}$ (in particular, for constant equation coefficients, $p = p_0$, $q = q_0$, and $s = s_0$) and gives a correct asymptotic behavior of (12.2.5.4) for any $p(x)$, $q(x)$, and $s(x)$. In addition, relation (12.2.5.7) gives two correct leading asymptotic terms as $n \to \infty$ if $p(x) = \text{const}$ and $s(x) = \text{const}$ [and also if $p(x)s(x) = \text{const}$].

6°. Suppose $p(x) = s(x) = 1$ and the function $q = q(x)$ has a continuous derivative. The following asymptotic relations hold for eigenvalues λ_n and eigenfunctions $y_n(x)$ as $n \to \infty$:

$$\sqrt{\lambda_n} = \frac{\pi n}{x_2 - x_1} + \frac{1}{\pi n} Q(x_1, x_2) + O\Big(\frac{1}{n^2}\Big),$$
$$y_n(x) = \sin\frac{\pi n(x - x_1)}{x_2 - x_1} - \frac{1}{\pi n}\Big[(x_1 - x)Q(x, x_2) + (x_2 - x)Q(x_1, x)\Big]\cos\frac{\pi n(x - x_1)}{x_2 - x_1} + O\Big(\frac{1}{n^2}\Big),$$

where

$$Q(u, v) = \frac{1}{2}\int_u^v q(x)\,dx. \tag{12.2.5.8}$$

7°. Let us consider the eigenvalue problem for the equation with a small parameter

$$y''_{xx} + [\lambda + \varepsilon q(x)]y = 0 \qquad (\varepsilon \to 0)$$

subject to the boundary conditions (12.2.5.5) with $x_1 = 0$ and $x_2 = 1$. We assume that $q(x) = q(-x)$.

This problem has the following eigenvalues and eigenfunctions:

$$\lambda_n = \pi^2 n^2 - \varepsilon A_{nn} + \frac{\varepsilon^2}{\pi^2}\sum_{k \neq n} \frac{A_{nk}^2}{n^2 - k^2} + O(\varepsilon^3), \qquad A_{nk} = 2\int_0^1 q(x)\sin(\pi n x)\sin(\pi k x)\,dx;$$
$$y_n(x) = \sqrt{2}\,\sin(\pi n x) - \varepsilon\frac{\sqrt{2}}{\pi^2}\sum_{k \neq n}\frac{A_{nk}}{n^2 - k^2}\sin(\pi k x) + O(\varepsilon^2).$$

Here, the summation is carried out over k from 1 to ∞. The next term in the expansion of y_n can be found in Nayfeh (1973).

12.2.5-4. Problems with boundary conditions of the second kind.

Let us note some special properties of the Sturm–Liouville problem that is the second boundary value problem for equation (12.2.5.1) with the boundary conditions

$$y'_x = 0 \quad \text{at} \quad x = x_1, \qquad y'_x = 0 \quad \text{at} \quad x = x_2.$$

1°. If $q > 0$, the upper estimate (12.2.5.6) is valid for the least eigenvalue, with $z = z(x)$ being any twice-differentiable function that satisfies the conditions $z'_x(x_1) = z'_x(x_2) = 0$. The equality in (12.2.5.6) is attained if $z = y_1(x)$, where $y_1(x)$ is the eigenfunction corresponding to the eigenvalue λ_1.

2°. Suppose $p(x) = s(x) = 1$ and the function $q = q(x)$ has a continuous derivative. The following asymptotic relations hold for eigenvalues λ_n and eigenfunctions $y_n(x)$ as $n \to \infty$:

$$\sqrt{\lambda_n} = \frac{\pi(n-1)}{x_2 - x_1} + \frac{1}{\pi(n-1)} Q(x_1, x_2) + O\Big(\frac{1}{n^2}\Big),$$

$$y_n(x) = \cos\frac{\pi(n-1)(x-x_1)}{x_2 - x_1} + \frac{1}{\pi(n-1)}\Big[(x_1 - x)Q(x, x_2) + (x_2 - x)Q(x_1, x)\Big] \sin\frac{\pi(n-1)(x-x_1)}{x_2 - x_1} + O\Big(\frac{1}{n^2}\Big),$$

where $Q(u, v)$ is given by (12.2.5.8).

12.2.5-5. Problems with boundary conditions of the third kind.

We consider the third boundary value problem for equation (12.2.5.1) subject to condition (12.2.5.2) with $\alpha_1 = \alpha_2 = 1$. We assume that $p(x) = s(x) = 1$ and the function $q = q(x)$ has a continuous derivative.

The following asymptotic formulas hold for eigenvalues λ_n and eigenfunctions $y_n(x)$ as $n \to \infty$:

$$\sqrt{\lambda_n} = \frac{\pi(n-1)}{x_2 - x_1} + \frac{1}{\pi(n-1)}\big[Q(x_1, x_2) - \beta_1 + \beta_2\big] + O\Big(\frac{1}{n^2}\Big),$$

$$y_n(x) = \cos\frac{\pi(n-1)(x-x_1)}{x_2 - x_1} + \frac{1}{\pi(n-1)}\Big\{(x_1 - x)\big[Q(x, x_2) + \beta_2\big] + (x_2 - x)\big[Q(x_1, x) - \beta_1\big]\Big\} \sin\frac{\pi(n-1)(x-x_1)}{x_2 - x_1} + O\Big(\frac{1}{n^2}\Big),$$

where $Q(u, v)$ is defined by (12.2.5.8).

12.2.5-6. Problems with mixed boundary conditions.

Let us note some special properties of the Sturm–Liouville problem that is the mixed boundary value problem for equation (12.2.5.1) with the boundary conditions

$$y'_x = 0 \quad \text{at} \quad x = x_1, \qquad y = 0 \quad \text{at} \quad x = x_2.$$

1°. If $q \geq 0$, the upper estimate (12.2.5.6) is valid for the least eigenvalue, with $z = z(x)$ being any twice-differentiable function that satisfies the conditions $z'_x(x_1) = 0$ and $z(x_2) = 0$. The equality in (12.2.5.6) is attained if $z = y_1(x)$, where $y_1(x)$ is the eigenfunction corresponding to the eigenvalue λ_1.

2°. Suppose $p(x) = s(x) = 1$ and the function $q = q(x)$ has a continuous derivative. The following asymptotic relations hold for eigenvalues λ_n and eigenfunctions $y_n(x)$ as $n \to \infty$:

$$\sqrt{\lambda_n} = \frac{\pi(2n-1)}{2(x_2 - x_1)} + \frac{2}{\pi(2n-1)} Q(x_1, x_2) + O\Big(\frac{1}{n^2}\Big),$$

$$y_n(x) = \cos\frac{\pi(2n-1)(x-x_1)}{2(x_2 - x_1)} + \frac{2}{\pi(2n-1)}\Big[(x_1 - x)Q(x, x_2) + (x_2 - x)Q(x_1, x)\Big] \sin\frac{\pi(2n-1)(x-x_1)}{2(x_2 - x_1)} + O\Big(\frac{1}{n^2}\Big),$$

where $Q(u, v)$ is defined by (12.2.5.8).

12.2.6. Theorems on Estimates and Zeros of Solutions

12.2.6-1. Theorems on estimates of solutions.

Let $f_n(x)$ and $g_n(x)$ $(n = 1, 2)$ be continuous functions on the interval $a \le x \le b$ and let the following inequalities hold:

$$0 \le f_1(x) \le f_2(x), \quad 0 \le g_1(x) \le g_2(x).$$

If $y_n = y_n(x)$ are some solutions to the linear equations

$$y_n'' = f_n(x)y_n + g_n(x) \qquad (n = 1, 2)$$

and $y_1(a) \le y_2(a)$ and $y_1'(a) \le y_2'(a)$, then $y_1(x) \le y_2(x)$ and $y_1'(x) \le y_2'(x)$ on each interval $a \le x \le a_1$, where $y_2(x) > 0$.

12.2.6-2. Sturm comparison theorem on zeros of solutions.

Consider the equation

$$[f(x)y']' + g(x)y = 0 \qquad (a \le x \le b), \tag{12.2.6.1}$$

where the function $f(x)$ is positive and continuously differentiable, and the function $g(x)$ is continuous.

THEOREM (COMPARISON, STURM). *Let $y_n = y_n(x)$ be nonzero solutions of the linear equations*

$$[f_n(x)y_n']' + g_n(x)y_n = 0 \qquad (n = 1, 2)$$

and let the inequalities $f_1(x) \ge f_2(x) > 0$ and $g_1(x) \le g_2(x)$ hold. Then the function y_2 has at least one zero lying between any two adjacent zeros, x_1 and x_2, of the function y_1 (it is assumed that the identities $f_1 \equiv f_2$ and $g_1 \equiv g_2$ are not satisfied on any interval simultaneously).

COROLLARY 1. *If $g(x) \le 0$ or there exists a constant k_1 such that*

$$f(x) \ge k_1 > 0, \quad g(x) < k_1\left(\frac{\pi}{b-a}\right)^2,$$

then every nontrivial solution to equation (12.2.6.1) has no more than one zero on the interval $[a, b]$.

COROLLARY 2. *If there exists a constant k_2 such that*

$$0 < f(x) \le k_2, \quad g(x) > k_2\left(\frac{\pi m}{b-a}\right)^2, \quad \text{where} \quad m = 1, 2, \ldots,$$

then every nontrivial solution to equation (12.2.6.1) has at least m zeros on the interval $[a, b]$.

12.2.6-3. Qualitative behavior of solutions as $x \to \infty$.

Consider the equation

$$y'' + f(x)y = 0, \tag{12.2.6.2}$$

where $f(x)$ is a continuous function for $x \geq a$.

1°. For $f(x) \leq 0$, every nonzero solution has no more than one zero, and hence $y \neq 0$ for sufficiently large x.

If $f(x) \leq 0$ for all x and $f(x) \not\equiv 0$, then $y \equiv 0$ is the only solution bounded for all x.

2°. Suppose $f(x) \geq k^2 > 0$. Then every nontrivial solution $y(x)$ and its derivative $y'(x)$ have infinitely many zeros, with the distance between any adjacent zeros remaining finite.

If $f(x) \to k^2 > 0$ for $x \to \infty$ and $f' \geq 0$, then the solutions of the equation for large x behave similarly to those of the equation $y'' + k^2 y = 0$.

3°. Let $f(x) \to -\infty$ for $|x| \to \infty$. Then every nonzero solution has only finitely many zeros, and $|y'/y| \to \infty$ as $|x| \to \infty$. There are two linearly independent solutions, y_1 and y_2, such that $y_1 \to 0$, $y_1' \to 0$, $y_2 \to \infty$, and $y_2' \to -\infty$ as $x \to -\infty$, and there are two linearly independent solutions, $\bar{y}_1$ and $\bar{y}_2$, such that $\bar{y}_1 \to 0$, $\bar{y}_1' \to 0$, $\bar{y}_2 \to \infty$, and $\bar{y}_2' \to \infty$ as $x \to \infty$.

4°. If the function f in equation (12.2.6.2) is continuous, monotonic, and positive, then the amplitude of each solution decreases (resp., increases) as f increases (resp., decreases).

12.3. Second-Order Nonlinear Differential Equations

12.3.1. Form of the General Solution. Cauchy Problem

12.3.1-1. Equations solved for the derivative. General solution.

A *second-order ordinary differential equation* solved for the highest derivative has the form

$$y''_{xx} = f(x, y, y'_x). \tag{12.3.1.1}$$

The general solution of this equation depends on two arbitrary constants, C_1 and C_2. In some cases, the general solution can be written in explicit form, $y = \varphi(x, C_1, C_2)$, but more often implicit or parametric forms of the general solution are encountered.

12.3.1-2. Cauchy problem. The existence and uniqueness theorem.

Cauchy problem: Find a solution of equation (12.3.1.1) satisfying the *initial conditions*

$$y(x_0) = y_0, \quad y'_x(x_0) = y_1. \tag{12.3.1.2}$$

(At a point $x = x_0$, the value of the unknown function, y_0, and its derivative, y_1, are prescribed.)

EXISTENCE AND UNIQUENESS THEOREM. *Let $f(x, y, z)$ be a continuous function in all its arguments in a neighborhood of a point (x_0, y_0, y_1) and let f have bounded partial derivatives f_y and f_z in this neighborhood, or the Lipschitz condition is satisfied: $|f(x, y, z) - f(x, \bar{y}, \bar{z})| \leq A\big(|y - \bar{y}| + |z - \bar{z}|\big)$, where A is some positive number. Then a solution of equation (12.3.1.1) satisfying the initial conditions (12.3.1.2) exists and is unique.*

12.3.2. Equations Admitting Reduction of Order

12.3.2-1. Equations not containing y explicitly.

In the general case, an equation that does not contain y implicitly has the form

$$F(x, y'_x, y''_{xx}) = 0. \tag{12.3.2.1}$$

Such equations remain unchanged under an arbitrary translation of the dependent variable: $y \to y + \text{const}$. The substitution $y'_x = z(x)$, $y''_{xx} = z'_x(x)$ brings (12.3.2.1) to a first-order equation: $F(x, z, z'_x) = 0$.

12.3.2-2. Equations not containing x explicitly (autonomous equations).

In the general case, an equation that does not contain x implicitly has the form

$$F(y, y'_x, y''_{xx}) = 0. \tag{12.3.2.2}$$

Such equations remain unchanged under an arbitrary translation of the independent variable: $x \to x + \text{const}$. Using the substitution $y'_x = w(y)$, where y plays the role of the independent variable, and taking into account the relations $y''_{xx} = w'_x = w'_y y'_x = w'_y w$, one can reduce (12.3.2.2) to a first-order equation: $F(y, w, ww'_y) = 0$.

Example 1. Consider the autonomous equation

$$y''_{xx} = f(y),$$

which often arises in the theory of heat and mass transfer and combustion. The change of variable $y'_x = w(y)$ leads to a separable first-order equation: $ww'_y = f(y)$. Integrating yields $w^2 = 2F(w) + C_1$, where $F(w) = \int f(w)\,dw$. Solving for w and returning to the original variable, we obtain the separable equation $y'_x = \pm\sqrt{2F(w) + C_1}$. Its general solution is expressed as

$$\int \frac{dy}{\sqrt{2F(w) + C_1}} = \pm x + C_2, \quad \text{where} \quad F(w) = \int f(w)\,dw.$$

Remark. The equation $y''_{xx} = f(y + ax^2 + bx + c)$ is reduced by the change of variable $u = y + ax^2 + bx + c$ to an autonomous equation, $u''_{xx} = f(u) + 2a$.

12.3.2-3. Equations of the form $F(ax + by, y'_x, y''_{xx}) = 0$.

Such equations are invariant under simultaneous translations of the independent and dependent variables in accordance with the rule $x \to x + bc$, $y \to y - ac$, where c is an arbitrary constant.

For $b = 0$, see equation (12.3.2.1). For $b \neq 0$, the substitution $bw = ax + by$ leads to equation (12.3.2.2): $F(bw, w'_x - a/b, w''_{xx}) = 0$.

12.3.2-4. Equations of the form $F(x, xy'_x - y, y''_{xx}) = 0$.

The substitution $w(x) = xy'_x - y$ leads to a first-order equation: $F(x, w, w'_x/x) = 0$.

12.3.2-5. Homogeneous equations.

1°. The *equations homogeneous in the independent variable* remain unchanged under scaling of the independent variable, $x \to \alpha x$, where α is an arbitrary nonzero number. In the general case, such equations can be written in the form

$$F(y, xy'_x, x^2y''_{xx}) = 0. \tag{12.3.2.3}$$

The substitution $z(y) = xy'_x$ leads to a first-order equation: $F(y, z, zz'_y - z) = 0$.

2°. The *equations homogeneous in the dependent variable* remain unchanged under scaling of the variable sought, $y \to \alpha y$, where α is an arbitrary nonzero number. In the general case, such equations can be written in the form

$$F(x, y'_x/y, y''_{xx}/y) = 0. \tag{12.3.2.4}$$

The substitution $z(x) = y'_x/y$ leads to a first-order equation: $F(x, z, z'_x + z^2) = 0$.

3°. The *equations homogeneous in both variables* are invariant under simultaneous scaling (dilatation) of the independent and dependent variables, $x \to \alpha x$ and $y \to \alpha y$, where α is an arbitrary nonzero number. In the general case, such equations can be written in the form

$$F(y/x, y'_x, xy''_{xx}) = 0. \tag{12.3.2.5}$$

The transformation $t = \ln|x|$, $w = y/x$ leads to an autonomous equation (see Paragraph 12.3.2-2): $F(w, w'_t + w, w''_{tt} + w'_t) = 0$.

Example 2. The homogeneous equation

$$xy''_{xx} - y'_x = f(y/x)$$

is reduced by the transformation $t = \ln|x|$, $w = y/x$ to the autonomous form: $w''_{tt} = f(w) + w$. For solution of this equation, see Example 1 in Paragraph 12.3.2-2 (the notation of the right-hand side has to be changed there).

12.3.2-6. Generalized homogeneous equations.

1°. The *generalized homogeneous equations* remain unchanged under simultaneous scaling of the independent and dependent variables in accordance with the rule $x \to \alpha x$ and $y \to \alpha^k y$, where α is an arbitrary nonzero number and k is some number. Such equations can be written in the form

$$F(x^{-k}y, x^{1-k}y'_x, x^{2-k}y''_{xx}) = 0. \tag{12.3.2.6}$$

The transformation $t = \ln x$, $w = x^{-k}y$ leads to an autonomous equation (see Paragraph 12.3.2-2):

$$F\big(w, w'_t + kw, w''_{tt} + (2k-1)w'_t + k(k-1)w\big) = 0.$$

2°. The most general form of representation of generalized homogeneous equations is as follows:

$$\mathcal{F}(x^n y^m, xy'_x/y, x^2y''_{xx}/y) = 0. \tag{12.3.2.7}$$

The transformation $z = x^n y^m$, $u = xy'_x/y$ brings this equation to the first-order equation

$$\mathcal{F}\big(z, u, z(mu + n)u'_z - u + u^2\big) = 0.$$

Remark. For $m \neq 0$, equation (12.3.2.7) is equivalent to equation (12.3.2.6) in which $k = -n/m$. To the particular values $n = 0$ and $m = 0$ there correspond equations (12.3.2.3) and (12.3.2.4) homogeneous in the independent and dependent variables, respectively. For $n = -m \neq 0$, we have an equation homogeneous in both variables, which is equivalent to equation (12.3.2.5).

12.3.2-7. Equations invariant under scaling–translation transformations.

1°. The equations of the form

$$F(e^{\lambda x}y, e^{\lambda x}y'_x, e^{\lambda x}y''_{xx}) = 0 \tag{12.3.2.8}$$

remain unchanged under simultaneous translation and scaling of variables, $x \to x + \alpha$ and $y \to \beta y$, where $\beta = e^{-\alpha\lambda}$ and α is an arbitrary number. The substitution $w = e^{\lambda x}y$ brings (12.3.2.8) to an autonomous equation: $F(w,\ w'_x - \lambda w,\ w''_{xx} - 2\lambda w'_x + \lambda^2 w) = 0$ (see Paragraph 12.3.2-2).

2°. The equation

$$F(e^{\lambda x}y^n, y'_x/y, y''_{xx}/y) = 0 \tag{12.3.2.9}$$

is invariant under the simultaneous translation and scaling of variables, $x \to x + \alpha$ and $y \to \beta y$, where $\beta = e^{-\alpha\lambda/n}$ and α is an arbitrary number. The transformation $z = e^{\lambda x}y^n$, $w = y'_x/y$ brings (12.3.2.9) to a first-order equation: $F\big(z, w, z(nw + \lambda)w'_z + w^2\big) = 0$.

3°. The equation

$$F(x^n e^{\lambda y}, xy'_x, x^2 y''_{xx}) = 0 \tag{12.3.2.10}$$

is invariant under the simultaneous scaling and translation of variables, $x \to \alpha x$ and $y \to y + \beta$, where $\alpha = e^{-\beta\lambda/n}$ and β is an arbitrary number. The transformation $z = x^n e^{\lambda y}$, $w = xy'_x$ brings (12.3.2.10) to a first-order equation: $F\big(z, w, z(\lambda w + n)w'_z - w\big) = 0$.

► *Some other second-order nonlinear equations are treated in Section T5.3.*

12.3.2-8. Exact second-order equations.

The second-order equation

$$F(x, y, y'_x, y''_{xx}) = 0 \tag{12.3.2.11}$$

is said to be exact if it is the total differential of some function, $F = \varphi'_x$, where $\varphi = \varphi(x, y, y'_x)$. If equation (12.3.2.11) is exact, then we have a first-order equation for y:

$$\varphi(x, y, y'_x) = C, \tag{12.3.2.12}$$

where C is an arbitrary constant.

If equation (12.3.2.11) is exact, then $F(x, y, y'_x, y''_{xx})$ must have the form

$$F(x, y, y'_x, y''_{xx}) = f(x, y, y'_x)y''_{xx} + g(x, y, y'_x). \tag{12.3.2.13}$$

Here, f and g are expressed in terms of φ by the formulas

$$f(x, y, y'_x) = \frac{\partial\varphi}{\partial y'_x}, \qquad g(x, y, y'_x) = \frac{\partial\varphi}{\partial x} + \frac{\partial\varphi}{\partial y}y'_x. \tag{12.3.2.14}$$

By differentiating (12.3.2.14) with respect to x, y, and $p = y'_x$, we eliminate the variable φ from the two formulas in (12.3.2.14). As a result, we have the following test relations for f and g:

$$\begin{aligned} f_{xx} + 2pf_{xy} + p^2 f_{yy} &= g_{xp} + pg_{yp} - g_y, \\ f_{xp} + pf_{yp} + 2f_y &= g_{pp}. \end{aligned} \tag{12.3.2.15}$$

Here, the subscripts denote the corresponding partial derivatives.

If conditions (12.3.2.15) hold, then equation (12.3.2.11) with F of (12.3.2.13) is exact. In this case, we can integrate the first equation in (12.3.2.14) with respect to $p = y'_x$ to determine $\varphi = \varphi(x, y, y'_x)$:

$$\varphi = \int f(x, y, p)\, dp + \psi(x, y), \tag{12.3.2.16}$$

where $\psi(x, y)$ is an arbitrary function of integration. This function is determined by substituting (12.3.2.16) into the second equation in (12.3.2.14).

Example 3. The left-hand side of the equation

$$yy''_{xx} + (y'_x)^2 + 2axyy'_x + ay^2 = 0 \tag{12.3.2.17}$$

can be represented in the form (12.3.2.13), where $f = y$ and $g = p^2 + 2axyp + ay^2$. It is easy to verify that conditions (12.3.2.15) are satisfied. Hence, equation (12.3.2.17) is exact. Using (12.3.2.16), we obtain

$$\varphi = yp + \psi(x, y). \tag{12.3.2.18}$$

Substituting this expression into the second equation in (12.3.2.14) and taking into account the relation $g = p^2 + 2axyp + ay^2$, we find that $2axyp + ay^2 = \psi_x + p\psi_y$. Since $\psi = \psi(x, y)$, we have $2axy = \psi_y$ and $ay^2 = \psi_x$. Integrating yields $\psi = axy^2 + \text{const}$. Substituting this expression into (12.3.2.18) and taking into account relation (12.3.2.12), we find a first integral of equation (12.3.2.17):

$$yp + axy^2 = C_1, \qquad \text{where} \quad p = y'_x.$$

Setting $w = y^2$, we arrive at the first-order linear equation $w'_x + 2axw = 2C_1$, which is easy to integrate. Thus, we find the solution of the original equation in the form:

$$y^2 = 2C_1 \exp\bigl(-ax^2\bigr) \int \exp\bigl(ax^2\bigr)\, dx + C_2 \exp\bigl(-ax^2\bigr).$$

12.3.3. Methods of Regular Series Expansions with Respect to the Independent Variable

12.3.3-1. Method of expansion in powers of the independent variable.

A solution of the Cauchy problem

$$y''_{xx} = f(x, y, y'_x), \tag{12.3.3.1}$$

$$y(x_0) = y_0, \quad y'_x(x_0) = y_1 \tag{12.3.3.2}$$

can be sought in the form of a Taylor series in powers of the difference $(x - x_0)$, specifically:

$$y(x) = y(x_0) + y'_x(x_0)(x - x_0) + \frac{y''_{xx}(x_0)}{2!}(x - x_0)^2 + \frac{y'''_{xxx}(x_0)}{3!}(x - x_0)^3 + \cdots. \tag{12.3.3.3}$$

The first two coefficients $y(x_0)$ and $y'_x(x_0)$ in solution (12.3.3.3) are defined by the initial conditions (12.3.3.2). The values of the subsequent derivatives of y at the point $x = x_0$ are determined from equation (12.3.3.1) and its derivative equations (obtained by successive differentiation of the equation) taking into account the initial conditions (12.3.3.2). In particular, setting $x = x_0$ in (12.3.3.1) and substituting (12.3.3.2), we obtain the value of the second derivative:

$$y''_{xx}(x_0) = f(x_0, y_0, y_1). \tag{12.3.3.4}$$

Further, differentiating (12.3.3.1) yields

$$y'''_{xxx} = f_x(x, y, y'_x) + f_y(x, y, y'_x)y'_x + f_{y'_x}(x, y, y'_x)y''_{xx}. \tag{12.3.3.5}$$

On substituting $x = x_0$, the initial conditions (12.3.3.2), and the expression of $y''_{xx}(x_0)$ of (12.3.3.4) into the right-hand side of equation (12.3.3.5), we calculate the value of the third derivative:

$$y'''_{xxx}(x_0) = f_x(x_0, y_0, y_1) + f_y(x_0, y_0, y_1)y_1 + f(x_0, y_0, y_1)f_{y'_x}(x_0, y_0, y_1).$$

The subsequent derivatives of the unknown are determined likewise.

The thus obtained solution (12.3.3.3) can only be used in a small neighborhood of the point $x = x_0$.

Example 1. Consider the following Cauchy problem for a second-order nonlinear equation:

$$y''_{xx} = yy'_x + y^3; \tag{12.3.3.6}$$

$$y(0) = y'_x(0) = 1. \tag{12.3.3.7}$$

Substituting the initial values of the unknown and its derivative (12.3.3.7) into equation (12.3.3.6) yields the initial value of the second derivative:

$$y''_{xx}(0) = 2. \tag{12.3.3.8}$$

Differentiating equation (12.3.3.6) gives

$$y'''_{xxx} = yy''_{xx} + (y'_x)^2 + 3y^2y'_x. \tag{12.3.3.9}$$

Substituting here the initial values from (12.3.3.7) and (12.3.3.8), we obtain the initial condition for the third derivative:

$$y'''_{xxx}(0) = 6. \tag{12.3.3.10}$$

Differentiating (12.3.3.9) followed by substituting (12.3.3.7), (12.3.3.8), and (12.3.3.10), we find that

$$y''''_{xxxx}(0) = 24. \tag{12.3.3.11}$$

On substituting the initial data (12.3.3.7), (12.3.3.8), (12.3.3.10), and (12.3.3.11) into (12.3.3.3), we arrive at the Taylor series expansion of the solution about $x = 0$:

$$y = 1 + x + x^2 + x^3 + x^4 + \cdots. \tag{12.3.3.12}$$

This geometric series is convergent only for $|x| < 1$.

12.3.3-2. Padé approximants.

Suppose the $k + 1$ leading coefficients in the Taylor series expansion of a solution to a differential equation about the point $x = 0$ are obtained by the method presented in Paragraph 12.3.3-1, so that

$$y_{k+1}(x) = a_0 + a_1x + \cdots + a_kx^k. \tag{12.3.3.13}$$

The partial sum (12.3.3.13) pretty well approximates the solution at small x but is poor for intermediate and large values of x, since the series can be slowly convergent or even divergent. This is also related to the fact that $y_k \to \infty$ as $x \to \infty$, while the exact solution can well be bounded.

In many cases, instead of the expansion (12.3.3.13), it is reasonable to consider a Padé approximant $P_M^N(x)$, which is the ratio of two polynomials of degree N and M, specifically,

$$P_M^N(x) = \frac{A_0 + A_1x + \cdots + A_Nx^N}{1 + B_1x + \cdots + B_Mx^M}, \qquad \text{where} \quad N + M = k. \tag{12.3.3.14}$$

The coefficients $A_1, \ldots, A_N$ and $B_1, \ldots, B_M$ are selected so that the $k+1$ leading terms in the Taylor series expansion of (12.3.3.14) coincide with the respective terms of the expansion (12.3.3.13). In other words, the expansions (12.3.3.13) and (12.3.3.14) must be asymptotically equivalent as $x \to 0$.

In practice, one usually takes $N = M$ (the diagonal sequence). It often turns out that formula (12.3.3.14) pretty well approximates the exact solution on the entire range of x (for sufficiently large N).

Example 2. Consider the Cauchy problem (12.3.3.6)–(12.3.3.7) again. The Taylor series expansion of the solution about $x = 0$ has the form (12.3.3.12). This geometric series is convergent only for $|x| < 1$.

The diagonal sequence of Padé approximants corresponding to series (12.3.3.12) is

$$P_1^1(x) = \frac{1}{1-x}, \quad P_2^2(x) = \frac{1}{1-x}, \quad P_3^3(x) = \frac{1}{1-x}. \tag{12.3.3.15}$$

It is not difficult to verify that the function $y(x) = \dfrac{1}{1-x}$ is the exact solution of the Cauchy problem (12.3.3.6)–(12.3.3.7). Hence, in this case, the diagonal sequence of Padé approximants recovers the exact solution from only a few terms in the Taylor series.

Example 3. Consider the Cauchy problem for a second-order nonlinear equation:

$$y''_{xx} = 2yy'_x; \qquad y(0) = 0, \quad y'_x(0) = 1. \tag{12.3.3.16}$$

Following the method presented in Paragraph 12.3.3-1, we obtain the Taylor series expansion of the solution to problem (12.3.3.16) in the form

$$y(x) = x + \tfrac{1}{3}x^3 + \tfrac{2}{15}x^5 + \tfrac{17}{315}x^7 + \cdots. \tag{12.3.3.17}$$

The exact solution of problem (12.3.3.16) is given by $y(x) = \tan x$. Hence it has singularities at $x = \pm\frac{1}{2}(2n+1)\pi$. However, any finite segment of the Taylor series (12.3.3.17) does not have any singularities.

With series (12.3.3.17), we construct the diagonal sequence of Padé approximants:

$$P_2^2(x) = \frac{3x}{3-x^2}, \quad P_3^3(x) = \frac{x(x^2-15)}{3(2x^2-5)}, \quad P_4^4(x) = \frac{5x(21-2x^2)}{x^4-45x^2+105}. \tag{12.3.3.18}$$

These Padé approximants have singularities (at the points where the denominators vanish):

$$\begin{aligned} &x \simeq \pm 1.732 && \text{for } P_2^2(x),\\ &x \simeq \pm 1.581 && \text{for } P_3^3(x),\\ &x \simeq \pm 1.571 \text{ and } x \simeq \pm 6.522 && \text{for } P_4^4(x). \end{aligned}$$

It is apparent that the Padé approximants are attempting to recover the singularities of the exact solution at $x = \pm\pi/2$ and $x = \pm 3\pi/2$.

In Fig. 12.2, the solid line shows the exact solution of problem (12.3.3.16), the dashed line corresponds to the four-term Taylor series solution (12.3.3.17), and the dot-and-dash line depicts the Padé approximants (12.3.3.18). It is evident that the Padé approximant $P_4^4(x)$ gives an accurate numerical approximation of the exact solution on the interval $|x| \le 2$; everywhere the error is less than 1%, except for a very small neighborhood of the point $x = \pm\pi/2$ (the error is 1% for $x = \pm 1.535$ and 0.84% for $x = \pm 2$).

12.3.4. Movable Singularities of Solutions of Ordinary Differential Equations. Painlevé Transcendents

12.3.4-1. Preliminary remarks. Singular points of solutions.

1°. Singular points of solutions to ordinary differential equations can be *fixed* or *movable*. The coordinates of fixed singular points remain the same for different solutions of an equation.* The coordinates of movable singular points vary depending on the particular solution selected (i.e., they depend on the initial conditions).

* Solutions of linear ordinary differential equations can only have fixed singular points, and their positions are determined by the singularities of the equation coefficients.

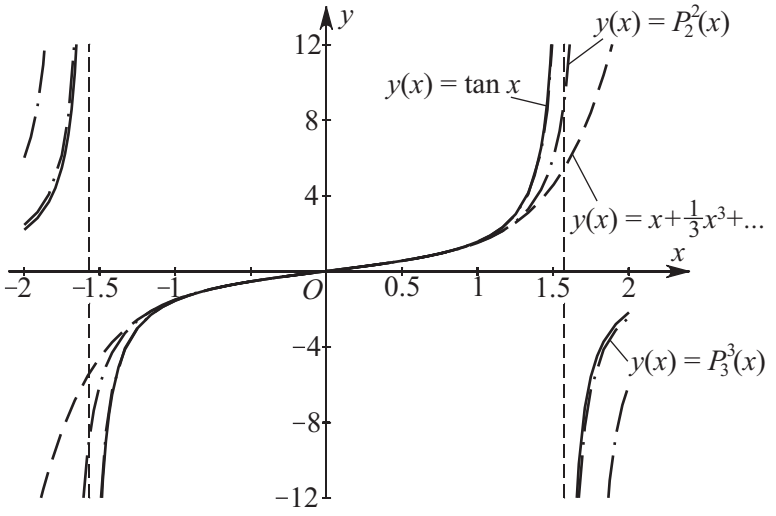


Figure 12.2. Comparison of the exact solution to problem (12.3.3.16) with the approximate truncated series solution (12.3.3.17) and associated Padé approximants (12.3.3.18).

Listed below are simple examples of first-order ordinary differential equations and their solutions having movable singularities:

Equation	*Solution*	*Solution's singularity type*
$y'_z = -y^2$	$y = 1/(z - z_0)$	movable pole
$y'_z = 1/y$	$y = 2\sqrt{z - z_0}$	algebraic branch point
$y'_z = e^{-y}$	$y = \ln(z - z_0)$	logarithmic branch point
$y'_z = -y \ln^2 y$	$y = \exp[1/(z - z_0)]$	essential singularity

Algebraic branch points, logarithmic branch points, and essential singularities are called *movable critical points*.

2°. The Painlevé equations arise from the classification of the following second-order differential equations over the complex plane:

$$y''_{zz} = R(z, y, y'_z),$$

where $R = R(z, y, w)$ is a function rational in y and w and analytic in z. It was shown by P. Painlevé (1897–1902) and B. Gambier (1910) that all equations of this type whose solutions do not have movable critical points (but are allowed to have fixed singular points and movable poles) can be reduced to 50 classes of equations. Moreover, 44 classes out of them are integrable by quadrature or admit reduction of order. The remaining 6 equations are irreducible; these are known as the *Painlevé equations* or *Painlevé transcendents*, and their solutions are known as the *Painlevé transcendental functions*.

The canonical forms of the Painlevé transcendents are given below in Paragraphs 12.3.4-2 through 12.3.4-6. Solutions of the first, second, and fourth Painlevé transcendents have movable poles (no fixed singular points). Solutions of the third and fifth Painlevé transcendents have two fixed logarithmic branch points, $z = 0$ and $z = \infty$. Solutions of the sixth Painlevé transcendent have three fixed logarithmic branch points, $z = 0$, $z = 1$, and $z = \infty$.

It is significant that the Painlevé equations often arise in mathematical physics.

12.3.4-2. First Painlevé transcendent.

1°. The *first Painlevé transcendent* has the form

$$y''_{zz} = 6y^2 + z. \tag{12.3.4.1}$$

The solutions of the first Painlevé transcendent are single-valued functions of z.

The solutions of equation (12.3.4.1) can be presented, in the vicinity of movable pole z_{p}, in terms of the series

$$y = \frac{1}{(z-z_{\mathrm{p}})^2} + \sum_{n=2}^{\infty} a_n (z-z_{\mathrm{p}})^n,$$

$$a_2 = -\tfrac{1}{10} z_{\mathrm{p}}, \quad a_3 = -\tfrac{1}{6}, \quad a_4 = C, \quad a_5 = 0, \quad a_6 = \tfrac{1}{300} z_{\mathrm{p}}^2,$$

where z_{p} and C are arbitrary constants; the coefficients a_j $(j \geq 7)$ are uniquely defined in terms of z_{p} and C.

2°. In a neighborhood of a fixed point $z = z_0$, the solution of the Cauchy problem for the first Painlevé transcendent (12.3.4.1) can be represented by the Taylor series (see Paragraph 12.3.3-1):

$$y = A + B(z-z_0) + \tfrac{1}{2}(6A^2 + z_0)(z-z_0)^2 + \tfrac{1}{6}(12AB+1)(z-z_0)^3 + \tfrac{1}{2}(6A^3 + B^2 + Az_0)(z-z_0)^4 + \cdots,$$

where A and B are initial data of the Cauchy problem, so that $y|_{z=z_0} = A$ and $y'_z|_{z=z_0} = B$.

Remark. The solutions of the Cauchy problems for the second and fourth Painlevé transcendents can be expressed likewise (fixed singular points should be excluded from consideration for the remaining Painlevé transcendents).

3°. For large values of $|z|$, the following asymptotic formula holds:

$$y \sim z^{1/2} \wp\big(\tfrac{4}{5} z^{5/4} - a;\ 12,\ b\big),$$

where the elliptic Weierstrass function $\wp(\zeta; 12, b)$ is defined implicitly by the integral

$$\zeta = \int \frac{d\wp}{\sqrt{4\wp^3 - 12\wp - b}};$$

a and b are some constants.

12.3.4-3. Second Painlevé transcendent.

1°. The *second Painlevé transcendent* has the form

$$y''_{zz} = 2y^3 + zy + \alpha. \tag{12.3.4.2}$$

The solutions of the second Painlevé transcendent are single-valued functions of z.

The solutions of equation (12.3.4.2) can be represented, in the vicinity of a movable pole z_{p}, in terms of the series

$$y = \frac{m}{z-z_{\mathrm{p}}} + \sum_{n=1}^{\infty} b_n (z-z_{\mathrm{p}})^n,$$

$$b_1 = -\tfrac{1}{6} m z_{\mathrm{p}}, \quad b_2 = -\tfrac{1}{4}(m+\alpha), \quad b_3 = C, \quad b_4 = \tfrac{1}{72} z_{\mathrm{p}}(m+3\alpha),$$

$$b_5 = \tfrac{1}{3024}\big[(27 + 81\alpha^2 - 2z_{\mathrm{p}}^3)m + 108\alpha - 216 C z_{\mathrm{p}}\big],$$

where z_{p} and C are arbitrary constants, $m = \pm 1$, and the coefficients b_n $(n \geq 6)$ are uniquely defined in terms of z_{p} and C.

2°. For fixed α, denote the solution by $y(z,\alpha)$. Then the following relation holds:

$$y(z,-\alpha) = -y(z,\alpha), \tag{12.3.4.3}$$

while the solutions $y(z,\alpha)$ and $y(z,\alpha-1)$ are related by the Bäcklund transformations:

$$\begin{aligned} y(z,\alpha-1) &= -y(z,\alpha) + \frac{2\alpha-1}{2y'_z(z,\alpha) - 2y^2(z,\alpha) - z}, \\ y(z,\alpha) &= -y(z,\alpha-1) - \frac{2\alpha-1}{2y'_z(z,\alpha-1) + 2y^2(z,\alpha-1) + z}. \end{aligned} \tag{12.3.4.4}$$

Therefore, in order to study the general solution of equation (12.3.4.2) with arbitrary α, it is sufficient to construct the solution for all α out of the band $0 \le \operatorname{Re}\alpha < \frac{1}{2}$.

Three solutions corresponding to α and $\alpha \pm 1$ are related by the rational formulas

$$y_{\alpha+1} = -\frac{(y_{\alpha-1}+y_\alpha)(4y_\alpha^3 + 2zy_\alpha + 2\alpha + 1) + (2\alpha-1)y_\alpha}{2(y_{\alpha-1}+y_\alpha)(2y_\alpha^2 + z) + 2\alpha - 1},$$

where y_α stands for $y(z,\alpha)$.

The solutions $y(z,\alpha)$ and $y(z,-\alpha-1)$ are related by the Bäcklund transformations:

$$\begin{aligned} y(z,-\alpha-1) &= y(z,\alpha) + \frac{2\alpha+1}{2y'_z(z,\alpha) + 2y^2(z,\alpha) + z}, \\ y(z,\alpha) &= y(z,-\alpha-1) - \frac{2\alpha+1}{2y'_z(z,-\alpha-1) + 2y^2(z,-\alpha-1) + z}. \end{aligned}$$

3°. For $\alpha = 0$, equation (12.3.4.2) has the trivial solution $y = 0$. Taking into account this fact and relations (12.3.4.3) and (12.3.4.4), we find that the second Painlevé transcendent with $\alpha = \pm 1, \pm 2, \ldots$ has the rational particular solutions

$$y(z,\pm 1) = \mp\frac{1}{z}, \quad y(z,\pm 2) = \pm\Big(\frac{1}{z} - \frac{3z^2}{z^3+4}\Big), \quad \ldots$$

For $\alpha = \frac{1}{2}$, equation (12.3.4.2) admits the one-parameter family of solutions:

$$y(z,\tfrac{1}{2}) = -\frac{w'_z}{w}, \qquad \text{where} \quad w = \sqrt{z}\,\Big[C_1 J_{1/3}\big(\tfrac{\sqrt{2}}{3}z^{3/2}\big) + C_2 Y_{1/3}\big(\tfrac{\sqrt{2}}{3}z^{3/2}\big)\Big]. \tag{12.3.4.5}$$

(Here, the function w is a solution of the Airy equation, $w''_{zz} + \frac{1}{2}zw = 0$.) It follows from (12.3.4.3)–(12.3.4.5) that the second Painlevé transcendent for all $\alpha = n + \frac{1}{2}$ with $n = 0, \pm 1, \pm 2, \ldots$ has a one-parameter family of solutions that can be expressed in terms of Bessel functions.

12.3.4-4. Third Painlevé transcendent.

1°. The *third Painlevé transcendent* has the form

$$y''_{zz} = \frac{(y'_z)^2}{y} - \frac{y'_z}{z} + \frac{1}{z}(\alpha y^2 + \beta) + \gamma y^3 + \frac{\delta}{y}. \tag{12.3.4.6}$$

In terms of the new independent variable ζ defined by $z = e^\zeta$, the solutions of the transformed equation will be single-valued functions of ζ.

In some special cases, equation (12.3.4.6) can be integrated by quadrature.

2°. Any solution of the Riccati equation

$$y'_z = ky^2 + \frac{\alpha - k}{kz}y + c, \tag{12.3.4.7}$$

where $k^2 = \gamma,\ c^2 = -\delta,\ k\beta + c(\alpha - 2k) = 0,$ is a solution of equation (12.3.4.6). Substituting $z = \lambda\tau,\ y = -\dfrac{u'_z}{ku},$ where $\lambda^2 = \dfrac{1}{kc},$ into (12.3.4.7), we obtain a linear equation

$$u''_{\tau\tau} + \frac{k - \alpha}{k\tau}u'_\tau + u = 0,$$

whose general solution is expressed in terms of Bessel functions:

$$u = \tau^{\frac{\alpha}{2k}}\Big[C_1 J_{\frac{\alpha}{2k}}(\tau) + C_2 Y_{\frac{\alpha}{2k}}(\tau)\Big].$$

12.3.4-5. Fourth Painlevé transcendent.

1°. The *fourth Painlevé transcendent* has the form

$$y''_{zz} = \frac{(y'_z)^2}{2y} + \frac{3}{2}y^3 + 4zy^2 + 2(z^2 - \alpha)y + \frac{\beta}{y}. \tag{12.3.4.8}$$

The solutions of the fourth Painlevé transcendent are single-valued functions of z.

The Laurent-series expansion of the solution of equation (12.3.4.8) in the vicinity of a movable pole z_{p} is given by

$$y = \frac{m}{z - z_{\mathrm{p}}} - z_{\mathrm{p}} - \frac{m}{3}(z_{\mathrm{p}}^2 + 2\alpha - 4m)(z - z_{\mathrm{p}}) + C(z - z_{\mathrm{p}})^2 + \sum_{j=3}^{\infty} a_j (z - z_{\mathrm{p}})^j,$$

where $m = \pm 1$; z_{p} and C are arbitrary constants; and the a_j $(j \geq 3)$ are uniquely defined in terms of $\alpha,\ \beta,\ z_{\mathrm{p}},$ and C.

2°. Two solutions of equation (12.3.4.8) corresponding to different values of the parameters α and β are related to each other by the Bäcklund transformations:

$$\widetilde{y} = \frac{1}{2sy}(y'_z - q - 2szy - sy^2), \qquad q^2 = -2\beta,$$
$$y = -\frac{1}{2s\widetilde{y}}(\widetilde{y}'_z - p + 2sz\widetilde{y} + s\widetilde{y}^2), \qquad p^2 = -2\widetilde{\beta},$$
$$2\beta = -(\widetilde{\alpha}s - 1 - \tfrac{1}{2}p)^2, \qquad 4\alpha = -2s - 2\widetilde{\alpha} - 3sp,$$

where $y = y(z, \alpha, \beta),\ \widetilde{y} = \widetilde{y}(z, \widetilde{\alpha}, \widetilde{\beta}),$ and s is an arbitrary parameter.

3°. If the condition $\beta + 2(1 + \alpha m)^2 = 0,$ where $m = \pm 1,$ is satisfied, then every solution of the Riccati equation

$$y'_z = my^2 + 2mzy - 2(1 + \alpha m)$$

is simultaneously a solution of the fourth Painlevé equation (12.3.4.8).

12.3.4-6. Fifth and sixth Painlevé transcendents.

1°. The *fifth Painlevé transcendent* has the form

$$y''_{zz} = \frac{3y-1}{2y(y-1)}(y'_z)^2 - \frac{y'_z}{z} + \frac{(y-1)^2}{z^2}\Big(\alpha y + \frac{\beta}{y}\Big) + \gamma\frac{y}{z} + \frac{\delta y(y+1)}{y-1}.$$

2°. The *sixth Painlevé transcendent* has the form

$$y''_{zz} = \frac{1}{2}\Big(\frac{1}{y} + \frac{1}{y-1} + \frac{1}{y-z}\Big)(y'_z)^2 - \Big(\frac{1}{z} + \frac{1}{z-1} + \frac{1}{y-z}\Big)y'_z + \frac{y(y-1)(y-z)}{z^2(z-1)^2}\Big[\alpha + \beta\frac{z}{y^2} + \gamma\frac{z-1}{(y-1)^2} + \delta\frac{z(z-1)}{(y-z)^2}\Big].$$

For details about these equations, see the list of references given at the end of the current chapter.

12.3.5. Perturbation Methods of Mechanics and Physics

12.3.5-1. Preliminary remarks. A summary table of basic methods.

Perturbation methods are widely used in nonlinear mechanics and theoretical physics for solving problems that are described by differential equations with a small parameter ε. The primary purpose of these methods is to obtain an approximate solution that would be equally suitable at all (small, intermediate, and large) values of the independent variable as $\varepsilon \to 0$.

Equations with a small parameter can be classified according to the following:

(i) the order of the equation remains the same at $\varepsilon = 0$;
(ii) the order of the equation reduces at $\varepsilon = 0$.

For the first type of equations, solutions of related problems* are sufficiently smooth (little varying as ε decreases). The second type of equation is said to be degenerate at $\varepsilon = 0$, or singularly perturbed. In related problems, thin boundary layers usually arise whose thickness is significantly dependent on ε; such boundary layers are characterized by high gradients of the unknown.

All perturbation methods have a limited domain of applicability; the possibility of using one or another method depends on the type of equations or problems involved. The most commonly used methods are summarized in Table 12.3 (the method of regular series expansions is set out in Paragraph 12.3.5-2). In subsequent paragraphs, additional remarks and specific examples are given for some of the methods. In practice, one usually confines oneself to few leading terms of the asymptotic expansion.

In many problems of nonlinear mechanics and theoretical physics, the independent variable is dimensionless time t. Therefore, in this subsection we use the conventional t $(0 \le t < \infty)$ instead of x.

* Further on, we assume that the initial and/or boundary conditions are independent of the parameter ε.

TABLE 12.3
Perturbation methods of nonlinear mechanics and theoretical physics
(the third column gives n leading asymptotic terms with respect to the small parameter ε).

Method name	Examples of problems solved by the method	Form of the solution sought	Additional conditions and remarks
Method of scaled parameters ($0 \le t < \infty$)	One looks for periodic solutions of the equation $y''_{tt} + \omega_0^2 y = \varepsilon f(y, y'_t)$; see also Paragraph 12.3.5-3	$y(t) = \sum_{k=0}^{n-1} \varepsilon^k y_k(z)$, $t = z\left(1 + \sum_{k=1}^{n-1} \varepsilon^k \omega_k\right)$	Unknowns: y_k and ω_k; $y_{k+1}/y_k = O(1)$; secular terms are eliminated through selection of the constants ω_k
Method of strained coordinates ($0 \le t < \infty$)	Cauchy problem: $y'_t = f(t, y, \varepsilon)$; $y(t_0) = y_0$ (f is of a special form); see also the problem in the method of scaled parameters	$y(t) = \sum_{k=0}^{n-1} \varepsilon^k y_k(z)$, $t = z + \sum_{k=1}^{n-1} \varepsilon^k \varphi_k(z)$	Unknowns: y_k and φ_k; $y_{k+1}/y_k = O(1)$, $\varphi_{k+1}/\varphi_k = O(1)$
Averaging method ($0 \le t < \infty$)	Cauchy problem: $y''_{tt} + \omega_0^2 y = \varepsilon f(y, y'_t)$, $y(0) = y_0$, $y'_t(0) = y_1$; for more general problems, see Paragraph 12.3.5-4, Item 2°	$y = a(t)\cos\varphi(t)$, the amplitude a and phase φ are governed by the equations $\frac{da}{dt} = -\frac{\varepsilon}{\omega_0} f_s(a)$, $\frac{d\varphi}{dt} = \omega_0 - \frac{\varepsilon}{a\omega_0} f_c(a)$	Unknowns: a and φ; $f_s = \frac{1}{2\pi}\int_0^{2\pi} \sin\varphi F\, d\varphi$, $f_c = \frac{1}{2\pi}\int_0^{2\pi} \cos\varphi F\, d\varphi$, $F = f(a\cos\varphi, -a\omega_0 \sin\varphi)$
Krylov–Bogolyubov–Mitropolskii method ($0 \le t < \infty$)	One looks for periodic solutions of the equation $y''_{tt} + \omega_0^2 y = \varepsilon f(y, y'_t)$; Cauchy problem for this and other equations	$y = a\cos\varphi + \sum_{k=1}^{n-1} \varepsilon^k y_k(a, \varphi)$, a and φ are determined by the equations $\frac{da}{dt} = \sum_{k=1}^{n} \varepsilon^k A_k(a)$, $\frac{d\varphi}{dt} = \omega_0 + \sum_{k=1}^{n} \varepsilon^k \Phi_k(a)$	Unknowns: y_k, A_k, Φ_k; y_k are 2π-periodic functions of φ; the y_k are assumed not to contain $\cos\varphi$
Method of two-scale expansions ($0 \le t < \infty$)	Cauchy problem: $y''_{tt} + \omega_0^2 y = \varepsilon f(y, y'_t)$, $y(0) = y_0$, $y'_t(0) = y_1$; for boundary value problems, see Paragraph 12.3.5-5, Item 2°	$y = \sum_{k=0}^{n-1} \varepsilon^k y_k(\xi, \eta)$, where $\xi = \varepsilon t$, $\eta = t\left(1 + \sum_{k=2}^{n-1} \varepsilon^k \omega_k\right)$, $\frac{d}{dt} = \varepsilon \frac{\partial}{\partial \xi} + \left(1 + \varepsilon^2 \omega_2 + \cdots\right)\frac{\partial}{\partial \eta}$	Unknowns: y_k and ω_k; $y_{k+1}/y_k = O(1)$; secular terms are eliminated through selection of ω_k
Multiple scales method ($0 \le t < \infty$)	One looks for periodic solutions of the equation $y''_{tt} + \omega_0^2 y = \varepsilon f(y, y'_t)$; Cauchy problem for this and other equations	$y = \sum_{k=0}^{n-1} \varepsilon^k y_k$, where $y_k = y_k(T_0, T_1, \ldots, T_n)$, $T_k = \varepsilon^k t$, $\frac{d}{dt} = \frac{\partial}{\partial T_0} + \varepsilon \frac{\partial}{\partial T_1} + \cdots + \varepsilon^n \frac{\partial}{\partial T_n}$	Unknowns: y_k; $y_{k+1}/y_k = O(1)$; for $n = 1$, this method is equivalent to the averaging method
Method of matched asymptotic expansions ($0 \le x \le b$)	Boundary value problem: $\varepsilon y''_{xx} + f(x, y) y'_x = g(x, y)$, $y(0) = y_0$, $y(b) = y_b$ (f assumed positive); for other problems, see Paragraph 12.3.5-6, Item 2°	Outer expansion: $y = \sum_{k=0}^{n-1} \sigma_k(\varepsilon) y_k(x)$, $O(\varepsilon) \le x \le b$; inner expansion ($z = x/\varepsilon$): $\widetilde{y} = \sum_{k=0}^{n-1} \widetilde{\sigma}_k(\varepsilon) \widetilde{y}_k(z)$, $0 \le x \le O(\varepsilon)$	Unknowns: y_k, $\widetilde{y}_k$, σ_k, $\widetilde{\sigma}_k$; $y_{k+1}/y_k = O(1)$, $\widetilde{y}_{k+1}/\widetilde{y}_k = O(1)$; the procedure of matching expansions is used: $y(x \to 0) = \widetilde{y}(z \to \infty)$
Method of composite expansions ($0 \le x \le b$)	Boundary value problem: $\varepsilon y''_{xx} + f(x, y) y'_x = g(x, y)$, $y(0) = y_0$, $y(b) = y_b$ (f assumed positive); boundary value problems for other equations	$y = Y(x, \varepsilon) + \widetilde{Y}(z, \varepsilon)$, $Y = \sum_{k=0}^{n-1} \sigma_k(\varepsilon) Y_k(x)$, $\widetilde{Y} = \sum_{k=0}^{n-1} \widetilde{\sigma}_k(\varepsilon) \widetilde{Y}_k(z)$, $z = \frac{x}{\varepsilon}$; here, $\widetilde{Y}_k \to 0$ as $z \to \infty$	Unknowns: Y_k, $\widetilde{Y}_k$, σ_k, $\widetilde{\sigma}_k$; $Y(b, \varepsilon) = y_b$, $Y(0, \varepsilon) + \widetilde{Y}(0, \varepsilon) = y_0$; two forms of representation of the equation (in terms of x and z) are used to obtain solutions

12.3.5-2. Method of regular (direct) expansion in powers of the small parameter.

We consider an equation of general form with a parameter ε:

$$y''_{tt} + f(t, y, y'_t, \varepsilon) = 0. \tag{12.3.5.1}$$

We assume that the function f can be represented as a series in powers of ε:

$$f(t, y, y'_t, \varepsilon) = \sum_{n=0}^{\infty} \varepsilon^n f_n(t, y, y'_t). \tag{12.3.5.2}$$

Solutions of the Cauchy problem and various boundary value problems for equation (12.3.5.1) with $\varepsilon \to 0$ are sought in the form of a power series expansion:

$$y = \sum_{n=0}^{\infty} \varepsilon^n y_n(t). \tag{12.3.5.3}$$

One should substitute expression (12.3.5.3) into equation (12.3.5.1) taking into account (12.3.5.2). Then the functions f_n are expanded into a power series in the small parameter and the coefficients of like powers of ε are collected and equated to zero to obtain a system of equations for y_n:

$$y''_0 + f_0(t, y_0, y'_0) = 0, \tag{12.3.5.4}$$

$$y''_1 + F(t, y_0, y'_0)y'_1 + G(t, y_0, y'_0)y_1 + f_1(t, y_0, y'_0) = 0, \quad F = \frac{\partial f_0}{\partial y'}, \quad G = \frac{\partial f_0}{\partial y}. \tag{12.3.5.5}$$

Here, only the first two equations are written out. The prime denotes differentiation with respect to t. To obtain the initial (or boundary) conditions for y_n, the expansion (12.3.5.3) is taken into account.

The success in the application of this method is primarily determined by the possibility of constructing a solution of equation (12.3.5.4) for the leading term y_0. It is significant to note that the other terms y_n with $n \geq 1$ are governed by linear equations with homogeneous initial conditions.

Example 1. The Duffing equation

$$y''_{tt} + y + \varepsilon y^3 = 0 \tag{12.3.5.6}$$

with initial conditions

$$y(0) = a, \quad y'_t(0) = 0$$

describes the motion of a cubic oscillator, i.e., oscillations of a point mass on a nonlinear spring. Here, y is the deviation of the point mass from the equilibrium and t is dimensionless time.

For $\varepsilon \to 0$, an approximate solution of the problem is sought in the form of the asymptotic expansion (12.3.5.3). We substitute (12.3.5.3) into equation (12.3.5.6) and initial conditions and expand in powers of ε. On equating the coefficients of like powers of the small parameter to zero, we obtain the following problems for y_0 and y_1:

$$y''_0 + y_0 = 0, \qquad y_0 = a, \quad y'_0 = 0;$$
$$y''_1 + y_1 = -y_0^3, \qquad y_1 = 0, \quad y'_1 = 0.$$

The solution of the problem for y_0 is given by

$$y_0 = a \cos t.$$

Substituting this expression into the equation for y_1 and taking into account the identity $\cos^3 t = \frac{1}{4}\cos 3t + \frac{3}{4}\cos t$, we obtain

$$y''_1 + y_1 = -\tfrac{1}{4}a^3(\cos 3t + 3\cos t), \qquad y_1 = 0, \quad y'_1 = 0.$$

Integrating yields

$$y_1 = -\tfrac{3}{8}a^3 t\sin t + \tfrac{1}{32}a^3(\cos 3t - 3\cos t).$$

Thus the two-term solution of the original problem is given by

$$y = a\cos t + \varepsilon a^3\left[-\tfrac{3}{8}t\sin t + \tfrac{1}{32}(\cos 3t - 3\cos t)\right] + O(\varepsilon^2).$$

Remark 1. The term $t\sin t$ causes $y_1/y_0 \to \infty$ as $t \to \infty$. For this reason, the solution obtained is unsuitable at large times. It can only be used for $\varepsilon t \ll 1$; this results from the condition of applicability of the expansion, $y_0 \gg \varepsilon y_1$.

This circumstance is typical of the method of regular series expansions with respect to the small parameter; in other words, the expansion becomes unsuitable at large values of the independent variable. This method is also inapplicable if the expansion (12.3.5.3) begins with negative powers of ε. Methods that allow avoiding the above difficulties are discussed below in Paragraphs 12.3.5-3 through 12.3.5-5.

Remark 2. Growing terms as $t \to \infty$, like $t\sin t$, that narrow down the domain of applicability of asymptotic expansions are called *secular*.

12.3.5-3. Method of scaled parameters (Lindstedt–Poincaré method).

We illustrate the characteristic features of the method of scaled parameters with a specific example (the transformation of the independent variable we use here as well as the form of the expansion are specified in the first row of Table 12.3).

Example 2. Consider the Duffing equation (12.3.5.6) again. On performing the change of variable

$$t = z(1 + \varepsilon\omega_1 + \cdots),$$

we have

$$y''_{zz} + (1 + \varepsilon\omega_1 + \cdots)^2(y + \varepsilon y^3) = 0. \tag{12.3.5.7}$$

The solution is sought in the series form $y = y_0(z) + \varepsilon y_1(z) + \cdots$. Substituting it into equation (12.3.5.7) and matching the coefficients of like powers of ε, we arrive at the following system of equations for two leading terms of the series:

$$y_0'' + y_0 = 0, \tag{12.3.5.8}$$

$$y_1'' + y_1 = -y_0^3 - 2\omega_1 y_0, \tag{12.3.5.9}$$

where the prime denotes differentiation with respect to z.

The general solution of equation (12.3.5.8) is given by

$$y_0 = a\cos(z + b), \tag{12.3.5.10}$$

where a and b are constants of integration. Taking into account (12.3.5.10) and rearranging terms, we reduce equation (12.3.5.9) to

$$y_1'' + y_1 = -\tfrac{1}{4}a^3\cos\left[3(z + b)\right] - 2a\left(\tfrac{3}{8}a^2 + \omega_1\right)\cos(z + b). \tag{12.3.5.11}$$

For $\omega_1 \neq -\frac{3}{8}a^2$, the particular solution of equation (12.3.5.11) contains a secular term proportional to $z\cos(z+b)$. In this case, the condition of applicability of the expansion $y_1/y_0 = O(1)$ (see the first row and the last column of Table 12.3) cannot be satisfied at sufficiently large z. For this condition to be met, one should set

$$\omega_1 = -\tfrac{3}{8}a^2. \tag{12.3.5.12}$$

In this case, the solution of equation (12.3.5.11) is given by

$$y_1 = \tfrac{1}{32}a^3\cos\left[3(z + b)\right]. \tag{12.3.5.13}$$

Subsequent terms of the expansion can be found likewise.

With (12.3.5.10), (12.3.5.12), and (12.3.5.13), we obtain a solution of the Duffing equation in the form

$$y = a\cos(\omega t + b) + \tfrac{1}{32}\varepsilon a^3\cos\left[3(\omega t + b)\right] + O(\varepsilon^2),$$

$$\omega = \left[1 - \tfrac{3}{8}\varepsilon a^2 + O(\varepsilon^2)\right]^{-1} = 1 + \tfrac{3}{8}\varepsilon a^2 + O(\varepsilon^2).$$

12.3.5-4. Averaging method (Van der Pol–Krylov–Bogolyubov scheme).

1°. The averaging method involved two stages. First, the second-order nonlinear equation

$$y''_{tt} + \omega_0^2 y = \varepsilon f(y, y'_t) \tag{12.3.5.14}$$

is reduced with the transformation

$$y = a\cos\varphi, \quad y'_t = -\omega_0 a \sin\varphi, \qquad \text{where} \quad a = a(t), \quad \varphi = \varphi(t),$$

to an equivalent system of two first-order differential equations:

$$\begin{aligned} a'_t &= -\frac{\varepsilon}{\omega_0} f(a\cos\varphi, -\omega_0 a\sin\varphi)\sin\varphi, \\ \varphi'_t &= \omega_0 - \frac{\varepsilon}{\omega_0 a} f(a\cos\varphi, -\omega_0 a\sin\varphi)\cos\varphi. \end{aligned} \tag{12.3.5.15}$$

The right-hand sides of equations (12.3.5.15) are periodic in φ, with the amplitude a being a slow function of time t. The amplitude and the oscillation character are changing little during the time the phase φ changes by 2π.

At the second stage, the right-hand sides of equations (12.3.5.15) are being averaged with respect to φ. This procedure results in an approximate system of equations:

$$\begin{aligned} a'_t &= -\frac{\varepsilon}{\omega_0} f_{\mathrm{s}}(a), \\ \varphi'_t &= \omega_0 - \frac{\varepsilon}{\omega_0 a} f_{\mathrm{c}}(a), \end{aligned} \tag{12.3.5.16}$$

where

$$\begin{aligned} f_{\mathrm{s}}(a) &= \frac{1}{2\pi}\int_0^{2\pi} \sin\varphi\, f(a\cos\varphi, -\omega_0 a\sin\varphi)\, d\varphi, \\ f_{\mathrm{c}}(a) &= \frac{1}{2\pi}\int_0^{2\pi} \cos\varphi\, f(a\cos\varphi, -\omega_0 a\sin\varphi)\, d\varphi. \end{aligned}$$

System (12.3.5.16) is substantially simpler than the original system (12.3.5.15) — the first equation in (12.3.5.16), for the oscillation amplitude a, is a separable equation and, hence, can readily be integrated; then the second equation in (12.3.5.16), which is linear in φ, can also be integrated.

Note that the Krylov–Bogolyubov–Mitropolskii method (see the fourth row in Table 12.3) generalizes the above approach and allows obtaining subsequent asymptotic terms as $\varepsilon \to 0$.

2°. Below we outline the general scheme of the averaging method. We consider the second-order nonlinear equation with a small parameter:

$$y''_{tt} + g(t, y, y'_t) = \varepsilon f(t, y, y'_t). \tag{12.3.5.17}$$

Equation (12.3.5.17) should first be transformed to the equivalent system of equations

$$\begin{aligned} y'_t &= u, \\ u'_t &= -g(t, y, u) + \varepsilon f(t, y, u). \end{aligned} \tag{12.3.5.18}$$

Suppose the general solution of the "truncated" system (12.3.5.18), with $\varepsilon = 0$, is known:

$$y_0 = \varphi(t, C_1, C_2), \qquad u_0 = \psi(t, C_1, C_2), \tag{12.3.5.19}$$

where C_1 and C_2 are constants of integration. Taking advantage of the method of variation of constants, we pass from the variables y, u in (12.3.5.18) to Lagrange's variables x_1, x_2 according to the formulas

$$y = \varphi(t, x_1, x_2), \qquad u = \psi(t, x_1, x_2), \tag{12.3.5.20}$$

where φ and ψ are the same functions that define the general solution of the "truncated" system (12.3.5.19). Transformation (12.3.5.20) allows the reduction of system (12.3.5.18) to the *standard form*

$$\begin{aligned} x_1' &= \varepsilon F_1(t, x_1, x_2), \\ x_2' &= \varepsilon F_2(t, x_1, x_2). \end{aligned} \tag{12.3.5.21}$$

Here, the prime denotes differentiation with respect to t and

$$F_1 = \frac{\varphi_2 f(t,\varphi,\psi)}{\varphi_2\psi_1 - \varphi_1\psi_2}, \quad F_2 = -\frac{\varphi_1 f(t,\varphi,\psi)}{\varphi_2\psi_1 - \varphi_1\psi_2}; \quad \varphi_k = \frac{\partial\varphi}{\partial x_k}, \quad \psi_k = \frac{\partial\psi}{\partial x_k},$$
$$\varphi = \varphi(t, x_1, x_2), \quad \psi = \psi(t, x_1, x_2).$$

It is significant to note that system (12.3.5.21) is equivalent to the original equation (12.3.5.17). The unknowns x_1 and x_2 are slow functions of time.

As a result of averaging, system (12.3.5.21) is replaced by a simpler, approximate autonomous system of equations:

$$\begin{aligned} x_1' &= \varepsilon \mathcal{F}_1(x_1, x_2), \\ x_2' &= \varepsilon \mathcal{F}_2(x_1, x_2), \end{aligned} \tag{12.3.5.22}$$

where

$$\mathcal{F}_k(x_1, x_2) = \frac{1}{T}\int_0^T F_k(t, x_1, x_2)\,dt \qquad \text{if } F_k \text{ is a } T\text{-periodic function of } t;$$
$$\mathcal{F}_k(x_1, x_2) = \lim_{T\to\infty}\frac{1}{T}\int_0^T F_k(t, x_1, x_2)\,dt \qquad \text{if } F_k \text{ is not periodic in } t.$$

Remark 1. The averaging method is applicable to equations (12.3.5.14) and (12.3.5.17) with nonsmooth right-hand sides.

Remark 2. The averaging method has rigorous mathematical substantiation. There is also a procedure that allows finding subsequent asymptotic terms. For this procedure, e.g., see the books by Bogolyubov and Mitropolskii (1974), Zhuravlev and Klimov (1988), and Arnold, Kozlov, and Neishtadt (1993).

12.3.5-5. Method of two-scale expansions (Cole–Kevorkian scheme).

1°. We illustrate the characteristic features of the method of two-scale expansions with a specific example. Thereafter we outline possible generalizations and modifications of the method.

Example 3. Consider the Van der Pol equation

$$y_{tt}'' + y = \varepsilon(1 - y^2)y_t'. \tag{12.3.5.23}$$

The solution is sought in the form (see the fifth row in Table 12.3):

$$\begin{aligned} &y = y_0(\xi,\eta) + \varepsilon y_1(\xi,\eta) + \varepsilon^2 y_2(\xi,\eta) + \cdots, \\ &\xi = \varepsilon t, \quad \eta = \bigl(1 + \varepsilon^2\omega_2 + \cdots\bigr)t. \end{aligned} \tag{12.3.5.24}$$

On substituting (12.3.5.24) into (12.3.5.23) and on matching the coefficients of like powers of ε, we obtain the following system for two leading terms:

$$\frac{\partial^2 y_0}{\partial \eta^2} + y_0 = 0, \tag{12.3.5.25}$$

$$\frac{\partial^2 y_1}{\partial \eta^2} + y_1 = -2\frac{\partial^2 y_0}{\partial \xi \partial \eta} + (1 - y_0^2)\frac{\partial y_0}{\partial \eta}. \tag{12.3.5.26}$$

The general solution of equation (12.3.5.25) is given by

$$y_0 = A(\xi)\cos\eta + B(\xi)\sin\eta. \tag{12.3.5.27}$$

The dependence of A and B on the slow variable ξ is not being established at this stage.

We substitute (12.3.5.27) into the right-hand side of equation (12.3.5.26) and perform elementary manipulations to obtain

$$\frac{\partial^2 y_1}{\partial \eta^2} + y_1 = \left[-2B'_\xi + \tfrac{1}{4}B(4 - A^2 - B^2)\right]\cos\eta + \left[2A'_\xi - \tfrac{1}{4}A(4 - A^2 - B^2)\right]\sin\eta$$
$$+ \tfrac{1}{4}(B^3 - 3A^2B)\cos 3\eta + \tfrac{1}{4}(A^3 - 3AB^2)\sin 3\eta. \tag{12.3.5.28}$$

The solution of this equation must not contain unbounded terms as $\eta \to \infty$; otherwise the necessary condition $y_1/y_0 = O(1)$ is not satisfied. Therefore the coefficients of $\cos\eta$ and $\sin\eta$ must be set equal to zero:

$$\begin{aligned} -2B'_\xi + \tfrac{1}{4}B(4 - A^2 - B^2) &= 0, \\ 2A'_\xi - \tfrac{1}{4}A(4 - A^2 - B^2) &= 0. \end{aligned} \tag{12.3.5.29}$$

Equations (12.3.5.29) serve to determine $A = A(\xi)$ and $B = B(\xi)$. We multiply the first equation in (12.3.5.29) by $-B$ and the second by A and add them together to obtain

$$r'_\xi - \tfrac{1}{8}r(4 - r^2) = 0, \qquad \text{where} \quad r^2 = A^2 + B^2. \tag{12.3.5.30}$$

The integration by separation of variables yields

$$r^2 = \frac{4r_0^2}{r_0^2 + (4 - r_0^2)e^{-\xi}}, \tag{12.3.5.31}$$

where r_0 is the initial oscillation amplitude.

On expressing A and B in terms of the amplitude r and phase φ, we have $A = r\cos\varphi$ and $B = -r\sin\varphi$. Substituting these expressions into either of the two equations in (12.3.5.29) and using (12.3.5.30), we find that $\varphi'_\xi = 0$ or $\varphi = \varphi_0 = \text{const}$. Therefore the leading asymptotic term can be represented as

$$y_0 = r(\xi)\cos(\eta + \varphi_0),$$

where $\xi = \varepsilon t$ and $\eta = t$, and the function $r(\xi)$ is determined by (12.3.5.31).

2°. The method of two-scale expansions can also be used for solving boundary value problems where the small parameter appears together with the highest derivative as a factor (such problems for $0 \le x \le a$ are indicated in the seventh row of Table 12.3 and in Paragraph 12.3.5-6). In the case where a boundary layer arises near the point $x = 0$ (and its thickness has an order of magnitude of ε), the solution is sought in the form

$$y = y_0(\xi, \eta) + \varepsilon y_1(\xi, \eta) + \varepsilon^2 y_2(\xi, \eta) + \cdots,$$
$$\xi = x, \quad \eta = \varepsilon^{-1}\left[g_0(x) + \varepsilon g_1(x) + \varepsilon^2 g_2(x) + \cdots\right],$$

where the functions $y_k = y_k(\xi, \eta)$ and $g_k = g_k(x)$ are to be determined. The derivative with respect to x is calculated in accordance with the rule

$$\frac{d}{dx} = \frac{\partial}{\partial \xi} + \eta'_x\frac{\partial}{\partial \eta} = \frac{\partial}{\partial \xi} + \frac{1}{\varepsilon}\left(g'_0 + \varepsilon g'_1 + \varepsilon^2 g'_2 + \cdots\right)\frac{\partial}{\partial \eta}.$$

Additional conditions are imposed on the asymptotic terms in the domain under consideration; namely, $y_{k+1}/y_k = O(1)$ and $g_{k+1}/g_k = O(1)$ for $k = 0, 1, \ldots$, and $g_0(x) \to x$ as $x \to 0$.

Remark. The two-scale method is also used to solve problems that arise in mechanics and physics and are described by partial differential equations.

12.3.5-6. Method of matched asymptotic expansions.

1°. We illustrate the characteristic features of the method of matched asymptotic expansions with a specific example (the form of the expansions is specified in the seventh row of Table 12.3). Thereafter we outline possible generalizations and modifications of the method.

Example 4. Consider the linear boundary value problem

$$\varepsilon y''_{xx} + y'_x + f(x)y = 0, \tag{12.3.5.32}$$

$$y(0) = a, \quad y(1) = b, \tag{12.3.5.33}$$

where $0 < f(0) < \infty$.

At $\varepsilon = 0$ equation (12.3.5.32) degenerates; the solution of the resulting first-order equation

$$y'_x + f(x)y = 0 \tag{12.3.5.34}$$

cannot meet the two boundary conditions (12.3.5.33) simultaneously. It can be shown that the condition at $x = 0$ has to be omitted in this case (a boundary layer arises near this point).

The leading asymptotic term of the outer expansion, $y = y_0(x) + O(\varepsilon)$, is determined by equation (12.3.5.34). The solution of (12.3.5.34) that satisfies the second boundary condition in (12.3.5.33) is given by

$$y_0(x) = b\exp\left[\int_x^1 f(\xi)\,d\xi\right]. \tag{12.3.5.35}$$

We seek the leading term of the inner expansion, in the boundary layer adjacent to the left boundary, in the following form (see the seventh row and third column in Table 12.3):

$$\widetilde{y} = \widetilde{y}_0(z) + O(\varepsilon), \qquad z = x/\varepsilon, \tag{12.3.5.36}$$

where z is the extended variable. Substituting (12.3.5.36) into (12.3.5.32) and extracting the coefficient of ε^{-1}, we obtain

$$\widetilde{y}''_0 + \widetilde{y}'_0 = 0, \tag{12.3.5.37}$$

where the prime denotes differentiation with respect to z. The solution of equation (12.3.5.37) that satisfies the first boundary condition in (12.3.5.33) is given by

$$\widetilde{y}_0 = a - C + Ce^{-z}. \tag{12.3.5.38}$$

The constant of integration C is determined from the condition of matching the leading terms of the outer and inner expansions:

$$y_0(x \to 0) = \widetilde{y}_0(z \to \infty). \tag{12.3.5.39}$$

Substituting (12.3.5.35) and (12.3.5.38) into condition (12.3.5.39) yields

$$C = a - be^{\langle f\rangle}, \qquad \text{where} \quad \langle f\rangle = \int_0^1 f(x)\,dx. \tag{12.3.5.40}$$

Taking into account relations (12.3.5.35), (12.3.5.36), (12.3.5.38), and (12.3.5.40), we represent the approximate solution in the form

$$y = \begin{cases} be^{\langle f\rangle} + \left(a - be^{\langle f\rangle}\right)e^{-x/\varepsilon} & \text{for } 0 \le x \le O(\varepsilon), \\ b\exp\left[\int_x^1 f(\xi)\,d\xi\right] & \text{for } O(\varepsilon) \le x \le 1. \end{cases} \tag{12.3.5.41}$$

It is apparent that inside the thin boundary layer, whose thickness is proportional to ε, the solution rapidly changes by a finite value, $\Delta = be^{\langle f\rangle} - a$.

To determine the function y on the entire interval $x \in [0, 1]$ using formula (12.3.5.41), one has to "switch" at some intermediate point $x = x_0$ from one part of the solution to the other. Such switching is not convenient and, in practice, one often resorts to a *composite solution* instead of using the double formula (12.3.5.41). In similar cases, a composite solution is defined as

$$y = y_0(x) + \widetilde{y}_0(z) - A, \qquad A = \lim_{x\to 0} y_0(x) = \lim_{z\to\infty} \widetilde{y}_0(z).$$

In the problem under consideration, we have $A = be^{\langle f\rangle}$ and hence the composite solution becomes

$$y = \left(a - be^{\langle f\rangle}\right)e^{-x/\varepsilon} + b\exp\left[\int_x^1 f(\xi)\,d\xi\right].$$

For $\varepsilon \ll x \le 1$, this solution transforms to the outer solution $y_0(x)$ and for $0 \le x \ll \varepsilon$, to the inner solution, thus providing an approximate representation of the unknown over the entire domain.

2°. We now consider an equation of the general form

$$\varepsilon y''_{xx} = F(x, y, y'_x) \tag{12.3.5.42}$$

subject to boundary conditions (12.3.5.33).

For the leading term of the outer expansion $y = y_0(x) + \cdots$, we have the equation

$$F(x, y_0, y'_0) = 0.$$

In the general case, when using the method of matched asymptotic expansions, the position of the boundary layer and the form of the inner (extended) variable have to be determined in the course of the solution of the problem.

First we assume that the boundary layer is located near the left boundary. In (12.3.5.42), we make a change of variable $z = x/\delta(\varepsilon)$ and rewrite the equation as

$$y''_{zz} = \frac{\delta^2}{\varepsilon} F\Big(\delta z, y, \frac{1}{\delta} y'_z\Big). \tag{12.3.5.43}$$

The function $\delta = \delta(\varepsilon)$ is selected so that the right-hand side of equation (12.3.5.43) has a nonzero limit value as $\varepsilon \to 0$, provided that z, y, and y'_z are of the order of 1.

Example 5. For $F(x, y, y'_x) = -kx^\lambda y'_x + y$, where $0 \le \lambda < 1$, the substitution $z = x/\delta(\varepsilon)$ brings equation (12.3.5.42) to

$$y''_{zz} = -\frac{\delta^{1+\lambda}}{\varepsilon} k z^\lambda y'_z + \frac{\delta^2}{\varepsilon} y.$$

In order that the right-hand side of this equation has a nonzero limit value as $\varepsilon \to 0$, one has to set $\delta^{1+\lambda}/\varepsilon = 1$ or $\delta^{1+\lambda}/\varepsilon = \text{const}$, where const is any positive number. It follows that $\delta = \varepsilon^{\frac{1}{1+\lambda}}$.

The leading asymptotic term of the inner expansion in the boundary layer, $y = \widetilde{y}_0(z) + \cdots$, is determined by the equation $\widetilde{y}''_0 + kz^\lambda \widetilde{y}'_0 = 0$, where the prime denotes differentiation with respect to z.

If the position of the boundary layer is selected incorrectly, the outer and inner expansions cannot be matched. In this situation, one should consider the case where an arbitrary boundary layer is located on the right (this case is reduced to the previous one with the change of variable $x = 1 - z$). In Example 5 above, the boundary layer is on the left if $k > 0$ and on the right if $k < 0$.

There is a procedure for matching subsequent asymptotic terms of the expansion (see the seventh row and last column in Table 12.3). In its general form, this procedure can be represented as

inner expansion of the outer expansion (y*-expansion for* $x \to 0$)
$=$ *outer expansion of the inner expansion* ($\widetilde{y}$*-expansion for* $z \to \infty$).

Remark 1. The method of matched asymptotic expansions can also be applied to construct periodic solutions of singularly perturbed equations (e.g., in the problem of relaxation oscillations of the Van der Pol oscillator).

Remark 2. Two boundary layers can arise in some problems (e.g., in cases where the right-hand side of equation (12.3.5.42) does not explicitly depend on y'_x).

Remark 3. The method of matched asymptotic expansions is also used for solving equations (in semi-infinite domains) that do not degenerate at $\varepsilon = 0$. In such cases, there are no boundary layers; the original variable is used in the inner domain, and an extended coordinate is introduced in the outer domain.

Remark 4. The method of matched asymptotic expansions is successfully applied for the solution of various problems in mathematical physics that are described by partial differential equations; in particular, it plays an important role in the theory of heat and mass transfer and in hydrodynamics.

12.3.6. Galerkin Method and Its Modifications (Projection Methods)

12.3.6-1. General form of an approximate solution.

Consider a boundary value problem for the equation

$$\mathfrak{F}[y] - f(x) = 0 \tag{12.3.6.1}$$

with linear homogeneous boundary conditions* at the points $x = x_1$ and $x = x_2$ $(x_1 \le x \le x_2)$. Here, $\mathfrak{F}$ is a linear or nonlinear differential operator of the second order (or a higher order operator); $y = y(x)$ is the unknown function and $f = f(x)$ is a given function. It is assumed that $\mathfrak{F}[0] = 0$.

Let us choose a sequence of linearly independent functions (called *basis functions*)

$$\varphi = \varphi_n(x) \qquad (n = 1, 2, \ldots, N) \tag{12.3.6.2}$$

satisfying the same boundary conditions as $y = y(x)$. According to all methods that will be considered below, an approximate solution of equation (12.3.6.1) is sought as a linear combination

$$y_N = \sum_{n=1}^{N} A_n \varphi_n(x), \tag{12.3.6.3}$$

with the unknown coefficients A_n to be found in the process of solving the problem.

The finite sum (12.3.6.3) is called an *approximation function*. The remainder term R_N obtained after the finite sum has been substituted into the left-hand side of equation (12.3.6.1),

$$R_N = \mathfrak{F}[y_N] - f(x). \tag{12.3.6.4}$$

If the remainder R_N is identically equal to zero, then the function y_N is the exact solution of equation (12.3.6.1). In general, $R_N \not\equiv 0$.

12.3.6-2. Galerkin method.

In order to find the coefficients A_n in (12.3.6.3), consider another sequence of linearly independent functions

$$\psi = \psi_k(x) \qquad (k = 1, 2, \ldots, N). \tag{12.3.6.5}$$

Let us multiply both sides of (12.3.6.4) by ψ_k and integrate the resulting relation over the region $V = \{x_1 \le x \le x_2\}$, in which we seek the solution of equation (12.3.6.1). Next, we equate the corresponding integrals to zero (for the exact solutions, these integrals are equal to zero). Thus, we obtain the following system of linear algebraic equations for the unknown coefficients A_n:

$$\int_{x_1}^{x_2} \psi_k R_N \, dx = 0 \qquad (k = 1, 2, \ldots, N). \tag{12.3.6.6}$$

Relations (12.3.6.6) mean that the approximation function (12.3.6.3) satisfies equation (12.3.6.1) "on the average" (i.e., in the integral sense) with weights ψ_k. Introducing

* Nonhomogeneous boundary conditions can be reduced to homogeneous ones by the change of variable $z = A_2x^2 + A_1x + A_0 + y$ (the constants A_2, A_1, and A_0 are selected using the method of undetermined coefficients).

the scalar product $\langle g, h\rangle = \int_{x_1}^{x_2} gh\,dx$ of arbitrary functions g and h, we can consider equations (12.3.6.6) as the condition of orthogonality of the remainder R_N to all weight functions ψ_k.

The Galerkin method can be applied not only to boundary value problems, but also to eigenvalue problems (in the latter case, one takes $f = \lambda y$ and seeks eigenfunctions y_n, together with eigenvalues λ_n).

Mathematical justification of the Galerkin method for specific boundary value problems can be found in the literature listed at the end of Chapter 12. Below we describe some other methods that are in fact special cases of the Galerkin method.

Remark. Most often, one takes suitable sequences of polynomials or trigonometric functions as $\varphi_n(x)$ in the approximation function (12.3.6.3).

12.3.6-3. Bubnov–Galerkin method, the moment method, the least squares method.

1°. The sequences of functions (12.3.6.2) and (12.3.6.5) in the Galerkin method can be chosen arbitrarily. In the case of equal functions,

$$\varphi_k(x) = \psi_k(x) \qquad (k = 1, 2, \ldots, N), \tag{12.3.6.7}$$

the method is often called the *Bubnov–Galerkin method*.

2°. The *moment method* is the Galerkin method with the weight functions (12.3.6.5) being powers of x,

$$\psi_k = x^k. \tag{12.3.6.8}$$

3°. Sometimes, the functions ψ_k are expressed in terms of φ_k by the relations

$$\psi_k = \mathfrak{F}[\varphi_k] \qquad (k = 1, 2, \ldots),$$

where $\mathfrak{F}$ is the differential operator of equation (12.3.6.1). This version of the Galerkin method is called the *least squares method*.

12.3.6-4. Collocation method.

In the collocation method, one chooses a sequence of points x_k, $k = 1, \ldots, N$, and imposes the condition that the remainder (12.3.6.4) be zero at these points,

$$R_N = 0 \quad \text{at} \quad x = x_k \qquad (k = 1, \ldots, N). \tag{12.3.6.9}$$

When solving a specific problem, the points x_k, at which the remainder R_N is set equal to zero, are regarded as most significant. The number of collocation points N is taken equal to the number of the terms of the series (12.3.6.3). This allows one to obtain a complete system of algebraic equations for the unknown coefficients A_n (for linear boundary value problems, this algebraic system is linear).

Note that the collocation method is a special case of the Galerkin method with the sequence (12.3.6.5) consisting of the Dirac delta functions:

$$\psi_k = \delta(x - x_k).$$

In the collocation method, there is no need to calculate integrals, and this essentially simplifies the procedure of solving nonlinear problems (although usually this method yields less accurate results than other modifications of the Galerkin method).

Example. Consider the boundary value problem for the linear variable-coefficient second-order ordinary differential equation

$$y''_{xx} + g(x)y - f(x) = 0 \tag{12.3.6.10}$$

subject to the boundary conditions of the first kind

$$y(-1) = y(1) = 0. \tag{12.3.6.11}$$

Assume that the coefficients of equation (12.3.6.10) are smooth even functions, so that $f(x) = f(-x)$ and $g(x) = g(-x)$. We use the collocation method for the approximate solution of problem (12.3.6.10)–(12.3.6.11).

1°. Take the polynomials

$$y_n(x) = x^{2n-2}(1 - x^2), \qquad n = 1, 2, \ldots N,$$

as the basis functions; they satisfy the boundary conditions (12.3.6.11), $y_n(\pm 1) = 0$.

Let us consider three collocation points

$$x_1 = -\sigma, \quad x_2 = 0, \quad x_3 = \sigma \qquad (0 < \sigma < 1) \tag{12.3.6.12}$$

and confine ourselves to two basis functions ($N = 2$), so that the approximation function is taken in the form

$$y(x) = A_1(1 - x^2) + A_2x^2(1 - x^2). \tag{12.3.6.13}$$

Substituting (12.3.6.13) in the left-hand side of equation (12.3.6.10) yields the remainder

$$R(x) = A_1\big[-2 + (1 - x^2)g(x)\big] + A_2\big[2 - 12x^2 + x^2(1 - x^2)g(x)\big] - f(x).$$

It must vanish at the collocation points (12.3.6.12). Taking into account the properties $f(\sigma) = f(-\sigma)$ and $g(\sigma) = g(-\sigma)$, we obtain two linear algebraic equations for the coefficients A_1 and A_2:

$$\begin{aligned} A_1\big[-2 + g(0)\big] + 2A_2 - f(0) &= 0 \quad (\text{at } x = 0),\\ A_1\big[-2 + (1 - \sigma^2)g(\sigma)\big] + A_2\big[2 - 12\sigma^2 + \sigma^2(1 - \sigma^2)g(\sigma)\big] - f(\sigma) &= 0 \quad (\text{at } x = \pm\sigma). \end{aligned} \tag{12.3.6.14}$$

2°. To be specific, let us take the following functions entering equation (12.3.6.10):

$$f(x) = -1, \quad g(x) = 1 + x^2. \tag{12.3.6.15}$$

On solving the corresponding system of algebraic equations (12.3.6.14), we find the coefficients

$$A_1 = \frac{\sigma^4 + 11}{\sigma^4 + 2\sigma^2 + 11}, \quad A_2 = -\frac{\sigma^2}{\sigma^4 + 2\sigma^2 + 11}. \tag{12.3.6.16}$$

In Fig. 12.3, the solid line depicts the numerical solution to problem (12.3.6.10)–(12.3.6.11), with the functions (12.3.6.15), obtained by the shooting method (see Paragraph 12.3.7-3). The dashed lines 1 and 2 show the approximate solutions obtained by the collocation method using the formulas (12.3.6.13), (12.3.6.16) with $\sigma = \frac{1}{2}$ (equidistant points) and $\sigma = \frac{\sqrt{2}}{2}$ (Chebyshev points, see Subsection 12.4.4), respectively. It is evident that both cases provide good coincidence of the approximate and numerical solutions; the use of Chebyshev points gives a more accurate result.

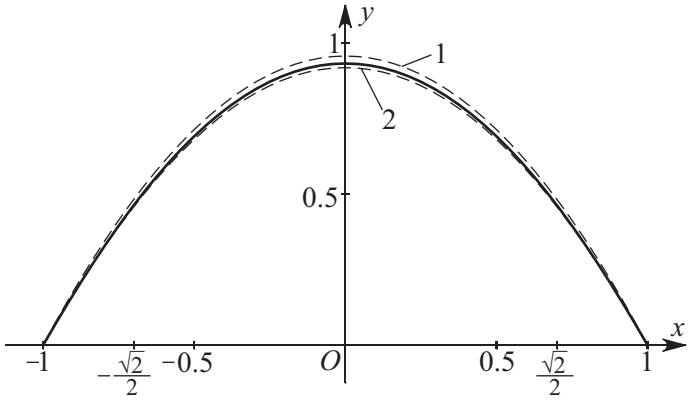


Figure 12.3. Comparison of the numerical solution of problem (12.3.6.10), (12.3.6.11), (12.3.6.15) with the approximate analytical solution (12.3.6.13), (12.3.6.16) obtained with the collocation method.

Remark. The theorem of convergence of the collocation method for linear boundary value problems is given in Subsection 12.4.4, where nth-order differential equations are considered.

12.3.6-5. Method of partitioning the domain.

The domain $V = \{x_1 \le x \le x_2\}$ is split into N subdomains: $V_k = \{x_{k1} \le x \le x_{k2}\}$, $k = 1, \ldots, N$. In this method, the weight functions are chosen as follows:

$$\psi_k(x) = \begin{cases} 1 & \text{for } x \in V_k, \\ 0 & \text{for } x \notin V_k. \end{cases}$$

The subdomains V_k are chosen according to the specific properties of the problem under consideration and can generally be arbitrary (the union of all subdomains V_k may differ from the domain V, and some V_k and V_m may overlap).

12.3.6-6. Least squared error method.

Sometimes, in order to find the coefficients A_n of the approximation function (12.3.6.3), one uses the least squared error method based on the minimization of the functional:

$$\Phi = \int_{x_1}^{x_2} R_N^2 \, dx \to \min. \tag{12.3.6.17}$$

For given functions φ_n in (12.3.6.3), the integral Φ is a function with respect to the coefficients A_n. The corresponding necessary conditions of minimum in (12.3.6.17) have the form

$$\frac{\partial \Phi}{\partial A_n} = 0 \qquad (n = 1, \ldots, N).$$

This is a system of algebraic (transcendental) equations for the coefficients A_n.

12.3.7. Iteration and Numerical Methods

12.3.7-1. Method of successive approximations (Cauchy problem).

The method of successive approximations is implemented in two steps. First, the Cauchy problem

$$y''_{xx} = f(x, y, y'_x) \qquad \text{(equation)}, \tag{12.3.7.1}$$

$$y(x_0) = y_0, \quad y'_x(x_0) = y'_0 \qquad \text{(initial conditions)} \tag{12.3.7.2}$$

is reduced to an equivalent system of integral equations by the introduction of the new variable $u(x) = y'_x$. These integral equations have the form

$$u(x) = y'_0 + \int_{x_0}^{x} f\big(t, y(t), u(t)\big) \, dt, \qquad y(x) = y_0 + \int_{x_0}^{x} u(t) \, dt. \tag{12.3.7.3}$$

Then the solution of system (12.3.7.3) is sought by means of successive approximations defined by the following recurrence formulas:

$$u_{n+1}(x) = y'_0 + \int_{x_0}^{x} f\big(t, y_n(t), u_n(t)\big) \, dt, \qquad y_{n+1}(x) = y_0 + \int_{x_0}^{x} u_n(t) \, dt; \qquad n = 0, 1, 2, \ldots$$

As the initial approximation, one can take $y_0(x) = y_0$ and $u_0(x) = y'_0$.

12.3.7-2. Runge–Kutta method (Cauchy problem).

For the numerical integration of the Cauchy problem (12.3.7.1)–(12.3.7.2), one often uses the Runge–Kutta method.

Let Δx be sufficiently small. We introduce the following notation:

$$x_k = x_0 + k\Delta x, \quad y_k = y(x_k), \quad y'_k = y'_x(x_k), \quad f_k = f(x_k, y_k, y'_k); \qquad k = 0, 1, 2, \ldots$$

The desired values y_k and y'_k are successively found by the formulas

$$y_{k+1} = y_k + y'_k\Delta x + \tfrac{1}{6}(f_1 + f_2 + f_3)(\Delta x)^2,$$
$$y'_{k+1} = y'_k + \tfrac{1}{6}(f_1 + 2f_2 + 2f_3 + f_4)\Delta x,$$

where

$$f_1 = f\big(x_k,\, y_k,\, y'_k\big),$$
$$f_2 = f\big(x_k + \tfrac{1}{2}\Delta x,\, y_k + \tfrac{1}{2}y'_k\Delta x,\, y'_k + \tfrac{1}{2}f_1\Delta x\big),$$
$$f_3 = f\big(x_k + \tfrac{1}{2}\Delta x,\, y_k + \tfrac{1}{2}y'_k\Delta x + \tfrac{1}{4}f_1(\Delta x)^2,\, y'_k + \tfrac{1}{2}f_2\Delta x\big),$$
$$f_4 = f\big(x_k + \Delta x,\, y_k + y'_k\Delta x + \tfrac{1}{2}f_2(\Delta x)^2,\, y'_k + f_3\Delta x\big).$$

In practice, the step Δx is determined in the same way as for first-order equations (see Remark 2 in Paragraph 12.1.10-3).

12.3.7-3. Shooting method (boundary value problems).

In order to solve the boundary value problem for equation (12.3.7.1) with the boundary conditions

$$y(x_1) = y_1, \qquad y(x_2) = y_2, \tag{12.3.7.4}$$

one considers an auxiliary Cauchy problem for equation (12.3.7.1) with the initial conditions

$$y(x_1) = y_1, \qquad y'_x(x_1) = a. \tag{12.3.7.5}$$

(The solution of this Cauchy problem can be obtained by the Runge–Kutta method or some other numerical method.) The parameter a is chosen so that the value of the solution $y = y(x, a)$ at the point $x = x_2$ coincides with the value required by the second boundary condition in (12.3.7.4):

$$y(x_2, a) = y_2.$$

In a similar way one constructs the solution of the boundary value problem with mixed boundary conditions

$$y(x_1) = y_1, \qquad y'_x(x_2) + ky(x_2) = y_2. \tag{12.3.7.6}$$

In this case, one also considers the auxiliary Cauchy problem (12.3.7.1), (12.3.7.5). The parameter a is chosen so that the solution $y = y(x, a)$ satisfies the second boundary condition in (12.3.7.6) at the point $x = x_2$.

12.3.7-4. Method of accelerated convergence in eigenvalue problems.

Consider the Sturm–Liouville problem for the second-order nonhomogeneous linear equation

$$[f(x)y'_x]'_x + [\lambda g(x) - h(x)]y = 0 \tag{12.3.7.7}$$

with linear homogeneous boundary conditions of the first kind

$$y(0) = y(1) = 0. \tag{12.3.7.8}$$

It is assumed that the functions f, f'_x, g, h are continuous and $f > 0$, $g > 0$.

First, using the Rayleigh–Ritz principle, one finds an upper estimate for the first eigenvalue λ_1^0 [this value is determined by the right-hand side of relation (12.2.5.6)]. Then, one solves numerically the Cauchy problem for the auxiliary equation

$$[f(x)y'_x]'_x + [\lambda_1^0 g(x) - h(x)]y = 0 \tag{12.3.7.9}$$

with the boundary conditions

$$y(0) = 0, \quad y'_x(0) = 1. \tag{12.3.7.10}$$

The function $y(x, \lambda_1^0)$ satisfies the condition $y(x_0, \lambda_1^0) = 0$, where $x_0 < 1$. The criterion of closeness of the exact and approximate solutions, λ_1 and λ_1^0, has the form of the inequality $|1 - x_0| \le \delta$, where δ is a sufficiently small given constant. If this inequality does not hold, one constructs a refinement for the approximate eigenvalue on the basis of the formula

$$\lambda_1^1 = \lambda_1^0 - \varepsilon_0 f(1)\frac{[y'_x(1)]^2}{\|y\|^2}, \quad \varepsilon_0 = 1 - x_0, \tag{12.3.7.11}$$

where $\|y\|^2 = \int_0^1 g(x)y^2(x)\,dx$. Then the value λ_1^1 is substituted for λ_1^0 in the Cauchy problem (12.3.7.9)–(12.3.7.10). As a result, a new solution y and a new point x_1 are found; and one has to check whether the criterion $|1 - x_1| \le \delta$ holds. If this inequality is violated, one refines the approximate eigenvalue by means of the formula

$$\lambda_1^2 = \lambda_1^1 - \varepsilon_1 f(1)\frac{[y'_x(1)]^2}{\|y\|^2}, \quad \varepsilon_1 = 1 - x_1, \tag{12.3.7.12}$$

and repeats the above procedure.

Remark 1. Formulas of the type (12.3.7.11) are obtained by a perturbation method based on a transformation of the independent variable x (see Paragraph 12.3.5-2). If $x_n > 1$, the functions f, g, and h are smoothly extended to the interval $(1, \xi]$, where $\xi \ge x_n$.

Remark 2. The algorithm described above has the property of accelerated convergence $\varepsilon_{n+1} \sim \varepsilon_n^2$, which ensures that the relative error of the approximate solution becomes 10^{-4} to 10^{-8} after two or three iterations for $\varepsilon_0 \sim 0.1$. This method is quite effective for high-precision calculations, is fail-safe, and guarantees against accumulation of roundoff errors.

Remark 3. In a similar way, one can compute subsequent eigenvalues λ_m, $m = 2, 3, \ldots$ (to that end, a suitable initial approximation λ_m^0 should be chosen).

Remark 4. A similar computation scheme can also be used in the case of boundary conditions of the second and the third kinds, periodic boundary conditions, etc. (see the reference below).

Example 1. The eigenvalue problem for the equation

$$y''_{xx} + \lambda(1 + x^2)^{-2}y = 0$$

with the boundary conditions (12.3.7.8) admits an exact analytic solution and has eigenvalues $\lambda_1 = 15$, $\lambda_2 = 63$, $\ldots$, $\lambda_n = 16n^2 - 1$.

According to the Rayleigh–Ritz principle, formula (12.2.5.6) for $z = \sin(\pi x)$ yields the approximate value $\lambda_1^0 = 15.33728$. The solution of the Cauchy problem (12.3.7.9)–(12.3.7.10) with $f(x) = 1$, $g(x) = \lambda(1 + x^2)^{-2}$, $h(x) = 0$ yields $x_0 = 0.983848$, $1 - x_0 = 0.016152$, $\|y\|^2 = 0.024585$, $y'_x(x_0) = -0.70622822$.

The first iteration for the first eigenvalue is determined by (12.3.7.11) and results in the value $\lambda_1^1 = 14.99245$ with the relative error $\Delta\lambda/\lambda_1^1 = 5 \times 10^{-4}$.

The second iteration results in $\lambda_1^2 = 14.999986$ with the relative error $\Delta\lambda/\lambda_1^2 < 10^{-6}$.

Example 2. Consider the eigenvalue problem for the equation

$$(\sqrt{1 + x}\, y'_x)'_x + \lambda y = 0$$

with the boundary conditions (12.3.7.8).

The Rayleigh–Ritz principle yields $\lambda_1^0 = 11.995576$. The next two iterations result in the values $\lambda_1^1 = 11.898578$ and $\lambda_1^2 = 11.898458$. For the relative error we have $\Delta\lambda/\lambda_1^2 < 10^{-5}$.

► *For more details about finite-difference methods and other numerical methods, see, for instance, the books by Lambert (1973), Keller (1976), Schiesser (1993), and Zwillinger (1997).*

12.4. Linear Equations of Arbitrary Order

12.4.1. Linear Equations with Constant Coefficients

12.4.1-1. Homogeneous linear equations.

An nth-order homogeneous linear equation with constant coefficients has the general form

$$y_x^{(n)} + a_{n-1}y_x^{(n-1)} + \cdots + a_1y'_x + a_0y = 0. \qquad (12.4.1.1)$$

The general solution of this equation is determined by the roots of the characteristic equation

$$P(\lambda) = 0, \qquad \text{where} \quad P(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0. \qquad (12.4.1.2)$$

The following cases are possible:

1°. All roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ of the characteristic equation (12.4.1.2) are real and distinct. Then the general solution of the homogeneous linear differential equation (12.4.1.1) has the form

$$y = C_1 \exp(\lambda_1 x) + C_2 \exp(\lambda_2 x) + \cdots + C_n \exp(\lambda_n x).$$

2°. There are m equal real roots $\lambda_1 = \lambda_2 = \cdots = \lambda_m$ $(m \le n)$, and the other roots are real and distinct. In this case, the general solution is given by

$$\begin{aligned} y = {} & \exp(\lambda_1 x)(C_1 + C_2 x + \cdots + C_m x^{m-1}) \\ & + C_{m+1} \exp(\lambda_{m+1} x) + C_{m+2} \exp(\lambda_{m+2} x) + \cdots + C_n \exp(\lambda_n x). \end{aligned}$$

3°. There are m equal complex conjugate roots $\lambda = \alpha \pm i\beta$ $(2m \le n)$, and the other roots are real and distinct. In this case, the general solution is

$$\begin{aligned} y = {} & \exp(\alpha x)\cos(\beta x)(A_1 + A_2 x + \cdots + A_m x^{m-1}) \\ & + \exp(\alpha x)\sin(\beta x)(B_1 + B_2 x + \cdots + B_m x^{m-1}) \\ & + C_{2m+1} \exp(\lambda_{2m+1} x) + C_{2m+2} \exp(\lambda_{2m+2} x) + \cdots + C_n \exp(\lambda_n x), \end{aligned}$$

where $A_1, \ldots, A_m, B_1, \ldots, B_m, C_{2m+1}, \ldots, C_n$ are arbitrary constants.

4°. In the general case, where there are r different roots $\lambda_1, \lambda_2, \ldots, \lambda_r$ of multiplicities $m_1, m_2, \ldots, m_r$, respectively, the left-hand side of the characteristic equation (12.4.1.2) can be represented as the product

$$P(\lambda) = (\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \ldots (\lambda - \lambda_r)^{m_r},$$

where $m_1 + m_2 + \cdots + m_r = n$. The general solution of the original equation is given by the formula

$$y = \sum_{k=1}^{r} \exp(\lambda_k x)(C_{k,0} + C_{k,1}x + \cdots + C_{k,m_k-1}x^{m_k-1}),$$

where $C_{k,l}$ are arbitrary constants.

If the characteristic equation (12.4.1.2) has complex conjugate roots, then in the above solution, one should extract the real part on the basis of the relation $\exp(\alpha \pm i\beta) = e^{\alpha}(\cos\beta \pm i \sin\beta)$.

Example 1. Find the general solution of the linear third-order equation

$$y''' + ay'' - y' - ay = 0.$$

Its characteristic equation is $\lambda^3 + a\lambda^2 - \lambda - a = 0$, or, in factorized form,

$$(\lambda + a)(\lambda - 1)(\lambda + 1) = 0.$$

Depending on the value of the parameter a, three cases are possible.

1. Case $a \neq \pm 1$. There are three different roots, $\lambda_1 = -a$, $\lambda_2 = -1$, and $\lambda_3 = 1$. The general solution of the differential equation is expressed as $y = C_1e^{-ax} + C_2e^{-x} + C_3e^{x}$.

2. Case $a = 1$. There is a double root, $\lambda_1 = \lambda_2 = -1$, and a simple root, $\lambda_3 = 1$. The general solution of the differential equation has the form $y = (C_1 + C_2x)e^{-x} + C_3e^{x}$.

3. Case $a = -1$. There is a double root, $\lambda_1 = \lambda_2 = 1$, and a simple root, $\lambda_3 = -1$. The general solution of the differential equation is expressed as $y = (C_1 + C_2x)e^{x} + C_3e^{-x}$.

Example 2. Consider the linear fourth-order equation

$$y''''_{xxxx} - y = 0.$$

Its characteristic equation, $\lambda^4 - 1 = 0$, has four distinct roots, two real and two pure imaginary,

$$\lambda_1 = 1, \quad \lambda_2 = -1, \quad \lambda_3 = i, \quad \lambda_4 = -i.$$

Therefore the general solution of the equation in question has the form (see Item 3°)

$$y = C_1e^{x} + C_2e^{-x} + C_3 \sin x + C_4 \cos x.$$

12.4.1-2. Nonhomogeneous linear equations. Forms of particular solutions.

1°. An nth-order nonhomogeneous linear equation with constant coefficients has the general form

$$y_x^{(n)} + a_{n-1}y_x^{(n-1)} + \cdots + a_1y'_x + a_0y = f(x). \tag{12.4.1.3}$$

The general solution of this equation is the sum of the general solution of the corresponding homogeneous equation with $f(x) \equiv 0$ (see Paragraph 12.4.1-1) and any particular solution of the nonhomogeneous equation (12.4.1.3).

If all the roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ of the characteristic equation (12.4.1.2) are different, equation (12.4.1.3) has the general solution:

$$y = \sum_{\nu=1}^{n} C_\nu e^{\lambda_\nu x} + \sum_{\nu=1}^{n} \frac{e^{\lambda_\nu x}}{P'_\lambda(\lambda_\nu)} \int f(x)e^{-\lambda_\nu x}\,dx \tag{12.4.1.4}$$

(for complex roots, the real part should be taken).

In the general case, if the characteristic equation (12.4.1.2) has multiple roots, the solution to equation (12.4.1.3) can be constructed using formula (12.4.2.5).

TABLE 12.4
Forms of particular solutions of the constant coefficient nonhomogeneous linear equation $y_x^{(n)} + a_{n-1}y_x^{(n-1)} + \cdots + a_1y_x' + a_0y = f(x)$ that correspond to some special forms of the function $f(x)$

Form of the function $f(x)$	Roots of the characteristic equation $\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = 0$	Form of a particular solution $y = \widetilde{y}(x)$
$P_m(x)$	Zero is not a root of the characteristic equation (i.e., $a_0 \neq 0$)	$\widetilde{P}_m(x)$
	Zero is a root of the characteristic equation (multiplicity r)	$x^r\widetilde{P}_m(x)$
$P_m(x)e^{\alpha x}$ (α is a real constant)	α is not a root of the characteristic equation	$\widetilde{P}_m(x)e^{\alpha x}$
	α is a root of the characteristic equation (multiplicity r)	$x^r\widetilde{P}_m(x)e^{\alpha x}$
$P_m(x)\cos\beta x + Q_n(x)\sin\beta x$	$i\beta$ is not a root of the characteristic equation	$\widetilde{P}_\nu(x)\cos\beta x + \widetilde{Q}_\nu(x)\sin\beta x$
	$i\beta$ is a root of the characteristic equation (multiplicity r)	$x^r[\widetilde{P}_\nu(x)\cos\beta x + \widetilde{Q}_\nu(x)\sin\beta x]$
$[P_m(x)\cos\beta x + Q_n(x)\sin\beta x]e^{\alpha x}$	$\alpha + i\beta$ is not a root of the characteristic equation	$[\widetilde{P}_\nu(x)\cos\beta x + \widetilde{Q}_\nu(x)\sin\beta x]e^{\alpha x}$
	$\alpha + i\beta$ is a root of the characteristic equation (multiplicity r)	$x^r[\widetilde{P}_\nu(x)\cos\beta x + \widetilde{Q}_\nu(x)\sin\beta x]e^{\alpha x}$

Notation: P_m and Q_n are polynomials of degrees m and n with given coefficients; $\widetilde{P}_m$, $\widetilde{P}_\nu$, and $\widetilde{Q}_\nu$ are polynomials of degrees m and ν whose coefficients are determined by substituting the particular solution into the basic equation; $\nu = \max(m,\ n)$; and α and β are real numbers, $i^2 = -1$.

2°. Table 12.4 lists the forms of particular solutions corresponding to some special forms of functions on the right-hand side of the linear nonhomogeneous equation.

3°. Consider the Cauchy problem for equation (12.4.1.3) subject to the homogeneous initial conditions

$$y(0) = y_x'(0) = \ldots = y_x^{(n-1)}(0) = 0. \tag{12.4.1.5}$$

Let $y(x)$ be the solution of problem (12.4.1.3), (12.4.1.5) for arbitrary $f(x)$ and let $u(x)$ be the solution of the auxiliary, simpler problem (12.4.1.3), (12.4.1.5) with $f(x) \equiv 1$, so that $u(x) = y(x)|_{f(x)\equiv 1}$. Then the formula

$$y(x) = \int_0^x f(t)u_x'(x-t)\,dt$$

holds. It is called the *Duhamel integral*.

12.4.1-3. Solution of the Cauchy problem using the Laplace transform.

Consider the Cauchy problem for equation (12.4.1.3) with arbitrary initial conditions

$$y(0) = y_0, \quad y_x'(0) = y_1, \quad \ldots, \quad y_x^{(n-1)}(0) = y_{n-1}, \tag{12.4.1.6}$$

where $y_0, y_1, \ldots, y_{n-1}$ are given constants.

Problem (12.4.1.3), (12.4.1.6) can be solved using the Laplace transform based on the formulas (for details, see Section 11.2)

$$\widetilde{y}(p) = \mathfrak{L}\{y(x)\}, \quad \widetilde{f}(p) = \mathfrak{L}\{f(x)\}, \quad \text{where} \quad \mathfrak{L}\{f(x)\} \equiv \int_0^\infty e^{-px} f(x)\,dx.$$

To this end, let us multiply equation (12.4.1.3) by e^{-px} and then integrate with respect to x from zero to infinity. Taking into account the differentiation rule

$$\mathfrak{L}\{y_x^{(n)}(x)\} = p^n \widetilde{y}(p) - \sum_{k=1}^{n} p^{n-k} y_x^{(k-1)}(+0)$$

and the initial conditions (12.4.1.6), we arrive at a linear algebraic equation for the transform $\widetilde{y}(p)$:

$$P(p)\widetilde{y}(p) - Q(p) = \widetilde{f}(p), \tag{12.4.1.7}$$

where

$$P(p) = p^n + a_{n-1}p^{n-1} + \cdots + a_1 p + a_0, \quad Q(p) = b_{n-1}p^{n-1} + \cdots + b_1 p + b_0,$$
$$b_k = y_{n-k-1} + a_{n-1}y_{n-k-2} + \cdots + a_{k+2}y_1 + a_{k+1}y_0, \quad k = 0,\ 1,\ \ldots,\ n-1.$$

The polynomial $P(p)$ coincides with the characteristic polynomial (12.4.1.2) at $\lambda = p$.

The solution of equation (12.4.1.7) is given by the formula

$$\widetilde{y}(p) = \frac{\widetilde{f}(p) + Q(p)}{\widetilde{P}(p)}. \tag{12.4.1.8}$$

On applying the Laplace inversion formula (see in Section 11.2) to (12.4.1.8), we obtain a solution to problem (12.4.1.3), (12.4.1.6) in the form

$$y(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{\widetilde{f}(p) + Q(p)}{\widetilde{P}(p)} e^{px}\,dp. \tag{12.4.1.9}$$

Since the transform $\widetilde{y}(p)$ (12.4.1.8) is a rational function, the inverse Laplace transform (12.4.1.9) can be obtained using the formulas from Paragraph 11.2.2-2 or the tables of Section T3.2.

Remark. In practice, the solution method for the Cauchy problem based on the Laplace transform leads to the solution faster than the direct application of general formulas like (12.4.1.4), where one has to determine the coefficients $C_1, \ldots, C_n$.

Example 3. Consider the following Cauchy problem for a homogeneous fourth-order equation:

$$y_{xxxx}'''' + a^4 y = 0; \qquad y(0) = y_x'(0) = y_{xxx}'''(0) = 0, \quad y_{xx}''(0) = b.$$

Using the Laplace transform reduces this problem to a linear algebraic equation for $\widetilde{y}(p)$: $(p^4 + a^4)\widetilde{y}(p) - bp = 0$. It follows that

$$\widetilde{y}(p) = \frac{bp}{p^4 + a^4}.$$

In order to invert this expression, let us use the table of inverse Laplace transforms T3.2.2 (see row 52) and take into account that a constant multiplier can be taken outside the transform operator to obtain the solution to the original Cauchy problem in the form

$$y(x) = \frac{b}{a^2}\sin\Big(\frac{ax}{\sqrt{2}}\Big)\sinh\Big(\frac{ax}{\sqrt{2}}\Big).$$

12.4.2. Linear Equations with Variable Coefficients

12.4.2-1. Homogeneous linear equations. Structure of the general solution.

The general solution of the nth-order homogeneous linear differential equation

$$f_n(x)y_x^{(n)} + f_{n-1}(x)y_x^{(n-1)} + \cdots + f_1(x)y'_x + f_0(x)y = 0 \tag{12.4.2.1}$$

has the form

$$y = C_1y_1(x) + C_2y_2(x) + \cdots + C_ny_n(x). \tag{12.4.2.2}$$

Here, $y_1(x)$, $y_2(x)$, …, $y_n(x)$ is a fundamental system of solutions (the y_k are linearly independent particular solutions, $y_k \not\equiv 0$); $C_1, C_2, \ldots, C_n$ are arbitrary constants.

12.4.2-2. Utilization of particular solutions for reducing the order of the equation.

1°. Let $y_1 = y_1(x)$ be a nontrivial particular solution of equation (12.4.2.1). The substitution

$$y = y_1(x)\int z(x)\,dx$$

results in a linear equation of order $n-1$ for the function $z(x)$.

2°. Let $y_1 = y_1(x)$ and $y_2 = y_2(x)$ be two nontrivial linearly independent solutions of equation (12.4.2.1). The substitution

$$y = y_1\int y_2w\,dx - y_2\int y_1w\,dx$$

results in a linear equation of order $n-2$ for $w(x)$.

3°. Suppose that m linearly independent solutions $y_1(x)$, $y_2(x)$, …, $y_m(x)$ of equation (12.4.2.1) are known. Then one can reduce the order of the equation to $n-m$ by successive application of the following procedure. The substitution $y = y_m(x)\int z(x)\,dx$ leads to an equation of order $n-1$ for the function $z(x)$ with known linearly independent solutions:

$$z_1 = \Big(\frac{y_1}{y_m}\Big)'_x, \quad z_2 = \Big(\frac{y_2}{y_m}\Big)'_x, \quad \ldots, \quad z_{m-1} = \Big(\frac{y_{m-1}}{y_m}\Big)'_x.$$

The substitution $z = z_{m-1}(x)\int w(x)\,dx$ yields an equation of order $n-2$. Repeating this procedure m times, we arrive at a homogeneous linear equation of order $n-m$.

12.4.2-3. Wronskian determinant and Liouville formula.

The *Wronskian determinant* (or simply, *Wronskian*) is the function defined as

$$W(x) = \begin{vmatrix} y_1(x) & \cdots & y_n(x) \\ y'_1(x) & \cdots & y'_n(x) \\ \cdots & \cdots & \cdots \\ y_1^{(n-1)}(x) & \cdots & y_n^{(n-1)}(x) \end{vmatrix}, \tag{12.4.2.3}$$

where $y_1(x)$, …, $y_n(x)$ is a fundamental system of solutions of the homogeneous equation (12.4.2.1); $y_k^{(m)}(x) = \dfrac{d^my_k}{dx^m}$, $m = 1, \ldots, n-1$; $k = 1, \ldots, n$.

The following *Liouville formula* holds:

$$W(x) = W(x_0)\exp\left[-\int_{x_0}^{x}\frac{f_{n-1}(t)}{f_n(t)}\,dt\right].$$

12.4.2-4. Nonhomogeneous linear equations. Construction of the general solution.

1°. The general nonhomogeneous nth-order linear differential equation has the form

$$f_n(x)y_x^{(n)} + f_{n-1}(x)y_x^{(n-1)} + \cdots + f_1(x)y'_x + f_0(x)y = g(x). \tag{12.4.2.4}$$

The general solution of the nonhomogeneous equation (12.4.2.4) can be represented as the sum of its particular solution and the general solution of the corresponding homogeneous equation (12.4.2.1).

2°. Let $y_1(x), \ldots, y_n(x)$ be a fundamental system of solutions of the homogeneous equation (12.4.2.1), and let $W(x)$ be the Wronskian determinant (12.4.2.3). Then the general solution of the nonhomogeneous linear equation (12.4.2.4) can be represented as

$$y = \sum_{\nu=1}^{n} C_\nu y_\nu(x) + \sum_{\nu=1}^{n} y_\nu(x) \int \frac{W_\nu(x)\,dx}{f_n(x)W(x)}, \tag{12.4.2.5}$$

where $W_\nu(x)$ is the determinant obtained by replacing the νth column of the matrix (12.4.2.3) by the column vector with the elements $0, 0, \ldots, 0, g$.

3°. Let $y(x, \sigma)$ be the solution to the Cauchy problem for the homogeneous equation (12.4.2.1) with nonhomogeneous initial conditions at $x = \sigma$:

$$y(\sigma) = y'_x(\sigma) = \cdots = y_x^{(n-2)}(\sigma) = 0, \quad y_x^{(n-1)}(\sigma) = 1,$$

where σ is an arbitrary parameter. Then a particular solution of the nonhomogeneous linear equation (12.4.2.4) with homogeneous boundary conditions

$$y(x_0) = y'_x(x_0) = \cdots = y_x^{(n-1)}(x_0) = 0$$

is given by the *Cauchy formula*

$$\bar{y}(x) = \int_{x_0}^{x} y(x, \sigma)\frac{g(\sigma)}{f_n(\sigma)}\,d\sigma.$$

4°. *Superposition principle.* The solution of a nonhomogeneous linear equation

$$\mathbf{L}[y] = \sum_{k=1}^{m} g_k(x), \qquad \mathbf{L}[y] \equiv f_n(x)y_x^{(n)} + f_{n-1}(x)y_x^{(n-1)} + \cdots + f_1(x)y'_x + f_0(x)y$$

is determined by adding together the solutions,

$$y = \sum_{k=1}^{m} y_k,$$

of m (simpler) equations,

$$\mathbf{L}[y_k] = g_k(x), \qquad k = 1, 2, \ldots, m,$$

corresponding to respective nonhomogeneous terms in the original equation.

12.4.2-5. Euler equation.

1°. The nonhomogeneous Euler equation has the form

$$x^n y_x^{(n)} + a_{n-1}x^{n-1}y_x^{(n-1)} + \cdots + a_1 x y_x' + a_0 y = f(x).$$

The substitution $x = be^t$ ($b \neq 0$) leads to a constant coefficient linear equation of the form (12.4.1.3).

2°. Particular solutions of the homogeneous Euler equation [with $f(x) \equiv 0$] are sought in the form $y = x^k$. If all k are real and distinct, its general solution is expressed as

$$y(x) = C_1|x|^{k_1} + C_2|x|^{k_2} + \cdots + C_n|x|^{k_n}.$$

Remark. To a pair of complex conjugate values $k = \alpha \pm i\beta$ there corresponds a pair of particular solutions: $y = |x|^\alpha \sin(\beta|x|)$ and $y = |x|^\alpha \cos(\beta|x|)$.

12.4.2-6. Solution of equations using the Laplace transform. Laplace equation.

1°. Some classes of equations (12.4.2.1) or (12.4.2.4) with polynomial coefficients

$$f_k(x) = \sum_{m=0}^{s_k} a_{km}x^m$$

may be solved using the Laplace transform (see Paragraph 12.4.1-3 and Section 11.2). To this end, one uses the following formula for the Laplace transform of the product of a power function and a derivative of the unknown function:

$$\mathfrak{L}\left\{x^m y_x^{(n)}(x)\right\} = (-1)^m \frac{d^m}{dp^m}\left[p^n \widetilde{y}(p) - \sum_{k=1}^{n} p^{n-k} y_x^{(k-1)}(+0)\right]. \qquad (12.4.2.6)$$

The right-hand side contains initial data $y_x^{(m)}(+0)$, $m = 0, 1, \ldots, n-1$ (specified in the Cauchy problem). As a result, one arrives at a linear ordinary differential equation, with respect to p, for the transform $\widetilde{y}(p)$; the order of this equation is equal to $\max\limits_{1\le k\le n}\{s_k\}$, the highest degree of the polynomials that determine the equation coefficients. In some cases, the equation for $\widetilde{y}(p)$ turns out to be simpler than the initial equation for $y(x)$ and can be solved in closed form. The desired function $y(x)$ is found by inverting the transform $\widetilde{y}(p)$ using the formulas from Paragraph 11.2.2-2 of the tables from Section T3.2.

2°. Consider the *Laplace equation*

$$(a_n + b_n x)y_x^{(n)} + (a_{n-1} + b_{n-1}x)y_x^{(n-1)} + \cdots + (a_1 + b_1 x)y_x' + (a_0 + b_0 x)y = 0, \qquad (12.4.2.7)$$

whose coefficients are linear functions of the independent variable x. The application of the Laplace transform, in view of formulas (12.4.2.6), brings it to a linear first-order ordinary differential equation for the transform $\widetilde{y}(p)$.

Example 1. Consider a special case of equation (12.4.2.7):

$$xy''_{xx} + y'_x + axy = 0. \tag{12.4.2.8}$$

Denote $y(0) = y_0$ and $y'_x(0) = y_1$. Let us apply the Laplace transform to this equation using formulas (12.4.2.6). On rearrangement, we obtain a linear first-order equation for $\widetilde{y}(p)$:

$$-(p^2\widetilde{y} - y_0 p - y_1)'_p + (p\widetilde{y} - y_0) - a\widetilde{y}'_p = 0 \qquad \Longrightarrow \qquad (p^2 + a)\widetilde{y}'_p + p\widetilde{y} = 0.$$

Its general solution is expressed as

$$\widetilde{y} = \frac{C}{\sqrt{p^2 + a}}, \tag{12.4.2.9}$$

where C is an arbitrary constant. Applying the inverse Laplace transform to (12.4.2.9) and taking into account formulas 19 and 20 from Subsection T3.2.3, we find a solution to the original equation (12.4.2.8):

$$y(x) = \begin{cases} CJ_0(x\sqrt{a}) & \text{if } a > 0, \\ CI_0(x\sqrt{-a}) & \text{if } a < 0, \end{cases} \tag{12.4.2.10}$$

where $J_0(x)$ is the Bessel function of the first kind and $I_0(x)$ is the modified Bessel function of the first kind.

In this case, only one solution (12.4.2.10) has been obtained. This is due to the fact that the other solution goes to infinity as $x \to 0$, and hence formula (12.4.2.6) cannot be applied to it; this formula is only valid for finite initial values of the function and its derivatives.

12.4.2-7. Solution of equations using the Laplace integral.

Solutions to linear differential equations with polynomial coefficients can sometimes be represented as a *Laplace integral* in the form

$$y(x) = \int_{\mathcal{K}} e^{px} u(p)\, dp. \tag{12.4.2.11}$$

For now, no assumptions are made about the domain of integration $\mathcal{K}$; it could be a segment of the real axis or a curve in the complex plane.

Let us exemplify the usage of the Laplace integral (12.4.2.11) by considering equation (12.4.2.7). It follows from (12.4.2.11) that

$$y_x^{(k)}(x) = \int_{\mathcal{K}} e^{px} p^k u(p)\, dp,$$

$$xy_x^{(k)}(x) = \int_{\mathcal{K}} xe^{px} p^k u(p)\, dp = \Big[e^{px} p^k u(p)\Big]_{\mathcal{K}} - \int_{\mathcal{K}} e^{px} \frac{d}{dp}\Big[p^k u(p)\Big]\, dp.$$

Substituting these expressions into (12.4.2.7) yields

$$\int_{\mathcal{K}} e^{px} \left\{ \sum_{k=0}^{n} a_k p^k u(p) - \sum_{k=0}^{n} b_k \frac{d}{dp}\Big[p^k u(p)\Big] \right\} dp + \sum_{k=0}^{n} b_k \Big[e^{px} p^k u(p)\Big]_{\mathcal{K}} = 0. \tag{12.4.2.12}$$

This equation is satisfied if the expression in braces vanishes, thus resulting in a linear first-order ordinary differential equation for $u(p)$:

$$u(p) \sum_{k=0}^{n} a_k p^k - \frac{d}{dp}\Big[u(p) \sum_{k=0}^{n} b_k p^k\Big] = 0. \tag{12.4.2.13}$$

The remaining term in (12.4.2.12) must also vanish:

$$\Big[\sum_{k=0}^{n} b_k e^{px} p^k u(p)\Big]_{\mathcal{K}} = 0. \tag{12.4.2.14}$$

This condition can be met by appropriately selecting the path of integration $\mathcal{K}$. Consider the example below to illustrate the aforesaid.

Example 2. The linear variable-coefficient second-order equation

$$xy''_{xx} + (x + a + b)y'_x + ay = 0 \qquad (a > 0,\ b > 0) \tag{12.4.2.15}$$

is a special case of equation (12.4.2.7) with $n = 2$, $a_2 = 0$, $a_1 = a + b$, $a_0 = a$, $b_2 = b_1 = 1$, and $b_0 = 0$. On substituting these values into (12.4.2.13), we arrive at an equation for $u(p)$:

$$p(p + 1)u'_p - [(a + b - 2)p + a - 1]u = 0.$$

Its solution is given by

$$u(p) = p^{a-1}(p + 1)^{b-1}. \tag{12.4.2.16}$$

It follows from condition (12.4.2.14), in view of formula (12.4.2.16), that

$$\left[e^{px}(p + p^2)u(p)\right]_\alpha^\beta = \left[e^{px}p^a(p + 1)^b\right]_\alpha^\beta = 0, \tag{12.4.2.17}$$

where a segment of the real axis, $\mathcal{K} = [\alpha, \beta]$, has been chosen to be the path of integration. Condition (12.4.2.17) is satisfied if we set $\alpha = -1$ and $\beta = 0$. Consequently, one of the solutions to equation (12.4.2.15) has the form

$$y(x) = \int_{-1}^{0} e^{px}p^{a-1}(p + 1)^{b-1}\,dp. \tag{12.4.2.18}$$

Remark 1. If a is noninteger, it is necessary to separate the real and imaginary parts in (12.4.2.18) to obtain real solutions.

Remark 2. By setting $\alpha = -\infty$ and $\beta = 0$ in (12.4.2.17), one can find a second solution to equation (12.4.2.15) (at least for $x > 0$).

12.4.3. Asymptotic Solutions of Linear Equations

This subsection presents asymptotic solutions, as $\varepsilon \to 0$ $(\varepsilon > 0)$, of some higher-order linear ordinary differential equations containing arbitrary functions (sufficiently smooth), with the independent variable being real.

12.4.3-1. Fourth-order linear equations.

1°. Consider the equation

$$\varepsilon^4 y''''_{xxxx} - f(x)y = 0$$

on a closed interval $a \le x \le b$. With the condition $f > 0$, the leading terms of the asymptotic expansions of the fundamental system of solutions, as $\varepsilon \to 0$, are given by the formulas

$$y_1 = [f(x)]^{-3/8} \exp\left\{-\frac{1}{\varepsilon}\int [f(x)]^{1/4}\,dx\right\}, \quad y_2 = [f(x)]^{-3/8} \exp\left\{\frac{1}{\varepsilon}\int [f(x)]^{1/4}\,dx\right\},$$

$$y_3 = [f(x)]^{-3/8} \cos\left\{\frac{1}{\varepsilon}\int [f(x)]^{1/4}\,dx\right\}, \quad y_4 = [f(x)]^{-3/8} \sin\left\{\frac{1}{\varepsilon}\int [f(x)]^{1/4}\,dx\right\}.$$

2°. Now consider the "biquadratic" equation

$$\varepsilon^4 y''''_{xxxx} - 2\varepsilon^2 g(x)y''_{xx} - f(x)y = 0. \tag{12.4.3.1}$$

Introduce the notation

$$D(x) = [g(x)]^2 + f(x).$$

In the range where the conditions $f(x) \ne 0$ and $D(x) \ne 0$ are satisfied, the leading terms of the asymptotic expansions of the fundamental system of solutions of equation (12.4.3.1) are described by the formulas

$$y_k = [\lambda_k(x)]^{-1/2}[D(x)]^{-1/4} \exp\left\{\frac{1}{\varepsilon}\int \lambda_k(x)\,dx - \frac{1}{2}\int \frac{[\lambda_k(x)]'_x}{\sqrt{D(x)}}\,dx\right\}; \quad k = 1,\ 2,\ 3,\ 4,$$

where

$$\lambda_1(x) = \sqrt{g(x) + \sqrt{D(x)}}, \qquad \lambda_2(x) = -\sqrt{g(x) + \sqrt{D(x)}},$$

$$\lambda_3(x) = \sqrt{g(x) - \sqrt{D(x)}}, \qquad \lambda_4(x) = -\sqrt{g(x) - \sqrt{D(x)}}.$$

12.4.3-2. Higher-order linear equations.

1°. Consider an equation of the form

$$\varepsilon^n y_x^{(n)} - f(x)y = 0$$

on a closed interval $a \le x \le b$. Assume that $f \ne 0$. Then the leading terms of the asymptotic expansions of the fundamental system of solutions, as $\varepsilon \to 0$, are given by

$$y_m = \big[f(x)\big]^{-\frac{1}{2}+\frac{1}{2n}} \exp\left\{\frac{\omega_m}{\varepsilon}\int \big[f(x)\big]^{\frac{1}{n}}\,dx\right\}\big[1+O(\varepsilon)\big],$$

where $\omega_1, \omega_2, \ldots, \omega_n$ are roots of the equation $\omega^n = 1$:

$$\omega_m = \cos\Big(\frac{2\pi m}{n}\Big) + i\sin\Big(\frac{2\pi m}{n}\Big), \qquad m = 1,\ 2,\ \ldots,\ n.$$

2°. Now consider an equation of the form

$$\varepsilon^n y_x^{(n)} + \varepsilon^{n-1} f_{n-1}(x) y_x^{(n-1)} + \cdots + \varepsilon f_1(x) y_x' + f_0(x)y = 0 \tag{12.4.3.2}$$

on a closed interval $a \le x \le b$. Let $\lambda_m = \lambda_m(x)$ $(m = 1, 2, \ldots, n)$ be the roots of the characteristic equation

$$P(x,\lambda) \equiv \lambda^n + f_{n-1}(x)\lambda^{n-1} + \cdots + f_1(x)\lambda + f_0(x) = 0.$$

Let all the roots of the characteristic equation be different on the interval $a \le x \le b$, i.e., the conditions $\lambda_m(x) \ne \lambda_k(x)$, $m \ne k$, are satisfied, which is equivalent to the fulfillment of the conditions $P_\lambda(x, \lambda_m) \ne 0$. Then the leading terms of the asymptotic expansions of the fundamental system of solutions of equation (12.4.3.2), as $\varepsilon \to 0$, are given by

$$y_m = \exp\left\{\frac{1}{\varepsilon}\int \lambda_m(x)\,dx - \frac{1}{2}\int [\lambda_m(x)]_x' \frac{P_{\lambda\lambda}\big(x,\lambda_m(x)\big)}{P_\lambda\big(x,\lambda_m(x)\big)}\,dx\right\},$$

where

$$P_\lambda(x,\lambda) \equiv \frac{\partial P}{\partial \lambda} = n\lambda^{n-1} + (n-1)f_{n-1}(x)\lambda^{n-2} + \cdots + 2\lambda f_2(x) + f_1(x),$$

$$P_{\lambda\lambda}(x,\lambda) \equiv \frac{\partial^2 P}{\partial \lambda^2} = n(n-1)\lambda^{n-2} + (n-1)(n-2)f_{n-1}(x)\lambda^{n-3} + \cdots + 6\lambda f_3(x) + 2f_2(x).$$

12.4.4. Collocation Method and Its Convergence

12.4.4-1. Statement of the problem. Approximate solution.

1°. Consider the linear boundary value problem defined by the equation

$$\mathrm{L}y \equiv y_x^{(n)} + f_{n-1}(x)y_x^{(n-1)} + \cdots + f_1(x)y_x' + f_0(x)y = g(x), \qquad -1 < x < 1, \tag{12.4.4.1}$$

and the boundary conditions

$$\sum_{j=0}^{n-1}\big[\alpha_{ij}y_x^{(j)}(-1) + \beta_{ij}y_x^{(j)}(1)\big] = 0, \qquad i = 1, \ldots, n. \tag{12.4.4.2}$$

2°. We seek an approximate solution to problem (12.4.4.1)–(12.4.4.2) in the form

$$y_m(x) = A_1\varphi_1(x) + A_2\varphi_2(x) + \cdots + A_m\varphi_m(x),$$

where $\varphi_k(x)$ is a polynomial of degree $n + k - 1$ that satisfies the boundary conditions (12.4.4.2). The coefficients A_k are determined by the linear system of algebraic equations

$$\big[\mathrm{L}y_m - g(x)\big]_{x=x_i} = 0, \qquad i = 1, \ldots, m, \tag{12.4.4.3}$$

with *Chebyshev nodes* $x_i = \cos\Big(\dfrac{2i-1}{2m}\pi\Big)$, $i = 1,\ \ldots,\ m$.

12.4.4-2. Convergence theorem for the collocation method.

THEOREM. *Let the functions $f_j(x)$ $(j = 0, \dots, n-1)$ and $g(x)$ be continuous on the interval $[-1, 1]$ and let the boundary value problem (12.4.4.1)–(12.4.4.2) have a unique solution, $y(x)$. Then there exists an m_0 such that system (12.4.4.3) is uniquely solvable for $m \geq m_0$; and the limit relations*

$$\max_{-1\leq x\leq 1} \left|y_m^{(k)}(x) - y^{(k)}(x)\right| \to 0, \qquad k = 0, 1, \dots, n-1;$$

$$\left\{\int_{-1}^{1} \frac{\left|y_m^{(n)}(x) - y^{(n)}(x)\right|^2}{\sqrt{1-x^2}}\, dx\right\}^{1/2} \to 0$$

hold for $m \to \infty$.

Remark. A similar result holds true if the nodes are roots of some orthogonal polynomials with some weight function. If the nodes are equidistant, the method diverges.

12.5. Nonlinear Equations of Arbitrary Order

12.5.1. Structure of the General Solution. Cauchy Problem

12.5.1-1. Equations solved for the highest derivative. General solution.

An *nth-order differential equation solved for the highest derivative* has the form

$$y_x^{(n)} = f(x, y, y'_x, \dots, y_x^{(n-1)}). \tag{12.5.1.1}$$

The general solution of this equation depends on n arbitrary constants $C_1, \dots, C_n$. In some cases, the general solution can be written in explicit form as

$$y = \varphi(x, C_1, \dots, C_n). \tag{12.5.1.2}$$

12.5.1-2. Cauchy problem. The existence and uniqueness theorem.

The *Cauchy problem*: find a solution of equation (12.5.1.1) with the *initial conditions*

$$y(x_0) = y_0, \quad y'_x(x_0) = y_0^{(1)}, \quad \dots, \quad y_x^{(n-1)}(x_0) = y_0^{(n-1)}. \tag{12.5.1.3}$$

(At a point x_0, the values of the unknown function $y(x)$ and all its derivatives of orders $\leq n-1$ are prescribed.)

EXISTENCE AND UNIQUENESS THEOREM. *Suppose the function $f(x, y, z_1, \dots, z_{n-1})$ is continuous in all its arguments in a neighborhood of the point $(x_0, y_0, y_0^{(1)}, \dots, y_0^{(n-1)})$ and has bounded derivatives with respect to y, z_1, …, z_{n-1} in this neighborhood. Then a solution of equation (12.5.1.1) satisfying the initial conditions (12.5.1.3) exists and is unique.*

12.5.1-3. Construction of a differential equation by a given general solution.

Suppose a general solution (12.5.1.2) of an unknown nth-order ordinary differential equation is given. The equation corresponding to the general solution can be obtained by eliminating the arbitrary constants $C_1, \ldots, C_n$ from the identities

$$\begin{aligned} y &= \varphi(x, C_1, \ldots, C_n), \\ y'_x &= \varphi'_x(x, C_1, \ldots, C_n), \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_x^{(n)} &= \varphi_x^{(n)}(x, C_1, \ldots, C_n), \end{aligned}$$

obtained by differentiation from formula (12.5.1.2).

12.5.1-4. Reduction of an nth-order equation to a system of n first-order equation.

The differential equation (12.5.1.1) is equivalent to the following system of n first-order equations:

$$y'_0 = y_1, \quad y'_1 = y_2, \quad \ldots, \quad y'_{n-2} = y_{n-1}, \quad y'_{n-1} = f(x, y_0, y_1, \ldots, y_{n-1}),$$

where the notation $y_0 \equiv y$ is adopted.

12.5.2. Equations Admitting Reduction of Order

12.5.2-1. Equations not containing $y, y'_x, \ldots, y_x^{(k)}$ explicitly.

An equation that does not explicitly contain the unknown function and its derivatives up to order k inclusive can generally be written as

$$F\big(x, y_x^{(k+1)}, \ldots, y_x^{(n)}\big) = 0 \qquad (1 \le k+1 < n). \tag{12.5.2.1}$$

Such equations are invariant under arbitrary translations of the unknown function, $y \to y + \text{const}$ (the form of such equations is also preserved under the transformation $u(x) = y + a_k x^k + \cdots + a_1 x + a_0$, where the a_m are arbitrary constants). The substitution $z(x) = y_x^{(k+1)}$ reduces (12.5.2.1) to an equation whose order is by $k+1$ smaller than that of the original equation, $F\big(x, z, z'_x, \ldots, z_x^{(n-k-1)}\big) = 0$.

12.5.2-2. Equations not containing x explicitly (autonomous equations).

An equation that does not explicitly contain x has in the general form

$$F\big(y, y'_x, \ldots, y_x^{(n)}\big) = 0. \tag{12.5.2.2}$$

Such equations are invariant under arbitrary translations of the independent variable, $x \to x + \text{const}$. The substitution $y'_x = w(y)$ (where y plays the role of the independent variable) reduces by one the order of an autonomous equation. Higher derivatives can be expressed in terms of w and its derivatives with respect to the new independent variable, $y''_{xx} = ww'_y$, $y'''_{xxx} = w^2 w''_{yy} + w(w'_y)^2$, ...

12.5.2-3. Equations of the form $F\big(ax+by, y'_x, \dots, y^{(n)}_x\big)=0$.

Such equations are invariant under simultaneous translations of the independent variable and the unknown function, $x \to x+bc$ and $y \to y-ac$, where c is an arbitrary constant.

For $b=0$, see equation (12.5.2.1). For $b\neq 0$, the substitution $w(x)=y+(a/b)x$ leads to an autonomous equation of the form (12.5.2.2).

12.5.2-4. Equations of the form $F\big(x, xy'_x-y, y''_{xx}, \dots, y^{(n)}_x\big)=0$.

The substitution $w(x)=xy'_x-y$ reduces the order of this equation by one.

This equation is a special case of the equation

$$F\big(x, xy'_x-my, y^{(m+1)}_x, \dots, y^{(n)}_x\big)=0, \qquad \text{where} \quad m=1,\,2,\,\dots,\,n-1. \tag{12.5.2.3}$$

The substitution $w(x)=xy'_x-my$ reduces by one the order of equation (12.5.2.3).

12.5.2-5. Homogeneous equations.

1°. *Equations homogeneous in the independent variable* are invariant under scaling of the independent variable, $x \to \alpha x$, where α is an arbitrary constant ($\alpha \neq 0$). In general, such equations can be written in the form

$$F\big(y, xy'_x, x^2y''_{xx}, \dots, x^n y^{(n)}_x\big)=0.$$

The substitution $z(y)=xy'_x$ reduces by one the order of this equation.

2°. *Equations homogeneous in the unknown function* are invariant under scaling of the unknown function, $y \to \alpha y$, where α is an arbitrary constant ($\alpha \neq 0$). Such equations can be written in the general form

$$F\big(x, y'_x/y, y''_{xx}/y, \dots, y^{(n)}_x/y\big)=0.$$

The substitution $z(x)=y'_x/y$ reduces by one the order of this equation.

3°. *Equations homogeneous in both variables* are invariant under simultaneous scaling (dilatation) of the independent and dependent variables, $x \to \alpha x$ and $y \to \alpha y$, where α is an arbitrary constant ($\alpha \neq 0$). Such equations can be written in the general form

$$F\big(y/x, y'_x, xy''_{xx}, \dots, x^{n-1}y^{(n)}_x\big)=0.$$

The transformation $t=\ln|x|$, $w=y/x$ leads to an autonomous equation considered in Paragraph 12.5.2-2.

12.5.2-6. Generalized homogeneous equations.

1°. *Generalized homogeneous equations* (*equations homogeneous in the generalized sense*) are invariant under simultaneous scaling of the independent variable and the unknown function, $x \to \alpha x$ and $y \to \alpha^k y$, where $\alpha \neq 0$ is an arbitrary constant and k is a given number. Such equations can be written in the general form

$$F\big(x^{-k}y, x^{1-k}y'_x, \dots, x^{n-k}y^{(n)}_x\big)=0.$$

The transformation $t=\ln x$, $w=x^{-k}y$ leads to an autonomous equation considered in Paragraph 12.5.2-2.

2°. The most general form of generalized homogeneous equations is

$$\mathcal{F}\big(x^n y^m, xy'_x/y, \ldots, x^n y_x^{(n)}/y\big) = 0.$$

The transformation $z = x^n y^m$, $u = xy'_x/y$ reduces the order of this equation by one.

12.5.2-7. Equations of the form $F\big(e^{\lambda x}y^n, y'_x/y, y''_{xx}/y, \ldots, y_x^{(n)}/y\big) = 0.$

Such equations are invariant under simultaneous translation and scaling of variables, $x \to x + \alpha$ and $y \to \beta y$, where $\beta = \exp(-\alpha\lambda/n)$ and α is an arbitrary constant. The transformation $z = e^{\lambda x}y^n$, $w = y'_x/y$ leads to an equation of order $n - 1$.

12.5.2-8. Equations of the form $F\big(x^n e^{\lambda y}, xy'_x, x^2 y''_{xx}, \ldots, x^n y_x^{(n)}\big) = 0.$

Such equations are invariant under simultaneous scaling and translation of variables, $x \to \alpha x$ and $y \to y+\beta$, where $\alpha = \exp(-\beta\lambda/n)$ and β is an arbitrary constant. The transformation $z = x^n e^{\lambda y}$, $w = xy'_x$ leads to an equation of order $n - 1$.

12.5.2-9. Other equations.

Consider the nonlinear differential equation

$$F\big(x, \mathrm{L}_1[y], \ldots, \mathrm{L}_k[y]\big) = 0, \tag{12.5.2.4}$$

where the $\mathrm{L}_s[y]$ are linear homogeneous differential forms,

$$\mathrm{L}_s[y] = \sum_{m=0}^{n_s} \varphi_m^{(s)}(x) y_x^{(m)}, \qquad s = 1, \ldots, k.$$

Let $y_0 = y_0(x)$ be a common particular solution of the linear equations

$$\mathrm{L}_s[y_0] = 0 \qquad (s = 1, \ldots, k).$$

Then the substitution

$$w = \varphi(x)\big[y_0(x)y'_x - y'_0(x)y\big] \tag{12.5.2.5}$$

with an arbitrary function $\varphi(x)$ reduces by one the order of equation (12.5.2.4).

Example. Consider the third-order equation

$$xy'''_{xxx} = f(xy'_x - 2y).$$

It can be represented in the form (12.5.2.4) with

$$k = 2, \quad F(x, u, w) = xu - f(w), \quad \mathrm{L}_1[y] = y'''_{xxx}, \quad \mathrm{L}_2[y] = xy'_x - 2y.$$

The linear equations $\mathrm{L}_k[y] = 0$ are

$$y'''_{xxx} = 0, \quad xy'_x - 2y = 0.$$

These equations have a common particular solution $y_0 = x^2$. Therefore, the substitution $w = xy'_x - 2y$ (see formula (12.5.2.5) with $\varphi(x) = 1/x$) leads to a second-order autonomous equation: $w''_{xx} = f(w)$. For the solution of this equation, see Example 1 in Subsection 12.3.2.

12.6. Linear Systems of Ordinary Differential Equations

12.6.1. Systems of Linear Constant-Coefficient Equations

12.6.1-1. Systems of first-order linear homogeneous equations. The general solution.

1°. In general, a homogeneous linear system of constant-coefficient first-order ordinary differential equations has the form

$$\begin{aligned} y_1' &= a_{11}y_1 + a_{12}y_2 + \cdots + a_{1n}y_n, \\ y_2' &= a_{21}y_1 + a_{22}y_2 + \cdots + a_{2n}y_n, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_n' &= a_{n1}y_1 + a_{n2}y_2 + \cdots + a_{nn}y_n, \end{aligned} \tag{12.6.1.1}$$

where a prime stands for the derivative with respect to x. In the sequel, all the coefficients a_{ij} of the system are assumed to be real numbers.

The homogeneous system (12.6.1.1) has the trivial particular solution $y_1 = y_2 = \cdots = y_n = 0$.

Superposition principle for a homogeneous system: any linear combination of particular solutions of system (12.6.1.1) is also a solution of this system.

The general solution of the system of differential equations (12.6.1.1) is the sum of its n linearly independent (nontrivial) particular solutions multiplied by an arbitrary constant.

System (12.6.1.1) can be reduced to a single homogeneous linear constant-coefficient nth-order equation; see Paragraph 12.7.1-3.

2°. For brevity (and clearness), system (12.6.1.1) is conventionally written in vector-matrix form:

$$\mathbf{y}' = \mathbf{a}\mathbf{y}, \tag{12.6.1.2}$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_n)^{\mathrm{T}}$ is the column vector of the unknowns and $\mathbf{a} = (a_{ij})$ is the matrix of the equation coefficients. The superscript T denotes the transpose of a matrix or a vector. So, for example, a row vector is converted into a column vector:

$$(y_1, y_2)^{\mathrm{T}} \equiv \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

The right-hand side of equation (12.6.1.2) is the product of the $n \times n$ square matrix $\mathbf{a}$ by the $n \times 1$ matrix (column vector) $\mathbf{y}$.

Let $\mathbf{y}_k = (y_{k1}, y_{k2}, \ldots, y_{kn})^{\mathrm{T}}$ be linearly independent particular solutions* of the homogeneous system (12.6.1.1), where $k = 1, 2, \ldots, n$; the first subscript in $y_{km} = y_{km}(x)$ denotes the number of the solution and the second subscript indicates the component of the vector solution. Then the general solution of the homogeneous system (12.6.1.2) is expressed as

$$\mathbf{y} = C_1\mathbf{y}_1 + C_2\mathbf{y}_2 + \cdots + C_n\mathbf{y}_n. \tag{12.6.1.3}$$

A method for the construction of particular solutions that can be used to obtain the general solution by formula (12.6.1.3) is presented below.

* This means that the condition $\det|y_{mk}(x)| \neq 0$ holds.

12.6.1-2. Systems of first-order linear homogeneous equations. Particular solutions.

Particular solutions to system (12.6.1.1) are determined by the roots of the characteristic equation

$$\Delta(\lambda) = 0, \quad \text{where} \quad \Delta(\lambda) \equiv \begin{vmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-\lambda & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn}-\lambda \end{vmatrix}. \tag{12.6.1.4}$$

The following cases are possible:

1°. Let $\lambda = \lambda_k$ be a simple real root of the characteristic equation (12.6.1.4). The corresponding particular solution of the homogeneous linear system of equations (12.6.1.1) has the exponential form

$$y_1 = A_1 e^{\lambda x}, \quad y_2 = A_2 e^{\lambda x}, \quad \dots, \quad y_n = A_n e^{\lambda x}, \tag{12.6.1.5}$$

where the coefficients $A_1, A_2, \dots, A_n$ are determined by solving the associated homogeneous system of algebraic equations obtained by substituting expressions (12.6.1.5) into the differential equation (12.6.1.1) and dividing by $e^{\lambda x}$:

$$\begin{aligned} &(a_{11}-\lambda)A_1 + a_{12}A_2 + \dots + a_{1n}A_n = 0, \\ &a_{21}A_1 + (a_{22}-\lambda)A_2 + \dots + a_{2n}A_n = 0, \\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ &a_{n1}A_1 + a_{n2}A_2 + \dots + (a_{nn}-\lambda)A_n = 0. \end{aligned} \tag{12.6.1.6}$$

The solution of this system is unique to within a constant factor.

If all roots of the characteristic equation $\lambda_1, \lambda_2, \dots, \lambda_n$ are real and distinct, then the general solution of system (12.6.1.1) has the form

$$\begin{aligned} y_1 &= C_1 A_{11} e^{\lambda_1 x} + C_2 A_{12} e^{\lambda_2 x} + \dots + C_n A_{1n} e^{\lambda_n x}, \\ y_2 &= C_1 A_{21} e^{\lambda_1 x} + C_2 A_{22} e^{\lambda_2 x} + \dots + C_n A_{2n} e^{\lambda_n x}, \\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ y_n &= C_1 A_{n1} e^{\lambda_1 x} + C_2 A_{n2} e^{\lambda_2 x} + \dots + C_n A_{nn} e^{\lambda_n x}, \end{aligned} \tag{12.6.1.7}$$

where $C_1, C_2, \dots, C_n$ are arbitrary constants. The second subscript in A_{mk} indicates a coefficient corresponding to the root λ_k.

2°. For each simple complex root, $\lambda = \alpha \pm i\beta$, of the characteristic equation (12.6.1.4), the corresponding particular solution is obtained in the same way as in the simple real root case; the associated coefficients $A_1, A_2, \dots, A_n$ in (12.6.1.5) will be complex. Separating the real and imaginary parts in (12.6.1.5) results in two real particular solutions to system (12.6.1.1); the same two solutions are obtained if one takes the complex conjugate root, $\bar{\lambda} = \alpha - i\beta$.

3°. Let λ be a real root of the characteristic equation (12.6.1.4) of multiplicity m. The corresponding particular solution of system (12.6.1.1) is sought in the form

$$y_1 = P_1(x)e^{\lambda x}, \quad y_2 = P_2(x)e^{\lambda x}, \quad \dots, \quad y_n = P_n(x)e^{\lambda x}, \tag{12.6.1.8}$$

where the $P_k(x) = \sum\limits_{i=0}^{m-1} B_{ki}x^i$ are polynomials of degree $m-1$. The coefficients of these polynomials result from the substitution of expressions (12.6.1.8) into equations (12.6.1.1); after dividing by $e^{\lambda x}$ and collecting like terms, one obtains n equations, each representing a polynomial equated to zero. By equating the coefficients of all resulting polynomials to zero, one arrives at a linear algebraic system of equations for the coefficients B_{ki}; the solution to this system will contain m free parameters.

4°. For a multiple complex, $\lambda = \alpha + i\beta$, of multiplicity m, the corresponding particular solution is sought, just as in the case of a multiple real root, in the form (12.6.1.8); here the coefficients B_{ki} of the polynomials $P_k(x)$ will be complex. Finally, in order to obtain real solutions of the original system (12.6.1.1), one separates the real and imaginary parts in formulas (12.6.1.8), thus obtaining two particular solutions with m free parameters each. The two solutions correspond to the complex conjugate roots $\lambda = \alpha \pm i\beta$.

5°. In the general case, where the characteristic equation (12.6.1.4) has simple and multiple, real and complex roots (see Items 1°–4°), the general solution to system (12.6.1.1) is obtained as the sum of all particular solutions multiplied by arbitrary constants.

Example 1. Consider the homogeneous system of two linear differential equations

$$y_1' = y_1 + 4y_2,$$
$$y_2' = y_1 + y_2.$$

The associated characteristic equations,

$$\begin{vmatrix} 1-\lambda & 4 \\ 1 & 1-\lambda \end{vmatrix} = \lambda^2 - 2\lambda - 3 = 0,$$

has distinct real roots:

$$\lambda_1 = 3, \quad \lambda_2 = -1.$$

The system of algebraic equations (12.6.1.6) for solution coefficients becomes

$$\begin{aligned} (1-\lambda)A_1 + 4A_2 &= 0, \\ A_1 + (1-\lambda)A_2 &= 0. \end{aligned} \tag{12.6.1.9}$$

Substituting the first root, $\lambda = 3$, into system (12.6.1.9) yields $A_1 = 2A_2$. We can set $A_1 = 2$ and $A_2 = 1$, since the solution is determined to within a constant factor. Thus the first particular solution of the homogeneous system of linear ordinary differential equations (12.6.1.9) has the form

$$y_1 = 2e^{3x}, \quad y_2 = e^{3x}. \tag{12.6.1.10}$$

The second particular solution, corresponding to $\lambda = -1$, is found in the same way:

$$y_1 = -2e^{-x}, \quad y_2 = e^{-x}. \tag{12.6.1.11}$$

The sum of the two particular solutions (12.6.1.10), (12.6.1.11) multiplied by arbitrary constants, C_1 and C_2, gives the general solution to the original homogeneous system of linear ordinary differential equations:

$$y_1 = 2C_1e^{3x} - 2C_2e^{-x}, \quad y_2 = C_1e^{3x} + C_2e^{-x}.$$

Example 2. Consider the system of ordinary differential equations

$$\begin{aligned} y_1' &= -y_2, \\ y_2' &= 2y_1 + 2y_2. \end{aligned} \tag{12.6.1.12}$$

The characteristic equation,

$$\begin{vmatrix} -\lambda & -1 \\ 2 & 2-\lambda \end{vmatrix} = \lambda^2 - 2\lambda + 2 = 0$$

has complex conjugate roots:

$$\lambda_1 = 1 + i, \quad \lambda_2 = 1 - i.$$

The algebraic system (12.6.1.6) for the complex coefficients A_1 and A_2 becomes

$$-\lambda A_1 - A_2 = 0,$$
$$2A_1 + (2-\lambda)A_2 = 0.$$

With $\lambda = 1 + i$, one nonzero solution is given by $A_1 = 1$ and $A_2 = -1 - i$. The corresponding complex solution to system (12.6.1.12) has the form

$$y_1 = e^{(1+i)x}, \quad y_2 = (-1-i)e^{(1+i)x}.$$

Separating the real and imaginary parts, taking into account the formulas

$$e^{(1+i)x} = e^x(\cos x + i\sin x) = e^x\cos x + ie^x\sin x,$$
$$(-1-i)e^{(1+i)x} = -(1+i)e^x(\cos x + i\sin x) = e^x(\sin x - \cos x) - ie^x(\sin x + \cos x),$$

and making linear combinations from them, one arrives at the general solution to the original system (12.6.1.12):

$$y_1 = C_1e^x\cos x + C_2e^x\sin x,$$
$$y_2 = C_1e^x(\sin x - \cos x) - C_2e^x(\sin x + \cos x).$$

Remark. Systems of two homogeneous linear constant-coefficient second-order differential equations are treated in detail in Paragraph 12.6.1-7.

12.6.1-3. Nonhomogeneous systems of linear first-order equations.

1°. In general, a nonhomogeneous linear system of constant-coefficient first-order ordinary differential equations has the form

$$\begin{aligned} y_1' &= a_{11}y_1 + a_{12}y_2 + \cdots + a_{1n}y_n + f_1(x),\\ y_2' &= a_{21}y_1 + a_{22}y_2 + \cdots + a_{2n}y_n + f_2(x),\\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ y_n' &= a_{n1}y_1 + a_{n2}y_2 + \cdots + a_{nn}y_n + f_n(x). \end{aligned} \qquad (12.6.1.13)$$

For brevity, the conventional vector notation will also be used:

$$\mathbf{y}' = \mathbf{a}\mathbf{y} + \mathbf{f}(x),$$

where $\mathbf{f}(x) = (f_1(x), f_2(x), \ldots, f_n(x))^{\mathrm{T}}$.

The general solution of this system is the sum of the general solution to the corresponding homogeneous system with $f_k(x) \equiv 0$ [see system (12.6.1.1)] and any particular solution of the nonhomogeneous system (12.6.1.13).

System (12.6.1.13) can also be reduced to a single nonhomogeneous linear constant-coefficient nth-order equation; see Paragraph 12.7.1-3.

2°. Let $\mathbf{y}_m = (D_{m1}(x), D_{m1}(x), \ldots, D_{mn}(x))^{\mathrm{T}}$ be particular solutions to the homogeneous linear system of constant-coefficient first-order ordinary differential equations (12.6.1.1) that satisfy the special initial conditions

$$y_m(0) = 1, \quad y_k(0) = 0 \quad \text{if} \quad k \neq m; \quad m, k = 1, \ldots, n.$$

Then the general solution to the nonhomogeneous system (12.6.1.13) is expressed as

$$y_k(x) = \sum_{m=1}^{n}\int_0^x f_m(t)D_{mk}(x-t)\,dt + \sum_{m=1}^{n} C_mD_{mk}(x), \quad k = 1, \ldots, n. \qquad (12.6.1.14)$$

Alternatively, the general solution to the nonhomogeneous linear system of equations (12.6.1.13) can be obtained using the formulas from Paragraph 12.6.2-2.

The solution of the Cauchy problem for the nonhomogeneous system (12.6.1.13) with arbitrary initial conditions,

$$y_1(0) = y_1^\circ, \quad y_2(0) = y_2^\circ, \quad \ldots, \quad y_n(0) = y_n^\circ, \qquad (12.6.1.15)$$

is determined by formulas (12.6.1.14) with $C_m = y_m^\circ$, $m = 1, \ldots, n$.

12.6.1-4. Homogeneous linear systems of higher-order differential equations.

An arbitrary system of homogeneous linear systems of constant-coefficient ordinary differential equations consists of n equations, each representing a linear combination of unknowns, y_k, and their derivatives, y_k', y_k'', …, $y_k^{(m_k)}$, $k = 1, 2, \ldots, n$.

The general solution of such systems is a linear combination of particular solutions multiplied by arbitrary constants. In total, such a system has $m_1 + m_2 + \cdots + m_n$ linearly independent particular solutions (the system is assumed to be consistent and nondegenerate, so that the constituent equations are linearly independent).

Particular solutions of the system are sought in the form (12.6.1.5). On substituting these expressions into the differential equations and dividing by $e^{\lambda x}$, one obtains a homogeneous linear algebraic system for coefficients $A_1, A_2, \ldots, A_n$. For this system to have nontrivial solutions, the determinant of the system must vanish. This results in an algebraic equation for the exponent λ; in physics, this equation is called a *dispersion equation*. To different roots of the dispersion equation there correspond different particular solutions of the original system of equations. For simple real and complex-conjugate roots, the procedure of finding particular solutions is the same as in the case of a linear system of first-order equations (12.6.1.1).

Example 3. Consider the linear system of constant-coefficient second-order equations

$$\begin{aligned} y_1'' + y_2' + a y_2 &= 0, \\ y_2'' + y_1' + a y_1 &= 0. \end{aligned} \tag{12.6.1.16}$$

Particular solutions are sought in the form

$$y_1 = A_1 e^{\lambda x}, \quad y_2 = A_2 e^{\lambda x}. \tag{12.6.1.17}$$

Substituting (12.6.1.17) into (12.6.1.16) yields a homogeneous linear algebraic system for the coefficients A_1 and A_2:

$$\begin{aligned} \lambda^2 A_1 + (\lambda + a) A_2 &= 0, \\ (\lambda + a) A_1 + \lambda^2 A_2 &= 0. \end{aligned} \tag{12.6.1.18}$$

For this system to have nontrivial solutions, its determinant must vanish. This results in the dispersion equation

$$\begin{vmatrix} \lambda^2 & \lambda + a \\ \lambda + a & \lambda^2 \end{vmatrix} = \lambda^4 - (\lambda + a)^2 = 0.$$

Its roots are

$$\lambda_{1,2} = \tfrac{1}{2} \pm \sqrt{\tfrac{1}{4} + a}, \quad \lambda_{3,4} = -\tfrac{1}{2} \pm \sqrt{\tfrac{1}{4} - a}. \tag{12.6.1.19}$$

Let us confine ourselves to the simplest case of $-\frac{1}{4} < a < \frac{1}{4}$, where all roots of the dispersion equation are real and distinct. It follows from the system of algebraic equations (12.6.1.18) that $A_1 = \lambda + a$ and $A_2 = -\lambda^2$, where $\lambda = \lambda_n$. Substituting these values into (12.6.1.17) yields the particular solutions $y_{1n} = (\lambda_n + a) e^{\lambda_n x}$, $y_{2n} = -\lambda_n^2 e^{\lambda_n x}$ ($n = 1, 2, 3, 4$). A linear combination of the particular solutions gives the general solution of system (12.6.1.16):

$$\begin{aligned} y_1 &= C_1(\lambda_1 + a) e^{\lambda_1 x} + C_2(\lambda_2 + a) e^{\lambda_2 x} + C_3(\lambda_3 + a) e^{\lambda_3 x} + C_4(\lambda_4 + a) e^{\lambda_4 x}, \\ y_2 &= -C_1 \lambda_1^2 e^{\lambda_1 x} - C_2 \lambda_2^2 e^{\lambda_2 x} - C_3 \lambda_3^2 e^{\lambda_3 x} - C_4 \lambda_4^2 e^{\lambda_4 x}, \end{aligned}$$

where C_1, C_2, C_3, and C_4 are arbitrary constants, and the roots λ_n are determined by formulas (12.6.1.19).

Remark. Paragraph 12.6.1-6 (see Item 2°) presents a method for the solution of systems of arbitrary homogeneous linear constant-coefficients ordinary differential equations using the Laplace transform.

12.6.1-5. Nonhomogeneous higher-order linear systems. D'Alembert's method.

Consider the system of two linear constant-coefficient mth-order differential equations

$$\begin{aligned} y_1^{(m)} &= a_{11}y_1 + a_{12}y_2 + f_1(x), \\ y_2^{(m)} &= a_{21}y_1 + a_{22}y_2 + f_2(x). \end{aligned} \tag{12.6.1.20}$$

Let us multiply the second equation of system (12.6.1.20) by k and add it termwise to the first equation to obtain, after rearrangement,

$$(y_1 + ky_2)^{(m)} = (a_{11} + ka_{21})\left(y_1 + \frac{a_{12} + ka_{22}}{a_{11} + ka_{21}}y_2\right) + f_1(x) + kf_2(x). \tag{12.6.1.21}$$

Let us take the constant k so that $\dfrac{a_{12} + ka_{22}}{a_{11} + ka_{21}} = k$, which results in a quadratic equation for k:

$$a_{21}k^2 + (a_{11} - a_{22})k - a_{12} = 0. \tag{12.6.1.22}$$

In this case, (12.6.1.21) is a nonhomogeneous linear constant-coefficient equation for $z = y_1 + ky_2$: $z^{(m)} = (a_{11} + ka_{21})z + f_1(x) + kf_2(x)$. Integrating this equation yields

$$y_1 + ky_2 = C_1\varphi_1(x, k) + \cdots + C_m\varphi_m(x, k) + \psi(x, k).$$

It follows that if the roots of the quadratic equation (12.6.1.22) are distinct, we have two relations,

$$\begin{aligned} y_1 + k_1y_2 &= C_1\varphi_1(x, k_1) + \cdots + C_m\varphi_m(x, k_1) + \psi(x, k_1), \\ y_1 + k_2y_2 &= C_{m+1}\varphi_1(x, k_2) + \cdots + C_{2m}\varphi_m(x, k_2) + \psi(x, k_2), \end{aligned}$$

which represent a linear algebraic system of equations for the functions y_1 and y_2.

Remark 1. The above method for the solution of system (12.6.1.20) is known as *D Alembert's method*. The quantity $z = y_1 + ky_2$ in the above reasoning gives an example of an integrable combination (see Paragraph 12.7.2-1).

Remark 2. The more complicated system where y_1 and y_2 on the right-hand side are replaced by the derivatives of the same order, $y_1^{(n)}$ and $y_2^{(n)}$, can be treated likewise.

Remark 3. System (12.6.1.20) can be solved using the Laplace transform (see Paragraph 12.6.1-6).

12.6.1-6. Usage of the Laplace transform for solving linear systems of equations.

1°. To solve the Cauchy problem for the nonhomogeneous linear system of differential equations (12.6.1.13) with the initial conditions (12.6.1.15), one can use the Laplace transform, based on the following formulas (for details, see Section 11.2):

$$\widetilde{y}_k(p) = \mathfrak{L}\left\{y_k(x)\right\}, \quad \widetilde{f}_k(p) = \mathfrak{L}\left\{f_k(x)\right\}, \quad \text{where} \quad \mathfrak{L}\left\{f(x)\right\} \equiv \int_0^\infty e^{-px}f(x)\,dx.$$

To this end, one should multiply each equation in (12.6.1.13) by e^{-px} and then integrate with respect to x from zero to infinity. In view of the differentiation rule $\mathfrak{L}\left\{y_k'(x)\right\} =$

$p\widetilde{y}_k(p) - y_k(0)$ and the initial conditions (12.6.1.15), one arrives at a nonhomogeneous linear system of algebraic equations for the transforms $\widetilde{y}_k(p)$:

$$\begin{aligned}
&(a_{11}-p)\widetilde{y}_1 + a_{12}\widetilde{y}_2 + \cdots + a_{1n}\widetilde{y}_n = -\widetilde{f}_1(p) - y_1^\circ,\\
&a_{21}\widetilde{y}_1 + (a_{22}-p)\widetilde{y}_2 + \cdots + a_{2n}\widetilde{y}_n = -\widetilde{f}_2(p) - y_2^\circ,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&a_{n1}\widetilde{y}_1 + a_{n2}\widetilde{y}_2 + \cdots + (a_{nn}-p)\widetilde{y}_n = -\widetilde{f}_n(p) - y_n^\circ.
\end{aligned} \tag{12.6.1.23}$$

The solution of this system is obtained by Kramer's rule and is given by

$$\widetilde{y}_k = \frac{\Delta_k(p)}{\Delta(p)}; \qquad k = 1, \ldots, n, \tag{12.6.1.24}$$

where $\Delta(p)$ is the determinant of the basic matrix of system (12.6.1.23), coinciding with the determinant in (12.6.1.4) with $\lambda = p$, and $\Delta_k(p)$ is the determinant of the matrix obtained from the basic matrix by replacing its kth column with the column of free terms of system (12.6.1.23).

On applying the Laplace inversion formula (see Section 11.2) to (12.6.1.24), one obtains a solution to the Cauchy problem (12.6.1.13), (12.6.1.15) in the form

$$y_k(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{\Delta_k(p)}{\Delta(p)} e^{px}\,dp; \qquad k = 1, \ldots, n.$$

The formulas from Paragraph 11.2.2-2 and tables from Section T3.2 can be used to find the inverse Laplace transform of the function $\Delta_k(p)/\Delta(p)$.

2°. The Laplace transform is also suitable for the solution of systems of constant-coefficient second- and higher-order ordinary differential equations.

Example 4. Consider the Cauchy problem for the nonhomogeneous linear system of constant-coefficient second-order differential equations

$$\sum_{k=1}^{n}\left(a_{mk}y_k'' + b_{mk}y_k' + c_{mk}y_k\right) = f_m(x), \qquad m = 1, 2, \ldots, n,$$

subject to the initial conditions

$$y_k(0) = \alpha_k, \quad y_k'(0) = \beta_k, \qquad k = 1, 2, \ldots, n.$$

The Laplace transform reduces this problem to a linear system of algebraic equations for the transform $\widetilde{y}_k(p)$:

$$\sum_{k=1}^{n}\left(a_{mk}p^2 + b_{mk}p + c_{mk}\right)\widetilde{y}_k(p) = \widetilde{f}_m(p) + \sum_{k=1}^{n}\left[(a_{mk}p + b_{mk})\alpha_k + \beta_k\right], \qquad m = 1, 2, \ldots, n.$$

The solution to this system can be obtained using Kramer's rule. By applying then the inverse Laplace transform to the resulting expressions of $\widetilde{y}_k(p)$, one obtains the solution to the Cauchy problem.

12.6.1-7. Classification of equilibrium points of two-dimensional linear systems.

Let us study the behavior of solutions near the equilibrium (also called stationary, steady-state, or fixed) point $x = y = 0$ for the system of two homogeneous linear constant-coefficient equations

$$\begin{aligned}
x_t' &= a_{11}x + a_{12}y,\\
y_t' &= a_{21}x + a_{22}y.
\end{aligned} \tag{12.6.1.25}$$

By convention, for clearness and convenience of interpretation of the results, t will be used to designate the independent variable and will be treated as time. A solution $x = x(t)$, $y = y(t)$ of system (12.6.1.25) plotted in the plane x, y (the phase plane) is called a (phase) trajectory of the system.

A solution to system (12.6.1.25) will be sought in the form

$$x = k_1 e^{\lambda t}, \quad y = k_2 e^{\lambda t}. \tag{12.6.1.26}$$

On substituting (12.6.1.26) into (12.6.1.25), one obtains the characteristic equation for the exponent λ:

$$\begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = 0, \quad \text{or} \quad \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21} = 0. \tag{12.6.1.27}$$

The coefficients k_1 and k_2 are found as

$$k_1 = Ca_{12}, \quad k_2 = C(\lambda - a_{11}), \tag{12.6.1.28}$$

where C is an arbitrary constant. To two different roots of the quadratic equation (12.6.1.27) there correspond two pairs of coefficients (12.6.1.28).

Denote the discriminant of the quadratic equation (12.6.1.27) by

$$D = (a_{11} - a_{22})^2 + 4a_{12}a_{21}. \tag{12.6.1.29}$$

Three situations are possible.

1°. If $D > 0$, the roots of the characteristic equation (12.6.1.27) are real and distinct ($\lambda_1 \neq \lambda_2$):

$$\lambda_{1,2} = \tfrac{1}{2}(a_{11} + a_{22}) \pm \tfrac{1}{2}\sqrt{D}.$$

The general solution of system (12.6.1.25) is expressed as

$$\begin{aligned} x &= C_1 a_{12} e^{\lambda_1 t} + C_2 a_{12} e^{\lambda_2 t}, \\ y &= C_1(\lambda_1 - a_{11}) e^{\lambda_1 t} + C_2(\lambda_2 - a_{11}) e^{\lambda_2 t}, \end{aligned} \tag{12.6.1.30}$$

where C_1 and C_2 are arbitrary constants. For $C_1 = 0$, $C_2 \neq 0$ and $C_2 = 0$, $C_1 \neq 0$, the trajectories in the phase plane x, y are straight lines. Four cases are possible here.

1.1. Two negative real roots, $\lambda_1 < 0$ and $\lambda_2 < 0$. This corresponds to $a_{11} + a_{22} < 0$ and $a_{11}a_{22} - a_{12}a_{21} > 0$. The equilibrium point is asymptotically stable and all trajectories starting within a small neighborhood of the origin tend to the origin as $t \to \infty$. To $C_1 = 0$, $C_2 \neq 0$ and $C_2 = 0$, $C_1 \neq 0$ there correspond straight lines passing through the origin. Figure 12.4 *a* depicts the arrangement of the phase trajectories near an equilibrium point called a *stable node* (or a *sink*). The direction of motion along the trajectories with increasing t is shown by arrows.

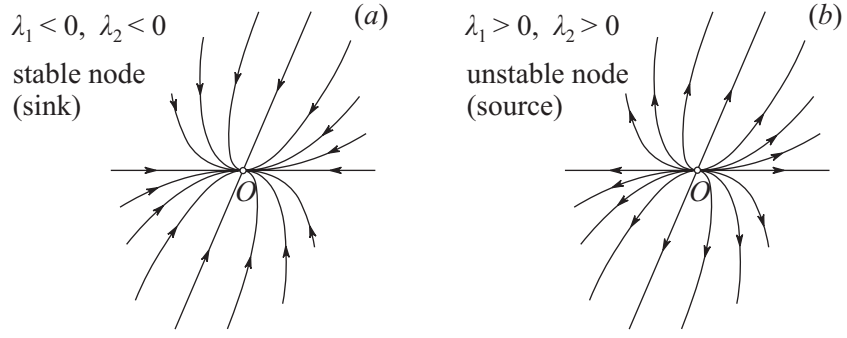


Figure 12.4. Phase trajectories of a system of differential equations near an equilibrium point of the node type.

1.2. $\lambda_1 > 0$ and $\lambda_2 > 0$. This corresponds to $a_{11} + a_{22} > 0$ and $a_{11}a_{22} - a_{12}a_{21} > 0$. The phase trajectories in the vicinity of the equilibrium point have the same pattern as in the preceding case; however, the trajectories go in the opposite direction, away from the equilibrium point; see Fig. 12.4 *b*. An equilibrium point of this type is called an *unstable node* (or a *source*).

1.3. $\lambda_1 > 0$ and $\lambda_2 < 0$ (or $\lambda_1 < 0$ and $\lambda_2 > 0$). This corresponds to $a_{11}a_{22} - a_{12}a_{21} < 0$. In this case, the equilibrium point is also unstable, since the trajectory (12.6.1.30) with $C_2 = 0$ goes beyond a small neighborhood of the origin as t increases. If $C_1C_2 \neq 0$, then the trajectories leave the neighborhood of the origin as $t \to -\infty$ and $t \to \infty$. An equilibrium point of this type is called a *saddle* (or a *hyperbolic point*); see Fig. 12.5.

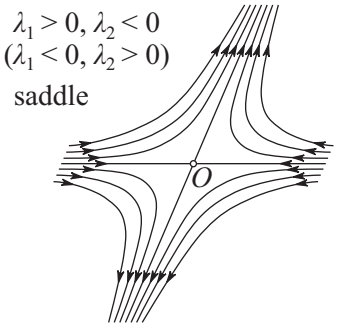


Figure 12.5. Phase trajectories of a system of differential equations near an equilibrium point of the saddle type.

1.4. $\lambda_1 = 0$ and $\lambda_2 = a_{11} + a_{22} \neq 0$. This corresponds to $a_{11}a_{22} - a_{12}a_{21} = 0$. The general solution of system (12.6.1.25) is expressed as

$$\begin{aligned} x &= C_1a_{12} + C_2a_{12}e^{(a_{11}+a_{22})t}, \\ y &= -C_1a_{11} + C_2a_{22}e^{(a_{11}+a_{22})t}, \end{aligned} \qquad (12.6.1.31)$$

where C_1 and C_2 are arbitrary constants. By eliminating time t from (12.6.1.31), one obtains a family of parallel lines defined by the equation $a_{22}x - a_{12}y = a_{12}(a_{11} + a_{22})C_1$. To $C_2 = 0$ in (12.6.1.31) there corresponds a one-parameter family of equilibrium points that lie on the straight line $a_{11}x + a_{12}y = 0$.

(i) If $\lambda_2 < 0$, then the trajectories approach the equilibrium point lying as $t \to \infty$; see Fig. 12.6. The equilibrium point $x = y = 0$ is stable (or neutrally stable) — there is no asymptotic stability.

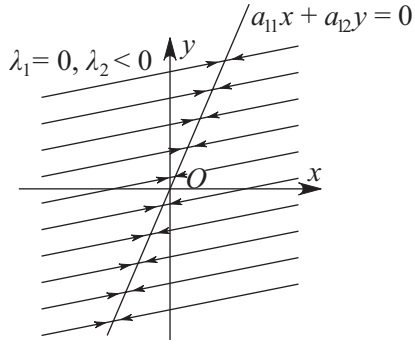


Figure 12.6. Phase trajectories of a system of differential equations near a set of equilibrium points located on a straight line.

(ii) If $\lambda_2 > 0$, the trajectories have the same pattern as in case (i), but they go, as $t \to \infty$, in the opposite direction, away from the equilibrium point. The point $x = y = 0$ is unstable.

2°. If $D < 0$, the characteristic equation (12.6.1.27) has complex-conjugate roots:

$$\lambda_{1,2} = \alpha \pm i\beta, \qquad \alpha = \tfrac{1}{2}(a_{11} + a_{22}), \quad \beta = \tfrac{1}{2}\sqrt{|D|}, \quad i^2 = -1.$$

The general solution of system (12.6.1.25) has the form

$$\begin{aligned} x &= e^{\alpha t}[C_1 \cos(\beta t) + C_2 \sin(\beta t)], \\ y &= e^{\alpha t}[C_1^* \cos(\beta t) + C_2^* \sin(\beta t)], \end{aligned} \tag{12.6.1.32}$$

where C_1 and C_2 are arbitrary constants, and C_1^* and C_2^* are defined by linear combinations of C_1 and C_2. The following cases are possible.

2.1. For $\alpha < 0$, the trajectories in the phase plane are spirals asymptotically approaching the origin of coordinates (the equilibrium point) as $t \to \infty$; see Fig. 12.7 *a*. Therefore the equilibrium point is asymptotically stable and is called a *stable focus* (also a *stable spiral point* or a *spiral sink*). A focus is characterized by the fact that the tangent to a trajectory changes its direction all the way to the equilibrium point.

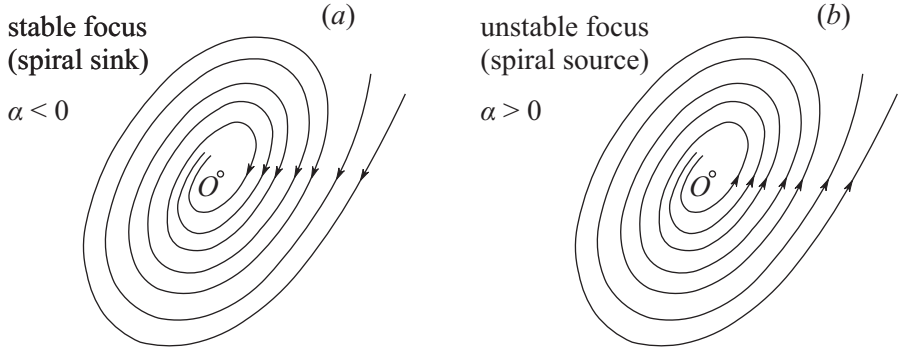


Figure 12.7. Phase trajectories of a system of differential equations near an equilibrium point of the focus type.

2.2. For $\alpha > 0$, the phase trajectories are also spirals, but unlike the previous case they spiral away from the origin as $t \to \infty$; see Fig. 12.7 *b*. Therefore such an equilibrium point is called an *unstable focus* (also an *unstable spiral point* or a *spiral source*).

2.3. At $\alpha = 0$, the phase trajectories are closed curves, containing the equilibrium point inside (see Fig. 12.8). Such an equilibrium point is called a *center*. A center is a stable equilibrium point. Note that there is no asymptotic stability in this case.

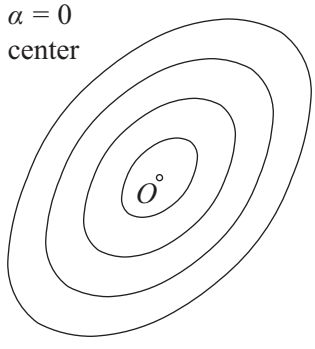


Figure 12.8. Phase trajectories of a system of differential equations near an equilibrium point of the center type.

3°. If $D = 0$, the characteristic equation (12.6.1.27) has a double real root, $\lambda_1 = \lambda_2 = \frac{1}{2}(a_{11} + a_{22})$. The following cases are possible.

3.1. If $\lambda_1 = \lambda_2 = \lambda < 0$, the general solution of system (12.6.1.25) has the form

$$\begin{aligned} x &= a_{12}(C_1 + C_2 t)e^{\lambda t}, \\ y &= [(\lambda - a_{11})C_1 + C_2 + C_2(\lambda - a_{11})t]e^{\lambda t}, \end{aligned} \tag{12.6.1.33}$$

where C_1 and C_2 are arbitrary constants.

Since there is a rapidly decaying factor, $e^{\lambda t}$, all trajectories tend to the equilibrium point as $t \to \infty$; see Fig. 12.9 *a*. To $C_2 = 0$ there corresponds a straight line in the phase plane x, y. The equilibrium point is asymptotically stable and is called a *stable node* (a *sink*). Such a node is in intermediate position between a node from Item 1.1 and a focus from Item 2.1.

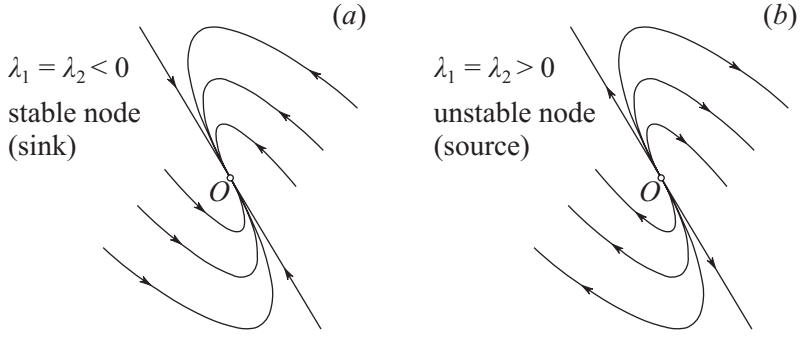


Figure 12.9. Phase trajectories of a system of differential equations near an equilibrium point of the node type in the case of a double root, $\lambda_1 = \lambda_2$.

3.2. If $\lambda_1 = \lambda_2 = \lambda > 0$, the general solution of system (12.6.1.25) is determined by formulas (12.6.1.33). The phase trajectories are similar to those from Item 3.1, but they go in the opposite direction, as $t \to \infty$, rapidly away from the equilibrium point. Such an equilibrium point is called an *unstable node* (a *source*); see Fig. 12.9 *b*.

3.3. If $\lambda_1 = \lambda_2 = 0$, which corresponds to

$$a_{11} + a_{22} = 0 \quad \text{and} \quad a_{11}a_{22} - a_{12}a_{21} = 0$$

simultaneously, the general solution of system (12.6.1.25) is obtained by substituting $\lambda = 0$ into (12.6.1.33) and has the form

$$\begin{aligned} x &= a_{12}C_1 + a_{12}C_2 t, \\ y &= C_2 - a_{11}C_1 - a_{11}C_2 t. \end{aligned}$$

For $a_{12} \neq 0$ all trajectories are parallel straight lines. As $t \to \pm\infty$, the trajectories go away from the origin. The equilibrium point is unstable.

For clearness, the classification results for equilibrium points of systems of two linear constant-coefficient differential equations (12.6.1.25) are summarized in Table 12.5.

Remark. For general definitions of a stable and an unstable equilibrium point, see Section 7.3.

TABLE 12.5
Classification of equilibrium points for systems of constant-coefficient equations (12.6.1.25); the symbols ∘ and ∗ indicate stable and unstable equilibrium points, respectively, where not clearly stated

Discriminant D, formula (12.6.1.29)	Roots of quadratic equation (12.6.1.27), λ_1 and λ_2	Conditions for coefficients of differential equations (12.6.1.25)	Type of equilibrium points or shape of phase trajectories
$D>0$	$\lambda_1<0,\ \lambda_2<0,\ \lambda_1\neq\lambda_2$ $\lambda_1>0,\ \lambda_2>0,\ \lambda_1\neq\lambda_2$ roots have unlike signs $\lambda_1=0,\ \lambda_2<0$ $\lambda_1=0,\ \lambda_2>0$	$a_{11}+a_{22}<0,\ a_{11}a_{22}-a_{12}a_{21}>0$ $a_{11}+a_{22}>0,\ a_{11}a_{22}-a_{12}a_{21}>0$ $a_{11}a_{22}-a_{12}a_{21}<0$ $a_{11}+a_{22}<0,\ a_{11}a_{22}-a_{12}a_{21}=0$ $a_{11}+a_{22}>0,\ a_{11}a_{22}-a_{12}a_{21}=0$	stable node unstable node saddle∗ parallel lines∘ parallel lines∗
$D<0$	$\lambda_{1,2}=\alpha\pm i\beta,\ \alpha>0$ $\lambda_{1,2}=\alpha\pm i\beta,\ \alpha<0$ $\lambda_{1,2}=\pm i\beta$, imaginary roots	$a_{11}+a_{22}<0,\ (a_{11}-a_{22})^2+4a_{12}a_{21}<0$ $a_{11}+a_{22}>0,\ (a_{11}-a_{22})^2+4a_{12}a_{21}<0$ $a_{11}+a_{22}=0,\ a_{11}a_{22}-a_{12}a_{21}>0$	stable focus unstable focus center∘
$D=0$	$\lambda_1=\lambda_2<0$ $\lambda_1=\lambda_2>0$ $\lambda_1=\lambda_2=0$	$a_{11}+a_{22}<0,\ (a_{11}-a_{22})^2+4a_{12}a_{21}=0$ $a_{11}+a_{22}>0,\ (a_{11}-a_{22})^2+4a_{12}a_{21}=0$ $a_{11}+a_{22}=0,\ a_{11}a_{22}-a_{12}a_{21}=0$	stable node unstable node saddle∗ parallel lines∗

12.6.2. Systems of Linear Variable-Coefficient Equations

12.6.2-1. Homogeneous systems of linear first-order equations.

1°. In general, a homogeneous linear system of variable-coefficient first-order ordinary differential equations has the form

$$\begin{aligned} y_1' &= f_{11}(x)y_1 + f_{12}(x)y_2 + \cdots + f_{1n}(x)y_n,\\ y_2' &= f_{21}(x)y_1 + f_{22}(x)y_2 + \cdots + f_{2n}(x)y_n,\\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ y_n' &= f_{n1}(x)y_1 + f_{n2}(x)y_2 + \cdots + f_{nn}(x)y_n, \end{aligned} \tag{12.6.2.1}$$

where the prime denotes a derivative with respect to x. It is assumed further on that the functions $f_{ij}(x)$ are continuous of an interval $a \le x \le b$ (intervals are allowed with $a=-\infty$ or/and $b=+\infty$).

Any homogeneous linear system of the form (12.6.2.1) has the trivial particular solution $y_1 = y_2 = \cdots = y_n = 0$.

Superposition principle for a homogeneous system: any linear combination of particular solutions to system (12.6.2.1) is also a solution to this system.

2°. Let

$$\mathbf{y}_k = (y_{k1}, y_{k2}, \ldots, y_{kn})^{\mathrm{T}}, \qquad y_{km} = y_{km}(x); \qquad k,\ m = 1,\ 2,\ \ldots,\ n \tag{12.6.2.2}$$

be nontrivial particular solutions of the homogeneous system of equations (12.6.2.1). Solutions (12.6.2.2) are linearly independent if the *Wronskian determinant* is nonzero:

$$W(x) \equiv \begin{vmatrix} y_{11}(x) & y_{12}(x) & \cdots & y_{1n}(x)\\ y_{21}(x) & y_{22}(x) & \cdots & y_{2n}(x)\\ \vdots & \vdots & \ddots & \vdots\\ y_{n1}(x) & y_{n2}(x) & \cdots & y_{nn}(x) \end{vmatrix} \neq 0. \tag{12.6.2.3}$$

If condition (12.6.2.3) is satisfied, the general solution of the homogeneous system (12.6.2.1) is expressed as

$$\mathbf{y} = C_1\mathbf{y}_1 + C_2\mathbf{y}_2 + \cdots + C_n\mathbf{y}_n, \tag{12.6.2.4}$$

where $C_1,\ C_2,\ \ldots,\ C_n$ are arbitrary constants. The vector functions $\mathbf{y}_1,\ \mathbf{y}_2,\ \ldots,\ \mathbf{y}_n$ in (12.6.2.4) are called *fundamental solutions* of system (12.6.2.1).

3°. Suppose condition (12.6.2.3) is met. Then the *Liouville formula*

$$W(x) = W(x_0)\exp\left[\int_{x_0}^{x}\left(\sum_{s=1}^{n} f_{ss}(t)\right)dt\right]$$

holds.

COROLLARY. *Particular solutions (12.6.2.2) are linearly independent on the interval* $[a, b]$ *if and only if there exists a point* $x_0 \in [a, b]$ *such that the Wronskian determinant is nonzero at* x_0*:* $W(x_0) \neq 0$.

4°. Suppose a nontrivial particular solution of system (12.6.2.1),

$$\mathbf{y}_1 = (u_1, u_2, \ldots, u_n)^{\mathrm{T}}, \quad u_m = u_m(x), \quad m = 1, 2, \ldots, n,$$

is known. Then the number of unknowns can be reduced by one. To this end, one considers the auxiliary homogeneous linear system of $n - 1$ equations

$$z_k' = \sum_{q=2}^{n}\left[f_{kq}(x) - \frac{u_k(x)}{u_1(x)}f_{1q}(x)\right]z_q, \qquad k = 2, \ldots, n. \tag{12.6.2.5}$$

Let

$$\mathbf{z}_p = (z_{p2}, z_{p3}, \ldots, z_{pn})^{\mathrm{T}}, \quad z_{mk} = z_{mk}(x); \quad p = 2, \ldots, n,$$

be a fundamental system of solutions to system (12.6.2.5) and let

$$\mathbf{Z}_p = (0, z_{p2}, z_{p3}, \ldots, z_{pn})^{\mathrm{T}}, \quad F_p(x) = \int\left(\frac{1}{u_1(x)}\sum_{s=2}^{n} f_{1s}(x)z_{ps}(x)\right)dx, \quad p = 2, \ldots, n,$$

the vector $\mathbf{Z}_p$ having an additional component compared with $\mathbf{z}_p$. Then the vector functions $\mathbf{y}_p = F_p(x)\mathbf{y}_1 + \mathbf{z}_p$ $(p = 2, \ldots, n)$ together with $\mathbf{y}_1$ form a fundamental system of solutions to the initial homogeneous system of equations (12.6.2.1).

12.6.2-2. Nonhomogeneous systems of linear first-order equations.

1°. In general, a nonhomogeneous linear system of variable-coefficient first-order differential equations has the form

$$\begin{aligned}
y_1' &= f_{11}(x)y_1 + f_{12}(x)y_2 + \cdots + f_{1n}(x)y_n + g_1(x),\\
y_2' &= f_{21}(x)y_1 + f_{22}(x)y_2 + \cdots + f_{2n}(x)y_n + g_2(x),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
y_n' &= f_{n1}(x)y_1 + f_{n2}(x)y_2 + \cdots + f_{nn}(x)y_n + g_n(x).
\end{aligned} \tag{12.6.2.6}$$

Alternatively, the system can be written in the short vector-matrix notation as

$$\mathbf{y}' = \mathbf{f}(x)\mathbf{y} + \mathbf{g}(x),$$

with $\mathbf{f}(x) = \big(f_{ij}(x)\big)$ being the matrix of equation coefficients and $\mathbf{g}(x) = \big(g_1(x), g_2(x), \ldots, g_n(x)\big)^{\mathrm{T}}$ being the vector function defining the nonhomogeneous part of the equations.

EXISTENCE AND UNIQUENESS THEOREM. *Let the functions $f_{ij}(x)$ and $g_i(x)$ be continuous on an interval $a < x < b$. Then, for any set of values $x^\circ, y_1^\circ, \ldots, y_n^\circ$, where $a < x^\circ < b$, there exists a unique solution $y_1 = y_1(x), \ldots, y_n = y_n(x)$ satisfying the initial conditions*

$$y_1(x^\circ) = y_1^\circ, \quad \ldots, \quad y_n(x^\circ) = y_n^\circ,$$

and this solution is defined on the whole interval $a < x < b$.

2°. Let

$$\bar{\mathbf{y}} = (\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_n)^{\mathrm{T}}, \quad \bar{y}_k = \bar{y}_k(x); \quad k = 1, 2, \ldots, n,$$

be a particular solution to the nonhomogeneous system of equations (12.6.2.6). The general solution of this system is the sum of the general solution of the corresponding homogeneous system (12.6.2.1), which corresponds to $g_k(x) \equiv 0$ in (12.6.2.6), and any particular solution of the nonhomogeneous system (12.6.2.6), or

$$\mathbf{y} = C_1\mathbf{y}_1 + C_2\mathbf{y}_2 + \cdots + C_n\mathbf{y}_n + \bar{\mathbf{y}}, \tag{12.6.2.7}$$

where $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ are linearly independent solutions of the homogeneous system (12.6.2.1).

3°. Given a fundamental system of solutions $y_{km}(x)$ (12.6.2.2) of the homogeneous system (12.6.2.1), a particular solution of the nonhomogeneous system (12.6.2.6) is found as

$$\bar{y}_k = \sum_{m=1}^{n} y_{mk}(x) \int \frac{W_m(x)}{W(x)}\, dx, \qquad k = 1, 2, \ldots, n,$$

where $W_m(x)$ is the determinant obtained by replacing the mth row in the Wronskian determinant (12.6.2.3) by the row of free terms, $g_1(x), g_2(x), \ldots, g_n(x)$, of equation (12.6.2.6). The general solution of the nonhomogeneous system (12.6.2.6) is given by (12.6.2.7).

4°. *Superposition principle for a nonhomogeneous system.* A particular solution, $\mathbf{y} = \bar{\mathbf{y}}$, of the nonhomogeneous system of linear differential equations,

$$\mathbf{y}' = \mathbf{f}(x)\mathbf{y} + \sum_{k=1}^{m} \mathbf{g}_k(x),$$

is given by the sum

$$\mathbf{y} = \sum_{k=1}^{m} \mathbf{y}_k,$$

where the $\mathbf{y}_k$ are particular solutions of m (simpler) systems of equations

$$\mathbf{y}_k' = \mathbf{f}(x)\mathbf{y}_k + \mathbf{g}_k(x), \qquad k = 1, 2, \ldots, m,$$

corresponding to individual nonhomogeneous terms of the original system.

12.6.2-3. Euler system of ordinary differential equations.

1°. A homogeneous Euler system is a homogeneous linear system of ordinary differential equations composed by linear combinations of the following terms:

$$y_k, \quad xy_k', \quad x^2y_k'', \quad \ldots, \quad x^{m_k}y_k^{(m_k)}; \qquad k = 1, 2, \ldots, n.$$

Such a system is invariant under scaling in the independent variable (i.e., it preserves its form under the change of variable $x \to \alpha x$, where α is any nonzero number). A nonhomogeneous Euler system contains additional terms, given functions.

The substitution $x = be^t$ ($b \neq 0$) brings an Euler system, both homogeneous and nonhomogeneous, to a constant-coefficient linear system of equations.

Example. In general, a nonhomogeneous Euler system of second-order equations has the form

$$\sum_{k=1}^{n}\left(a_{mk}x^2\frac{d^2y_k}{dx^2} + b_{mk}x\frac{dy_k}{dx} + c_{mk}y_k\right) = f_m(x), \qquad m = 1, 2, \ldots, n. \tag{12.6.2.8}$$

The substitutions $x = \pm e^t$ bring this system to a constant-coefficient linear system,

$$\sum_{k=1}^{n}\left[a_{mk}\frac{d^2y_k}{dt^2} + (b_{mk} - a_{mk})\frac{dy_k}{dt} + c_{mk}y_k\right] = f_m(\pm e^t), \qquad m = 1, 2, \ldots, n,$$

which can be solved using, for example, the Laplace transform (see Example 4 from Paragraph 12.6.1-6).

2°. Particular solutions to a homogeneous Euler system (for system (12.6.2.8), corresponding to $f_m(x) \equiv 0$) are sought in the form of power functions:

$$y_1 = A_1x^\sigma, \quad y_2 = A_2x^\sigma, \quad \ldots, \quad y_n = A_nx^\sigma, \tag{12.6.2.9}$$

where the coefficients $A_1, A_2, \ldots, A_n$ are determined by solving the associated homogeneous system of algebraic equations obtained by substituting expressions (12.6.2.9) into the differential equations of the system in question and dividing by x^σ. Since the system is homogeneous, for it to have nontrivial solutions, its determinant must vanish. This results in a dispersion equation for the exponent σ.

12.7. Nonlinear Systems of Ordinary Differential Equations

12.7.1. Solutions and First Integrals. Uniqueness and Existence Theorems

12.7.1-1. Systems solved for the derivative. A solution and the general solution.

We will be dealing with a system of first-order ordinary differential equations solved for the derivatives

$$y_k' = f_k(x, y_1, \ldots, y_n), \qquad k = 1, \ldots, n. \tag{12.7.1.1}$$

Here and henceforth throughout the current chapter, the prime denotes a derivative with respect to the independent variable x.

A set of numbers $x, y_1, \ldots, y_n$ is convenient to treat as a point in the $(n+1)$-dimensional space.

For brevity, system (12.7.1.1) is conventionally written in vector form:

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}),$$

where $\mathbf{y}$ and $\mathbf{f}$ are the vectors defined as $\mathbf{y} = (y_1, \ldots, y_n)^T$ and $\mathbf{f} = (f_1, \ldots, f_n)^T$.

A *solution* (also an *integral* or an *integral curve*) *of a system of differential equations* (12.7.1.1) is a set of functions $y_1 = y_1(x), \ldots, y_n = y_n(x)$ such that, when substituted into all equations (12.7.1.1), they turn them into identities. The *general solution of a system of differential equations* is the set of all its solutions. In the general case, the general solution of system (12.7.1.1) depends on n arbitrary constants.

12.7.1-2. Existence and uniqueness theorems.

THEOREM (EXISTENCE, PEANO). *Let the functions $f_k(x, y_1, \ldots, y_n)$ $(k = 1, \ldots, n)$ be continuous in a domain G of the $(n+1)$-dimensional space of the variables $x, y_1, \ldots, y_n$. Then there is at least one integral curve passing through every point $M(x^\circ, y_1^\circ, \ldots, y_n^\circ)$ in G. Each of such curves can be extended on both ends up to the boundary of any closed domain completely belonging to G and containing the point M inside.*

Remark. If there is more than one integral curve passing through the point M, there are infinitely many integral curves passing through M.

THEOREM (UNIQUENESS). *There is a unique integral curve passing through the point $M(x^\circ, y_1^\circ, \ldots, y_n^\circ)$ if the functions f_k have partial derivatives with respect to all y_m, continuous in $x, y_1, \ldots, y_n$ in the domain G, or if each function f_k in G satisfies the Lipschitz condition:*

$$|f_k(x, \bar{y}_1, \ldots, \bar{y}_n) - f_k(x, y_1, \ldots, y_n)| \le A \sum_{m=1}^{n} |\bar{y}_m - y_m|,$$

where A is some positive number.

12.7.1-3. Reduction of systems of equations to a single equation.

Suppose the right-hand sides of equations (12.7.1.1) are n times differentiable in all variables. Then system (12.7.1.1) can be reduced to a single nth-order equation. Indeed, using the chain rule, let us differentiate the first equation of system (12.7.1.1) with respect to x to get

$$y_1'' = \frac{\partial f_1}{\partial x_1} + \frac{\partial f_1}{\partial y_1} y_1' + \cdots + \frac{\partial f_1}{\partial y_n} y_n'. \tag{12.7.1.2}$$

Then change the first derivatives y_k' in (12.7.1.2) to $f_k(x, y_1, \ldots, y_n)$ [the right-hand sides of equations (12.7.1.1)] to obtain

$$y_1'' = F_2(x, y_1, \ldots, y_n), \tag{12.7.1.3}$$

where $F_2(x, y_1, \ldots, y_n) \equiv \dfrac{\partial f_1}{\partial x_1} + \dfrac{\partial f_1}{\partial y_1} f_1 + \cdots + \dfrac{\partial f_1}{\partial y_n} f_n$. Now differentiate equation (12.7.1.3) with respect to x and replace the first derivatives y_k' on the right-hand side of the resulting equation by f_k. As a result, we obtain

$$y_1''' = F_3(x, y_1, \ldots, y_n),$$

where $F_3(x, y_1, \ldots, y_n) \equiv \dfrac{\partial F_2}{\partial x_1} + \dfrac{\partial F_2}{\partial y_1} f_1 + \cdots + \dfrac{\partial F_2}{\partial y_n} f_n$. Repeating this procedure as many times as required, one arrives at the following system of equations:

$$\begin{aligned}
y_1' &= F_1(x, y_1, \ldots, y_n),\\
y_1'' &= F_2(x, y_1, \ldots, y_n),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
y_1^{(n)} &= F_n(x, y_1, \ldots, y_n),
\end{aligned}$$

where

$$F_1(x, y_1, \ldots, y_n) \equiv f_1(x, y_1, \ldots, y_n), \quad F_{k+1}(x, y_1, \ldots, y_n) \equiv \frac{\partial F_k}{\partial x_1} + \frac{\partial F_k}{\partial y_1} f_1 + \cdots + \frac{\partial F_k}{\partial y_n} f_n.$$

Expressing $y_2, y_3, \ldots, y_n$ from the $n-1$ first equations of this system in terms of $x, y_1, y_1', \ldots, y_1^{(n-1)}$ and then substituting the resulting expressions into the last equation of system (12.7.1.1), one finally arrives at an nth-order equation:

$$y_1^{(n)} = \Phi(x, y_1, y_1', \ldots, y_1^{(n-1)}). \tag{12.7.1.4}$$

Remark 1. If (12.7.1.1) is a linear system of first-order differential equations, then (12.7.1.4) is a linear nth-order equation.

Remark 2. Any equation of the form (12.7.1.4) can be reduced to a system on n first-order equations (see Paragraph 12.5.1-4).

12.7.1-4. First integrals. Using them to reduce system dimension.

1°. A relation of the form

$$\Psi(x, y_1, \ldots, y_n) = C, \tag{12.7.1.5}$$

where C is an arbitrary constant, is called a *first integral* of system (12.7.1.1) if its left-hand side Φ, generally not identically constant, is turned into a constant by any particular solution, $y_1, \ldots, y_n$, of system (12.7.1.1). In the sequel, we consider only continuously differentiable functions $\Psi(x, y_1, \ldots, y_n)$ in a given domain of variation of its arguments.

THEOREM. *An expression of the form (12.7.1.5) is a first integral of system (12.7.1.1) if and only if the function $\Psi = \Psi(x, y_1, \ldots, y_n)$ satisfies the relation*

$$\frac{\partial \Psi}{\partial x} + \sum_{k=1}^{n} \frac{\partial \Psi}{\partial y_k} f_k(x, y_1, \ldots, y_n) = 0.$$

This relation may be treated as a first-order partial differential equation for Ψ.

Different first integrals of system (12.7.1.1) are called *independent* if the Jacobian of their left-hand sides is nonzero.

System (12.7.1.1) admits n independent first integrals if the conditions of the uniqueness theorem from Paragraph 12.7.1-2 are met.

2°. Given a first integral (12.7.1.5) of system (12.7.1.1), it may be treated as an implicit specification of one of the unknowns. Solving (12.7.1.5), for example, for y_n yields $y_n = G(x, y_1, \ldots, y_{n-1})$. Substituting this expression into the first $n-1$ equations of system (12.7.1.1), one obtains a system in $n-1$ variables with one arbitrary constant.

Likewise, given m independent first integrals of system (12.7.1.1),

$$\Psi_k(x, y_1, \ldots, y_n) = C_k, \quad k = 1, \ldots, m \quad (m < n),$$

the system may be reduced to a system of $n-m$ first-order equations in $n-m$ unknowns.

12.7.2. Integrable Combinations. Autonomous Systems of Equations

12.7.2-1. Integrable combinations.

In some cases, first integrals of systems of differential equations may be obtained by finding *integrable combinations*. An integrable combination is a differential equation that is easy to integrate and is a consequence of the equations of the system under consideration. Most commonly, an integrable combination is an equation of the form

$$d\,\Psi(x, y_1, \ldots, y_n) = 0 \tag{12.7.2.1}$$

or an equation reducible by a change of variables to one of the integrable types of equations in one unknown.

Example 1. Consider the nonlinear system

$$ay_1' = (b-c)y_2y_3, \quad by_2' = (c-a)y_1y_3, \quad cy_3' = (a-b)y_1y_2, \tag{12.7.2.2}$$

where a, b, and c are some constants. Such systems arise in the theory of motion of a rigid body.

Let us multiply the first equation by y_1, the second by y_2, and the third by y_3 and add together to obtain

$$ay_1y_1' + by_2y_2' + cy_3y_3' = 0 \quad \Longrightarrow \quad d(ay_1^2 + by_2^2 + cy_3^2) = 0.$$

Integrating yields a first integral:

$$ay_1^2 + by_2^2 + cy_3^2 = C_1. \tag{12.7.2.3}$$

Now multiply the first equation of the system by ay_1, the second by by_2, and the third by cy_3 and add together to obtain

$$a^2y_1y_1' + b^2y_2y_2' + c^2y_3y_3' = 0 \quad \Longrightarrow \quad d(a^2y_1^2 + b^2y_2^2 + c^2y_3^2) = 0.$$

Integrating yields another first integral:

$$a^2y_1^2 + b^2y_2^2 + c^2y_3^2 = C_2. \tag{12.7.2.4}$$

If the case $a = b = c$, where system (12.7.2.2) can be integrated directly, does not take place, the above two first integrals (12.7.2.3) and (12.7.2.4) are independent. Hence, using them, one can express y_2 and y_3 in terms of y_1 and then substitute the resulting expressions into the first equation of system (12.7.2.2). As a result, one arrives at a single separable first-order differential equation for y_1.

In this example, the integrable combinations have the form (12.7.2.1).

Example 2. A specific example of finding an integrable combination reducible with a change of variables to a simpler, integrable equation in one unknown can be found in Paragraph 12.6.1-5.

12.7.2-2. Autonomous systems and their reduction to systems of lower dimension.

1°. A system of equations is called *autonomous* if the right-hand sides of the equations do not depend explicitly on x. In general, such systems have the form

$$y_k' = f_k(y_1, \ldots, y_n), \qquad k = 1, \ldots, n. \tag{12.7.2.5}$$

If $\mathbf{y}(x)$ is a solution of the autonomous system (12.7.2.5), then the function $\mathbf{y}(x + C)$, where C is an arbitrary constant, is also a solution of this system.

A point $\mathbf{y}^\circ = (y_1^\circ, \ldots, y_n^\circ)$ is called an *equilibrium point* (or a *stationary point*) of the autonomous system (12.7.2.5) if

$$f_k(y_1^\circ, \ldots, y_n^\circ) = 0, \qquad k = 1, \ldots, n.$$

To an equilibrium point there corresponds a special, simplest particular solution when all unknowns are constant:

$$y_1 = y_1^\circ, \quad \ldots, \quad y_n = y_n^\circ, \qquad k = 1, \ldots, n.$$

2°. Any n-dimensional autonomous system of the form (12.7.2.5) can be reduced to an $(n-1)$-dimensional system of equations independent of x. To this end, one should select one of the equations and divide the other $n-1$ equations of the system by it.

Example 3. The autonomous system of two first-order equations

$$y'_x = f_1(y, z), \quad z'_x = f_2(y, z) \tag{12.7.2.6}$$

is reduced by dividing the first equation by the second to a single equation for $y = y(z)$:

$$y'_z = \frac{f_1(y, z)}{f_2(y, z)}. \tag{12.7.2.7}$$

If the general solution of equation (12.7.2.7) is obtained in the form

$$y = \varphi(z, C_1), \tag{12.7.2.8}$$

then $z = z(x)$ is found in implicit form from the second equation in (12.7.2.6) by quadrature:

$$\int \frac{dz}{f_2(\varphi(z, C_1), z)} = x + C_2. \tag{12.7.2.9}$$

Formulas (12.7.2.8)–(12.7.2.9) determine the general solution of system (12.7.2.6) with two arbitrary constants, C_1 and C_2.

Remark. The dependent variables y and z in the autonomous system (12.7.2.6) are often called *phase variables*; the plane y, z they form is called a *phase plane*, which serves to display integral curves of equation (12.7.2.7).

12.7.3. Elements of Stability Theory

12.7.3-1. Lyapunov stability. Asymptotic stability.

1°. In many applications, time t plays the role of the independent variable, and the associated system of differential equations is conventionally written in the following notation:

$$x'_k = f_k(t, x_1, \ldots, x_n), \qquad k = 1, \ldots, n. \tag{12.7.3.1}$$

Here the $x_k = x_k(t)$ are unknown functions that may be treated as coordinates of a moving point in an n-dimensional space.

Let us supply system (12.7.3.1) with initial conditions

$$x_k = x_k^\circ \quad \text{at} \quad t = t^\circ \qquad (k = 1, \ldots, n). \tag{12.7.3.2}$$

Denote by

$$x_k = \varphi_k(t; t^\circ, x_1^\circ, \ldots, x_n^\circ), \qquad k = 1, \ldots, n, \tag{12.7.3.3}$$

the solution of system (12.7.3.1) with the initial conditions (12.7.3.2).

A solution (12.7.3.3) of system (12.7.3.1) is called *Lyapunov stable* if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that if

$$|x_k^\circ - \widetilde{x}_k^\circ| < \delta, \qquad k = 1, \ldots, n, \tag{12.7.3.4}$$

then the following inequalities hold for $t^\circ \le t < \infty$:

$$\left|\varphi_k(t; t^\circ, x_1^\circ, \ldots, x_n^\circ) - \varphi_k(t; t^\circ, \widetilde{x}_1^\circ, \ldots, \widetilde{x}_n^\circ)\right| < \varepsilon, \qquad k = 1, \ldots, n.$$

Any solution which is not stable is called *unstable*. Solution (12.7.3.3) is called unperturbed and the solution $\varphi_k(t; t^\circ, \widetilde{x}_1^\circ, \ldots, \widetilde{x}_n^\circ)$ is called perturbed. Geometrically, Lyapunov stability means that the trajectory of the perturbed solution stays at all times $t \ge t^\circ$ within a small neighborhood of the associated unperturbed solution.

2°. A solution (12.7.3.3) of system (12.7.3.1) is called *asymptotically stable* if it is Lyapunov stable and, in addition, with inequalities (12.7.3.4) met, satisfies the conditions

$$\lim_{t\to\infty} \left|\varphi_k(t; t^\circ, x_1^\circ, \ldots, x_n^\circ) - \varphi_k(t; t^\circ, \widetilde{x}_1^\circ, \ldots, \widetilde{x}_n^\circ)\right| = 0, \qquad k = 1, \ldots, n. \tag{12.7.3.5}$$

Remark. Condition (12.7.3.5) alone does not suffice for the solution to be Lyapunov stable.

3°. In stability analysis, it is normally assumed, without loss of generality, that $t^\circ = x_1^\circ = \cdots = x_n^\circ = 0$ (this can be achieved by shifting each of the variables by a constant value). Further, with the changes of variables

$$z_k = x_k - \varphi_k(t; t^\circ, x_1^\circ, \ldots, x_n^\circ) \qquad (k = 1, \ldots, n),$$

the stability analysis of any solution is reduced to that of the zero solution $z_1 = \cdots = z_n = 0$.

12.7.3-2. Theorems of stability and instability by first approximation.

In studying stability of the trivial solution $x_1 = \cdots = x_n = 0$ of system (12.7.3.1) the following method is often employed. The right-hand sides of the equations are approximated by the principal (linear) terms of the expansion into Taylor series about the equilibrium point:

$$f_k(t, x_1, \ldots, x_n) \approx a_{k1}(t)x_1 + \cdots + a_{kn}(t)x_n, \quad a_{km}(t) = \left.\frac{\partial f_k}{\partial x_m}\right|_{x_1=\cdots=x_n=0}, \qquad k = 1, \ldots, n.$$

Then a stability analysis of the resulting simplified, linear system is performed. The question arises: Is it possible to draw correct conclusions about the stability of the original nonlinear system (12.7.3.1) from the analysis of the linearized system? Two theorems stated below give a partial answer to this question.

THEOREM (STABILITY BY FIRST APPROXIMATION). *Suppose in the system*

$$x_k' = a_{k1}x_1 + \cdots + a_{kn}x_n + \psi_k(t, x_1, \ldots, x_n), \qquad k = 1, \ldots, n, \tag{12.7.3.6}$$

the functions ψ_k are defined and continuous in a domain $t \ge 0$, $|x_k| \le b$ $(k = 1, \ldots, n)$ and, in addition, the inequality

$$\sum_{k=1}^{n} |\psi_k| \le A \sum_{k=1}^{n} |x_k| \tag{12.7.3.7}$$

holds for some constant A. In particular, this implies that $\psi_k(t, 0, \ldots, 0) = 0$, and therefore

$$x_1 = \cdots = x_n = 0 \tag{12.7.3.8}$$

is a solution of system (12.7.3.6). Suppose further that

$$\frac{\sum\limits_{k=1}^{n} |\psi_k|}{\sum\limits_{k=1}^{n} |x_k|} \to 0 \quad \text{as} \quad \sum_{k=1}^{n} |x_k| \to 0 \quad \text{and} \quad t \to \infty, \tag{12.7.3.9}$$

and the real parts of all roots of the characteristic equation

$$\det |a_{ij} - \lambda \delta_{ij}| = 0, \qquad \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \ne j \end{cases} \tag{12.7.3.10}$$

are negative. Then solution (12.7.3.8) is stable.

Remark 1. Necessary and sufficient conditions for the real parts of all roots of the characteristic equation (12.7.3.10) to be negative are established by Hurwitz's theorem, which allows avoiding its solution.

Remark 2. In the above system, the a_{ij}, x_k, and ψ_k may be complex valued.

THEOREM (INSTABILITY BY FIRST APPROXIMATION). *Suppose conditions (12.7.3.7) and (12.7.3.9) are met and the conditions for the functions ψ_k from the previous theorem are also met. If at least one root of the characteristic equation (12.7.3.10) has a positive real part, then the equilibrium point (12.7.3.8) of system (12.7.3.6) is unstable.*

Example 1. Consider the following two-dimensional system of the form (12.7.3.6) with real coefficients:

$$\begin{aligned} x_t' &= a_{11}x + a_{12}y + \psi_1(t, x, y), \\ y_t' &= a_{21}x + a_{22}y + \psi_2(t, x, y). \end{aligned} \tag{12.7.3.11}$$

We assume that the functions ψ_1 and ψ_2 satisfy conditions (12.7.3.7) and (12.7.3.9).

The characteristic equation of the linearized system (obtained by setting $\psi_1 = \psi_2 = 0$) is given by

$$\lambda^2 - b\lambda + c = 0, \quad \text{where} \quad b = a_{11} + a_{22}, \quad c = a_{11}a_{22} - a_{12}a_{21}. \tag{12.7.3.12}$$

1. Using the theorem of stability by first approximation and examining the roots of the quadratic equation (12.7.3.12), we obtain two sufficient stability conditions for system (12.7.3.11):

$b < 0, \quad 0 < \frac{1}{4}b^2 < c$ (complex roots with negative real part);

$b < 0, \quad 0 < c < \frac{1}{4}b^2$ (negative real roots).

The two conditions can be combined into one:

$$b < 0, \quad c > 0, \qquad \text{or} \qquad a_{11} + a_{22} < 0, \quad a_{11}a_{22} - a_{12}a_{21} > 0.$$

These inequalities define the second quadrant in the plane b, c; see Fig. 12.10.

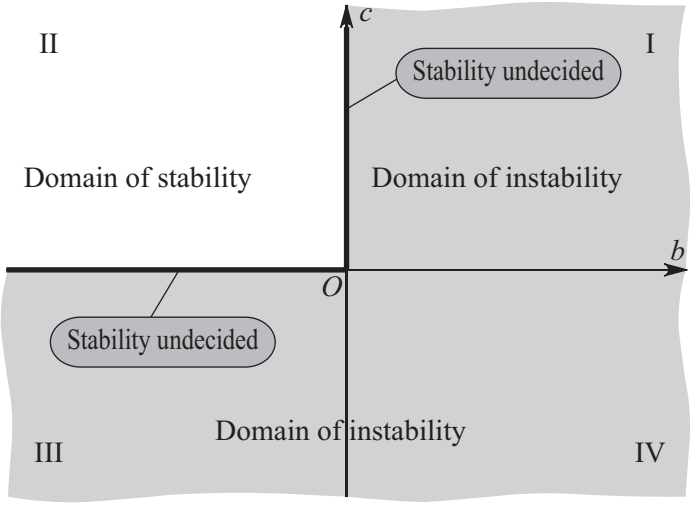


Figure 12.10. Domains of stability and instability of the trivial solution of system (12.7.3.11).

2. Using the theorem of instability by first approximation and examining the roots of the quadratic equation (12.7.3.12), we get three sufficient instability conditions for system (12.7.3.11):

$b > 0, \quad 0 < \frac{1}{4}b^2 < c$ (complex roots with positive real part);

$b > 0, \quad 0 < c < \frac{1}{4}b^2$ (positive real roots);

$c < 0, \quad b$ is any (real roots with different signs).

The first two conditions can be combined into one:

$$b > 0, \quad c > 0, \qquad \text{or} \qquad a_{11} + a_{22} > 0, \quad a_{11}a_{22} - a_{12}a_{21} > 0.$$

The domain of instability of system (12.7.3.11) covers the first, third, and fourth quadrants in the plane b, c (shaded in Fig. 12.10).

3. The conditions obtained above in Items 1 and 2 do not cover the whole domain of variation of the parameters a_{ij}. Stability or instability is not established for the boundary of the second quarter (shown by thick solid line in Fig. 12.10). This corresponds to the cases

$b = 0, \quad c \geq 0$ (two pure imaginary or two zero roots);

$c = 0, \quad b \leq 0$ (one zero root and one negative real or zero root).

Specific examples of such systems are considered below in Paragraph 12.7.3-3.

Remark. When the conditions of Item 1 or 2 hold, the phase trajectories of the nonlinear system (12.7.3.11) have the same qualitative arrangement in a neighborhood of the equilibrium point $x = y = 0$ as that of the phase trajectories of the linearized system (with $\psi_2 = \psi_1 = 0$). A detailed classification of equilibrium points of linear systems with associated arrangements of the phase trajectories can be found in Paragraph 12.6.1-7.

12.7.3-3. Lyapunov function. Theorems of stability and instability.

In the cases where the theorems of stability and instability by first approximation fail to resolve the issue of stability for a specific system of nonlinear differential equations, more subtle methods must be used. Such methods are considered below.

A *Lyapunov function* for system of equations (12.7.3.1) is a differentiable function $V = V(x_1, \ldots, x_n)$ such that

$$1) \quad V > 0 \quad \text{if} \quad \sum_{k=1}^{n} x_k^2 \neq 0, \qquad V = 0 \quad \text{if} \quad x_1 = \cdots = x_n = 0;$$

$$2) \quad \frac{dV}{dt} = \sum_{k=1}^{n} f_k(t, x_1, \ldots, x_n) \frac{\partial V}{\partial x_k} \leq 0 \quad \text{for} \quad t \geq 0.$$

Remark. The derivative with respect to t in the definition of a Lyapunov function is taken along an integral curve of system (12.7.3.1).

THEOREM (STABILITY, LYAPUNOV). *Let system (12.7.3.1) have the trivial solution $x_1 = x_2 = \cdots = x_n = 0$. This solution is stable if there exists a Lyapunov function for the system.*

THEOREM (ASYMPTOTIC STABILITY, LYAPUNOV). *Let system (12.7.3.1) have the trivial solution $x_1 = \cdots = x_n = 0$. This solution is asymptotically stable if there exists a Lyapunov function satisfying the additional condition*

$$\frac{dV}{dt} \leq -\beta < 0 \quad \textit{with} \quad \sum_{k=1}^{n} x_k^2 \geq \varepsilon_1 > 0, \quad t \geq \varepsilon_2 \geq 0,$$

where ε_1 and ε_2 are any positive numbers.

Example 2. Let us perform a stability analysis of the two-dimensional system

$$x_t' = -ay - x\varphi(x, y), \qquad y_t' = bx - y\psi(x, y),$$

where $a > 0$, $b > 0$, $\varphi(x, y) \geq 0$, and $\psi(x, y) \geq 0$ (φ and ψ are continuous functions).

A Lyapunov function will be sought in the form $V = Ax^2 + By^2$, where A and B are constants to be determined. The first condition characterizing a Lyapunov function will be satisfied automatically if $A > 0$ and $B > 0$ (it will be shown later that these inequalities do hold). To verify the second condition, let us compute the derivative:

$$\frac{dV}{dt} = f_1(x, y)\frac{\partial V}{\partial x} + f_2(x, y)\frac{\partial V}{\partial y} = -2Ax[ay + x\varphi(x, y)] + 2By[bx - y\psi(x, y)]$$
$$= 2(Bb - Aa)xy - 2Ax^2\varphi(x, y) - 2By^2\psi(x, y).$$

Setting here $A = b > 0$ and $B = a > 0$ (thus satisfying the first condition), we obtain the inequality

$$\frac{dV}{dt} = -2bx^2\varphi(x, y) - 2ay^2\psi(x, y) \leq 0.$$

This means that the second condition characterizing a Lyapunov function is also met. Hence, the trivial solution of the system in question is stable.

Example 3. Let us perform a stability analysis for the trivial solution of the nonlinear system

$$x_t' = -xy^2, \qquad y_t' = yx^4.$$

Let us show that the $V(x, y) = x^4 + y^2$ is a Lyapunov function for the system. Indeed, both conditions are satisfied:

$$1) \quad x^4 + y^2 > 0 \quad \text{if} \quad x^2 + y^2 \neq 0, \qquad V(0, 0) = 0 \quad \text{if} \quad x = y = 0;$$

$$2) \quad \frac{dV}{dt} = -4x^4y^2 + 2x^4y^2 = -2x^4y^2 \leq 0.$$

Hence the trivial solution of the system is stable.

Remark. No stability analysis of the systems considered in Examples 2 and 3 is possible based on the theorem of stability by first approximation.

THEOREM (INSTABILITY, CHETAEV). *Suppose there exists a differentiable function* $W = W(x_1, \ldots, x_n)$ *that possesses the following properties:*

1. In an arbitrarily small domain R *containing the origin of coordinates, there exists a subdomain* $R_+ \subset R$ *in which* $W > 0$, *with* $W = 0$ *on part of the boundary of* R_+ *in* R.

2. The condition

$$\frac{dW}{dt} = \sum_{k=1}^{n} f_k(t, x_1, \ldots, x_n)\frac{\partial W}{\partial x_k} > 0$$

holds in R_+ *and, moreover, in the domain of the variables where* $W \geq \alpha > 0$, *the inequality* $\dfrac{dW}{dt} \geq \beta > 0$ *holds.*

Then the trivial solution $x_1 = \cdots = x_n = 0$ *of system (12.7.3.1) is unstable.*

Example 4. Perform a stability analysis of the nonlinear system

$$x'_t = y^3\varphi(x, y, t) + x^5, \qquad y'_t = x^3\varphi(x, y, t) + y^5,$$

where $\varphi(x, y, t)$ is an arbitrary continuous function.

Let us show that the $W = x^4 - y^4$ satisfies the conditions of the Chetaev theorem. We have:

1. $W > 0$ for $|x| > |y|$, $\quad W = 0$ for $|x| = |y|$.
2. $\dfrac{dW}{dt} = 4x^3[y^3\varphi(x, y, t) + x^5] - 4y^3[x^3\varphi(x, y, t) + y^5] = 4(x^8 - y^8) > 0$ for $|x| > |y|$.

Moreover, if $W \geq \alpha > 0$, we have $\dfrac{dW}{dt} = 4\alpha(x^4 + y^4) \geq 4\alpha^2 = \beta > 0$. It follows that the equilibrium point $x = y = 0$ of the system in question is unstable.

References for Chapter 12

Akulenko, L. D. and Nesterov, S. V., *High Precision Methods in Eigenvalue Problems and Their Applications*, Chapman & Hall/CRC Press, Boca Raton, 2005.

Arnold, V. I., Kozlov, V. V., and Neishtadt, A. I., *Mathematical Aspects of Classical and Celestial Mechanics, Dynamical System III*, Springer-Verlag, Berlin, 1993.

Bakhvalov, N. S., *Numerical Methods: Analysis, Algebra, Ordinary Differential Equations*, Mir Publishers, Moscow, 1977.

Bogolyubov, N. N. and Mitropol'skii, Yu. A., *Asymptotic Methods in the Theory of Nonlinear Oscillations* [in Russian], Nauka Publishers, Moscow, 1974.

Boyce, W. E. and DiPrima, R. C., *Elementary Differential Equations and Boundary Value Problems, 8th Edition*, John Wiley & Sons, New York, 2004.

Braun, M., *Differential Equations and Their Applications, 4th Edition*, Springer-Verlag, New York, 1993.

Cole, G. D., *Perturbation Methods in Applied Mathematics*, Blaisdell Publishing Company, Waltham, MA, 1968.

Dormand, J. R., *Numerical Methods for Differential Equations: A Computational Approach*, CRC Press, Boca Raton, 1996.

El'sgol'ts, L. E., *Differential Equations*, Gordon & Breach, New York, 1961.

Fedoryuk, M. V., *Asymptotic Analysis. Linear Ordinary Differential Equations*, Springer-Verlag, Berlin, 1993.

Finlayson, B. A., *The Method of Weighted Residuals and Variational Principles*, Academic Press, New York, 1972.

Grimshaw, R., *Nonlinear Ordinary Differential Equations*, CRC Press, Boca Raton, 1991.

Gromak, V. I., *Painlevé Differential Equations in the Complex Plane*, Walter de Gruyter, Berlin, 2002.

Gromak, V. I. and Lukashevich, N. A., *Analytical Properties of Solutions of Painlevé Equations* [in Russian], Universitetskoe, Minsk, 1990.

Ince, E. L., *Ordinary Differential Equations*, Dover Publications, New York, 1964.

Kamke, E., *Differentialgleichungen: Lösungsmethoden und Lösungen, I, Gewöhnliche Differentialgleichungen*, B. G. Teubner, Leipzig, 1977.

Kantorovich, L. V. and Krylov, V. I., *Approximate Methods of Higher Analysis* [in Russian], Fizmatgiz, Moscow, 1962.

Keller, H. B., *Numerical Solutions of Two Point Boundary Value Problems*, Society for Industrial & Applied Mathematics, Philadelphia, 1976.

Kevorkian, J. and Cole, J. D., *Multiple Scale and Singular Perturbation Methods*, Springer-Verlag, New York, 1996.

Korn, G. A. and Korn, T. M., *Mathematical Handbook for Scientists and Engineers, 2nd Edition*, Dover Publications, New York, 2000.

Lambert, J. D., *Computational Methods in Ordinary Differential Equations*, Cambridge University Press, New York, 1973.

Lee, H. J. and Schiesser, W. E., *Ordinary and Partial Differential Equation Routines in C, C++, Fortran, Java, Maple, and MATLAB*, Chapman & Hall/CRC Press, Boca Raton, 2004.

Levitan, B. M. and Sargsjan, I. S., *Sturm–Liouville and Dirac Operators*, Kluwer Academic, Dordrecht, 1990.

Marchenko, V. A., *Sturm–Liouville Operators and Applications*, Birkhäuser Verlag, Basel, 1986.

Murphy, G. M., *Ordinary Differential Equations and Their Solutions*, D. Van Nostrand, New York, 1960.

Nayfeh, A. H., *Introduction to Perturbation Techniques*, John Wiley & Sons, New York, 1981.

Nayfeh, A. H., *Perturbation Methods*, Wiley-Interscience, New York, 1973.

Petrovskii, I. G., *Lectures on the Theory of Ordinary Differential Equations* [in Russian], Nauka Publishers, Moscow, 1970.

Polyanin, A. D. and Zaitsev, V. F., *Handbook of Exact Solutions for Ordinary Differential Equations, 2nd Edition*, Chapman & Hall/CRC Press, Boca Raton, 2003.

Schiesser, W. E., *Computational Mathematics in Engineering and Applied Science: ODEs, DAEs, and PDEs*, CRC Press, Boca Raton, 1993.

Tenenbaum, M. and Pollard, H., *Ordinary Differential Equations*, Dover Publications, New York, 1985.

Wasov, W., *Asymptotic Expansions for Ordinary Differential Equations,* John Wiley & Sons, New York, 1965.

Zhuravlev, V. Ph. and Klimov, D. M., *Applied Methods in Oscillation Theory* [in Russian], Nauka Publishers, Moscow, 1988.

Zwillinger, D., *Handbook of Differential Equations, 3rd Edition*, Academic Press, New York, 1997.

Chapter 13
First-Order Partial Differential Equations

13.1. Linear and Quasilinear Equations

13.1.1. Characteristic System. General Solution

13.1.1-1. Equations with two independent variables. General solution. Examples.

1°. A *first-order quasilinear partial differential equation with two independent variables* has the general form

$$f(x,y,w)\frac{\partial w}{\partial x} + g(x,y,w)\frac{\partial w}{\partial y} = h(x,y,w). \tag{13.1.1.1}$$

Such equations are encountered in various applications (continuum mechanics, gas dynamics, hydrodynamics, heat and mass transfer, wave theory, acoustics, multiphase flows, chemical engineering, etc.).

If two independent integrals,

$$u_1(x,y,w) = C_1, \qquad u_2(x,y,w) = C_2, \tag{13.1.1.2}$$

of the *characteristic system*

$$\frac{dx}{f(x,y,w)} = \frac{dy}{g(x,y,w)} = \frac{dw}{h(x,y,w)} \tag{13.1.1.3}$$

are known, then the *general solution* of equation (13.1.1.1) is given by

$$\Phi(u_1,u_2) = 0, \tag{13.1.1.4}$$

where $\Phi(u,v)$ is an arbitrary function of two variables. With equation (13.1.1.4) solved for u_1 or u_2, we often specify the general solution in the form

$$u_k = \Psi(u_{3-k}),$$

where $k = 1, 2$ and $\Psi(u)$ is an arbitrary function of one variable.

2°. For linear equations (13.1.1.1) with the functions f, g, and h independent of the unknown w, the first integrals (13.1.1.2) of the characteristic system (13.1.1.3) have a simple structure (one integral is independent of w and the other is linear in w):

$$U(x,y) = C_1, \qquad w - V(x,y) = C_2.$$

In this case the general solution can be written in explicit form

$$w = V(x,y) + \Psi(U(x,y)),$$

where $\Psi(U)$ is an arbitrary function of one variable.

TABLE 13.1
General solutions to some special types of linear and quasilinear first-order partial differential equations; $\Psi(u)$ is an arbitrary function. The subscripts x and y indicate the corresponding partial derivatives

No.	Equations	General solutions	Notation, remarks
1	$w_x + [f(x)y + g(x)]w_y = 0$	$w = \Psi\left(e^{-F}y - \int e^{-F}g(x)\,dx\right)$	$F = \int f(x)\,dx$
2	$w_x + [f(x)y + g(x)y^k]w_y = 0$	$w = \Psi\left(e^{-F}y^{1-k} - (1-k)\int e^{-F}g(x)\,dx\right)$	$F = (1-k)\int f(x)\,dx$
3	$w_x + [f(x)e^{\lambda y} + g(x)]w_y = 0$	$w = \Psi\left(e^{-\lambda y}E + \lambda\int f(x)E\,dx\right)$	$E = \exp\left(\lambda\int g(x)\,dx\right)$
4	$f(x)w_x + g(y)w_y = 0$	$w = \Psi\left(\int \frac{dx}{f(x)} - \int \frac{dy}{g(y)}\right)$	
5	$aw_x + bw_y = f(x)g(w)$	$\int \frac{dw}{g(w)} = \frac{1}{a}\int f(x)\,dx + \Psi(bx - ay)$	solution in implicit form
6	$f(x)w_x + g(y)w_y = h(w)$	$\int \frac{dw}{h(w)} = \int \frac{dx}{f(x)} + \Psi(u)$	$u = \int \frac{dx}{f(x)} - \int \frac{dy}{g(y)}$
7	$w_x + f(w)w_y = 0$	$y = xf(w) + \Psi(w)$	solution in implicit form
8	$w_x + [f(w) + ay]w_y = 0$	$x = \frac{1}{a}\ln\lvert ay + f(w)\rvert + \Psi(w),\ a \neq 0$	solution in implicit form
9	$w_x + [f(w) + g(x)]w_y = 0$	$y = xf(w) + \int g(x)\,dx + \Psi(w)$	solution in implicit form

Example 1. Consider the linear constant coefficient equation

$$\frac{\partial w}{\partial x} + a\frac{\partial w}{\partial y} = 0.$$

The characteristic system for this equation is

$$\frac{dx}{1} = \frac{dy}{a} = \frac{dw}{0}.$$

It has two independent integrals:

$$y - ax = C_1, \quad w = C_2.$$

Hence, the general solution of the original equation is given by $\Phi(y - ax,\ w) = 0$. On solving this equation for w, one obtains the general solution in explicit form

$$w = \Psi(y - ax).$$

It is the traveling wave solution.

Example 2. Consider the quasilinear equation

$$\frac{\partial w}{\partial x} + aw\frac{\partial w}{\partial y} = 1.$$

The characteristic system

$$\frac{dx}{1} = \frac{dy}{aw} = \frac{dw}{1}$$

has two independent integrals:

$$x - w = C_1, \quad 2y - aw^2 = C_2.$$

Hence, the general solution of the original equation is given by

$$\Phi(x - w,\ 2y - aw^2) = 0.$$

3°. Table 13.1 lists general solutions to some linear and quasilinear first-order partial differential equations in two independent variables.

► *In Sections T7.1–T7.2, many more first-order linear and quasilinear partial differential equations in two independent variables are considered than in Table 13.1.*

13.1.1-2. Construction of a quasilinear equations when given its general solution.

Given a set of functions

$$w = F\big(x, y, \Psi(G(x, y))\big), \tag{13.1.1.5}$$

where $F(x, y, \Psi)$ and $G(x, y)$ are prescribed and $\Psi(G)$ is arbitrary, there exists a quasilinear first-order partial differential equation such that the set of functions (13.1.1.5) is its general solution. To prove this statement, let us differentiate (13.1.1.5) with respect to x and y and then eliminate the partial derivative Ψ_G from the resulting expression to obtain

$$\frac{w_x - F_x}{G_x} = \frac{w_y - F_y}{G_y}. \tag{13.1.1.6}$$

On solving the relation $w = F(x, y, \Psi)$ [see (13.1.1.5)] for Ψ and substituting the resulting expression into (13.1.1.6), one arrives at the desired partial differentiable equation.

Example 3. Let us construct a partial differential equation whose general solution is given by

$$w = x^k \Psi(ax^n + by^m), \tag{13.1.1.7}$$

where $\Psi(z)$ is an arbitrary function.

Differentiating (13.1.1.7) with respect to x and y yields the relations $w_x = kx^{k-1}\Psi + anx^{k+n-1}\Psi_z$ and $w_y = bmx^k y^{m-1}\Psi_z$. Eliminating Ψ_z from them gives

$$\frac{w_x - kx^{k-1}\Psi}{anx^{n-1}} = \frac{w_y}{bmy^{m-1}}. \tag{13.1.1.8}$$

Solving the original relation (13.1.1.7) for Ψ, we get $\Psi = x^{-k}w$. Substituting this expression into (13.1.1.8) and rearranging, we arrive at the desired equation

$$bmxy^{m-1}\frac{\partial w}{\partial x} - anx^n \frac{\partial w}{\partial y} = bkmy^{m-1},$$

whose general solution is the function (13.1.1.7).

13.1.1-3. Equations with n independent variables. General solution.

A first-order quasilinear partial differential equation with n independent variables has the general form

$$f_1(x_1, \ldots, x_n, w)\frac{\partial w}{\partial x_1} + \cdots + f_n(x_1, \ldots, x_n, w)\frac{\partial w}{\partial x_n} = g(x_1, \ldots, x_n, w). \tag{13.1.1.9}$$

Let n independent integrals,

$$u_1(x_1, \ldots, x_n, w) = C_1, \quad \ldots, \quad u_n(x_1, \ldots, x_n, w) = C_n,$$

of the characteristic system

$$\frac{dx_1}{f_1(x_1, \ldots, x_n, w)} = \cdots = \frac{dx_n}{f_n(x_1, \ldots, x_n, w)} = \frac{dw}{g(x_1, \ldots, x_n, w)}$$

be known. Then the general solution of equation (13.1.1.9) is given by

$$\Phi(u_1, \ldots, u_n) = 0,$$

where Φ is an arbitrary function of n variables.

13.1.2. Cauchy Problem. Existence and Uniqueness Theorem

13.1.2-1. Cauchy problem.

Consider two formulations of the Cauchy problem.

1°. *Generalized Cauchy problem.* Find a solution $w = w(x, y)$ of equation (13.1.1.1) satisfying the initial conditions

$$x = h_1(\xi), \quad y = h_2(\xi), \quad w = h_3(\xi), \tag{13.1.2.1}$$

where ξ is a parameter ($\alpha \le \xi \le \beta$) and the $h_k(\xi)$ are given functions.

Geometric interpretation: find an integral surface of equation (13.1.1.1) passing through the line defined parametrically by equation (13.1.2.1).

2°. *Classical Cauchy problem.* Find a solution $w = w(x, y)$ of equation (13.1.1.1) satisfying the initial condition

$$w = \varphi(y) \quad \text{at} \quad x = 0, \tag{13.1.2.2}$$

where $\varphi(y)$ is a given function.

It is convenient to represent the classical Cauchy problem as a generalized Cauchy problem by rewriting condition (13.1.2.2) in the parametric form

$$x = 0, \quad y = \xi, \quad w = \varphi(\xi). \tag{13.1.2.3}$$

13.1.2-2. Procedure of solving the Cauchy problem.

The procedure of solving the Cauchy problem (13.1.1.1), (13.1.2.1) involves several steps. First, two independent integrals (13.1.1.2) of the characteristic system (13.1.1.3) are determined. Then, to find the constants of integration C_1 and C_2, the initial data (13.1.2.1) must be substituted into the integrals (13.1.1.2) to obtain

$$u_1\big(h_1(\xi), h_2(\xi), h_3(\xi)\big) = C_1, \qquad u_2\big(h_1(\xi), h_2(\xi), h_3(\xi)\big) = C_2. \tag{13.1.2.4}$$

Eliminating C_1 and C_2 from (13.1.1.2) and (13.1.2.4) yields

$$\begin{aligned} u_1(x, y, w) &= u_1\big(h_1(\xi), h_2(\xi), h_3(\xi)\big), \\ u_2(x, y, w) &= u_2\big(h_1(\xi), h_2(\xi), h_3(\xi)\big). \end{aligned} \tag{13.1.2.5}$$

Formulas (13.1.2.5) are a parametric form of the solution of the Cauchy problem (13.1.1.1), (13.1.2.1). In some cases, one may succeed in eliminating the parameter ξ from relations (13.1.2.5), thus obtaining the solution in an explicit form.

Example 1. Consider the Cauchy problem for linear equation

$$\frac{\partial w}{\partial x} + a\frac{\partial w}{\partial y} = bw \tag{13.1.2.6}$$

subjected to the initial condition (13.1.2.2).

The corresponding characteristic system for equation (13.1.2.6),

$$\frac{dx}{1} = \frac{dy}{a} = \frac{dw}{bw},$$

has two independent integrals

$$y - ax = C_1, \qquad we^{-bx} = C_2. \tag{13.1.2.7}$$

Represent the initial condition (13.1.2.2) in parametric form (13.1.2.3) and then substitute the data (13.1.2.3) into the integrals (13.1.2.7). As a result, for the constants of integration we obtain $C_1 = \xi$ and $C_2 = \varphi(\xi)$. Substituting these expressions into (13.1.2.7), we arrive at the solution of the Cauchy problem (13.1.2.6), (13.1.2.2) in parametric form:

$$y - ax = \xi, \qquad we^{-bx} = \varphi(\xi).$$

By eliminating the parameter ξ from these relations, we obtain the solution of the Cauchy problem (13.1.2.6), (13.1.2.2) in explicit form:

$$w = e^{bx}\varphi(y - ax).$$

13.1.2-3. Existence and uniqueness theorem.

Let $\mathfrak{G}_0$ be a domain in the xy-plane and let $\mathfrak{G}$ be a cylindrical domain of the xyw-space obtained from $\mathfrak{G}_0$ by adding the coordinate w, with the condition $|w| < A_1$ being satisfied. Let the coefficients f, g, and h of equation (13.1.1.1) be continuously differentiable functions of x, y, and w in $\mathfrak{G}$ and let $x = h_1(\xi)$, $y = h_2(\xi)$, and $w = h_3(\xi)$ be continuously differentiable functions of ξ for $|\xi| < A_2$ defining a curve C in $\mathfrak{G}$ with a simple projection C_0 onto $\mathfrak{G}_0$. Suppose that $(h_1')^2 + (h_2')^2 \neq 0$ (the prime stands for the derivative with respect to ξ) and $fh_2' - gh_1' \neq 0$ on C. Then there exists a subdomain $\overline{\mathfrak{G}_0} \subset \mathfrak{G}_0$ containing C_0 where there exists a continuously differentiable function $w = w(x, y)$ satisfying the differential equation (13.1.1.1) in $\mathfrak{G}_0$ and the initial condition (13.1.2.1) on C_0. This function is unique.

It is important to note that this theorem has a local character, i.e., the existence of a solution is guaranteed in some "sufficiently narrow," unknown neighborhood of the line C (see the remark at the end of Example 2).

Example 2. Consider the Cauchy problem for Hopf's equation

$$\frac{\partial w}{\partial x} + w\frac{\partial w}{\partial y} = 0 \tag{13.1.2.8}$$

subject to the initial condition (13.1.2.2).

First, we rewrite the initial condition (13.1.2.2) in the parametric form (13.1.2.3). Solving the characteristic system

$$\frac{dx}{1} = \frac{dy}{w} = \frac{dw}{0}, \tag{13.1.2.9}$$

we find two independent integrals,

$$w = C_1, \quad y - wx = C_2. \tag{13.1.2.10}$$

Using the initial conditions (13.1.2.3), we find that $C_1 = \varphi(\xi)$ and $C_2 = \xi$. Substituting these expressions into (13.1.2.10) yields the solution of the Cauchy problem (13.1.2.8), (13.1.2.2) in the parametric form

$$w = \varphi(\xi), \tag{13.1.2.11}$$

$$y = \xi + \varphi(\xi)x. \tag{13.1.2.12}$$

The characteristics (13.1.2.12) are straight lines in the xy-plane with slope $\varphi(\xi)$ that intersect the y-axis at the points ξ. On each characteristic, the function w has the same value equal to $\varphi(\xi)$ (generally, w takes different values on different characteristics).

For $\varphi'(\xi) > 0$, different characteristics do not intersect and, hence, formulas (13.1.2.11) and (13.1.2.12) define a unique solution. As an example, we consider the initial profile

$$\varphi(\xi) = \begin{cases} w_1 & \text{for } \xi \le 0, \\ \dfrac{w_2\xi^2 + \varepsilon w_1}{\xi^2 + \varepsilon} & \text{for } \xi > 0, \end{cases} \tag{13.1.2.13}$$

where $w_1 < w_2$ and $\varepsilon > 0$. Formulas (13.1.2.11)–(13.1.2.13) give a unique smooth solution in the entire half-plane $x > 0$. In the domain filled by the characteristics $y = \xi + w_1 x$ (for $\xi \le 0$), the solution is constant, i.e.,

$$w = w_1 \quad \text{for} \quad y/x \le w_1. \tag{13.1.2.14}$$

For $\xi > 0$, the solution is determined by relations (13.1.2.11)–(13.1.2.13).

Let us look how this solution is transformed in the limit case $\varepsilon \to 0$, which corresponds to the piecewise-continuous initial profile

$$\varphi(\xi) = \begin{cases} w_1 & \text{for } \xi \le 0, \\ w_2 & \text{for } \xi > 0, \end{cases} \qquad \text{where} \quad w_1 < w_2. \tag{13.1.2.15}$$

We further assume that $\xi > 0$ [for $\xi \le 0$, formula (13.1.2.14) is valid]. If $\xi = \text{const} \neq 0$ and $\varepsilon \to 0$, it follows from (13.1.2.13) that $\varphi(\xi) = w_2$. Hence, in the domain filled by the characteristics $y = \xi + w_2 x$ (for $\xi > 0$), the solution is constant, i.e., we have

$$w = w_2 \quad \text{for} \quad y/x \ge w_2 \qquad (\text{as } \varepsilon \to 0). \tag{13.1.2.16}$$

For $\xi \to 0$, the function φ can assume any value between w_1 and w_2 depending on the ratio of the small parameters ε and ξ; the first term on the right-hand side of equation (13.1.2.12) can be neglected. As a result,

we find from equations (13.1.2.11) and (13.1.2.12) that the solution has the following asymptotic behavior in explicit form:

$$w = y/x \quad \text{for} \quad w_1 \le y/x \le w_2 \qquad (\text{as } \varepsilon \to 0). \tag{13.1.2.17}$$

By combining relations (13.1.2.14), (13.1.2.16), and (13.1.2.17) together, we obtain the solution of the Cauchy problem for equation (13.1.2.8) subject to the initial conditions (13.1.2.15) in the form

$$w(x,y) = \begin{cases} w_1 & \text{for } y \le w_1 x, \\ y/x & \text{for } w_1 x \le y \le w_2 x, \\ w_2 & \text{for } y \ge w_2 x. \end{cases} \tag{13.1.2.18}$$

Figure 13.1 shows characteristics of equation (13.1.2.8) that satisfy condition (13.1.2.15) with $w_1 = \frac{1}{2}$ and $w_2 = 2$. This figure also depicts the dependence of w on y (for $x = x_0 = 1$). In applications, such a solution is referred to as a centered rarefaction wave (see also Subsection 13.1.3).

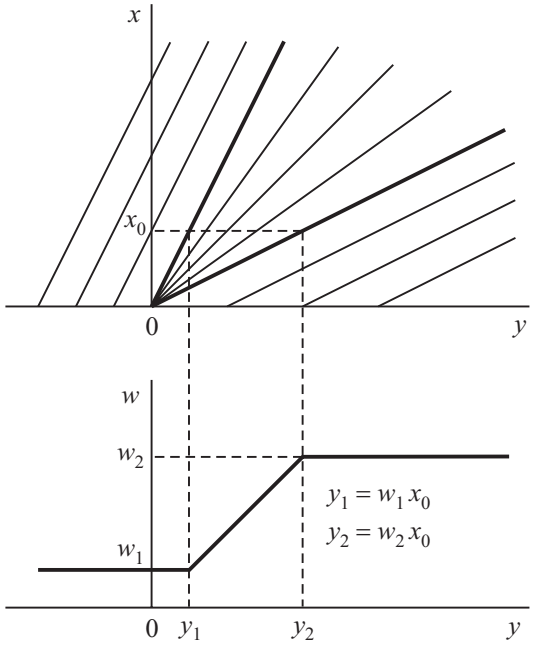


Figure 13.1. Characteristics of the Cauchy problem (13.1.2.8), (13.1.2.2) with the initial profile (13.1.2.15) and the dependence of the unknown w on the coordinate y for $w_1 = \frac{1}{2}$, $w_2 = 2$, and $x_0 = 1$.

Remark. If there is an interval where $\varphi'(\xi) < 0$, then the characteristics intersect in some domain. According to equation (13.1.2.11), at the point of intersection of two characteristics defined by two distinct values ξ_1 and ξ_2 of the parameter, the function w takes two distinct values equal to $\varphi(\xi_1)$ and $\varphi(\xi_2)$, respectively. Therefore, the solution is not unique in the domain of intersecting characteristics. This example illustrates the local character of the existence and uniqueness theorem. These issues are discussed in Subsections 13.1.3 and 13.1.4 in more detail.

13.1.3. Qualitative Features and Discontinuous Solutions of Quasilinear Equations

13.1.3-1. Model equation of gas dynamics.

Consider a quasilinear equation of the special form*

$$\frac{\partial w}{\partial x} + f(w)\frac{\partial w}{\partial y} = 0, \tag{13.1.3.1}$$

* Equations of the general form are discussed in Subsection 13.1.4.

which represents a conservation law of mass (or another quantity) and is often encountered in continuum mechanics, gas dynamics, hydrodynamics, wave theory, acoustics, multiphase flows, and chemical engineering. This equation is a model for numerous processes of mass transfer: sorption and chromatography, two-phase flows in porous media, flow of water in river, street traffic development, flow of liquid films along inclined surfaces, etc. The independent variables x and y in equation (13.1.3.1) usually play the role of time and spatial coordinate, respectively, w is the density of the quantity being transferred, and $f(w)$ is the rate of w.

13.1.3-2. Solution of the Cauchy problem. Rarefaction wave. Wave "overturn."

1°. The solution $w = w(x, y)$ of the Cauchy problem for equation (13.1.3.1) subject to the initial condition

$$w = \varphi(y) \quad \text{at} \quad x = 0 \qquad (-\infty < y < \infty) \tag{13.1.3.2}$$

can be represented in the parametric form

$$y = \xi + \mathcal{F}(\xi)x, \quad w = \varphi(\xi), \tag{13.1.3.3}$$

where $\mathcal{F}(\xi) = f\big(\varphi(\xi)\big)$.

Consider the characteristics $y = \xi + \mathcal{F}(\xi)x$ in the yx-plane for various values of the parameter ξ. These are straight lines with slope $\mathcal{F}(\xi)$. Along each of these lines, the unknown function is constant, $w = \varphi(\xi)$. In the special case $f = a = \text{const}$, the equation in question is linear; solution (13.1.3.3) can be written explicitly as $w = \varphi(y - ax)$, thus representing a traveling wave with a fixed profile. The dependence of f on w leads to a typical nonlinear effect: distortion of the profile of the traveling wave.

We further consider the domain $x \geq 0$ and assume* that $f > 0$ for $w > 0$ and $f'_w > 0$. In this case, the greater values of w propagate faster than the smaller values. If the initial profile satisfies the condition $\varphi'(y) > 0$ for all y, then the characteristics in the yx-plane that come from the y-axis inside the domain $x > 0$ are divergent lines, and hence there exists a unique solution for all $x > 0$. In physics, such solutions are referred to as rarefaction waves.

Example 1. Figures 13.2 and 13.3 illustrate characteristics and the evolution of a rarefaction wave for Hopf's equation [for $f(w) = w$ in (13.1.3.1)] with the initial profile

$$\varphi(y) = \frac{4}{\pi}\arctan(y - 2) + 2. \tag{13.1.3.4}$$

It is apparent that the solution is smooth for all $x > 0$.

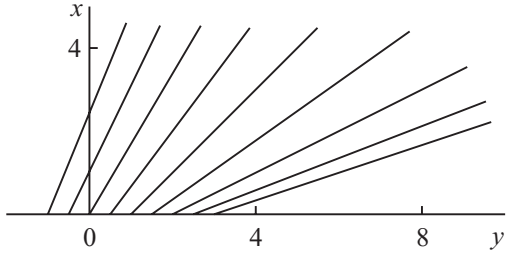


Figure 13.2. Characteristics for the Hopf's equation (13.1.2.8) with the initial profile (13.1.3.4).

* By the change $x = -\widetilde{x}$ the consideration of the domain $x \leq 0$ can be reduced to that of the domain $\widetilde{x} \geq 0$. The case $f < 0$ can be reduced to the case $f > 0$ by the change $y = -\widetilde{y}$.

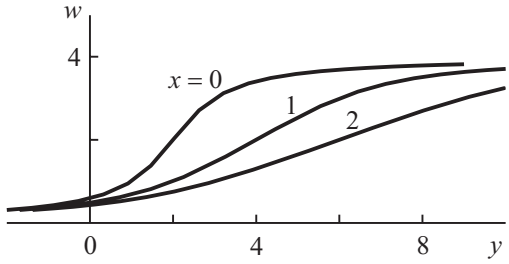


Figure 13.3. The evolution of a rarefaction wave for the Hopf's equation (13.1.2.8) with the initial profile (13.1.3.4).

2°. Let us now look at what happens if $\varphi'(y) < 0$ on some interval of the y-axis. Let y_1 and y_2 be points of this interval such that $y_1 < y_2$. Then $f(y_1) > f(y_2)$. It follows from the first relation in (13.1.3.3) that the characteristics issuing from the points y_1 and y_2 intersect at the "time instant"

$$x_* = \frac{y_2 - y_1}{f(w_1) - f(w_2)}, \qquad \text{where} \quad w_1 = \varphi(y_1), \quad w_2 = \varphi(y_2).$$

Since w has different values on these characteristics, the solution cannot be continuously extended to $x > x_*$. If $\varphi'(y) < 0$ on a bounded interval, then there exists $x_{\min} = \min\limits_{y_1, y_2} x_*$ such that the characteristics intersect in the domain $x > x_{\min}$ (see Fig. 13.4). Therefore, the front part of the wave where its profile is a decreasing function of y will "overturn" with time. The time $x_{\min}$ when the overturning begins is defined by

$$x_{\min} = -\frac{1}{\mathcal{F}'(\xi_0)},$$

where ξ_0 is determined by the condition $|\mathcal{F}'(\xi_0)| = \max|\mathcal{F}'(\xi)|$ for $\mathcal{F}'(\xi) < 0$, and the wave is also said to *break*. A formal extension of the solution to the domain $x > x_{\min}$ makes this solution nonunique. The boundary of the uniqueness domain in the yx-plane is the envelope of the characteristics. This boundary can be represented in parametric form as

$$y = \xi + \mathcal{F}(\xi)x, \quad 0 = 1 + \mathcal{F}'(\xi)x.$$

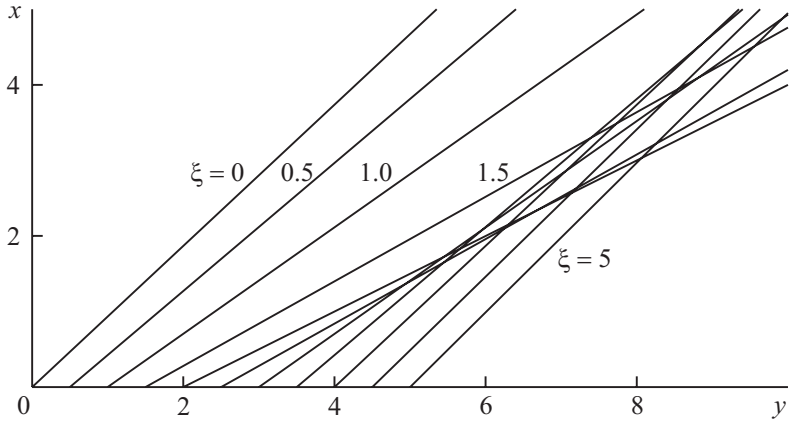


Figure 13.4. Characteristics for the Hopf's equation (13.1.2.8) with the initial profile (13.1.3.5).

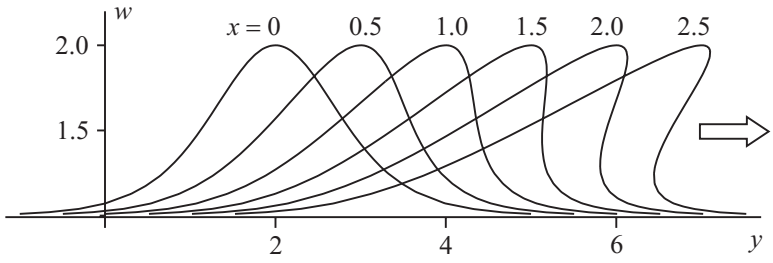


Figure 13.5. The evolution of a solitary wave for the Hopf's equation (13.1.2.8) with the initial profile (13.1.3.5).

Example 2. Figure 13.5 illustrates the evolution of a solitary wave with the initial profile

$$\varphi(y) = \cosh^{-2}(y-2) + 1 \tag{13.1.3.5}$$

for equation (13.1.3.1) with $f(w) = w$. It is apparent that for $x > x_{\min}$, where $x_{\min} = \frac{3}{4}\sqrt{3} \approx 1.3$, the wave "overturns" (the wave profile becomes triple-valued).

13.1.3-3. Shock waves. Jump conditions.

In most applications where the equation under consideration is encountered, the unknown function $w(x, y)$ is the density of a medium and must be unique for its nature. In these cases, one has to deal with a generalized (nonsmooth) solution describing a step-shaped shock wave rather than a continuous smooth solution. The many-valued part of the wave profile is replaced by an appropriate discontinuity, as shown in Fig. 13.6. It should be emphasized that a discontinuity can occur for arbitrarily smooth functions $f(w)$ and $\varphi(y)$ entering equation (13.1.3.1) and the initial condition (13.1.3.2).

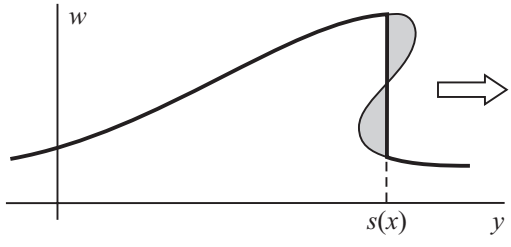


Figure 13.6. Replacement of the many-valued part of the wave profile by a discontinuity that cuts off domains with equal areas (shaded) from the profile of a breaking wave.

In what follows, we assume that $w(x, y)$ experiences a jump discontinuity at the line $y = s(x)$ in the yx-plane. On both sides of the discontinuity the function $w(x, y)$ is smooth and single-valued; as before, it is described by equations (13.1.3.3). The speed of propagation of the discontinuity, V, is expressed as $V = s'$ (the prime stands for the derivative) and must satisfy the condition

$$V = \frac{F(w_2) - F(w_1)}{w_2 - w_1}, \qquad F(w) = \int f(w)\,dw, \tag{13.1.3.6}$$

where the subscript 1 refers to values before the discontinuity and the subscript 2 to those after the discontinuity. In applications, relation (13.1.3.6), expressing a conservation law at discontinuity, is conventionally referred to as the Rankine–Hugoniot jump condition (this condition is derived below in Paragraph 13.1.3-4).

The continuous wave "overturns" (breaks), thus resulting in a discontinuity if and only if the propagation velocity $f(w)$ decreases as y increases, i.e., the inequalities

$$f(w_2) < V < f(w_1) \tag{13.1.3.7}$$

are satisfied. Conditions (13.1.3.7) have the geometric meaning that the characteristics issuing from the x-axis (these characteristics "carry" information about the initial data) must intersect the line of discontinuity (see Fig. 13.7). In this case, the discontinuous solution is stable with respect to small perturbations of the initial profile (i.e., the corresponding solution varies only slightly).

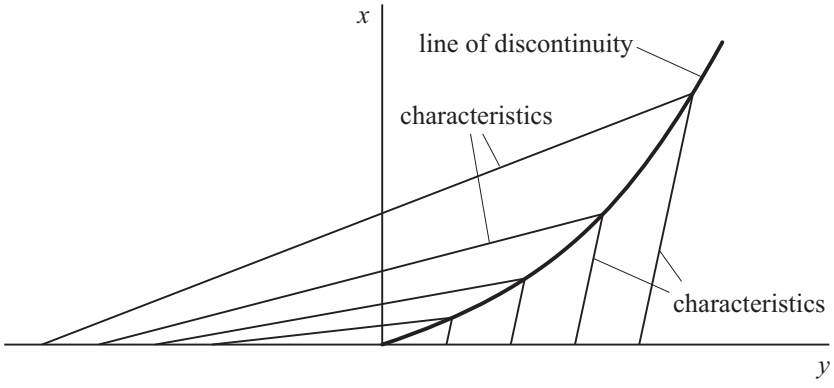


Figure 13.7. Mutual arrangement of characteristics and lines of discontinuity in the case of a stable shock wave.

The position of the point of discontinuity in the yw-plane may be determined geometrically by following Whitham's rule: the discontinuity must cut off domains with equal areas from the overturning wave profile (these domains are shaded in Fig. 13.6). Mathematically, the position of the point of discontinuity can be determined from the equations

$$\begin{aligned} s(x) &= \xi_1 + \mathcal{F}_1 x, \\ s(x) &= \xi_2 + \mathcal{F}_2 x, \\ w_2\mathcal{F}_2 - w_1\mathcal{F}_1 &= F(w_2) - F(w_1) + \frac{\mathcal{F}_2 - \mathcal{F}_1}{\xi_2 - \xi_1} \int_{\xi_1}^{\xi_2} w \, d\xi. \end{aligned} \tag{13.1.3.8}$$

Here, w and $\mathcal{F}$ are defined as functions of ξ by $w = \varphi(\xi)$ and $\mathcal{F} = f(w)$, the function $F(w)$ is introduced in equation (13.1.3.6), and the subscripts 1 and 2 refer to the values of the corresponding quantities at $\xi = \xi_1$ and ξ_2. Equations (13.1.3.8) permit one to determine the dependences $s = s(x)$, $\xi_1 = \xi_1(x)$, and $\xi_2 = \xi_2(x)$. It is possible to show that the jump condition (13.1.3.6) follows from the last equation in (13.1.3.8).

Example 3. For Hopf's equation, which corresponds to $f(w) = w$ in equation (13.1.3.1), the jump condition (13.1.3.6) can be represented as

$$V = \frac{w_1 + w_2}{2}.$$

Here, we take into account the relation $F(w) = \frac{1}{2}w^2$. System (13.1.3.8), which determines the position of the point of discontinuity, becomes

$$\begin{aligned} s(x) &= \xi_1 + \varphi(\xi_1)x, \\ s(x) &= \xi_2 + \varphi(\xi_2)x, \\ \frac{\varphi(\xi_1) + \varphi(\xi_2)}{2} &= \frac{1}{\xi_2 - \xi_1} \int_{\xi_1}^{\xi_2} \varphi(\xi) \, d\xi, \end{aligned}$$

where the function $\varphi(\xi)$ specifies the initial wave profile.

Figure 13.8 illustrates the formation of the shock wave described by the generalized solution of Hopf's equation with $f(w) = w$ and generated from a solitary wave with the smooth initial profile (13.1.3.5).

The nonsmooth "step" curves depicted in Fig. 13.8 (for $x = 0.15$, 0.20, and 0.25) are obtained from the smooth (but many-valued) curves shown in Fig. 13.5 by means of Whitham's rule of equal areas.

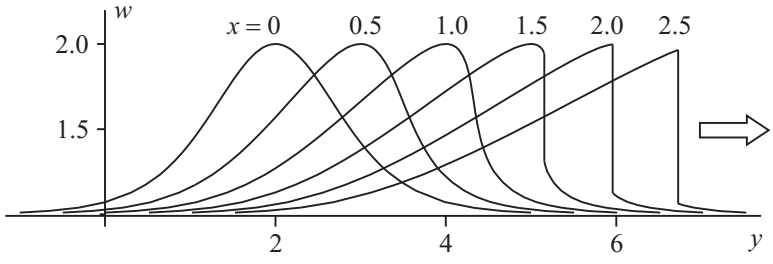


Figure 13.8. The formation of a shock wave generated from a solitary wave with the smooth initial profile.

13.1.3-4. Utilization of integral relations for determining generalized solutions.

Generalized solutions which are described by piecewise-smooth (piecewise-continuous) functions may formally be introduced by considering the following equation written in an integral form:

$$-\iint_D \left[w\frac{\partial \psi}{\partial x} + F(w)\frac{\partial \psi}{\partial y} \right] dy\,dx = 0. \tag{13.1.3.9}$$

Here, D is an arbitrary rectangle in the yx-plane, $\psi = \psi(x, y)$ is any "test" function with continuous first derivatives in D that is zero at the boundary of D, and the function $F(w)$ is defined in equation (13.1.3.6). If w and $F(w)$ are continuously differentiable, then equation (13.1.3.9) is equivalent to the original differential equation (13.1.3.1). Indeed, multiplying equation (13.1.3.1) by ψ, integrating over the domain D, and then integrating by parts, we obtain equation (13.1.3.9). Conversely, integrating (13.1.3.9) by parts yields

$$\iint_D \left[\frac{\partial w}{\partial x} + \frac{\partial F(w)}{\partial y} \right] \psi\, dy\,dx = 0.$$

Since this equation must be satisfied for any test function ψ and since $F'(w) = f(w)$, we obtain the original equation (13.1.3.1). However, equation (13.1.3.9) has a wider class of solutions since the admissible functions $w(x, y)$ need not necessarily be differentiable. The functions $w(x, y)$ satisfying the integral relation (13.1.3.9) for all test functions ψ are referred to as generalized (or weak) solutions of equation (13.1.3.1).

The use of generalized solutions is convenient for the description of discontinuities, since it permits one to obtain jump conditions automatically. Consider a solution of equation (13.1.3.9) continuously differentiable in two portions D_1 and D_2 of the rectangle D, which has a jump discontinuity at the interface Γ between D_1 and D_2. Integrating equation (13.1.3.9) by parts in each of the subdomains D_1 and D_2 yields

$$\iint_{D_1} \left[\frac{\partial w}{\partial x} + \frac{\partial F(w)}{\partial y} \right] \psi\, dy\,dx + \iint_{D_2} \left[\frac{\partial w}{\partial x} + \frac{\partial F(w)}{\partial y} \right] \psi\, dy\,dx + \int_\Gamma \big\{ [w]\, dy - [F(w)]\, dx \big\} \psi = 0,$$

where $[w] = w_2 - w_1$ and $[F(w)] = F(w_2) - F(w_1)$ are jumps of w and $F(w)$ across Γ. The curvilinear integral over Γ is formed by the boundary terms of the integrals over D_1 and D_2 that result from the integration by parts. Since the relation obtained must be valid for all

test functions ψ, it follows that equation (13.1.3.1) is valid inside each of the subdomains D_1 and D_2 and, moreover, the relation

$$[w]\,dy - [F(w)]\,dx = 0 \qquad (\text{on } \Gamma)$$

must hold. Assuming as before that the line of discontinuity is defined by the equation $y = s(x)$, we arrive at the jump condition (13.1.3.6).

It is worth noting that condition (13.1.3.7) does not follow from the integral relation (13.1.3.9) but is deduced from the additional condition of stability of the solution.

13.1.3-5. Conservation laws. Viscosity solutions.

Point out also other ways of introducing generalized solutions.

1°. Generalized solution may be introduced using the conservation law

$$\frac{d}{dx}\int_{y_1}^{y_2} w\,dy + F(w_2) - F(w_1) = 0, \tag{13.1.3.10}$$

where $w = w(x, y)$ and $w_n = w(x, y_n)$ $(n = 1, 2)$. Just as in equation (13.1.3.6), the function $F(w)$ is defined as $F(w) = \int f(w)\,dw$. Relation (13.1.3.10) is assumed to hold for any y_1 and y_2. It has a simple physical interpretation: the rate of change of the total value of w distributed over the interval (y_1, y_2) is compensated for by the "flux" of the function $F(w)$ through the endpoints of the interval.

Let w be a continuously differentiable solution of the conservation law. Then, differentiating equation (13.1.3.10) with respect to y_2 and setting $y_2 = y$, we arrive at equation (13.1.3.1). The conservation law (13.1.3.10) is convenient for the reason that is admits discontinuous solutions. It is not difficult to show that in this case the jump condition (13.1.3.6) must hold. For this reason, conservation laws like (13.1.3.10) are sometimes used as the basis for determining generalized solutions.

2°. An alternative approach to determining generalized solutions involves the consideration of an auxiliary equation of the parabolic type of the form

$$\frac{\partial w}{\partial x} + f(w)\frac{\partial w}{\partial y} = \varepsilon\frac{\partial^2 w}{\partial x^2}, \qquad \varepsilon > 0. \tag{13.1.3.11}$$

The generalized solution of the Cauchy problem (13.1.3.1), (13.1.3.2) (for a finite initial profile) is defined as the limit of the solution of equation (13.1.3.11) with the same initial condition (13.1.3.2) as $\varepsilon \to 0$. It is shown by Oleinik (1957) and Gelfand (1959) that the above definitions of the generalized solution leads to the same results.

The parameter ε plays the role of "viscosity" (by analogy with the viscosity of a fluid), which "smooths out" the jump, thus making the profile of the unknown w continuous. Therefore, the above construction, based on proceeding to the limit as $\varepsilon \to 0$, is called the method of vanishing viscosity and the limit function obtained is called the viscosity solution. Equation (13.1.3.11) with small ε is not infrequently used as a basis for numerical simulation of discontinuous solutions of equation (13.1.3.1); in this case, one need not specially separate in the numerical scheme a domain of discontinuity.

Remark. In specific problems, first-order quasilinear equations are often a consequence of integral conservation laws, having clear physical interpretation. In such cases, one should introduce generalized solutions on the basis of these conservation laws; for example, see Whitham (1974) and Rozhdestvenskii and Yanenko (1983). The thus obtained nonsmooth generalized solutions may differ from those described above.

13.1.3-6. Hopf's formula for the generalized solution.

Below we give a general formula for a generalized solution of the Cauchy problem (13.1.3.1), (13.1.3.2), describing discontinuous solutions that satisfy the stability condition (13.1.3.7). As above, we assume that $x \ge 0$ and $f > 0$ for $w > 0$; $f'_w > 0$.

Consider the function

$$Z(s) = \min_w \{ws - F(w)\}, \quad \text{where} \quad F(w) = \int f(w)\,dw. \tag{13.1.3.12}$$

We set

$$H(x, y, \eta) = \int_0^\eta \varphi(\bar\eta)\,d\bar\eta + xZ\left(\frac{y-\eta}{x}\right). \tag{13.1.3.13}$$

This is a continuous function of η for fixed x and y. It can be shown that for fixed x and with the exception of a countable set of values of y, function (13.1.3.13) has a unique minimum with respect to η. Denote the position of this minimum by $\eta = \xi$, where $\xi = \xi(x, y)$. The stable generalized solution of equation (13.1.3.1) subject to the initial condition (13.1.3.2) is given by

$$w(x, y) = \mathcal{Z}\left(\frac{y-\xi}{x}\right), \quad \text{where} \quad \mathcal{Z}(s) = \frac{dZ}{ds}. \tag{13.1.3.14}$$

The function $Z = Z(s)$ defined by relation (13.1.3.12) can be represented in the parametric form

$$s = f(w), \qquad Z = ws - \int f(w)\,dw. \tag{13.1.3.15}$$

Hence follows the parametric representation for its derivative $\mathcal{Z} = \mathcal{Z}(s)$:

$$s = f(w), \qquad \mathcal{Z} = w. \tag{13.1.3.16}$$

The position of the minimum $\eta = \xi(x, y)$ of function (13.1.3.13) is determined by the condition $H_\eta = 0$, which results in the following equation for ξ:

$$\varphi(\xi) - \mathcal{Z}\left(\frac{y-\xi}{x}\right) = 0. \tag{13.1.3.17}$$

To illustrate the utilization of the above formulas, we consider two cases.

1°. Let the algebraic (or transcendental) equation (13.1.3.17) have a unique solution $\xi = \xi(x, y)$ in some domain of the xy-plane. We set $s = (y-\xi)/x$ in (13.1.3.16) and consider these relations in conjunction with equation (13.1.3.17). Eliminating the functions $f(w)$ and $\mathcal{Z}$ from these equations yields a solution of the problem in the parametric form (13.1.3.3). In this case, we obtain a smooth (classical) solution describing a rarefaction wave.

2°. Let the algebraic (or transcendental) equation (13.1.3.17) have two different solutions, ξ_1 and ξ_2, that are functions of x and y. For both cases, solution (13.1.3.3) is valid, where either $\xi = \xi_1$ or $\xi = \xi_2$. At each point (x, y), we choose that solution ξ_n $(n = 1, 2)$ which minimizes the function $H(x, y, \xi_n)$ defined by equation (13.1.3.13). In this case, we obtain a discontinuous (generalized) solution describing a shock wave.

13.1.3-7. Problem of propagation of a signal.

In the problem of propagation of a signal and other physical applications, one seeks a solution of equation (13.1.3.1) subject to the conditions

$$\begin{aligned} &w = w_0 \quad \text{at} \quad x = 0 \quad \text{(initial condition)},\\ &w = g(x) \quad \text{at} \quad y = 0 \quad \text{(boundary condition)}, \end{aligned} \tag{13.1.3.18}$$

where w_0 is some constant and $g(x)$ is a prescribed function. One considers the domain $x > 0$, $y > 0$, where x plays the role of time and y the role of the spatial coordinate. It is assumed that $f(w) > 0$.

The characteristics of this problem issue from the positive semiaxis y and the positive semiaxis x (see Fig. 13.9). We have $w = w_0$ at the characteristics issuing from the y-axis. Hence, these characteristics are straight lines defined by $y - a_0 x = \text{const}$, where $a_0 = f(w_0)$. It follows that

$$w = w_0 \quad \text{for} \quad y > a_0 x. \tag{13.1.3.19}$$

As far as the characteristics issuing from the x-axis are concerned, we assume that one of the characteristics starts from a point $x = \tau$. The solution of equation (13.1.3.1) subject to conditions (13.1.3.18) can be represented in the parametric form

$$y = \mathcal{G}(\tau)(x - \tau), \quad w = g(\tau), \tag{13.1.3.20}$$

where $\mathcal{G}(\tau) = f\big(g(\tau)\big)$. This solution can be related to solution (13.1.3.3) of the Cauchy problem (13.1.3.1), (13.1.3.2) by setting

$$\xi = -\tau\mathcal{G}(\tau), \quad \varphi(\xi) = g(\tau), \quad \mathcal{F}(\xi) = \mathcal{G}(\tau). \tag{13.1.3.21}$$

This corresponds to the continuation of characteristics through the points $y = 0$, $x = \tau$ to the y-axis and to the designation of the points of intersection by $y = \xi$. In this case, the problem of propagation of a signal is formulated as a Cauchy problem.

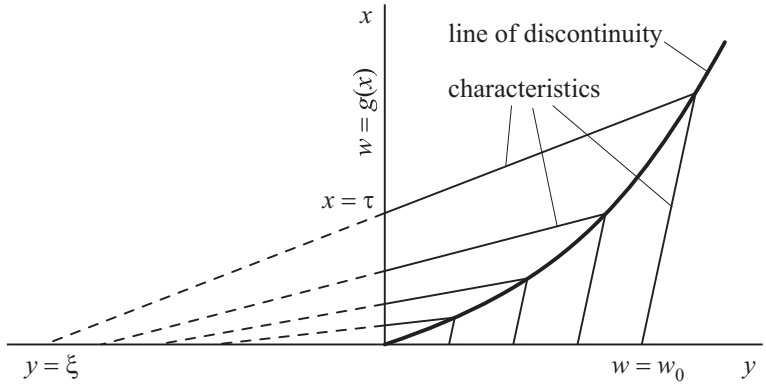


Figure 13.9. Characteristics of the problem of propagation of a signal.

Each domain of nonuniqueness in solution (13.1.3.20) should be replaced by a jump discontinuity. If

$$\mathcal{G}(+0) > a_0, \quad \text{where} \quad a_0 = f(w_0),$$

such a domain arises instantaneously, since the first characteristic $y = \mathcal{G}(+0)x$ is ahead of the last characteristic $y = a_0 x$ of the unperturbed domain. In this case, the discontinuity appears at the origin of coordinates and the relation

$$G - G_0 = (w - w_0)\mathcal{G} - \frac{1}{x - \tau}\int_0^\tau \big[G(\bar\tau) - G_0\big]\, d\bar\tau \tag{13.1.3.22}$$

holds. The quantities w, $\mathcal{G}$, and G are functions of τ in the domain behind the discontinuity and are given by

$$w = g(\tau), \quad \mathcal{G} = f\big(g(\tau)\big), \quad G = F\big(g(\tau)\big).$$

The subscript 0 refers to the values of these variables ahead of the discontinuity, $w = w_0$, $\mathcal{G}_0 = f(w_0)$, and $G_0 = F(w_0)$.

Relations (13.1.3.20) describe the solution in the perturbed domain behind the discontinuity. Equation (13.1.3.22) serves to determine $\tau(x)$ at the point of discontinuity; by setting this value into relations (13.1.3.20), we find both the position of the discontinuity and the value of w immediately behind it.

If $g(x)$ remains constant and equal to w_c, then for $a_c > a_0$, where $a_c = f(w_c)$, the solution has a jump discontinuity propagating at a constant velocity and separating two homogeneous domains with $w = w_c$ and $w = w_0$.

13.1.4. Quasilinear Equations of General Form. Generalized Solution, Jump Condition, and Stability Condition

13.1.4-1. Quasilinear equations in conservative form.

In the general case, the quasilinear equation

$$\frac{\partial w}{\partial x} + f(x, y, w)\frac{\partial w}{\partial y} = g(x, y, w) \tag{13.1.4.1}$$

can be represented in an equivalent, conservative form as

$$\frac{\partial w}{\partial x} + \frac{\partial}{\partial y}F(x, y, w) = G(x, y, w), \tag{13.1.4.2}$$

where

$$F(x, y, w) = \int_{w_0}^{w} f(x, y, t)\, dt, \quad G(x, y, w) = g(x, y, w) + \int_{w_0}^{w} \frac{\partial}{\partial y}\big[f(x, y, t)\big]\, dt, \tag{13.1.4.3}$$

and w_0 is an arbitrary number. In what follows we assume that the functions f and g are continuous and have continuous first derivatives.

As was shown by examples in Subsections 13.1.2 and 13.1.3, characteristics of equation (13.1.4.1) can intersect in some domain, which results in the nonuniqueness of the solution and the absence of a physical interpretation of this solution. For this reason, one has to make use of a generalized solution, described by a discontinuous function instead of a classical smooth solution.

We consider the class of functions $w(x, y) \in \mathcal{K}$ satisfying the following conditions:

1°. In any bounded portion of the half-plane $x \geq 0$, there exists a finite number of lines and points of discontinuity; outside these lines and points, the function $w(x, y)$ is continuous and has continuous first derivatives.

2°. At the lines of discontinuity, $y = y(x)$, the left and right limit values of w exist: $w(x, y-0)$ and $w(x, y+0)$.

13.1.4-2. Generalized solution. Jump condition and stability condition.

A generalized solution may be introduced in the following manner. Let $\psi(x,y) \in \mathcal{C}_1$ be a continuous finite function (which vanishes outside a finite portion of the xy-plane) having continuous first derivatives.

Multiply equation (13.1.4.1) by $\psi(x,y)$ and integrate the resulting relation over the half-plane $\Omega = \{0 \le x < \infty,\ -\infty < y < \infty\}$. On integrating by parts, we obtain

$$\iint_{\Omega}\left[w\frac{\partial\psi}{\partial x} + F(x,y,w)\frac{\partial\psi}{\partial y} + G(x,y,w)\psi(x,y)\right]dy\,dx + \int_{-\infty}^{\infty} w(0,y)\psi(0,y)\,dy = 0. \tag{13.1.4.4}$$

The function $F(x,y,w)$ is defined in equation (13.1.4.3). The integral relation (13.1.4.4) does not contain derivatives of the unknown function and does not lose its meaning for discontinuous $w(x,y)$. The function $w(x,y) \in \mathcal{K}$ will be called the generalized solution of equation (13.1.4.1) if inequalities (13.1.4.4) hold for any finite $\psi(x,y) \in \mathcal{C}_1$.

Basic properties of the stable generalized solution:

1°. In the domain where the solution w is continuously differentiable, equations (13.1.4.1) and (13.1.4.4) are equivalent.

2°. Let $y = y(x)$ be the equation of a discontinuity line of $w(x,y)$. Then the Rankine–Hugoniot jump condition must hold. It expresses the speed of motion of the discontinuity line via the solution parameters ahead of and behind the discontinuity as

$$V = \frac{[F(x,y,w)]}{[w]} \equiv \frac{F\big(x,y(x),w_2(x)\big) - F\big(x,y(x),w_1(x)\big)}{w_2(x) - w_1(x)}, \tag{13.1.4.5}$$

where

$$V = y'(x), \quad w_1(x) = w\big(x,y(x)-0\big), \quad w_2(x) = w\big(x,y(x)+0\big).$$

3°. For $f'_w(x,y,w) \ne 0$, the generalized solution stable with respect to small perturbations of the initial profile (it is stable solutions that are physically realizable) must satisfy the condition

$$f\big(x,y(x),w_2(x)\big) \le V \le f\big(x,y(x),w_1(x)\big). \tag{13.1.4.6}$$

The stability condition (13.1.4.6) has the geometrical meaning that the characteristics issuing from the x-axis (these characteristics "carry" information about the initial data) must intersect the discontinuity line (see Fig. 13.7). This condition is very important since it allows for the existence of a stable generalized solution and provides its uniqueness.

The properties of Items 1° and 2° follow from the integral relation (13.1.4.4), and the condition of Item 3° is additional [it cannot be deduced from the integral relation (13.1.4.4)]. If the stability condition of Item 3° is not imposed, then various generalized solutions satisfying Items 1° and 2° may be constructed.

Example 4. Consider the Cauchy problem for equation (13.1.2.8) with the initial condition (13.1.2.15). We set

$$\overline{w}(x,y) = \begin{cases} w_1 & \text{for } y < Vx, \\ w_2 & \text{for } y > Vx, \end{cases} \qquad \text{where} \quad V = \frac{w_1 + w_2}{2}. \tag{13.1.4.7}$$

This function is constant from the left and right of the discontinuity line $y = Vx$, where the jump condition (13.1.4.5) is met [since $F(x,y,w) = \frac{1}{2}w^2$], and satisfies the initial condition (13.1.2.15). Hence, $\overline{w}$ is a generalized solution.

Figure 13.10 shows the discontinuity line and characteristics corresponding to solution (13.1.4.7). One can see that the characteristics "issue" from the discontinuity line and do not intersect the x-axis. Therefore, solution (13.1.4.7) is unstable, does not satisfy condition (13.1.4.6), and is not physically realizable. A stable solution of this problem was constructed earlier; see relation (13.1.2.18).

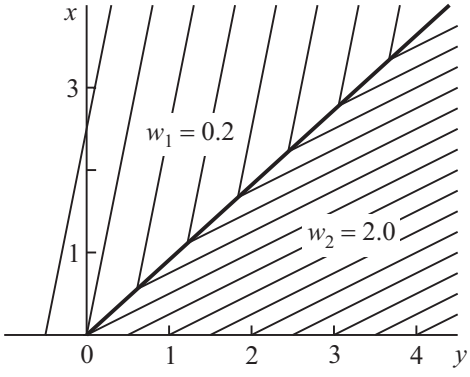


Figure 13.10. Characteristics and the discontinuity line for an unstable discontinuous solution (13.1.4.7).

If $f'_w(x, y, w)$ is not a function of fixed sign, the stability condition for the generalized solution becomes more complicated:

$$\frac{F(x,y,w_*)-F(x,y,w_2)}{w_*-w_2} \le V \le \frac{F(x,y,w_*)-F(x,y,w_1)}{w_*-w_1},$$
$$y = y(x), \quad w_1 < w_* < w_2,$$

where w_* is any value from the interval (w_1, w_2).

Remark. Point out also other ways of defining generalized solutions (using conservation laws and viscosity solutions).

13.1.4-3. Method for constructing stable generalized solutions.

Consider the Cauchy problem for the quasilinear equation

$$\frac{\partial w}{\partial x} + \frac{\partial}{\partial y}F(x, y, w) = 0 \tag{13.1.4.8}$$

subject to the initial condition

$$w = \varphi(y) \quad \text{at} \quad x = 0. \tag{13.1.4.9}$$

It is assumed that the function $F(x, y, w)$ is continuously differentiable with respect to all its arguments for $x \ge 0$, $-\infty < y < \infty$ and any bounded w. We also assume that the second derivative F_{ww} is positive. Let the functions $\varphi(y)$ and $\varphi'(y)$ be piecewise-continuous for any finite y.

The characteristic system for equation (13.1.4.8) has the form

$$y'_x = F_w(x, y, w), \quad w'_x = -F_y(x, y, w), \tag{13.1.4.10}$$

where F_w and F_y are the partial derivatives of the function F with respect to w and y.

Suppose the functions

$$y(x) = Y(x, \tau, \xi, \eta), \quad w(x) = W(x, \tau, \xi, \eta) \tag{13.1.4.11}$$

are solutions of system (13.1.4.10) satisfying the boundary conditions

$$y(0) = \eta, \quad y(\tau) = \xi. \tag{13.1.4.12}$$

Here, η and ξ are arbitrary numbers and $\tau > 0$. We assume that problem (13.1.4.10), (13.1.4.12) has a unique bounded solution.

The stable generalized solution of the Cauchy problem (13.1.4.8), (13.1.4.9) is given by

$$\begin{aligned} w(x, y-0) &= W\big(x, x, y, \xi_-(x, y)\big), \\ w(x, y+0) &= W\big(x, x, y, \xi_+(x, y)\big), \end{aligned} \tag{13.1.4.13}$$

where $\xi_-(x, y)$ and $\xi_+(x, y)$ denote, respectively, the greatest lower bound and the least upper bound of the set of values $\{\xi = \xi_n\}$ for which the function

$$I(x, y, \xi) = \int_0^{\xi} \big[\varphi(\eta) - W(0, x, y, \eta)\big]\, d\eta \tag{13.1.4.14}$$

takes the minimum value for fixed x and y ($x > 0$). If function (13.1.4.14) takes the minimum value for a single $\xi = \xi_1$, then $\xi_- = \xi_+$ and relation (13.1.4.14) describes the classical smooth solution.

13.2. Nonlinear Equations

13.2.1. Solution Methods

13.2.1-1. Complete, general, and singular integrals.

A *nonlinear first-order partial differential equation* with two independent variables has the general form

$$F(x, y, w, p, q) = 0, \qquad \text{where} \quad p = \frac{\partial w}{\partial x}, \quad q = \frac{\partial w}{\partial y}. \tag{13.2.1.1}$$

Such equations are encountered in analytical mechanics, calculus of variations, optimal control, differential games, dynamic programming, geometric optics, differential geometry, and other fields.

In this subsection, we consider only smooth solutions $w = w(x, y)$ of equation (13.2.1.1), which are continuously differentiable with respect to both arguments (Subsection 13.2.3 deals with nonsmooth solutions).

1°. Let a particular solution of equation (13.2.1.1),

$$w = \Xi(x, y, C_1, C_2), \tag{13.2.1.2}$$

depending on two parameters C_1 and C_2, be known. The two-parameter family of solutions (13.2.1.2) is called a *complete integral* of equation (13.2.1.1) if the rank of the matrix

$$M = \begin{pmatrix} \Xi_1 & \Xi_{x1} & \Xi_{y1} \\ \Xi_2 & \Xi_{x2} & \Xi_{y2} \end{pmatrix} \tag{13.2.1.3}$$

is equal to two in the domain being considered (for example, this is valid if $\Xi_{x1}\Xi_{y2} - \Xi_{x2}\Xi_{y1} \neq 0$). In equation (13.2.1.3), Ξ_n denotes the partial derivative of Ξ with respect to C_n ($n = 1, 2$), Ξ_{xn} is the second partial derivative with respect to x and C_n, and Ξ_{yn} is the second partial derivative with respect to y and C_n.

In some cases, a complete integral can be found using the method of undetermined coefficients by presetting an appropriate structure of the particular solution sought. (The complete integral is determined by the differential equation nonuniquely.)

Example 1. Consider the equation

$$\frac{\partial w}{\partial x} = a\left(\frac{\partial w}{\partial y}\right)^n + b.$$

We seek a particular solution as the sum $w = C_1 y + C_2 + C_3 x$. Substituting this expression into the equation yields the relation $C_3 = aC_1^n + b$ for the coefficients C_1 and C_3. With this relation, we find a complete integral in the form $w = C_1 y + \big(aC_1^n + b\big)x + C_2$.

A complete integral of equation (13.2.1.1) is often written in implicit form:*

$$\Xi(x, y, w, C_1, C_2) = 0. \tag{13.2.1.4}$$

2°. The *general integral* of equation (13.2.1.1) can be represented in parametric form by using the complete integral (13.2.1.2) [or (13.2.1.4)] and the two equations

$$\begin{aligned} &C_2 = f(C_1), \\ &\frac{\partial \Xi}{\partial C_1} + \frac{\partial \Xi}{\partial C_2} f'(C_1) = 0, \end{aligned} \tag{13.2.1.5}$$

where f is an arbitrary function and the prime stands for the derivative. In a sense, the general integral plays the role of the general solution depending on an arbitrary function (the questions whether it describes all solutions calls for further analysis).

Example 2. For the equation considered in the first example, the general integral can be written in parametric form by using the relations

$$w = C_1 y + \big(aC_1^n + b\big)x + C_2, \quad C_2 = f(C_1), \quad y + anC_1^{n-1}x + f'(C_1) = 0.$$

Eliminating C_2 from these relations and renaming C_1 by C, one can represent the general integral in a more graphic manner in the form

$$\begin{aligned} w &= Cy + \big(aC^n + b\big)x + f(C), \\ y &= -anC^{n-1}x + f'(C). \end{aligned}$$

3°. *Singular integrals* of equation (13.2.1.1) can be found without invoking a complete integral by eliminating p and q from the following system of three algebraic (or transcendental) equations:

$$F = 0, \quad F_p = 0, \quad F_q = 0,$$

where the first equation coincides with equation (13.2.1.1).

13.2.1-2. Method of separation of variables. Equations of special form.

The method of separation of variables implies searching for a complete integral as the sum or product of functions of various arguments. Such solutions are called *additive separable* and *multiplicative separable*, respectively. Presented below are structures of complete integrals for some classes of nonlinear equations admitting separation of variables.

1°. If the equation does not depend explicitly on y and w, i.e.,

$$F(x, w_x, w_y) = 0,$$

then one can seek a complete integral in the form of the sum of two functions with different arguments

$$w = C_1 y + C_2 + u(x).$$

The new unknown function u is determined by solving the following ordinary differential equation:

$$F(x, u'_x, C_1) = 0.$$

Expressing u'_x from this equation in terms of x, one arrives at a separable differential equation for $u = u(x)$.

* In equations (13.2.1.2) and (13.2.1.4), the symbol Ξ denotes different functions.

2°. Consider an equation with separated variables

$$F_1(x, w_x) = F_2(y, w_y).$$

Then one can seek a complete integral as the sum of two functions with different arguments,

$$w = u(x) + v(y) + C_1,$$

which are determined by the following two ordinary differential equations:

$$F_1(x, u'_x) = C_2,$$
$$F_2(y, v'_y) = C_2.$$

3°. Let the equation have the form (generalizes the equation of Item 2°)

$$F_1(x, w_x) + F_2(y, w_y) = aw.$$

Then one can seek a complete integral as the sum of two functions with different arguments,

$$w = u(x) + v(y) + C_1,$$

which are determined by the following two ordinary differential equations:

$$F_1(x, u'_x) - au = aC_1 + C_2,$$
$$F_2(y, v'_y) - av = -C_2,$$

where C_1 is an arbitrary constant.

4°. Suppose the equation can be rewritten in the form

$$F\big(\varphi(x, w_x),\ y, w_y\big) = 0.$$

Then one can seek a complete integral as the sum of two functions with different arguments,

$$w(x, y) = u(x) + v(y) + C_1,$$

which are determined by the following two ordinary differential equations:

$$\varphi(x, u'_x) = C_2,$$
$$F(C_2, y, v'_y) = 0,$$

where C_2 is an arbitrary constant.

5°. Let the equation have the form

$$F_1(x, w_x/w) = w^k F_2(y, w_y/w).$$

Then one can seek a complete integral in the form of the product of two functions with different arguments,

$$w = u(x)v(y),$$

which are determined by the following two ordinary differential equations:

$$F_1(x, u'_x/u) = C_1 u^k,$$
$$F_2(y, v'_y/v) = C_1 v^{-k},$$

where C_1 is an arbitrary constant.

6°. Table 13.2 lists complete integrals of the above and some other nonlinear equations of general form involving arbitrary functions with several arguments.

► *Section T7.3 presents complete integrals for many more nonlinear first-order partial differential equations with two independent variables than in Table 13.2.*

TABLE 13.2
Complete integrals for some special types of nonlinear first-order partial differential equations; C_1 and C_2 are arbitrary constants

No.	Equations and comments	Complete integrals	Auxiliary equations
1	$F(w_x, w_y) = 0$, does not depend on x, y, and w implicitly	$w = C_1 + C_2x + C_3y$	$F(C_2, C_3) = 0$
2	$F(x, w_x, w_y) = 0$, does not depend on y and w implicitly	$w = C_1y + C_2 + u(x)$	$F(x, u'_x, C_1) = 0$
3	$F(w, w_x, w_y) = 0$, does not depend on x and y implicitly	$w = u(z),\ z = C_1x + C_2y$	$F(u, C_1u'_z, C_2u'_z) = 0$
4	$F_1(x, w_x) = F_2(y, w_y)$, separated equation	$w = u(x) + v(y) + C_1$	$F_1(x, u'_x) = C_2$, $F_2(y, v'_y) = C_2$
5	$F_1(x, w_x) + F_2(y, w_y) = aw$, generalizes equation 4	$w = u(x) + v(y)$	$F_1(x, u'_x) - au = C_1$, $F_2(y, v'_y) - av = -C_1$
6	$F_1(x, w_x) = e^{aw}F_2(y, w_y)$, generalizes equation 4	$w = u(x) + v(y)$	$F_1(x, u'_x) = C_1e^{au}$, $F_2(y, v'_y) = C_1e^{-av}$
7	$F_1(x, w_x/w) = w^kF_2(y, w_y/w)$, can be reduced to equation 6 by the change of variable $w = e^z$	$w = u(x)v(y)$	$F_1(x, u'_x/u) = C_1u^k$, $F_2(y, v'_y/v) = C_1v^{-k}$
8	$w = xw_x + yw_y + F(w_x, w_y)$, Clairaut equation	$w = C_1x + C_2y + F(C_1, C_2)$	—
9	$F(x, w_x, w_y, w - yw_y) = 0$, generalizes equation 2	$w = C_1y + u(x)$	$F\big(x, u'_x, C_1, u\big) = 0$
10	$F(w, w_x, w_y, xw_x + yw_y) = 0$, generalizes equations 3 and 8	$w = u(z),\ z = C_1x + C_2y$	$F(u, C_1u'_z, C_2u'_z, zu'_z) = 0$
11	$F\big(\varphi(x, w_x),\ y, w_y\big) = 0$, generalizes equation 4	$w = u(x) + v(y) + C_1$	$\varphi(x, u'_x) = C_2$, $F(C_2, y, v'_y) = 0$

13.2.1-3. Lagrange–Charpit method.

Suppose that a first integral,

$$\Phi(x, y, w, p, q) = C_1, \tag{13.2.1.6}$$

of the characteristic system of ordinary differential equations

$$\frac{dx}{F_p} = \frac{dy}{F_q} = \frac{dw}{pF_p + qF_q} = -\frac{dp}{F_x + pF_w} = -\frac{dq}{F_y + qF_w} \tag{13.2.1.7}$$

is known. Here,

$$p = \frac{\partial w}{\partial x}, \quad q = \frac{\partial w}{\partial y}, \quad F_x = \frac{\partial F}{\partial x}, \quad F_y = \frac{\partial F}{\partial y}, \quad F_w = \frac{\partial F}{\partial w}, \quad F_p = \frac{\partial F}{\partial p}, \quad F_q = \frac{\partial F}{\partial q}.$$

We assume that solution (13.2.1.6) and equation (13.2.1.1) can be solved for the derivatives p and q, i.e.,

$$p = \varphi_1(x, y, w, C_1), \quad q = \varphi_2(x, y, w, C_1). \tag{13.2.1.8}$$

The first equation of this system can be treated as an ordinary differential equation with independent variable x and parameter y. On finding the solution of this equation depending on an arbitrary function $\psi(y)$, one substitutes this solution into the second equation to arrive at an ordinary differential equation for ψ. On determining $\psi(y)$ and on substituting it into the general solution of the first equation of (13.2.1.8), one finds a complete integral of equation (13.2.1.1). In a similar manner, one can start solving system (13.2.1.8) with the second equation, treating it as an ordinary differential equation with independent variable y and parameter x.

Example 3. Consider the equation

$$ywp^2 - q = 0, \qquad \text{where} \quad p = \frac{\partial w}{\partial x}, \quad q = \frac{\partial w}{\partial y}.$$

In this case, the characteristic system (13.2.1.7) has the form

$$\frac{dx}{2ywp} = -\frac{dy}{1} = \frac{dw}{2ywp^2 - q} = -\frac{dp}{yp^3} = -\frac{dq}{wp^2 + yp^2q}.$$

By making use of the original equation, we simplify the denominator of the third ratio to obtain an integrable combination: $dw/(ywp^2) = -dp/(yp^3)$. This yields the first integral $p = C_1/w$. Solving the original equation for q, we obtain the system

$$p = \frac{C_1}{w}, \quad q = \frac{C_1^2 y}{w}.$$

The general solution of the first equation has the form $w^2 = 2C_1x + \psi(y)$, where $\psi(y)$ is an arbitrary function. With this solution, it follows from the second equation of the system that $\psi'(y) = 2C_1^2 y$. Thus, $\psi(y) = C_1^2y^2 + C_2$. Finally, we arrive at a complete integral of the form

$$w^2 = 2C_1x + C_1^2y^2 + C_2.$$

Note that the general solution of the completely integrable Pfaff equation (see Subsection 15.14.2)

$$dw = \varphi_1(x, y, w, C_1)\,dx + \varphi_2(x, y, w, C_1)\,dy \tag{13.2.1.9}$$

is a complete integral of equation (13.2.1.1). Here, the functions φ_1 and φ_2 are the same as in system (13.2.1.8).

Remark. The relation $F(x, y, w, p, q) = C$ is an obvious first integral of the characteristic system (13.2.1.7). Hence, the function Φ determining the integral (13.2.1.6) must differ from F. However, the use of relation (13.2.1.1) makes it possible to reduce the order of system (13.2.1.7) by one.

13.2.1-4. Construction of a complete integral with the aid of two first integrals.

Suppose two independent first integrals,

$$\Phi(x, y, w, p, q) = C_1, \quad \Psi(x, y, w, p, q) = C_2, \tag{13.2.1.10}$$

of the characteristic system of ordinary differential equations (13.2.1.7) are known. Assume that the functions F, Φ, and Ψ determining equation (13.2.1.1) and the integrals (13.2.1.10) satisfy the two conditions

$$\begin{aligned} &\text{(a)} \quad J \equiv \frac{\partial(F, \Phi, \Psi)}{\partial(w, p, q)} \not\equiv 0, \\ &\text{(b)} \quad [\Phi, \Psi] \equiv \begin{vmatrix} \Phi_p & \Phi_x + p\Phi_w \\ \Psi_p & \Psi_x + p\Psi_w \end{vmatrix} + \begin{vmatrix} \Phi_q & \Phi_y + q\Phi_w \\ \Psi_q & \Psi_y + q\Psi_w \end{vmatrix} \equiv 0, \end{aligned} \tag{13.2.1.11}$$

where J is the Jacobian of F, Φ, and Ψ with respect to w, p, and q, and $[\Phi, \Psi]$ is the *Jacobi–Mayer bracket*. In this case, relations (13.2.1.1) and (13.2.1.10) form a parametric representation of the complete integral of equation (13.2.1.1) (p and q are considered to be parameters). Eliminating p and q from equations (13.2.1.1) and (13.2.1.10) followed by solving the obtained relation for w yields a complete integral in an explicit form $w = w(x, y, C_1, C_2)$.

Example 4. Consider the equation

$$pq - aw = 0, \qquad \text{where} \quad p = \frac{\partial w}{\partial x}, \quad q = \frac{\partial w}{\partial y}.$$

The characteristic system (13.2.1.7) has the form

$$\frac{dx}{q} = \frac{dy}{p} = \frac{dw}{2pq} = \frac{dp}{ap} = \frac{dq}{aq}.$$

Equating the first ratio with the fifth one and the second ratio with the fourth one, we obtain the first integrals

$$q - ax = C_1, \qquad p - ay = C_2.$$

Thus, $F = pq - aw$, $\Phi = q - ax$, and $\Psi = p - ax$. These functions satisfy conditions (13.2.1.11). Solving the equation and the first integrals for w yields a complete integral of the form

$$w = \frac{1}{a}(ax + C_1)(ay + C_2).$$

13.2.1-5. Case where the equation does not depend on w explicitly.

Suppose the original equation does not contain the unknown first explicitly, i.e., it has the form

$$F(x, y, p, q) = 0. \tag{13.2.1.12}$$

1°. Given a one-parameter family of solutions $w = \overline{\Xi}(x, y, C_1)$ such that $\overline{\Xi}'_1 \not\equiv \text{const}$, a complete integral is given by $w = \overline{\Xi}(x, y, C_1) + C_2$.

2°. The first integral may be sought in the form

$$\Phi(x, y, p, q) = C_1$$

similar to that of equation (13.2.1.12). In this case, the characteristic system (13.2.1.7) is represented as

$$\frac{dx}{F_p} = \frac{dy}{F_q} = -\frac{dp}{F_x} = -\frac{dq}{F_y}.$$

The corresponding Pfaff equation (13.2.1.9) becomes

$$dw = \varphi_1(x, y, C_1)\,dx + \varphi_2(x, y, C_1)\,dy.$$

One may integrate this equation in quadrature, thus arriving at the following expression for the complete integral:

$$w = \int_{x_0}^{x} \varphi_1(t, y, C_1)\,dt + \int_{y_0}^{y} \varphi_2(x_0, s, C_1)\,ds + C_2, \tag{13.2.1.13}$$

where x_0 and y_0 are arbitrary numbers.

3°. Suppose that equation (13.2.1.12) can be solved for p or q, for example,

$$p = -\mathcal{H}(x, y, q).$$

Then, by differentiating this relation with respect to y, we obtain a quasilinear equation for the derivative q in the form

$$\frac{\partial q}{\partial x} + \frac{\partial}{\partial y}\mathcal{H}(x, y, q) = 0, \qquad q = \frac{\partial w}{\partial y}.$$

This equation is simpler than the original one; qualitative features of it and solution methods can be found in Section 13.1.1.

13.2.1-6. Hamilton–Jacobi equation.

Equation (13.2.1.1) solved for one of the derivatives, e.g.,

$$p + \mathcal{H}(x, y, w, q) = 0, \qquad \text{where} \quad p = \frac{\partial w}{\partial x}, \quad q = \frac{\partial w}{\partial y}, \tag{13.2.1.14}$$

is commonly referred to as the *Hamilton–Jacobi equation** and the function $\mathcal{H}$ as the *Hamiltonian*. Equations of the form (13.2.1.14) are frequently encountered in various fields of mechanics, control theory, and differential games, where the variable x usually plays the role of time and the variable y the role of the spatial coordinate. To the Hamilton–Jacobi equation (13.2.1.14) there corresponds the function $F(x, y, w, p, q) = p + \mathcal{H}(x, y, w, q)$ in equation (13.2.1.1).

The characteristic system (13.2.1.7) for equation (13.2.1.14) can be reduced, by taking into account the relation $p = -\mathcal{H}$, to a simpler system consisting of three differential equations,

$$y'_x = \mathcal{H}_q, \quad w'_x = q\mathcal{H}_q - \mathcal{H}, \quad q'_x = -q\mathcal{H}_w - \mathcal{H}_y, \tag{13.2.1.15}$$

which are independent of p; the left-hand sides of these equations are derivatives with respect to x.

13.2.2. Cauchy Problem. Existence and Uniqueness Theorem

13.2.2-1. Statement of the problem. Solution procedure.

Consider the Cauchy problem for equation (13.2.1.1) subject to the initial conditions

$$x = h_1(\xi), \quad y = h_2(\xi), \quad w = h_3(\xi), \tag{13.2.2.1}$$

where ξ is a parameter ($\alpha \le \xi \le \beta$) and the $h_k(\xi)$ are given functions.

The solution of this problem is carried out in several steps:

1°. First, one determines additional initial conditions for the derivatives,

$$p = p_0(\xi), \quad q = q_0(\xi). \tag{13.2.2.2}$$

To this end, one must solve the algebraic (or transcendental) system of equations

$$F\big(h_1(\xi), h_2(\xi), h_3(\xi), p_0, q_0\big) = 0, \tag{13.2.2.3}$$

$$p_0 h'_1(\xi) + q_0 h'_2(\xi) - h'_3(\xi) = 0 \tag{13.2.2.4}$$

for p_0 and q_0. Equation (13.2.2.3) results from substituting the initial data (13.2.2.1) into the original equation (13.2.1.1). Equation (13.2.2.4) is a consequence of the dependence of w on x and y and the relation $dw = p\,dx + q\,dy$, where dx, dy, and dw are calculated in accordance with the initial data (13.2.2.1).

2°. One solves the autonomous system

$$\frac{dx}{F_p} = \frac{dy}{F_q} = \frac{dw}{pF_p + qF_q} = -\frac{dp}{F_x + pF_w} = -\frac{dq}{F_y + qF_w} = d\tau, \tag{13.2.2.5}$$

which is obtained from (13.2.1.7) by introducing the additional variable τ (playing the role of time).

* The Hamilton–Jacobi equation often means equation (13.2.1.14) that does not depend on w explicitly, i.e., the equation $p + \mathcal{H}(x, y, q) = 0$.

3°. Finally, one determines the constant of integration from the initial conditions

$$x = h_1(\xi), \quad y = h_2(\xi), \quad w = h_3(\xi), \quad p = p_0(\xi), \quad q = q_0(\xi) \quad \text{at} \quad \tau = 0, \qquad (13.2.2.6)$$

which are the combination of conditions (13.2.2.1) and (13.2.2.2). This results in the three functions

$$x = x(\tau, \xi), \quad y = y(\tau, \xi), \quad w = w(\tau, \xi), \qquad (13.2.2.7)$$

which give the solution of the stated Cauchy problem in parametric form (τ and ξ are parameters).

13.2.2-2. Existence and uniqueness theorem.

Let the function $F = F(x, y, w, p, q)$, determining equation (13.2.1.1), be twice continuously differentiable with respect to all five arguments (in the domain being considered), with $F_p^2 + F_q^2 \neq 0$. Let the functions $h_1(\xi)$, $h_2(\xi)$, and $h_3(\xi)$, determining the initial conditions (13.2.2.1), be twice differentiable with respect to ξ, with $(h_1')^2 + (h_2')^2 \neq 0$. Assume that the function $p_0(\xi)$ and $q_0(\xi)$, determining the additional initial condition (13.2.2.2), satisfy the system (13.2.2.3), (13.2.2.4). Moreover, we adopt the *transversality condition*

$$\Delta \equiv F_p h_2' - F_q h_1' \neq 0,$$

where the functions h_1, h_2, p, and q are defined by equations (13.2.2.1) and (13.2.2.2), and the prime denotes the derivative with respect to ξ. Under the adopted assumptions, there exists a unique, twice continuously differentiable solution of equation (13.2.1.1) satisfying the initial conditions (13.2.2.1) and (13.2.2.2).

Remark 1. This theorem has a local character, i.e., the existence of a unique solution of the Cauchy problem is merely guaranteed in some neighborhood of the line defined by the initial data (13.2.2.1) together with the additional conditions (13.2.2.2).

Remark 2. The algebraic (or transcendental) system (13.2.2.3), (13.2.2.4) can have several solutions (see Example 2 at the end of this subsection) that result in distinct additional conditions (13.2.2.2) for the derivatives p and q. Each pair of these additional conditions will generate a solution of the Cauchy problem (13.2.1.1), (13.2.2.1).

Remark 3. For nonlinear equations, the global solution of the Cauchy problem (13.2.1.1), (13.2.2.1) may be nonunique for another reason — because the characteristics in the xy-plane may intersect (see Example 1 in Subsection 13.2.3). Such a situation is discussed in Subsections 13.1.3 and 13.1.4, where quasilinear equations are considered.

13.2.2-3. Cauchy problem for the Hamilton–Jacobi equation.

The initial condition for the Hamilton–Jacobi equation (13.2.1.14) is usually stated in the form

$$w = \varphi(y) \quad \text{at} \quad x = L. \qquad (13.2.2.8)$$

In this case, the solution of the Cauchy problem is reduced to the solution of the characteristic system (13.2.1.15) subject to the initial conditions

$$y = \xi, \quad w = \varphi(\xi), \quad q = \varphi'(\xi) \quad \text{at} \quad x = L, \qquad (13.2.2.9)$$

where the prime stands for the derivative with respect to the parameter ξ.

13.2.2-4. Examples of solving the Cauchy problem.

Consider a few specific examples.

Example 1. Find a solution of the equation

$$aw = pq, \qquad \text{where} \quad p = \frac{\partial w}{\partial x}, \quad q = \frac{\partial w}{\partial y}, \tag{13.2.2.10}$$

satisfying the initial condition

$$w = f(y) \quad \text{at} \quad x = 0. \tag{13.2.2.11}$$

Rewrite the initial condition (13.2.2.11) in parametric form:

$$x = 0, \quad y = \xi, \quad w = f(\xi). \tag{13.2.2.12}$$

The system of equations (13.2.2.3), (13.2.2.4) for $p_0(\xi)$ and $q_0(\xi)$ is written as

$$af(\xi) = p_0 q_0, \quad q_0 - f'(\xi) = 0.$$

It follows that

$$p_0 = a\frac{f(\xi)}{f'(\xi)}, \quad q_0 = f'(\xi). \tag{13.2.2.13}$$

If $F = pq - aw$, the characteristic system (13.2.2.5) is written as

$$\frac{dx}{q} = \frac{dy}{p} = \frac{dw}{2pq} = \frac{dp}{ap} = \frac{dq}{aq} = d\tau. \tag{13.2.2.14}$$

Its solution is given by the relations

$$p = C_1 e^{a\tau}, \quad q = C_2 e^{a\tau}, \quad x = \frac{C_2}{a}e^{a\tau} + C_3, \quad y = \frac{C_1}{a}e^{a\tau} + C_4, \quad w = \frac{C_1 C_2}{a}e^{2a\tau} + C_5; \tag{13.2.2.15}$$

it is easy to integrate the last two equations first.

By using the initial conditions (13.2.2.12), (13.2.2.13), which must be satisfied at $\tau = 0$, we find the constants of integration in the form

$$C_1 = a\frac{f(\xi)}{f'(\xi)}, \quad C_2 = f'(\xi), \quad C_3 = -\frac{f'(\xi)}{a}, \quad C_4 = \xi - \frac{f(\xi)}{f'(\xi)}, \quad C_5 = 0.$$

Substituting theses values into (13.2.2.15) yields a solution of the Cauchy problem (13.2.2.10), (13.2.2.11) in the parametric form

$$x = \frac{1}{a}f'(\xi)\big(e^{a\tau} - 1\big), \quad y = \frac{f(\xi)}{f'(\xi)}\big(e^{a\tau} - 1\big) + \xi, \quad w = f(\xi)e^{2a\tau}.$$

Example 2. Find a solution of the equation

$$\left(\frac{\partial w}{\partial x}\right)^2 + \left(\frac{\partial w}{\partial y}\right)^2 = a^2, \tag{13.2.2.16}$$

passing through the circle

$$x^2 + y^2 = b^2, \quad w = 0. \tag{13.2.2.17}$$

By introducing a parameter ξ, we rewrite the equation of the circle in the form

$$x = b\sin\xi, \quad y = b\cos\xi, \quad w = 0. \tag{13.2.2.18}$$

In this case, the equations (13.2.2.3), (13.2.2.4) for determining additional initial conditions are

$$p_0^2 + q_0^2 = a^2, \quad p_0\cos\xi - \sin\xi q_0 = 0.$$

Whence,

$$p_0 = \varepsilon a\sin\xi, \quad q_0 = \varepsilon a\cos\xi, \qquad \text{where} \quad \varepsilon = \pm 1. \tag{13.2.2.19}$$

For $F = p^2 + q^2 - a^2$, system (13.2.2.5) is represented as

$$\frac{dx}{2p} = \frac{dy}{2q} = \frac{dw}{2(p^2+q^2)} = -\frac{dp}{0} = -\frac{dq}{0} = d\tau. \tag{13.2.2.20}$$

The general solution is given by (the last two equations are integrated first)

$$p = C_1, \quad q = C_2, \quad x = 2C_1\tau + C_3, \quad y = 2C_2\tau + C_4, \quad w = 2(C_1^2 + C_2^2)\tau + C_5. \tag{13.2.2.21}$$

Using the initial conditions (13.2.2.18), (13.2.2.19), which must be satisfied at $\tau = 0$, we find the constants of integration

$$C_1 = \varepsilon a \sin\xi, \quad C_2 = \varepsilon a \cos\xi, \quad C_3 = b \sin\xi, \quad C_4 = b \cos\xi, \quad C_5 = 0, \qquad \text{where} \quad \varepsilon = \pm 1.$$

Substituting these values into (13.2.2.21) yields a solution of the Cauchy problem (13.2.2.16), (13.2.2.17) in the parametric form

$$x = (2\varepsilon a\tau + b)\sin\xi, \quad y = (2\varepsilon a\tau + b)\cos\xi, \quad w = 2a^2\tau.$$

On eliminating the parameters ξ and τ, we can rewrite this solution in a more graphic form,

$$a^2(x^2 + y^2) = (ab \pm w)^2. \tag{13.2.2.22}$$

Relation (13.2.2.22) describes two circular coaxial cones in the space (x, y, w). The circle (13.2.2.17) is a base of the cones. The common axis coincides with the w-axis. The vertices of the cones have the coordinates $w = \pm ab$.

It is significant that solution (13.2.2.22) is a many-valued function.

13.2.3. Generalized Viscosity Solutions and Their Applications

13.2.3-1. Preliminary remarks.

Subsections 13.2.1 and 13.2.2 dealt with classical smooth solutions $w = w(x, y)$, having continuous derivatives with respect to both arguments. However, in optimal control, differential games, and some other applications, problems arise whose solutions are continuous but nonsmooth functions. To describe and construct generalized solutions of this sort, approaches other than those outlined above are required. It should be noted that for determining generalized solutions to nonlinear equations of the general forms (13.2.1.1) and (13.2.1.14), it turns out to be ineffective to use graphic constructs such as integral relations and conservation laws, which are frequently encountered in the theory of quasilinear equations (see Subsections 13.1.3 and 13.1.4).

Note that nonsmoothness of a solution can be caused by (i) the intersection of characteristics in the xy-plane (see Example 1 below), (ii) nonsmoothness of initial conditions, and/or (iii) nonsmoothness of the functions F and $\mathcal{H}$ determining equations (13.2.1.1) and (13.2.1.14).

13.2.3-2. Viscosity solutions based on the use of a parabolic equation.

The solution of the Cauchy problem for equation (13.2.1.14) subject to the initial condition

$$w = \varphi(y) \quad \text{at} \quad x = 0 \tag{13.2.3.1}$$

may be approximated by the solution of the following second-order partial differential equation of the parabolic type:

$$\frac{\partial u}{\partial x} + \mathcal{H}\left(x, y, u, \frac{\partial u}{\partial y}\right) = \varepsilon \frac{\partial^2 u}{\partial y^2} \qquad (\varepsilon > 0). \tag{13.2.3.2}$$

This equation is subject to the same initial condition (13.2.3.1). For a fairly wide class of functions $\mathcal{H}$ and φ, the Cauchy problem for equation (13.2.3.2) is known to have a unique solution. In the theory of Hamilton–Jacobi equations, this fact is used to determine the solution of the Cauchy problem (13.2.1.14), (13.2.3.1) as a limit of the solution of problem

(13.2.3.2), (13.2.3.1): $w(x,y) = \lim\limits_{\varepsilon\to 0} u(x,y)$. Just as in the theory of quasilinear equations (see Subsection 13.1.3), this structure, based on proceeding to the limit as $\varepsilon \to 0$, is referred to as the method of vanishing viscosity, and the limit function as the viscosity solution of the Hamilton–Jacobi equation.

The vanishing viscosity method can be implemented, for example, by a numerical solution of problem (13.2.3.2), (13.2.3.1) for sufficiently small ε (in this case, one need not search for singular points, at which the smoothness of the solution is violated). However, this method is difficult to use for constructing analytical solutions, since one has to treat a more complex, second-order partial differential equation.

13.2.3-3. Viscosity solutions based on test functions and differential inequalities.

A continuous function $w = w(x,y)$ is called the viscosity solution of the initial value problem (13.2.1.1), (13.2.3.1) in a layer $0 \le x \le L$ if the following two conditions are satisfied:

1°. The function $w = w(x,y)$ satisfies the initial condition (13.2.3.1).

2°. Let $\psi(x,y)$ be any continuously differentiable test function. If (x°, y°) is a local extremum point of the difference

$$w(x,y) - \psi(x,y), \tag{13.2.3.3}$$

then the following relations hold at this point:

$$\begin{aligned} &F\big(x^\circ, y^\circ, w^\circ, \psi_x^\circ, \psi_y^\circ\big) \ge 0 \qquad \text{if } (x^\circ, y^\circ) \text{ is a local minimum point,} \\ &F\big(x^\circ, y^\circ, w^\circ, \psi_x^\circ, \psi_y^\circ\big) \le 0 \qquad \text{if } (x^\circ, y^\circ) \text{ is a local maximum point.} \end{aligned} \tag{13.2.3.4}$$

Only those local extremum points that lie within the layer in question ($0 < x^\circ < L$) are to be examined.

Note that it is not necessary that a test function $\psi(x,y)$ exists for which the difference (13.2.3.3) has a local extremum. But if such a function does exist, then condition (13.2.3.4) must hold.

If the Cauchy problem has a smooth classical solution, then it coincides with the viscosity generalized solution.

In optimal control and differential games, terminal value problems are encountered apart from initial value problems. In terminal value problems, the solution of equations (13.2.1.1) and (13.2.1.14) is sought in the layer $0 \le x \le L$ and the "initial" condition is set at the right endpoint $x = L$. For these problems, the inequalities in (13.2.3.4) must be changed for the opposites. The terminal value problems can be reduced to initial value problems by introducing a new independent variable $z = L - x$ instead of x.

13.2.3-4. Local structure of generalized viscosity solutions.

A generalized solution, $w(x,y)$, consists of regular and singular points. In a neighborhood of a regular point, the function $w(x,y)$ is a solution in the classical sense (such twice continuously differentiable solutions are discussed in the existence and uniqueness theorem in Subsection 13.2.2). All points that are not regular are called singular points.

Let D be a sufficiently small neighborhood of a singular point (x_*, y_*). Usually, it happens that the singular points form a smooth curve Γ, which passes through (x_*, y_*) and

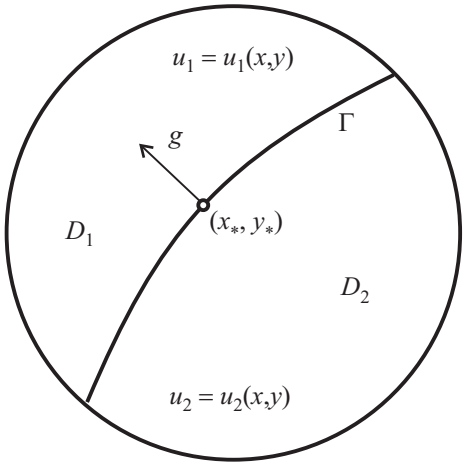


Figure 13.11. Generalized viscosity solution undergoes a discontinuity along the singular curve Γ; the solution is smooth in the subdomains D_1 and D_2.

divides the domain D into two subdomains D_1 and D_2 (see Fig. 13.11). On both sides of Γ, the generalized solution w is defined by different classical solutions u_1 and u_2, i.e.,

$$w(x,y) = \begin{cases} u_1(x,y) & \text{if } (x,y) \in D_1, \\ u_2(x,y) & \text{if } (x,y) \in D_2. \end{cases} \tag{13.2.3.5}$$

The solutions u_1 and u_2 join together along the interface Γ in a continuous but nonsmooth manner. When crossing Γ, the derivatives of the generalized solution, $\partial w/\partial x$ and $\partial w/\partial y$, undergo a discontinuity. We assume that the smooth components u_1 and u_2 of the generalized solution are smoothly expended to the entire domain D. Then the equation of the curve Γ, formed by the singular points, can be represented as

$$g(x,y) = 0, \qquad \text{where} \quad g(x,y) = u_2(x,y) - u_1(x,y). \tag{13.2.3.6}$$

The gradient of g, directed along the normal to Γ, is given by

$$\nabla g = (p_2 - p_1)\mathbf{e}_x + (q_2 - q_1)\mathbf{e}_y, \qquad p_n = \frac{\partial u_n}{\partial x}, \quad q_n = \frac{\partial u_n}{\partial y},$$

where $\mathbf{e}_x$ and $\mathbf{e}_y$ are the direction cosines along the x- and y-axes, respectively. Two situations are possible.

1°. The vector ∇g is directed from D_2 toward D_1. In this case, the following statements are valid:

(A) The generalized solution in D can be represented as $w = \min[u_1, u_2]$; see Fig. 13.12.

(B) There is no smooth test function $\psi(x,y)$ such that a local minimum of the difference (13.2.3.3) is attained at singular points comprising Γ.

(C) For the one-parameter family of test functions

$$\psi(x,y) = \lambda u_1(x,y) + (1-\lambda)u_2(x,y), \qquad 0 \le \lambda \le 1, \tag{13.2.3.7}$$

the maximum

$$\max_{(x,y)\in D} \big[w(x,y) - \psi(x,y)\big] \tag{13.2.3.8}$$

is attained at $(x,y) \in \Gamma$.

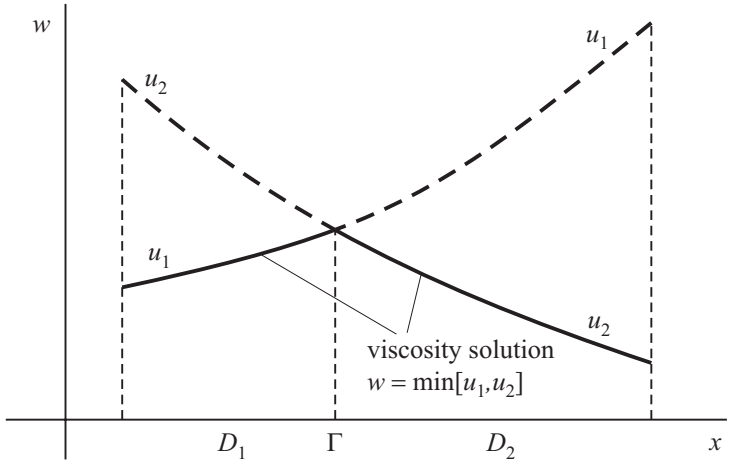


Figure 13.12. Graphic construction of a generalized viscosity solution by joining together smooth solutions.

Remark. For the generalized solution of the form $w = \min[u_1, u_2]$, one need not check the first inequality of (13.2.3.4); as far as the second inequality of (13.2.3.4) is concerned, it suffices to check this inequality against the one-parameter family of test functions defined by equation (13.2.3.7).

2°. The vector ∇g is directed from D_1 toward D_2. In this case, the generalized solution can be represented in the form $w = \max[u_1, u_2]$ and one need to check only the first inequality of (13.2.3.4) against the one-parameter family of test functions of equation (13.2.3.7).

13.2.3-5. Generalized classical method of characteristics.

Consider the Cauchy problem for the Hamilton–Jacobi equation of the form

$$\begin{aligned} &\frac{\partial w}{\partial x} + \mathcal{H}\left(x, y, \frac{\partial w}{\partial y}\right) = 0, \\ &w = \varphi(y) \quad \text{at} \quad x = L. \end{aligned} \tag{13.2.3.9}$$

It is assumed here that (i) the function $\mathcal{H}(x, y, q)$ is convex with respect to q for all $x \in (0, L]$, $y \in \mathbb{R}$; (ii) the function $\mathcal{H}(x, y, q)$ is continuously differentiable with respect to x, y, and q; and (iii) the second derivatives $\mathcal{H}_{xy}$ and $\mathcal{H}_{xq}$ exist.

By

$$y = Y(x, \xi), \quad w = W(x, \xi), \quad q = Q(x, \xi) \tag{13.2.3.10}$$

we denote a solution of the characteristic system (13.2.1.15) satisfying condition (13.2.2.9).

Let $\{\xi_n = \xi_n(x, y)\}$ be the set of functions obtained by solving the first equation of (13.2.3.10) for the parameter ξ. The subscript n indicates the number of such functions.

The classical method of characteristics can be used to construct a generalized viscosity solution with the help of the relation

$$w(x, y) = \max_{\xi \in \{\xi_n\}} W(x, \xi) \tag{13.2.3.11}$$

for all $x \in (0, L]$, $y \in \mathbb{R}$. To the value $n = 1$ there corresponds the classical smooth solution.

13.2.3-6. Examples of viscosity (nonsmooth) solutions.

Below we solve two problems that are encountered in the theory of differential games.

Example 1. Consider the terminal value problem for the Hamilton–Jacobi equation

$$\frac{\partial w}{\partial x}+\sqrt{1+\Big(\frac{\partial w}{\partial y}\Big)^2}=0 \tag{13.2.3.12}$$

with the initial condition

$$w=\tfrac{1}{2}y^2 \quad \text{at} \quad x=L. \tag{13.2.3.13}$$

The solution is sought in the domain $0\le x\le L$.

The characteristic system (13.2.1.15) for equation (13.2.3.12) with the Hamiltonian $\mathcal{H}(x,y,w,q)=\sqrt{1+q^2}$ has the form

$$y'_x=\frac{q}{\sqrt{1+q^2}}, \quad w'_x=-\frac{1}{\sqrt{1+q^2}}, \quad q'_x=0. \tag{13.2.3.14}$$

The initial conditions are obtained from (13.2.2.9) for $\varphi(\xi)=\frac{1}{2}\xi^2$:

$$y=\xi, \quad w=\tfrac{1}{2}\xi^2, \quad q=\xi \quad \text{at} \quad x=L. \tag{13.2.3.15}$$

Integrating equation (13.2.3.14) and applying conditions (13.2.3.15), we obtain a solution of the Cauchy problem (13.2.3.12), (13.2.3.13) of the form

$$y=\frac{\xi(x-L)}{\sqrt{1+\xi^2}}+\xi, \quad w=\frac{L-x}{\sqrt{1+\xi^2}}+\tfrac{1}{2}\xi^2. \tag{13.2.3.16}$$

Figure 13.13 displays the characteristics $y(x,\xi)$ in the xy-plane for $L=2$ and $\xi=0,\pm0.2,\pm0.4,\ldots,\pm1.0$. It is apparent that characteristics intersect. In this example, it is possible to construct a local classical solution of problem (13.2.3.12), (13.2.3.13). But this solution cannot be extended to the entire layer $0\le x\le L$, that is, there is no global classical solution. Pay attention to the fact that the Hamiltonian $\mathcal{H}=\sqrt{1+q^2}$ of equation (13.2.3.12) and the function determining the initial condition (13.2.3.13) are infinitely differentiable functions.

The viscosity solution of the Cauchy problem (13.2.3.12), (13.2.3.13) has the form

$$w(x,y)=\max_{q\in\mathbb{R}}\big[qy+(L-x)\sqrt{1+q^2}-\tfrac{1}{2}q^2\big], \tag{13.2.3.17}$$

where $0\le x\le L$ and y is any number. Level curves of this function are depicted in Fig. 13.14. The heavy line indicates the set of singular points, at which the solution is nondifferentiable.

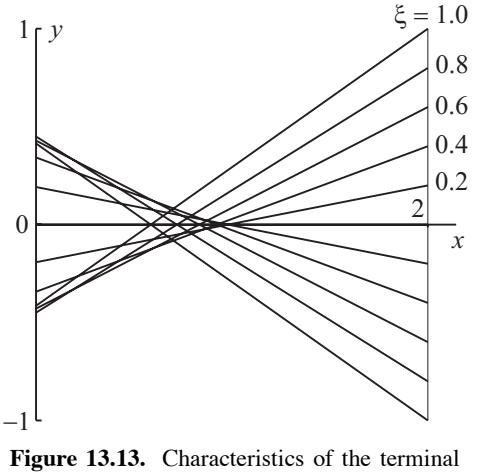


Figure 13.13. Characteristics of the terminal value problem (13.2.3.12)–(13.2.3.13) for the Hamilton–Jacobi equation.

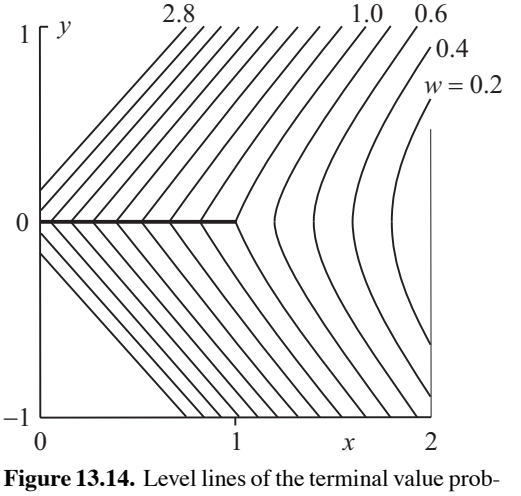


Figure 13.14. Level lines of the terminal value problem (13.2.3.12)–(13.2.3.13) for the Hamilton–Jacobi equation.

Example 2. Consider the terminal value problems for the more general Hamilton–Jacobi equation

$$\frac{\partial w}{\partial x} + \mathcal{H}\left(\frac{\partial w}{\partial y}\right) = 0 \qquad (0 \le x \le L) \tag{13.2.3.18}$$

with an arbitrary initial condition

$$w = \varphi(y) \quad \text{at} \quad x = L. \tag{13.2.3.19}$$

The following two statements hold:

1°. Let the Hamiltonian satisfy the Lipschitz condition

$$|\mathcal{H}(q_2) - \mathcal{H}(q_1)| \le \beta|q_2 - q_1| \qquad \text{for any} \quad q_1, q_2 \in \mathbb{R}, \tag{13.2.3.20}$$

and let the function $\varphi(y)$ be convex. Then the function

$$w(x, y) = \sup_{q\in\mathbb{R}}\big[qy + (L - x)\mathcal{H}(q) - \varphi^*(q)\big]$$

is the viscosity solution of problem (13.2.3.18), (13.2.3.19). The function φ^* is the conjugate of φ, i.e.,

$$\varphi^*(q) = \sup_{x\in\mathbb{R}}\big[qx - \varphi(x)\big].$$

2°. Let the Hamiltonian $\mathcal{H}$ be convex and satisfy the Lipschitz condition (13.2.3.20). Let the function $\varphi(y)$ be continuous. Then the function

$$w(x, y) = \sup_{t\in\mathbb{R}}\big[\varphi(y + (L - x)t) - (L - x)\mathcal{H}^*(t)\big]$$

is the viscosity solution of problem (13.2.3.18), (13.2.3.19). The function

$$\mathcal{H}^*(t) = \sup_{q\in\mathbb{R}}\big[qt - \mathcal{H}(q)\big]$$

is the conjugate of the Hamiltonian.

References for Chapter 13

Bardi, M. and Dolcetta, I. C., *Optimal Control and Viscosity Solutions of Hamilton–Jacobi–Bellman Equations*, Birkhäuser Verlag, Boston, 1998.

Barron, E. N. and Jensen, R., Generalized viscosity solutions for Hamilton–Jacobi equations with time-measurable Hamiltonians, *J. Different. Equations*, Vol. 68, No. 1, pp. 10–21, 1987.

Courant, R. and Hilbert, D., *Methods of Mathematical Physics, Vol. 2*, Wiley-Interscience, New York, 1989.

Crandall, M. G., Evans, L. C., and Lions, P.-L., Some properties of viscosity solutions of Hamilton–Jacobi equations, *Trans. Amer. Math. Soc.,* Vol. 283, No. 2, pp. 487–502, 1984.

Gelfand, I. M., Some problems of the theory of quasi-linear equations, *Uspekhi Matem. Nauk*, Vol. 14, No. 2, pp. 87–158, 1959 [*Amer. Math. Soc. Translation, Series 2*, pp. 295–381, 1963].

Hopf, E., Generalized solutions of nonlinear equations of first order, *J. Math. Mech.,* Vol. 14, pp. 951–973, 1965.

Jeffery, A., *Quasilinear Hyperbolic Systems and Waves*, Pitman, London, 1976.

Kamke, E., *Differentialgleichungen: Lösungsmethoden und Lösungen, II, Partielle Differentialgleichungen Erster Ordnung für eine gesuchte Funktion*, Akad. Verlagsgesellschaft Geest & Portig, Leipzig, 1965.

Kruzhkov, S. N., Generalized solutions of nonlinear first order equations with several variables [in Russian], *Mat. Sbornik*, Vol. 70, pp. 394–415, 1966.

Lax, P. D., *Hyperbolic systems of conservation laws and the mathematical theory of shock waves* [reprint from the classical paper of 1957], Society for Industrial & Applied Mathematics, Philadelphia, 1997.

Melikyan, A. A., *Generalized Characteristics of First Order PDEs: Applications in Optimal Control and Differential Games*, Birkhäuser Verlag, Boston, 1998.

Oleinik, O. A., Discontinuous solutions of nonlinear differential equations, *Uspekhi Matem. Nauk*, Vol. 12, No. 3, pp. 3–73, 1957 [*Amer. Math. Soc. Translation, Series 2*, Vol. 26, pp. 95–172, 1963].

Petrovskii, I. G., *Lectures on the Theory of Ordinary Differential Equations* [in Russian], Nauka Publishers, Moscow, 1970.

Polyanin, A. D., Zaitsev, V. F., and Moussiaux, A., *Handbook of First Order Partial Differential Equations*, Taylor & Francis, London, 2002.

Rhee, H., Aris, R., and Amundson, N. R., *First Order Partial Differential Equations, Vols. 1 and 2*, Prentice Hall, Englewood Cliffs, New Jersey, 1986 and 1989.

Rozhdestvenskii, B. L. and Yanenko, N. N., *Systems of Quasilinear Equations and Their Applications to Gas Dynamics*, American Mathematical Society, Providence, Rhode Island, 1983.

Smoller, J., *Shock Waves and Reaction-Diffusion Equations*, Springer-Verlag, New York, 1994.

Subbotin, A. I., *Generalized Solutions of First Order PDEs: the Dynamical Optimization Perspective*, Birkhäuser Verlag, Boston, 1995.

Whitham, G. B., *Linear and Nonlinear Waves*, Wiley, New York, 1974.

Chapter 14
Linear Partial Differential Equations

14.1. Classification of Second-Order Partial Differential Equations

14.1.1. Equations with Two Independent Variables

14.1.1-1. Examples of equations encountered in applications.

Three basic types of partial differential equations are distinguished—*parabolic*, *hyperbolic*, and *elliptic*. The solutions of the equations pertaining to each of the types have their own characteristic qualitative differences.

The simplest example of a *parabolic equation* is the *heat equation*

$$\frac{\partial w}{\partial t} - \frac{\partial^2 w}{\partial x^2} = 0, \tag{14.1.1.1}$$

where the variables t and x play the role of time and the spatial coordinate, respectively. Note that equation (14.1.1.1) contains only one highest derivative term. Frequently encountered particular solutions of equation (14.1.1.1) can be found in Paragraph T8.1.1-1.

The simplest example of a *hyperbolic equation* is the *wave equation*

$$\frac{\partial^2 w}{\partial t^2} - \frac{\partial^2 w}{\partial x^2} = 0, \tag{14.1.1.2}$$

where the variables t and x play the role of time and the spatial coordinate, respectively. Note that the highest derivative terms in equation (14.1.1.2) differ in sign.

The simplest example of an *elliptic equation* is the *Laplace equation*

$$\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} = 0, \tag{14.1.1.3}$$

where x and y play the role of the spatial coordinates. Note that the highest derivative terms in equation (14.1.1.3) have like signs. Frequently encountered particular solutions of equation (14.1.1.3) can be found in Paragraph T8.3.1-1.

Any linear partial differential equation of the second-order with two independent variables can be reduced, by appropriate manipulations, to a simpler equation that has one of the three highest derivative combinations specified above in examples (14.1.1.1), (14.1.1.2), and (14.1.1.3).

14.1.1-2. Types of equations. Characteristic equations.

Consider a second-order partial differential equation with two independent variables that has the general form

$$a(x,y)\frac{\partial^2 w}{\partial x^2} + 2b(x,y)\frac{\partial^2 w}{\partial x \partial y} + c(x,y)\frac{\partial^2 w}{\partial y^2} = F\left(x, y, w, \frac{\partial w}{\partial x}, \frac{\partial w}{\partial y}\right), \tag{14.1.1.4}$$

where a, b, c are some functions of x and y that have continuous derivatives up to the second-order inclusive.*

Given a point (x, y), equation (14.1.1.4) is said to be

$$\begin{aligned} &\textit{parabolic} && \text{if } b^2 - ac = 0,\\ &\textit{hyperbolic} && \text{if } b^2 - ac > 0,\\ &\textit{elliptic} && \text{if } b^2 - ac < 0 \end{aligned}$$

at this point.

In order to reduce equation (14.1.1.4) to a canonical form, one should first write out the characteristic equation

$$a\,(dy)^2 - 2b\,dx\,dy + c\,(dx)^2 = 0,$$

which splits into two equations

$$a\,dy - \left(b + \sqrt{b^2 - ac}\right) dx = 0 \tag{14.1.1.5}$$

and

$$a\,dy - \left(b - \sqrt{b^2 - ac}\right) dx = 0, \tag{14.1.1.6}$$

and then find their general integrals.

Remark. The characteristic equations (14.1.1.5)–(14.1.1.5) may be used if $a \not\equiv 0$. If $a \equiv 0$, the simpler equations

$$\begin{aligned} dx &= 0,\\ 2b\,dy - c\,dx &= 0 \end{aligned}$$

should be used; the first equation has the obvious general solution $x = C$.

14.1.1-3. Canonical form of parabolic equations (case $b^2 - ac = 0$).

In this case, equations (14.1.1.5) and (14.1.1.6) coincide and have a common general integral,

$$\varphi(x, y) = C.$$

By passing from x, y to new independent variables ξ, η in accordance with the relations

$$\xi = \varphi(x, y), \qquad \eta = \eta(x, y),$$

where $\eta = \eta(x, y)$ is any twice differentiable function that satisfies the condition of nondegeneracy of the Jacobian $\frac{D(\xi,\eta)}{D(x,y)}$ in the given domain, we reduce equation (14.1.1.4) to the canonical form

$$\frac{\partial^2 w}{\partial \eta^2} = F_1\left(\xi, \eta, w, \frac{\partial w}{\partial \xi}, \frac{\partial w}{\partial \eta}\right). \tag{14.1.1.7}$$

As η, one can take $\eta = x$ or $\eta = y$.

It is apparent that the transformed equation (14.1.1.7) has only one highest-derivative term, just as the heat equation (14.1.1.1).

Remark. In the degenerate case where the function F_1 does not depend on the derivative $\partial_\xi w$, equation (14.1.1.7) is an ordinary differential equation for the variable η, in which ξ serves as a parameter.

* The right-hand side of equation (14.1.1.4) may be nonlinear. The classification and the procedure of reducing such equations to a canonical form are only determined by the left-hand side of the equation.

14.1.1-4. Canonical forms of hyperbolic equations (case $b^2 - ac > 0$).

The general integrals

$$\varphi(x, y) = C_1, \qquad \psi(x, y) = C_2$$

of equations (14.1.1.5) and (14.1.1.6) are real and different. These integrals determine two different families of real characteristics.

By passing from x, y to new independent variables ξ, η in accordance with the relations

$$\xi = \varphi(x, y), \qquad \eta = \psi(x, y),$$

we reduce equation (14.1.1.4) to

$$\frac{\partial^2 w}{\partial \xi \partial \eta} = F_2\left(\xi, \eta, w, \frac{\partial w}{\partial \xi}, \frac{\partial w}{\partial \eta}\right).$$

This is the so-called *first canonical form of a hyperbolic equation.*

The transformation

$$\xi = t + z, \qquad \eta = t - z$$

brings the above equation to another canonical form,

$$\frac{\partial^2 w}{\partial t^2} - \frac{\partial^2 w}{\partial z^2} = F_3\left(t, z, w, \frac{\partial w}{\partial t}, \frac{\partial w}{\partial z}\right),$$

where $F_3 = 4F_2$. This is the so-called *second canonical form of a hyperbolic equation.* Apart from notation, the left-hand side of the last equation coincides with that of the wave equation (14.1.1.2).

In some cases, reduction of an equation to a canonical form allows finding its general solution.

Example. The equation

$$kx\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial x \partial y} = 0$$

is a special case of equation (14.1.1.4) with $a = kx$, $b = \frac{1}{2}$, $c = 0$, and $F = 0$. The characteristic equations

$$kx\,dy - dx = 0,$$
$$dy = 0$$

have the general integrals $ky - \ln|x| = C_1$ and $y = C_2$. Switching to the new independent variables

$$\xi = ky - \ln|x|, \qquad \eta = y$$

reduces the original equation to the canonical form

$$\frac{\partial^2 w}{\partial \xi \partial \eta} = k\frac{\partial w}{\partial \xi}.$$

Integrating with respect to ξ yields the linear first-order equation

$$\frac{\partial w}{\partial \eta} = kw + f(\eta)$$

where $f(\eta)$ is an arbitrary function. Its general solution is expressed as

$$w = e^{k\eta} g(\xi) + e^{k\eta} \int e^{-k\eta} f(\eta)\, d\eta,$$

where $g(\xi)$ is an arbitrary function.

14.1.1-5. Canonical form of elliptic equations (case $b^2 - ac < 0$).

In this case the general integrals of equations (14.1.1.5) and (14.1.1.6) are complex conjugate; these determine two families of complex characteristics.

Let the general integral of equation (14.1.1.5) have the form

$$\varphi(x, y) + i\psi(x, y) = C, \qquad i^2 = -1,$$

where $\varphi(x, y)$ and $\psi(x, y)$ are real-valued functions.

By passing from x, y to new independent variables ξ, η in accordance with the relations

$$\xi = \varphi(x, y), \qquad \eta = \psi(x, y),$$

we reduce equation (14.1.1.4) to the canonical form

$$\frac{\partial^2 w}{\partial \xi^2} + \frac{\partial^2 w}{\partial \eta^2} = F_4\left(\xi, \eta, w, \frac{\partial w}{\partial \xi}, \frac{\partial w}{\partial \eta}\right).$$

Apart from notation, the left-hand side of the last equation coincides with that of the Laplace equation (14.1.1.3).

14.1.1-6. Linear constant-coefficient partial differential equations.

1°. When reduced to a canonical form, linear homogeneous constant-coefficient partial differential equations

$$a\frac{\partial^2 w}{\partial x^2} + 2b\frac{\partial^2 w}{\partial x \partial y} + c\frac{\partial^2 w}{\partial y^2} + p\frac{\partial w}{\partial x} + q\frac{\partial w}{\partial y} + sw = 0 \tag{14.1.1.8}$$

admit further simplifications. In general, the substitution

$$w(x, y) = \exp(\beta_1 \xi + \beta_2 \eta) u(\xi, \eta) \tag{14.1.1.9}$$

can be used. Here, ξ and η are new variables used to reduce equation (14.1.1.8) to a canonical form (see Paragraphs 14.1.1-3 to 14.1.1-5); the coefficients β_1 and β_2 in (14.1.1.9) are chosen so that there is only one first derivative remaining in a parabolic equation or both first derivatives vanish in a hyperbolic or an elliptic equation. For final results, see Table 14.1.

2°. The coefficients k and k_1 in the reduced hyperbolic and elliptic equations (see the third, fourth, and fifth rows in Table 14.1) are expressed as

$$k = \frac{2bpq - aq^2 - cp^2}{16a(b^2 - ac)^2} - \frac{s}{4a(b^2 - ac)}, \qquad k_1 = \frac{s}{4b^2} + \frac{cp^2 - 2bpq}{16b^4}. \tag{14.1.1.10}$$

If the coefficients in equation (14.1.1.8) satisfy the relation

$$2bpq - aq^2 - cp^2 - 4s(b^2 - ac) = 0,$$

then $k = 0$; in this case with $a \neq 0$, the general solution of the corresponding hyperbolic equation has the form

$$w(x, y) = \exp(\beta_1 \xi + \beta_2 \eta)\big[f(\xi) + g(\eta)\big], \qquad D = b^2 - ac > 0,$$
$$\xi = ay - \big(b + \sqrt{D}\,\big)x, \qquad \eta = ay - \big(b - \sqrt{D}\,\big)x,$$
$$\beta_1 = \frac{aq - bp}{4aD} + \frac{p}{4a\sqrt{D}}, \qquad \beta_2 = \frac{aq - bp}{4aD} - \frac{p}{4a\sqrt{D}},$$

where $f(\xi)$ and $g(\xi)$ are arbitrary functions.

TABLE 14.1
Reduction of linear homogeneous constant-coefficient partial differential equations (14.1.1.8) using transformation (14.1.1.9); the constants k and k_1 are given by formulas (14.1.1.10)

Type of equation, conditions on coefficients	Variables ξ and η in transformation (14.1.1.9)	Coefficients β_1 and β_2 in transformation (14.1.1.9)	Reduced equation
Parabolic equation, $a=b=0,\ c\neq 0,\ p\neq 0$	$\xi=-\dfrac{c}{p}x,\ \eta=y$	$\beta_1=\dfrac{4cs-q^2}{4c^2},\ \beta_2=-\dfrac{q}{2c}$	$u_\xi-u_{\eta\eta}=0$
Parabolic equation, $b^2-ac=0$ $(aq-bp\neq 0,\ \lvert a\rvert+\lvert b\rvert\neq 0)$	$\xi=\dfrac{a(ay-bx)}{bp-aq},\ \eta=x$	$\beta_1=\dfrac{4as-p^2}{4a^2},\ \beta_2=-\dfrac{p}{2a}$	$u_\xi-u_{\eta\eta}=0$
Hyperbolic equation, $a\neq 0,\ D=b^2-ac>0$	$\xi=ay-\left(b+\sqrt{D}\right)x,$ $\eta=ay-\left(b-\sqrt{D}\right)x$	$\beta_{1,2}=\dfrac{aq-bp}{4aD}\pm\dfrac{p}{4a\sqrt{D}}$	$u_{\xi\eta}+ku=0$
Hyperbolic equation, $a=0,\ b\neq 0$	$\xi=x,$ $\eta=2by-cx$	$\beta_1=\dfrac{cp-2bq}{4b^2},\ \beta_2=-\dfrac{p}{4b^2}$	$u_{\xi\eta}+k_1u=0$
Elliptic equation, $D=b^2-ac<0$	$\xi=ay-bx,$ $\eta=\sqrt{\lvert D\rvert}\,x$	$\beta_1=\dfrac{aq-bp}{2aD},\ \beta_2=-\dfrac{p}{2a\sqrt{\lvert D\rvert}}$	$u_{\xi\xi}+u_{\eta\eta}+4ku=0$
Ordinary differential equation, $b^2-ac=0,\ aq-bp=0$	$\xi=ay-bx,$ $\eta=x$	$\beta_1=\beta_2=0$	$aw_{\eta\eta}+pw_\eta+sw=0$

3°. In the degenerate case $b^2-ac=0$, $aq-bp=0$ (where the original equation is reduced to an ordinary differential equation; see the last row in Table 14.1), the general solution of equation (14.1.1.8) is expressed as

$$w=\exp\left(-\frac{px}{2a}\right)\left[f(ay-bx)\exp\left(\frac{x\sqrt{\lambda}}{2a}\right)+g(ay-bx)\exp\left(-\frac{x\sqrt{\lambda}}{2a}\right)\right]\quad\text{if}\quad\lambda=p^2-4as>0,$$

$$w=\exp\left(-\frac{px}{2a}\right)\left[f(ay-bx)\sin\left(\frac{x\sqrt{|\lambda|}}{2a}\right)+g(ay-bx)\cos\left(\frac{x\sqrt{|\lambda|}}{2a}\right)\right]\quad\text{if}\quad\lambda=p^2-4as<0,$$

$$w=\exp\left(-\frac{px}{2a}\right)\left[f(ay-bx)+xg(ay-bx)\right]\quad\text{if}\quad 4as-p^2=0,$$

where $f(z)$ and $g(z)$ are arbitrary functions.

14.1.2. Equations with Many Independent Variables

Let us consider a second-order partial differential equation with n independent variables $x_1,\ldots,x_n$ that has the form

$$\sum_{i,j=1}^{n}a_{ij}(\mathbf{x})\frac{\partial^2 w}{\partial x_i\partial x_j}=F\left(\mathbf{x},w,\frac{\partial w}{\partial x_1},\ldots,\frac{\partial w}{\partial x_n}\right),\qquad(14.1.2.1)$$

where the a_{ij} are some functions that have continuous derivatives with respect to all variables to the second-order inclusive, and $\mathbf{x}=\{x_1,\ldots,x_n\}$. [The right-hand side of equation (14.1.2.1) may be nonlinear. The left-hand side only is required for the classification of this equation.]

At a point $\mathbf{x}=\mathbf{x}_0$, the following quadratic form is assigned to equation (14.1.2.1):

$$Q=\sum_{i,j=1}^{n}a_{ij}(\mathbf{x}_0)\xi_i\xi_j.\qquad(14.1.2.2)$$

TABLE 14.2
Classification of equations with many independent variables

Type of equation (14.1.2.1) at a point $\mathbf{x} = \mathbf{x}_0$	Coefficients of the canonical form (14.1.2.4)
Parabolic (in the broad sense)	At least one coefficient of the c_i is zero
Hyperbolic (in the broad sense)	All c_i are nonzero and some c_i differ in sign
Elliptic	All c_i are nonzero and have like signs

By an appropriate linear nondegenerate transformation

$$\xi_i = \sum_{k=1}^{n} \beta_{ik}\eta_k \qquad (i = 1, \ldots, n) \tag{14.1.2.3}$$

the quadratic form (14.1.2.2) can be reduced to the canonical form

$$Q = \sum_{i=1}^{n} c_i \eta_i^2, \tag{14.1.2.4}$$

where the coefficients c_i assume the values 1, –1, and 0. The number of negative and zero coefficients in (14.1.2.4) does not depend on the way in which the quadratic form is reduced to the canonical form.

Table 14.2 presents the basic criteria according to which the equations with many independent variables are classified.

Suppose all coefficients of the highest derivatives in (14.1.2.1) are constant, $a_{ij} = \text{const}$. By introducing the new independent variables $y_1, \ldots, y_n$ in accordance with the formulas $y_i = \sum_{k=1}^{n} \beta_{ik} x_k$, where the β_{ik} are the coefficients of the linear transformation (14.1.2.3), we reduce equation (14.1.2.1) to the canonical form

$$\sum_{i=1}^{n} c_i \frac{\partial^2 w}{\partial y_i^2} = F_1\left(\mathbf{y}, w, \frac{\partial w}{\partial y_1}, \ldots, \frac{\partial w}{\partial y_n}\right). \tag{14.1.2.5}$$

Here, the coefficients c_i are the same as in the quadratic form (14.1.2.4), and $\mathbf{y} = \{y_1, \ldots, y_n\}$.

Remark 1. Among the parabolic equations, it is conventional to distinguish the parabolic equations in the narrow sense, i.e., the equations for which only one of the coefficients, c_k, is zero, while the other c_i is the same, and in this case the right-hand side of equation (14.1.2.5) must contain the first-order partial derivative with respect to y_k.

Remark 2. In turn, the hyperbolic equations are divided into normal hyperbolic equations — for which all c_i but one have like signs — and ultrahyperbolic equations — for which there are two or more positive c_i and two or more negative c_i.

Specific equations of parabolic, elliptic, and hyperbolic types will be discussed further in Section 14.2.

14.2. Basic Problems of Mathematical Physics

14.2.1. Initial and Boundary Conditions. Cauchy Problem. Boundary Value Problems

Every equation of mathematical physics governs infinitely many qualitatively similar phenomena or processes. This follows from the fact that differential equations have infinitely

many particular solutions. The specific solution that describes the physical phenomenon under study is separated from the set of particular solutions of the given differential equation by means of the initial and boundary conditions.

Throughout this section, we consider linear equations in the n-dimensional Euclidean space $\mathbb{R}^n$ or in an open domain $V \in \mathbb{R}^n$ (exclusive of the boundary) with a sufficiently smooth boundary $S = \partial V$.

14.2.1-1. Parabolic equations. Initial and boundary conditions.

In general, a linear second-order partial differential equation of the parabolic type with n independent variables can be written as

$$\frac{\partial w}{\partial t} - L_{\mathbf{x},t}[w] = \Phi(\mathbf{x}, t), \tag{14.2.1.1}$$

where

$$L_{\mathbf{x},t}[w] \equiv \sum_{i,j=1}^{n} a_{ij}(\mathbf{x}, t)\frac{\partial^2 w}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(\mathbf{x}, t)\frac{\partial w}{\partial x_i} + c(\mathbf{x}, t)w, \tag{14.2.1.2}$$

$$\mathbf{x} = \{x_1, \ldots, x_n\}, \quad \sum_{i,j=1}^{n} a_{ij}(\mathbf{x}, t)\xi_i\xi_j \geq \sigma \sum_{i=1}^{n} \xi_i^2, \quad \sigma > 0.$$

Parabolic equations govern unsteady thermal, diffusion, and other phenomena dependent on time t.

Equation (14.2.1.1) is called homogeneous if $\Phi(\mathbf{x}, t) \equiv 0$.

Cauchy problem ($t \geq 0$, $\mathbf{x} \in \mathbb{R}^n$). Find a function w that satisfies equation (14.2.1.1) for $t > 0$ and the initial condition

$$w = f(\mathbf{x}) \quad \text{at} \quad t = 0. \tag{14.2.1.3}$$

*Boundary value problem** ($t \geq 0$, $\mathbf{x} \in V$). Find a function w that satisfies equation (14.2.1.1) for $t > 0$, the initial condition (14.2.1.3), and the boundary condition

$$\Gamma_{\mathbf{x},t}[w] = g(\mathbf{x}, t) \quad \text{at} \quad \mathbf{x} \in S \qquad (t > 0). \tag{14.2.1.4}$$

In general, $\Gamma_{\mathbf{x},t}$ is a first-order linear differential operator in the space variables $\mathbf{x}$ with coefficient dependent on $\mathbf{x}$ and t. The basic types of boundary conditions are described in Subsection 14.2.2.

The initial condition (14.2.1.3) is called homogeneous if $f(\mathbf{x}) \equiv 0$. The boundary condition (14.2.1.4) is called homogeneous if $g(\mathbf{x}, t) \equiv 0$.

14.2.1-2. Hyperbolic equations. Initial and boundary conditions.

Consider a second-order linear partial differential equation of the hyperbolic type with n independent variables of the general form

$$\frac{\partial^2 w}{\partial t^2} + \varphi(\mathbf{x}, t)\frac{\partial w}{\partial t} - L_{\mathbf{x},t}[w] = \Phi(\mathbf{x}, t), \tag{14.2.1.5}$$

* *Boundary value problems* for parabolic and hyperbolic equations are sometimes called *mixed* or *initial-boundary value problems*.

where the linear differential operator $L_{\mathbf{x},t}$ is defined by (14.2.1.2). Hyperbolic equations govern unsteady wave processes, which depend on time t.

Equation (14.2.1.5) is said to be homogeneous if $\Phi(\mathbf{x}, t) \equiv 0$.

Cauchy problem ($t \geq 0$, $\mathbf{x} \in \mathbb{R}^n$). Find a function w that satisfies equation (14.2.1.5) for $t > 0$ and the initial conditions

$$\begin{aligned} w &= f_0(\mathbf{x}) \quad \text{at} \quad t = 0, \\ \partial_t w &= f_1(\mathbf{x}) \quad \text{at} \quad t = 0. \end{aligned} \tag{14.2.1.6}$$

Boundary value problem ($t \geq 0$, $\mathbf{x} \in V$). Find a function w that satisfies equation (14.2.1.5) for $t > 0$, the initial conditions (14.2.1.6), and boundary condition (14.2.1.4).

The initial conditions (14.2.1.6) are called homogeneous if $f_0(\mathbf{x}) \equiv 0$ and $f_1(\mathbf{x}) \equiv 0$.

Generalized Cauchy problem. In the generalized Cauchy problem for a hyperbolic equation with two independent variables, values of the unknown function and its first derivatives are prescribed on a curve in the (x, t) plane. Alternatively, values of the unknown function and its derivative along the normal to this curve may be prescribed. For more details, see Paragraph 14.8.4-4.

Goursat problem. On the characteristics of a hyperbolic equation with two independent variables, the values of the unknown function w are prescribed; for details, see Paragraph 14.8.4-5).

14.2.1-3. Elliptic equations. Boundary conditions.

In general, a second-order linear partial differential equation of elliptic type with n independent variables can be written as

$$-L_{\mathbf{x}}[w] = \Phi(\mathbf{x}), \tag{14.2.1.7}$$

where

$$L_{\mathbf{x}}[w] \equiv \sum_{i,j=1}^{n} a_{ij}(\mathbf{x}) \frac{\partial^2 w}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(\mathbf{x}) \frac{\partial w}{\partial x_i} + c(\mathbf{x}) w, \tag{14.2.1.8}$$

$$\sum_{i,j=1}^{n} a_{ij}(\mathbf{x}) \xi_i \xi_j \geq \sigma \sum_{i=1}^{n} \xi_i^2, \quad \sigma > 0.$$

Elliptic equations govern steady-state thermal, diffusion, and other phenomena independent of time t.

Equation (14.2.1.7) is said to be homogeneous if $\Phi(\mathbf{x}) \equiv 0$.

Boundary value problem. Find a function w that satisfies equation (14.2.1.7) and the boundary condition

$$\Gamma_{\mathbf{x}}[w] = g(\mathbf{x}) \quad \text{at} \quad \mathbf{x} \in S. \tag{14.2.1.9}$$

In general, $\Gamma_{\mathbf{x}}$ is a first-order linear differential operator in the space variables $\mathbf{x}$. The basic types of boundary conditions are described below in Subsection 14.2.2.

The boundary condition (14.2.1.9) is called homogeneous if $g(\mathbf{x}) \equiv 0$. The boundary value problem (14.2.1.7)–(14.2.1.9) is said to be homogeneous if $\Phi \equiv 0$ and $g \equiv 0$.

TABLE 14.3
Boundary conditions for various boundary value problems specified by parabolic and hyperbolic equations in two independent variables ($x_1 \le x \le x_2$)

Type of problem	Boundary condition at $x = x_1$	Boundary condition at $x = x_2$
First boundary value problem	$w = g_1(t)$	$w = g_2(t)$
Second boundary value problem	$\partial_x w = g_1(t)$	$\partial_x w = g_2(t)$
Third boundary value problem	$\partial_x w + \beta_1 w = g_1(t)$ $(\beta_1 < 0)$	$\partial_x w + \beta_2 w = g_2(t)$ $(\beta_2 > 0)$
Mixed boundary value problem	$w = g_1(t)$	$\partial_x w = g_2(t)$
Mixed boundary value problem	$\partial_x w = g_1(t)$	$w = g_2(t)$

14.2.2. First, Second, Third, and Mixed Boundary Value Problems

For any (parabolic, hyperbolic, and elliptic) second-order partial differential equations, it is conventional to distinguish four basic types of boundary value problems, depending on the form of the boundary conditions (14.2.1.4) [see also the analogous condition (14.2.1.9)]. For simplicity, here we confine ourselves to the case where the coefficients a_{ij} of equations (14.2.1.1) and (14.2.1.5) have the special form

$$a_{ij}(\mathbf{x}, t) = a(\mathbf{x}, t)\delta_{ij}, \quad \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \ne j. \end{cases}$$

This situation is rather frequent in applications; such coefficients are used to describe various phenomena (processes) in isotropic media.

The function $w(\mathbf{x}, t)$ takes prescribed values at the boundary S of the domain:

$$w(\mathbf{x}, t) = g_1(\mathbf{x}, t) \quad \text{for} \quad \mathbf{x} \in S. \tag{14.2.2.1}$$

Second boundary value problem. The derivative along the (outward) normal is prescribed at the boundary S of the domain:

$$\frac{\partial w}{\partial N} = g_2(\mathbf{x}, t) \quad \text{for} \quad \mathbf{x} \in S. \tag{14.2.2.2}$$

In heat transfer problems, where w is temperature, the left-hand side of the boundary condition (14.2.2.2) is proportional to the heat flux per unit area of the surface S.

Third boundary value problem. A linear relationship between the unknown function and its normal derivative is prescribed at the boundary S of the domain:

$$\frac{\partial w}{\partial N} + k(\mathbf{x}, t)w = g_3(\mathbf{x}, t) \quad \text{for} \quad \mathbf{x} \in S. \tag{14.2.2.3}$$

Usually, it is assumed that $k(\mathbf{x}, t) = \text{const}$. In mass transfer problems, where w is concentration, the boundary condition (14.2.2.3) with $g_3 \equiv 0$ describes a surface chemical reaction of the first order.

Mixed boundary value problems. Conditions of various types, listed above, are set at different portions of the boundary S.

If $g_1 \equiv 0$, $g_2 \equiv 0$, or $g_3 \equiv 0$, the respective boundary conditions (14.2.2.1), (14.2.2.2), (14.2.2.3) are said to be homogeneous.

Boundary conditions for various boundary value problems for parabolic and hyperbolic equations in two independent variables x and t are displayed in Table 14.3. The equation coefficients are assumed to be continuous, with the coefficients of the highest derivatives being nonzero in the range $x_1 \le x \le x_2$ considered.

Remark. For elliptic equations, the first boundary value problem is often called the *Dirichlet problem*, and the second boundary value problem is called the *Neumann problem*.

14.3. Properties and Exact Solutions of Linear Equations

14.3.1. Homogeneous Linear Equations and Their Particular Solutions

14.3.1-1. Preliminary remarks.

For brevity, in this paragraph a homogeneous linear partial differential equation will be written as

$$\mathcal{L}[w] = 0. \tag{14.3.1.1}$$

For second-order linear parabolic and hyperbolic equations, the linear differential operator $\mathcal{L}[w]$ is defined by the left-hand side of equations (14.2.1.1) and (14.2.1.5), respectively. It is assumed that equation (14.3.1.1) is an arbitrary homogeneous linear partial differential equation of any order in the variables $t, x_1, \ldots, x_n$ with sufficiently smooth coefficients.

A linear operator $\mathcal{L}$ possesses the properties

$$\mathcal{L}[w_1 + w_2] = \mathcal{L}[w_1] + \mathcal{L}[w_2],$$
$$\mathcal{L}[Aw] = A\mathcal{L}[w], \qquad A = \text{const}.$$

An arbitrary homogeneous linear equation (14.3.1.1) has a trivial solution, $w \equiv 0$.

A function w is called a *classical solution* of equation (14.3.1.1) if w, when substituted into (14.3.1.1), turns the equation into an identity and if all partial derivatives of w that occur in (14.3.1.1) are continuous; the notion of a classical solution is directly linked to the range of the independent variables. In what follows, we usually write "solution" instead of "classical solution" for brevity.

14.3.1-2. Usage of particular solutions for the construction of other solutions.

Below are some properties of particular solutions of homogeneous linear equations.

1°. Let $w_1 = w_1(\mathbf{x}, t)$, $w_2 = w_2(\mathbf{x}, t)$, $\ldots$, $w_k = w_k(\mathbf{x}, t)$ be any particular solutions of the homogeneous equation (14.3.1.1). Then the linear combination

$$w = A_1 w_1 + A_2 w_2 + \cdots + A_k w_k \tag{14.3.1.2}$$

with arbitrary constants $A_1, A_2, \ldots, A_k$ is also a solution of equation (14.3.1.1); in physics, this property is known as the *principle of linear superposition*.

Suppose $\{w_k\}$ is an infinite sequence of solutions of equation (14.3.1.1). Then the series $\sum_{k=1}^{\infty} w_k$, irrespective of its convergence, is called a *formal solution* of (14.3.1.1). If the solutions w_k are classical, the series is uniformly convergent, and the sum of the series has all the necessary particular derivatives, then the sum of the series is a classical solution of equation (14.3.1.1).

2°. Let the coefficients of the linear differential operator $\mathcal{L}$ be independent of time t. If equation (14.3.1.1) has a particular solution $\widetilde{w} = \widetilde{w}(\mathbf{x}, t)$, then the partial derivatives of $\widetilde{w}$ with respect to time,*

$$\frac{\partial \widetilde{w}}{\partial t}, \quad \frac{\partial^2 \widetilde{w}}{\partial t^2}, \quad \ldots, \quad \frac{\partial^k \widetilde{w}}{\partial t^k}, \quad \ldots,$$

are also solutions of equation (14.3.1.1).

* Here and in what follows, it is assumed that the particular solution $\widetilde{w}$ is differentiable sufficiently many times with respect to t and $x_1, \ldots, x_n$ (or the parameters).

3°. Let the coefficients of the linear differential operator $\mathcal{L}$ be independent of the space variables $x_1, \ldots, x_n$. If equation (14.3.1.1) has a particular solution $\widetilde{w} = \widetilde{w}(\mathbf{x}, t)$, then the partial derivatives of $\widetilde{w}$ with respect to the space coordinates

$$\frac{\partial \widetilde{w}}{\partial x_1}, \quad \frac{\partial \widetilde{w}}{\partial x_2}, \quad \frac{\partial \widetilde{w}}{\partial x_3}, \quad \ldots, \quad \frac{\partial^2 \widetilde{w}}{\partial x_1^2}, \quad \frac{\partial^2 \widetilde{w}}{\partial x_1 \partial x_2}, \quad \ldots, \quad \frac{\partial^{k+m} \widetilde{w}}{\partial x_2^k \partial x_3^m}, \quad \ldots$$

are also solutions of equation (14.3.1.1).

If the coefficients of $\mathcal{L}$ are independent of only one space coordinate, say x_1, and equation (14.3.1.1) has a particular solution $\widetilde{w} = \widetilde{w}(\mathbf{x}, t)$, then the partial derivatives

$$\frac{\partial \widetilde{w}}{\partial x_1}, \quad \frac{\partial^2 \widetilde{w}}{\partial x_1^2}, \quad \ldots, \quad \frac{\partial^k \widetilde{w}}{\partial x_1^k}, \quad \ldots$$

are also solutions of equation (14.3.1.1).

4°. Let the coefficients of the linear differential operator $\mathcal{L}$ be constant and let equation (14.3.1.1) have a particular solution $\widetilde{w} = \widetilde{w}(\mathbf{x}, t)$. Then any particular derivatives of $\widetilde{w}$ with respect to time and the space coordinates (inclusive mixed derivatives)

$$\frac{\partial \widetilde{w}}{\partial t}, \quad \frac{\partial \widetilde{w}}{\partial x_1}, \quad \ldots, \quad \frac{\partial^2 \widetilde{w}}{\partial x_2^2}, \quad \frac{\partial^2 \widetilde{w}}{\partial t \partial x_1}, \quad \ldots, \quad \frac{\partial^k \widetilde{w}}{\partial x_3^k}, \quad \ldots$$

are solutions of equation (14.3.1.1).

5°. Suppose equation (14.3.1.1) has a particular solution dependent on a parameter μ, $\widetilde{w} = \widetilde{w}(\mathbf{x}, t; \mu)$, and the coefficients of the linear differential operator $\mathcal{L}$ are independent of μ (but can depend on time and the space coordinates). Then, by differentiating $\widetilde{w}$ with respect to μ, one obtains other solutions of equation (14.3.1.1),

$$\frac{\partial \widetilde{w}}{\partial \mu}, \quad \frac{\partial^2 \widetilde{w}}{\partial \mu^2}, \quad \ldots, \quad \frac{\partial^k \widetilde{w}}{\partial \mu^k}, \quad \ldots$$

Example 1. The linear equation

$$\frac{\partial w}{\partial t} = a \frac{\partial^2 w}{\partial x^2} + bw$$

has a particular solution

$$w(x, t) = \exp[\mu x + (a\mu^2 + b)t],$$

where μ is an arbitrary constant. Differentiating this equation with respect to μ yields another solution

$$w(x, t) = (x + 2a\mu t) \exp[\mu x + (a\mu^2 + b)t].$$

Let some constants $\mu_1, \ldots, \mu_k$ belong to the range of the parameter μ. Then the sum

$$w = A_1 \widetilde{w}(\mathbf{x}, t; \mu_1) + \cdots + A_k \widetilde{w}(\mathbf{x}, t; \mu_k), \tag{14.3.1.3}$$

where $A_1, \ldots, A_k$ are arbitrary constants, is also a solution of the homogeneous linear equation (14.3.1.1). The number of terms in sum (14.3.1.3) can be both finite and infinite.

6°. Another effective way of constructing solutions involves the following. The particular solution $\widetilde{w}(\mathbf{x}, t; \mu)$, which depends on the parameter μ (as before, it is assumed that the coefficients of the linear differential operator $\mathcal{L}$ are independent of μ), is first multiplied by an arbitrary function $\varphi(\mu)$. Then the resulting expression is integrated with respect to μ over some interval $[\alpha, \beta]$. Thus, one obtains a new function,

$$\int_\alpha^\beta \widetilde{w}(\mathbf{x}, t; \mu) \varphi(\mu) \, d\mu,$$

which is also a solution of the original homogeneous linear equation.

The properties listed in Items 1°–6° enable one to use known particular solutions to construct other particular solutions of homogeneous linear equations of mathematical physics.

TABLE 14.4
Homogeneous linear partial differential equations that admit multiplicative separable solutions

No.	Form of equation (14.3.1.1)	Form of particular solutions
1	Equation coefficients are constant	$w(\mathbf{x},t)=A\exp(\lambda t+\beta_1x_1+\cdots+\beta_nx_n)$, $\lambda,\beta_1,\ldots,\beta_n$ are related by an algebraic equation
2	Equation coefficients are independent of time t	$w(\mathbf{x},t)=e^{\lambda t}\psi(\mathbf{x})$, λ is an arbitrary constant, $\mathbf{x}=\{x_1,\ldots,x_n\}$
3	Equation coefficients are independent of the coordinates $x_1,\ldots,x_n$	$w(\mathbf{x},t)=\exp(\beta_1x_1+\cdots+\beta_nx_n)\psi(t)$, $\beta_1,\ldots,\beta_n$ are arbitrary constants
4	Equation coefficients are independent of the coordinates $x_1,\ldots,x_k$	$w(\mathbf{x},t)=\exp(\beta_1x_1+\cdots+\beta_kx_k)\psi(t,x_{k+1},\ldots,x_n)$, $\beta_1,\ldots,\beta_k$ are arbitrary constants
5	$L_t[w]+L_{\mathbf{x}}[w]=0$, operator L_t depends on only t, operator $L_{\mathbf{x}}$ depends on only $\mathbf{x}$	$w(\mathbf{x},t)=\varphi(t)\psi(\mathbf{x})$, $\varphi(t)$ satisfies the equation $L_t[\varphi]+\lambda\varphi=0$, $\psi(\mathbf{x})$ satisfies the equation $L_{\mathbf{x}}[\psi]-\lambda\psi=0$
6	$L_t[w]+L_1[w]+\cdots+L_n[w]=0$, operator L_t depends on only t, operator L_k depends on only x_k	$w(\mathbf{x},t)=\varphi(t)\psi_1(x_1)\ldots\psi_n(x_n)$, $\varphi(t)$ satisfies the equation $L_t[\varphi]+\lambda\varphi=0$, $\psi_k(x_k)$ satisfies the equation $L_k[\psi_k]+\beta_k\psi_k=0$, $\lambda+\beta_1+\cdots+\beta_n=0$
7	$f_0(x_1)L_t[w]+\sum_{k=1}^{n}f_k(x_1)L_k[w]=0$, operator L_t depends on only t, operator L_k depends on only x_k	$w(\mathbf{x},t)=\varphi(t)\psi_1(x_1)\ldots\psi_n(x_n)$, $L_t[\varphi]+\lambda\varphi=0$, $L_k[\psi_k]+\beta_k\psi_k=0$, $k=2,\ldots,n$, $f_1(x_1)L_1[\psi_1]-\Big[\lambda f_0(x_1)+\sum_{k=2}^{n}\beta_kf_k(x_1)\Big]\psi_1=0$
8	$\dfrac{\partial w}{\partial t}+L_{1,t}[w]+\cdots+L_{n,t}[w]=0$, where $L_{k,t}[w]=\sum_{s=0}^{m_k}f_{ks}(x_k,t)\dfrac{\partial^s w}{\partial x_k^s}$	$w(\mathbf{x},t)=\psi_1(x_1,t)\psi_2(x_2,t)\ldots\psi_n(x_n,t)$, $\dfrac{\partial\psi_k}{\partial t}+L_{k,t}[\psi_k]=\lambda_k(t)\psi_k$, $k=1,\ldots,n$, $\lambda_1(t)+\lambda_2(t)+\cdots+\lambda_n(t)=0$

14.3.1-3. Multiplicative and additive separable solutions.

1°. Many homogeneous linear partial differential equations have solutions that can be represented as the product of functions depending on different arguments. Such solutions are referred to as *multiplicative separable solutions*; very commonly these solutions are briefly, but less accurately, called just *separable solutions*.

Table 14.4 presents the most commonly encountered types of homogeneous linear differential equations with many independent variables that admit exact separable solutions. Linear combinations of particular solutions that correspond to different values of the separation parameters, λ, β_1, $\ldots$, β_n, are also solutions of the equations in question. For brevity, the word "operator" is used to denote "linear differential operator."

For a constant coefficient equation (see the first row in Table 14.4), the separation parameters must satisfy the algebraic equation

$$D(\lambda,\ \beta_1,\ \ldots,\beta_n)=0, \tag{14.3.1.4}$$

which results from substituting the solution into equation (14.3.1.1). In physical applications, equation (14.3.1.4) is usually referred to as a *dispersion equation*. Any n of the $n+1$ separation parameters in (14.3.1.4) can be treated as arbitrary.

Example 2. Consider the linear equation

$$\frac{\partial^2 w}{\partial t^2} + k\frac{\partial w}{\partial t} = a^2\frac{\partial^2 w}{\partial x^2} + b\frac{\partial w}{\partial x} + cw.$$

A particular solution is sought in the form

$$w = A\exp(\beta x + \lambda t).$$

This results in the dispersion equation $\lambda^2 + k\lambda = a^2\beta^2 + b\beta + c$, where one of the two parameters β or λ can be treated as arbitrary.

For more complex multiplicative separable solutions to this equation, see Subsection 14.4.1.

Note that constant coefficient equations also admit more sophisticated solutions; see the second and third rows, the last column.

The eighth row of Table 14.4 presents the case of *incomplete separation of variables* where the solution is separated with respect to the space variables $x_1, \ldots, x_n$, but is not separated with respect to time t.

Remark 1. For stationary equations that do not depend on t, one should set $\lambda = 0$, $L_t[w] \equiv 0$, and $\varphi(t) \equiv 1$ in rows 1, 6, and 7 of Table 14.4.

Remark 2. Multiplicative separable solutions play an important role in the theory of linear partial differential equations; they are used for finding solutions to stationary and nonstationary boundary value problems; see Sections 14.4 and 14.7–14.9.

2°. Linear partial differential equations of the form

$$L_t[w] + L_{\mathbf{x}}[w] = f(\mathbf{x}) + g(t),$$

where L_t is a linear differential operator that depends on only t and $L_{\mathbf{x}}$ is a linear differential operator that depends on only $\mathbf{x}$, have solutions that can be represented as the sum of functions depending on different arguments

$$w = u(\mathbf{x}) + v(t).$$

Such solutions are referred to as *additive separable solutions*.

Example 3. The equation from Example 2 admits an exact additive separable solution $w = u(x) + v(t)$ with $u(x)$ and $v(t)$ described by the linear constant-coefficient ordinary differential equations

$$a^2u''_{xx} + bu'_x + cu = C,$$
$$v''_{tt} + kv'_t - cv = C,$$

where C is an arbitrary constant, which are easy to integrate. A more general partial differential equation with variable coefficients $a = a(x)$, $b = b(x)$, $k = k(t)$, and $c =$ const also admits an additive separable solution.

14.3.1-4. Solutions in the form of infinite series in t.

1°. The equation

$$\frac{\partial w}{\partial t} = M[w],$$

where M is an arbitrary linear differential operator of the second (or any) order that only depends on the space variables, has the formal series solution

$$w(\mathbf{x}, t) = f(\mathbf{x}) + \sum_{k=1}^{\infty}\frac{t^k}{k!}M^k[f(\mathbf{x})], \qquad M^k[f] = M\big[M^{k-1}[f]\big],$$

where $f(\mathbf{x})$ is an arbitrary infinitely differentiable function. This solution satisfies the initial condition $w(\mathbf{x}, 0) = f(\mathbf{x})$.

Example 4. Consider the heat equation

$$\frac{\partial w}{\partial t} = a\frac{\partial^2 w}{\partial x^2}.$$

In this case we have $M = a\frac{\partial^2}{\partial x^2}$. Therefore the formal series solution has the form

$$w(x,t) = f(x) + \sum_{k=1}^{\infty} \frac{(at)^k}{k!} f_x^{(2n)}(x), \qquad f_x^{(m)} = \frac{d^m}{dx^m} f(x).$$

If the function $f(x)$ is taken as a polynomial of degree n, the solution will also be a polynomial of degree n. For example, setting $f(x) = Ax^2 + Bx + C$, we obtain the particular solution

$$w(x,t) = A(x^2 + 2at) + Bx + C.$$

2°. The equation

$$\frac{\partial^2 w}{\partial t^2} = M[w],$$

where M is a linear differential operator, just as in Item 1°, has a formal solution represented by the sum of two series as

$$w(\mathbf{x},t) = \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!} M^k[f(\mathbf{x})] + \sum_{k=0}^{\infty} \frac{t^{2k+1}}{(2k+1)!} M^k[g(\mathbf{x})],$$

where $f(\mathbf{x})$ and $g(\mathbf{x})$ are arbitrary infinitely differentiable functions. This solution satisfies the initial conditions $w(\mathbf{x},0) = f(\mathbf{x})$ and $\partial_t w(\mathbf{x},0) = g(\mathbf{x})$.

14.3.2. Nonhomogeneous Linear Equations and Their Particular Solutions

14.3.2-1. Simplest properties of nonhomogeneous linear equations.

For brevity, we write a nonhomogeneous linear partial differential equation in the form

$$\mathcal{L}[w] = \Phi(\mathbf{x},t), \tag{14.3.2.1}$$

where the linear differential operator $\mathcal{L}$ is defined above (see the beginning of Paragraph 14.3.1-1).

Below are the simplest properties of particular solutions of the nonhomogeneous equation (14.3.2.1).

1°. If $\widetilde{w}_\Phi(\mathbf{x},t)$ is a particular solution of the nonhomogeneous equation (14.3.2.1) and $\widetilde{w}_0(\mathbf{x},t)$ is a particular solution of the corresponding homogeneous equation (14.3.1.1), then the sum

$$A\widetilde{w}_0(\mathbf{x},t) + \widetilde{w}_\Phi(\mathbf{x},t),$$

where A is an arbitrary constant, is also a solution of the nonhomogeneous equation (14.3.2.1). The following, more general statement holds: The general solution of the nonhomogeneous equation (14.3.2.1) is the sum of the general solution of the corresponding homogeneous equation (14.3.1.1) and any particular solution of the nonhomogeneous equation (14.3.2.1).

2°. Suppose w_1 and w_2 are solutions of nonhomogeneous linear equations with the same left-hand side and different right-hand sides, i.e.,

$$\mathcal{L}[w_1] = \Phi_1(\mathbf{x},t), \qquad \mathcal{L}[w_2] = \Phi_2(\mathbf{x},t).$$

Then the function $w = w_1 + w_2$ is a solution of the equation

$$\mathcal{L}[w] = \Phi_1(\mathbf{x},t) + \Phi_2(\mathbf{x},t).$$

14.3.2-2. Fundamental and particular solutions of stationary equations.

Consider the second-order linear stationary (time-independent) nonhomogeneous equation

$$L_{\mathbf{x}}[w] = -\Phi(\mathbf{x}). \tag{14.3.2.2}$$

Here, $L_{\mathbf{x}}$ is a linear differential operator of the second (or any) order of general form whose coefficients are dependent on $\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^n$.

A distribution $\mathscr{E} = \mathscr{E}(\mathbf{x}, \mathbf{y})$ that satisfies the equation with a special right-hand side

$$L_{\mathbf{x}}[\mathscr{E}] = -\delta(\mathbf{x} - \mathbf{y}) \tag{14.3.2.3}$$

is called a *fundamental solution* corresponding to the operator $L_{\mathbf{x}}$. In (14.3.2.3), $\delta(\mathbf{x})$ is an n-dimensional Dirac delta function and the vector quantity $\mathbf{y} = \{y_1, \ldots, y_n\}$ appears in equation (14.3.2.3) as an n-dimensional free parameter. It is assumed that $\mathbf{y} \in \mathbb{R}^n$.

The n-dimensional Dirac delta function possesses the following basic properties:

1. $\delta(\mathbf{x}) = \delta(x_1)\delta(x_2)\ldots\delta(x_n)$,
2. $\displaystyle\int_{\mathbb{R}^n} \Phi(\mathbf{y})\delta(\mathbf{x} - \mathbf{y})\, d\mathbf{y} = \Phi(\mathbf{x})$,

where $\delta(x_k)$ is the one-dimensional Dirac delta function, $\Phi(\mathbf{x})$ is an arbitrary continuous function, and $d\mathbf{y} = dy_1 \ldots dy_n$.

For constant coefficient equations, a fundamental solution always exists; it can be found by means of the n-dimensional Fourier transform (see Paragraph 11.4.1-4).

The fundamental solution $\mathscr{E} = \mathscr{E}(\mathbf{x}, \mathbf{y})$ can be used to construct a particular solution of the linear stationary nonhomogeneous equation (14.3.2.2) for arbitrary continuous $\Phi(\mathbf{x})$; this particular solution is expressed as follows:

$$w(\mathbf{x}) = \int_{\mathbb{R}^n} \Phi(\mathbf{y})\mathscr{E}(\mathbf{x}, \mathbf{y})\, d\mathbf{y}. \tag{14.3.2.4}$$

Remark 1. The fundamental solution $\mathscr{E}$ is not unique; it is defined up to an additive term $w_0 = w_0(\mathbf{x})$, which is an arbitrary solution of the homogeneous equation $L_{\mathbf{x}}[w_0] = 0$.

Remark 2. For constant coefficient differential equations, the fundamental solution possesses the property $\mathscr{E}(\mathbf{x}, \mathbf{y}) = \mathscr{E}(\mathbf{x} - \mathbf{y})$.

Remark 3. The right-hand sides of equations (14.3.2.2) and (14.3.2.3) are often prefixed with the plus sign. In this case, formula (14.3.2.4) remains valid.

Remark 4. Particular solutions of linear nonstationary nonhomogeneous equations can be expressed in terms of the fundamental solution of the Cauchy problem; see Section 14.6.

Example 1. For the two- and three-dimensional Poisson equations, the fundamental solutions have the forms

Equations		*Fundamental solutions*
$\dfrac{\partial^2 w}{\partial x_1^2} + \dfrac{\partial^2 w}{\partial x_2^2} = -\Phi(x_1, x_2)$	$\Longrightarrow$	$\mathscr{E}(x_1, x_2, y_1, y_2) = \dfrac{1}{2\pi} \ln \dfrac{1}{\rho}$,
$\dfrac{\partial^2 w}{\partial x_1^2} + \dfrac{\partial^2 w}{\partial x_2^2} + \dfrac{\partial^2 w}{\partial x_3^2} = -\Phi(x_1, x_2, x_3)$	$\Longrightarrow$	$\mathscr{E}(x_1, x_2, x_3, y_1, y_2, y_3) = \dfrac{1}{4\pi r}$,

where $\rho = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$ and $r = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}$.

Example 2. The two-dimensional Helmholtz equation

$$\frac{\partial^2 w}{\partial x_1^2} + \frac{\partial^2 w}{\partial x_2^2} + \lambda w = -\Phi(x_1, x_2)$$

has the following fundamental solutions:

$$\mathscr{E}(x_1, x_2, y_1, y_2) = \frac{1}{2\pi} K_0(k\rho) \quad \text{if} \quad \lambda = -k^2 < 0,$$
$$\mathscr{E}(x_1, x_2, y_1, y_2) = -\frac{i}{4} H_0^{(2)}(k\rho) \quad \text{if} \quad \lambda = k^2 > 0,$$

where $\rho = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$, $K_0(z)$ is the modified Bessel function of the second kind, $H_0^{(2)}(z)$ is the Hankel functions of the second kind of order 0, $k > 0$, and $i^2 = -1$.

Example 3. The three-dimensional Helmholtz equation

$$\frac{\partial^2 w}{\partial x_1^2} + \frac{\partial^2 w}{\partial x_2^2} + \frac{\partial^2 w}{\partial x_3^2} + \lambda w = -\Phi(x_1, x_2, x_3)$$

has the following fundamental solutions:

$$\mathscr{E}(x_1, x_2, x_3, y_1, y_2, y_3) = \frac{1}{4\pi r} \exp(-kr) \quad \text{if} \quad \lambda = -k^2 < 0,$$
$$\mathscr{E}(x_1, x_2, x_3, y_1, y_2, y_3) = \frac{1}{4\pi r} \exp(-ikr) \quad \text{if} \quad \lambda = k^2 > 0,$$

where $r = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}$, $k > 0$, and $i^2 = -1$.

14.3.3. General Solutions of Some Hyperbolic Equations

14.3.3-1. D'Alembert's solution for the wave equation.

The wave equation

$$\frac{\partial^2 w}{\partial t^2} - a^2 \frac{\partial^2 w}{\partial x^2} = 0 \tag{14.3.3.1}$$

has the general solution

$$w = \varphi(x + at) + \psi(x - at), \tag{14.3.3.2}$$

where $\varphi(x)$ and $\psi(x)$ are arbitrary twice continuously differentiable functions. Solution (14.3.3.2) has the physical interpretation of two *traveling waves* of arbitrary shape that propagate to the right and to the left along the x-axis with a constant speed a.

14.3.3-2. Laplace cascade method for hyperbolic equation in two variables.

A general linear hyperbolic equation with two independent variables can be reduced to an equation of the form (see Subsection 14.1.1):

$$\frac{\partial^2 w}{\partial x \partial y} + a(x, y)\frac{\partial w}{\partial x} + b(x, y)\frac{\partial w}{\partial y} + c(x, y)w = f(x, y). \tag{14.3.3.3}$$

Sometimes it is possible to obtain formulas determining all solutions to equation (14.3.3.3). First consider two special cases.

1°. *First special case.* Suppose the identity

$$g \equiv \frac{\partial a}{\partial x} + ab - c \equiv 0 \tag{14.3.3.4}$$

is valid; for brevity, the arguments of the functions are omitted. Then equation (14.3.3.3) can be rewritten in the form

$$\frac{\partial u}{\partial x} + bu = f, \tag{14.3.3.5}$$

where

$$u = \frac{\partial w}{\partial y} + aw. \tag{14.3.3.6}$$

Equation (14.3.3.5) is a linear first-order ordinary differential equation in x for u (the variable y appears in the equation as a parameter) and is easy to integrate. Further substituting u into (14.3.3.6) yields a linear first-order ordinary differential equation in y for w (now x appears in the equation as a parameter). On solving this equation, one obtains the general solution of the original equation (14.3.3.3) subject to condition (14.3.3.4):

$$w = \exp\Big(-\int a\,dy\Big)\Big\{\varphi(x) + \int\Big[\psi(y) + \int f\exp\Big(\int b\,dx\Big)dx\Big]\exp\Big(\int a\,dy - \int b\,dx\Big)dy\Big\},$$

where $\varphi(x)$ and $\psi(y)$ are arbitrary functions.

2°. *Second special case.* Suppose the identity

$$h \equiv \frac{\partial b}{\partial y} + ab - c \equiv 0 \tag{14.3.3.7}$$

holds true. Proceeding in the same ways as in the first special case, one obtains the general solution to (14.3.3.3):

$$w = \exp\Big(-\int b\,dx\Big)\Big\{\psi(y) + \int\Big[\varphi(x) + \int f\exp\Big(\int a\,dy\Big)dy\Big]\exp\Big(\int b\,dx - \int a\,dy\Big)dx\Big\}.$$

3°. *Laplace cascade method.* In the case $g \neq 0$, consider the new equation of the form (14.3.3.3),

$$\mathrm{L}_1[w_1] \equiv \frac{\partial^2 w_1}{\partial x\partial y} + a_1(x,y)\frac{\partial w_1}{\partial x} + b_1(x,y)\frac{\partial w_1}{\partial y} + c_1(x,y)w_1 = f_1(x,y), \tag{14.3.3.8}$$

where

$$a_1 = a - \frac{\partial \ln g}{\partial y}, \quad b_1 = b, \quad c_1 = c - \frac{\partial a}{\partial x} + \frac{\partial b}{\partial y} - b\frac{\partial \ln g}{\partial y}, \quad f_1 = \left(a - \frac{\partial \ln g}{\partial y}\right)f.$$

If one manages to find w_1, the corresponding solution to the original equation (14.3.3.3) can be obtained by the formula

$$w = \frac{1}{g}\left(\frac{\partial w_1}{\partial x} + bw_1 - f\right).$$

For equation (14.3.3.8), the functions similar to (14.3.3.4) and (14.3.3.7) are expressed as

$$g_1 = 2g - h - \frac{\partial^2 \ln g}{\partial x\partial y}, \quad h_1 = h.$$

If $g_1 \equiv 0$, the function w_1 can be found using the technique described above. If $g_1 \not\equiv 0$, one proceeds to the construction, in the same way as above, of the equation $\mathrm{L}_2[w_2] = f_2$, and so on. In the case $h \not\equiv 0$, a similar chain of equations may be constructed: $\mathrm{L}_1^*[w_1^*] = f_1^*$, $\mathrm{L}_2^*[w_2^*] = f_2^*$, etc.

If at some step, g_k or h_k vanishes, it is possible to obtain the general solution of equation (14.3.3.3).

Example. Consider the *Euler–Darboux equation*

$$\frac{\partial^2 w}{\partial x \partial y} - \frac{\alpha}{x-y}\frac{\partial w}{\partial x} + \frac{\beta}{x-y}\frac{\partial w}{\partial y} = 0.$$

We will show that its general solution can be obtained if at least one of the numbers α or β is integer.

With the notation adopted for equation (14.3.3.3) and function (14.3.3.4), we have

$$a(x,y) = -\frac{\alpha}{x-y}, \quad b(x,y) = \frac{\beta}{x-y}, \quad c(x,y) = f(x,y) = 0, \quad g = \frac{\alpha(1-\beta)}{(x-y)^2},$$

which means that $g \equiv 0$ if $\alpha = 0$ or $\beta = 1$. If $g \not\equiv 0$, we construct equation (14.3.3.8), where

$$a_1 = -\frac{2+\alpha}{x-y}, \quad b_1 = \frac{\beta}{x-y}, \quad c_1 = -\frac{\alpha+\beta}{(x-y)^2}.$$

If follows that

$$g_1 = \frac{(1+\alpha)(2-\beta)}{(x-y)^2},$$

and hence $g_1 \equiv 0$ at $\alpha = -1$ or $\beta = 2$. Similarly, it can be shown that $g_k \equiv 0$ at $\alpha = -k$ or $\beta = k+1$ ($k = 0, 1, 2, \ldots$). If we use the other sequence of auxiliary equations, $\mathrm{L}_k^*[w_k^*] = f_k^*$, it can be shown that the above holds for $\alpha = 1, 2, \ldots$ and $\beta = 0, -1, -2, \ldots$

14.4. Method of Separation of Variables (Fourier Method)

14.4.1. Description of the Method of Separation of Variables. General Stage of Solution

14.4.1-1. Scheme of solving boundary value problems by separation of variables.

Many linear problems of mathematical physics can be solved by separation of variables. Figure 14.1 depicts the scheme of application of this method to solve boundary value problems for second-order homogeneous linear equations of the parabolic and hyperbolic type with homogeneous boundary conditions and nonhomogeneous initial conditions. For simplicity, problems with two independent variables x and t are considered, with $x_1 \le x \le x_2$ and $t \ge 0$.

The scheme presented in Fig. 14.1 applies to boundary value problems for second-order linear homogeneous partial differential equations of the form

$$\alpha(t)\frac{\partial^2 w}{\partial t^2} + \beta(t)\frac{\partial w}{\partial t} = a(x)\frac{\partial^2 w}{\partial x^2} + b(x)\frac{\partial w}{\partial x} + \big[c(x) + \gamma(t)\big]w \qquad (14.4.1.1)$$

with homogeneous linear boundary conditions,

$$\begin{aligned} s_1 \partial_x w + k_1 w &= 0 \quad \text{at} \quad x = x_1, \\ s_2 \partial_x w + k_2 w &= 0 \quad \text{at} \quad x = x_2, \end{aligned} \qquad (14.4.1.2)$$

and arbitrary initial conditions,

$$w = f_0(x) \quad \text{at} \quad t = 0, \qquad (14.4.1.3)$$

$$\partial_t w = f_1(x) \quad \text{at} \quad t = 0. \qquad (14.4.1.4)$$

For parabolic equations, which correspond to $\alpha(t) \equiv 0$ in (14.4.1.1), only the initial condition (14.4.1.3) is set.

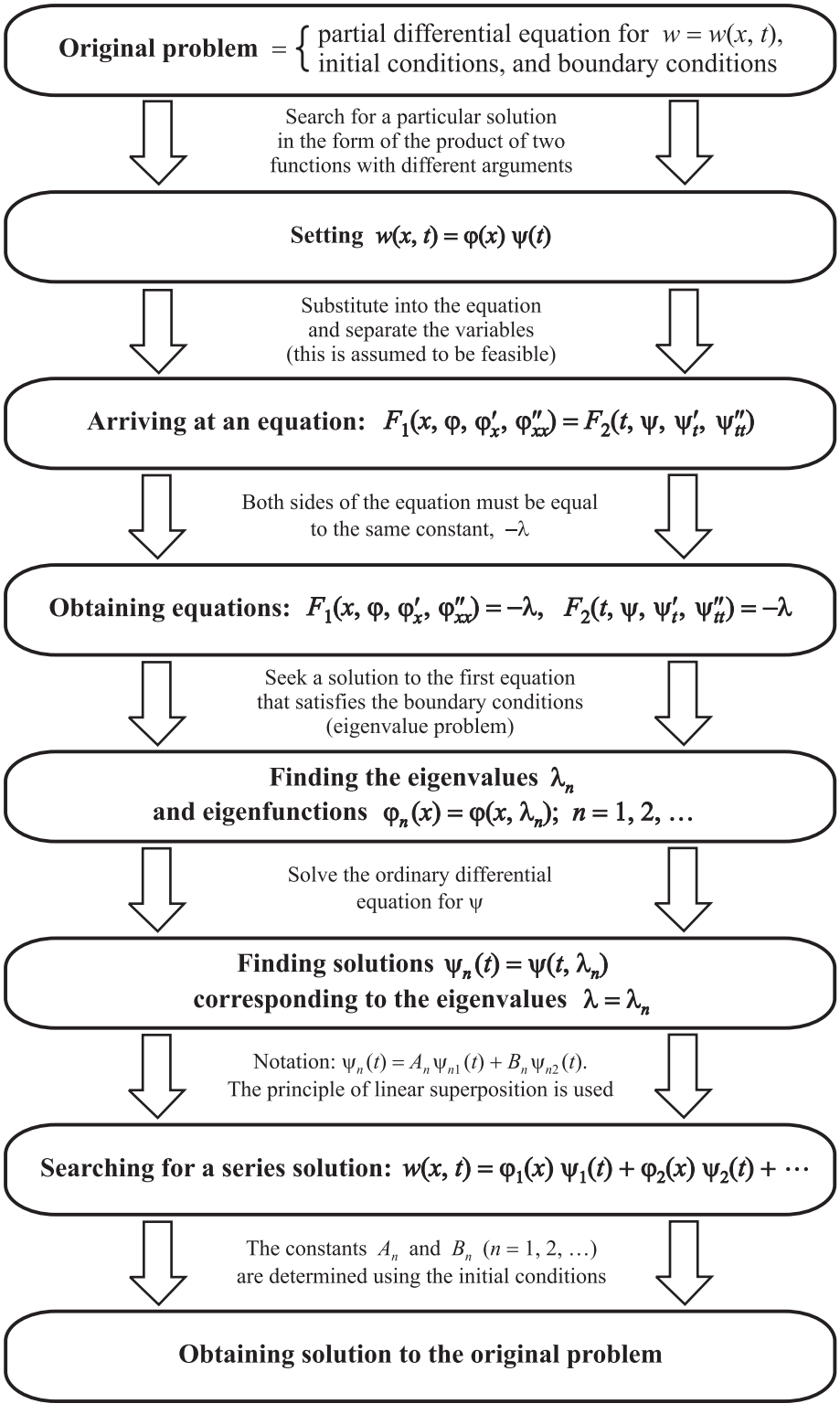


Figure 14.1. Scheme of solving linear boundary value problems by separation of variables (for parabolic equations, the function F_2 does not depend on ψ''_{tt}, and all $B_n = 0$).

Below we consider the basic stages of the method of separation of variables in more detail. We assume that the coefficients of equation (14.4.1.1) and boundary conditions (14.4.1.2) meet the following requirements:

$$\alpha(t),\ \beta(t),\ \gamma(t),\ a(x),\ b(x),\ c(x) \text{ are continuous functions,}$$
$$\alpha(t) \ge 0, \quad a(x) > 0, \quad |s_1| + |k_1| > 0, \quad |s_2| + |k_2| > 0.$$

Remark 1. The method of separation of variables is also used to solve linear boundary value problems for elliptic equations of the form (14.4.1.1), provided that $\alpha(t) < 0$, $a(x) > 0$ and with the boundary conditions (14.4.1.2) in x and similar boundary conditions in t. In this case, all results obtained for the general stage of solution, described in Paragraphs 14.4.1-2 and 14.4.1-3, remain valid; for details, see Subsection 14.4.4.

Remark 2. In various applications, equations of the form (14.4.1.1) may arise with the coefficient $b(x)$ going to infinity at the boundary, $b(x) \to \infty$ as $x \to x_1$, with the other coefficients being continuous. In this case, the first boundary condition in (14.4.1.2) should be replaced with a condition of boundedness of the solution as $x \to x_1$. This may occur in spatial problems with central or axial symmetry where the solution depends only on the radial coordinate.

14.4.1-2. Derivation of equations and boundary conditions for particular solutions.

The approach is based on searching for particular solutions of equation (14.4.1.1) in the product form

$$w(x,t) = \varphi(x)\,\psi(t). \tag{14.4.1.5}$$

After separation of the variables and elementary manipulations, one arrives at the following linear ordinary differential equations for the functions $\varphi = \varphi(x)$ and $\psi = \psi(t)$:

$$a(x)\varphi''_{xx} + b(x)\varphi'_x + [\lambda + c(x)]\varphi = 0, \tag{14.4.1.6}$$
$$\alpha(t)\psi''_{tt} + \beta(t)\psi'_t + [\lambda - \gamma(t)]\psi = 0. \tag{14.4.1.7}$$

These equations contain a free parameter λ called the separation constant. With the notation adopted in Fig. 14.1, equations (14.4.1.6) and (14.4.1.7) can be rewritten as follows: $\varphi F_1(x, \varphi, \varphi'_x, \varphi''_{xx}) + \lambda\varphi = 0$ and $\psi F_2(t, \psi, \psi'_t, \psi''_{tt}) + \lambda\psi = 0$.

Substituting (14.4.1.5) into (14.4.1.2) yields the boundary conditions for $\varphi = \varphi(x)$:

$$\begin{aligned} s_1\varphi'_x + k_1\varphi = 0 \quad &\text{at} \quad x = x_1, \\ s_2\varphi'_x + k_2\varphi = 0 \quad &\text{at} \quad x = x_2. \end{aligned} \tag{14.4.1.8}$$

The homogeneous linear ordinary differential equation (14.4.1.6) in conjunction with the homogeneous linear boundary conditions (14.4.1.8) makes up an eigenvalue problem.

14.4.1-3. Solution of eigenvalue problems. Orthogonality of eigenfunctions.

Suppose $\widetilde{\varphi}_1 = \widetilde{\varphi}_1(x, \lambda)$ and $\widetilde{\varphi}_2 = \widetilde{\varphi}_2(x, \lambda)$ are linearly independent particular solutions of equation (14.4.1.6). Then the general solution of this equation can be represented as the linear combination

$$\varphi = C_1\widetilde{\varphi}_1(x, \lambda) + C_2\widetilde{\varphi}_2(x, \lambda), \tag{14.4.1.9}$$

where C_1 and C_2 are arbitrary constants.

Substituting solution (14.4.1.9) into the boundary conditions (14.4.1.8) yields the following homogeneous linear algebraic system of equations for C_1 and C_2:

$$\begin{aligned} \varepsilon_{11}(\lambda)C_1 + \varepsilon_{12}(\lambda)C_2 = 0, \\ \varepsilon_{21}(\lambda)C_1 + \varepsilon_{22}(\lambda)C_2 = 0, \end{aligned} \tag{14.4.1.10}$$

where $\varepsilon_{ij}(\lambda) = \left[s_i(\widetilde{\varphi}_j)'_x + k_i\widetilde{\varphi}_j\right]_{x=x_i}$. For system (14.4.1.10) to have nontrivial solutions, its determinant must be zero; we have

$$\varepsilon_{11}(\lambda)\varepsilon_{22}(\lambda) - \varepsilon_{12}(\lambda)\varepsilon_{21}(\lambda) = 0. \tag{14.4.1.11}$$

Solving the transcendental equation (14.4.1.11) for λ, one obtains the *eigenvalues* $\lambda = \lambda_n$, where $n = 1, 2, \ldots$ For these values of λ, equation (14.4.1.6) has nontrivial solutions,

$$\varphi_n(x) = \varepsilon_{12}(\lambda_n)\widetilde{\varphi}_1(x, \lambda_n) - \varepsilon_{11}(\lambda_n)\widetilde{\varphi}_2(x, \lambda_n), \tag{14.4.1.12}$$

which are called *eigenfunctions* (these functions are defined up to a constant multiplier).

To facilitate the further analysis, we represent equation (14.4.1.6) in the form

$$[p(x)\varphi'_x]'_x + [\lambda\rho(x) - q(x)]\varphi = 0, \tag{14.4.1.13}$$

where

$$p(x) = \exp\left[\int \frac{b(x)}{a(x)}\,dx\right], \quad q(x) = -\frac{c(x)}{a(x)}\exp\left[\int \frac{b(x)}{a(x)}\,dx\right], \quad \rho(x) = \frac{1}{a(x)}\exp\left[\int \frac{b(x)}{a(x)}\,dx\right]. \tag{14.4.1.14}$$

It follows from the adopted assumptions (see the end of Paragraph 14.4.1-1) that $p(x)$, $p'_x(x)$, $q(x)$, and $\rho(x)$ are continuous functions, with $p(x) > 0$ and $\rho(x) > 0$.

The eigenvalue problem (14.4.1.13), (14.4.1.8) is known to possess the following properties:

1. All eigenvalues $\lambda_1, \lambda_2, \ldots$ are real, and $\lambda_n \to \infty$ as $n \to \infty$; consequently, the number of negative eigenvalues is finite.

2. The system of eigenfunctions $\varphi_1(x), \varphi_2(x), \ldots$ is orthogonal on the interval $x_1 \le x \le x_2$ with weight $\rho(x)$, i.e.,

$$\int_{x_1}^{x_2} \rho(x)\varphi_n(x)\varphi_m(x)\,dx = 0 \quad \text{for} \quad n \ne m. \tag{14.4.1.15}$$

3. If

$$q(x) \ge 0, \quad s_1k_1 \le 0, \quad s_2k_2 \ge 0, \tag{14.4.1.16}$$

there are no negative eigenvalues. If $q \equiv 0$ and $k_1 = k_2 = 0$, the least eigenvalue is $\lambda_1 = 0$ and the corresponding eigenfunction is $\varphi_1 = \text{const}$. Otherwise, all eigenvalues are positive, provided that conditions (14.4.1.16) are satisfied; the first inequality in (14.4.1.16) is satisfied if $c(x) \le 0$.

Subsection 12.2.5 presents some estimates for the eigenvalues λ_n and eigenfunctions $\varphi_n(x)$.

► *The procedure of constructing solutions to nonstationary boundary value problems is further different for parabolic and hyperbolic equations; see Subsections 14.4.2 and 14.4.3 below for results (elliptic equations are treated in Subsection 14.4.4).*

14.4.2. Problems for Parabolic Equations: Final Stage of Solution

14.4.2-1. Series solutions of boundary value problems for parabolic equations.

Consider the problem for the parabolic equation

$$\frac{\partial w}{\partial t} = a(x)\frac{\partial^2 w}{\partial x^2} + b(x)\frac{\partial w}{\partial x} + \left[c(x) + \gamma(t)\right]w \tag{14.4.2.1}$$

(this equation is obtained from (14.4.1.1) in the case $\alpha(t) \equiv 0$ and $\beta(t) = 1$) with homogeneous linear boundary conditions (14.4.1.2) and initial condition (14.4.1.3).

First, one searches for particular solutions to equation (14.4.2.1) in the product form (14.4.1.5), where the function $\varphi(x)$ is obtained by solving an eigenvalue problem for the ordinary differential equation (14.4.1.6) with the boundary conditions (14.4.1.8). The solution of equation (14.4.1.7) with $\alpha(t) \equiv 0$ and $\beta(t) = 1$ corresponding to the eigenvalues $\lambda = \lambda_n$ and satisfying the normalizing conditions $\psi_n(0) = 1$ has the form

$$\psi_n(t) = \exp\left[-\lambda_n t + \int_0^t \gamma(\xi)\, d\xi\right]. \tag{14.4.2.2}$$

Then the solution of the nonstationary boundary value problem (14.4.2.1), (14.4.1.2), (14.4.1.3) is sought in the series form

$$w(x,t) = \sum_{n=1}^{\infty} A_n \varphi_n(x)\, \psi_n(t), \tag{14.4.2.3}$$

where the A_n are arbitrary constants and the functions $w_n(x,t) = \varphi_n(x)\, \psi_n(t)$ are particular solutions (14.4.2.1) satisfying the boundary conditions (14.4.1.2). By the principle of linear superposition, series (14.4.2.3) is also a solution of the original partial differential equation that satisfies the boundary conditions.

To determine the coefficients A_n, we substitute series (14.4.2.3) into the initial condition (14.4.1.3), thus obtaining

$$\sum_{n=1}^{\infty} A_n \varphi_n(x) = f_0(x).$$

Multiplying this equation by $\rho(x)\varphi_n(x)$, where the weight function $\rho(x)$ is defined in (14.4.1.14), then integrating the resulting relation with respect to x over the interval $x_1 \le x \le x_2$, and taking into account the properties (14.4.1.15), we find

$$A_n = \frac{1}{\|\varphi_n\|^2} \int_{x_1}^{x_2} \rho(x)\varphi_n(x) f_0(x)\, dx, \quad \|\varphi_n\|^2 = \int_{x_1}^{x_2} \rho(x)\varphi_n^2(x)\, dx. \tag{14.4.2.4}$$

Relations (14.4.2.3), (14.4.2.2), (14.4.2.4), and (14.4.1.12) give a formal solution of the nonstationary boundary value problem (14.4.2.1), (14.4.1.2), (14.4.1.3).

Example. Consider the first (Dirichlet) boundary value problem on the interval $0 \le x \le l$ for the heat equation

$$\frac{\partial w}{\partial t} = \frac{\partial^2 w}{\partial x^2} \tag{14.4.2.5}$$

with the general initial condition (14.4.1.3) and the homogeneous boundary conditions

$$w = 0 \quad \text{at} \quad x = 0, \qquad w = 0 \quad \text{at} \quad x = l. \tag{14.4.2.6}$$

The function $\psi(t)$ in the particular solution (14.4.1.5) is found from (14.4.2.2), where $\gamma(t) = 0$:

$$\psi_n(t) = \exp(-\lambda_n t). \tag{14.4.2.7}$$

The functions $\varphi_n(x)$ are determined by solving the eigenvalue problem (14.4.1.6), (14.4.1.8) with $a(x) = 1$, $b(x) = c(x) = 0$, $s_1 = s_2 = 0$, $k_1 = k_2 = 1$, $x_1 = 0$, and $x_2 = l$:

$$\varphi''_{xx} + \lambda\varphi = 0; \qquad \varphi = 0 \quad \text{at} \quad x = 0, \qquad \varphi = 0 \quad \text{at} \quad x = l.$$

So we obtain the eigenfunctions and eigenvalues:

$$\varphi_n(x) = \sin\left(\frac{n\pi x}{l}\right), \quad \lambda_n = \left(\frac{n\pi}{l}\right)^2, \quad n = 1, 2, \ldots \tag{14.4.2.8}$$

The solution to problem (14.4.2.5)–(14.4.2.6), (14.4.1.3) is given by formulas (14.4.2.3), (14.4.2.4). Taking into account that $\|\varphi_n\|^2 = l/2$, we obtain

$$w(x,t) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi x}{l}\right) \exp\left(-\frac{n^2\pi^2 t}{l^2}\right), \qquad A_n = \frac{2}{l}\int_0^l f_0(\xi)\sin\left(\frac{n\pi\xi}{l}\right) d\xi. \tag{14.4.2.9}$$

If the function $f_0(x)$ is twice continuously differentiable and the compatibility conditions (see Paragraph 14.4.2-2) are satisfied, then series (14.4.2.9) is convergent and admits termwise differentiation, once with respect to t and twice with respect to x. In this case, formula (14.4.2.9) gives the classical smooth solution of problem (14.4.2.5)–(14.4.2.6), (14.4.1.3). [If $f_0(x)$ is not as smooth as indicated or if the compatibility conditions are not met, then series (14.4.2.9) may converge to a discontinuous function, thus giving only a generalized solution.]

Remark. For the solution of linear nonhomogeneous parabolic equations with nonhomogeneous boundary conditions, see Section 14.7.

14.4.2-2. Conditions of compatibility of initial and boundary conditions.

Suppose the function w has a continuous derivative with respect to t and two continuous derivatives with respect to x and is a solution of problem (14.4.2.1), (14.4.1.2), (14.4.1.3). Then the boundary conditions (14.4.1.2) and the initial condition (14.4.1.3) must be consistent; namely, the following compatibility conditions must hold:

$$[s_1 f_0' + k_1 f_0]_{x=x_1} = 0, \quad [s_2 f_0' + k_2 f_0]_{x=x_2} = 0. \tag{14.4.2.10}$$

If $s_1 = 0$ or $s_2 = 0$, then the additional compatibility conditions

$$\begin{aligned} [a(x)f_0'' + b(x)f_0']_{x=x_1} &= 0 \quad \text{if} \quad s_1 = 0, \\ [a(x)f_0'' + b(x)f_0']_{x=x_2} &= 0 \quad \text{if} \quad s_2 = 0 \end{aligned} \tag{14.4.2.11}$$

must also hold; the primes denote the derivatives with respect to x.

14.4.3. Problems for Hyperbolic Equations: Final Stage of Solution

14.4.3-1. Series solution of boundary value problems for hyperbolic equations.

For hyperbolic equations, the solution of the boundary value problem (14.4.1.1)–(14.4.1.4) is sought in the series form

$$w(x,t) = \sum_{n=1}^{\infty} \varphi_n(x)\left[A_n\psi_{n1}(t) + B_n\psi_{n2}(t)\right]. \tag{14.4.3.1}$$

Here, A_n and B_n are arbitrary constants. The functions $\psi_{n1}(t)$ and $\psi_{n2}(t)$ are particular solutions of the linear equation (14.4.1.7) for ψ (with $\lambda = \lambda_n$) that satisfy the conditions

$$\psi_{n1}(0) = 1, \quad \psi_{n1}'(0) = 0; \qquad \psi_{n2}(0) = 0, \quad \psi_{n2}'(0) = 1. \tag{14.4.3.2}$$

The functions $\varphi_n(x)$ and λ_n are determined by solving the eigenvalue problem (14.4.1.6), (14.4.1.8).

Substituting solution (14.4.3.1) into the initial conditions (14.4.1.3)–(14.4.1.4) yields

$$\sum_{n=1}^{\infty} A_n\varphi_n(x) = f_0(x), \qquad \sum_{n=1}^{\infty} B_n\varphi_n(x) = f_1(x).$$

Multiplying these equations by $\rho(x)\varphi_n(x)$, where the weight function $\rho(x)$ is defined in (14.4.1.14), then integrating the resulting relations with respect to x on the interval $x_1 \le x \le x_2$, and taking into account the properties (14.4.1.15), we obtain the coefficients of series (14.4.3.1) in the form

$$A_n = \frac{1}{\|\varphi_n\|^2}\int_{x_1}^{x_2} \rho(x)\varphi_n(x)f_0(x)\,dx, \quad B_n = \frac{1}{\|\varphi_n\|^2}\int_{x_1}^{x_2} \rho(x)\varphi_n(x)f_1(x)\,dx. \quad (14.4.3.3)$$

The quantity $\|\varphi_n\|$ is defined in (14.4.2.4).

Relations (14.4.3.1), (14.4.1.12), and (14.4.3.3) give a formal solution of the nonstationary boundary value problem (14.4.1.1)–(14.4.1.4) for $\alpha(t) > 0$.

Example. Consider a mixed boundary value problem on the interval $0 \le x \le l$ for the wave equation

$$\frac{\partial^2 w}{\partial t^2} = \frac{\partial^2 w}{\partial x^2} \quad (14.4.3.4)$$

with the general initial conditions (14.4.1.3)–(14.4.1.4) and the homogeneous boundary conditions

$$w = 0 \quad \text{at} \quad x = 0, \qquad \partial_x w = 0 \quad \text{at} \quad x = l. \quad (14.4.3.5)$$

The functions $\psi_{n1}(t)$ and $\psi_{n2}(t)$ are determined by the linear equation [see (14.4.1.7) with $\alpha(t) = 1$, $\beta(t) = \gamma(t) = 0$, and $\lambda = \lambda_n$]

$$\psi''_{tt} + \lambda\psi = 0$$

with the initial conditions (14.4.3.2). We find

$$\psi_{n1}(t) = \cos\bigl(\sqrt{\lambda_n}\,t\bigr), \quad \psi_{n2}(t) = \frac{1}{\sqrt{\lambda_n}}\sin\bigl(\sqrt{\lambda_n}\,t\bigr). \quad (14.4.3.6)$$

The functions $\varphi_n(x)$ are determined by solving the eigenvalue problem (14.4.1.6), (14.4.1.8) with $a(x) = 1$, $b(x) = c(x) = 0$, $s_1 = k_2 = 0$, $s_2 = k_1 = 1$, $x_1 = 0$, and $x_2 = l$:

$$\varphi''_{xx} + \lambda\varphi = 0; \qquad \varphi = 0 \quad \text{at} \quad x = 0, \qquad \varphi'_x = 0 \quad \text{at} \quad x = l.$$

So we obtain the eigenfunctions and eigenvalues:

$$\varphi_n(x) = \sin(\mu_n x), \quad \mu_n = \sqrt{\lambda_n} = \frac{\pi(2n-1)}{2l}, \quad n = 1, 2, \ldots \quad (14.4.3.7)$$

The solution to problem (14.4.3.4)–(14.4.3.5), (14.4.1.3)–(14.4.1.4) is given by formulas (14.4.3.1) and (14.4.3.3). Taking into account that $\|\varphi_n\|^2 = l/2$, we have

$$\begin{aligned} w(x,t) &= \sum_{n=1}^{\infty}\bigl[A_n\cos(\mu_n t) + B_n\sin(\mu_n t)\bigr]\sin(\mu_n x), \quad \mu_n = \frac{\pi(2n-1)}{2l}, \\ A_n &= \frac{2}{l}\int_0^l f_0(x)\sin(\mu_n x)\,dx, \quad B_n = \frac{2}{l\mu_n}\int_0^l f_1(x)\sin(\mu_n x)\,dx. \end{aligned} \quad (14.4.3.8)$$

If $f_0(x)$ and $f_1(x)$ have three and two continuous derivatives, respectively, and the compatibility conditions are met (see Paragraph 14.4.3-2), then series (14.4.3.8) is convergent and admits double termwise differentiation. In this case, formula (14.4.3.8) gives the classical smooth solution of problem (14.4.3.4)–(14.4.3.5), (14.4.1.3)–(14.4.1.4).

Remark. For the solution of linear nonhomogeneous hyperbolic equations with nonhomogeneous boundary conditions, see Section 14.8.

14.4.3-2. Conditions of compatibility of initial and boundary conditions.

Suppose w is a twice continuously differentiable solution of problem (14.4.1.1)–(14.4.1.4). Then conditions (14.4.2.10) and (14.4.2.11) must hold. In addition, the following conditions of compatibility of the boundary conditions (14.4.1.2) and initial condition (14.4.1.4) must be satisfied:

$$[s_1 f_1' + k_1 f_1]_{x=x_1} = 0, \qquad [s_2 f_1' + k_2 f_1]_{x=x_2} = 0.$$

14.4.4. Solution of Boundary Value Problems for Elliptic Equations

14.4.4-1. Solution of special problem for elliptic equations.

Now consider a boundary value problem for the elliptic equation

$$a(x)\frac{\partial^2 w}{\partial x^2} + \alpha(y)\frac{\partial^2 w}{\partial y^2} + b(x)\frac{\partial w}{\partial x} + \beta(y)\frac{\partial w}{\partial y} + \big[c(x) + \gamma(y)\big]w = 0 \tag{14.4.4.1}$$

with homogeneous boundary conditions (14.4.1.2) in x and the following mixed (homogeneous and nonhomogeneous) boundary conditions in y:

$$\begin{aligned} &\sigma_1 \partial_y w + \nu_1 w = 0 \quad &&\text{at} \quad y = y_1, \\ &\sigma_2 \partial_y w + \nu_2 w = f(x) \quad &&\text{at} \quad y = y_2. \end{aligned} \tag{14.4.4.2}$$

We assume that the coefficients of equation (14.4.4.1) and boundary conditions (14.4.1.2) and (14.4.4.2) meet the following requirements:

$$\begin{aligned} &a(x),\ b(x),\ c(x)\ \ \alpha(y),\ \beta(y),\ \text{and}\ \ \gamma(t)\ \ \text{are continuous functions,} \\ &a(x) > 0,\ \alpha(y) > 0,\ |s_1| + |k_1| > 0,\ |s_2| + |k_2| > 0,\ |\sigma_1| + |\nu_1| > 0,\ |\sigma_2| + |\nu_2| > 0. \end{aligned}$$

The approach is based on searching for particular solutions of equation (14.4.4.1) in the product form

$$w(x, y) = \varphi(x)\,\psi(y). \tag{14.4.4.3}$$

As before, we first arrive at the eigenvalue problem (14.4.1.6), (14.4.1.8) for the function $\varphi = \varphi(x)$; the solution procedure is detailed in Paragraph 14.4.1-3. Further on, we assume the λ_n and $\varphi_n(x)$ have been found. The functions $\psi_n = \psi_n(y)$ are determined by solving the linear ordinary differential equation

$$\alpha(y)\psi_{yy}'' + \beta(y)\psi_y' + [\gamma(y) - \lambda_n]\psi = 0 \tag{14.4.4.4}$$

subject to the homogeneous boundary condition

$$\sigma_1 \partial_y \psi + \nu_1 \psi = 0 \quad \text{at} \quad y = y_1, \tag{14.4.4.5}$$

which is a consequence of the first condition (14.4.4.2). The functions ψ_n are determined up to a constant factor.

Taking advantage of the principle of linear superposition, we seek the solution to the boundary value problem (14.4.4.1), (14.4.4.2), (14.4.1.2) in the series form

$$w(x, y) = \sum_{n=1}^{\infty} A_n \varphi_n(x)\psi_n(y), \tag{14.4.4.6}$$

where A_n are arbitrary constants. By construction, series (14.4.4.6) will satisfy equation (14.4.4.1) with the boundary conditions (14.4.1.2) and the first boundary condition (14.4.4.2). In order to find the series coefficients A_n, substitute (14.4.4.6) into the second boundary condition (14.4.4.2) to obtain

$$\sum_{n=1}^{\infty} A_n B_n \varphi_n(x) = f(x), \qquad B_n = \sigma_2 \frac{d\psi_n}{dy}\bigg|_{y=y_2} + \nu_2 \psi_n(y_2). \tag{14.4.4.7}$$

Further, we follow the same procedure as in Paragraph 14.4.2-1. Specifically, multiplying (14.4.4.7) by $\rho(x)\varphi_n(x)$, then integrating the resulting relation with respect to x over the interval $x_1 \le x \le x_2$, and taking into account the properties (14.4.1.15), we obtain

$$A_n = \frac{1}{B_n \|\varphi_n\|^2} \int_{x_1}^{x_2} \rho(x)\varphi_n(x) f(x)\, dx, \quad \|\varphi_n\|^2 = \int_{x_1}^{x_2} \rho(x)\varphi_n^2(x)\, dx, \tag{14.4.4.8}$$

where the weight function $\rho(x)$ is defined in (14.4.1.14).

Example. Consider the first (Dirichlet) boundary value problem for the Laplace equation

$$\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} = 0 \tag{14.4.4.9}$$

subject to the boundary conditions

$$\begin{aligned} &w = 0 \quad \text{at} \quad x = 0, \qquad w = 0 \qquad \text{at} \quad x = l_1; \\ &w = 0 \quad \text{at} \quad y = 0, \qquad w = f(x) \quad \text{at} \quad y = l_2 \end{aligned} \tag{14.4.4.10}$$

in a rectangular domain $0 \le x \le l_1$, $0 \le y \le l_2$.

Particular solutions to equation (14.4.4.9) are sought in the form (14.4.4.3). We have the following eigenvalue problem for $\varphi(x)$:

$$\varphi''_{xx} + \lambda\varphi = 0; \qquad \varphi = 0 \quad \text{at} \quad x = 0, \qquad \varphi = 0 \quad \text{at} \quad x = l_1.$$

On solving this problem, we find the eigenfunctions with respective eigenvalues

$$\varphi_n(x) = \sin(\mu_n x), \quad \mu_n = \sqrt{\lambda_n} = \frac{\pi n}{l_1}, \quad n = 1, 2, \ldots \tag{14.4.4.11}$$

The functions $\psi_n = \psi_n(y)$ are determined by solving the following problem for a linear ordinary differential equation with homogeneous boundary conditions:

$$\psi''_{yy} - \lambda_n \psi = 0; \qquad \psi = 0 \quad \text{at} \quad y = 0. \tag{14.4.4.12}$$

It is a special case of problem (14.4.4.4)–(14.4.4.5) with $\alpha(y) = 1$, $\beta(y) = \gamma(y) = 0$, $\sigma_1 = 0$, and $\nu_1 = 1$. The nontrivial solutions of problem (14.4.4.12) are expressed as

$$\psi_n(y) = \sinh(\mu_n y), \quad \mu_n = \sqrt{\lambda_n} = \frac{\pi n}{l_1}, \quad n = 1, 2, \ldots \tag{14.4.4.13}$$

Using formulas (14.4.4.6), (14.4.4.8), (14.4.4.11), (14.4.4.13) and taking into account the relations $B_n = \psi_n(l_2) = \sinh(\mu_n l_2)$, $\rho(x) = 1$, and $\|\varphi_n\|^2 = l/2$, we find the solution of the original problem (14.4.4.9)–(14.4.4.10) in the form

$$w(x,y) = \sum_{n=1}^{\infty} A_n \sin(\mu_n x)\sinh(\mu_n y), \qquad A_n = \frac{2}{l_1 \sinh(\mu_n l_2)} \int_0^{l_1} f(x)\sin(\mu_n x)\, dx, \quad \mu_n = \frac{\pi n}{l_1}.$$

TABLE 14.5
Description of auxiliary problems for equation (14.4.4.1) and problems for associated functions $\varphi(x)$ and $\psi(y)$ that determine particular solutions of the form (14.4.4.3). The abbreviation HBC below stands for a "homogeneous boundary condition"

Auxiliary problem	Functions vanishing in the boundary conditions (14.4.4.14)	Eigenvalue problem with homogeneous boundary conditions	Another problem with one homogeneous boundary condition (for λ_n found)
Problem 1	$f_2(y) = f_3(x) = f_4(x) = 0$, function $f_1(y)$ prescribed	functions $\psi_n(y)$ and values λ_n to be determined	functions $\varphi_n(x)$ satisfy an HBC at $x = x_2$
Problem 2	$f_1(y) = f_3(x) = f_4(x) = 0$, function $f_2(y)$ prescribed	functions $\psi_n(y)$ and values λ_n to be determined	functions $\varphi(x)$ satisfy an HBC at $x = x_1$
Problem 3	$f_1(y) = f_2(y) = f_4(x) = 0$, function $f_3(x)$ prescribed	functions $\varphi_n(x)$ and values λ_n to be determined	functions $\psi(y)$ satisfy an HBC at $y = y_2$
Problem 4	$f_1(y) = f_2(y) = f_3(x) = 0$, function $f_4(x)$ prescribed	functions $\varphi_n(x)$ and values λ_n to be determined	functions $\psi(y)$ satisfy an HBC at $y = y_1$

14.4.4-2. Generalization to the case of nonhomogeneous boundary conditions.

Now consider the linear boundary value problem for the elliptic equation (14.4.4.1) with general nonhomogeneous boundary conditions

$$\begin{aligned} &s_1\partial_x w + k_1 w = f_1(y) \quad \text{at} \quad x = x_1, \qquad s_2\partial_x w + k_2 w = f_2(y) \quad \text{at} \quad x = x_2, \\ &\sigma_1\partial_y w + \nu_1 w = f_3(x) \quad \text{at} \quad y = y_1, \qquad \sigma_2\partial_y w + \nu_2 w = f_4(x) \quad \text{at} \quad y = y_2. \end{aligned} \tag{14.4.4.14}$$

The solution to this problem is the sum of solutions to four simpler auxiliary problems for equation (14.4.4.1), each corresponding to three homogeneous and one nonhomogeneous boundary conditions in (14.4.4.14); see Table 14.5. Each auxiliary problem is solved using the procedure given in Paragraph 14.4.4-1, beginning with the search for solutions in the form of the product of functions with different arguments (14.4.4.3), determined by equations (14.4.1.6) and (14.4.4.4). The separation parameter λ is determined by the solution of a eigenvalue problem with homogeneous boundary conditions; see Table 14.5. The solution to each of the auxiliary problems is sought in the series form (14.4.4.6).

Remark. For the solution of linear nonhomogeneous elliptic equations subject to nonhomogeneous boundary conditions, see Section 14.9.

14.5. Integral Transforms Method

Various integral transforms are widely used to solve linear problems of mathematical physics. The Laplace transform and the Fourier transform are in most common use (its and other integral transforms are considered in Chapter 11 in detail).

14.5.1. Laplace Transform and Its Application in Mathematical Physics

14.5.1-1. Laplace and inverse Laplace transforms. Laplace transforms for derivatives.

The Laplace transform of an arbitrary (complex-valued) function $f(t)$ of a real variable t ($t \ge 0$) is defined by

$$\widetilde{f}(p) = \mathfrak{L}\{f(t)\}, \quad \text{where} \quad \mathfrak{L}\{f(t)\} \equiv \int_0^\infty e^{-pt} f(t)\,dt, \tag{14.5.1.1}$$

where $p = s + i\sigma$ is a complex variable, $i^2 = -1$.

Given the transform $\widetilde{f}(p)$, the function $f(t)$ can be found by means of the inverse Laplace transform

$$f(t) = \mathfrak{L}^{-1}\{\widetilde{f}(p)\}, \quad \text{where} \quad \mathfrak{L}^{-1}\{\widetilde{f}(p)\} \equiv \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \widetilde{f}(p)e^{pt}\,dp, \tag{14.5.1.2}$$

where the integration path is parallel to the imaginary axis and lies to the right of all singularities of $\widetilde{f}(p)$, which corresponds to $c > \sigma_0$.

In order to solve nonstationary boundary value problems, the following Laplace transform formulas for derivatives will be required:

$$\begin{aligned} \mathfrak{L}\{f'(t)\} &= p\widetilde{f}(p) - f(0), \\ \mathfrak{L}\{f''(t)\} &= p^2\widetilde{f}(p) - pf(0) - f'(0), \end{aligned} \tag{14.5.1.3}$$

where $f(0)$ and $f'(0)$ are the initial conditions.

More details on the properties of the Laplace transform and the inverse Laplace transform can be found in Section 11.2. The Laplace transforms of some functions are listed in Section T3.1. Tables of inverse Laplace transforms are listed in Section T3.2. Such tables are convenient to use in solving linear problems for partial differential equations.

14.5.1-2. Solution procedure for linear problems using the Laplace transform.

Figure 14.2 shows schematically how one can utilize the Laplace transforms to solve boundary value problems for linear parabolic or hyperbolic equations with two independent variables in the case where the equation coefficients are independent of t (the same procedure can be applied for solving linear problems characterized by higher-order equations). Here and henceforth, the short notation $\widetilde{w}(x,p) = \mathfrak{L}\{w(x,t)\}$ will be used; the arguments of $\widetilde{w}$ may be omitted.

It is significant that with the Laplace transform, the original problem for a partial differential equation is reduced to a simpler problem for an ordinary differential equation with parameter p; the derivatives with respect to t are replaced by appropriate algebraic expressions taking into account the initial conditions; see formulas (14.5.1.3).

14.5.1-3. Solving linear problems for parabolic equations with the Laplace transform.

Consider a linear nonstationary boundary value problem for the parabolic equation

$$\frac{\partial w}{\partial t} = a(x)\frac{\partial^2 w}{\partial x^2} + b(x)\frac{\partial w}{\partial x} + c(x)w + \Phi(x,t) \tag{14.5.1.4}$$

with the initial condition (14.4.1.3) and the general nonhomogeneous boundary conditions

$$\begin{aligned} s_1\partial_x w + k_1 w &= g_1(t) \quad \text{at} \quad x = x_1, \\ s_2\partial_x w + k_2 w &= g_2(t) \quad \text{at} \quad x = x_2. \end{aligned} \tag{14.5.1.5}$$

The application of the Laplace transform results in the problem defined by the ordinary differential equation for x (p is treated as a parameter)

$$a(x)\frac{\partial^2 \widetilde{w}}{\partial x^2} + b(x)\frac{\partial \widetilde{w}}{\partial x} + [c(x) - p]\widetilde{w} + f_0(x) + \widetilde{\Phi}(x,p) = 0 \tag{14.5.1.6}$$

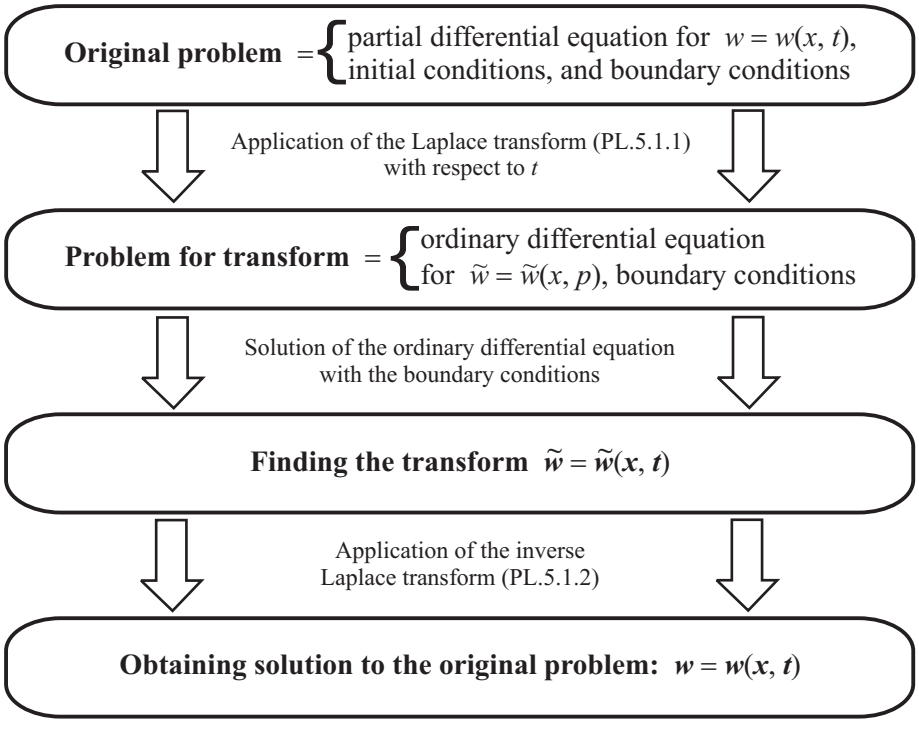


Figure 14.2. Solution procedure for linear boundary value problems using the Laplace transform.

with the boundary conditions

$$\begin{aligned} s_1\partial_x\widetilde{w} + k_1\widetilde{w} &= \widetilde{g}_1(p) \quad \text{at} \quad x = x_1, \\ s_2\partial_x\widetilde{w} + k_2\widetilde{w} &= \widetilde{g}_2(p) \quad \text{at} \quad x = x_2. \end{aligned} \tag{14.5.1.7}$$

Notation employed: $\widetilde{\Phi}(x,p) = \mathfrak{L}\left\{\Phi(x,t)\right\}$ and $\widetilde{g}_n(p) = \mathfrak{L}\left\{g_n(t)\right\}$ $(n = 1, 2)$. On solving problem (14.5.1.6)–(14.5.1.7), one should apply to the resulting solution $\widetilde{w} = \widetilde{w}(x,p)$ the inverse Laplace transform (14.5.1.2) to obtain the solution, $w = w(x,t)$, of the original problem.

Example 1. Consider the first boundary value problem for the heat equation:

$$\begin{aligned} &\frac{\partial w}{\partial t} = \frac{\partial^2 w}{\partial x^2} && (x > 0,\ t > 0), \\ &w = 0 \quad \text{at} \quad t = 0 && \text{(initial condition)}, \\ &w = w_0 \quad \text{at} \quad x = 0 && \text{(boundary condition)}, \\ &w \to 0 \quad \text{at} \quad x \to \infty && \text{(boundary condition)}. \end{aligned}$$

We apply the Laplace transform with respect to t. Let us multiply the equation, the initial condition, and the boundary conditions by e^{-pt} and then integrate with respect to t from zero to infinity. Taking into account the relations

$\mathfrak{L}\{\partial_t w\} = p\widetilde{w} - w|_{t=0} = p\widetilde{w}$ (used are first property (14.5.1.3) and the initial condition),

$\mathfrak{L}\{w_0\} = w_0\mathfrak{L}\{1\} = w_0/p$ (used are property 1 of Subsection T3.1.1 and the relation $\mathfrak{L}\{1\} = 1/p$, see Property 1 of Subsection T3.1.2)

we arrive at the following problem for a second-order linear ordinary differential equation with parameter p:

$$\begin{aligned} &\widetilde{w}''_{xx} - p\widetilde{w} = 0, \\ &\widetilde{w} = w_0/p \quad \text{at} \quad x = 0 && \text{(boundary condition)}, \\ &\widetilde{w} \to 0 \quad \text{at} \quad x \to \infty && \text{(boundary condition)}. \end{aligned}$$

Integrating the equation yields the general solution $\widetilde{w} = A_1(p)e^{-x\sqrt{p}} + A_2(p)e^{x\sqrt{p}}$. Using the boundary conditions, we determine the constants, $A_1(p) = w_0/p$ and $A_2(p) = 0$. Thus, we have

$$\widetilde{w} = \frac{w_0}{p}e^{-x\sqrt{p}}.$$

Let us apply the inverse Laplace transform to both sides of this relation. We refer to the table in Subsection T3.2.5, row 20 (where x must be replaced by t and then a by x^2), to find the inverse transform of the right-hand side. Finally, we obtain the solution of the original problem in the form

$$w = w_0 \operatorname{erfc}\left(\frac{x}{2\sqrt{t}}\right).$$

14.5.1-4. Solving linear problems for hyperbolic equations by the Laplace transform.

Consider a linear nonstationary boundary value problem defined by the hyperbolic equation

$$\frac{\partial^2 w}{\partial t^2} + \varphi(x)\frac{\partial w}{\partial t} = a(x)\frac{\partial^2 w}{\partial x^2} + b(x)\frac{\partial w}{\partial x} + c(x)w + \Phi(x,t) \tag{14.5.1.8}$$

with the initial conditions (14.4.1.3), (14.4.1.4) and the general boundary conditions (14.5.1.5). The application of the Laplace transform results in the problem defined by the ordinary differential equation for x (p is treated as a parameter)

$$a(x)\frac{\partial^2 \widetilde{w}}{\partial x^2} + b(x)\frac{\partial \widetilde{w}}{\partial x} + [c(x) - p^2 - p\varphi(x)]\widetilde{w} + f_0(x)[p + \varphi(x)] + f_1(x) + \widetilde{\Phi}(x,p) = 0 \tag{14.5.1.9}$$

with the boundary conditions (14.5.1.7). On solving this problem, one should apply the inverse Laplace transform to the resulting solution, $\widetilde{w} = \widetilde{w}(x,p)$.

14.5.2. Fourier Transform and Its Application in Mathematical Physics

14.5.2-1. Fourier transform and its properties.

The Fourier transform is defined as follows:

$$\widetilde{f}(u) = \mathfrak{F}\left\{f(x)\right\}, \quad \text{where} \quad \mathfrak{F}\left\{f(x)\right\} \equiv \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(x)e^{-iux}\,dx, \qquad i^2 = -1. \tag{14.5.2.1}$$

This relation is meaningful for any function $f(x)$ absolutely integrable on the interval $(-\infty, \infty)$.

Given $\widetilde{f}(u)$, the function $f(x)$ can be found by means of the inverse Fourier transform

$$f(x) = \mathfrak{F}^{-1}\left\{\widetilde{f}(u)\right\}, \quad \text{where} \quad \mathfrak{F}^{-1}\left\{\widetilde{f}(u)\right\} \equiv \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \widetilde{f}(u)e^{iux}\,du, \tag{14.5.2.2}$$

where the integral is understood in the sense of the Cauchy principal value.

The main properties of the correspondence between functions and their Fourier transforms are gathered in Table 11.4.

14.5.2-2. Solving linear problems of mathematical physics by the Fourier transform.

The Fourier transform is usually employed to solve boundary value problems for linear partial differential equations whose coefficients are independent of the space variable x, $-\infty < x < \infty$.

The scheme for solving linear boundary value problems with the help of the Fourier transform is similar to that used in solving problems with help of the Laplace transform. With the Fourier transform, the derivatives with respect to x in the equation are replaced by appropriate algebraic expressions; see Property 4 or 5 in Table 11.4. In the case of two independent variables, the problem for a partial differential equation is reduced to a simpler problem for an ordinary differential equation with parameter u. On solving the latter problem, one determines the transform. After that, by applying the inverse Fourier transform, one obtains the solution of the original boundary value problem.

Example 2. Consider the following Cauchy problem for the heat equation:

$$\frac{\partial w}{\partial t} = \frac{\partial^2 w}{\partial x^2} \qquad (-\infty < x < \infty),$$
$$w = f(x) \quad \text{at} \quad t = 0 \qquad \text{(initial condition)}.$$

We apply the Fourier transform with respect to the space variable x. Setting $\widetilde{w} = \mathfrak{F}\{w(x,t)\}$ and taking into account the relation $\mathfrak{F}\{\partial_{xx} w\} = -u^2 \widetilde{w}$ (see Property 4 in Table 11.4), we arrive at the following problem for a linear first-order ordinary differential equation in t with parameter u:

$$\widetilde{w}'_t + u^2 \widetilde{w} = 0,$$
$$\widetilde{w} = \widetilde{f}(u) \quad \text{at} \quad t = 0,$$

where $\widetilde{f}(u)$ is defined by (14.5.2.1). On solving this problem for the transform $\widetilde{w}$, we find

$$\widetilde{w} = \widetilde{f}(u) e^{-u^2 t}.$$

Let us apply the inversion formula to both sides of this equation. After some calculations, we obtain the solution of the original problem in the form

$$w = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \widetilde{f}(u) e^{-u^2 t} e^{iux}\, du = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(\xi) e^{-iu\xi} d\xi \right] e^{-u^2 t + iux}\, du$$
$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\xi)\, d\xi \int_{-\infty}^{\infty} e^{-u^2 t + iu(x-\xi)} du = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} f(\xi) \exp\left[-\frac{(x-\xi)^2}{4t} \right] d\xi.$$

At the last stage we used the relation $\displaystyle\int_{-\infty}^{\infty} \exp\left(-a^2 u^2 + bu\right) du = \frac{\sqrt{\pi}}{|a|} \exp\left(\frac{b^2}{4a^2} \right)$.

14.6. Representation of the Solution of the Cauchy Problem via the Fundamental Solution

14.6.1. Cauchy Problem for Parabolic Equations

14.6.1-1. General formula for the solution of the Cauchy problem.

Let $\mathbf{x} = \{x_1, \ldots, x_n\}$ and $\mathbf{y} = \{y_1, \ldots, y_n\}$, where $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$.

Consider a nonhomogeneous linear equation of the parabolic type with an arbitrary right-hand side,

$$\frac{\partial w}{\partial t} - L_{\mathbf{x},t}[w] = \Phi(\mathbf{x}, t), \tag{14.6.1.1}$$

where the second-order linear differential operator $L_{\mathbf{x},t}$ is defined by relation (14.2.1.2).

The solution of the Cauchy problem for equation (14.6.1.1) with an arbitrary initial condition,

$$w = f(\mathbf{x}) \quad \text{at} \quad t = 0,$$

can be represented as the sum of two integrals,

$$w(\mathbf{x}, t) = \int_0^t \int_{\mathbb{R}^n} \Phi(\mathbf{y}, \tau) \mathscr{E}(\mathbf{x}, \mathbf{y}, t, \tau)\, d\mathbf{y}\, d\tau + \int_{\mathbb{R}^n} f(\mathbf{y}) \mathscr{E}(\mathbf{x}, \mathbf{y}, t, 0)\, d\mathbf{y}, \quad d\mathbf{y} = dy_1 \ldots dy_n.$$

Here, $\mathscr{E} = \mathscr{E}(\mathbf{x}, \mathbf{y}, t, \tau)$ is the fundamental solution of the Cauchy problem that satisfies, for $t > \tau \geq 0$, the homogeneous linear equation

$$\frac{\partial \mathscr{E}}{\partial t} - L_{\mathbf{x},t}[\mathscr{E}] = 0 \tag{14.6.1.2}$$

with the nonhomogeneous initial condition of special form

$$\mathscr{E} = \delta(\mathbf{x} - \mathbf{y}) \quad \text{at} \quad t = \tau. \tag{14.6.1.3}$$

The quantities τ and $\mathbf{y}$ appear in problem (14.6.1.2)–(14.6.1.3) as free parameters, and $\delta(\mathbf{x}) = \delta(x_1) \ldots \delta(x_n)$ is the n-dimensional Dirac delta function.

Remark 1. If the coefficients of the differential operator $L_{\mathbf{x},t}$ in (14.6.1.2) are independent of time t, then the fundamental solution of the Cauchy problem depends on only three arguments, $\mathscr{E}(\mathbf{x}, \mathbf{y}, t, \tau) = \mathscr{E}(\mathbf{x}, \mathbf{y}, t - \tau)$.

Remark 2. If the differential operator $L_{\mathbf{x},t}$ has constant coefficients, then the fundamental solution of the Cauchy problem depends on only two arguments, $\mathscr{E}(\mathbf{x}, \mathbf{y}, t, \tau) = \mathscr{E}(\mathbf{x} - \mathbf{y}, t - \tau)$.

14.6.1-2. Fundamental solution allowing incomplete separation of variables.

Consider the special case where the differential operator $L_{\mathbf{x},t}$ in equation (14.6.1.1) can be represented as the sum

$$L_{\mathbf{x},t}[w] = L_{1,t}[w] + \cdots + L_{n,t}[w], \tag{14.6.1.4}$$

where each term depends on a single space coordinate and time,

$$L_{k,t}[w] \equiv a_k(x_k, t) \frac{\partial^2 w}{\partial x_k^2} + b_k(x_k, t) \frac{\partial w}{\partial x_k} + c_k(x_k, t) w, \qquad k = 1, \ldots, n.$$

Equations of this form are often encountered in applications. The fundamental solution of the Cauchy problem for the n-dimensional equation (14.6.1.1) with operator (14.6.1.4) can be represented in the product form

$$\mathscr{E}(\mathbf{x}, \mathbf{y}, t, \tau) = \prod_{k=1}^{n} \mathscr{E}_k(x_k, y_k, t, \tau), \tag{14.6.1.5}$$

where $\mathscr{E}_k = \mathscr{E}_k(x_k, y_k, t, \tau)$ are the fundamental solutions satisfying the one-dimensional equations

$$\frac{\partial \mathscr{E}_k}{\partial t} - L_{k,t}[\mathscr{E}_k] = 0 \qquad (k = 1, \ldots, n)$$

with the initial conditions

$$\mathscr{E}_k = \delta(x_k - y_k) \quad \text{at} \quad t = \tau.$$

In this case, the fundamental solution of the Cauchy problem (14.6.1.5) admits incomplete separation of variables; the fundamental solution is separated in the space variables $x_1, \ldots, x_n$ but not in time t.

Example 1. Consider the two-dimensional heat equation

$$\frac{\partial w}{\partial t} = \frac{\partial^2 w}{\partial x_1^2} + \frac{\partial^2 w}{\partial x_2^2}.$$

The fundamental solutions of the corresponding one-dimensional heat equations are expressed as

Equations		*Fundamental solutions*
$\dfrac{\partial w}{\partial t} = \dfrac{\partial^2 w}{\partial x_1^2}$	$\Longrightarrow$	$\mathscr{E}_1(x_1, y_1, t, \tau) = \dfrac{1}{2\sqrt{\pi(t-\tau)}} \exp\left[-\dfrac{(x_1 - y_1)^2}{4(t-\tau)}\right],$
$\dfrac{\partial w}{\partial t} = \dfrac{\partial^2 w}{\partial x_2^2}$	$\Longrightarrow$	$\mathscr{E}_2(x_2, y_2, t, \tau) = \dfrac{1}{2\sqrt{\pi(t-\tau)}} \exp\left[-\dfrac{(x_2 - y_2)^2}{4(t-\tau)}\right].$

Multiplying $\mathscr{E}_1$ and $\mathscr{E}_2$ together gives the fundamental solution of the two-dimensional heat equation:

$$\mathscr{E}(x_1, x_2, y_1, y_2, t, \tau) = \frac{1}{4\pi(t-\tau)} \exp\left[-\frac{(x_1 - y_1)^2 + (x_2 - y_2)^2}{4(t-\tau)}\right].$$

Example 2. The fundamental solution of the equation

$$\frac{\partial w}{\partial t} = \sum_{k=1}^{n} a_k(t) \frac{\partial^2 w}{\partial x_k^2}, \qquad 0 < a_k(t) < \infty,$$

is given by formula (14.6.1.5) with

$$\mathscr{E}_k(x_k, y_k, t, \tau) = \frac{1}{2\sqrt{\pi T_k}} \exp\left[-\frac{(x_k - y_k)^2}{4T_k}\right], \quad T_k = \int_\tau^t a_k(\eta)\, d\eta.$$

In the derivation of this formula it was taken into account that the corresponding one-dimensional equations could be reduced to the ordinary constant-coefficient heat equation by passing from x_k, t to the new variables x_k, T_k.

14.6.2. Cauchy Problem for Hyperbolic Equations

Consider a nonhomogeneous linear equation of the hyperbolic type with an arbitrary right-hand side,

$$\frac{\partial^2 w}{\partial t^2} + \varphi(\mathbf{x}, t) \frac{\partial w}{\partial t} - L_{\mathbf{x},t}[w] = \Phi(\mathbf{x}, t), \tag{14.6.2.1}$$

where the second-order linear differential operator $L_{\mathbf{x},t}$ is defined by relation (14.2.1.2) with $\mathbf{x} \in \mathbb{R}^n$.

The solution of the Cauchy problem for equation (14.6.2.1) with general initial conditions,

$$w = f_0(\mathbf{x}) \quad \text{at} \quad t = 0,$$
$$\partial_t w = f_1(\mathbf{x}) \quad \text{at} \quad t = 0,$$

can be represented as the sum

$$w(\mathbf{x}, t) = \int_0^t \int_{\mathbb{R}^n} \Phi(\mathbf{y}, \tau) \mathscr{E}(\mathbf{x}, \mathbf{y}, t, \tau)\, d\mathbf{y}\, d\tau - \int_{\mathbb{R}^n} f_0(\mathbf{y}) \left[\frac{\partial}{\partial \tau} \mathscr{E}(\mathbf{x}, \mathbf{y}, t, \tau)\right]_{\tau=0} d\mathbf{y}$$
$$+ \int_{\mathbb{R}^n} \big[f_1(\mathbf{y}) + f_0(\mathbf{y}) \varphi(\mathbf{y}, 0)\big] \mathscr{E}(\mathbf{x}, \mathbf{y}, t, 0)\, d\mathbf{y}, \qquad d\mathbf{y} = dy_1 \ldots dy_n.$$

Here, $\mathscr{E} = \mathscr{E}(\mathbf{x}, \mathbf{y}, t, \tau)$ is the fundamental solution of the Cauchy problem that satisfies, for $t > \tau \geq 0$, the homogeneous linear equation

$$\frac{\partial^2 \mathscr{E}}{\partial t^2} + \varphi(\mathbf{x}, t) \frac{\partial \mathscr{E}}{\partial t} - L_{\mathbf{x},t}[\mathscr{E}] = 0 \tag{14.6.2.2}$$

with the semihomogeneous initial conditions of special form

$$\begin{aligned} \mathscr{E} &= 0 && \text{at} \quad t = \tau, \\ \partial_t \mathscr{E} &= \delta(\mathbf{x} - \mathbf{y}) && \text{at} \quad t = \tau. \end{aligned} \tag{14.6.2.3}$$

The quantities τ and $\mathbf{y}$ appear in problem (14.6.2.2)–(14.6.2.3) as free parameters ($\mathbf{y} \in \mathbb{R}^n$).

Remark 1. If the coefficients of the differential operator $L_{\mathbf{x},t}$ in (14.6.2.2) are independent of time t, then the fundamental solution of the Cauchy problem depends on only three arguments, $\mathscr{E}(\mathbf{x},\mathbf{y},t,\tau)=\mathscr{E}(\mathbf{x},\mathbf{y},t-\tau)$. Here, $\frac{\partial}{\partial\tau}\mathscr{E}(\mathbf{x},\mathbf{y},t,\tau)\big|_{\tau=0}=-\frac{\partial}{\partial t}\mathscr{E}(\mathbf{x},\mathbf{y},t)$.

Remark 2. If the differential operator $L_{\mathbf{x},t}$ has constant coefficients, then the fundamental solution of the Cauchy problem depends on only two arguments, $\mathscr{E}(\mathbf{x},\mathbf{y},t,\tau)=\mathscr{E}(\mathbf{x}-\mathbf{y},t-\tau)$.

Example 1. For the one-, two-, and three-dimensional wave equations, the fundamental solutions have the forms

Equations *Fundamental solutions*

$$\frac{\partial^2 w}{\partial t^2}-\frac{\partial^2 w}{\partial x^2}=0 \quad\Longrightarrow\quad \mathscr{E}(x,t,y,\tau)=\frac{1}{2}\vartheta(t-\tau-|x-y|),\quad \vartheta(z)=\begin{cases}1 & \text{if } z\ge 0,\\ 0 & \text{if } z<0;\end{cases}$$

$$\frac{\partial^2 w}{\partial t^2}-\frac{\partial^2 w}{\partial x_1^2}-\frac{\partial^2 w}{\partial x_2^2}=0 \quad\Longrightarrow\quad \mathscr{E}(x_1,x_2,t,y_1,y_2,\tau)=\frac{\vartheta(t-\tau-\rho)}{2\pi\sqrt{(t-\tau)^2-\rho^2}};$$

$$\frac{\partial^2 w}{\partial t^2}-\frac{\partial^2 w}{\partial x_1^2}-\frac{\partial^2 w}{\partial x_2^2}-\frac{\partial^2 w}{\partial x_3^2}=0 \quad\Longrightarrow\quad \mathscr{E}(x_1,x_2,x_3,t,y_1,y_2,y_3,\tau)=\frac{1}{2\pi}\delta\big((t-\tau)^2-r^2\big),$$

where $\vartheta(z)$ is the Heaviside unit step function ($\vartheta=0$ for $z<0$ and $\vartheta=1$ for $z\ge 0$), $\rho=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2}$, $r=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+(x_3-y_3)^2}$, and $\delta(z)$ is the Dirac delta function.

Example 2. The one-dimensional Klein-Gordon equation

$$\frac{\partial^2 w}{\partial t^2}=\frac{\partial^2 w}{\partial x^2}-bw$$

has the fundamental solutions

$$\mathscr{E}(x,t,y,\tau)=\tfrac{1}{2}\vartheta\big(t-\tau-|x-y|\big)\,J_0\big(k\sqrt{(t-\tau)^2-(x-y)^2}\,\big)\quad\text{for}\quad b=k^2>0,$$

$$\mathscr{E}(x,t,y,\tau)=\tfrac{1}{2}\vartheta\big(t-\tau-|x-y|\big)\,I_0\big(k\sqrt{(t-\tau)^2-(x-y)^2}\,\big)\quad\text{for}\quad b=-k^2<0,$$

where $\vartheta(z)$ is the Heaviside unit step function, $J_0(z)$ is the Bessel function, $I_0(z)$ is the modified Bessel function, and $k>0$.

Example 3. The two-dimensional Klein-Gordon equation

$$\frac{\partial^2 w}{\partial t^2}=\frac{\partial^2 w}{\partial x_1^2}+\frac{\partial^2 w}{\partial x_2^2}-bw$$

has the fundamental solutions

$$\mathscr{E}(x_1,x_2,t,y_1,y_2,\tau)=\vartheta(t-\tau-\rho)\frac{\cos\big(k\sqrt{(t-\tau)^2-\rho^2}\,\big)}{2\pi\sqrt{(t-\tau)^2-\rho^2}}\quad\text{for}\quad b=k^2>0,$$

$$\mathscr{E}(x_1,x_2,t,y_1,y_2,\tau)=\vartheta(t-\tau-\rho)\frac{\cosh\big(k\sqrt{(t-\tau)^2-\rho^2}\,\big)}{2\pi\sqrt{(t-\tau)^2-\rho^2}}\quad\text{for}\quad b=-k^2<0,$$

where $\vartheta(z)$ is the Heaviside unit step function and $\rho=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2}$.

14.7. Boundary Value Problems for Parabolic Equations with One Space Variable. Green's Function

14.7.1. Representation of Solutions via the Green's Function

14.7.1-1. Statement of the problem ($t\ge 0$, $x_1\le x\le x_2$).

In general, a nonhomogeneous linear differential equation of the parabolic type with variable coefficients in one dimension can be written as

$$\frac{\partial w}{\partial t}-L_{x,t}[w]=\Phi(x,t),\qquad(14.7.1.1)$$

where

$$L_{x,t}[w] \equiv a(x,t)\frac{\partial^2 w}{\partial x^2} + b(x,t)\frac{\partial w}{\partial x} + c(x,t)w, \qquad a(x,t) > 0. \tag{14.7.1.2}$$

Consider the nonstationary boundary value problem for equation (14.7.1.1) with an initial condition of general form

$$w = f(x) \quad \text{at} \quad t = 0, \tag{14.7.1.3}$$

and arbitrary nonhomogeneous linear boundary conditions

$$\alpha_1\frac{\partial w}{\partial x} + \beta_1 w = g_1(t) \quad \text{at} \quad x = x_1, \tag{14.7.1.4}$$

$$\alpha_2\frac{\partial w}{\partial x} + \beta_2 w = g_2(t) \quad \text{at} \quad x = x_2. \tag{14.7.1.5}$$

By appropriately choosing the coefficients $\alpha_1, \alpha_2, \beta_1$, and β_2 in (14.7.1.4) and (14.7.1.5), we obtain the first, second, third, and mixed boundary value problems for equation (14.7.1.1).

14.7.1-2. Representation of the problem solution in terms of the Green's function.

The solution of the nonhomogeneous linear boundary value problem (14.7.1.1)–(14.7.1.5) can be represented as

$$w(x,t) = \int_0^t \int_{x_1}^{x_2} \Phi(y,\tau)G(x,y,t,\tau)\,dy\,d\tau + \int_{x_1}^{x_2} f(y)G(x,y,t,0)\,dy + \int_0^t g_1(\tau)a(x_1,\tau)\Lambda_1(x,t,\tau)\,d\tau + \int_0^t g_2(\tau)a(x_2,\tau)\Lambda_2(x,t,\tau)\,d\tau. \tag{14.7.1.6}$$

Here, $G(x,y,t,\tau)$ is the Green's function that satisfies, for $t > \tau \geq 0$, the homogeneous equation

$$\frac{\partial G}{\partial t} - L_{x,t}[G] = 0 \tag{14.7.1.7}$$

with the nonhomogeneous initial condition of special form

$$G = \delta(x-y) \quad \text{at} \quad t = \tau \tag{14.7.1.8}$$

and the homogeneous boundary conditions

$$\alpha_1\frac{\partial G}{\partial x} + \beta_1 G = 0 \quad \text{at} \quad x = x_1, \tag{14.7.1.9}$$

$$\alpha_2\frac{\partial G}{\partial x} + \beta_2 G = 0 \quad \text{at} \quad x = x_2. \tag{14.7.1.10}$$

The quantities y and τ appear in problem (14.7.1.7)–(14.7.1.10) as free parameters, with $x_1 \leq y \leq x_2$, and $\delta(x)$ is the Dirac delta function.

The initial condition (14.7.1.8) implies the limit relation

$$f(x) = \lim_{t\to\tau}\int_{x_1}^{x_2} f(y)G(x,y,t,\tau)\,dy$$

for any continuous function $f = f(x)$.

TABLE 14.6
Expressions of the functions $\Lambda_1(x,t,\tau)$ and $\Lambda_2(x,t,\tau)$ involved in the integrands of the last two terms in solution (14.7.1.6)

Type of problem	Form of boundary conditions	Functions $\Lambda_m(x,t,\tau)$
First boundary value problem ($\alpha_1 = \alpha_2 = 0,\ \beta_1 = \beta_2 = 1$)	$w = g_1(t)$ at $x = x_1$ $w = g_2(t)$ at $x = x_2$	$\Lambda_1(x,t,\tau) = \partial_y G(x,y,t,\tau)\big\|_{y=x_1}$ $\Lambda_2(x,t,\tau) = -\partial_y G(x,y,t,\tau)\big\|_{y=x_2}$
Second boundary value problem ($\alpha_1 = \alpha_2 = 1,\ \beta_1 = \beta_2 = 0$)	$\partial_x w = g_1(t)$ at $x = x_1$ $\partial_x w = g_2(t)$ at $x = x_2$	$\Lambda_1(x,t,\tau) = -G(x,x_1,t,\tau)$ $\Lambda_2(x,t,\tau) = G(x,x_2,t,\tau)$
Third boundary value problem ($\alpha_1 = \alpha_2 = 1,\ \beta_1 < 0,\ \beta_2 > 0$)	$\partial_x w + \beta_1 w = g_1(t)$ at $x = x_1$ $\partial_x w + \beta_2 w = g_2(t)$ at $x = x_2$	$\Lambda_1(x,t,\tau) = -G(x,x_1,t,\tau)$ $\Lambda_2(x,t,\tau) = G(x,x_2,t,\tau)$
Mixed boundary value problem ($\alpha_1 = \beta_2 = 0,\ \alpha_2 = \beta_1 = 1$)	$w = g_1(t)$ at $x = x_1$ $\partial_x w = g_2(t)$ at $x = x_2$	$\Lambda_1(x,t,\tau) = \partial_y G(x,y,t,\tau)\big\|_{y=x_1}$ $\Lambda_2(x,t,\tau) = G(x,x_2,t,\tau)$
Mixed boundary value problem ($\alpha_1 = \beta_2 = 1,\ \alpha_2 = \beta_1 = 0$)	$\partial_x w = g_1(t)$ at $x = x_1$ $w = g_2(t)$ at $x = x_2$	$\Lambda_1(x,t,\tau) = -G(x,x_1,t,\tau)$ $\Lambda_2(x,t,\tau) = -\partial_y G(x,y,t,\tau)\big\|_{y=x_2}$

The functions $\Lambda_1(x,t,\tau)$ and $\Lambda_2(x,t,\tau)$ involved in the integrands of the last two terms in solution (14.7.1.6) can be expressed in terms of the Green's function $G(x,y,t,\tau)$. The corresponding formulas for $\Lambda_m(x,t,\tau)$ are given in Table 14.6 for the basic types of boundary value problems.

It is significant that the Green's function G and the functions Λ_1, Λ_2 are independent of the functions Φ, f, g_1, and g_2 that characterize various nonhomogeneities of the boundary value problem.

If the coefficients of equation (14.7.1.1)–(14.7.1.2) are independent of time t, i.e., the conditions

$$a = a(x), \quad b = b(x), \quad c = c(x) \tag{14.7.1.11}$$

hold, then the Green's function depends on only three arguments,

$$G(x,y,t,\tau) = G(x,y,t-\tau).$$

In this case, the functions Λ_m depend on only two arguments, $\Lambda_m = \Lambda_m(x,t-\tau)$, $m = 1, 2$.

Formula (14.7.1.6) also remains valid for the problem with boundary conditions of the third kind if $\beta_1 = \beta_1(t)$ and $\beta_2 = \beta_2(t)$. Here, the relation between Λ_m ($m = 1, 2$) and the Green's function G is the same as that in the case of constants β_1 and β_2; the Green's function itself is now different.

The condition that the solution must vanish at infinity, $w \to 0$ as $x \to \infty$, is often set for the first, second, and third boundary value problems that are considered on the interval $x_1 \le x < \infty$. In this case, the solution is calculated by formula (14.7.1.6) with $\Lambda_2 = 0$ and Λ_1 specified in Table 14.6.

14.7.2. Problems for Equation $s(x)\dfrac{\partial w}{\partial t} = \dfrac{\partial}{\partial x}\left[p(x)\dfrac{\partial w}{\partial x}\right] - q(x)w + \Phi(x,t)$

14.7.2-1. General formulas for solving nonhomogeneous boundary value problems.

Consider linear equations of the special form

$$s(x)\frac{\partial w}{\partial t} = \frac{\partial}{\partial x}\left[p(x)\frac{\partial w}{\partial x}\right] - q(x)w + \Phi(x,t). \tag{14.7.2.1}$$

TABLE 14.7
Expressions of the functions $\Lambda_1(x,t)$ and $\Lambda_2(x,t)$ involved in the integrands of the last two terms in solutions (14.7.2.2) and (14.8.2.2); the modified Green's function $\mathcal{G}(x,\xi,t)$ for parabolic equations of the form (14.7.2.1) are found by formula (14.7.2.3), and that for hyperbolic equations of the form (14.8.2.1), by formula (14.8.2.3)

Type of problem	Form of boundary conditions	Functions $\Lambda_m(x,t)$
First boundary value problem ($\alpha_1=\alpha_2=0,\ \beta_1=\beta_2=1$)	$w = g_1(t)$ at $x = x_1$ $w = g_2(t)$ at $x = x_2$	$\Lambda_1(x,t) = \partial_\xi \mathcal{G}(x,\xi,t)\big\|_{\xi=x_1}$ $\Lambda_2(x,t) = -\partial_\xi \mathcal{G}(x,\xi,t)\big\|_{\xi=x_2}$
Second boundary value problem ($\alpha_1=\alpha_2=1,\ \beta_1=\beta_2=0$)	$\partial_x w = g_1(t)$ at $x = x_1$ $\partial_x w = g_2(t)$ at $x = x_2$	$\Lambda_1(x,t) = -\mathcal{G}(x,x_1,t)$ $\Lambda_2(x,t) = \mathcal{G}(x,x_2,t)$
Third boundary value problem ($\alpha_1=\alpha_2=1,\ \beta_1<0,\ \beta_2>0$)	$\partial_x w + \beta_1 w = g_1(t)$ at $x = x_1$ $\partial_x w + \beta_2 w = g_2(t)$ at $x = x_2$	$\Lambda_1(x,t) = -\mathcal{G}(x,x_1,t)$ $\Lambda_2(x,t) = \mathcal{G}(x,x_2,t)$
Mixed boundary value problem ($\alpha_1=\beta_2=0,\ \alpha_2=\beta_1=1$)	$w = g_1(t)$ at $x = x_1$ $\partial_x w = g_2(t)$ at $x = x_2$	$\Lambda_1(x,t) = \partial_\xi \mathcal{G}(x,\xi,t)\big\|_{\xi=x_1}$ $\Lambda_2(x,t) = \mathcal{G}(x,x_2,t)$
Mixed boundary value problem ($\alpha_1=\beta_2=1,\ \alpha_2=\beta_1=0$)	$\partial_x w = g_1(t)$ at $x = x_1$ $w = g_2(t)$ at $x = x_2$	$\Lambda_1(x,t) = -\mathcal{G}(x,x_1,t)$ $\Lambda_2(x,t) = -\partial_\xi \mathcal{G}(x,\xi,t)\big\|_{\xi=x_2}$

They are often encountered in heat and mass transfer theory and chemical engineering sciences. Throughout this subsection, we assume that the functions s, p, p'_x, and q are continuous and $s > 0$, $p > 0$, and $x_1 \le x \le x_2$.

The solution of equation (14.7.2.1) under the initial condition (14.7.1.3) and the arbitrary linear nonhomogeneous boundary conditions (14.7.1.4)–(14.7.1.5) can be represented as the sum

$$\begin{aligned} w(x,t) = {} & \int_0^t \int_{x_1}^{x_2} \Phi(\xi,\tau)\mathcal{G}(x,\xi,t-\tau)\,d\xi\,d\tau + \int_{x_1}^{x_2} s(\xi)f(\xi)\mathcal{G}(x,\xi,t)\,d\xi \\ & + p(x_1)\int_0^t g_1(\tau)\Lambda_1(x,t-\tau)\,d\tau + p(x_2)\int_0^t g_2(\tau)\Lambda_2(x,t-\tau)\,d\tau. \end{aligned} \tag{14.7.2.2}$$

Here, the modified Green's function is given by

$$\mathcal{G}(x,\xi,t) = \sum_{n=1}^{\infty} \frac{y_n(x)y_n(\xi)}{\|y_n\|^2}\exp(-\lambda_n t), \qquad \|y_n\|^2 = \int_{x_1}^{x_2} s(x)y_n^2(x)\,dx, \tag{14.7.2.3}$$

where the λ_n and $y_n(x)$ are the eigenvalues and corresponding eigenfunctions of the following Sturm–Liouville problem for a second-order linear ordinary differential equation:

$$\begin{aligned} &[p(x)y'_x]'_x + [\lambda s(x) - q(x)]y = 0, \\ &\alpha_1 y'_x + \beta_1 y = 0 \quad \text{at} \quad x = x_1, \\ &\alpha_2 y'_x + \beta_2 y = 0 \quad \text{at} \quad x = x_2. \end{aligned} \tag{14.7.2.4}$$

The functions $\Lambda_1(x,t)$ and $\Lambda_2(x,t)$ that occur in the integrands of the last two terms in solution (14.7.2.2) are expressed via the Green's function (14.7.2.3). The corresponding formulas for $\Lambda_m(x,t)$ are given in Table 14.7 for the basic types of boundary value problems.

14.7.2-2. Properties of Sturm–Liouville problem (14.7.2.4). Heat equation example.

1°. There are infinitely many eigenvalues. All eigenvalues are real and different and can be ordered so that $\lambda_1 < \lambda_2 < \lambda_3 < \cdots$, with $\lambda_n \to \infty$ as $n \to \infty$ (therefore, there can exist only finitely many negative eigenvalues). Each eigenvalue is of multiplicity 1.

2°. The different eigenfunctions $y_n(x)$ and $y_m(x)$ are orthogonal with weight $s(x)$ on the interval $x_1 \le x \le x_2$:

$$\int_{x_1}^{x_2} s(x)y_n(x)y_m(x)\,dx = 0 \quad \text{for} \quad n \ne m.$$

3°. If the conditions

$$q(x) \ge 0, \quad \alpha_1\beta_1 \le 0, \quad \alpha_2\beta_2 \ge 0 \tag{14.7.2.5}$$

are satisfied, there are no negative eigenvalues. If $q \equiv 0$ and $\beta_1 = \beta_2 = 0$, then $\lambda_1 = 0$ is the least eigenvalue, to which there corresponds the eigenfunction $\varphi_1 = \text{const}$. Otherwise, all eigenvalues are positive, provided that conditions (14.7.2.5) are satisfied.

Other general and special properties of the Sturm–Liouville problem (14.7.2.4) are given in Subsection 12.2.5; various asymptotic and approximate formulas for eigenvalues and eigenfunctions can also be found there.

Example. Consider the first boundary value problem in the domain $0 \le x \le l$ for the heat equation with a source

$$\frac{\partial w}{\partial t} = a\frac{\partial^2 w}{\partial x^2} - bw$$

under the initial condition (14.7.1.3) and boundary conditions

$$\begin{aligned} w &= g_1(t) \quad \text{at} \quad x = 0, \\ w &= g_2(t) \quad \text{at} \quad x = l. \end{aligned} \tag{14.7.2.6}$$

The above equation is a special case of equation (14.7.2.1) with $s(x) = 1$, $p(x) = a$, $q(x) = b$, and $\Phi(x, t) = 0$. The corresponding Sturm–Liouville problem (14.7.2.4) has the form

$$ay''_{xx} + (\lambda - b)y = 0, \qquad y = 0 \quad \text{at} \quad x = 0, \qquad y = 0 \quad \text{at} \quad x = l.$$

The eigenfunctions and eigenvalues are found to be

$$y_n(x) = \sin\left(\frac{\pi n x}{l}\right), \qquad \lambda_n = b + \frac{a\pi^2 n^2}{l^2}, \qquad n = 1, 2, \ldots$$

Using formula (14.7.2.3) and taking into account that $\|y_n\|^2 = l/2$, we obtain the Green's function

$$\mathcal{G}(x, \xi, t) = \frac{2}{l}e^{-bt}\sum_{n=1}^{\infty} \sin\left(\frac{\pi n x}{l}\right)\sin\left(\frac{\pi n \xi}{l}\right)\exp\left(-\frac{a\pi^2 n^2}{l^2}t\right).$$

Substituting this expression into (14.7.2.2) with $p(x_1) = p(x_2) = s(\xi) = 1$, $x_1 = 0$, and $x_2 = l$ and taking into account the formulas

$$\Lambda_1(x, t) = \partial_\xi \mathcal{G}(x, \xi, t)\big|_{\xi = x_1}, \qquad \Lambda_2(x, t) = -\partial_\xi \mathcal{G}(x, \xi, t)\big|_{\xi = x_2}$$

(see the first row in Table 14.7), one obtains the solution to the problem in question.

► *Solutions to various boundary value problems for parabolic equations can be found in Section T8.1.*

14.8. Boundary Value Problems for Hyperbolic Equations with One Space Variable. Green's Function. Goursat Problem

14.8.1. Representation of Solutions via the Green's Function

14.8.1-1. Statement of the problem ($t \ge 0$, $x_1 \le x \le x_2$).

In general, a one-dimensional nonhomogeneous linear differential equation of hyperbolic type with variable coefficients is written as

$$\frac{\partial^2 w}{\partial t^2} + \varphi(x,t)\frac{\partial w}{\partial t} - L_{x,t}[w] = \Phi(x,t), \tag{14.8.1.1}$$

where the operator $L_{x,t}[w]$ is defined by (14.7.1.2).

Consider the nonstationary boundary value problem for equation (14.8.1.1) with the initial conditions

$$\begin{aligned} w &= f_0(x) \quad \text{at} \quad t = 0, \\ \partial_t w &= f_1(x) \quad \text{at} \quad t = 0 \end{aligned} \tag{14.8.1.2}$$

and arbitrary nonhomogeneous linear boundary conditions

$$\alpha_1\frac{\partial w}{\partial x} + \beta_1 w = g_1(t) \quad \text{at} \quad x = x_1, \tag{14.8.1.3}$$

$$\alpha_2\frac{\partial w}{\partial x} + \beta_2 w = g_2(t) \quad \text{at} \quad x = x_2. \tag{14.8.1.4}$$

14.8.1-2. Representation of the problem solution in terms of the Green's function.

The solution of problem (14.8.1.1), (14.8.1.2), (14.8.1.3), (14.8.1.4) can be represented as the sum

$$\begin{aligned} w(x,t) = {} & \int_0^t \int_{x_1}^{x_2} \Phi(y,\tau) G(x,y,t,\tau)\, dy\, d\tau \\ & - \int_{x_1}^{x_2} f_0(y)\left[\frac{\partial}{\partial\tau} G(x,y,t,\tau)\right]_{\tau=0} dy + \int_{x_1}^{x_2} \big[f_1(y) + f_0(y)\varphi(y,0)\big] G(x,y,t,0)\, dy \\ & + \int_0^t g_1(\tau) a(x_1,\tau)\Lambda_1(x,t,\tau)\, d\tau + \int_0^t g_2(\tau) a(x_2,\tau)\Lambda_2(x,t,\tau)\, d\tau. \end{aligned} \tag{14.8.1.5}$$

Here, the Green's function $G(x,y,t,\tau)$ is determined by solving the homogeneous equation

$$\frac{\partial^2 G}{\partial t^2} + \varphi(x,t)\frac{\partial G}{\partial t} - L_{x,t}[G] = 0 \tag{14.8.1.6}$$

with the semihomogeneous initial conditions

$$G = 0 \quad \text{at} \quad t = \tau, \tag{14.8.1.7}$$

$$\partial_t G = \delta(x-y) \quad \text{at} \quad t = \tau, \tag{14.8.1.8}$$

and the homogeneous boundary conditions

$$\alpha_1 \frac{\partial G}{\partial x} + \beta_1 G = 0 \quad \text{at} \quad x = x_1, \tag{14.8.1.9}$$

$$\alpha_2 \frac{\partial G}{\partial x} + \beta_2 G = 0 \quad \text{at} \quad x = x_2. \tag{14.8.1.10}$$

The quantities y and τ appear in problem (14.8.1.6)–(14.8.1.8), (14.8.1.9), (14.8.1.10) as free parameters ($x_1 \le y \le x_2$), and $\delta(x)$ is the Dirac delta function.

The functions $\Lambda_1(x,t,\tau)$ and $\Lambda_2(x,t,\tau)$ involved in the integrands of the last two terms in solution (14.8.1.5) can be expressed via the Green's function $G(x,y,t,\tau)$. The corresponding formulas for $\Lambda_m(x,t,\tau)$ are given in Table 14.6 for the basic types of boundary value problems.

It is significant that the Green's function G and Λ_1, Λ_2 are independent of the functions Φ, f_0, f_1, g_1, and g_2 that characterize various nonhomogeneities of the boundary value problem.

If the coefficients of equation (14.8.1.1) are independent of time t, then the Green's function depends on only three arguments, $G(x,y,t,\tau) = G(x,y,t-\tau)$. In this case, one can set $\frac{\partial}{\partial\tau}G(x,y,t,\tau)\big|_{\tau=0} = -\frac{\partial}{\partial t}G(x,y,t)$ in solution (14.8.1.5).

14.8.2. Problems for Equation $s(x)\dfrac{\partial^2 w}{\partial t^2} = \dfrac{\partial}{\partial x}\left[p(x)\dfrac{\partial w}{\partial x}\right] - q(x)w + \Phi(x,t)$

14.8.2-1. General relations for solving nonhomogeneous boundary value problems.

Consider linear equations of the special form

$$s(x)\frac{\partial^2 w}{\partial t^2} = \frac{\partial}{\partial x}\left[p(x)\frac{\partial w}{\partial x}\right] - q(x)w + \Phi(x,t). \tag{14.8.2.1}$$

It is assumed that the functions s, p, p'_x, and q are continuous and the inequalities $s > 0$, $p > 0$ hold for $x_1 \le x \le x_2$.

The solution of equation (14.8.2.1) under the general initial conditions (14.8.1.2) and the arbitrary linear nonhomogeneous boundary conditions (14.8.1.3)–(14.8.1.4) can be represented as the sum

$$\begin{aligned} w(x,t) = {} & \int_0^t \int_{x_1}^{x_2} \Phi(\xi,\tau)\mathcal{G}(x,\xi,t-\tau)\,d\xi\,d\tau \\ & + \frac{\partial}{\partial t}\int_{x_1}^{x_2} s(\xi)f_0(\xi)\mathcal{G}(x,\xi,t)\,d\xi + \int_{x_1}^{x_2} s(\xi)f_1(\xi)\mathcal{G}(x,\xi,t)\,d\xi \\ & + p(x_1)\int_0^t g_1(\tau)\Lambda_1(x,t-\tau)\,d\tau + p(x_2)\int_0^t g_2(\tau)\Lambda_2(x,t-\tau)\,d\tau. \end{aligned} \tag{14.8.2.2}$$

Here, the modified Green's function is determined by

$$\mathcal{G}(x,\xi,t) = \sum_{n=1}^{\infty} \frac{y_n(x)y_n(\xi)\sin\left(t\sqrt{\lambda_n}\right)}{\|y_n\|^2\sqrt{\lambda_n}}, \qquad \|y_n\|^2 = \int_{x_1}^{x_2} s(x)y_n^2(x)\,dx, \tag{14.8.2.3}$$

where the λ_n and $y_n(x)$ are the eigenvalues and corresponding eigenfunctions of the Sturm–Liouville problem for the second-order linear ordinary differential equation

$$\begin{aligned} &[p(x)y'_x]'_x + [\lambda s(x) - q(x)]y = 0, \\ &\alpha_1 y'_x + \beta_1 y = 0 \quad \text{at} \quad x = x_1, \\ &\alpha_2 y'_x + \beta_2 y = 0 \quad \text{at} \quad x = x_2. \end{aligned} \tag{14.8.2.4}$$

The functions $\Lambda_1(x,t)$ and $\Lambda_2(x,t)$ that occur in the integrands of the last two terms in solution (14.8.2.2) are expressed in terms of the Green's function of (14.8.2.3). The corresponding formulas for $\Lambda_m(x,t)$ are given in Table 14.7 for the basic types of boundary value problems.

14.8.2-2. Properties of the Sturm–Liouville problem. The Klein–Gordon equation.

The general and special properties of the Sturm–Liouville problem (14.8.2.4) are given in Subsection 12.2.5; various asymptotic and approximate formulas for eigenvalues and eigenfunctions can also be found there.

Example. Consider the second boundary value problem in the domain $0 \le x \le l$ for the Klein–Gordon equation

$$\frac{\partial^2 w}{\partial t^2} = a^2 \frac{\partial^2 w}{\partial x^2} - bw,$$

under the initial conditions (14.8.1.2) and boundary conditions

$$\begin{aligned} &\partial_x w = g_1(t) \quad \text{at} \quad x = 0, \\ &\partial_x w = g_2(t) \quad \text{at} \quad x = l. \end{aligned}$$

The Klein–Gordon equation is a special case of equation (14.8.2.1) with $s(x) = 1$, $p(x) = a^2$, $q(x) = b$, and $\Phi(x,t) = 0$. The corresponding Sturm–Liouville problem (14.8.2.4) has the form

$$a^2 y''_{xx} + (\lambda - b)y = 0, \qquad y'_x = 0 \quad \text{at} \quad x = 0, \qquad y'_x = 0 \quad \text{at} \quad x = l.$$

The eigenfunctions and eigenvalues are found to be

$$y_{n+1}(x) = \cos\left(\frac{\pi n x}{l}\right), \qquad \lambda_{n+1} = b + \frac{a\pi^2 n^2}{l^2}, \qquad n = 0,\ 1,\ \ldots$$

Using formula (14.8.2.4) and taking into account that $\|y_1\|^2 = l$ and $\|y_n\|^2 = l/2$ $(n = 1,\ 2,\ \ldots)$, we find the Green's function:

$$\mathcal{G}(x,\xi,t) = \frac{1}{l\sqrt{b}} \sin\bigl(t\sqrt{b}\,\bigr) + \frac{2}{l} \sum_{n=1}^{\infty} \cos\left(\frac{\pi n x}{l}\right) \cos\left(\frac{\pi n \xi}{l}\right) \frac{\sin\bigl(t\sqrt{(a\pi n/l)^2 + b}\,\bigr)}{\sqrt{(a\pi n/l)^2 + b}}.$$

Substituting this expression into (14.8.2.3) with $p(x_1) = p(x_2) = s(\xi) = 1$, $x_1 = 0$, and $x_2 = l$ and taking into account the formulas

$$\Lambda_1(x,t) = -\mathcal{G}(x,x_1,t), \qquad \Lambda_2(x,t) = \mathcal{G}(x,x_2,t)$$

(see the second row in Table 14.7), one obtains the solution to the problem in question.

► *Solutions to various boundary value problems for hyperbolic equations can be found in Section T8.2.*

14.8.3. Problems for Equation $\dfrac{\partial^2 w}{\partial t^2} + a(t)\dfrac{\partial w}{\partial t} = b(t)\left\{\dfrac{\partial}{\partial x}\left[p(x)\dfrac{\partial w}{\partial x}\right] - q(x)w\right\} + \Phi(x,t)$

14.8.3-1. General relations to solve nonhomogeneous boundary value problems.

Consider the generalized telegraph equation of the form

$$\frac{\partial^2 w}{\partial t^2} + a(t)\frac{\partial w}{\partial t} = b(t)\left\{\frac{\partial}{\partial x}\left[p(x)\frac{\partial w}{\partial x}\right] - q(x)w\right\} + \Phi(x,t). \tag{14.8.3.1}$$

It is assumed that the functions p, p'_x, and q are continuous and $p > 0$ for $x_1 \le x \le x_2$.

The solution of equation (14.8.3.1) under the general initial conditions (14.8.1.2) and the arbitrary linear nonhomogeneous boundary conditions (14.8.1.3)–(14.8.1.4) can be represented as the sum

$$\begin{aligned} w(x,t) = {} & \int_0^t \int_{x_1}^{x_2} \Phi(\xi,\tau) G(x,\xi,t,\tau)\, d\xi\, d\tau \\ & - \int_{x_1}^{x_2} f_0(\xi)\left[\frac{\partial}{\partial\tau} G(x,\xi,t,\tau)\right]_{\tau=0} d\xi + \int_{x_1}^{x_2} \big[f_1(\xi) + a(0) f_0(\xi)\big] G(x,\xi,t,0)\, d\xi \\ & + p(x_1)\int_0^t g_1(\tau) b(\tau) \Lambda_1(x,t,\tau)\, d\tau + p(x_2)\int_0^t g_2(\tau) b(\tau) \Lambda_2(x,t,\tau)\, d\tau. \end{aligned} \tag{14.8.3.2}$$

Here, the modified Green's function is determined by

$$G(x,\xi,t,\tau) = \sum_{n=1}^{\infty} \frac{y_n(x) y_n(\xi)}{\|y_n\|^2} U_n(t,\tau), \qquad \|y_n\|^2 = \int_{x_1}^{x_2} y_n^2(x)\, dx, \tag{14.8.3.3}$$

where the λ_n and $y_n(x)$ are the eigenvalues and corresponding eigenfunctions of the Sturm–Liouville problem for the following second-order linear ordinary differential equation with homogeneous boundary conditions:

$$\begin{aligned} & [p(x)y'_x]'_x + [\lambda - q(x)]y = 0, \\ & \alpha_1 y'_x + \beta_1 y = 0 \quad \text{at} \quad x = x_1, \\ & \alpha_2 y'_x + \beta_2 y = 0 \quad \text{at} \quad x = x_2. \end{aligned} \tag{14.8.3.4}$$

The functions $U_n = U_n(t,\tau)$ are determined by solving the Cauchy problem for the linear ordinary differential equation

$$\begin{aligned} & U''_n + a(t)U'_n + \lambda_n b(t) U_n = 0, \\ & U_n\big|_{t=\tau} = 0, \quad U'_n\big|_{t=\tau} = 1. \end{aligned} \tag{14.8.3.5}$$

The prime denotes the derivative with respect to t, and τ is a free parameter occurring in the initial conditions.

The functions $\Lambda_1(x,t)$ and $\Lambda_2(x,t)$ that occur in the integrands of the last two terms in solution (14.8.3.2) are expressed in terms of the Green's function of (14.8.3.3). The corresponding formulas will be specified below when studying specific boundary value problems.

The general and special properties of the Sturm–Liouville problem (14.8.3.4) are detailed in Subsection 12.2.5. Asymptotic and approximate formulas for eigenvalues and eigenfunctions are also presented there.

14.8.3-2. First, second, third, and mixed boundary value problems.

1°. *First boundary value problem.* The solution of equation (14.8.3.1) with the initial conditions (14.8.1.2) and boundary conditions (14.8.1.3)–(14.8.1.4) for $\alpha_1 = \alpha_2 = 0$ and $\beta_1 = \beta_2 = 1$ is given by relations (14.8.3.2) and (14.8.3.3), where

$$\Lambda_1(x,t,\tau) = \frac{\partial}{\partial\xi}G(x,\xi,t,\tau)\Big|_{\xi=x_1}, \quad \Lambda_2(x,t,\tau) = -\frac{\partial}{\partial\xi}G(x,\xi,t,\tau)\Big|_{\xi=x_2}.$$

2°. *Second boundary value problem.* The solution of equation (14.8.3.1) with the initial conditions (14.8.1.2) and boundary conditions (14.8.1.3)–(14.8.1.4) for $\alpha_1 = \alpha_2 = 1$ and $\beta_1 = \beta_2 = 0$ is given by relations (14.8.3.2) and (14.8.3.3) with

$$\Lambda_1(x,t,\tau) = -G(x,x_1,t,\tau), \quad \Lambda_2(x,t,\tau) = G(x,x_2,t,\tau).$$

3°. *Third boundary value problem.* The solution of equation (14.8.3.1) with the initial conditions (14.8.1.2) and boundary conditions (14.8.1.3)–(14.8.1.4) for $\alpha_1 = \alpha_2 = 1$ and $\beta_1\beta_2 \neq 0$ is given by relations (14.8.3.2) and (14.8.3.3) in which

$$\Lambda_1(x,t,\tau) = -G(x,x_1,t,\tau), \quad \Lambda_2(x,t,\tau) = G(x,x_2,t,\tau).$$

4°. *Mixed boundary value problem.* The solution of equation (14.8.3.1) with the initial conditions (14.8.1.2) and boundary conditions (14.8.1.3)–(14.8.1.4) for $\alpha_1 = \beta_2 = 0$ and $\alpha_2 = \beta_1 = 1$ is given by relations (14.8.3.2) and (14.8.3.3) with

$$\Lambda_1(x,t,\tau) = \frac{\partial}{\partial\xi}G(x,\xi,t,\tau)\Big|_{\xi=x_1}, \quad \Lambda_2(x,t,\tau) = G(x,x_2,t,\tau).$$

5°. *Mixed boundary value problem.* The solution of equation (14.8.3.1) with the initial conditions (14.8.1.2) and boundary conditions (14.8.1.3)–(14.8.1.4) for $\alpha_1 = \beta_2 = 1$ and $\alpha_2 = \beta_1 = 0$ is given by relations (14.8.3.2) and (14.8.3.3) with

$$\Lambda_1(x,t,\tau) = -G(x,x_1,t,\tau), \quad \Lambda_2(x,t,\tau) = -\frac{\partial}{\partial\xi}G(x,\xi,t,\tau)\Big|_{\xi=x_2}.$$

14.8.4. Generalized Cauchy Problem with Initial Conditions Set Along a Curve

14.8.4-1. Statement of the generalized Cauchy problem. Basic property of a solution.

Consider the general linear hyperbolic equation in two independent variables which is reduced to the first canonical form (see Paragraph 14.1.1-4):

$$\frac{\partial^2 w}{\partial x\partial y} + a(x,y)\frac{\partial w}{\partial x} + b(x,y)\frac{\partial w}{\partial y} + c(x,y)w = f(x,y), \tag{14.8.4.1}$$

where $a(x,y)$, $b(x,y)$, $c(x,y)$, and $f(x,y)$ are continuous functions.

Let a segment of a curve in the xy-plane be defined by

$$y = \varphi(x) \qquad (\alpha \le x \le \beta), \tag{14.8.4.2}$$

where $\varphi(x)$ is continuously differentiable, with $\varphi'(x) \neq 0$ and $\varphi'(x) \neq \infty$.

The *generalized Cauchy problem* for equation (14.8.4.1) with initial conditions defined along a curve (14.8.4.2) is stated as follows: find a solution to equation (14.8.4.1) that satisfies the conditions

$$w(x,y)\big|_{y=\varphi(x)} = g(x), \qquad \frac{\partial w}{\partial x}\Big|_{y=\varphi(x)} = h_1(x), \qquad \frac{\partial w}{\partial y}\Big|_{y=\varphi(x)} = h_2(x), \tag{14.8.4.3}$$

where $g(x)$, $h_1(x)$, and $h_2(x)$ are given continuous functions, related by the compatibility condition

$$g'_x(x) = h_1(x) + h_2(x)\varphi'_x(x). \tag{14.8.4.4}$$

Basic property of the generalized Cauchy problem: the value of the solution at any point $M(x_0, y_0)$ depends only on the values of the functions $g(x)$, $h_1(x)$, and $h_2(x)$ on the arc AB, cut off on the given curve (14.8.4.2) by the characteristics $x = x_0$ and $y = y_0$, and on the values of $a(x,y)$, $b(x,y)$, $c(x,y)$, and $f(x,y)$ in the curvilinear triangle AMB; see Fig. 14.3. The domain of influence on the solution at $M(x_0, y_0)$ is shaded for clarity.

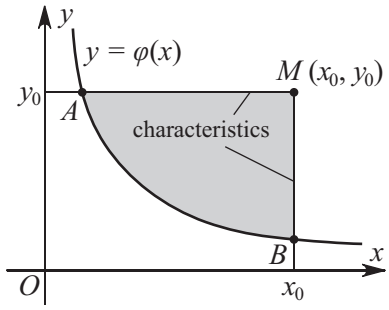


Figure 14.3. Domain of influence of the solution to the generalized Cauchy problem at a point M.

Remark 1. Rather than setting two derivatives in the boundary conditions (14.8.4.3), it suffices to set either of them, with the other being uniquely determined from the compatibility condition (14.8.4.4).

Remark 2. Instead of the last two boundary conditions in (14.8.4.3), the value of the derivative along the normal to the curve (14.8.4.2) can be used:

$$\frac{\partial w}{\partial n}\Big|_{y=\varphi(x)} \equiv \frac{1}{\sqrt{1+[\varphi'_x(x)]^2}}\left[\frac{\partial w}{\partial y} - \varphi'_x(x)\frac{\partial w}{\partial x}\right]_{y=\varphi(x)} = h_3(x). \tag{14.8.4.5}$$

Denoting $w_x\big|_{y=\varphi(x)} = h_1(x)$ and $w_y\big|_{y=\varphi(x)} = h_2(x)$, we have

$$h_2(x) - \varphi'_x(x)h_1(x) = h_3(x)\sqrt{1+[\varphi'_x(x)]^2}. \tag{14.8.4.6}$$

The functions $h_1(x)$ and $h_2(x)$ can be found from (14.8.4.4) and (14.8.4.6). Further substituting their expressions into (14.8.4.3), one arrives at the standard formulation of the generalized Cauchy problem, where the compatibility condition for initial data (14.8.4.4) will be satisfied automatically.

14.8.4-2. Riemann function.

A *Riemann function*, $\mathcal{R} = \mathcal{R}(x, y; x_0, y_0)$, corresponding to equation (14.8.4.1) is defined as a solution to the equation

$$\frac{\partial^2 \mathcal{R}}{\partial x \partial y} - \frac{\partial}{\partial x}\Big[a(x,y)\mathcal{R}\Big] - \frac{\partial}{\partial y}\Big[b(x,y)\mathcal{R}\Big] + c(x,y)\mathcal{R} = 0 \tag{14.8.4.7}$$

that satisfies the conditions

$$\mathcal{R} = \exp\left[\int_{y_0}^{y} a(x_0,\xi)\,d\xi\right] \quad \text{at} \quad x = x_0, \qquad \mathcal{R} = \exp\left[\int_{x_0}^{x} b(\xi,y_0)\,d\xi\right] \quad \text{at} \quad y = y_0 \tag{14.8.4.8}$$

at the characteristics $x = x_0$ and $y = y_0$. Here, (x_0, y_0) is an arbitrary point from the domain of equation (14.8.4.1). The x_0 and y_0 appear in problem (14.8.4.7)–(14.8.4.8) as parameters in the boundary conditions only.

THEOREM. *If the functions a, b, c and the partial derivatives a_x, b_y are all continuous, then the Riemann function $\mathcal{R}(x, y; x_0, y_0)$ exists. Moreover, the function $\mathcal{R}(x_0, y_0, x, y)$, obtained by swapping the parameters and the arguments, is a solution to the homogeneous equation (14.8.4.1), with $f = 0$.*

Remark. It is significant that the Riemann function depends on neither the shape of the curve (14.8.4.2) nor the initial data set on it (14.8.4.3).

Example 1. The Riemann function for the equation $w_{xy} = 0$ is just $\mathcal{R} \equiv 1$.

Example 2. The Riemann function for the equation

$$w_{xy} + cw = 0 \qquad (c = \text{const}) \tag{14.8.4.9}$$

is expressed via the Bessel function $J_0(z)$ as

$$\mathcal{R} = J_0\big(\sqrt{4c(x_0 - x)(y_0 - y)}\big).$$

Remark. Any linear constant-coefficient partial differential equation of the parabolic type in two independent variables can be reduced to an equation of the form (14.8.4.9); see Paragraph 14.1.1-6.

14.8.4-3. Solution of the generalized Cauchy problem via the Riemann function.

Given a Riemann function, the solution to the generalized Cauchy problem (14.8.4.1)–(14.8.4.3) at any point (x_0, y_0) can written as

$$w(x_0,y_0) = \frac{1}{2}(w\mathcal{R})_A + \frac{1}{2}(w\mathcal{R})_B + \frac{1}{2}\int_{AB}\left(\mathcal{R}\frac{\partial w}{\partial x} - w\frac{\partial \mathcal{R}}{\partial x} + 2bw\mathcal{R}\right)dx$$
$$- \frac{1}{2}\int_{AB}\left(\mathcal{R}\frac{\partial w}{\partial y} - w\frac{\partial \mathcal{R}}{\partial y} + 2aw\mathcal{R}\right)dy + \iint_{\Delta AMB} f\mathcal{R}\,dx\,dy.$$

The first two terms on the right-hand side are evaluated as the points A and B. The third and fourth terms are curvilinear integrals over the arc AB; the arc is defined by equation (14.8.4.2), and the integrands involve quantities defined by the initial conditions (14.8.4.3). The last integral is taken over the curvilinear triangular domain AMB.

14.8.5. Goursat Problem (a Problem with Initial Data of Characteristics)

14.8.5-1. Statement of the Goursat problem. Basic property of the solution.

The *Goursat problem* for equation (14.8.4.1) is stated as follows: find a solution to equation (14.8.4.1) that satisfies the conditions at characteristics

$$w(x,y)\big|_{x=x_1} = g(y), \qquad w(x,y)\big|_{y=y_1} = h(x), \tag{14.8.5.1}$$

where $g(y)$ and $h(x)$ are given continuous functions that match each other at the point of intersection of the characteristics, so that

$$g(y_1) = h(x_1).$$

Basic properties of the Goursat problem: the value of the solution at any point $M(x_0, y_0)$ depends only on the values of $g(y)$ at the segment AN (which is part of the characteristic $x = x_1$), the values of $h(x)$ at the segment BN (which is part of the characteristic $y = y_1$), and the values of the functions $a(x, y)$, $b(x, y)$, $c(x, y)$, and $f(x, y)$ in the rectangle $NAMB$; see Fig. 14.4. The domain of influence on the solution at the point $M(x_0, y_0)$ is shaded for clarity.

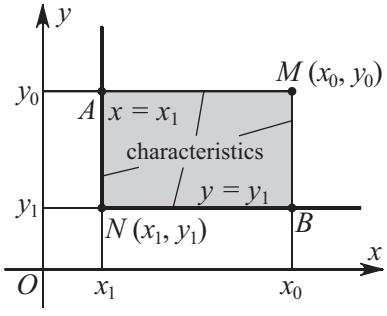


Figure 14.4. Domain of influence of the solution to the Goursat problem at a point M.

14.8.5-2. Solution representation for the Goursat problem via the Riemann function.

Given a Riemann function (see Paragraph 14.8.4-2), the solution to the Goursat problem (14.8.4.1), (14.8.5.1) at any point (x_0, y_0) can be written as

$$w(x_0, y_0) = (w\mathcal{R})_N + \int_N^A \mathcal{R}\left(g'_y + bg\right) dy + \int_N^B \mathcal{R}\left(h'_x + ah\right) dx + \iint_{NAMB} f\mathcal{R}\, dx\, dy.$$

The first term on the right-hand side is evaluated at the point of intersection of the characteristics (x_1, y_1). The second and third terms are integrals along the characteristics $y = y_1$ $(x_1 \le x \le x_0)$ and $x = x_1$ $(y_1 \le y \le y_0)$; these involve the initial data of (14.8.5.1). The last integral is taken over the rectangular domain $NAMB$ defined by the inequalities $x_1 \le x \le x_0$, $y_1 \le y \le y_0$.

The Goursat problem for hyperbolic equations reduced to the second canonical form (see Paragraph 14.1.1-4) is treated similarly.

Example. Consider the Goursat problem for the wave equation

$$\frac{\partial^2 w}{\partial t^2} - a^2 \frac{\partial^2 w}{\partial x^2} = 0$$

with the boundary conditions prescribed on its characteristics

$$\begin{aligned} w &= f(x) \quad \text{for} \quad x - at = 0 \quad (0 \le x \le b), \\ w &= g(x) \quad \text{for} \quad x + at = 0 \quad (0 \le x \le c), \end{aligned} \tag{14.8.5.2}$$

where $f(0) = g(0)$.

Substituting the values set on the characteristics (14.8.5.2) into the general solution of the wave equation, $w = \varphi(x - at) + \psi(x + at)$, we arrive to a system of linear algebraic equations for $\varphi(x)$ and $\psi(x)$. As a result, the solution to the Goursat problem is obtained in the form

$$w(x, t) = f\left(\frac{x + at}{2}\right) + g\left(\frac{x - at}{2}\right) - f(0).$$

The solution propagation domain is the parallelogram bounded by the four lines

$$x - at = 0, \quad x + at = 0, \quad x - at = 2c, \quad x + at = 2b.$$

14.9. Boundary Value Problems for Elliptic Equations with Two Space Variables

14.9.1. Problems and the Green's Functions for Equation $a(x)\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + b(x)\frac{\partial w}{\partial x} + c(x)w = -\Phi(x, y)$

14.9.1-1. Statements of boundary value problems.

Consider two-dimensional boundary value problems for the equation

$$a(x)\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + b(x)\frac{\partial w}{\partial x} + c(x)w = -\Phi(x, y) \tag{14.9.1.1}$$

with general boundary conditions in x,

$$\begin{aligned} \alpha_1\frac{\partial w}{\partial x} - \beta_1 w &= f_1(y) \quad \text{at} \quad x = x_1, \\ \alpha_2\frac{\partial w}{\partial x} + \beta_2 w &= f_2(y) \quad \text{at} \quad x = x_2, \end{aligned} \tag{14.9.1.2}$$

and different boundary conditions in y. It is assumed that the coefficients of equation (14.9.1.1) and the boundary conditions (14.9.1.2) meet the requirements

$$a(x),\ b(x),\ c(x) \text{ are continuous } (x_1 \le x \le x_2); \quad a > 0, \quad |\alpha_1| + |\beta_1| > 0, \quad |\alpha_2| + |\beta_2| > 0.$$

14.9.1-2. Relations for the Green's function.

In the general case, the Green's function can be represented as

$$G(x, y, \xi, \eta) = \rho(\xi)\sum_{n=1}^{\infty}\frac{u_n(x)u_n(\xi)}{\|u_n\|^2}\Psi_n(y, \eta; \lambda_n). \tag{14.9.1.3}$$

Here,

$$\rho(x) = \frac{1}{a(x)}\exp\left[\int\frac{b(x)}{a(x)}\,dx\right], \quad \|u_n\|^2 = \int_{x_1}^{x_2}\rho(x)u_n^2(x)\,dx, \tag{14.9.1.4}$$

and the λ_n and $u_n(x)$ are the eigenvalues and eigenfunctions of the homogeneous boundary value problem for the ordinary differential equation

$$a(x)u''_{xx} + b(x)u'_x + [\lambda + c(x)]u = 0, \tag{14.9.1.5}$$

$$\alpha_1 u'_x - \beta_1 u = 0 \quad \text{at} \quad x = x_1, \tag{14.9.1.6}$$

$$\alpha_2 u'_x + \beta_2 u = 0 \quad \text{at} \quad x = x_2. \tag{14.9.1.7}$$

The functions Ψ_n for various boundary conditions in y are specified in Table 14.8.

Equation (14.9.1.5) can be rewritten in self-adjoint form as

$$[p(x)u'_x]'_x + [\lambda\rho(x) - q(x)]u = 0, \tag{14.9.1.8}$$

TABLE 14.8
The functions Ψ_n in (14.9.1.3) for various boundary conditions.* Notation: $\sigma_n = \sqrt{\lambda_n}$

Domain	Boundary conditions	Function $\Psi_n(y,\eta;\lambda_n)$
$-\infty < y < \infty$	$\lvert w\rvert < \infty$ for $y \to \pm\infty$	$\dfrac{1}{2\sigma_n} e^{-\sigma_n\lvert y-\eta\rvert}$
$0 \le y < \infty$	$w = 0$ for $y = 0$	$\dfrac{1}{\sigma_n}\begin{cases} e^{-\sigma_n y}\sinh(\sigma_n\eta) & \text{for } y > \eta, \\ e^{-\sigma_n \eta}\sinh(\sigma_n y) & \text{for } \eta > y \end{cases}$
$0 \le y < \infty$	$\partial_y w = 0$ for $y = 0$	$\dfrac{1}{\sigma_n}\begin{cases} e^{-\sigma_n y}\cosh(\sigma_n\eta) & \text{for } y > \eta, \\ e^{-\sigma_n \eta}\cosh(\sigma_n y) & \text{for } \eta > y \end{cases}$
$0 \le y < \infty$	$\partial_y w - \beta_3 w = 0$ for $y = 0$	$\dfrac{1}{\sigma_n(\sigma_n+\beta_3)}\begin{cases} e^{-\sigma_n y}[\sigma_n\cosh(\sigma_n\eta)+\beta_3\sinh(\sigma_n\eta)] & \text{for } y > \eta, \\ e^{-\sigma_n \eta}[\sigma_n\cosh(\sigma_n y)+\beta_3\sinh(\sigma_n y)] & \text{for } \eta > y \end{cases}$
$0 \le y \le h$	$w = 0$ at $y = 0$, $w = 0$ at $y = h$	$\dfrac{1}{\sigma_n\sinh(\sigma_n h)}\begin{cases} \sinh(\sigma_n\eta)\sinh[\sigma_n(h-y)] & \text{for } y > \eta, \\ \sinh(\sigma_n y)\sinh[\sigma_n(h-\eta)] & \text{for } \eta > y \end{cases}$
$0 \le y \le h$	$\partial_y w = 0$ at $y = 0$, $\partial_y w = 0$ at $y = h$	$\dfrac{1}{\sigma_n\sinh(\sigma_n h)}\begin{cases} \cosh(\sigma_n\eta)\cosh[\sigma_n(h-y)] & \text{for } y > \eta, \\ \cosh(\sigma_n y)\cosh[\sigma_n(h-\eta)] & \text{for } \eta > y \end{cases}$
$0 \le y \le h$	$w = 0$ at $y = 0$, $\partial_y w = 0$ at $y = h$	$\dfrac{1}{\sigma_n\cosh(\sigma_n h)}\begin{cases} \sinh(\sigma_n\eta)\cosh[\sigma_n(h-y)] & \text{for } y > \eta, \\ \sinh(\sigma_n y)\cosh[\sigma_n(h-\eta)] & \text{for } \eta > y \end{cases}$

where the functions $p(x)$ and $q(x)$ are given by

$$p(x) = \exp\left[\int \frac{b(x)}{a(x)}\,dx\right], \quad q(x) = -\frac{c(x)}{a(x)}\exp\left[\int \frac{b(x)}{a(x)}\,dx\right],$$

and $\rho(x)$ is defined in (14.9.1.4).

The eigenvalue problem (14.9.1.8), (14.9.1.6), (14.9.1.7) possesses the following properties:

1°. All eigenvalues $\lambda_1, \lambda_2, \ldots$ are real and $\lambda_n \to \infty$ as $n \to \infty$.

2°. The system of eigenfunctions $\{u_1(x), u_2(x), \ldots\}$ is orthogonal on the interval $x_1 \le x \le x_2$ with weight $\rho(x)$, that is,

$$\int_{x_1}^{x_2} \rho(x)u_n(x)u_m(x)\,dx = 0 \quad \text{for} \quad n \ne m.$$

3°. If the conditions

$$q(x) \ge 0, \quad \alpha_1\beta_1 \ge 0, \quad \alpha_2\beta_2 \ge 0 \tag{14.9.1.9}$$

are satisfied, there are no negative eigenvalues. If $q \equiv 0$ and $\beta_1 = \beta_2 = 0$, then the least eigenvalue is $\lambda_0 = 0$ and the corresponding eigenfunction is $u_0 = \text{const}$; in this case, the summation in (14.9.1.3) must start with $n = 0$. In the other cases, if conditions (14.9.1.9) are satisfied, all eigenvalues are positive; for example, the first inequality in (14.9.1.9) holds if $c(x) \le 0$.

Subsection 12.2.5 presents some relations for estimating the eigenvalues λ_n and eigenfunctions $u_n(x)$.

* For unbounded domains, the condition of boundedness of the solution as $y \to \pm\infty$ is set; in Table 14.8, this condition is omitted.

Example. Consider a boundary value problem for the Laplace equation

$$\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} = 0$$

in a strip $0 \le x \le l$, $-\infty < y < \infty$ with mixed boundary conditions

$$w = f_1(y) \quad \text{at} \quad x = 0, \qquad \frac{\partial w}{\partial x} = f_2(y) \quad \text{at} \quad x = l.$$

This equation is a special case of equation (14.9.1.1) with $a(x) = 1$ and $b(x) = c(x) = \Phi(x, t) = 0$. The corresponding Sturm–Liouville problem (14.9.1.5)–(14.9.1.7) is written as

$$u''_{xx} + \lambda y = 0, \qquad u = 0 \quad \text{at} \quad x = 0, \qquad u'_x = 0 \quad \text{at} \quad x = l.$$

The eigenfunctions and eigenvalues are found as

$$u_n(x) = \sin\left[\frac{\pi(2n-1)x}{l}\right], \quad \lambda_n = \frac{\pi^2(2n-1)^2}{l^2}, \quad n = 1, 2, \ldots$$

Using formulas (14.9.1.3) and (14.9.1.4) and taking into account the identities $\rho(\xi) = 1$ and $\|y_n\|^2 = l/2$ ($n = 1, 2, \ldots$) and the expression for Ψ_n from the first row in Table 14.8, we obtain the Green's function in the form

$$G(x, y, \xi, \eta) = \frac{1}{l}\sum_{n=1}^{\infty} \frac{1}{\sigma_n} \sin(\sigma_n x)\sin(\sigma_n \xi) e^{-\sigma_n |y-\eta|}, \quad \sigma_n = \sqrt{\lambda_n} = \frac{\pi(2n-1)}{l}.$$

14.9.2. Representation of Solutions to Boundary Value Problems via the Green's Functions

14.9.2-1. First boundary value problem.

The solution of the first boundary value problem for equation (14.9.1.1) with the boundary conditions

$$w = f_1(y) \quad \text{at} \quad x = x_1, \qquad w = f_2(y) \quad \text{at} \quad x = x_2,$$
$$w = f_3(x) \quad \text{at} \quad y = 0, \qquad w = f_4(x) \quad \text{at} \quad y = h$$

is expressed in terms of the Green's function as

$$\begin{aligned} w(x, y) = {} & a(x_1)\int_0^h f_1(\eta)\left[\frac{\partial}{\partial \xi} G(x, y, \xi, \eta)\right]_{\xi = x_1} d\eta - a(x_2)\int_0^h f_2(\eta)\left[\frac{\partial}{\partial \xi} G(x, y, \xi, \eta)\right]_{\xi = x_2} d\eta \\ & + \int_{x_1}^{x_2} f_3(\xi)\left[\frac{\partial}{\partial \eta} G(x, y, \xi, \eta)\right]_{\eta = 0} d\xi - \int_{x_1}^{x_2} f_4(\xi)\left[\frac{\partial}{\partial \eta} G(x, y, \xi, \eta)\right]_{\eta = h} d\xi \\ & + \int_{x_1}^{x_2}\int_0^h \Phi(\xi, \eta) G(x, y, \xi, \eta)\, d\eta\, d\xi. \end{aligned}$$

14.9.2-2. Second boundary value problem.

The solution of the second boundary value problem for equation (14.9.1.1) with boundary conditions

$$\partial_x w = f_1(y) \quad \text{at} \quad x = x_1, \qquad \partial_x w = f_2(y) \quad \text{at} \quad x = x_2,$$
$$\partial_y w = f_3(x) \quad \text{at} \quad y = 0, \qquad \partial_y w = f_4(x) \quad \text{at} \quad y = h$$

is expressed in terms of the Green's function as

$$w(x,y) = -a(x_1)\int_0^h f_1(\eta)G(x,y,x_1,\eta)\,d\eta + a(x_2)\int_0^h f_2(\eta)G(x,y,x_2,\eta)\,d\eta$$
$$-\int_{x_1}^{x_2} f_3(\xi)G(x,y,\xi,0)\,d\xi + \int_{x_1}^{x_2} f_4(\xi)G(x,y,\xi,h)\,d\xi$$
$$+\int_{x_1}^{x_2}\int_0^h \Phi(\xi,\eta)G(x,y,\xi,\eta)\,d\eta\,d\xi.$$

14.9.2-3. Third boundary value problem.

The solution of the third boundary value problem for equation (14.9.1.1) in terms of the Green's function is represented in the same way as the solution of the second boundary value problem (the Green's function is now different).

► *Solutions of various boundary value problems for elliptic equations can be found in Section T8.3.*

14.10. Boundary Value Problems with Many Space Variables. Representation of Solutions via the Green's Function

14.10.1. Problems for Parabolic Equations

14.10.1-1. Statement of the problem.

In general, a nonhomogeneous linear differential equation of the parabolic type in n space variables has the form

$$\frac{\partial w}{\partial t} - L_{\mathbf{x},t}[w] = \Phi(\mathbf{x},t), \tag{14.10.1.1}$$

where

$$L_{\mathbf{x},t}[w] \equiv \sum_{i,j=1}^{n} a_{ij}(\mathbf{x},t)\frac{\partial^2 w}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(\mathbf{x},t)\frac{\partial w}{\partial x_i} + c(\mathbf{x},t)w,$$
$$\mathbf{x} = \{x_1,\ldots,x_n\}, \quad \sum_{i,j=1}^{n} a_{ij}(\mathbf{x},t)\xi_i\xi_j \ge \sigma\sum_{i=1}^{n}\xi_i^2, \quad \sigma > 0. \tag{14.10.1.2}$$

Let V be some simply connected domain in $\mathbb{R}^n$ with a sufficiently smooth boundary $S = \partial V$. We consider the nonstationary boundary value problem for equation (14.10.1.1) in the domain V with an arbitrary initial condition,

$$w = f(\mathbf{x}) \quad \text{at} \quad t = 0, \tag{14.10.1.3}$$

and nonhomogeneous linear boundary conditions,

$$\Gamma_{\mathbf{x},t}[w] = g(\mathbf{x},t) \quad \text{for} \quad \mathbf{x} \in S. \tag{14.10.1.4}$$

In the general case, $\Gamma_{\mathbf{x},t}$ is a first-order linear differential operator in the space coordinates with coefficients dependent on $\mathbf{x}$ and t.

14.10.1-2. Representation of the problem solution in terms of the Green's function.

The solution of the nonhomogeneous linear boundary value problem defined by (14.10.1.1)–(14.10.1.4) can be represented as the sum

$$w(\mathbf{x},t)=\int_0^t\int_V \Phi(\mathbf{y},\tau)G(\mathbf{x},\mathbf{y},t,\tau)\,dV_y\,d\tau+\int_V f(\mathbf{y})G(\mathbf{x},\mathbf{y},t,0)\,dV_y$$
$$+\int_0^t\int_S g(\mathbf{y},\tau)H(\mathbf{x},\mathbf{y},t,\tau)\,dS_y\,d\tau, \tag{14.10.1.5}$$

where $G(\mathbf{x},\mathbf{y},t,\tau)$ is the Green's function; for $t>\tau\ge 0$, it satisfies the homogeneous equation

$$\frac{\partial G}{\partial t}-L_{\mathbf{x},t}[G]=0 \tag{14.10.1.6}$$

with the nonhomogeneous initial condition of special form

$$G=\delta(\mathbf{x}-\mathbf{y})\quad\text{at}\quad t=\tau \tag{14.10.1.7}$$

and the homogeneous boundary condition

$$\Gamma_{\mathbf{x},t}[G]=0\quad\text{for}\quad \mathbf{x}\in S. \tag{14.10.1.8}$$

The vector $\mathbf{y}=\{y_1,\dots,y_n\}$ appears in problem (14.10.1.6)–(14.10.1.8) as an n-dimensional free parameter ($\mathbf{y}\in V$), and $\delta(\mathbf{x}-\mathbf{y})=\delta(x_1-y_1)\dots\delta(x_n-y_n)$ is the n-dimensional Dirac delta function. The Green's function G is independent of the functions Φ, f, and g that characterize various nonhomogeneities of the boundary value problem. In (14.10.1.5), the integration is performed everywhere with respect to $\mathbf{y}$, with $dV_y=dy_1\dots dy_n$.

The function $H(\mathbf{x},\mathbf{y},t,\tau)$ involved in the integrand of the last term in solution (14.10.1.5) can be expressed via the Green's function $G(\mathbf{x},\mathbf{y},t,\tau)$. The corresponding formulas for $H(\mathbf{x},\mathbf{y},t,\tau)$ are given in Table 14.9 for the three basic types of boundary value problems; in the third boundary value problem, the coefficient k can depend on $\mathbf{x}$ and t. The boundary conditions of the second and third kind, as well as the solution of the first boundary value problem, involve operators of differentiation along the conormal of operator (14.10.1.2); these operators act as follows:

$$\frac{\partial G}{\partial M_x}\equiv\sum_{i,j=1}^n a_{ij}(\mathbf{x},t)N_j\frac{\partial G}{\partial x_i},\qquad \frac{\partial G}{\partial M_y}\equiv\sum_{i,j=1}^n a_{ij}(\mathbf{y},\tau)N_j\frac{\partial G}{\partial y_i}, \tag{14.10.1.9}$$

where $\mathbf{N}=\{N_1,\dots,N_n\}$ is the unit outward normal to the surface S. In the special case where $a_{ii}(\mathbf{x},t)=1$ and $a_{ij}(\mathbf{x},t)=0$ for $i\ne j$, operator (14.10.1.9) coincides with the ordinary operator of differentiation along the outward normal to S.

TABLE 14.9
The form of the function $H(\mathbf{x},\mathbf{y},t,\tau)$ for the basic types of nonstationary boundary value problems

Type of problem	Form of boundary condition (14.10.1.4)	Function $H(\mathbf{x},\mathbf{y},t,\tau)$
First boundary value problem	$w=g(\mathbf{x},t)$ for $\mathbf{x}\in S$	$H(\mathbf{x},\mathbf{y},t,\tau)=-\dfrac{\partial G}{\partial M_y}(\mathbf{x},\mathbf{y},t,\tau)$
Second boundary value problem	$\dfrac{\partial w}{\partial M_x}=g(\mathbf{x},t)$ for $\mathbf{x}\in S$	$H(\mathbf{x},\mathbf{y},t,\tau)=G(\mathbf{x},\mathbf{y},t,\tau)$
Third boundary value problem	$\dfrac{\partial w}{\partial M_x}+kw=g(\mathbf{x},t)$ for $\mathbf{x}\in S$	$H(\mathbf{x},\mathbf{y},t,\tau)=G(\mathbf{x},\mathbf{y},t,\tau)$

If the coefficient of equation (14.10.1.6) and the boundary condition (14.10.1.8) are independent of t, then the Green's function depends on only three arguments, $G(\mathbf{x}, \mathbf{y}, t, \tau) = G(\mathbf{x}, \mathbf{y}, t - \tau)$.

Remark. Let S_i $(i = 1, \ldots, p)$ be different portions of the surface S such that $S = \sum_{i=1}^{p} S_i$ and let boundary conditions of various types be set on the S_i,

$$\Gamma^{(i)}_{\mathbf{x},t}[w] = g_i(\mathbf{x}, t) \quad \text{for} \quad \mathbf{x} \in S_i, \qquad i = 1, \ldots, p. \tag{14.10.1.10}$$

Then formula (14.10.1.5) remains valid but the last term in (14.10.1.5) must be replaced by the sum

$$\sum_{i=1}^{p} \int_0^t \int_{S_i} g_i(\mathbf{y}, \tau) H_i(\mathbf{x}, \mathbf{y}, t, \tau)\, dS_y\, d\tau. \tag{14.10.1.11}$$

14.10.2. Problems for Hyperbolic Equations

14.10.2-1. Statement of the problem.

The general nonhomogeneous linear differential hyperbolic equation in n space variables can be written as

$$\frac{\partial^2 w}{\partial t^2} + \varphi(\mathbf{x}, t)\frac{\partial w}{\partial t} - L_{\mathbf{x},t}[w] = \Phi(\mathbf{x}, t), \tag{14.10.2.1}$$

where the operator $L_{\mathbf{x},t}[w]$ is explicitly defined in (14.10.1.2).

We consider the nonstationary boundary value problem for equation (14.10.2.1) in the domain V with arbitrary initial conditions,

$$w = f_0(\mathbf{x}) \quad \text{at} \quad t = 0, \tag{14.10.2.2}$$

$$\partial_t w = f_1(\mathbf{x}) \quad \text{at} \quad t = 0, \tag{14.10.2.3}$$

and the nonhomogeneous linear boundary condition (14.10.1.4).

14.10.2-2. Representation of the problem solution in terms of the Green's function.

The solution of the nonhomogeneous linear boundary value problem defined by (14.10.2.1)–(14.10.2.3), (14.10.1.4) can be represented as the sum

$$\begin{aligned} w(\mathbf{x}, t) = {} & \int_0^t \int_V \Phi(\mathbf{y}, \tau) G(\mathbf{x}, \mathbf{y}, t, \tau)\, dV_y\, d\tau - \int_V f_0(\mathbf{y}) \left[\frac{\partial}{\partial \tau} G(\mathbf{x}, \mathbf{y}, t, \tau)\right]_{\tau=0} dV_y \\ & + \int_V \big[f_1(\mathbf{y}) + f_0(\mathbf{y})\varphi(\mathbf{y}, 0)\big] G(\mathbf{x}, \mathbf{y}, t, 0)\, dV_y \\ & + \int_0^t \int_S g(\mathbf{y}, \tau) H(\mathbf{x}, \mathbf{y}, t, \tau)\, dS_y\, d\tau. \end{aligned} \tag{14.10.2.4}$$

Here, $G(\mathbf{x}, \mathbf{y}, t, \tau)$ is the Green's function; for $t > \tau \geq 0$ it satisfies the homogeneous equation

$$\frac{\partial^2 G}{\partial t^2} + \varphi(\mathbf{x}, t)\frac{\partial G}{\partial t} - L_{\mathbf{x},t}[G] = 0 \tag{14.10.2.5}$$

with the semihomogeneous initial conditions

$$G = 0 \quad \text{at} \quad t = \tau,$$

$$\partial_t G = \delta(\mathbf{x} - \mathbf{y}) \quad \text{at} \quad t = \tau,$$

and the homogeneous boundary condition (14.10.1.8).

If the coefficients of equation (14.10.2.5) and the boundary condition (14.10.1.8) are independent of time t, then the Green's function depends on only three arguments, $G(\mathbf{x}, \mathbf{y}, t, \tau) = G(\mathbf{x}, \mathbf{y}, t - \tau)$. In this case, one can set $\frac{\partial}{\partial \tau} G(\mathbf{x}, \mathbf{y}, t, \tau)\Big|_{\tau=0} = -\frac{\partial}{\partial t} G(\mathbf{x}, \mathbf{y}, t)$ in solution (14.10.2.4).

The function $H(\mathbf{x}, \mathbf{y}, t, \tau)$ involved in the integrand of the last term in solution (14.10.2.4) can be expressed via the Green's function $G(\mathbf{x}, \mathbf{y}, t, \tau)$. The corresponding formulas for H are given in Table 14.9 for the three basic types of boundary value problems; in the third boundary value problem, the coefficient k can depend on $\mathbf{x}$ and t.

Remark. Let S_i ($i = 1, \ldots, p$) be different portions of the surface S such that $S = \sum_{i=1}^{p} S_i$ and let boundary conditions of various types (14.10.1.10) be set on the S_i. Then formula (14.10.2.4) remains valid but the last term in (14.10.2.4) must be replaced by the sum (14.10.1.11).

14.10.3. Problems for Elliptic Equations

14.10.3-1. Statement of the problem.

In general, a nonhomogeneous linear elliptic equation can be written as

$$-L_{\mathbf{x}}[w] = \Phi(\mathbf{x}), \tag{14.10.3.1}$$

where

$$L_{\mathbf{x}}[w] \equiv \sum_{i,j=1}^{n} a_{ij}(\mathbf{x}) \frac{\partial^2 w}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(\mathbf{x}) \frac{\partial w}{\partial x_i} + c(\mathbf{x}) w. \tag{14.10.3.2}$$

Two-dimensional problems correspond to $n = 2$ and three-dimensional problems to $n = 3$.

We consider equation (14.10.3.1)–(14.10.3.2) in a domain V and assume that the equation is subject to the general linear boundary condition

$$\Gamma_{\mathbf{x}}[w] = g(\mathbf{x}) \quad \text{for} \quad \mathbf{x} \in S. \tag{14.10.3.3}$$

The solution of the stationary problem (14.10.3.1)–(14.10.3.3) can be obtained by passing in (14.10.1.5) to the limit as $t \to \infty$. To this end, one should start with equation (14.10.1.1), whose coefficients are independent of t, and take the homogeneous initial condition (14.10.1.3), with $f(\mathbf{x}) = 0$, and the stationary boundary condition (14.10.1.4).

14.10.3-2. Representation of the problem solution in terms of the Green's function.

The solution of the linear boundary value problem (14.10.3.1)–(14.10.3.3) can be represented as the sum

$$w(\mathbf{x}) = \int_V \Phi(\mathbf{y}) G(\mathbf{x}, \mathbf{y})\, dV_y + \int_S g(\mathbf{y}) H(\mathbf{x}, \mathbf{y})\, dS_y. \tag{14.10.3.4}$$

Here, the Green's function $G(\mathbf{x}, \mathbf{y})$ satisfies the nonhomogeneous equation of special form

$$-L_{\mathbf{x}}[G] = \delta(\mathbf{x} - \mathbf{y}) \tag{14.10.3.5}$$

with the homogeneous boundary condition

$$\Gamma_{\mathbf{x}}[G] = 0 \quad \text{for} \quad \mathbf{x} \in S. \tag{14.10.3.6}$$

The vector $\mathbf{y} = \{y_1, \ldots, y_n\}$ appears in problem (14.10.3.5), (14.10.3.6) as an n-dimensional free parameter ($\mathbf{y} \in V$). Note that G is independent of the functions Φ and g characterizing various nonhomogeneities of the original boundary value problem.

The function $H(\mathbf{x}, \mathbf{y})$ involved in the integrand of the second term in solution (14.10.3.4) can be expressed via the Green's function $G(\mathbf{x}, \mathbf{y})$. The corresponding formulas for H are given in Table 14.10 for the three basic types of boundary value problems. The boundary conditions of the second and third kind, as well as the solution of the first boundary value problem, involve operators of differentiation along the conormal of operator (14.10.3.2); these operators are defined by (14.10.1.9); in this case, the coefficients a_{ij} depend on $\mathbf{x}$ only.

TABLE 14.10
The form of the function $H(\mathbf{x}, \mathbf{y})$ involved in the integrand of the last term in solution (14.10.3.4) for the basic types of stationary boundary value problems

Type of problem	Form of boundary condition (14.10.3.3)	Function $H(\mathbf{x}, \mathbf{y})$
First boundary value problem	$w = g(\mathbf{x})$ for $\mathbf{x} \in S$	$H(\mathbf{x}, \mathbf{y}) = -\dfrac{\partial G}{\partial M_y}(\mathbf{x}, \mathbf{y})$
Second boundary value problem	$\dfrac{\partial w}{\partial M_x} = g(\mathbf{x})$ for $\mathbf{x} \in S$	$H(\mathbf{x}, \mathbf{y}) = G(\mathbf{x}, \mathbf{y})$
Third boundary value problem	$\dfrac{\partial w}{\partial M_x} + kw = g(\mathbf{x})$ for $\mathbf{x} \in S$	$H(\mathbf{x}, \mathbf{y}) = G(\mathbf{x}, \mathbf{y})$

Remark. For the second boundary value problem with $c(\mathbf{x}) \equiv 0$, the thus defined Green's function must not necessarily exist; see Polyanin (2002).

14.10.4. Comparison of the Solution Structures for Boundary Value Problems for Equations of Various Types

Table 14.11 lists brief formulations of boundary value problems for second-order equations of elliptic, parabolic, and hyperbolic types. The coefficients of the differential operators $L_{\mathbf{x}}$ and $\Gamma_{\mathbf{x}}$ in the space variables $x_1, \ldots, x_n$ are assumed to be independent of time t; these operators are the same for the problems under consideration.

TABLE 14.11
Formulations of boundary value problems for equations of various types

Type of equation	Form of equation	Initial conditions	Boundary conditions
Elliptic	$-L_{\mathbf{x}}[w] = \Phi(\mathbf{x})$	not set	$\Gamma_{\mathbf{x}}[w] = g(\mathbf{x})$ for $\mathbf{x} \in S$
Parabolic	$\partial_t w - L_{\mathbf{x}}[w] = \Phi(\mathbf{x}, t)$	$w = f(\mathbf{x})$ at $t = 0$	$\Gamma_{\mathbf{x}}[w] = g(\mathbf{x}, t)$ for $\mathbf{x} \in S$
Hyperbolic	$\partial_{tt} w - L_{\mathbf{x}}[w] = \Phi(\mathbf{x}, t)$	$w = f_0(\mathbf{x})$ at $t = 0$, $\partial_t w = f_1(\mathbf{x})$ at $t = 0$	$\Gamma_{\mathbf{x}}[w] = g(\mathbf{x}, t)$ for $\mathbf{x} \in S$

Below are the respective general formulas defining the solutions of these problems with

zero initial conditions ($f = f_0 = f_1 = 0$):

$$w_0(\mathbf{x}) = \int_V \Phi(\mathbf{y})G_0(\mathbf{x},\mathbf{y})\,dV_y + \int_S g(\mathbf{y})\mathcal{H}\big[G_0(\mathbf{x},\mathbf{y})\big]\,dS_y,$$

$$w_1(\mathbf{x},t) = \int_0^t\int_V \Phi(\mathbf{y},\tau)G_1(\mathbf{x},\mathbf{y},t-\tau)\,dV_y\,d\tau + \int_0^t\int_S g(\mathbf{y},\tau)\mathcal{H}\big[G_1(\mathbf{x},\mathbf{y},t-\tau)\big]\,dS_y\,d\tau,$$

$$w_2(\mathbf{x},t) = \int_0^t\int_V \Phi(\mathbf{y},\tau)G_2(\mathbf{x},\mathbf{y},t-\tau)\,dV_y\,d\tau + \int_0^t\int_S g(\mathbf{y},\tau)\mathcal{H}\big[G_2(\mathbf{x},\mathbf{y},t-\tau)\big]\,dS_y\,d\tau,$$

where the G_n are the Green's functions, and the subscripts 0, 1, and 2 refer to the elliptic, parabolic, and hyperbolic problem, respectively. All solutions involve the same operator $\mathcal{H}[G]$; it is explicitly defined in Subsections 14.10.1–14.10.3 (see also Section 14.7) for different boundary conditions.

It is apparent that the solutions of the parabolic and hyperbolic problems with zero initial conditions have the same structure. The structure of the solution to the problem for a parabolic equation differs from that for an elliptic equation by the additional integration with respect to t.

14.11. Construction of the Green's Functions. General Formulas and Relations

14.11.1. Green's Functions of Boundary Value Problems for Equations of Various Types in Bounded Domains

14.11.1-1. Expressions of the Green's function in terms of infinite series.

Table 14.12 lists the Green's functions of boundary value problems for second-order equations of various types in a bounded domain V. It is assumed that $L_{\mathbf{x}}$ is a second-order linear self-adjoint differential operator (e.g., see Zwillinger, 1997) in the space variables $x_1,\ldots,x_n$, and $\Gamma_{\mathbf{x}}$ is a zeroth- or first-order linear boundary operator that can define a boundary condition of the first, second, or third kind; the coefficients of the operators $L_{\mathbf{x}}$ and $\Gamma_{\mathbf{x}}$ can depend on the space variables but are independent of time t. The coefficients λ_k and the functions $u_k(\mathbf{x})$ are determined by solving the homogeneous eigenvalue problem

$$L_{\mathbf{x}}[u] + \lambda u = 0, \tag{14.11.1.1}$$

$$\Gamma_{\mathbf{x}}[u] = 0 \quad \text{for} \quad \mathbf{x} \in S. \tag{14.11.1.2}$$

It is apparent from Table 14.12 that, given the Green's function in the problem for a parabolic (or hyperbolic) equation, one can easily construct the Green's functions of the corresponding problems for elliptic and hyperbolic (or parabolic) equations. In particular, the Green's function of the problem for an elliptic equation can be expressed via the Green's function of the problem for a parabolic equation as follows:

$$G_0(\mathbf{x},\mathbf{y}) = \int_0^\infty G_1(\mathbf{x},\mathbf{y},t)\,dt. \tag{14.11.1.3}$$

Here, the fact that all λ_k are positive is taken into account; for the second boundary value problem, it is assumed that $\lambda = 0$ is not an eigenvalue of problem (14.11.1.1)–(14.11.1.2).

TABLE 14.12
The Green's functions of boundary value problems for equations of various types in bounded domains. In all problems, the operators $L_{\mathbf{x}}$ and $\Gamma_{\mathbf{x}}$ are the same; $\mathbf{x} = \{x_1, \ldots, x_n\}$

Equation	Initial and boundary conditions	Green's function
Elliptic equation $-L_{\mathbf{x}}[w] = \Phi(\mathbf{x})$	$\Gamma_{\mathbf{x}}[w] = g(\mathbf{x})$ for $\mathbf{x} \in S$ (no initial condition required)	$G(\mathbf{x},\mathbf{y}) = \sum\limits_{k=1}^{\infty} \dfrac{u_k(\mathbf{x})u_k(\mathbf{y})}{\|u_k\|^2\lambda_k}, \quad \lambda_k \neq 0$
Parabolic equation $\partial_t w - L_{\mathbf{x}}[w] = \Phi(\mathbf{x},t)$	$w = f(\mathbf{x})$ at $t = 0$ $\Gamma_{\mathbf{x}}[w] = g(\mathbf{x},t)$ for $\mathbf{x} \in S$	$G(\mathbf{x},\mathbf{y},t) = \sum\limits_{k=1}^{\infty} \dfrac{u_k(\mathbf{x})u_k(\mathbf{y})}{\|u_k\|^2} \exp\bigl(-\lambda_k t\bigr)$
Hyperbolic equation $\partial_{tt} w - L_{\mathbf{x}}[w] = \Phi(\mathbf{x},t)$	$w = f_0(\mathbf{x})$ at $t = 0$ $w = f_1(\mathbf{x})$ at $t = 0$ $\Gamma_{\mathbf{x}}[w] = g(\mathbf{x},t)$ for $\mathbf{x} \in S$	$G(\mathbf{x},\mathbf{y},t) = \sum\limits_{k=1}^{\infty} \dfrac{u_k(\mathbf{x})u_k(\mathbf{y})}{\|u_k\|^2\sqrt{\lambda_k}} \sin\bigl(t\sqrt{\lambda_k}\,\bigr)$

14.11.1-2. Some remarks and generalizations.

Remark 1. Formula (14.11.1.3) can also be used if the domain V is infinite. In this case, one should make sure that the integral on the right-hand side is convergent.

Remark 2. Suppose the equations given in the first column of Table 14.12 contain $-L_{\mathbf{x}}[w] - \beta w$ instead of $-L_{\mathbf{x}}[w]$, with β being a free parameter. Then the λ_k in the expressions of the Green's function in the third column of Table 14.12 must be replaced by $\lambda_k - \beta$; just as previously, the λ_k and $u_k(\mathbf{x})$ were determined by solving the eigenvalue problem (14.11.1.1)–(14.11.1.2).

Remark 3. The formulas for the Green's functions presented in Table 14.12 will also hold for boundary value problems described by equations of the fourth or higher order in the space variables; provided that the eigenvalue problem for equation (14.11.1.1) subject to appropriate boundary conditions is self-adjoint.

14.11.2. Green's Functions Admitting Incomplete Separation of Variables

14.11.2-1. Boundary value problems for rectangular domains.

1°. Consider the parabolic equation

$$\frac{\partial w}{\partial t} = L_{1,t}[w] + \cdots + L_{n,t}[w] + \Phi(\mathbf{x},t), \tag{14.11.2.1}$$

where each term $L_{m,t}[w]$ depends on only one space variable, x_m, and time t:

$$L_{m,t}[w] \equiv a_m(x_m,t)\frac{\partial^2 w}{\partial x_m^2} + b_m(x_m,t)\frac{\partial w}{\partial x_m} + c_m(x_m,t)w, \qquad m = 1, \ldots, n.$$

For equation (14.11.2.1) we set the initial condition of general form

$$w = f(\mathbf{x}) \quad \text{at} \quad t = 0. \tag{14.11.2.2}$$

Consider the domain $V = \{\alpha_m \le x_m \le \beta_m,\ m = 1, \ldots, n\}$, which is an n-dimensional parallelepiped. We set the following boundary conditions at the faces of the parallelepiped:

$$\begin{aligned} s_m^{(1)}\frac{\partial w}{\partial x_m} + k_m^{(1)}(t)w &= g_m^{(1)}(\mathbf{x},t) \quad \text{at} \quad x_m = \alpha_m, \\ s_m^{(2)}\frac{\partial w}{\partial x_m} + k_m^{(2)}(t)w &= g_m^{(2)}(\mathbf{x},t) \quad \text{at} \quad x_m = \beta_m. \end{aligned} \tag{14.11.2.3}$$

By appropriately choosing the coefficients $s_m^{(1)}$, $s_m^{(2)}$ and functions $k_m^{(1)} = k_m^{(1)}(t)$, $k_m^{(2)} = k_m^{(2)}(t)$, we can obtain the boundary conditions of the first, second, or third kind. For infinite domains, the boundary conditions corresponding to $\alpha_m = -\infty$ or $\beta_m = \infty$ are omitted.

2°. The Green's function of the nonstationary n-dimensional boundary value problem (14.11.2.1)–(14.11.2.3) can be represented in the product form

$$G(\mathbf{x}, \mathbf{y}, t, \tau) = \prod_{m=1}^{n} G_m(x_m, y_m, t, \tau), \tag{14.11.2.4}$$

where the Green's functions $G_m = G_m(x_m, y_m, t, \tau)$ satisfy the one-dimensional equations

$$\frac{\partial G_m}{\partial t} - L_{m,t}[G_m] = 0 \qquad (m = 1, \ldots, n)$$

with the initial conditions

$$G_m = \delta(x_m - y_m) \quad \text{at} \quad t = \tau$$

and the homogeneous boundary conditions

$$s_m^{(1)} \frac{\partial G_m}{\partial x_m} + k_m^{(1)}(t) G_m = 0 \quad \text{at} \quad x_m = \alpha_m,$$
$$s_m^{(2)} \frac{\partial G_m}{\partial x_m} + k_m^{(2)}(t) G_m = 0 \quad \text{at} \quad x_m = \beta_m.$$

Here, y_m and τ are free parameters ($\alpha_m \le y_m \le \beta_m$ and $t \ge \tau \ge 0$), and $\delta(x)$ is the Dirac delta function.

It can be seen that the Green's function (14.11.2.4) admits incomplete separation of variables; it separates in the space variables $x_1, \ldots, x_n$ but not in time t.

Example. Consider the boundary value problem for the two-dimensional nonhomogeneous heat equation

$$\frac{\partial w}{\partial t} = \frac{\partial^2 w}{\partial x_1^2} + \frac{\partial^2 w}{\partial x_2^2} + \Phi(x_1, x_2, t)$$

with initial condition (14.11.2.2) and the nonhomogeneous mixed boundary conditions

$$w = g_1(x_2, t) \quad \text{at} \quad x_1 = 0, \qquad w = h_1(x_2, t) \quad \text{at} \quad x_1 = l_1;$$
$$\frac{\partial w}{\partial x_2} = g_2(x_1, t) \quad \text{at} \quad x_2 = 0, \qquad \frac{\partial w}{\partial x_2} = h_2(x_1, t) \quad \text{at} \quad x_2 = l_2.$$

The Green's functions of the corresponding homogeneous one-dimensional heat equations with homogeneous boundary conditions are expressed as

Equations and boundary conditions *Green's functions*

$$\frac{\partial w}{\partial t} = \frac{\partial^2 w}{\partial x_1^2}, \quad w = 0 \text{ at } x_1 = 0, l_1 \implies G_1 = \frac{2}{l_1} \sum_{m=1}^{\infty} \sin(\lambda_m x_1) \sin(\lambda_m y_1) e^{-\lambda_m^2 (t-\tau)}, \quad \lambda_m = \frac{m\pi}{l_1};$$

$$\frac{\partial w}{\partial t} = \frac{\partial^2 w}{\partial x_2^2}, \quad \frac{\partial w}{\partial x_2} = 0 \text{ at } x_2 = 0, l_2 \implies G_2 = \frac{1}{l_2} + \frac{2}{l_2} \sum_{n=1}^{\infty} \sin(\sigma_n x_2) \sin(\sigma_n y_2) e^{-\sigma_n^2 (t-\tau)}, \quad \sigma_n = \frac{n\pi}{l_2}.$$

Multiplying G_1 and G_2 together gives the Green's function for the original two-dimensional problem:

$$G(x_1, x_2, y_1, y_2, t, \tau) = \frac{4}{l_1 l_2} \left[\sum_{m=1}^{\infty} \sin(\lambda_m x_1) \sin(\lambda_m y_1) e^{-\lambda_m^2 (t-\tau)} \right] \left[\frac{1}{2} + \sum_{n=1}^{\infty} \sin(\sigma_n x_2) \sin(\sigma_n y_2) e^{-\sigma_n^2 (t-\tau)} \right].$$

14.11.2-2. Boundary value problems for a arbitrary cylindrical domain.

1°. Consider the parabolic equation

$$\frac{\partial w}{\partial t} = L_{\mathbf{x},t}[w] + M_{z,t}[w] + \Phi(\mathbf{x}, z, t), \tag{14.11.2.5}$$

where $L_{\mathbf{x},t}$ is an arbitrary second-order linear differential operator in $x_1, \ldots, x_n$ with coefficients dependent on $\mathbf{x}$ and t, and $M_{z,t}$ is an arbitrary second-order linear differential operator in z with coefficients dependent on z and t.

For equation (14.11.2.5) we set the general initial condition (14.11.2.2), where $f(\mathbf{x})$ must be replaced by $f(\mathbf{x}, z)$.

We assume that the space variables belong to a cylindrical domain $V = \{\mathbf{x} \in D, z_1 \le z \le z_2\}$ with arbitrary cross-section D. We set the boundary conditions*

$$\begin{aligned}
&\Gamma_1[w] = g_1(\mathbf{x}, t) \quad \text{at} \quad z = z_1 \quad (\mathbf{x} \in D),\\
&\Gamma_2[w] = g_2(\mathbf{x}, t) \quad \text{at} \quad z = z_2 \quad (\mathbf{x} \in D),\\
&\Gamma_3[w] = g_3(\mathbf{x}, z, t) \quad \text{for} \quad \mathbf{x} \in \partial D \quad (z_1 \le z \le z_2),
\end{aligned} \tag{14.11.2.6}$$

where the linear boundary operators Γ_k ($k = 1, 2, 3$) can define boundary conditions of the first, second, or third kind; in the last case, the coefficients of the differential operators Γ_k can be dependent on t.

2°. The Green's function of problem (14.11.2.5)–(14.11.2.6), (14.11.2.2) can be represented in the product form

$$G(\mathbf{x}, \mathbf{y}, z, \zeta, t, \tau) = G_L(\mathbf{x}, \mathbf{y}, t, \tau) G_M(z, \zeta, t, \tau), \tag{14.11.2.7}$$

where $G_L = G_L(\mathbf{x}, \mathbf{y}, t, \tau)$ and $G_M = G_M(z, \zeta, t, \tau)$ are auxiliary Green's functions; these can be determined from the following two simpler problems with fewer independent variables:

Problem on the cross-section D:

$$\begin{cases}
\dfrac{\partial G_L}{\partial t} = L_{\mathbf{x},t}[G_L] & \text{for} \quad \mathbf{x} \in D,\\
G_L = \delta(\mathbf{x} - \mathbf{y}) & \text{at} \quad t = \tau,\\
\Gamma_3[G_L] = 0 & \text{for} \quad \mathbf{x} \in \partial D,
\end{cases}$$

Problem on the interval $z_1 \le z \le z_2$:

$$\begin{cases}
\dfrac{\partial G_M}{\partial t} = M_{z,t}[G_M] & \text{for} \quad z_1 < z < z_2,\\
G_M = \delta(z - \zeta) & \text{at} \quad t = \tau,\\
\Gamma_k[G_M] = 0 & \text{at} \quad z = z_k \ \ (k = 1, 2).
\end{cases}$$

Here, $\mathbf{y}$, ζ, and τ are free parameters ($\mathbf{y} \in D$, $z_1 \le \zeta \le z_2$, $t \ge \tau \ge 0$).

It can be seen that the Green's function (14.11.2.7) admits incomplete separation of variables; it separates in the space variables $\mathbf{x}$ and z but not in time t.

14.11.3. Construction of Green's Functions via Fundamental Solutions

14.11.3-1. Elliptic equations. Fundamental solution.

Consider the elliptic equation

$$L_{\mathbf{x}}[w] + \frac{\partial^2 w}{\partial z^2} = \Phi(\mathbf{x}, z), \tag{14.11.3.1}$$

* If $z_1 = -\infty$ or $z_2 = \infty$, the corresponding boundary condition is to be omitted.

where $\mathbf{x} = \{x_1, \dots, x_n\} \in \mathbb{R}^n$, $z \in \mathbb{R}^1$, and $L_{\mathbf{x}}[w]$ is a linear differential operator that depends on $x_1, \dots, x_n$ but is independent of z. For subsequent analysis it is significant that the homogeneous equation (with $\Phi \equiv 0$) does not change under the replacement of z by $-z$ and z by $z + \text{const}$.

Let $\mathscr{E} = \mathscr{E}(\mathbf{x}, \mathbf{y}, z - \zeta)$ be a fundamental solution of equation (14.11.3.1), which means that

$$L_{\mathbf{x}}[\mathscr{E}] + \frac{\partial^2 \mathscr{E}}{\partial z^2} = \delta(\mathbf{x} - \mathbf{y})\delta(z - \zeta).$$

Here, $\mathbf{y} = \{y_1, \dots, y_n\} \in \mathbb{R}^n$ and $\zeta \in \mathbb{R}^1$ are free parameters.

The fundamental solution of equation (14.11.3.1) is an even function in the last argument, i.e.,

$$\mathscr{E}(\mathbf{x}, \mathbf{y}, z) = \mathscr{E}(\mathbf{x}, \mathbf{y}, -z).$$

Below, Paragraphs 14.11.3-2 and 14.11.3-3 present relations that permit one to express the Green's functions of some boundary value problems for equation (14.11.3.1) via its fundamental solution.

14.11.3-2. Domain: $\mathbf{x} \in \mathbb{R}^n$, $0 \le z < \infty$. Problems for elliptic equations.

1°. *First boundary value problem.* The boundary condition:

$$w = f(\mathbf{x}) \quad \text{at} \quad z = 0.$$

Green's function:

$$G(\mathbf{x}, \mathbf{y}, z, \zeta) = \mathscr{E}(\mathbf{x}, \mathbf{y}, z - \zeta) - \mathscr{E}(\mathbf{x}, \mathbf{y}, z + \zeta). \tag{14.11.3.2}$$

Domain of the free parameters: $\mathbf{y} \in \mathbb{R}^n$ and $0 \le \zeta < \infty$.

Example 1. Consider the first boundary value problem in the half-space $-\infty < x_1, x_2 < \infty$, $0 \le x_3 < \infty$ for the three-dimensional Laplace equation

$$\frac{\partial^2 w}{\partial x_1^2} + \frac{\partial^2 w}{\partial x_2^2} + \frac{\partial^2 w}{\partial x_3^2} = 0$$

under boundary condition

$$w = f(x_1, x_2) \quad \text{at} \quad x_3 = 0.$$

The fundamental solution for the Laplace equation has the form

$$\mathscr{E} = \frac{1}{4\pi\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}}.$$

In terms of the notation adopted for equation (14.11.3.1) and its fundamental solution, we have $x_3 = z$, $y_3 = \zeta$, and $\mathscr{E} = \mathscr{E}(x_1, y_1, x_2, y_2, z - \zeta)$. Using formula (14.11.3.2), we obtain the Green's function for the first boundary value problem in the half-space:

$$\begin{aligned} G(x_1, y_1, x_2, y_2, z, \zeta) &= \mathscr{E}(x_1, y_1, x_2, y_2, z - \zeta) - \mathscr{E}(x_1, y_1, x_2, y_2, z + \zeta) \\ &= \frac{1}{4\pi\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}} - \frac{1}{4\pi\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 + y_3)^2}}. \end{aligned}$$

2°. *Second boundary value problem.* The boundary condition:

$$\partial_z w = f(\mathbf{x}) \quad \text{at} \quad z = 0.$$

Green's function:

$$G(\mathbf{x},\mathbf{y},z,\zeta)=\mathscr{E}(\mathbf{x},\mathbf{y},z-\zeta)+\mathscr{E}(\mathbf{x},\mathbf{y},z+\zeta).$$

Example 2. The Green's function of the second boundary value problem for the three-dimensional Laplace equation in the half-space $-\infty<x_1,x_2<\infty$, $0\le x_3<\infty$ is expressed as

$$G(x_1,y_1,x_2,y_2,z,\zeta)=\frac{1}{4\pi\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+(x_3-y_3)^2}}+\frac{1}{4\pi\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+(x_3+y_3)^2}}.$$

It is obtained using the same reasoning as in Example 1.

3°. *Third boundary value problem.* The boundary condition:

$$\partial_z w-kw=f(\mathbf{x})\quad\text{at}\quad z=0.$$

Green's function:

$$\begin{aligned}G(\mathbf{x},\mathbf{y},z,\zeta)&=\mathscr{E}(\mathbf{x},\mathbf{y},z-\zeta)+\mathscr{E}(\mathbf{x},\mathbf{y},z+\zeta)-2k\int_0^\infty e^{-ks}\mathscr{E}(\mathbf{x},\mathbf{y},z+\zeta+s)\,ds\\&=\mathscr{E}(\mathbf{x},\mathbf{y},z-\zeta)+\mathscr{E}(\mathbf{x},\mathbf{y},z+\zeta)-2k\int_{z+\zeta}^\infty e^{-k(\sigma-z-\zeta)}\mathscr{E}(\mathbf{x},\mathbf{y},\sigma)\,d\sigma.\end{aligned}$$

14.11.3-3. Domain: $\mathbf{x}\in\mathbb{R}^n$, $0\le z\le l$. Problems for elliptic equations.

1°. *First boundary value problem.* Boundary conditions:

$$w=f_1(\mathbf{x})\quad\text{at}\quad z=0,\qquad w=f_2(\mathbf{x})\quad\text{at}\quad z=l.$$

Green's function:

$$G(\mathbf{x},\mathbf{y},z,\zeta)=\sum_{n=-\infty}^{\infty}\big[\mathscr{E}(\mathbf{x},\mathbf{y},z-\zeta+2nl)-\mathscr{E}(\mathbf{x},\mathbf{y},z+\zeta+2nl)\big].\qquad(14.11.3.3)$$

Domain of the free parameters: $\mathbf{y}\in\mathbb{R}^n$ and $0\le\zeta\le l$.

2°. *Second boundary value problem.* Boundary conditions:

$$\partial_z w=f_1(\mathbf{x})\quad\text{at}\quad z=0,\qquad \partial_z w=f_2(\mathbf{x})\quad\text{at}\quad z=l.$$

Green's function:

$$G(\mathbf{x},\mathbf{y},z,\zeta)=\sum_{n=-\infty}^{\infty}\big[\mathscr{E}(\mathbf{x},\mathbf{y},z-\zeta+2nl)+\mathscr{E}(\mathbf{x},\mathbf{y},z+\zeta+2nl)\big].\qquad(14.11.3.4)$$

3°. *Mixed boundary value problem.* The unknown function and its derivative are prescribed at the left and right end, respectively:

$$w=f_1(\mathbf{x})\quad\text{at}\quad z=0,\qquad \partial_z w=f_2(\mathbf{x})\quad\text{at}\quad z=l.$$

Green's function:

$$G(\mathbf{x},\mathbf{y},z,\zeta)=\sum_{n=-\infty}^{\infty}(-1)^n\big[\mathscr{E}(\mathbf{x},\mathbf{y},z-\zeta+2nl)-\mathscr{E}(\mathbf{x},\mathbf{y},z+\zeta+2nl)\big].\qquad(14.11.3.5)$$

4°. *Mixed boundary value problem.* The derivative and the unknown function itself are prescribed at the left and right end, respectively:

$$\partial_z w=f_1(\mathbf{x})\quad\text{at}\quad z=0,\qquad w=f_2(\mathbf{x})\quad\text{at}\quad z=l.$$

Green's function:

$$G(\mathbf{x},\mathbf{y},z,\zeta)=\sum_{n=-\infty}^{\infty}(-1)^n\big[\mathscr{E}(\mathbf{x},\mathbf{y},z-\zeta+2nl)+\mathscr{E}(\mathbf{x},\mathbf{y},z+\zeta+2nl)\big].\qquad(14.11.3.6)$$

Remark. One should make sure that series (14.11.3.3)–(14.11.3.6) are convergent; in particular, for the three-dimensional Laplace equation, series (14.11.3.3), (14.11.3.5), and (14.11.3.6) are convergent and series (14.11.3.4) is divergent.

14.11.3-4. Boundary value problems for parabolic equations.

Let $\mathbf{x} \in \mathbb{R}^n$, $z \in \mathbb{R}^1$, and $t \geq 0$. Consider the parabolic equation

$$\frac{\partial w}{\partial t} = L_{\mathbf{x},t}[w] + \frac{\partial^2 w}{\partial z^2} + \Phi(\mathbf{x}, z, t), \tag{14.11.3.7}$$

where $L_{\mathbf{x},t}[w]$ is a linear differential operator that depends on $x_1, \ldots, x_n$ and t but is independent of z.

Let $\mathscr{E} = \mathscr{E}(\mathbf{x}, \mathbf{y}, z-\zeta, t, \tau)$ be a fundamental solution of the Cauchy problem for equation (14.11.3.7), i.e.,

$$\begin{aligned} &\frac{\partial \mathscr{E}}{\partial t} = L_{\mathbf{x},t}[\mathscr{E}] + \frac{\partial^2 \mathscr{E}}{\partial z^2} \quad \text{for} \quad t > \tau, \\ &\mathscr{E} = \delta(\mathbf{x}-\mathbf{y})\,\delta(z-\zeta) \quad \text{at} \quad t = \tau. \end{aligned}$$

Here, $\mathbf{y} \in \mathbb{R}^n$, $\zeta \in \mathbb{R}^1$, and $\tau \geq 0$ are free parameters.

The fundamental solution of the Cauchy problem possesses the property

$$\mathscr{E}(\mathbf{x}, \mathbf{y}, z, t, \tau) = \mathscr{E}(\mathbf{x}, \mathbf{y}, -z, t, \tau).$$

Table 14.13 presents formulas that permit one to express the Green's functions of some nonstationary boundary value problems for equation (14.11.3.7) via the fundamental solution of the Cauchy problem.

TABLE 14.13
Representation of the Green's functions of some nonstationary boundary value problems in terms of the fundamental solution of the Cauchy problem

Boundary value problems	Boundary conditions	Green's functions
First problem $\mathbf{x} \in \mathbb{R}^n$, $z \in \mathbb{R}^1$	$G = 0$ at $z = 0$	$G(\mathbf{x}, \mathbf{y}, z, \zeta, t, \tau) = \mathscr{E}(\mathbf{x}, \mathbf{y}, z-\zeta, t, \tau) - \mathscr{E}(\mathbf{x}, \mathbf{y}, z+\zeta, t, \tau)$
Second problem $\mathbf{x} \in \mathbb{R}^n$, $z \in \mathbb{R}^1$	$\partial_z G = 0$ at $z = 0$	$G(\mathbf{x}, \mathbf{y}, z, \zeta, t, \tau) = \mathscr{E}(\mathbf{x}, \mathbf{y}, z-\zeta, t, \tau) + \mathscr{E}(\mathbf{x}, \mathbf{y}, z+\zeta, t, \tau)$
Third problem $\mathbf{x} \in \mathbb{R}^n$, $z \in \mathbb{R}^1$	$\partial_z G - kG = 0$ at $z = 0$	$G(\mathbf{x}, \mathbf{y}, z, \zeta, t, \tau) = \mathscr{E}(\mathbf{x}, \mathbf{y}, z-\zeta, t, \tau) + \mathscr{E}(\mathbf{x}, \mathbf{y}, z+\zeta, t, \tau) - 2k \int_0^\infty e^{-ks}\, \mathscr{E}(\mathbf{x}, \mathbf{y}, z+\zeta+s, t, \tau)\, ds$
First problem $\mathbf{x} \in \mathbb{R}^n$, $0 \leq z \leq l$	$G = 0$ at $z = 0$, $G = 0$ at $z = l$	$G(\mathbf{x}, \mathbf{y}, z, \zeta, t, \tau) = \sum_{n=-\infty}^{\infty} \big[\mathscr{E}(\mathbf{x}, \mathbf{y}, z-\zeta+2nl, t, \tau) - \mathscr{E}(\mathbf{x}, \mathbf{y}, z+\zeta+2nl, t, \tau)\big]$
Second problem $\mathbf{x} \in \mathbb{R}^n$, $0 \leq z \leq l$	$\partial_z G = 0$ at $z = 0$, $\partial_z G = 0$ at $z = l$	$G(\mathbf{x}, \mathbf{y}, z, \zeta, t, \tau) = \sum_{n=-\infty}^{\infty} \big[\mathscr{E}(\mathbf{x}, \mathbf{y}, z-\zeta+2nl, t, \tau) + \mathscr{E}(\mathbf{x}, \mathbf{y}, z+\zeta+2nl, t, \tau)\big]$
Mixed problem $\mathbf{x} \in \mathbb{R}^n$, $0 \leq z \leq l$	$G = 0$ at $z = 0$, $\partial_z G = 0$ at $z = l$	$G(\mathbf{x}, \mathbf{y}, z, \zeta, t, \tau) = \sum_{n=-\infty}^{\infty} (-1)^n \big[\mathscr{E}(\mathbf{x}, \mathbf{y}, z-\zeta+2nl, t, \tau) - \mathscr{E}(\mathbf{x}, \mathbf{y}, z+\zeta+2nl, t, \tau)\big]$
Mixed problem $\mathbf{x} \in \mathbb{R}^n$, $0 \leq z \leq l$	$\partial_z G = 0$ at $z = 0$, $G = 0$ at $z = l$	$G(\mathbf{x}, \mathbf{y}, z, \zeta, t, \tau) = \sum_{n=-\infty}^{\infty} (-1)^n \big[\mathscr{E}(\mathbf{x}, \mathbf{y}, z-\zeta+2nl, t, \tau) + \mathscr{E}(\mathbf{x}, \mathbf{y}, z+\zeta+2nl, t, \tau)\big]$

14.12. Duhamel's Principles in Nonstationary Problems

14.12.1. Problems for Homogeneous Linear Equations

14.12.1-1. Parabolic equations with two independent variables.

Consider the problem for the homogeneous linear equation of parabolic type

$$\frac{\partial w}{\partial t} = a(x)\frac{\partial^2 w}{\partial x^2} + b(x)\frac{\partial w}{\partial x} + c(x)w \tag{14.12.1.1}$$

with the homogeneous initial condition

$$w = 0 \quad \text{at} \quad t = 0 \tag{14.12.1.2}$$

and the boundary conditions

$$s_1\partial_x w + k_1 w = g(t) \qquad \text{at} \quad x = x_1, \tag{14.12.1.3}$$

$$s_2\partial_x w + k_2 w = 0 \qquad \text{at} \quad x = x_2. \tag{14.12.1.4}$$

By appropriately choosing the values of the coefficients s_1, s_2, k_1, and k_2 in (14.12.1.3) and (14.12.1.4), one can obtain the first, second, third, and mixed boundary value problems for equation (14.12.1.1).

The solution of problem (14.12.1.1)–(14.12.1.4) with the nonstationary boundary condition (14.12.1.3) at $x = x_1$ can be expressed by the formula (*Duhamel's first principle*)

$$w(x,t) = \frac{\partial}{\partial t}\int_0^t u(x, t-\tau)\, g(\tau)\, d\tau = \int_0^t \frac{\partial u}{\partial t}(x, t-\tau)\, g(\tau)\, d\tau \tag{14.12.1.5}$$

in terms of the solution $u(x,t)$ of the auxiliary problem for equation (14.12.1.1) with the initial and boundary conditions (14.12.1.2) and (14.12.1.4), for u instead of w, and the following simpler stationary boundary condition at $x = x_1$:

$$s_1\partial_x u + k_1 u = 1 \quad \text{at} \quad x = x_1. \tag{14.12.1.6}$$

Remark. A similar formula also holds for the homogeneous boundary condition at $x = x_1$ and a nonhomogeneous nonstationary boundary condition at $x = x_2$.

Example. Consider the first boundary value problem for the heat equation

$$\frac{\partial w}{\partial t} = \frac{\partial^2 w}{\partial x^2} \tag{14.12.1.7}$$

with the homogeneous initial condition (14.12.1.2) and the boundary condition

$$w = g(t) \quad \text{at} \quad x = 0. \tag{14.12.1.8}$$

(The second boundary condition is not required in this case; $0 \le x < \infty$.)

First consider the following auxiliary problem for the heat equation with the homogeneous initial condition and a simpler boundary condition:

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad u = 0 \quad \text{at} \quad t = 0, \qquad u = 1 \quad \text{at} \quad x = 0.$$

This problem has a self-similar solution of the form

$$w = w(z), \qquad z = xt^{-1/2},$$

where the function $w(z)$ is determined by the following ordinary differential equation and boundary conditions:

$$u''_{zz} + \tfrac{1}{2}zu'_z = 0, \qquad u = 1 \quad \text{at} \quad z = 0, \qquad u = 0 \quad \text{at} \quad z = \infty.$$

Its solution is expressed as

$$u(z) = \operatorname{erfc}\left(\frac{z}{2}\right) \quad \Longrightarrow \quad u(x,t) = \operatorname{erfc}\left(\frac{x}{2\sqrt{t}}\right),$$

where $\operatorname{erfc} z = \dfrac{2}{\sqrt{\pi}}\displaystyle\int_z^\infty \exp(-\xi^2)\, d\xi$ is the complementary error function. Substituting the obtained expression of $u(x,t)$ into (14.12.1.5), we obtain the solution to the first boundary value problem for the heat equation (14.12.1.7) with the initial condition (14.12.1.2) and an arbitrary boundary condition (14.12.1.8) in the form

$$w(x,t) = \frac{x}{2\sqrt{\pi}}\int_0^t \exp\left[-\frac{x^2}{4(t-\tau)}\right]\frac{g(\tau)\, d\tau}{(t-\tau)^{3/2}}.$$

14.12.1-2. Hyperbolic equations with two independent variables.

Consider the problem for the homogeneous linear hyperbolic equation

$$\frac{\partial^2 w}{\partial t^2} + \varphi(x)\frac{\partial w}{\partial t} = a(x)\frac{\partial^2 w}{\partial x^2} + b(x)\frac{\partial w}{\partial x} + c(x)w \tag{14.12.1.9}$$

with the homogeneous initial conditions

$$\begin{aligned} w &= 0 \quad \text{at} \quad t = 0, \\ \partial_t w &= 0 \quad \text{at} \quad t = 0, \end{aligned} \tag{14.12.1.10}$$

and the boundary conditions (14.12.1.3) and (14.12.1.4).

The solution of problem (14.12.1.9), (14.12.1.10), (14.12.1.3), (14.12.1.4) with the nonstationary boundary condition (14.12.1.3) at $x = x_1$ can be expressed by formula (14.12.1.5) in terms of the solution $u(x,t)$ of the auxiliary problem for equation (14.12.1.9) with the initial conditions (14.12.1.10) and boundary condition (14.12.1.4), for u instead of w, and the simpler stationary boundary condition (14.12.1.6) at $x = x_1$.

In this case, the remark made in Paragraph 14.12.1-1 remains valid.

14.12.1-3. Second-order equations with several independent variables.

Duhamel's first principle can also be used to solve homogeneous linear equations of the parabolic or hyperbolic type with many space variables,

$$\frac{\partial^k w}{\partial t^k} = \sum_{i,j=1}^{n} a_{ij}(\mathbf{x})\frac{\partial^2 w}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(\mathbf{x})\frac{\partial w}{\partial x_i} + c(\mathbf{x})w, \tag{14.12.1.11}$$

where $k = 1, 2$ and $\mathbf{x} = \{x_1, \ldots, x_n\}$.

Let V be some bounded domain in $\mathbb{R}^n$ with a sufficiently smooth surface $S = \partial V$. The solution of the boundary value problem for equation (14.12.1.11) in V with the homogeneous initial conditions (14.12.1.2) if $k = 1$ or (14.12.1.10) if $k = 2$, and the nonhomogeneous linear boundary condition

$$\Gamma_{\mathbf{x}}[w] = g(t) \quad \text{for} \quad \mathbf{x} \in S, \tag{14.12.1.12}$$

is given by

$$w(\mathbf{x},t) = \frac{\partial}{\partial t}\int_0^t u(\mathbf{x}, t-\tau)\, g(\tau)\, d\tau = \int_0^t \frac{\partial u}{\partial t}(\mathbf{x}, t-\tau)\, g(\tau)\, d\tau.$$

Here, $u(\mathbf{x},t)$ is the solution of the auxiliary problem for equation (14.12.1.11) with the same initial conditions, (14.12.1.2) or (14.12.1.10), for u instead of w, and the simpler stationary boundary condition

$$\Gamma_{\mathbf{x}}[u] = 1 \quad \text{for} \quad \mathbf{x} \in S.$$

Note that (14.12.1.12) can represent a boundary condition of the first, second, or third kind; the coefficients of the operator $\Gamma_{\mathbf{x}}$ are assumed to be independent of t.

14.12.2. Problems for Nonhomogeneous Linear Equations

14.12.2-1. Parabolic equations.

The solution of the nonhomogeneous linear equation

$$\frac{\partial w}{\partial t} = \sum_{i,j=1}^{n} a_{ij}(\mathbf{x})\frac{\partial^2 w}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(\mathbf{x})\frac{\partial w}{\partial x_i} + c(\mathbf{x})w + \Phi(\mathbf{x}, t)$$

with the homogeneous initial condition (14.12.1.2) and the homogeneous boundary condition

$$\Gamma_{\mathbf{x}}[w] = 0 \quad \text{for} \quad \mathbf{x} \in S \tag{14.12.2.1}$$

can be represented in the form (*Duhamel's second principle*)

$$w(\mathbf{x}, t) = \int_0^t U(\mathbf{x}, t-\tau, \tau)\, d\tau. \tag{14.12.2.2}$$

Here, $U(\mathbf{x}, t, \tau)$ is the solution of the auxiliary problem for the homogeneous equation

$$\frac{\partial U}{\partial t} = \sum_{i,j=1}^{n} a_{ij}(\mathbf{x})\frac{\partial^2 U}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(\mathbf{x})\frac{\partial U}{\partial x_i} + c(\mathbf{x})U$$

with the boundary condition (14.12.2.1), in which w must be substituted by U, and the nonhomogeneous initial condition

$$U = \Phi(\mathbf{x}, \tau) \quad \text{at} \quad t = 0,$$

where τ is a parameter.

Note that (14.12.2.1) can represent a boundary condition of the first, second, or third kind; the coefficients of the operator $\Gamma_{\mathbf{x}}$ are assumed to be independent of t.

14.12.2-2. Hyperbolic equations.

The solution of the nonhomogeneous linear equation

$$\frac{\partial^2 w}{\partial t^2} + \varphi(\mathbf{x})\frac{\partial w}{\partial t} = \sum_{i,j=1}^{n} a_{ij}(\mathbf{x})\frac{\partial^2 w}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(\mathbf{x})\frac{\partial w}{\partial x_i} + c(\mathbf{x})w + \Phi(\mathbf{x}, t)$$

with the homogeneous initial conditions (14.12.1.10) and homogeneous boundary condition (14.12.2.1) can be expressed by formula (14.12.2.2) in terms of the solution $U = U(\mathbf{x}, t, \tau)$ of the auxiliary problem for the homogeneous equation

$$\frac{\partial^2 U}{\partial t^2} + \varphi(\mathbf{x})\frac{\partial U}{\partial t} = \sum_{i,j=1}^{n} a_{ij}(\mathbf{x})\frac{\partial^2 U}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(\mathbf{x})\frac{\partial U}{\partial x_i} + c(\mathbf{x})U$$

with the homogeneous initial and boundary conditions, (14.12.1.2) and (14.12.2.1), where w must be replaced by U, and the nonhomogeneous initial condition

$$\partial_t U = \Phi(\mathbf{x}, \tau) \quad \text{at} \quad t = 0,$$

where τ is a parameter.

Note that (14.12.2.1) can represent a boundary condition of the first, second, or third kind.

14.13. Transformations Simplifying Initial and Boundary Conditions

14.13.1. Transformations That Lead to Homogeneous Boundary Conditions

A linear problem with arbitrary nonhomogeneous boundary conditions,

$$\Gamma^{(k)}_{\mathbf{x},t}[w] = g_k(\mathbf{x}, t) \quad \text{for} \quad \mathbf{x} \in S_k, \tag{14.13.1.1}$$

can be reduced to a linear problem with homogeneous boundary conditions. To this end, one should perform the change of variable

$$w(\mathbf{x}, t) = \psi(\mathbf{x}, t) + u(\mathbf{x}, t), \tag{14.13.1.2}$$

where u is a new unknown function and ψ is any function that satisfies the nonhomogeneous boundary conditions (14.13.1.1),

$$\Gamma^{(k)}_{\mathbf{x},t}[\psi] = g_k(\mathbf{x}, t) \quad \text{for} \quad \mathbf{x} \in S_k. \tag{14.13.1.3}$$

Table 14.14 gives examples of such transformations for linear boundary value problems with one space variable for parabolic and hyperbolic equations. In the third boundary value problem, it is assumed that $k_1 < 0$ and $k_2 > 0$.

TABLE 14.14
Simple transformations of the form $w(x,t) = \psi(x,t) + u(x,t)$ that lead to homogeneous boundary conditions in problems with one space variables ($0 \le x \le l$)

No.	Problems	Boundary conditions	Function $\psi(x,t)$
1	First boundary value problem	$w = g_1(t)$ at $x = 0$ $w = g_2(t)$ at $x = l$	$\psi(x,t) = g_1(t) + \dfrac{x}{l}\big[g_2(t) - g_1(t)\big]$
2	Second boundary value problem	$\partial_x w = g_1(t)$ at $x = 0$ $\partial_x w = g_2(t)$ at $x = l$	$\psi(x,t) = xg_1(t) + \dfrac{x^2}{2l}\big[g_2(t) - g_1(t)\big]$
3	Third boundary value problem	$\partial_x w + k_1 w = g_1(t)$ at $x = 0$ $\partial_x w + k_2 w = g_2(t)$ at $x = l$	$\psi(x,t) = \dfrac{(k_2x - 1 - k_2l)g_1(t) + (1 - k_1x)g_2(t)}{k_2 - k_1 - k_1k_2l}$
4	Mixed boundary value problem	$w = g_1(t)$ at $x = 0$ $\partial_x w = g_2(t)$ at $x = l$	$\psi(x,t) = g_1(t) + xg_2(t)$
5	Mixed boundary value problem	$\partial_x w = g_1(t)$ at $x = 0$ $w = g_2(t)$ at $x = l$	$\psi(x,t) = (x - l)g_1(t) + g_2(t)$

Note that the selection of the function ψ is of a purely algebraic nature and is not connected with the equation in question; there are infinitely many suitable functions ψ that satisfy condition (14.13.1.3). Transformations of the form (14.13.1.2) can often be used at the first stage of solving boundary value problems.

14.13.2. Transformations That Lead to Homogeneous Initial and Boundary Conditions

A linear problem with nonhomogeneous initial and boundary conditions can be reduced to a linear problem with homogeneous initial and boundary conditions. To this end, one should introduce a new dependent variable u by formula (14.13.1.2), where the function ψ must satisfy nonhomogeneous initial and boundary conditions.

Below we specify some simple functions ψ that can be used in transformation (14.13.1.2) to obtain boundary value problems with homogeneous initial and boundary conditions. To be specific, we consider a parabolic equation with one space variable and the general initial condition

$$w = f(x) \quad \text{at} \quad t = 0. \tag{14.13.2.1}$$

1. First boundary value problem: the initial condition is (14.13.2.1) and the boundary conditions are given in row 1 of Table 14.14. Suppose that the initial and boundary conditions are compatible, i.e., $f(0) = g_1(0)$ and $f(l) = g_2(0)$. Then, in transformation (14.13.1.2), one can take

$$\psi(x,t) = f(x) + g_1(t) - g_1(0) + \frac{x}{l}\big[g_2(t) - g_1(t) + g_1(0) - g_2(0)\big].$$

2. Second boundary value problem: the initial condition is (14.13.2.1) and the boundary conditions are given in row 2 of Table 14.14. Suppose that the initial and boundary conditions are compatible, i.e., $f'(0) = g_1(0)$ and $f'(l) = g_2(0)$. Then, in transformation (14.13.1.2), one can set

$$\psi(x,t) = f(x) + x\big[g_1(t) - g_1(0)\big] + \frac{x^2}{2l}\big[g_2(t) - g_1(t) + g_1(0) - g_2(0)\big].$$

3. Third boundary value problem: the initial condition is (14.13.2.1) and the boundary conditions are given in row 3 of Table 14.14. If the initial and boundary conditions are compatible, then, in transformation (14.13.1.2), one can take

$$\psi(x,t) = f(x) + \frac{(k_2 x - 1 - k_2 l)[g_1(t) - g_1(0)] + (1 - k_1 x)[g_2(t) - g_2(0)]}{k_2 - k_1 - k_1 k_2 l} \qquad (k_1 < 0,\ k_2 > 0).$$

4. Mixed boundary value problem: the initial condition is (14.13.2.1) and the boundary conditions are given in row 4 of Table 14.14. Suppose that the initial and boundary conditions are compatible, i.e., $f(0) = g_1(0)$ and $f'(l) = g_2(0)$. Then, in transformation (14.13.1.2), one can set

$$\psi(x,t) = f(x) + g_1(t) - g_1(0) + x\big[g_2(t) - g_2(0)\big].$$

5. Mixed boundary value problem: the initial condition is (14.13.2.1) and the boundary conditions are given in row 5 of Table 14.14. Suppose that the initial and boundary conditions are compatible, i.e., $f'(0) = g_1(0)$ and $f(l) = g_2(0)$. Then, in transformation (14.13.1.2), one can take

$$\psi(x,t) = f(x) + (x - l)\big[g_1(t) - g_1(0)\big] + g_2(t) - g_2(0).$$

References for Chapter 14

Akulenko, L. D. and Nesterov, S. V., *High Precision Methods in Eigenvalue Problems and Their Applications*, Chapman & Hall/CRC Press, Boca Raton, 2004.

Butkovskiy, A. G., *Green's Functions and Transfer Functions Handbook*, Halstead Press–John Wiley & Sons, New York, 1982.

Carslaw, H. S. and Jaeger, J. C., *Conduction of Heat in Solids*, Clarendon Press, Oxford, 1984.
Constanda, C., *Solution Techniques for Elementary Partial Differential Equations*, Chapman & Hall/CRC Press, Boca Raton, 2002.
Courant, R. and Hilbert, D., *Methods of Mathematical Physics, Vol. 2*, Wiley-Interscience, New York, 1989.
Dezin, A. A., *Partial Differential Equations. An Introduction to a General Theory of Linear Boundary Value Problems,* Springer-Verlag, Berlin, 1987.
Ditkin, V. A. and Prudnikov, A. P., *Integral Transforms and Operational Calculus*, Pergamon Press, New York, 1965.
Duffy, D. G., *Transform Methods for Solving Partial Differential Equations, 2nd Edition*, Chapman & Hall/CRC Press, Boca Raton, 2004.
Farlow, S. J., *Partial Differential Equations for Scientists and Engineers*, John Wiley & Sons, New York, 1982.
Guenther, R. B. and Lee, J. W., *Partial Differential Equations of Mathematical Physics and Integral Equations*, Dover Publications, New York, 1996.
Haberman, R., *Elementary Applied Partial Differential Equations with Fourier Series and Boundary Value Problems*, Prentice-Hall, Englewood Cliffs, New Jersey, 1987.
Hanna, J. R. and Rowland, J. H., *Fourier Series, Transforms, and Boundary Value Problems,* Wiley-Interscience, New York, 1990.
Kanwal, R. P., *Generalized Functions. Theory and Technique,* Academic Press, Orlando, 1983.
Leis, R., *Initial-Boundary Value Problems in Mathematical Physics,* John Wiley & Sons, Chichester, 1986.
Mikhlin, S. G. (Editor), *Linear Equations of Mathematical Physics*, Holt, Rinehart and Winston, New York, 1967.
Miller, W., Jr., *Symmetry and Separation of Variables*, Addison-Wesley, London, 1977.
Moon, P. and Spencer, D. E., *Field Theory Handbook, Including Coordinate Systems, Differential Equations and Their Solutions, 3rd Edition*, Springer-Verlag, Berlin, 1988.
Morse, P. M. and Feshbach, H., *Methods of Theoretical Physics, Vols. 1 and 2*, McGraw-Hill, New York, 1953.
Petrovsky, I. G., *Lectures on Partial Differential Equations*, Dover Publications, New York, 1991.
Polyanin, A. D., *Handbook of Linear Partial Differential Equations for Engineers and Scientists*, Chapman & Hall/CRC Press, Boca Raton, 2002.
Sneddon, I. N., *Fourier Transformations*, Dover Publications, New York, 1995.
Stakgold, I., *Boundary Value Problems of Mathematical Physics. Vols. 1 and 2*, Society for Industrial & Applied Mathematics, Philadelphia, 2000.
Strauss, W. A., *Partial Differential Equations. An Introduction*, John Wiley & Sons, New York, 1992.
Tikhonov, A. N. and Samarskii, A. A., *Equations of Mathematical Physics*, Dover Publications, New York, 1990.
Vladimirov, V. S., *Equations of Mathematical Physics*, Dekker, New York, 1971.
Zauderer, E., *Partial Differential Equations of Applied Mathematics*, Wiley-Interscience, New York, 1989.
Zwillinger, D., *Handbook of Differential Equations, 3rd Edition*, Academic Press, New York, 1997.