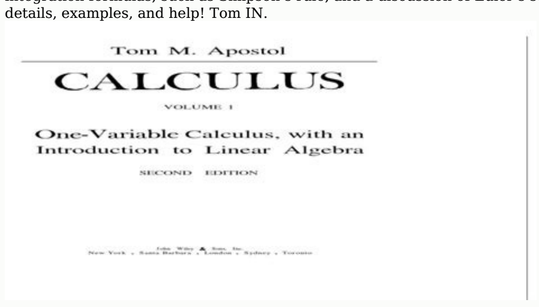


I'm not robot!

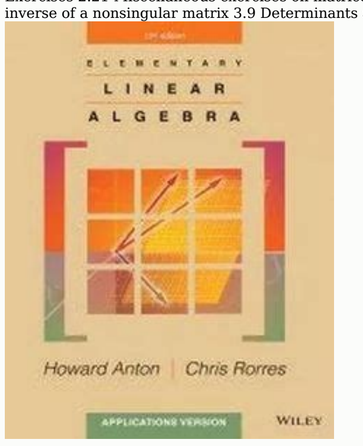
Tom m apostol calculus vol 2 pdf

Calculus vol. 2 by tom m. apostol pdf.

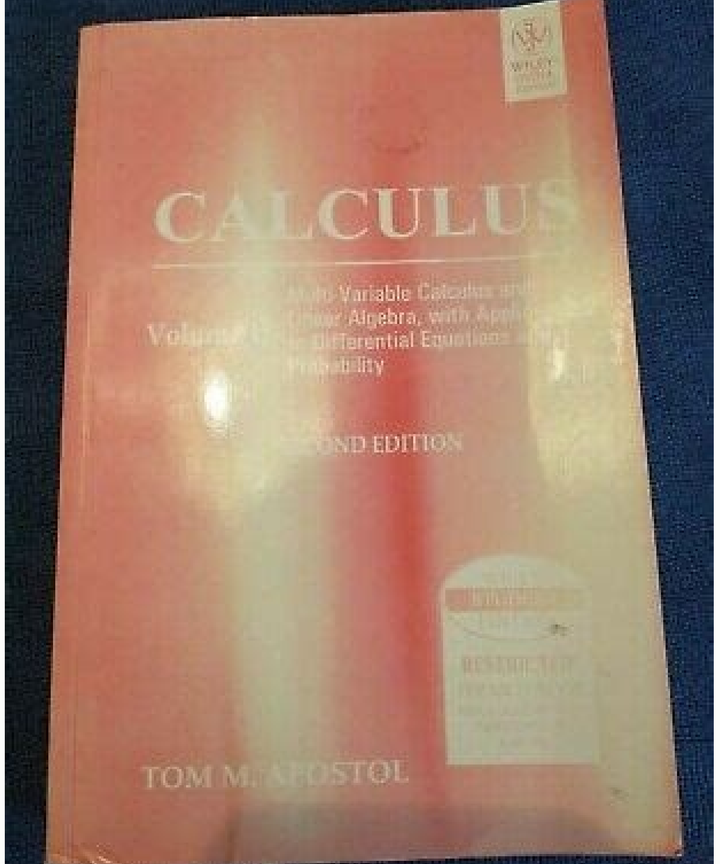
PREFACE This book is a continuation of the author’s Calculus, Volume I, Second Edition. The present volume has been written with the same underlying philosophy that prevailed in the first. Sound training in technique is combined with a strong theoretical development. Every effort has been made to convey the spirit of modern mathematics without undue emphasis on formalization. As in Volume I, historical remarks are included to give the student a sense of participation in the evolution of ideas. The second volume is divided into three parts, entitled Linear Analysis, Nonlinear Ana!ysis, and Special Topics. The last two chapters of Volume I have been repeated as the first two chapters of Volume II so that all the material on linear algebra will be complete in one volume. Part 1 contains an introduction to linear algebra, including linear transformations, matrices, determinants, eigenvalues, and quadratic forms. Applications are given to analysis, in particular to the study of linear differential equations.
Systems of differential equations are treated with the help of matrix calculus. Existence and uniqueness theorems are proved by Picard’s method of successive approximations, which is also cast in the language of contraction operators. Part 2 discusses the calculus of functions of several variables. Differential calculus is unified and simplified with the aid of linear algebra. It includes chain rules for scalar and vector fields, and applications to partial differential equations and extremum problems. Integral calculus includes line integrals, multiple integrals, and surface integrals, with applications to vector analysis. Here the treatment is along more or less classical lines and does not include a formal development of differential forms. The special topics treated in Part 3 are Probability and Numerical Analysis. The material on probability is divided into two chapters, one dealing with finite or countably infinite sample spaces; the other with uncountable sample spaces, random variables, and dis- tribution functions. The use of the calculus is illustrated in the study of both one- and two-dimensional random variables. The last chapter contains an introduction to numerical analysis, the chief emphasis being on different kinds of polynomial approximation. Here again the ideas are unified by the notation and terminology of linear algebra. The book concludes with a treatment of approximate integration formulas, such as Simpson’s rule, and a discussion of Euler’s summation formula. Academia.edu uses cookies to personalize content, tailor ads and improve the user experience. By using our site, you agree to our collection of information through the use of cookies. To learn more, view our Privacy Policy. Want more? Advanced embedding details, examples, and help! Tom IN.



Apostol CALCULUS VOLUME II Mlul ti Variable Calculus and Linear Algebra, with Applications to DifFeren tial Equations and Probability SECOND EDITION John Wiley & Sons New York London Sydney Toronto Want more? Advanced embedding details, examples, and help! Tom IN. Apostol CALCULUSVOLUME II Mlul ti Variable Calculus and LinearAlgebra, with Applications to DifFeren tial Equations and Probability SECOND EDITION John Wiley & Sons New York London Sydney Toronto C O N S U L T I N G E D I T O R George Springer, Indiana University COPYRIGHT 0 1969 BY XEROX CORPORATION.All rights reserved. No part of the material covered by this copyright may be produced in any form, or by any means of reproduction.Previous edition copyright 0 1962 by Xerox Corporation. Librar of Congress Catalog Card Number: 67-14605ISBN 0 471 00007 8 Printed in the United States of America. 1 0 9 8 7 6 5 4 3 2 To Jane and Stephen PREFACE This book is a continuation of the author’s Calculus, Volume I, Second Edition. Thepresent volume has been written with the same underlying philosophy that prevailed in thefirst. Sound training in technique is combined with a strong theoretical development.Every effort has been made to convey the spirit of modern mathematics without undueemphasis on formalization. As in Volume I, historical remarks are included to give thestudent a sense of participation in the evolution of ideas. The second volume is divided into three parts, entitled Linear Analysis, NonlinearAna!ysis, and Special Topics. The last two chapters of Volume I have been repeated as thefirst two chapters of Volume II so that all the material on linear algebra will be completein one volume. Part 1 contains an introduction to linear algebra, including linear transformations,matrices, determinants, eigenvalues, and quadratic forms. Applications are given toanalysis, in particular to the study of linear differential equations. Systems of differentialequations are treated with the help of matrix calculus. Existence and uniqueness theoremsare proved by Picard’s method of successive approximations, which is also cast in thelanguage of contraction operators. Part 2 discusses the calculus of functions of several variables. Differential calculus isunified and simplified with the aid of linear algebra. It includes chain rules for scalar andvector fields, and applications to partial differential equations and extremum problems.Integral calculus includes line integrals, multiple integrals, and surface integrals, withapplications to vector analysis. Here the treatment is along more or less classical lines anddoes not include a formal development of differential forms. The special topics treated in Part 3 are Probability and Numerical Analysis. The materialon probability is divided into two chapters, one dealing with finite or countably infinitesample spaces; the other with uncountable sample spaces, random variables, and dis-tribution functions. The use of the calculus is illustrated in the study of both one- andtwo-dimensional random variables. The last chapter contains an introduction to numerical analysis, the chief emphasisbeing on different kinds of polynomial approximation. Here again the ideas are unifiedby the notation and terminology of linear algebra.
The book concludes with a treatment ofapproximate integration formulas, such as Simpson’s rule, and a discussion of Euler’ssummation formula. \‘I11 Preface There is ample material in this volume for a full year’s course meeting three or four timesper week. It presupposes a knowledge of one-variable calculus as covered in most first-yearcalculus courses. The author has taught this material in a course with two lectures and tworecitation periods per week, allowing about ten weeks for each part and omitting thestarred sections. This second volume has been planned so that many chapters can be omitted for a varietyof shorter courses. For example, the last chapter of each part can be skipped withoutdisrupting the continuity of the presentation. Part 1 by itself provides material for a com-bined course in linear algebra and ordinary differential equations. The individual instructorcan choose topics to suit his needs and preferences by consulting the diagram on the nextpage which shows the logical interdependence of the chapters. Once again I acknowledge with pleasure the assistance of many friends and colleagues.In preparing the second edition I received valuable help from Professors Herbert S.Zuckerman of the University of Washington, and Basil Gordon of the University ofCalifornia, Los Angeles, each of whom suggested a number of improvements. Thanks arealso due to the staff of Blaisdell Publishing Company for their assistance and cooperation.
As before, it gives me special pleasure to express my gratitude to my wife for the manyways in which she has contributed. In grateful acknowledgement I happily dedicate thisbook to her. T. M.
A.Pasadena, CaliforniaSeptember 16, 1968 Logical Interdependence of the Chapters ix 1LINEARS P A C E S I 6LINEAR I I 7 T R A N S F O R M A T I O N SAND MATRICES D E T E R M I N A N T S \DIFFERENTIAL EQUATIONS4 E I G E N V A L U E S, I A N D EIGENVECTORSSYSTEMS OF DIFFERENTIALEQUATIONS I- INTRODUCTIONTO NUMERICAL A N A L Y S I S I I 10 1 1 13DIFFERENTIAL H LINE SET FUNCTIONSCALCULUS OF I N T E G R A L S AND ELEMENTARYSCALAR AND PROBABILITY VECTOR FIELDS OPERATORS ACTINGON EUCLIDEAN 1 S P A C E S 9’ APPLICATIONS OF DIFFERENTIALC A L C U L U S 1 P R OBABILITI E S 1 I N T E G R A L S CONTENTS PART 1. LINEAR ANALYSIS 1.
LINEAR SPACES 1.1 Introduction 1.2 The definition of a linear space 1.3 Examples of linear spaces 1.4 Elementary consequences of the axioms 1.5 Exercises 1.6 Subspaces of a linear space 1.7 Dependent and independent sets in a linear space 1.8 Bases and dimension 1.9 Components 1.10 Exercises 1.11 Inner products, Euclidean spaces. Norms 1.12 Orthogonality in a Euclidean space 1.13 Exercises 1.14 Construction of orthogonal sets. The Gram-Schmidt process 1.15 Orthogonal complements. Projections 1.16 Best approximation of elements in a Euclidean space by elements in a finite- 3 3 4 6 7 8 9 1 2 1 3 1 3 14 18 20 22 2 6 dimensional subspace 2 8 1.17 Exercises 3 0 2. LINEAR TRANSFORMATIONS AND MATRICES 2.1 Linear transformations 2.2 Null space and range 2.3 Nullity and rank 3 1 32 3 4 xi xii Contents 2.4 Exercises 2.5 Algebraic operations on linear transformations 2.6 Inverses 2.7 One-to-one linear transformations 2.8 Exercises 2.9 Linear transformations with prescribed values 2.10 Matrix representations of linear transformations 2.11 Construction of a matrix representation in diagonal form 2.12 Exercises 2.13 Linear spaces of matrices 2.14 Tsomorphism between linear transformations and matrices 2.15 Multiplication of matrices 2.16 Exercises 2.17 Systems of linear equations 2.18 Computation techniques 2.19 Inverses of square matrices 2.20 Exercises 2.21 Miscellaneous exercises on matrices 3. DETERMINANTS 3.1 Introduction 3.2 Motivation for the choice of axioms for a determinant function 3.3 A set of axioms for a determinant function 3.4 Computation of determinants 3.5 The uniqueness theorem 3.6 Exercises 3.7 The product formula for determinants 3.8 The determinant of the inverse of a nonsingular matrix 3.9 Determinants and independence of vectors 3.10 The determinant of a block-diagonal matrix 3.11 Exercises 3.12 Expansion formulas for determinants.



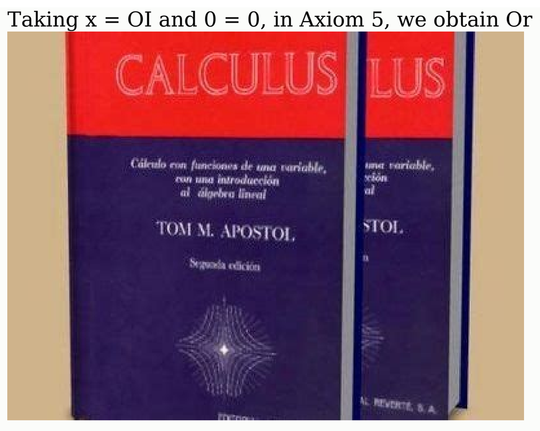
Minors and cofactors 3.13 Existence of the determinant function 3.14 The determinant of a transpose 3.15 The cofactor matrix 3.16 Cramer’s rule 3.17 Exercises 35 36 38 41 4 2 44 4 5 4 8 50 5 1 52 5 4 57 58 6 1 6 5 6 7 6 8 71 7 2 73 7 6 7 9 7 9 8 1 83 83 8 4 8 5 86 9 0 9 1 9 2 9 3 94 Contents. . . XIII 4. EIGENVALUES AND EIGENVECTORS 4.1 Linear transformations with diagonal matrix representations4.2 Eigenvectors and eigenvalues of a linear transformation4.3 Linear independence of eigenvectors corresponding to distinct eigenvalues4.4 Exercises *4.5 The finite-dimensional case. Characteristic polynomials4.6 Calculation of eigenvalues and eigenvectors in the finite-dimensional case4.7 Trace of a matrix4.8 Exercises4.9 Matrices representing the same linear transformation. Similar matrices4.10 Exercises 969 7 10010110210310610710811 2 5. EIGENVALUES OF OPERATORS ACTING ONEUCLIDEAN SPACES 5.1 Eigenvalues and inner products5.2 Hermitian and skew-Hermitian transformations5.3 Eigenvalues and eigenvectors of Hermitian and skew-Hermitian operators5.4 Orthogonality of eigenvectors corresponding to distinct eigenvalues5.5 Exercises5.6 Existence of an orthonormal set of eigenvectors for Hermitian and 114115117117118 skew-Hermitian operators acting on finite-dimensional spaces5 . 7 Matrix representations for Hermitian and skew-Hermitian operators5.8 Hermitian and skew-Hermitian matrices. The adjoint of a matrix5 . 9 Diagonalization of a Hermitian or skew-Hermitian matrix5.10 Unitary matrices. Orthogonal matrices5.11 Exercises5.12 Quadratic forms5.13 Reduction of a real quadratic form to a diagonal form5.14 Applications to analytic geometry5.15 Exercises A5.16 Eigenvalues of a symmetric transformation obtained as values of itsquadratic form k5.17 Extremal properties of eigenvalues of a symmetric transformationk5.18 The finite-dimensional case 5.19 Unitary transformations5.20 Exercises 120121122122123124126128130134 135136137138141 xiv Contents 6. LINEAR DIFFERENTIAL EQUATIONS 6.1 Historical introduction 6.2 Review of results concerning linear equations of first and second orders 6.3 Exercises 6.4 Linear differential equations of order n 6.5 The existence-uniqueness theorem 6.6 The dimension of the solution space of a homogeneous linear equation 6.7 The algebra of constant-coefficient operators 6.8 Determination of a basis of solutions for linear equations with constant coefficients by factorization of operators 6.9 Exercises6.10 The relation between the homogeneous and nonhomogeneous equations 6.11 Determination of a particular solution of the nonhomogeneous equation. The method of variation of parameters 6.12 Nonsingularity of the Wronskian matrix of n independent solutions of a homogeneous linear equation 6.13 Special methods for determining a particular solution of the nonhomogeneous equation. Reduction to a system of first-order linear equations 6.14 The annihilator method for determining a particular solution of the nonhomogeneous equation 6.15 Exercises 6.16 Miscellaneous exercises on linear differential equations 6.17 Linear equations of second order with analytic coefficients 6.18 The Legendre equation 6.19 The Legendre polynomials 6.20 Rodrigues‘ formula for the Legendre polynomials 6.21 Exercises 6.22 The method of Frobenius 6.23 The Bessel equation 6.24 Exercises 7. SYSTEMS OF DIFFERENTIAL EQUATIONS 7.1 Introduction 1 9 1 7 .2 Calculus of matrix functions 193 7.3 Infinite series of matrices. Norms of matrices 194 7.4 Exercises 195 7.5 The exponential matrix 197 142 143 144 145 147 147 148 150 154 156 157 1 6 1 163 163 166 167 169 1 7 1 174 176 177 180 182 188 Contents x v 7.6 The differential equation satisfied by etA 7.7 Uniqueness theorem for the matrix differential equation F’(t) = AF(t)7.8 The law of exponents for exponential matrices 7.9 Existence and uniqueness theorems for homogeneous linear systems 197 198 199 with constant coefficients 200 7.10 The problem of calculating etA4 201 7.11 The Cayley-Hamilton theorem 203 7.12 Exercises 205 7.13 Putzer’s method for calculating eta 205 7.14



Alternate methods for calculating et-’ in special cases 208 7.15 Exercises 211 7.16 Nonhomogeneous linear systems with constant coefficients 213 7.17 Exercises 215 7.18 The general linear system Y’(t) = P(t) Y(t) + Q(t) 217 7.19 A power-series method for solving homogeneous linear systems 220 7.20 Exercises 221 7.21 Proof of the existence theorem by the method of successive approximations 222 7.22 The method of successive approximations applied to first-order nonlinear systems 227 7.23 Proof of an existence-uniqueness theorem for first-order nonlinear systems 229 7.24 Exercises 230 *7.25 Successive approximations and fixed points of operators 232 *7.26 Normed linear spaces 233 k7.27 Contraction operators 234 k7.28 Fixed-point theorem for contraction operators 235 A7.29 Applications of the fixed-point theorem 237 PART 2. NONLINEAR ANALYSIS 8.1 8 .2 8.3 8 .4 8.5 8 .6 8.7 8.8 8.9 8. DIFFERENTIAL CALCULUS OF SCALAR ANDVECTOR FIELDS Functions from R” to R”. Scalar and vector fields Open balls and open sets Exercises Limits and continuity ExercisesThe derivative of a scalar field with respect to a vector Directional derivatives and partial derivatives Partial derivatives of higher order Exercises 243 244 245 247 251 252 254 255 255 xvi Contents 8.10 Directional derivatives and continuity 8.11 The total derivative 8.12 The gradient of a scalar field 8.13 A sufficient condition for differentiability 8.14 Exercises 8.15 A chain rule for derivatives of scalar fields 8.16 Applications to geometry. Level sets. Tangent planes 8.17 Exercises 8.18 Derivatives of vector fields 8. I9 Differentiability implies continuity 8.20 The chain rule for derivatives of vector fields 8.21 Matrix form of the chain rule 8.22 Exercises A8.23 Sufficient conditions for the equality of mixed partial derivatives 8.24 Miscellaneous exercises 251 258 259 261 262 263 266 268 269 271 272 273 275 277 281 9. APPLICATIONS OF THE DIFFERENTIAL CALCULUS 9.1 Partial differential equations 283 9.2 A first-order partial differential equation with constant coefficients 284 9.3 Exercises 286 9.4 The one-dimensional wave equation 288 9.5 Exercises 292 9.6 Derivatives of functions defined implicitly 294 9.7 Worked examples 298 9.8 Exercises 302 9.9 Maxima, minima, and saddle points 303 9.

IO Second-order Taylor formula for scalar fields 308 9.11 The nature of a stationary point determined by the eigenvalues of the Hessian matrix 310 9.12 Second-derivative test for extrema of functions of two variables 312 9.13 Exercises 313 9.14 Extrema with constraints.
Lagrange’s multipliers 314 9. I5 Exercises 318 9.16 The extreme-value theorem for continuous scalar fields 319 9.17 The small-span theorem for continuous scalar fields (uniform continuity) 321 10. LINE INTEGRALS10.1 Introduction 10.2 Paths and line integrals 323 323 Contents xvii 10.3 Other notations for line integrals 324 10.4 Basic properties of line integrals 326 10.5 Exercises 328 10.6 The concept of work as a line integral 328 10.7 Line integrals with respect to arc length 329 10.8 Further applications of line integrals 330 10.9 Exercises 331 10.10 Open connected sets. Independence of the path 332 10.11 The second fundamental theorem of calculus for line integrals 333 10.12 Applications to mechanics 335 10.13 Exercises 336 10.14 The first fundamental theorem of calculus for line integrals 331 10.15 Necessary and sufficient conditions for a vector field to be a gradient 339 10.16 Necessary conditions for a vector field to be a gradient 340 10.17 Special methods for constructing potential functions 342 10.18 Exercises 345 10.19 Applications to exact differential equations of first order 346 10.20 Exercises 349 10.21 Potential functions on convex sets 350 11. MULTIPLE INTEGRALS11 .l Introduction 353 11.2 Partitions of rectangles. Step functions 353 11.3 The double integral of a step function 355 11.4 The definition of the double integral of a function defined and bounded on a rectangle 11.5 Upper and lower double integrals 11.6 Evaluation of a double integral by repeated one-dimensional integration 11.7 Geometric interpretation of the double integral as a volume 11.8 Worked examples 11.9 Exercises 11.10 Integrability of continuous functions 11 .I 1 Integrability of bounded functions with discontinuities 11.12 Double integrals extended over more general regions 11.13 Applications to area and volume 11.14 Worked examples 11.
I5 Exercises11.16 Further applications of double integrals 11.17 Two theorems of Pappus 11.18 Exercises 357 357 358 359 360 362 363 364 365 368 369 371 373 376 377 .
. .xv111 Contents 11.19 Green’s theorem in the plane 11.20 Some applications of Green’s theorem 11.21 A necessary and sufficient condition for a two-dimensional vector field to be a gradient 11.22 Exercises kll.23 Green’s theorem for multiply connected regions *11.24 The winding number *I 1.25 Exercises 11.26 Change of variables in a double integral 11.27 Special cases of the transformation formula 11.28 Exercises11.29 Proof of the transformation formula in a special case 11.30 Proof of the transformation formula in the general case 11.31 Extensions to higher dimensions 11.32 Change of variables in an n-fold integral 11.33 Worked examples 11.34 Exercises 12. SURFACE INTEGRALS12.1 Parametric representation of a surface 12.2 The fundamental vector product 12.3 The fundamental vector product as a normal to the surface 12.4 Exercises 12.5 Area of a parametric surface 12.6 Exercises 12.7 Surface integrals 12.8 Change of parametric representation 12.9 Other notations for surface integrals 12.10 Exercises 12.11 The theorem of Stokes 12.12 The curl and divergence of a vector field 12.13 Exercises 12.14 Further properties of the curl and divergence 12.15 Exercises *12.16 Reconstruction of a vector field from its curl *12.17 Exercises 12.18 Extensions of Stokes’ theorem 12.19 The divergence theorem (Gauss’ theorem:)12.20 Applications of the divergence theorem 12.21 Exercises 378 382 383 385 387 389391 392 396 399 401 403 405 407 409 413 417 420 423 424 424 429 430 432 434 436 438 440 442 443 447 448 452 453 457 460 462 Contents x ix PART 3.
SPECIAL TOPICS 13. SET FUNCTIONS AND ELEMENTARY PROBABILITY 13.1 Historical introduction 13.2 Finitely additive set functions 13.3 Finitely additive measures 13.4 Exercises 13.5 The definition of probability for finite sample spaces 13.6 Special terminology peculiar to probability theory 13.7 Exercises 13.8 Worked examples 13.9 Exercises 13.10 Some basic principles of combinatorial analysis 13.11 Exercises 13.12 Conditional probability 13.13 Independence 13.14 Exercises 13.15 Compound experiments 13.16 Bernoulli trials 13.17 The most probable number of successes in n Bernoulli trials 13.18 Exercises 13.19 Countable and uncountable sets 13.20 Exercises 13.21 The definition of probability for countably infinite sample spaces 13.22 Exercises 13.23 Miscellaneous exercises on probability 14. CALCULUS OF PROBABILITIES 14.1 The definition of probability for uncountable sample spaces 14.2 Countability of the set of points with positive probability 14.3 Random variables 14.4 Exercises 14.5 Distribution functions 14.6 Discontinuities of distribution functions 14.7 Discrete distributions. Probability mass functions 14.8 Exercises 14.9 Continuous distributions. Density functions 469 470 471 472 473 475 477 477 479 481 485 486 488 490 492 495 497 499 501 504 506 507 507 510 511 512 513514 517 520 523 525 xx Contents 14.10 Uniform distribution over an interval 14.11 Cauchy’s distribution 14.12 Exercises 14.13 Exponential distributions 14.14 Normal distributions 14.15 Remarks on more general distributions 14.16 Exercises 14.17 Distributions of functions of random variables 14.18 Exercises14.19 Distributions of two-dimensional random variables 14.20 Two-dimensional discrete distributions 14.21 Two-dimensional continuous distributions. Density functions 14.22 Exercises 14.23 Distributions of functions of two random variables 14.24 Exercises 14.25 Expectation and variance 14.26 Expectation of a function of a random variable 14.27 Exercises 14.28 Chebyshev’s inequality 14.29 Laws of large numbers 14.30 The central limit theorem of the calculus of probabilities 14.3 1 Exercises Suggested References 15. INTRODUCTION TO NUMERICAL ANALYSIS 15.1 Historical introduction 15.2 Approximations by polynomials 15.3 Polynomial approximation and normed linear spaces 15.4 Fundamental problems in polynomial approximation 15.5 Exercises 15.6 Interpolating polynomials 15.7 Equally spaced interpolation points 15.8 Error analysis in polynomial interpolation 15.9 Exercises 15.10 Newton’s interpolation formula 15.11 Equally spaced interpolation points. The forward difference operator 15.12 Factorial polynomials 15.13 Exercises 15.14 A minimum problem relative to the max norm 526 530 532 533 535 539 540 541 542 543 545 546 548 550 553 556 559 560 562 564 566 568 569 571 572 574 575 577 579 582 583 585 588 590 592 593 595 Contents xxi 15.15 Chebyshev polynomials 15.16 A minimal property of Chebyshev polynomials 15.17 Application to the error formula for interpolation 15.18 Exercises15.19 Approximate integration. The trapezoidal rule 15.20 Simpson’s rule 15.21 Exercises 15.22 The Euler summation formula 15.23 ExercisesSuggested References Answers to exercises Index 596 598 599 600 602 605 610613 618 621 622 665 Calculus PART 1LINEAR ANALYSIS 1 LINEAR SPACES 1.1 Introduction Throughout mathematics we encounter many examples of mathematical objects thatcan be added to each other and multiplied by real numbers. First of all, the real numbersthemselves are such objects. Other examples are real-valued functions, the complexnumbers, infinite series, vectors in n-space, and vector-valued functions. In this chapter wediscuss a general mathematical concept, called a linear space, which includes all theseexamples and many others as special cases. Briefly, a linear space is a set of elements of any kind on which certain operations (calledaddition and multiplication by numbers) can be performed. In defining a linear space, wedo not specify the nature of the elements nor do we tell how the operations are to beperformed on them. Instead, we require that the operations have certain properties whichwe take as axioms for a linear space. We turn now to a detailed description of these axioms. 1.2 The definition of a linear space Let V denote a nonempty set of objects, called elements. The set V is called a linearspace if it satisfies the following ten axioms which we list in three groups. Closure axioms AXIOM 1. CLOSURE UNDER ADDITION. For every pair of elements x and y in V therecorresponds a unique element in V called the sum of x and y, denoted by x + y . AXIOM 2. CLOSURE UNDER MULTIPLICATION BY REAL NUMBERS. For every x in V andevery real number a there corresponds an element in V called the product of a and x, denotedby ax. Axioms for addition AXIOM 3. COMMUTATIVE LAW. For all x and y in V, we have x + y = y + x. AXIOM 4. ASSOCIATIVELAW. Forallx,y,andzinV,wehave(x+y) + z =x +(y+z). ? 4 Linear spaces AXIOM 5. EXISTENCEOFZEROELEMENT. There is an element in V, denoted by 0, such that x+0=x forallxin V . AXIOM 6. EXISTENCEOFNEGATIVES. For every x in V, the element (- 1)x has the property x+(-1)x= 0 . Axioms for multiplication by numbers AXIOM 7. ASSOCIATIVE LAW. For every x in V and all real numbers a and b, we have a(bx) = (ab)x. AXIOM 8. DISTRIBUTIVE LAW FOR ADDITION IN V. For all x andy in V and all real a,we hare a(x + y) = ax + ay . AXIOM 9. DISTRIBUTIVE LAW FOR ADDITION OF NUMBERS. For all x in V and all reala and b, we have (a + b)x = ax + bx. AXIOM 10. EXISTENCE OF IDENTITY. For every x in V, we have lx = x. Linear spaces, as defined above, are sometimes called real linear spaces to emphasizethe fact that we are multiplying the elements of V by real numbers. If real number isreplaced by complex number in Axioms 2, 7, 8, and 9, the resulting structure is called acomplex linear space. Sometimes a linear space is referred to as a linear vector space orsimply a vector space; the numbers used as multipliers are also called scalars.
A real linearspace has real numbers as scalars; a complex linear space has complex numbers as scalars.Although we shall deal primarily with examples of real linear spaces, all the theorems arevalid for complex linear spaces as well. When we use the term linear space without furtherdesignation, it is to be understood that the space can be real or complex.
1.3 Examples of linear spaces If we specify the set V and tell how to add its elements and how to multiply them bynumbers, we get a concrete example of a linear space. The reader can easily verify thateach of the following examples satisfies all the axioms for a real linear space. EXAMPLE 1. Let V = R , the set of all real numbers, and let x + y and ax be ordinaryaddition and multiplication of real numbers. EXAMPLE 2. Let V = C, the set of all complex numbers, define x + y to be ordinaryaddition of complex numbers, and define ax to be multiplication of the complex number x Examples of linear spaces by the real number a. Even though the elements of V are complex numbers, this is a reallinear space because the scalars are real.
EXAMPLE 3. Let V’ = V,, the vector space of all n-tuples of real numbers, with additionand multiplication by scalars defined in the usual way in terms of components. EXAMPLE 4. Let V be the set of all vectors in V, orthogonal to a given nonzero vectorIV. If n = 2, this linear space is a line through 0 with N as a normal vector. If n = 3,it is a plane through 0 with N as normal vector. The following examples are called function spaces. The elements of V are real-valuedfunctions, with addition of two functions f and g defined in the usual way: (f + g)(x) =f(x) + g(x) for every real x in the intersection of the domains off and g. Multiplication of a functionf by a real scalar a is defined as follows: af is that function whose value at each x in thedomain off is af (x). The zero element is the function whose values are everywhere zero.The reader can easily verify that each of the following sets is a function space. EXAMPLE 5. The set of all functions defined on a given interval. EXAMPLE 6. The set of all polynomials. EXAMPLE 7. The set of all polynomials of degree 5 n, where n is fixed. (Whenever weconsider this set it is understood that the zero polynomial is also included.) The set ofall polynomials of degree equal to IZ is not a linear space because the closure axioms are notsatisfied. For example, the sum of two polynomials of degree n need not have degree n. EXAMPLE 8. The set of all functions continuous on a given interval. If the interval is[a, b], we denote this space by C(a, b). EXAMPLE 9. The set of all functions differentiable at a given point. EXAMPLE 10. The set of all functions integrable on a given interval. EXAMPLE 11. The set of all functions f defined at 1 with f(1) = 0. The number 0 isessential in this example. If we replace 0 by a nonzero number c, we violate the closureaxioms. EXAMPLE 12. The set of all solutions of a homogeneous linear differential equationy” + ay’ + by = 0, where a and b are given constants. Here again 0 is essential. The setof solutions of a nonhomogeneous differential equation does not satisfy the closure axioms. These examples and many others illustrate how the linear space concept permeatesalgebra, geometry, and analysis. When a theorem is deduced from the axioms of a linearspace, we obtain, in one stroke, a result valid for each concrete example. By unifying 6 Linear spaces diverse examples in this way we gain a deeper insight into each. Sometimes special knowl-edge of one particular example helps to anticipate or interpret results valid for otherexamples and reveals relationships which might otherwise escape notice. 1.4 Elementary consequences of the axioms The following theorems are easily deduced from the axioms for a linear space. THEOREM 1.1. UNIQUENESS OF THE ZERO ELEMENT. in any linear space there is oneand only one zero element. Proof. Axiom 5 tells us that there is at least one zero element. Suppose there were two,say 0, and 0,.

Taking x = OI and 0 = 0, in Axiom 5, we obtain Or + O2 = 0,.Similarly, taking x = 02 and 0 = 0,) we find 02 + 0, = 02. But Or + 02 = 02 + 0,because of the commutative law, so 0, = 02. THEOREM 1.2. UNIQUENESS OF NEGATIVE ELEMENTS.



In any linear space every elementhas exactly one negative. That is, for every x there is one and only one y such that x + y = 0. Proof. Axiom 6 tells us that each x has at least one negative, namely (- 1)x. Supposex has two negatives, say y1 and yZ. Then x + yr = 0 and x + yZ = 0. Adding yZ to bothmembers of the first equation and using Axioms 5, 4, and 3, we find that and Y2 + (x + yd = (y2 + x) + y1 = 0 + y, = y1 + 0 = y,, Therefore y1 = y2, so x has exactly one negative, the element (- 1)x. Notation. The negative of x is denoted by -x. The difference y - x is defined to bethe sum y + (-x) . The next theorem describes a number of properties which govern elementary algebraicmanipulations in a linear space. THEOREM 1.3. In a given linear space, let x and y denote arbitrary elements and let a and b denote arbitrary scalars. Then we'have the following properties: (a) Ox = 0.(b) a 0 = 0 .(c) (- a) x = - (a x) = a (- x) .(d) I f a x = O , t h e n e i t h e r a = O o r x = O .(e) Ifax=ayanda#O, t h e n x = y .(f) Ifax=bxandx#O,thena=b. (g> -(x + y) = (-4 + C-y) = --x - y. (h) x + x = 2x, x + x +x = 3x, andingeneral, &x = nx. Exercises 7 We shall prove (a), (b), and (c) and leave the proofs of the other properties as exercises. Proof of (a). Let z = Ox. We wish to prove that z = 0. Adding z to itself and usingAxiom 9, we find that Now add -z to both members to get z = 0. Proof of(b). Let z = a0, add z to itself, and use Axiom 8. Proof of(c).
Let z = (-a)x. Adding z to ax and using Axiom 9, we find that z+ax=(-a)x+ax=(-a+a)x=Ox=O, so z is the negative of ax, z = -(ax). Similarly, if we add a(-~) to ax and use Axiom 8and property (b), we find that a(-~) = -(ax). 1.5 Exercises In Exercises 1 through 28, determine whether each of the given sets is a real linear space, ifaddition and multiplication by real scalars are defined in the usual way. For those that are not,tell which axioms fail to hold. The functions in Exercises 1 through 17 are real-valued. In Exer-cises 3, 4, and 5, each function has domain containing 0 and 1. In Exercises 7 through 12, eachdomain contains all real numbers. 1. All rational functions.2. All rational functionsflg, with the degree off < the degree ofg (including f = 0).3 . Allfwithf(0) = f (l) . 8.
All even functions.4 . Allfwith2f(O) = f (l) . 9. All odd functions.5. Allfwithf(1) = 1 +f(O). 10. All bounded functions.6. All step functions defined on [0, 11. 11. All increasing functions.7. Allfwithf(x)-Oasx+ +a. 12. All functions with period 2a. 13.



All f integrable on [0, l] with Ji f(x) dx = 0.14. All f integrable on [0, l] with JA f(x) dx > 0.15. All f satisfyingf(x) = f(l - x) for all x.16. All Taylor polynomials of degree < n for a fixed n (including the zero polynomial).17. All solutions of a linear second-order homogeneous differential equation' y" + P(x)y' + Q(x)y = 0, where P and Q are given functions, continuous everywhere.18. All bounded real sequences.
20. All convergent real series.19. All convergent real sequences. 21. All absolutely convergent real series.22.
All vectors (x, y, z) in V, with z = 0.23.
All vectors (x, y, z) in V, with x = 0 or y = 0.24. All vectors (x, y, z) in V, with y = 5x.25. All vectors (x, y, z) in V, with 3x + 4y = 1, z = 0.26. All vectors (x, y, z) in V, which are scalar multiples of (1,2, 3).27. All vectors (x, y, z) in V, whose components satisfy a system of three linear equations of the form : allx + a,,y + a13z = 0, azlx + a,,y + uz3z = 0 , CZ31X + U33Y + U33Z = 0. 8 Linear spaces 28.
All vectors in V,, that are linear combinations of two given vectors A and B.29. Let V = R+, the set of positive real numbers. Define the "sum" of two elements x and y in V to be their product x y (in the usual sense), and define "multiplication" of an element xin V by a scalar c to be xc. Prove that V is a real linear space with 1 as the zero element. 30. (a) Prove that Axiom 10 can be deduced from the other axioms.(b) Prove that Axiom 10 cannot be deduced from the other axioms if Axiom 6 is replaced by Axiom 6': For every x in V there is an element y in V such that x + y = 0.3 1. Let S be the set of all ordered pairs (x1, xZ) of real numbers. In each case determine whether or not S is a linear space with the operations of addition and multiplication by scalars definedas indicated. If the set is not a linear space, indicate which axioms are violated.(4 (x1,x2) + (y19y2) = (x1 +y1,x2 +y,), 4x1, x2) = @Xl) 0).(b) (-99x2) + (y1,y,) = (~1 +yl,O), 4X1,X,) = (ax,, ax,>. cc> (Xl, x2) + cy1,y2> = (Xl, x2 +y2>9 4x1, x2> = (~17 QX2>. (4 @1,x2) + (yl,y2) = (Ix, + x,l,ly1 +y,l)t 4x1, x2) = (lql, lq!l) f 32. Prove parts (d) through (h) of Theorem 1.3. 1.6 Subspaces of a linear space Given a linear space V, let S be a nonempty subset of V. If S is also a linear space, withthe same operations of addition and multiplication by scalars, then S is called a subspaceof V. The next theorem gives a simple criterion for determining whether or not a subset ofa linear space is a subspace.
THEOREM 1.4. Let S be a nonempty subset of a linear space V. Then S is a subspaceif and only if S satisfies the closure axioms. Proof. If S is a subspace, it satisfies all the axioms for a linear space, and hence, inparticular, it satisfies the closure axioms. Now we show that if S satisfies the closure axioms it satisfies the others as well. Thecommutative and associative laws for addition (Axioms 3 and 4) and the axioms formultiplication by scalars (Axioms 7 through 10) are automatically satisfied in S becausethey hold for all elements of V. It remains to verify Axioms 5 and 6, the existence of a zeroelement in S, and the existence of a negative for each element in S. Let x be any element of S. (S has at least one element since S is not empty.) By Axiom2, ax is in S for every scalar a. Taking a = 0, it follows that Ox is in S. But Ox = 0, byTheorem 1.3(a), so 0 E S, and Axiom 5 is satisfied. Taking a = - 1, we see that (-1)xis in S. But x + (- 1)x = 0 since both x and (- 1)x are in V, so Axiom 6 is satisfied inS. Therefore S is a subspace of V. DEFINITION. Let S be a nonempty subset of a linear space V. An element x in V of theform k x = z: cixi)i=l where x1,. . . , xk are all in S and cl, . .
.
, ck are scalars, is called a$nite linear combinationof elements of S. The set of alljnite linear combinations of elements of S satisjies the closureaxioms and hence is a subspace of V. We call this the subspace spanned by S, or the linearspan of S, and denote it by L(S). If S is empty, we dejne L(S) to be {0}, the set consistingof the zero element alone. Dependent and independent sets in a linear space 9 Different sets may span the same subspace. For example, the space V, is spanned byeach of the following sets of vectors: {i,j}, {i,j, i +j}, (0, i, -i,j, -j, i + j}. The spaceof all polynomialsp(t) of degree < n is spanned by the set of n + 1 polynomials (1, t, t2, . . .) P}. It is also spanned by the set {I, t/2, t2/3, . .
. , t"/(n + l>>, and by (1, (1 + t), (1 + t)", . . . ,(1 + t)"}. The space of all polynomials is spanned by the infinite set of polynomials(1, t, t2, . . .}. A number of questions arise naturally at this point. For example, which spaces can bespanned by a finite set of elements? If a space can be spanned by a finite set of elements,what is the smallest number of elements required? To discuss these and related questions,we introduce the concepts of dependence, independence, bases, and dimension. These ideaswere encountered in Volume I in our study of the vector space V, . Now we extend themto general linear spaces. 1.7 Dependent and independent sets in a linear space DEFINITION. A set S of elements in a linear space V is called dependent if there is a-finiteset of distinct elements in S, say x1, . . . , xg, and a corresponding set of scalars cl, . . . , c,,not all zero, such that An equation 2 c,x(= 0 with not all ci = 0 is said to be a nontrivial representation of 0.The set S is called independent ifit is not dependent. In this case, for all choices of distinctelements x1, . . . , xk in S and scalars cl, . . . , ck, ii cixi = O implies c1=c2=..*=ck=o. Although dependence and independence are properties of sets of elements, we also applythese terms to the elements themselves. For example, the elements in an independent setare called independent elements. If S is a finite set, the foregoing definition agrees with that given in Volume I for thespace V,. However, the present definition is not restricted to finite sets. EXAMPLE 1. If a subset T of a set S is dependent, then S itself is dependent. This islogically equivalent to the statement that every subset of an independent set is independent. EXPMPLE 2. If one element in S is a scalar multiple of another, then S is dependent. EXAMPLE 3. If 0 E S, then S is dependent. EXAMPLE 4. The empty set is independent, 10 Linear spaces Many examples of dependent and independent sets of vectors in V,, were discussed inVolume I. The following examples illustrate these concepts in function spaces. In eachcase the underlying linear space V is the set of all real-valued functions defined on the realline. EXAMPLE 5. Let ui(t) = co? t , uz(t) = sin2 t , us(f) = 1 for all real t. The Pythagoreanidentity shows that u1 + u2 - uQ = 0, so the three functions ui, u2, u, are dependent. EXAMPLE 6. Let uk(t) = tk for k = 0, 1,2. . . . , and t real. The set S = {u,, , ui, u2, . . .}is independent. To prove this, it suffices to show that for each n the n + 1 polynomialsu,, 4, *. * 3 u, are independent. A relation of the form 1 c,u, = 0 means that (1.1) -&tk = 0k=O for all real 1.
When t = 0, this gives co = 0 . Differentiating (1.1) and setting t = 0,we find that c1 = 0. Repeating the process, we find that each coefficient ck is zero. EXAMPLE 7. If a,,..., a, are distinct real numbers, the n exponential functions q(x) = ea@, . . . , u,(x) = eanr are independent.
We can prove this by induction on n. The result holds trivially whenn = 1 . Therefore, assume it is true for n - 1 exponential functions and consider scalarsCl, * *. 7 c, such that (1.2) tckeakx = 0.k=l Let alIf be the largest of the n numbers a,, . .
. , a,.
Multiplying both members of (1.2)by e-a~x, we obtain If k # M, the number ak - aAl is negative.
Therefore, when x + + co in Equation (1.3),each term with k # M tends to zero and we find that cnl = 0. Deleting the Mth term from(1.2) and applying the induction hypothesis, we find that each of the remaining n - 1coefficients c, is zero. THEOREM 1.5. Let s = {X1,..., xk} be an independent set consisting of k elements in alinear space V and let L(S) be the subspace spanned by S. Then every set of k + 1 elementsin L(S) is dependent. Proof. The proof is by induction on k, the number of elements in S. First supposek = 1. Then, by hypothesis, S consists of one element xi, where x1 # 0 since S isindependent. Now take any two distinct elements y1 and yZ in L(S). Then each is a scalar Dependent and independent sets in a linear space 1 1 multiple of x1, say y1 = clxl and yZ = cZxl, where c, and c2 are not both 0. Multiplyingy1 by c2 and y, by c1 and subtracting, we find that c2.h - c,y, = 0. This is a nontrivial representation of 0 soy, and y2 are dependent. This proves the theoremwhen k = 1 . Now we assume the theorem is true for k - 1 and prove that it is also true for k. Takeany set of k + 1 elements in L(S), say T = {yr , yZ, . . I , yk+r} . We wish to prove that Tisdependent. Since each yi is in L(S) we may write (1.4)k yi = 2 a, jxji-1 f o r e a c h i = 1,2,... , k + 1 . We examine all the scalars ai, that multiply x1 and split theproof into two cases according to whether all these scalars are 0 or not. CASE 1. ai, = 0 for every i = 1,2, .
. . , k + 1 . In this case the sum in (1.4) does notinvolve x1, so each yi in T is in the linear span of the set S' = {x2, . . . , xk} . But S' isindependent and consists of k - 1 elements. By the induction hypothesis, the theorem istrue for k - 1 so the set T is dependent.
This proves the theorem in Case 1. CASE 2. Not all the scaIars ai, are zero.
Let us assume that a,, # 0. (If necessary, wecan renumber the y's to achieve this.) Taking i = 1 in Equation (1.4) and multiplying bothmembers by ci, where ci = ail/all, we get k Ciyl = ailxl + 1 CiUl jXj .j=2 From this we subtract Equation (1.4) to get k C,yl - yi = x(Cial j - aij>xj 3j=2 fori=2,..., k + 1 . This equation expresses each of the k elements ciy, - yi as a linearcombination of the k - 1 independent elements .x2, . .
. , xk . By the induction hypothesis,the k elements ciy, - yi must be dependent. Hence, for some choice of scalars t,, . . . ,tk+l, not all zero, we have kfl iz2ti(ciYl - Yi) = O 9 from which we find But this is a nontrivial linear combination of y,, . . . , yh.+l which represents the zero ele-ment, so the elements y1 , . . . , yri.r must be dependent. This completes the proof. 1 2 Linear spaces 1.8 Bases and dimension DEFINITION. A jinite set S of elements in a linear space V is called aJnite basis .for Vif S is independent and spans V. The space V is called$nite-dimensional if it has a jinitebasis, or if V consists of 0 alone. Otherwise, V is called injinite-dimensional. THEOREM 1.6. Let V be a jnite-dimensional linear space. Then every jnite basis for Vhas the same number of elements. Proof. Let S and T be two finite bases for V. Suppose S consists of k elements and Tconsists of m elements. Since S is independent and spans V, Theorem 1.5 tells us thatevery set of k + 1 elements in Vis dependent. Therefore, every set of more thank elementsin V is dependent. Since T is an independent set, we must have m 5 k. The same argu-ment with S and T interchanged shows that k < m . Therefore k = m . DEFINITION. If a linear space V has a basis of n elements, the integer n is called thedimension of V. We write n = dim V. If V = {O}!, we say V has dimension 0. EXAMPLE 1. The space V, has dimension n. One basis is the set of n unit coordinatevectors. EXAMPLE 2. The space of all polynomials p(t) of degree < n has dimension n + 1 . Onebasis is the set of n + 1 polynomials (1, t, t2, .
. . , t'"}. Every polynomial of degree 5 n is alinear combination of these n + 1 polynomials. EXAMPLE 3. The space of solutions of the differential equation y" - 2y' - 3y = 0 hasdimension 2. One basis consists of the two functions ul(x) = e-", u2(x) = e3x, Everysolution is a linear combination of these two. EXAMPLE 4.
The space of all polynomials p(t) is infinite-dimensional. Although theinfinite set (1, t, t2, . . .} spans this space, no$nite set of polynomials spans the space. THEOREM 1.7. Let V be a jinite-dimensional linear space with dim V = n. Then wehave the following: (a) Any set of independent elements in V is a s&set of some basis for V.(b) Any set of n independent elements is a basisf;pr V. Proof. To prove (a), let S = {x1, . . . , xk} be any independent set of elements in V.If L(S) = V, then S is a basis. If not, then there is some element y in V which is not inL(S). Adjoin this element to S and consider the new set S' = {x1, . . . , xk, y} . If thisset were dependent there would be scalars cl, . .
. , c~+~, not all zero, such that izlCiXi + cktly = 0 * But Ck+l # 0 since xi, . .
. , xk are independent. Hence, we could solve this equation fory and find that y E L(S), contradicting the fact that y is not in L(S). Therefore, the set S' Exercises 13 is independent but contains k + 1 elements. If L(S') = V, then S' is a basis and, since Sis a subset of S', part (a) is proved. If S' is not a basis, we can argue with S' as we didwith S, getting a new set S" which contains k + 2 elements and is independent. If S" is abasis, then part (a) is proved. If not, we repeat the process. We must arrive at a basis ina finite number of steps, otherwise we would eventually obtain an independent set withit + 1 elements, contradicting Theorem 1.5. Therefore part (a) is proved. To prove (b), let S be any independent set consisting of II elements. By part (a), S is asubset of some basis, say B. But by Theorem 1.6, the basis B has exactly n elements, soS= B. 1.9 Components Let V be a linear space of dimension II and consider a basis whose elements e, , . . . , e,are taken in a given order. We denote such an ordered basis as an n-tuple (e,, . . . , e,).If x E V, we can express x as a linear combination of these basis elements: (1.5) x = $ ciei .i=l The coefficients in this equation determine an n-tuple of numbers (c,, . . . , CJ that isuniquely determined by x.
In fact, if we have another representation of x as a linearcombination of e,, . . .
, e,, say x = I7z1 d,e,, then by subtraction from (1.5), we find that& (ci - d,)e, = 0. Bu since the basis elements are independent, this implies ci = ditforeachi,sowehave(c, ,..., c,)=(d, ,..., d,). The components of the ordered n-tuple (c,, . . .
, CJ determined by Equation (1.5) arecalled the components of x relative to the ordered basis (e, , . . . , e,). 1.10 Exercises In each of Exercises 1 through 10, let S denote the set of all vectors (x, y, z) in V3 whose com-ponents satisfy the condition given. Determine whether S is a subspace of V3. If S is a subspace,compute dim S. 1. X = 0.
6.x=y o r x=z.2 . x + y = o . 7. x2-y2=o.3.x+y+z=o. 8. x fy = 1.4 . x = y . 9..y=2x a n d z=3x.5 . x =y =z. 10. x + J + z = 0 and x - y - z = 0.
Let P, denote the linear space of all real polynomials of degree < it, where n is fixed. In eachof Exercises 11 through 20, let S denote the set of all polynomials f in P, satisfying the conditiongiven. Determine whether or not S is a subspace of P, . If S is a subspace, compute dim S. 11. f(0) = 0. 16. f(0) = f(2) .12. j-'(O) = 0. 17. f is even.13. j-"(O) = 0. 18. f is odd.14. f(O) +f'(o> = 0. 19. f has degree _< k, where k < n, or f = 0.15. f(0) =f(l). 20.
f has degree k, where k < n , or f = 0. 21. In the linear space of all real polynomials p(t), describe the subspace spanned by each of thefollowing subsets of polynomials and determine the dimension of this subspace.6-4 (1, t2, t4>; (b) {t, t3, t5>; cc> 0, t2> ; (d) { 1 + t, (1 + t,"}. 14 Linear spaces 22. In this exercise, L(S) denotes the subspace spanned by a subset S of a linear space V. Proveeach of the statements (a) through (f).(a) S G L(S).(b) If S G T G V and if T is a subspace of V, then L(S) c T. This property is described by saying that L(S) is the smallest subspace of V which contains 5'.(c) A subset S of V is a subspace of V if and only if L(S) = S.(d) If S c T c V, then L(S) c L(T).(e) If S and Tare subspaces of V, then so is S n T.(f) If S and Tare subsets of V, then L(S n T) E L(S) n L(T).(g) Give an example in which L(S n T) # L(S) ~-1 L(T). 23. Let V be the linear space consisting of all real-valued functions defined on the real line.Determine whether each of the following subsets of V is dependent or independent. Computethe dimension of the subspace spanned by each set. I%: i',ea2,ebz},a #b. (f) {cos x, sin x>. ear, xeax}. (g) {cosz x, sin2 x}. iz il, eaz, xeaz). (h) {'I, cos 2x, sin2 x}.eax, xeax, x2eax}. (i) {sin x, sin 2x}. (e) {e", ec", cash x}. (j) {e" cos x, eP sin x}.24. Let V be a finite-dimensional linear space, and let S be a subspace of V. Prove each of the following statements.(a) S is finite dimensional and dim S 2 dim V.(b) dim S = dim V if and only if S = V.(c) Every basis for S is part of a basis for V.(d) A basis for V need not contain a basis for S. 1.11 Inner products, Euclidean spaces. Norms In ordinary Euclidean geometry, those properties that rely on the possibility of measuringlengths of line segments and angles between lines are called metric properties. In our studyof V,, we defined lengths and angles in terms of the dot product. Now we wish to extendthese ideas to more general linear spaces. We shall introduce first a generalization of thedot product, which we call an inner product, and then define length and angle in terms of theinner product. The dot product x *y of two vectors x = (x1, . . . , x,) and y = (ul, . . . , yn) in V, wasdefined in Volume I by the formula (1.6) x * y = i x,y,.i=I In a general linear space, we write (x, JJ) instead of x * y for inner products, and we definethe product axiomatically rather than by a specific formula. That is, we state a number ofproperties we wish inner products to satisfy and we regard these properties as axioms. DEFINITION. A real linear space V is said to have an inner product if for each pair ofelements x and y in V there corresponds a unique real number (x, y) satisfying the followingaxioms for all choices of x, y, z in V and all real scalars c. (1) (XT y> = oi, 4 (commutativity, or symmetry).
(2) (x, y + z> = (x, y> + (x3 z> (distributivity, or linearity). (3) 4x2 .Y> = (cx, Y> (associativity, or homogeneity). (4) (x3 x> > 0 if x#O (positivity). Inner products, Euclidean spaces.
Norms 1 5 A real linear space with an inner product is called a real Euclidean space. Note: Taking c = 0 in (3), we find that (0,~) = 0 for all y. In a complex linear space, an inner product (x, y) is a complex number satisfying thesame axioms as those for a real inner product, except that the symmetry axiom is replacedby the relation (1') (X>Y> = (YP 4, (Hermitian?
symmetry) where (y, x) denotes the complex conjugate of (y, x). In the homogeneity axiom, the scalarmultiplier c can be any complex number. From the homogeneity axiom and (l'), we getthe companion relation - _ _(3') (x, cy) = (cy, x) = Q, x) = qx, y).
A complex linear space with an inner product is called a complex Euclidean 'space.(Sometimes the term unitary space is also used.) One example is complex vector spaceV,(C) discussed briefly in Section 12.16 of Volume I. Although we are interested primarily in examples of real Euclidean spaces, the theoremsof this chapter are valid for complex Euclidean spaces as well. When we use the termEuclidean space without further designation, it is to be understood that the space can bereal or complex.
The reader should verify that each of the following satisfies all the axioms for an innerproduct.
EXAMPLE 1.
In I', let (x, y) = x . y , the usual dot product of x and y. EXAMPLE 2. If x = (xi, XJ and y = (yi , yJ are any two vectors in V,, define (x, y) bythe formula (x3 Y) = %Yl + XlY2 + X2Yl + X2Y2 * This example shows that there may be more than one inner product in a given linear space. EXAMPLE 3. Let C(a, b) denote the linear space of all real-valued functions continuouson an interval [a, b]. Define an inner product of two functions f and g by the formula CL d = jab J-(&At) dt . This formula is analogous to Equation (1.6) which defines the dot product of two vectorsi n I!,. The function values f(t) and g(t) play the role of the components xi and yi , andintegration takes the place of summation. t In honor of Charles Hermite (1822-1901), a French mathematician who made many contributions toalgebra and analysis. 1 6 Linear spaces EXAMPLE 4. In the space C(a, b), define u-3 d = jab W(W(Od~) dt, where w is a fixed positive function in C(a, b). The function w is called a weightfunction.In Example 3 we have w(t) = 1 for all t. EXAMPLE 5. In the linear space of all real polynomials, define CL d = jam e-tf(MO dt.Because of the exponential factor, this improper integral converges for every choice ofpolynomials /and g. THEOREM 1.8. In a Euclidean space V, every inner product satisfies the Cauchy-Schwarzinequality: I(x,y)12 5 (x, x)(y, y) for all x andy in V. Moreover, the equality sign holds lyand only if x and y are dependent.
Proof. If either x = 0 or y = 0 the result holds trivially, so we can assume that bothx and y are nonzero. Let z = ax + by, where a and b are scalars to be specified later. Wehave the inequality (z, z) >_ 0 for all a and b. When we express this inequality in terms of xand y with an appropriate choice of a and b we will obtain the Cauchy-Schwarz inequality. To express (z, z) in terms of x and y we use properties (l'), (2) and (3') to obtain (z,Z> = (ax + by, ax + by) = (ax, ax) + (ax, by) + (by, ax) + (by, by) = a@, x> + a&x, y) + bii(y, x) + b&y, y) 2 0. Taking a = (y, y) and cancelling the positive factor (J, y) in the inequality we obtain 01, y>(x, 4 + 6(x, y> + Ny, xl + b6 2 0. Now we take b = -(x, y) . Then 6 = - (y, x) and the last inequality simplifies to (Y, y)(x, x) 2 (x, y>c.Y9 x> = I(& yv. This proves the Cauchy-Schwarz inequality. The equality sign holds throughout the proofif and only if z = 0. This holds, in turn, if and only if x and y are dependent. EXAMPLE. Applying Theorem 1.8 to the space C(a, b) with the inner product (f, g) =j,bf(t)g(t) dt , we find that the Cauchy-Schwarz inequality becomes (jbf(MO dt)' I (jabfZW dt)(jab g"(t) dl).a Inner products, Euclidean spaces. Norms 17 The inner product can be used to introduce the metric concept of length in any Euclideanspace. DEFINITION. In a Euclidean space V, the nonnegative number IIx I/ deJned by the equation llxjl = (x, x)" is called the norm of the element x. When the Cauchy-Schwarz inequality is expressed in terms of norms, it becomes IGGY)I 5 llxll M . Since it may be possible to define an inner product in many different ways, the normof an element will depend on the choice of inner product. This lack of uniqueness is to beexpected. It is analogous to the fact that we can assign different numbers to measure thelength of a given line segment, depending on the choice of scale or unit of measurement.The next theorem gives fundamental properties of norms that do not depend on the choiceof inner product. THEOREM 1.9. In a Euclidean space, every norm has the following properties for allelements x and y and all scalars c: (4 II-4 = 0 if x=0. @I II4 > 0 if x#O (positivity).
cc> Ilcxll = IcIll4 (homogeneity). (4 Ilx + YII I l/x/I + Ilyll (triangle inequality).The equality sign holds in (d) if x = 0, ify = 0, or if y = cxfor some c > 0. Proof. Properties (a), (b) and (c) follow at once from the axioms for an inner product.To prove (d), we note that Il.~+yl12=(~+y,~+y>=~~,~~+~y,y>+~~,y>+cv,~> = lIxl12 + llyl12 + (x3 y> + t-x, y>. The sum (x, y) + (x, y) is real. The Cauchy-Schwarz inequality shows that 1(x, y)l 5 II-4 llyll and IGG y)I I I I4 llyll , so w e have /lx + yl12 I lIxl12 + llYl12 + 2llxll llyll = Wll + llyll>". This proves (d). When y = cx , where c > 0, we have /lx +yII = IIX + cxll = (1 + c> IL-II = llxll + IICXII = I I4 + llyll . 1 8 Linear spaces DEFINITION. In a real Euclidean space V, the angle between two nonzero elements x andy is dejned to be that number t9 in the interval 0 5 8 < TT which satisfies the equation (1.7)(x9 Y)cos e = - IIXII llvll ' Note: The Cauchy-Schwarz inequality shows that the quotient on the right of (1.7)lies in the interval [- 1 , 11, so there is exactly one 0 in [0, 7~1 whose cosine is equalto this quotient.
1.12 Orthogonality in a Euclidean space DEFINITION. In a Euclidean space V, two elements x and y are called orthogonal if theirinner product is zero. A subset S of V is calIed an orthogonal set if (x, y) = 0 for every pairof distinct elements x and y in S. An orthogonal set is called orthonormal if each of itselements has norm 1. The zero element is orthogonal to every element of V; it is the only element orthogonal toitself. The next theorem shows a relation between orthogonality and independence. THEOREM 1.10. In a Euclidean space V, every orthogonal set of nonzero elements isindependent. In particular, in a jinite-dimensional Euclidean space with dim V = n, everyorthogonal set consisting of n nonzero elements is a basis for V.
Proof. Let S be an orthogonal set of nonzero elements in V, and suppose some finitelinear combination of elements of S is zero, say where each xi E S.
Taking the inner product of each member with x1 and using the factthat (xi , xi) = 0 if i # 1 , we find that cl(xl, x1) = 0. But (x1, x,) # 0 since xi # 0 soc1 = 0. Repeating the argument with x1 replaced by xi, we find that each cj = 0. Thisproves that S is independent. If dim V = n and if S consists of n elements, Theorem 1.7(b)shows that S is a basis for V. EXAMPLE. In the real linear space C(O,27r) with the inner product (f, g) = JiBf(x)g(x) dx,let S be the set of trigonometric functions {u,, ul, u2, .
. .} given by %&4 = 1, uznpl(x) = cos nx, uZn(x) = sin nx, f o r n = 1,2,.... If m # n, we have the orthogonality relations s 2n 0u~(x)u,(x) dx = 0, Orthogonality in a Euclidean space 1 9 so S is an orthogonal set. Since no member of S is the zero element, S is independent. Thenorm of each element of S is easily calculated. We have (u,, uO) = j'$' dx = 27r and, forn 2 1, we have (~1~~~~) u2,r-l -wo cos2 nx dx = T,1 - i"" (uzT1, uzvr) =Ib?" sin2 nx dx = T. Therefore, iluOll = J% and llu,/l = V% for n 2 1 . Dividing each zd, by its norm, weobtain an orthonormal set {pO, cpi , yz, . . .} where ~j~ = u,/~Iu,lI Thus, we have q+)(x) = 1J2n ' cp2,Ax) = y > ql,,(x) = s=) \'Gf o r n>l In Section 1.14 we shall prove that every finite-dimensional Euclidean space has anorthogonal basis. The next theorem shows how to compute the components of an elementrelative to such a basis. THEOREM I .l 1. Let V he a finite-dimerwionai Euclidean space with dimension n, andwume that S = {El, . . . , e,>IS an orthogonal basis,fbr V. [fan element x is expressed asg linear combination of the basis elements, say :1.8) .x = f c,e, ,c=l then its components relative to the ordered basis (e, , . . . , e,> are given by the,formulas '1.9) rn particular, if S is an orthonormal basis, each cj is given by 11.10) cj = (x, ej) . Proof. Taking the inner product of each member of (1.8) with ej, we obtain (X, ej) = i c,(ei, ej) = c,(ej, ej)i=-;* Ante (ei, eJ = 0 if i #j. This implies (1.9), and when (ej, ej) = 1, we obtain (1.10). If {e,, . . . , e,} is an orthonormal basis, Equation (1.9) can be written in the form 1.11) X = f (x, ei)ei .i=l The next theorem shows that in a finite-dimensional Euclidean space with an orthonormaloasis the inner product of two elements can be computed in terms of their components. 20 Linear spaces THEOREM 1.12.
Let V be a$nite-dimensional Euclidean space of dimension n, and assumefhat {e,, . . . , e,} is an orthonormal basis for V. Then for every puir of elements x and y in V,we have (1.12) (Parseval's formula). In particular, when x = y , we have Proof, Taking the inner product of both members of Equation (1.11) withy and usingthe linearity property of the inner product, we obtain (1.12). When x = y, Equation(1.12) reduces to (1.13).
Note: Equation (1.12) is named in honor of hf.
A. ParsevaI (circa 1776-1836), whoobtained this type of formula in a special function space.
Equation (1.13) is ageneralization of the theorem of Pythagoras. 1.13 Exercises 1. Let x = (x1, . . . , x,) andy = (yl, .
. . , yn) be arbitrary vectors in V, . In each case, determinewhether (x, y) is an inner product for V,, if (x, y) is defined by the formula given. In case(x, y) is not an inner product, tell which axioms are not satisfied. (4 (A y> = 5 xi lyili=l (4 (x, y) = (i&:yf)"2 .
(4 0, y> = j$ (xi + yd2 - t$lx1 - $IIK. 2. Suppose we retain the first three axioms for a real inner product (symmetry, linearity, andhomogeneity but replace the fourth axiom by a new axiom (4'): (x, x) = 0 if and only ifx = 0. Prove that either (x, x) > 0 for all x # 0 or else (x, x) < 0 for all x # 0. [Hint: Assume (x, x) > 0 for some x # 0 and (y, y) < 0 for some y # 0. In thespace spanned by {x, y), find an element z # 0 with (z, z) = 0.1 Prove that each of the statements in Exercises 3 through 7 is valid for all elements x and y in areal Euclidean space. 3. (x, y) = 0 if and only if /Ix + yll = l/x - yl/ .4. (x, y) = 0 if and only if 11x + yj12 = j/x112 + 11~11~.5. (x, y) = 0 if and only if 11x + cyll >]jxll for all real c.6. (x + y, x - y) = 0 if and only if (Ix/I = liyjj.7. If x and y are nonzero elements making an angle 0 with each other, then IIX - yl12 = llxl12 + Ilyl12 - 2 IIXII llyll cos 0. Exercises 8. In the real linear space C(l, e), define an inner product by the equation 2 1 Cf, g> = 1: (log xv-Wg(4 dx . (a) Iff(x) = G, compute lifi1.(b) Find a linear polynomial g(x) = a + bx that is orthogonal to the constant functionf(x) = 1. 9. In the real linear space C(- 1, l), let (f, g) = j', f(t)g(t) dt .
Consider the three functionsUl, u2, u, given by u,(t) = 1 , u2(t) = t 7 Y+(t) = 1 + t.
Prove that two of them are orthogonal, two make an angle n/3 with each other, and twomake an angle n/6 with each other. 10. In the linear space P, of all real polynomials of degree 5 n, define (f3g) = -$ f(3 g(k).k=O (a) Prove that (f,g) is an inner product for P,, .(b) Compute (f, g) whenf(t) = t and g(t) = at + b .(c) Iff(t) = t , find all linear polynomials g orthogonal tof. Il. In the linear space of all real polynomials, define (f,g) = sr e&f(t)g(t) dt .(a) Prove that this improper integral converges absolutely for all polynomials f and g.(b)Ifx,(t)=t"forn=0,1,2,...,provethat(x,,;r,)=(m+n)!.(c) Compute (f,g) when f(t) = (t + 1)2 andg(t) = t2 + 1 .(d) Find all linear polynomialsg(t) = a + bt orthogonal tof(t) = 1 + t. 12. In the linear space of all real polynomials, determine whether or not (f, g) is an inner productif (f,g) is defined by the formula given. In case (f,g) is not an inner product, indicate whichaxioms are violated.
In (c), f' and g ' denote derivatives. (a> (f,g) =f(l)g(l). (b) (f,g) = (j";f(t)g(t) dt (. (cl (f,g> = /;f'(t)g'(t)dt. Cd) Cftg) = (j-;f(f, df)(j-;g(t) dtj . 13. Let Vconsist of all infinite sequences {x,} of real numbers for which the series Cx: converges.If x = {x~} and y = {y,} are two elements of V, define (a) Prove that this series converges absolutely. [Hint: Use the Cauchy-Schwarz inequality to estimate the sum 2:&t Ixnynl.] (b) Prove that V is a linear space with (x, y) as an inner product.(c) Compute (x, y) if x, = l/n andy, = l/(n + 1) for n 2 1.(d) Compute (x, v) if x, =2"andy, = l/n!forn 2 1. 14. Let V be the set of all real functions f continuous on [0, + a) and such that the integralsc e-Ff2(t) dt converges. Define (f,g) = j"r ebf(t)g(t) dt .
22 Linear spaces (a) Prove that the integral for (f, g) converges absolutely for each pair of functions f and gin V. [Hint: Use the Cauchy-Schwarz inequality to estimate the integral jf e-t 1 f (t)g(t)l dt.] (b) Prove that V is a linear space with (f, g) as an inner product.(c) Compute (f,g) iff(t) = e& and&t) = P, where n = 0, 1,2, ... 15.
In a complex Euclidean space, prove that the inner product has the following properties forall elements X, y and z, and all complex a and b.(4 (ax, by) = &x, y). (b) (x, ay + bz) = rf(x, y) + 6(x, z). 16.
Prove that the following identities are valid in every Euclidean space.(a) Ilx +yl12 = l/xl12 + llyl12 + (x,y> + (y, x).(b) I/x + yl12 - lb - yl12 = 2(x, y) + xy, 4.(4 l/x + yl12 + lx - yl12 = 2 llxl12 + 2 IIy112. 17.
Prove that the space of all complex-valued functions continuous on an interval [a, b] becomesa unitary space if we define an inner product by the formula (fvg) = s: Wf(QgO4 where w is a fixed positive function, continuous on [a, b]. 1.14 Construction of orthogonal sets. The Gram-Scltmidt process Every finite-dimensional linear space has a finite basis. If the space is Euclidean, we canalways construct an orthogonal basis. This result will be deduced as a consequence of ageneral theorem whose proof shows how to construct orthogonal sets in any Euclideanspace, finite or infinite dimensional.
The construction is called the Gram-Schmidt orthog-onalizationprocess, in honor of J. P. Gram (1850-1916) and E. Schmidt (18451921). THEOREM 1.13. ORTHOGONALIZATION THEOREM. Let x1,x2,. .., be ajinite or intnitesequence of elements in a Euclidean space V, and let L(x,, . . . , xk) denote the subspacespanned by thejrst k of these elements. Then there is a corresponding sequence of elementsy1,y2, * * * 9 in V which has the following properties for each integer k: (a) The element yr is orthogonal to every element in the subspace L(yl, . . . , yk-J.(b) The subspace spanned by yl, . . . , yk is the same as that spanned by x1, . .
. , x, : uyl,. . . ,yJ = L(x,, . . . , XTJ. (c) The sequence yl, y, , . . . , is unique, except for scalar factors. That is, ifyi , y: , . . .
, isanother sequence of elements in V satisfying properties (a) and (b) for all k, then for each kthere is a scalar ck such that y; = ckyr . Proof. We construct the elements yr, y2, . . . , by induction. To start the process, wetake yr = x1. Now assume we have constructed yl, . . . , y,. so that (a) and (b) are satisfiedwhen k = r . Then we define y,.+r by the equation (1.14) Yr+1 = Xr+l - & ad+ y Construction of orthogonal sets.
The Gram-Schmidt process 2 3 where the scalars a,, . . .
, a, are to be determined. For j < r, the inner product of yI+rwith yj is given by since (yi, yj) = 0 if i #j . If yj # 0, we can make yr+r orthogonal to yj by taking (1.15) a _ (XT+1 7 .Yi)3 (Yj, Yi) ' If yj = 0, then yr+i is orthogonal to yj for any choice of aj, and in this case we chooseaj=O. Thus, the element Y?+~ is well defined and is orthogonal to each of the earlierelements yr , . .
. , y, . Therefore, it is orthogonal to every element in the subspace This proves (a) when k = r + 1.To prove (b) when k = r + 1, we must show that L(y,, .
.
.
, y,.+J = L(x, , . . .
, x,+r), given that L(y,, . . . , yr) = L(x,, . . . , x,) . The first r elements yl, . . . , y,. are in and hence they are in the larger subspace L(x, , . . . , x,+~). The new element yrsl given by(1.14) is a difference of two elements in ,5(x,, . . . , , , x,+~) so it, too, is in L(x,, . . . , x,+r).This proves that Equation (1.14) shows that x,+i is the sum of two elements in LQ, , . . . , yr+r) so a similarargument gives the inclusion in the other direction: UXl, . . . 9 x,+1) s uyl, . . . ,y7+1).
This proves (b) when k = r + 1. Therefore both (a) and (b) are proved by induction on k.Finally we prove (c) by induction on k. The case k = 1 is trivial. Therefore, assume (c) is true for k = r and consider the element y:+r . Because of (b), this element is in so we can write Yk+* =;ciyi = ZT + Cr+lYr+l, where z, E L(y,, . . . , y,.) . We wish to prove that z, = 0. By property (a), both vi+,, and~,+ry,.+~ are orthogonal to z, . Therefore, their difference, z,, is orthogonal to z, . In otherwords, z, is orthogonal to itself, so z, = 0.
This completes the proof of the orthogonaliza-tion theorem. 24 Linear spaces In the foregoing construction, suppose we have Y?,.~ = 0 for some r. Then (1.14)shows that x,+~ is a linear combination of yl, . .
. ,y,, and hence of x1, . . . , xc, so theelements x1, . . . , x,.,, are dependent.
In other words, if the first k elements x1,. . . , x,are independent, then the corresponding elements y1 , . . . , yk are nonzero.
In this case thecoefficients ai in (1.14) are given by (1.15), and the formulas defining y, , .
. .
, yk become (1.16) y, = x1, Yr+l = %+1 - for r = 1,2, . . . , k - 1. These formulas describe the Gram-Schmidt process for constructing an orthogonal set ofnonzero elements yl, . . . , y, which spans the same subspace as a given independent setXl,. . . ,x,. In particular, if x1, .
. . , x, is a basis for a finite-dimensional Euclidean space, theny,, . . . ,yk is an orthogonal basis for the same space. We can also convert this to anorthonormal basis by normalizing each element yi, that is, by dividing it by its norm.Therefore, as a corollary of Theorem 1.13 we have the following.
THEOREM 1.14. Every$nite-dimensional Euclidean space has an orthonormal basis. If x and y are elements in a Euclidean space, withy # 0, the element is called the projection of x along y. In the Gram-Schmidt process (1.16), we constructthe element Y,.+~ by subtracting from x,.+~ the projection of x,+r along each of the earlierelements yl, . . . , yr. Figure 1.1 illustrates the construction geometrically in the vectorspace V,. FrooaE 1.1 The Gram-Schmidt process in V3 . An orthogonal set {yl, yZ , y3} isconstructed from a given independent set {x1, x2, ~3. Construction of orthogonal sets. The Gram-Schmidt process 25 EXAMPLE 1. In Vd, find an orthonormal basis for the subspace spanned by the threevectors x1 = (1, -1, 1, -l), xZ = (5, 1, 1, l), and xQ = (-3, -3, 1, -3). Solution.
Applying the Gram-Schmidt process, we find y, = x1 = (1, -1, 1, -l), (x2 9 Yl)~ Yl = x2 - Yl = (4,2,0,2), yz = xp - (Yl, Yl) (x3 3 Yl) (x3 3 Y2>~ Y2 = x3 - Yl + Y2 = (0, 0, 0, 0). y3=x3-(Yl~Yl)y1-(Y2~Y2) Since y3 = 0, the three vectors x1, x2, x3 must be dependent. But since y1 and y2 arenonzero, the vectors x1 and x2 are independent.
Therefore L(x,, x2, x3) is a subspace ofdimension 2. The set {y1,y2} is an orthogonal basis for this subspace. Dividing each ofy1 and y2 by its norm we get an orthonormal basis consisting of the two vectors EXAMPLE 2. The Legendre polynomials. In the linear space of all polynomials, with theinner product (x, y) = ST, x(t)y(t) dt , consider the infinite

sequence x,, , Xl, x2, * * * , wherex,(t) = tn. When the orthogonalization theorem is applied to this sequence it yieldsanother sequence of polynomials y,, y,, y2, . . . , first encountered by the French mathe-matician A. M. Legendre (1752-1833) in his work on potential theory. The first fewpolynomials are easily calculated by the Gram-Schmidt process.
First of all, we haveye(t) = x0(t) = 1 . Since (yo,yo)=j:,dt=2 and (~1, ~0) = jtl t dt = 0, we find that yl(t) = x1(t) - E ye(t) = x1(t) = t.07 0 Next, we use the relations (xzz,yo)= f_llt2dt=$, b,yl)= jJ1t3dt=0, (yl,yl)= jT1t2dt+, to obtain y2(t) = x2(t) - (E; ye(t) - E$l; yl(t) = t2 - 8.03 0 17 1 Similarly, we find that y3(t) = t3 - St, y4(t) = t4 - $t2 + -&, y5(t) = t5 - -
gt3 + &t. 26 Linear spaces We shall encounter these polynomials again in Chapter 6 in our further study of differentialequations, and we shall prove that n! d"yn(f) = (2n)! dt" - - (t2 - 1)". The polynomials P, given by p?st> = (2n)!- y,(t) = h -$ 0" - 1)"2"(n !)" are known as the Legendrepolynomials. The polynomials in the corresponding orthonormal
sequence fro, ply v2, . . . , given by ~7~ = y,/llynll are called the normalized Legendre poly-nomials. From the formulas for yo, . . . , y5 given above, we find that f&l(t) = Ji) q%(t) = 4 t) q*(t) = &JS (3t2 - 1)) Q)3(t) = &J3 (5t3 - 3t), p4(t) = +Ji (35th - 30t* + 3), p5(t) = $& (63t" - 70t3 + 19). 1.15. Orthogonal complements. Projections Let V be a
Euclidean space and let S be a finite-dimensional subspace.
We wish toconsider the following type of approximation problem: Given an element x in V, to deter-mine an element in S whose distance from x is as small as possible. The distance betweentwo elements x and y is defined to be the norm 11x - yII . Before discussing this problem in its general form, we consider a special case, illustratedin Figure 1.2.
Here V is the vector space V, and S is a two-dimensional subspace, a planethrough the origin. Given x in V, the problem is to find, in the plane S, that point snearest to x. If x E S, then clearly s = x is the solution. If x is not in S, then the nearest point sis obtained by dropping a perpendicular from x to the plane. This simple example suggestsan
approach to the general approximation problem and motivates the discussion thatfollows. DEFINITION. Let S be a subset of a Euclidean space V. An element in V is said to beorthogonal to S if it is orthogonal to every element of S. The set of' all elements orthogonalto S is denoted by S-' and is called "S perpendicular." It is a simple exercise to verify
that Sl is a subspace of V, whether or not S itself is one.In case S is a subspace, then S1 is called the orthogonal complement of S.
EXAMPLE. If S is a plane through the origin, as shown in Figure 1.2, then S1 let S a linethrough the origin perpendicular to this plane. This example also gives a geometric inter-pretation for the next theorem. Orthogonal complements. Projections 2 7 FIGURE 1.2 Geometric interpretation of the orthogonal decomposition theorem in V,. THEOREM 1.15.
ORTHOGONAL DECOMPOSITION THEOREM. Let V be a Euclidean spaceand let S be ajnite-dimensional subspace of V. Then every element x in V can be representeduniquely as a sum of two elements, one in S and one in Sl. That is, we have (1.17) x=s+s', where s E S and d- E 5-l.
Moreover, the norm of x is given by the Pythagorean formula (1.18) llxl12 = IIsl12 + Il~1112. Proof. First we prove that an orthogonal decomposition (1.17) actually exists. SinceS is finite-dimensional, it has a finite orthonormal basis, say {e, , . . . , e,}. Given x, definethe elements s and sL as follows: (1.19) s = i (x, ei)ei, SI = x - s .i=l Note that each
term (x, e,)e, is the projection of x along e, . The element s is the sum of theprojections of x along each basis element. Since s is a linear combination of the basiselements, s lies in S. The definition of & shows that Equation (1.17) holds. To prove thatd lies in Sl, we consider the inner product of sL and any basis element ej . We have (S', ej> = (x - s, ej)
= (x, e,) - (s, ei) . But from (1.19), we find that (s, eJ = (x, e,), so sL is orthogonal to ei. Therefore sLis orthogonal to every element in S, which means that sL E SL . Next we prove that the orthogonal decomposition (1.17) is unique. Suppose that xhas two such representations, say (1.20) x=s+sl- and x=t+tl, 28 Linear spaces where s and t are in S, and
sL and t' are in SI. We wish to prove that s = t and sL = t1 .From (1.20), we have s - t = t1 - sL, so we need only prove that s - I = 0. Buts - t E Sand t1 - s1 E SL so s - t is both orthogonal to tl - .sL and equal to t1 - & .Since the zero element is the only element orthogonal to itself, we must have s - t = 0.This shows that the decomposition is unique.
Finally, we prove that the norm of x is given by the Pythagorean formula.
We have llxll2 = (x, x) = (s + &, s + sl) = (s, s) + (sl, s'->, the remaining terms being zero since s and sL are orthogonal. This proves (1.18). DEFINITION.
Let S be a Jnite-dimensional subspace of a Euclidean space V, and let {e,, . . . , e,} be an orthonormal basis for S. If x E V, the element s dejned by the equation s = 2 (x, ei)eii=l is called the projection of x on the subspace S. We prove next that the projection of x on S is the solution to the approximation problemstated at the beginning of this section.
1.16 Best approximation of elements ,in a Euclidean space by elements in a finite-dimensional subspace THEOREM 1.16. APPROXIMATION THEOREM. Let S be a ,finite-dimensional subspace ofa Euclidean space V, and let x be any element of V.
Then the projection of x on S is nearer tox than any other element of S. That is, [f s is the projection of x on S, we have llx - $11 I IIX - tll for all t in S; the equality sign holds if and only if t = s. Proof. By Theorem 1.15 we can write x = s + sL, where s E S and s1 E SL. Then,for any t in S, we have x - t = (x - s) + (s - t) . Since s - t E S and x - s = s-l E SL,
this is an orthogonal decomposition of x - t, soits norm is given by the Pythagorean formula lb - tl12 = IIX - sly + l/s - tll2. But IIs - tlj2 2 0, so we have IIx - tl12 2 [Ix - sl12, with equality holding if and only ifs = t.
This completes the proof. Best approximation of elements in a Euclidean space 29 EXAMPLE 1. Approximation of continuous functions on [0,2n] by trigonometric polynomials.Let V = C(O,27r), the linear space of all real functions continuous on the interval [0,2~],and define an inner product by the equation (f, g) = Ji" f(x)g(x) dx . In Section 1.12we
exhibited an orthonormal set of trigonometric functions pO, all, yz, .
. . , where (1.21) PO(X) = $9 cos kxP)'m--1(X) = - JG 'Q)'@(x) = sE-!k) JGf o r k>l. 7-r The 2n + 1 elements vO, all,. .
. , pzn span a subspace S of dimension 2n + 1. The ele-ments of S are called trigonometric polynomials. IffE C(O,27r), letf, denote the projection off on the subspace S. Then we have (1.22) where (f, PJ = ~oz"fb)plp(x) dx. The numbers (f, Q)& are called Fourier coefJicients off. Using the formulas in (1.21), wecan rewrite (1.22) in the form (1.23) where
f,(x) = ba, +$:ak cos kx + b, sin kx), ak = $ L'>(X) cos kx dx ,1 s 2nb, = - f(x) sin kx dx i-r 0 fork=0,1,2 ,..., n. The approximation theorem tells us that the trigonometric poly-nomial in (1.23) approximates ,f better than any other trigonometric polynomial in S, inthe sense that the norm l\f - f,ll is as small as possible. EXAMPLE 2. Approximation of
continuous functions on [- 1, I] by polynomials ofdegree < n. Let V = C(- 1, 1)) the space of real continuous functions on [- 1, 11, and let (f, g) = S', f(x)&) dx. The n + 1 normalized Legendre polynomials qo, pl, .
. . , P)~,introduced in Section 1.14, span a subspace S of dimension n + 1 consisting of all poly-nomials of degree < n. Iffe C(- 1, 1)) let f, denote the projection off on S. Then wehave where Cf, Q)A = i_',f (Ovdt> dt . This is the polynomial of degree < n for which the norm Ilf - f,ll is smallest. For example,whenf(x) = sin TX, the coefficients (f, vk) are
given by (f, q&j = ST, sin nt &t) dt. In particular, we have (f, po) = 0 and 30 Linear spaces Therefore the linear polynomialf,(t) which is nearest to sin nt on [- 1, l] is Since (f, & = 0, this is also the nearest quadratic approximation. 1.17 Exercises 1. 2. 3. 4. 5. 6.
7. 8. 9. 10. In each case, find an orthonormal basis for the subspace of V, spanned by the given vectors.(a) x1 = (1, 1, l), x2 = (l,O,l), x3 = (3,2,3). (b) xl = (1, 1, l), x2 = (-l,l, -l), x3 = (l,O, 1).In each case, find an orthonormal basis for the subspace of V, spanned by the given vectors.(a> x1 = (1, l,O, 01, x.2 = (%I, 1, o>, x3 = (0, 0, 1, l), xq = (l,O,O, 1)
 XI = (1, l,O, 11, x2 = (1, 0,2, 11, x3 = (1,2, -2, 1).In the real linear space C(0, n), with inner product (x, y) = j; x(t)y(t) dt, let x,(t) = cos ntforn =0,1,2 ,.... Prove that the functions yO, y,, y,, .
. . , given by 2and y%(f) = ; cos nt for II 2 1, form an orthonormal set spanning the same subspace as x0, x1, x2, ...In the linear space of all real polynomials, with inner product (x,~) = & x(t)y(r) dt, letx,(t) = tn for n = 0, 1, 2, ... Prove that the functions y,(t) = 11 y1(t) = 6 (2t - l), y,(t) = 6 (6? - 6t + 1) form an orthonormal set spanning the same
subspace as {x,, , x1, x2}.Let V be the linear space of all real functions f continuous on [0, + a) and such that theintegral j: eetf2(f) dt converges. Define (f,g) = jr ePtf(t)g(t) dt, and let y,, y, ,y,, . . . , bethe set obtained by applying the Gram-Schmidt process to x0, xi, x2, . . . , where x,(t) = tnfor n 2 0. Prove that v,,(t) = 1, I+ = t - 1, y2(t) = t2 - 4t + 2,
y3(t) = t3 - 9t2 +18t - 6.In the real linear space C(1, 3) with inner product (f,g) = jf f (x)g(x) dx, let f(x) = l/xand show that the constant polynomial g nearest to f is g = 4 log 3. Compute lig -f II2 forthis g.In the real linear space C(0, 2) with inner product (f, g) = j"i f (x)g(x) dx, let f(x) = en andshow that the constant polynomial g nearest to f is g = $(e" -
1). Compute iig -f II2 forthis g.In the real linear space C(- 1, 1) with inner product (f, g) = ST1 f (x)g(x) dx , let f(x) = e5and find the linear polynomial g nearest to f.
Compute ilg - f II2 for this g.In the real linear space C(0,27~) with inner product (f,g) = jiT f(x)g(x) dx, let f(x) = x.In the subspace spanned by u,,(x) = 1 , ul(x) = cos x, u2(x) = sin x, find the trigonometricpolynomial nearest to JIn the linear space V of Exercise 5, let f (x) = e-" and find the linear polynomial that is nearesttoJ 2 LINEAR
TRANSFORMATIONS AND MATRICES 2.1 Linear transformations One of the ultimate goals of analysis is a comprehensive study of functions whosedomains and ranges are subsets of linear spaces. Such functions are called transformations,mappings, or operators. This chapter treats the simplest examples, called linear transforma-tions, which occur
in all branches of mathematics. Properties of more general transforma-tions are often obtained by approximating them by linear transformations. First we introduce some notation and terminology concerning arbitrary functions. LetV and W be two sets.
The symbol T:V+W will be used to indicate that T is a function whose domain is V and whose values are in W.For each x in V, the element T(x) in W is called the image of x under T, and we say that Tmaps x onto T(x). If A is any subset of V, the set of all images T(x) for x in A is called theimage of A under T and is denoted by T(A). The image of the
domain V, T(V), is the rangeof T. Now we assume that V and Ware linear spaces having the same set of scalars, and wedefine a linear transformation as follows. D E F I N I T I O N . If V and Ware linear spaces, a function T: V + W is called a linear trans-formation of V into W if it has the.following two properties: (a) T(x + y) = T(x) + T(y) for all x and y
in V,(b) T(cx) = CT(X) for all x in V and all scalars c. These properties are verbalized by saying that T preserves addition and multiplication byscalars. The two properties can be combined into one formula which states that T(ax + by) = aT(x) + bTQ) for all x,y in V and all scalars a and 6. By induction, we also have the more general relation for any n
elements x1, . . . , x,inVandanynscalarsa,,...,a,. 31 32 Linear transformations and matrices The reader can easily verify that the following examples are linear transformations. EXAMPLE 1. The identity transformation. The transformation T: V + V, where T(x) = xfor each x in V, is called the identity transformation and is denoted by Z or by IV.
EXAMPLE 2. The zero transformation. The transformation T: V--f V which maps eachelement of V onto 0 is called the zero transformation and is denoted by 0. EXAMPLE 3. Multiplication by ajxed scalar c. Here we have T: V + V, where T(x) = cxfor all x in V. When c = 1 , this is the identity transformation. When c = 0, it is the zerotransformation.
EXAMPLE 4. Linear equations. Let V = V,, and W = V, . Given mn real numbers aiL,wherei= 1,2,...,mandk= 1,2,...,n,defineT: V,+V,asfollows: T m a p s e a c hvector x = (x1, . . . , x,) in V, onto the vector y = (,vl, . . . , ym) in V, according to theequations Yi = l$ aikXk for i=l,2 ,..., m.k=l EXAMPLE 5. Inner product with afixed element. Let V be a real
Euclidean space. For afixed element z in V, define T: V -+ R as follows: If x E V, then T(x) = (x, z), the innerproduct of x with z. EXAMPLE 6. Projection on a subspace. Let V be a Euclidean space and let S be a finite-dimensional subspace of V. Define T: V + S as follows: If x E V, then T(x) is theprojection of x on S. EXAMPLE 7.
The dzferentiation operator. Let V be the linear space of all real functionsf differentiable on an open interval (a, b). The linear transformation which maps eachfunctionfin V onto its derivativef' is called the differentiation operator and is denoted byD. Thus, we have D: V + W, where D (f) = f' for each f in V. The space W consists ofall derivatives f'.
EXAMPLE 8. The integration operator. Let V be the linear space of all real functionscontinuous on an interval [a, b].
IffE V, define g = T(f) to be that function in V given by g(x) = JaZfW dt i f alf c # 0, the null space contains only 0.If c = 0, the null space is V. EXAMPLE 4. Linear equations. The null space consists of all vectors (xi, . . . , x,) in V,for which for i=l,2 ,..., m.
EXAMPLE 5. Inner product with ajxed element z. The null space consists of all elementsin V orthogonal to z. EXAMPLE 6. Projection on a subspace S. If x E V, we have the unique orthogonaldecomposition x = s + sL (by Theorem I .15). ,'mce T(x) = s, we have T(x) = 0if and only if x = sL . Therefore, the null space is zl, the orthogonal complement of S.
EXAMPLE 7. DifSerentiation operator. The null space consists of all functions that areconstant on the given interval. EXAMPLE 8. Integration operator. The null space contains only the zero function. 3 4 2.3 Nullity and rank Linear transformations and matrices Again in this section T denotes a linear transformation of a linear space V into a
linearspace W. We are interested in the relation between the dimensionality of V, of the nullspace N(T), and of the range T(V). If V is finite-dimensional, then the null space is alsofinite-dimensional since it is a subspace of I'. The dimension of N(T) is called the nullityof T.
In the next theorem, we prove that the range is also finite-dimensional; its dimensionis called the rank of T. THEOREM 2.3. NULLITY PLUS RANK THEOREM. If V is finite-dimensional, then T(V) isalso finite-dimensional, and we have (2.1) dim N(T> + dim T(V) = dim V . In other words, the nullity plus the rank of a linear transformation is equal to the
dimensionof its domain. Proof. Let n = dim V and let e, , . . . , e,beabasisforN(T),wherek = dimN(T)< n.By Theorem 1.7, these elements are part of some basis for V, say the basis (2.2) e,, . . . , ek9 ekflp .
. . ? ek+7y where k + r = n . We shall prove that the r elements (2.3) T(e,+A . . . 7 T(%,-r) form a basis for T(V), thus proving that dim T(V) = r .
Since k + r = n , this also proves(2.1). First we show that the r elements in (2.3) span T(V). If y E T(V), we have y = T(x)for some x in V, and we can write x = clel + * * * + ck+,.ek+,. . Hence, we have since T(e,) = * *'* = T(e,) = 0. This shows that the elements in (2.3) span T(V).Now we show that these elements are independent. Suppose that there
are scalars ck+19 * * - 3 Ck+T such that k+r 2 ciT(ei) = 0 Ii=k+l This implies that SO the element X = Ck+lek+l + ' ' ' + ck+&k+r is in the null space N(T). This means there Exercises 35 are scalars cl, . .
.
, c, such that x = clel + * . * + c,e,, so we have k ktr x - x = 2 ciei - 2 ciei = 0 .i=l i=kfl But since the elements in (2.2) are indepecdent, this implies that all the scalars ci are zero.Therefore, the elements in (2.3) are independent. Note: If V is infinite-dimensional, then at least one of N(T) or T(V) is infinite-dimensional. A proof of of this fact is outlined in
Exercise 30 of Section 2.4. 2.4 Exercises In each of Exercises 1 through 10, a transformation T: V, -+ VZ is defined by the formula givenfor T(x, y), where (x, y) is an arbitrary point in VZ . In each case determine whether Tis linear. IfT is linear, describe its null space and range, and compute its nullity and rank. 1. W, y> = Cy, 4. 6. T(x, y) = (e5, ev) .2.
T(x, y> = (x, -y> . 7. T(x, y) = (x, 1).3. T(x,y) = (x,0). 8. T(x,y) = (x + 1,~ + 1).4. T(x, y) = (x, x) . 9. T(x,y) = (x -y,x +y).5. T(x,y) = (x2,y2). 10. T(x,y) =(2x -y,x +y). Do the same as above for each of Exercises 11 through 15 if the transformation T: V, --f V,is described as indicated. 11. T rotates every point through the same angle q about the origin. That
is, T maps a pointwith polar coordinates (r, 0) onto the point with polar coordinates (r, 0 + v), where q~ isfixed. Also, T maps 0 onto itself. 12. T maps each point onto its reflection with respect to a fixed line through the origin.13. T maps every point onto the point (1, 1).14. T maps each point with polar coordinates (r, 0) onto the point with polar
coordinates (2r, 0). Also, T maps 0 onto itself.15. T maps each point with polar coordinates (r, 0) onto the point with polar coordinates (r, 20). Also, T maps 0 onto itself.
Do the same as above in each of Exercises 16 through 23 if a transformation T: V,+ V, isdefined by the formula given for T(x, y, z), where (x, y, z) is an arbitrary point of V, . 16. T@,y,z) = (z,y,x). 20.
T(x,y,z) =(x + l,y + l,z - 1).17. W,y,z) = ky,O). 21. T(x,y,z) = (x + 1,y + 2,z + 3).18. T(x, y, z) = (x, 2y, 3~).
22. T(x, y, z) = (x, y2,z3).19. T(x,y,z) = (x,y, 1). 23. T(x, y, z) = (x + z, 0, x + y). In each of Exercises 24 through 27, a transformation T: V'+ V is described as indicated. Ineach case, determine whether T is linear.
If T is linear, describe its null space and range, andcompute the nullity and rank when they are finite.
24. Let V be the linear space of all real polynomialsp(x) of degree I; n . Ifp E V, y = T(p) meansthat q(x) =p(x + 1) for all real x. 25. Let V be the linear space of all real functions differentiable on the open interval (- 1, 1).IffE v, g = T(f) means that g(x) = xf'(x) for all x in (- 1, 1). 36 Linear transformations and matrices 26. Let V be the linear space of all
real functions continuous on [a, b]. Iffg V, g = T(f) meansthat g(x) =/If(t) sin (x - t) dt for a < x 5 b. 27. Let V be the space of all real functions twice differentiable on an open interval (a, b). Ify E V, define 7"(y) = y" + Py' + Qy , where P and Q are fixed constants. 28. Let V be the linear space of all real convergent sequences {x,}. Define a
transformationT: V--t V as follows: If x = {x,} is a convergent sequence with limit a, let T(x) = {y,},where yn = a - x, for n 2 1 . Prove that Tis linear and describe the null space and range of T. 29. Let V denote the linear space of all real functions continuous on the interval [-n, ~1. Let Sbe that subset of Vconsisting of all f satisfying the three equations
J:gf(t)costdt = 0 , j:rf(t)sintdt =O. (a) Prove that S is a subspace of V.(b) Prove that S contains the functions f (x) = cos nx and f (x) = sin nx for each n = 2,3, ...(c) Prove that S is infinite-dimensional. Let T: V-t V be the linear transformation defined as follows: Iff E V,g = T(f) means that g(x) = j;, (1 + cos (x - t)}f (t) dt . (d) Prove that T(V), the range of
T, is finite-dimensional and find a basis for T(V).(e) Determine the null space of T.(f) Find all real c # 0 and all nonzero f in V such that T(f) = cJ (Note that such an flies in the range of 7'.) 30. Let T: V+ W be a linear transformation of a linear space V into a linear space W. If V isinfinite-dimensional, prove that at least one of T(V) or N(T) is infinite-
dimensional. [Hint: Assume dim N(T) = k , dim T(V) = r , let e, , . . . , ek be a basis for N(T) andlet e,, . .
. , ek, e,+r, . . . , ek+n be independent elements in V, where n > r. Theelements T(e,+,), . . .
, T(erc+,) are dependent since n > r.
Use this fact to obtain acontradiction.] 2.5 Algebraic operations on linear transformations Functions whose values lie in a given linear space W can be added to each other and canbe multiplied by the scalars in W according to the following definition. DEFINITION . Let S: V+ W and T: V + W be two functions with a common domain Vand with values in
a linear space W. If c is any scalar in W, we define the sum S + T and theproduct CT by the equations (2.4) (S + T)(x) = S(x) + T(x) 3 (CT)(X) = CT(X) for all x in V. Algebraic operations on linear transformations 37 We are especially interested in the case where V is also a linear space having the samescalars as W. In this case we denote by .Y(V, W)
the set of all linear transformations of Vinto W. If S an'd Tare two linear transformations in =.Y(V, W), it is an easy exercise to verify thatS + T and CT are also linear transformations in LZ'(V, W). More than this is true. Withthe opera.tions just defined, the set L?(V, W) itself becomes a new linear space. The zerotransformation serves as the zero
element of this space, and the transformation (-1)Tis the neg,ative of T. It is a straightforward matter to verify that all ten axioms for a linearspace are satisfied.
Therefore, we have the following. THEOREM 2.4. The set Z'(V, W) of all linear transformations of V into W is a linearspace with the operations of addition and multiplication by scalars de$ned as in (2.4). A more interesting algebraic operation on linear ,transformations is composition ormultiplication of transformations. This operation makes no use of
the algebraic structureof a linear space and can be defined quite generally as follows.
FIGURE 2.1 Illustrating the composition of two transformations. DEFINITION. Let U, V, W be sets. Let T: II -+ V be a function with domain U andvalues in V, a.nd let S: V--f W be another function with domain V and values in W. Thenthe composition ST is the function ST: U---f W defined b-y the equation (ST)(x) = S[T(x)] for every x in U. Thus, to map
x by the composition ST, we first map x by T and then map T(x) by S.This is illustrated in Figure 2.1. Compo8sition of real-valued functions has been encountered repeatedly in our study ofcalculus, and we have seen that the operation is, in general, not commutative. However,as in the (case of real-valued functions, composition does satisfy an
associative law. THEOREM 2.5. If T: u -+ V , S: V + W, and R: W +- X are three functions, then we have R(ST) = (RS)T. 38 Linear transformations and matrices Proof. Both functions R(ST) and (RS)T have domain U and values in X. For each xin U, we have DWT)ICd = R[W"Ml = R[WWll and WS'X4 = W)LWl = RLW(x)ll, which proves that R(ST) =
(RS)T. DEFINITION. Let T: V-t V be a function which maps V into itself. We dejne integralpowers of T inductively as follows: To= I , T" = TT"-1 f o r n>l. Here I is the identity transformation. The reader may verify that the associative lawimplies the law of exponents T"T" = T m+n for all nonnegative integers m and n. The next theorem shows that the
composition of linear transformations is again linear. THEOREM 2.6. If U, V, W are linear spaces with the same scalars, and tf T: U -+ Vand S: V -+ W are linear transformations, then the composition ST: U + W is linear. Proof. For all x, y in U and all scalars a and b, we have (ST)(ax + by) = S[T(ax + by)] = S[aT(x) + bT(y)] = aST(x) + bST(y) .
Composition can be combined with the algebraic operations of addition and multiplica-tion of scalars in 9(V, W) to give us the following. THEOREM 2.7. Let U, V, W be linear spaces with the same scalars, assume S and T arein Z(V, W), and let c be any scalar. (a) For any function R with values in V, we have (S + T)R = SR + TR and (cS)R = c(SR) . (b)
For any linear transformation R: W + U, we have R (S + T)=RS+RT and R(cS) = c(RS) . The proof is a straightforward application of the definition of composition and is left asan exercise. 2.6 Inverses In our study of real-valued functions we learned how to construct new functions byinversion of monotonic functions. Now we wish to extend the
process of inversion to amore general class of functions. Inverses 39 Given a function T, our goal is to find, if possible, another function S whose compositionwith T is the identity transformation. Since composition is in general not commutative,we have to distinguish between ST and TS. Therefore we introduce two kinds of inverseswhich we call left
and right inverses. DEFINITION. Given two sets V and Wand a function T: V + W. A function S: T(V) + Vis called a IejYt inverse of T tj'S[T(x)] = x for all x in V, that is, if where r, is the ide,ntity transformation on V. A function R: T(V) + V is called a right inverseof T if T[R(y)] = y for ally in T(V), that is, if TR = IT(V), where I, (vj is the identity
transformation on T(V). EXAMPLE. A function with no left inverse but with two right inverses. Let V = (1, 2)and let W = (0). Define T: V-+ Was follows: T(1) = T(2) = 0. This function has tworight inverses R: W+ V and R' : W---f V given by R(0) = 1, R'(0) = 2. It cannot have a left inverse S since this would require 1 = S[T(l)] = S(0) and 2 = S[T(2)] =
S(0). This simple example shows that left inverses need not exist and that right inverses need notbe unique. Every function T: V --f W has at least one right inverse. In fact, each y in T(V) has theform y = T(x) for at least one x in V.
If we select one such x and define R(y) = x , thenT[R(y)] = T(x) = y for each y in T(V), so R is a right inverse.
Nonuniqueness may occurbecause there may be more than one x in V which maps onto a given y in T(V). We shallprove presently (in Theorem 2.9) that if each y in T(V) is the image of exactly one x in V,then right inverses are unique. First we prove that if a left inverse exists it is unique and, at the same time, is a rightinverse. THEOREM 2.8. A
function T: V + W can have at most one left inverse. If T has a leftinverse S, then ,S is also a right inverse. Proof. Assume T has two left inverses, S: T(V)+ Vand S': T(V)+ V. Choose anyy in T(V). We shall prove that S(y) = S'(y) . Now y = T(x) for some x in V, so we have S[T(x)] = x and S' [T(x)] = x , 40 Linear transformations and matrices since both S
and S' are left inverses.
Therefore S(y) = x and S'(y) = x , so S(y) = S'@)for all y in T(V). Therefore S = S' which proves that left inverses are unique. Now we prove that every left inverse S is also a right inverse. Choose any element y inr(V). We shall prove that T[S(y)] = y . Since y E 7'(I') , we have y = T(x) for some x inV. But S is a left inverse, so x = S[T(x)] = S(y).
Applying T, we get T(x) = T[S@)] . But y = T(x), so y = T[S(y)] , which completes theproof. The next theorem characterizes all functions having left inverses. THEOREM 2.9. A function T: V + W has a left inverse if and only if T maps distinctelements of V onto distinct elements of W; that is, if and only if, for all x and y in V, (2.5) X#Y implies T(x) # T(y).
Note: Condition (2.5) is equivalent to the statement (2.6) T(x) = T(y) implies x = y .
A function T satisfying (2.5) or (2.6) for all x and y in V is said to be one-to-one on V. Proof. Assume T has a left inverse S, and assume that T(x) = T(y). We wish to provethat x = y . Applying S, we find S[T(x)] = S[T(y)] S'mce S[T(x)] = x and S[T(y)] = y,this implies x = y. This proves that a function with a left inverse is one-to-one on itsdomain. Now we
prove the converse. Assume Tis one-to-one on V. We shall exhibit a functionS: T(V) --f V which is a left inverse of T. If y E T(V) , then y = T(x) for some x in V. By(2.6), there is exactly one x in V for which y = T(x). Define S(y) to be this x. That is,we define S on T(V) as follows: S(y) = x means that T(x) = y . Then we have S[T(x)] = x for each x in V, so ST
= I,. Therefore, the function S sodefined is a left inverse of T. DEFINITION. Let T: V -+ W be one-to-one on V. The unique left inverse of T (whichwe know is also a right inverse) is denoted by T-l. We say that T is invertible, and we callT-l the inverse of T. The results of this section refer to arbitrary functions. Now we apply these ideas tolinear
transformations.
One-to-one linear transformations 2.7 One-to-one linear transformations 41 In this section, V and W denote linear spaces with the same scalars, and T: V-t Wdenotes a linear transformation in Z?(V, W). The linearity of T enables us to express theone-to-one property in several equivalent forms. THEOREM 2.10~. Let T: V -+ W be a linear
transformation in Z(V, W). Then the followingstatements are equivalent. (a) T is one-to-one on V.(b) T is invertible and its inverse T-l : T(V) + V is linear.(c) For all x in V, T(x) = 0 implies x = 0. That is, the null space N(T) contains only the zero element of V. Proof. We shall prove that (a) implies (b), (b) implies (c), and (c) implies (a). Firstassume (a)
holds. Then T has an inverse (by Theorem 2.'9), and we must show that T-lis linear. Take any two elements u and v in T(V). Then u = T(x) and v = T(y) for somex and y in V. For any scalars a and b, we have a* •k bv = aT(x) + bT(y) = T(ax + by), since T is linear. Hence, applying T-l, we have T-Yau + bv) = ax + by = aT-l(u) + bT-l(v), so T-l is linear..
Therefore (a) implies (b).Next assume that (b) holds. Take any x in V for which T(x) = 0. Applying T-l, we find that x = T-l(O) = 0, since T-l is linear. Therefore, (b) implies (c).Finally, assu.me (c) holds. Take any two elements u and v in V with T(u) = T(v). By linearity, we have T(u - v) = T(u) - T(v) = 0, so u - v = 0. Therefore, Tis one-to-oneon V, and the
proof of the theorem is complete. When V is finite-dimensional, the one-to-one property can be formulated in terms ofindependence and dimensionality, as indicated by the next theorem. THEOREM 2.1 I. Let T: V + W be a linear transformatton in 3(V, W) and assume thatV is jinite-dime,nsional, say dim V = n . Then the following statements are
equivalent. (a) T is one-to-one on V. (b)Zfe,,..., e, are independent elements in V, then T(e,), . . . , T(e,) are independentelements in T(V). (c) dim T(V) = n . (4 lfh,..., e,} is a basis for V, then {T(e,), . . . , T(e,)} is a basis for T(V). Proof. We shall prove that (a) implies (b), (b) implies (c), (c) implies (d), and (d) implies(a). Assume (a) holds. Let e,, . . . , e, be
independent elements of V and consider the 42 Linear transformations and matrices elements T(e,), . . . , T(e,) in T(V).
Suppose that i$ciT(eA = 0for certain scalars q, . . . , c,.
By linearity, we obtain and hence since T is one-to-one. But e,, .
. . , e, are independent, so c1 = * * . = c, = 0.
Therefore(a) implies (b). Now assume (b) holds. Let {e,, . . . , e,} be a basis for V. By (b), the n elements T(e,), . . . , T(e,) in T(V) are independent. Therefore, dim T(V) 2 n . But, by Theorem2.3, we have dim T(V) < n . Therefore dim T(V) = n, so (b) implies (c). Next, assume (c) holds and let {e,, . . . , e,} be a basis for V.
Take any element y inT(V). Then y = T(x) for some x in V, so we have x = i ciei,61 and hence y = T(x) = i ciT(ei).i=l Therefore {T(e,), .
. . , T(e,)} spans T(V). But we are assuming dim T(V) = n, so{Wd, . . . , T(e,)} is a basis for T(V). Therefore (c) implies (d). Finally, assume (d) holds. We will prove that T(x) = 0 implies x = 0. Let {e,, . . . , e,}be a basis for V. If x E V, we may write x = i ciei ,i=l and hence T(X) = itlciT(eJ .
IfT(x) = O,thenc, = .* * = c, = 0, since the elements T(e,), .
. . , T(e,) are independent.Therefore x = 0, so Tis one-to-one on V. Thus, (d) implies (a) and the proof is complete. 2.8 Exercises 1. Let V = (0, l} . Describe all functions T: V-t V . There are four altogether. Label them asTl , Tz , T3, T4 and make a multiplication table showing the composition of each pair.
Indicatewhich functions are one-to-one on V and give their inverses. 2.
Let V = (0, 1,2}. Describe all functions T: V+ V for which T(V) = V . There are sixaltogether.
Label them as T,, . . . , Ts and make a multiplication table showing the com-position of each pair. Indicate which functions are one-to-one on V, and give their inverses. In each of Exercises 3 through 12, a function T: V2 4 V, is defined by the formula given forT(x, JJ), where (x, y) is an arbitrary point in Vz . In each case determine whether T is one-to-
oneon V,. If it is, describe its range T(Vz); for each point (u, v) in T(V,), let (x, y) = T-l(u, v) andgive formulas for determining x and y in terms of u and v. 3. Thy) = (y, 4. 8. T(x, y) = (e", eY).4. T(x, y) = (x, --.I'). 9. T(x,y) = (x, 1).5. T(x, y) = (x, 0). 10. T(x,y) = (x + 1,~ + 1).6. T(x, y) = (x, x) . 11. T(x,y) = (x -y,x +y).7. T(x, y) = (x2, y2). 12. T(x,y) =(2x -y,x
+y). Exercises 43 In each of Elxercises 13 through 20, a function T: Va+ I', is defined by the formula given forT(x, y, z), where (x, y, z) is an arbitrary point in Vs. In each case, determine whether Tis one-to-one on Vs. If it is, describe its range T(V,); for each point (u, v, W) in T(V,), let (x,~, Z) =T-'(u, v, w) and give formulas for determining x, y, and z in
terms of U, u, and W. 13. T(x, y, z) = (z. y, x) 17.
T(x,y,z) =: (x + 1,~ + 1,z - 1).14. T(x,y, z) = (x,y, 0). 18.
T(x,y,z) =:(x + 1,~ +2,z +3).15. T(x, y, z) = (x, 2y, 32).
19. T(x, y, z) =: (x, x + y, x + y + z).16. T(x,y,z)=(x,y,x+y+z). 20. T(x,y,z) =:(x +y,y +z,x +z). 21. Let T: V -+ V be a function which maps V into itself. Powers are defined inductively by theformulas 'To = I, T" = TTn-l for n 2 1 .
Prove that the associative law for compositionimplies the law Iof exponents: TmTn = Tm+n.
If Tis invertiblle, prove that T" is also invertibleand that (Tn)-l = (T-l)n. In Exercises, 22 through 25, S and T denote functions with domain V and values in V. Ingeneral, ST # TS. If ST = TS, we say that S and T commute. 22. If S and T commute, prove that (ST)n = SnTn for all integers n 2 0.23. If S and Tare invertible, prove that ST is also invertible
and that (ST)-1 = T-lS-l. In other words, the inverse of ST is the composition of inverses, taken in reverse order.24. If S and Tare invertible and commute, prove that their inverses also commute.25. Let V be a line,ar space. If S and T commute, prove that (S + T)' = S2 + 2ST + T2 and (S + T)3 = S3 + 3S2T + 3ST2 + T3. Indicate how these formulas
must be altered if ST # TS.26. Let S and T be the linear transformations of V, into V, defined by the formulas S(x, y, z) = (z, y, x) and T(x, y, z) = (x, x + y, x + y + z), where (x, ;v, z) is an arbitrary point of V,.(a) Determine the image of (x, y, z) under each of the following transformations : ST, TS,ST - TS., S2, T2, (ST)2, (TS)2, (ST - TS)2.(b) Prove that S
and Tare one-to-one on V, and find the image of (u, v, w) under each of thefollowing transformations : S-l, T-l, (ST)-l, (TS)-l.(c) Find the image of (x, y, z) under (T - Z)n for each n 2 1. 27. Let V be the linear space of all real polynomialsp(x). Let D dlenote the differentiation operatorand let T denote the integration operator which maps each polynomial p
onto the polynomialq given by q(x) = jgp(t) dt . Prove that DT = ZV but that TD # Z,. Describe the null spaceand range of TD.
28.
Let Vbe the linear space of all real polynomialsp(x). Let D dsenote the differentiation operatorand let T be the linear transformation that maps p(x) onto x,n'(x).(a) Let p(x) = 2 + 3x - x2 + 4x3 and determine the image ofp under each of the followingtransformatiom: D, T, DT, TD, DT - TD, T2D2 - D2T2.(b) Determine those p in V for which T(p) = p .(c)
Determine thosep in V for which (DT - 2D)(p) = 0.(d) Determine thosep in V for which (DT - TD)%(p) = D"(p). 29. Let V and1 D be as in Exercise 28 but let T be the linear transformation that maps p(x) ontoxp(x). Prove tihat DT - TD = Z and that DTn - TnD = nTn-l for n 2 2. 30. Let S and T be in y(V, V) and assume that ST - TS = I.
Prove that STn - TnS = nTn-lfor all II ;z 1 .
3 1. Let V be the linear space of all real polynomialsp(x). Let R, S, T be the functions which mapan arbitrary polynomialp(x) = co + clx + . . . + c,xn in V onto the polynomials u(x), s(x), 44 Linear transformations and matrices and t(x), respectively, where 44 = p(O) ,k=l t(x) = i CkX*+l.Rx0 (a) Let p(x) = 2 + 3x - x2 + x3 and determine the image of p
under each of the followingtransformations: R, S, T, ST, TS, (TS)2, T2S2, S2T2, TRS, RST.(b) Prove that R, S, and Tare linear and determine the null space and range of each.(c) Prove that T is one-to-one on V and determine its inverse.(d) If n 2 1, express (TS)n and SnTn in terms of I and R. 32.
Refer to Exercise 28 of Section 2.4. Determine whether T is one-to-one on V. If it is, describeits inverse. 2.9 Linear transformations with prescribed values If V is finite-dimensional, we can always construct a linear transformation T: V-+ Wwith prescribed values at the basis elements of V, as described in the next theorem.
THEOREM 2.12. Let e,,..., e, be a basis for an n-dimensional linear space V. LetUl,... 7 u, be n arbitrary elements in a linear space W. Then there is one and only one lineartransformation T: V + W such that (2.7) T(%> = u/c for k=l,2 ,..., n. This T maps an arbitrary element x in V as follows: (2.8) I f x =ix,e,, then T(x) = i xkuk.k=l k=l Proof.
Every x in V can be expressed uniquely as a linear combination of e,, . . . , e, ,the multipliers x1, . .
. , x, being the components of x relative to the ordered basis(e,, . . . , e,). If we define T by (2.8), it is a straightforward matter to verify that T islinear. If x = ek for some k, then all components of x are 0 except the kth, which is 1, so(2.8) gives T(e,) = uk, are required. To prove that there is only one linear transformation satisfying (2.7), let T' be
anotherand compute T'(x). We find that =k$IxkT'(eb) = i xkuk = T(x).k=l Since T'(x) = T(x) for all x in V, we have T' = T, which completes the proof. EXAMPLE. Determine the linear transformation T: V, -+ V, which maps the basis elementsi = (1,O) and j = (0, 1) as follows: T(i) = i +i, T(j) = 2i -j. Matrix representations of linear transformations 45-__-
Solution. Nf x == x,i + xz j is an arbitrary element of V,, then T(x) is given by T(x) = x,T(i) + xzT(j) = Xl(i +j> + x,(2i - j) = (x, + 2x& + (Xl - x2)j. 2.10 Matrix representations of linear transformations Theorem 2.12 shows that a linear transformation T: V-+ W of a finite-dimensionallinear space V is completely determined by its action on a given set of
basis elementse,, . . . , e,. Now, suppose the space W is also finite-dimensional, say dim W = m, and let 6, * * *, w, be a basis for W. (The dimensions n and m may or may not be equal.) Since Thas values in IV, each element T(e,) can be expressed uniquely as a linear combination of thebasis elements wl, . . . , w,, say T(e,) = 2 tikwi,i=l where t,, , . . . , t,,
are the components of T(e,) relative to the: ordered basis (w,, . . . , w,).We shall display the m-tuple (t,, , , . . . , tmk) vertically, as follows: hk (2.9) This array is called a column vector or a column matrix. We have such a column vector foreach of the n elements T(e,), . .
.
, T(e,).
We place them side by side and enclose them inone pair of brackets to obtain the following rectangular array: -t11 t,, *-* t1, t21 t22 - * * t 2n. .1 :-. . . . t ml tnl2 * * * trim- This array is ca.lled a matrix consisting of m rows and n columns. We call it an m by n matrix,or an m x n matrix. The first row is the 1 x n matrix (tl-, , t,,, . . . , tl,). The m x 1matrix
displayed in (2.9) is the kth column. The scalars tik are indexed so the first subscripti indicates the row:, and the second subscript k indicates the column in which tik occurs.We call tik the ik-entry or the ik-element of the matrix. The more compact notation is also used to denote the matrix whose ik-entry is tik . 46 Linear transformations and matrices
Thus, every linear transformation T of an n-dimensional space V into an m-dimensionalspace W gives rise to an m x n matrix (tik) whose columns consist of the components ofT(e,>, . . . , T(e,) relative to the basis (wl, . . . , w,). We call this the matrix representationof T relative to the given choice of ordered bases (e,, .
. . , e,) for Vand (wl, .
. . , w,) forW. Once we know the matrix (tik), the components of any element T(x) relative to thebasis (w,, . . .
, w,) can be determined as described in the next theorem. THEOREM 2.13. Let T be a linear transformation in 9(V, W), bchere dim V = n anddim W = m. Let (e,, . . .
, e,,) and (wl, . . . , w,) be OrderedbasesJor Vand W, respectively,and let (tik) be the m x n matrix whose entries are determined by the equations (2.10) T(e,) = 2 tikWi 3 for k=1,2 ,.,., n.i=l Then an arbitrary element (2.11)n x = zxkekk = l in V with components (x1, . . .
, x,) relative to (e,, .
, , e,) is mapped by T onto the element (2.12) T(x) = 2 yiwii=l in W with components (yl, . . . , y,) relative to (w, , . . . , w,). The yi are related to thecomponents of x by the linear equations (2.13) Y, = i tikXk for i=l,2 ,.,., m.k=l Proof. Applying T to each member of (2.11) and using (2.10), we obtain where each yi is given by (2.13). This completes the
proof. Having chosen a pair of bases (e,, . . . , e,) and (+, . . . , MI,) for V and W, respectively,every linear transformation T: V + W has a matrix representation (tik). Conversely, ifwe start with any ~FIII scalars arranged as a rectangular matrix (rik) and choose a pair ofordered bases for V and W, then it is easy to prove that there is exactly one linear
trans-formation T: V-t W having this matrix representation. We simply define Tat the basiselements of V by the equations in (2.10). Then, by Theorem 2.12, there is one and onlyone linear transformation T: V + W with these prescribed values.
The image T(x) of anarbitrary point x in V is then given by Equations (2.12) and (2.13). Matrix representations of linear transformations 4 7_ _ _ _ _ _ _ EXAMPLE. 1. Construction of a linear transformation from a given matrix. Suppose westart with the 2 x 3 matrix r3 1 -21 Choose the usual bases of unit coordinate vectors for V, and V,. Then the given
matrixrepresents a linear transformation T: V, ---f V, which maps an arbitrary vector (xi, x2, x8)in V3 onto the vector Q1, yZ) in V, according to the linear equations y, = 3x, + xg - 2X,) yz = Xl + ox , + 4x,. EXAMPLE 2. (?OnStrUCtiOn of a matrix representation ?f a giVt?a linear transformation.Let V be the linear space of all real polynomials p(x) of
degree < 3. This space has dimen-sion 4, and we choose the basis (I, x, x2, x3). Let D be the differentiation operator whichmaps each polynomial p(x) in V onto its derivative p'(x). We can regard D as a lineartransformation of V into W, where W is the 3-dimensional space of all real polynomialsof degree <: 2'. In W we choose the basis (1, x, x2). To find
the matrix representation of Drelative to this (choice of bases, we transform (differentiate) each basis element of V andexpress it as a linear combination of the basis elements of W. Thus, we find that D(1) = 0 = 0 + ox + 0x2, D(x)=1 =1+0x+0x2, D(x2) = 2 x = 0 + 2 X + 0X2, D(x3)=3x2=0+0x+3X2. The coefficients of these polynomials determine the
columns of the matrix representation ofD. Therefore, the required representation is given by the following 3 x 4 matrix: 0 1 0 0 i I0 0 2 0 . 0 0 0 3 To emphasize that the matrix representation depends not only on the basis elements butalso on their order, let us reverse the order of the basis elements in Wand use, instead, theordered basis ($3, x, 1).
Then the basis elements of V are transformed into the same poly-nomials obtained above, but the components of these polynomials relative to the newbasis (x2, x., 1) appear in reversed order. Therefore, the matrix representation of D now becomes0 0 0 3 [I0 0 2 0 . 0 1 0 0 48 Linear transformations and matrices Let us compute a third matrix
representation for D, usingthebasis (1, 1 + x , 1 + x + x2,1 + x + x2 + x3) for V, and the basis (1, x, x2) for W. The basis elements of Vare trans-formed as follows: D(1) = 0, D(1 + x) = 1, D (l + x + x 2) = 1 + 2 x , D (l +x+x2+x3)=1 + 2 x + 3 x 2 , so the matrix representation in this case is 0 1 1 1[10 0 2 2 . 0 0 0 3 2.11 Construction of a matrix
representation in diagonal form Since it is possible to obtain different matrix representations of a given linear transforma-tion by different choices of bases, it is natural to try to choose the bases so that the resultingmatrix will have a particularly simple form.
The next theorem shows that we can makeall the entries 0 except possibIy along the diagonal starting from the upper left-hand cornerof the matrix. Along this diagonal there will be a string of ones followed by zeros, thenumber of ones being equal to the rank of the transformation. A matrix (tik) with allentries ti, = 0 when i # k is said to be a diagonal
matrix. THEOREM 2.14. Let V and W be finite-dimensional linear spaces, with dim V = n anddim W = m .
Assume T E Z(V, W) and let r = dim T(V) denote the rank of T. Then thereexists a basis (e, , . . .
, e,) for V and a basis (wl, .
. . , w,)for W such that (2.14) T(eJ = wi for i=1,2 ,..., r, and (2.15) T(e,)=O f o r i=r+l,...,n. Therefore, the matrix (tir) of T relative to these bases has all entries zero except for the rdiagonal entries tll = tzz = ' * ' = t,, = 1 . Proof. First we construct a basis for W. Since T(V) is a subspace of W with dim T(V) =r, the space T(V) has a basis of r elements in
W, say wr, . . . , w,, By Theorem 1.7, theseelements form a subset of some basis for W. Therefore we can adjoin elements w,+~, . . . ,w, so that (2.16) (Wl,. * * 7 WT, Wr+l, * . .
2 w,) is a basis for W. Construction of a matrix representation in diagonal form 49 Now we construct a basis for V. Each of the first r elements u'~ in (2.16) is the image of atleast one element in V. Choose one such element in V and call it e, . Then T(e,) = wji fori= 1,2,..., r so (2.14) is satisfied. Now let k be the dimension of the null space N(T).By
Theorem 2.3 we have n = k + r. Since dim N(T) =: k, the space N(T) has a basisconsisting of k elements in V which we designate as e,,, , . . . , er.tlc. For each of theseelements, Equation (2.15) is satisfied. Therefore, to comlplete the proof, we must showthat the ordered set (2.17) (e,, . .
. , e,, ql, . . . , q.+k) is a basis flar V. Since dim V = n = r + k, we need only show that these elements areindependent. Suppose that some linear combination of th,em is zero, say (2.18) Applying 7' and using Equations (2.14) and (2.19, we find that But wl,...,w, are independent, and hence c1 = * * * = c, = 0. Therefore, the first rterms in (2.18) a.re
zero, so (2.18) reduces to r+k i&lc,ei = O. But e,,, , * * * , er+kare independent since they form a basis for N(T), and hence c,+~ =. ..=c r+k = 0. Therefore, all the ci in (2.18) are zero, S,D the elements in (2.17) form abasis for V. This completes the proof. EXAMPLE. We refer to Example 2 of Section 2.10, where D is the differentiation operatorwhich
maps the space V of polynomials of degree < 3 into the space W of polynomials ofdegree 52. In this example, the range T(V) = W, so T has rank 3. Applying the methodused to prove Theorem 2.14, we choose any basis for W, for example the basis (1, x, x2).A set of polynomials in V which map onto these elements is given by (x, +x2, +x3). Weextend
this set to get a basis for V by adjoining the constant polynomial 1, which is a basisfor the null space of D. lTherefore, if we use the basis (x, Zx 2, &x3, 1) for V and the basis(1, x, x2) for W, the corresponding matrix representation for D has the diagonal form 1 0 0 0 1 10 1 0 0 . 10 0 1 01 50 Linear transformations and matrices 2.12 Exercises In all
exercises involving the vector space V, , the usual basis of unit coordinate vectors is to bechosen unless another basis is specifically mentioned. In exercises concerned with the matrix ofa linear transformation T: Y + W where V = W, we take the same basis in both Y and W unlessanother choice is indicated. 1. Determine the matrix of each of the
following linear transformations of V, into V, :(a) the identity transformation,(b) the zero transformation,(c) multiplication by a fixed scalar c. 2. Determine the matrix for each of the following projections.(a) T: V,- V,, where T(x, , x2, x3) = (x1, xJ .(b) T: V,+ 7',, where Th, x2, x2) = (x2, x2).(c) T: V,+ V,, where Wl,x2,x3,x4,x5) =(x2,x2,x4). 3.
A linear transformation T: V, + V, maps the basis vectors i andj as follows: T(i) = i + j, T(j)=2i-j. (a) Compute T(3i - 4j) and T2(3i - 4j) in terms of i and j.(b) Determine the matrix of T and of T2.(c) Solve part (b) if the basis (i, j) is replaced by (er , e,), where e, = i - j, e2 = 3i + j. 4. A linear transformation T: V, + V2 is defined as follows: Each vector (x, y)
is reflected inthe y-axis and then doubled in length to yield T(x, v). Determine the matrix of T and of T2. 5.
Let T: Va-t V, be a linear transformation such that T(k) =2i+3j+5k, T (j + k) =i, T (i + j + k) = j - k . (a) Compute T(i + 2j + 3k) and determine the nullity and rank of T.(b) Determine the matrix of T. 6. For the linear transformation in Exercise 5, choose both bases to be (e,, e2, e,), where e, =(2, 3, 5)) e2 = (1, 0, 0)) e3 = (0, 1, -1)) and determine the
matrix of T relative to the newbases. 7. A linear transformation T: Va -+ V2 maps the basis vectors as follows: T(i) = (0, 0)) T(j) =(1, I), T(k) = (1, -1).(a) Compute T(4i -j + k) and determine the nullity and rank of T.(b) Determine the matrix of T.(c) Use the basis (i, j, k) in V, and the basis (wr , we) in V,, where wr = (1, 1), w2 = (1,2).Determine the
matrix of T relative to these bases.(d) Find bases (er , e2, ea) for V, and (wr , wZ) for Y2 relative to which the matrix of Twill bein diagonal form. 8.
A linear transformation T: V2 --f V's maps the basis vectors as follows: T(i) = (I, 0, l),T(j) = (-l,O, 1).(a) Compute T(2i - 3j) and determine the nullity and rank of T.(b) Determine the matrix of T.(c) Find bases (e,, e,) for V, and (w,, w2, wa) for V, relative to which the matrix of Twill bein diagonal form.
9. Solve Exercise 8 if T(i) = (1, 0, 1) and T(j) = (1, 1, 1).10. Let Vand W belinear spaces, each with dimension 2 and each with basis (er , e2). Let T: V-t W be a linear transformation such that T(e, + e2) = 3e, + 9e,, T(3e, + 2e,) = 7e, + 23e,.(a) Compute T(e, - er) and determine the nullity and rank of T.(b) Determine the matrix of T relative to the given
basis. Linear spaces of matrices 5 1 (c) Use the basis (q , e,) for V and find a new basis of the form (el + uez, 2e, + be,) for W,relative to which the matrix of Twill be in diagonal form. In the linear space of all real-valued functions, each of the following sets is independent andspans a finite-dimensional subspace V. Use the given set as a basis for V and
let D: VA V bethe differentiation operator. In each case, find the matrix of D and of D2 relative to this choiceof basis. 11. (sin x, cos x). 15. (-cos x, sin x).12. (1, x, e5). 16. (sin x, cos x., x sin x, x cos x).13. (1, I -t x, I + x + er). 17. (eX sin x, e2 cos x).14. (e", xe"). 18. (ezz sin 3x, ezs cos 3x).
19. Choose the basis (1, x, x2, x3) in the linear space V of all real polynomials of degree 13.Let D denote the differentiation operator and let T: V + 1' be the linear transformationwhich map'sp(x:, onto x$(x). Relative to the given basis, determine the matrix of each of thefollowingtransformations: (a) T; (b) DT; (c) TD; (d) TD - DT; (e) T2; (f) T2D2 - D2T2.
20. Refer to Exercise 19. Let W be the image of Vunder TD. Find bases for Vand for W relativeto which the matrix of TD is in diagonal form. 2.13 Linear spaces of matrices We have seen how matrices arise in a natural way as representations of linear trans-formations. Matrices can also be considered as objects existing in their own right,
withoutnecessarily being connected to linear transformations. As such, they form another class ofmathematical objects on which algebraic operations can be defined. The connectionwith linear transformations serves as motivation for these definitions, but this connectionwill be ignored for the moment. Let nz and n be two positive integers, and let I,,,
be the set of all pairs of integers (i,j)such that 1 I; i 5 m, 1 Any function A whose domain is I,,, is called an m x nmatrix. The flJnction value A(i, j) is called the g-entry or ij-element of the matrix and willbe denoted also by aij . It is customary to display all the function values in a rectangulararray consisting of m rows and n columns, as follows: azl az2 **a
azn. . The elements a;j may be arbitrary objects of any kind. Usually they will be real or complexnumbers, but sometimes it is convenient to consider matrices whose elements are otherobjects, for example, functions. We also denote matrices by the more compact notation A = (u,,)~;~~ or A = (aij). If m = n , the matr.ix is said to be a square matrix. A
1 x n matrix is called a row matrix;an m x 1 matrix is called a column matrix. 52 Linear transformations and matrices Two functions are equal if and only if they have the same domain and take the samefunction value at each element in the domain.
Since matrices are functions, two matricesA = (a,J and B = (bJ are equal if and only if they have the same number of rows, thesame number of columns, and equal entries aij = bij for each pair (i,j). Now we assume the entries are numbers (real or complex) and we define addition ofmatrices and multiplication by scalars by the same method used for
any real- or complex-valued functions. DEFINITION. If A = (aij) and B = (bij) are two m x n matrices and if c is any scalar,we define matrices A + B and CA as follows: A + B = (a,j + b,j>, CA = (caij). The sum is defined only when A and B have the same size. EXAMPLE. If 1 2 - 3A = -1 0 4 I and B = [: -1 i], then we have A+.=[; -; -:]. 2A= [-'z ; -;], (-1)B=
[I; ; I:]. We define the zero matrix 0 to be the m x n matrix all of whose elements are 0. -Withthese definitions, it is a straightforward exercise to verify that the collection of all m x nmatrices is a linear space. We denote this linear space by M,,,. If the entries are realnumbers, the space M,,, is a real linear space. If the entries are complex, M,,, is a
complexlinear space. It is also easy to prove that this space has dimension mn. In fact, a basis forM consists of the mn matrices having one entry equal to 1 and all others equal to 0.Fo?"example, the six matrices form a basis for the set of all 2 x 3 matrices. 2.14 Isomorphism between linear transformations and matrices We return now to the
connection between matrices and linear transformations. Let Vand W be finite-dimensional linear spaces with dim V = n and dim W = m. Choose abasis (e,, . . . , e,) for V and a basis (wl, . .
. , w,) for W. In this discussion, these bases arekept fixed. Let p(V, W) denote the linear space of all linear transformations of V intoW. If T E P(V, W), let m(T) denote the matrix of T relative to the given bases. We recallthat m(T) is defined as follows. Zsomorphism between linear transformations and matrices 53 The image of each basis element e, is
expressed as a linear combination of the basiselements in IV: (2.19) T(e,) = g tikWi for k=l,2 ,..., n.i=l The scalar multipliers ci, are the ik-entries of m(T). Thus, we have (2.20) 47 = (4,>~&. Equation (2.20) defines a new function m whose domain is L?'(V, W) and whose valuesare matrices in M,, ~. Since every m x n matrix is the,matrix m(T) for some
Tin A?(V, W),the range of m is ii,,, . The next theorem shows that the transformation m: LZ(V, W) +M112.11 is linear and one-to-one on LY(V, W). THEOREM 2.15. ISOMORPHISM THEOREM. For all S and 1" in P(V, W) and all scalarsc, we have 4s + T> = m(S) + m(T) and m(cT) = cm(T).Moreover, m(S) = m(T) implies S = T, so m is one-to-one on
9(V, W). Proof. The matrix m(T) is formed from the multipliers ti, in (2.19). Similarly, thematrix m(S) is formed from the multipliers sik in the equations (2.21) Since we have S(e,) = 5 SikWi for k=1,2 ,..., n.i=l and we obtainm(S + 73 = (sik + tile) = m(S) + m(T) and m(cT) = (ctJ = cm(T). This provesthat m is linear. To prove that m is one-to-one,
suppose that m(S) = m(Z3, where S = (sik) and T =(t&. Equations (2.19) and (2.21) show that S(e,) = T(e,) for each basis element e,,so S(x) = T(x) for all x in V, and hence S = T. Note: The function m is called an isomorphism. For a given choice of bases, mestablishes a one-to-one correspondence between the set of linear transformationsU(V, W) and
the set of m x n matrices M,,, .
The operations of addition and multipli-cation by scalars are preserved under this correspondence. The linear spaces -Y(V, W) and Mm,, are said to be isomorphic.
Incidentally, Theorem 2.11 shows that the domainof a one-to-one linear transformation has the same dimension as its range. Therefore,dim Y(V, IV) = dim M,,, = mn . If V = Wand if we choose the same basis in both V and W, then the matrix m(Z) whichcorresponds to the identity transformation I: V ---f V is an n x n diagonal matrix with each 5 4 Linear
transformations and matrices diagonal entry equal to 1 and all others equal to 0. This is called the identity or unit matrixand is denoted by I or by I,. 2.15 Multiplication of matrices Some linear transformations can be multiplied by means of composition. Now we shalldefine multiplication of matrices in such a way that the product of two matrices
correspondsto the composition of the linear transformations they represent. We recall that if T: U--f V and S: V + Ware linear transformations, their compositionST: U--j W is a linear transformation given by ST(x) = S[T(x)] f o r a l l x i n U. Suppose that U, V, and Ware finite-dimensional, say dim U = n, dim V=p, dim W= m. Choose bases for U, V, and
W. Relative to these bases, the matrix m(s) is an m x pmatrix, the matrix T is a p x n matrix, and the matrix of ST is an m x n matrix. Thefollowing definition of matrix multiplication will enable us to deduce the relation m(ST) =m(S)m(T). This extends the isomorphism property to products. DEFINITION. Let A be any m x p matrix, and let B be any p x n
matrix, say A = (u,,);;Z)~ and B = (bij);,8jn=1.
Then the product AB is defined to be the m x n matrix C = (cJ whose ij-entry is given by (2.22) Note: The product AB is not defined unless the number of columns of A is equal tothe number of rows of B.
If we write Ai for the ith row of A, and Bi for thejth column of B, and think of these asp-dimensional vectors, then the sum in (2.22) is simply the dot product Ai * Bj. In otherwords, the q-entry of AB is the dot product of the ith row of A with thejth column of B: AB = (Ai . B');3"1. Thus, matrix multiplication can be regarded as a generalization of the dot

product. ExAMPLnl.
LetA= [-: : i]andB= [--!I. S i n c e A i s 2 x 3andBis3 x 2 , Multiplication of matrices 55 the product AB is the 2 x 2 matrix The entries of AB are computed as follows: A,.B1=3.4+1.5+2.0=17, A, .
B2 = 3*6+1.(-1)+2.2=21, A, h B1 = (-1) * 4 + 1 -5 + 0.0 = 1, A,*B2=(---1)*6+1.(-1)+0.2=-7. EXAMPLE 2. Let A= [: : -3 and B= -i . [I Here A is; 2 x 3 and B is 3 x 1, so AB is the 2 x I matrix given by A B = [;::;I] = [-;I, since A, * B1 = 2.(-2)+1.1+(-3).2= -9andA,nBr= l-(-2)+2*1 +4*2=8. EXAMPLE 3. If A and B are both square matrices of the same size,
then both AB and BAare defined. For example, if 3 4A = and B = [15 2' we find that AB=[1: -3, BA=[-: ;;I. This example shows that in general AB # BA . If AB = BA , we say A and B commute. EXAMPLE 4. If 1, is the p x p identity matrix, then I,Aand BI, = B for every m x p matrix B.
For example, = A for every p x n matrix A, 0 1 [1 2 3 0 = 4 51 I6 ' Now we prove that the matrix of a composition ST is the product of the matrices m(S)and m(r). 56 Linear transformations and matrices THEOREM 2.16. Let T: U + V and S: V --f W be linear transformations, where U, V, Ware jnite-dimensional linear spaces. Then, for a$xed choice of
bases, the matrices of S, T,and ST are related by the equation Proof. Assume dim U = n, dim V = p, dim W = m. Let (ul, . . . , u,) be a basis for U,h,..., v,) a basis for V, and (wl, . . . , w,) a basis for W. Relative to these bases, wehave and m(S) = (sii)~;El, where S(v,) = 5 SikWi for k = 1,2, . .
. , p,i=l m(T) = (fij)Y;kl, where T(uJ = i tkjvR for j=1,2 ,..., n.k=l Therefore, we have so we find that = m(S)m(T). We have already noted that matrix multiplication does not always satisfy the com-mutative law. The next theorem shows that it does satisfy the associative and distributivelaws. THEOREM 2.17. ASSOCIATIVE AND DISTRIBUTIVE LAWS
FOR MATRIX MULTIPLICATION. Given matrices A, B, C.(a) If the products A(BC) and (AB)C are meaningful, we have A(K) = (AB)C (associative law).
(b) Assume A and B are of the same size. If AC and BC are meaningful, we have (A + B)C = AC + BC (right distributive law), whereas if CA and CB are meaningjiil, we have C(A + B) = CA + CB (left distributive law). Proof. These properties can be deduced directly from the definition of matrix multi-plication, but we prefer the following type of
argument.
Introduce finite-dimensionallinear spaces U, V, W, X and linear transformations T: U---f V, S: V---f W, R: W-t Xsuch that, for a fixed choice of bases, we have A = m(R), B = m(S), C=m(T). Exercises 5 7 By Theorem 2.16, we have m(M) = AB and m(ST) = BC From the associative law forcomposition, we find that R(ST) = (RS)T. Applying Theorem 2.16
once more to thisequation, we obtain m(R)m(ST) = m(RS)m(T) or A(X) = (AB)C, which proves (a). Theproof of (b) can be given by a similar type of argument. DEFINITION. VA is a square matrix, we define integral powers of A inductively as follows: A” = I, A” = AA-1 f o r n>l. 2.16 Exercises ,.,,.=[_: -1 -;I, B+ -3. c= BA, AC, CA, .4(2B - 3C). i-0 11
compute B + C, AB, 2.LetA=1 J0 2 .
Find all 2 x 2 matrices B such that (a) .4B = 0 ; (b) BA = 0. 3. In each case find a, b, c, d to satisfy the given equation. (4a b c d (b) [11 4 9 2 4. Calculate AB - BA in each case. (b)A=[; ; ;], B=[-; -f ---]. 5. If A is a square matrix, prove that A”Am = Am-tn for all integers m 2 0, n 2 0. 1 and compute A”.
Compute A3 and A4. Guess a general formula for A” and prove it by induction, 58 Linear transformations and matrices 1 09. Let A = [I-1 1 *Prove that A2 = 2A - Z and compute Aloo. 10. Find all 2 x 2 matrices A such that A2 = 0.11. (a) Prove that a 2 x 2 matrix A commutes with every 2 x 2 matrix if and only if A commutes with each of the four
matrices (b) Find all such matrices A.12. The equation A2 = Z is satisfied by each of the 2 x 2 matrices where b and c are arbitrary real numbers. Find all 2 x 2 matrices A such that A2 = I. U.IfA=[-: -t] andB=[i 11, find 2 x 2 matrices C and D such that AC = B a n d D A = B .14.
(a) Verify that the algebraic identities (A + B)2 = A2 + 2AB + B2 and (A + B)(A - B) = A2 - B2 do not hold for the 2 x 2 matrices A = [t -t]andB=[: J. (b) Amend the right-hand members of these identities to obtain formulas valid for all squarematrices A and B.(c) For which matrices A and B are the identities valid as stated in (a)? 2.17 Systems of linear
equations Let A = (aJ be a given m x n matrix of numbers, and let cl, . . . , c, be m furthernumbers. A set of m equations of the form (2.23) $Fixxk = ci for i = 1, 2, . . . , m, is called a system of m linear equations in n unknowns. Here x1, . . .
, X, are regarded asunknown. A solution of the system is any n-tuple of numbers (x1, . . . , x,) for which all theequations are satisfied. The matrix A is called the coefJicient-matrix of the system. Linear systems can be studied with the help of linear transformations. Choose the usualbases of unit coordinate vectors in V, and in V,,. The coefficient-matrix A
determines a Systems of linear equations 59 linear transformation, T: V, + V,, which maps an arbitrary vector x = (xi, . . .
, x,) in V,onto the vector y = (yl, . .
.
, ym) in V, given by the m linear equations yi = i aikXk for i=1,2 ,..., m.k=l Let c = (cr , . . . , c,) be the vector in V, whose components are the numbers appearing insystem (2.23). This system can be written more simply as; T(x) = c. The system has a solution if and only if c is in the range of T. If exactly one x in V,, mapsonto c, the system has exactly
one solution. If more than one x maps onto c, the systemhas more than one solution. EXAMPLE: 1. A system with no solution. The system x + y = 1, x + y = 2 has nosolution. ‘The sum of two numbers cannot be both I and 2. EXAMPLE 2.
.4 system with exactly one solution. The system x + y = 1, x - y = 0 hasexactly one solution: (x, y) = (4, 4). EXAMPLE: 3. .4 system with more than one solution. The system x + y = 1 , consistingof one equation in two unknowns, has more than one solution. Any two numbers whosesum is 1 gives ii solution. With each linear system (2.23), we can
associate another system &kXk = 0 for i = 1, 2, . . . , m, obtained by replacing each ci in (2.23) by 0. This is called the homogeneous system corre-sponding to (2.23). If c Z 0, system (2.23) is called a nonhomogeneous system. A vectorx in V, will satisfy the homogeneous system if and only if T(x) = 0, where T is, the linear transformation determined by
the coefficient-matrix. The homogene-ous system always has one solution, namely x = 0, but it may have others. The set ofsolutions of the homogeneous system is the null space of 7: The next theorem describes therelation between solutions of the homogeneous system and those of the nonhomogeneoussystem. THEOREM 2.18. Assume the
nonhomogeneous system (2.23) has a solution, say 6.(a) If a vector x is a solution of the nonhomogeneous system, then the vector v = x - b is a solution of the corresponding homogeneous system.(b) If a vector v is a solution of the homogeneous+ system, then the vector x = v + b is a solution #of the nonhomogeneous system. 60 Linear
transformations and matrices Proof. Let T: V, + V, be the linear transformation determined by the coefficient-matrix, as described above. Since b is a solution of the nonhomogeneous system we haveT(b) = c . Let x and v be two vectors in V, such that v = x - b. Then we have T(u) = T(x - b) = T(x) - T(b) = T(x) - c.
Therefore T(x) = c if and only if T(v) = 0. This proves both (a) and (b). This theorem shows that the problem of finding all solutions of a nonhomogeneoussystem splits naturally into two parts: (1) Finding all solutions v of the homogeneoussystem, that is, determining the null space of T; and (2) finding one particular solution b ofthe nonhomogeneous
system. By adding b to each vector v in the null space of T, we therebyobtain all solutions x = v + b of the nonhomogeneous system. Let k denote the dimension of N(T) (the nullity of T). If we can find k independentsolutions vi, . . . , vk of the homogeneous system, they will form a basis for N(T), and wecan obtain every v in N(T) by forming all possible
linear combinations u = t,v, + * * * + t,v, , where t,, . . . , t, are arbitrary scalars. This linear combination is called the general solution of the homogeneous system. If b is one particular solution of the nonhomogeneous system,then all solutions x are given by x = b + t,V, + * * . + t,V,. This linear combination is called the general solution of the
nonhomogeneous system.Theorem 2.18 can now be restated as follows. THEOREM 2.19. Let T: V, -+ V, be the linear transformation such that T(x) = y, wherex = (x1,*--, x,), y = Q1, . . . , ym> and Yi = &ikXk f o r i=l,2 ,,.., m.k = l Let k denote the nullity of T. If vl, . .
. , vk are k independent solutions of the homogeneous system T(x) = 0 , and if b is one particular solution of the nonhomogeneous system T(x) = c ,then the general solution of the nonhomogeneous system is x = b + t,v, + . . . + tkvk, where t,, . . . , t, are arbitrary scalars.
This theorem does not tell us how to decide if a nonhomogeneous system has a particularsolution b, nor does it tell us how to determine solutions vl, . . . , vk of the homogeneoussystem. It does tell us what to expect when the nonhomogeneous system has a solution.The following example, although very simple, illustrates the theorem. Computation
techniques 6 1 EXAMPLE. The system x + y = 2 has for its associated homogeneous system the equationx + y = 0. Therefore, the null space consists of all vectors in V, of the form (t, -t),where t is arbitrary. Since (t, -t) = t(1, -I), this is a one-dimensional subspace of V,with basis (1, -- 1). A particular solution of the nonhomogeneous system is (0,2).
There-fore the general solution of the nonhomogeneous system is given by (x,y> = (0,2) + t(1, -1) or x = t, y=2-tt, where t is arbitrary. 2.18 Computation techniques We turn now to the problem of actually computing the solutions of a nonhomogeneouslinear system. Although many methods have been developed for attacking this problem,all of them
require considerable computation if the system is large. For example, to solvea system of ten equations in as many unknowns can require several hours of hand com-putation, (even with the aid of a desk calculator.
We shall discuss a widely-used method, known as the Gauss-Jordan elimination method,which is relatively simple and can be easily programmed for high-speed electronic computingmachines. The method consists of applying three basic types of operations on the equationsof a linear system: (1) Interchanging two equations;(2) Mui’tiply,ing all the
terms of an equation by a nonzero scalar:(3) Adding tlg one equation a multiple of another. Each time we perform one of these operations on the system we obtain a new system havingexactly the same solutions.
Two such systems are called equivalent. By performing theseoperations over and over again in a systematic fashion we finally arrive at an equivalentsystem which can be solved by inspection. We shall illustrate the method with some specific examples. It will then be clear how themethod is to be applied in general. EXAMPLE 1. .4 system with a
unique solution. Consider .the system 2x-5y+42= - 3x-2y+ z=5x-4y+6z=lO. This partiizular system has a unique solution, x = 124, y = 75, z = 31 , which we shallobtain by the Gauss-Jordan elimination process.
To save labor we do not bother to copythe letters x, y,, z and the equals sign over and over again, but work instead with the aug-mented matrix 2 -5 4 -3 (2.24) l-2 1 5 1 -4 6 10 6 2 Linear transformations and matrices obtained by adjoining the right-hand members of the system to the coefficient matrix. Thethree basic types of operation mentioned
above are performed on the rows of the augmentedmatrix and are called row operations. At any stage of the process we can put the lettersx, y, z back again and insert equals signs along the vertical line to obtain equations. Ourultimate goal is to arrive at the augmented matrix 1 0 0 124 (2.25) [10 10 75 0 0 1 3 1 after a succession of row operations.
The corresponding system of equations is x = 124,y = 75, z = 31 , which gives the desired solution. The first step is to obtain a 1 in the upper left-hand corner of the matrix.
We can do thisby interchanging the first row of the given matrix (2.24) with either the second or thirdrow. Or, we can multiply the first row by h.
Interchanging the first and second rows, we get l-2 1 5 ! I2 -5 4 -3 . 1 -4 6 10 The next step is to make all the remaining entries in the first column equal to zero, leavingthe first row intact. To do this we multiply the first row by -2 and add the result to thesecond row. Then we multiply the first row by -1 and add the result to the third row.After these
two operations, we ob tstin (2.26) ‘1 -2 1 0 -1 2 - 1 3 . o - 2 5 5 15Now we repeat the process on the smaller matrix [1: : / -‘:I which appears adjacent to the two zeros. We can obtain a 1 in its upper left-hand corner by multiplyingthe second row of (2.26) by - 1 . This gives us the matrixi 0 o-2 l-2 1 -2 5 1 13 5 5 1 . Multiplying the second row by 2 and
adding the result to the third, we get l - 2 1 5 (2.27) [0 1 - 2 13 I . 0 0 1 3 1 Computation techniques At this stage, the corresponding system of equations is given by 63 x-2y+ z= 5 y - 22 = 13 2 = 31.
These equations can be solved in succession, starting with the third one and workingbackwards, to give us z = 31, y = 13 + 22 = 13 + 62 = 75, x = 5 + 2~ - z = 5 + 150 - 31 = 124.
Or, we can continue the Gauss-Jordan process by making all the entries zero above thediagonal elernents in the second and third columns. MultiFllying the second row of (2.27)by 2 and adding rhe result to the first row, we obtain [0 0 1 0 0 1 -3 -2 1 31 31 13 1 . Finally, we multiply the third row by 3 and add the result to the first row, and then
multiplythe third row by :2 and add the result to the second row to get the matrix in (2.25). EXAMPLE 2. A system with more than one solution. Consider the following system of 3equations in 5 unknowns: 2x-55y+4z+ u-vu-3 (2.28) x-2y+ z - u+v=5 x-44y+62+2u--v=lO. The corresponding augmented matrix is [2-5 1 1 -2 -4 4 6 l-l 2 1 -1 -1 1 -3 10 5. 1
The coefficients of x, y, z and the right-hand members are the same as those in Example 1.If we perform the same row operations used in Example 1, we: finally arrive at the augmentedmatrix 1 0 0 - 1 6 19 124 0 1 0 -9 1 1 7 5 1 . 0 0 1 -3 4 3 1 64 Linear transformations and matrices The corresponding system of equations can be solved for x, y, and z
in terms of u and u,giving us x = 124 + 162~ - 190 y = 75+ 9U-110 z = 3 1 + 3u- 4v. If we let u = t, and u = t,, where t, and t2 are arbitrary real numbers, and determinex, y, z by these equations, the vector (x, y, z, U, u) in V, given by (x,y,z, u, u) = (124 + 16t, - 19t,, 75 + 9t, - llt,, 31 + 3t, - 4t,, t,, t,) is a solution.
By separating the parts involving t, and t,, we can rewrite this as follows: (x,y,z, u, u) = (124, 75, 31,0,0) + t1(16, 9,3, 1,0) + &C-19, -11, -4,O, 1)s This equation gives the general solution of the system.
The vector (124, 75, 31,0,0) is aparticular solution of the nonhomogeneous system (2.28). The two vectors (16, 9, 3, 1,0)and (-19 , -11, -4 , 0, 1) are solutions of the corresponding homogeneous system. Sincethey are independent, they form a basis for the space of all solutions of the homogeneoussystem. EXAMPLE 3. A system with no solution.
Consider the system 2x-5y+4z= - 3 (2.29) x-2y+ z = 5 x-4y+5z= 1 0 . This system is almost identical to that of Example 1 except that the coefficient of z in the third equation has been changed from 6 to 5. The corresponding augmented matrix is i 2 1 l-2 -5 -4 4 5 1 -3 10 5.IApplying the same row operations used in Example 1 to transform (2.24) into
(2.27), wearrive at the augmented matrix (2.30) i 0 0 l-2 0 1 -2 0 1 3 1 1 3 5 1 . When the bottom row is expressed as an equation, it states that 0 = 31. Therefore theoriginal system has no solution since the two systems (2.29) and (2.30) are equivalent. Inverses of square matrices 65 In each of the foregoing examples, the number of equations did not
exceed the numberof unknowns. If there are more equations than unknowm, the Gauss-Jordan process isstill applicable. For example, suppose we consider the syste:m of Example 1, which has thesolution x = 124 , , y = 75, z = 31. If we adjoin a new equation to this system which is alsosatisfied by the same triple, for example, the equation 2x - 3~ + z
= 54, then theelimination process leads to the agumented matrix 0 10 75 0 0 1 3 1 000 0 with a row of zeros along the bottom. But if we adjoin a new equation which is not satisfiedby the triple (124, 75, 31), for example the equation x + y + z = 1, then the eliminationprocess leads to an augmented matrix of the form 1 0 0 124 where a # 0. The last
row now gives a contradictory equation 0 = a which shows thatthe system has no solution. 2.19 Inverses of square matrices Let A = (a,J be a square n x n matrix. If there is another n x n matrix B such thatBA = I, where Z is the n x n identity matrix, then A is called nonsinpdar and B is called aleft inverse of A. Choose the usual basis of unit coordinate
vectors in V, and let T: V, + V, be the lineartransformation with matrix m(7’) = A. Then we have the following. THEOREM 2.20. The matrix A is nonsingular if and only if T is invertible. If BA = Ithen B = m(T-I). Proof. Assume that A is nonsingular and that BA = I. We shall prove that T(x) = 0implies x = 0.
Given x such that T(x) = 0, let X be the n x 1 column matrix formedfrom the components of x. Since T(x) = 0, the matrix product AX is an n x 1 columnmatrix consisting of zeros, so B(AX) is also acolumn matrixof zeros. But B(AX) = (BA)X =IX = X, so every component of x is 0. Therefore, Tis invertible, and the equation TT-l= Zimplies that m(73m(T-l) =
Z or Am(T-l) = I. Multiplying on the left by B, we findm(T-l) = B , , Conversely, if T is invertible, then T-lT is the identity transformation som(T-l)m(T)i.s the identity matrix. Therefore A is nonsingular and m(F)A = I. 66 Linear transformations and matrices All the properties of invertible linear transformations have their counterparts for non-singular
matrices. In particular, left inverses (if they exist) are unique, and every leftinverse is also a right inverse. In other words, if A is nonsingular and BA = I, thenAB = I. We call B the inverse of A and denote it by A-l. The inverse A-’ is also non-singular and its inverse is A. Now we show that the problem of actually determining the entries of the inverse of
anonsingular matrix is equivalent to solving n separate nonhomogeneous linear systems. Let A = (a,J be nonsingular and let A-l = (hij) be its inverse. The entries of A andA-l are related by the II? equations (2.31) where aij = 1 if i = j , and dij = 0 if i # j.
For each fixed choice of j, we can regard thisas a nonhomogeneous system of n linear equations in n unknowns bli, bzj, .
. .
, b,, .
SinceA is nonsingular, each of these systems has a unique solution, the jth column of B. Allthese systems have the same coefficient-matrix A and differ only in their right members.For example, if A is a 3 x 3 matrix, there are 9 equations in (2.31) which can be expressedas 3 separate linear systems having the following augmented matrices : If we apply
the Gauss-Jordan process, we arrive at the respective augmented matrices In actual practice we exploit the fact that all three systems have the same coefficient-matrixand solve all three systems at once by working with the enlarged matrix all i 62 al3 a21 az2 a23 a3, a32 a33 The elimination process then leads to 1 0 0 0 1 0 .
0 0 1 1 Exercises 67 The matrix on the right of the vertical line is the required inverse. The matrix on the leftof the line is the 3 x 3 identity matrix. It is not necessary to know in advance whether A is nonsingular. If A is singular (notnonsingular), we can still apply the Gauss-Jordan method, but somewhere in the processone of the diagonal elements
will become zero, and it w.ill not be possible to transform Ato the identity matrix. A system of n linear equations in n unknowns, say i = 1, 2, .
. . , n , can be written more simply as a matrix equation, AX= C, where A := (aij) is the coefficient matrix, and X and C are column matrices, Xl Cl X2 c2 x= :)1.. ::c= 1 . X73 cn If A is nonsingular there is a unique solution of the system given by X = A-T. 2.20 Exercises Apply the Gauss-Jordan elimination process to each of the following systems. If a
solutionexists, determine the general solution. 1. x+y+3z= 52x -;v + 4z = 11 -y + z = 3.2. 3x + 2y + z = 1 5x :3y 32 =+ + 2xs y - z = l . 3. 3x + 2y + z = 15x :3y 32 =+ + 27x + 4y + sz = 3.
4. 3x + 2y + z = 15x 3y 32 =+ + 27x + 4y + 5z = 3 x + y - z = o . 5 . 3X - 2J1 + 5Z + U = 1X + J’ - 3Z + 2U = 26x+ JI-42+3U=7. 6. x+y-3z+ u=52x-y+ z-2u=27x+y-72+3u=3. 7. X + J’ + 22 + 3U + 4U = 02X + 2J, + 72 + 1lU + 14U = 03x + 3J' + 62 + 10~ + 15~ =O. 8. X - 21’ + Z + 2U = - 22X + 3J, - Z - 5U = 94 x - Jj+Z-U= 55X - 31’ + 22 + U = 3.
9.Provethatthesystemx+y+2~=2,2x-y+3~=2,5x--y+az=6,hasauniquesolution if IZ # 8. Find all solutions when a = 8. 10. (a) Determine all solutions of the system 5x+2y-62+2u= --1 x- y+ z- u=--2. 68 Linear transformations and matrices (b) Determine all solutions of the system 5x +2y -6z+2u= - 1 x - y + z - u = - 2 x+y+z = 6./ 11. This exercise tells
how to determine all nonsingular 2 x 2 matrices. Prove that a bDeduce that [1c d is nonsingular if and only if ad - bc # 0, in which case its inverse is Determine the inverse of each of the matrices in Exercises 12 through 16. 13. 14.
15. 16. ‘ 1 2 3 4 0 1 2 3 10 0 12’ Ip 0 0 1_1 -0 1 0 0 0 o- 2 0 2 0 0 0 0 3 0 1 0 0 _ ! 0 0 1 0 2 0 ’ 0 0 0 3 0 1 0 0 0 0 2 0 2.21 Miscellaneous exercises on matrices 1. If a square matrix has a row of zeros or a column of zeros, prove that it is singular.2. For each of the following statements about II x n matrices, give a proof or exhibit a counter example.(a) If
AB + BA = 0, then A2B3 = B3A2.(b) If A and B are nonsingular, then A + B is nonsingular.(c) If A and B are nonsingular, then AB is nonsingular.(d) If A, B, and A + B are nonsingular, then A - B is nonsingular.(e) If A3 = 0, then A - Z is nonsingular.(f) If the product of k matrices A, .
. . A, is nonsingular, then each matrix Ai is nonsingular.
Miscellaneous exercises on mawices 69 find a nonsingular matrix P such that P-‘AP = a i4. The matrix A = [I , where i2 = i b-1,a =$(l + $$,andb =$(l - &),hastheprop- erty that A2 = A. Describe completely all 2 x 2 matrices A with complex entries such thatA2=A. 5.
If A2 = A, prove that (A + I)” = Z f (2k - l)A.6. The special theory of relativity makes use of a set of equations of the form x’ = a(x - ut), y’ = y, z’ = z, t’ = a(t - vx/c2). Here v represents the velocity of a moving object, c thespeed of light, and a = c/Jc2 - v2, where Ju] < c . The linear transformation which maps thetwo-dimensional vector (x, t) onto (x’, t’)
is called a L,orentz transformation. Its matrixrelative to the usual bases is denoted by L(v) and is given by 1 - vL(u) = a[I,,--DC-Z 1 Note that L(v) is nonsingular and that L(O) = I. Prove that L(v)L(u) = L(w), where w =(u + v)c2/(uv + c2). In other words, the product of two Lorentz transformations is anotherLorentz transformation.
7. If we interchange the rows and columns of a rectangular matrix A, the new matrix so obtainedis called the transpose of A and is denoted by At. For example, if we have [1 2 3 A = 4 5 16 ’ Prove that transposes have the following properties :(a) (At)t = A(d) (AB)t = B”A”. (b) (A + B)t = At + Bt. (c) (cA)~ = cAt,(e) (At)-l = (A-l)t if A is nonsingular. 8.
A square matrix A is called an orthogonal matrix if AA t := I. Verify that the 2 x 2 matrix [::e” z]is orthogonal for each real 0. If A is any n X II orthogonal matrix, prove that its rows, considered as vectors in V, , form an orthonormal set.9. For each of the following statements about n x n matrices, give a proof or else exhibit a counter example.(a) If A
and B are orthogonal, then A + B is orthogonal.(b) If A and B are orthogonal, then AB is orthogonal.(c) If A and B are orthogonal, then B is orthogonal. 10. Hadumard matrices, named for Jacques Hadamard (1865-1963), are those n x n matriceswith the following properties: I. Each entry is 1 or -1 .II. Each row, considered as a vector in V, , has length
&. III. The dot product of any two distinct rows is 0.Hadamard matrices arise in certain problems in geometry and the theory of numbers, andthey have been applied recently to the construction of optimum code words in space com-munication. In spite of their apparent simplicity, they present many unsolved problems. The 70 Linear transformations
and matrices main unsolved problem at this time is to determine all n for which an n x n Hadamard matrixexists. This exercise outlines a partial solution.(a) Determine all 2 x 2 Hadamard matrices (there are exactly 8).(b) This part of the exercise outlines a simple proof of the following theorem: ZfA is an n x nHadamard matrix, where n > 2, then n is a
multiple of 4. The proof is based on two very simplelemmas concerning vectors in n-space.
Prove each of these lemmas and apply them to therows of Hadamard matrix to prove the theorem. LEMMA 1. If X, Y, Z are orthogonal vectors in V, , then we have (x+ Y)*(X+Z)= lpq2. LEMMA 2.
Write X=(x, ,..., xn), Y=(yl ,..., yn), Z=(z, ,..., zn). Zf eachcomponent xi, yi, zi is either 1 or - 1 , then the product (xi + yi)(xi + zi) is either 0 or 4. 3DETERMINANTS 3.1 Introduction In many applications of linear algebra to geometry and analysis the concept of adeterminant plays an important part. This chapter studies the basic properties of determi-
nants and some of their applications. Determinants of order two and three were intoduced in Volume I as a useful notation forexpressing certain formulas in a compact form. We recall ‘that a determinant of order twowas defined by the formula (3.1)Qll QlzI I = Q l l Q 2 2 - Ql2Q21 *Q21 a 2 2 Despite similarity in notation, the determinantQll I I‘I2
(written with vertical bars) is a21 Q22 conceptually distinct from the matrix (‘ttwn en with square brackets).
The determinant is ;i number assigned to the matrix according to Formula (3.1). To emphasizethis connection we also write Determinants of order three were defined in Volume I in terms of second-order determi-nants by the formula This chapter treats the more general case, the determinant of a square matrix of order nfor any integer n 2 1 . Our
point of view is to treat the: determinant as a function which 71 72 Determinants assigns to each square matrix A a number called the determinant of A and denoted by det A.It is possible to define this function by an explicit formula generalizing (3.1) and (3.2).This formula is a sum containing n! products of entries of A. For large n the formula
isunwieldy and is rarely used in practice. It seems preferable to study determinants fromanother point of view which emphasizes more clearly their essential properties. Theseproperties, which are important in the applications, will be taken as axioms for a determi-nant function. Initially, our program will consist of three parts: (1) To motivate
thechoice of axioms. (2) To deduce further properties of determinants from the axioms.(3) To prove that there is one and only one function which satisfies the axioms. 3.2 Motivation for the choice of axioms for a determinant function In Volume I we proved that the scalar triple product of three vectors A,, A,, A, in 3-spacecan be expressed as the
determinant of a matrix whose rows are the given vectors. Thuswe have [1all al2 al3A, x A, - A, = det aZ1 aZ2 aZ3 , a31 a32 a33 where Al = (a,, , a12, a13>, A2 = (azl , a22, a23>, and A3 = (a31, a32T a33) - If the rows are linearly independent the scalar triple product is nonzero; the absolutevalue of the product is equal to the volume of the
parallelepiped determined by the threevectors A,, A,, A,. If the rows are dependent the scalar triple product is zero.
In this casethe vectors A,, A,, A, are coplanar and the parallelepiped degenerates to a plane figure ofzero volume. Some of the properties of the scalar triple product will serve as motivation for the choiceof axioms for a determinant function in the higher-dimensional case. To state theseproperties in a form suitable for generalization, we consider the
scalar triple product as afunction of the three row-vectors A,, A,, A,. We denote this function by d; thus, d(A,,A,,A,)=A, X A2*A3. We focus our attention on the following properties:(a) Homogeneity in each row. For example, homogeneity in the first row states that dOA,, A,, A31 = t 44, A2 9 A31 for every scalar t .
(b) Additivity in each row.
For example, additivity in the second row states that 441, 4 + C, ~43) = 44, A,, A31 + d(A,, C, A31 for every vector C.(c) The scalar triple product is zero if two of the rows are equal.(d) Normalization: d(i,j, k) = 1, w h e r e i=(l,O,O), j=(O,l,O), k=(O,O,l). A set of axioms for a determinant function 73 Each of these properties can be easily verified from
properties of the dot and crossproduct. Some of these properties are suggested by the geometric relation between the scalartriple product and the volume of the parallelepiped determined by the geometric vectors A,, A,, A:,. The geometric meaning of the additive prc’perty (b) in a special case is ofparticular interest. If we take C = A, in (b) the second
term on the right is zero because of(c), and relation (b) becomes (3.3) 44, A, + A,, 4) = d(A , A,, A&. This property is illustrated in Figure 3.1 which shows a parallelepiped determined byA,, A,, A,, and another parallelepiped determined by A,, A, + AZ, A,. Equation (3.3)merely states that these two parallelepipeds have equal volumes. This is evident
geometri-cally because the parallelepipeds have equal altitudes and bases of equal area. Volume = d(A , , A 2r A 3) Volume =: d(A ,, A, + A,, AZ) FIGURE 3.1 Geometric interpretation of the property d(A,, A,, AS) = d(A,, A, + A,, A&.
Thetwo parallelepipeds have equal volumes. 3.3 A set of axioms for a determinant function The properties of the scalar triple product mentioned in the foregoing section can besuitably generalized and used as axioms for determinants of order n. If A = (a,J is ann x n matrix with real or complex entries, we denote its rows by A,, . . . , A,, . Thus, theith
row of .4 is a vector in n-space given by Ai = (ail, ai2, . . . , a,,J. We regard the determinant as a function of the n rows A,, . . . , A,, and denote its values by d(A,,..., A,) or by det A. AXIOMATIC DEFINITION OF A DETERMINANT FUNCTION. A real- or complex-valuedfunctiond, dejinedfor each ordered n-tuple of vectors A,, . . . , A, ir;l n-space, is called a
determinantfunction of order n if it satisjes the following axioms for al’1 choices of vectors A,, . . . , A,and C in n-space: 7 4 Determinants AXIOM 1. HOMOGENEITY IN EACH ROW. If the kth row A, is multiplied by a scalar t,then the determinant is also multiplied by t: d(. . . , tA,, .
. .) = td(. .
. , A,, . . .). A X I O M 2 . A D D I T I V I T Y I N E A C H R O W . For each k we have d(A, ,..., A,+C ,..., A,)=d(A, ,..., A, ,..., A,)+&(A, ,..., C ,..., A,).
AXIOM 3. THE DETERMINANT VANISHES IF ANY TWO ROWS ARE EQUAL: d(A,,...,A,)=O if Ai = Ai for some i and j with i # j. AXIOM 4. THE DETERMINANT OF THE IDENTITY MATRIX IS EQUAL TO 1: d(I,, . . . , Z,) = 1 , where I, is the kth unit coordinate vector. The first two axioms state that the determinant of a matrix is a linear function of each
ofits rows. This is often described by saying that the determinant is a multilinear function ofits rows. By repeated application of linearity in the first row we can write &,C,,A,,...,A,k=l =~~hWwb,. . . ,A,), where tl, .
. . , t, are scalars and C,, . . . , C, are any vectors in n-space.Sometimes a weaker version of Axiom 3 is used: AXIOM 3’. THE DETERMINANT VANISHES IF TWO ADJACENT ROWS ARE EQUAL: d(A,, . . . , A,) = 0 if Ak = A,+, forsomek=1,2 ,..., n - l . It is a remarkable fact that for a given n there is one and only one function d whichsatisfies Axioms 1, 2,
3’ and 4. The proof of this fact, one of the principal results of thischapter, will be given later. The next theorem gives properties of determinants deducedfrom Axioms 1,2, and 3’ alone. One of these properties is Axiom 3. It should be noted thatAxiom 4 is not used in the proof of this theorem. This observation will be useful later whenwe prove
uniqueness of the determinant function. THEOREM 3.1. A determinant function satisfying Axioms 1, 2, and 3’ has the followingfurther properties: (a) The determinant vanishes if some row is 0: d(A,, . . .,A,)=0 if A,=0 for some k . A set of axioms for a determinant function 7 5 (b) The determinant changes sign if two adjacent rows are interchanged: 4..
. , A,, &+I,. . .> = -d(. . .
, AB+:,, A,, . . .). (c) The determinant changes sign if any two rows Ai and A) with i # j are interchanged.(d) The determinant vanishes if any two rows are equal: d(A,., . . .,A,)=0 i f Ai = Aj for some i and j with i # j.
(e) The determinant vanishes $ its rows are dependent.
Proof. To prove (a) we simply take t = 0 in Axiom 1. To prove (b), let B be a matrixhaving the same rows as A except for row k and row k + 11. Let both rows Bk and BKflbe equal to /I, + A,,, . Then det B = 0 by Axiom 3’. Thus we may write d(. . . , Ak+Ak+l,Ak+Ak+l,...)=O. Applying the additive property to row k and to row k + 1 we can rewrite this
equation asfollows : d(. .
. . A,,& ,...)+d(...,A,A,+, ,...)+d(. . .
. &+I,& ,...) +d(.
. . . Ak+l,Ak+l,...)=O.
The first and fourth terms are zero by Axiom 3’. Hence the second and third terms arenegatives of each other, which proves (b). To prove (c) we can assume that i < j . We can interchange rows Ai and Aj by performingan odd number of interchanges of adjacent rows. First we interchange row A, successivelywith the earlier adjacent rows Aj+l, Ajpz,.
. . , Ai. This requires j - i interchanges.Then we interchan,ge row Ai successively with the later adjacent rows Ai+l, Ai+z, . . . , Ajpl.This requires j - i - I further interchanges. Each interchange of adjacent rows reversesthe sign of the determinant. Since there are (j - i) + (j -- i - 1) = 2(j - i) - 1 inter-changes altogether (an odd number), the determinant
changes, sign an odd number of times,which proves (c). To prove (d), let B be the matrix obtained from A by interchanging rows Ai and Ad.Since Ai = Aj we have B = A and hence det B = det A. But by (c), det B = -det A.Therefore det A = 0. To prove (e) suppose scalars cr, . . . , c, exist, not all zero, such that I;=:=, ckA, = 0.Then any row A, with cK # 0
can be expressed as a linear combination of the other rows.For simplicity, suppose that A, is a linear combination of the others, say A, = zE=, t,A, .By linearity of the first row we have d(A,,A,, t.. , &Ak,A2,...,Ank=2 =kt2ft 44x., A,, . . . , A,). But each term d(A,, A,, . .
. , A,) in the last sum is zero since A, is equal to at least one ofA A2,“‘, n* Hence the whole sum is zero. If row A, is a linear combination of the otherrows we argue the same way, using linearity in the ith row. This proves (e). 7 6 3.4 Computation of determinants Determinants At this stage it may be instructive to compute some determinants, using only
the axiomsand the properties in Theorem 3.1, assuming throughout that determinant functions exist.In each of the following examples we do not use Axiom 4 until the very last step in thecomputation. EXAMPLE 1. Determinant of a 2 x 2 matrix. We shall prove that (3.4) detall al2[1 = a11a22 - a12a21.azl a22 Write the row vectors as linear
combinations of the unit coordinate vectors i = (1,O) andj = (0, 1): Al = (all, al,) = alli + ad, A2 = (a,, , a22) = a,+ + a22 j .Using linearity in the first row we have WI, A2) = d(alli + a12j, A,) = a,, d(i, A,) + a,, d(j, A,). Now we use linearity in the second row to obtain d(i, A,) = d(i, a,,i + a2zj) = u21 d(i, i) + a22 d(i,j) = a22 d(i,j), since d(i, i) = 0. Similarly
we find d(j, A,) = d(j, a,,i + az2j) = a21 d(j, 9 = -a21 dG,j). Hence we obtain 44, A,) = Gw22 - a12a2d 4&j) . But d(i,j) = 1 by Axiom 4, so d(A,, A,) = a11u22 - ur2u2r, as asserted.This argument shows that if a determinant function exists for 2 x 2 matrices, then it must necessarily have the form (3.4). Conversely, it is easy to verify that this formula
does,indeed, define a determinant function of order 2.
Therefore we have shown that there is oneand only one determinant function of order 2. EXAMPLE 2. Determinant of a diagonal matrix. A square matrix of the form a11 0 . . . 0 0 a22 0 A=' * *I I. . . . _ 0 0 - * - arm Computation of determinants 77 is called a diagonal matrix. Each entry ai, off the main diag,onal (i #j) is zero. We shallprove that the
det.erminant of A is equal to the product of its diagonal elements, (3.5) det A = a11az2 * * - arm , The kth row of .A is simply a scalar multiple of the kth unit ‘coordinate vector, A, = a,,Z, .Applying the homogeneity property repeatedly to factor out the scalars one at a time we get det A = d(Al, . . . , A,) = %A, . . . , a,,Z,J = all * * * arm d(Z,, . . . , I,). This
formula can be written in the form det A = a,, - - * a,,,, det I, where Z is the identity matrix.
Axiom 4 tells us that det Z = 1 so we obtain (3.5). EXAMPLE 3. Determinant of an upper triangular matrix. A square matrix of the form is called an upper triangular matrix. All the entries below the main diagonal are zero. Weshall prove that the determinant of such a matrix is equal to the product of its diagonalelements, det U = ullu22 * * * u,, . First
we prove that det U = 0 if some diagonal element Q = 0. If the last diagonalelement u nn is zero., then the last row is 0 and det U = 0 by Theorem 3.1 (a). Suppose, then,that some earlier diagonal element uii is zero. To be specific, say uz2 = 0. Then each ofthe n - 1 row-vectors U,, . . . , U, has its first two components zero. Hence these vectorsspan a
subspace of dimension at most n - 2. Therefore these n - 1 rows (and hence allthe rows) are dependent. By Theorem 3.1(e), det U = 0. In the same way we find thatdet U = 0 if any diagonal element is zero. Now we treat the general case. First we write the first row UL as a sum of two row-vectors, where V, = [ull, 0, . .
. , 0] and Vi = [O, Q, u,,, . .
.
, uln]. By linearity in the first rowwe have det. U = det (V,, U,, . . . , U,) + det (Vi, U2,. . . , U,).
78 Determinants But det (Vi, U,, . . . , U,) = 0 since this is the determinant of an upper triangular matrixwith a diagonal element equal to 0. Hence we have (3.6) detU=det(T/,,U, ,..., U,,). Now we treat row-vector U, in a similar way, expressing it as a sum, where v2 = LO, u22,0, * * .
3 01 and v; = [o, O, u23, . . . , UZn] * We use this on the right of (3.6) and apply linearity in the second row to obtain (3.7) det U = det (V,, V,, U,, . , . , U,,), since det (V,, Vi, U3, . . .
, U,) = 0.
Repeating the argument on each of the succeedingrows in the right member of (3.7) we finally obtain detU=det(VI,V, ,..., V,), where (V,, V,, . . .
, V,) is a diagonal matrix with the same diagonal elements as U. There-fore, by Example 2 we have det U = ullu22 . + * u,,, as required. EXAMPLE 4. Computation by the Gauss-Jordan process.
The Gauss-Jordan eliminationprocess for solving systems of linear equations is also one of the best methods for com-puting determinants. We recall that the method consists in applying three types of operationsto the rows of a matrix: (1) Interchanging two rows.(2) Multiplying all the elements in a row by a nonzero scalar.(3) Adding to one row a
scalar multiple of another. By performing these operations over and over again in a systematic fashion we can transformany square matrix A to an upper triangular matrix U whose determinant we now knowhow to compute. It is easy to determine the relation between det A and det U. Each timeoperation (1) is performed the determinant changes
sign. Each time (2) is performed witha scalar c # 0, the determinant is multiplied by c. Each time (3) is performed the deter-minant is unaltered. Therefore, if operation (1) is performedp times and if c1 , . . . , C~ arethe nonzero scalar multipliers used in connection with operation (2), then we have (3.8) det A = (- 1)P(c,c2 * * * c,)-l det U. Again we note
that this formula is a consequence of the first three axioms alone. Its proofdoes not depend on Axiom 4. 3.5 The uniqueness theorem Exercises 7 9 In Example 3 of the foregoing section we showed that Axioms 1, 2, and 3 imply theformula det U == ulluz2 * * . u,, det I. Combining this with (3.8) we see that for every n x nmatrix A there is a scalar c
(depending on A) such that (3.9) d(A,) . . .) A,) = c d(Z,) . . .) I,). Moreover, this formula is a consequence of Axioms 1, 2, and 3 alone. From this we caneasily prove that there cannot be more than one determinant function. THEOREM 3.2, UNIQUENESS THEOREM FOR DETERMINANTS. Let d be afunction satisfyingall four axioms for a determinant
function of order n, and let f be another function satisfyingAxioms 1, 2, and 3. Then for every choice of vectors A,, . . . , A, in n-space we have (3.10) .f(A,>. . . , A,) = 44,. . . , A,)f(Z,, . . . , Z,). In particular, ifJf also satisjies Axiom 4 we have f (A,, , .
. , A,,) = d(A, , . . . , A,,). Proof.
Let g(A,, . . . , A,) = f(A,, . . . , A,) - d(A,, . . . , A,)f(Z,, .
. . , I,). We willprove that g(A,, . . . , A,,) = 0 for every choice of A,, . . . , A,. Since both d and f satisfyAxioms 1,2, and 3 the same is true of g. Hence g also satisfies Equation (3.9) since this wasdeduced from the first three axioms alone.
Therefore we can write (3.11) g(4,. . . , A,) = c&Z,, . . . , fn), where c is a sca1a.r depending on A. Taking A = Z in the dlefinition of g and noting that dsatisfies Axiom 4 we find gu,, * * * , L> =f(4, * * * , Z,) -f(Z,, *. * , Z,) = 0. Therefore Equation (3.11) becomes g(A,, . .
. , A,) = 0. This completes the proof. 3.6 Exercises In this set of exercises you may assume existence of a determinant function. Determinants oforder 3 may be computed by Equation (3.2). 1. Compute each of the following determinants 2 1 1 3 0 8 (4 1 4 -4, (b) 5 0 7 1 0 2 - 1 4 2 a 1 0 , (c) 2 a 2 0 1 a 80 Determinants = 1, compute the determinant of
each of the following matrices: 3. (a) Prove that a b c = (b - a)(c - a)(c - b). a2 b2 c2 (b) Find corresponding formulas for the determinants 1 1 1 a b c and a3 b3 c3 1 1 1 a2 b2 c2 a3 b3 c3 4 . Compute the determinant of each of the following matrices by transforming each of them to anupper triangular matrix.
I1 a a2 u3 1 1 1 b c d 1 b2 c2 d2 ’(4 b3 c3 d3 -1 1 1 1 a b c d a2 b2 c2 1 d2 ’ -a4 b4 c4 d4 :: 11 - 1- 1 1 11 1 11 - 1- 1 -1 -1-1 -1 Ca)Ca) 11 11 -1 -1-1 -1 11 ’’ (b)(b) 11 11 1 - l1 - l (4 - a 1 0 0 o- 4 a 2 0 0 0 3 a 3 0 0 0 2 a 4 -0 0 0 1 a - , Cd ‘ 1 1 1 1 1 1 1 1 -1 - 1 - 1 1 1 -1 -1 1 1 1 - 1 -1 1 - 1 1 1 - 1 l - l 1 1 !* 1 1 - 1 -1 1 1 - 1 5. A lower triangular matrix A = (Q)
is a square matrix with all entries above the main diagonalequal to 0; that is, aij = 0 whenever i . . , A,)d(Z,B, . .
. , Z,B). We keep B fixed and introduce a function f defined by the formula f(A,, . . . , A,) = d(A,B, . . . , A,B). The equation we wish to prove states that (3.13) f(A,, , . .
,A,) = W,,..., A,)fU,, .
. . ,4>. Determinants and independence of vectors 83 But now it is a simple matter to verify that f satisfies Axioms 1, 2, and 3 for a determinantfunction so, by the uniqueness theorem, Equation (3.13) holds for every matrix A.
Thiscompletes the proof. Applications of the product formula are given in the next two sections. 3.8 The determinant of the inverse of a nonsingular matrix We recall that a.
square matrix A is called nonsingular if it has a left inverse B such thatBA = I. If a left inverse exists it is unique and is also a right inverse, AB = I. We denotethe inverse by A-l. The relation between det A and det A-’ is as natural as one could expect. THEOREM 3.5. If matrix A is nonsingular, then det A # 0 and we have (3.14) &t A-l = 1det A ’ Proqfi
From the product formula we have (det A)(det A-l) = det (AA-‘) = det .I = 1. Hence det A # 0 and (3.14) holds.
Theorem 3.5 shows that nonvanishing of det A is a necessary condition for A to be non-singular. Later we will prove that this condition is also sufficient. That is, if det A # 0then A--l exists. 3.9 Determinants and independence of vectors A simple criterion for testing independence of vectors can be deduced from Theorem 3.5. THEOREM 3.6. A set of n
vectors A,, . . . , A, in n-space is independent if and only ifW, , .
. . , A,) # 0. Proof. We already proved in Theorem 3.2(e) that dependence implies d(A, , . . . , A,) =0. To prove the converse, we assume that A,, . . .
, A, are independent and prove thatd(A,, . . . , A,) # 0. Let Vn denote the linear space of n-tuples of scalars. Since A’, , .
. .
, A, are n independentelements in an n-dimensional space they form a basis for V, . By Theorem 2.12 there is alinear transformation T: V, + V, which maps these n vectors onto the unit coordinatevectors, TV,) = Ire for k=l,2 ,..., n.
Therefore there is an n x n matrix B such that A,B = I, for k=l,2 ,..., n. 8 4 Determinants But by Lemma 3.3 we have A,B = (AB), , where A is the matrix with rows A,, . . . , A,.Hence AB = I, so A is nonsingular and det A # 0. 3.10 The determinant of a block-diagonal matrix A square matrix C of the formA 0 c= 0 B’[1 where A and B are square matrices
and each 0 denotes a matrix of zeros, is called a block-diagonal matrix with two diagonal blocks A and B. An example is the 5 x 5 matrix 0 0 4 5 6 The diagonal blocks in this case are1 0 A= [1 and 0 1 The next theorem shows that the determinant of a block-diagonal matrix is equal to theproduct of the determinants of its diagonal blocks. THEOREM
3.7. For any two square matrices A and B we have (3.15) det = (det A)(det B). Proof. AssumeAisnxnandBismxm. We note that the given block-diagonalmatrix can be expressed as a product of the form where I,, and I,,, are identity matrices of orders n and m, respectively. Therefore, by theproduct formula for determinants we have (3.16) det[t I]=det[t
ZI]det[z Og]. --- Exercises 85 A 0Now we regard the determinant det [I as a function of the n rows of A. This is 0 Anpossible because of the block of zeros in the upper right-hand corner. It is easily verifiedthat this function satisfies all four axioms for a determinant function of order n.
Therefore,by the uniqueness theorem, we must have det = det A.
A similar argument shows that det 3.11 Exercises = det B. Hence (3.16) implies (3.15). 1. For each of the rfollowing statements about square matrices, give a proof or exhibit a counterexample.(a) det(A +B) =detA +detB.(b) det {(A + B)12} = {det (A + B)}2(c) det {(A + B)2} = det (A2 + 2AB + B2)(d) det {(A + B)2} = det (A2 + B2). 2. (a) Extend
‘Theorem 3.7 to block-diagonal matrices with three diagonal blocks: = (det A)(det B)(det C). (b) State and prove a generalization for block-diagonal matrices with any number of diagonalblocks.
3.L,etA=[i! i, B=i i 21. ProvethatdetA=det[i t]andthat a $bdet B = det [1e .f ’ 4. State and prove a generalization of Exercise 3 for n x n matrices.a b 0 0 c d 0 05,LetA= I 1 efgh’ProvethatdetA=det[I z]detK L].
x y .z w6. State and prove a generalization of Exercise 5 for II x n matrices of the form A = where B, C, D denote square matrices and 0 denotes a matrix of zeros. 8 6 Determinants 7. Use Theorem 3.6 to determine whether the following sets of vectors are linearly dependent orindependent.(a) A, = (1, -l,O), A, = (O,l, -l), A, = (2,3, -1).(b)A,=
(1,-1,2,1),A,=(-1,2,-1,O),Ag=(3,-1,1,0),A~=(1,0,0,1). Cc) A,=(1,0,0,0,1),A,=(1,1,0,0,O),As=(1,0,1,0,1),Aq=(1,1,0,1,1),A, = (O,l,O, 1,O). 3.12 Expansion formulas for determinants. Minors and cofactors We still have not shown that a determinant function actually exists, except in the 2 x 2case. In this section we exploit the linearity property and the
uniqueness theorem to showthat if determinants exist they can be computed by a formula which expresses every deter-minant of order n as a linear combination of determinants of order n - 1 . Equation (3.2)in Section 3.1 is an example of this formula in the 3 x 3 case. The general formula willsuggest a method for proving existence of determinant
functions by induction. Every row of an n x n matrix A can be expressed as a linear combination of the n unitcoordinate vectors II, . . . , Z, . For example, the first row A, can be written as follows: A, = $a,J, .61 Since determinants are linear in the first row we have ia,$,,A ,..., A,j=l =$;u d(l,, AZ,. . . , A,). Therefore to compute det A it suffices to compute
d(Z,, A,, . . . , A,) for each unit coordinatevector Zj . Let us use the notation ,4ij to denote the matrix obtained from A by replacing the firstrow A, by the unit vector Zj . For example, if IZ = 3 there are three such ma&ices: Note that det Aij = d(Zj, AZ, . . . , A,). Equation (3.17) can now be written in the form (3.18) det A = i all det Ali.j=l This is called an
expansion formula; it expresses the determinant of A as a linear combina-tion of the elements in its first row. The argument used to derive (3.18) can be applied to the kth row instead of the first row.The result is an expansion formula in terms of elements of the kth row. Expansion formulas for determinants. Minors and cofactors 8 7 THEOREM 3.8.
EXPANSION BY COFACTORS. Let Akj denote the matrix obtained from Aby replacing the kth row A, by the unit coordinate vector Ij. Then we have the expansionformula (3.19) det A = takj det Akjj,l which expresses the determinant of A as a linear combination of the elements of the kth row.The number det Akj is called the cofactor of entry akj.
In the next theorem we shall prove that each cofactor is, except for a plus or minus sign,equal to a determinant of a matrix of order n - 1 . These smaller matrices are calledminors. DEFINITION. Given a square matrix A of order n 2 2, the square matrix of order n - 1obtained by deleting the kth row and the jth column of A is called the k, j minor of A
and isdenoted by A,. . EXAMPLE .
A matrix A = (alei) of order 3 has nine minors. Three of them are &I=[;;: ;;;]y Alz=[;:: ;;l]r &a=[;;; ;:]- Equation (3.2) expresses the determinant of a 3 x 3 matrix as a linear combination ofdeterminants of these three minors. The formula can be written as follows: det A = a,, det A,, - aI det A,, + aI3 det A,, .
The next theorem extends this formula to the n x n case for any n 2 2. THEOREM 3.9. EXPANSION BY kTH-ROW MINORS. For any n x n matrix A, n > 2, thecofactor of ski is related to the minor Aki by the formula (3.20) det ALj = (- l)kfj det Aki.
Therefore the expansion of det A in terms of elements of the kth row is given by (3.21) det A = $(- l)“+jakj det A,$.j=l Proof. We illustrate the idea of the proof by considering first the special case k = j = 1 . The matrix ,4& has the form A;, = 1 0 0 *a* o- azl az2 az3 *** azn a31 a32 a33 * *a a3n .a,, an2 an3 * * * an+ 88 Determinants- By applying
elementary row operations of type (3) we can make every entry below the 1 inthe first column equal to zero, leaving all the remaining entries intact. For example, if wemultiply the first row of Ai, by -azl and add the result to the second row, the new secondrow becomes (0, az2, az3, . . . , a2J. By a succession of such elementary row operationswe obtain
a new matrix which we shall denote by A& and which has the form 0a 2 2 a32 Since row operations of type (3) leave the determinant unchanged we have (3.22) det A!, = det A;, . But A,O, is a block-diagonal matrix so, by Theorem 3.7, we have det A!, = det AlI, where A,, is the 1, 1 minor of A, Therefore det Ai, = det AlI, which proves (3.20) for k = j =
1.We consider next the special case k = 1 ,i arbitrary, and prove that (3.23) det Aij = (- l)j-’ det A,$. Once we prove (3.23) the more general formula (3.20) follows at once because matrix AJ,can be transformed to a matrix of the form Bij by k - 1 successive interchanges of adjacentrows. The determinant changes sign at each interchange so we have
(3.24) det Akj = (- l)“-1 det B&, where B is an n x n matrix whose first row is Ij and whose 1 ,i minor BIj is Akj. By (3.23),we have det Bij = (- l)‘-l det Blj = (- l)j-’ det Aki,so (3.24) gives us det ALj = (- l)“-l(- l)j-’ det Akj = (- l)“+j det Akj. Therefore if we prove (3.23) we also prove (3.20). Expansion formulas for determinants. Minors and cofactors 89 We
turn now to the proof of (3.23). The matrix Aij has the form 0 . . . 1 . . . 0azl . . . azj ..’ I* 1 azn Aii = : a nl ... anj a** a 11 . . nn By elementar,y row operations of type (3) we introduce a column of zeros below the 1 andtransform this to 0 . . . 0 1 0 .‘* 0azl . . . a2,j-1 0 a2,j+l * ’ . aPn Ayi = : . * -an1 . * * a,,j-l 0 a,,j+l . * . h- As before, the determinant is
unchanged so det Ayj = det Aij.
The 1, j minor AIj has the 1form a2n Now we regard det AL as a function of the n - 1 rows of AIj, say det AL = f(A,J . Thefunctionfsatisfies the first three axioms for a determinant function of order n - 1. There-fore, by the uniqueness theorem we can write (3.25) f(Alj) = f(J) det Alj 3 where J is the identity matrix of order n - 1. Therefore, to prove (3.23)
we must show thatf(J) = (- l)i-l. Now f(J) is by definition the determinant of the matrix -0 . . . 0 1 0 *** o- 1 0 0 0 0 c= :\. . ..
. . \... . 0 . . . 1 0 0 *** 0 0 . . . 0 0 1 *.* 0.
. .\. . . . . . \:0 . . . 0 0 0 *a- 1 t -jth column c jth row 90 Determinants The entries along the sloping lines are all 1. The remaining entries not shown are all 0.By interchanging the first row of C successively with rows 2, 3, . . . ,j we arrive at the n x nidentity matrix I after j - 1 interchanges. The determinant changes sign at each inter-change, so det C = (-
l)j-‘.
Hencef(J) = (- l)j-‘, which proves (3.23) and hence (3.20). 3.13 Existence of the determinant function In this section we use induction on n, the size of a matrix, to prove that determinantfunctions of every order exist. For n = 2 we have already shown that a determinant functionexists. We can also dispense with the case n = 1 by defining det [aa,,] = a,,
. Assuming a determinant function exists of order n - 1, a logical candidate for a deter-minant function of order n would be one of the expansion formulas of Theorem 3.9, forexample, the expansion in terms of the first-row minors.
However, it is easier to verify theaxioms if we use a different but analogous formula expressed in terms of the first-columnminors. THEOREM 3.10. Assume determinants of order n - 1 exist. For any n X n matrixA = (aik) , let f be the function dejned by the formula (3.26) f(A,, . .
. , A,) = i(- l)j+‘aj, det Aj,.j=l Then f satisjes allfour axiomsfor a determinantfunction of order n. Therefore, by induction,determinants of order n exist for every n. Proof. We regard each term of the sum in (3.26) as a function of the rows of A and wewrite .fW,,.. . , A,) = (- l)jflaj, det Aj, . If we verify that eachf, satisfies Axioms 1 and 2 the same will be
true forf.Consider the effect of multiplying the first row of A by a scalar t. The minor A,, is not affected since it does not involve the first row. The coefficient a,, is multiplied by t, so wehave f,(tA, 9 A,z 3 . . . > A,) = tall det A,, = tf,(A,, . . . , A,). Ifj > 1 the first row of each minor Aj, gets multiplied by t and the coefficient ajI is notaffected, so again we
have &@A,, A,, .
. .
, A,) = tY&h, A,, .
.
. , A,). Therefore each fj is homogeneous in the first row.If the kth row of A is multiplied by t, where k > 1, the minor A,, is not affected but akl is multiplied by t, so fk is homogeneous in the kth row. If j # k, the coefficient aj, is notaffected but some row of Aj, gets multiplied by t. Hence every J; is homogeneous in thekth row. The determinant of a

transpose 91 A similar argument shows that each fi is additive in every row, so f satisfies Axioms 1 and2 . We prove next that f satisfies Axiom 3’, the weak version of Axiom 3. From Theorem3.1, it then follows that f satisfies Axiom 3. To verify that. f satisfies Axiom 3’, assume two adjacent rows of A are equal, say A, =A k+l' Then, except for minors
Akl and A,+, 1, each mi,nor A), has two equal rows sodet .4,, = 0.
Therefore the sum in (3.26) consists only of the two terms corresponding toJ’=kand,i=k+ 1, (3.27) f(A,>..., An) = (-l)k+lakl det A,, + (-1)“i-2ak+,,, det A,,, , , . But A,, = A,,,,, and akl = a,,,,, since A, = A,,, . Therefore the two terms in (3.27)differ only in sign, so f (A,, . . .
, A,) = 0. Thus, f satisfies Axiom 3’. Finally, we verify that f satisfies Axiom 4. When A = 1 we have a,, = 1 and a,, = 0 fori > 1. Also, A,, is the identity matrix of order n - 1, so each term in (3.26) is zero exceptthe first, which is equal to 1. Hence f (f, , . . . , 1,) = 1 sgof satisfies Axiom 4. In the foregoing proof we could just as well have used a.
functionfdefined in terms of thekth-column minors Aik instead of the first-column minors Aj, . In fact, if we let (3.28) f(Al,..., A,) = 2 (- l)ii-kajl, det Ajk ,j=l exactly the same type of proof shows that this f satisfies all four axioms for a determinantfunction. Since determinant functions are unique, the expansion formulas in (3.28) andthose in (3.21) are all
equal to det A. The expansion formulas (3.28) not only establish the existence of determinant functionsbut also reveal a new aspect of the theory of determinants-a connection between row-properties and column-properties. This connection is discussed further in the next section. 3.14 The determinant of a transpose Assoc:iated with each matrix A is
another matrix called the transpose of A and denotedby At. The rows of At are the columns of A. For example, if A = A formal definition may be given as follows. DEFI’VITION OF TRANSPOSE. The transpose of an m X: n matrix A = (a&~& is the n X mmatrix At whose i, j entry is aji . Although transposition can be applied to any rectangular matrix we
shall be concernedprimarily with square matrices. We prove next that transposition of a square matrix doesnot alter its determinant, 9 2 Determinants THEOREM 3. Il. For any n x n matrix A we have det A = det At. Proof. The proof is by induction on n. For n = 1 and n = 2 the result is easily verified.Assume, then, that the theorem is true for matrices
of order n - 1 . Let A = (aij) and letB = At = (6,). Expanding det A by its first-column minors and det B by its first-rowminors we have det A = &-l)‘flaj, det Ajl, det B = i (- l)‘+‘blj det Blj.kl i=l But from the definition of transpose we have bli = a,, and Blj = (AJ”. Since we areassuming the theorem is true for matrices of order n - 1 we have det Blj = det
Aj,.Hence the foregoing sums are equal term by term, so det A = det B. 3.15 The cofactor matrix Theorem 3.5 showed that if A is nonsingular then det A # 0. The next theorem provesthe converse. That is, if det A # 0, then A-l exists. Moreover, it gives an explicit formulafor expressing A-l in terms of a matrix formed from the cofactors of the entries of
A. In Theorem 3.9 we proved that the i, j cofactor of aij is equal to (- l)i+i det Aii , whereAii is the i, j minor of A. Let us denote this cofactor by cof aij. Thus, by definition, cof aij = (- l)ifj det Aij . DEFINITION OF THE COFACTOR MATRIX. The matrix whose i, j entry is cof aii is called thecofactor matrix? of A and is denoted by cof A. Thus, we have cof A
= (cof aij)& = ((- l)i+i det Ai,)tj=, . The next theorem shows that the product of A with the transpose of its cofactor matrix is,apart from a scalar factor, the identity matrix I. THEOREM 3.12. For any n x n matrix A with n 2 2 we have (3.29) A(cof A)t = (det A)Z. In particular, if det A # 0 the inverse of A exists and is given by A-l = bA (cof A)t. t In much
of the matrix literature the transpose of the cofactor matrix is called the adjugate of A. Some ofthe older literature calls it the adjoint of A. However, current nomenclature reserves the term adjoint foran entirely different object, discussed in Section 5.8. Cramer.? rule 93 Proof. Using Theorem 3.9 we express det A in terms of its kth-row cofactors by
theformula (3.30) det A = i akj cof akj.j=l Keep k fixed and apply this relation to a new matrix B whose ith row is equal to the kthrow of A for some i # k, and whose remaining rows are the same as those of A. Thendet B = 0 because the ith and kth rows of B are equal. Expressing det B in terms of itsith-row cofactors we have (3.31) det B = 5 bij cof bij
= 0.j=l But since the ith row of B is equal to the kth row of A wc have bii = akj and cof bij = cof aij for every j . Hence (3.31) states that (3.32) if k + i. Equations ((3.30) and (3.32) together can be written as follows: (3.33)12cakj cof aij = det A i f i = k0j=l i f i#k. But the sum appearing on the left of (3.33) is the k, i entry of the product A(cof #. There-
fore (3.33) implies (3.29). As a direct corollary of Theorems 3.5 and 3.12 we have the following necessary andsufficient condition for a square matrix to be nonsingular. THEOREM 3.13. A square matrix A is nonsingular if and CRAMER'S RULE. If a system of n &near equations in n unknowns 94 Determinants has a nonsingular coeficient-matrix A =
(aij) , then there is a unique solution for the system&en by the formulas (3.34)1 n xj =det k=lcb, cof akj , for j=l,2 ,..., n. Proof. The system can be written as a matrix equation, ’ AX=B, where X and B are column matrices, X = Since A is nonsingular there is a unique solution X given by (3.35) X = A-‘B = bA (cof A)tB. The formulas in (3.34) follow by
equating components in (3.35). It should be noted that the formula for xj in (3.34) can be expressed as the quotient of twodeterminants, det C.x.=2,Idet A where Cj is the matrix obtained from A by replacing the jth column of A by the columnmatrix B. 3.17 Exercises 1. Determine the cofactor matrix of each of the following matrices : 2. Determine the
inverse of each of the nonsingular matrices in Exercise 1.3. Find all values of the scalar A for which the matrix II - A is singular, if A is equal to Exercises 9 5 4. If A is an n x n matrix with n 2 2, prove each of the following, properties of its cofactor matrix:(a) cof (At) = (cof #. (b) (cof #A = (det A)Z.(c) A(cof ~1)~ = (cof #A (A commutes with the
transpose of its cofactor matrix).
5. Use Cramer’s rule to solve each of the following systems:(a) x + 5 + 3.z = 8, 2 x - y +4z =7, -y + z = 1.(b) x +y +2z =O, 3x-y-z=3, 2x +5y +3z =4. 6. (a) Explain why each of the following is a Cartesian equation for a straight line in the xy-planepassing thtough two distinct points (x1, yl) and (x2, yz). (b) State and prove corresponding relations for a
plane in 3-space passing through threedistinct points.(c) State alnd prove corresponding relations for a circle in the xy-plane passing through threenoncolinear points. 7. Given n2 functions fij, each differentiable on an interval (a, i5), define F(x) = det [j&x)] foreach x in (a, b).
Prove that the derivative F’(x) is a sum of n determinants, F’(x) = 2 det A,(x),i=l where Ai is the matrix obtained by differentiating the fun’ctions in the ith row of [f&(x)]. 8. An n x n matrix of functions of the form W(x) = [uji-l’(x)], in which each row after the firstis the derivative Iof the previous row, is called a Wronskian matrix in honor of the Polish
mathe-matician J. M.
II. Wronski (1778-1853). Prove that the derivative of the determinant of W(x)is the determinant of the matrix obtained by differentiating each entry in the last row of W(X). [Hint: Use Exercise 7.1 4EIGENVALUES AND EIGENVECTORS 4.1 Linear transformations with diagonal matrix representations Let T: V+ V be a linear transformation on a finite-
dimensional linear space V. Thoseproperties of T which are independent of any coordinate system (basis) for V are calledintrinsicproperties of T. They are shared by all the matrix representations of T. If a basiscan be chosen so that the resulting matrix has a particularly simple form it may be possibleto detect some of the intrinsic properties directly
from the matrix representation. Among the simplest types of matrices are the diagonal matrices. Therefore we might askwhether every linear transformation has a diagonal matrix representation. In Chapter 2we treated the problem of finding a diagonal matrix representation of a linear transfor-mation T: V+ W, where dim V = n and dim W = m . In
Theorem 2.14 we proved thatthere always exists a basis (el, . . . , e,) for V and a basis (M.‘, , . . . , w,) for W such that thematrix of T relative to this pair of bases is a diagonal matrix. In particular, if W = Vthe matrix will be a square diagonal matrix. The new feature now is that we want to use thesame basis for both Vand W. With this restriction it is not
always possible to find a diagonalmatrix representation for T. We turn, then, to the problem of determining which trans-formations do have a diagonal matrix representation. Notation: If A = (aij) is a diagonal matrix, we write A = diag (all, az2, . . . , a,,). It is easy to give a necessary and sufficient condition for a linear transformation to have adiagonal
matrix representation. THEOREM 4.1. Given a linear transformation T: V--f V, where dim V = n. If T has adiagonal matrix representation, then there exists an independent set of elements ul, . . . , u,in V and a corresponding set of scalars A,, . . . , A, such that (4.1) T&J = &curt f o r k=1,2 ,..., n. Conversely, i f there is an independent set ul, . . . , u, in V
and a corresponding set of scalarsA I,“‘, 1, satisfying (4.1), then the matrix A =diag(&,...,&) is a representation of T relative to the basis (uI , .
. . , u,). 96 Eigenvectors and eigenvalues of a linear transformation 97 Proof. Assume first that T has a diagonal matrix representation A = (ai& relative tosome basis (e,, . . . , e,).
The action of T on the basis elements is given by the formula T(eJ = i aikei = akBeki=l since aik = 01 for i Z k . This proves (4.1) with uk = ek and 1, = a,, .Now suppose independent elements ul, .
. . , U, and scalars &, . .
. , A, exist satisfying (4.1). Since zfl, . . . , u, are independent they form a basis for V. If we define a,, = 1, andaik = 0 for i # Ic, then the matrix A = (a& is a diagonal matrix which represents Trelative to the basis (ul, . . . , u,). Thus the probkm of finding a diagonal matrix representation of a linear transformationhas been transformed to another
problem, that of finding independent elements ul, . . . , U,and scalars ;I,, . . . , 1, to satisfy (4.1). Elements uk and scalars 1, satisfying (4.1) are calledeigenvectors and eigenvalues of T, respectively. In the next section we study eigenvectorsand eigenvaluesf in a more general setting. 4.2 Eigenvectors and eigenvalues of a linear transformation In this
discussion V denotes a linear space and S denotes a subspace of V. The spaces Sand V are not required to be finite dimensional. DEFINITION. Let T: S 4 V be a linear transformation of S into V. A scalar ;i is called aneigenvalue of‘ T if’there is a nonzero element x in S such that (4.2) T(x) = Ix. The element .x is called an eigenvector of Tbelonging to 1.
The scalar 1 is called an eigenvaluecorresponding to .x. There is exactly one eigenvalue corresponding to a given eigenvector x. In fact, if wehave T(x) = iix and T(x) = ,UX for some x # 0, then 1x q = yx so il = ,u .
Note: Although Equation (4.2) always holds for x = 0 and any scalar I, the definitionexcludes 0 as an eigenvector. One reason for this prejudice against 0 is to have exactly oneeigenvalue), associated with a given eigenvector x. The following examples illustrate the meaning of these concepts. EXAMPLE 1. Multiplication by a fixed scalar. Let T: S + V
be the linear transformationdefined by the equation T(x) = cx for each x in S, where c is a fixed scalar. In this exampleevery nonzero element of S is an eigenvector belonging to the scalar c. t The words eigenvector and eigenvalue are partial translations of the German words Egenvektor andEigenwert, respectively.
Some authors use the terms characteristic vector, or proper vector as synonyms foreigenvector. Eigenv,alues are also called characteristic values, proper values, or latent roots. 98 Eigenvalues and eigenvectors EXAMPLE 2. The eigenspace E(A) consisting of all x such that T(x) = Ax. Let T: S -+ Vbe a linear transformation having an eigenvalue 1. Let
E(A) be the set of all elements x inS such that T(x) = 1.~. This set contains the zero element 0 and all eigenvectors belongingto il. It is easy to prove that E(I) is a subspace of S, because if x and y are in E(1) we have T(aX + by) = aT(x) + bT(y) = ailx + bAy = A(ax + by) for all scalars a and b.
Hence (ax + by) E E(1) so E(A) is a subspace.
The space E(1) iscalled the eigenspace corresponding to 1.
It may be finite- or infinite-dimensional. If E(A)is finite-dimensional then dim E(I) 2 1 , since E(1)contains at least one nonzero element xcorresponding to 1. EXAMPLE 3. Existence of zero eigenvalues. If an eigenvector exists it cannot be zero, bydefinition. However, the zero scalar can be an eigenvalue. In fact, if 0 is an eigenvalue forx then T(x) = Ox
= 0, so x is in the null space of T. Conversely, if the null space of Tcontains any nonzero elements then each of these is an eigenvector with eigenvalue 0. Ingeneral, E(A) is the null space of T - AZ. EXAMPLE 4. Rejection in the xy-plane. Let S = V = V,(R) and let T be a reflection inthe xy-plane. That is, let Tact on the basis vectors i, j, k as follows: T(i) =
i, T(j) = j,T(k) = -k. Every nonzero vector in the xy-plane is an eigenvector with eigenvalue 1.The remaining eigenvectors are those of the form ck, where c # 0 ; each of them haseigenvalue - 1 . EXAMPLE 5. Rotation of theplane through ajxedangle u. This example is of special interestbecause it shows that the existence of eigenvectors may depend on
the underlying field ofscalars. The plane can be regarded as a linear space in two different ways: (1) As a 2-dimensional real linear space, V = V,(R), with two basis elements (1, 0) and (0, l), andwith real numbers as scalars; or (2) as a l-dimensional complex linear space, V = V,(C),with one basis element 1, and complex numbers as scalars. Consider
the second interpretation first. Each element z # 0 of V,(C) can be expressedin polar form, z = yei*. If T rotates z through an angle cc then T(z) = reicefa) = eiaz.Thus, each z # 0 is an eigenvector with eigenvalue il = eia. Note that the eigenvalue ei” isnot real unless c4 is an integer multiple of r. Now consider the plane as a real linear space, V,(R).
Since the scalars of V,(R) are realnumbers the rotation T has real eigenvalues only if tc is an integer multiple of n. In otherwords, if CI is not an integer multiple of v then T has no real eigenvalues and hence noeigenvectors. Thus the existence of eigenvectors and eigenvalues may depend on the choiceof scalars for V. EXAMPLE 6. The dzyerentiation
gperator. Let V be the linear space of all real functions fhaving derivatives of every order on a given open interval. Let D be the linear transfor-mation which maps each f onto its derivative, D(f) = f ‘ . The eigenvectors of D are thosenonzero functions f satisfying an equation of the form Eigenvectors and eigenvalues of a linear transformation 99 f’ = If
for some real 1. This is a first order linear differential equation.
All its solutions are givenby the formula f(x) = ce”“, where c is an arbitrary real constant. Therefore the eigenvectors of D are all exponentialfunctions f(x) == ceAx with c # 0 . The eigenvalue corresponding to f(x) = celZ is 1. Inexamples li’ke this one where V is a function space the eigenvectors are called eigenfunctions. EXAMPLE 7. The integration
operator.
Let V be the linear space of all real functionscontinuous on a finite interval [a, b]. Iffg V define g = T(f) to be that function given by g(x) = azf(t) dts i f aIn fact, if y = cx we Ihave T(y) = T(cx) = CT(X) = c(Ax) = l(cx) = Ay. If c # 0 then y # 0 so every nonzero element y of L(x) is also an eigenvector belongingto 1. A subspace U of S is called invariant
under Tif Tmaps each element of U onto an elementof U. We have j ust shown that the subspace spanned by an eigenvector is invariant under T. 100 Eigenvalues and eigenvectors 4.3 Linear independence of eigenvectors corresponding to distinct eigenvalues One of the most important properties of eigenvalues is described in the next theorem.
Asbefore, S denotes a subspace of a linear space V. THEOREM 4.2. Let ul, . . . , uk be eigenvectors of a linear transformation T: S + V, andassume that the corresponding eigenvalues A,, . . . , ii, are distinct. Then the eigenvectors %, * *. 9 uk are independent.
Proof. The proof is by induction on k. The result is trivial when k = 1 . Assume, then,that it has been proved for every set of k - 1 eigenvectors. Let ul, . . . , uk be k eigen-vectors belonging to distinct eigenvalues, and assume that scalars ci exist such that (4.4) iciui = O * Applying T to both members of (4.4) and using the fact that T(uJ = ii,u, we find
Multiplying (4.4) by 1., and subtracting from (4.5) we obtain the equation k-l isCi(A - ‘kj”i = 0 * But since ul, . . . , uk-l are independent we must have ci(ili - jlk) = 0 for each i = 1, 2, . . . ,k - 1 . Since the eigenvalues are distinct we have ai #)Lk for i # k so ci = 0 for i = 1, 2,. . .) k - 1 . From (4.4) we see that ck is also 0 so the eigenvectors ul, .
. .
, uk are inde-pendent.
Note that Theorem 4.2 would not be true if the zero element were allowed to be an eigen-vector. This is another reason for excluding 0 as an eigenvector. Warning: The converse of Theorem 4.2 does not hold. That is, if T has independenteigenvectors ul, . . . , uk, then the corresponding eigenvalues A,, . . .
, 2, need not be dis-tinct. For example, if T is the identity transformation, T(x) = x for all x, then everyx # 0 is an eigenvector but there is only one eigenvalue, ii = 1 .
Theorem 4.2. has important consequences for the finite-dimensional case. THEOREM 4.3. If dim V = n , every linear transformation T: V + V has at most n distincteigenvalues. If T has exactly n distinct eigenvalues, then the corresponding eigenvectors forma basis for V and the matrix of T relative to this basis is a diagonal matrix with the
eigenvaluesas diagonal entries. Proof. If there were n + 1 distinct eigenvalues then, by Theorem 4.2, V would containn + 1 independent elements. This is not possible since dim V = n . The second assertionfollows from Theorems 4.1 and 4.2. Exercises 101 Note: Theorem 4.3 tells us that the existence of n distinct eigenvalues is a for
Tto have a diagonal matrix representation. This condition is not necessary.There exist linear transformations with less than n distinct eigenvalues that can berepresented by diagonal matrices. The identity transformation is an example. All its eigen-values are eqiual to 1 but it can be represented by the identity matrix. Theorem 4.1tells us that the
existence of n independent eigenvectors is necessary and sufJient forT to have a diagonal matrix representation. 4.4 Exercises 1. (a) If T has an eigenvalue 1, prove that aT has the eigenvalue al.(b) If x is an eigenvector for both Tl and T, , prove that x is also an eigenvector for aT, + bT, .How are the eigenvalues related? 2. Assume ‘T: V-- Vhas an
eigenvector x belonging to an eigenvalue 1. Prove that xGs an eigen-vector of T2 belonging to A2 and, more generally, x is an eigenvector of Tn belonging to 2”.Then use the result of Exercise 1 to show that if P is a polynomial, then x is an eigenvector ofP(T) belongin,g to P(A). 3. Consider the plane as a real linear space, V = V,(R) , and let T be a
rotation of V through anangle of ~12 radians.
Although T has no eigenvectors, prove that every nonzero vector is aneigenvector for T2. 4. If T: V--c V has the property that T2 has a nonnegative eigenvalue A2, prove that at least oneof 3, or --3, is an eigenvalue for T. [Hint: T2 - A21 = (T + lZ)(T - AZ) .] 5. Let V be the linear space of all real functions differentiable on (0, 1). If f E V, define g =T(f) to
mean i:hatg(t) = tf ‘(t) for all t in (0, 1). Prove that every real 1 is an eigenvalue forT, and determine the eigenfunctions corresponding to A. 6. Let V be the linear space of all real polynomials p(x) of degree < n.
If p E V, define q =T(p) to mean that q(t) = p(t + 1) for all real t. Prove that T has only the eigenvalue 1. Whatare the eigenfunctions belonging to this eigenvalue? 7. Let V be the linear space of all functions continuous on (- 00, + to) and such that the integralJ?., f(t) dt exists for all real x. If f E V, let g = T(f) be defined by the equation g(x) =5”. co f (t)
dt . F’rove that every positive I is an eigenvalue for Tand determine the eigenfunctionscorresponding to 1. 8. Let V be lthe linear space of all functions continuous on (- ~1, + a) and such that the integralI”., t f(t) dt exists for all real x.
If f 6 Y let g = T(f) be defined by the equationg(x) =JY m t f (t) dt . Prove that every negative 1 is aneigenvalue for T and determine the eigenfunc-tions corresponding to 1. 9. Let V = C(0, n) be the real linear space of all real functions continuous on the interval [0, ~1.Let S be the subspace of all functions f which have a continuous second derivative in
linearand which also satisfy the boundary conditions f(0) =f”(r) = 0. Let T: S + V be the lineartransformation which maps each f onto its second derivative, T(f) = f II. Prove that theeigenvalues of Tare the numbers of the form -n2, where n = 1, 2, .
. . , and that the eigen-functions corresponding to -n2 are f(t) = c, sin nt , where L’, # 0. 10. Let V be the linear space of all real convergent sequences {x,}. Define T: V---t V as follows:If x = {x,} is a convergent sequence with limit a, let T(x) = {y,}, where yS = a - x, forn 2 1 . Prove that Thas only two eigenvalues, 1 = 0 and 3, = -1, and determine the
eigen-vectors belonging to each such 1. 11. Assume that a linear transformation T has two eigenvectors x and y belonging to distincteigenvalues i, and p. If ax + by is an eigenvector of T, prove that a = 0 or b = 0. 12. Let T: S -+ Vbe a linear transformation such that every nonzero element of S is an eigenvector.Prove that therle exists a scalar c such
that T(x) = cx . In other words, the only transformationwith this property is a scalar times the identity. [Hint: Use Exercise 11.1 102 EigenvaIues and eigenvectors 4.5 The finite-dimensional case. Characteristic polynomials If dim V = it, the problem of finding the eigenvalues of a linear transformation T: V+ Vcan be solved with the help of
determinants. We wish to find those scalars I such that theequation T(x) = 4x has a solution x with x # 0. The equation T(x) = Ax can be writtenin the form (U- T)(x) = 0, where Zis the identity transformation. If we let TA = AZ - T, then 1 is an eigenvalue if andonly if the equation (4.6) T,(x) = 0 has a nonzero solution x, in which case TA is not invertible
(because of Theorem 2.10).Therefore, by Theorem 2.20, a nonzero solution of (4.6) exists if and only if the matrix ofT, is singular.
If A is a matrix representation for T, then AZ - A is a matrix representationfor TL. By Theorem 3.13, the matrix AZ - A is singular if and only if det (AZ - A) = 0.Thus, if il is an eigenvalue for Tit satisfies the equation (4.7) det (AZ- A) = 0. Conversely, any 1 in the underlying jield of scalars which satisfies (4.7) is an eigenvalue.This suggests that we should
study the determinant det (AZ - A) as a function of A. THEOREM 4.4. Zf A is any n x n matrix and lyZ is the n x n identity matrix, thefinctionfdefined by the equation ,f(ii) = det (AZ - A) = det (AZ - A) is a polynomial in 1 of degree n. Moreover, the term of highest degree is I”, and the constantterm isf(0) = det (-A) = (- 1)” det A. Proof. The statement f(0) = det (-A)
follows at once from the definition of J Weprove thatf is a polynomial of degree n only for the case n 5 3. The proof in the generalcase can be given by induction and is left as an exercise. (See Exercise 9 in Section 4.8.) For n = 1 the determinant is the linear polynomialf(l) = 1 - alI.
For n = 2 we have det (AZ - A) =1 - alI --al2 = (a - ada - a,,) - a12a21- azl a - a22 = a2 - (all + az2M + (w22 - a12azl>) a quadratic polynomial in 1. For n = 3 we have a - all -al2 -al3det (ill - A) = -a21 I - a22 - a23 -4, -a32 1 - a,, Calculation of eigenvalues and eigenvectors in thejnite-dimensional case 103 = 0 - all>1 - a22 - az3 I I --a,, - az3 azl a - az2 1 -
a33+ al2 -a32 -a31 1 -. a,, -a31 -a32 The last two .terms are linear polynomials in 1. The first term is a cubic polynomial, theterm of highest degree being L3. DEFINITION. If A is an n x n matrix the determinant f(A) = det (AZ - A) is called the characteristic polynomial of A. The roots of the characteristic polynomial of A are complex numbers, some of
whichmay be real. If we let F denote either the real field R or the complex field C, we have thefollowing theorem,. THEOREM 43. Let T: V+ V be a linear transformation, where V has scaIars in F, anddim V = n . Let A be a matrix representation of T. Then the set of eigenvalues of T consistsof those roots of the characteristic polynomial of A which lie in
F. Proof. The discussion preceding Theorem 4.4 shows that every eigenvalue of T satisfiesthe equation det (AZ - A) = 0 and that any root of the characteristic polynomial of Awhich lies in 17 is an eigenvalue of T. , The matrix .4 depends on the choice of basis for V, but the eigenvalues of T were definedwithout reference t’o a basis. Therefore, the set of
roots of the characteristic polynomial ofA must be independent of the choice of basis. More than this is true. In a later section weshall prove that thie characteristic polynomial itself is independent of the choice of basis.We turn now to the problem of actually calculating the eigenvalues and eigenvectors in thefinite-dimensional case.
4.6 Calculation of’ eigenvalues and eigenvectors in the finite-dimensional case In the finite-dimensional case the eigenvalues and eigenvectors of a linear transformationTare also called eig,envalues and eigenvectors of each matrix representation of T. Thus, theeigenvalues of a square matrix A are the roots of the characteristic polynomial f(A) =det
(AZ - A). The eigenvectors corresponding to an eigenvalue A are those nonzerovectorsX= (x.r,. . . , x,) regarded as n x 1 column matrices satisfying the matrix equation AX = IX, or (AI- A)X= 0 . This is a system of n linear equations for the components x1, . . . , x,. Once we know ;1we can obtain the eigenvectors by solving this system.
We illustrate with three examplesthat exhibit difrerent features. 104 Eigenvalues and eigenvectors EXAMPLE 1. A matrix with all its eigenvalues distinct. The matrix 2 1 1 A = [2 3- 1 - 1 4 1- 2 has the characteristic polynomial det (AZ - A) = det [:’ t3 ;,:d =(n-l>(n+l)(n-3), so there are three distinct eigenvalues: 1 , - 1 , and 3. To find the eigenvectors
correspondingto 1 = 1 we solve the system AX = X, or This gives us which can be rewritten as -x1 - x2 - 3x, = 0. Adding the first and third equations we find xQ = 0, and all three equations then reduce tox1 + x2 = 0 .
Thus the eigenvectors corresponding to 1 = 1 are X = t(l, - 1 , 0) , wheret is any nonzero scalar. By similar calculations we find the eigenvectors X = t(0, 1 , -1) corresponding to L =- 1, and X = t(2,3, - 1) corresponding to il = 3, with t any nonzero scalar. Since theeigenvalues are distinct the correspondingeigenvectors (1, - 1 , 0), (0, 1, - l), and (2,3, -
1)are independent. The results can be summarized in tabular form as follows. In the thirdcolumn we have listed the dimension of the eigenspace E(1). Eigenvalue L Eigenvectors dim E(l) 1 t(1, -1, O), t # 0 1- 1 t(O, 1, -11, t # 0 1 3 r(2; 3, -1), t # 0 1 Calculation of eigenvalues and eigenvectors in thejnite-dimensional case 105 EXAMPLE 2. A matrix with
repeated eigenvalues. The matrix 2 - 1 1 A = [0 32 1 -1 I3 has the characteristic polynomial det (AZ - A) = det The eigenvalues are 2, 2, and 4. (We list the eigenvalue 2 twice to emphasize that it is adouble root of the characteristic polynomial.) To find the eigenvectors corresponding toA = 2 we solve the system AX = 2X, which reduces to -x2 + XQ =
0 x.2 - xg = 0 2x, + x2 + x3 = 0. This has the solution x2 = xg = -x1 so the eigenvectors corresponding to i = 2 aret(-1, 1, l), where t # 0. Similarly we find the eigenvectors t(1, - 1, 1) correspondingto the eigenvalue: 1 = 4. The results can be summarized as follows: Eigenvalue Eigenvectors dim E(il) 2, 2 t(-1, 1, l), t # 0 14 t(l, --I, l), t # 0 1 EXAMPLE 3.
Another matrix with repeated eigenvalues. The matrix has the characteristic polynomial (1 - 1)(1 - l)(A - 7). When L = 7 the system AX =7X becomes 5x, - X’J - x3 = 0 -2x, + 4x, - 2x, = 0 -3x, - 3x, + 3x, = 0. This has the solution x2 = 2x,, xs = 3x,, so the eigenvectors corresponding to 1 = 7 aret(l,2, 3), where t # 0. For the eigenvalue il = 1, the system
AX = X consists of theequation 106 Eigenvalues and eigenvectors Xl + x2 + x3 = 0 repeated three times. To solve this equation we may take x1 = a, xZ = b , where a and bare arbitrary, and then take x3 = -a - b . Thus every eigenvector corresponding to il = 1has the form (a, b, -a - b) = a(l,O, -1) + b(0, 1, -1), where a # 0 or b # 0. This means that the
vectors (1, 0, -1) and (0, 1, -1) form a basisfor E(1).
Hence dim E(A) = 2 when il = 1. The results can be summarized as follows: Eigenvalue Eigenvectors dim E(l) 7 t(l,2,3), t#O 1131 a(1, 0, - 1) + b(0, 1, -l), a, b not both 0 . 2 Note that in this example there are three independent eigenvectors but only two distincteigenvalues. 4.7 Trace of a matrix Letf(il) be the characteristic polynomial of an n x n
matrix A. We denote the n roots off(4 by 4, .
. . 3 A,, with each root written as often as its multiplicity indicates. Then wehave the factorization f(l) = (1 - A,) * * * (A - I,). We can also writef(A) in decreasing powers of I as follows, f(A) = A” + c,-lIz”-l + - * * + Cl2 + co. Comparing this with the factored form we find that the constant term c,, and thecoefficient ofAn-r are given by the formulas Since
we also have c,, = (-1)” det A, we see that 1, * * - 1, = det A. That is, theproduct of the roots of the characteristicpolynomialof A is equal to the determinantofA. The sum of the roots off(A) is called the trace of A, denoted by tr A. Thus, by definition, tr A = 5 Ii.i=l The coefficient of 1*-l is given by c,-r = -tr A. We can also compute this coefficientfrom the
determinant form forf(il) and we find that cnel = -(all + es* + ad. Exercises 107 (A proof of this formula is requested in Exercise 12 of Section 4.8.) The two formulas forc,,-~ show that tr A =iaii. That is, the trace of A is also equal to the sum of the diagonal elements of A.Since the sum of the diagonal elements is easy to compute it can be used as a
numerical check in calculations of eigenvalues. Further properties of the trace are described in thenext set of exerci;ses. 4.8 Exercises Determine the eigenvalues and eigenvectors of each of the matrices in Exercises 1 through 3.Also, for eaclh eigenvalue A compute the dimension of the eigenspace E(1). 1 a 2* b 1[I , a > 0, b > 0. 4.ThematricesP,=[T
i], Pz=[p -i-J, P,=[:: -T]occurinthequantum mechanical theory of electron spin and are called Pauli spin matrices, in honor of the physicistWolfgang Pauli (1900-1958).
Verify that they all have eigenvalues 1 and - 1 . Then determineall 2 x 2 matrices with complex entries having the two eigenvalues 1 and -1 . 5. Determine all 2 x 2 matrices with real entries whose eigenvalues are (a) real and distinct, (b)real and equal, (c) complex conjugates. 6.
Determine a, b, c, d, e, f, given that the vectors (1, 1, l), (1, 0, - l), and (1, - 1 , 0) are eigen-vectors of the matrix 7.
Calculate the eigenvalues and eigenvectors of each of the following matrices, Also, computethe dimension of the eigenspaceE(A) for each eigenvalue L 1 5 - 6 - 6 - 1 4 2 3 - 6 - 4I8. Calculate the eligenvalues of each of the five matrices (4 9 (b) 108 Eigenvalues and eigenvectors These are called Dirnc matrices in honor of Paul A. M. Dirac (1902-), the
Englishphysicist. They occur in the solution of the relativistic wave equation in quantum mechanics. 9. If A and B are n x n matrices, with B a diagonal matrix, prove (by induction) that the deter-minant f(n) = det (AB - A) is a polynomial in 1 with f(0) = (-1)” det A, and with thecoefficient of I” equal to the product of the diagonal entries of B. 10.
Prove that a square matrix A and its transpose At have the same characteristic polynomial.11. If A and B are n x n matrices, with A nonsingular, prove that AB and BA have the same set of eigenvalues.
Note: It can be shown that AB and BA have the same characteristic poly-nomial, even if A is singular, but you are not required to prove this. 12. Let A be an n x n matrix with characteristic polynomialf(l). Prove (by induction) that thecoefficient of An-l inf(A) is -tr A. 13. Let A and B be II x n matrices with det A = det B and tr A = tr B. Prove that A and B
havethe same characteristic polynomial if n = 2 but that this need not be the case if n > 2. 14. Prove each of the following statements about the trace. (a) t r (A + B) = t r A + t r B .(b) tr (CA) = c tr A.(c) tr (AB) = tr (BA).(d) tr At = tr A.
4.9 Matrices representing the same linear transformation. Similar matrices In this section we prove that two different matrix representations of a linear trans-formation have the same characteristic polynomial. To do this we investigate more closelythe relation between matrices which represent the same transformation. Let us recall how matrix
representations are defined. Suppose T: V-t W is a linearmapping of an n-dimensional space V into an m-dimensional space W. Let (e,, . . . , e,)and(w,,..., w,) be ordered bases for V and W respectively. The matrix representationof T relative to this choice of bases is the m x n matrix whose columns consist of the com-ponents of T(e,), . .
. ,T(e,) relative to the basis (wl, . . . , w,). Different matrix represen-tations arise from different choices of the bases. We consider now the case in which V = W, and we assume that the same ordered basish,..., e,) is used for both V and W. Let A = (a+) be the matrix of T relative to thisbasis. This means that we have (4.8) T(e,) = %tl aikei for k=l,2 ,..., n.
Now choose another ordered basis (q , . . . , u,) for both V and Wand let B = (bkj) be thematrix of T relative to this new basis.
Then we have Matrices representing the same linear transformation. Similar matrices 109 (4.9) T(ui) = 5 bkju, for j=l,2 ,..., n.k=l Since each ui is in the space spanned by e,, . . . , e, we can write (4.10) uj =$2jek for j = 1,2, . . , n, for some set of scalars cki . The n x n matrix C = (ckJ determined by these scalars is non-singular because it represents a
linear transformation which maps a basis of V onto anotherbasis of V. ,4pplying T to both members of (4.10) we also have the equations (4.11) T(uj) = i ckiT(ek) for j=l,2 ,..., n.k=l The systems of equations in (4.8) through (4.11) can be written more simply in matrixform by introducing matrices with vector entries.
Let E = [e,,...,e,] and u = [u, , * . .) u,] be 1 x n row matrices whose entries are the basis elements in question. Then the set ofequations in (4.10) can be written as a single matrix equation, (4.12) U=EC.
Similarly, if we introduce E’ q = [T(e,), . . . , T(e,)] and U’ = [T(q), . . . , T(u,)] , Equations (4.8), (4.9), and (4.11) become, respectively, (4.13) E’=EA, U’= UB, U’ = E’C. From (4.12) we also have E = UC-‘. To find the relation between A and B we express U’ in two ways in terms of U. From(4.13) we have U’ = UB and U’ = E’C = EAC = UC-IAC,
Therefore UB = ZJCYAC. But each entry in this matrix equation is a linear combination 110 Eigenvalues and eigenvectors of the basis vectors ul, . . . , u,, . Since the ui are independent we must have B = C-‘AC. Thus, we have proved the following theorem. THEOREM 4.6. If two n x n matrices A and B represent the same linear transformation T,then
there is a nonsingular matrix C such that B = C-lAC. Moreover, if A is the matrix of T relative to a basis E = [el , . , .
, e,] and if B is the matrixof T relative to a basis U = [ul, . . .
, u,], then for C we can take the nonsingular matrixrelating the two bases according to the matrix equation U = EC. The converse of Theorem 4.6 is also true. THEOREM 4.7. Let A and B be two n x n matrices related by an equation of the formB = C-IAC, where C is a nonsingular n x n matrix. Then A and B represent the samelinear transformation.
Proof. Choose a basis E = [e,, . .
.
, e,] for an n-dimensional space V. Let u,, . . . , u,be the vectors determined by the equations (4.14)n U j = xckjek for j=l,2 ,..., n,+l where the scalars c,, are the entries of C. Since C is nonsingular it represents an invertiblelinear transformation, so U = [uI, . . . , u,,] is also a basis for V, and we have U = EC. Let Tbe the linear transformation having the
matrix representation A relative to the basisE, and let S be the transformation having the matrix representation B relative to the basis U.Then we have (4.15) T(ek) = fi aike{ for k = 1,2, . .
. , nf=l and (4.16) s(U,) = 5 b,,u, for j = 1,2, . . .
, n.k = = l We shall prove that S = T by showing that T(u,) = S(u,) for eachj.Equations (4.15) and (4.16) can be written more simply in matrix form as follows, Matrices representing the same linear transformation. Similar matrices 111 [TM . . . , %Jl = E-4, [SW, - * . , S(u,)] = UB. Applying 7’to (4.14) we also obtain the relation T(z+) = 2 cJ(e,), or [T(q), .
.
. , T(u,)] = EAC. But we have UB = ECB = EC(C-lAC) = EAC, which shows tha.t T(uJ = S(q) for each j. Therefore T(X) = S(x) for each x in V, so T =S. In other words, the matrices A and B represent the same linear transformation. DEFINITION.
‘Two n x n matrices A and B are called similar tf there is a nonsingularmatrix C such that B = C-IAC. Theorems 4.6 and 4.7 can be combined to give us THEOREM ,4.8. Two n x n matrices are similar tfand only tf they represent the same Iineartransformation. Similar matrices share many properties. For example, they have the same determinantsince
det (CYAC) = det (C-l)(det A)(det C) = det A. This property gives us the following theorem. THEOREM 4.9. Similar matrices have the same characteristic polynomial and therefore thesame eigenvalues. Proof. If A and B are similar there is a nonsingular matrix C such that B = C-lAC.Therefore we have ill .- B = LI - C-IAC = ilC-=IC - C-lAC := C-l(ilZ -
A)C. This shows that .ill - B and AZ - A are similar, so det (Al- B) = det (AZ - A). Theorems 4.8 and 4.9 together show that all matrix representations of a given lineartransformation T have the same characteristic polynomial. This polynomial is also calledthe characteristic polynomial of T. The next theorem is a combination of Theorems 4.5, 4.2, and 4.6.
In Theorem 4.10, Fdenotes either the real field R or the complex field C. 112 Eigenvalues and eigenvectors THEOREM 4.10. Let T: V + V be a linear transformation, where V has scalars in F, anddim V = n.
Assume that the characteristic polynomial of T has n distinct roots I,, . . . , 1, inF. Then we have: (a) The corresponding eigenvectors ul, . . . , u, form a basis for V.(b) The matrix of T relative to the ordered basis U = [ul, . . . , u,] is the diagonal matrix A having the eigenvalues as diagonal entries: A=diag(il,,...,il,). (c) If A is the matrix of T relative to
another basis E = [e,, . . .
, e,], then A = C-lAC, where C is the nonsingular matrix relating the two bases by the equation U=EC. Proof. By Theorem 4.5 each root Ai is an eigenvalue. Since there are n distinct roots,Theorem 4.2 tells us that the corresponding eigenvectors ul, . . . , u, are independent.Hence they form a basis for V. This proves (a). Since T(uJ = &ui the matrix of T
relativeto U is the diagonal matrix A, which proves (b). To prove (c) we use Theorem 4.6. Note: The nonsingular matrix C in Theorem 4.10 is called a diagonalizing matrix. Ifh,..., e,) is the basis of unit coordinate vectors (I,, . . . , I,), then the equation U =EC in Theorem 4.10 shows that the kth column of C consists of the components of theeigenvector
ule relative to (Z1, . . . , I,). If the eigenvalues of A are distinct then A is similar to a diagonal matrix. If the eigen-values are not distinct then A still might be similar to a diagonal matrix. This will happenif and only if there are k independent eigenvectors corresponding to each eigenvalue ofmultiplicity k. Examples occur in the next set of exercises. 4.10
Exercises Prove that the matrices [i :] and [IO T] have the same eigenvalues but are not similar. In each case find a nonsingular matrix C such that C-IAC is a diagonal matrix or explain whyno such C exists. Three bases in the plane are given. With respect to these bases a point has components (x,, x,),(yl,yzh and (q, Z-A respectively. Suppose that [yl,
yzl = [xl, x&, h 4 = [xl, x,lB, and[z,, z2] = [yl, y,]C, where A, B, C are 2 x 2 matrices. Express C in terms of A and B. Exercises 113 4. 5. 6. 7. 8.
In each case, slhow that the eigenvalues of A are not distinct but that A has three independenteigenvectors. Find a nonsingular matrix C such that C-‘AC is a diagonal matrix. Show that none of the following matrices is similar to a diagonal matrix, but that each is similarA 0 to a triangular matrix of the form[I1 I’ where A is an eigenvalue. 0 -1 0
Determine the eigenvalues and eigenvectors of the matrix ! I0 0 1 and thereby showthat it is not similar to a diagonal matrix. -. 1 -3 3 (a) Prove that a. square matrix A is nonsingular if and only if 0 is not an eigenvalue of A.(b) If A is nonsingular, prove that the eigenvalues of A-l are the reciprocals of the eigenvaluesofA. Given an n x n matrix A with
real entries such that A2 = -I. Prove the following statementsabout A.(a) A is nonsingular.(b) n is even.(c) A has no real eigenvalues.(d) det A q = 1.
5EIGENVALUES OF OPERATORS ACTING ON EUCLIDEAN SPACES 5.1 Eigenvalues and inner products This chapter describes some properties of eigenvalues and eigenvectors of linear trans-formations that operate on Euclidean spaces, that is, on linear spaces having an innerproduct. We recall the fundamental properties of inner products. In a
real Euclidean space an inner product (x, u) of two elements x andy is a real numbersatisfying the following properties : (1) (X>Y) = 09 4 (symmetry) (2) 6 + z, Y> = (XT Y> + (z9 r> (linearity) (3) (WY) = C(XPY> (homogeneity) (4) (x, 4 > 0 i f x#O (positivity).In a complex Euclidean space the inner product is a complex number satisfying the same
properties, with the exception that symmetry is replaced by Hermitian symmetry, (1’) (x3 JJ> = (y, 4, where the bar denotes the complex conjugate. In (3) the scalar c is complex. From (1’)and (3) we obtain (3’) (X, Cy) = f(& J,) 3 which tells us that scalars are conjugated when taken out of the second factor. Takingx = y in (1’) we see that (x, x) is real
so property (4) is meaningful if the space is complex. When we use the term Euclidean space without further designation it is to be understoodthat the space can be real or complex. Although most of our applications will be to finite-dimensional spaces, we do not require this restriction at the outset. The first theorem shows that eigenvalues (if they
exist) can be expressed in terms of theinner product. THEOREM 5.1. Let E be a Euclidean space, let V be a subspace of E, and let T: V--t E be alinear transformation having an eigenvalue 31 with a corresponding eigenvector x. Then we have (5.1)11A Hermitian and skew-Hermitian transformations 115 Proof. Since T(x) = iix we have (T(x), x) = (Ax, x)
= 1(x, x).
Since x # 0 we can divide by (x, x) to get (5.1).
Several properties of eigenvalues are easily deduced from Equation (5.1).
For example,from the Hermitian symmetry of the inner product we have the companion formula (5.2)2 = (x2 T(x)) (x9 x> for the complex conjugate of A. From (5.1) and (5.2) we see that il is real (1 = 1) if andonly if (T(x), x) is real, that is, if and only if (T(x), 4 = (x, T(x)) for the eigenvector x . (This condition is trivially satisfied in a real Euclidean
space.) Also, ;i is pure imaginary(A = -2) if and only if (T(x , x 1s)) ’ pure imaginary, that is, if and only if (T(x), 4 = -6~ T(x)) for the eigenvector x. 5.2 Hermitian and skew-Hermitian transformations In this section we introduce two important types of linear operators which act on Euclid-ean spaces. These operators have two categories of names,
depending on whether theunderlying Euclidean space has a real or complex inner product. In the real case the trans-formations are called symmetric and skew-symmetric. In the complex case they are calledHermitian and skew-Hermitian. These transformations occur in many different applications.For example, Hermitian operators on infinite-
dimensional spaces play an important role inquantum mechanics. We shall discuss primarily the complex case since it presents no addeddifficulties. DEFINITION. Let E be a Euclidean space and let V be a sub..l;aace of E. A linear trans-formation T: V--f E is called Hermitian on V if (T(x), y) = (x, T(y)) .for all x and y in V. Operator T is called skew-
Hermitian on V if (T(X), y) = - (x, T(y)> .for all x and y in V. In other words, a Hermitian operator Tcan be shifted from one factor of an inner productto the other without changing the value of the product. Shifting a skew-Hermitian operatorchanges the sign of the product. Note: As already mentioned, if Eis a reaLEuclidean space, H’ermitian
transformationsare also called symmetric; skew-Hermitian transformations are called skew-symmetric. 116 Eigenvalues of operators acting on Euclidean spaces EXAMPLE 1. Symmetry and skew-symmetry in the space C(a, b). Let C(a, b) denote thespace of all real functions continuous on a closed interval [a, b], with the real inner product U-Y s> =
jabf(Os(t) dt.Let Vbe a subspace of C(a, b). If T: V -+ C(a, b) is a linear transformation then cf, T(g)) =ji f(t)Tg(t) dt , where we have written Q(t) for T(g)(t). Therefore the conditions for sym-metry and skew-symmetry become (5.3) and (5.4) s,” {f(t)Tg(t) - g(t)Tf(t)} dt = 0 if T is symmetric, jab {f(t)Tg(t) + g(t)Tf(t)} dt = 0 if T is skew-symmetric.
EXAMPLE 2. Multiplication by a$xedfunction. In the space C(a, b) of Example 1, choosea fixed functionp and define T(f) = pf, the product ofp and f. For this T, Equation (5.3)is satisfied for all f and g in C(a, b) since the integrand is zero. Therefore, multiplicationby a fixed function is a symmetric operator. EXAMPLE 3. The dyerentiation operator. In
the space C(a, b) of Example 1, let Y be thesubspace consisting of all functions f which have a continuous derivative in the open interval(a, b) and which also satisfy the boundary condition f (a) = f(b). Let D: V --+ C(a, b) be thedifferentiation operator given by D(f) = f' . It is easy to prove that D is skew-symmetric.In this case the integrand in (5.4) is the
derivative of the product fg, so the integral is equalto Iab (h)‘(t) dt = @) - f(a>g(a) . Since bothfand g satisfy the boundary condition, we have f (b)g(b) - f (a)g(a) = 0. Thus,the boundary condition implies skew-symmetry of D. The only eigenfunctions in the sub-space V are the constant functions. They belong to the eigenvalue 0.
EXAMPLE 4. Sturm-Liouville operators. This example is important in the theory of linearsecond-order differential equations.
We use the space C(a, b) of Example I once more andlet V be the subspace consisting of all f which have a continuous second derivative in [a, b]and which also satisfy the two boundary conditions (5.5) p(4f (4 = 0 p p(blf(b) = 0, wherep is a fixed function in C(a, b) with a continuous derivative on [a, b]. Let q be anotherfixed function in C(a, b) and let T:
V ---f C(a, b) be the operator defined by the equation T(f) = Cpf 7' + qf. Orthogonality of eigenvectors corresponding to distinct eigenvalues 117 This is called a Sturm-Liouville operator. To test for symmetry we note that j?(g) - gT(f) = f(pg’)’ - dpf’)’ ’ Using this in (5.3) and integrating both ji f * (pg’)’ dt andJ,” g * (pf’)’ dt by parts, we find jab if%) -
gT(f)) dt = fm’ I;- Iab pg’f’ dt - gpf’ [I+ 1,” pf’g’ dt = 0, since bothfand g satisfy the boundary conditions (5.5). Hence T is symmetric on V. Theeigenfunctions of Tare those nonzero f which satisfy, for some real 1, a differential equationof the form Q?fY’ + qf = v on [a, b], and also satisfy the boundary conditions (5.5). 5.3 Eigenvalues and eigenvectors of
Hermitian and skew-Hermitian operators Regarding eigenvalues we have the following theorem; THEOREM 5.2. Assume T has an eigenvalue il. Then we have:(a) If T is Hermitian, 1 is real: il = 2.(b) If T is skew-Hermitian, R is pure imaginary: .1 = -1. Proof. Let x be an eigenvector corresponding to 1. Then we have 2 _ (T(x), x>- - and - ___-j _ (x9 T(x))
(x, x> (x,x> * If Tis Hermitian we have (T(x), x) = (x, T(x)) so A = 1. If Tis skew-Hermitian we have G’W, 4 = -(x, T(x)) so il = -2. Note: If T is symmetric, Theorem 5.2 tells us nothing new about the eigenvalues of Tsince all the eigenvalues must be real if the inner product is real..
If T is skew-symmetric,the eigenvalues of T must be both real and pure imaginary. Hence all the eigenvaluesof a skew-symmetric operator must be zero (if any exist). 5.4 Orthogonality of eigenvectors corresponding to distinct eigenvalues Distinct eigenvalues of any linear transformation correspond to independent eigen-vectors (by Theorem 4.2). For
Hermitian and skew-Hermitian transformations more istrue. THEOREM 5.3. Let T be a Hermitian or skew-Hermitian transformation, and let L and ,abe distinct eigenvalues of T with corresponding eigenvectors x and y. Then x and y areorthogonal; that is, (x, y) = 0. ProoJ We write T(x) = Lx, T(y) = ,uy and compare the two inner products (T(x), y)and
(x, T(y)). We have (T(x), y> = (k y> = W, y> and (x, To)) = (x, luy> = /Xx, y).
118 Eigenvalues of operators acting on Euclidean spaces If T is Hermitian this gives us 2(x, y) = ,L?(x, y) = ,u(x,~) since ,,v = ,ii. Therefore(x, y) = 0 since il # p. If T is skew-Hermitian we obtain 1(x, y) = -,E(x, y) = ,M(x, y)which again implies (x, J) = 0. EXAMPLE. We apply Theorem 5.3 to those nonzero functions which satisfy a differentialequation of
the form (5.6) w-7' + qf= Af on an interval [a, b], and which also satisfy the boundary conditionsp(a)f(a) = p(b)f (b) =0. The conclusion is that ;any two solutions f and g corresponding to two distinct values of;I are orthogonal. For example, consider the differential equation of simple harmonicmotion, f" + k2f= 0 on the interval [0, 7~1, where k + 0.
This has the form (5.6) with p = 1, q = 0, andA = -k2. All solutions are given byf(t) = c1 cos kt + c2 sin kt . The boundary conditionf (0) = 0 implies c1 = 0. The second boundary condition, f (7r) = 0, implies cg sin krr = 0.Since c2 # 0 for a nonzero solution, we must have sin kr = 0, which means that k is aninteger. In other words, nonzero solutions
which satisfy the boundary conditions exist ifand only if k is an integer. These solutions are f(t) = sin nt , n = f 1, i-2, ... Theorthogonality condition irnplied by Theorem 5.3 now becomes the familiar relation s: sin nt sin mt dt = 0 if m2 and n2 are distinct integers. 5.5 Exercises1. Let E be a Euclidean space, let V be a subspace, and let T: V -+ E be a
given linear trans- formation. Let ii be a scalar and x a nonzero element of V. Prove that I is an eigenvalue of Twith x as an eigenvector if and only if (TW,y) = %y) for every y in E. 2. Let T(x) = cx for every x in a linear space V, where c is a fixed scalar. Prove that T is sym-metric if V is a real Euclidean space. 3. Assume T: I/ -+ V is a Hermitian
transformation.(a) Prove that T” is Hermitian for every positive integer n, and that T-l is Hermitian if T isinvertible.(b) What can you conclude about Tn and T-l if T is skew-Hermitian? 4. Let T,: V -+ E and T,: V -+ E be two Hermitian transformations.(a) Prove that aT, + bT, is Hermitian for all real scalars a and b.(b) Prove that the product
(composition) TIT, is Hermitian if TI and T, commute, that is, ifTIT, = T,T,. 5. Let V = Vs(R) with the usual dot product as inner product.
Let T be a reflection in the xy-plane; that is, let T(!) = i, T(i) = j, and T(k) = -k. Prove that T is symmetric. Exercises 119 6. Let C(0, 1) be the real linear space of all real functions continuous on [0, l] with inner product(f,g) =]A f(t)g(t) dt. Let V be the subspace of all f such that]: f(t) dt = 0. Let T: V --fC(0, 1) be the integration operator defined by Tf (x)
= fi f(t) dt . Prove that T is skew-sym-metric. 7. Let V be the real Euclidean space of all real polynomials with the inner product (f,g) =]’r f (t)g(t) dt . Determine which of the following transformations T: V + V is symmetric orskew-symmetric:(a) TfW =f(-4. cc> rf(x =f(x) +f(-4.(b) Tf(4 =f(x)f(-4. Cd) ?fCf(x) =f(x) -f(-4. 8. Refer to Example 4 of Section 5.2.
Modify the inner product as follows: (f,g> = j: f(W)W dt, where w is a fixed positive function in C(a, b). Modify the Sturm-Liouville operator T bywriting T(f) - cpf ‘)’ +- qf.W Prove that the modified operator is symmetric on the subspace V.9. Let V be a subspace of a complex Euclidean space E. Let T: V + E be a linear transformation and define a scalar-
valued function Q on V as follows : Q(x) = (T(x), 4 for all x in V. (a) If T is Hermitian on I’, prove that Q(x) is real for all x.(b) If T is skew-Hermitian, prove that Q(x) is pure imaginary for all x.(c) Prove that Q(tx) = tie(x) for every scalar 1.(d) Prove that Q(x + y) = Q(x) + Q(r) + (T(x).
,v) + (T(y), x), and find a correspondingformula for Q(x + tr).(e) If Q(x) = 0 for all x prove that T(x) = 0 for all x.(f) If Q(x) is real for all x prove that T is Hermitian. [Hint: Use the fact that Q(x + ty)equals its conjugate for every scalar t.] 10. This exercise shows that the Legendre polynomials (introduced in Section 1.14) are eigen-functions of a Sturm-
Liouville operator. The Legendre polynomials are defined by the equation pn(d = &,/y(1), where f,(t) = (t2 - 1)“. (a) Verify that (t2 - 1) f’,(t) = 2ntf,(t).(b) Differentiate the equation in (a) n + 1 times, using Leibniz’s formula (see p.
222 of VolumeI) to obtain (t2 - 1)f :+2)(t) + 2t(n + 1)f p+qt> + n(n + 1)f y(t) = 2ntf k+qt) + 2n(n + 1)f 2’(r). (c) Show that the equation in (b) can be rewritten in the form w2 - l)Ph(f)]’ = n(n + l)P&). This shows that P,(t) is an eigenfunction of the Sturm-Liouville operator T given on the 120 Eigenvalues of operators acting on Euclidean spaces interval
[-1 , l] by r(f) = (pf’)‘, wherep(t) = t2 - 1 . The eigenfunction P,(t) belongs tothe eigenvalue A = n(n -b 1). In this example the boundary conditions for symmetry areautomatically satisfied since p(l) = p(- 1) = 0. 5.6 Existence of an orthonormal set of eigenvectors for Hermitian and skew-Hermitianoperators acting on finite-dimensional spaces Both
Theorems 5.2 and 5.3 are based on the assumption that T has an eigenvalue.
Aswe know, eigenvalues need not exist. However, if T acts on ajnite-dimensional complexspace, then eigenvalues always exist since they are the roots of the characteristic polynomial.If T is Hermitian, all the eigenvalues are real. If T is skew-Hermitian, all the eigenvaluesare pure imaginary. We also know that two distinct eigenvalues belong to
orthogonal eigenvectors if T isHermitian or skew-Hermitian. Using this property we can prove that T has an orthonormalset of eigenvectors which spans the whole space. (We recall that an orthogonal set is calledorthonorm.al if each of its (elements has norm 1.) THEOREM 5.4. Assume dim V = n and let T: V ---f V be Hermitian or skew-
Hermitian.Then there exist n eigenvectors ul, . . . , u, of T which form an orthonormal basis for V.Hence the matrix of T relative to this basis is the diagonal matrix A = diag (1,) . . . , 1,))where lk is the eigenvalue belonging to ule. Proof. We use inductian on the dimension n. If n = 1, then T has exactly one eigen-value. Any eigenvector u1 of norm 1 is
an orthonormal basis for V. Now assume the theorem is true for every Euclidean space of dimension n - 1 . Toprove it is also true for V we choose an eigenvalue 1, for T and a corresponding eigenvectoru1 of norm 1. Then T(u,) = &u, and (IuJ = 1 . Let S be the subspace spanned by ul.We shall apply the induction hypothesis to the subspace Sl
consisting of all elements in Ywhich are orthogonal to u:~, P = {x) x E v, (x, 241) = O}. To do this we need to know that dim Sl = n - 1 and that T maps Sl into itself.From Theorem 1.7(a) we know that u1 is part of a basis for V, say the basis (%,VZ,..., vJ. We can assume, without loss in generality, that this is an orthonormalbasis. (If not, we apply the
Gram-Schmidt process to convert it into an orthonormalbasis, keeping u1 as the first basis element.) Now take any x in S1 and write Then x1 = (x, uJ = 0 since the basis is orthonormal, so x is in the space spanned byv2, * . .) v,. Hence dim SI = n - 1. Next we show that T maps Sl into itself. Assume T is Hermitian. If x E Sl we have (T(x), 4 = 6, W4) =
(x, hu,> = Mx., uJ = 0, Matrix representations for Hermitian and skew-Hermitian operators 121 so T(x) E SL . Since T is Hermitian on SL we can apply the induction hypothesis to findthat T has n - 1 eigenvectors u2, . . . , U, which form an orthonormal basis for Sl.Therefore the orthogonal set ur, . . : , U, is an orthonormal basis for V. This proves
thetheorem if T is Hermitian. A similar argument works if T is skew-Hermitian. 5.7 Matrix representations for Hermitian and skew-Hermitian operators In this section we assume that V is a finite-dimensional Euclidean space. A Hermitianor skew-Hermitian transformation can be characterized in terms of its action on theelements of any basis.
THEOREM 5.5. Let (f?,, . . . , e,) be a basis for V and let T: V -+ V be a linear transfor-mation. Then we have: (a) T is Hermitian if and on4 if (T(e)), ei) = (ej, T(ei)) for all i andj.(b) T is skew-Hermitian if and onfy if (T(e,), ei) = - (ej, T(ei))for all i and j. Proof. Take any two elements x and y in V and express each in terms of the basiselements, say x = 1 xjej
and y = 2 yiei. Then we have Similarly we find (x, T(Y)) = i 2 qiii(q, T(4).+I i=l Statements (a) and (b) following immediately from these equations. Now we characterize these concepts in terms of a matrix representation of T. THEOREM 5.6. Let (e, , . . . , e,) be an orthonormal basis for V, and let A = (aij) be thematrix representation of a linear
transformation T: V-+ V relative to this basis. Then wehave: (a) T is Hermitian if and only if aij = riji for all i andj.(b) T is skew-Hermitian if and only ifaij = -Cji for all i andj. Proof. Since A is the matrix of T we have T(eJ = z;==, akjek . Taking the inner productof T(eJ with ei and using the linearity of the inner product we obtain But (ek , ei) = 0 unless
k = i, so the last sum simplifies to aij(ei, ei) = aij since (ei , ei) = 1 .Hence we have aij = (Wj), 4 for all i, j. 122 Eigenvalues of operators acting on Euclidean spaces Interchanging i and j, taking conjugates, and using the Hermitian symmetry of the innerproduct, ‘we find aji = (ej, T(4) for all i, j. Now we apply Theorem 5.5 to complete the proof. 5.8
Hermitian and skew-Hermitian matrices. The adjoint of a matrix The following definition is suggested by Theorem 5.6. DEFINITION. A square matrix A = (adj) is called Hermitian if aij = cji for all i and j.Matrix A is called skew-hrermitian if aij = -cjji for all i and j. Theorem 5.6 states that a transformation T on a finite-dimensional space V is
Hermitianor skew-Hermitian according as its matrix relative to an orthonormal basis is Hermitianor skew-Hermitian. These matrices can be described in another way. Let 2 denote the matrix obtained byreplacing each entry of A by its complex conjugate.
Matrix A is called the conjugate of A.Matrix A is Hermitian if and only if it is equal to the transpose of its conjugate, A = At.It is skew-Hermitian if A = -Jt. The transpose of the conjugate is given a special name. DEFINITION OF THE ADJOI‘NT OF A MATRIX.
For any matrix A, the transpose of the conjugate,xt, is also called the adjoint of A and is denoted by A*. Thus, a square matrix A is Hermitian if A = A*, and skew-Hermitian if A = -A*.A Hermitian matrix is also called self-adjoint. Note: Much of the older matrix literature uses the term adjoint for the transpose of thecofactor matrix, an entirely different
object. The definition given here conforms to thecurrent nomenclature in the theory of linear operators. 5.9 Diagonalization of a Hermitian or skew-Hermitian matrix THEOREM 5.7. Every n x n Hermitian or skew-Hermitian matrix A is similar to thediagonal matrix A = diag (2,) . . . , I,) of its eigenvalues. Moreover, we have A = C-IAC, where C is a
nonsingular matrix whose inverse is its adjoint, C-l = C* . Proof. Let V be the space of n-tuples of complex numbers, and let (e,, . .
. , e,) be theorthonormal basis of unit coordinate vectors. If x = 2 xiei and y = zyyiei let the innerproduct be given by (x, y) = 2 XJi . For the given matrix A, let T be the transformationrepresented by A relative to the chosen basis. Then Theorem 5.4 tells us that V has an Unitary matrices.
Orthogonal matrices 123 orthonormal basis of eigenvectors (ul, . . . , u,), relative to which T has the diagonalmatrix representation A = diag (1r, . . . , A,), where 1, is the eigenvalue belonging to u,.Since both A! and A represent T they are similar, so we have A = CFAC, where C = (cu)is the nonsingular matrix relating the two bases: [#I 3 * * * 9 u,] =
[el , . . . , e,]C. This equation shows that thejth column of C consists of the components of uj relative to(el, . . .
, e,). Therefore cij is the ith component of ui. The inner product of uj and Ui isgiven by Since {ur , . . . , u,} is an orthonormal set, this shows that CC* = I, so C-l = C* . Note: The proof of Theorem 5.7 also tells us how to determine the diagonalizingmatrix C. We find an orthonormal set of eigenvectors ul, . . . , u, and then use thecomponents of uj
(relative to the basis of unit coordinate vectors) as the entries of thejthcolumn of C. 2 2EXAMPLE 1, The real Hermitian matrix A = [12 5 has eigenvalues il, = 1 and 1, = 6.
The eigenvectors belonging to 1 are t(2, -l), t # 0. Those belonging to 6 are t(l, 2), t # 0. The two eigenvectors u1 = t(2, -1) and u2 = t(l, 2) with t = l/J5 form anorthonormal set. Therefore the matrix 2 1 C=ljT -1 2[1is a diagonalizing matrix for A. In this case C* = Ct since C is real. It is easily verified 1 0that PAC = [10 6’ EXAMPLE 2. If A is already a
diagonal matrix, then the diagonalizing matrix C of Theorem5.7 either leaves A unchanged or m&ely rearranges the diagonal entries. 5.10 Unitary matrices. Orthogonal matrices DEFINITION. A square matrix A is called unitary if AA* = I.

It is called orthogonal ifAAt = I. Note: Every real unitary matrix is orthogonal since A* = At. Theorem 5.7 tells us that a Hermitian or skew-Hermitian matrix can always be diago-nalized by a unitary matrix. A real Hermitian matrix has real eigenvalues and the corre-sponding eigenvectors can be taken real. Therefore a real Hermitian matrix can be

124 Eigenvalues of operators acting on Euclidean spaces diagonalized by a real orthogonal matrix. This is not true for real skew-Hermitian matrices.(See Exercise 11 in Section 5.11.) We also have the following related concepts. DEFINITION.

A square matrix A with real or complex entries is called symmetric if A =At ; it is called skew-symmetric if A = -At. EXAMPLE 3. If A is real, its adjoint is equal to its transpose, A* = At. Thus, every realHermitian matrix is symmetric, but a symmetric matrix need not be Hermitian.

1 -I- i 2,EXAMPLE: 4. If A = [I3 -- i 4i 'thenA=[i+: -ti], At=[Izi 3il] andA*=[lii 3_+4:I. EXAMPLE: 5. Both matrices [: metric, the second is not. :] and [2 y i 2 : i] are Hermitian. The first is sym- EXAMPLE 6. Both matrices [i -3 and L ,2] are skew-Hermitian. The first is skew-symmetric, the second is not. EXAMPLE 7. All the diagonal elements of a

Hermitian matrix are real. All the diagonalelements of a skew-Hermitian matrix are pure imaginary. All the diagonal elements of askew-symmetric matrix are zero. EXAMPLE 8. For any square matrix A, the matrix B = &(A + A*) is Hermitian, and thematrix C = $(A - A*) is skew-Hermitian. Their sum is A. Thus, every square matrixA can be expressed

as a sum A = B + C, where B is Hermitian and C is skew-Hermitian.It is an easy exercise to verify that this decomposition is unique.

Also every square matrixA can be expressed uniquely as the sum of a symmetric matrix, &4 + At), and a skew-symmetric matrix, $(A - At). EXAMPLE 9. If A is orthogonal we have 1 = det (AAt) = (det A)(det At) = (det A)2, sodetA= f l . 5.11 Exercises 1. Determine which of the following matrices are symmetric, skew-symmetric, Hermitian, skew-

Hermitian. Exercises 125 2. (a) Verify that the 2 x 2 matrix A = is an orthogonal matrix. (b) Let T be the linear transformation with the above matrix A relative to the usual basis{i,i}. Prove that T maps each point in the plane with polar coordinates (r, a) onto the pointwith polar coordinates (r, w + 0). Thus, T is a rotation of the plane about the origin, 0

beingthe angle of rotation. 3. Let V be real 3-space with the usual basis vectors i, j, k. Prove that each of the followingmatrices is orthogonal and represents the transformation indicated. 1 0 0 (a) 0 1 0[1 (reflection in the xy-plane). 0 0 -1 (reflection through the x-axis). -1 0 0 Cc) [I o-1 0 (reflection through the origin). 0 0 -1 (d) 1: co:0 -s.o] (rotation

about the x-axis). 10 sin 0 cos 8 1 - 1 0 0 (4 10 cos 19 -sin 19 (rotation about x-axis followed by reflection in the yz-plane). 1 0 sin 8 cos e 14. A real orthogonal matrix A is called proper if det A = 1 , and improper if det A = -1 . (a) If A is a proper 2 x 2 matrix, prove that A =a rotation through an angle 8. for some 0. This represents (b) Prove that [A 01]

and [-10 (I)] are improper matrices. The first matrix represents a reflection of the xy-plane through the x-axis; the second represents a reflection through they-axis. Find all improper 2 x 2 matrices. In each of Exercises 5 through 8, find (a) an orthogonal set of eigenvectors for A, and (b)a unitary matrix C such that C-lAC is a diagonal matrix. 9 125. A

= [I 0 -2 12 16 *6. A = [I2 0' 1 3 4 %A= [34 10.

10 1 126 Eigenvalues of operators acting on Euclidean spaces 9. Determine which of the following matrices are unitary, and which are orthogonal (a, b, 0 real). 10. The special theory of relativity makes use of the equations x' = a(x -- ut), y'= y, z'= z, t' = a(t - vx/c2), Here u is the velocity o'f a moving object, c the speed of light, and a = c/m.

Thelinear transformation which maps (x, y, z, t) onto (x', y', z', t') is called a Lorentz transformation.(a> Let 63, ~2, x3, xq) q = (x, y, z, ict) and (xi, xi, xi, xi) = (x', y', z', ict'). Show that thefour equations can be written as one matrix equation, a 0 0 -iav/c 0 1 0 0 b;,x;;,x;, x;1 = [Xl, x2,x,, xql i 0 0 1 0 1 * liavjc 0 0 a J (b) Prove that the 4 x 4 matrix in (a) is

orthogonal but not unitary.0 a 11. Let a be a nonzero real number and let A be the skew-symmetric matrix A = [I.--a 0(a) Find an orthonormal set of eigenvectors for A.(b) Find a unitary matrix C such that @AC is a diagonal matrix.(c) Prove that there is no real orthogonal matrix C such that CplAC is a diagonal matrix.

12. If the eigenvalues of a Hermitian or skew-Hermitian matrix A are all equal to c, prove thatA = c l . 13. If A is a real skew-symmetric matrix, prove that both Z - A and Z + A are nonsingular andthat (I - A)(Z + A)-1 is orthogonal. 14. For each of the following statements about n x n matrices, give a proof or exhibit a counterexample.(a) If A and B are

unitary, then A + B is unitary.(b) If A and B are unitary, then AB is unitary.(c) If A and AB are unitary, then B is unitary.(d) If A and B are unitary, then A + B is not unitary.

5.12 Quadratic forms Let V be a real Euclidea:n space and let T: V+ V be a symmetric operator. This meansthat T can be shifted from one factor of an inner product to the other, V(x), y) = (x, T(y)) f o r a l l x a n d y i n V. Given T, we define a real-valued function Q on V by the equation Q(x) = UW 4. Quadratic forms 127 The function Q is called the

quadratic form associated with T. The term "quadratic" issuggested by the following theorem which shows that in the finite-dimensional case Q(x)is a quadratic polynomial in the components of x. THEOREM 5.8. Let (e,, . . . , e,) be an orthonormal basis for a real Euclidean space V.Let T: V + V be a symmetric transformation, and let A = (aii) be the

matrix of T relativeto this basis. Then the quadratic form Q(x) = (T(x), x) is related to A as follows: (5.7) Q(X) =i~~~i+ixi !f x =i2xieie Proof. By linearity we have T(x) = 2 xiT(eJ .

Therefore This proves (5.7) since aij = aji = (T(ei), ej). The sum appearing in (5.7) is meaningful even if the matrix A is not symmetric. DEFINITION. Let V be any real Euclidean space with an orthonormal basis (e, , . . . , e,), andlet A = (aij) by any n x n matrix of scalars. The scalar-valued function Q defined at eachelement x = 2 xiei of V by the double

sum (5.8) Q(X) = iz $FilXiXi is called the quadratic form associated with A. If A is a diagonal matrix, then aij = 0 if i # j so the sum in (5.8) contains only squaredterms and can be written more simply as Q(x) = i$l w,2 . In this case the quadratic form is called a diagonalform.The double sum appearing in (5.8) can also be expressed as a product of three

matrices. THEOREM 5.9. Let x = [Xl,.. . , x,] be a 1 x n row matrix, and let A = (a,J be ann x n matrix. Then XAXt is a 1 x 1 matrix with entry (5.9) Proof. Th.e product XA is a 1 x II matrix, XA = [yl, . . . , y,] , where entry yj is the dotproduct of X with thejth column of A, yi = i xiaij.i=l 128 Eigenvalues of operators acting on Euclidean spaces Therefore

the product XLF is a 1 x 1 matrix whose single entry is the dot product Note: It is customary to identify the 1 x 1 matrix XAXt with the sum in (5.9) and tocall the product XAXt a. quadratic form. Equation (5.8) is written more simply as follows: Q(x) = XAXt . 1 - 1EXAMPLE 1. Let z'i = [1 X = [x1, x2]. Then we have .- 3 5' 1 - 1XA = I'X,) x2] [1 = [Xl -

3x,, -x1 + 5x,1, - 3 5and hence XAXt = [Xl -- 3x,, -x1 + 5x,] = x; - 3x,x, - XlX'J + 5x;. EXAMPLE 2. LetB = X = [x1, x2]. Then we have XBXt = [xl,,x2][-; -3[::1 = x; -2x,x, -2x,x, + 5x2,. In both Examples 1 and 2 the two mixed product terms add up to -4x1x, so XAF =X&V.

These examples show that different matrices can lead to the same quadratic form.Note that one of these matrices is symmetric. This illustrates the next theorem. THEOREM 5.10. For any n x n matrix A and any 1 x n row matrix X we have XAXt =XBXt where B is the symmetric matrix B = +(A + At). Proof. Since XAXt is a 1 x 1 matrix it is equal to its

transpose, XAXt = (XAXt)t.But the transpose of a product is the product of transposes in reversed order, so we have(XAXt)t = XAtXt .

Therefore XAXt = &XAXt + *XAtXt = XBXt . 5.13 Reduction of a real quadratic form to a diagonal form A real symmetric matrix A is Hermitian. Therefore, by Theorem 5.7 it is similar to thediagonal matrix A = diag (jll, . . . , A,) of its eigenvalues. Moreover, we have A = CtAC,where C is an orthogonal matrix.

Now we show that C can be used to convert the quadraticform XAXt to a diagonal form. Reduction of a real quadratic form to a diagonal form 129 THEOREM 5.11. Let XAF be the quadratic form associated with a real symmetric matrixA, and let C be an orthogonal matrix that converts A to a diagonal matrix A = FAC.Then we have XAXt = YAYt =

i&y;,i=l where Y = [;vI, . ,I .

, y,] is the row matrix Y = XC, and AI, . . . ,I,, are the eigenvalues of A. Proof. Since C is orthogonal we have C-l = Ct. Therefore the equation Y = XCimplies X = YCt , and we obtain XAXt = (YCt)A(YCt)t = Y(CtAC) Yf = YA Yt . Note: Theorem 5.11 is described by saying that the linear transformation Y = XCreduces the quadratic form X,4X1 to a

diagonal form YA Yt. EXAMPLE 1. The quadratic form belonging to the identity matrix is XI&v =tglxf = I(Xl12, the square of the length of the vector X = (x1, . , , , x,) . A linear transformation Y = XC,where C is an orthogonal matrix, gives a new quadratic form YAYt with A = CZCt =CC?

= I. Since XZXt = YZYt we have I(X(12 = I(Y112, so Y has the same length as X. Alinear transformation which preserves the length of each vector is called an isometry.These transformations are discussed in more detail in Section 5.19. EXAMPLE 2.. Determine an orthogonal matrix C which reduces the quadratic form Q(x) =2x,2 + 4x,x, + 5x-i to a

diagonal form. 2 2Solution.

We write Q(x) = XAXt, where A = [12 5. This symmetric matrix was diagonalized in Example 1 following Theorem 5.7. It has the eigenvalues 1, = 1, AZ = 6,and an orthonormal set of eigenvectors ul, u2, where u1 = t(2, -l), u2 = t(l,2), t = l/J?. An orthogonal diagonalizing matrix is C = t2 1[1 The corresponding diagonal form is -1 2 . YAYt = il,y; + I,y;

= y; + 6-y;.

The result of Example 2 has a simple geometric interpretation, illustrated in Figure 5.1.The linear transformation Y =: XC can be regarded as a rotation which maps the basis i,j onto the new basis ul, u2. A point with coordinates (x1, xJ relative to the first basis hasnew coordinates (yl, yJ relative to the second basis. Since XAF = YAYt , the set ofpoints

(x1, x2) satisfying the equation XAXt = c for some c is identical with the set ofpoints (yl, y2) satisfying YA Yt = c . The second equation, written as y: + 6~: = c, is 130 Eigenvalues of operators acting on Euclidean spaces to to basis i,j basis u,, FIGURE 5.1 Rotation of axes by an orthogonal matrix. The ellipse has Cartesian equationXAXt = 9 in the jc,x,-

system, and equation YA Yt = 9 in the y&-system. the Cartesian equation of an ellipse if c > 0. Therefore the equation XL%? = c, writtenas 2x: + 4x,x, + 5x: = c, represents the same ellipse in the original coordinate system.Figure 5.1 shows the ellipse corresponding to c = 9. 5.14 Applications to analytic geometry The reduction of a quadratic form to

a diagonal form can be used to identify the set ofall points (x, v) in the pla.ne which satisfy a Cartesian equation of the form (5.10) ax2 + bxy + cy2 + dx + ey +f= 0.

We shall find that this set is always a conic section, that is, an ellipse, hyperbola, parabola,or one of the degenerate cases (the empty set, a single point, or one or two straight lines).The type of conic is governed by the second-degree terms, that is, by the quadratic formax2 + bxy + cy2. To conform with the notation used earlier, we write x1 for X, x2

for y,and express this quadratic form as a matrix product, XAXt = ax,2 + bx,x2 + cxi, where X = [x1, x2] and 11 = By a rotation Y = XC we reduce this form to a diagonal form L,y,2 + 3c2yi, where 1,) A2 are the eigenvalues of A. An orthonormal setof eigenvectors ul, u2 determines a new set of coordinate axes, relative to which the Cartesianequation

(5.10) becomes (5.11) 4Yf + i& + d'y, + e'y, + f = 0, with new coefficients d' and e' in the linear terms.

In this equation there is no mixed productterm yly2, so the type of conic is easily identified by examining the eigenvalues 1, and il, . Applications to analytic geometry 1 3 1 If the conic is not degenerate, Equation (5.11) represents an ellipse if 1,) 1, have the samesign, a hyperbola if &, 1, have opposite signs, and a parabola if either 31, or ;1, is zero..

Thethree cases correspond to ;Z,;i., > 0, I,& < 0, and J,J, = 0. We illustrate with somespecific examples. EXAMPLE 1. 2x" + 4xy + 5y2 + 4x + 13y - a = 0. We rewrite this as (5.12) 2~,2+4x~x~+5x~+4x~+13x~-~=O. The quadratic form 2x; -l- 4x,.x, + 5x: is the one treated in Example 2 of the foregoingsection. Its matrix has eigenvalues 1, = 1 , 1, =

6, and an orthonormal set of eigenvectors u1 = t(2, - I), u2 = t(l,2), where t = 1/45. An orthogonal diagonalizing matrix is2 1 C=t[I-1 2 * This reduces the quadratic part of (5.12) to the form yf + 6~:. To determine the effect on the linear part we write the equation of rotation Y = XC in theform X = YCt and obtain 2[Xl? x21 = 1- 1 J5 Lh ? Y21 [I1 2 9

x1= + Y2), x2 = +1 + 2Y2). Therefore th.e linear part 4x, + 13x, is transformed to $(2'.' + Y2) + $(-,I + 2~2) = -$y, + 643~~~. The transformed Cartesian equation becomes By completing the squares in y1 and y, we rewrite this as follows: (yl - bh2 + 6(y2 + $&)2 = 9. This is the (equation of an ellipse with its center at the point (f&,

-&) in the y1y2-system. The positive directions of the yr and y2 axes are determined by the eigenvectors urand u2, as indicated in Figure 5.2. We can simplify the equation further by writing I1==y1-&5, z2=y2+4J5. Geometrically, this is the same as introducing a new system of coordinate axes parallel tothe yly2 axes but with the new origin at the

center of the ellipse. In the z,=,-system theequation of the ellipse is simply 2:: + 6z", = 9, or The ellipse and all three coordinate systems are shown in Figure 5.2. 132 Eigenualues of operators acting on Euclidean spaces F IGURE 5.2 Rotation and translation of coordinate axes. The rotation Y = XC is followed by the translation z1 = y1 - 2lJS,z, =y2

+tJi. EXAMPLE 2. 2x2 - 4xy .- y2 - 4x + 1Oy - 13 = 0. We rewrite this as 2x:! - 4x,x, - x; -44x,+10x,-13=0. 2 -2The quadratic part is XAF, where A = [1-2 -1 . This matrix has the eigenvalues 1, = 3, il, = -2. An orthonormal set of eigenvectors is q = t(2, -l), q, = t(l, 2), where t = l/ JS. An orthogonal diagonalizing matrix is C = tof rotation X = YCt gves us

The equation Xl = -L (2YI + Y2>,J5 Xz=+y,+2y,).J5 Therefore the transformed equation becomes 3Y,2 - 2Yz" -~~(2Y~+Y2)+$(-YI+2y~)-13=0, or 3Y; - 2y; - 18 y, + 16 y, J5 Js --'=O' By completing the squares in y1 and yz we obtain the equation 3C.h - SJS)2 - 2(y2 - +Js>2 = 12, Applications to analytic geometry 133 (a) Hyperbola: 3~: - 2~; = 12 (b)

Parabola: y: + y, = 0 FIGURE 5.3 The curves in Examples 2 and 3. which represents a hyperbola with its center at ($JS, +J5) in theyly2-system. The trans-lation z 1 = y1 - a&, z2 = yz - +& simplifies this equation further to 2 23z,2 - 22; = 12, or 2 - 2 = 1. 4 6 The hyperbola is shown in Figure 5.3(a). The eigenvectors u1 and u2 determine the direc-tions

of the positive y1 and y2 axes. EXAMPLE 3.

9x2 + 24xy + 16y2 - 20x + 15~ = 0. We rewrite this as 9x," + 24x,x, + 16x; - 20x, + 15X, = 0. 9 12The symmetric matrix for the quadratic part is A = [112 16 * Its eigenvalues are 1, = 25, il2 = 0. An orthonormal set of eigenvectors is u 1 = $(3,4), u2 = +(-4,3). An orthog-3 - 4 onal diagonalizing matrix is C = $, [14 3'The equation of rotation X = YCt

gives us x1 = Wyl - 4y2), x2 = %(4yl + 3y2) I Therefore the transformed Cartesian equation becomes 25~; - a:(3y, - 4y,) + ':(4Y, + 3Y2) = O. This simplifies toy; + y2 = 0, the equation of a parabola with its vertex at the origin. Theparabola is shown in Figure 5.3(b). 134 Eigenvalues of operators acting on Euclidean spaces EXAMPLE 4. Degenerate

cases. A knowledge of the eigenvalues alone does not revealwhether the Cartesian equation represents a degenerate conic section. For example, thethree equations x2 + 2y2 = 1, x2 + 2y2 = 0, and x2 + 2y2 = - 1 all have the sameeigenvalues; the first represents a nondegenerate ellipse, the second is satisfied only by(x, y) = (0, 0)) and the third

represents the empty set. The last two can be regarded asdegenerate cases of the ellipse. The graph of the equationy2 = 0 is the x-axis. The equation y2 - 1 = 0 represents thetwo parallel lines y = 1 a.nd y = - 1 . These can be regarded as degenerate cases of theparabola. The equation x2 - 4y2 = 0 represents two intersecting lines since it is satisfiedif

either x - 2y = 0 or x + 2~ = 0. This can be regarded as a degenerate case of thehyperbola. However, if the Cartesian equation ax2 + bxy + cy2 + dx + ey + f = 0 representsa nondegenerate conic section, then the type of conic can be determined quite easily. Thecharacteristic polynomial of the matrix of the quadratic form ax2 + bxy + cy2 is d e t

[16/l yz]= A2 - (a + c)/l + (ac - abz) = (I - A,)(1 - A,). Therefore the product of .the eigenvalues is AlA = ac - $b2 =)(4ac - b2). Since the type of conic is determined by the algebraic sign of the product A,il, , we see thatthe conic is an ellipse, hyperbola, or parabola, according as 4ac - b2 is positive, negative, orzero. The number 4ac - b2 is called the

discriminant of the quadratic form ax2 + bxy +cy2 . In E,xamples 1, 2 and 3 the discriminant has the values 34, -24, and 0, respectively. 5.15 Exercises In each of Exercises 1 through 7, find (a) a symmetric matrix A for the quadratic form; (b) theeigenvalues of A; (c) an orthonormal set of eigenvectors; (d) an orthogonal diagonalizing matrix C. 1. 4X;

+4X,X, +X,". 5. Xi +X1X2 +X1X3 + X3X3.2. x1x2. 6. 2x; + 4x,x, + x; - x;.3. x; + 2x,x, - xi. 7.

3x; + 4x,x, + 8x,x, + 4x,x, + 3x;.4. 34x; - 24x,x, + 41x;. In each of Exercises 8 through 18, identify and make a sketch of the conic section represented bythe Cartesian equation. 8. y2 - 2xy+2x2-5=0.

14. 5x2 +6xy +5y2 -2 =O.9. y2-2xy+5x=o. 15.x2+2xy+y2-2x+2y+3=0. 10. y2-2xy+x2-5x=0. 16. 2x2 + 4xy + 5y2 - 2x - y - 4 = 0.11.5x2-4xy+2y2-6=:0. 17. x2 + 4xy - 2y2 - 12 = 0.12. 19x2 + 4x.y + 16y2 - 212x + 104~ = 356.13. 9x2 + 24xy + 16y2 - 5:2x + 14y = 6. 18. xy +y -2x -2 =O. 19. For what value (or values) of c will the graph of the Cartesian

equation 2xy - 4x + 7y + c =0 be a pair of lines ? 20. If the equation ax2 + buy + cy2 = 1 represents an ellipse, prove that the area of the region it___.bounds is 2~144ac - 6:".

This gives a geometric meaning to the discriminant 4ac - b2. Eigenvalues of a symmetric transformation obtained as values of its quadratic form 135 k 5.16t Eigenvalues of a symmetric transformation obtained as values of its quadratic form Now we drop the requirement that V be finite-dimensional and we find a relation betweenthe eigenvalues of a

symmetric operator and its quadratic form. Suppose x is an eigenvector with norm 1 belonging to an eigenvalue A.

Then T(x) = Axso we have (5.13) Q(x) = (T(x), x) = (Ax, x) = A(x, x) = 2, since (x, x) =I 1. The set of all x in V satisfying (x, x) = 1 is called the unit sphere in V.Equation (5.13) proves the following theorem. THEOREM 5.12. Let T: V + V be a symmetric transformation on a real Euclidean space V,and let Q(x) q = (T(x), x) . Then the eigenvalues of T

(ifany exist) are to be found among thevalues that Q takes on the unit sphere in V. EXAMPLE. L,et V = V,(R) with the usual basis (i,j) and the usual dot product as inner4 0 product. Let T be the symmetric transformation with matrix A = [1 Then thequadratic form of T is given by 0 8' Q(X) = i iaijxixj = 4x,2 + 8x:.i=l j=l The eigenvalues of,T are il, = 4,

1, = 8. It is easy to see that these eigenvalues are,respectively, the minimum and maximum values which Q takes on the unit circle xf +xi = 1 . In fact, on this circle we have Q(x) = 4(x; + x;) + 4x; = 4 + 4x;, where - 1 IX,&. This has its smallest value, 4, when x2 = 0 and its largest value, 8, when x2 = f 1 .Figure 5.4 shows the unit circle and two

ellipses. The inner ellipse has the Cartesian equation 4~; -+ 8x: = 4. It consists of all points x = (x1, x,) in the plane satisfyingQ(x) = 4. The outer ellipse has Cartesian equation 4x,2 + 8x: = 8 and consists of allpoints satisfying Q(x) = 8. The points (* I, 0) where the inner ellipse touches the unitcircle are eigenvectors belonging to the eigenvalue 4. The

points (0, fl) on the outerellipse are eigenvectors belonging to the eigenvalue 8. The foregoing example illustrates extremal properties of eigenvalues which hold moregenerally. In the next section we will prove that the smallest and largest eigenvalues (ifthey exist) are always the minimum and maximum values which Q takes on the unit sphere.Our

discussion of these extremal properties will make use of the following theorem onquadratic forms. It should be noted that this theorem does not require that V be finitedimensional. t Starred sections can be omitted or postponed without loss in continuity. 136 Eigenvalues of operators acting on Euclidean spaces- t X2 /Eigenvector belonging to 8 Q(x)

= 8 on this ellipse tXl Eigenvector belonging Q(x) = 4 on this ellipse to 4 FIG~JRE 5.4 Geometric relation between the eigenvalues of T and the values of Qon the unit sphere, illustrated with a two-dimensional example. THEORE:M 5.13. Let T: V + V be a symmetric transformation on a real Euclidean space Vwith quadratic form Q(x) = (T(x), x) .

Assume that Q does not change sign on V.

Then ifQ(x) = 0 for some x in V we also have T(x) = 0. In other words, ifQ does not change sign,then Q vanishes only on the null space of T. Proof. Assume Q(x) = 0 for some x in V and let y be any element in V.

Choose anyreal t and consider Q(x + ty).

Using linearity of T, linearity of the inner product, andsymmetry of T, we have Q = at + bt2, where a = 2(T(x), y) and b = Q(y). If Q is nonnegative on V we have the inequality at + bt2 > 0 for all real t . In other words, the quadratic polynomial p(t) = at + bt2 has its minimum at t = 0.Hence p'(0) = 0. But p'(0) = a = 2(T(x), y), so (T(x), y) = 0.

Since y was arbitrary,we can in particular take y = T(x), getting (T(x), T(x)) = 0.

This proves that T(x) = 0. If Q is nonpositive on V we get p(t) = at + bt2 < 0 for all t, sop has its maximum att = 0, and hencep'(0) =: 0 as before. * 5.17 Extremal properties of eigenvalues of a symmetric transformation Now we shall prove that the extreme values of a quadratic form on the unit sphere areeigenvalues'. THEOREM 5.14.

Let T: V--f V be a symmetric linear transformation on a real Euclideanspace V, and let Q(x) = (T(x), x) . Among all values that Q takes on the unit sphere, assume The finite-dimensional case 137 there is an exlremumt (maximum or minimum) at apoint u with (u, u) = 1 . Then uisaneigen-vector for T; the corresponding eigenvalue is Q(u), the extreme

value of Q on the unit sphere. Proof. Assume Q has a minimum at u. Then we have (5.14) Q(x) 2 Q(u) for all x with (x, x) = 1. Let jz = Q(u).

If (x, x) = 1 we have Q(u) = 1(x, x) = (lx, x) so inequality (5.14) can bewritten as (5.15) (T(x), xk 2 @x, 4 provided (x, ,x) = 1.

Now we prove that (5.15) is valid for all x in V. Suppose llxll = a.Then x = ay , where llyll = 1. Hence (T(x), x> = ( a.v> = a2(TCy),y) and (k 4 = a"<+, y>.

But (TCy), y> 2 (iiy, y) since (y, y> = 1. Multiplying both members of this inequality bya2 we get (5.15) for x = ay .

Since (T(x), x) - (Ax, x) = (T(x) - 1x, x , we can rewrite inequality (5.15) in the form)(T(x) - lx, x) 2 0, or (5.16) (S(x), 4 2 0, where S = T - 2.

When x = u 'we have equality in (5.14) and hence also in (5.16). The linear transformationS is symmetric. Inequality (5.16) states that the quadratic form Q1 given by Q,(x) =(S(x), x) is nonnegative on V. When x = u we have Q,(u) = 0. Therefore, by Theorem5.13 we must have S(U) = 0. In other words, T(u) = iiu, so u is an eigenvector for T, and1 =

Q(u) is the corresponding eigenvalue. This completes the proof if Q has a minimumat u. If there is a maximum at u all the inequalities in the foregoing proof are reversed and weapply Theorem 5.13 to the nonpositive quadratic form Q, . t 5.18 The finitedimensional case Suppose now that dim V = n .

Then T has n real eigenvalues which can be arranged inincreasing order, say n,138~_ Eigenvalues of operators acting on Euclidean spaces Let u1 be an eigenvector on the unit sphere which minimizes Q. Then 1, = Q(uJ . If ilis an eigenvalue different from 1, any eigenvector belonging to 1 must be orthogonal to ul.Therefore it is natural to search for

such an eigenvector on the orthogonal complement ofthe subspace spanned by ul. Let S be the subspace spanned by ul. The orthogonal complement S'- consists of allelements in V orthogonal to ul. In particular, S'- contains all eigenvectors belonging toeigenvalues il # 1,. It is easily verified that dim SI = n - 1 and that T maps Sl intoitself.? Let Snwl

denote the unit sphere in the (n - l)-dimensional subspace S'. (Theunit sphere S,-, is a subset of the unit sphere in V.) Applying Theorem 5.14 to the subspaceSl we find that 1, = Q(uJ , where u2 is a point which minimizes Q on S,-l. The next eigenvector il, can be obtained in a similar way as the minimum value of Q onthe unit sphere S,-Z in the (n - 2)-

dimensiontil space consisting of those elements orthog-onal to both u1 and u2. Continuing in this manner we find that each eigenvalue & is theminimum value which Q takes on a unit sphere Sn--k+1 in a subspace of dimension n - k + 1.The largest of these minima, A,, is also the maximum value which Q takes on each of thespheres Sn--k+l. The

corresponding set of eigenvectors ul, .

. . , u, form an orthonormalbasis for V. 5.19 Unitary transformations We conclude this chapter with a brief discussion of another important class of trans-formations known as unitary transformations. In the finite-dimensional case they arerepresented by unitary matrices. DEFINITION. Let E be a Euclidean space and V a subspace of E. A linear

transformationT: V + E is called unitary on V if we have (5.17) (T(x), T(y)) = (x, y) for all x and y in V.

When E is a real Euclidean space a unitary transformation is also called an orthogonaltransformation. Equation (5.17) is described by saying that T preserves inner products. Therefore it isnatural to expect that T also preserves orthogonality and norms, since these are derivedfrom the inner product. THEOREM 5.15. IfT: V + E is a unitary

transformation on V, then for all x and y in V wehave: (a) (x, y) = 0 implies (T(x), T(y)) = 0 (Tpreserves orthogonality). @I IIW)ll = llxll (Tpreserves norms).Cc> IIT(x) - T(y)/I = llx - yll (Tpreserves distances). .(d) T is invertible, and T-l is unitary on T(V). t This was done in the proof of Theorem 5.4, Section 5.6. Unitary transformations 139 Proof. Part (a)

follows at once from Equation (5.17).

Part (b) follows by taking x = yin (5.17). Part (c) follows from (b) because T(x) - T(y) = T(x - y).

To prove (d) we use (b) which shows that T(x) = 0 implies x = 0, so T is invertible.If x E T(V) and y E: T(V) we can write x = T(u), y = T(v), so we have (T-W> Wy>> = (u, u) = (T(u), T(v)) = (x,y). Therefore T-l is unitary on T(V). Regarding eigenvalues and eigenvectors we have the following theorem. THEOREM 5.116. Let T: V -+ E' be a unitary

transformation on V.(a) Zf T has an eigenvalue I, then 11) = 1 .(b) Zf x and y are eigenvectors belonging to distinct eigenvalues 1 and ~1, then x and y are orthogonal.(c) Zf V = E and dim V = n, and if V is a complex space, then there exist eigenvectors Ul,... , u, cLf T which form an orthonormal basis for V. The matrix of T relative tothis basis is the

diagonal matrix A = diag (Al, .

. . , A,), u>here 1, is the eigenvaluebelonging to uk . Proof. To prove (a), let x be an eigenvector belonging to il. Then x # 0 and T(x) = 31x.Taking y = x in Equation (5.17) we get (lx, Ax) = (x, x) or nqx, x) = (x, x). Since (x, x) > 0 and Ai = 1112, this implies 111 = 1.To prove (b), write T(x) = Lx, T(y) = ,uy and compute the inner product (T(x), T(y)) in

two ways. We have (T(x), T(Y)) = CGY)since T is unitary. We also have (WA T(y)> = Gk py) = 44x, y) since x and y are eigenvectors. Therefore 1,6(x, y) = (x, y) , so (x, y) = 0 unless l/i = 1 .

But AX = 1 by (a), so if we had A/i = 1 we would also have 21 = Ap ,x = p , A = ,u, whichcontradicts the assumption that A and ,U are distinct. Therefore Ap # 1 and (x, y) = 0. Part (c) is proved by induction on n in much the same way that we proved Theorem 5.4,the corresponding result for Hermitian operators. The only change required is in that

partof the proof which shows that T maps SL into itself, where P-={xIxEV, (x,u1)=0}. Here u1 is an eigenvector of T with eigenvalue &. From the equation T(u,) = i,u, we find Ul = a;T(u,) = XIT(ul) 140 Eigenvalues of operators acting on Euclidean spaces since A,& = 11,12 = 1 . Now choose any x in SL and note that Hence T(x) E SL if x E SL, so T

maps S' into itself. The rest of the proof is identical withthat of Theorem 5.4, so we shall not repeat the details. The next two theorems describe properties of unitary transformations on a finite-dimensional space. We give only a brief outline of the proofs. THEOKEM 5.17. Assume dim V = n and let E = (e,, . .

.

, e,) be a jixed basis for V.Then a linear transformation T: V + V is unitary if and only if (5.18) (T(eJ, T(e,)) = (ei , ej) for all i and j . In particular, ifE is orthonormal then T is unitary ifand only ifT maps E onto an orthonormalbasis. Sketch of proof. Write x = 2 xiei , y = c yjej . Then we have Now compare (x, y) with (T(x), T(y)). THEOREM 5.18. Assume

dim V = n and let (e, , . . . , e,) be an orthonormal basis for V.Let A = (aij) be the matrix representation of a linear transformation T: V + V relative tothis basis. Then T is unitary if and only ifA is unitary, that is, if and only if (5.19) A*A = I. Sketch of proof. Since (ei, eJ is the g-entry of the identity matrix, Equation (5.19)implies (5.20) Since A is the

matrix of T we have T(eJ = x;=, a,,e,, T(eJ = cFzt=, a,+e,, so Now compare this with (5.20) and use Theorem 5.17. Exercises 1 4 1 THEOREM 5.19. Every unitary matrix A has the following properties:(a) A is nonsingular and A-l = A*.(b) Each of At, /i, and A* is a unitary matrix.(c) The eigenvalues of A are complex numbers of absolute value 1.(d) ldet

Al == 1; if A is real, then det A = f 1 . The proof oif Theorem 5.19 is left as an exercise for the reader. 5.20 Exercises 1.

(a) Let T: I' --t I' be the transformation given by T(x) = cx, where c is a fixed scalar. Provethat T is unitary if and only if ICI = 1 .(b) If V is one-dimensional, prove that the only unitary transformations on V are those des-cribed in (a). In particular, if I/is a real one-dimensional space, there are only two orthogonaltransformanons, T(x) = x and T(x) = --x.

2.

Prove each of the following statements about a real orthogonal n x n matrix A.(a) If 1 is a real eigenvalue of A, then 1 = 1 or 1 = - 1 .(b) If I is a complex eigenvalue of A, then the complex conjugate 1 is also an eigenvalue of A.In other words, the nonreal eigenvalues of A occur in conjugate pairs.(c) If n is odd, then A has at least one real eigenvalue. 3.

Let V be a real Euclidean space of dimension n. An orthogonal transformation T: V + Vwith determinant 1 is called a rotation. If n is odd, prove that 1 is an eigenvalue for T.

Thisshows that every rotation of an odd-dimensional space has a fixed axis. [Hint: Use Exercise 2.1 4. Given a real orthogonal matrix A with -1 as an eigenvalue of multiplicity k. Prove that det A =(-1)". 5. If T is linear and norm-preserving, prove that T is unitary.6. If T: V - V is both unitary and Hermitian, prove that T2 = I.7. Let (e,, .

. . , e,) and (u,, . . . , u,) be two orthonormal bases for a Euclidean space V.

Prove that there is a unitary transformation T which maps one of these bases onto the other.8. Find a real (I such that the following matrix is unitary: 9.

If A is a skew-Hermitian matrix, prove that both Z - A and Z + A are nonsingular and that(Z - A)(Z + A)-l is unitary. 10. If A is a unitary matrix and if Z + A is nonsingular, prove that (I - A)(Z + A)-l is skew-Hermitian. 11. If A is Hermitian, prove that A - iZ is nonsingular and that (A - iZ)-l(A + U) is unitary.12. Prove that any unitary matrix can be

diagonalized by a unitary matrix.13. A square ma.trix is called normal if AA* = A*A . Determine which of the following types of matrices are normal.(a) Hermitian matrices. (d) Skew-symmetric matrices.(b) Skew-Hermitian matrices. (e) Unitary matrices.(c) Symmetric matrices. (f) Orthogonal matrices. 14. If A is a normal matrix (AA* = A*A) and if Cr is

a unitary matrix, prove that U*AU is normal. 6 LINEAR DIFFERENTIAL EQUATIONS 6.1 Historical introduction The history of differential equations began in the 17th century when Newton, Leibniz, andthe Bernoullis solved some simple differential equations of the first and second order arisingfrom problems in geometry and mechanics. These early

discoveries, beginning about 1690,seemed to suggest that the solutions of all differential equations based on geometric andphysical problems could be expressed in terms of the familiar elementary functions ofcalculus. Therefore, much of the early work was aimed at developing ingenious techniquesfor solving differential equations by elementary

means, that is to say, by addition, sub-traction, multiplication, division, composition, and integration, applied only a finitenumber of times to the familiar functions of calculus. Special methods such as separation of variables and the use of integrating factors weredevised more or less haphazardly before the end of the 17th century. During the

18thcentury, more systematic procedures were developed, primarily by Euler, Lagrange, andLaplace. It soon became apparent that relatively few differential equations could be solvedby elementary means. Little by little, mathematicians began to realize that it was hopelessto try to discover methods for solving all differential equations.

Instead, they found it morefruitful to ask whether or not a given differential equation has any solution at all and,when it has, to try to deduce properties of the solution from the differential equation itself.Within this framework, mathematicians began to think of differential equations as newsources of functions. An important phase in the theory

developed early in the 19th century, paralleling thegeneral trend toward a more rigorous approach to the calculus. In the 1820's, Cauchyobtained the first "existence theorem" for differential equations. He proved that everyfirst-order equation of the form Y' = fc? .v) has a solution whenever the right member, S(x,r), satisfies certain general

conditions.One important example is the Ricatti equation y' = P(x>y" + QWy + W-4, where P, Q, and R are given functions. Cauchy's work implies the existence of a solutionof the Ricatti equation in any open interval (-r, r) about the origin, provided P, Q, and 1 4 2 Review of results concerning linear equations of first and second orders 143 R have

power-series expansions in (-r, r).

In 1841 Joseph Liouville (1809-1882) showedthat in some cases this solution cannot be obtained by elementary means.

Experience has shown that it is difficult to obtain results of much generality aboutsolutions of differential equations, except for a few types. Among these are the so-calledlinear differential equations which occur in a great variety of scientific problems. Somesimple types were discussed in Volume I-linear equations of first order and linear equationsof

second order with constant coefficients. The next section gives a review of the principalresults concerning these equations. 6.2 Review of results concerning linear equations of first and second orders A linear differential equation of first order is one of the form (6.1) Y' + P(x>Y = Q(x) 7 where P and Q are given functions.

In Volume I we proved an existence-uniquenesstheorem for this equation (Theorem 8.3) which we restate here. THEOREM 6.1. Assume P and Q are continuous on an open interval J.

Choose any point ain J and let b be any real number. Then there is one and on4 one function y = f (x) whichsatisfies the difSere)rltiaI equation (6.1) and the initial condition f(a) = b .

This function isgiven by the explicit formula (6.2) f(x) = be-A(") + e-ddsaz Q(t)tict' dt , where A(x) = jz P(t) dt . Linear equations of second order are those of the form P,,(x)y" + P&)y' + P,(x)y = R(x). If the coefficients PO, PI, P, and the right-hand member R are continuous on some intervalJ, and if P,, is never zero on J, an existence theorem (discussed

in Section 6.5) guaranteesthat solutions always exist over the interval J. Nevertheless, there is no general formulaanalogous to (6.2) for expressing these solutions in terms of PO, P, , P,, and R. Thus, evenin this relatively simple generalization of (6.1), the theory is far from complete, except inspecial cases.

If the coefficients are constants and if R is zero, all the solutions can bedetermined explicitly in terms of polynomials, exponential and trigonometric functions bythe following theorem which was proved in Volume I (Theorem 8.7). THEOREM 6.2. Consider the dyerential equation (6.3) y"+ay'+by=O, 144~~ Linear differential equations where a and b

are given real constants. Let d = a2 - 4b. Then every solution of (6.3) on theinterval (- co, + 00) has the form (6.4) y = e-~r'2[c1u1(x) + c2u2(x)l, where c1 and c2 are constants, and the functions u1 and u2 are determined according to thealgebraic sign of d as follows: (a) If d = 0, then ul(x) = 1 and u2(x) = x . (b) If d > 0, then ul(x) = elcr and u2(x) =

eWkx, where k = &6.

_ (c) If d < 0, then ul(x) = cos kx and u2(x) = sin kx, where k = &f-d. The number d = a2 - 4b is the discriminant of the quadratic equation (6.5) r2+ar+b=0. This is called the characteristic equation of the differential equation (6.3). Its roots are given bY- a + Jd - a - Jcl rl = 2 , r2= 2 . The algebraic sign of d determines the nature of these roots. If d > 0

both roots are realand the solution in (6.4) can be expressed in the form y = cleT1" + c2erzx. If d < 0, the roots rl and r2 are conjugate complex numbers. Each of the complexexponential functions fi(x) = er12 and fJ,x) = e7zz is a complex solution of the differentialequation (6.3). We obtain real solutions by examining the real and imaginary parts

offiand f2. Writing rl = -ia + ik , r2 = -&a - ik, where k = &J-d, we have fi(x) =: erlx = e-ax12eikz = e-azI2cos kx + ie-arl'sin kx andf2(x) = e'?zx = e--nzfze--iks = e-a"'2cos kx _ ie-a"'"sin kx. The general solution appearing in Equation (6.4) is a linear combination of the real andimaginary parts of fi(x) and f2(x). 6.3 Exercises These exercises have been

selected from Chapter 8 in Volume I and are intended as a review ofthe introductory material on linear differential equations of first and second orders. Linear equations offirst order. In Exercises 1,2,3, solve the initial-value problem on the specifiedinterval. 1.y'-3y=e2Zon(-C0, +co),withy =Owhenx

=O.2.xy'-2y=x50n(0,+oo),withy=1whenx=1.3.y'+ytanx=sin2xon(-6r,$a),withy=2whenx=O. Linear deferential equations of order n 145 4. If a strain of bacteria grows at a rate proportional to the amount present and if the populationdoubles in one hour, by how much will it increase at the end of two hours? 5. A curve with Cartesian equation y =f(x)

passes through the origin. Lines drawn parallel tothe coordinate axes through an arbitary point of the curve form a rectangle with two sides onthe axes. The curve divides every such rectangle into two regions A and B, one of which has anarea equal to n times the other. Find the functionf.

6. (a) Let u be a nonzero solution of the second-order equation y" + P(x)/ + Q(x)y = 0.Show that the substitution y = uv converts the equation y" + P(x),' + Q(x,y = R(x) into a first-order linear equation for v' .(b) Obtain a nonzero solution of the equation y' - 4y' + x2(y' - 4y) = 0 by inspection anduse the method of part (a) to find a solution of yfl - 4y' +

x2 (y' - 4y) = 2xe-x3/3 suchthaty=Oandy'=4whenx=O.

Linear equations ofsecond order with constant coeflcients. In each of Exercises 7 through 10, findall solutions on (- lm, + m). 7. yn - 4 y = o .

9. yn -2y'+sy =o.8. y" + 4y = 0. 10. y" +2y'+y=o. 11. Find all values of the constant k such that the differential equation y" + ky = 0 has a nontrivialsolution y =f&) for which&(O) =fk(l) = 0. For each permissible value of k, determine thecorresponding solution y =fk(x) . Consider both positive and negative values of k. 12. If (a, b) is a given point in the

plane and if m is a given real number, prove that the differentialequation y" + k"y = 0 has exactly one solution whose graph passes through (a, b) and hasslope m there. Discuss separately the case k = 0. 13. In each case, find a linear differential equation of second order satisfied by ur and u2.(a> ul(x) = e" , u2(x) = e-".(b) ul(x) = e2x, u2(x) = xe2z.(c)

ul(x) = eP'2 cos x, u2(x) = eh2 sin x.(d) z+(x) = sin (2x + l), u2(x) = sin (2x + 2).(e) ul(x) = cash x, u2w = sinh x . 14. A particle undergoes simple harmonic motion.

Initially its displacement is 1, its velocity is 1and its acceleration is - 12. Compute its displacement and acceleration when the velocity is 4s. 6.4 Linear differential equations of order n A linear differential equation of order n is one of the form (6.6) P,(x)y'n) + P1(x)y'"-1' + 9 * * + P,(x)y = R(x) * The functions P,, , P, , . . . , P, multiplying the various

derivatives of the unknown functiony are called the coeficients of the equation. In our general discussion of the linear equationwe shall assume that all the coefficients are continuous on some interval J. The word"interval" will refer either to a bounded or to an unbounded interval. 146~_ Linear equations In the differential equation

(6.6) the leading coefficient P, plays a special role, since itdetermines the order of the equation. Points at which P,(x) = 0 are called singular pointsof the equation.

The presence of singular points sometimes introduces complications thatrequire special investigation. To avoid these difficulties we assume that the function P,is never zero on J. Then we can divide both sides of Equation (6.6) by PO and rewrite thedifferential equation in a form with leading coefficient 1. Therefore, in our general dis-cussion we shall

assume that the differential equation has the form (6.7) y@) + P,(x)y-l) + * * * + P,(x)y = R(x). The discussion of linear equations can be simplified by the use of operator notation. LetV(J) denote the linear space consisting of all real-valued functions continuous on an intervalJ. Let F"(J) denote the subspace consisting of all functions f whose first n

derivativesj-',f", . I. ,ftn) exist and are continuous on J. Let P,, . . . , P, be n given functions in%?(J) and consider the operator L: V"(J) -+ V(J) given by L(f) = f (n) + PJ'"-1' + * * * + PJ. The operator L itself is sometimes written as where Dk denotes the kth derivative operator. In operator notation the differential equationin (6.7) is written simply as 1

(6.8) L(y) = R. A solution of this equation is any function y in V'"(J) which satisfies (6.8) on the interval J.It is easy to verify that L(yl + yz) = L(yl) + L(y&, and that L(cy) = CL(~) for every constant c. Therefore L is a linear operator. This is why the equation L(y) = R is referredto as a linear equation. The operator L is called a linear d$erential operator

of order n. With each linear equation L(y) = R we may associate the equation L(y) = 0, in which the right-hand side has been replaced by zero.

This is called the homogeneousequation corresponding to L(y) = R. When R is not identically zero, the equation L(y) = Ris called a nonhomogeneous equation. We shall find that we can always solve the non-homogeneous equation whenever we can solve the corresponding homogeneous equation.Therefore, we begin our study with the homogeneous

case. The set of solutions of the homogeneous equation is the null space N(L) of the operatorL. This is also called the solution space of the equation. The solution space is a subspaceof V:"(J). Although GF?(.~) .is infinite-dimensional, it turns out that the solution space N(L)is always finite-dimensional. In fact, we shall prove that (6.9) dim N(L) = n, The

dimension of the solution space of a homogeneous linear equation 147 where n is the order of the operator L. Equation (6.9) is called the dimensionality theoremfor linear differential operators. The dimensionality theorem will be deduced as a conse-quence of an existence-uniqueness theorem which we discuss next. 6.5 The existence-uniqueness

theorem THEOREM 6.3. EXISTENCE-UNIQUENESS THEOREM FOR LINEAR EQUATIONS OF ORDER n. LetPl,P,, * * . , P, be continuous functions on an open interval J, alzd let L be the linear operator L=Dn+PIDn-l+...+P,. IfxOEJandifk,,k,,..., knpl are n given real numbers, then there exists one and only onefunction y = f(x) which

satisfies the homogeneous dlrerential equation L(y) = 0 on J andwhich also satisfies the initial conditions f(x,) = k, ,f'(x,) = kl, . . . ,f +l)(x,,) = k,-, . Note: The vector in n-space given by (f(x,),f'(x,), . . . ,f(n-l)(~O)) is called theinitial-value vector off at x0, Theorem 6.3 tells us that if we choose a point x0 in J andchoose a vector in n-space, then the

homogeneous equation L(y) = 0 has exactly onesolution y = f (x) on J with this vector as initial-value vector at x0. For example, whenn = 2 there is exactly one solution with prescribed value f(xo) and prescribed derivativef'(xo) at a prescribed point x0. The proof of the existence-uniqueness theorem will be obtained as a corollary of moregeneral

existence-uniqueness theorems discussed in Chapter 7. An alternate proof for thecase of equations with constant coefficients is given in Section 7.9. 6.6 The dimension of the solution space of a homogeneous linear equation THEOREM 6.4. DIMENSIONALITY THEOREM. Let L: F(J) + 9?(J) be a linear of order n given by (6.10) L = D"

+PIDn-' + *.* +P,. Then the solution space of the equation L(y) = 0 has dimension n. Proof. Let V, denote the n-dimensional linear space of n-tuples of scalars. Let T be thelinear transformation that maps each function f in the solution space N(L) onto the initial-value vector off at x0, T(f) = (f(x,,>>f 'Cd, .

. . J-'"-l'(~,>) 3 where x,, is a fixed point in J. The uniqueness theorem tells us that T(f) = 0 implies f = 0.Therefore, by Theorem 2.10, T is one-to-one on N(L). Hence T-l is also one-to-one andmaps V, onto N(L)!. and Theorem 2.11 shows that dim N(L) = dim V, = n . 148 Linear diferential equations Now that we know that the solution space has

dimension n, any set of n independentsolutions will serve as a ibasis. Therefore, as a corollary of the dimensionality theorem wehave : THEOREM 6.5. Let L: 'g"(J) -+ V(J) be a linear dtyerential operator of order n. If ul, . . .

,u, are n independent solutions of the homogeneous differential equation L(y) = 0 on J, thenevery solution y = f (x) on J can be expressed in the form (6.11) where cl, . .

. , c, are constants.

Note: Since all solutions of the differential equation L(y) = 0 are contained in formula(6.1 I), the linear combination on the right, with arbitrary constants cl, . . . , c,, issometimes called the general solution of the differential equation. The dimensionality theorem tells us that the solution space of a homogeneous lineardifferential equation of order n

always has a basis of n solutions, but it does not tell us howto determine such a basis. In fact, no simple method is known for determining a basis ofsolutions for every linear equation. However, special methods have been devised forspecial equations. Among these are differential equations with constant coefficients towhich we turn now. 6.7 The

algebra of constant-coefficient operators A constant-coefficient operator A is a linear operator of the form (6.12) .4 = a,,Dn + aIDn-l + * * * + a,-,D + a,, where D is the derivative operator and a,, a,, . . . , a, are real constants. If a, # 0 theoperator is said to have order n. The operator A can be applied to any function y with nderivatives on some interval,

the result being a function A(y) given by A(y) = a,,y(") + aIy(+l) + * * * + a,-,y' + any. In this section, we restrict our attention to functions having derivatives of every order on(- co, + co). The set of all such functions will be denoted by 59" and will be referred to asthe class of injnitely di#erentiable functions. If y E ?Zrn then A(y) is also in Ym. The

usual algebraic operations on linear transformations (addition, ,multiplication byscalars, and composition or multiplication) can be applied, in particular, to constant-coefficient operators. Let A and B be two constant-coefficient operators (not necessarilyof the same order). Since the sum A + B and all scalar ,multiples IA are also constant-coefficient

operators, the set of all constant-coefficient operators is a linear space. Theproduct of A and B (in either order) is also a constant-coefficient operator. Therefore,sums, products, and scalar multiples of constant-coefficient operators satisfy the usualcommutative, associative, and distributive laws satisfied by all linear transformations.Also, since we

have DrDs = DSDT for all positive integers r and s, any two constant-coefficient operators commute; AB = BA . The algebra of constant-coejicient operators 149 With each constant-coefficient operator A we associate a polynomial pA4 called thecharacteristic polynomial of A. If A is given by (6.12),p, is that polynomial which has thesame coefficients as

A.

That is, for every real r we have PA(r) = a/ + aIF + * * * + a,.

Conversely, given any real polynomial p, there is a corresponding operator A whosecoefficients are the same as those ofp. The next theorem shows that this associationbetween operators and polynomials is a one-to-one correspondence. Moreover, thiscorrespondence associates with sums, products, and scalar multiples of operators therespective

sums, products, and scalar multiples of their characteristic polynomials. THEOREM 6.6. Let A and B denote constant-coeficient operators with characteristicpolynomials pA and pB , respectively, and let 1 be a real number. Then ~'e have: (C) PAB~~~:BY (4 PAA ' * Proof. We consider part (a) first. AssumepA = pB . We wish to prove that A(J) = B(y)for

every y in v". Since pA = pB, both polynomials have the same degree and the samecoefficients. Therefore A and B have the same order and the same coefficients, so A(y) =B(y) for every y in V". Next we prove that A = B impliesp, = pB . The relation A = B means that A(y) = B(y)for every y in %?a.

Take y = er", where r is a constant. Since y(l;) = rkerr for every k 2 0,we have 4.~9 = pA(de'" and B(y) = pB(r)e". The equation A(y) = B(j) implies pA4(r) = pB(r). S ince r is arbitrary we must havepA = pB . This completes the proof of part (a). Parts (b), (c), and (d) follow at once from the definition of the characteristic polynomial.

From Theorem 6.6 it follows that every algebraic relation involving sums, products, andscalar multiples of polynomials pA4 and pB also holds for the operators A and B. Inparticular, if the characteristic polynomial pAI can be factored as a product of two or morepolynomials, each factor must be the characteristic polynomial of some constant-

coefficientoperator, so, by Theorem 6.6, there is a corresponding factorization of the operator A.For example, ifp,(r) = pB(r)pc(r), then A = BC. Ifpe4(r) can be factored as a product ofn linear factors, say (6.13) h(r) = a& - rl)@ - r2) - * * (r - IT,), the corresponding factorization of A takes the form A = a,(D - r,)(L) - r2) * * * (D - rJ. 150 Linear dtyerential

equations The fundamental theorem of algebra tells us that every polynomial PA(r) of degreen 2 1 has a factorization of the form (6.13), where rl, r2, . . .

, rn are the roots of the equa-tion, PAr) = 0, called the characteristic equation of A. Each root is written as often as its multiplicityindicates. The roots may be real or complex. SincepA has real coefficients, the complexroots occur in conjugate pairs, tl + i/3, u - i/I, if fl Z 0. The two linear factorscorresponding to each such pair can be combined to give

one quadratic factor r2 - 2ur +cc2 + ,P whose coefficients are real. Therefore, every polynomial p,(r) can be factored asa product of linear and quadratic polynomials with real coejicients. This gives a corre-sponding factorization of the operator A as a product of first-order and second-orderconstant-coefficient operators with real coefficients.

EXAMPLE I. Let A = D2 - 5D + 6. Since the characteristic polynomial pa(r) has thefactorization r2 - 5r + 6 = (r - 2)(r - 3), the operator A has the factorization D2 - 5D + 6 = (D - 2)(D - 3). EXAMPLE 2.

Let A = D4 - 2D3 + 2D2 - 20 + 1. The characteristic polynomial PA(r)has the factorization r4 - 2r3 + 2r2 - 2r + 1 = (r - l)(r - l)(r2 + I), so A has the factorizatia'n A = (D - l)(D - 1)(D2 + 1). 6.8 Determination of a basis of solutions for linear equations with constant coefficientsby factorization of operators The next theorem shows how factorization of

constant-coefficient operators helps us tosolve linear differential equations with constant coefficients. THEOREM 6.1. Let L be a constant-coejicient operator which can be factored as a productof constant-coefJicient operators, say L = A,A2*.*A,. Then the solution space qf the linear difSerentia1 equation L(y) = 0 contains the solution spaceof each

dtyerential equation Ai = 0. In other words, (6.14) N(AJc N(L) foreachi= 1,2 ,..., k. Proof. If u is in the null space of the last factor A, we have A,(u) = 0 so L(u) = &A, * * * A,)(u) = (A, . * * A&A,(u) = (A, * * * AkJ(0) = 0. Determination of a basis of solutions for linear equations 1 5 1 Therefore the null space of L contains the null space of the last factor

A,.

But sinceconstant-coefficient operators commute, we can rearrange the factors so that any one ofthem is the last factor. This proves (6.14). If L(u) = 0, the operator L is said to annihilate U. Theorem 6.7 tells us that if a factor Aiof L annihilates u:, then L also annihilates U. We illustrate how the theorem can be used to solve homogeneous differential

equationswith constant coefficients. We have chosen examples to illustrate different features,depending on the nature of the roots of the characteristic equation. CASE I. Real distinct roots. EXAMPLE 1. Find a basis of solutions for the differential equation (6.15) (D3 - 70+6)y=O. Solution. This lhas the form L(y) = 0 with L =D3-7D+6=(D-l)(D-2)(0+3).

The null space of 11 - 1 contains ur(x) = t?, that of D - 2 contains uZ(x) = ezZ, and thatof D + 3 contains u3(x) = e-3x. In Chapter 1 (p. 10) we proved that u1 , u2, us are inde-pendent. Since three independent solutions of a third order equation form a basis for thesolution space, the general solution of (6.15) is given by Y = cleZ + c2e2" + c3ew3'. The

method used to solve Example 1 enables us to find a basis for the solution space ofany constant-coefficient operator that can be factored into distinct linear factors.

THEOREM 6.8. Let L be a constant coeflcient operator whose characteristic equationpL(r) = 0 has n distinct real roots rI , r2, . . . , r, . Then the general solution of the difSerentia1equation L(y) = 0 on the interval (- co, + KI) is given by the formula (6.16) y = j$ ckerk" .k=l Proof. We have the factorization L = a,(D - rI)(D - r2).

* * (D - r,).

Since the null space of (D - rJ contains uk(x) = erkr, the null space of L contains the nfunctions (6.17) U1(x) = f?, uz(X) = er@, . . . , u,(x) = eTnr, 152 Linear equations In Chapter 1 (p. 10) we proved that these functions are independent. Therefore they forma basis for the solution space of the equation L(y) = 0, so the general solution

is given by(6.16). CASE II. Real roots, some of which are repeated.If all the roots are real but not distinct, the functions in (6.17) are not independent and therefore do not form a basis for the solution space. If a root r occurs with multiplicity m,then (D - r)"' is a factor of L. The next theorem shows how to obtain m independentsolutions in the null

space of this factor. THEOREM 6.9. The m jimctions ul(x) = e'", u2(x) = xerr, . . . , u,(x) = xm-%?* are m independent elements annihilated by the operator (D - r)" . Proof. The independence of these functions follows from the independence of thepolynomials 1, x, x2, . .

. , x+-l. To prove that ur, u2, . . . , u, are annihilated by (D - r)"we use induction on m. If m = 1 there is only one function, ul(x) = eF2, which is clearly annihilated by (D - r) .Suppose, then, that the theorem is true for m - 1. This means that the functions ul, . . . ,u,-~ are annihilated by (D - r)m-l. Since (D - r)" = (D - r)(D - r)+' the functions u1 , . . . , u,-~

are also annihilated by (D - r)" . To complete the proof wemust show that (D - r)m annihilates u, .

Therefore we consider We have(D - r)%, = (D - r)m-l(D - r)(xm-1e'3c). (D - r)(x"-le'") = D(x"-lp) - rxm-leTz = (m - 1)xm--2p + - rx~-lp = (m - 1)xm-2erz = (m - l)u,-,(x).

When we apply (D - r:Y'-' to both members of this last equation we get 0 on the rightsince (D - r)m-l annihilates u,-r . Hence (D - r)%, = 0 so u, is annihilated by(D - r)".

This completes the proof. EXAMPLE 2. Find the general solution of the differential equation L(y) = 0, whereL=D3-D2-8D+12. Solution. The operator L has the factorization L = (D - 2)2(D + 3). Determination of a basis of solutions for linear equations By Theorem 6.9, the two functions 153 q(x) = e2*) z&(x) = xe2= are in the null space of (D - 2)2. The

function us(x) = e+ is in the null space of (D + 3).Since ul, u2, ua are independent (see Exercise 17 of Section 6.9) they form a basis for thenull space of L, so the general solution of the differential equation is y = cle2s + c2xe2x + c3ee3S. Theorem 6.9 tells us how to find a basis of solutions for any nth order linear equationwith constant coefficients

whose characteristic equation has only real roots, some of whichare repeated. If the distinct roots are rl, r2, . . . , r, and if they occur with respectivemultiplicities m, , m2, . . . , mk, that part of the basis corresponding to rs is given by the m,functions u,,,(x) = x*-leQr, where q= I,2 ,..., mM,. Asptakesthevalues1,2,...,kwegetm,+*** + mk functions

altogether.

In Exercise17 of Section 6.9 we outline a proof showing that all these functions are independent. Sincethe sum of the multiplicities m, + * * * + m, is equal to n, the order of the equation, thefunctions uB,Q form a basis for the solution space of the equation. EXAMPLE 3. Solve the equation (D6 + 2D5 - 2D3 - D2)y = 0.

Solution.

We have D6 + 2D5 - 2D3 - D2 = D2(D - l)(D + I)". The part of thebasis corresponding to the factor D2 is ul(x) = 1, u2(x) = x ; the part corresponding tothe factor (D - 1) is u3(x) = e" ; and the part corresponding to the factor (D + 1)" isu4(x) = e-", us(x) = xe-" , Us = x2ePr. The six functions ul, . . . , us are independentso the general solution of the

equation is y = cl + c2x + c3e5 + (c4 + c,x + c6x2)emz. CASE III.

Complex roots.If complex exponentials are used, there is no need to distinguish between real and complex roots of the characteristic equation of the differential equation L(y) = 0. If real-valuedsolutions are desired, we factor the operator L into linear and quadratic factors with realcoefficients. Each pair of conjugate complex roots u + i,!?, u - i/3

corresponds to aquadratic factor, (6.18) D2-2uD+cc2+/12. The null space of this second-order operator contains the two independent functionsU(X) = eaz cos px and v(x) = ear sin px. If the pair of roots dc f i/3 occurs with multi-plicity m, the quadratic factor occurs to the mth power. The null space of the operator [D2 - 2uD + u2 + p21rn 154 Linear equations contains 2m independent functions, u,(x) = X*-lezz cos px) u*(x) = x*--lea2 sin /3x, q=l,2 ,..., m. These facts can be easily proved by induction on m. (Proofs are outlined in Exercise 20of Section 6.9.) The following examples illustrate some of the possibilities. EXAMPLE 4. y"I - 4y" + l3y' = 0.
The characteristic equation, r3 - 49 + 13r = 0 , has the roots 0, 2 f 3i; the general solution is y = cl + ezx(c2 cos 3x + c3 sin 3x). EXAMPLE 5. y"I - 2~" + 4;~' - 8y = 0 . The characteristic equation is r3 - 2r2 + 4r - 8 = (r - 2)(r2 + 4) = 0 ; its roots are 2, 2i, -2i, so the general solution of the differential equation is y = c1e2" + c2 cos 2x + c3 sin 2x.
EXAMPLE 6. y (5) - gyt41 $. 34y" _ 66~" + 65~' - 25~ = 0. The characteristic equationcan be written as (r - l)(r2 - 4r + 5)2 = 0 ; its roots are 1, 2 f i, 2 f I;, so the general solution of the differential equation is y = c,e" + e2z[(c2 + c3x) cos x + (c4 + c5x) sin x] . 6.9 Exercises Find the general solution of each of the differential equations in Exercises 1 through 12. 1 . y"' -2y" -3y'=o.
7 . yt4' + 16y = 0.2 . y"' -y' =o. 8. y"' - y = o .3. y#' + 4y" + 4y' = 0.
-3yN+3y'-y=o.9. yc4) + 4y" + 8~" + 8y' + 4y = 0. 4 . y"'5. yt4' + 4 ~ " ' + 6~" + 4y' -t- y = 0. 10. y(4) + 2y" + y = 0.11. y(6) + 4y(4) + 4y" = 0.
6. Y t4) - 16y = 0. 12. y@) + 8yt4) + 16~" = 0. 13. If m is a positive constant, find that particular solution y =f(x) of the differential equation y"' - my" + m2y' - &y = 0 which satisfies the conditionf(0) =f'(O) = O,f"(O) = 1 .14. A linear differential equation with constant coefficients has characteristic equationf(r) = 0. If all the roots of the characteristic equation are negative, prove that every solution of the differ-ential equation approaches zero as x + + to. What can you conclude about the behavior of allsolutions on the interval [0, + co) if all the roots of the characteristic equation are nonpositive? Exercises 155 1.5. In each case, find a linear differential equation with constant coefficients satisfied by all thegiven functions.(4 ~~(4 = e5, uz(x) = .e-, u,(x) = e2z, u4(x) = eC2x. 16. 17. 18. (b) ur(x) = e-2x, u2(x) = xe-2z, u&x) = x2eC2".(cl Ul(X) = 1 3 u2(4 = x, z+(x) = e5, u4(x) = xex. (d) ur(x) = zc, u2(x) = e", u2(x) = xe" .(e) ul(x) = x2, u2(x) = e", u2(x) = xe5.(f) ur(x) = e-as cos 3x, u2(x) = e-2r sin 3x, u2(x) = e-2x, q(x) = xe-&.(g) ul(x) = cash x, u2(x) = sinh x, u2(x) = x cash x, u4(x) = x sinh x .(h) ur(x) = cash x sin x, u2(x) = sinh x cos x, us(x) = x.Letr,,..., rn be n distinct real numbers, and let Q, , . . . , Q, be n polynomials, none of whichis the zero polynomial. Prove that the n functions ul(x) = Ql(x)e'15, . . . , u,(x) = Qn(x)er@ are independent.Outline of pro06 Use induction on n.
For n = 1 and n = 2 the result is easily verified. Assume the :statement is true for n =p and let cr, . . . , cD.+r be p + 1 real scalars such that Multiply both sides by e-'p+l" and differentiate the resulting equation. Then use the inductionhypothesis to show that all the scalars ck are 0. An alternate proof can be given based on orderof magnitude as x -+ + CC, as was done in Example 7 of Section 1.7 (p. 10).Letm,,m,,,.., mk be k positive integers, let rl, r2, . . . , rk be k distinct real numbers, andletn =m, t,... +m,. For each pair of integersp, q satisfying 1 < p < k, 1 5 q 5 mz,, let U,,,(x) = x*-V+.
For example, when p = 1 the corresponding functions are u~,~(x) = erlz, uz,l(x) = xerlz, . . . , u,l,,(x) = x"l-'e'lz. Prove that the n functions u*,~ so defined are independent. [Hint: Use Exercise 16.1Let L be a constant-coefficient linear differential operator of order n with characteristic poly-nomialp(r). Let L' be the constant-coefficient operator whose characteristic polynomial is thederivative p'(r). For example, if L = 20 2 - 30 + 1 then L' = 40 - 3. More generally,define the mth derivative L(?") to be the operator whose characteristic polynomial is the mthderivative p c'n)(r).
(The operator L (nl) should not be confused with the mth power Lm.)(a) If u has ,n derivatives, prove that k=O (b) If u has n - m derivatives, prove that n--m p+"'(())L(m)(u) = 2 7 u (8) for m=0,1,2,:..,n, k=O where L(O) == L.
156 Linear equations 19. Refer to the notation of E.xercise 18. If u and u have n derivatives, prove that k=O [Hint: Use Exercise 18 along with Leibniz's formula for the kth derivative of a product: (m)(k) =r~o(:,u(k-w~) .] 20. (a) I&p(t) = q(t)?(t), where q and r are polynomials and m is a positive integer.
Prove thatp'(t) = q(t)"-is(t), where s is a polynomial.(b) Let 1, be a constant-coefficient operator which annihilates u, where u is a given function ofx . LetM=Lm,the mth power of L, where m > 1. Prove that each of the derivatives M',M", . . .) M("-l) also annihilates u.(c) Use part (b) and Exercise 19 to prove that M annihilates each of the functions u, xu, . .
. ,xm-Ill.(d) Use part (c) to show that the operator (D2 - 2ctD + a2 + p2)" annihilates each of thefunctions xQe"z sin px and x*eaz cos px for q = 1, 2, . . . , m - 1 . 21. Let L be a constant-coefficient operator of order n with characteristic polynomial p(r). If c(is constant and if u has n derivatives, prove that L(e"%(x)) = eazc . n p? U(k)(X). k=O 6.10 The relation between the homogeneous and nonhomogeneous equations We return now to the general linear differential equation of order n with coefficients thatare not necessarily constant. The next theorem describes the relation between solutions ofa homogeneous equation L(;v) = 0 and those of a nonhomogeneous equation L(y) = R(x). THEOREM 6.10. Let L 9*(J) * V(J) be a linear difSerentia1 operator of order n. Let ul, . . . ,u, be n independent solutions of the homogeneous equation L(y) = 0, and let yl be a particularsolution of the nonhomogeneous equation L(y) = R, where R E V(J). Then every solutiony = f (x) of the nonhomogeneous equation has the form (6.19) .I-@> = Yl(X> +~ykw where cl, . . .
, c, are constants. Proof. By linearity we have L(f - yr) = L(f) - L(yr) = R - R = 0. Therefore f - y1 is in the solution space of the homogeneous equation L(y) = 0, so f - y1 is a linearcombination of ul, . . . , u, , sayf - y, = clul + . . - + c,u, .
This proves (6.19). Since all solutions of L(y) = R are found in (6.19), the sum on the right of (6.19) (witharbitrary constants cl, c2, . . . , c,) is called the general solution of the nonhomogeneous Determination of a particular solution of the nonhomogeneous equation 1 5 7 equation. Theorem 6.10 states that the general solution of the nonhomogeneous equationis obtained by adding toy, the general solution of the homogeneous equation. Note: Theorem 6.10 has a simple geometric analogy which helps give an insight intoits meaning. To determine all points on a plane we find a particular point on the planeand add to it all points on the parallel plane through the origin.
To find all solutions ofL(y) = R we find a particular solution and add to it all solutions of the homogeneousequation L(y) = 0. The set of solutions of the nonhomogeneous equation is analogousto a plane through a particular point. The solution space of the homogeneous equationis analogous to a parallel plane through the origin. To use Theorem 6.10 in practice we must solve two problems: (1) Find the generalsolution of the homogeneous equation L(y) = 0, and (2) find a particular solution of thenonhomogeneous equation L(y) = R.
In the next section we show that we can alwayssolve problem (2) if we can solve problem (1). 6.11 Determination of a particular solution of the nonhomogeneous equation. The methodof variation of parameters We turn now to the problem of determining one particular solution yr of the nonhomo-geneous equation L(y) = R . We shall describe a method known as variation ofparameterswhich tells us how to determine y1 if we know n independent solutions ul, . . . , u, of thehomogeneous equation L(y) = 0. The method provides a particular solution of the form (6.20) y1= VI%+ --- + v,u,, wherev,, . . . , v, are functions that can be calculated in terms of u1 , . . . , u, and the right-hand member R. The method leads to a system of n linear algebraic equations satisfied bythe derivatives vi, .
. . , v:. This system can always be solved because it has a nonsingularcoefficient matrix. Integration of the derivatives then gives the required functions vi, . .
.
,v, . The method was first used by Johann Bernoulli to solve linear equations of first order,and then by Lagrange in 1774 to solve linear equations of second order.
For the nth order case the details can be simplified by using vector and matrix notation.The right-hand member of (6.20) can be written as an inner product, (6.21) Yl = (0, 4, where v and u are n-dimensional vector functions given by v = (Vl, - * * 9 %I, u = (Ul,. . .) UJ. We try to choose v so that the inner product defining y1 will satisfy the nonhomogeneousequation L(y) = R, given that L(u) = 0, where L(u) = (L(u,), . . . , L(u,)). We begin by calculating the first derivative of yr . We find (6.22) y; = (0, u') + (u', u). 158 Linear equations We have n functions vl, . . . , v, to determine, so we should be able to put n conditions onthem. If we impose the condition that the second term on the right of (6.22) should vanish,the formula for yi simplifies to y; =: (v, U')) provided that (v', u) = 0. Differentiating the relation for y; we find yf = (v, 24") + (v', u') . If we can choose v so that (Iv', u') = 0 then the formula for yy also simplifies and we get y; = (v, u")) provided that also (v', u') = 0. If we continue in this manner for the first n - 1 derivatives of y1 we find ,:,-1) = (v, &L-l))) provided that also (v', u(+')) = 0. So far we have put n - 1 conditions on v. Differentiating once more we get y:"' = (v, ZP) + (v', ZP-1)). This time we impose the condition (v', u @-l)) = R(x), and the last equation becomes y:"' = (v, P) + R(X)) provided that also (v', u(~-')) = R(x). Suppose, for the moment, that we can satisfy the n conditions imposed on v. Let L =D" + P,(x)D"-l + * . . + P,(x). When we apply L to y1 we find I&) = y:"' + P,(x)Jpl' + . . .
+ P,(X>Yl = {(u, U(n:') + R(x)} + Pl(X)(U, LP-l)) + . . . + P,(x)(u, u) = (v, L(u)) + R(x) = (v, 0) + R(x) = R(x). Thus L(yJ = R(x), so y1 is a solution of the nonhomogeous equation.The method will succeed if we can satisfy the n conditions we have imposed on v. These conditions state that (v', u("') = 0 for k = 0, 1, . . . , n - 2, and that (v', zJn-l)) = R(x).We can write these n equations as a single matrix equation, - - (6.23) W(x)v'(x) = R(x) 0 'i ' II1 Determination of a particular solution of the nonhomogeneous equation 159 where v'(x) is regarded as an n x 1 column matrix, and where W is the n x n matrixfunction whose rows consist of the components of u and its successive derivatives: w= % l.lz ..* un4 4
... 11; . . (n-1)-Ul (n-1)uz . .
..IU(n-l)n The matrix W is called the Wronskian matrix of ul, . . .
, u,, after J. M. H. Wronski(1778-1853). In the next section we shall prove that the Wronskian matrix is nonsingular. Thereforewe can multiply both sides of (6.23) by W(x)-l to obtain 0u'(x) = R(x)W(x)-l 0 .II1Choose two points c and x in the interval J under consideration and integrate this vectorequation over the interval from c to x to obtain u(x) = u(c) + where The formula y1 = (u, v) for the particular solution now becomes y1 = (4 v) = (%0(C) + 2) = (t&u(c)) + (u,z).
The first term (u, v(c)) satisfies the homogeneous equation since it is a linear combination of241,. . .) u,. Therefore we can omit this term and use the second term (u, z) as a particularsolution of the nonhomogeneous equation. In other words, a particular solution of 160 Linear dtrerential equations L(y) = R is given by the inner product 0 (u(x), z(x)> = (u(x),lZR(t)W(t)-' ; dt) . ;I1 Note that it is not necessary that the function R be continuous on the interval J. All that isrequired is that R be integrable on [c, x]. We can summarize the results of this section by the following theorem. THEOREM 6.11. Let Ml,...
, u, be n independent solutions of the homogeneous nth orderlinear dtxerential equation L,(y) = 0 on an interval J. Then a particular solution y1 of thenonhomogeneous equation L(y) = R is given by the formula where vl, .
. . , v, are the entries in the n x 1 column matrix v determined-by the equation (6.24) In this formula, W is the Wronskian matrix of ul, . . . , u,, and c is any point in J. Note: The definite integral in (6.24) can be replaced by any indefinite integral 0sR(x)W(x)-' ; dx.Ij1EXAMPLE 1. Find the general solution of the differential equation))" - y = 21 + e" on the interval (- co, + co). Solution. The homogeneous equation, (D2 - 1)y = 0 has the two independent solutionsul(x) = eZ, uZ(x) = e-" . The Wronskian matrix of u1 and u2 is ez e-zW(x) = [1e" -e-" * Nonsingularity of the Wronskian matrix 161 Since det W(x) = -2, the matrix is nonsingular and its inverse is given by Therefore and we have 0 R(x)W(x)-' [1 1 21= - i 1+ [-eP 1 =e"e --z 1 1. 1 + er -e"Ll + eZ1 Integrating each component of the vector on the right we find s -22q(x) = AL- dx =1 + e" f (e-" - 1 + ?ff--1 + e" dx = -ee-" - x + log(1 + ez) and ~z(X) = sL dx = -log (1 + eZ).1 + e" Therefore the general solution of the differential equation is y = Cl%(X) + wzc4 + s(x)4W + %(X>&)= cleX + c,e-" - 1 - xez + (ez - e-2) log (1 + e") . 6.12 Nonsingularity of the Wronskian matrix uf n independent solutions of a homogeneouslinear equation In this section we prove that the Wronsklan matrix W of n independent solutions4,.
. . 7 u, of a homogeneous equation L(y) = 0 is nonsingular. We do this by provingthat the determinant of W is an exponential function which is never zero on the interval Junder consideration. Let w(x) = det W(x) for each x in J, and assume that the differential equation satisfiedbyu,,..., u, has the form (6.25) y(n) + P1(x)y(-') + * * * + P,(x)y = 0. Then we have: THEOREM 6.12. The Wronskian determinant satisfies the @St-order deferential equation (6.26) w' + Pl(X)W = 0 162 Linear equations on J. Therefore, if c E J we have (6.27) w(x) = w(c) exp [-6 PI(t) dt] Moreover, w(x) # 0 for all x in J. (Abel'sformula). Proof. Let u be the row-vector u = (ul, . . . , UJ . Since each component of u satisfiesthe differential equation (6.25) the same is true of u. The rows of the Wronskian matrix Ware the vectors u, u', . . . , utn-l) .
Hence we can write w = det W = det (u, u', . . . , ~(~-l)), The derivative of w is the determinant of the matrix obtained by differentiating the last rowof W (see Exercise 8 of Section 3.17).
That is w' .= det (u, u', . . . , I&~-~), ucn)). Multiplying the last row of w by PI(x) we also have P,(x:)w = det (u, u', . . . , z&"-~), P~(x)u("-~)). Adding these last two equations we find w' + Pl(x)w = det (u, u', . . . , u(+~), U(~) + Pl(x)u'"-'9. But the rows of this last determinant are dependent since u satisfies the differential equation(6.25). Therefore the determinant is zero, which means that w satisfies (6.26). Solving(6.26) we obtain Abel's formula (6.27). Next we prove that w(c) # 0 for some c in J. We do this by a contradiction argument.Suppose that w(t) = 0 for all t in J. Choose a fixed t in J, say t = t,, and consider thelinear system of algebraic equations w(t,)x = 0 , where X is a column vector.
Since det W(t,) = 0, the matrix W(t,) is singular so thissystem has a nonzero solution, say X = (cl, . . .
, c,) Z (0, . . . , 0) . Using the componentsof this nonzero vector, let f be the linear combination f(t) = c,u,(t) + * * * + c&&(t). The function f so defined satisfies L(f) = 0 on J since it is a linear combination ofUl,...,U,. The matrix equation W(t,)X = 0 implies that f(t,) = f '(t,) = * * * = f (n-l)(tJ = 0. Therefore f has the initial-value vector 0 at t = t, so, by the uniqueness theorem, f is thezero solution. This means c1 = * * * = c, = 0, which is a contradiction. Thereforew(t) # 0 for some t in J. Taking c to be this t in Abel's formula we see that w(x) # 0 forall x in J. This completes the proof of Theorem 6.12. The annihilator method for determining a particular solution 163 6.13 Special methods for determining a particular solution of the nonhomogeneous equation.Reduction to a system of first-order linear equations Although variation of parameters provides a general method for determining a particularsolution of ,5,(y) = R, special methods are available that are often easier to apply when theequation has certain special forms. For example, if the equation has constant coefficientswe can reduce the problem to that of solving a succession of linear equations of first order.The general method is best illustrated with a simple example. EXAMPLE 1. Find a particular solution of the equation (6.28) (D - l)(D - 2)~ = xe"+"' . Solution. Let u = (D - 2)~. Then the equation becomes (D - 1)~ = xez+"' . This is a first-order linear equation in u which can be solved using Theorem 6.1. A particularsolution is u = $es+x2 Substituting this in the equation u = (D - 2)~ we obtain a first-order linear equation for y. Solving this by Theorem 6.1 we find that a particularsolution (withy,(O) = 0) is given by yl(x) = frezr ," et'-' dt .I Although the integral cannot be evaluated in terms of elementary functions we considerthe equation as having been solved, since the solution is expressed in terms of integrals offamiliar functions. The general solution of (6.28) is y = cle2 + c2ezx + +ezx * et2-tdt.i0 6.14 The annihilator method for determining a particular solution of the nonhomogeneousequation We describe next a method which can be used if the equation L(y) = R has constantcoefficients and if the right-hand member R is itself annihilated by a constant-coefficientoperator, say A(R) = 0.
In principle, the method is very simple. We apply the operatorA to both members of the differential equation L(y) = R and obtain a new equationAL(y) = 0 which must be satisfied by all solutions of the original equation. Since AL isanother constant-coefficient operator we can determine its null space by calculating theroots of the characteristic equation of AL. Then the problem remains of choosing from 164 Linear equations this null space a particular function y1 that satisfies L(yl) = R.
The following examplesillustrate the process. EXAMPLE 1. Find a particular solution of the equation (~4- 16)y=x4+x+ 1. Solution. The right-hand member, a polynomial of degree 4, is annihilated by theoperator D5. Therefore any solution of the given equation is also a solution of the equation (6.29) D5(D4 - 16)~ = 0. The roots of the characteristic equation are 0, 0, 0, 0, 0,2, -2,2i, -2i, so all the solutionsof (6.29) are to be found in the linear combiuation y = c1 + c2x + c3x2 + c4x3 + c5x4 + cgezs + c7cz2 + G3COS2x + cg sin2x* We want to choose the ci so that L(y) = x4 + x + 1, where L = D4 - 16. Since the lastfour terms are annihilated by L, we can take cs = c, = c8 = cg = 0 and try to findCl,. * * 9 c5 so that L(c, + c2x + c3x3 + c,x3 + c5x4) = x4 + x + 1. In other words, we seek a particular solution y1 which is a polynomial of degree 4 satisfyingL(yl) = x4 + x + 1 .
To simplify the algebra we write 16y,=ax4+bx3+cx2+dx+e.
This gives us 16# = 24a, so yy' = 3a/2. Substituting in the differential equationL(yJ = x4 + x + 1, we must determine a, b, c, d, e to satisfy $a - ax4 - bx3 - cx2 - dx - e = x4 + X + 1. Equating coefficients of like powers of x we obtain a = - 1 , b=c=O, d= - 1 , e= -Q, so the particular solution y,, is given by EXAMPLE 2. Solve the differential equation y" - 5~' + 6~ = xe". Solution. The differential equation has the form (6.30) Uy) = R, The annihilator method for determining a particular solution 165 where R(x) = xe' and L = D2 - 5D + 6. The corresponding homogeneous equationcan be written as (D - 2)(D - 3)y = 0; it has the independent solutions ul(x) = e2', u2(x) = e3Z. Now we seek a particularsolution y1 of the nonhomogeneous equation. We recognize the function R(x) = xe3Eas a solution of the homogeneous equation (D - 1)2y = 0 . Therefore, if we operate on both sides of (6.30) with the operator (D - 1)2 we find that anyfunction which satisfies (6.30) must also satisfy the equation (D - 1)2(D - 2)(D - 3)y = 0. This differential equation has the characteristic roots 1, 1, 2, 3, so all its solutions are to befound in the linear combination y = sex + bxe" + ce2' + de3", where a, b, c, d are constants. We want to choose a, b, c, d so that the resulting solutionyl satisfies L(yl) = xe" . Since L(ce2" + de3") = 0 for every choice of c and d, we need onlychoose a and b so that L(ae" + bxe") = xe" and take c = d = 0. If we put we havey1 = ae" + bxe", D&) = (a + b)e" + bxe", D2(y1) = (a + 2b)e" + bxe", so the equation (D2 - 5 D + 6)yl = xe" becomes (2a - 3b)e" + 2bxe" = xe". Canceling ex and equating coefficients of like powers of x we find a = $, b = 4. Thereforey1 = $eZ + 4x8 and the general solution of L(y) = R is given by the formula y = cle2x + c2e3' + ze' + *X8. The method used in the foregoing examples is called the annihilator method. It willalways work if we can find a constant coefficient operator A that annihilates R. From ourknowledge of homogeneous linear differential equations with constant coefficients, we knowthat the only real-valued functions annihilated by constant-coefficient operators are linearcombinations of terms of the form x ,TS-leaX X *-lea' COs /lx, xm-leax sin /TX, where m is a positive integer and u and /-I are real constants. The function y = xm--lezris a solution of a differential equation with a characteristic root M.
having multiplicity m. 166 Linear equations Therefore, this function has the annihilator (D - M)~. Each of the functions y =xm-leax cos ,6x and y = x "--lerrx sin ,8x is a solution of a differential equation with complexcharacteristic roots tc f i/3 :, each occurring with multiplicity m, so they are annihilated bythe operator [P - 2crD +- (E" + p2)]". For ease of reference, we list these annihilatorsin Table 6.1, along with some of their special cases. TABLE 6.1 Function Annihilator Y = P--ly = eax Y = x"--le"lZy = cos~x or y = sin /3xy = x"--l cos /3x or y = xm-l sin Bx Y= eax cos jx or y = eaz sin /?xy = x"--lea" cos /lx or y = x"-leas sin Bx D " . (; I :,m D2 + ji3"CD2 + B2P D2 - 2crD + (cc" + p2) [D2 - 2uD + (cc2 + p2)]" Although the annihilator method is very efficient when applicable, it is limited toequations whose right members R have a constant-coefficient annihilator. If R(x) has theform ex2, log x, or tan x, the method will not work; we must then use variation ofparameters or some other method to find a particular solution. 6.15 Exercises In each of Exercises 1 through 10, find the general solution on the interval (- to, + co). 1. y" -y' =x2. 6. y"' - y' = e".2. y" - 4y = e2x.
7. y"' -y' =e" +e-".3. y" + 2-v' = 3xe". 8. y"' + 3y" + 3y' + y = xecz.4. y" + 4y = sin x. 9.
y" + y = xe" sin 2x.5. y" - 2y' + y = ex + e2%. 10. yt4) - y = x2eP . 11. If a constant-coefficient operator A annihilates f and if a constant-coefficient operator Bannihilates g, show that the product AB annihilates f + g . 12. Let A be a constant-coefficient operator with characteristic polynomial pA .(a) Use the annihilator method to prove that the differential equation ,4(y) = em% has aparticular solution of the form eaxYl = p-4 (a) if CC is not a zero of the polynomial pA .(b) If a is a simple zero of pA4 (multiplicity l), prove that the equation A(y) = eaz has theparticular solution xezxy1 = p>(u) * (c) Generalize the results of (a) and (b) when a is a zero of pA with multiplicity m. Miscellaneous exercises on linear diflerenfial equations 167 13. Given two constant-coefficient operators A and B whose characteristic polynomials have nozeros in common. Let C = AB.(a) Prove that every solution of the differential equation C(y) = 0 has the form y = y1 + ys,where A(yJ = 0 and B(yJ = 0.(b) Prove that the functions yr and ye in part (a) are uniquely determined. That is, for agiven y satisfying C(y) = 0 there is only one pair y1 , yz with the properties in part (a). 14. If L(y) = y" + uy' + by, where u and b are constants, let f be that particular solution ofL(y) = 0 satisfying the conditions f(0) = 0 and f'(0) = 1 . Show that a particular solutionof L(y) = R is given by the formula for any choice of c. In particular, if the roots of the characteristic equation are equal, say rl =r2 = m , show that the formula for y,(x) becomes yl(x) = ema - t)epmtR(t) dt. 15.
Let Q be the operator "multiplication by x." That is, Q(y)(x) = x . y(x) for each y in class %"and each real x. Let Z denote the identity operator, defined by Z(y) = y for each y in %?".(a) Prove that DQ - QZ) = I.(b) Show'that D2Q - QDz is a constant-coefficient operator of first order, and determine thisoperator explicitly as a linear polynomial in D.(c) Show that D3Q - QD3 is a constant-coefficient operator of second order, and determinethis operator explicitly as a quadratic polynomial in D.(d) Guess the generalization suggested for the operator D"Q - QDn, and prove your resultby induction. In each of Exercises 16 through 20, find the general solution of the differential equation in thegiven interval. 16. y" -y = l/x, (0, +a). 17. y" + 4y = set 2x, (- $, a 1 . 18. y" -y = sec3x -7-r 77 secx, (-Tj,j 1 * 19. y" - 2y' + y = ee"(ez - 1)2, (-co, +a).20. y" - 7y" + 14y' - sy = log x, (0, +a). 6.16 Miscellaneous exercises on linear differential equations 1. An integral curvey = U(X) of the differential equation yfl - 3~1' - 4y = 0 intersects an integralcurve y = v(x) of the differential equation y"+ 4y' - 5y = 0 at the origin. Determine thefunctions u and v if the two curves have equal slopes at the origin and if [a414 5;y&j- =6. 2. An integral curve y = u(x) of the differential equation y" - 4y' + 29y = 0 intersects an integralcurve y = v(x) of the differential equation y" + 4y' + 13y = 0 at the origin. The two curveshave equal slopes at the origin. Determine u and v if u'(rr/2) = 1 . 168 Linear equations 3. Given that the differential equation y" + 4xy' + Q(x)y = 0 has two solutions of the form'0; ; u(x) and y2 = 4x>, where u(0) = 1 . Determine both u(x) and Q(x) explicitly in terms 4. Let L(y) = y' + Ply' + P,y . To solve the nonhomogeneous equation L(y) = R by variationof parameters, we need to know two linearly independent solutions of the homogeneousequation. This exercise shows that if one solution aI of L(y) = 0 is known, and if aI is neverzero on an interval J, a second solution u2 of the homogeneous equation is given by the formula s ' Q(t) dtu2(4 = 4(x) c [Ul(t)12 9 where Q(x) = ~~r'l(~)~~, and c is any point in J. These two solutions are independent on J.(a) Prove that the function u2 does, indeed, satisfy L(y) = 0.(b) Prove that a1 and u2 are independent on J. 5. Find the general solution of the equation xy " - 2(x + 1)~' + (X + 2)y = x3e2~ for x > 0, given that the homogeneous equation has a solution of the form y = emr .6. Obtain one nonzero solution by inspection and then find the general solution of the differential equation(y" - 4y') + x2(y' - 4y) = 0. 7. Find the general solution of the differential equation 4x2y" +4xy'-y =o, given that there is a particular solution of the form y = x" for x > 0.8. Find a solution of the homogeneous equation by trial, and then find the general solution of the equationx(1 - xI)y" - (1 - 2x)y' + (x2 - 3x + 1)y = (1 - x)3. 9. Find the general solution of the equation (2x - 3x3)y" + 4y' + 6xy = 0, given that it has a solution that is a polynomial in x.10. Find the general solution of the equation x2(1 - x)y" + 2x(2 - x)y' + 2(1 + x)y = x2) given that the homogeneous equation has a solution of the form y = XC.11. Let g(x) = JT et/t dt if x > 0. (Do not attempt to evaluate this integral.) Find all values of the constant a such that the function f defined by f(x) = ieag(z) Linear equations of second order with analytic coefficients 169 satisfies the linear differential equation x2yn + (3x - x2)/ + (1 - x - e2,c)y = 0. Use this information to determine the general solution of the Iequation on the interval (0, + to). 6.17 Linear equations of second order with analytic coefficients A functionfis said to be analytic on an interval (x0 - r, x0 + r) iffhas a power-seriesexpansion in this interval, convergent for Ix - x0] < r. If the coefficients of a homogeneous linear differentialequation y@) + P1(x)y(-) + - * * + P,(x)y = 0 are analytic in an interval (x,, - r, x,, + r), then it can be shown that there exist n inde-pendent solutions ul, . . . , u,, each of which is analytic o.n the same interval. We shallprove this theorem for equations of second order and then discuss an important examplethat occurs in many applications. THEOREM 6.13.
Let P, and P, be analytic on an open interval (x, - r, x0 + r), say P,($ =n~oUx - XOY 3 P2(x) = i c;,(x - XJ .12=0 Then the deferential equation (6.31) y" + P,(x)y' + P2<y = 0 has two independent solutions u1 and u2 which are analytic on the same interval. Proof. We try to find a power-series solution of the form (6.32) y = 2 a& - xo)lz,?I=0 convergent in the given interval. To do this, we substitute the given series for P, and P, inthe differential equation and then determine relations which the coefficients a,, must satisfyso that the function y given by (6.32) will satisfy the equation. The derivatives y' and y" can be obtained by differentiating the power series for y termby term (see Theorem 11.9 in Volume I).
This gives us y' =$;a& - xoY--l =zJn + l)a,+lb - -WY y" =z;(n - l)a,(x - x0)+' =gJn + 2)(n -I- l)a,+,(x - xoY. 170 Linear equations The products P,(x)y' and Pz(x)v are given by the power series? and When these series are substituted in the differential equation (6.31) we find n + 2)(n + l)a,+, + i Kk + 1) Uk+lbn-k + a,&-,] (x - &J)~ = 0.k=O I Therefore the differential equation will be satisfied if we choose the coefficients a, so thatthey satisfy the recursion formula (6.33) (n + 2)(n + lb,,, = -$Jck + l)uk+lbn-k -k akcn-kl for n=0,1,2 ,.... This formula expresses an+2 in terms of the earlier coefficients ao,~l,...,~,+l and the coefficients of the given functions P, and P2. We choose arbitraryvalues of the first two coefficients a, and a, and use the recursion formula to define theremaining coefficients a2, u3, . . .
, in terms of a, and a,. This guarantees that the powerseries in (6.32) will satisfy the differential equation (6.31). The next step in the proof is toshow that the series so defined actually converges for every x in the interval (x0 - r, x0 + r).This is done by dominating the series in (6.32) by another power series known to converge.Finally, we show that we can choose a, and a, to obtain two independent solutions.
We prove now that the series (6.32) whose coefficients are defined by (6.33) converges inthe required interval.
Choose a fixed point x1 # x0 in the interval (x0 - r, x0 + r) and let t = Ix1 - x01.Since the series for P, and P, converge absolutely for x = x1 the terms of these series arebounded, say @ki tk 5 Ml and icki tk 5 M,~ for some MI > 0, Mz > 0. Let A4 be the larger of M, and tM, . Then we have The recursion formula implies the inequality (n + 2)(n + 1) la n+sl $& + 1) bk+ll Fk + bkl 5) k=O nM=- tn+l lx (k + 1) bk+ll tk+'+ &,I/ t"+' + boi - b,+,l tn+l k=O k=O 1 I j$ -&+ 2) Ia k+ll tk+' + luolk=O I = -$ T@ + 1) l&l tk.k=O t Those readers not familiar with multiplication of power series may consult Exercise 7 of Section 6.21. The Legendre equation 1 7 1 Now let A, = laOI, A, = (all, and define A,, A,, . . . successively by the recursionformula (6.34) (n + 2)(n + lM,+, = zl $$k + lM,tkt k=O for it > 0. Then Ia,\ 5 A, for all n 2 0, so the series 2 a,(x - x~)~ is dominated by theseries 1 A, Ix - xOln. Now we use the ratio test to show that x A, Ix - xOln converges iflx - x01 < t. Replacing n by 12 - 1 in (6.34) and subtracting t-l times the resulting equation from (6.34)we find that (n + 2)(n + l)A,+, - t-'(n + l)nA,+, = M(n + 2)A.+, . Therefore A - An+1(n + l>n + (n + m4t n+2 -(n + un + 1>t - ' and we find An+2 lx - ~ol~+~ = (n + l)n + (n + 2)Mt lx - XOIAn+1 Ix - xor+l (n + 2)(n + l>t Ix - x0/ --+ ~t as n --f co. This limit is less than 1 if Ix - x01 < t . Hence 1 a,(x - x~)~ converges ifIx-x,lThis completes the proof.
6.18 The Legendre equation In this section we find power-series solutions for the Legendre equation, (6.35) (1 - x2)y" - 2xy' + cr(a + 1)y =: 0, where c(is any real constant. This equation occurs in problems of attraction and in heat-flow problems with spherical symmetry. When CI is a positive integer we shall find that theequation has polynomial solutions called Legendre polynomials. These are the samepolynomials we encountered earlier in connection with the Gram-Schmidt process (Chapter1, page 26).
The Legendre equation can be written as which has the form[(x" - 1)y'l' = M(cX + 1)y , T(Y)=+, 172 Linear equations where T is a Sturm-Liouville operator, T(f) = (pf')', with p(x) = x2 - 1 and il =cr(a + 1). Therefore the nonzero solutions of the Legendre equation are eigenfunctions ofT belonging to the eigenvalue CC(M + 1).
Sincep(x) satisfies the boundary conditions P(l) =p(-1) = 0, the operator T is symmetric with respect to the inner product The general theory of symmetric operators tells us that eigenfunctions belonging to distincteigenvalues are orthogonal (Theorem 5.3). In the differential equation treated in Theorem 6.13 the coefficient of y" is 1. TheLegendre equation can be put in this form if we divide through by 1 - x2. From (6.35) weobtain yn + P1(x)y' + P2W.Y = 0, where2x Pi(X) = - - and P2(4 =x(x + 1) 1 - x2 l-x2 ' if x2 # 1 . Since l/(1 - x2) = zzEO x2" for 1x1 < 1, both PI and P2 have power-seriesexpansions in the open interval (- I, 1) so Theorem 6.13 is applicable. To find the recursionformula for the coefficients it is simpler to leave the equation in the form (6.35) and try tofind a power-series solution of the form y =ganxnTl=O valid in the open interval (- 1, 1). Differentiating this series term by term we obtain y' = i na,x+l and y" = 2 n(n - l)a,x+'.n-=1 n=2 Therefore we have 2xy' = z 2na,x" = $ 2na,xn,n=l 7l=O and (1 - x2)y" =nx;(n - l)a,xnP2 - n$2n(n - l)a,xn:.
=$+I + 2)(n + l)anf2xn -z;(n - l)a# =zoL(n + 2)(n + lb,,, - n(n - l>a,lx". If we substitute these series in the differential equation (6.35), we see that the equationwill be satisfied if, and only if, the coefficients satisfy the relation (n + 2)(n + l)a,+, - n(n - I)a, - 2na, + LX(CC + l)a, = 0 The Legendre equation 173 for all n 2 0.
This equation is the same as or(n + W + l)a,+, - (n - ct)(n + 1 + &)a, = 0, (6.36) a(E - n)(u + n + 1) nt2 = -(n + l)(n + 2) a'L' This relation enables us to determine a2, a,, a,, . . . , successively in terms of a,. Similarly,we can compute as, as, a,, .
. . , in terms of a,. For the coefficients with even subscriptswe have c((tc + 1)a2= -Tao, a =_(a-2)(a+3)a2=(_l)2a(a-2)(a+l)(a+3)ao,4 3.4 4! and, in general, a,, = (-l)n cc(a - 2, * .
. (cc - 2n + 2) * (a + l)(a + 3).
. * (a + 2n - 1) a, (2n)! This can be proved by induction. For the coefficients with odd subscripts we find a 2n+l = (-1)" (a - l)Ccr - 3, . . * (a - 2n + 1) * (CC + 2)(cx + 4) * * . (a + 2n) a, (2n + l)! Therefore the series for y can be written as (6.37) Y = aoul(x> + alu2(x>, where (6.38) ul(x) = 1 +~on 4c.c - 2) * * * (a - 2% + 2) -:; ,'r l)(a + 3) f * * (a + 2n - 1) X2n ?7=1 n . and (6.39) u2(x) = x + O"c (- 11% (a - l)(a - 3). . * (E - 2T2nf:);:pl + 2)(a + 4). . * (a + 2n) X2n+l. ?l=l The ratio test shows that each of these series converges for 1x1 < 1. Also, since therelation (6.36) is satisfied separately by the even and odd coefficients, each of u1 and u2 is asolution of the differential equation (6.35). These solutions satisfy the initial conditions G9 = 1, u;(o) = 0, u,(O) = 0, u;(o) = 1.
174 Linear deferential equations Since u1 and u2 are independent, the general solution of the Legendre equation (6.35) overthe open interval (- 1, 1) is given by the linear combination (6.37) with arbitrary constantsa, and a,. When t(is 0 or a positive even integer, say K = 2m, the series for uI(x) becomes apolynomial of degree 2m containing only even powers of x. Since we have cc(t(- 2) . . . (cr-2n+2)=2m(2m-2)~~~(2m-2n+2)=, 2"m!(m - n)! and (a + l)(or + 3) * s . (0~ + 2n - 1) = (2m + 1)(2m + 3) . * ' (2m + 2n - 1) (2m + 2n)! m!= 2"(2m)! (m + n)! the formula for or in this case becomes (6.40) q(x) = 1 + @I? 2(-l)"(2m + 2k)! (2m 9 k=l Cm - k)! (m + k)! (2k)! '""* For example, when cc = 0, 2, 4, 6 (m = 0, 1,2, 3) the corresponding polynomials are/ q(x) = 1 3 1 -' 3x3, 1 - 10X2 + 335x*) 1 - 21x2 + 63x4 - zQL~6. The series for uZ(x) is not a polynomial when CI is even because the coefficient of x271+1 isnever zero. When CC is an odd positive integer, the roles of u1 and u2 are reversed; the series foruZ(x) becomes a polynomial and the series for ul(x) is not a polynomial. Specifically if,M. = 2m + 1 we have (6.41)(m!)' m c(2m + 2k + l)! x 2k+l u2(X)=X+(2m+1)!X=l(-1)k(m-k)!(m+k)!(2k+1)! ' For example, when CI = 1, 3, 5 (m = 0, 1,2), the corresponding polynomials are uz(x) = x > x - Bx33 7 x _ L3Sx3 + +Lx5.
6.19 The Legendre polynomials Some of the properties of the polynomial solutions of the Legendre equation can bededuced directly from the differential equation or from the formulas in (6.40) and (6.41).Others are more easily deduced from an alternative formula for these polynomials whichwe shall now derive.
First we shall obtain a single formula which contains (aside from constant factors) boththe polynomials in (6.40) and (6.41).
Let (6.42) P,(x) = $2rn'2' (-1)'(2n - 2r)! Xn-2rr=O r! (n - r)! (n - 2r)! ' The Legendre polynomials 175 where [n/2] denotes the greatest integer 5 n/2. We will show presently that this is theLegendrepolynomial of degree n introduced in Chapter 1. When n is even, it is a constantmultiple of the polynomial ul(x) in Equation (6.40); when n is odd, it is a constant multipleof the polynomial uz(x) in (6.41).? The first seven Legendre polynomials are given by theformulas Po(x> = 1 7 PI(X) = x, P2(x) = &(3x2 - I), P3(x) = $(5x3 - 3x), PJ(X) = $(35x4 - 30x2 + 3), P5(x) = &(63x5 - 70x3 + 15x), , P6(x) = ,+(231x6 - 315x4 + 105x2 - 5). Figure 6.1 shows the graphs of the first five of these functions over the interval [-I , I]. Y P2 P, P, - - I FIGURE 6.1 Graphs of Legendre polynomials over the interval [- 1 , 11. Now we can show that, except for scalar factors, the Legendre polynomials are thoseobtained by applying the Gram-Schmidt orthogonalization process to the sequence ofpolynomials 1, x, x2, . . . , with the inner product t When n is even, say n = 2m, we may replace the index of summation k in Equation (6.40) by a new indexr, where Y = m - k ; we find that the sum in (6.40) is a constant multiple of P,(x). Similarly, when n isodd, a change of index transforms the sum in (6.41) to a constant multiple of P,(x). 176 Linear dlflerential equations First we note that if m # n the polynomials P, and P, are orthogonal because they areeigenfunctions of a symmetric operator belonging to distinct eigenvalues. Also, since P,has degree n and P,, = 1, the polynomials P,(x), PI(X), . . .
, P,(x) span the same subspaceas 1,x ,..., x". In Section 1.14, Example 2, we constructed another orthogonal set ofpolynomials y, , y, , y,, . . . , such that y,(x), ul(x), . .
. , yJx> spans the same subspace as1,x,..., x" for each n. The orthogonalization theorem (Theorem 1.13) tells us that,except for scalar factors, there is only one set of orthogonal functions with this property.Hence we must have P,(x) = w,(x) for some scalars c, . The coefficient of xn in y,(x) is 1, so c, is the coefficient of xn in P,(x).From (6.42) we see that c _ On)!n 2"(n !)2 . 6.20 Rodrigues' formula for the Legendre polynomials In the sum (6.42) defining P,(x) we note that (2n-22r)! _xn d"2r= - X2n-2P and 1 1 n=-(n - 2r)! dx" r! (n - r)! dx" r! (n - r)! 0n! r ' where (;) is the binomial coefficient, and we write the sum in the form P,(x) = & j.g (4)'(;)x2"2'. r=o When [n/2] < r 5 n, the term x2n-2r has degree less than n, so its nth derivative is zero.Therefore we do not alter the sum if we allow r to run from 0 to n. This gives us p,(x) = 1 dn -+2"n! dx" /+(- 1)' ", X2n-2r-7=0 0 Now we recognize the sum on the right as the binomial expansion of (x2 - 1)". Thereforewe have This is known as Rodrigues'formula, in honor of Olinde Rodrigues (17941851), a Frencheconomist and reformer. Using Rodrigues' formula and the differential equation, we can derive a number ofimportant properties of the Legendre polynomials.
Some of these properties are listedbelow. Their proofs are outlined in the next set of exercises. For each n 2 0 we have P,(l) = 1. Exercises Moreover, P,(x) is the only polynomial which satisfies the Legendre equation 177 (1 - x2)yC - 24 + n(n + lly = 0 and has the value 1 when x = 1.For each n 2 0 we have P,(-x) = (-l)"P,(x). This shows that P, is an even function when n is even, and an odd function when n is odd.We have already mentioned the orthogonality relation, s-1, P,(x)P,(x) dx = 0 i f m#n. When m = n we have the norm relation lIPnIl = s l- 1 [P,(x)]' dx = & . Every polynomial of degree n can be expressed as a linear combination of the Legendrepolynomials PO, P, , . . . , P, .
In fact, iff is a polynomial of degree n we have where From the orthogonality relation it follows that s1, g(x)P,(x) dx = 0for every polynomial g of degree less than n. This property can be used to prove that theLegendre polynomial P, has n distinct real zeros and that they all lie in the open interval (-1 , 1). 6.21 Exercises 1. The Legendre equation (6.35) with cc = 0 has the polynomial solution ul(x) = 1 and a solutionu2, not a polynomial, given by the series in Equation (6.41).(a) Show that the sum of the series for u2 is given by 1 +xG4 = ; log rx for lx] X 1.
(b) Verify directly that the function u2 in part (a) is a solution of the Legendre equation whena = o . 178 Linear equations 2. Show that the function f defined by the equation f(X) = 1 - ; log s for 1x1 < 1 satisfies the Legendre equation (6.35) with G(= 1 . Express this function as a linearcombination of the solutions ur and u2 given in Equations (6.38) and (6.39). 3.
The Legendre equation (6.35) can be written in the form [(x2 - l)y']' - cc(c(+ 1)y = 0. (a) If a, b, c are constants with a > b and 4c + 1 > 0, show that a differential equation ofthe type [(x - 4(x - b)y']' - cy = 0 can be transformed to a Legendre equation by a change of variable of the form x = At + B,with A > 0.
Determine A and B in terms of a and b.(b) Use the method suggested in part (a) to transform the equation (x2 - x)f + (2x - I)/ - 2r = 0 to a Legendre equation.4. Find two independent power-series solutions of the Hermite equation y" - 2xy' f 2ary = 0 on an interval of the form (-r , r). Show that one of these solutions is a polynomial when cc is anonnegative integer. 5.
Find a power-series solution of the differential equation xy" + (3 + xyy + 3xzy = 0 valid for all x. Find a second solution of the form .y = x? 2 a,x" valid for all x # 0.6. Find a power-series solution of the differential equation x2/' + x2/ - (ax + 2)y = 0 valid on an interval of the form (-r , r) .7. Given two functions A and B analytic on an interval (x0 - r , x0 + r) , say A(x) = 2 a,(x - x0)%, B(x) = 2 b,(x - x0)".71=0 n-0 It can be shown that the product C(x) = A(x)B(x) is also analytic on (x0 - r , x0 + r). Thisexercise shows that C has the power-series expansion C(x) = 2 c,(x - x0)",T&=0 where c, =$oaA-a. Exercises 179 (a) Use Leibniz's rule for the nth derivative of a product to show that the nth derivative of C isgiven by C(n)(x) = 2 (;pqx)B("-")(.x)).k=O (b) Now use the fact that At")(x,) = k! uk and B lnpk)(xo) = (n - k)! bnwk to obtain @(x0) = n ! 2 ukbnvk.k=O Since Ctn)(xo) = n! c, , this proves the required formula for c, . In Exercises 8 through 14, P,(x) denotes the Legendre polynomial of degree n. These exercisesoutline proofs of the properties of the Legendre polynomials described in Section 6.20. 8. (a) Use Rodrigues' formula to show that p,(X) = ; (x + 1)" + (x - l)Q,,(x), where en(x) is a polynomial.(b) Prove that P,(1) = 1 and that P,(- 1) = (- 1)" .(c) Prove that P,(x) is the only polynomial solution of Legendre's equation (with G(= n)having the value 1 when x = 1. 9. (a) Use the differential equations satisfied by P, and P,,, to show that [(I - x2)(PnP:, - PAPm)l' = [n(n + 1) - m(m + l)IP,P,. (b) If m # n, integrate the equation in (a) from -1 to 1 to give an alternate proof of theorthogonality relation s~lPn(x)Pm(x)dx = 0. 10. (a) Letf(x) = (x2 - 1)". Use integration by parts to show that ~p(xp(x) dx = -~~lf(.+l)(x)f(.-l)(x) d,y, Apply this formula repeatedly to deduce that the integral on the left is equal to 2(2n)!s:(l - x2jndx. (b) The substitution x = cos t transforms the integral Ji (1 - x~)~ dx to J,"12 sin2n+1t dt.
Usethe relation s n/2 2n(2n - 2). .'2 0Sin2n+1fdt = (2n + l)(zn _ 1)...3. 1 and Rodrigues' formula to obtain s~JW)12 dx = & . 180 Linear equations 11. (a) Show that P,(x) = &$ X" + Q,(x), where Q,(x) is a polynomial of degree less than n.(b) Express the polynomialf(x) = x4 as a linear combination of P,, , P, , Pz , P, , and P4.(c) Show that every polynomialf of degree II can be expressed as a linear combination of theLegendre polynomials P,,, P,, . . . , P,. 12. (a) Iffis a polynomial of degree II, write [This is possible because of Exercise 11 (c).] For a fixed m, 0 5 m 5 n , multiply both sides ofthis equation by P,(x) and integrate from - 1 to 1. Use Exercises 9(b) and 10(b) to deduce therelation 2m+l lc, = - 2 spxn(x) dx. 13. Use Exercises 9 and 11 to show that ~~,g(x)P,(x) dx = 0 for every polynomial g of degreeless than IZ. 14. (a) Use Rolle's theorem to show that P, cannot have any multiple zeros in the open interval(-1 , 1). In other words, any zeros of P, which lie in (-1 , 1) must be simple zeros.(b) Assume P, has m zeros in the interval (- 1 , 1). If m = 0, let QO(x) = 1.
If m 2 1, let Q,W = (x - x1)(x - x2) . . .
(x - xm), wherex,,x,, . . . , x,arethemzerosofP,in(-1, 1). Showthat,ateachpointxin(-1 , l),Q,(x) has the same sign as P,(x).(c) Use part (b), along with Exercise 13, to show that the inequality m < n leads to a con-tradiction. This shows that P, has n distinct real zeros, all of which lie in the open intervalC-1 , 1). 15. (a) Show that the value of the integral j?, P, (x)Ph+,(x) dx is independent of n.(b) Evaluate the integral JL1 x P,(x)PGl(x) dx. 6.22 The method of Frobenius In Section 6.17 we learned how to find power-series solutions of the differential equation (6.43) y" + P1(x>y' + P,(x)y = 0 in an interval about a point x,, where the coefficients P, and P2 are analytic. If either PI orP2 is not analytic near x,, , power-series solutions valid near x0 may or may not exist.
Forexample, suppose we try to find a power-series solution of the differential equation (6.44) x2y" -y'-y=o near x0 = 0. If we assume that a solution y = 2 akxk exists and substitute this series in thedifferential equation we are led to the recursion formula n2-n-la n+1 = n+la,. The method of Frobenius ia1 Although this gives us a power series y = 2 akxk which formally satisfies (6.44), the ratiotest shows that this power series converges only for x = 0. Thus, there is no power-seriessolution of (6.44) valid in any open interval about x,, := 0. This example does not violateTheorem 6.13 because when we put Equation (6.44) in the form (6.43) we find that thecoefficients PI and P, are given by PI(X) = - 1x2 and P,(x) = - L .X2 These functions do not have power-series expansions about the origin. The difficulty hereis that the coefficient of y" in (6.44) has the value 0 when x = 0 ; in other words, thedifferential equation has a singular point at x = 0. A knowledge of the theory of functions of a complex variable is needed to appreciate thedifficulties encountered in the investigation of differential equations near a singular point.However, some important special cases of equations with singular points can be treated byelementary methods. For example, suppose the differential equation in (6.43) is equivalentto an equation of the form (6.45) (x - x,)2y" + (x - x,)P(x)y' + ec.x>v = 0 9 where P and Q have power-series expansions in some open interval (x,, - r, x0 + r). Inthis case we say that x0 is a regular singular point of the equation. If we divide both sidesof (6.45) by (x - x0)2 the equation becomes Y" + -J'(x) y, + Q(x) x - x 0 (x - x0)2 y = O for x # x0. If P(x,,) # 0 or Q(xJ # 0, or if Q(x,,) = 0 and Q'(x,) # 0, either the co-efficient ofy' or the coefficient ofy will not have a power-series expansion about the point x0,so Theorem 6.13 will not be applicable. In 1873 the German mathematician Georg Fro-benius (1849-1917) developed a useful method for treating such equations. We shalldescribe the theorem of Frobenius but we shall not present its proof.? In the next sectionwe give the details of the proof for an important special case, the Bessel equation. Frobenius' theorem splits into two parts, depending on the nature of the roots of thequadratic equation (6.46) t(t - 1) + P(x& + Q(xo> = 0. This quadratic equation is called the indicial equation of the given differential equation(6.45). The coefficients P(x,J and Q(x,) are the constant terms in the power-series ex-pansions of P and Q. Let tcl and x2 denote the roots of the indicial equation, These rootsmay be real or complex, equal or distinct. The type of solution obtained by the Frobeniusmethod depends on whether or not these roots differ by an integer. t For a proof see E. Hille, Analysis, Vol. II, Blaisdell Publishing Co., 1966, or E. A. Coddington, AnIntroduction to Ordinary Dijkrential Equations, Prentice-Hall, 1961. 182 Linear equations THEOREM 6.14. FIRST CASE OF FROBENIUS' THEOREM. Let u1 and u2 be the roots of theindicial equation and assume that uI - u2 is not an integer. Then the difSerentia1 equation(6.45) has two independent solutions u1 and u2 of the form (6.47) and with a, = 1, uz(x) = Ix - x01+ b,(x - xoy, with b, = 1.?L=O Both series converge in the interval Ix - x01 < r , and the differential equation is satisfied forOThe constant C is nonzero if N = 0. If N > 0, the constant C may or maynot be zero. As in Case 1, both series converge in the interval Ix - x01 < r , and the solutionsare validfor 0 < Ix - x0(< r . 6.23 The Bessel equation In this section we use the method suggested by Frobenius to solve the Bessel equation x2y" + xy' + (x2 - u")y = 0 , where 0: is a nonnegative constant. This equation is used in problems concerning vibrationsof membranes, heat flow in cylinders, and propagation of electric currents in cylindricalconductors. Some of its solutions are known as Besselfunctions. Bessel functions also arisein certain problems in Analytic Number Theory. The equation is named after the Germanastronomer F. W. Bessel (1784-1846), although it appeared earlier in the researches ofDaniel Bernoulli (1732) and Euler (1764). The Bessel equation has the form (6.45) with x0 = 0, P(X) = 1, and Q(X) = x2 - u2,so the point x0 is a regular singular point. Since P and Q are analytic on the entire real line,we try to find solutions of the form (6.50) y = Ixltza,xn,71=0 with a, # 0, valid for all real x with the possible exception of x = 0.First we keep x > 0, so that lxlt = xt . Differentiation of (6.50) gives us y' = txt-1 5 a,xn + xt 2 na,xn-' = xt--l?l=O TL=O ZJn + W$. The Bessel equation 183 Similarly, we obtain y" = xt-2n$o(n + t)(n + t - l)a,x". If L(y) = x2$' + xy' + (x2 - a2)y, we find L(y) = xt f$ (n + t)(n + t - l)a,xn + x'z (n + t)a,x"n==O n=O + x1$ anxni2 - x'f a2a,xR = x'z [(n + t)" - cc2]anxn + x15 anxn+2.VI=0 ?I=0 ?I=0 VI=0 Now we put L(y) = 0, cancel 9, and try to determine the a,, so that the coefficient of eachpower of x will vanish. For the constant term we need (t2 - a2)ao = 0. Since we seek asolution with a, # 0, this requires that (6.51) t2 - u2 = 0.
This is the indicial equation. Its roots a and -a are the only possible values of t that cangive us a solution of the desired type. Consider first the choice t = u, For this t the remaining equations for determining thecoefficients become (6.52) [(l + m)" - a2]al = 0 and [(n + a)" - a2]a, + a,, = 0 for n 2 2. Since CL 2 0, the first of these implies that a, = 0. The second formula canbe written as (6.53) a, = - (n +';s2- tc2 = an-2-n(n+)' so a3 = a5 = a, = * * * = 0. For the coefficients with even subscripts we have --a0a2 = 2(2 + 2a) = --a0 -a2 (- 112a0 271 + a) ' a4 = 4(4 + 2a) = 242!(1+a)(2+a)' and, in general, -a4a6 = 6(6 + 2cr) = (- U3a0 263! (1 + a)(2 + a)(3 + a) ' a - (- lYa0 2n - 22nn ! (1 + a)(2 + a) * * . (n ,+T) ' Therefore the choice t = u gives us the solution m y = aox' 1 +(2 (-1),x2, 1n=,22nn!(1+cr)(2+a).*~(n+a) ' The ratio test shows that the power series appearing in this formula converges for all real x.
184 Linear equations In this discussion we assumed that x > 0. If x < 0 we can repeat the discussion withxt replaced by (-x)" . We again find that t must satisfy the equation t2 - c? = 0. Takingt = u we then obtain the same solution, except that the outside factor xa is replaced by(-x)".
Therefore the functionf, given by the equation (6.54) fAx) = ao IxIa is a solution of the Bessel equation valid for all real x # 0.
For those values of cc for whichf:(O) andfl(O) exist the solution is also valid for x = 0. Now consider the root t = -LX of the indicial equation. We obtain, in place of (6.52), theequations [(l - cc)" - c?]al = 0 and Kn - 4" - cr2]a, + an-2 = 0, which become (1 - 20()a, = 0 and 0 - 2a)a, + ane2 = 0. If 2u is not an integer these equations give us a, = 0 and a, = - anm2n(n - 2cc) for n 2 2.
Since this recursion formula is the same as (6.53), with CI replaced by -CI, weare led to the solution (6.55) f-Ax) = ao W" l + 2& 22nn((1 _ &y;. . . (n _ u) valid for all real x # 0.The solution f-, was obtained under the hypothesis that 2cr is not a positive integer. However, the series forf_, is meaningful even if 2u is a positive integer, so long as CC isnot a positive integer. It can be verified thatf-, satisfies the Bessel equation for such CC.Therefore, for each M 2 0 we have the series solution f,, given by Equation (6.54); andif cI is not a nonnegative integer we have found another solutionf-, given by Equation (6.55).The two solutions f, and f-, are independent, since one of them --to3 as x + 0, and theother does not. Next we shall simplify the form of the solutions. To do this we need someproperties of Euler's gamma function, and we digress briefly to recall these properties. For each real s > 0 we define I'(s) by the improper integral l?(s) = lo: ts-le-t dt . The Bessel equation 185 This integral converges if s > 0 and diverges if s 5 0. Integration by parts leads to thefunctional equation (6.56) lT(s + 1) = s I?(s). This implies that lys + 2) = (s + l)l?(s + 1) = (s + 1)s l'(s), r(s + 3) = ($ + 2)r(s + 2) = (s + 2) ($ + qs r(s),and, in general, (6.57) r(s + n) = (S + n - 1). . .
(S + 1:)~ r(#) for every positive integer n. Since I'(1) = jr e-t dt = 1, when we put s = 1 in (6.57) wefind r(n + 1) = n!. Thus, the gamma function is an extension of the factorial function from integers to positivereal numbers. The functional equation (6.56) can be used to extend the definition of l?(s) to negativevalues of s that are not integers. We write (6.56) in the form (6.58) qs) = Us + l)s The right-hand member is meaningful ifs + 1 > 0 and s # 0. Therefore, we can use thisequation to define I'(s) if - 1 < s < 0. The right-hand member of (6.58) is now meaning-ful if s + 2 > 0, s # - 1, s # 0, and we can use this equation to define l?(s) for -2 ~~* (n + x). We can express this product interms of the gamma function by taking s = 1 + c(in (6.57). This gives us r(n + 1 + ~4)(1 + a)(2 + a> * * * (n + a> = ~+a) . Therefore, if we choose a,, = 2-"/I'(1 + a) in Equation (6.54) and denote the resultingfunctionf,(x) by J,(x) when x > 0, the solution for x > 0 can be written as (6.59) J,(x) = ia20(-ly '5 Zn* 0n=,n!Vn+ 1 +a) 2 1 8 6 Linear difSerentia1 equations The function J, defined by this equation for x > 0 and M 2 0 is called the Bessel/unctionof thejrst kind of order u. When u is a nonnegative integer, say CI = p , the Bessel functionJ, is given by the power series J,(x) = z. ., :;yp,, (5"'" (p = 0, 1,2, . . .). This is also a solution of the Bessel equation for x < 0. Extensive tables of Bessel functionshave been constructed.~~
~~The graphs of the two functions Jo and J1 are shown in Figure 6.2. tJo FIGURE 6.2 Graphs of the Bessel functions Jo and J1.~~
~~We can define a new function I, by replacing CI by -u in Equation (6.59), if u is suchthat P(n + 1 - u) is meaningful; that is, if cc is not a positive integer. Therefore, if x > 0andu>O,u# 1,2,3 ,..., wedeline J-,(x) = 05 -& r(;--'; _ .,($'.~~
~~Taking s = 1 - u in (6.57) we obtain r(n + 1 - LX) = (1 - ct) (2 - U) . . . (n - U) I?(1 - U) and we see that the series for J-,(x) is the same as that for f-,(x) in Equation (6.55) witha, = 2"/I'(l - u), x > 0. Therefore, if u is not a positive integer, J-, is a solution ofthe Bessel equation for x > 0. If u is not an integer, the two solutions J,(x) and J-,(x) are linearly independent on thepositive real axis (since their ratio is not constant) and the general solution of the Besselequation for x > 0 is y = c,J,(x) + c,J&) . If u is a nonnegative integer, say cx = p , we have found only the solution J, and its con-stant multiples valid for x > 0. Another solution, independent of this one, can be found The Bessel equation 187 by the method described in Exercise 4 of Section 6.16. This states that if u1 is a solution ofy" + Ply' + P,y = 0 that never vanishes on an interval I, a second solution u2 independentof u1 is given by the integral where Q(x) = e-J1'l(s)dr. For the Bessel equation we have PI(x) = l/x, so Q(x) = l/xand a second solution u2 is given by the formula (6.60) if c and x lie in an intervalThis second solution can may write up(x) = J,(x) s"1 dt *c w&v I in which J, does not vanish.be put in other forms. For example, from Equation (6.59) we 1= r s,(t), [4(01" tzn where g,(O) # 0. In the interval Z the function g, has a power-series expansion s,(t) = 2 Ant"VI=0 which could be determined by equating coefficients in the identity g,(t) [J,(t)]2 = Pp. Ifwe assume the existence of such an expansion, the integrand in (6.60) takes the form 1 1 m f[J,(f>]" = t2"+1 n=OcAnt".~~
~~Integrating this formula term by term from c to x we obtain a logarithmic term A,, log x(from the power t-l) plus a series of the form xP2P 2 B,?.~~
~~Therefore Equation (6.60)takes the form u2(x) = A2,J,(x) log x + Jy(x)x-2D~ B,x" .?k=O It can be shown that the coefficient A,, # 0. If we multiply u2(x) by l/A,, the resultingsolution is denoted by K,(x) and has the form K,(x) = J,(x) log x + x-p2 c,,xn.7L=O This is the form of the solution promised by the second case of Frobenius' theorem.Having arrived at this formula, we can verify that a solution of this form actually exists by substituting the right-hand member in the Bessel equation and determining the co-efficients C, so as to satisfy the equation. The details of this calculation are lengthy and 188 Linear dxerential equations will be omitted. The final result can be expressed as where ho = 0 and h, = 1 + 4 + * * * + l/n for n 2 1. The series on the right convergesfor all real x. The function K, defined for x > 0 by this formula is called the Besselfunctionof the second kind of orderp.~~
~~Since K, is not a constant multiple of J, , the general solutionof the Bessel equation in this case for x > 0 is Y = c,J,(x) + c,K&). Further properties of the Bessel functions are discussed in the next set of exercises. 6.24 Exercises 1. (a) Let f be any solution of the Bessel equation of order a and let g(x) = x'hf (x) for x > 0.Show that g satisfies the differential equation (1 -44a2y" + 1 + 4x2 1y =o . (b) When 4a2 = 1 the differential equation in (a) becomes y" + y = 0 ; its general solutionis y = A cos x + B sin x. Use this information and the equation? I'(h) = d, to show that,forx >O, and (c) Deduce the formulas in part (b) directly from the series for Jim(x) and Il,h(x).2. Use the series representation for Bessel functions to show that (a) 2 (x~J,W> = x~J,-~W, (b) $ (x-="J,(x)) = -xP-r~+,(x) 3. Let F,(x) = x"J,(x) and 'G,(x) = xaaJ,(x) for x > 0. Note that each positive zero of J, is azero of F, and is also a zero of G, . Use Rolle's theorem and Exercise 2 to prove that the posi-tive zeros of J, and J,+r interlace. That is, there is a zero of J, between each pair of positivezeros of Ja+l, and a zero of J,+l between each pair of positive zeros of J,. (See Figure 6.2.) t The change of variable I = u* gives us r(t) = Jo; t-'/&-t dt = 2s ")e-d du = 4,.0 (See Exercise 16 of Section 11.28 for a proof that 2 Jo" e-u2 du = 2/G.) Exercises 189 4. (a) From the relations in Exercise 2 deduce the recurrence relations f J,(x) + J;(x) = Ja&(X) and f J,(x) - J;(x) = J,+,(x) . I (b) Use the relations in part (a) to deduce the formulas Ja-I(X) + J,+,(x) = ; J,(x) and J-,(x) - J,+,(x) = 2J;(x). 5. Use Exercise 1 (b) and a suitable recurrence formula to show that Find a similar formula for J&(x).~~
~~Note: J,(x) is an elementary function for every c(which ishalf an odd integer. 6. Prove that ; ;; (J,~(x) + J,ztl(x)) = ; J;(x) - "$ J;+l(x) and ; (xJ,WJ,+,(xN = x(J,2(4 - J:+I(x>). 7. (a) Use the identities in Exercise 6 to show that J;(x) + 2 2 J;(x) = 1 and 2 (2n + l)J,(x)J,+,(x) = &x.n=l Tl=O (b) From part (a), deduce that IJo 5 1 and jJ,(x)l 5 44% for n = 1,2, 3, . . .~~

and allx 20. 8. Let g,(x) = x'~~~(ux") for x > 0, where a and b are nonzero constants. Show that g, satisfiesthe differential equation 2y" + (aWx2b + 2 - aW)y = 0 if, and only if, fa is a solution of the Bessel equation of order a.9. Use Exercise 8 to express the general solution of each of the following differential equations in terms of Bessel functions for x > 0.(a) y" +xy =O. (c) y" + x"y = 0.(b)y" +x2y =O. (d) x"y" + (x4 + &)y = 0. 10. Generalize Exercise 8 when fm and g, are related by the equation g,(x) = xcf,(u9) for x > 0.Then find the general solution of each of the following equations in terms of Bessel functionsforx >O.(a) xy" + 6y' + y = 0. (c) xy" + 6y' + x*y = 0.(b)xy"+6y'+xy=O. (d) x2y" - xy' + (x + 1)y = 0. 11. A Bessel function identity exists of the form J,(x) -Jo(x) = aJ,"W, where a and c are constants. Determine a and c. 190 Linear difSerentia1 equations 12. Find a power series solution of the differential equation xy" + y' + y = 0 convergent for- - t o < x < +a. Show that for x > 0 it can be expressed in terms of a Bessel function. 13. Consider a linear second-order differential equation of the form X2&)yn + xP(x)y' + Q(x)y = 0, where A(x), P(x), and Q(x) have power series expansions, A(x) = 2 UkXk)k=O P(x) = ygPkXk,k=O with a, # 0, each convergent in an open interval (-r, r) . If the differential equation has a series solution of the form y = xt 2 CnXll)?I=0 valid for 0 < x < Y, show that t satisfies a quadratic equation of the form t2 + bt + c = 0,and determine b and c in terms of coefficients of the series for A(x), P(x), and Q(x). 14. Consider a special case of Exercise 13 in which A(x) = 1 - x, P(X) = $, and Q(x) = -8~.Find a series solution with t not an integer. 15. The differential equation 2x2y" + (x2 - x)y' + y = 0 has two independent solutions of theform y =Xt~C,,x",T&=0 valid for x > 0. Determine these solutions.16. The nonlinear differential equation y" + y + my 2 = 0 is only "mildly" nonlinear if c(is a small nonzero constant. Assume there is a solution which can be expressed as a power series in aof the form y = g U,(X)@ (valid in some interval 0 < CL < r)?I=0 and that this solution satisfies the initial conditions y = 1 and y' = 0 when x = 0. To con-form with these initial conditions, we try to choose the coefficients U,(X)SO that U,(O) = 1,u,(O) = 0 and u,(O) = u;(O) = 0 for n 2 1 . Substitute this series in the differential equation,equate suitable powers of c(and thereby determine uo(x) and ut(x). 7SYSTEMS OF DIFFERENTIAL EQUATIONS 7.1 Introduction Although the study of differential equations began in the 17th century, it was not untilthe 19th century that mathematicians realized that relatively few differential equations couldbe solved by elementary means. The work of Cauchy, Liouville, and others showed theimportance of establishing general theorems to guarantee the existence of solutions tocertain specific classes of differential equations. Chapter 6 illustrated the use of an existence-uniqueness theorem in the study of linear differential equations. This chapter is concernedwith a proof of this theorem and related topics. Existence theory for differential equations of higher order can be reduced to the first-order case by the introduction of systems of equations. For example, the second-orderequation (7.1) y" + 22~' - y = et can be transformed to a system of two first-order equations by introducing two unknownfunctions y, and yZ , where Yl=YY Y2 = YI. Then we have y6 = yl = y", so (7.1) can be written as a system of two first-order equations : (7.2)Y; = Y2 y] = y, - 2ty2 + et. We cannot solve the equations separately by the methods of Chapter 6 because each of theminvolves two unknown functions. In this chapter we consider systems consisting of n linear differential equations of firstorder involving n unknown functions yI, . . . , y, . These systems have the form YI = Pll(OY1 + P12COY2 + * * . + Pln(OYn + q1(0 (7.3) Y:, = I&JOY1 + P,2(9Y2 + * * * + Pm(9Yn + q,(t). 191 192 Systems of deferential equations The functionsp, and qi which appear in (7.3) are considered as given functions defined ona given interval J.

The functions yr, . . .

. , yn are unknown functions to be determined.Systems of this type are called$rst-order linear systems.

In general, each equation in thesystem involves more than one unknown function so the equations cannot be solvedseparately. A linear differential equation of order n can always be transformed to a linear system.Suppose the given nth order equation is (7.4) ytn) + ) + * - * + a,y = R(t), where the coefficients ai are given functions. To transform this to a system we write y1 = yand introduce a new unknown function for each of the successive derivatives of y. That is,we put Yl=Y, Y2 = Y;, Y3 = Yi!, .

.

. 3 Yn = YL, and rewrite (7.4) as the system YI = Y2 Y;; = Y3 (7.5) I Yn-1 = Yn y; = -any1 - a,-,y2 - * * * - alyn + R(t). The discussion of systems may be simplified considerably by the use of vector and matrixnotation. Consider the general system (7.3) and introduce vector-valued functions Y = cY17.. . ,Y,), Q = (41,. . . .,q,J, and a matrix-valued function P = [pij], defined by theequations y(t) = O1(t>, * * * 9 y,(t)>, Q(t) = @l(t), . . . 9 qn(t>>, PC0 = h(t)1 for each t in J. We regard the vectors as n x 1 column matrices and write the system (7.3)in the simpler form (7.6) Y' = P(t) Y + Q(t).

For example, in system (7.2) we have Y= K;], P(t)= [; -:,I, Q(t)= [,j. Calculus of matrix functions 193 In system (7.5) we have Y = r1 - Y2 . Yn- 2 P(t) = -0 1 0 . .

. 0- 0 0 1 . . . 0 0 0 0 . . . 1 ---a, -a,-, -an-2 * ** - a , - , Q(t) = 'o- 0 0 .R(t)- An initial-value problem for system (7.6) is to find a vector-valued function Y whichsatisfies (7.6) and which also satisfies an initial condition of the form Y(a) = B, whereaEJandB=(b,... , b,) is a given n-dimensional vector. In the case n = 1 (the scalar case) we know from Theorem 6.1 that, if P and Q arecontinuous on J, all solutions of (7.6) are given by the explicit formula (7.7) Y(X) = &(')Y(a) + eA(') ," e-act)Q(f) dt,s where A(x) = j; P(t) dt , and a is any point in J. We will show that this formula can besuitably generalized for systems, that is, when P(t) is an n x n matrix function and Q(t)is an n-dimensional vector function. To do this we must assign a meaning to integrals ofmatrices and to exponentials of matrices. Therefore, we digress briefly to discuss thecalculus of matrix functions. 7.2 Calculus of matrix functions The generalization of the concepts of integral and derivative for matrix functions isstraightforward. If P(t) = [pi,(t)], we define the integral jz P(t) dt by the equation That is, the integral of matrix P(t) is the matrix obtained by integrating each entry of P(t),assuming of course, that each entry is integrable on [a, b]. The reader can verify that thelinearity property for integrals generalizes to matrix functions. Continuity and differentiability of matrix functions are also defined in terms of theentries. We say that a matrix function P = [pij] is continuous at t if each entry pij iscontinuous at t. The derivative P' is defined by differentiating each entry, whenever all derivatives pi,(t) exist. It is easy to verify the basic differentiation rules forsums and products. For example, if P and Q are differentiable matrix functions, we have (P + Q)' = P' + Q' 194 Systems of equations if P and Q are of the same size, and we also have (PQ)' = PQ' + P'Q if the product PQ is defined. The chain rule also holds. That is, if F(t) = P[g(t)] , where Pis a differentiable matrix function and g is a differentiable scalar function, then F'(t) =g'(t)P'[g(t)] . The zero-derivative theorem, and the first and second fundamental theoremsof calculus are also valid for matrix functions. Proofs of these properties are requested inthe next set of exercises. The definition of the exponential of a matrix is not so simple and requires furtherpreparation.

This is discussed in the next section. 7.3 Infinite series of matrices. Norms of matrices Let A = [aij] be an n x n matrix of real or complex entries. We wish to define theexponential en in such a way that it possesses some of the fundamental properties of theordinary real or complex-valued exponential. In particular, we shall require the law ofexponents in the form (7.8) for all real s and t , and the relation (7.9) e"=I, where 0 and Z are the n x n zero and identity matrices, respectively. It might seem naturalto define eA to be the matrix []. However, this is unacceptable since it satisfies neither ofproperties (7.8) or (7.9). Instead, we shall define eA by means of a power series expansion, We know that this formula holds if A is a real or complex number, and we will prove thatit implies properties (7.8) and (7.9) if A is a matrix. Before we can do this we need toexplain what is meant by a convergent series of matrices. DEFINITION OF CONVERGENT SERIES OF MATRICES. Given an infkite sequence of m X nmatrices {Ck} whose entries are real or complex numbers, denote the ij-entry of C, by cl,"' . Zfall mn series (7.10) $6' (i=l,..., m,j=l,..., n) are convergent, then we say the series of matrices z& C, is convergent, and its sum is deJnedto be the m x n matrix whose ij-entry is the series in (7.10). Exercises 195 A simple and useful test for convergence of a series of matrices can be given in terms of thenorm of a matrix, a generalization of the absolute value of a number. DEFINITION OF NORM OF A MATRIX. If A = [aij] is an m x n matrix of real or complexentries, the norm of A, denoted by l\All, is defined to be the nonnegative number given by theformula (7.11) In other words, the norm of A is the sum of the absolute values of all its entries. Thereare other definitions of norms that are sometimes used, but we have chosen this one becauseof the case with which we can prove the following properties. THEOREM 7.1. FUNDAMENTAL PROPERTIES OF NORMS. For rectangular matrices A and B,and all real or complex scalars c we have II-4 + AI I IIAII + IIBII, IWll I II4 IIBII 3 II4 = ICI IIAII . Proof. We prove only the result for II ABll , assuming that A is m x n and B is n x p .The proofs of the others are simpler and are left as exercises.

Writing A = [aJ , B = [bkj] , we have AB = [z;==, ai,bkj] , so from (7.11) we obtain IIABII = 2 I$ 1 ~airhj 1 I$ ~laikl~lhjl 52 ~I4 IIBII = IIAII II4 .i=l j,l J+l i=lk=l j=l i=l k=l Note that in the special case B = A the inequality for [lABI/ becomes llA211 5 llAl12.By induction we also have llA"II I II4" for k=l,2,3 ,.... These inequalities will be useful in the discussion of the exponential matrix.The next theorem gives a useful sufficient condition for convergence of a series of matrices. THEOREM 7.2. TEST FOR CONVERGENCE OF A MATRIX SERIES. If {c,} is a sequence Of Wl X nmatrices such that I;=1 II C,ll converges, then the matrix series zF-'em=l C, also converges. Proof. Let the q-entry of C, be denoted by c$'. Since lcj,k'l 5 IlC,(l , convergence of121 Ilckll pl'im ies absolute convergence of each series xTZl cjj"'. Hence each series 2:-l c$'is convergent, so the matrix series z;Cmcl C, is convergent. 7.4 Exercises 1. Verify that the linearity property of integrals also holds for integrals of matrix functions.2. Verify each of the following differentiation rules for matrix functions, assuming P and Q are differentiable. In (a), P and Q must be of the same size so that P + Q is meaningful. In 196 Systems of equations (b) and (d) they need not be of the same size provided the products are meaningful. In (c)and (d), Q is assumed to be nonsingular.(a) (P + Q)' = P' + Q' . (c) (Q-l)' = -Q-lQ'Q-l.(b) (PQ)' = PQ' + P'Q. (d) (PQ-')' = -PQ-lQ'Q-l + P'Q-l. 3.

(a) Let P be a differentiable matrix function. Prove that the derivatives of P2 and P3 are givenby the formulas (P2)' = PP' + P'P, (P3)' = P2P' + PP'P + P'P2. (b) Guess a general formula for the derivative of Pk and prove it by induction.4. Let P be a differentiable matrix function and let g be a differentiable scalar function whose range is a subset of the domain of P. Define the composite function F(t) = P[g(t)] and provethe chain rule, F'(t) =g'(t)P'[g(t)J . 5. Prove the zero-derivative theorem for matrix functions: If P'(t) = 0 for every t in an openinterval (a, b), then the matrix function P is constant on (a, b). 6. State and prove generalizations of the first and second fundamental theorems of calculus formatrix functions. 7. State and prove a formula for integration by parts in which the integrands are matrix functions.8. Prove the following properties of matrix norms : IM + WI I IIA I I + IIBII , IlcAII = ICI IMII . 9. If a matrix function P is integrable on an interval [a, b] prove that 10. Let D be an n x n diagonal matrix, say D = diag (Al, . . . , A,). Prove that the matrix seriesz:km_s D"/k! converges and is also a diagonal matrix, m D"c - = diag (ear, . . . , e"n),k! k=O (The term corresponding to k = 0 is understood to be the identity matrix I.)11. Let D be an n x n diagonal matrix, D = diag (jlr , . . . , I,). If the matrix series I:=, ckDk converges, prove that 2 CkDk = diag 2 c&, . .k=O k=O * ' z. c&) * 12. Assume that the matrix series IF=1 ck converges, where each ck is an n x n matrix. Prove thatthe matrix series I:& (AC&I) also converges and that its sum is the matrix Here A and B are matrices such that the products ACkB are meaningful. The diferential equation satisjed by et' 197 7.5 The exponential matrix Using Theorem 7.2 it is easy to prove that the matrix series (7.12)m Akckc0 72 converges for every square matrix A with real or complex entries. (The term correspondingto k = 0 is understood to be the identity matrix Z.) The norm of each term satisfies theinequality Since the series 2 a"/k! converges for every real a, Theorem 7.2 implies that the series in(7.12) converges for every square matrix A. DEFINITION OF THE EXPONENTIAL MATRIX. For any n x n matrix A with real or complexentries we dejne the exponential & to be the n x n matrix given by the convergent series in(7.12). That is, Note that this definition implies e* = I, where 0 is the zero matrix. Further propertiesof the exponential will be developed with the help of differential equations. 7.6 The differential equation satisfied by et* Let t be a real number, let A be an n x n matrix, and let E(t) be the n x n matrix given by E(t) = d" . We shall keep A fixed and study this matrix as a function oft. First we obtain a differentialequation satisfied by E. T H E O R E M 7.3. For every real t the matrix function E defined by E(t) = etA satisfies thematrix diferential equation E'(t) = E(t)A = AE(t). Proof. From the definition of the exponential matrix we have 198 Systems of difSerentia1 equations Let ci,"' denote the ij-entry of Ak. Then the ij-entry of tkAklk! is t"c$'lk! . Hence, from thedefinition of a matrix series, we have (7.13) Each entry on the right of (7.13) is a power series in t, convergent for all t.

Therefore itsderivative exists for all t and is given by the differentiated series m ktk-1c- $9 = k=lk! " This shows that the derivative E'(t) exists and is given by the matrix series A = lqt)A. In the last equation we used the property Ak+l = A"A . Since A commutes with A' we couldalso have written A"+l = AAL to obtain the relation E'(t) = AE(t).
This completes theproof.

Note: The foregoing proof also shows that A commutes with etA . 7.7 Uniqueness theorem for the matrix differential equation F'(t) = M(t) In this section we prove a uniqueness theorem which characterizes all solutions of thematrix differential equation F'(t) = AF(t) . The proof makes use of the following theorem.
THEOREM 7.4. NONSINGULARITY OF etA. For any n x n matrix A and any scalar t we have (7.14) Hence etA is nonsingular, and its inverse is e-tA. Proof.

Let F be the matrix function defined for all real t by the equation F(t) = etAcetA. We shall prove that F(t) is the identity matrix I by showing that the derivative F'(t) is thezero matrix. Differentiating F as a product, using the result of Theorem 7.3, we find F'(t) = etA(e--tA)f + (etA>'e-tA = &4(-A&A) + A&f+' = -AetAe-tA + AetAe-tA = 0, since A commutes with etA. Therefore, by the zero-derivative theorem, F is a constantmatrix. But F(0) = eoAeoA = I, so F(t) = I for all t. This proves (7.14). The law of exponents for exponential matrices 199 THEOREM 7.5. UNIQUENESS THEOREM. Let A and B be given n x n constant matrices.Then the only n x n matrix function F satisfying the initial-value problem F'(t) = AF(t), F(0) = B for - a3 < t < + 00 is (7.15) F(t) = etaB .

Proof. First we note that etAB is a solution. Now let F be any solution and consider thematrix function G(t) = eWtaF(t) . Differentiating this product we obtain G'(t) = ePtAF'(t) - AePtaF(t) = ePtAAF(t) - eTtAAF(t) = 0. Therefore G(t) is a constant matrix, G(t) = G(0) = F(0) = B. In other words, e-taF(t) = B. Multiplying by eta and using (7.14) we obtain (7.15). Note: The same type of proof shows that F(t) = BetA is the only solution of theinitial-value problem F(t) = FQ)A, F(0) = B. 7.8 The law of exponents for exponential matrices The law of exponents eAeB = eAfB is not always true for matrix exponentials. A counterexample is given in Exercise 13 of Section 7.12. However, it is not difficult to prove thatthe formula is true for matrices A and B which commute. THEOREM 7.6. Let A and B be two n x n matrices which commute, AB = BA . Then wehave (7.16) I+ + ~ = eAeB. Proof. From the equation AB = BA we find that A2B = A(BA) = (AB)A = (BA)A = BA2, so B commutes with A2. By induction, B commutes with every power of A. By writingetA as a power series we find that B also commutes with eta for every real t. Now let F be the matrix function defined by the equation I = $(A+B) - etAetBe 200 Systems of differential equations Differentiating F(t) and using the fact that B commutes with eta we find F'(t) = (A + B)et'A+B' - AetAetB - etABetB = (A + B)et'A+B' - (A + B)etAetB = (A + B)F(t) . By the uniqueness theorem we have F(t) = et(d+B)F(0). But F(0) = 0, so F(t) = 0 for all t. Hence $(A+B) = etAetB When t = 1 we obtain (7.16). EXAMPLE. The matrices SA and tA commute for all scalars s and t. Hence we have @etA = e (s+t)A . 7.9 Existence and uniqueness theorems for homogeneous linear systems with constantcoefficients The vector differential equation Y'(t) = A Y(t), where A is an n x n constant matrix andY is an n-dimensional vector function (regarded as an n x 1 column matrix) is called ahomogeneous linear system with constant coeficients. We shall use the exponential matrixto give an explicit formula for the solution of such a system. THEOREM 7.7. Let A be a given n x n constant matrix and let B be a given n-dimensionalvector. Then the initial-value problem (7.17) Y'(t) = AY(t), Y(0) = B, has a unique solution on the interval - CO < t < + CO. This solution is given by the formula (7.18) Y(t) = etAB. More generally, the unique solution of the initial value problem is Y(t) = ectaJAB.Y'(t) = A Y(t), Y(a) = B, Proof. Differentiation of (7.18) gives us Y'(t) = ActAB = A Y(t). Since Y(0) = B ,this is a solution of the initial-value problem (7.17). To prove that it is the only solution we argue as in the proof of Theorem 7.5. Let Z(t)be another vector function satisfying Z'(t) = AZ(t) with Z(0) = B , and let G(t) = emtAZ(t).Then we easily verify that G'(t) = 0, so G(t) = G(0) = Z(0) = B. In other words, The problem of calculating et' 201 etAZ(t) = B, so Z(t) = etAB = Y(t). The more general case with initial value Y(a) = Bis treated in exactly the same way. 7.10 The problem of calculating erA Although Theorem 7.7 gives an explicit formula for the solution of a homogeneoussystem with constant coefficients, there still remains the problem of actually computingthe exponential matrix etA. If we were to calculate etA directly from the series definition wewould have to compute all the powers AL for k = 0, 1,2, , . . , and then compute the sumof each series z& tkc$)/k! , where c:j"' is the q-entry of Ak. In general this is a hopelesstask unless A is a matrix whose powers may be readily calculated. For example, if A is adiagonal matrix, say A = diag (1,) . . . , 1,)) then every power of A is also a diagonal matrix, in fact, A'=diag(;I:,...,iZE).

Therefore in this case etA is a diagonal matrix given by eiA = diag = diag (et"', . . . , et""). Another easy case to handle is when A is a matrix which can be diagonalized. Forexample, if there is a nonsingular matrix C such that PAC is a diagonal matrix, sayC-lAC = D , then we have A = CDC-l , from which we find and, more generally, A2 = (CDC-l)(CDC-l) = CDW-l, Ak = CD"C-I. Therefore in this case we have etA = 2 $ik= 2 +; CDkC-1 = c k=O * k=O ' Here the difficulty lies in determining C and its inverse. Once these are known, etA is easilycalculated. Of course, not every matrix can be diagonalized so the usefulness of theforegoing remarks is limited. 5 4EXAMPLE 1. Calculate etA for the 2 x 2 matrix A = [11 2 ' Solution. This matrix has distinct eigenvalues 1, = 6, 1, = 1 , so there is a nonsingular matrix C = such that C-lAC = D, where D = diag (&, 1,) = To 202 Systems of deferential equations determine C we can write AC = CD, or Multiplying the matrices, we find that this equation is satisfied for any scalars a, b, c, d witha=4c,b= - d . Takingc=d=lwechoose Therefore etA EXAMPLE 2. Solve the linear system YI = 5Yl + 4Y,Y;; = Yl + 2Y, subject to the initial conditions ~~(0) = 2, ~~(0) = 3. Solution. In matrix form the system can be written as Y'(t) = A Y(t), where A = By Theorem 7.7 the solution is Y(t) = etA Y(0). Using the matrix etA calculated in Example1 we find 4eet + et 4e6t - 4et 2I[1e6t - et est + 4et 3from which we obtain yl = 4e6t - 2et, yz = eat + 2et. There are many methods known for calculating etA when A cannot be diagonalized.Most of these methods are rather complicated and require preliminary matrix transforma-tions, the nature of which depends on the multiplicities of the eigenvalues of A. In a latersection we shall discuss a practical and straightforward method for calculating etA whichcan be used whether or not A can be diagonalized.

It is valid for all matrices A and requiresno preliminary transformations of any kind. This method was developed by E. J. Putzerin a paper in the American Mathematical Monthly, Vol. 73 (1966), pp. 2-7.

It is based on afamous theorem attributed to Arthur Cayley (1821-1895) and William Rowan Hamilton The Cayley-Hamilton theorem 203 (18051865) which states that every square matrix satisfies its characteristic equation.First we shall prove the Cayley-Hamilton theorem and then we shall use it to obtainPutzer's formulas for calculating etA. 7.11 The Cayley-Hamilton theorem THEOREM 7.8. CAYLEY-HAMILTON THEOREM. Let A be an n x n matrix and let (7.19) f(a) = det (az - A) = an + C,-lan-l + . e. + c,a + C, be its characteristic polynomial. Then f (A) = 0. In other words, A satisfies the equation (7.20) A" + c,~A'@ + . * * + c,A + co1 = 0. Proof. The proof is based on Theorem 3.12 which states that for any square matrix Awe have (7.21) A (cof A)t = (det A)I. We apply this formula with A replaced by ;ZI - A . Since det (AZ - A) = f(a), Equation(7.21) becomes (7.22) (a1 - A)(cof (al - ~))t =f(w. This equation is valid for all real 1. The idea of the proof is to show that it is also validwhen ;3. is replaced by A. The entries of the matrix cof (AZ - A) are the cofactors of LI - A. Except for a factorf 1 , each such cofactor is the determinant of a minor of 2I- A of order n - 1. Thereforeeach entry of cof (U - A), and hence of (cof (U - A)}*, is a polynomial in j2 of degree<_n - 1. Therefore n-1 +0f (ar - ~)y = zakBk,k=O where each coefficient Bk is an n x n matrix with scalar entries. Using this in (7.22) weobtain the relation (7.23)11-l (aI - A)Z:~~B, =f(a)zk=O which can be rewritten in the form (7.24)n - 1 S-l a"& , + - A&) - AB, = a"1 + 2 akckl i- CoI.k=l 204 Systems of diflerential equations At this stage we equate coefficients of like powers of 3, in (7.24) to obtain the equations B, , = I &-a - AB, l = c, J. (7.25) B, - AB, = c,Z -AB, = c,Z. Equating coefficients is permissible because (7.24) is equivalent to n2 scalar equations, ineach of which we may equate coefficients of like powers of il. Now we multiply the equa-tions in (7.25) in succession by A", An-l, . .
. , A, I and add the results. The terms on theleft cancel and we obtain 0 = A" + c,elAn-l + . . * + c,A + c,,I. This proves the Cayley-Hamilton theorem. Note: Hamilton proved the theorem in 1853 for a special class of matrices. A fewyears later, Cayley announced that the theorem is true for all matrices, but gave no proof. EXAMPLE. The matrix A = has characteristic polynomial f(l) = (3, - l)(L - 2)(1 - 6) = L3 - 912 + 201 - 12. The Cayley-Hamilton theorem states that A satisfies the equation (7.26) A3-9A2+20A- 12I=O. This equation can be used to express A3 and all higher powers of A in terms of I, A, andAZ. For example, we have A3 = 9A2 - 20A + 121, A4 = 9A3 - 20A2 + 12A = 9(9A2 - 20A + 121) - 20A2 + 12A = 61A2 - 168A + 1081. It can also be used to express A-l as a polynomial in A. From (7.26) we writeA(A2 - 9A + 201) = 121, and we obtain A-1 = ,+&2 - 9A + 201). Putzer's methodfor calculating etA 205 7.12 Exercises In each of Exercises 1 through 4, (a) express A-1 , A2 and all higher powers of A as a linearcombination of Z and A. (The Cayley-Hamilton theorem can be of help.) (b) Calculate et&. 5. (a) If A = , prove that et-4 a b(b) Find a corresponding formula for etA when A = [I-b a , a, b real. t t - l6. If F(t) = [I0 1 ' prove that eFtt) = eF(et-l) . 7. If A(t) is a scalar function of t, the derivative of eA(t) is eA(t)A'(t). Compute the derivative of1 teAct) when A(t) = [I0 0 and show that the result is not equal to either of the two productsetAtt)A'(t) or A'(t)eAct) . In each of Exercises 8,9, 10, (a) calculate An, and express A3 in terms of I, A, A2.
(b) CalculateetA . Lo 0 0-l I , express etA as a linear combination of I, A, A2. x2 xy y2 12.IfA= 1 2 0 ,provethateA= X2+Y2 ,wherex=coshlandy=sinhl.
0 1 2 1 0 [2Xy y2 xy 2xy 1 x2 13. This example shows that the equation eA+B = A B.e e is not always true for matrix exponentials. Compute each of the matrices eAeB , eBeA , eA+B when A = [i IO] andB = [t -i],and note that the three results are distinct. 7.13 Putzer's method for calculating etA The Cayley-Hamilton theorem shows that the nth power of any n x n matrix A can be:xpressed as a linear combination of the lower powers I, A, A2, . . . , An-l. It follows that:ach of the higher powers A"+l, An+2, . . .
, can also be expressed as a linear combination of 206 Systems of dzrerential equations I, A, AZ, . . . , An-l.

Therefore, in the infinite series defining etA, each term tkAk/k! withk 2 n is a linear combination of t"I, t"A, tkAz, . . . , t"A'+l. Hence we can expect that elAshould be expressible as a polynomial in A of the form (7.27)n-1 etA = ;oqk(f)Ak) where the scalar coefficients qk(t) depend on t. Putzer developed two useful methods forexpressing etn as a polynomial in A. The next theorem describes the simpler of the twomethods. , THEOREM 7.9. Let A, . . . , 1, be the eigenvalues of an n x n matrix A, and dejne asequence of polynomials in A as follows: (7.28) PO(A) = 1, P,(A) = fi (A - LO, for k=l,2 ,..., n.TiL=l Then we have (7.29)n-1 etA = ;dt)Pk(A) 2 where the scalar coefJicients r,(t), . . . , r,(t) are determined recursively from the system of linear differential equations G(t) = Q,(t) , r,(O) = 1 3(7.30) d+dt> = Ak+lrk+l(t) + rk(t), rk+,(o) = o 9 (k=1,2 ,..., n - l) .

Note: Equation (7.29) does not express E tA directly in powers of A as indicated in(7.271, but as a linear combination of the polynomials P,(A), P,(A), . . . , I',-l(A). Thesepolynomials are easily calculated once the eigenvalues of A are determined. Also themultipliers vi(t), . . . , r,(t) in (7.30) are easily calculated. Although this requires solvinga system of linear differential equations, this particular system has a triangular matrixand the solutions can be determined in succession. Proof. Let rl(t), . . . , r,(t) be the scalar functions determined by (7.30) and define amatrix function F by the equation (7.31)n - l Fl(t> = 2 rk+l(t)Pk(A) *k=O Note that F(0) = r,(O)P,(A) = I. We will prove that F(t) = etA by showing that Fsatisfies the same differential equation as etA, namely, F'(t) = AF(t) . Differentiating (7.31) and using the recursion formulas (7.30) we obtain w-1 n-1 F'(t) = 2 rL+l(t)Pk(A) = 2 irkct) + Ak+lrk+l(t))Pk(A) ?k=O k=O Putzer's method for calculating etA 207 where r,,(t) is defined to be 0. We rewrite this in the form n-2 n-1 F'(t) = 2 rk+l(fPk+l(4 + 2 ~k+lrk+l(f)Pk(49k=O k=O then subtract &F(t) = 2;~: ;I,r,,(t)P,(A) to obtain the relation (7.32)n-2 F'(t) - nnF(t> = c rk+l(t){Pk+l(A) + (',+I - An)Pk(A)} .k=O But from (7.28) we see that Pk+l(A) = (A - &+,Z)P,(A) , so 'k+l@) + @k+l - 'dPk(d = CA - Ak+,z)Pk(A) + @k+l - A,)Pk(& = (A - il,Z)P,(A). Therefore Equation (7.32) becomes n-2 F'(t) - &F(t) = (A - kz> 2 rk+l(f)Pk(A) = (A - &z){F(t) - rn(t)Pn-l(d)k=O = (A - &Z)F(t) - r,(t)P,(A). The Cayley-Hamilton theorem implies that P,(A) = 0, so the last equation becomes F'(t) - &F(t) = (A - il,Z)F(t) = AF(t) - &F(t), from which we find F'(t) = AF(t).
Since F(0) = I, the uniqueness theorem (Theorem 7.7)shows that F(t) = efA .

EXAMPLE I. Express etA as a linear combination of Zand A if A is a 2 x 2 matrix with bothits eigenvalues equal to il. Solution. Writing A, = 2, = 1, we are to solve the system of differential equations d(t) = J-rAO, GO = Ar,(t) + rdt), rdo) = 1, r,(O) = 0. Solving these first-order equations in succession we find rl(t) = eat, r2(t) = te"'. Since P,(A) = Z and P,(A) = A - AZ, the required formula for etA is (7.33) etA = eAtZ + teAt(A - AZ) = eAt(l - iltZ + teAtA. 208 Systems of equations EXAMPLE 2. Solve Example 1 if the eigenvalues of A are 1 and ,D, where 31 # ,u. Solution. In this case the system of differential equations is 4(t) = lr,(t>,d(t) = vdf) + rdt>, r,(O) = 1 2r,(O) = 0. Its solutions are given by rl(t) = eat,eat - eat rz(t) = ~A-p a Since P,(A) = Z and P,(A) = A - AZ the required formula for etA is (7.34) etA = eatI +eat - ept -(A-i~)=~~~~~~p~""z+~A. A-P P If the eigenvalues 1, ,LJ are complex numbers, the exponentials eLt and ept will also becomplex numbers. But if 3, and ,u are complex conjugates, the scalars multiplying Z and A in(7.34) will be real.

For example, suppose l=a+@, ~=a-$, B#O. Then 1, - ,U = 2i/I? so Equation (7.34) becomes etA = e(a+iP)tz +e(a+il)t _ eb-iPH 2iB[A - (a + i/WI e ipt _ e-iSt= eat eiStz + 2iB (A - aZ - W>I = eat((cos /?t + i sin Pt)Z + @f (A - al - iaZ)J . The terms involving i cancel and we get (7.35) etA = fjt {(/I cos /It - a sin pt)Z + sin pt A}. 7.14 Alternate methods for calculating et* in special cases Putzer's method for expressing etA as a polynomial in A is completely general because itis valid for all square matrices A. A general method is not always the simplest method touse in certain special cases. In this section we give simpler methods for computing etA inthree special cases: (a) When all the eigenvalues of A are equal, .(b) when all the eigenvaluesof A are distinct, and (c) when A has two distinct eigenvalues, exactly one of which hasmultiplicity 1. Alternate methods for calculating etA in special cases 209 THEOREM 7.10. If A is an n x n matrix with all its eigenvalues equal to 1, then we have (7.36) Proof.
Since the matrices ;3tI and t(A - AZ) commute we have m t"etA = eltlet(A-") = (eAtI)~o z (A - ill)". The Cayley-Hamilton theorem implies that (A - AZ)" = 0 for every k 2 n , so the theoremis proved.
THEOREM 7.11. If A is an n x n matrix with n distinct eigenvalues Al, 12, . .
. . , &, , thenwe have etA =$6+L,(A), where LJA) is a polynomial in A of degree n - 1 given by the formula Lk(A) = fl $+f for k = 1,2,.
. . , n. j+kk 3 Note: The polynomials L,(A) are called Lagrange interpolation coeficients. Proof. We define a matrix function F by the equation (7.37) F(t) = 2 etAkLk(A)k=l and verify that F satisfies the differential equation F'(t) = AF(t) and the initial conditionF(0) = I. From (7.37) we see that AF(t) - F'(t) = j$ etAk(A - A,I)L,(A) .k=l By the Cayley-Hamilton theorem we have (A - l&L,(A) = 0 for each k, so F satisfiesthe differential equation F'(t) = AF(t) .

To complete the proof we need to show that F satisfies the initial condition F(0) = I,which becomes (7.38) k21Lkca, = I . A proof of (7.38) is outlined in Exercise 16 of Section 7.15. The next theorem treats the case when A has two distinct eigenvalues, exactly one ofwhich has multiplicity 1. 210 Systems of diferential equations THEOREM 7.12. Let A be an n x n matrix (n 2 3) with two distinct eigenvalues 3, and ,u,where il has multiplicity n - 1 and p has multiplicity 1. Then we have etA = ,Uept eAt n-2 k ((A - V-l - (p - T1 k=O k!c t (,a - I,)" (A - IZ)n-" Proof. As in the proof of Theorem 7.10 we begin by writing m tk n-2m tk etA = ,itc ~(A-IZ)k=e"t~$(A-rll)k+e"'~h=O ' k=O ' F(A -AZ)"k=n-1 ' = ents i (A - II)" + eLt 7z (n Lny'; r) , (A - 3,Z)n-1+'.k=O ' Now we evaluate the series over r in closed form by using the Cayley-Hamilton theorem.Since we findA - ,uZ = A - izZ - (,a - J)Z (A - nZ)n-1(A - ,uZ) = (A - IZ)n - (,u - ;1)(A - ,IZ)'+l. The left member is 0 by the Cayley-Hamilton theorem so (A - U)" = (,a - ;Z)(A - IZ)+l. Using this relation repeatedly we find (A - j[Z)~l+r = (,a - /?))'(A - IZ)'+l. Therefore the series over r becomes ccc p--l+?. 7=. (n - 1 + r>! cp _A)T(A _nz)"-1 = 1 2 _4_" (p - I)k(A - AZ)'+'(P - r1 k=n-l IQ! This completes the proof. The explicit formula in Theorem 7.12 can also be deduced by applying Putzer's method,but the details are more complicated. The explicit formulas in Theorems 7.10, 7.11 and 7.12 cover all matrices of order n < 3 .Since the 3 x 3 case often arises in practice, the formulas in this case are listed below foreasy reference. CASE 1. If a 3 x 3 matrix A has eigenvalues 1, 1, 1, then etA = eat{Z + t(A - AZ) + &@A - U)2}. Exercises 211 CASE 2. If a 3 x 3 matrix A has distinct eigenvalues 1, ,u, v, then eta = eat (A - PNA - ~1) + ept (A - WA - ~1) + evt (A - WA - ~1)*(1 - Pu)Q - 4 (P - m - 4 (v - WV - P) CASE 3.

If a 3 x 3 matrix A has eigenvalues A, 1, ,u, with ;Z # ~1, then etA = eAt{Z + t(A - AZ)) + ~ept - 8 (A - AZ)2 - te"t (A - 11)2.(P - 9" P-1 EXAMPLE. Compute etAi when A = Solution. The eigenvalues of A are 1, 1, 2, so the formula of Case 3 gives us (7.39) etA = et{Z + t(A - Z)} + (e2t - et)(A - I)" - tet(A - I)". By collecting powers of A we can also write this as follows, (7.40) etA = (-2tet + e2t)Z + ((3t + 2)et - 2e2t}A - {(t + l)et - ezt}A2. At this stage we can calculate (A - Z)2 or A2 and perform the indicated operations in (7.39)or (7.40) to write the result as a 3 x 3 matrix, -2tet + ezt (3t + 2)et - 2e2t - (t + l)et + e2t et-4 = -2(t + l)et + 2e2t (3t + 5)et - 4e2t -(t + 2)et + 2e2t . -2(t + 2)et + 4e2t (3t + 8)et - 8e2t -(t + 4)& + 4e2t 17.15 Exercises For each of the matrices in Exercises 1 through 6, express etA as a polynomial in A. 1. A = I5 -'1.L4 -1 l ,A=[: ;I. 3.A=[% :]. 1. 5.A=l 1; :]. 6.A=[; ; 3. -6 -11 -6 7. (a) A 3 x 3 matrix A is known to have all its eigenvalues equal to 1. Prove that eta = &eit{(A2t2 - 21t + 2)Z + (-21t2 + 2t)A + t2A2}, (b) Find a corresponding formula if A is a 4 x 4 matrix with all its eigenvalues equal to A. 212 Systems of diferential equations In each of Exercises 8 through 15, solve the system Y' = A Y subject to the given initial 3 10. A = i 2 - 1 1 0 1 1 , 1 - 1 2 condition. ,.A=[; -;I, -6 -11 -6 1 1 0 0 0 2 1 0 14. A = I 10 0 3 0' 0 o-o 4 0 0 2 0 1 0 0 215. A = I = 0 0 01 4 'Y(0) 0 0 1 0 16. This exercise outlines a proof of Equation (7.38) used in the proof of Theorem 7.11,be the polynomial in A of degree n - 1 defined by the equation .8 . I 0 . 0 '1 0 2 * .1 1Let L,(I) i#k where 1,) . . . , I, are n distinct scalars.(a) Prove that (0 if li # A,,Lk(A) = 1 if li = A,.
(b) Letyl,..., yn be n arbitrary scalars, and let p@) = $ykLk(@).k=l Prove that p(A) is the only polynomial of degree I n - 1 which satisfies the n equations p('k) = yk for k=1,2 ,..., n. (c) Prove that zzCl &(A) = 1 for every 1, and deduce that for every square matrix A we have where Z is the identity matrix. Nonhomogeneous linear systems with constant coe$cients 213 7.16 Nonhomogeneous linear systems with constant coefficients We consider next the nonhomogeneous initial-value problem (7.41) Y'(t) = A Y(t) + Q(t), Y(a) = B, on an interval J. Here A is an n x n constant matrix, Q is an n-dimensional vector function(regarded as an n x I column matrix) continuous on J, and a is a given point in J. We canobtain an explicit formula for the solution of this problem by the same process used totreat the scalar case. First we multiply both members of (7.41) by the exponential matrix e-tA and rewrite thedifferential equation in the form (7.42) eMtA{ Y'(t) - AY(t)) = ePtAQ(t). The left member of (7.42) is the derivative of the product ePtAY(t). Therefore, if weintegrate both members of (7.42) from a to x, where x E J, we obtain e-"AY(x) - PAY(a) = s," eetAQ(t) dt . Multiplying by ezA we obtain the explicit formula (7.43) which appears in the followingtheorem. THEOREM 7.13. Let A be an n x n constant matrix and let Q be an n-dimensional vectorfunction continuous on an interval J. Then the initial-value problem Y'(t) = A Y(t) + Q(t), Y(a) = B, has a unique solution on J given by the explicit formula (7.43) Y(x) = eczmajAB + esA ,' e&Q(t) dt.s As in the homogeneous case, the difficulty in applying this formula in practice lies in thecalculation of the exponential matrices. Note that the first term, e(zpa)AL3, is the solution of the homogeneous problem Y'(t) =A Y(t), Y(a) = B. The second term is the solution of the nonhomogeneous problem Y'(t) = A Y(t) + Q(t), Y(a) = 0. We illustrate Theorem 7.13 with an example. EXAMPLE. Solve the initial-value problem Y'(t) = AY(t) + Q(t), Y(0) = B, 214 Systems of difSerentia1 equations on the interval (- co , + co), where Solution. According to Theorem 7.13, the solution is given by (7.44) Y(x) = exa lo' (t) dt = IO' ecz-t'AQ(t) dt . The eigenvalues of A are 2, 2, and 4. To calculate exa we use the formula of Case 3,Section 7.14, to obtain e 'A = e'"{Z + x(A - 21)) + $(e4% - e'")(A - 21)' - +xe'"(A - 21)2 = e2"{Z + x(A - 21) + $(e2' - 2x - l)(A - 21)'). We can replace x by x - t in this formula to obtain e cr-tln. Therefore the integrand in(7.44) is ecz-t'AQ(t) = ezczpt'{Z + (x - t)(A - 2Z) + i[e2("-t) - 2(x - t) - l](A - 21)2} Q(t) Integrating, we find Y(x) = ze'"-t'"Q(t) dt = e2xs0 [$I + (A - 20e2'[-~-] + $A +e2x -$-x-x" - 2Z)2e2' 0*,23: _* - ix - ix2 - 4x3 Since we have we find The rows of this matrix are the required functions yl, y2, y3. Exercises 215 7.17 Exercises 1. Let 2 be a solution of the nonhomogeneous system Z'(r) = AZ(r) + Q(r), on an interval J with initial value Z(a). Prove that there is only one solution of the non-homogeneous system Y'(t) = AY(t) + Q(t) on J with initial value Y(a) and that it is given by the formula Y(t) = Z(t) + d-4{ Y(a) - Z(a)}. Special methods are often available for determining a particular solution Z(t) which resemblesthe given function Q(t). Exercises 2, 3, 5, and 7 indicate such methods for Q(t) = C, Q(r) =&C, Q(t) = tmC, and Q(r) = (cos cct)C + (sin ar)D, where C and D are constant vectors.If the particular solution Z(t) so obtained does not have the required initial value, we modifyZ(t) as indicated in Exercise 1 to obtain another solution Y(r) with the required initialvalue. 2. (a) Let A be a constant n x n matrix, B and C constant n-dimensional vectors. Prove that thesolution of the system Y'(f) = AY(f) + c, Y(a) = B, on (- 00, + to) is given by the formula Y(x) = e(=-a)AB + C . (b) If A is nonsingular, show that the integral in part (a) has the value {ecz--a)A - Z}A-l.(c) Compute Y(x) explicitly when A=[-: :I, C=[:], B=[:], a=O. 3. Let A be an n x n constant matrix, let B and C be n-dimensional constant vectors, and let abe a given scalar.(a) Prove that the nonhomogeneous system Z'(t) = AZ(t) + eW has a solution of the formZ(t) = eatB if, and only if, (aZ - A)B = C.(b) If a is not an eigenvalue of A, prove that the vector B can always be chosen so that thesystem in (a) has a solution of the form Z(t) = eztB .(c) If cc is not an eigenvalue of A, prove that every solution of the system Y'(t) = ,4 Y(i) + eZtChas the form Y(t) = elA(Y(0) - B) + eatB, where B = (al - A)-lC. 4. Use the method suggested by Exercise 3 to find a solution of the nonhomogeneous systemY'(t) = AY(t) + eztC, with A=[: ;I, c=[-;I. Y(@=[,]. 216 Systems of difSerential equations 5. Let A be an n x n constant matrix, let B and C be n-dimensional constant vectors, and letm be a positive integer.(a) Prove that the nonhomogeneous system Y'(t) = A Y(t) + PC, Y(0) = B, has a particularsolution of the form Y(f) = B, + tB, + t2B, + ... + PB,, whereB,,B,,..., B, are constant vectors, if and only if 1C = - - Am+'B?72! Determine the coefficients B,, B, , . . . , B, for such a solution.(b) If A is nonsingular, prove that the initial vector B can always be chosen so that the systemin (a) has a solution of the specified form. 6. Consider the nonhomogeneous system y; = 3yl + yz + t3 y;: = 2Yl + 2y, + t3. (a) Find a particular solution of the form Y(f) = B, + tB, + t2B2 + t3B3.(b) Find a solution of the system with ~~(0) = ~~(0) = 1 . 7.

Let A be an II x n constant matrix, let B, C, D be n-dimensional constant vectors, and letu be a given nonzero real number. Prove that the nonhomogeneous system Y'(t) = A Y(t) + (cos cct)C + (sin at)D, Y(0) = B, has a particular solution of the form Y(f) = (cos ut)E + (sin ur)F, where E and Fare constant vectors, if and only if (A2 + a2Z)B = -(AC + MD). Determine E and Fin terms of A, B, C for such a solution. Note that if A2 + a21 is nonsingular,the initial vector B can always be chosen so that the system has a solution of the specifiedform. 8. (a) Find a particular solution of the nonhomogeneous system (b) Find a solution of the system with J+(O) = ~~(0) = 1 . The general linear system Y'(t) = P(t) Y(t) + Q(t) 217 In each of Exercises 9 through 12, solve the nonhomogeneous system Y'(t) = A Y(t) + Q(t)subject to the given initial condition.

7.18 The general linear system Y'(t) = P(QY(t) + Q(t) Theorem 7.13 gives an explicit formula for the solution of the linear system Y'(t) = A Y(t) + Q(t), Y(u) = B, where A is a constant n x IZ matrix and Q(t), Y(t) are n x 1 column matrices. We turnnow to the more general case (7.45) Y'(t) = W> Y(t) + Q(t) > Y(a) = B, where the IZ x IZ matrix P(t) is not necessarily constant.If P and Q are continuous on an open interval J, a general existence-uniqueness theorem which we shall prove in a later section tells us that for each a in J and each initial vector Bthere is exactly one solution to the initial-value problem (7.45).

In this section we use thisresult to obtain a formula for the solution, generalizing Theorem 7.13. In the scalar case (n = 1) the differential equation (7.45) can be solved as follows. Welet A(x) = J; P(t) dt, then multiply both members of (7.45) by e--A't) to rewrite the differ-ential equation in the form (7.46) emAct){ Y'(t) - P(t)Y(t)} = eeAct)Q(t). Now the left member is the derivative of the product e- ACt) Y(t). Therefore, we can integrateboth members from a to x, where a and x are points in J, to obtain e-A(r)y(x) - epAta)y(a) = s cc eeA'"'Q(t) dt.aMultiplying by eAfz) we obtain the explicit formula (7.47) Y(x) = e~(de-~(dy(u) + &kd ,; e--B(t)Q(t) &.

218 Systems of equations The only part of this argument that does not apply immediately to matrix functions isthe statement that the left-hand member of (7.46) is the derivative of the product e--A(tJ Y(t).At this stage we used the fact that the derivative of e-nctt is -P(t)e-A'l'. In the scalar casethis is a consequence of the following formula for differentiating exponential functions : I f E(t) = eaft), then E'(t) = Ar(t)eAc Unfortunately, this differentiation formula is not always true when A is a matrix function.1 t For example, it is false for the 2 x 2 matrix function A(t) = [10 0 . (See Exercise 7 of Section 7.12.) Therefore a modified argument is needed to extend Equation (7.47) to thematrix case. Suppose we multiply each member of (7.45) by an unspecified n x n matrix F(t). Thisgives us the relation F(t) Y'(t) = F(t)P(t) Y(t) + F(t>Q(t> . Now we add F'(t)Y(t) to both members in order to transform the left member to thederivative of the product F(t) Y(t). If we do this, the last equation gives us {F(t) r(t)}' = {F'(t) + f'Ct)f'(t)} Y(t) + FWQ(t) . If we can choose the matrix F(t) so that the sum {F'(t) + F(t)P(t)} on the right is the zeromatrix, the last equation simplifies to {F(t) Y(t))' = FWQ(t) . Integrating this from a to x we obtain F(x)Y(x) - F(a)Y(a) = 1: F(t)Q(t) dt . If, in addition, the matrix F(x) is nonsingular, we obtain the explicit formula (7.48) Y(x) = F(x)-'F(a) Y(a) + F(x)-' J: F(t)Q(t) dt . This is a generalization of the scalar formula (7.47). The process will work if we can find an n x n matrix function F(t) which satisfies the matrix differential equation and which is nonsingular. F'(t) = -F(t)P(t) Note that this differential equation is very much like the original differential equation(7.45) with Q(t) = 0, except that the unknown function F(t) is a square matrix instead of acolumn matrix. Also, the unknown function is multiplied on the right by -P(t) instead ofon the left by P(t). We shall prove next that the differential equation for F always has a nonsingular solution.The proof will depend on the following existence theorem for homogeneous linear systems. The general linear system Y'(t) = P(t) Y(t) + Q(t) 219 THEOREM 7.14. Assume A(t) is an n x n matrix function continuous on an open intervalJ. If a E J and if B is a given n-dimensional vector, the homogeneous linear system Y'(t) = A(t) Y(t), Y(a) = B, has an n-dimensional vector solution Y on J. A proof of Theorem 7.14 appears in Section 7.21. With the help of this theorem we canprove the following. THEOREM 7.15. Given an n x n matrix function P, continuous on an open interval J,and given any point a in J, there exists an n x n matrix function F which satisj?es the matrixdifSerentia1 equation (7.49) F'(x) = -F(x)P(x) on J with initial value F(a) = I. Moreover, F(x) is nonsingular for each x in J. Proof. Let Y,(x) be a vector solution of the differential equation y;(x) = -P(x)"Y,(x) on J with initial vector Y,(a) = Z, , where Ik is the kth column of the n x n identity matrix I.Here P(x)" denotes the transpose of P(x). Let G(x) be the n x n matrix whose kth columnis Y,(x). Then G satisfies the matrix differential equation (7.50) G'(x) = -PUG on J with initial value G(a) = I. Now take the transpose of each member of (7.50). Sincethe transpose of a product is the product of transposes in reverse order, we obtain {G'(x)}" = -G(x)"P(x). Also, the transpose of the derivative G' is the derivative of the transpose Gt. Therefore thematrix F(x) = G(x)l satisfies the differential equation (7.49) with initial value F(a) = I. Now we prove that F(x) is nonsingular by exhibiting its inverse.

Let H be the n x nmatrix function whose kth column is the solution of the differential equation Y'(x) = P(x) Y(x) with initial vector Y(a) = I,, the kth column of I. Then H satisfies the initial-value problem H'(x) = P(x)H(x), H(a) = I, on J.
The product F(x)H(x) has derivative F(x)H'(x) + k'(x)H(x) = F(x)P(x)H(x) - F(x)P(x)H(x) = 0 220 Systems of dtrerential equations for each x in J. Therefore the product F(x)H(x) is constant, F(x)H(x) = F(a)H(a) = I,so H(x) is the inverse of F(X). This completes the proof. The results of this section are summarized in the following theorem. THEOREM 7.16. Given an n x n matrix function P and an n-dimensional vector functionQ, both continuous on an open interval J, the solution of the initial-value problem (7.51) r'(x) = P(x) Y(x) + Q(x), Y(a) = B, on J is given by the formula (7.52) Y(x) = F(x)-'Y(a) + F(x)-l i% F(t)Q(t) dt. The n x n matrix F(x) is the transpose of the matrix whose kth column is the solution of theinitial-value problem (7.53) Y'(x) = -P(x)"Y(x), Y(a) = 4, where I, is the kth column of the identity matrix I. Although Theorem 7.16 provides an explicit formula for the solution of the generallinear system (7.51), the formula is not always a useful one for calculating the solutionbecause of the difficulty involved in determining the matrix function F. The determinationof F requires the solution of n homogeneous linear systems (7.53).

The next sectiondescribes a power-series method that is sometimes used to solve homogeneous linearsystems. We remind the reader once more that the proof of Theorem 7.16 was based on Theorem7.14, the existence theorem for homogeneous linear systems, which we have not yet proved.
7.19 A power-series method for solving homogeneous linear systems Consider a homogeneous linear system (7.54) Y'(x) = A(x) Y(x)) Y(0) = B, in which the given n x n matrix A(x) has a power-series expansion in x convergent insome open interval containing the origin, say A(x) = A, + XL41 + X2A, + * ' * + X'CA, + .
. . , for 1x1 < rl,where the coefficients AO, A,, A,, . . . are given n x n matrices. Let us try to find apower-series solution of the form Y(X) = B, + xB, + x2B, + * * * + x"Bk + * * * , Exercises 221 with vector coefficients B,,, B,, B,, ... Since Y(0) = B,, the initial condition will besatisfied by taking B, = B, the prescribed initial vector. To determine the remainingcoefficients we substitute the power series for Y(x) in the differential equation and equatecoefficients of like powers of x to obtain the following system of equations: (7.55) B, = A,B, (k + l)&+, = i A,~,-r f o r k= 1,2,....7=0 These equations can be solved in succession for the vectors B,, B,, .
.

. . If the resultingpower series for Y(x) converges in some interval 1x1 < r2, then Y(x) will be a solution of theinitial-value problem (7.54) in the interval 1x1 < r, where r = min {rl, r2}. For example, if A(x) is a constant matrix A, then A, = A and A, = 0 for k 2 1, so thesystem of equations in (7.55) becomes B,=AB, (k + l)B,+, = AB, f o r k>l. Solving these equations in succession we find B, = -& AkB f o r k>l.

Therefore the series solution in this case becomes ycx) = B +kz$AkB = ezAB. This agrees with the result obtained earlier for homogeneous linear systems with constantcoefficients. 7.20 Exercises 1. Let p be a real-valued function and Q an it x 1 matrix function, both continuous on an intervalJ. Let A be an n x n constant matrix. Prove that the initial-value problem has the solutionY'(x) = ~(4-4 Y(x) + Qt.4 , Y(u) = E, y(x) = edz)AB + @(%)As r e-~(t)AQ(t) &a on J, where q(x) = Jzp(t) dt.2. Consider the special case of Exercise 1 in which A is nonsingular, a = 0, p(x) = 2x, and Q(x) = XC, where C is a constant vector. Show that the solution becomes Y(x) = ezzA(B + $A-%') - $A-lC. 3. Let A(t) be an n x n matrix function and let E(t) = eBtt). Let Q(t), Y(r), and B be n x 1column matrices. Assume that E'(t) = A'(t)E(t) 222 Systems of difSerentia1 equations on an open interval J. If a E J and if A' and Q are continuous on J, prove that the initial-valueproblem Y ' (t) = A'(t)Y(t) + Q(t), Y(u) = B, has the following solution on J: y(x) = eAh)e-Ah)B + eAh) ' e-A(t)Q(t) &.ia Let E(t) = eBtt). This exercise describes examples of matrix functions A(t) for which E'(t) =A'(t)E(t).(a) Let A(t) = l"A, where A is an n x n constant matrix and r is a positive integer. Prove thatE'(t) = A'(t)E(t) on (- co, w).(b) Let A(t) be a polynomial in t with matrix coefficients, say A(t) = 2 PA,,r=O where the coefficients commute, A,A, = A,A, for all r and S. Prove that E'(t) = A'(t)E(t) on(- a, a).(c) Solve the homogeneous linear system Y'(r) = (I + tA)Y(t), Y(0) = B on the interval (- KJ, co), where A is an II x n constant matrix.Assume that the n x n matrix function A(x) has a power-series expansion convergent for1x1 Cr. Develop a power-series procedure for solving the following homogeneous linear systemof second order: Y"(x) = A(x)Y(x), with Y(0) = B, Y'(0) = c. Consider the second-order system Y"(x) + A Y(x) = 0, with Y(0) = B, Y'(0) = C, where Ais a constant II x n matrix. Prove that the system has the power-series solution convergent for - co < x < + * I 7.21 Proof of the existence theorem by the method of successive approximations In this section we prove the existence and uniqueness of a solution for any homogeneouslinear system (7.56) Y'(t) = A(t) Y(t), where A(t) is an n x n matrix function, continuous on an open interval J. We shall provethat for any point a in J and any given initial-vector B there exists exactly one solution Y(t)on J satisfying the initial condition Y(a) = B . Proof of the existence theorem by the method of successive approximations 223 We shall use the method of successive approximations, an iterative method which also hasapplications in many other problems.
The method was first published by Liouville in 1838in connection with the study of linear differential equations of second order. It was laterextended by J. CaquC in 1864, L. Fuchs in 1870, and G. Peano in 1888 to the study of linearequations of order n. In 1890 Smile Picard (18561941) extended the method to encompassnonlinear differential equations as well. In recognition of his fundamental contributions,some writers refer to the method as Picard's method. The method is not only of theoreticalinterest but can also be used to obtain numerical approximations to solutions in some cases. The method begins with an initial guess at a solution of the equation (7.56). We take asinitial guess the given initial vector B, although this is not essential. We then substitutethis guess in the right-hand member of the equation and obtain a new differential equation, Y'(t) = A(t)B. In this equation the right-hand member no longer contains the unknown function, so theequation can be solved immediately by integrating both members from a to x, where x isan arbitrary point in J. This equation has exactly one solution Y, on J satisfying the initialcondition Y,(a) = B, namely Y,(x) = B + j-; A(t)B dt. Now we replace Y(t) by Yl(t) in the right-hand member of the original differentialequation (7.56) to obtain a new differential equation Y'(t) = A(t)Y,(t). This equation has a unique solution Y, on J with Y, (a) = B, (7.57) Ydx) = B + j-; A(t)Y,(t) dt. We then substitute Yz in the right-hand member of (7.56) and solve the resulting equationto determine Ya with Y,(a) = B, and so on. This process generates a sequence of functionsy,, y,, yz, * * * > where Y, = B and where Yti+i is determined from Y, by the recursionformula (7.58) Y,c+dx) = B + j-;, A(t)Y,(t) dt for k = 0, 1, 2, ... Our goal is to prove that the sequence of functions so defined converges to a limitfunction Ywhich is a solution of the differential equation (7.56) on Jand which also satisfiesthe initial condition Y(a) = B. The functions Y,, Y, , Yz , . . . are called successive approxi-mations to Y. Before we investigate the convergence of the process we illustrate the methodwith an example. EXAMPLE. Consider the initial-value problem Y'(t) = A Y(t), Y(0) = B, where A is aconstant IZ x n matrix. We know that the solution is given by the formula Y(x) = exAIBfor all real x. We will show how this solution can be obtained by the method of successiveapproximations. 224 Systems of equations The initial guess is Y,(x) = B. The recursion formula (7.58) gives us Y,(x) = B + j;ABdt = B + xAB, Y,(x) = B + j; AY,(t) dt = B + jOz (AB + tA2B) dt = B + xAB + +x2A2B, By induction we find Y,(x) = B + xAB + $x2A2B + . 1 . The sum on the right is a partial sum of the series for e"-j. Therefore when k + CO we find lim Y,(x) = eeriBk-tm for all x. Thus, in this example we can show directly that the successive approximationsconverge to a solution of the initial-value problem on (- co, + oo). Proof of convergence of the sequence of successive approximations. We return now to thegeneral sequence defined by the recursion formula (7.58). To prove that the sequenceconverges we write each term Y,(x) as a telescoping sum, (7.59)k - l yk(x) = y,cx) + m&' ym+l(x) - y,n} * To prove that Yk(x) tends to a limit as k + co we shall prove that the infinite series (7.60) converges for each x in J. For this purpose it suffices to prove that the series (7.61) f, II Ym+l(X) - Ymc4ll converges. In this series we use the matrix norm introduced in Section 7.3; the norm of amatrix is the sum of the absolute values of all its entries. Consider a closed and bounded subinterval J1 of J containing a. We shall prove that forevery x in J1 the series in (7.61) is dominated by a convergent series of constants inde-pendent of x. This implies that the series converges uniformly on J1, To estimate the size of the terms in (7.61) we use the recursion formula repeatedly.Initially, we have Y,(x) - Ye(x) = j,z A(t)B dt. Proof of the existence theorem by the method of successive approximations 225 For simplicity, we assume that a < x . Then we can write (7.62) Since each entry of A(t) is continuous on J, each entry is bounded on the closed boundedinterval J1. Therefore jlA(t)ll 5 M, where A4 is the sum of the bounds of all the entries ofA(t) on the interval J1. (The number A4 depends on J1 .) Therefore the integrand in (7.62)is bounded by llB\l M, so we have ll~(~> - Y,(x)/I I jaz 11~~1 M dt = IIBII M(x - a) for all x > a in J1.Now we use the recursion formula once more to express the difference Y, - Y, in terms of Y, - Y,, and then use the estimate just obtained for Y1 - Y, to obtain lj Y,(X) - Y,(x)II = 11 f 4O{Y,(t) - Yo(O} dt 11 I 1," IIA(t>ll IIBII M(t - a> dl <))B)) M2 1," (t - a) dt = jJB/J M2(x2; a)2 for all x > a in J1. By induction we find IIY,+,(x) - Y,(x)ll 5 IlBll M"+l(x - a)m+l for(m + I)! m = 0, 1, 2, . . . , and for all x > a in J1. If x < a a similar argument gives the same inequality with Ix - alappearing instead of (x - a). If we denote by L the length of the interval J1, then we haveIx - al < L for all x in J1 so we obtain the estimate I I Y,+,(x) - Y&N I IIBII yn;;;,' for m = 0, 1, 2, . .
. , and for all x in J1. Therefore the series in (7.61) is dominated by the convergent series This proves that the series in (7.61) converges uniformly on J1.The foregoing argument shows that the sequence of successive approximations always converges and the convergence is uniform on J1. Let Y denote the limit function. That is,define Y(x) for each x in J1 by the equation Y(x) = lim Y,(x).h--+m 226 Systems of d$erential equations We shall prove that Y has the following properties:(a) Y is continuous on Jr.(b) Y(x) = B + &4(t)Y(t)dt for all x in J1.(c) Y(a) = B and Y'(x) = A(x)Y(x) for all x in J1. Part (c) shows that Y is a solution of the initial-value problem on J1. Proof of (a). Each function Y, is a column matrix whose entries are scalar functions,continuous on J1. Each entry of the limit function Y is the limit of a uniformly convergentsequence of continuous functions so, by Theorem 11.1 of Volume I, each entry of Y is alsocontinuous on J1. Therefore Y itself is continuous on J1. Proof of(b). The recursion formula (7.58) states that Therefore y,+,(x) = B + j; A(t)Y,(t) dt .
Y(x) = lim Yk+r(x) = B + limk-too 1" A(t)Y,(t) dt = B + 1," A(t)lim Yk(t) dtk+cc a k-tm = B + j; A(t)Y(t) dt . The interchange of the limit symbol with the integral sign is valid because of the uniformconvergence of the sequence {Y,} on J1. Proof of (c). The equation Y(a) = B follows at once from (b). Because of (a), theintegrand in (b) is continuous on J1 so, by the first fundamental theorem of calculus, Y'(x)exists and equals A(x) Y(x) on J1. The interval J1 was any closed and bounded subinterval of J containing a. If J1 isenlarged, the process for obtaining Y(x) doesn't change because it only involves integrationfrom a to x. Since for every x in J there is a closed bounded subinterval of J containing aand x, a solution exists over the full interval J. THEOREM 7.17. UNIQUENESS THEOREM FOR HOMOGENEOUS LINEAR SYSTEMS. Zf A(t) iscontinuous on an open interval J, the difSerentia1 equation Y'(t) = A(t) Y(t) has at most one solution on J satisfying a given initial condition Y(a) = B . Proof. Let Y and Z be two solutions on J. Let J1 be any closed and bounded subintervalof J containing a. We will prove that Z(x) = Y(x) for every x in J1. This implies thatZ = Y on the full interval J. Since both Y and Z are solutions we have Z'(t) - Y'(t) = A(t){Z(t) - Y(t)}. The method of successive approximations applied tojrst-order nonlinear systems 227 Choose x in J1 and integrate this equation from a to x to obtain Z(x) - Y(x) = j-l A(t){Z(t) - Y(t)} dt. This implies the inequality (7.63) IIZW - Y(x)ll I ~4 11," llz(t> - Y(t)ll dt 1, where Mis an upper bound for \lA(t)li onJ,. Let MI be an upper bound for the continuousfunction /Z(t) - Y(t)11 on J1. Then the inequality (7.63) gives us (7.64) IlZ(4 - Y(x)ll I MM1 Ix - 4.
Using (7.64) in the right-hand member of (7.63) we obtain ljZ(x) - Y(x)]1< M2M, 16 It - al dt 1 = M'M, y. By induction we find (7.65) IIZ(x) - Y(x)11 5 MrnMl 7. When m --f co the right-hand member approaches 0, so Z(x) = Y(x). This completes theproof. The results of this section can be summarized in the following existence-uniquenesstheorem.
THEOREM 7.18. Let A be an n x n matrix function continuous on an open interval J.If a E J and if B is any n-dimensional vector, the homogeneous linear system Y'(t) = A(t) Y(t), Y(a) = B, has one and only one n-dimensional vector solution on J. 7.22 The method of successive approximations applied to first-order nonlinear systems The method of successive approximations can also be applied to some nonlinear systems.Consider a first-order system of the form (7.66) Y' = F(t, Y), where Fis a given n-dimensional vector-valued function, and Yis an unknown n-dimensionalvector-valued function to be determined. We seek a solution Y which satisfies the equation Y'(t) = F[t, Y(t)] 228 Systems of d$erential equations for each t in some interval J and which also satisfies a given initial condition, say Y(a) = B,where a E J and B is a given n-dimensional vector. In a manner parallel to the linear case, we construct a sequence of successive approxi-mations Y,, Y,, Y,, . . . , by taking Y, = B and defining Y,,, in terms of Y, by therecursion formula (7.67)
z',+,(x) = B + 1: Fit, y,(t)1 dt for k = 0, 1, 2, . . .

. Under certain conditions on F, this sequence will converge to a limit function Y which willsatisfy the given differential equation and the given initial condition. Before we investigate the convergence of the process we discuss some one-dimensionalexamples chosen to illustrate some of the difficulties that can arise in practice. EXAMPLE 1. Consider the nonlinear initial-value problem y' = x2 + y2 withy = 0 whenx = 0. We shall compute a few approximations to the solution. We choose Y,,(x) = 0and determine the next three approximations as follows: Y,(x) =s0 't2 dt = 5 , Yz(x)=/jz+ Y;(t)]dt=[(t'+:)dt++;, v,(x)=[[t'+ (;+di']dt=:+$+g+-&. It is now apparent that a great deal of labor will be needed to compute further approxi-mations. For example, the next two approximations Y, and Y, will be polynomials ofdegrees 3 1 and 63, respectively. The next example exhibits a further difficulty that can arise in the computation of thesuccessive approximations. EXAMPLE 2. Consider the nonlinear initial-value problem y' = 2x + e" , withy = 0 whenx = 0.
We begin with the initial guess Y,(x) = 0 and we find Ye = if (2t + 1) dt = x2 + x, Y2(x).
= 6 (2t + eztt) dt = x' + 1," et2+t dt . Here further progress is impeded by the fact that the last integral cannot be evaluated interms of elementary functions. However, for a given x it is possible to calculate a numericalapproximation to the integral and thereby obtain an approximation to Y,(x). Because of the difficulties displayed in the last two examples, the method of successiveapproximations is sometimes not very useful for the explicit determination of solutionsin practice.

The real value of the method is its use in establishing existence theorems. Proof of an existence-uniqueness theorem for @t-order nonlinear systems 229 7.23 Proof of an existence-uniqueness theorem for first-order nonlinear systems We turn now to an existence-uniqueness theorem for first-order nonlinear systems. Byplacing suitable restrictions on the function which appears in the right-hand member of thedifferential equation y' = Ft'(x, Y>, we can extend the method of proof used for the linear case in Section 7.21.Let Jdenote the open interval over which we seek a solution. Assume a E J and let B be a given n-dimensional vector. Let S denote a set in (n + 1)-space given by S = {(x, Y) I Ix - al I h, II Y - BII I k) , where h > 0 and k > 0. [If n = 1 this is a rectangle with center at (a, B) and with base 2hand altitude 2k.l We assume that the domain of F includes a set S of this type and that F isbounded on S, say (7.68) Ills(x, VII I ~4 for all (x, Y) in S, where M is a positive constant.Next, we assume that the composite function G(x) = F(x, Y(x)) is continuous on the interval (a - h , a + h) for every function Y which is continuous on (a - h , a + h) andwhich has the property that (x, Y(x)) E S for all x in (a - h , a + h). This assumptionguarantees the existence of the integrals that occur in the method of successive approxi-mations, and it also implies continuity of the functions so constructed. Finally, we assume that F satisfies a condition of the form IIF(x> Y> - F(x, 2111 I A II Y - 41 for every pair of points (x, Y) and (x, 2) in S, where A is a positive constant. This is calleda condition in honor of Rudolph Lipschitz who first introduced it in 1876. ALipschitz condition does not restrict a function very seriously and it enables us to

extend theproof of existence and uniqueness from the linear to the nonlinear case. THEOREM 7.19. EXISTENCE AND UNIQUENESS OF SOLUTIONS TO FIRST ORDER NONLINEAR SYSTEMS. Assume F satisjies the boundedness, continuity, and Lipschitz conditions speciJiedabove on a set S.

Let I denote the open interval (a - c , a + c), where c = min {h, k/M}.Then there is one and onZy one function Y de$ned on Z with Y(a) = B such that (x, Y(x)) E Sand Y'(x) = F(x, Y(x)) for each x in I. Proof. Since the proof is analogous to that for the linear case we sketch only theprincipal steps. We let Y,,(x) = B and define vector-valued functions Y, , Y,, . . . on Z bythe recursion formula (7.69) Y,+,(x) = B + j: FLt, Ym(Ol dt for m = 0, 1,2, ... 230 Systems of differential equations For the recursion formula to be meaningful we need to know that (x, Y,(x)) E S for eachx in I. This is easily proved by induction on m. When m = 0 we have (x, Y,,(x)) = (x, B),which is in S. Assume then that (x, Y,(x)) E S for some m and each x in I. Using (7.69)and (7.68) we obtain llY, +dx) - BII I J;" llF[t, L(t>lll dt 1 I M 1 1: dt 1 = M Ix - al. Since (x - al < c for x in I, this implies that which shows that (x, Y,+,(X)) E S for each x in I. Therefore the recursion formula ismeaningful for every m 2 0 and every x in I. The convergence of the sequence {Y,(x)} is now established exactly as in Section 7.21.We write k-l yk(x) = y,tx) + z: iym+l(.x> - ym(x)}Wh=O and prove that Y,(x) tends to a limit as k --f 00 by proving that the infinite series g Ym+*(x) - Ym(x>ll converges on I.

This is deduced from the inequality llym+l(x) - Y&>l/ s MA;;x+-I;,lmi- 1 IMAmcm+' (m + 1)! which is proved by induction, using the recursion formula and the Lipschitz condition.We then define the limit function Y by the equation Y(x) = lim Y,(x)m-m for each x in I and verify that it satisfies t,he integral equation Y(x) = B + J^," F[t, Y(t)] dt, exactly as in the linear case. This proves the existence of a solution. The uniqueness maythen be proved by the same method used to prove Theorem 7.17. 7.24 Exercises 1. Consider the linear initial-value problem y'+y =2e", with y = 1 when x = 0. Exercises 2 3 1 (a) Find the exact solution Y of this problem.(b) Apply the method of successive approximations, starting with the initial guess YO(x) = 1.Determine Y,(x) explicitly and show that for all real x. lim Y,(x) = Y(x)ne+'x 2. Apply the method of successive approximations to the nonlinear initial-value problem y' =x +y2, with y =Owhenx =O. Take Y,(x) = 0 as the initial guess and compute Y,(x).3. Apply the method of successive approximations to the nonlinear initial-value problem y'= 1 +xya, with y =Owhenx =O. Take Y,(x) = 0 as the initial guess and compute Y,(x).4. Apply the method of successive approximations to the nonlinear initial-value problem y'=x2 +y2, with y=Owhenx =O. Start with the "bad" initial guess Y,(x) = 1, compute Ys(x), and compare with the results ofExample 1 in Section 7.22.

5. Consider the nonlinear initial-value problem y' =x2 +y2, w i t h y=lwhenx=O. (a) Apply the method of successive approximations, starting with the initial guess Y,(x) = 1,and compute Y,(x).(b) Let R = [- 1 , I] x [- 1 , 11. Find the smallest Msuch that jf(x, y)l 5 Mon R. Find aninterval Z = (-c , c) such that the graph of every approximating function Y, over Z will liein R.(c) Assume the solution y = Y(x) has a power-series expansion in a neighborhood of theorigin. Determine the first six nonzero terms of this expansion and compare with the result ofpart (a). 6. Consider the initial-value problem y'=l +ye, with y =Owhenx =O. (a) Apply the method of successive approximations, starting with the initial guess Y,(x) = 0,and compute Y,(x).(b) Prove that every approximating function Y, is defined on the entire real axis.(c) Use Theorem 7.19 to show that the initial-value problem has at most one solution in anyinterval of the form (-h , h).(d) Solve the differential equation by separation of variables and thereby show that there isexactly one solution Y of the initial-value problem on the interval (n/2 , n/2) and no solutionon any larger interval.

In this example, the successive approximations are defined on the entirereal axis, but they converge to a limit function only on the interval (--n/2 , n/2). 7. We seek two functions y = Y(x) and z = Z(x) that simultaneously satisfy the system ofequations y' =z, z' = xyy + z) 232 Systems of difSerentia1 equations with initial conditions y = 1 and z = l/2 when x = 0. Start with the initial guesses Y,(x) = 1,Z,(x) = l/2, and use the method of successive approximations to obtain the approximatingfunctions Y,(x) = 1 + ; + ; + ; -t g2) Z,(x) = ; + y + ; + 2 + g + g. 8. Consider the system of equations y'=2x fz, z' = 3xy + xzz, with initial conditions y = 2 and z = 0 when x = 0. Start with the initial guesses Y,(x) = 2,Z,,(x) = 0, use the method of successive approximations, and determine Y,(x) and Z,(x). 9. Consider the initial-value problem y' = x2y' + x4y, with y = 5 and y' = 1 whenx = 0. Change this problem to an equivalent problem involving a system of two equations for twounknown functions y = Y(x) and z = Z(x), where z = y'.

Then use the method of successiveapproximations, starting with initial guesses Y,(x) = 5 and Z&j = 1, and determine Y,(x)and Z,(x). 10. Letfbe defined on the rectangle R = [-1 , l] x [-1 , l] as follows: 0 i f x = 0 , fc?

y) =2YlX i f x+0 a n d IyI x2, - 2 x i f x#O a n d y < x2.

(a) Prove that If(x, y)I < 2 for all (x,y) in R.(b) Show thatfdoes not satisfy a Lipschitz condition on R.(c) For each constant C satisfying ICI 5 1 , show that y = Cx2 is a solution of the initial-valueproblem y' =,f(x, y), with y = 0 when x = 0. Show also that the graph of each of thesesolutions over (- 1 , 1) lies in R.(d) Apply the method of successive approximations to &his initial-value problem, startingwith initial guess Y,(x) = 0. Determine Y,(x) and show that the approximations converge toa solution of the problem on the interval (- 1 , 1).(e) Repeat part (d), starting with initial guess Y,,(x) = x. Determine Y,(x) and show thatthe approximating functions converge to a solution different from any of those in part (c).(f) Repeat part (d), starting with the initial guess Y,,(x) = x3.(g) Repeat part (d), starting with the initial guess Y,,(x) = x1j3. * 7.25 Successive approximations and fixed points of operators The basic idea underlying the method of successive approximations can be used not onlyto establish existence theorems for differential equations but also for many other importantproblems in analysis. The rest of this chapter reformulates the method of successiveapproximations in a setting that greatly increases the scope of its applications. Normed linear spaces 233 In the proof of Theorem 7.18 we constructed a sequence of functions {Y,} according tothe recursion formula Y,+,(x) = B + 1: AY,(t) dt. The right-hand member of this formula can be regarded as an operator T which convertscertain functions Y into new functions T(Y) according to the equation T(Y) = B + j; AY(t) dt. In the proof of Theorem 7.18 we found that the solution Y of the initial-value problemY'(t) = A Y(r), Y(a) = B, satisfies the integral equation Y = B + j;AY(t)dt. In operator notation this states that Y = T(Y). In other words, the solution Y remainsunaltered by the operator T. Such a function Y is called afixedpoint of the operator T. Many important problems in analysis can be formulated so their solution depends on theexistence of a fixed point for some operator. Therefore it is worthwhile to try to discoverproperties of operators that guarantee the existence of a fixed point. We turn now to asystematic treatment of this problem. * 7.26 Normed linear spaces To formulate the method of successive approximations in a general form it is convenientto work within the framework of linear spaces. Let S be an arbitrary linear space. Whenwe speak of approximating one element x in S by another element y in S, we consider thedifference x - y , which we call the error of the approximation. To measure the size ofthis error we introduce a norm in the space. DEFINITION OF A NORM. Let S be any linear space. A real-valuedfunction N dejned on Sis called a norm if it has the following properties: (a) N(x) 2 0 for each x in S.(b) N(cx) = ICI N(x) for each x in S and each scalar c.(c) N(x + y) < N(x) + N(y) for all x and y in S.(d) N(x) = 0 implies x = 0. A linear space with a norm assigned to it is called a normed linear space. The norm of x is sometimes written llxll instead of N(x). In this notation the fundamentalproperties become : (a) llxll 2 0 for all x in S.(b) llcxll = ICI llxll for all x in S and all scalars c.Cc) llx + yll I llxll + llyll for all x and y in S.(d) llxll = 0 implies x = 0.If x and y are in S, we refer to IIx - y 11 as the distance from x to y, 234 Systems of equations If the space S is Euclidean, then it always has a norm which it inherits from the innerproduct, namely, llxll = (x, x)x. However, we shall be interested in a particular normwhich does not arise from an inner product. EXAMPLE. The max norm. Let C(J) denote the linear space of real-valued functionscontinuous on a closed and bounded interval J. If q E C(J), define where the symbol on the right stands for the maximum absolute value of v on J. Thereader may verify that this norm has the four fundamental properties. The max norm is not derived from an inner product. To prove this we show that themax norm violates some property possessed by all inner-product norms.

For example, if anorm is derived from an inner product, then the "parallelogram law" lb + yl12 + lb - yl12 = 2 11x112 + 2 lly112 holds for all x and y in S. (See Exercise 16 in Section 1.13.) The parallelogram law is notalways satisfied by the max norm. For example, let x and y be the functions on the interval[0, I] given by x(t) = t, y (t)= 1 -t. Then we have llx\l = lly]l = /Ix + yI[= IIx - yJI = 1, so the parallelogram law is violated. * 7.27 Contraction operators In this section we consider the normed linear space C(J) of all real functions continuouson a closed bounded interval J in which II ~11 is the max norm. Consider an operator T: C(J) --f C(J) whose domain is C(J) and whose range is a subset of C(J). That is, if q is continuous on J,then T(p) is also continuous on J. The following formulas illustrate a few simple examplesof such operators. In each case pl is an arbitrary function in C(J) and T(v)(x) is defined foreach x in J by the formula given: T(v)(x) = wx> > where 2 is a fixed real number, UP)(X) = t ~(0 dt, where c is a given point in J, T(v)(x) = b + j)[t, ~(01 dt, where b is a constant and the compositionf[t, p(t)] is continuous on J.We are interested in those operators T for which the distance 11 T(v) - T(y)\1 is less than a fixed constant multiple tc < 1 of I/p - y/l . These are called contraction operators; theyare defined as follows. Fixed-point theorem for contraction operators 235 DEFINITION OF A CONTRACTION OPERATOR. An operator T: C(J)- C(J) is called acontraction operator if there is a constant 3: satisjying 0 < (y. < 1 such that for every pair offunctions a, and y in C(J) we have (7.70) II T(v) - T(y)11 < tl.

II P - Y II . The constant tc is called a contraction constant for T. Note: Inequality (7.70) holds if and only if we have IT(v)(x) - T(Y)(X)1 I a l/P - w/l for every x in J. EXAMPLE 1. Let T be the operator defined by T(v)(x) = A&x), where 1.

is constant.Since IT(Y)(X) - T(y)(x)l = I4 IPW - YWI 3 we have 1) T(q) - T(yl)I/ = lill II q - y/I . Therefore this operator is a contraction operatorif and only if Jill < 1 , in which case lill may be used as a contraction constant. EXAMPLE 2. Let T(p)(x) = b + jzf[t, p(t)] dt, where f satisfies a Lipschitz conditionof the form If(x,y) -f(x, 41 I Kly - 4 for all x in J and all real y and z; here K is a positive constant. Let L(J) denote the lengthof the interval J. If KL(J) < 1 we can easily show that T is a contraction operator withcontraction constant KL(J). In fact, for every x in J we have IT(v)(x) - T(ly)(x)l = 1 s," {f[t, y(t)1 -f[4 y(f)l]- dt 1 I K 1 j" Ip(t) - y(t)/ dt (I~lIrp-~ll/~~~~~~~I~~~J~lIp, yll. If KL(J) < 1, then Tis a contraction operator with contraction constant r. = KL(J).

* 7.28 Fixed-point theorem for contraction operators The next theorem shows that every contraction operator has a unique fixed point. THEOREM 7.20. Let T: C(J) 4 C(J) be a contraction operator. Then there exists one andonly one function q in C(J) such that (7.71) T(v) = v. Proof. Let p10 be any function in C(J) and define a sequence of functions {pl,} by therecursion formula P)n+l = T(G) for n = 0, 1,2, ... 236 Systems of dlyerential equations Note that P)~+ ~ E C(J) for each n. We shall prove that the sequence {q,} converges to alimit function p in C(J). The method is similar to that used in the proof of Theorem 7.18.We write each qn as a telescoping sum, n-1 (7.72) %(X> = vob) +&bk+lb) - PkW and prove convergence of {p,} by showing that the infinite series (7.73) converges uniformly on J. Then we show that the sum of this series is the required fixedpoint. The uniform convergence of the series will be established by comparing it with theconvergent geometric series M&k=O where M = II P)~II + (IqlII , and c(is a contraction constant for T.

The comparison isprovided by the inequality iP)k+~(~) - pk(X)i < M&k which holds for every x in J and every k 2 1 . To prove (7.74) we note that Therefore the inequality in (7.74) will be proved if we show that (7.75) IIPk - Pk-111 5 Mak-l for every k 2 1. We now prove (7.75) by induction. For k = 1 we have II%_ ~011 I IIPIII + lIploll = M, which is the same as (7.75). To prove that (7.75) holds for k + 1 if it holds for k we notethat Ivk+~(~) - Q)k(X)I = IT(vk)(x) - T(f²'k-&)I 5 a iIP)k - ~k-lii 5 Ma". Since this is valid for each x in J we must also have b~k+l - vkll 5 Ma". This proves (7.75) by induction. Therefore the series in (7.73) converges uniformly on J.If we let v(x) denote its sum we have (7.76) dx) ';irnm%dx) = e)Ocx> +k~obk+l(x) - vk> . Applications of thejxed-point theorem 2 3 7 The function p is continuous on J because it is the sum of a uniformly convergent seriesof continuous functions. To prove that 9 is a fixed point of T we compare T(v) with a) n+l = T(~I~). Using the contraction property of T we have (7.77) IT(gl)(x) - ~n+lCdI = IT(v)(x) - Al I a Id4 - R&~I .

But from (7.72) and (7.76) we find where in the last step we used (7.74).

Therefore (7.77) implies IT(P)(x) - f²L+dx)I I M 2 d+l*k=n When n -+ co the series on the right tends to 0, so ~~+r(x) - T(p)(x).

But since pn+r(x) -p(x) as n + co, this proves that q(x) = T(v)(x) for each x in J. Therefore p = T(v),so v is a fixed point. Finally we prove that the fixed point q~ is unique. Let y be another function in C(J) suchthat T(y) = y. Then we have IIF - YII = IIT - T(y)II I a IIP - YII. This gives us (1 - a) 11 CJI - yI[5 0. Since a < 1 we may divide by 1 - a to obtain theinequality 11~~ - yll < 0.

But since we also have II v - yII 2 0 this means that 11~~ - yII =0, and hence pl - y = 0. The proof of the fixed-point theorem is now complete. r 7.29 Applications of the fixed-point theorem To indicate the broad scope of applications of the fixed point theorem we use it to provetwo important theorems. The first gives a sufficient condition for an equation of the formf(x, r) = 0 to define y as a function of x. THEOREM 7.21. AN IMPLICIT-FUNCTION THEOREM. Let f be defined on a rectangular stripR of the form R={(x,y)Ia~xIb,-co<<++]. Assume that the partial derivative D,f (x, y) exists7 and satisfies an inequality of the form (7.78) OFor this function Y we have Y = T(Y)which means Y(x) = Y(x) - -$, Y(x)] for each x in [a, b]. This gives us (7.79), as required.To show that T is a contraction operator we consider the difference T(F)(X) - T(y)(x) = P(X) - Y(X) - fbk dx>l - f[% v(x)1M By the mean-value theorem for derivatives we have fk p?c41 -fh YWI = &fb? 44l[dx) - Y(X)1 3 where z(x) lies between q(x) and y(x). Therefore (7.80) gives us T(q)(x) - T(y)(x) = [y(x) - y(x)] (1 - "j[; '(x") . The hypothesis (7.78) implies that 0 < 1 _ wk z(x)1- MApplications of the jxed-point theorem 239 The next application of the fixed-point theorem establishes an existence theorem for theintegral equation (7.83) yW = y(x) + 1 Jab W, Oy(t) dt . Here y is a given function, continuous on [a, b], L is a given constant, and K is a givenfunction defined and bounded on the square The function K is called the kernel of the integral equation. Let C be the linear space ofcontinuous functions on [a, b]. We assume that the kernel K is such that the operator Tgiven by T(F)(X) = Y(X) + 2 s," K(x> t>dt> dt maps C into C. In other words, we assume that T(p) E C whenever 9 E C.

A solution ofthe integral equation is any function ~1 in C that satisfies (7.83). THEOREM 7.22. AN EXISTENCETHEOREMFOR INTEGRAL EQUATIONS. If,undertheforegoingconditions. we have (7.84) IW,YI I M for all (x, y) in S, where A4 > 0, then for each i such that (7.85) 111 < lM(b - a) there is one and only one function 91 in C that satisjes the integral equation (7.83). Proof. We shall prove that T is a contraction operator. Take any two functions v1 andqz in C and consider the difference T(eld(x> - T(P&) = A J^," K(x, t>[dt) - dt)l dt. Using the inequality (7.84) we may write IT(R)(X) - Al I I4 Wb - a) 11~ - PAI = CC Iv1 - qd , where CI = Iill M(b - a).

Because of (7.85) we have 0 < tc < 1, so T is a contractionoperator with contraction constant u. Therefore T has a unique fixed point q~ in C. Thisfunction v satisfies (7.83). PART 2NONLINEAR ANALYSIS 8 DIFFERENTIAL CALCULUS OF SCALAR AND VECTOR FIELDS 8.1 Functions from R" to R".

Scalar and vector fields Part 1 of this volume dealt primarily with linear transformations T:V+W from one linear space V into another linear space W. In Part 2 we drop the requirementthat T be linear but restrict the spaces V and W to be finite-dimensional. Specifically, weshall consider functions with domain in n-space R" and with range in m-space R". When both n and m are equal to 1, such a function is called a real-valued function of areal variable. When it = 1 and m > 1 it is called a vector-valued function of a real variable.Examples of such functions were studied extensively in Volume I. In this chapter we assume that n > 1 and m 2 1 . When m = 1 the function is called areal-valued function of a vector variable or, more briefly, a scalar$eld.

When m > 1 it iscalled a vector-valued function of a vector variable, or simply a vectorJeld. This chapter extends the concepts of limit, continuity, and derivative to scalar andvector fields. Chapters 10 and 11 extend the concept of the integral.

Notation: Scalars will be denoted by light-faced type, and vectors by bold-faced type.If f is a scalar field defined at a point x = (x1, . . . , XJ in R", the notations f(x) and"f(x1,. . * 3 x,) are both used to denote the value offat that particular point. Iffis a vectorfield we write f(x) or f(xl, . . . , x,) for the function value at x. We shall use the innerproduct nx-y =xxry, k=l and the corresponding norm \lxll = (x * x)'h, where x = (x1, . . .

) XJ and y = e;",). , y,). Points in R2 are usually denoted by (x, JJ) instead of (x1, x,); points in R3 X, , z ins t e a d o f (x x x)Scalar and vector field!s'd&~d on subsets of R2 and R3 occur frequently in the appli- cations of mathematics to science and engineering. For example, if at each point x of theatmosphere we assign a real numberf(x) which represents the temperature at x, the function 243 244 DifSerential calculus of scalar and vectorjelds f so defined is a scalar field.

If we assign a vector which represents the wind velocity atthat point, we obtain an example of a vector field. In physical problems dealing with either scalar or vector fields it is important to knowhow the field changes as we move from one point to another. In the one-dimensional casethe derivative is the mathematical tool used to study such changes. Derivative theory inthe one-dimensional case deals with functions defined on open intervals. To extend thetheory to R" we consider generalizations of open intervals called open sets. 8.2 Open balls and open sets Let a be a given point in R" and let r be a given positive number. The set of all points xin R" such that lb - all < r is called an open n-ball of radius r and center a. We denote this set by B(a) or by B(a; r).The ball B(a; r) consists of all points whose distance from a is less than r.

In R1 this is simply an open interval with center at a. In R2 it is a circular disk, and in R3 it is a sphericalsolid with center at II and radius r. DEFINITION OF AN INTERIOR POINT. Let S be a subset of R", and assume that a E S . Thena is called an interior point of S if there is an open n-ball with center at a, all of whose pointsbelong to S. In other words, every interior point a of S can be surrounded by an n-ball B(a) such thatB(a) G S. The set of all interior points of S is called the interior of S and is denoted byint S. An open set containing a point a is sometimes called a neighborhood of a. DEFINITION OF AN OPEN SET.

A set S in R" is called open ifall its points are interior points.In other words, S is open if and only if S = int S.

EXAMPLES. In R1 the simplest type of open set is an open interval. The union of two ormore open intervals is also open. A closed interval [a, b] is not an open set because neitherendpoint of the interval can be enclosed in a l-ball lying within the given interval. The 2-ball S = B(O; 1) shown in Figure 8.1 is an example of an open set in R2. Everypoint a of S is the center of a disk lying entirely in S. For some points the radius of thisdisk is very small.

Some open sets in R2 can be constructed by taking the Cartesian product of open sets inR1. If Al and A, are subsets of R1, their Cartesian product A, x A, is the set in R2 defined bYA , x A2= {(al,a2)ja,EAl and a,EA,}. An example is shown in Figure 8.2. The sets A, and A2 are intervals, and A, x A2 is arectangle. 1 f A, and A, are open subsets of R', then AI x A, will be an open subset of R2. To provethis, choose any point a = (a,, a2) in A, x A,. We must show that a is an interior point Exercises 245 Circular disk FIGURE 8.1 The disk B(0; 1) is an openset in R2. FIGURE 8.2 The Cartesian product of twoopen intervals is an open rectangle. of A, x A,. Since Ai and A2 are open in R1 there is a l-ball B(a,; rl) in A1 and a l-ballB(a, ; rJ in A,. Let r = min {rl, r2}. We can easily show that the 2-ball B(a; r) cA1 x A,. In fact, if x = (x1, x2) is any point of B(a; r) then 1(x - u/J < r, so Ix1 - a,! < rland lx, - a,] < r2. Hence x1 E B(a, ; rl) and x2 E B(a, ; r2). Therefore x1 E A, andx2 ~A29 so (x,,x,)~A, x A,. This proves that every point of B(a; r) is in A, x A*.Therefore every point of A, x A? is an interior point, so A, x A2 is open. The reader should realize that an open subset of R1 is no longer an open set when it isconsidered as a subset of R2, because a subset of R' cannot contain a 2-ball. DEFINITIONS OF EXTERIOR AND BOUNDARY. A point x is said to be exterior to a set S inR" ifthere is an n-b&l B(x) containing no points of S. The set of allpoints in R" exterior to Sis culled the exterior of 5' and is denoted by ext S. A point which is neither exterior to S nor aninterior point of S is called a boundary point of S. The set qf ail boundary points of S is calledthe boundary of S and is denoted by 8s. These concepts are illustrated in Figure 8.1. The exterior of S is the set of all x withllxjl > 1. The boundary of S consists of all x with l/xl/ = 1.

8.3 Exercises1. Letf be a scalar field defined on a set S and let c be a given real number. The set of all points x in S such thatf(x) = c is called a level set off. (Geometric and physical problems dealingwith level sets will be discussed later in this chapter.) For each of the following scalar fields.S is the whole space Rfl. Make a sketch to describe the level sets corresponding to the givenvalues of c.G-4 .f(x, y) = x2 + y2, c =o, 1,4,9.(b) .f(x, y) = eZy, c = ee2 , e-l, 1, c, e2, e3. (4 f(x, y) = cm (x I- y) 9 c = -1,0,&g& 1.(d),f(x,y,z) =x +y +z, c = -1,O, 1.(e) ,f(x, y, z) = x2 + 2y2 + 3z2, c =0,6,12.

(f) ,f(x, y, z) = sin (x" + -y2 + z2), c = -1, -3, o, +,b, 1. 246 DifSerential calculus of scalar and vectorjelds 2. In each of the following cases, let S be the set of all points (x,y) in the plane satisfying thegiven inequalities. Make a sketch showing the set S and explain, by a geometric argument,whether or not S is open. Indicate the boundary of S on your sketch.(a) x2 + y2 < 1 . (h) 1 Ix 12and3 (i) 1 O.(c) 1x1 < 1 and lyl < 1. Cj) x2y.(d) x 2 0 and y > 0 . 04 x >y.(e) 1x1 5 1 and lyl I 1.

(1) y > x2 and [xl < 2.(f) x >Oandy 0.(g) xy < 1 . (n) (2x -x2 -y2)(x2 +y2 -x) >O. 3. In each of the following, let S be the set of all points (x, y, z) in 3-space satisfying the giveninequalities and determine whether or not S is open.(a) z* -x2 -y* - 1 >O.(b) 1x1 < 1, lyl < 1, and Iz/ < 1.(c) x + y + z < 1.(d) 1x1 I 1, lyl < 1, and Izl < 1.(e) x+y+zO,y>O,z>O.(f) x2 + 4y2 + 4z2 - 2x + 16y +4Oz + 113 5. Prove the following properties of open sets in Rn:(a) The empty set 0 is open.(b) R" is open.(c) The union of any collection of open sets is open.(d) The intersection of a finite collection of open sets is open.(e) Give an example to show that the intersection of an infinite collection of open sets is notnecessarily open.

Closed sets. A set Sin Rn is called closed if its complement Rn - S is open. The next threeexercises discuss properties of closed sets. 6. In each of the following cases, let S be the set of all points (x, y) in R2 satisfying the givenconditions. Make a sketch showing the set Sand give a geometric argument to explain whetherS is open, closed, both open and closed, or neither open nor closed.(a) x2 +y2 2 0. (g) 1 Continuity of polynomials in n variables. A scalar field P defined on R" by aformula of the form P(x) =jgO.

. .,r, Ck,. . +,x: * * * xk,,1 " is called a polynomial in n variables x1, . . . , x, . A polynomial is continuous everywherein R" because it is a finite sum of products of scalar fields continuous everywhere in R".For example, a polynomial in two variables x and y, given by is continuous at every point (x, y) in R2. ExAMeLE 5.

Continuity of rational functions. A scalar field f given by f (x) = P(x)/Q(x) ,where P and Q are polynomials in the components of x, is called a rational function. Arational function is continuous at each point where Q(x) # 0. Further examples of continuous function can be constructed with the help of the nexttheorem, which is concerned with continuity of composite functions. 2 5 0 calculus of scalar and vectorjelds THEOREM 8.2. Let f and g be functions such that the composite function f 0 g is dejned ata, where (fo d(x) = fW>l * If g is continuous at a and if f is continuous at g(a), then the composition f 0 g is continuousat a. Proof. Let y = f(x) and 6 = g(a). Then we have fM41 - fM41 =fW -f(b). By hypothesis, y +basx-ta,sowehave ,,x!-~;_ollfk(~)]l -fk(a>lII ,liy ollf(y) -f(b)ll = 0. Therefore tin f[g(x)] = f[g(a)] , so f 0 g is continuous at a.

EXAMPLE 6. The foregoing theorem implies continuity of scalar fields h, where h(x, y) isgiven by formulas such as X+2/sin (x"y), log (2 + y">, x - - X+Y'log [cos (x" + y2)]. These examples are continuous at all points at which the functions are defined. The first iscontinuous at all points in the plane, and the second at all points except the origin.

Thethird is continuous at all points (x, y) at which x + y # 0, and the fourth at all points atwhich x2 + y2 is not an odd multiple of n/2. [The set of (x, y) such that x2 + y2 = m-/2,n = 1) 3, 5, . . . , is a family of circles centered at the origin.] These examples show thatthe discontinuities of a function of two variables may consist of isolated points, entirecurves, or families of curves. EXAMPLE 7. A function of two variables may be continuous in each variable separatelyand yet be discontinuous as a function of the two variables together.the following example: This is illustrated by fb, Y>= e2 if (x, Y) Z (0, 01, f(0, 0) = 0 . For points (x, y) on the x-axis we have y = 0 andf(x, y) = f(i, 0) = 0, so the functionhas the constant value 0 everywhere on the x-axis. Therefore, if we put y = 0 and think off as a function of x alone,fis continuous at x = 0. Similarly,fhas the constant value 0at all points on the y-axis, so if we put x = 0 and think off as a function of y alone, f iscontinuous at JJ = 0. However, as a function of two variables,fis not continuous at theorigin.

In fact, at each point of the line y = x (except at the origin) the function has theconstant value 4 because f(x, x) = x*/(2x2) = 4 ; since there are points on this linearbitrarily close to the origin and since f (0,O) # 4, the function is not continuous at (0, 0). 8.5 Exercises Exercises 251 The exercises in this section are concerned with limits and continuity of scalar fields defined onsubsets of the plane. 1. In each of the following examples a scalar fieldfis defined by the given equation for all points(x, y) in the plane for which the expression on the right is defined. In each example determinethe set of points (x, y) at whichfis continuous. (a) f(x, y) = x4 + y4 - 4x2y2. (f) f(x, y) = arcsin X .Jx" + y2 (b) f(x, y) = log (x2 + y2).x +Y(g) f(x, y) = arctan - 1 -xy' (h) f(x, y) = -A.Jx" + y2 (i) f(x, y) = x("'). (e) f(x, y) = arctan :. (j) f(x, y) = arccos Jx/y . 2.

If lim f(x, y) = L, and if the one-dimensional limits(r,yl+h,bJ lim fz+a v-b both exist, prove that lim [limf(x, y)] = lim [limf(x, y)] = L.z+a ll+b 2/+b ~+a The two limits in this equation are called iterated limits; the exercise shows that the existence ofthe two-dimensional limit and of the two one-dimensional limits implies the existence and equalityof the two iterated limits. (Th e converse is not always true. A counter example is given inExercise 4.) 3. Letf(x, y) = (x - y)/(x + y) if x + y Z 0. Show that lim [limf(x, y)] = 1e - o u-0 but that lim [limJ'(x, y)] = -1.u - o x-.0 Use this result with Exercise 2 to deduce thatf(x, y) does not tend to a limit as (x, y) -+ (0, 0).4. Let Show that j-(X, y) = xzy"x"y" + (x -y)2 whenever X'J? + (x - Y)~ # 0. lim [limf(x, y)] = lim [limf(x, y)] = 0x-0 v-0 1/r+o s+o but thatf(x, y) does not tend to a limit as (x, y) + (0,O). [Hint: Examinefon the line y = x .]This example shows that the converse of Exercise 2 is not always true. 252 Dlrerential calculus of scalar and vector$elds 5. 6. Let l 1x sin - f = yi f y#O, 0 i f y=O. Show thatf(x, y) + 0 as (x, y) + (0,O) but that lim [limf(x, u)l # lim [limf(x, u)] .u-0 x40 x7+0 w-+0 Explain why this does not contradict Exercise 2. If (x, y) + @,O), letfk y) = (x2 -v2)/(x2 + v2). Find the limit off(x, v) as (x, y) + (0,O)along the line y = mx. Is it possible to definef(0, 0) so as to makefcontinuous at (0, O)?Letf(x, y) = 0 ify < 0 or ify > x2 and letf(x, y) = 1 if 0 < y < x2. Show thatf(x, y) - 0as (X,JJ) + (0,O) along any straight line through the origin. Find a curve through the originalong which (except at the origin)f(x, JJ) has the constant-value 1. Isfcontinuous at the origin?If f(x, y) = [sin (x2 + v2)]/(x2 + u2) when (x, u) # (0,O) how must f(0, 0) be defined so asto makefcontinuous at the origin?Letfbe a scalar field continuous at an interior point a of a set S in R". Iff(a) # 0, prove thatthere is an n-ball B(a) in whichfhas the same sign asf(a). 8.6 The derivative of a scalar field with respect to a vector This section introduces derivatives of scalar fields. Derivatives of vector fields are dis-cussed in Section 8.18. Let f be a scalar field defined on a set S in R", and let a be an interior point of S. Wewish to study how the field changes as we move from a to a nearby point. For example,supposef(a) is the temperature at a point a in a heated room with an open window. If wemove toward the window the temperature will decrease, but if we move toward the heaterit will increase. In general, the manner in which a field changes will depend on the directionin which we move away from I(. Suppose we specify this direction by another vector y. That is, suppose we move from atoward another point a + y along the line segment joining a and a + y . Each point on thissegment has the form a + hy, where h is a real number. An example is shown in Figure8.3. The distance from a to a + hy is llhyll = lhl llyll . Since a is an interior point of S, there is an n-ball B(a; r) lying entirely in S. If h is chosenso that Ihl llyll < r , the segment from a to a + hy will lie in S. (See Figure 8.4.) We keep ' B(a;r) FIGURE 8.3 The point a + hv lies on the line FIGURE 8.4 The point a + hy lies in thethrough a parallel to JJ. n-ball B(a; r) if jlhyjl < r. The derivative of scalarjeld with respect to a vector 253 h # 0 but small enough to guarantee that a + hy E S and we form the difference quotient (8.3)f@ + b) -f(a) h ' The numerator of this quotient tells us how much the function changes when we movefrom a to a + hy . The quotient itself is called the average rate of change off over the linesegment joining a to a + hy .

We are interested in the behavior of this quotient as h + 0.This leads us to the following definition. DEFINITION OF THE DERIVATIVE OF A SCALAR FIELD WITH RESPECT TO A VECTOR. Given ascaIar$eldf: S + R , where S c R" . Let a be an interior point of S and let y be an arbitrarypoint in R". The derivative off at a with respect toy is denoted by the symbolf '(a; y) and isdejned by the equation (8.4)fr(a; y) = limf(a + W -.I (4 h-t0 h when the limit on the right exists. EXAMPLE 1. If y = 0, the difference quotient (8.3) is 0 for every h # 0, so f'(u; 0)always exists and equals 0. EXAMPLE 2. Derivative of a linear transformation. Iff: S + R is linear, then f (a + hy) =f(a) + hf(y) and the difference quotient (8.3) is equal tof(y) for every h # 0. In this case,f'(a; y) always exists and is given by f'@; Y> =f(y) for every a in S and every y in R". In other words, the derivative of a linear transformationwith respect to y is equal to the value of the function at y. TO study how f behaves on the line passing through a and a + y for y # 0 we introducethe function g(t) = f(a + 09. The next theorem relates the derivatives g'(t) and f '(a + ty; y). T H E O R E M 8.3. Let g(t) =f(a + ty). If one of the derivatives g'(t) or f '(a + ty; y)exists then the other also exists and the two are equal, (8.5) g'(t) =f'(a + ty; y>. In particular, when t = 0 we have g'(O) = f '(a; y) . 2 5 4 DifSerential calculus of scalar and vector$elds Proof. Forming the difference quotient for g, we have At + h) - g(t) = f(a + tY + hY> - f@ + ty>h h Letting h + 0 we obtain (8.5). EXAMPLE 3. Compute f '(a; y) if f (x) = 11.x /I 2 for all x in R". Solution. We let g(t)=f(a+tY)=(a+tY).(a+tY)=a.a+2taeY+t2YaY.Therefore g'(t) = 2u . y + 2ty . y , so g'(0) = f '(u; Y) = 2~ . y . A simple corollary of Theorem 8.3 is the mean-value theorem for scalar fields. THEOREM 8.4. MEAN-VALUE THEOREM FOR DERIVATIVES OF SCALAR FIELDS. Assume thederivativef'(u + ty; y) exists for each t in the interval 0 5 t < 1. Then for some real 19 inthe open interval 0 < 0 < 1 we have flu + Y) -f(a) =f'(z;y), where z=u+By. Proof. Let g(t) = f (a + ty) . Applying the one-dimensional mean-value theorem to gon the interval [0, l] we have g(l) - g(O) = g'(W 3 where 0 < 6' < 1. Since g(l) - g(0) = f(u + y) -f(u) and g'(0) = f '(a + 19y; y), this completes the proof. 8.7 Directional derivatives and partial derivatives In the special case when y is a unit vector, that is, when 11 yll = 1 , the distance between aand a + hy is Ihl. In this case the difference quotient (8.3) represents the average rate ofchange offper unit distance along the segment joining u to a + hy ; the derivative f '(a; y)is called a directional derivative. DEFINITION OF DIRECTIONAL AND PARTIAL DERIVATIVES. If y is a unit vector, the derivativef '(a, y) is called the directional derivative off at a in the direction of y.

In particular, ify = ek (the kth unit coordinate vector) the directional derivative f '(a; e,J is called the partialderivative with respect to ek and is also denoted by the symbol Dkf (a). Thus, &f (4 = f '(a ; ek) . The following notations are also used for the partial derivative Dkf (a): Dkf(al, . . . , a,), and f&(a,, . . . , a,). Exercises 255 Sometimes the derivative fLk is written without the prime as f,, or even more simply ash.In R2 the unit coordinate vectors are denoted by i and j. If a = (a, b) the partial derivatives f '(a; i) and f '(a; j) are also written as E (a, b) and ; (0, b), respectively. In R3, if a = (u, b, c) the partial derivatives D&u), D&(u), and Df(u) arealso denoted by E (a, b, c> > ; (a, b, 4, and z (a, b, 4.

8.8 Partial, derivatives of higher order Partial differentiation produces new scalar fields DJ, . . . , DJfrom a given scalar fieldJ The partial derivatives of DJ, .

. . , Dnf are called second-order partial derivatives off.For functions of two variables there are four second-order partial derivatives, which arewritten as follows: a"fDd4f) = -axay'atf D,(D,f) = ax 3atf YD,(W) = j -y. Note that D1(D2f) means the partial derivative of D,f with respect to the first variable.We sometimes use the notation Di,jf for the second-order partial derivative Di(Djf). Forexample, D,,,f = D,(D,f). In the ¬ation we indicate the order of derivatives bywriting a2f a afzFy=&iy(1 This may or may not be equal to the other mixed partial derivative, a"f a afayax = G Z *(1 In Section 8.23 we shall prove that the two mixed partials D,(D2f) and D,(D,f) are equalat a point if one of them is continuous in a neighborhood of the point. Section 8.23 alsocontains an example in which D1(D2f) # D2(DIf) at a point. 8.9 Exercises 1. A scalar field f is defined on Rn by the equationf(x) = a . x, where a is a constant vector.Computef'(x; y) for arbitrary x and y. 256 Dzrerential calculus of scalar and vectorjelds 2. (a) Solve Exercise 1 whenf(x) = ll~11~.(b) Take n = 2 in (a) and find all points (x, y) for whichf'(2i + 3j; xi + yj) = 6.(c) Taken = 3 in (a) and find all points (x, y, z) for whichf'(i + 2j + 3k; xi + yj + zk) = 0. 3. Let T: Rn + Rn be a given linear transformation. Compute the derivativef'(x; y) for thescalar field defined on Rn by the equationf(x) = x . T(x). In each of Exercises 4 through 9, compute all first-order partial derivatives of the given scalar field.The fields in Exercises 8 and 9 are defined on Rn. 4. f(x, y) = x2 + y2 sin (xy) . 7.&y) =z, x #y. 5. f(X, y) = Jx" + y2. 8. f(x) = a . x , II fixed.

6.f(x,y) = 5,Jx" + y2 (x, y) # (0,O).

9. f(x) = 2 2 UijXiXi, adj = aji.i-1 j=l In each of Exercises 10 through 17, compute all first-order partial derivatives. In each of Exercises10, 11, and 12 verify that the mixed partials D,(D,f) and D,(D,f) are equal. 10. f(X, y) = x4 + y4 - 4xzy2. 14. f(x, y) = arctan (y/x), x #O. 11. f(X, y) = log (x2 + y2),x +Y(x, y) # (0,O). 15. f(x, y) = arctan -1 -xy' xy # 1. 12. J-(x, y) = $ cos X2) y #O. 16. f(x, y) = , x >o . 13. fUse themean-value theorem to prove thatfis constant on B(a).(b) Suppose that f'(x; JJ) = 0 for a $xed vector y and for every x in B(a). What can youconclude about f in this case? 21. A set S in Rn is called convex if for every pair of points a and b in S the line segment from ato b is also in S; in other words, ta + (1 - t)b E S for each t in the interval 0 < t < 1 .(a) Prove that every n-ball is convex.(b) Iff'(x; JJ) = 0 for every x in an open convex set S and for every y in Rn, prove thatfisconstant on S. 22. (a) Prove that there is no scalar fieldfsuch thatf'(a; y) > 0 for a fixed vector a and everynonzero vector y.(b) Give an example of a scalar field f such thatf'(x; y) > 0 for a fixed vector y and everyvector x.

Directional derivatives and continuity 257 8.10 Directional derivatives and continuity In the one-dimensional theory, existence of the derivative of a function f at a pointimplies continuity at that point. This is easily proved by choosing an h # 0 and writing f(a+h)-f(a)=f(a+h)--f(a).h.h As h + 0 the right side tends to the limit f'(a) * 0 = 0 and hence f(a + h) --f(a). Thisshows that the existence off'(a) implies continuity off at a. Suppose we apply the same argument to a general scalar field. Assume the derivativef'(a; y) exists for some y. Then if h # 0 we can write f(a + hy) -f(a) = j(' + '; -f(u). h. As h -+ 0 the right side tends to the limitf'(u; JJ) .O = 0 ; hence the existence off'(u; y)for a given y implies that limf(a + hy) = f(a)h+O for the same y. This means thatf(x) -f(a as x --f a along a straight line through u having)the direction y. Iff'(u; y) exists for every vector y, thenf(x) +f(u) as x + a along everyline through a. This seems to suggest thatfis continuous at u. Surprisingly enough, thisconclusion need not be true.

The next example describes a scalar field which has a direc-tional derivative in every direction at 0 but which is not continuous at 0. EXAMPLE.

Letfbe the scalar field defined on R2 as follows: f(x, Y) = --g-j$ i f x#O, f(O, Y> = 0. Let a = (0,O) and let y = (a, b) be any vector. If a # 0 and h # 0 we have f@ + hy> -f(u) = .fW -f(O) = f(ha, hb) = ab2h h h a2 + h2b4 * Letting h + 0 we find f '(0; y) = b2/a. If y = (0, b) we find, in a similar way, thatf'(0; y) = 0. Therefore f'(0; y) exists for all directions y. Also, f(x) + 0 as x + 0along any straight line through the origin. However, at each point of the parabola x = y2(except at the origin) the functionfhas the value +. Since such points exist arbitrarily closeto the origin and since f (0) = 0, the function f is not continuous at 0. The foregoing example shows that the existence of all directional derivatives at a pointfails to imply continuity at that point. For this reason, directional derivatives are a some-what unsatisfactory extension of the one-dimensional concept of derivative. A more suitablegeneralization exists which implies continuity and, at the same time, permits us to extendthe principal theorems of one-dimensional derivative theory to the higher dimensional case.This is called the total derivative. 258 Deferential calculus of scalar and vectorjells 8.11 The total derivative We recall that in the one-dimensional case a function f with a derivative at a can beapproximated near a by a linear Taylor polynomial. Iff '(a) exists we let E(a, h) denote thedifference (8.6) E(a, h) = f(a + h, - f(a) - f'(a)h i f h#O, and we define E(a, 0) = 0. From (8.6) we obtain the formula f(a + h) =f(a) + f'(a>h + hE(a,h), an equation which holds also for h = 0. This is the first-order Taylor formula for approxi-mating f (a + h) - f(a) by f '(a)h. The error committed is hE(a, h). From (8.6) we seethat E(a, h) --f 0 as h + 0. Therefore the error hE(a, h) is of smaller order than h forsmall h. This property of approximating a differentiable function by a linear function suggests away of extending the concept of differentiability to the higher-dimensional case. Let f: S + R be a scalar field defined on a set S in R". Let a be an interior point of S,and let B(a; r) be an n-ball lying in S. Let u be a vector with llujl < r, so thata + uEB(a;r). DEFINITION OF A DIFFERENTIABLE SCALAR FIELD. We say that f is dtrerentiable at a if thereexists a linear transformation T,:R"+R from R" to R, and a scalar function E(a, u) such that (8.7) f (a + v) = f (a) + T,(v) + II 4 -W, 4, for II VII < r , where E(a, u) + 0 as II ~11 + 0. The linear transformation T, is called the totalderivative off at a.

Note: The total derivative T, is a linear transformation, not a number. The functionvalue T,(v) is a real number; it is defined for every point u in Rn. The total derivative wasintroduced by W. H. Young in 1908 and by M. Frechet in 1911 in a more general context. Equation (8.7), which holds for II uII < r, is called aflrst-order Taylorformula forf (a + u).It gives a linear approximation, T,(u), to the difference f(a + u) -f(u). The error in theapproximation is II uII E(a, u), a term which is of smaller order than II uII as 11 VII + 0 ; that is,W, 4 = o(ll4> as Ilull - 0. The next theorem shows that if the total derivative exists it is unique. It also tells us howto compute T,(y) for every y in R". The gradient of a scalarjield 259 THEOREM 8.5. Assume f is differentiable at a with total derivative T,.

Then the derivativef'(a; y) exists for every y in R" and we have w3) T,(Y) =.f'(a;y). Moreover, f '(a; y) is a linear combination of the components ofy. Infact, $y = (yI, . . .

, y,,) ,we have (8.9) Proof. Equation (8.8) holds trivially if y = 0 since both T,(O) = 0 andf'(a; 0) = 0.Therefore we can assume that y # 0. Since f is differentiable at a we have a Taylor formula, (8.10) f(a + v) =f(a) + T,(u) + II41 E(a, 4 for ((u/l < r for some r > 0, where E(a, u) + 0 as Ij~11 -+ 0. In this formula we take21 = hy, where h # 0 and lhl llyll < r.

Then Ilull . , D&a) exist. We also write grad f for Vf. The symbolV is pronounced "del." 260 Dlrerential calculus of scalar and vector fields The first-order Taylor ormula (8.10) can now be written in the form (8.12) f(+ 4 =m + VfW * u + II 4 J% 4 5 where E(a, u) ---f 0 as II uI[-+ 0. In this form it resembles the one-dimensional Taylorformula, with the gradie t vector V'(u) playing the role of the derivativef'(u). From the Taylor form la we can easily prove that differentiability implies continuity. THEOREM 8.6. iIf a sea rjeldf is difSerentiable at a, then f is continuous at a. Proof. From (8.12) we have If@ + 4 -f@>l = IV@> * u + 1141 WY 41. Applying the triangle inequality and the Cauchy-Schwarz inequality we find 0 I If(a + 4 - .fWl I IlWa)ll Ilull + Ilull LW, 41. This shows thatf(u + u) --f( )u as 11 II2, + 0, so f is continuous at u.

FIGURE 8.5 Geometric relation of the directional derivative to the gradient vector. When y is a unit vector the directional derivativef'(u; y) has a simple geometric relationto the gradient vector. Assume that Of(u) # 0 and let 0 denote the angle between y andVf(u). Then we have ~'(u;Y) = VW 3 = IIVWII lbll ~0s e = IIVWII ~0s 0. This shows that the directional derivative is simply the component of the gradient vectorin the direction of y. Figure 8.5 shows the vectors Vf(u) and y attached to the point u.The derivative is largest when cos 8 = 1 , that is, when y has the same direction as Vf(u).In other words, at a given point a, the scalar field undergoes its maximum rate of change inthe direction of the gradient vector; moreover, this maximum is equal to the length of thegradient vector. When Vf( )u is orthogonal to y, the directional derivative f'(a; y) is 0.

A su$icient condition for diferentiability In 2-space the gradient vector is often written as 261 Vfk Y> = -jyVk Y> i + wxt Y$-aY * In 3-space the corresponding formula is v-(x, Y, 4 =af(x' " ') i + af(x' Y, ')j + af(x, y, ') k .&aY aZ 8.!3 A sufficient condition for differentiability Iffis differentiable at a, then all partial derivatives Dj(a), . . . , D&a) exist. However,the existence of all these partials does not necessarily imply that f is differentiable at (I.A counter example is provided by the function .a Y> = -$$ i f x#O, f@, Y) = 0, discussed in Section 8.10.

For this function, both partial derivatives D&O) and D&O)exist but f is not continuous at 0, hence f cannot be differentiable at 0. The next theorem shows that the existence of continuous partial derivatives at a pointimplies differentiability at that point. THEOREM 8.7. A SUFFICIENT CONDITION FOR DIFFERENTIABILITY. Assume that thepartialderivatives DJ, . . . , D,,f exist in some n-ball B(a) and are continuous at a.

Then f is dif-ferentiable at a. Note: A scalar field satisfying the hypothesis of Theorem 8.7 is said to be continuouslydiflerentiable at a. Proof. The only candidate for T,(V) is Of (a) * v . We will show that f (a + 4 - f (4 = Vf (4 * v + II 41 Jm 4 > where E(a, v) + 0 as 11 VII -+ 0. This will prove the theorem.Let il = I(v((. Then v = iu , where ((u((= 1. We keep iz small enough so that a + v lies in the ball B(a) in which the partial derivatives D& . . .

, DJexist. Expressing u in termsof its components we have where e,, . . . , e, are the unit coordinate vectors. Now we write the difference f (a + v) -f(a) as a telescoping sum, (8.13) f(a + v) - f(a) = f(u + 1~) - f(a) =*&(a + h) - f(a + k,)) 9 262 Differential calculus of scalar and vectorjelds where v,, , ul, . . . , o, are any vectors in R" such that I+, = 0 and u,, = u.

We choose thesevectors so they satisfy the recurrence relation uk = u~...~ + ukek .

That is, we take u()= 0 , u1 = wl, u2 = ulel + u2e2, . . . , u, = ulel + . * * + u,e, . Then the kth term of the sum in (8.13) becomes f(a + Iu, , + h4 - f(a + h J = .f(h + J.w,J - f(h) , where 6, = a + Au, , . The two points b, and b, + lu,:e, differ only in their kth component.Therefore we can apply the mean-value theorem of differential calculus to write (8.14) where ck lies on the line segment joining 6, to 6, + lu,e,. Note that 6, + a and henceck+aasA-+O. Using (8.14) in (8.13) we obtain But Cf(u) * u = 1 V'(u) . u = /I z;==, D,f(u)u,, so f(a + u> -f(a) - V(a) . u = 6 {W-(4 - W-(4}u, = Ilull @a, ~1,h = l where Since clc -+ a as 11 u/I + 0, and since each partial derivative Dkf is continuous at a, we seethat E(u, u) 4 0 as 11 uII + 0. This completes the proof. 8.14 Exercises 1. Find the gradient vector at each point at which it exists for the scalar fields defined by thefollowing equations : (a) f(x, y) = x2 + y2 sin (xy) . (d) f(x, y, z) = x2 - y2 + 2z2.(b) J-(x, y) = e5 cos ,v . (e) f(x, y, z) = log (x2 + 2y2 - 3~~).(c) f(X, y, z) = x2y3z4. (0 f(X,.Y, z> = xy"* 2. Evaluate the directional derivatives of the following scalar fields for the points and directionsgiven :(a) f(x, y, z) = x2 + 2y2 + 3z2 at (1, 1,O) in the direction of i -j + 2k.(b) f(x, y, z) = (x/y)" at (1, 1, 1) in the direction of 2i + j - k. 3. Find the points (x, y) and the directions for which the directional derivative of f(x, y) =3x2 + y2 has its largest value, if (x, y) is restricted to be on the circle x2 + y2 = 1 . 4. A differentiable scalar fieldfhas, at the point (1, 2), directional derivatives +2 in the directiontoward (2,2) and -2 in the direction toward (1, 1). Determine the gradient vector at (1, 2)and compute the directional derivative in the direction toward (4, 6). 5. Find values of the constants a, b, and c such that the directional derivative off(x, y, z) =axy2 + byz + cz2.? at the point (1,2, -1) has a maximum value of 64 in a direction parallelto the z-axis. A chain rule for derivatives of scalar$elds 263 6.

Given a scalar field differentiable at a point a in R2. Su ppose thatf'(a; JJ) = 1 andf'(a; z) = 2,where y = 2i + 3j and z = i + j. Make a sketch showing the set of all points (x, v) for whichf'(a; xi + vj) = 6. Also, calculate the gradient V'(a).

7. Let f andg denote scalar fields that are differentiable on an open set S. Derive the followingproperties of the gradient:(a) grad f = 0 iff is constant on S.(b) grad (f + g) = gradf + gradg .(c) grad (cf) = c gradfif c is a constant.(4 grad (fg) = fgradg + g gradf. (e) grad f &wav-ftPd~08 g2 at points at which g # 0.

8. In R3 let r(x, y, z) = xi + rj + zk, and let r(x, y, z) = Ilr(x,Y, z)il .(a) Show that Vr(x, y, z) is a unit vector in the direction of r(x, y, z).(b) Show that V(P) = nrnv2r if n is a positive integer. [Hint: Use Exercise 7(d).](c) Is the formula of part (b) valid when n is a negative integer or zero?(d) Find a scalar field f such that Vf = r . 9. Assume f is differentiable at each point of an n-ball B(a). If f '(x; y) = 0 for n independentvectors y, , . . . , y, and for every x in B(a), prove that f is constant on B(a). 10. Assume f is differentiable at each point of an n-ball B(u).(a) If Vf (x) = 0 for every x in B(u), prove that f is constant on B(a).(b) Iff(x) I f(a) for all x in B(u), prove that of(a) = 0. 11. Consider the following six statements about a scalar field f: S - R, where S c Rn and u isan interior point of S.(a) f is continuous at a.(b) f is differentiable at a.(c) f '(a; y) exists for every y in Rn.(d) All the first-order partial derivatives off exist in a neighborhood of a and are continuousat a.(e) Of (a) = 0.(f) f(x) = I/x - alI for all x in R". a b c d e f----__- a TIn a table like the one shown here, mark T - - - - - - - in the appropriate square if the statement in b T row (x) always implies the statement in - - - - - - - column (y). For example, if (a) always implies C T (b), mark T in the second square of the first----____- row.

The main diagonal has already been d T filled in for you.---____- e T-----__- f T 8.15 A chain rule for derivatives of scalar fields In one-dimensional derivative theory, the chain rule enables us to compute the derivativeof a composite function g(t) =f[r(t)] by the formula g'(t) = f '[r(t)] * r'(t). 264 DifSerential calculus of scalar and vectorjelds This section provides an extension of the formula when f is replaced by a scalar fielddefined on a set in n-space and r is replaced by a vector-valued function of a real variablewith values in the domain off. In a later section we further extend the formula to cover the case in which both f and rare vector fields. It is easy to conceive of examples in which the composition of a scalar field and a vectorfield might arise. For instance, supposef(x) measures the temperature at a point x of asolid in 3-space, and suppose we wish to know how the temperature changes as the point xvaries along a curve C lying in the solid.

If the curve is described by a vector-valuedfunction r defined on an interval [a, b], we can introduce a new function g by means of theformula g(t) = fb+(t)l i f aSince S is open there isan n-ball B(a) lying in S. We take h # 0 but small enough so that r(t + h) lies in B(a),and we let y = r(t + h) - r(t). Note that y + 0 as h --f 0. Now we have g(t + h) -g(t) =fMt + h)l -f P(t)1 = fh + Y> -f(a) Applying the first-order Taylor formula for f we have S(a + Y> -f(a) = Vf(4.Y + llYll GPY)9 where E(u, y) + 0 as I/y/(---f 0. Since y = r(t + h) - r(t) this gives us dt i- h; - g(t) = vff(u) . r(t + hj - 40 + IIr(t + hl - r(t)I(ECU, Y> . Letting h + 0 we obtain (8.15). A chain rule for derivatives of scaIar$elds 265 EXAMPLE 1. Directional derivative along a curve. When the function r describes a curveC, the derivative r' is the velocity vector (tangent to the curve) and the derivative g' inEquation (8.15) is the derivative off with respect to the velocity vector, assuming thatr' # 0 . If T(t) is a unit vector in the direction of r'(t) (Tis the unit tangent vector), the dotproduct Vf [r(t)] * T(t) is called the directional derivative off along the curve C or in thedirection of C. For a plane curve we can write T(t) = cos cx(t) i + cos /3(t)j, where cl(t) and ,B(t) are the angles made by the vector T(t) and the positive x- and y-axes;the directional derivative off along C becomes Of [r(t)] * T(t) = D&(t)] cos a(t) + D&r(t)] cm P(t). This formula is often written more briefly as Vf.T=$cos~+afcos~.ay Some authors write df/ds for the directional derivative Vf * T. Since the directionalderivative along C is defined in terms of T, its value depends on the parametric representa-tion chosen for C. A change of the representation could reverse the direction of T; this,in turn, would reverse the sign of the directional derivative.

EXAMPLE 2. Find the directional derivative of the scalar field f(x, y) = x2 - 3xy alongthe parabola y = x2 - x + 2 at the point (1,2). . .' Solution. At an arbitrary point (x, r) the gradient vector is afVf(X, y) = $Ji + j = (2x - 3y)i - 3xj.ay At the point (1,2) we have Vf(l, 2) = -4i - 3j. The parabola can be representedparametrically by the vector equation r(t) = ti + (t2 - t + 2)j. Therefore, r(1) = i + 2j,r'(t) = i + (2t - l)j, and r'(l) = i + j.

For this representation of C the unit tangent vector T(1) is (i + j)/%h and the required directional derivative is Vf(l ,2) * T(1) = -7/d?. EXAMPLE 3. Let f be a nonconstant scalar field, differentiable everywhere in the plane,and let c be a constant. Assume the Cartesian equation f (x, y) = c describes a curve Chaving a tangent at each of its points. Prove that f has the following properties at eachpoint of C: (a) The gradient vector of Of is normal to C.(b) The directional derivative off is zero along C.(c) The directional derivative off has its largest value in a direction normal to C. 266 DifSerentiaI calculus of scalar and vectorjelds Solution. If T is a unit tangent vector to C, the directional derivative off along C is thedot product Of. T.

This product is zero if Vf is perpendicular to T, and it has its largestvalue if Of is parallel to T. Therefore both statements (b) and (c) are consequences of (a).To prove (a), consider any plane curve r with a vector equation of the form r(t) =X(t)i + Y(t)j and introduce the function g(t) = f [r(t)]. By the chain rule we haveg'(t) = Of [r(t)] . r'(t). When r = C, the function g has the constant value c so g'(t) = 0if r(t) E C. Since g' = Of. r', this shows that Of is perpendicular to r' on C; hence Ofis normal to C.

8.16 Applications to geometry. Level sets. Tangent planes The chain rule can be used to deduce geometric properties of the gradient vector.

Let fbe a scalar field defined on a set S in R" and consider those points x in S for which f (x) hasa constant value, say f (x) = c. Denote this set by L(c), so that The set L(c) is called a levelset off. In R2, L(c) is called a Zevel curve; in R3 it is called alevel surface. FIGURE 8.6 The dotted curves are iso-thermals: f (x, y) = c. The gradient vectorVf points in the direction of the lines of flow. FIGURE 8.7 The gradient vector Vfis normalto each curve I? on the level surface j-(x, y, z> = c- Families of level sets occur in many physical applications. For example, if f(x, y)represents temperature at (x, y), the level curves off (curves of constant temperature) arecalled isothermals. The how of heat takes place in the direction of most rapid change intemperature. As was shown in Example 3 of the foregoing section, this direction is normalto the isothermals. Hence, in a thin flat sheet the flow of heat is along a family of curves Applications to geometry. Level sets.

Tangent planes 2 6 7 orthogonal to the isothermals. These are called the lines of~?ow; they are the orthogonaltrajectories of the isothermals. Examples are shown in Figure 8.6. Now consider a scalar fieldfdifferentiable on an open set S in R3, and examine one of itslevel surfaces, L(c). Let u be a point on this surface, and consider a curve I which lies onthe surface and passes through u, as suggested by Figure 8.7. We shall prove that thegradient vector V'(a) is normal to this curve at u. That is, we shall prove that Vf(a) isperpendicular to the tangent vector of I' at a. For this purpose we assume that I is VL a normal vector I/ Tangent plane /Level surface L(c) FIGURE 8.8 The gradient vector Vf is normal to the tangent plane of a level surfacef(x, y, z) = c. described parametrically by a differentiable vector-valued function r defined on some intervalJ in R1.

Since I? lies on the level surface L(c), the function r satisfies the equation fEr(Ql = c foralltinJ. If g(r) =f[r(t)] for t in J, the chain rule states that g'(t) = Vf[r(t)] - r' (t) . Since g is constant on J, we have g'(t) = 0 on J.we find that In particular, choosing t, so that g(tJ = a Vf(u) * r'(tJ = 0 . ,In other words, the gradient off at a is perpendicular to the tangent vector r'(tJ, asasserted. 268 DifSerential calculus of scalar atld vectorjelds Now. we take a family of curves on the level surface L(c), all passing through the pointa.

According to the foregoing discussion, the tangent vector? of all these curves are per-pendicular to the gradient vector V'(a). If Vf( )a IS not the zero vector, these tangent vectorsdetermine a plane, and the gradient V'(a) is normal to this plane. (See Figure 8.8.). Thisparticular plane is called the tangentpzane of the surface L(c) at a. We know from Volume I that a plane through a with normal vector N consists of allpoints x in R3 satisfying N.

(x - a) = 0. Therefore the tangent plane to the level surfaceL(c) at u consists of all x in R3 satisfying q (u)+-u)=O. To obtain a Cartesian equation for this plane we express x, u, and V'(u) in terms of theircomponents. Writing x = (x, y, z), a = (x1 ,yl , zl), and Vf(4 = 4fW + D2f(4i + D3fW 2 we obtain the Cartesian equation 4f(4(x - XI) + Dzf(4O) - yJ + D,f(+ - 4 = 0. A similar discussion applies to scalar fields defined in R2. In Example 3 of the foregoingsection we proved that the gradient vector Vf(u) at a point u of a level curve is perpendicularto the tangent vector of the curve at a. Therefore the tangent line of the level curve L(c) ata point a = (x1 ,rJ has the Cartesian equation W-(4(x - 4 + Dzf(4(y - UI) = 0. 8.17 Exercises 1. In this exercise you may assume the existence and continuity of all derivatives under con-sideration. The equations u = f(x, y) , x = X(t), y = Y(t) define u as a function of t, sayu = F(t).(a) Use the chain rule to show that af afF(t) = GX'(f) + qJ Y'(l), where af/lax and aflay are to be evaluated at [X(t), Y(r)].(b) In a similar way, express the second derivative F"(t) in terms of derivatives off, X, andY. Remember that the partial derivatives in the formula of part (a) are composite functionsgiven by z = 4f[m), W)l, af- = D,J'w(th W)l.aY 2. Refer to Exercise 1 and compute F'(t) and F"(t) in terms of t for each of the following specialcases :(a) f(x, y) = x2 + y2, X(t) = t, Y(t) = t2.(b) f(x, y) = exy cos (xy"), X(t) = cos t , Y(t) = sin t .(c) f (x, y) = log [(l + f&)/(1 + eY2)], X(t) = e, Y(t)t = emt.

Derivatives of vector fields 269 3. In each case, evaluate the directional derivative off for the points and directions specified:(a) f(x, y, z) = 3x - Sy + 22 at (2,2, 1) in the direction of the outward normal to the spherex2 + y2 + 22 = 9. (b) f(x> y, z) = x2 - y2 at a general point of the surface x2 $- y2 + z2 = 4 in the directionof the outward normal at that point. (cl f(X> y, z) = x2 + y2 - z2 at (3,4, 5) along the curve of intersection of the two surfaces2x2 +2yz-z2 =25andx2 +y2 =z2. 4. (a) Find a vector V(x, y, z) normal to the surface 2 = Jx" + y2 + (x2 + y2p at a general point (x, y, z) of the surface, (x, y, z) # (0, 0,O).(b) Find the cosine of the angle 0 between V(x, y, z) and the z-axis and determine the limit ofcos 8 as (x, y, z) + (0, 0, 0). 5. The two equations e" cos v = x and eU sin v = y define u and v as functions of x and y, sayu = U(x, y) and v = V(x, y). Find explicit formulas for U(x, y) and V(x, y), valid for x > 0,and show that the gradient vectors VU(x, y) and V V(x, y) are perpendicular at each point 6, y).6. Letfk y> = &yl . (a) Verify that af/ax and af/lay are both zero at the origin.(b) Does the surface z =f(x, y) have a tangent plane at the origin? [Hint: Consider thesection of the surface made by the plane x = y .] 7. If (x0, yO, z,,) is a point on the surface z = xy , then the two lines z = y,x , y = y,, and z = xOy,x = x,, intersect at (x0, ye, zO) and lie on the surface. Verify that the tangent plane to

thissurface at the point (x0, ye, z,,) contains these two lines. 8. Find a Cartesian equation for the tangent plane to the surface xyz = a3 at a general point(x0, y,, , zJ. Show that the volume of the tetrahedron bounded by this plane and the threecoordinate plane is 9a3/2. 9. Find a pair of linear Cartesian equations for the line which is tangent to both the
surfaces~~+y~+2~~=4andz=e~~atthepoint(l,l,l). 10. Find a constant c such that at any point of intersection of the two spheres (x - c)2 + y2 + 22 = 3 and x2 + (y - 1)s + zz = 1 the corresponding tangent planes will be perpendicular to each other.11. If r1 and r2 denote the distances from a point (x, y) on an ellipse to its foci, show that the equation
r1 + r2 = constant (satisfied by these distances) implies the relation T. V(r, + r2) = 0, where T is the unit tangent to the curve. Interpret this result geometrically, thereby showingthat the tangent makes equal angles with the lines joining (x, y) to the foci. 12. If Vf(x, y, z) is always parallel to xi + yj + zk, show thatfmust assume equal values at
thepoints (0, 0, a) and (0, 0, -a). 8.18 Derivatives of vector fields Derivative theory for vector fields is a straightforward extension of that for scalar fields. Letfi S + Rm be a vector field defined on a subset S of R". If a is an interior point of S 2 7 0 D(fferentia1 calculus of scalar and vectorjelds and if y is any vector in R'" we define the derivativef'(a; y)
by the formula f'(a; y) = lim f@ + hY) - f(4h-0 h ' whenever the limit exists. The derivativef'(a; y) is a vector in R".Letf, denote the kth component off. We note that the derivativef'(u; JJ) exists if and only iffi(u; y) exists for each k = 1, 2, . .
. , m, in which case we have f'(a; y> = (flYa; Y>, . . * , fAda; Y>> =lgKu; Ykk 3 where ek is the kth unit coordinate vector.We say that f is at an interior point a if there is a linear transformation such that (8.16) f(u + u) = f(a) + T,(u) + II u II E(a, 4 7 where E(u, a) --f 0 as u + 0. The first-order Taylor formula (8.16) is to hdd for all
uwith /(u I/< r for some r > 0. The term E(u, u) is a vector in R" . The linear trans formation T, is called the total derivative off at a. For scalar fields we proved that T,(y) is the dot product of the gradient vector Vf(u)withy. For vector fields we will prove that T,(y) is a vector whose kth component is the dotproduct Vfk(u) .
JJ . THEOREM 8.9. Assume f is at a with total derivative T,. Then the derivativef '(a; y) exists jbr every u in R", and tile have (8.17) T,(Y) = f '(u; Y> . Moreover, iff = (fi , . . . , f,) and ify = (yI , . . .
, ,vJ , we have (8.18) T,(Y) = 5 vfk(u> ' Y ek = (vfi(a> ' Y, .
. . , C7fm(a> ' Y> .k=l Proof.

We argue as in the scalar case. Ify= 0, thenf'(u,y) = Oand T,(O) = 0.Therefore we can assume that y # 0. Taking u = lzy in the Taylor formula (8.16) we have f(u + hy) -f(u) = T,,(hy) + llhyli Ecu, 8) = hT,b) + Ihl 11~11 E(u, 8). Dividing by h and letting h + 0 we obtain (8.17).To prove (8.18) we simply note that f'@;y) =k~lf;(u;y) ek =kg 'jkb> 'ye,.
DifSerentiability implies continuity 271 Equation (8.18) can also be written more simply as a matrix product, T,(Y) = Df(a>v > where Df(a) is the m x n matrix whose kth row is Vfk(a), and where y is regarded as ann x 1 column matrix. The matrix Df (a) is called the Jacobian matrix off at a. Its kj entryis the partial derivative Djfk(a).
Thus, we have Df(a) = The Jacobian matrix Df(a) is defined at each point where the mn partial derivativesD,,&(a) exist. The total derivative T,, is also written as f '(a). The derivative f'(a) is a linear trans-formation; the Jacobian Df(a) is a matrix representation for this transformation. The first-order Taylor formula takes the form (8.19) f(a + 4 =f (a) +f
'(u)(s) + II4 E(a, u), where E(a, v) -f 0 as u -f 0. This resembles the one-dimensional Taylor formula. Tocompute the components of the vector f '(a)( )u we can use the matrix product Df(a)u orformula (8.18) of Theorem 8.9. 8.19 Differentiability implies continuity THEOREM 8.10. If a vector$eldf is difSerentiable at a, then f is continuous at a. Proof. As
in the scalar case, we use the Taylor formula to prove this theorem. If we letu + 0 in (8.19) the error term II u/I E(a, u + 0. The linear part f '(a)(u) also tends to 0)because linear transformations are continuous at 0. This completes the proof. At this point it is convenient to derive an inequality which will be used in the proof ofthe chain rule in the next
section. The inequality concerns a vector field f differentiable ata; it states that (8.20) IIf'M~Nl 5 $64 Ilull 9 where Mj(4 = If IIVf,(4ll.k=l To prove this we use Equation (8.18) along with the triangle inequality and the Cauchy-Schwarz inequality to obtain lif'(a>(u>ll =II (") * u ek(I 5 f IVfk(a> * uI 5 2 Iivfk(a>li ilull .k=l x=1 272
Dtxerential calculus of scalar and vector fields 8.20 The chain rule for derivatives of vector fields THEOREM 8.11. CHAIN RULE. Let f and g be vector fields such that the compositionh = f 0 g is defined in a neighborhood of a point a. Assume that g is dtyerentiable at a, with total derivative g'(a). Let b = g(a) and assume that f is dtflerentiable at b,
with totalderivative f '(b). Then h is dtifSentiable at a, and the total derivative h'(a) is given by h'(a) =f 'W Og'W, the composition of the linear transformations f '(b) and g'(a). Proof We consider the difference h(a + y) - h(a) for small IIy II, and show that we havea first-order Taylor formula. From the definition of h we have (8.21) hh(a + Y) - 44 = f Ma +
~91 - f Cd41 = f (b + 8) - f (4 y where u = g(u + y) - g(a). The Taylor formula for g(a + y) gives us (8.22) u = g'(a)(y) + IIYII E&, Y> > w h e r e E,(a, y)- 0 a s y - 0 . The Taylor formula for f(b + v) gives us (8.23) f (b + u) - f(b) = f '(b)(u) + II u II E,(b, 4 3 where EJb, u) ---f 0 as u ---f 0. Using (8.22) in (8.23) we obtain (8.24) f(b + 4 -f(b) =f'(b)d(a)(y) +f
'(b)(Ilyll &(a, Y>> + II 41 E,@, u)= f '(W(a)(y) + IIY I I E(a, Y> 3 where E(a, 0) = 0 and (8.25) E(a, Y> = f '(b)(E(&, YN + 1 E,(b, 4 i f y#O.

To complete the proof we need to show that E(a, y) - 0 as y - 0.The first term on the right of (8.25) tends to 0 as y - 0 because E,(a, y) - 0 as y - 0 and linear transformations are continuous at 0.In the second term on the right of (8.25) the factor Ef(b, u) - 0 because u - 0 as y-f 0. The quotient llull/IlyII remains bounded because, by (8.22) and (8.20) we
have Therefore both terms on the right of (8.25) tend to 0 as y + 0, so E(a, y) - 0.Thus, from (8.24) and (8.21) we obtain the Taylor formula hh(a + Y> - 44 = f '(b)d(a)(y) + I IY I I Way Y) y Matrix form of the chain rule 273 where E(a, y) - 0 as y - 0. This proves that h is differentiable at a and that the totalderivative K(a) is equal to the compositionf'(b) 0
g'(u). 8.21 Matrix form of the chain rule Let It = f 3 g , where g is differentiable at a andfis differentiable at b = g(u). The chainrule states that h'(a) = f'(b) 0 g'(u). We can express the chain rule in terms of the Jacobian matrices DA(u), Df(b), and Dg(u)which represent the linear transformations h'(a), f'(b), and g'(u), respectively. Sincecomposition of
linear transformations corresponds to multiplication of their matrices, weobtain (8.26) where b = g(u). This is called the matrix form of the chain rule. It can also be written as a set of scalarequations by expressing each matrix in terms of its entries. Suppose that a E RP , 6 = g(u) E R" , and f(b) E Rm. Then h(u) E R" and we can write g=(g,,...,g,), f=
cfl~ * ~ * ,fm>, h = (h,, . . . , h,).

Then D/z(u) is an m x p matrix, Df(b) is an m x n matrix, and Dg(u) is an n x p matrix,given by The matrix equation (8.26) is equivalent to mp scalar equations, DM4 = i QJXWigd4 2 f o r i=l,2,...,m and j=l,2)...) p,k=l These equations express the partial derivatives of the components of h in terms of thepartial derivatives of the components off and g.
EXAMPLE 1. Extended chain rule for scalar jelds. Suppose 1" is a scalar field (m = 1).Then h is also a scalar field and there arep equations in the chain rule, one for each of thepartial derivatives of h : D&4 = i ,g,(4 > for j = 1,2, . . . , p.k = l The special casep = 1 was already considered in Section 8.15. In this case we get only
oneequation, k'(a) = i PJVM(4.k=l 274 calculus of scalar and vectorjelds Nowtakep=2andn=2. Write a = (s, t) and b = (x, JJ) . Then the components xand y are related to s and t by the equations x = g,(s, t) 7 Y = g,(s, t>.
The chain rule gives a pair of equations for the partial derivatives of h: D,h(s, t> = W-k y> D,g,(s, 0 + D,f(x, Y> 4g& t) 3D,h(s, t) = D,f(x, y) D,g,(s, t) + W-(x, y) D,g&, t>. In the ¬ation, this pair of equations is usually written as (8.27)ah-= afam+zay ah-=--as ax as ay as ' at EXAMPLE 2. Polar coordinates. The temperature of a thin plate is described by
a scalarfieldf, the temperature at (x, u) beingf(x, y). Polar coordinates x = r cos 8, y = r sin 8are introduced, and the temperature becomes a function of r and 19 determined by thequation y(r, 0) = f(r cos 8, r sin e) . Express the partial derivatives a&% and in terms of the partial derivatives aflaxand aflay. Solution. We use the chain
rule as expressed in Equation (8.27), writing (r, 0) instead of(s, t), and p instead of h. The equations give usx = rc0s 8, y = r sin e k,y $=cose, $=sinO, g= -rsinfI, $=rcose. Substituting these formulas in (8.27) we obtain (8.28) afz = E cos 8 + ay sin 8, aQ? af afas = -rzsine+r-c0sO. ay These are the required formulas for and i?p/aO.
EXAMPLE 3. Second-order partial derivatives.

Refer to Example 2 and express the second-order partial derivative a2pl/a02 in terms of partial derivatives off. Solution. We begin with the formula for &p/a0 in (8.28) and differentiate with respectto 0, treating r as a constant. There are two terms on the right, each of which must be Exercises differentiated as a product. Thus we have 275 3%-=(8 .29)
ae2 + r af a(cos e, + r- -ay ae r sin e af + r cos 8ay To compute the derivatives of aflax and aflay with respect to 8 we must keep in mind that,as functions of r and 8, af/lax and afjay are composite functions given by af = D f(r cos 8 r sin e)ax ' ' and af- = D, f (r cos 8, r sin e) .ay Therefore, their derivatives with respect to 0 must be determined by use of
the chain rule.We again use (8.27), withfreplaced by DJ, to obtain a af uhf) ay- (1 =- - -ae ax a(D,f) ax +ax ae zf c r sin 0) + 37ay ae=ax2 - (r cos e) .dyax Similarly, using (8.27) withfreplaced by D,f, we find a af(1 a(D,f) ax + a(D,f) ay ?f - = - - - -=-ae ay ax ae ay ae axay (-rsinO)fazf(reOsO).aY2 When these formulas are used in (8.29) we obtain a2q as
w atf =a82 -rcosO- - + r2sin28- - r2 sinOcostl----ax ax2 ayax - r sin 8 Lf - Y' sin 8 cos 8 a"f + r2 cos' 8 aYay axay -.w This is the required formula for az~/a02. Analogous formulas for the second-order partialderivatives , a2p/(ar a@, and a2yl/(a0 &) are requested in Exercise 5 of the nextsection. 8.22 Exercises In these exercises you may
assume differentiability of all functions under consideration. 1. The substitution t = g(x, y) converts F(t) into f (x, y), wheref(x, y) = F[g(x, y)] .(a) Show that z = F'rg(x, y>l $ and (b) Consider the special case F(t) = Fin t, g(x, y) = cos (x2 + y2). Compute af/lax andaflay by use of the formulas in part (a). To check your result, determine f (x, y) explicitly
interms of x and y and compute af/ ax and af/ ay directly fromJ: 2 7 6 D$erential calculus of scalar and vector fields 2. The substitution u = (x - y)/2, v = (x + y)/2 changesf (u, v) into F(x, y).
Use an appropriateform of the chain rule to express the partial derivatives aF/ ax and aF/ ay in terms of the partialderivatives af/ au and afj au. 3. The equations u =f(x, y), x = X(s, t), y = Y(s, t) define u as a function of s and t, sayu = F(s, t).(a) Use an appropriate form of the chain rule to express the partial derivatives aF/as andaqat in terms of aflax,
aflay, ax/as, ax/at, atlas, arlat.(b) If azf/c ?X ay) = azfl(ay ax), show that a2F af a2x a2f ax 2ax%+% z +2---(1 ax ay a2f af a2y a2f ay 2-=a.32 as as ax ay +dyas2+ay2ds.(1 (c) Find similar formulas for the partial derivatives a2F/(as at) and a2F/ at2.4. Solve Exercise 3 in each of the following special cases: (a) X(s, t) = s + t, Y(s, t) = st.(b) X(s, t) = st,
Y(s, t) = s/t.ce> as, t> = (s - tY2, Y(s, t) = (s + ty2. 5. The introduction of polar coordinates changes f(x, y) into p(r, .9), where x = r cos 8 andy = r sin 0.

Express the second-order partial derivatives a"qlar2, a2q/(au a@), and a*p/(80 ar)in terms of the partial derivatives off. You may use the formulas derived in Example 2 ofSection 8.21. 6. The equations u = f(x, y, z), x = X(r, s, t), y = Y(r, s, t), and z = Z(r, s, ti define u as afunction ofr, s, and t, say u = F(r, s, t) . Use an appropriate form of the chain
rule to expressthe partial derivatives aF/ au, aF/ as, and aF/ at in terms of partial derivatives off, X, Y, and Z. 7. Solve Exercise 6 in each of the following special cases:(a) X(r, s, t) = r + s + t, Y(r, s, t) = r - 2s + 3t, Z(r, s, t) = 2r + s - t.(b) X(r, s, t) = r2 + s2 + t2, Y(r, s, t) = r2 - s2 - t2, Z(r, s, t) = r2 - s2 + t2. 8. The equations u = f (x, y, z), x = X(s, t), y
= Y(s, t) , z = Z(s, t) define u as a function ofs and t, say u = F(s, t) . Use an appropriate form of the chain rule to express the partial de-rivatives aF/ as and aF/ at in terms of partial derivatives off, X, Y, and Z. 9. Solve Exercise 8 in each of the following special cases: 10. 11. 1 2 . 1 3 . (a) X(s, t) = s2 + t2, Y(s, 1) = s2 - t2, Z(s, t) = 2st.(b) X(s, t) = s + t,
Y(s, t) = s - t, Z(s, t) = st.The equations u = f (x, y), x = X(r, s, t), y = Y(r, s, t) define u as a function of r, s, and t, say u = F(r, s, t). Use an appropriate form of the chain rule to express the partial derivativesaF/ar, aF/ as, and aFlat in terms of partial derivatives off, X, and Y.Solve Exercise 10 in each of the following special cases:(a) X(r,s,t) = r +s, Y(r,
s, t) = t.(b) X (r , s , t) = r + s + t , Y(r, s, t) = r2 + s2 + t2. Cc) X(r, s, t> = r/s, Y(r, s, t) = s/t. Let 0) = ,fk(x)l , where g = (gr , . . . , g,) is a vector field differentiable at a, and f is ascalar field differentiable at b = s(a).
Use the chain rule to show that the gradient of h canbe expressed as a linear combination of the gradient vectors of the components of s, asfollows : Vh(u) = 5 &f (b)Vg&) .k=l (a) If f(x, y, z) = xi + yi + zk, prove that the Jacobian matrix Df(x, y, z) is the identitymatrix of order 3. Suficient conditions for the equality of mixed partial derivatives 277 (b)
Find all differentiable vector fieldsf: R3 - R3 for which the Jacobian matrix Df(x, y, z)is the identity matrix of order 3.(c) Find all differentiable vector fieldsf: R3 - R3 for which the Jacobian matrix is a diagonalmatrix of the form diag (p(x), q(y), r(z)), wherep, q, and r are given continuous functions. 14. Let f: R2 + R2 and g: R3 --+ R2 be two vector fields
defined as follows: f(x, y) = exfzYi + sin 0, + 2x)j, g(u, v, w) = (u + 2u2 + 3w3)i + (20 - u2)j. (a) Compute each of the Jacobian matrices Df(x, y) and Dg(u, u, w).(b) Compute the composition h(u, v, W) =f[g(u, v, w)].(c) Compute the Jacobian matrix DR(l , - 1, 1). 15. Let f: R3 + R2 and g: R3 + R3 be two vector fields defined as follows: f(x,y,z) = (x2 +y +
z)i + (2x + y + z2)j, g(u, 21, w) = uu"w2i + w2 sin v j + Auk. (a) Compute each of the Jacobian matrices Df(x, y, z) and Dg(u, v, w).(b) Compute the composition h(u, u, w) =f[g(u, u, w)].(c) Compute the Jacobian matrix DR(u, 0, w). k8.23 Sufficient conditions for the equality of mixed partial derivatives Iffis a real-valued function of two variables, the two
mixed partial derivatives D,,,fand D,,,Sare not necessarily equal. By D,,zf we mean D,(D2f) = azfi(ax ay), and byD,,,fwe mean D2(D1f) = azfl(ay ax). F or example, ifJ'is defined by the equations .0x, Y> = XY f for (x3 Y> Z (0, 0)) f(O, 0) = 0 2 it is easy to prove that D,,J(O, 0) = - 1 and D,,,f(O, 0) = 1. This may be seen as follows:The definition of
D2,1f(0, 0) states that (8.30) D2,J(0, 0) = lim W(O, k) - D,f(o, 0)k-0 k -' Now we have D,f(O, 0) = limf(k 0) - f@, 0) = oh-t0 h and, if (x, y) # (O,O), we find 4m Y> =y(x4 + 4x29 - y") (x2 + y2)2 * Therefore, if k # 0 we have D,f (0, k) = -k5/k4 = -k and hence W(o, k) - W-(0,0) = -lk 278 DifSerential calculus of scalar and vectorjelds Using this in (8.30) we
find that D,,,f(O, 0) = -1 . A similar argument shows that%f(O, 0) = 1, and hence &,,f(O, 0) Z D~f(0,0). In the example just treated the two mixed partials D,,,f and D,,,f are not both con-tinuous at the origin. It can be shown that the two mixed partials are equal at a point(a, b) if at least one of them is continuous in a neighborhood of the point. We shall
provefirst that they are equal if both are continuous. More precisely, we have the followingtheorem. THEOREM 8.12. A SUFFICIENT CONDITION FOR EQUALITY OF MIXED PARTIAL DERIVATIVES. Assumef is a scalar$eldsuch that the partial derivatives Dlf, D,f, D,,,f, and D,,,f exist onan open set S. If (a, b) is apoint in S at which both Dl,zf and D,,,f
are continuous, we have (8.31) Dl,2f@, b) = D2,1f(a9 b) I Proof. Choose nonzero h and k such that the rectangle R(h, k) with vertices (a, b),(a + h, b), (a + h, b + k), and (a, b + k) lies in S. (An example is shown in Figure 8.9.) (a, h + k) (a + h, h + k) (4 6 :ll + h, b) FIGURE 8.9 A@, k) is a combination of values off at the vertices. Consider the
expression A(h,k)=f(a+h,b+k)-ff(a+h,b)-f(a,b+k)+f(a,b). This is a combination of the values off at the vertices of R(h, k) , taken with the algebraicsigns indicated in Figure 8.9. We shall express A(h, k) in terms of D,,,f and also in termsof Dd We consider a new function G of one variable defined by the equation G(x) =f(x, b + 4 -f&b) for all x between a
and a + h . (Geometrically, we are considering the values off at thosepoints at which an arbitrary vertical line cuts the horizontal edges of R(h, k).) Then wehave (8.32) A(h, k) = G(a + h) - G(a).
SufJicient conditions for the equality of mixed partial derivatives 279 Applying the one-dimensional mean-value theorem to the right-hand member of (8.32) weobtain G(a + h) - G(a) = hG'(x,), where x1 lies between a and a + h. Since G'(x) =W-(x, b + k) - D$(x, b), Equation (8.32) becomes (8.33) Nh, k) = h[D,fh b + k) - 4fh b)l. Applying the mean-value
theorem to the right-hand member of (8.33) we obtain (8.34) A@, 4 = hkhfhy,), where y1 lies between b and b + k. The point (x,, yJ lies somewhere in the rectangleR@, 4. Applying the same procedure to the function H(y) = f(a + h, y) - f(a, y) we find asecond expression for A(h, k), namely, (8.35) A@, k> = hk%fb,,yd, where (xz, yJ also lies in R(h, k) .
Equating the two expressions for A(h, k) and cancellinghk we obtain %f(Xl,Yl) = DZ,lf(&YJ * Now we let (h, k) -+ (0,O) and use the continuity of D,,,f and D,,,f to obtain (8.31).

The foregoing argument can be modified to prove a stronger version of Theorem 8.12. THEOREM 8.13. Letf be a scalarfield such that the partial derivatives Dlf, D,f, and D,,,fexist on an open set S containing (a, b). Assume further that D,,,f is continuous on S. Thenthe derivative D,,,f (a, b) exists and we have D,,,f (a, b) = Dz,S (a, b) . Proof. We define
A(h, k) as in the proof of Theorem 8.12. The part of the proof leadingto Equation (8.34) is still valid, giving us (8.36) A@, 4- = 4,,f(xl, ~1)hk for some (x1, yJ in the rectangle R(h, k). The rest of the proof is not applicable since itrequires the existence of the derivative D,,,f(a, b), which we now wish to prove. The definition of DI,,f(a, b) states that (8.37)
280 DifSerential calculus of scalar and vector fields We are to prove that this limit exists and has the value D,,,f (a, b). From the definition ofD,f we have and D fca b) = limf(a, b + k) - f(a, b)2 9 k-0 k DJ(a + h, b) = lim f(a + h, b + k) - f(a + h, b) k-t0 k Therefore the difference quotient in (8.37) can be written as &f(a + k b) - W(a, b) = lim '(h, k)h k'Ohk'
Using (8.36) we can rewrite this as (8.38) D&a + h, b) - W(a, b)h = lim 4,Jh 9 Yd .k-+0 To complete the proof we must show that (8.39). [lim lim &J(x, , vl> = D2J(a, b) .h-0 k-0 1 When k -+ 0, the point y1 -+ b, but the behavior of x1 as a function of k is unknown.If we knew that x1 approached some limit, say R, as k -+ 0, we could use the continuity
ofDzsIf to deduce that lim D2J(x,, Y,) = &f(% b) .k+O Since the limit f would have to lie in the interval a 5 1 < a + h , we could then let h + 0and deduce (8.39). However, the fact that 3 depends on k in an unknown fashion makes aslightly more involved argument necessary. Because of Equation (8.38) we know that the following limit exists: lim
D2,J(xl~ Y,) .k r0 Let us denote this limit by F(h). To complete the proof we must show that lim F(h) = D,,,f(a, b) .h - 1 0 For this purpose we appeal to the definition of continuity of D,,,f at (a, b).Let E be a given positive number.

Continuity of D,,,f at (a, b) means that there is an open disk N with center (a, b) and radius 6, say, such that ID2,Jh Y> - 4,&, WI < z whenever (x, y) E N.L Miscellaneous Exercises 281 If we choose h and k so that Ihl < S/2 and Ikl < 612, the entire rectangle shown in Figure8.9 will lie in the neighborhood N and, specifically, the point (xi, y,) will be in
N. Therefore(8.40) is valid when (x, r) = (x,, vi) and we can write (8.41) 0 5 I4Jb,, Y,) - 4JXa, b)l < 5 . Now keep h fixed and let k --f 0. The term D,,,f(x,, y,) approaches F(h) and the otherterms in (8.41) are independent of k. Therefore we have 0 I F'(h) - 4,,f(a, b)l I i < E, provided that 0 < IhJ < 8/2. But this is precisely the meaning of the statement lim
F(h) = D,,,f(a, b)h-0 and, as we have already remarked, this completes the proof. Note: It should be observed that the theorem is also valid if the roles of the twoderivatives D,*,f and D,,,f are interchanged. 8.24 Miscellaneous exercises 1. 2. 3. 4. 5. 6 . Find a scalar field f satisfying both the following conditions:(a) The partial derivatives D,f(O, 0) and
D,f(O, 0) exist and are zero.(b) The directional derivative at the origin in the direction of the vector i + j exists and hasthe value 3. Explain why such an f cannot be differentiable at (0,O).Let f be defined as follows : f(x,y) =yg$ if thy) Z (0, O>, f(O,O) = 0. Compute the following partial derivatives, when they exist: Df (0, 0), D,f (0, 0), D,,,f (0, 0), D,,,f (0,
0). Let f (x, y) = xy3- if (x, u) # (0,O) , and define f (0,O) = 0.x3 + y6 (a) Prove that the derivative f '(0; a) exists for every vector a and compute its value in termsof the components of a.(b) Determine whether or not f is continuous at the origin.Define f (x, y) = ~ ~ ~ ~e-t2 dt for x > 0, y > 0. Compute af/ ax in terms of x and y.Assume that the equations u
= f (x, y) , x = X(t), y = Y(t) define u as a function of t, sayu = F(t). Compute the third derivative F"'(t) in terms of derivatives off, X, and Y.The change of variables x = u + u , y = uv2 transforms f (x, JJ) into g(u, u). Compute thevalue of azg/(av au) at the point at which u = 1, v = 1 , given that af azf azf azf azfay=ax3=ay2--=ax ay -=1ay ax at that point.
282 DifSerentiaI calculus of scalar and vector fields 7. The change of variables x = uu, y = i(u2 - 02) transformsf(x, y) to g(u, 0).(a) Calculate g/au, ag/au, and a2g/(& ao) in terms of partial derivatives of & (You mayassume equality of mixed partials.) (b) If II Vf(x, yW = 2 for all x and y, determine constants a and b such that 8. Two functions F and G of
one variable and a function z of two variables are related by theequation [F(x) + G(y)12 eZ(Z,y) = 2F'(x)G'(y) whenever F(x) + G(y) # 0.

Show that the mixed partial derivative D,,,z(x, y) is never zero.(You may assume the existence and continuity of all derivatives encountered.) 9. A scalar field f is bounded and continuous on a rectangle R = [a, b] x [c, d] .
A new scalarfield g is detined on R as follows : gh 0) = s,"[/;.fk y) dx] dy . (a) It can be shown that for each fixed u in [a, b] the function A defined on [c, d] by theequation A(y) = JE f(x, y) dx is continuous on [c, d]. Use this fact to prove that ag/ au existsand is continuous on the open rectangle S = (a, b) x (c, d) (the interior of R).(b) Assume that .I' D :'
;f(x, 4') dx] dy = j-; [I'-fhr) dy] dx for all (u, v) in R. Prove that g is differentiable on S and that the mixed partial derivativesD,,,g(u, u) and &,g(u, v) exist and are equal tof(u, v) at each point of S. 10. Refer to Exercise 9. Suppose u and v are expressed parametrically as follows: u = A(t),v = B(t) ; and let q(t) = g[A(t), B(t)].(a) Determine q'(t) in terms off,
A', and B'.(b) Compute p)'(t) in terms of t whenf(x, y) = e'+3/ and A(t) = B(t) = t2. (Assume R liesin the first quadrant.) 11. Iff(x,y, z) = (r x A) . (u x B), where r = xi + yj + zk and A and B are constant vectors,showthat Vf(x,y,z) = B x (r x A) +A x (r x B). 12. Let r = xi + yj + zk and let r = J/r/J . If A and B are constant vectors, show that: (a) A.I' i = -7.0
(b) B. $A.
V(i)) = 3A';; r -7, 13.

Find the set of all points (a, b, c) in 3-space for which the two spheres (x - a)" + (y - b)" +(z -c)~ = 1 andx2 +y2 +z 2 = 1 intersect orthogonally. (Their tangent planes should beperpendicular at each point of intersection.) 14. A cylinder whose equation is y =f(x is tangent to the surface z2 + 2xz + y = 0 at all)points common to the two surfaces.
Findf(x). 9APPLICATIONS OF DIFFERENTIAL CALCULUS 9.1 Partial differential equations The theorems of differential calculus developed in Chapter 8 have a wide variety ofapplications. This chapter illustrates their use in some examples related to partial differen-tial equations, implicit functions, and extremum problems. We begin with some
elemen-tary remarks concerning partial differential equations. An equation involving a scalar fieldfand its partial derivatives is called a partial dlferen-tiul equation. Two simple examples in whichf is a function of two variables are the first-order equation (9.1) and the second-order equation (9.4w(x3 Y) + axx, Y) -=o axa ay2 . Each of these is a
homogeneous linear partial differential equation. That is, each has theform L(f) = 0, where L is a linear differential operator involving one or more partialderivatives.

Equation (9.2) is called the two-dimensional Luphce equation. Some of the theory of linear ordinary differential equations can be extended to partialdifferential equations. For example, it is easy to verify that for each of Equations (9.1)and (9.2) the set of solutions is a linear space. However, there is an important differencebetween ordinary and partial
linear differential equations that should be realized at theoutset. We illustrate this difference by comparing the partial differential equation (9.1)with the ordinary differential equation (9.3) f'(x) = 0. The most general function satisfying (9.3) isJ'(x) = C, where C is an arbitrary constant:In other words, the solution-space of (9.3) is one-dimensional. But
the most generalfunction satisfying (9.1) is fh u> = gw 3 283 284 Applications of difSerentia1 calculus where g is any function of y. Since g is arbitrary we can easily obtain an infinite set ofindependent solutions. For example, we can take g(y) = ecu and let c vary over all realnumbers. Thus, the solution-space of (9.1) is inznite-dimensional. In some
respects this example is typical of what happens in general. Somewhere in theprocess of solving a first-order partial differential equation, an integration is required toremove each partial derivative. At this step an arbitrary function is introduced in thesolution. This results in an infinite-dimensional solution space. In many problems involving partial
differential equations it is necessary to select from thewealth of solutions a particular solution satisfying one or more auxiliary conditions. Asmight be expected, the nature of these conditions has a profound effect on the existence oruniqueness of solutions. A systematic study of such problems will not be attempted in thisbook. Instead, we will treat
some special cases to illustrate theideas introducedinChapter8. 9.2 A first-order partial differential equation with constant coefficients Consider the first-order partial differential equation (9.4)3 am Y) + 2 am Y>_ o ax ay All the solutions of this equation can be found by geometric considerations. We express theleft member as a dot product, and write
the equation in the form (3i + 2j) * Of (x, y) = 0. This tells us that the gradient vector Of (x, y) is orthogonal to the vector 3i + 2j at eachpoint (x, y). But we also know that V'(x, y) is orthogonal to the level curves of.6 Hencethese level curves must be straight lines parallel to 3i + 2j. In other words, the level curvesoff are the lines 2x - 3y = c.
Thereforef(x, y) is constant when 2x - 3y is constant.

This suggests that (9.5) f(x, y> = g(2x - 3y) for some function g.Now we verify that, for each differentiable function g, the scalar fieldf defined by (9.5) does, indeed, satisfy (9.4). Using the chain rule to compute the partial derivatives off wefind af- = 2g'(2x - 3y), afax - = -3g'(2x - 3y))ay 3 E + 2 ; = 6g'(2x - 3y) - 6g'(2x - 3y) = 0. Therefore, f satisfies
(9.4). A first-order partial difSerentia1 equation with constant coeficients 285 Conversely, we can show that every differentiablefwhich satisfies (9.4) must necessarilyhave the form (9.5) for some g. To do this, we introduce a linear change of variables, (9.6) x = Au + Bv, y=Cu+ Dv. This transformsf'(x, v) into a function of u and v, say h(u, v) =f(Au +
Bv, CM + Dv) . We shall choose the constants A, B, C, D so that h satisfies the simpler equation (9.7)wu, VI (),-= au Then we shall solve this equation and show that f has the required form.Using the chain rule we find ah afax af ay af af-=--au axaz,+--=GA+-C.ay au 8)) Sincef satisfies (9.4) we have afjay = - (3/2)(af/ax), so the equation for ah/au
becomes Therefore, h will satisfy (9.7) if we choose A = {C. Taking A = 3 and C = 2 we find (9.8) x = 311 + BP, y = 2u + Dz>. For this choice of A and C, the function h satisfies (9.7), so h(u, z,) is a function of c alone, sayMu, 1)) = g(v) for some function g. To express 1' in terms of x and J we eliminate u from (9.8) and obtain2x-3y=(2B-3D)r.
NowwechooseBandDtomake2B-3D=l,sayB=2,D = 1. For this choice the transformation (9.6) is nonsingular; we have c = 2x - 3~,and hence .f(x, y) = h(u, r) = g(c) = g(2x - 3y). This shows that every differentiable solutionfof (9.4) has the form (9.5).Exactly the same type of argument proves the following theorem for first-order equations with constant
coefficients. THEOREM 9.1. Let g he difSerentiadle on RI, and let f be the scalarjeld dejined on R2 b)vthe equation (9.9) f(x, y) = g(bx - ay), 2 8 6 Applications of d$erential calculus sphere a and b are constants, not both zero. ThenJ'satis$es thejrst-order partial dlJ%erentialequation (9.10)m, Y) + b V(X> Y) _ o a-z-- ayeverywhere in R2. Converseb,
every dljierentiable solution of (9.10) necessarily has the form(9.9) for some g. 9.3 Exercises In this set of exercises you may assume differentiability of all functions under consideration.

1. Determine that solution of the partial differential equation 4 af(x7Y) + 3 WGy) _ oax aY which satisfies the conditionf(x, 0) = sin x for all x.2. Determine that solution of the partial differential equation ?m Y) _ 2 af(x, .v) _ os - - - - - - - ax aY which satisfies the conditionsf(0, 0) = 0 and D&x, 0) = e" for all x.3. (a) If u(x, y) =f(xy), prove that u satisfies the
partial differential equation au au "ax -y- = OSaY Find a solution such that u(x, x) = x4erZ for all x.(b) If V(X, y) = f(x/y) for y # 0, prove that u satisfies the partial-differential equation Find a solution such that ~(1, 1) = 2 and D,u(x, l/x) = l/x for all x # 0.4. Ifg(u, v) satisfies the partial differential equation prove thatg(u, v) = p,(u) + IJ+(v), where P)~
(N) is a function of u alone and pZ(v) is a functionof 2~ alone. 5. Assume f satisfies the partial differential equation apt. a2f azf--ax2 2- -3- -o.ax ay ay Introduce the linear change of variables, x = Au + Bv, y = Cu + Dv, where A, B, C, Dare constant, and let g(u, v) = f(Au + Bv , Cu + Dv) . Compute nonzero integer values of AJirst-order partial dlxerential
equation with constant coejjicients 287 A, B, C, D such thatg satisfies azg/(au au) = 0. Solve this equation forg and thereby deter-minef. (Assume equality of the mixed partials.) 6. A function u is defined by an equation of the form x +Y4&y) = xyf ~(1XY * Show that u satisfies a partial differential equation of the form au"*aX -Y za" = G(x,y)u,ay and
find G(x, y).7. The substitution x = es, y = et convertsf(x, y) into g(s, t), whereg(s, t) =f(e", et). Iffis known to satisfy the partial differential equation azt- a*f af afX2%* +y*dy2 +xX +Yry =O, show that g satisfies the partial-differential equation azg azgs +$ =o. 8 . Let,fbe a scalar field that is differentiable on an open set Sin R". We say that f is
homogeneousof degree p over S if f(t4 = tpfw for every t > 0 and every x in S for which tx E S. For a homogeneous scalar field of degreepshow that we have x . Vf(x> = p f (x) for each x in S . This is known as Euler's theorem for homogeneous functions.
If x = (x1, . . . , x,) it can beexpressed as afx1- f...8x1 + .,; =pf(xl,...,xn).n [Hint: For fixed x, defineg(t) = f(tx) and computeg'(0).] 9. Prove the converse of Euler's theorem. That is, if f satisfies x Cf (x) = p f (x) for all x in anopen set S, then ,f must be homogeneous of degree p over S. [Hint: For fixed x, defineg(t) =f(tx) - t"f(x) and computeg'(t).] 10. Prove
the following extension of Euler's theorem for homogeneous functions of degree p inthe 2-dimensional case. (Assume equality of the mixed partials.) a2fx2 j~i2 + 2xy5 + y2$ =p(p - 1,f. 288 Applications of difSerentia1 calculus 9.4 The one-dimensional wave equation Imagine a string of infinite length stretched along the x-axis and allowed to vibrate in
thexq'-plane.

We denote by y = f(x, t) the vertical displacement of the string at the point xat time t. We assume that, at time t = 0, the string is displaced along a prescribed curve,y = F(x). An example is shown in Figure 9.1(a). Figures 9.1(b) and (c) show possibledisplacement curves for later values oft. We regard the displacementf(x, t) as an unknownfunction of x
and t to be determined. A mathematical model for this problem (suggestedby physical considerations which we shall not discuss here) is the partial differential equation where c is a positive constant depending on the physical characteristics of the string. Thisequation is called the one-dimensional waue equation. We will solve this equation subject
tocertain auxiliary conditions. Yt Y (a) t = 0 (b) t = 1 FI G U R E 9.1 The displacement curve y = J'(x, t) shown for various values of t. Since the initial displacement is the prescribed curve y = F(x), we seek a solutionsatisfying the condition .f(x, 0) = F(x). We also assume that ay/af, the velocity of the vertical displacement, is prescribed at timet = 0, say
4f(x, 0) = G(x), where G is a given function. It seems reasonable to expect that this information shouldsuffice to determine the subsequent motion of the string. We will show that, indeed, this istrue by determining the functionfin terms of F and G. The solution is expressed in a formgiven by Jean d'Alembert (1717-1783), a French mathematician and
philosopher. THEOREM 9.2. D'ALEMBERT'S SOLUTION OF THE WAVE EQUATION. Let Fand G be givenfunctions such that C is dlferentiable and F is twice difSerentiable on R1. Then the function fgiven by the formula (9.11) f'(x, g = F(x + cf> + R-u - ct) + _L2 2c I z'~'tc(s) ds*-ct The one-dimensional waue equation satisjies the one-dimensional
lvace equation 289 (9.12) and the initial conditions (9.13) "f (x, 0) = F(x) > D&x, 0) = G(x). Conversely, any function ,f' with equal mixed partials Mshich satisjes (9.12) and (9.13) neces-sari/y has the form (9.11). Proof. It is a straightforward exercise to verify that the function f given by (9.11)satisfies the wave equation and the given initial conditions.
This verification is left to thereader. We shall prove the converse. One way to proceed is to assume thatfis a solution of the wave equation, introduce alinear change of variables, x=Au+Bv, t=Cu+ Dv, which transformsf(x, t) into a function of u and v, say g(u, v) = f'(Au + Bv, Cu + Dv) * and choose the constants A, B, C, D so that g satisfies the simpler
equation B-0allav * Solving this equation for g we find that g(u, v) = vr(u) + ~l.Jc), where ml(u) is a functionof u alone and pi,(c) is a function of v alone. The constants '4, B, C, D can be chosen sothat u = x + ct, v = x - ct , from which we obtain (9.14) f(x, t) = f&(x + Cl> + yz(x - ct>. Then we use the initial conditions (9.13) to determine the functions
q1 and va in terms ofthe given functions F and G. We will obtain (9.14) by another method which makes use of Theorem 9.1 and avoids thechange of variables.

First we rewrite the wave equation in the form where L, and L, are the first-order linear differential operators given by Letf be a solution of (9.15) and let 4x9 t> = w-(x, t) * 290 Applications of differential calculus Equation (9.15) states that u satisfies the first-order equation L,(u) = 0. Hence, byTheorem 9.1 we have u(x, t) = 9(x + ct) for some
function ~1. Let 0 be any primitive of 9, say @(JJ) = jz v(s) ds, and let u(x, t) = h 0(x + ct) . We will show that L2(tl) = L,(f). We have S O au= 12, @'(x + co andax g = iO'(x + ct), L,v = g + c g = W(x + ct) = p(x + ct) = u(x, t) = LJ . In other words, the difference f - v satisfies the first-order equation L,(f- v) = 0. By Theorem 9.1 we must havef(x, t) - u(x,
t) = y(x - ct) for some function y. There-fore f(x, t) = v(x, t) + y(x - ct) = i 0(x + ct) + y(x - ct) . This proves (9.14) with q1 = & (Ii and p2 = y.

Now we use the initial conditions (9.13) to determine the functions v1 and v2 in terms ofthe given functions F and G. The relationf(x, 0) = F(x) implies (9.16) Q)~(x) + PZ(X> = F(x). The other initial condition, D,f(x, 0) = G(x), implies (9.17) q;(x) - up;,(x) = G(x). Differentiating (9.16) we obtain (9.18) P;(x) + P;(x) = F'(x) * Solving (9.17) and (9.18) for p;
(x) and q;(x) we find m,(x) = ; F'(x) + h G(x), &x) = ; F'(x) - $ G(x). The one-dimensional wave equation 291 Integrating these relations we get ml(x) - 940) = F(x) ; F(o) + ; o"e(r) ds,f Q)&) - p,(O) = F(x) ; F(o) - h 3(s) ds.s In the first equation we replace x by x + ct; in the second equation we replace x byx - ct. Then we add the two resulting equations
and use the fact that ~~(0) + ~~(0) =F(O) to obtain f(X, t) = pl(X + ct) + &X - ct) = F(x + ct) ; F(x - ct) + ; j";'&) ds.+ ct This completes the proof. EXAMPLE. Assume the initial displacement is given by the formula 1 + cos 7rX forF(x) = -lThe graph of F is shown in Figures 9.1(a) and 9.2(a). Suppose that the initial velocityG(x) = 0 for all x. Then the
resulting solution of the wave equation is given by the formula fcx, t) = F(x + ct> + F(x - ct)2 Figures 9.1 and 9.2 show the curve y = f(x, t) for various values of t. The figures illustratethat the solution of the wave equation is a combination of two standing waves, one travelingto the right, the other to the left, each with speed c. Further examples
illustrating the use of the chain rule in the study of partial differentialequations are given in the next set of exercises. 292 Applications of difSerentia1 calculus 9.5 Exercises 1. 2. 3. In this set of exercises you may assume differentiability of all functions under consideration. If k is a positive constant and&, t) = $x/Jkt, let f(x, I) = Ji'""' ecu2 du. af ag(a)
Show that z = e g8 - a n d af -g* agax at'" a t ' (b) Show thatfsatisfies the partial differential equation az/- afk ax2 = z (the heat equation). Consider a scalar fieldfdefined in R* such thatf(x, y) depends only on the distance r of (x, y)from the origin, say f(x, y) = g(r), where r = (x2 + y*)" .(a) Prove that for (x, y) # (0,O) we have g< + $ = i g'(r) + g"(r). (b)
Now assume further thatfsatisfies Laplace's equation, a2f a*fs+y=o, au for all (x, y) # (0,O). Use part (a) to prove that f (x, yj = a log (x2 + y*) + b for (x, Y) #(0, 0)) where a and b are constants.Repeat Exercise 2 for the n-dimensional case, where n > 3. That is, assume that f(x) =fk,..., x,) =g(r), where r = 1IxlI. Show that azf azf n-1jp +... +j-g = 1 n Y
g'(r) + g"(r) for x # 0. If f satisfies the n-dimensional Laplace equation, azfax2 + a2f1 ... +j-g =o,n for all x # 0, deduce that f(x) = a ljx112-n + b for x # 0, where a, b are constants. Note: The linear operator 'i2 defined by the equation r2f=!Y+...+$1 n is called the n-dimensional Laplacian. Exercises 293 4. 5 . 6. 7 . Two-dimensicInal Laplacian in polar
coordinates. The introduction of polar coordinates x =r cos 0 ,)I = r sin 8, converts f(x, y) into g(r, 19). Verify the following formulas: (a) /[Yf(r cos 0, r sin e>llz = ;i;ca7 Yagr + y2 jj Three-dimensional Laplacian in spherical coordinates. The introduction of spherical coordinates x = pcos0sin 92, y = p sin 0 sin p, z = pcos p, transformsf(x, y, z) to F(p, 0,
q). This exercise shows how to express the Laplacian V2f interms of partial derivatives of F.(a) First introduce polar coordinates x = r cos 0 ,y = r sin 0 to transformf (x, y, z) tog@, 0, z).Use Exercise 4 to show that a?? I a2g 1 ag a2g""f =g +ps +;z +jgz. (b) Now transform g(r, 0, z) to F(p, 0, 9) by taking z = p cos p, r = p sin p. Note that,except for a
change in notation, this transformation is the same as that used in part (a). Deduce thatcos v aF 1 a2F This exercise shows how Legendre's differential equation arises when we seek solutions ofLaplace's equation having a special form. Letfbe a scalar field satisfying the three-dimensionalLaplaceequation, T2f = 0. Introduce spherical coordinates as in
Exercise 5 and let F(p, 0, p) = f(X> y> z).(a) Suppose we seek solutions f of Laplace's equation such that F(p, 0, 9) is independent of 0and has the special form F(p, 0, ~7) = p"G(p,) . Show that f satisfies Laplace's equation if Gsatisfies the second-order equation d2Gdpii+cot~$+n(n+l)G=O. (b) The change of variable x = cos 9 (q = arccos x, -1 < x 5 1)
transforms G(v) to g(x).Show that g satisfies the Legendre equation Two-dimensional wave equation. A thin flexible membrane is stretched over the xy-plane andallowed to vibrate. Let z = f(x, y, t) denote the vertical displacement of the membrane at thepoint (x, y) at time t. Physical considerations suggest that,f satisfies the two-dimensional
waveequation, 2 9 4 Applications of calculus where c is a positive constant depending on the physical characteristics of the membrane. Thisexercise reveals a connection between this equation and Bessel's differential equation.(a) Introduce polar coordinates x = r cos 0, y = r sin 0, and let F(r, 0, t) = f(r cos 0, r sin 0, t) .Iffsatisfies
the wave equation show that F satisfies the equation a2F 2 i a% i aF-= at2c2 dl"+2z+Td,-.

c ar2 1 (b) If F(r, 0, t) is independent of 8, say F(r, 0, t) = q(r, t) the equation in (a) simplifies to Now let p be a solution such that qfr, t) factors into a function of r times a function of t, sayp(r, t) = R(r)T(t) .
Show that each of the functions R and T satisfies an ordinary linear differ-ential equation of second order.(c) If the function T in part (b) is periodic with period 2n/c, show that R satisfies the Besselequation r2R" + rR' + r2R = 0. 9.6 Derivatives of functions defined implicitly Some surfaces in 3-space are described by Cartesian equations of the form
F(x,y, z) = 0. An equation like this is said to provide an implicit representation of the surface. Forexample, the equation x2 + y2 + z2 - 1 = 0 represents the surface of a unit sphere withcenter at the origin. Sometimes it is possible to solve the equation F(x, y, z) = 0 for oneof the variables in terms of the other fwo, say for z in terms of x and y. This
leads to one ormore equations of the form z =f(x,y). For the sphere we have two solutions, z=dl -x2-y2 and z=--1/1-&y27 one representing the upper hemisphere, the other the lower hemisphere.In the general case it may not be an easy matter to obtain an explicit formula for z in terms of x and y. For example, there is no easy method for solving for z
in the equationy2 + xz + z2 - ez - 4 = 0.

Nevertheless, a judicious use of the chain rule makes itpossible to deduce various properties of the partial derivatives af/ax and aflay without anexplicit knowledge off(x, y). The procedure is described in this section. We assume that there is a functionf(x, y) such that (9.19) ax, y,f(x, y)l = 0 for all (x, y) in some open set S, although we may not have
explicit formulas for calculatingf(x, y).

We describe this by saying that the equation F(x, y, z) = 0 defines z implicitly as a Derivatives of functions dejined implicitly 295 function of x and y, and we write z =f'(x,y>. Now we introduce an auxiliary function g defined on S as follows: g(x, Y> = F[x, y,f(x, .Y)l . Equation (9.19) states that g(x, y) = 0 on S; hence the partial derivatives ag/ax andi?
g/ay are also 0 on S.

But we can also compute these partial derivatives by the,chain rule.To do this we write g(x, Y> = F[Ul(X, y>, dx, y>, u,(x, VII > where u,(x, y) = x, u,(x, y) = y, and u&x, y) =f(x, y).

The chain rule gives us theformulas ag-= DlFa2 + D,Faz+ D,F?ax and ag- = D,& + D,Fa* + D,Fa3ay ay ay ay ' where each partial derivative D,F is to be evaluated at (x, y,J'(x, y)). Since we have au,- 1 au, o au3 af-=-ax ' XL' ' ax ax' and ag = 0ax 2 the first of the foregoing equations becomes DF+DFaf=O1 3ax *Solving this for aflax we obtain (9.20) af
D,F[x, y,f(x, ~11-=-ax D,F[x, Y, fk Y)l at those points at which D,F [x, y, f(x, y)] # 0. By a similar argument we obtain a corre-sponding formula for aflay: (9.21) !x- wk YYfbf Y)l- -ay D,FbG Ytf(X, Y)l at those points at which D,F[x, y,f(x, y)] # 0. These formulas are usually written morebriefly as follows: af aFlax af aFlay-=--ax aqaz G ---=z* EXAMPLE.
Assume that the equation y2 + xz + z2 - ez - c = 0 defines z as a functionof x and y, say z =f(x, y).

Find a value of the constant c such that ,f(O, e) = 2, andcompute the partial derivatives af/jax and @/lay at the point (x, y) = (0, e) . 296 Applications of difSerential calculus Solution. Whenx=O,y=e,andz=2,theequationbecomese2+4-e2-c=0,and this is satisfied by c = 4. Let F(X, y, z) = y2 + xz + z2 - eZ - 4. From (9.20)and (9.21) we have af Z af 2Y-=-
ax x + 22 - e" -=-ay x + 22 - ez ' When x = 0, y = e, and z = 2 we find aflax = 2/(e2 - 4) and af/jay = 2e/(e2 - 4).Note that we were able to compute the partial derivatives afiax and af;iay using only thevalue off(x, y) at the single point (0, e). The foregoing discussion can be extended to functions of more than two variables. THEOREM 9.3. Let F be a
scalarjeld difSerentiable on an open set Tin R". Assume thatthe equation F(x,, ...
, x,) = 0 dejines x, implicitly as a function of x1, . . . , x, ~, say x, =f(x,, * *. 3 x, J, for allpoints (x1, . . . , x,_~) in some open set S in R"-I. Then for each k = 1,2, . . . , n - 1,the partial derivative Dlcf is given by the formula (9.22) Dkf= -E71 at those points at which D,F # 0. The partial derivatives D,F and D,F which appear in(9.22) are to
be evaluated at the point (x1, x2, . . . , x,-~, f(x, , .
. . , x, J). The proof is a direct extension of the argument used to derive Equations (9.20) and(9.21) and is left to the reader. The discussion can be generalized in another way.

Suppose we have two surfaces withthe following implicit representations : (9.23) F(x, y, z) = 0, G(x,y,z) = o . If these surfaces intersect along a curve C, it may be possible to obtain a parametricrepresentation of C by solving the two equations in (9.23) simultaneously for two of thevariables in terms of the third, say for x and y in terms of z. Let us
suppose that it ispossible to solve for x and y and that solutions are given by the equations x = X(z), Y = Y(z) for all z in some open interval (a, b). Then when x and y are replaced by X(z) and Y(z),respectively, the two equations in (9.23) are identically satisfied. That is, we can write Derivatives of functions dejned implicitly 297 F[X(z), Y(z), z] = 0 and
G[X(z), Y(z), z] = 0 for all z in (a, b).

Again, by using thechain rule, we can compute the derivatives X'(z) and Y'(z) without an explicit knowledge ofX(z) and Y(z). To do this we introduce new functionsf and g by means of the equations f(z) = FW(z), Y(z), zl and g(z) = G [X(z), Y(z), z]. Then f(z) = g(z) = 0 for every z in (a, b) and hence the derivatives f'(z) and g'(z) arealso zero on (a, b). By
the chain rule these derivatives are given by the formula f'(z) = fz X'(z) + E Y'(z) + z)ax ay aZ g'(z) = E X'(z) + E Y'(z) + a; . Sincef'(z) and g'(z) are both zero we can determine X'(z) and Y'(z) by solving the follow-ing pair of simultaneous linear equations: aF-&X'(z) + E Y'(z) = - g) , g X'(z) + g Y'(z) = - g . At those points at which the determinant of the
system is not zero, these equations have aunique solution which can be expressed as follows, using Cramer's rule: (9.24) X'(z) = - aF aF zay ac ac-aZ ay aF aF- -ax ayaG ac- -ax ay aF aF- -ax aZaG aG- -ax aZ 2 Y'(z) = -aF aF-Z ayaG aG- -ax ay -.
The determinants which appear in (9.24) are determinants of Jacobian matrices and arecalled Jacobian determinants. A special notation is often used to denote Jacobian deter-minants. We write 3, af,- / - . . .ax, ax, 298 Applications of diflerential calculus In this notation, the formulas in (9.24) can be expressed more briefly in the form (9.25) (The minus
sign has been incorporated into the numerators by interchanging the columns.)The method can be extended to treat more general situations in which m equations in n variables are given, where n > m and we solve for m of the variables in terms of theremaining n - m variables. The partial derivatives of the new functions so defined can beexpressed
as quotients of Jacobian determinants, generalizing (9.25). An example withm = 2 and n = 4 is described in Exercise 3 of Section 9.8. 9.7 Worked examples In this section we illustrate some of the concepts of the foregoing section by solvingvarious types of problems dealing with functions defined implicitly. EXAMPLE 1. Assume that the equation g(x,
y) = 0 determines y as a differentiablefunction of x, say y = Y(x) for all x in some open interval (a, b). Express the derivativeY'(x) in terms of the partial derivatives of g. Solution. Let G(x) = g[x, Y(x)] for x in (a, b).
Then the equation g(x, y) = 0 impliesG(x) = 0 in (a, b). By the chain rule we have from which we obtain G'(x) = $a 1 + %' Y'(x),ay (9.26) at those points x in (a, b) at which ag/ay # 0. The partial derivatives ag/ax and ag/ayare given by the formulas ag/ax = D,g[x, Y(x)] and ag/ay = D,g[x, Y(x)]. EXAMPLE 2. When y is eliminated from the two equations z
=f(x, y) and g(x, y) = 0,the result can be expressed in the form z = h(x). Express the derivative h'(x) in terms ofthe partial derivatives off and g. Solution. Let us assume that the equation g(x, y) = 0 may be solved for y in terms ofx and that a solution is given by y = Y(x) for all x in some open interval (a, b).
Thenthe function h is given by the formula h(x) =fb, Y(x)1 if x E (a, b) . Applying the chain rule we have a-h'(x) = E + ay Y'(x). Worked examples 299 Using Equation (9.26) of Example 1 we obtain the formula wf af ag- - _ - - h'(x) =ay ax ay ax as .ay The partial derivatives on the right are to be evaluated at the point (x, Y(x)). Note thatthe numerator can
also be expressed as a Jacobian determinant, giving us h'(x) = a(fT dia(x3 Y)aday . EXAMPLE 3. The two equations 2x = v2 - u2 and y = MU define u and ~1 as functions0f x and y.
Find formulas for au/ax, au/ay, au/ax, av/ay.

Solution. If we hold y fixed and differentiate the two equations in question with respectto x, remembering that u and v are functions of x and y, we obtain ~=&k-&?ax ax and Solving these simultaneously for au/ax and au/ax we find all u-=--2 + uz and - = 2auax ax u2 + v2' On the other hand, if we hold x fixed and differentiate the two given equations
with respectto y we obtain the equations 04$-&!!ay anday -~=uLv,vauay ay * Solving these simultaneously we find au vand - = 2au-== ay u2 + v2 ay u2 + 2' EXAMPLE 4. Let u be defined as a function of x and y by means of the equation u = F(x + u,yu). Find au/ax and au/ay in terms of the partial derivatives of F. Solution. Suppose that u = g(x, y) for
all (x, y) in some open set S. Substitutingg(x, y) for u in the original equation we must have (9.27) g(x,u> = F[u,(x,y), u26, y)]l, 300 Applications of dlrerential calculus where u,(x, y) = x + g(x, y) and u,(x, y) = y g(x, y) . Now we hold y fixed and differ-entiate both sides of (9.27) with respect to x, using the chain rule on the right, to obtain (9.28)
ag-=.,au,+.,au,8X l ax 2 ax' But &,/ax = 1 + ag/ax, and &,/ax = y ag/ax. Hence (9.28) becomes Solving this equation for ag/ax (and writing au/ax for ag/ax) we obtain In a similar way we find au -DD,F-=ax D,F + y D,F - 1 '_= D,F au, + D,F f!!Z = D,F ag + D,Fag .ay ay ay ay y ' + g(x, Y>ay This leads to the equationau -dx, Y> Way= D,F + y D,F - 1 ' The
partial derivatives D,F and D,F are to be evaluated at the point (x + g(x, y), y g(x, y)). EXAMPLE 5. When u is eliminated from the two equations x = u + v and y = uv2, weget an equation of the form F(x, y, v) = 0 which defines v implicitly as a function of xand y, say v = h(x, y) . Prove that ah h(x, Y)-=ax 3h(x, y) - 2x and find a similar formula for ahlay.
Solution. Eliminating u from the two given equations, we obtain the relation xv2 - v3 -y=o. Let F be the function defined by the equation F(x, y, 4 = xv2 - v3 - y . The discussion in Section 9.6 is now applicable and we can write (9.29)ah aFlax-=--ax aqav and ah amy- = - -ay aqav * Worked examp Ies 301 ~ ~ ~ aFlax = 9, aFlav = 2xv - 3v2, and
 = -1. Hence the equations in (9.29)become ah V2 11 h(x, Y) =- -= ax 2xv - 3v2 = 2x - 3v 3h(x, y) - 2xand ah - 1 1-= ay 2xv - 3v2 = 2xh(x, y) - 3h2(x, y) ' EXAMPLE 6. The equation F(x, y, z) = 0 defines z implicitly as a function of x and y,say z =f(x,y). Assuming that a2r;l(ax az) = a2F/(az ax), show that (9.30)ay- (5) (Zj - 2(iZ) (IF) (3 +
(iFi'- = -ax2 aF 3 i 1 2 z where the partial derivatives on the right are to be evaluated at (x, y,f(x, y)). Solution. By Equation (9.20) of Section 9.6 we have (9.31) af aFlax a x - - - i @ % * We must remember that this quotient really means _W'b, y,fk Y)ID,Fb, Y?f(xYYN' Let us introduce G(x,y) = D,F[x,y,f(x,y)] and H(x,y) = D,F[x,y,f(x,y)]. Ourobject is to
evaluate the partial derivative with respect to x of the quotient af G(x, Y)=--ax WG Y) ' holding y fixed. The rule for differentiating quotients gives us -_aG -aH (9.32) azf H--G-ax ax=-- ax2 H2 * Since G and H are composite functions, we use the chain rule to compute the partialderivatives aG/ax and aH/ax. For aG/&x we have a2F + a2F aj=- - -ax2 az
ax ax * 302 Similarly, we find Applications of d$erential calculus aH af-= D,(D,F) . 1 + D,(D,F) * 0 + D,(D,F) . -gax ?F a2F af=ax+ +aZ::a.X Substituting these in (9.32) and replacing aflax by the quotient in (9.31) we obtain theformula in (9.30). 9.8 Exercises In the exercises in this section you may assume the existence and continuity of all
derivativesunder consideration. 1. The two equations x + y = uv and xy = u - v determine x and y implicitly as functions ofu and v, say x = X(u, v) and y = Y(u, v). Show that ax/au = (XV - 1)/(x - y) if x # y ,and find similar formulas for ax/ au, 8 Y/au, 8 Y/au. 2. The two equations x + y = uv and xy = u - v determine x and v as functions of u and y,say x =
X(u, y) and v = V(u, y). Show that ax/au = (u + v)/(l + yu) if 1 + yu # 0, andfind similar formulas for ax/ ay, aV/ au, 8 V/ ay. 3. The two equations F(x, y, u, v) = 0 and G(x, y, u, v) = 0 determine x and y implicitly asfunctions of u and v, say x = X(u, v) and y = Y(u, v). Show that ax a(& C>/ a& U)-=au a@', G)/ a@, Y) at points at which the Jacobian a(F,
G)/a(x, y) # 0, and find similar formulas for the partialderivatives ax/ au, a Y/au, and 8 Y/au. 4. The intersection of the two surfaces given by the Cartesian equations 2x' + 3y2 - z2 = 25and x2 +y 2 = z2 contains a curve C passing through the point P = (47, 3,4). Theseequations may be solved for x and y in terms of z to give a parametric representation
of Cwith z as parameter.(a) Find a unit tangent vector T to C at the point] P without usingtan explicit knowledge ofthe parametric representation.(b) Check the result in part (a) by determining a parametric representation of C with z asparameter. 5. The three equations F(u, v) = 0, u = xy, and v = dm define a surface in xyz-space.Find a normal vector
to this surface at the point x = 1, y = 1, z = fi if it is known thatD,F(l, 2) = 1 and D,F(l ,2) = 2. 6. The three equationsx2-ycos(uv) +22=0, x2 + y2 - sin (uv) + 2z2 = 2, xy -sinucosv +z =O, define x, y, and z as functions of u and v. Compute the partial derivatives ax/ au and ax/au atthepointx =y = 1,~ = ~12,~ =O,z =O. Maxima, minima, and saddle
points 303 7. The equationf(y/x, z/x) = 0 defines z implicitly as a function of x and y, say z = g(x, y).Show that xg +y% =g(x,y)aY at those points at which DPf[y/x, g(x, y)/x] is not zero.8. Let F be a real-valued function of two real variables and assume that the partial derivatives D,F and D,F are never zero. Let u be another real-valued function of two
real variables suchthat the partial derivatives au/ ax and &A/ 8y are related by the equation F(au/ ax, au/ ay) = 0.Prove that a constant n exists such that a% a% a% *--=ax2 ay2 (1ms \ and find n. Assume that a2u/(ax ay) = a2u/(ay ax).9. The equation x + z + (y + z)~ = 6 defines z implicitly as a function of x and y, say z =f(x, y) . Compute the partial
derivatives af/lx, af/lay, and a2f(ax ay) in terms of x,y, and z. 10 . 11. 12. The equation sin (x + y) + sin (y + z) = 1 defines z implicitly as a function of x and y, sayz =f(x, y) . Compute the second derivative D,,,fin terms of x, y, and z.The equation F(x + y + z, x2 + y2 + z2) = 0 defines z implicitly as a function of x and y,say z =f(x, y) . Determine the
partial derivatives af/ ax and af/ ay in terms of x, y, z and thepartial derivatives DIF and D,F.Letfandg be functions of one real variable and define F(x, y) = f[x + g(y)]. Find formulasfor all the partial derivatives of Fof first and second order, expressed in terms of the derivativesoff and g. Verify the relation aF a2F aF a2F- -ax ax ay = 6 axi * 9.9 Maxima,
minima, and saddle points A surface that is described explicitly by an equation of the form z =f(x, y) can bethought of as a level surface of the scalar field F defined by the equation F(x, y, 4 = f(x, y>- - z . Iffis differentiable, the gradient of this field is given by the vector vF=Ei+afj-k.ay A linear equation for the tangent plane at a point P1 = (x1, y, , zr)
can be written in the form where2 - z1 = A(x - x1) + B(y - y1) 7 A = Rf(x19yd and B = D,f(x, 3 rd. When both coefficients A and B are zero, the point PI is called a stationary point of thesurface and the point (x1, yJ is called a stationarypoint or a criticalpoint of the function f. 304 Applications of differential calculus The tangent plane is horizontal at a
stationary point. The stationary points of a surface areusually classified into three categories : maxima, minima, and saddle points. If the surfaceis thought of as a mountain landscape, these categories correspond, respectively, tomountain tops, bottoms of valleys, and mountain passes.
The concepts of maxima, minima, and saddle points can be introduced for arbitrary scalarfields defined on subsets of R". DEFINITION. A scalarfieldf is said to have an absolute maximum at a point a of a set Sin R" if (9.33) for allx in S. The number f (a) is called the absolute maximum value off on S. The function fis said to have a relative maximum at a
if the inequality in (9.33) is satisjiedfor every x insome n-ball B(a) lying in S. In other words, a relative maximum at a is the absolute maximum in some neighborhoodof a. The terms absolute minimum and relative minimum are defined in an analogousfashion, using the inequality opposite to that in (9.33). The adjectives global and local aresometimes
used in place of absolute and relative, respectively. DEFINITION. A number which is either a relative maximum or a relative minimum off iscalled an extremum off. Iff has an extremum at an interior point a and is differentiable there, then all first-orderpartial derivatives D&a), . . . , D&a) must be zero. In other words, Vf(a) = 0.
(Thisis easily proved by holding each component fixed and reducing the problem to the one-dimensional case.) In the case n = 2, this means that there is a horizontal tangent planeto the surface z = f(x, y) at the point (a, f (a)). On the other hand, it is easy to find examplesin which the vanishing of all partial derivatives at a does not necessarily imply an
extremumat a. This occurs at the so-called saddlepoints which are defined as follows. DEFINITION. Assume f is differentiable at a. If Of (a) = 0 the point a is called a stationarypoint of J A stationary point is called a saddle point if every n-ball B(a) contains points xsuch that f (x) < f(a) and other points such that f (x) > f (a).
The situation is somewhat analogous to the one-dimensional case in which stationarypoints of a function are classified as maxima, minima, and points of inflection. Thefollowing examples illustrate several types of stationary points.
In each case the stationarypoint in question is at the origin.

EXAMPLE 1. Relative maximum. z = f (x, y) = 2 - x2 - ~2. This surface is a parabo loid of revolution. In the vicinity of the origin it has the shape shown in Figure 9,3(a). Its Maxima, minima, and saddle poirlts 305 Z Y I bX ,I---b, /I XJ (a) z = 2 - x2 - y* (b) Level curves: x' + y' = c Example 1. Relative maximum at the origin. XJ (c) z = x" + ,!J2 Example 2.
Relative minimum at the origin.

FIGURE 9.3 Examples 1 and 2. level curves are circles, some of which are shown in Figure 9.3(b). Since f(x, y) =2 - (x2 + y") 5 2 = f(0, 0) for all (x, y), it follows thatfnot only has a relative maximumat (0,O) but also an absolute maximum there. Both partial derivatives af/ax and aflayvanish at the origin. EXAMPLE 2. Relative minimum. z = f(x, y) = x2
+ y2. This example, another parabo loid of revolution, is essentially the same as Example 1, except that there is a minimum atthe origin rather than a maximum. The appearance of the surface near the origin isillustrated in Figure 9.3(c) and some of the level curves are shown in Figure 9.3(b). 306 Applications of difSerential calculus EXAMPLE 3.
Saddle point. z = f(x, y) = xy. This surface is a hyperbolic paraboloid.Near the origin the surface is saddle shaped, as shown in Figure 9.4(a). Both partialderivatives aflax and aflay are zero at the origin but there is neither a relative maximumnor a relative minimum there. In fact, for points (x, y) in the first or third quadrants,x and y have the same
sign, giving us f(x, y) > 0 = f(0, 0) , whereas for points in thesecond and fourth quadrants x and y have opposite signs, giving us f (x, y) < 0 = f (0,O).Therefore, in every neighborhood of the origin there are points at which the function isless thanf(O, 0) and points at which the function exceedsf(0, 0), so the origin is a saddle (a) z = XL' (b) Level curves:
xy = c FIGURE 9.4 Example 3. Saddle point at the origin. point. The presence of the saddle point is also revealed by Figure 9.4(b), which showssome of the level curves near (0,O). These are hyperbolas having the x- and y-axes asasymptotes. EXAMPLE 4. Saddle point. z = f(.x, y) = x3 - 3xy2. Near the origin, this surface hasthe appearance of a
mountain pass in the vicinity of three peaks. This surface, sometimesreferred to as a "monkey saddle," is shown in Figure 9.5(a). Some of the level curvesare illustrated in Figure 9.5(b). It is clear that there is a saddle point at the origin. EXAMPLE 5. Relative minimum. z = f(x, y) = 2x y2. This surface has the appearance ofa valley surrounded by four
mountains, as suggested by Figure 9.6(a). There is an absoluteminimum at the origin, sincef(x, y) >f(O, 0) for all (x, y). The level curves [shown inFigure 9.6(b)] are hyperbolas having the x- and y-axes as asymptotes. Note that these levelcurves are similar to those in Example 3. In this case, however, the function assumes onlynonnegative values on all
its level curves. EXAMPLE 6. Relative maximum. z =f(x, y) = 1 - x2. In this case the surface is acylinder with generators parallel to the y-axis, as shown in Figure 9.7(a). Cross sectionscut by planes parallel to the x-axis are parabolas. There is obviously an absolute maximumat the origin becausef(x, y) = 1 - x2 < 1 = f(0, 0) for all (x, y). The level curves
form afamily of parallel straight lines as shown in Figure 9.7(b). (a) ; zz .Y.' - 3 Xj". (b) Level curves: s.' - 3 q" = c. FIGURE 9.5 Example 4. Saddle point at the origin. Ji (a) z = Yy' (b) Level curves: x'y' = c FIGURE 9.6 Example 5. Relative minimum at the origin. ; ?
Tangent plane at (O,O,l) / (a) i = 1 - xz (b) Level curves: 1 - x' = c FIGURE 9.7 Example 6. Relative maximum at the origin. 307 308 Applications of d$erentiaI calculus 9.10 Second-order Taylor formula for scalar fields If a differentiable scalar field f has a stationary point at a, the nature of the stationarypoint is determined by the algebraic sign of the
difference f'(x) -f(a) for x near a. Ifx = CI + y , we have the first-order Taylor formula f(a + Y) - f(a) = Vf(a) * Y + IIY II J%& Y> > where E(a,y)+O as y+ 0. At a stationary point, Vf(u) = 0 and the Taylor formula becomes f(a + Y) -f(u) = IIY II -m Y) * To determine the algebraic sign of f(u + y) - f(u we need more information about the)error term IIy 11
E(u, y). The next theorem shows that if f has continuous second-orderpartial derivatives at a, the error term is equal to a quadratic form, plus a term of smaller order than IIy \12. The coefficients of the quadratic form are thesecond-order partial derivatives Diif = Di(Df) , evaluated at a. The n x n matrix ofsecond-order derivatives Dijf(x) is called the
Hessian matrix? and is denoted by H(x).Thus, we have whenever the derivatives exist.

The quadratic form can be written more simply in matrixnotation as follows: where y = (yl, . . . ,u,J is considered as a 1 x n row matrix, and yt is its transpose, ann x 1 column matrix. When the partial derivatives Dijf are continuous we have Dijf =Djif and the matrix H(u) is symmetric. Taylor's formula, giving a quadratic approximation to f(u + y) - f(u),
now takes thefollowing form. THEOREM 9.4. SECOND-ORDER TAYLOR FORMULA FOR SCALAR FIELDS. Let f be a scalarjield with continuous second-order partial derivatives Dijf in an n-ball B(u). Then for ally in R" such that a + y E B(u) we have (9.34) f(a + y) -f(a) = v 64 'Y + $YfJ(U + CY>Y", where OHere we have used Lagrange's form of the

remainder (see Section 7.7 of Volume I).Since g is a composite function given by g(u) =f[v(u)] , where r(u) = u + uy, we can compute its derivative by the chain rule. We have r'(u) = y so the chain rule gives us g'(u) = Vf[r(u>l . r'(u) = Vf[r(u)l *Y =jiiI~jflr(uNui9 provided r(u) E B(u). In particular, g'(0) = Of(u) * y .
Using the chain rule once morewe find g"(u) = 2 Di i Djf[r(u>lvii 1 Yi = 5 i Rif[r(u)lwj = yfWu)lyt.i=l j=l i-1 j=l Hence g"(c) = yH(u + cy)yt, so Equation (9.36) becomes (9.34).To prove (9.35) we define E,(u, y) by the equation (9.37) llyl12 E,&Y) = ~YW(U + CY) - W4)yt i f y#O, and let E,(u, 0) = 0. Then Equation (9.34) takes the form f(u +y) -f(u) =
Vf(u)·y + $ yH(u)y" 4 (lyl12 E,(~,Y). To complete the proof we need to show that E2(u, y) -+ 0 as y -+ 0.From (9.37) we find that llyl12 IE,@, y)l = ; x 3 {DJ(u + cy) - Dtjf(u>} YiYj2=19=1 310 Applications of calculus Dividing by l\y (I2 we obtain the inequality I&k Y)I I 5 i' $lDiJ(u + cy) - DiJ(u>l2=13=1 for y # 0. Since each second-order partial derivative Dijf is continuous at a, we have D&a + CY> - D&a> as Y -+ 0, so E,(a, y) -+ 0 as y + 0. This completes the proof. 9.11 The nature of a stationary point determined by the eigenvalues of the Hessian matrix At a stationary point we have V'(a) = 0, so the Taylor formula in Equation (9.35)becomes f(a + Y) -f(a) = +YH(4Yt + IIY II2 -
%(a, Y> * Since the error term ((y/j2 E,(a, y) tends to zero faster than lIylj2, it seems reasonable toexpect that for small y the algebraic sign of f(a + y) -f(a) is the same as that of thequadratic form yH(a)yt; hence the nature of the stationary point should be determined bythe algebraic sign of the quadratic form. This section is devoted to a proof of this fact. First we give a connection between the algebraic sign of a quadratic form and its eigen-values. THEOREM 9.5. Let A = [aij] be an n x n realsymmetric matrix, and let Then we have: Q(Y) = y 4~ ~ = 2 i ww j.i=l j=l (a) Q(y) > 0 for ally # 0 ifand onb ifall the eigenvalues of A are positive.(b) Q(y) < 0 for ally # 0 if and only ifall the eigenvalues of A are negative. Note: In case (a), the quadratic form is calledpositive definite; in case (b) it is callednegative definite.

Proof. According to Theorem 5.11 there is an orthogonal matrix C that reduces thequadratic form yAyt to a diagonal form. That is (9.38) Q(y) = YAY~ = &: wherex=(x,,..., x,) is the row matrix x = yC, and 1,) . . . , 1, are the eigenvalues ofA. The eigenvalues are real since A is symmetric. If all the eigenvalues are positive, Equation (9.38) shows that Q(y) > 0 wheneverx # 0. But since x = yCwe have y = xc-l, so x # 0 if and only if y # 0. ThereforeQ(y)>Oforally# 0 . Conversely, if Q(y) > 0 for all y # 0 we can choose y so that x = yC is the kth co-ordinate vector ek . For this y, Equation (9.38) gives us Q(y) = &, so each L, > 0. Thisproves part (a). The proof of (b) is entirely analogous. Nature of a stationary point determined by eigenvalues of Hessian matrix 311 The next theorem describes the nature of a stationary point in terms of the algebraicsign of the quadratic form yH(a)$. THEOREM 9.6. Let f be a scalar field with continuous second-order partial derivativesDijf in an n-ball B(a), and let H(a) denote the Hessian matrix at a stationary point a. Thenwe have: (a) Ifall the eigenvalues of H(a) are positive, f has a relative minimum at a.(b) If all the eigenvalues of H(a) are negative, f has a relative maximum at a.(c) If H(a) has both positive and negative eigenvalues, then f has a saddle point at a. Proof.

Let Q(y) = yH(u)yt. The Taylor formula gives us (9.39) f@ + Y) - f (4 = BQ(Y) + IIY II2 E,(a, Y), where E,(a, y) + 0 as y - f 0. We will prove that there is a positive number Y such that, if0 < llyll < r, the algebraic sign off(a + y) -f(u) is th[e same as that of Q(y). Assume first that all the eigenvalues A1, . . . , 1, of H(a) are positive. Let h be thesmallest eigenvalue. If u < h, the n numbers are also positive. These numbers are the eigenvalues of the real symmetric matrix H(a) -uI, where I is the n x n identity matrix. By Thseorem 9.5 , the quadrat ic formy[H(a) - u1Jy" is positive definite, and hence y[H(a) - uZJyt > 0 for all y # 0. There-fore yH(4yt > YW)Y" = u lbl12 for all real u < h . Taking u = frh we obtain the inequality Q(Y) > $3 IIYII~ for all y # 0.

Since E,(a, JJ) --f 0 as y + 0, there is a positive number r such thatIE2(a, y)I < ah whenever 0 < IIyII < r. For such y we have 0 I Ilull" IEz(a, Y)I < $h Ilyll" <: &Q(y), and Taylor's formula (9.39) shows that f (a + Y> - f (4 2 &Q(Y) - I IY I I 2 IE2k y)l > 0 - Therefore f has a relative minimum at a, which proves part (a). To prove (b) we can use asimilar argument, or simply apply part (a) to -f. To prove (c), let I, and L, be two eigenvalues of H(u) of opposite signs. Let h =min {l&l, l&l} . Then for each real u satisfying -h < u ,< h the numbers 2, - 24 and /I, - u 312 Applications of differential calculus are eigenvalues of opposite sign for the matrix H(a) - ~1. Therefore, if u E (--h, h) , thequadratic form y[H(a) - uZ]yt takes both positive and negative values in every neighbor-hood of y = 0.

Choose r > 0 as above so that]&(a, y)(< fh whenever 0 < (lyll < r.Then, arguing as above, we see that for such y the algebraic sign off(a + y) -f(a) is thesame as that of Q(y). Since both positive and negative values occur as y -+ 0, f has asaddle point at u. This completes the proof. Note: If all the eigenvalues of H(a) are zero, Theorem 9.6 gives no informationconcerning the stationary point. Tests involving higher order derivatives can be used totreat such examples, but we shall not discuss them here. 9.12 Second-derivative test for extrema of functions of two variables In the case n = 2 the nature of the stationary point can also be determined by thealgebraic sign of the second derivative D,,,f(u) and the determinant of the Hessian matrix. THEOREM 9.7. Let a be a stationary point of a scalarjeldf (x,, XJ with continuous second-order partial derivatives in a 2-ball B(u). Let and let A = det H(u) = det = A C - B2. Then we have:(a) If A < 0, f has a saddle point at u.(b) If A > 0 and A > 0, f has a relative minimum at a.(c) If A > 0 and A < 0, f has a relative maximum at a.(d) If A = 0, the test is inconclusive. Proof. In this case the characteristic equation det [A1 - H(u)] = 0 is a quadraticequation, The eigenvalues iI, 1, are related to the coefficients by the equations A,+&=A+C, A,& = A. If A < 0 the eigenvalues have opposite signs, so f has a saddle point at a, which proves (a).If A > 0, the eigenvalues have the same sign. In this case AC > B2 > 0, so A and C havethe same sign. This sign must be that of J., and il, since A + C = A1 + 1,. This proves(b) and (c). To prove (d) we refer to Examples 4 and 5 of Section 9.9. In both these examples we haveA = 0 at the origin. In Example 4 the origin is a saddle point, and in Example 5 it is arelative minimum. Exercises 313 Even when Theorem 9.7 is applicable it may not be the simplest way to determine thenature of a stationary point. For example, when f(x, y) = el'g(z*y), where g(x, y) =x2 + 2 + CO? y - 2 cos y , the test is applicable, but the computations are lengthy. In thiscase we may express g(x, y) as a sum of squares by writing g(x, y) = 1 + x2 + (1 - cos y)" .We see at once thatf has relative maxima at the points at which x2 = 0 and (1 - cos y)" =0.

These are the points (0,2nn), when n is any integer. 9.13 Exercises In Exercises 1 through 15, locate and classify the stationary points (if any) of the surfaces havingthe Cartesian equations given. 1.2 .3.4 .5 .6. 13. 14. 15. 16. 17. 18. z = x2 + 0, - 1)2. 7. z = x3 - 3xy2 + y3.z = x2 - (y - 1)2. 8. z = x2y3(6 '- x - y).z=l +x2--2. 9. z =x3 +y'l - 3xy.z = (x - y + 1)2. 10.

z = sin x cash y .z=2x2-xy-3y2-3x+7y. 11. z = e2r+3y(8~2 - 6xy + 3y2).z=x2-xy+y2-2x+y.

12. z = (5x + '7y - 25)e-(2a+zy+y*).z =sinxsinysin(x +y), OIxIn,OIyo. z = (x2 + y2)e-(z2+t!*). Letf(x, y) = 3x4 - 4x2y + y2. Show that on every line y =: mx the function has a minimumat (0, 0), but that there is no relative minimum in any two-dimensional neighborhood of theorigin. Make a sketch indicating the set of points (x, y) at whichf(x, y) > 0 and the set atwhichf(x, y) < 0.Letf(x, y) = (3 --x)(3 - y)(x + y - 3).(a) Make a sketch indicating the set of points (x, y) at which f (x, y) 2 0.(b) Find all points (x, y) in the plane at which DJ (x, y) = D2f (x, y) = 0. [H in t : D&x, y)has (3 - y) as a factor.](c) Which of the stationary points are relative maxima.7 Which are relative minima? Whichare neither? Give reasons for your answers.(d) Does f have an absolute minimum or an absolute maximum on the whole plane? Givereasons for your answers.Determine all the relative and absolute extreme values and the saddle points for the functionf(x7y)=xy(l-x2-_ _ _ _ _y2)onthesquare0 Let f(x, y) = Ax2 + 2Bxy + Cy2 + 2Dx + 2Ey + F, where A > 0 and B2 < AC. (a) Prove that a point (xX, yl) exists at which f has a minimum. [Hint: Transform the quad-ratic part to a sum of squares.](b) Prove that f(x,,y,) = Dxl + Ey, + Fat this minimum.(c) Show that IA B DI f(x1vy1) = AC '- B2 /B C El. 314 Applications of calculus 21. Method of least squares. Given n distinct numbers x1, . . . , x, and n further numbers yl, . . . ,yn (not necessarily distinct), it is generally impossible to find a straight linef(x) = ax + bwhich passes through all the points (xi, yi), that is, such thatf(x,) = yi for each i. However,we can try a linear function which makes the "total square error" a minimum. Determine values of a and 6 which do this.22.

Extend the method of least squares to 3-space. That is, find a linear functionf(x,y) = ax + by + c which minimizes the total square error Q, 6, c) = 2 if&, yi) - zi12,i=l where (xi, yi) are IZ given distinct points and zl, . . . , z, are 72 given real numbers.23. Letzl,..., zn be n distinct points in m-space. If x E R" , define f(X) =k$l lb - Zkl12. Prove thatfhas a minimum at the point a = i12 czli (the centroid). k=l 24. Let a be a stationary point of a scalar fieldfwith continuous second-order partial derivativesin an n-ball B(a). Prove thatfhas a saddle point at a if at least two of the diagonal entriesof the Hessian matrix H(a) have opposite signs. 25. Verify that the scalar fieldf(x, y, z) = x4 + y4 + z4 - 4xyz has a stationary point at (1, 1, l),and determine the nature of this stationary point by computing the eigenvalues of its Hessianmatrix. 9.14 Extrema with constraints. Lagrange's multipliers We begin this section with two examples of extremum problems with constraints. EXAMPLE 1. Given a surface S not passing through the origin, determine those points ofS which are nearest to the origin. EXAMPLE 2. Iff(x,y, z) denotes the temperature at (x, y, z), determine the maximumand minimum values of the temperature on a given curve C in 3-space. Both these examples are special cases of the following general problem: Determine theextreme values of a scalarJieldf ( )x w h en x is restricted to lie in a given subset of the domainoff. In Example 1 the scalar field to be minimized is the distance function, f(x, y, 2) = (x2 + y2 + + ; Extrema with constraints. Lagrange's jmultipliers 315 the constraining subset is the given surface S. In Example 2 the constraining subset is thegiven curve C. Constrained extremum problems are often very difficult; no general method is knownfor attacking them in their fullest generality. Special methods are available when theconstraining subset has a fairly simple structure, for instance, if it is a surface as in Example1, or a curve as in Example 2. This section discusses the method of Lagrange's multipliersfor solving such problems. First we describe the method in its general form, and then wegive geometric arguments to show why it works in the two examples mentioned above. The method of Lagrange's multipliers. If a scalar field f(xl, . . .

, x,) has a relativeextremum when it is subject to m constraints, say (9.40) glh,...,&)=O, . . . , g,(~l,....,x,)=o, where m < n , then there exist m scalars 1,) . . . , 1, such that (9.41) Vf=il,Vg,+ -+A,Vg, at each extremum point. To determine the extremum points in practice we consider the system of n + m equationsobtained by taking the m constraint equations in (9.40) allong with the n scalar equationsdetermined by the vector relation (9.41). These equations are to be solved (if possible) forthe n + m unknowns xi, . . . , x, and 3r,, . . . , 1,. The points (x1, . . . , x,) at whichrelative extrema occur are found among the solutions to these equations.

The scalars Jr, . . . , 1, which are introduced to help us solve this type of problem arecalled Lagrange's multipliers. One multiplier is introduced for each constraint. The scalarfield f and the constraint functions gl, . . . , g, are assumed to be differentiable. Themethod is valid if the number of constraints, m, is less than the number of variables, n, andif not all the Jacobian determinants of the constraint functions with respect to m of thevariables x1, . .

., x, are zero at the extreme value in question. The proof of the validity ofthe method is an important result in advanced calculus and will not be discussed here.(See Chapter 7 of the author's Mathematical Analysis, Adldison-Wesley, Reading, Mass.,1957.) Instead we give geometric arguments to show why the method works in the twoexamples described at the beginning of this section. Geometric solution of Example 1. We wish to determine: those points on a given surfaceS which are nearest to the origin. A point (x, y, z) in 3-space lies at a distance r from theorigin if and only if it lies on the sphere x2 + y2 + z2 = r2. This sphere is a level surface of the function f (x, y, z) = (x2 + y2 + z2)lm which is beingminimized.

If we start with r = 0 and let r increase until the corresponding level surfacefirst touches the given surface S, each point of contact will be a point of S nearest to theorigin. 316 Applications of calculus To determine the coordinates of the contact points we assume that S is described by aCartesian equation g(x, y, z) = 0. If S has a tangent plane at a point of contact, this planemust also be tangent to the contacting level surface. Therefore the gradient vector of thesurface g(x, y, z) = 0 must be parallel to the gradient vector of the contacting level surfacef(x,y, 4 = I+. Hence there is a constant il such that at each contact point. This is the vector equation (9.41) provided by Lagrange's methodwhen there is one constraint. Geometric solution to Example 2. We seek the extreme values of a temperature functionf(x, y, z) on a given curve C. If we regard the curve C as the intersection of two surfaces, sayg,b, y, 4 = 0 and g,(x,y, 4 = 0, we have an extremum problem with two constraints. The two gradient vectors Vgl andVg, are normals to these surfaces, hence they are also normal to C, the curve of inter-section. (See Figure 9.8.) We show next that the gradient vector Of of the temperature FIGURE 9.8 The vectors Vgr, Vgz, and Vfshown lying in the same plane. FIGURE 9.9 The gradient vector Vf lies in aplane normal to C. function is also normal to C at each relative extremum on C. This implies that Vflies in thesame plane as Vgr and Vg,; hence if Vg, and Vg2 are independent we can express Vf as alinear combination of Vg, and Vg, , say Vf = a,vg, + A,Vg,. This is the vector equation (9.41) provided by Lagrange's method when there are twoconstraints. Extrema with constraints. Lagrange's multipliers 317 To show that Of is normal to C at an extremum we imagine C as being described by avector-valued function a(t), where t varies over an interval [a, 61. On the curve C thetemperature becomes a function oft, say q(t) = f [a(t)]. If fp has a relative extremum at aninterior point tl of [a, b] we must have cp'(tl) = 0. On the other hand, the chain rule tellsus that v'(t) is given by the dot product q'(t) = Of [a(t)] . a'(t). This dot product is zero at t, , hence Of is perpendicular to a'(tJ . But a'(tJ is tangent to C,so Of [a(t,)] lies in the plane normal to C, as shown in Figure 9.9. The two gradient vectors Vgl and Vgz are independent if a.nd only if their cross product isnonzero. The cross product is given by i i k vg1 x vg2 = &l f&l f& = %% 9 g2> i + 3% 3 g2) j + a(.& 9 g2) k .ax ay aZ ah 4 acz, 4 ah Y)"& ag, agzax ay aZ Therefore, independence of Vgl and VgZ means that not all three of the Jacobian deter-minants on the right are zero. As remarked earlier, Lagrange's method is applicablewhenever this condition is satisfied. If Vg, and Vg, are dependent the method may fail. For example, suppose we try to applyLagrange's method to find the extreme values of f(x,y, z) = x2 + y2 on the curve ofintersection of the two surfaces g,(x,y, z) = 0 and g,(x, J', Z) = 0, where g,(x, y, z) = zand g,(x, y, z) = z2 - (y - 1)". The two surfaces, a plane and a cylinder, intersect alongthe straight line C shown in Figure 9.10. The problem obviously has a solution, because FIGURE 9.10 An example where Lagrange's method is not applicable. 318 Applications of diferential calculus f(x, y, Z) represents the distance of the point (x, y, z) from the z-axis and this distance is aminimum on C when the point is at (0, 1, 0). However, at this point the gradient vectorsare Ygl = k , Vgs = 0, and Of = 2j, and it is clear that there are no scalars il, and 1, thatsatisfy Equation (9.41). 9.15 Exercises 1.

Find the extreme values of z = xy subject to the condition x + y = 1.7.. Find the maximum and minimum distances from the origin to the curve 5x2 + 6xy + 5y2 = 8.3. Assume a and b are fixed positive numbers.

(a) Find the extreme values of z = x/a + y/b subject to the condition x2 + y2 = 1 .(b) Find the extreme values of z = x2 + y2 subject to the condition x/a + y/b = 1 .In each case, interpret the problem geometrically. 4. Find the extreme values of z = cos2 x + cos2 y subject to the side condition x - y = n/4.5. Find the extreme values of the scalar fieldf(x, y, z) = x - 2y + 22 on the sphere x2 + y2 + z2 = 1.6. Find the points of the surface z2 - xy = 1 nearest to the origin.7. Find the shortest distance from the point (1,O) to the parabola y2 = 4x.8. Find the points on the curve of intersection of the two surfaces x2 - xy + y2 - 22 = 1 and x2 + y2 = 1 which are nearest to the origin.9. If n, b, and c are positive numbers, find the maximum value off (x, y, z) = xaybzc subject to the side condition x + y + z = 1 .10. Find the minimum volume bounded by the planes x = 0, y = 0, z = 0, and a plane which is tangent to the ellipsoid atapointintheoctantx>O,y>O,z>O.11. Find the maximum of log x + logy + 3 log z on that portion of the sphere x2 + y2 + ~2 = 5r2 where x > 0, y > 0, z > 0. Use the result to prove that for real positive numbers a, b, cw e h a v e abc3 <27(a +z +'r. 12. Given the conic section Ax2 + 2Bxy + Cy2 = 1, where A > 0 and B2 < AC. Let m and Mdenote the distances from the origin to the nearest and furthest points of the conic. Show that J42= A + C + &A - C)2 + 4B22(AC - B2) and find a companion formula for m2.13. Use the method of Lagrange's multipliers to find the greatest and least distances of a point on the ellipse x2 + 4y2 = 4 from the straight line x + y = 4.14. The cross section of a trough is an isosceles trapezoid. If the trough is made by bending up the sides of a strip of metal c inches wide, what should be the angle of inclination of the sidesand the width across the bottom if the cross-sectional area is to be a maximum? The extreme-value theorem for continuous scalar$elds 319 9.16 The extreme-value theorem for continuous scalar fields The extreme-value theorem for real-valued functions continuous on a closed and boundedinterval can be extended to scalar fields. We consider scalar fields continuous on a closedn-dimensional interval. Such an interval is defined as the Cartesian product of n one-dimensional closed intervals. If a = (a,, . . . , a,) and 6 = (b,, . . . , b,) we write b,bl=b,,Ux--x [a,,b,l={(x,,...,x,)Ix,E[a,,b,l,...,x,Eta,,b,l} . For example, when n = 2 the Cartesian product [a, b] is a rectangle.The proof of the extreme-value theorem parallels the proof given in Volume I for the l-dimensional case.

First we prove that continuity offimplies boundedness, then we provethatfactually attains its maximum and minimum values somewhere in [a, b]. THEOREM 9.8. BOUNDEDNESS THEOREM FOR CONTINUOUS SCALAR FIELDS. Zf fis a scaZarjield continuous at each point of a closed interval [a, b] in R", then f is bounded on [a, b].That is, there is a number C > 0 such that If(x)1 < Cfor all x in [a, b]. Proof. We argue by contradiction, using the method of successive bisection.

Figure9.11 illustrates the method for the case n = 2. Assumefis unbounded on [a, b]. Let 1") = [a, b] and let 12' = [a,, b,], so that p = I(l) x . . .

x p1 n . Bisect each one-dimensional interval Zk(l) to form two subintervals, a left half 1:': and a'right half Z,$. Now consider all possible Cartesian pr0duct.s of the form I;';, x . . . x p.n,3n ' FIGURE 9.11 Illustrating the method of successive bisection in the plane. 320 Applications of differential calculus where cachj, = I or 2. There are exactly 2" such products. Each product is an n-dimen-sional subinterval of [a, b], and their union is equal to [a, h]. The functionf'is unboundedin at least one of these subintervals (if it were bounded in each of them it would also bebounded on [a, b]).
One of these we denote by It2' which we express as where each I;,' is one of the one-dimensional subintervals of 1;") of length h(h, - al,).We now proceed with I"' as we did with I(", bisecting each one-dimensional component interval I,:,' and arriving at an n-dimensional interval 1t3' in which f' is unbounded. Wecontinue the process, obtaining an infinite set of n-dimensional intervals in each of whichj'is unbounded. The n'th interval Pm' can be expressed in the form Since each one-dimensional interval 1;"' is obtained by m - 1 successive bisections of[a,, /I,], if we write IL"" = [a:"', hj;'n'] we have (9.42) b(m) _ a(m) b, - akk k =-y-1 ' for k=l,2 ,..., n. For each fixed k, the supremum of all left endpoints u:."~' (m = 1,2, . .) must therefore beequal to the infimum of all right endpoints hkm' (m = 1, 2, . .

.), and their common valuewe denote by t,. The point t = (tl, . . . , t,) lies in [a, b]. By continuity off'at t there is ann-ball B(t, r) in which we have for all x in B(t; r) n [a, b] . This inequality implies IfWl < 1 + I./w for all x in B(t; r) f? [a, 61, so f'is bounded on the set B(t; r) n [a, b] . But this set contains the entire interval ZCm' when m is large enough so that each of the n numbers in (9.42) is less than r/x/n. Thereforefor such m the function f is bounded on Itm', contradicting the fact that f is unbounded onIfm'. This contradiction completes the proof. If f is bounded on [a, b], the set of all function values f'(x) is a set of real numbersbounded above and below. Therefore this set has a supremum and an infimum which wedenote by sup f and inff, respectively. That is, we write supf = sup (f(x) 1 x E [a, 61)) inff'= inf (f(x) 1 x t [u, b]). The small-span theorem for continuous sculurjelds (uniform continuit),) 3 2 1 Now we prove that a continuous function takes on both values inff'and supfsomewherein [a, 61. THEOREM 9.9 EXTREMFVALUE THEOREM FOR CONTINUOUS SCALAR FIELDS. Iff is con-tinuous on a closed interval [a, b] in R", then there exist points c and d in [a, b] such that f(c) = supf and ,f(d) = inff. Proof. It suffices to prove that f attains its supremum in [a, b] . The result for theinfimum then follows as a consequence because the infimum off is the supremum of -f. Let M = sup f. We shall assume that there is no x in [la, b] for which ,f(x) = M andobtain a contradiction. Let g(x) = M - .f'(x).

Then g(x) > 0 for all x in [a, b] so thereciprocal l/g is continuous on [a, b]. By the boundedness theorem, l/g is bounded on[a, b], say l/g(x) < C for all x in [a, b], where C > 0. This implies M -f(x) > l/C, sothatf(x) < A4 - l/C for all x in [a, b]. This contradicts the fact that M is the least upperbound off on [a, b]. Hence f(x) = M for at least one x in [a, b].
9.17 The small-span theorem for continuous scalar fields (uniform continuity) Let f be continuous on a bounded closed interval [a, b] in R", and let M(f) and m(f)denote, respectively, the maximum and minimum values off on [a, b]. The difference M(f) -m(f) is called the span off on [a, b]. As in the one-dimensional case we have a small-spantheorem for continuous functions which tells us that the interval [a, b] can be partitioned sothat the span off in each subinterval is arbitrarily small. Write [a, b] = [a,, b,] x . . . x [a,, 6,], and let P, be a partition of the interval[a,, bk].

That is, Pr, is a set of points such that ak = x0 < x1 The small-span theorem, also called the theoremon uniform continuity, now takes the following form.

THEOREM 9.10. Let f be a scalarfield continuous on a closed interval [a, b] in R". Thenfor every E > 0 there is a partition of [a, b] into afinite number of subintervals such that thespan off in every subintercal is less than 6. Proof. The proof is entirely analogous to the one-dimensional case so we only outlinethe principal steps.
We argue by contradiction, using the method of successive bisection.We assume the theorem is false; that is, we assume that for some E,, the interval [a, b] 3 2 2 Applications of diferential calculus cannot be partitioned into a finite number of subintervals in each of which the span off isless than Q. By successive bisection we obtain an infinite set of subintervals I('), I@), . . . ,in each of which the span off is at least E,,. By considering the least upper bound of theleftmost endpoints of the component intervals of Z(l), Z@), . . . we obtain a point t in [a, b]lying in all these intervals.
By continuity off at t there is an n-ball B(t; Y) such that the spanoff is less than $x0 in B(t; r)·| [a, b] . But, when m is sufficiently large, the interval Zfrn)lies in the set B(t; r) n [a, b] , so the span off is no larger than &E,, in Ztrn), contradicting thefact that the span off is at least q, in Pm). 10 LINE INTEGRALS 10.1 Introduction In Volume I we discussed the integral js f(x) dx, first for real-valued functions definedand bounded on finite intervals, and then for unbounded functions and infinite intervals.The concept was later extended to vector-valued functions and, in Chapter 7 of Volume II,to matrix functions. This chapter extends the notion of integral in another direction. The interval [a, b] isreplaced by a curve in n-space described by a vector-valued function a, and the integrandis a vector fieldfdefined and bounded on this curve. The resulting integral is called a lineintegral, a curvilinear integral, or a contour integral, and is denoted by j' f * da or by somesimilar symbol. The dot is used purposely to suggest an inner product of two vectors. Thecurve is called a path of integration. Line integrals are of fundamental importance in both pure and applied mathematics.They occur in connection with work, potential energy, heat flow, change in entropy,circulation of a fluid, and other physical situations in which the behavior of a vector orscalar field is studied along a curve. 10.2 Paths and line integrals Before defining line integrals we recall the definition of curve given in Volume 1. Let abe a vector-valued function defined on a finite closed interval J = [a, b] As t runs throughJ, the function values a(t) trace out a set of points in n-space called thegruph of the function.If a is continuous on J the graph is called a curve; more specifically, the curve describedby a. In our study of curves in Volume I we found that different functions can trace out thesame curve in different ways, for example, in different directions or with different velocities.In the study of line integrals we are concerned not only with the set of points on a curve butwith the actual manner in which the curve is traced out, that is, with the function a itself.Such a function will be called a continuous path. DEFINITION. Let J = [a, b] be ajnite closed interval in R1. A function a: JA R" whichis continuous on J is called a continuous path in n-space. The path is called smooth if thederivative a' exists and is continuous in the open interval (a, 6). The path is calledpiecewise 323 324 Line integrals smooth if the interval [a, b] can be partitioned into ajinite number of subintervals in each ofwhich the path is smooth.
Figure 10.1 shows the graph of a piecewise smooth path. In this example the curve has atangent line at all but a finite number of its points. These exceptional points subdivide thecurve into arcs, along each of which the tangent line turns continuously. FIGURE 10.1 The graph of a piecewise smooth path in the plane. DEFINITION OF LINE INTEGRAL.
Let a be a piecewise smooth path in n-space dejned on aninterval [a, b], and let f be a vector field defined and bounded on the graph of a. The lineintegral qf f along a is denoted by the symbol j" f. da and is defined by the equation (10.1) jf. da = jabf[a(t)] . a'(t) dt, whenever the integral on the right exists, either as a proper or improper integral. Note: In most examples that occur in practice the dot productf[a(t)] - a'(t)is boundedon [a, 61 and continuous except possibly at a finite number of points, in which case theintegral exists as a proper integral. 10.3 Other notations for line integrals If C denotes the graph of a, the line integral J f.
da is also written as SC f * da and iscalled the integrat off along C. If a = a(a) and b = a(b) denote the end points of C, the line integral is sometimeswritten as j: for as jf: f. da and is called the line integral off from o to b along a. Whenthe notation Jt f is used it should be kept in mind that the integral depends not only on theend points a and b but also on the path a joining them. When u = b the path is said to be closed. The symbol 4 is often used to indicate integra-tion along a closed path. When f and a are expressed in terms of their components, say f = (fi,. . . ,f,) and a = (a,,. . . , a,), the integral on the right of (10.1) becomes a sum of integrals, In this case the line integral is also written as J fi da, + * * * + f, da,. Other notations for line integrals 3 2 5 In the two-dimensional case the path 01 is usually desc:ribed by a pair of parametricequations, x = %(t> 3 y = @z(t) 3 and the line integral Jof* da is written as Jcfi dx +fi dy , or as Jcf,(x,y) dx +fi(x,y) 4~.In the three-dimensional case we use three parametric equations, x = %(t>, y = az(t> 9 z = Q(t) and write the line integral SC f * da as Jo fi dx + fZ dy + J; dz , or as I&(x,y,z)dx +fi(x,y,z)dy +fdx,y,z)dz. EXAMPLE. Let f be a two-dimensional vector field given by f(x, y) = Jy i + (x3 + Y)i for all (x, y) withy 2 0. Calculate the line integral offrom (0, 0) to (1, 1) along each ofthe following paths : (a) the line with parametric equations x = t , y = t , 0 < t 5 1;(b) the path with parametric equations x = t2, y = t3, 0 < t 5 1 . Solution. For the path in part (a) we take a(t) = ti + f,i. Then a'(t) = i + i and f[a(t)] = &i + (t3 + t)j. Therefore the dot product of f[a(t)] and a'(t) is equal to l/i + t3 + t and we find fff. da = I' (Ji + t3 + t) dt = E. For the path in part (b) we take a(t) = t2i + t3j. Then a'(t) = 2ti + 3ty andf[a(t)] = t?'% + (t" + t3)j. Therefore sof[a(r)] * a'(t) = 2@ + 3t* + 3f5, [(l"'f. da = J' (2t52 + 3P + 3t") dt = 42 .SF/. (0.0) 0 This example shows that the integral from one point to another may depend on the pathjoining the two points. Now let us carry out the calculation for part (b) once more, using the same curve butwith a different parametric representation. The same curve can be described by thefunction P(t) = fi + t?sj, where 0 < f < 1. 3 2 6 This leads to the relation Line integrals f[P(t)] . P'(t) = (txj + (t3 + ts)j) * (i +it!$j) = t3/: + $tpi + $t2, the integral of which from 0 to 1 is 59/42, as before. This calculation illustrates that thevalue of the integral is independent of the parametric representation used to describe thecurve. This is a general property of line integrals which is proved in the next section. 10.4 Basic properties of line integrals Since line integrals are defined in terms of ordinary integrals, it is not surprising to findthat they share many of the properties of ordinary integrals. For example, they have alinearity property with respect to the integrand, J(af+bg).da=alf.da+bJg.da, and an additiveproperty with respect to the path of integration: Jc f'. da = Jc, f. da + 1^,, J'. da, where the two curves C, and C, make up the curve C. That is, C is described by a functiona defined on an interval [a, b], and the curves C, and C, are those traced out by a(t) as tvaries over subintervals [a, c] and. [c, b], respectively, for some c satisfying a < c < b.The proofs of these properties follow immediately from the definition of the line integral;they are left as exercises for the reader. Next we examine the behavior of line integrals under a change of parameter. Let a be acontinuous path defined on an interval [a, b], let u be a real-valued function that is differen-tiable, with U' never zero on an interval [c, d], and such that the range of u is [a, b]. Thenthe function (3 defined on [c, d] by the equation P(t) = ab(t)l is a continuous path having the same graph as a. Two paths a and p so related are calledequivalent. They are said to provide different parametric representations of the same curve.The function u is said to define a change of parameter. Let C denote the common graph of two equivalent paths a and p. If the derivative of u is always positive on [c, d] the function u is increasing and we say that the two paths a and f3trace out C in the same direction. If the derivative of u is always negative we say that aand p trace out C in opposite directions. In the first case the function u is said to beorientation-preserving; in the second case u is said to be orientation-reversing. An exampleis shown in Figure 10.2. The next theorem shows that a line integral remains unchanged under a change of param-eter that preserves orientation; it reverses its sign if the change of parameter reversesorientation. We assume both intergals jf* da and Jf. df3 exist. Basic properties of fine integrals 327 U 4 ---- ----- ----! :: j h IIc d :t -IIc II +d t (4 (b)FIGURE 10.2 A change of parameter defined by u = h(r). In (a), the function h preserves orientation. In (b), the function h reverses the orientation. THEOREM 10.1. BEHAVIOR OF A LINE INTEGRAL UNDER A CHANGE OF PARAMETER. Leta and p be equivalent piecewise smooth paths. Then we have s, f. da = Jo f.
dBif a and f3 trace out C in the same direction; and if a and p trace out C in opposite directions.
Proof , I t suffices to prove the theorem for smooth paths; then we invokethe additive property with respect to the path of integration to deduce the result for piece-wise smooth paths. The proof is a simple application of the chain rule. The paths a and p are related by anequation of the form P(f) = a[u(t)], where u is defined on an interval [c, d] and a is definedon an interval [a, b]. From the chain rule we have Therefore we findp'(t) = a'[u(t)]u'(t). Jc f. dP = J:f [P(t)1 . P'(t) dt = j~f.(a[u(t)l) . a'[u(t)lu'(t> dt. In the last integral we introduce the substitution v = u(t), dv = u'(t) dt to obtain jc f. dp = j~~'p,'f(a(v)) . a'(v) du = & j.f(a(v)) . a'(v) dv = hjc f. day 328 Line integrals where the + sign is used if a = u(c) and b = u(d), and the - sign is used if a = u(d) andb = u(c). The first case occurs if a and p trace out C in the same direction, the second ifthey trace out C in opposite directions. 10.5 Exercises In each of Exercises 1 through 8 calculate the line integral of the vector fieldfalong the pathdescribed. 1. f(x,y) = (x2 - 2xy)i + (y" - 2xy)j, from (- 1, 1) to (1, 1) along the parabola y = x2.2. f(x,y) = (2a -y)i + xj, along the path described by a(r) = a(f - sin f)i + a(1 - cos t)j, 0 < l 2 2ir.3. f(x, y, z) = (y" - z2)i + 2yzj - x*k, along the path described by a(t) = ti + r? + t3k, o. . , IX,) be a path joining a and 6, say a(a) = a and a(b) = b, and we Line integrals with respect to arc length 329 write c = (c,, . . . , c,). Assume a' is continuous on [a, b]. Then the work done by f isequal to sf. da = ick Jab LX,(~) dt = icJ~(b) - a,(a)] = c . [a(b) - a(a)] = c . (b - a).K=l k=l For this force field the work depends only on the end points a and b and not on the curvejoining them. Not all force fields have this property. Those which do are called conservative.The example on p. 325 is a nonconservative force field. In a later section we shall determineall conservative force fields.
EXAMPLE 2. The principle of work and energy. A particle of mass m moves along a curveunder the action of a force field f.
If the speed of the particle at time t is v(t), its kineticenergy is defined to be &mu"(t). Prove that the change in kinetic energy in any time intervalis equal to the work done by f during this time interval. Solution. Let r(t) denote the position of the particle at time t. The work done by fduring a time interval [a, b] is J:c', f * dr . We wish to prove that sRyan f. dr = $mv2(b) - &mu"(a). From Newton's second law of motion we have f [r(t)] = mr"(t) = mu'(t), where u(t) denotes the velocity vector at time t. The speed is the length of the velocityvector, v(t) = Ilu(t .
Therefore f [r(t)] * r'(t) = f [r(t)] . u(t) = mu'(t) f u(t) = irn $ (u(t) f u(t)) = $m i (v"(t)), Integrating from a to b we obtain jr::, f * dr = i," f [r(t)] * r'(t) dt = &mv2(t)ll = +mv2(b) - imu"(as required. 10.7 Line integrals with respect to arc length Let a be a path with a' continuous on an interval [a, b]. The graph of a is a rectifiablecurve. In Volume I we proved that the corresponding arc-length function s is given by theintegral 40 = s,' I I a'(u>II du .
The derivative of arc length is given by s'(t) = IIa'(OII .
330 Line integrals Let y be a scalar field defined and bounded on C, the graph of a. The line integral of ywith respect to arc length along C is denoted by the symbol SC 9 ds and is defined by theequation sc p ds = s," da(Ols'(O dt, whenever the integral on the right exists.Now consider a scalar field q given by q[a(t)] =f[a(t)] * T(t), the dot product of a vector fieldfdefined on C and the unit tangent vector T(t) = (da/ds) . In this case the lineintegral SC y ds is the same as the line integral Jcf* da because f[a(t)l . a'(t) = f[a(t)l * 2 2 = f[a(t)] * T(t)s'(t) = q[a(t)]s'(t) .
When f denotes a velocity field, the dot product f. T is the tangential component ofvelocity, and the line integral SC f * T ds is called thejow integral off along C.
When C is aclosed curve the flow integral is called the circulation off along C. These terms are commonlyused in the theory of fluid flow. 10.8 Further applications of line integrals Line integrals Line integrals with respect to arc length also occur in problems cmcerned with massdistribution along a curve. For example, think of a curve C in 3-space as a wire made of athin material of varying density. Assume the density is described by a scalar field M, whereCJJ(X, y, z) is the massper unit length at the point (x, y, z) of C. Then the total mass M ofthe wire is defined to be the line integral of ~1 with respect to arc length: i'vf = jc (x, Y, z> ds.The center of mass of the wire is defined to be the point (Z,?, Z) whose coordinates are ' determined by the equations iM = sc X&G Y, z> ds, JTM = sc IT@, Y, z> ds 9 zM = sc zdx, Y, z> ds. A wire of constant density is called uniform. In this case the center of mass is also called thecentroid. EXAMPLE 1. Compute the mass M of one coil of a spring having the shape of the helixwhose vector equation is a(t)=acosti+asintj+btk if the density at (x, y, z) is x2 + y2 + z2. Solution. The integral for A4 is M = jc (x2 + y2 + z") ds = j02r (a" cos2 t + a2 sin2 t + b2t2)s'(t) dt . Exercises 331 Since s'(t) = ~la'(t)~ ~ and a'(f) = -a sin t i + a cos tj + bk, we have s'(t) = da2 + b2and hence M = Ja2 + bZS2' (a' + b2t') dt = da2 + b2 .oi 2na2 + i r3b' In this example the z-coordinate z of the center of mass is given by ZM = lC z(x" + y2 + z") ds = Ja* + b2 .r.?" bt(a* + b't') df = bJa* + b" (2r2a2 + 4rr4b2). The coordinates R and jj are requested in Exercise 15 of Section 10.9. Line integrals can be used to define the moment of inertia of a wire with respect to anaxis. If 6(x, y, z) represents the perpendicular distance from a point (x, y, z) of C to anaxis L, the moment of inertia IL is defined to be the line integral 1, = i, 6*(x, Y, Mx, Y, z> ds,where ~(x, y, z) is the density at (x, y, z). The moments of inertia about the coordinateaxes are denoted by I,, I,, and I,. EXAMPLE 2. Compute the moment of inertia Z, of the spring coil in Example 1.
Solution. Here 6*(x, y, z) = x2 + y* = a2 and ~(x, y, z) = x2 + y2 + z*, so we have I, = Jc (x2 + y")(x* + y2 + z") ds = a2 IC (x2 + y2 + z") ds = Ma*, where M is the mass, as computed in Example 1. 10.9 Exercises 1. A force field f in 3-space is given by f(x, y, z) = xi + yj + (xz - y)k .
Compute the workdone by this force in moving a particle from (0, 0,O) to (1, 2,4) along the line segment joiningthese two points. 2. Find the amount of work done by the forcef(x, y) = (x2 - y2)i + 2xyj in moving a particle(in a counterclockwise direction) once around the square bounded by the coordinate axes andthelinesx =aandy =u,u > O . 3. A two-dimensional force fieldfis given by the equationf(x, y) = cxyi + x"yy, where c is apositive constant. This force acts on a particle which must move from (0, 0) to the line x = 1along a curve of the form y = axb, where a > 0 and b >O. Find a value of a (in terms of c) such that the work done by this force is independent of b.4. A force fieldfin 3-space is given by the formulaf(x, y, z) = yzi + xzj + x0, + 1)k. Calculate the work done by fin moving a particle once around the triangle with vertices (0, 0, 0), (1, 1, l),(-1, 1, -1) in that order. 332 Line integrals 5. Calculate the work done by the force fieldf(x, y, .z) = (y - z)i + (z - x)j + (x - y)k alongthe curve of intersection of the sphere x2 + y2 + z2 = 4 and the plane z = y tan 0, where0 < 0 < v/2.
The path is transversed in a direction that appears counterclockwise whenviewed from high above the xy-plane. 6. Calculate the work done by the force field f(x, I', z) = y2i + z? + x2k along the curve ofintersection of the sphere x2 + y2 + z2 = u2 and the cylinder x2 + y2 = ax, where z 2 0 anda > 0. The path is traversed in a direction that appears clockwise when viewed from highabove the xy-plane. Calculate the line integral with respect to arc length in each of Exercises 7 through IO. 7. SC (x + r) ds, where C is the triangle with vertices (0, 0), (1, 0), and (0, l), traversed in acounterclockwise direction. 8.
Joy2 ds, where C has the vector equation a(t) = a(t - sin t)i + a(1 - cos t).j, 0 9 provided that 9' is continuous on some open interval containing both a and b.
To extendthis result to line integrals we need a slightly stronger version of the theorem in whichcontinuity of q' is assumed only in the open interval (a, b). THEOREM 10.2. Let 9 be a realfunction that is continuous on a closed interval [a, b] andassume that the integral ji y'(t) dt exists. If q~' is continuous on the open interval (a, b), wehave i," ~'(0 dt = p(b) - da>. Proof.
For each x in [a, b] define f(x) = Jz v'(t) dt . We wish to prove that (10.2) f(b) = v(b) - Ha>.
t For a further discussion of connectedness, see Chapter 8 of the author's Mafhenxdenl Analysis, Addison-Wesley, Reading, Mass., 1957.
334 Line integrals By Theorem 3.4 of Volume I,Sis continuous on the closed interval [a, b]. By Theorem 5.1of Volume 1,fis differentiable on the open interval (a, b), withy'(x) = q'(x) for each xin (a, b). Therefore, by the zero-derivative theorem (Theorem 5.2 of Volume I), thedifference f - v is constant on the open interval (a, b). By continuity, f - y is alsoconstant on the closed interval [a, b]. In particular, f(b) - q(b) =f(a) - v(a). Butsincef(a) = 0, this proves (10.2). THEOREM 10.3. SECOND FUNDAMENTAL THEOREM OF CALCULUS FOR LINE INTEGRALS. Let ~1 be a scalarJield with a continuous gradient Ve, on an open connected setS in R". Then for any two points a and b joined by a piecewise smooth path a in S we have J," VP,. da = v(b) - v(a). Proof. Choose any two points a and b in S and join them by a piecewise smooth path ain S defined on an interval [a, b]. Assume first that a is smooth on [a, b]. Then the lineintegral of Vpl from a to b along a is given by jab Vpl * da = s," Vv[a(t)] * a'(t) dt By the chain rule we have Vp,[a(t)l * a'(t) = g'(t), where g is the composite function defined on [a, b] by the formula g(t) = da(t)1 . The derivative g' is continuous on the open interval (a, b) because Vg, is continuous on Sand a is smooth. Therefore we can apply Theorem 10.2 to g to obtain f VP, * da = s," g'(t) dt = g(b) - s(a) = da(b)1 - Ma>1 = db) - da>.
This proves the theorem if a is smooth.When a is piecewise smooth we partition the interval [a, b] into a finite number (say r) of subintervals [tkpl , tk], in each of which a is smooth, and we apply the result just provedto each subinterval. This gives us VP = i bb(td -vb(LI)l) = v(b) - &I, as required. As a consequence of Theorem 10.3 we see that the line integral of a gradient is inde-pendent of the path in any open connected set S in which the gradient is continuous. For aclosed path we have b = a, so v(b) - q(a) = 0. In other words, the line integral of a -Applications to mechanics 335 continuous gradient is zero around every piecewise smooth closed path in S. In Section 10.14we shall prove (in Theorem 10.4) that gradients are the only continuous vector fields withthis property. 10.12 Applications to mechanics If a vector field f is the gradient of a scalar field p, then y is called a potential functionforf. In 3-space, the level sets of q are called equipotential surfaces; in 2-space they arecalled equipotential lines. (If p denotes temperature, the word "equipotential" is replacedby "isothermal"; if p denotes pressure the word "isobaric" is used.) EXAMPLE 1. In 3-space, let ~(x, y, z) = P, where r = (x2 + y2 + z")%. For everyinteger n we have V(P) = nrne2r, where r = xi + yi + zk. (See Exercise 8 of Section 8.14.) Therefore 9 is a potential ofthe vector field f(x, y, z) = nr"-2r. The equipotential surfaces of p are concentric spheres centered at the origin. EXAMPLE 2. The Newtonian potential. Newton's gravitation law states that the force fwhich a particle of mass A4 exerts on another particle of mass m is a vector of lengthGmMlr2, where G is a constant and r is the distance between the two particles. Place theorigin at the particle of mass M, and let r = xi + yj + zk be the position vector from theorigin to the particle of mass m. Then r = 11~11 and -Y/r is a unit vector with the samedirection as f, so Newton's law becomes f = -GmMr% . Taking n = - 1 in Example 1 we see that the gravitational force f is the gradient of thescalar field given by ~(x, y, z) = GmMr-l. This is called the Newtonian potential.The work done by the gravitational force in moving the particle of mass m from (x1, yly zl> to (x2, y2, z2> is T&Y Yl? Zl> - v(x2,y2,z2)=GmM where rl = (~4 + yl + .+)'A and r2 = (xi + yi + ~3". If the two points lie on the sameequipotential surface then rl = r2 and no work is done. EXAMPLE 3. Theprinciple of conservation of mechanical energy. Let f be a continuous forcefield having a potential v in an open connected set S. Theorem 10.3 tells us that the workdone by fin moving a particle from u to x along any piecewise smooth path in S is 3 3 6 Line integrals q(x) - ~(a), the change in the potential function. In Example 2 of Section 10.6 we provedthat this work is also equal to the change in kinetic energy of the particle, k(x) - k(a)where k(x) denotes the kinetic energy of the particle when it is located at x. Thus, we have or (10.3) k(x) - v(x) = w - fP(u>*The scalar - p(x) is called the potential energy? of the particle. If a is kept fixed and n is allowed to vary over the set S, Equation (10.3) tells us that thesum of k(x) and -v(x) is constant. In other words, if a forcefield is a gradient, the sumof the kinetic andpotential energies of aparticle moving in thisfield is constant. In mechanicsthis is called the principle of conservation of (mechanical) energy.
A force field with apotential function is said to be conservative because the total energy, kinetic plus potential,is conserved. In a conservative field, no net work is done in moving a particle around aclosed curve back to its starting point. A force field will not be conservative if friction orviscosity exists in the system, since these tend to convert mechanical energy into heatenergy. 10.13 Exercises 1. Determine which of the following open sets S in R2 are connected. For each connected set,choose two arbitrary distinct points in S and explain how you would find a piecewise smooth curve in S connecting the two points.(a) S=~(x,y)Ix2+y220). (cl s = {(x, y) I x2 + y2 < 1) .(b) S = {(x,y) I x2 +y2 > O}. (4 S = {(x, y) I 1 < x2 + y2 < 2).(e) S={(x,y)Ix2+y2>1 a n d (~-3)~+y">l].(f) S={(x,y)Ix2+y2<1 o r (~-3)~+y~ f(x, y) = yi - xi.(b) fb, y) = yi + by - x>i.
4. Given a three-dimensional vector field f(x, y, z> = P(x, y, z)i + Qk y, z)i + Nx, y, z)k, t Some authors refer to --9, as the potential function offso that the potential energy at x will be equal to thevalue of the potential function Q, at x. The$rst fundamental theorem of calculus]'or line integrals 337 where the partial derivativesap ap aQ aQ aR aRay' T$i 'dx 'dz ' ax '7; ' are continuous on an open set S. Iffis the gradient of some potential function ~7, prove that at each point of S. ap aQ ap aR aQ aRay=y$ dz=z' a,; = -q 5. For each of the following vector fields, use the result of Exercise 4 to prove that f is not agradient. Then find a closed path C such that $o f # 0.(a)f(x,y,z) =yi +xj + x k .(b) f(x, y, z) = xyi + (x2 + 1)j + z2k. 6.
A force field f is defined in 3-space by the equation f(x,y,z) =yi + zj +yzli. (a) Determine whether or not fis conservative.(b) Calculate the work done in moving a particle along the curve described by as t runs from 0 to X. a(t) = cos ti + sin ~j + etk 7. A two-dimensional force field F is described by the equation F(x, y) = (x + y)i + (x - u)i.
(a) Show that the work done by this force in moving a particle along a curve a(t) = f(t>i + g(t>i, a = f (4.
To extend this theorem to line integrals we begin with a vector fieldf, continuous on anopen connected set S, and integrate it along a piecewise smooth curve C from a fixedpoint a in S to an arbitrary point x.
Then we let 9 denote the scalar field defined by the lineintegral &> = /;f * da, where a describes C. Since S is connected, each point x in S can be reached by such acurve. For this definition of gj(x) to be unambiguous, we need to know that the integraldepends only on x and not on the particular path used to join a to x. Therefore, it is naturalto require the line integral off to be independent of the path in S. Under these conditions,the extension of the first fundamental theorem takes the following form: THEOREM 10.4. FIRST FUNDAMENTAL THEOREM FOR LINE INTEGRALS. Let f be a vectorjieId that is continuous on an open connected set S in R", and assume that the line integral offis independent of the path in S.
Let a be aJixedpoint of S and dejne a scalarJieId q on S bythe equation &I = /Jf * da, where a is any piecewise smooth path in S joining a to x. Then the gradient of y exists and isequal to f; that is, VT(X) = f(x) for every x in S. Proof, We shall prove that the partial derivative DAv(x) exists and is equal to fk(x),the kth component of f(x), for each k = 1, 2, . . . , n and each x in S. Let B(x; r) be an n-ball with center at x and radius r lying in S. If y is a unit vector, thepoint x + hu also lies in S for every real h satisfying 0 < Ihl < r, and we can form thedifference quotient dx + h.v) - ~(4 Because of the additive property of line integrals, the numerator of this quotient can bewritten as rp(x + hu) - v(x) = Ji+"'f. da, and the path joining x to x + hu can be any piecewise smooth path lying in S. In particular, Necessary and su$icient conditions for a vector,field to be a gradient 339 we can use the line segment described by a(t) = x + thy, where 0 5 t < 1. Since a'(t) = hy , the difference quotient becomes (10.4) dx + hY) - v(x)h = 'f(x + thy).ydt.s0 Now we take y = eL, the kth unit coordinate vector, and note that the integrand becomesf(x + thy) . y = f,(x + the,).
Then we make the change (of variable u = ht , du = h dt ,and we write (10.4) in the form (10.5)& + he,) - y(x) 1=- hh + uek) du =: g(h) i g(o) , where g is the function defined on the open interval (-r, r) by the equation At> = S,:XCx + ueJ du .
Since each component ,f, is continuous on S, the first fundamental theorem for ordinaryintegrals tells us that g'(t) exists for each t in (-r, r) and that g'(t) =.fk(x + te,) . In particular, g'(0) =,fk(x). Therefore, if we let h + 0 in (10.5) we find that lirn dx + heA - dx> = lim g(h) - g(O)h+O h h =: g'(0) = f&x).h + 0 This proves that the partial derivative L&&x) exists and equalsf,(x), as asserted. 10.15 Necessary and sufficient conditions for a vector field to be a gradient The first and second fundamental theorems for line integrals together tell us that anecessary and sufficient condition for a continuous vector field to be a gradient on an openconnected set is for its line integral between any two points to be independent of the path.We shall prove now that this condition is equivalent to the lstatement that the line integralis zero around every piecewise smooth closed path. All these conditions are summarized inthe following theorem. THEOREM 10.5. NECESSARY AND SUFFICIENT CONDITIONS FOR A VECTOR FIELD TO BE AG R A D I E N T .
Let f be a vector$eld continuous on an open connected set S in R". Then thejollowing three statements are equivalent. (a) f is the gradient of somepotentialfinction in S.(b) The line integral off is independent of the path in S.(c) The line integral off is zero around every piecewise smooth closedpath in S. 340 Line integrals ProoJ We shall prove that (b) implies (a), (a) implies (c), and (c) implies (b). State-ment (b) implies (a) because of the first fundamental theorem. The second fundamentaltheorem shows that (a) implies (c). To complete the proof we show that (c) implies (b). Assume (c) holds and let C, and C,be any two piecewise smooth curves in S with the same end points. Let C1 be the graph of afunction a defined on an interval [a, b], and let C, be the graph of a function f3 defined on [c, 4.Define a new function y as follows: y(t) =a(t) i f altlb, P(b + d - t) i f b, n and every x in S. Necessary conditions for a vector field to be a gradient 341 Proof. Iffis a gradient, then f = Vg, for some potential function q~. This means that "fj = D,P, foreachj= I,2 ,..., n. Differentiating both members of this equation with respect to xiwe find Difj = DiD,p, . Similarly, we have DJ = D*Di~. Since the partial derivatives D& and D& are continuous on S, the two mixed partialderivatives DiD,p and DjD,pl must be equal on S. This proves (10.6). EXAMPLE 1. Determine whether or not the vector field f(x, y) = 3x2yi + x3yj is a gradient on any open subset of R2. Solution. Here we have fi(x, y> = 3x2y, fi(x, y) = xy. The partial derivatives D2fi and Dlf2 are given by 4fky) = 3x2, D&(x, y) = 3x2y. Since D2flb,y) # D&G, y) except when x = 0 or y = 1, this vector field is not agradient on any open subset of R2. The next example shows that the conditions of Theorem 10.6 are not always sufficientfor a vector field to be a gradient.
EXAMPLE 2. Let S be the set of all (x, y) # (0,O) in R2, and let f be the vector fielddefined on S by the equation f'(x,y) = e2i+ k2j. Show that Dlf2 = D2fi everywhere on S but that, nevertheless, f is not a gradient on S. Solution. The reader can easily verify that DIf2(x, y) = D,f,(x, y) for all (x, y) in S.(See Exercise 17 in Section 10.18.) To prove thatfis not a gradient on S we compute the line integral offaround the unitcircle given by cr(t)=costi+sintj, oSince the line integral around this closed path is not zero, f is not a gradient on S. Furtherproperties of this vector field are discussed in Exercise 18 of Section 10.18. At the end of this chapter we shall prove that the necessary conditions of Theorem 10.6are also sufficient if they are satisfied on an open comex set. (See Theorem 10.9.) 10.17 Special methods for constructing potential functions The first fundamental theorem for line integrals also gives us a method for constructingpotential functions. If f is a continuous gradient on an open connected set S, the lineintegral off is independent of the path in S. Therefore we can find a potential 9 simply byintegrating f from some fixed point a to an arbitrary point x in S, using any piecewisesmooth path lying in S. The scalar field so obtained depends on the choice of the initialpoint a.
If we start from another initial point, say b, we obtain a new potential y. But,because of the additive property of line integrals, q and q~ can differ only by a constant,this constant being the integral off from a to b. The following examples illustrate the use of different choices of the path of integration. (x, b) FIGURE 10.5 Two polygonal paths from (a, b) to (x, u). EXAMPLE 1. Construction of a potential on an open rectangle. If f is a continuous gradienton an open rectangle in R", a potential q~ can be constructed by integrating from a fixedpoint to an arbitrary point along a set of line segments parallel to the coordinate axes.A two-dimensional example is shown in Figure 10.5. We can integrate first from (a, b) to(x, b) along a horizontal segment, then from (x, b) to (x, v) along a vertical segment.Along the horizontal segment we use the parametric representation a(t) = ti + bj, a Y) = jr fk 0 dt + j; fdt, Y) dt . Both formulas (10.7) and (10.8) give the same value fair ~(x, y) because the line integralof a gradient is independent of the path. EXAMPLE 2. Construction of a potentialfunction by the use of indefkite integrals. The useof indefinite integrals often helps to simplify the calculation of potential functions. Forexample, suppose a three-dimensional vector field f = (fi ,fi ,.f3) is the gradient of apotential function 97 on an open set S in R3. Then we ha.ve ?Lfax " a3 = fi)ay a%? = fa2 3 everywhere on S. Using indefinite integrals and integrating the first of these equations withrespect to x (holding y and z constant) we find dx, Y, z> = s fib Y, z> dx + 4y, z> , where A(y, z) is a "constant of integration" to be determined. Similarly, if we integrate theequation apl/ay = fi with respect to y and the equation acp/az = f3 with respect to z weobtain the further relations and dx, Y, z) = j f3k Y, z> dz + Cyx, Y> , where B(x, z) and C(x, y) are functions to be determined. Finding 9; means finding threefunctions A(y, z), B(x, z), and C(x, y) such that all three equations for &x, y, z) agree intheir right-hand members. In many cases this can be done by inspection, as illustrated bythe following example. 344 Line integrals EXAMPLE 3. Find a potential function v for the vector field defined on R3 by the equation f(x, y, 4 = (2xyz + z2 - 2y2 + 1)i + (X2Z - 4xy)j + (x2y + 2xz - 2)k. Solution. Without knowing in advance whether or not f has a potential function v, wetry to construct a potential as outlined in Example 2, assuming that a potential q exists. Integrating the componentf, with respect to x we find &, y, z) = i (2xyz + z2 - 2y2 + 1) dx + A(y, z) = x2yz + xz2 - 2xy2 + x + A(y, z). Integrating f2 with respect to y, and thenf, with respect to z, we find cp(x, Y, z> = I(x2z - 4xy) dy + B(x, z) = x2yz - 2xy2 + B(X, z), ~6, Y, z> = s (x2y + 2xz - 2) dz + C(x, y) = x2yz + xz2 - 22 + c(x, y), By inspection we see that the choices A(y, z) = -2z, B(x, z) = xz2 + x - 2z, andC(x, y) = x - 2xy2 will make all three equations agree; hence the function 9 given by theequation fp(x, y, 2) = x2yz + xz2 - 2xy2 + x - 2z is a potential for f on R3. EXAMPLE 4. Construction of a potential on a convex set. A set S in R" is called convexif every pair of points in S can be joined by a line segment, all of whose points lie in S.An example is shown in Figure 10.6. Every open convex set is connected. Iff is a continuous gradient on an open convex set, then a potential q~ can be constructedby integratingf from a fixed point u in S to an arbitrary point x along the line segmentjoining II to X. The line segment can be parametrized by the function a(t) = a + t(x - a), where 0 5 t 5 1. This gives us a'(t) = x - a, so the corresponding potential is given by the integral (10.9) p(x) = Jb'f(u + t(x - a)) * (x - a) dt. a/ S Convex FIGURE 10.6 In a convex set S, the segment joining a and b is in S for all points a andb in S. Exercises If S contains the origin we can take a = 0 and write (10.9) more simply as 345 (10.10) p,(x) = S,ff(tx) - x dt. 10.18 Exercises In each of Exercises 1 through 12, a vector fieldfis defined by the formulas given. In each casedetermine whether or not f is the gradient of a scalar field. When f is a gradient, find a correspond-ing potential function p. 1. f(x,y) = xi +yj.2. f(x, y) = 3x2yi + x"j.3. f(x, y) = (2xe" + y)i + (x2eu + x - 2y)j.4. f(x,y) = (siny -ysinx + x)i + (cosx + xcosy +v)j.5. f(x, y) = [sin (xy) + xy cos (xy)]i + x2 cos (xy)j.6. f(x,y,z) = xi +yj + zk.7..f(x,y,z)=(x+z)i-((y+z)j+(x--)k.8. f(x, y, z) = 2xy3i + x2z3j + 3x2yz2k.9. f(x, y, z) = 3y4z2i + 4x3z2j - 3x2y2k.
10. f(x,y, z) = (2x2 + 8xy2)i + (3x3y - 3xy)j - (4y2z2 + 2x3z)k.11. f(x, y, z) = 01" cos x + z3)i - (4 - 2~ sin x)j + (3xz2 + 2)k.12. f(x, y, z) = (4xy - 3x2z2 + 1)i + 2(x2 + 1)j - (2x32 + 3z")k. 13. A fluid flows in the xy-plane, each particle moving directly away from the origin. If a particleis at a distance r from the origin its speed is am, where a and n are constants.(a) Determine those values of a and n for which the velocity vector field is the gradient ofsome scalar field.(b) Find a potential function of the velocity whenever the velocity is a gradient. The casen = -1 should be treated separately. 14. If both q and w are potential functions for a continuous vector field f on an open connectedset S in R", prove that q - YJ is constant on S. 15. Let S be the set of all x # 0 in Rn. Let r = /1x(1 , and let f be the vector field defined on S by the equation f(x) = rpx, where p is a real constant. Find a potential function for f on S. The case p = -2 should betreated separately.
16. Solve Exercise 15 for the vector field defined by the equation f(x) 2Xx,r where g is a real function with a continuous derivative everywhere on R'. The following exercises deal with the vector field f defined on the set S of all points (x, y) # (0,O)in R2 by the equation f(x,y> = -h2i +$$j. 346 Line integrals In Example 2 of Section 10.16 we showed that f is not a gradient on S, even though DLfi = Dzfieverywhere on S. 17. Verify that for every point (x, y) in S we have 2 24./'&,y) = ~./-I(x~Y) = (z2 +-2,2 . 18. This exercise shows that f is a gradient on the set T=R2--[(x,y)ly=o, x50} , consisting of all points in the xy-plane except those on the nonpositive x-axis.(a) If (x, y) E T, express x and y in polar coordinates, x =rcosO, y = r sin e, wherer > Oand -m < 0 < r. Prove that 0 is given by the formulas i Yarctan - if x > 0,X e= g2 i f x=0, Yarctan; + n if x < 0. [Recall the definition of the arc tangent function: For any real t, arctan t is the unique realnumber p which satisfies the two conditions tan q = t and -n/2 < q < 42.](b) Deduce that ae Y ae xax=-- - =-- ~2 +y2 3 ay x2 + y2 for all (x, y) in T. This proves that 0 is a potential function for f on the set T:This exercise illustrates that the existence of a potential function depends not only on the vector field f but also on the set in which the vector field is defined. 10.19 Applications to exact differential equations of first order Some differential equations of the first order can be solved with the aid of potentialfunctions. Suppose we have a first-order differential equation of the form Y' = fb, y> . If we multiply both sides by a nonvanishing factor Q(x, y) we transform this equationto the form Q(x, y)y' - f(x, y)Q(x, y) = 0. If we then write P(x, y) for -f(x, y)Q(x, y)and use the Leibniz notation for derivatives, writing dy/dx for y', the differential equationtakes the form (10.11) P(x, y) dx + Q(x, y> dy = 0. Applications to exact deferential equations ofjirst order 347 We assume that P and Q are continuous on some open connected set S in the plane.With each such differential equation we can associate a vector field V, where JW y> = P(x, y)i + Qk y:)i. The components P and Q are the coefficients of dx and dy in Equation (10.11). The differ-ential equation in (10.11) is said to be exact in S if the vector field V is the gradient of apotential; that is, if V(x, y) = Vv(x, y) for each point (x, y) in S, where p is some scalarfield. When such a 91 exists we have ayllax = P and ay/ay = Q, and the differentialequation in (10.11) becomes We shall prove now that each solution of this differenti,al equation satisfies the relationm(x, y) = C, where C is a constant. More precisely, assume there is a solution Y of thedifferential equation (10.11) defined on an open interval (a, b) such that the point (x, Y(x))is in S for each x in (a, b). We shall prove that vb, Y(x)1 = c for some constant C. For this purpose we introduce the composite function g defined on(a, b) by the equation g(x) = pl[x, Y(x)1 - By the chain rule, the derivative of g is given by (10.12) g'(x) = &Y'[x, Y(x)1 + &v[x, Y(x)1 Y'(x) = P(x, y) + QThis proves that everysolution y satisfies the equation ~(x, y) = C. Now we may turn this argument around to find a solution

of the differential equation.Suppose the equation (10.13) hY> = c defines y as a differentiable function of x, say y = Y(x) for x in an interval (a, b), andlet g(x) = pl[x, Y(x)]. Equation (10.13) implies that g is constant on (a, b). Hence, by(10.12), P(x, y) + Q(x, y)y' = 0, so y is a solution. Therefore, we have proved thefollowing theorem : THEOREM 10.7.
Assume that the dl&%erential equation (10.14) P(x, y> dx + Q(x, y) dy =: 0 348 Line integrals is exact in an open connected set S, and let 9 be a scalarjeld satisfying !?2 = p and aYax -Q iij- everywhere in S. Then every solution y = Y(x) of (10.14) whose graph lies in S satisjes theequation ~[x, Y(x)] = C for some constant C. Conversely, if the equation
dejnes y implicitly as a differentiable function of x, then this function is a solution of thedzrerential equation (10.14). The foregoing theorem provides a straightforward method for solving exact differentialequations of the first order.
We simply construct a potential function q and then writethe equation. ~(x, y) = C, where C is a constant. Whenever this equation defines yimplicitly as a function of x, the corresponding y satisfies (10.14).
Therefore we can useEquation (10.13) as a representation of a one-parameter family of integral curves. Ofcourse, the only admissible values of C are those for which ~(x,, y,,) = C for some (x,, , yO)in S. EXAMPLE 1. Consider the differential equation dy 3x2 + 6xy2 TX=- 6x'y + 4y3 * Clearing the fractions we may write the equation as (3x2 + 6xy2) dx + (6x2y + 4~3) dy = 0. This is now a special case of (10.14) with P(x, y) = 3x2 + 6xy2 and Q(x, y) = 6x2y + 4y3.Integrating P with respect to x and Q with respect toy and comparing results, we find thata potential function q~ is given by the formula y(x, y) = x3 + 3x2y2 + y4. By Theorem 10.7, each solution of the differential equation satisfies x3 + 3x2y2 + y4 = c for some C. This provides an implicit representation of a family of integral curves. Inthis particular case the equation is quadratic in y2 and can be solved to give an explicitformula for y in terms of x and C. EXAMPLE 2. Consider the first-order differential equation (10.15) ydx+2xdy=O. Show that the differential equations in Exercises 1 through 5 are exact, and in each case find aone-parameter family of integral curves. 1.2 .3 .4 .5 . 6 . (x + 2y) dx + (2x + y) dy = 0.2xydx + x"dy = 0.(x2 -y)dx - (x + siGy)dy = 0.4sinxsin3ycosxdx -3cos3ycos2xdy = O .(3x2y + 8xy2) dx + (x3 + 8x2y + 12yeY) dy = 0.
Show that a linear first-order equation, y' + P(x)y = Q(x)i, has the integrating factor p(x) =elPtz)dz.
Use this to solve the equation.Let ~(x, y) be an integrating factor of the differential equation P(x, y) dx + Q(x, y) dy = 0.Show that ap aQ- - -ay ax = Qfxlogl/I -P;logl/4. Use this equation to deduce the following rules for finding integrating factors:(a) If (aP/ay - aQ/ax)/Q is a function of x alone, say j"(x), then e'ft2)d2: is an integratingfactor.(b) If (aQ/ ax - aP/ ay)/P is a function ofy alone, sayg(y), then elg(y)dy is an integrating factor.Use Exercise 7 to find integrating factors for the following equations, and determine a one-parameter family of integral curves.(a) y dx - (2x + y) dy = 0.(b) (x3 + y3) dx - xy2 dy = 0.If aP/ay - aQ/ax =f(x)Q(x, y) - g(y>P(x, y), show thait exp {Jf(x> dx + Jg(y) dy} is anintegrating factor of the differential equation P(x, y) dx + Q(x, y) dy = 0. Find such anintegrating factor for each of the following equations and obtain a one-parameter family ofintegral curves.
Exercises 349 Here P(x, y) = y and Q(x, y) = 2x. Since lIPlay = 1 a.nd aQ/ax = 2, this differentialequation is not exact. However, if we multiply both sides by y we obtain an equationthat is exact: (10.16) y2dx + 2xydy = 0. A potential of the vector field y2i + 2xyj is 9(x, y) = xy2, and every solution of (10.16)satisfies the relation xy2 = C. This relation also represents a family of integral curves forEquation (10.15).
The multiplier y which converted (10.15) into an exact equation is called an integratingfactor.
In general, if multiplication of a first-order 1inea.r equation by a nonzero factor~(x, y) results in an exact equation, the multiplier ~(x, y) is called an integrating factor ofthe original equation.
A differential equation may have more than one integrating factor.For example, /~(x, y) = 2xy3 is another integrating factor of (10.15). Some special differ-ential equations for which integrating factors can easily be found are discussed in thefollowing set of exercises.
10.20 Exercises (a) (2x2y + y2) dx + (2x3 - xy) dy = 0.(b) (e"secy - tany)dx +dy = 0.
350 Line integrals 10. The following differential equations have an integrating factor in common. Find such anintegrating factor and obtain a one-parameter family of integral curves for each equation.
(3y + 4xy2) dx + (4x + 5x2y) dy = 0, (6y + x"y") dx + (8x + x"y) dy = 0. 10.21 Potential functions on convex sets In Theorem 10.6 we proved that the conditions are necessary for a continuously differentiable vector fieldf = (fi , . . . ,f,) to be a gradienton an open set S in R". We then showed, by an example, that these conditions are notalways sufficient. In this section we prove that the conditions are sufficient whenever theset S is convex. The proof will make use of the following theorem concerning differentiationunder the integral sign. THEOREM 10.8. Let S be a closed interval in Rn with nonempty interior and let J = [a, b]be a closed interval in R1. Let Jn+l be the closed interval S x J in Rn+l. Write each point inJ,+l as (x, t), where x E S and t E J, Assume that y is a scalarjeld dejned on J,+l such that the partial derivative D,v is continuous on J,+l p wherekisoneof 1,2,..., n. De$ne a scalarjeld p on S by the equation ~(4 = j-," Y(X, 0 dt . Then the partial derivative D,q exists at each interior point of S and is given by the formula In other words, we have 4cp?W = Jab D,w(x, 0 dt. D, Jab 14x, 0 dt = s," &,4x, 0 dt . Note: This theorem is usually described by saying that we can differentiate under theintegral sign.
Proof. Choose any x in the interior of S. Since int S is open, there is an r > 0 such thatx + he, E int S for all h satisfying 0 < Ihl < r. Here ek is the kth unit coordinate vectorin R". For such h we have (10.17) P(X + bed - dx> = 1," {y - Y(X, t) = h &AZ, t>, where z lies on the line segment joining x and x + he,. Hence (10.17) becomes Therefore dx + he,) - ~(4 = s bh D,Y(Z, 0 dt *a '@ + hek) - &) -h s bD,y(w, t) dt = b[D,y(z, t) - D,y(x, t)] dt.a sa The last integral has absolute value not exceeding .r," I~ Y,(z, 0 - QY(x, 0 I dt I (b - a> max IQ&z, 0 - &,4x, 01 where the maximum is taken for all z on the segment joining x to x + he,, and all t in[a, b]. Now we invoke the uniform continuity of D, on S :K J (Theorem 9.10) to concludethat for every E > 0 there is a 6 > 0 such that this maximum is <~/(b - a), whenever0 < Ihl < 6. Therefore dx + bed - ~(4 s bh -a D,y(x, t) dt < E wlhenever 0 < Ihl < 6.
This proves that I&Q)(X) exists and equals Ji D,y(x, t) dt, as required. Now we shall use this theorem to give the following necessary and sufficient condition fora vector field to be a gradient on a convex set.
THEOREM 10.9. Let f = (fl, . . . , f,J be a continuously dlyerentiable vector jield on anopen convex set S in R". Then f is a gradient on S if and only if we have (10.18) wx;.(x) = WC(x) foreachxinSandallk,j= 1,2 ,..., n. Proof. We know, from Theorem 10.6, that the conditions are necessary. To provesufficiency, we assume (10.18) and construct a potential v on S. For simplicity, we assume that S contains the origin. Let v(x) be the integral offalongthe line segment from 0 to an arbitrary point x in S. As shown earlier in Equation (10.10)we have P(X) = j,'f(tx) * x dt = JO1 y(x, t) dt, 352 Line integrals where y(x, t) = f(tx) * x . There is a closed n-dimensional subinterval T of S with non-empty interior such that y satisfies the hypotheses of Theorem 10.8 in T x J, whereJ = [0, I]. Therefore the partial derivative D,pl(x) exists for each k = 1, 2, . . . , n andcan be computed by differentiating under the integral sign, To compute D,+v(x, t), we differentiate the dot productf(tx) . x and obtain D,y(x, t) =f(tx) * D,x + Dk{fW>] * x= f(k) - ek + t(Dkfi(tx>, * . .
, Dkfn(rX)) * x =f,(tx) + t(&fk(f.r), .
., D&(tX)) * x, where in the last step we used the relation (10.18). Therefore we have D,y(X, t) =fk(tX) + tvfk(tX) . X. Now let g(t) = f,(tx) . By the chain rule we have g'(t) = Vfk(tx) * x so the last formula for D,y(x, t) becomes D,Y(& t> = g(t) + tg'(t) . Integrating this from 0 to 1 we find (10.19) &V(x) = j; D,y(x, t> = j,' g(t) dt + j; Q'(t) dt . We integrate by parts in the last integral to obtain jol Q'(t) dt = Q(t) I;- j,' g(t) dt = g(l) - j; g(t) dt. Therefore (10.19) becomes h?'(X) = g(l) = fk(x). This shows that VP, = f on S, which completes the proof. 11 MULTIPLE INTEGRALS 11.1 Introduction Volume I discussed integrals jP, ,f(x) dx, first for functions defined and bounded on finiteintervals, and later for unbounded functions and infinite intervals.
Chapter 10 of VolumeII generalized the concept by introducing line integrals. This chapter extends the concept inyet another direction. The one-dimensional interval [a, b] is replaced by a two-dimensionalset Q, called the region of integration. First we consider rectangular regions; later weconsider more general regions with curvilinear boundaries.
The integrand is a scalar field fdefined and bounded on Q. The resulting integral is called a double integral and is denotedby the symbol SI f9 or by SJ' f(x, Y> dx &.Q Q As in the one-dimensional case, the symbols dx and dy play no role in the definition of thedouble integral; however, they are useful in computations and transformations of integrals.
The program in this chapter consists of several stages. First we discuss the definition ofthe double integral. The approach here is analogous to the one-dimensional case treated inVolume I. The integral is defined first for step functions and then for more generalfunctions. As in the one-dimensional case, the definition does not provide a usefulprocedure for actual computation of integrals. We shall find that most double integralsoccurring in practice can be computed by repeated one-dimensional integration. We shallalso find a connection between double integrals and line integrals. Applications of doubleintegrals to problems involving area, volume, mass, center of mass, and related conceptsare also given. Finally, we indicate how the concepts can be extended to n-space. 11.2 Partitions of rectangles. Step functions Let Q be a rectangle, the Cartesian product of two closed intervals [a, b] and [c, d], Q = b, bl x k, 4 = 1(x, y) 1 x E [a, bl and y E [c, 4). An example is shown in Figure 11 .I. Consider two partitions P, and P, of [a, b] and [c, d], 3 5 3 354 Multiple integrals Y Y d_----- d_ _ _ - - - - 11 ('_ _- --1 I II I II LX I CX b a b FIGURE 11.1 A rectangle Q, the Cartesian FIGURE 11.2 A partition of a rectangle Q.product of two intervals. respectively, say P~= {Xo,X1,...,Xn-l,Xn} and pz = {Y"~Yl, * * * ~Yrn-l&J? where x,, = a, x, = b, yO = c, ym = d.
The Cartesian product P, x P, is said to be apartition of Q. Since P, decomposes [a, b] into n subintervals and P, decomposes [c, d] intom subintervals, the partition P = P, x P, decomposes Q into mn subrectangles. Figure11.2 illustrates an example of a partition of Q into 30 subrectangles. A partition P' of Qis said to be finer than P if P c P', that is, if every point in P is also in P'. The Cartesian product of two open subintervals of P, and P, is a subrectangle with itsedges missing. This is called an open subrectangle of P or of Q. DEFINITION OF STEP FUNCTION. A function f dejined on a rectangle Q is said to be a stepfunction if a partition P of Q exists such that f is constant on each of the open subrectanglesOfP. The graph of an example is shown in Figure 11.3. Most of the graph consists of horizontalrectangular patches. A step function also has well-defined values at each of the boundary FIGURE 1 I .3 The graph of a step function defined on a rectangle Q. The double integral of a step function 355 points of the subrectangles, but the actual values at these points are not relevant to integra-tion theory. Iff and g are two step functions defined on a given rectangle Q, then the linear combina-tion cJ + czg is also a step function. In fact, if P and P' are partitions of Q such that f isconstant on the open subrectangles of P and g is constant on the open subrectangles of P',then qf + czg is constant on the open subrectangles of the union P U P' (which we maycall a common refinement of P and P'). Thus, the set of step functions defined on Q formsa linear space. 11.3 The double integral of a step function Let P = P, x P, be a partition of a rectangle Q into mn subrectangles and letfbe a stepfunction that is constant on the open subrectangles of Q. Let the subrectangle determinedby [xi-r, xi] and [Y~ i - ~, yj] be denoted by Qij and let cij denote the constant value that ftakes at the interior points of Qij. Iffis positive, the volume of the rectangular box withbase Qij and altitude Cij is the product Gj ' Cxi - xi-l>(Yj - yj-1). For any step functionf, positive or not, the sum of all these products is defined to be thedouble integral off over Q. Thus, we have the following definition. DEFINITION OF THE DOUBLE INTEGRAL OF A STEP FUNCTION. Let f be a step function whichtakes the constant value cij on the open subrectangle (xieI, xi) x (yjpl, yj) @a rectangle Q.The double integral off over Q is dejined by the formula (11.1) j j f = i: $ c*j ' Cxi - xi-l)(Yj - Yj-1). Q z=lj=l As in the one-dimensional case, the value of the integral does not change if the partitionP is replaced by any finer partition P'. Thus, the value of the integral is independent of thechoice of P so long as f is constant on the open subrectangles of Q. For brevity, we sometimes write Axi instead of (xi - xi-r) and Ayj instead of (yj - yjP1),and the sum in (11.1) becomes To remind ourselves how this sum is formed, we write the symbol for the integral as This symbol is merely an alternative notation for JjJ Note that if f is constant on the interior of Qy say f(x, y) = k when a < x < b andc < y < d, we have (11.2) SJ' f = k(b - a)(d - c),Q 356 Multiple integrals regardless of the values off on the edges of Q. Since we have b - a = 1," dx and d - c = s,"dyt formula (11.2) can also be written as (11.3) ij f = j; [jab fk Y> dx] dy = j," [jzf(x, Y> dy] dx . The integrals which appear on the right are one-dimensional integrals, and the formulais said to provide an evaluation of the double integral by repeated or iterated integration.In particular, when f is a step function of the type described above, we can write fk Y) dx] dy = jzml[j:- , Ax, Y> dy] dx. Summing on i and i and using (11. l), we find that (11.3) holds for step functions.The following further properties of the double integral of a step function are generaliza- tions of the corresponding one-dimensional theorems. They may be proved as directconsequences of the definition in (11.1) or by use of formula (11.3) and the companiontheorems for one-dimensional integrals. In the following theorems the symbols s and tdenote step functions defined on a rectangle Q. To avoid trivial special cases we assumethat Q is a nondegenerate rectangle; in other words, that Q is not merely a single pointor a line segment. THEOREM 11.1. LINEARITY PROPERTY. For every real c1 and c2 we have j j k&, Y) + 4x, Y)I dx dy = ~1 j j s(x, Y) dx 4v + cz j j t(x, Y) dx dyeQ Q Q THEOREM 11.2. ADDITIVE PROPERTY. If Q is subdivided into two rectangles Q, and Q2,then j j s(x, y) dx 4v = j j 4x, Y> dx dy + j j 6 Y> dx dy .Q Ql QZ THEOREM 11.3 COMPARISON THEOREM. Zfs(x,y) 5 t(x,y)for every(x,y)in Q, wehave j j 4x, Y) dx dy I j j t(x, Y> dx dy .Q Q In particular, if t(x, y) 2 0 for every (x, y) in Q, then SI t(x,y)dxdy>O.Q The proofs of these theorems are left as exercises. Upper and lower double integrals 357 11.4 The definition of the double integral of a function defined and bounded on a rectangle Let f be a function that is defined and bounded on a rectangle Q; specifically, supposethat If(X~Y>I I fv if (x,y)~Q. Then f may be surrounded from above and from below by two constant step functions sand t, where s(x, y) = --M and t(x, y) = M for all (x, y) in Q. Now consider any twostep functions s and t, defined on Q, such that (11.4) 4x3 y) is fk y> I t(x, y) for every point (x, u) in Q . DEFINITION OF THE INTEGRAL OF A BOUNDED FUNCTION OVER A RECTANGLE. If there is oneand only one number Z such that (11.5) s.i sAssume f is bounded on a rectangle Q and let s and t be step functions satisfying (11.4).We say that s is below f, and t is above f, and we write s < f < t . Let S denote the set of allnumbers jj s obtained ass runs through all step functions below f, and let T be the set of all numbek jj t obtained as t runs through all step functions above f. Both sets S and TQ are nonempty since f is bounded. Also, jj s 2 jj t ifs 5 f 5 t, so every number in S is less than every number in T. Therefore S hQas a su&emum, and T has an infimum, and theysatisfy the inequalities jjsIsupSThus, we have J(f)=sup([jIsI/], I(f)=infl[jtlf"'). The foregoing argument proves the following theorem. THEOREM 11.4. Every function f which is bounded on a rectangle Q has a lower integralJ(f) and an upper integral i(f) satisfying the inequalities j j s I i(f) s R(f) 5 j j tQ Q ,for all step functions s and t with s If 5 t . The function f is integrable on Q if and only ifits upper and lower integrals are equal, in which case we have I'! f = m = i(f).Q Since the foregoing definitions are entirely analogous to the one-dimensional case, it isnot surprising to learn that the linearity property, the additive property, and the comparisontheorem as stated for step functions in Section 11.3, also hold for double integrals in general.The proofs of these statements are analogous to those in the one-dimensional case and willbe omitted. 11.6 Evaluation of a double integral by repeated one-dimensional integration In one-dimensional integration theory, the second fundamental theorem of calculusprovides a practical method for calculating integrals. The next theorem accomplishes thesame result in the two-dimensional theory; it enables us to evaluate certain double integralsby means of two successive one-dimensional integrations. The result is an extension offormula (11.3), which we have already proved for step functions. THEOREM 11.5. Let f be defined and bounded on a rectangle Q = [a, b] x [c, d] , andassume that f is integrable on Q. For each$xed y in [c, d] assume that the one-dimensionalintegral ji f(x, y) dx exists, and denote its value by A(y). If the integral jf A(y) dy exists itis equal to the double integral jj f. In other words, we have the formula Q (11.6) ?l?^f(x, Y> dx dy = s," [[;lk Y> dx] dy. Proof. Choose any two step functions s and t satisfying s If < t on Q. Integratingwith respect to x over the interval [a, b] we have jab s(x, Y) dx I A(Y) I jab t(x, Y> dx. Geometric interpretation of the double integral as a volume 359 Since the integral Jf A(y) dy exists, we can integrate both these inequalities with respect toy over [c, d] and use Equation (11.3) to obtain Therefore jz A(y) dy is a number which lies between Jj s and jj t for all step functions sQ Q and t approximating f from below and from above, respectively. Since f is integrable onQ, the only number with this property is the double integral off over Q. Thereforeji A(y) dy = vf, which proves Equation (11.6). Formula (11.6) is said to provide an evaluation of the double integral by repeated oriterated integration. The process is described by saying that first we integrate f with respectto x from a to b (holding y fixed), and then we integrate the result with respect to y fromc to d. If we interchange the order of integration, we have a similar formula, namely, (11.7) [j fb, Y) dx dy = jab [jedf(x, Y> dy] dx , which holds if we assume that jzf(x, y) dy exists for each fixed x in [a, b] and is integrableon [a, b]. 11.7 Geometric interpretation of the double integral as a volume Theorem 11.5 has a simple geometric interpretation, illustrated in Figure 11.4. If f isnonnegative, the set S of points (x, y, z) in 3-space with (x, y) in Q and 0 5 z 5 f(x, y) is The ordinate set S of f'over Q Cross section with area A(y)= (a) 0-J) FIGURE 11.4 The volume of S is the integral of the cross-sectional area: 360 Multiple integrals called the ordinate set off over Q. It consists of those points between the rectangle Q andthe surface z = f (x, y) . (See Figure 11.4(a).) For each y in the interval [c, d], the integral is the area of the cross section of S cut by a plane parallel to the xz-plane (the shaded regionin Figure 11.4(b)). Since the cross-sectional area A(y) is integrable on [c, d], Theorem 2.7of Volume 1 tells us that the integral j: A(y) dy is equal to v(S), the volume of S. Thus, fornonnegative integrands, Theorem 11.5 shows that the volume of the ordinate set off over Qis equal to the double integral JjJ Equation (11.7) gives anothe: way of computing the volume of the ordinate set. Thistime we integrate the area of the cross sections cut by planes parallel to the yz-plane. 11.8 Worked examples In this section we illustrate Theorem 11.5 with two numerical examples. EXAMPLE 1. If Q = [-1, 1] X [O,?'r/2], evaluate Jj (x sin y - ye") dx dy, given that the integral exists. The region of integration is showr?in Figure 11.5. Solution. Integrating first with respect to x and calling the result A(y), we have 2NY) = s l(x sin y - ye") dx = x sin y X=1 -1 2- y e " = - e y + y / e .)I x=-l FIGURE 11.5 The region of integration for FIGURE 11.6 The region of integration forExample 1. Example 2. Worked examples 3 6 1 Applying Theorem 11.5 we find Ii (x sin y - ye") dx dy = ji"4~) dy = ji'" (- ey + y/e) dyQ = (l/e - e) i:" y dy = (l/e - e)r2/8. As a check on the calculations we may integrate first with respect to y: SI (x sin y - ye") dx dy = l:, [Jf" (x sin y - .ye"> dy] dxQ = J'`J-x cos y - $y2ez) IrII'" dx = s', (- 2e"/8 + x) dx = (l/e - e).rr2/8 .
EXAMPLE 2. If Q = [- 1, l] x [0,2], evaluate the double integral jj Jly i x21 dx dy ,given that it exists. Q Solution. If we integrate first with respect toy and call the result H(x), we have H(x) =jz Jly - x21 dy . The region of integration is the rectangle shown in Figure 11.6. Theparabola y = x2 is also shown because of the presence of ly - x21 in the integrand. Abovethis parabola we have y > x2 and below it we have y < x2. This suggests that we split theintegral for H(x) as follows: We remember that x is treated as a constant in each of these integrals. In the first integralwe make the change of variable t = x2 - y and in the second we put t = y - x2. Thisgives us H(X) = s," Jly - x21 dy = -&& dt + j,'-= 'Jt dt = +x3 + $(2 - - ')~"a Applying Theorem 11.5 we find ss JIY - x"l dx dy =Q s: C% ir x3 + 4 (2 - x2)3/21 dx = +s ' (2 - x2)3'2dx0 '*=-x(2 -x2)3/2 + 3xJ2 - x2 + 6 arcsin = 4 + T .3 0 3 2 The same result may be obtained by integrating first with respect to x, but the calculationsare more complicated. 362 Multiple integrals 11.9 Exercises Evaluate the double integrals in Exercises 1 through 8 by repeated integration, given that eachintegral exists. 1.1s xy(x + y) dx dy , where Q = [0, l] x [0, 11.Q 2. II (x3 + 3x5~ + y3> dx dy , where Q = [O, 11 x CO, 11.Q 3. ij (JV + x - 3xy2) dx dy , where Q = [0, 11 x 11, 31. 4 !jsin2 x sin2 y dx dy , where Q = 10, ~1 X [O, ~1. 5 .i s sin(x +y)dxdy, where Q = 10, n/2] X 10, p/2].Q 6. lj lcos (x + y)l dx dy , where Q = 10, ~1 X 10, ~1. 7. j jf(x + y) dx dy, where Q = [0,2] x [0,2], and f(t) denotes the greatest integer 5 t.Q 8. jj yp3 etrJy dx dy, where Q = [O,t] x [l,t],t >O.Q 9. If Q is a rectangle, show that a double integral of the form jjf(x)g(y) dx dy is equal to the product of two one-dimensional integrals.
State the assumpti%is about existence that you areusing.
10. Letfbe defined on the rectangle Q = [0, l] x [0, l] as follows: j (x, y> =1 - x - y i f x+yll,o otherwise. Make a sketch of the ordinate set off over Q and compute the volume of this ordinate set bydouble integration. (Assume the integral exists.) 11. Solve Exercise 10 when i f x2 ly <2x2, otherwise. 12. Solve Exercise 10 when Q = [-1, l] x [-1, l] and f(X,Y) = f' + y2(if x2 + y2 i; 1,otherwise Integrability of continuous functions 363 13. Let f be defined on the rectangle Q = [l, 21 x [l, 41 as follows: (xf(x, y) = + yv2 i f xLet f be defined on the:ectangle Q = [0, 11 x [0, l] as follows: f(x,y)=i if xzy(i f x=y, Prove that the double integral Lj j-exists and equals 0. 11.10 Integrability of continuous functions
The small-span theorem (Theorem 9.10) can be used to prove integrability of a functionwhich is continuous on a rectangle.
THEOREM 11.6. INTEGRABILITY OF CONTINUOUS FUNCTIONS. Ifafinction f is continuouson a rectangle Q = [a, b] x [c, d], then f is integrable on Q. Moreover, the value of theintegral can be obtained by iterated integration, 'Ijf = j," [Is," 0, Y> dx] dy = j; [s:'f(x, Y> dy] dx. Proof. Theorem 9.8 shows that f is bounded on Q, so f has an upper integral and alower integral. We shall prove that -I(f) = I(f). Choose E > 0.
By the small-spantheorem, for this choice of E there is a partition P of Q into a finite number (say n) of sub-rectangles QI, . . . , Q, such that the span offin every subrectangle is less than E. Denote by Mdf) and m,(f), respectively, the absolute maximum and minimum values off inQ,. Then we have Mrdf> - we(f) < 6 foreachk=1,2,..., n. Now let s and t be two step functions defined on the interior ofeach Qk as follows: 44 = m,(f), t(x) = WSf) i f xEintQ,. At the boundary points we define s(x) = m and t(x) = M, where m and Mare, respectively,the absolute minimum and maximum values off on Q. Then we have s If < t for all xin Q. Also, we have and jj t =$lMi,(f)a(Qk) 9Q . 364 Multiple integrals where a@,) is the area of rectangle Q, . The difference of these two integrals is ff t - f/ S = i {Mk(.f) - mk(f))a(Qk> < l i a(Qk) = es(Q),Q Q k=l .k=l where a(Q) is the area of Q. Since Jj s < I(f) < r(f) I jj t, we obtain the inequalityQ Q 0 5 1C.f) - I(f) II 4Q>.
Letting E -+ 0 we see that I(f) = I(f), so f is integrable on Q.Next we prove that the double integral is equal to the first iterated integral in (11.8). For each fixed y in [c, d] the one-dimensional integral ji f (x, JJ) 4x exists since the integrandis continuous on Q. Let A(y) = Ji f (x, y) dx . We shall prove that A is continuous on[c, d]. Ify andyI are any two points in [c, d] we have A(Y) - &YI) = s," {f(x, Y> - fk Y,>> dx from which we find. IA(Y) - A(I (b - a~, - - - -,Ifb, Y> - f(x, ydl = @ - a) If(xly Y> - f(xlp YJI where x1 is a point in [a, b] where]f(x, y) - f (x, u,)] attains its maximum. This inequalityshows that A(y) -+ A&) as y -+yl, so A is continuous at yl. Therefore the integralj% A(y) dy exists and, by Theorem 11.5, it is equal to JjJ A similar argument works whenthe iteration is taken in the reverse order. 0 11.11 Integrability of bounded functions with discontinuities Let f be defined and bounded on a rectangle Q. In Theorem 11.6 we proved that thedouble integral off over Q exists if f is continuous everywhere on Q. In this section weprove that the integral also exists if f has discontinuities in Q, provided the set of dis-continuities is not too large. To measure the size of the set of discontinuities we introducethe following concept. DEFINITION OF A BOUNDED SET OF CONTENT ZERO. Let A be a bounded subset of the plane.The set A is said to have content zero $for every E > 0 there is ajnite set of rectangles whoseunion contains A and the sum of whose areas does not exceed E. In other words, a bounded plane set of content zero can be enclosed in a union of rec-tangles whose total area is arbitrarily small. The following statements about bounded sets of content zero are easy consequences ofthis definition. Proofs are left as exercises for the reader. (a) Any finite set of points in the plane has content zero.(b) The union of a finite number of bounded sets of content zero is also of content zero.(c) Every subset of a set of content zero has content zero.(d) Every line segment has content zero. Double integrals extended over more general regions 365 THEOREM 11.7. Let f be dejned and bounded on a rectangle Q = [a, b] x [c, d]. If theset of discontinuities off in Q is a set of content zero then the double integral JJ f exists. Q Proof. Let M > 0 be such that If 1 < M on Q. Let D denote the set of discontinuitiesoff in Q. Given 6 > 0, let P be a partition of Q such that the sum of the areas of all thesubrectangles of P which contain points of D is less than S. (This is possible since D hascontent zero.) On these subrectangles define step functions s and t as follows: s (x) = -M, t(x) = M. On the remaining subrectangles of P define s and t as was done in the proof of Theorem 11.6.Then we have s If 5 t throughout Q. By arguing as in the proof of Theorem 11.6 weobtain the inequality (11.9) b-ss s < es(Q) + 2M6.Q Q The first term, Es(Q), comes from estimating the integral of t - s over the subrectanglescontaining only points of continuity off; the second term, 2M6, comes from estimatingthe integral of t - s over the subrectangles which contain points of D.
From (11.9) weobtain the inequality 0 I r(f) - I(f) I a(Q) + 246. Letting E + 0 we find 0 5 i(f) - r(S) < 2M6. Since 6 is arbitrary this implies i(f) =J(f), so f is integrable on Q. 11.12 Double integrals extended over more general regions Up to this point the double integral has been defined only for rectangular regions ofintegration. However, it is not difficult to extend the concept to more general regions. Let S be a bounded region, and enclose S in a rectangle Q. Let f be defined and boundedon S. Define a new function p on Q as follows : (11.10) Rx y) = if (x,y> ES,3i f (x,y)~Q - S . In other words, extend the definition off to the whole rectangle Q by making the functionvalues equal to 0 outside S. Now ask whether or not the extended function f is integrableon Q. If so, we say that f is integrable on S and that, by dejinition, First we consider sets of points S in the xy-plane described as follows: S = {(x,y) 1 a I x I b and Q)~(x) 2 y 2 C&C)>, 366 Multiple integrals where ql and p2 are functions continuous on a closed interval [a, 61 and satisfying ~ZJ~ < qZ.An example of such a region, which we call a region of Type I, is shown in Figure 11.7.In a region of Type I, for each point t in [a, 61 the vertical line x = t intersects S in a linesegment joining the curve y = vl(x) to y = qZ(x). Such a region is bounded because q1and p2 are continuous and hence bounded on [a, h]. Another type of region T (Type II) can be described as follows: T= {(x,y)Iclyld and Y~(.Y> 5 x I Ye>, where y1 and yZ are continuous on an interval [c, d] with w1 5 wZ. An example is shownin Figure 11.8. In this case horizontal lines intersect T in line segments. Regions of TypeII are also bounded.
All the regions we shall consider are either of one of these two types orcan be split into a finite number of pieces, each of which is of one of these two types. - FIGURE 11.7 A region S of Type I FIGURE 11.8 A region T of Type 11. Let f be defined and bounded on a region S of Type I. Enclose S in a rectangle Q anddefinefon Q as indicated in Equation (11.10).
The discontinuities offin Q will consist ofthe discontinuities off in S plus those points on the boundary of S at which f is nonzero.The boundary of S consists of the graph of cpl, the graph of pjp, and two vertical linesegments joining these graphs. Each of the line segments has content zero. The nexttheorem shows that each of the graphs also has content zero. THEOREM 11.8. Let p be a real-valued function that is continuous on an interval [a, b].Then the graph of v has content zero. Proof. Let A denote the graph of ~j, that is, let ~={(x,y)ly=rp(x) and alxlb).
Double integrals extended over more general regions 367 Choose any E > 0. We apply the small-span theorem (Theorem 3.13 of Volume I) toobtain a partition P of [a, b] into a finite number of subintervals such that the span of q inevery subinterval of P is less than c/(b - a). Therefore, above each subinterval of P thegraph of p lies inside a rectangle of height E/(b - a). Hence the entire of graph v lieswithin a finite union of rectangles, the sum of whose areas is E. (An example is shown inFigure 11.9.) This proves that the graph of CJI has content zero. b FIGURE 11.9 Proof that the graph of a continuous function p has content zero. The next theorem shows that the double integral JJf exists iffis continuous on int S, theinterior of S. This is the set S int S = {G c, y) 1 a < x < b and Q)~(x) < y < q+,> . THEOREM 11.9. Let S be a region of Type I, between the graphs of q1 and p2. Assume thatf is dejned and bounded on S and that f is continuous on the interior of S. Then the double integral jj f exists and can be evaluated by repeated one-dimensional integration,27 (11.11) Proof. Let Q = [a, b] x [c, d] be a rectangle which contains S, and let y be defined by(11.10).
The only possible points of discontinuity offare the boundary points of S. Sincethe boundary of S has content zero,fis integrable on Q.
For each fixed x in (a, 6) the one-dimensional integral Jty(x, JJ) dy exists since the integrand has at most two discontinuitieson [c, d].
Applying the version of Theorem 11.5 given by Equation (11.7) we have (11.12) 368 Multiple integrals Now.%, y) = f(x, y> if Q)~(x) < y I d-4, andfl(x, y) = 0 for the remaining values of yin [c, d]. Therefore jcd3k Y> dy = j;~',x, Y> dy so Equation (11.12) implies (11.11). There is, of course, an analogous theorem for a region T of Type II. Iffis defined andbounded on T and continuous on the interior of T, then f is integrable on T and the formulafor repeated integration becomes [j fk Y> dx dy = s," [j;,(r)' j (x, y) dx] dy . Some regions are of both Type I and Type II. (Regions bounded by circles and ellipsesare examples.) In this case the order of integration is immaterial and we may write s," [j;l:, f(x, Y> dy] dx = j," f s,"i:'fCx, Y> dx] dy. In some cases one of these integrals may be much easier to compute than the other; it isusually worthwhile to examine both before attempting the actual evaluation of a doubleintegral. 11.13 Applications to area and volume Let S be a region of Type I given by s = {(x, y) 1 a I x 5 b and pll(xl I Y I R(X)). Applying Theorem 11.9 with f (x, y) = 1 for all (x, v) in S we obtain j j dx dy = jab [vz(x> - dx]l dx.S By Theorem 2.1 of Volume I, the integral on the right is equal to the area of the region S.Thus, double integrals can be used to compute areas. If f is nonnegative, the set of points (x, y, z) in 3-space such that (x, y) E S and0 5 z 5 f (x, u) is called the ordinate set off over S. An example is shown in Figure 11.10.Iff is nonnegative and continuous on S, the integral represents the area of a cross-section of the ordinate set cut by a plane parallel to theyz-plane (the shaded region in Figure II. 10). Formula (11.11) of Theorem 11.9 shows thatthe double integral off over S is equal to the integral of the cross-sectional area. Hence Worked examples 369 X section = FIGURE 11.10 The integralI~ ~ ~ ~ ~'.f(x, y) dy is th e area of a cross section of the ordinate set. The iterated integral dx is the volume of the ordinate set. the double integral jjfis equal to the volume of the ordinate set offover S. (See Theorem 2.7 of Volume I, p.sll3.)More generally, if f and g are both continuous on S with f I g , then the double integral s,J' (g - f) is equal to the volume of the solid lying between the graphs of the functions f and g. Similar remarks apply, of course, to regions of Type II. 11.14 Worked examples EXAMPLE 1. Compute the volume of the solid enclosed by the ellipsoid Solution.
The solid in question lies between the graphs of two functions f and g, where g(X, Y) = cd - x2/u2 - y2/b2 and f (x, y> = -g(x, y>.
Here (x, y) E S, where S is the elliptical region given by s = (x, y) f + ;I I 1 .(I I Applying Theorem 11.9 and using the symmetry we find that the volume V of the ellipsoidalsolid is given by V = JJ (g -f) = 2 /i g = 8c 1: [~~~'/'-'"" Jrxx2/a2 - y2/b2 dy] dx.S S 370 Multiple integrals -___Let A = Jl - x2/a". Then the inner integral is jb"" JA2 - ye/b2 tiy = A jb"" Ji- y?lo2 dy .
Using the change of variable y = Ab sin t, dy = Ab cos t dt , we find that the last integralis equal to A2b Therefore In the special case a = b = c the solid is a sphere of radius a and the volume is +na3. EXAMPLE 2. The double integral of a positive function f, jjf(x, y) dx dy, reduces to therepeated integral S j,' [j; j(x, Y) dy] dx. Determine the region S and interchange the order of integration. Solution. For each fixed x between 0 and 1, the integration with respect to y is overthe interval from x2 to x. This means that the region is of Type I and lies between the twocurves y = x2 andy = x. The region S is the set of points between these two curves andabove the interval [0, I]. (See Figure 11 .l 1.) Since S is also of Type II we may interchange FIGURE 11.11 Example 2. FIGURE 11.12 Example 3. the order of integration to obtain Exercises 371 EXAMPLE 3. A double integral of a positive function f, jJ.f(x, y) dx dy, reduces to therepeated integral : s ~3 s [104;;5~%x, Y> dx] dy .
Determine the region S and interchange the order of integration. Solution. For each fixed y between 0 and 3, the integration with respect to x is over the interval from 4y/3 to 425 - y2. Therefore region S is of Type II and lies between the two curves x = 4y/3 and x = J25 - y2. This region, shown in Figure 11.12, is asector of a circle.
When the order of integration is reversed the region must be split intotwo regions of Type I; the result is the sum of two integrals: jo4 [jo3"'"f(x> Y> dy] dx + J^.j [jod25 zzf(~, y) dy] dx. 11.15 Exercises In Exercises 1 through 5, make a sketch of the region of integration and evaluate the doubleintegral. 1.J s x cos (x + y) dx dy , where S is the triangular region whose vertices are (0, 0), (TT, 0), (x, x).s 2 .s.i (1 + x) siny dx dy, where S is the trapezoid with vertices (0, 0), (1, 0), (1,2), (0, 1).S 3 .ss e5+y dx dy, where S = {(x, y) I 1x1 + lyl I 1} .Sn n 4 .JJ xzy2dxdy, where S is the bounded portion of the first quadrant lying between the two hsyperbolas xy = 1 and xy = 2 and the two straight lines y = x and y = 4x. 5 . (x" - y2) dx dy , where S is bounded by the curve y = sin x and the interval [0, ~1.S 6. A pyramid is bounded by the three coordinate planes and the plane x + 2y + 32 = 6. Makea sketch of the solid and compute its volume by double integration. 7. A solid is bounded by the surface z = x2 - y2, the xy-plane, and the planes x = 1 and x = 3.Make a sketch of the solid and compute its volume by double integration. 8. Compute, by double integration, the volume of the ordinate set offover S if:(a) f(x, y> = x2 + y2 and s = {O,y>0}.Cc) f(x,y) = y + 2 x +20 and S =((x,y))x2 +y2 516). 372 Multiple integrals In Exercises 9 through 18 assume that the double integral of a positive function f extended overa region S reduces to the given repeated integral. In each case, make a sketch of the region S andinterchange the order of integration. 9. j;[j~fk,d~l 4. 10. j; [j;:m Y) d-g dY * 11. I," [j:; f(X? Y) dY] dx * 12. j;[j2~f(x,y)dy] dx.213*-6s [I;,lalj4 l-(X, Y) dY] dx * 1/14. J;"[jFzf(x,y)dy] d x . 15. f,I [i1-2 f(xt y) dy] dx . 16. j; [j= f(x, y> dy] dx . 17. j; [j:s;r:(z,z) J-k Y) du] dx * '48. j; [j;, f(x, y> dx] dy . 19. When a double integral was set up for the volume Yof the solid under the paraboloid z = x2 +y2 and above a region S of the xy-plane, the following sum of iterated integrals was obtained: v = j; [so" (x2 + y2) dx] dy + j; [j;-' (x2 + y2) dx] dy . Sketch the region S and express V as an iterated integral in which the order of integration isreversed.
Also, carry out the integration and compute V. 20.
When a double integral was set up for the volume V of the solid under the surface z = f(x, y)and above a region S of the xy-plane, the following sum of iterated integrals was obtained: V= j;"inc[j~f(x,y)dx] dy + jI:;;:[j,'zf(x,y)dx] dy. Given that 0 < a < b and 0 < c < ~12, sketch the region S, giving the equations of all curveswhich form its boundary.
21. When a double integral was set up for the volume V of the solid under the surface z = f(x, y)and above a region S of the xy-plane, the following sum of iterated integrals was obtained: v = I," [j:" J-(X, y) dy] dx + j; [j," fk Y) du] dx. (a) Sketch the region S and express V as an iterated integral in which the order of integrationis reversed.(b) Carry out the integration and compute V whenf(x, y) = e'(x/~)'~'. 22. Let A = JA e-@ dt and B = Jy e-@ dt . Evaluate the iterated integral in terms of A and B. These are positive integers 1'72 and n such that Z = mA - nB + ,e-' - ew1i4.
Use this fact to check your answer. Further applications of double integrals 313 23. A solid cone is obtained by connecting every point of a plane region S with a vertex not in theplane of S. Let A denote the area of S, and let h denote the altitude of the cone.
Prove that:(a) The cross-sectional area cut by a plane parallel to the base and at a distance t from thevertex is (t/h)2A, if 0 < f < h .(b) The volume of the cone is &Ah. 24.
Reverse the order of integration to derive the formula as u' em(a-x)f(x) dx] dy = s," (a - x)emcaFr)f(x) dx,0 0where a and m are constants, a > 0. 11.16 Further applications of double integrals We have already seen that double integrals can be used to compute volumes of solidsand areas of plane regions. Many other concepts such as mass, center of mass, and momentof inertia can be defined and computed with the aid of double integrals. This sectioncontains a brief discussion of these topics.
They are of special importance in physics andengineering. Let P denote the vector from the origin to an arbitrary point in 3-space. If n positivemasses m,, m2, . .
. , m, are located at points P, , P2, . . . * P,, respectively, the center ofmass of the system is defined to be the point C determined by the vector equation The denominator, z mk, is called the total mass of the system.If each mass m, is translated by a given vector A to a new point Q, where Qk = Pk + A, the center of mass is also translated by A, since we have 2 mkQk _ z mk(Pk + A) _ 2 mkPk I ti = c + A zmk Irnk Irnk This may also be described by saying that the location of the center of mass depends onlyOnthepointsP,, P2,. . . , P, and the masses, and not on the origin. The center of mass isa theoretically computed quantity which represents, so to speak, a fictitious "balancepoint" of the system. If the masses lie in a plane at points with coordinates (x1 , ~r), . . . , (x,, y,), and if thecenter of mass has coordinates (5, j), the vector relation which defines C can be expressedas two scalar equations and p= 1 mkyk 2 *lk In the numerator of the quotient defining R, the kth term of the sum, mkxk, is called themoment of the mass mk about the y-axis.
If a mass m equal to the total mass of the systemwere placed at the center of mass, its moment about the y-axis would be equal to themoment of the system, II m2 =C mkxk.k=l 374 Multiple integrals When we deal with a system whose total mass is distributed throughout some regionin the plane rather than at a finite number of discrete points, the concepts of mass, centerof mass, and moment are defined by means of integrals rather than sums. For example,consider a thin plate having the shape of a plane region S. Assume that matter isdistributed over this plate with a known density (mass per unit area). By this we meanthat there is a nonnegative function j- defined on S and that f (x, y) represents the massper unit area at the point (x, y). If the plate is made of a homogeneous material the densityis constant.
In this case the total mass of the plate is defined to be the product of the densityand the area of the plate. When the density varies from point to point we use the double integral of the density asthe definition of the total mass. In other words, if the density function f is integrable overS, we define the total mass m(S) of the plate by the equation m(S) = J-s f(x, Y) dx dy .s The quotientmaSS $'fCx, Y> dx dy-=area s,s dx dY is called the average density of the plate. If S is thought of as a geometric configurationrather than as a thin plate, this quotient is called the average or mean value of the functionf over the region S. In this case we do not require f to be nonnegative. By analogy with the finite case, we define the center of mass of the plate to be the point(a,?) determined by the equations (11.14) Zm(s> = if xf(x, Y> dx dy and .WS> = s] yf(x, Y> dx dy .s s The integrals on the right are called the moments of the plate about the y-axis and thex-axis, respectively. When the density is constant, say f(x, y) = c, a factor c cancels ineach of Equations (11.14) and we obtain Ta(S) = s] x dx dy and ya(S) = jj y dx dy ,s s where a(S) is the area of S. In this case the point (2, j) is called the centroid of the plate(or of the region S). If L is a line in the plane of the plate, let 6(x, y) denote the perpendicular distance froma point (x, y) in S to the line L. Then the number IL defined by the equation I, = j j. @b, YMX, Y> dx dys is called the moment of inertia of the plate about L. When f(x, y) = 1, I, is called themoment of inertia or second moment of the region S about L. The moments of inertia Further applications of double integrals 315 about the x- and y-axes are denoted by Z, and I,, respectively, and they are given by theintegrals 1, = j j ~7-k Y> dx dy and I, = j j x7(x, y) dx dyes S The sum of these two integrals is called the polar moment of inertia I,, about the origin : zo = I, + Z, = j j (x2 + y2)f(x, Y> dx dy .s Note: The mass and center of mass of a plate areproperties of the body and are inde-pendent ofthe location of the origin and of the directions chosen for the coordinates axes.The polar moment of inertia depends on the location of the origin but not on the directionschosen for the axes. The moments and the moments of inertia about the x- and y-axesdepend on the location of the origin and on the directions chosen for the axes. If a plate ofconstant density has an axis of symmetry, its centroid will lie on this axis. If there are twoaxes of symmetry, the centroid will lie on their intersection. These facts, which can beproven from the foregoing definitions, often help to simplify calculations involving centerof mass and moment of inertia.
EXAMPLE 1. A thin plate of constant density c is bounded by two concentric circles withradii a and b and center at the origin, where 0 < b < a. Compute the polar moment ofinertia. Solution.
The integral for I0 is I, = css (x2 + y") dx dy ,S where S = {(x, y)) b2 < x2 + y2 < a"} . TO simplify the computations we write the integralas a difference of two integrals, I,=cjj(x2+y2)dxdy-cjj(x2+y2)dxdy:S(a) El(b) where S(a) and S(b) are circular disks with radii a and b, respectively. We can use iteratedintegration to evaluate the integral over S(a), and we find a(x2 + y") dy dy = 4s cs '= (x2 + y") dy0 0 I dx = $. S(a) (We have omitted the details of the computation because this integral can be evaluatedmore easily with the use of polar coordinates, to be discussed in Section 11.27.) Therefore 1, = y (a4 - b4) = n4a2 _ b2) ('" z b2) = m !I?$!, .where m = rc(a" - b2), the mass of the plate.
376 Multiple integrals EXAMPLE 2. Determine the centroid of the plane region bounded by one arch of a sinecurve. Solution.
We take the region S bounded by the curve y = sin x and the intervalO~x~n. By symmetry, the x-coordinate of the centroid is R = 3712. The y-coordinate,j, is given by Ljydxdy S;[Jf""ydy]dx jg&sin2xdx ~ ' = Ljdx d y = jg s i n x dx- = 2 = i* 11.17 Two theorems of Pappus Pappus of Alexandria, who lived around 300 A.D., was one of the last geometers of theAlexandrian school of Greek mathematics. He wrote a compendium of eight bookssummarizing much of the mathematical knowledge of his time. (The last six and a part ofthe second are extant.) Pappus discovered a number of interesting properties of centroids,two of which are described in this section.
The first relates the centroid of a plane regionwith the volume of the solid of revolution obtained by rotating the region about a line inits plane. Consider a plane region Q lying between the graphs of two continuous functions f and gover an interval [a, b], where 0 < g - g2Cd1 dx and that J is given by the formula XQ) = jjy dy dx = s," [i:,':,' y dy] dx = 5," ;[f'(x) - g2(x)] dx.Q Comparing these two formulas we immediately obtain (11.15). EXAMPLE 1. Volume of a torus. Let S be the torus generated by rotating a circular disk Qof radius R about an axis at a distance b > R from the center of Q. The volume of S iseasily calculated by Pappus' theorem. We have ,j = b and a(Q) = .rrR2, so u(S) = 27rjja(Q) = 2.rr2R2b. The next example shows that Pappus' theorem can also be used to calculate centroids. Exercises 317 EXAMPLE 2. Centroid of a semicircular disk. Let j denote the y-coordinate of the centroidof the semicircular disk The area of Q is &R2. When Q is rotated about the x-axis it generates a solid sphere ofvolume QrrR3. By Pappus' formula we have &rR3 = ~~~TJ~(&TR~),4 R soJ=-.37r The next theorem of Pappus states that the centroid of the union of two disjoint planeregions A and B lies on the line segment joining the centroid of A and the centroid of B.More generally, let A and B be two thin plates that are either disjoint or intersect in a set ofcontent zero. Let m(A) and m(B) denote their masses and let C, and C, denote vectorsfrom an origin to their respective centers of mass. Then the union A u B has massm(A) + m(B) and its center of mass is determined by the vector C, where (11.16) c = m(A)G + m(WB*m(A) + m(B) The quotient for C is a linear combination of the form aC, + bC,, where a and b arenonnegative scalars with sum 1. A linear combination of this form is called a convexcombination of C, and C, .
The endpoint of C lies on the line segment joining the endpointsof C, and C, . Pappus' formula (11.16) follows at once from the definition of center of mass given in(11.14). The proof is left as an exercise for the reader. The theorem can be extended in anobvious way to the union of three or more regions. It is especially useful in practice when aplate of constant density is made up of several pieces, each of which has geometric symmetry.We determine the centroid of each piece and then form a suitable convex combination tofind the centroid of the union. Examples are given in Exercise 21 of the next section. 11.18 Exercises In Exercises 1 through 8 a region S is bounded by one or more curves described by the givenequations.
In each case sketch the region S and determine the coordinates 1 and 7 of the centroid. 1. y = x2, x+y=2.2. y2=x +3, y2=5 - x .3. x -2y +8 =o, x+3y+5=0, x = -2 9 x =4.4. y = sin2x, y =o, oy = sinsx, y= -sin2x, -nO; f(x, y) = 1 * 13. (x - r)2 + (y - r)2 = r2, x =o, y =o, O = 1.14. xy = 1, xy =2, x =2y, y =2x , x >o, y >o; fky) = 1.15. y = ez2 y=o, OFind the average distance from one comer of a square of side h to points inside the square.20. Let 6 denote the distance from an arbitrary point P inside a circle of radius r to a fixed point P,, whose distance from the center of the circle is h. Find the average of the function fi2 overthe region enclosed by the circle.
21. Let A, B, C denote the following rectangles in the xy-plane: A = PA41 x 10, 11, B = t&31 x U,31, C = [2,4] x [3,4]. Use a theorem of Pappus to determine the centroid of each of the following figures:(a) A u B. (c) B u C.(b) A u C. (d) A uB UC.
22. An isosceles triangle T has base 1 and altitude h. The base of T coincides with one edge of arectangle R of base 1 and altitude 2. Find a value of h so that the centroid of R u Twill lieon the edge common to R and T. 23. An isosceles triangle T has base 2r and altitude h. The base of T coincides with one edge of asemicircular disk D of radius r.
Determine the relation that must hold between r and h so thatthe centroid of T U D will lie inside the triangle. 11.19 Green's theorem in the plane The second fundamental theorem of calculus for line integrals states that the line integralof a gradient Vf along a path joining two points a and b may be expressed in terms of thefunction values f (a) and f (b). There is a two-dimensional analog of the second fundamental Green's theorem in the plane 379 theorem which expresses a double integral over a plane region R as a line integral takenalong a closed curve forming the boundary of R. This theorem is usually referred to asGreen's theorem.?
It can be stated in several ways; the most common is in the form of theidentity : (11.17) The curve C which appears on the right is the boundary of the region R, and the integra- tion symbol $ indicates that the curve is to be traversed in the counterclockwise direction, as suggested by the example shown in Figure 11.13. FIGURE 11.13 The curve C is the boundary of R, traversed in a counterclockwisedirection. Two types of assumptions are required for the validity of this identity.
First, conditionsare imposed on the functions P and Q to ensure the existence of the integrals. The usualassumptions are that P and Q are continuously differentiable on an open set S containingthe region R. This implies continuity of P and Q on C as well as continuity of aPlay andaQ/ax on R, although the theorem is also valid under less stringent hypotheses. Second,there are conditions of a geometric nature that are imposed on the region R and its boundarycurve C. The curve C may be any rectijable simple closed curve. The term "rectifiable"means, of course, that C has a finite arc length. To explain what is meant by a simpleclosed curve, we refer to the vector-valued function which describes the curve. Suppose C is described by a continuous vector-valued function a defined on an interval[a, b]. If a(a) = a(b), the curve is closed. A closed curve such that a(tl) # a(&) for everypair of values t, # t, in the half-open interval (a, b] is called a simple closed curve. Thismeans that, except for the end points of the interval [a, b], distinct values of t lead todistinct points on the curve. A circle is the prototype of a simple closed curve. Simple closed curves that lie in a plane are usually called Jordan curves in honor ofCamille Jordan (1838-1922), a famous French mathematician who did much of the pioneer-ing work on such concepts as simple closed curves and arc length. Every Jordan curve C t In honor of George Green (1793-1841), an English mathematician who wrote on the applications ofmathematics to electricity and magnetism, fluid flow, and the reflection and refraction of light and sound.The theorem which bears Green's name appeared earlier in the researches of Gauss and Lagrange. 380 Multiple integrals decomposes the plane into two disjoint open connected sets having the curve C as theircommon boundary.
One of these regions is bounded and is called the interior (or innerregion) of C. (An example is the shaded region in Figure 11.13.) The other is unboundedand is called the exterior (or outer region) of C. For some familiar Jordan curves such ascircles, ellipses, or elementary polygons, it is intuitively evident that the curve dividesthe plane into an inner and an outer region, but to prove that this is true for an arbitraryJordan curve is not easy. Jordan was the first to point out that this statement requiresproof; the result is now known as the Jordan curve theorem. Toward the end of the 19thcentury Jordan and others published incomplete proofs.
In 1905 the American mathe-matician Oswald Veblen (1880-1960) gave the first complete proof of this theorem.Green's theorem is valid whenever C is a rectifiable Jordan curve, and the region R is theunion of C and its interior.? Since we have not defined line integrals along arbitraryrectifiable curves, we restrict our discussion here to piecewise smooth curves. There is another technical difficulty associated with the formulation of Green's theorem.We have already remarked that, for the validity of the identity in (11.17), the curve Cmust be traversed in the counterclockwise direction. Intuitively, this means that a manwalking along the curve in this direction always has the region R to his left. Again, forsome familiar Jordan curves, such as those mentioned earlier, the meaning of the expression"traversing a curve in the counterclockwise direction" is intuitively evident. However, in astrictly rigorous treatment of Green's theorem one would have to define this expression incompletely analytic terms, that is, in terms of the vector-valued function a that describesthe curve. One possible definition is outlined in Section 11.24. Having pointed out some of the difficulties associated with the formulation of Green'stheorem, we shall state the theorem in a rather general form and then indicate briefly whyit is true for certain special regions. In this discussion the meaning of "counterclockwise"will be intuitive, so the treatment is not completely rigorous. THEOREM 11.10. GREEN'S THEOREM FOR PLANE REGIONS BOUNDED BY PIECEWISE SMOOTHJORDAN CURVES. Let P and Q be scalar fields that are continuously differentiable on anopen set S in the xy-plane.
Let C be apiecewise smooth Jordan curve, and let R denote theunion of C and its interior. Assume R is u subset of S. Then we have the identity (11.18) jji~ ~ ~ ~dxdg= ~ ~ ~ ~Pdx+Qd,, R where the line integral is taken around C in the counterclockwise direction. Note: The identity in (11.18) is equivalent to the two formulas (11.19)JJ gdxdy = R t A proof of Green's theorem for regions of this generality can be found in Chapter 10 of the author'sMathematical Analysis. Green's theorem in the plane 381 and (11.20) In fact, if both of these are true, (11.18) follows by addition. Conversely, if (I 1.18) is truewe may obtain (11.19) and (11.20) as special cases by taking P = 0 and Q = 0, respectively. Proof for special regions.
We shall prove (1 I .20) for a region R of Type I. Such a regionhas the form R = {(x3 y) 1 a I x I b and .fW I y I g(411 where f and g are continuous on [a, b] with f 5 g. The boundary C of R consists of fourparts, a lower arc C, (the graph of,f), an upper arc C, (the graph ofg), and two vertical linesegments, traversed in the directions indicated in Figure 11.14. Y, G:Y = g(x) c' 1t 7 ; c,:y = f(x) / FIGURE 11.14 Proof of Green's theorem for aspecial region. V FIGURE 11.15 Proof of Green's theorem fora more general region. First we evaluate the double integral -Lj (aPlay) dx 4~ by iterated integration. Integrat-ing first with respect to y we have = Jab Pb,f(x>l dx - j-abPb, s(x>l dx. On the other hand, the line integral SC P dx can be written as follows: jcPdx = ?;,Pdx + j-c,Pdx,since the line integral along each vertical segment is zero. To evaluate the integral along 382 Multiple integrals~-- C, we use the vector representation a(t) = ti + ,f'(t).j and obtain To evaluate the integral along C, we use the representation a(t) = ti + g(t)j and take intoaccount the reversal in direction to obtain Jcz P dx = -1; P[t, g(t)] dt. Therefore we have .r, I' dx = /:' P[t, f(t)] dt -- f P[t, g(t)] dt.
Comparing this equation with the formula in (11.21) we obtain (11.20).A similar argument can be used to prove (11.19) for regions of Type II. In this way a proof of Green's theorem is obtained for regions that are of both Type I and Type II.Once this is done, the theorem can be proved for those regions R that can be decomposedinto a finite number of regions that are of both types. "Crosscuts" are introduced as shownin Figure 11.15, the theorem is applied to each subregion, and the results are addedtogether. The line integrals along the crosscuts cancel in pairs, as suggested in the figure,and the sum of the line integrals along the boundaries of the subregions is equal to the lineintegral along the boundary of R.
11.20 Some applications of Green's theorem The following examples illustrate some applications of Green's theorem. EXAMPLE 1. Use Green's theorem to compute the work done by the force fieldf(x, y) = (y + 3x)i + (2-y - x)j in moving a particle once around the ellipse 4x2 + y2 = 4in the counterclockwise direction. Solution.
TheworkisequaltoS,Pdx+Qd,v,whereP=y+3x,Q=2y-xx,andC is the ellipse.
Since aQ/ax - aP/ay = -2, Green's theorem gives us s, P dx + Q dy = s.r (-2) dx dy = -2a(R),R where a(R) is the area of the region enclosed by the ellipse. Since this ellipse has semiaxesa = 1 and b = 2, its area is nab = 2n and the value of the line integral is -44~. EXAMPLE 2. Evaluate the line integral j, (5 - xy - y") dx - (2xy - x2) dy , where C isthe square with vertices (0, 0), (1, 0), (1, l), (0, l), traversed counterclockwise. Solution. Here P = 5 - xy - y2, Q = x2 - 2.x~) and aQ/ax - aP/+ = 3x.
Hence,by Green's theorem, we have JcPdx+Qdy=3j?'xdxdy=3Z,R A necessary and sujficient condition for a vector jeld to be a gradient 383 where X is the x-coordinate of the centroid of the square. Since f is obviously i, the valueof the line integral is $.
EXAMPLE 3. Area expressed as a line integral. The double integral for the area a(R) of aregion R can be expressed in the form a(R)=jJdxdy=/jg-$)dkdy, R R where P and Q are such that aQ/ax - aP/ay = 1. For example, we can take Q(x, y) = -hxand P(x, y) = --$y . If R is the region enclosed by a Jordan curve C we can apply Green'stheorem to express a(R) as a line integral, a(R) =s Pdx+Qdy=Is2c -yddx+ xdy.c If the boundary curve C is described by parametric equations, say x = X(t), y = w>, a = --IL i + g-$x2 + y2 satisfies (11.22) everywhere on the set S = R2 - ((0, O)}, but f is not a gradient on S.In Theorem 10.9 we proved that condition (11.22) is both necessary and sufficient for fto be a gradient on S if the set S is convex. With the help of Green's theorem we can extendthis result to a more general class of plane sets known as simpZy connected sets. They aredefined as follows. 384 Multiple integrals DEFINITION OF A SIMPLY CONNECTED PLANE SET. Let S be an open connected set in theplane.
Then S is called simply connected {f, for every Jordan curve C which lies in S, the innerregion of C is also a subset of S. An annulus (the set of points lying between two concentric circles) is not simply connectedbecause the inner region of a circle concentric with the bounding circles and of radiusbetween theirs is not a subset of the annulus.
Intuitively speaking, we say that S is simplyconnected when it has no "holes." Another way to describe simple connectedness is to saythat a curve C, in S connecting any two points may be continuously deformed into anyother curve Cz in S joining these two points, with all intermediate curves during thedeformation lying completely in S. An alternative definition, which can be shown to beequivalent to the one given here, states that an open connected set S is simply connected ifits complement (relative to the whole plane) is connected.
For example, an annulus is notsimply connected because its complement is disconnected. An open connected set that isnot simply connected is called multiply connected.
THEOREM 11.11. Letf(x, y) = P(x, y)i + Q(x, y)j be a vectorjeld that is continuouslydifferentiable on an open simply connected set S in the plane. Then f is a gradient on S if andonly if we have (11.23)aP aQ-=-ay ax everywhere on S . Proof. As already noted, condition (11.23) is necessary for f to be a gradient. We shallprove now that it is also sufficient.
It can be shown that in any open connected plane set S, every pair of points a and xcan be joined by a simple step-polygon, that is, by a polygon whose edges are parallel tothe coordinate axes and which has no self-intersections. If the line integral offfrom a to xhas the same value for every simple step-polygon in S joining a to x, then exactly the

sameargument used to prove Theorem 10.4 shows that f is a gradient on S. Therefore, we need J[=TjI I 1 t i 1' \, +II ! 2G I + T , I I I 4 11 4r I 1 I I Ir zzL_.I_"_I I I I I f I II a + FIGURE 11.16 Independence of the path in a simply connected region. Exercises 385 only verify that the line integral off from II to x has the same value for every simple step-polygon in S joining a to x. Let C, and C, be two simple step-polygons in S joining a to x. Portions of these polygonsmay coincide along certain line segments. The remaining portions will intersect at most afinite number of times, and will form the boundaries of a finite number of polygonalregions, say RI, . . . , R,.

Since S is assumed to be simply connected, each of the regionsR, is a subset of S. An example is shown in Figure 11.16. The solid line represents C, ,the dotted line represents C,, and the shaded regions represent RI, . . . , R, . (These twoparticular polygons coincide along the segment pq.) We observe next that the line integral off from a to x along C, plus the integral from xback to a along C, is zero because the integral along the closed path is a sum of integralstaken over the line segments common to C, and C, plus a sum of integrals taken aroundthe boundaries of the regions R,. The integrals over the common segments cancel in pairs,since each common segment is traversed twice, in opposite directions, and their sum iszero. The integral over the boundary IJk of each region R, is also zero because, by Green'stheorem we may write and the integrand of the double integral is zero because of the hypothesis aQ/ax = aPlay.It follows that the integral from a to x along C, is equal to that along C,. As we havealready noted, this implies that f is a gradient in S. 11.22 Exercises Use Green's theorem to evaluate the line integral c y2 dx + x dy when (a) C is the square with vertices (0, 0), (2,0), (2,2), (0,2).(b) C is the square with vertices (f 1, f 1).(c) C is the square with vertices (f2,0), (0, f2).(d) C is the circle of radius 2 and center at the origin.(e) C has the vector equation a(t) = 2 cos3 t i + 2 sin3 t j, 0 5 t 5 27. If P(x, y) = xe+' and Q(x, y) = -x2y e- Y2 + 1/(x2 + y2), evaluate the line integral Pdx + Q dy around the boundary of the square of side 2a determined by the inequalities 1x1 < a andlyl I a.Let C be a simple closed curve in the xy-plane and let Z, denote the moment of inertia (aboutthe z-axis) of the region enclosed by C. Show that an integer n exists such that nz, = Cx3dy -y3dx. Given two scalar fields u and v that are continuously differentiable on an open set containingthe circular disk R whose boundary is the circle x2 + y2 = 1 . Define two vector fieldsfand gas follows: f(x, y) = 4x, y)i + 4x, y)i, g(x.yl=(G-JJi+(g-$)j. Find the value of the double integral JJf. g dx dy if it is known that on the boundary of R wehave u(x, y) = 1 and u(x, y) = y . IL 386 Multiple integrals 5.

If f and g are continuously differentiable in an open connected set S in the plane, show that +$c fVg'da = -$,gVf.d a for every piecewise smooth Jordan curve C in S.6. Let u and v be scalar fields having continuous first- and second-order partial derivatives in an open connected set Sin the plane. Let R be a region in S bounded by a piecewise smooth Jordancurve C. Show that: (a) 4? .vdx+.vdy=---(v(~-;) +u(&;))dxdy. ~b++(v~-a$x+(a;-v;)dy =jj(a+$v~)dxdy. R Normalderivatives. In Section 10.7 we defined line integrals with respect to arc length in such away that the following equation holds: lcPdx + Qdy = J, f. Tds,i where f = Pi + Qj and T is the unit tangent.vector to C. (The dot product f T is called thetangential component off along C.) If C is a Jordan curve described by a continuously differ-entiable function a, say a(t) = X(t)i + Y(t)j, the unit outer normal n of C is defined by theequation n(t) = |~a':tj~~~ (Y'(t)i - X'(t)j) whenever |la'(t)l| # 0. If p is a scalar field with gradient VP on C, the normalderivative aq/an isdefined on C by the equation ap7- = Vq2.n.an This is, of course, the directional derivative of p in the direction of n.

These concepts occur in theremaining exercises of this section. If f = Qi - Pj, show that cPdx + Qdy = .c, f.nds.r (The dot product f. n is called the normal component off along C.)Let f and g be scalar fields with continuous first- and second-order partial derivatives on anopen set S in the plane. Let R denote a region (in S) whose boundary is a piecewise smoothJordan curve C. Prove the following identities, where V% = a2u/ax2 + a2u/aY2. (a) $o z ds = j/V'gdXdY.

(b) $)fj+ = jj" (fv2g + Vf . Vg) dx dy . cc) i(f; -g$ ds = //(f"ze -gV2f)dx+. R Green's theorem for multiply connected regions 381 The identity in (c) is known as Green's formula; it shows that whenever f and g are both harmonic on R (that is, when V2f = V2g = 0 on R).9. Suppose the differential equation P(x, y) dx + Q(x, y) dy = 0 has an integrating factor ,u(x, y) which leads to a one-parameter family of solutions of the form~(x, y) = C. If the slope of the curve ~(x, y) = C at (x, y) is tan 0, the unit normal vector n istaken to mean n =sin0i -cosej. There is a scalar field g(x, y) such that the normal derivative of v is given by the formula - = b4-T y>g(x, y) 3an where aq/ an = V q . n . Find an explicit formula for g(x, y) in terms of P(x, u) and Q(x, y). k11.23 Green's theorem for multiply connected regions Green's theorem can be generalized to apply to certain multiply connected regions. THEOREM 11.12. GREEN'S THEOREM FOR MULTIPLY CONNECTED REGIONS.

Let C1, . . . , C,be n piecewise smooth Jordan curves having the following properties: (a) No two qf the curves intersect.(b) The curves C,, . . . , C, all lie in the interior of C, ,(c) Curve Ci lies in the exterior of curve Cj for each i # j, i > 1, j > 1 . Let R denote the region which consists of the union of C, with that portion of the interior of C,that is not inside any of the curves C, , C,, . . .

, C, .

(An example of such a region is shownin Figure 11 .17.) Let P and Q be continuously diflerentiable on an open set S containing R.Then we have the,following identity. (11.24)~!~~~)dxdy= $,.R (Pdx+Qdy)~2~~!Pdx+Qdv,.. The theorem can be proved by introducing crosscuts which transform R into a unionof a finite number of simply connected regions bounded by Jordan curves. Green's theoremis applied to each part separately, and the results are added together. We shall illustratehow this proof may be carried out when n = 2. The more general case may be dealt withby using induction on the number n of curves. The idea of the proof when n = 2 is illustrated by the example shown in Figure 1 I .18,where C, and C, are two circles, C, being the larger circle. Introduce the crosscuts ABand CD, as shown in the figure. Let KI denote the Jordan curve consisting of the upper 388 Multiple integrals FIGURE 11.17 A multiply connected region. FIGURE 11.18 Proof of Green's theoremfor a multiply connected region. half of C,, the upper half of C, , and the segments AB and CD. Let K, denote the Jordancurve consisting of the lower half of Cr, the lower half of C,, and the two segments ABand CD. Now apply Green's theorem to each of the regions bounded by K; and K, andadd the two identities so obtained.

The line integrals along the crosscuts cancel (since eachcrosscut is traversed once in each direction), resulting in the equation (Pdx+Qdy)- The minus sign appears because of the direction in which C, is traversed. This is Equation(11.24) when n = 2. For a simply connected region, the condition aP/ay = aQ/ax implies that the lineintegral j P dx + Q dy is independent of the path (Theorem 11. II). As we have alreadynoted, if S is not simply connected, the condition aP/ay = aQ/ax does not necessarilyimply independence of the path. However, in this case there is a substitute for inde-pendence that can be deduced from Theorem 11.12. THEOREM 11.13. INVARIANCE OF A LINE INTEGRAL UNDER DEFORMATION OF THE PATH. Let P and Q be continuously diflerentiable on an open connected set S in the plane, and assumethat aqay = aQ/a x everywhere on S. Let C, and C, be two piecewise smooth Jordan curveslying in S and satisfying the following conditions: (a) C, lies in the interior of C, .(b) Those points inside C, which lie outside C, are in S. (Figure 11.19 shows an example.) Then we have (11.25) where both curves are traversed in the same direction. Proof.

Under the conditions stated, Equation (11.24) is applicable when n = 2. Theregion R consists of those points lying between the two curves C, and C, and the curves The winding number 389 FIGURE 11.19. Invariance of a line integral under deformation of the path. themselves. Since aP/ay = aQ/a x in S, the left member of Equation (11.24) is zero andwe obtain (11.25).

Theorem 11.13 is sometimes described by saying that if 3P/~?y = aQ/ax in S the value ofa line integral along a piecewise smooth simple closed curve in S is unaltered if the path isdeformed to any other piecewise smooth simple closed curve in S, provided all intermediatecurves remain within the set S during the deformation. The set S is assumed to be openand connected-it need not be simply connected. *11.24 The winding number We have seen that the value of a line integral often depends both on the curve alongwhich the integration takes place and on the direction in which the curve is traversed.For example, the identity in Green's theorem requires the line integral to be taken in thecounterclockwise direction. In a completely rigorous treatment of Green's theorem itwould be necessary to describe analytically what it means to traverse a closed curve inthe "counterclockwise direction." For some special curves this can be done by makingspecific statements about the vector-valued function a which describes the curve. Forexample, the vector-valued function a defined on the interval [0,27r] by the equation (11.26) a (t) = (a COS t + x,)i + (a sin t + yJj describes a circle of radius a with center at (x,, y,). This particular function is said todescribe the circle in a positive or counterclockwise direction. On the other hand, if wereplace t by -t on the right of (11.26) we obtain a new function which is said to describethe circle in a negative or clockwise direction. In this way we have given a completelyanalytical description of "clockwise" and "counterclockwise" for a circle. However, itis not so simple to describe the same idea for an arbitrary closed curve. For piecewisesmooth curves this may be done by introducing the concept of the winding number, ananalytic device which gives us a mathematically precise way of counting the number oftimes a radius vector a "winds around" a given point as it traces out a given closed curve.In this section we shall describe briefly one method for introducing the winding number. 390 Multiple integrals Then we shall indicate how it can be used to assign positive and negative directions toclosed curves. Let C be a piecewise smooth closed curve in the plane described by a vector-valuedfunction u defined on an interval [a, b], say u(t) = X(t)i + Y(t)j i f a dYzic (x - .J2 + (Y - YA2 * It can be shown that the value of this integral is always an integer, positive, negative, orzero. Moreover, if C is a Jordan curve (simple closed curve) this integer is 0 if PO is outsideC and has the value + 1 or - 1 if P,, is inside C. (See Figure 11.20.) Furthermore, W(a; P,,) Winding number + 1 Winding number - I Winding number 0 FIGURE 11.20 Illustrating the possible values of the winding number of a Jordancurve C with respect to PO. is either + 1 for every point P, inside C or it is - 1 for every such point. This enables usto define positive and negative orientations for C as follows: If the winding numberW(u; P,) is + 1 for every point PO inside C we say that a traces out C in the positive orcounterclockwise direction. If the winding number is -1 we say that a traces out C inthe negative or clockwise direction. [An example of the integral in (11.28) with x,, = y,, = 0was encountered earlier in Example 2 of Section 10.16.1 To prove that the integral for the winding number is always +l or - 1 for a simpleclosed curve enclosing (x0, yO) we use Theorem 11.13. Let S denote the open connectedregion consisting of all points in the plane except (x,, , y,). Then the line integral in (11.28)may be written as jc P dx + Q dy , and it is easy to verify that aP/?ly = aQ/ax every-where in S. Therefore, if (x,, , y,,) is inside C, Theorem 11.13 tells us that we may replacethe curve C by a circle with center at (x,, yO) without changing the value of the integral.Now we verify that for a circle the integral for the winding number is either + 1 or - 1, Exercises 391 depending on whether the circle is positively or negatively oriented. For a positivelyoriented circle we may use the representation in Equation (11.26). In this case we have X(t) = aces t + x0, Y(t) = a sin t + yo, and the integrand in (11.27) is identically equal to 1. Therefore we obtain W(a;P,) = -!-s 2nldt = 1. 27?

0 By a similar argument we find that the intergal is -1 when C is negatively oriented.

Thisproves that the winding number is either +l or -1 for a simple closed curve enclosingthe point (x0, yo). k11.25 Exercises 1. Let S = {(x,y) 1x2 + y2 > O}, and let Yw, y) = -x2 +y2' Q(x, y> = s2 if (x, y) E S . Let C be a piecewise smooth Jordan curve lying in S.(a) If (0,O) is inside C, show that the line integral SC P dx + Q dy has the value f25r, andexplain when the plus sign occurs.(b) Compute the value of the line integral Jo P dx + Q dy when (0,O) is outside C. 2. If r = xi + yj and r = ljrl) , let Wg r) .fky) = -g - Wag r) . ax J for r > 0. Let C be a piecewise smooth Jordan curve lying in the annulus 1 < x2 + y2 < 25,and find all possible values of the line integral offalong C. 3. A connected plane region with exactly one "hole" is called doubly connected.

(The annulus1 < x2 + y2 < 2 is an example.) If P and Q are continuously differentiable on an open doublyconnected region R, and if aP/ay = aQ/ax everywhere in R, how many distinct values arepossible for line integrals so P dx + Q dy taken around piecewise smooth Jordan curves in R? 4. Solve Exercise 3 for triply connected regions, that is, for connected plane regions with exactlytwo holes. 5. Let P and Q be two scalar fields which have continuous derivatives satisfying aP/ ay = aQ/ axeverywhere in the plane except at three points.

Let C, , C, , C, be three nonintersecting circles having centers at these three points, as shown in Figure 11.21, and let Zk =Assume that Z, = 12, I2 = 10, Z, = 15. Tfc, P dx + Q dy .li (a) Find the value of SC P dx + Q dy , where C is the figure-eight curve shown.(b) Draw another closed curve JY along which S P dx + Q dy = 1 . Indicate on your drawingthe direction in which I is traversed.(c) If Z, = 12, Z, = 9, and Z3 = 15, show that there is no closed curve I along which 5 P dx +Qdy = 1. 392 Multiple integrals FIGURE 11.21 Exercise 5. FIGURE 11.22 Exercise 6. 6. Let Zk =f-Ck Pdx + Qdy, wheie 1 1 1P(x,u)= - Y [(x-1)2+y2+x2+Y2+ 1(x + 1Y + y2 andx - l X x+1 Q(x,Y) = (x_ 1>2 + y 2 + x2 + y 2 + (x + 112 + y2' In Figure 11.22, C, is the smallest circle, x2 + y2 = 4 (traced counterclockwise), C, is the largestcircle, x2 + y2 = 4 (traced counterclockwise), and C, is the curve made up of the three inter-mediate circles (x - 1)2 + y2 = a, x2 + y2 = $, and (x + 1)2 + y2 = $ traced out as shown.If I2 = 6n and 1s = 277, find the value of Z1. 11.26 Change of variables in a double integral In one-dimensional integration theory the method of substitution often enables us toevaluate complicated integrals by transforming them into simpler ones or into types thatcan be more easily recognized. The method is based on the formula where a = g(c) and b = g(d). We proved this formula (in Volume I) under the assumptionsthat g has a continuous derivative on an interval [c, d] and that f is continuous on the setof values taken by g(t) as t runs through the interval [c, d].

There is a two dimensional analogue of (11.29) called the formula for making a change ofvariables in a double integral.

It transforms an integral of the form jj f(x, y) dx dy, extended over a region S in the xy-plane, into another double integral :J F(u, v) du dv, extended over a new region Tin the uv-plane. The exact relationship bet;een the regionsS and T and the integrands f (x, y) and F(u, v) will be discussed presently. The method of Change of variables in a double integral 395 V A J u = constantx = x(u,u) 1: = Y(u,u)- CM - | - v, = "'i' u u-curve FIGURE 11.24 A u-curve and a corresponding velocity vector. with respect to u and v, respectively. That is, define and These vectors may be interpreted geometrically as follows: Consider a horizontal linesegment in the uv-plane (v is constant on such a segment). The vector function r maps thissegment onto a curve (called a u-curve) in the xy-plane, as suggested in Figure 11.24. If wethink of u as a parameter representing time, the vector V, represents the velocity of theposition r and is therefore tangent to the curve traced out by the tip of r. In the same way,each vector V, represents the velocity vector of a v-curve obtained by setting u = constant.A u-curve and a v-curve pass through each point of the region 5'. Consider now a small rectangle with dimensions Au and Au, as shown in Figure 11.25.If Au is the length of a small time interval, then in time Au a point of a u-curve movesalong the curve a distance approximately equal to the product 11 V, 11 Au (since /I V, /I repre-sents the speed and Au the time). Similarly, in time Au a point on a v-curve moves a FIGURE 11.25 The image of a rectangular region in the uv-plane is acurvilinear parallelogram in the x),-plane. 396 Multiple integrals distance nearly equal to II V,ll Au. Hence the rectangular region with dimensions Au andAu in the uv-plane is traced onto a portion of the xy-plane that is nearly a parallelogram,whose sides are the vectors VI Au and V, AU, as suggested by Figure 11.25. The area ofthis parallelogram is the magnitude of the cross product of the two vectors V, Au andV, Au; this is equal to llV,W x V',Wll = IIV, x V,ll AuAv. If we compute the cross product VI x V, in terms of the components of V, and V, we find i axauax7% j ar au arav ka x 0 au = ax 0 au ayau ayk = J(u, u)k. -au Therefore the magnitude of VI x V, is exactly lJ(u, v)l and the area of the curvilinearparallelogram in Figure 11.25 is nearly equal to lJ(u, v)l Au Au. If J(u, V) = 1 for all points in T, then the "parallelogram" has the same area as therectangle and the mapping preserves areas. Otherwise, to obtain the area of the parallelo-gram we must multiply the area of the rectangle by lJ(u, u)l. This suggests that the Jacobianmay be thought of as a "magnification factor" for areas. Now let P be a partition of a large rectangle R enclosing the entire region T and considera typical subrectangle of P of, say, dimensions Au and AU. If Au and Au are small, theJacobian function J is nearly constant on the subrectangle and hence J acts somewhat like astep function on R.

(We define J to be zero outside T.) If we think of J as an actual stepfunction, then the double integral of IJI over R (and hence over T) is a sum of productsof the form IJ(u, v)l Au Au and the above remarks suggest that this sum is nearly equal tothe area of S, which we know to be the double integral js dx dy. This geometric discussion, which merely suggests why'we might expect an equation like(11.33) to hold, can be made the basis of a rigorous proof, but the details are lengthy andrather intricate. As mentioned above, a proof of (11.33), based on an entirely differentapproach, will be given in a later section. If J(u, v) = 0 at a particular point (u, v), the two vectors VI and V, are parallel (sincetheir cross product is the zero vector) and the parallelogram degenerates into a line segment.Such points are called singular points of the mapping. As we have already mentioned,transformation formula (11.32) is also valid whenever there are only a finite number of suchsingular points or, more generally, when the singular points form a set of content zero.This is the case for all the mappings we shall use. In the next section we illustrate the use offormula (11.32) in two important examples. 11.27 Special cases of the transformation formula EXAMPLE 1. Polar coordinates.

In this case we write r and 0 instead of u and v anddescribe the mapping by the two equations: x = rcose, y = r sin e , Change of variables in a double integral- 397 That is, X(r, 0) = r cos 6' and Y(r, 0) = r sin 0. To obtain a one-to-one mapping wekeep r > 0 and restrict 0 to lie in an interval of the form I!& 5 8 < 19~ + 2~. For example,the mapping is one-to-one on any subset of the rectangle (0, a] x [0, 2~) in the r0-plane.The Jacobian determinant of this mapping is ax ayarar cos 8 sin e J(r, e) =ax ay = -rsinO rcose = r(cos" e + sin2 e) = r . asae Hence the transformation formula in (11.32) becomes ljf(x, y) dx dy = sjj(r cos 0, r sin 0) r dr de.T The r-curves are straight lines through the origin and the &curves are circles centered at theorigin. The image of a rectangle in the r&plane is a "parallelogram" in the xy-plane I9 I Ie = constant Y x = rcose y = r sin ec ._____-___, 1I III III + r = constant I 0I cr - 0 r-curve (0 = constant) FIGURE 11.26 Transformation by polar coordinates. bounded by two radial lines and two circular arcs, as shown in Figure 11.26. The Jacobiandeterminant vanishes when r = 0, but this does not affect the validity of the transformationformula because the set of points with r = 0 has content zero. Since V, = cos 8 i + sin ej, we have II VI II = 1, so there is no distortion of distances alongthe r-curves. On the other hand, we have V,= -rsin8i+rcosBj, IIV211 = r, so distances along the e-curves are multiplied by the factor r.Polar coordinates are particularly suitable when the region of integration has boundaries along which r or 0 is constant. For example, consider the integral for the volume of one 398 MuItiple integrals octant of a sphere of radius a, 0 Ja2 - x2 - y"dx dy,s where the region S is the first quadrant of the circular disk x2 + y2 < a2. In polar co-ordinates the integral becomes JP a2 - r2 r dr d&i,T where the region of integration T is now a rectangle [0, a] x [0, in]. Integrating first withrespect to 8 and then with respect to r we obtain Ja2 - r2 r dr d0 = iI T Oa rJ= dr = f .

(a" T:'% /a- 7;' .0 The same result can be obtained by integrating in rectangular coordinates but the calculationis more complicated. EXAMPLE 2. Linear transformations.

Consider a linear transformation defined by a pair ofequations of the form (11.34) x = Au + Bv, y=Cu+ Dv, where A, B, C, D are given constants. The Jacobian determinant is J(u, v) = AD - BC, and in order to have an inverse we assume that AD - BC # 0. This assures us that thetwo linear equations in (11.34) can be solved for u and v in terms of x and y. Linear transformations carry parallel lines into parallel lines. Therefore the image ofa rectangle in the uv-plane is a parallelogram in the xy-plane, and its area is that of therectangle multiplied by the factor lJ(u, v)] = IAD - BCI . Transformation formula (11.32)becomes jjj(x, y) dx dy = IAD - BCI jjf(At.4 + Bv, Cu + Dv) du dv.S T To illustrate an example in which a linear change of variables is useful, let us considerthe integral sse(Y-X)/(Y+X) dx dy , S where S is the triangle bounded by the line x + y = 2 and the two coordinate axes. (SeeFigure 11.27.) The presence of y - x and y + x in the integrand suggests the change ofvariables u=y-x, v = y + x . Exercises 399 FIGURE 11.27 Mapping by a linear transformation. Solving for x and y we find U-U v+uXC- and2 y=---2 * The Jacobian determinant is J(u, v) = -4. To find the image T of S in the uu-plane wenote that the lines x = 0 and y = 0 map onto the lines u = v and u = -v, respectively;the line x + y = 2 maps onto the line u = 2. Points inside S satisfy 0 < x + y < 2 andthese are carried into points of T satisfying 0 < v < 2. Therefore the new region ofintegration T is a triangular region, as shown in Figure 11.27. The double integral inquestion becomes 1e(~-X)/(%'+X) dx dy = _ e"" du dv . Integrating first with respect to u we find 11.28 Exercises In each of Exercises 1 through 5, make a sketch of the region S and express the double integral.l/ f(x, y) dx dy as an iterated integral in polar coordinates. 1. s = {(x,y)] x2 +y2 I a2}, where a>O.2. s = ((x,y) 1 x2 +y2 2 2x).3. s = {(x,y> Iu2 I x2 +y2 I b2}, where 0 ls [I0 -f(x, y) dy] dx.1-X 11. j;[jf"fC~x' + yz)dy] dx. 13. j;[j~'f(x,yWy] dx. 14. Use a suitable linear transformation to evaluate the double integral IS (x - y)" sin2 (x + y) dx dyS where S is the parallelogram with vertices (71, 0), (2~, n), (r, 2~), (0, n).15. A parallelogram S in the xy-plane has vertices (O,O), (2, lo), (3, 17), and (1, 7). (a) Find a linear transformation u = ax + by, u = cx + dy, which maps S onto a rectangleR in the uu-plane with opposite vertices (0,O) and (4,2). The vertex (2, 10) should map ontoa point on the u-axis.(b) Calculate the double integral IJ xy dx dy by transforming it into an equivalent integralover the rectangle R of part (a). S 16.

If r > 0, let Z2(r) = j~7e-U2du.(a) Show that Z2(r) = s.f e-rze+Y2) dx dy , where R is the square R = [-r, r] x [-r, r]. R (b) If C, and C, are the circular disks inscribing and circumscribing R, show that jj e-(r2+v2) dx dy < Z2(r) < 11 e-@+@) dx dy . C l ca (c) Express the integrals over C1 and C, in polar coordinates and use (b) to deduce thatZ(r) + JG as r - 00 . This proves that j; e-u2 du = ,/n/2. (d) Use part (c) to deduce that I'($) = &, where I is the gamma function.17. Consider the mapping defined by the equations x=u+u, y = v -2. (a) Compute the Jacobian determinant J(u, a).(b) A triangle T in the uv-plane has vertices (0, 0), (2,0), (0,2). Describe, by means of asketch, its image S in the xy-plane.(c) Calculate the area of S by a double integral extended over S and also by a double integralextended over T.(d) Evaluate 5 1 (x - y $ l) * dx dy . 18.

Consider the kapping defined by the two equations x = u2 - v2. y = 2uv.(a) Compute the Jacobian determinant J(u, V) . Proof of the transformation formula in a special case 401 (b) Let T denote the rectangle in the uu-plane with vertices (1, l), (2, I), (2, 3), (1, 3). De-scribe, by means of a sketch, the image S in the xy-plane.(c) Evaluate the double integral JJ xy dx dy by making the change of variables x = u2 - u*, y=2uv,whereC={(x,y)/x2~y2<1}.19. Evaluate the double integral I@, r) =dx dy (p" + x2 + y2)PR over the circular disk R = {(x, y) 1 x2 + y2 5 r"} .

Determine those values of p for whichI@, r) tends to a limit as r + + co. In Exercises 20 through 22, establish the given equations by introducing a suitable change ofvariables in each case. 20. jjfb +y)dxdy = j -lfWu, where S = {(x, y) 1 I4 + lyl 5 1).s 21.

jjf(ax + by + c) dx dy = 2 j!1 ~-~f(~&z" + b2 + c) du,s where S = {(x, y) 1x2+y2j1} a n d a2+b2#0. 22. jjf(xy) dx dy = log 2 s,2f(u) du, where S is the region in the first quadrant bounded byS thecurvesxy = 1,xy =2,y =x,y =4x. 11.29 Proof of the transformation formula in a special case As mentioned earlier, the transformation formula (11.35) j jfk Y> dx dy = j j f[X(u, 01, Vu, @I Mu, 91 du du T can be deduced as a consequence of the special case in which S is a rectangle and f isidentically 1. In this case the formula simplifies to (11.36) jj dx dy = jj IJ(u, u)l du du.R R* Here R denotes a rectangle in the xy-plane and R* denotes its image in the Mu-plane (seeFigure 11.28) under a one-to-one mapping u = Uky), 2, = V(x,y). The inverse mapping is given by x = X(24, u), y = Y(u,v), 402 Multiple integrals Y VA FIGURE 11.28 The transformation law for double integrals derived from Green's theorem. and J(u, v) denotes the Jacobian determinant, .I(u, v) = ax axZ JaUi3Y aYau-z In this section we use Green's theorem to prove (11.36), and in the next section we deducethe more general formula (11.35) from the special case in (11.36). For the proof we assume that the functions Xand Y have continuous second-order partialderivatives and that the Jacobian is never 0 in R *.

ThenJ(u, v) is either positive everywhereor negative everywhere. The significance of the sign of J(u, v) is that when a point (x, y)traces out the boundary of R in the counterclockwise direction, the image point (u, v)traces out the boundary of R* in the counterclockwise direction if J(u, v) is positive and inthe opposite direction if J(u, v) is negative. In the proof we shall assume that J(u, v) > 0. The idea of the proof is to express each double integral in (11.36) as a line integral, usingGreen's theorem.

Then we verify equality of the two line integrals by expressing each inparametric form. We begin with the double integral in the xy-plane, writing j 1dxdy=j Je-$) dxdy, R R where Q(x, y) = x and P(x, y) = 0. By Green's theorem this double integral is equal tothe line integral Here C is the boundary of R traversed in a counterclockwise direction. Proof of the transformation formula in the general case 403 Similarly, we transform the double integral in the uv-plane into a line integral around theboundary C* of R*. The integrand, J(u, o), can be written as follows: axar axar axar avI(u,v)=-----=--+x--Xa2y_aXaU au au au au au au avau au au -gq). Applying Green's theorem to the double integral over R* we find J(u, v) du dv =R * Therefore, to complete the proof of (11.36) we need only verify that (11.37) We introduce a parametrization of C* and use this to find a representation of C. SupposeC* is described by a function a defined on an interval [a, b], say Leta(t) = U(t)i + V(t)j. P(t) = m-J(t), v(w + m.J(t), I'(f>lj. Then as t varies over the interval [a, b], the vector a(t) traces out the curve C* and g(t)traces out C. By the chain rule, the derivative of @ is given by Hence P'(t) = Lxdy=L (au)X[U(t), V(t)] !?- U'(t) + $: V'(t) dt. The last integral over [a, b] is also obtained by parametrizing the line integral over C* in(11.37). Therefore the two line integrals in (11.37) are equal, which proves (11.36). 11.30 Proof of the transformation formula in the general case In this section we deduce the general transformation formula (11.38) j j-Sk Y) dx dy = j j-f[X[u, V), Vu, v>] IJ(u, VII du dvs T 404 Multiple integrals from the special case treated in the foregoing section, (11.39) j j dx dy = j j IJ(u, v>l du dv,R R* where R is a rectangle and R* is its image in the Mu-plane.First we prove that we have (11.40) ss s(x, y) dx dy = IS s[X(u, 4, Vu, v>l IJ(u, 4 du dv,R R * where s is any step function defined on R. For this purpose, let P be a partition of R intomn subrectangles Rij of dimensions Axi and Ayj, and let cij be the constant value that stakes on the open subrectangle Rij . Applying (11.39) to the rectangle Rij we find Axi Ayj = /I dx dy = /I IJ(Uy V)l du dv.Ril Rij* Multiplying both sides by ci, and summing on i and j we obtain (11.41) Since s is a step function, this is the same as (11.42) izl jzlci, Axi Ayj = i 2 jj S[X(u> V>, Y(u, v>I I J(u> ')I du dv.i=lj=l Ri,* Using the additive property of double integrals we see that (11.42) is the same as (11.40).Thus, (11.40) is a consequence of (11.39). Next we show that the step function s in (11.40) can be replaced by any function f forwhich both sides of (11.40) exist. Let f be integrable over a rectangle R and choose stepfunctions s and t satisfying the inequalities (11.43) 4x, Y> 5 f(x, y> < t(x, y), for all points (x, y) in R. Then we also have (11.44) SF(~ u>, Y(u, UN , Y(u, v>l < W(u, v), Y(u, u)l for every point (u, u) in the image R*. For brevity, write S(u, v) for s[X(u, v), Y(u, u)] anddefine F(u, v) and T(u, v) similarly. Multiplying the inequalities in (11.44) by [J(u, v)]and integrating over R* we obtain jj S(u, v> IJ(u, 41 du dv I jJ F(u, 4 IJ(u, v>l du dv I jj T(u, u> IJ(u, v>l du du.R * R * R * Extensions to higher dimensions 405 Because of (11.40), the foregoing inequalities are the same as jj S(x, d dx dy_< jj Fk 4 LO, u)l du do I 11 t(x, Y> dx dy .R R * R Therefore jj P(u, u)]J(u, II)] du dv is a number which lies between the integrals jS s(x, y) dx cand j[t(x, y) dx dy for all choices of step functions s and t satisfying (11.43). &rce f is integrable,?his implies that i/f@, Y) dx dy = j j F(u, u> Mu, 41 du duR' and hence (11.38) is valid for integrable functions defined over rectangles.Once we know that (11.38) is valid for rectangles we can easily extend it to more general regions S by the usual procedure of enclosing S in a rectangle R and extending the functionf to a new functionfwhich agrees with f on S and has the value 0 outside S. Then we notethat jjf= jjf= jjfLW v>, Y(u, 41 Mu, v>l du dv = jj F(u, v> IJ(u, v>l du doS R R* T and this proves that (11.38) is, indeed, a consequence of (11.39). 11.31 Extensions to higher dimensions The concept of multiple integral can be extended from 2-space to n-space for any n 2 3. Since the development is entirely analogous to the case n = 2 we merely sketch the principalresults. The integrand is a scalar field f defined and bounded on a set S in n-space. The integraloff over S, called an n-fold integral, is denoted by the symbols s 1. . . f, or s I... f(xl,...,xn:)dxl**.dx,,9 S with n integral signs, or more simply with one integral sign, Js f(x) dx, where x =(-%,***,x,). When n = 3 we write (x, y, 7) instead of (x1, x2, x3) and denote tripleintegrals by JJI f, or .Tf l f(x, y, z) dx dy dz .s S First we define the n-fold integral for a step function defined on an n-dimensional interval.We recall that an n-dimensional closed interval [a, b] is the Cartesian product of n closedone-dimensional intervals [a,, bk], where a = (a,, . . . , a,) and b = (b,, . . . , h,). Ann-dimensional open interval (a, 6) is the Cartesian product of n open intervals (ak, bk).The volume of [a, b] or of (a, 6) is defined to be the product of the lengths of the componentintervals, (b, - a,) * * * (b, - a,). 406 Multiple integrals IfPI,. . . , P, are partitions of [a,, b,], .

. . , [a,, b,], respectively, the Cartesian productP=P, X"' x P, is called a partition of [e, b].

A functionfdefined on [a, b] is called a stepfunction if it is constant on each of the open subintervals determined by some partition P.The n-fold integral of such a step function is defined by the formula where ci is the constant value thatftakes on the ith open subinterval and vi is the volume ofthe ith subinterval. The sum is a finite sum extended over all the subintervals of P. Having defined the n-fold integral for step functions, we define the integral for moregeneral bounded functions defined on intervals, following the usual procedure. Let s and tdenote step functions such that s 0 there is a finite collectionof n-dimensional intervals whose union includes S and the sum of whose volumes does notexceed E. To define the n-fold integral of a bounded function f over a more general bounded set S,we extend f to a new functionfwhich agrees with f on S and has the value zero outside S;the integral off over S is defined to be the integral of -? over some interval containing S. Some multiple integrals can be calculated by using iterated integrals of lower dimension.For example, suppose S is a set in 3-space described as follows: (11.45) S = 1(x, y, 4 1 (x, r> E Q and R(-G Y> I z I pdx, ~91, where Q is a two-dimensional region, called the projection of S on the xy-plane, and pl,Q;~ are continuous on S. (An example is shown in Figure 11.29.) Sets of this type are boundedby two surfaces with Cartesian equations z = ql(x,v), and z = qz(x,y) and (perhaps) aportion of the cylinder generated by a line moving parallel to the z-axis along the boundaryof Q.

Lines parallel to the z-axis intersect this set in line segments joining the lower surfaceto the upper one.

Iff is continuous on the interior of S we have the iteration formula (11.46) j j j Rx, Y, z> dx dy dz = j j [j<;;;:;;; Rx, Y, z> dz] dx dy .S & Change of variables in an n-fold integral 407 FIGURE 11.29 A solid S and its projection Q in the xy-plane. That is, for fixed x andy, the first integration is performed with respect to z from the lowerboundary surface to the upper one. This reduces the calculation to a double integral over theprojection Q, which can be treated by the methods discussed earlier.

There are two more types of sets analogous to those described by (11.45) in which the x-'and y-axes play the role of the z-axis, with the projections taken in the yz- or xz-planes,respectively. Triple integrals over such sets can be computed by iteration, with formulasanalogous to (11.46). Most 3-dimensional sets that we shall encounter are either of one ofthe three types just mentioned or they can be split into a finite number of pieces, each ofwhich is of one of these types. Many iteration formulas exist for n-fold integrals when n > 3. For example, if Q is ak-dimensional interval and R an m-dimensional interval, then an (m + ,$)-fold integral overQ x R is the iteration of an m-fold integral with a k-fold integral, j*.jf = j-s j [j- jfdx,v.fx,] dx, +l.-dx,+k,QxR Q R provided all the multiple integrals in question exist. Later in this chapter we illustrate theuse of iterated multiple integrals in computing the volume of an n-dimensional sphere. 11.32 Change of variables in an n-fold integral The formula for making a change of variables in a double integral has a direct extensionfor n-fold integrals. We introduce new variables ul, . . . , u, related to x1, . . . , x, by nequations of the form x,=X,(u, ...) UJ, . . .) x, = XJU,, . . .) UJ. 408 Multiple irltegrals Let x = (-vi, . . . , .xn), u = (u,, . . . , un), and X = (X,, . . . , X,,). Then these equationsdefine a vector-valued mapping X: T -+S from a set Tin n-space to another set S in n-space. We assume the mapping X is one-to-oneand continuously differentiable on T. The transformation formula for n-fold integralsassumes the form (11.47) /J(x) dx = ~JLWN ldet DW4I du, where DX(u) = [D,X,(u)] is the Jacobian matrix of the vector field X. In terms of com-ponents we have DX(u) = 1: * 1 I*&x&) &G(u) * * * ~,Uu) As in the two-dimensional case, the transformation formula is valid if Xis one-to-one onT and if the Jacobian determinant J(u) = det DX(u) is never zero on T. It is also valid ifthe mapping fails to be one-to-one on a subset of T having n-dimensional content zero, or ifthe Jacobian determinant vanishes on such a subset. For the three-dimensional case we write (x, y, z) for (xi, x2, xg), (u, u, w) for (ur , u2, us),and (X, Y, Z) for (X1, X2, X,).

The transformation formula for triple integrals takes theform (11.48) jjjfb, Y, z> dx dy dzS =Iss f[X(u, 0, w>, Y(u, v, w), Z(u, v, $1 Mu, v, w>l du dv dw , T where J(u, v, PC) is the Jacobian determinant, J(u, v, w) = ax au azzauauax ar azzauavax ay azawawaw In 3-space the Jacobian determinant can be thought of as a magnification factor for volumes.In fact, if we introduce the vector-valued function r defined by the equation r(z4, 2), w) = X(u, u, w)i + Y(u, u, w)j + Z(u, u, w)k, Worked examples 409 and the vectors pl =ar=ax;+auj+az,2 au au au azl ' an argument similar to that given in Section 11.26 suggests that a rectangular parallelepipedof dimensions Au, Au, Aw in uvw-space is carried onto a solid which is nearly a curvilinear"parallelepiped" in xyz-space determined by the three vectors VI Au, V, Au, and V, Aw.(See Figure 11.30.) The boundaries of this solid are surfaces obtained by setting u =constant, v = constant, and w = constant, respectively. The volume of a parallelepipedis equal to the absolute value of the scalar triple product of the three vectors which determineit, so the volume of the curvilinear parallelepiped is nearly equal to I(Vl Au). (V, Au) x (V,Aw]l = IV, -V, x V,I Au Au Aw = IJ(u,v, w)l Au Au Aw. 11.33 Worked examples Two important special cases of (11.48) are discussed in the next two examples. EXAMPLE 1. Cylindrical coordinates. Here we write r, 0, z for u, v, w and define themapping by the equations (11.49) x =rcos6, y = r sin 8, z = z. ;A z = constant7---- --____|_ / x = r cos 0'A| . y = r sin 0 ;; ; - __ _ z ';;z=z *. 'L,;; r = constant ep /y 0 = constant r x FIGURE 11.30 Transformation by cylindrical coordinates. 410 Multiple integrals In other words, we replace x and y by their polar coordinates in the xy-plane and leavez unchanged. Again, to get a one-to-one mapping we must keep r > 0 and restrict e tobe in an interval of the form 8, 5 6 < 8, + 27~. Figure 11.30 shows what happens to arectangular parallelepiped in the r&-space. The Jacobian determinant of the mapping in (11.49) is cog 8 sin 8 0 J(r, 8, Z) = -r sin e r cos 8 0 = r(cosz e + sin2 e) = r, 0 0 1 and therefore the transformation formula in (11.48) becomes jjj f(x, y, z) dx dy dz = j/j f(r cos 0, r sin 8, z)r dr dt3 dz .T The Jacobian determinant vanishes when r = 0, but this does not affect the validity of thetransformation formula because the set of points with r = 0 has 3-dimensional content 0.

EXAMPLE 2. Spherical coordinates. In this case the symbols p, 8, q are used instead ofu, v, w and the mapping is defined by the equations X= pcosesing,, y=psinesinq, z = pcos q3. The geometric meanings of p, 8, and p are shown in Figure 11.31. To get a one-to-onemapping we keep p > 0, 0 < 6 < 27r, and 0 < pl < rr. The surfaces p = constant arespheres centered at the origin, the surfaces 8 = constant are planes passing through thez-axis, and the surfaces q = constant are circular cones with their axes along the z-axis.Therefore a rectangular box in pQ-space is mapped onto a solid of the type shown inFigure 11.31. P z x = p sin Y cos Ry = p sin P sin 8z = p cos cp t 0 = constantP = constant kYY4|*11 i+ P),,';,/q /' III Q1= constant ,' 1/+0 . /"-p)\, 1 cos (p X p sin p |.J FIGURE 11.31 Transformation by spherical coordinates. Worked examples 411 The Jacobian determinant of the mapping is cos 19 sin q2 sin 0 sin q2 cos fp J(p,O,q)= -psin8sinp pcosesinp 0 = -p2sing,. p cos e cos qI psin8cospl -psinpl Since sin q 2 0 if 0 5 v < r, we have IJ(p, 8, v)] = p2 sin pl and the formula for trans-forming triple integrals becomes jjjf (x, Y, 4 dx dy dz = jjj F(P, 0, Y)P' sin P dp do dp,,T where F(p, 9, q) =f(p cos 8 sin q, p sin f3 sin q, p cos cp). Although the Jacobian deter-minant vanishes when p = 0 the transformation formula is still valid because the set ofpoints with v = 0 has 3-dimensional content 0. The concept of volume can be extended to certain classes of sets (called measurable sets)in n-space in such a way that if S is measurable then its volume is equal to the integral of theconstant function 1 over S. That is, if u(S) denotes the volume of S, we have v (S) = ... dxI...dx,.i ss We shall not attempt to describe the class of sets for which this formula is valid. Instead,we illustrate how the integral can be calculated in some special cases. EXAMPLE 3. Volume of' an n-dimensional interval. If S is an n-dimensional interval, sayS = [a1,hl x 0.. x [a,, b,], the multiple integral for c(S) is the product of n one-dimensional integrals, v(S) = jGy dx, * . 1 ja: dx, = (b, - a,). * * (b, - a,). This agrees with the formula given earlier for the volume of an n-dimensional interval. EXAMPLE 4.

The volume of an n-dimensional sphere. Let S,(a) denote the n-dimensionalsolid sphere (or n-ball) of radius a given by and letS,(a) = {(x,, . . . , x,) 1 x", + . * * + x", 5 a"}, V,(a) = j;n;;) 1 dx, * . * dx, , the volume of L?,(a). We shall prove that (11.50) V,(a) =~ n/2 r(+n + 1) "" 412 Multiple integrals where I? is the gamma function. For n = 1 the formula gives V,(a) = 2a, the length of theinterval [-a, a]. For n = 2 it gives V,(a) = vu2, the area of a circular disk of radius a.We shall prove (11.50) for n 2 3. First we prove that for every a > 0 we have (11.51) V,(a) = a"V,(I). In other words, the volume of a sphere of radius a is an times the volume of a sphere ofradius 1. To prove this we use the linear change of variable x = au to map S,(l) ontoS,(a). The mapping has Jacobian determinant an. Hence V,(a) = f . . . 1 dxI. . . dx, = j . .

3 j an du, . . . du, = a"V,(l).S,(a) &fl) This proves (11.51). Therefore, to prove (11.50) it suffices to prove that (11.52) 1/,(l) =$,piz Iy+n + 1) . First we note that x: + * . . + xz 5 1 if and only if xf + . . * + xi-2 I 1 - x:-1 - xi and xi-1 + x: I 1. Therefore we can write the integral for Vn(l) as the iteration of an (n - 2)-fold integral anda double integral, as follows: (11.53) V,(l) = j j [j . . . j dx, .

. . dx, ;l dx,hI dx, . z:-1+r2ni12 2 x*+.'.+z,-2"1--x~-1-~~ The inner integral is extended over the (n - 2)-dimensional sphere S,-,(R), where R = X/I - xiP1 - xi, so it is equal to K,(R) = 2R"-2V&1) = (1 - x;dl - x,) n'2-1 V+,(l) . Now we write x for x, ~ and y for x, . Then (11.53) becomes 1/,(l) = v,-2u> Ji (1z=+Y2 5 1 - x2 - ~~)~"-l dx dy . We evaluate the double integral by transforming it to polar coordinates and obtain 2R1/,(l) = V, ,(l)ss ' (1 - r2)n'2-1 Y dr d0 = V+,(l) e .0 0 n In other words, the numbers V,(l) satisfy the recursion formula Vn(l) = h v,-2(l) ifn 2 3 .n Exercises But the sequence of numbers {f(n)} defined by 413 f(n) =&2 Wn + 1) satisfies the same recursion formula because I'(s + 1) = SF(S). Also, I'(+) = Ji (see Exercise 16, Section 11.28), so I?(+) = &J nandf(1) = V,(l) = 2. Also, f(2) = Vz(1) = n,hence we havef'(n) = I',(l) for all n > 1. This proves (11.52). 11.34 Exercises Evaluate each of the triple integrals in Exercises 1 through 5. Make a sketch of the region ofintegration in each case. You may assume the existence of all the integrals encountered. I. JJJxy2z3dxdydz,s where S is the solid bounded by the surface z = xy and the planes y =x,x = 1,andz = O .2. JAJ (1 +x + y + ~) ~dxdydz, where S is the solid bounded by the three coordinate planes and the plane x + y + z = I.3. JJJxyzdxdydz, where S = {(x, y, z) 1xz+yz+zzI1,x20,y20,z20}. 6x2 2 22a2 +$ +2 dxdydz, where S is the solid bounded by the ellipsoid X2a2+$+;=l. 5. JAJ Jx" + y2 dx dy dz, where S is the solid formed by the upper nappe of the cone z2 =x2 +y2andtheplanez = 1. In Exercises 6,7, and 8, a triple integral J J J f(x, y, z) dx dy dz of a positive function reduces to the iterated integral given. In each caseSdescribe the region of integration S by means of asketch, showing its projection on the xy-plane. Then express the triple integral as one or moreiterated integrals in which the first integration is with respect toy. 6. j;(j;-"[j;+"f(x,y, z)dz]dy) dx.

7.

j'l,(j~~~~[j:,,f(x~~.z)dz] dv)dx. 8. j;(I,' [jr2 fb, y, z> dz] dy) dx . 9. Show that: Evaluate the integrals in Exercises IO, 11, and I2 by transforming to cylindrical coordinates. Youmay assume the existence of all integrals encountered. 10.

JJJ (x2 + y2) dx dy dz, where S is the solid bounded by the surface x2 + y2 = 2z and the pfilanez = 2. 414 Multiple integrals 11. jjjdxdydz, where S is the solid bounded by the three coordinate planes, the surface :=x2+y2,andtheplanex +y = 1.12. jjj (y" + r2) dx dy dz, where S is a right circular cone of altitude h with its base, of radius Sa, in the xy-plane and its axis along the z-axis. Evaluate the integrals in Exercises 13, 14, and 15 by transforming to spherical coordinates. 13. jjj dx dy dz, where S is a solid sphere of radius a and center at the origin. 14. ;j dx dy dz, where S is the solid bounded by two concentric spheres of radii a and b,S where 0 < a < b, and the center is at the origin.15. jjj [(x - a)2 + (y - b)2 + (z - c)2 ~] -% dx dy dz, where S is a solid sphere of radius R sand center at the origin, and (a, b, c) is a fixed point outside this sphere. 16. Generalized spherical coordinates may be defined by the following mapping: x = ap cos" 0 sin" p, y = bp sin" 0 sinn p, z = cp COP qJ, where a, b, c, m, and n are positive constants. Show that the Jacobian determinant is equal to -abump cos m--1 0 sin"-1 0 cos"-l p sin2+l pl. Triple integrals can be used to compute volume, mass, center of mass, moment of inertia, andother physical concepts associated with solids. If S is a solid, its volume V is given by the tripleintegral v =SIS dxdydz.S If the solid is assigned a densityf(x, y, z) at each of its points (x, y, z) (mass per unit volume), itsmass M is defined to be &f= -jisf(x,y,z)dxdydz, s and its center of mass the point (2, J, Z) with coordinates 1xc- M xf(x,y,ddxdydz, S and so on. The moment of inertia ZXy about the xy-plane is detined by the equation L = SII z"f (x, y, z) dx dy dzS and similar formulas are used to define Zzlz and Z,, .

The moment of inertia IL about a line L isdetined to be IL =0 J S2(x, y, df(x, y, z> dx dy dz, S Exercises 4 1 5 where 6(x, y, z) denotes the perpendicular distance from a general point (x, y, z) of S to the line L.

17. Show that the moments of inertia about the coordinate axes are 4! = L + L 9 4! = z,, + z,2, z, = zzz + z,, .181 Find the volume of the solid bounded above by the sphere x2 + y2 + z2 = 5 and below by the paraboloid x2 + y2 = 42.19.

Find the volume of the solid bounded by the xy-plane, the cylinder x2 + y2 = 2x, and the cone z = J-x2 + y2.20. Compute the mass of the solid lying between two concentric spheres of radii a and b , where 0 < a < b , if the density at each point is equal to the square of the distance of this point fromthe center. 21. A homogeneous solid right circular cone has altitude h.

Prove that the distance of its centroidfrom the base is h/4. 22. Determine the center of mass of a right circular cone of altitude h if its density at each pointis proportional to the distance of this point from the base. 23. Determine the center of mass of a right circular cone of altitude h if i.ts density at each point isproportional to the distance of this point from the axis of the cone. 24. A solid is bounded by two concentric hemispheres of radii a and b, where 0 < a < b. If thedensity is constant, find the center of mass. 25. Find the center of mass of a cube of side h if its density at each point is proportional to thesquare of the distance of this point from one corner of the base. (Take the base in the xy-planeand place the edges on the coordinate axes.) 26. A right circular cone has altitude h, radius of base a, constant density, and mass M. Find itsmoment of inertia about an axis through the vertex parallel to the base. 27.

Find the moment of inertia of a sphere of radius R and mass M about a diameter if the densityis constant. 28. Find the moment of inertia of a cylinder of radius a and mass M if its density at each point isproportional to the distance of this point from the axis of the cylinder. 29.

The stem of a mushroom is a right circular cylinder of diameter 1 and length 2, and its capis a hemisphere of radius R. If the mushroom is a homogeneous solid with axial symmetry,and if its center of mass lies in the plane where the stem joins the cap, find R. 30. A new space satellite has a smooth unbroken skin made up of portions of two circular cylindersof equal diameters D whose axes meet at right angles. It is proposed to ship the satellite toCape Kennedy in a cubical packing box of inner dimension D. Prove that one-third of the box will be waste space.31. Let S,(u) denote the following set in n-space, where a > 0: When n = 2 the set is a square with vertices at (0, +ZU) and (fu, 0). When n = 3 it is anoctahedron with vertices at (0, 0, &.a), (0, fu, 0), and (fu, 0,O).

Let l',(u) denote the volumeof S,(u), given by V,(u)=j... jdx,...dx,.f&(a) (a) Prove that V,(u) = @V,(l).(b) For IZ 2 2, express the integral for V,(l) as an iteration of a one-dimensional integral andan (n - 1)-fold integral and show that s 1 V,(l) = V,-,(l)-1 (1 - l~l)~--l dx = ; V,,(l). 416 32. (c) Use parts (a) and (b) to deduce that V,(a) = 7. Let S,(a) denote the following set in n-space, where a > 0 and n 2 2 : &(a> = {(Xl, * .

. ,x,) 1 lxil + Ix,1 5 a for each i = 1,. . . , n - 1}. (a) Make a sketch of S,(l) when n = 2 and when n = 3.(b) Let V,(a) = j .

. . j dx, . . . dx,, and show that V,(u) = anVn(l) . S,(a) 33. (c) Express the integral for V,(l) as an iteration of a one-dimensional integral and an (n - l)-fold integral and deduce that V,(u) = 2W/n.(a) Refer to Example 4, p. 411.

Express the integral for V,(l), the volume of the n-dimensionalunit sphere, as the iteration of an (n - l)-fold integral and a one-dimensional integral andthereby prove that Multiple integrals V,(l) = 2 I'&l)s l (1 - x2)(+-1)'2 dx.0 (b) Use part (a) and Equation (11.52) to deduce that 12SURFACE INTEGRALS 12.1 Parametric representation of a surface This chapter discusses surface integrals and their applications. A surface integral can bethought of as a two-dimensional analog of a line integral where the region of integration is asurface rather than a curve. Before we can discuss surface integrals intelligently, we mustagree on what we shall mean by a surface. Roughly speaking, a surface is the locus of a point moving in space with two degreesof freedom. In our study of analytic geometry in Volume I we discussed two methodsfor describing such a locus by mathematical formulas. One is the implicit representationin which we describe a surface as a set of points (x, y, z) satisfying an equation of theform F(x, y, z) = 0. Sometimes we can solve such an equation for one of the coordinatesin terms of the other two, say for z in terms of x and y. When this is possible we obtainan explicit representation given by one or more equations of the form z = f (x, y) . Forexample, a sphere of radius 1 and center at the origin has the implicit representationx2 + y2 + z2 - 1 = 0. When this equation is solved for z it leads to two solutions, z=Jl -x2- y2and z= -41 -x2-y2. The first gives an explicit representationof the upper hemisphere and the second of the lower hemisphere. A third method for describing surfaces is more useful in the study of surface integrals;this is the parametric or vector representation in which we have three equations expressingx, y, and z in terms of two parameters u and v: (12.1) x = X(u, v), y = Y(u, v) > z = Z(u, v). Here the point (u, v) is allowed to vary over some two-dimensional connected set T inthe #v-plane, and the corresponding points (x, y, z) trace out a surface in xyz-space. Thismethod for describing a surface is analogous to the representation of a space curve by threeparametric equations involving one parameter. The presence of the two parameters in(12.1) makes it possible to transmit two degrees of freedom to the point (x, y, z), as suggestedby Figure 12.1. Another way of describing the same idea is to say that a surface is the imageof a plane region T under the mapping defined by (12.1).

If we introduce the radius vector r from the origin to a general point (x, y, z) of thesurface, we may combine the three parametric equations in (12.1) into one vector equation 417 Surface integrals FIGURE 12.1 Parametric representation of a surface. of the form (12.2) r(u, v) = X(u, u>i + Y(u, u)j + Z(u, zl)k, where (u, v) E T. This is called a vector equation for the surface.There are, of course, many parametric representations for the same surface. One of these can always be obtained from an explicit form z = f(x, y) by taking X(u, v) = U,Y(u, u) = u, Z(u, U) = f(~, u). On the other hand, if we can solve the first two equationsin (12.1) for u and u in terms of x and y and substitute in the third-we obtain an explicitrepresentation z = f(x, y) . EXAMPLE 1.

A parametric representation of a sphere. The three equations (12.3) x = acosucosv, y = a s i n u c o s v , z = asinu FIGURE 12.2 Parametric representation of a surface D FIGURE 12.3 Deformation of a rectangle into a hemisphere. serve as parametric equations for a sphere of radius a and center at the origin. If wesquare and add the three equations in (12.3) we find x2 + y2 + z2 = a2, and we see thatevery point (x, y, z) satisfying (12.3) lies on the sphere. The parameters u and u in thisexample may be interpreted geometrically as the angles shown in Figure 1 2.2. If we letthe point (u, u) vary over the rectangle T = [0, 2x] x [-1 zz-, +T], the points determinedby (12.3) trace out the whole sphere. The upper hemisphere is the image of the rectangle[0, 2~1 x [0, &T] and the lower hemisphere is the image of [0,2~] x [-&T, 01. Figure 12.3gives a concrete idea of how the rectangle [0, 2~1 x [0, 3x1 is mapped onto the upperhemisphere. Imagine that the rectangle is made of a flexible plastic material capable ofbeing stretched or shrunk. Figure 12.3 shows the rectangle being deformed into a hemi-sphere. The base AB eventually becomes the equator, the opposite edges AD and BC arebrought into coincidence, and the upper edge DC shrinks to a point (the North Pole). EXAMPLE 2. A parametric representation of a cone. The vector equation Y(U, u) = 2, sin M cos 24 i + u sin CI sin uj + v cos tc k represents the right circular cone shown in Figure 12.4, where u denotes half the vertexangle. Again, the parameters u and v may be given geometric interpretations; u is thedistance from the vertex to the point (x, y, z) on the cone, and u is the polar-coordinateangle. When (u, v) is allowed to vary over a rectangle of the form [0, 2~1 x [0, h], thecorresponding points (x, y, z) trace out a cone of altitude h cos CC. A plastic rectangle FIGURE 12.4 Parametric representation of a FIGURE 12.5 Deformation of a rectanglecone. into a cone. 420 Surface integrals may be physically deformed into the cone by bringing the edges AD and BC into coin-cidence, as suggested by Figure 12.5, and letting the edge AB shrink to a point (the vertexof the cone). The surface in Figure 12.5 shows an intermediate stage of the deformation. In the general study of surfaces, the functions X, Y, and 2 that occur in the parametricequations (12.1) or in the vector equation (12.2) are assumed to be continuous on T. Theimage of T under the mapping r is called a parametric surface and will be denoted by thesymbol r(T). In many of the examples we shall discuss, T will be a rectangle, a circulardisk, or some other simply connected set bounded by a simple closed curve. If the functionr is one-to-one on T, the image r(T) will be called a simple parametric surface. In such acase, distinct points of T map onto distinct points of the surface. In particular, everysimple closed curve in T maps onto a simple closed curve lying on the surface. A parametric surface r(T) may degenerate to a point or to a curve.

For example, if allthree functions X, Y, and 2 are constant, the image r(T) is a single point. If X, Y, and Zare independent of v, the image r(T) is a curve. Another example of a degenerate surfaceoccurs when X(u, v) = u + v, Y(u, v) = (U + v)~, and Z(u, v) = (U + v)~, whereT = [0, I] x [0, I].

If we write t = u + v we see that the surface degenerates to the spacecurve having parametric equations x = 1, y = t2, and z = t3, where 0 < t < 2. Suchdegeneracies can be avoided by placing further restrictions on the mapping function r,as described in the next section.

12.2 The fundamental vector product Consider a surface described by the vector equation r(u, v) = X(u, v)i + Y(u, v)j + Z(u, v)k, where (u, v) E T. If X, Y, and Z are differentiable on T we consider the two vectors and ar ax-=-au au i+azj+a;k & a x ay . az-=-av au i+-&J+--k.au The cross product of these two vectors &/au x &/au will be referred to as the fundamentalvector product of the representation r. Its components can be expressed as Jacobiandeterminants.

In fact, we have i j k ax ar az- - -(12.4) E X E = au au au ax ay az- - -au au au = ar azauaua y az z-z i+ az a x-aU auaz a x-Z au j+ a x a y- -au au ax ar- -au av k m z) i + a(z, X) . acx, Y) k=- ~ -ah 4 acu, v) ' + acu, v) . The fundamental vector product 421 - V Au r AU T tlA X/ ar ar au au %r au k:ar au in_, +Y ar ar . ar arF IGURE 12.6 Geometric interpretation of the Vectors z , a~ , and z X a~ If (u, v) is a point in T at which &/au and at /& are continuous and the fundamentalvector product is nonzero, then the image point r(u, v) is called a regular point of r.Points at which ar/au or &/au fails to be continuous or &/au x &/au = '0 are calledsingular points of r.

A surface r(T) is called smooth if all its points are regular points.Every surface has more than one parametric representation.

Some of the examplesdiscussed below show that a point of a surface may be a regular point for one repre-sentation but a singular point for some other representation. The geometric significanceof regular and singular points can be explained as follows: Consider a horizontal line segment in T. Its image under r is a curve (called a u-curve)lying on the surface r(T). For fixed v, think of the parameter u as representing time. Thevector ar/au is the velocity vector of this curve. When u changes by an amount Au, a pointoriginally at r(u, v) moves along a u-curve a distance approximately equal to l|ar/au]j Ausince |I&/&|| represents the speed along the u-curve. Similarly, for fixed u a point of av-curve moves in time Au a distance nearly equal to llar/avll Au. A rectangle in T havingarea Au Au is traced onto a portion of r(T) which we shall approximate by the parallelogramdetermined by the vectors (au/au) Au and (iYr/ih) Au. (See Figure 12.6.) The area of theparallelogram spanned by (au/au) Au and (&/au) Au is the magnitude of their crossproduct, Therefore the length of the fundamental vector product may be thought of as a localmagnification factor for areas. At the points at which this vector product is zero theparallelogram collapses to a curve or point, and degeneracies occur. At each regularpoint the vectors &/au and &/au determine a plane having the vector &/au x &/au asa normal. In the next section we shall prove that &/au x i?r/av is normal to every smoothcurve on the surface; for this reason the plane determined by &/au and &/au is called thetangent plane of the surface. Continuity of &/au and ar/av implies continuity of &/au xiYr/av; this, in turn, means that the tangent plane varies continuously on a smooth surface. 422 Surface integrals Thus we see that continuity of &/au and &/au prevents the occurrence of sharp edges orcorners on the surface; the nonvanishing of &/au x &/au prevents degeneracies. EXAMPLE 1. Surfaces with an explicit representation, z = f (x, y) .

For a surface with anexplicit representation of the form z = f(x, y) , we can use x and y as the parameters,which gives us the vector equation r(x, y) = xi + yj +f(x, yP. This representation always gives a simple parametric surface. The region T is called theprojection of the surface on the xy-plane. (An example is shown in Figure 12.7, p. 425.)To compute the fundamental vector product we note that ar_=i+afk andar af- = j + - k ,ay ayax ax iffis differentiable. This gives us (12.5) ij k & i% 10%-x-x axax ay 0 1 afay Since the z-component of &/ax x &lay is 1, the fundamental vector product is never zero.Therefore the only singular points that can occur for this representation are points atwhich at least one of the partial derivatives i?f/lax or af/lay fails to be continuous. A specific case is the equation z = 41 - x2 - 7, which represents a hemisphere ofradius 1 and center at the origin, if x2 + y2 5 1. The vector equation r(x,y)=xi+yj+dl -x2-y2k maps the unit disk T = {(x, y) 1 x2 + y2 5 1} onto the hemisphere in a one-to-onefashion. The partial derivatives &/i3x and &/i3y exist and are continuous everywhere inthe interior of this disk, but they do not exist on the boundary. Therefore every pointon the equator is a singular point of this representation. EXAMPLE 2. We consider the same hemisphere as in Example 1, but this time as the imageof the rectangle T = [0,2~] x [0, $T] under the mapping r (u ,v)=acosucosvi+as inucosvj+as invk . The fundamental vector product as a normal to the surface 423 The vectors &/au and at+/& are given by the formulas arau-- a s i n u c o s v i + a c o s u c o s v j , at-- =au -acosusinui-asinusinvj+acosvk. An easy calculation shows that their cross product is equal to & i%z x au = acosor(u,v). The image of T is not a simple parametric surface because this mapping is not one-to-oneon T. In fact, every point on the line segment v = &r, 0 < u 5 277, is mapped ontothe point (0, 0, a) (the North Pole). Also, because of the periodicity of the sine andcosine, r takes the same values at the points (0, v) and (27r, v), so the right and left edgesof Tare mapped onto the same curve, a circular arc joining the North Pole to the point(a, 0,O) on the equator. (See Figure 12.3.) The vectors &/au and i%/ih are continuouseverywhere in T. Since [I&/au x i?r/avll = a2 cos v , the only singular points of this repre-sentation occur when cos v = 0, The North Pole is the only such point. 12.3 The fundamental vector product as a normal to

~~the surface Consider a smooth parametric surface r(T), and let C* be a smooth curve in T. Thenthe image C = r(C*) is a smooth curve lying on the surface. We shall prove that at eachpoint of C the vector %/au x &/au is normal to C, as illustrated in Figure 12.6. Suppose that C* is described by a function a defined on an interval [a, b], say a(t) = U(t)i + V(t)j. Then the image curve C is described by the composite function p(t) = r[a(t)] = X[a(t)]i + Y[a(t)]j + Z[a(t)]k. We wish to prove that the derivative p'(t) is perpendicular to the vector &/au x &/auwhen the partial derivatives &/jJu and &/au are evaluated at (U(t), v(t)). To computep'(t) we differentiate each component of p(t) by the chain rule (Theorem 8.8) to obtain (12.6) p'(t) = VX. a'(t)i + VY - a'(t)j + VZ * a'(t)k, where the gradient vectors OX, V Y, and VZ are evaluated at (U(t), V(t)). Equation (12.6)can be rewritten as p'(t) = 5 U'(t) + k V(t), where the derivatives &/au and &/au are evaluated at (U(t), V(t)). Since &/au and &/auare each perpendicular to the cross product %/au x &/au, the same is true of p'(t). This 424 Surface integrals proves that &/au x &/au is normal to C, as asserted. For this reason, the vector product&/au x &/au is said to be normal to the surface r(T). At each regular point P of r(T) thevector &/au x arlav is nonzero; the plane through P having this vector as a normal iscalled the tangent plane to the surface at P. 12.4 Exercises In Exercises 1 through 6, eliminate the parameters u and v to obtain a Cartesian equation, thusshowing that the given vector equation represents a portion of the surface named. Also, computethe fundamental vector product ar/ au x ar/ au in terms of u and v. 1. Plane:r(u, v) = (x0 + a,u + b,v)i + (yO + a,u + b,v)j + (z,, + a+ + b,v)k.~~

~~2. Elliptic paraboloid:r(u,v) =aucosvi+businvj+u2k. 3. Ellipsoid:r(u,v) =asinucosvi + bsinusinvj + c c o s u k .~~

~~4. Surface of revolution:r(u, v) = u cos v i + u sin v j + f (u)k . 5. Cylinder:r(u, v) = ui + a sin v j + a cos v k. 6. Torus:r (u , v) = (a + b c o s u) s i n v i + (a + b c o s u) c o s v j + b s i n u k , w h e r e O < b < a . W h a t a r e the geometric meanings of a and b? In Exercises 7 through 10 compute the magnitude of the vector product &/au x ar/av. 7. r(u,v) =asinucoshvi + bcosucoshvj +csinhvk.8. r(u, v) = (u + v)i + (u - v)j + 4v2k.9. r(u, v) = (u + v)i + (u2 + v2)j + (u3 + v3)k. 10. r(u, v) = u cos v i + u sin v j + $4" sin 2v k. 12.5 Area of a parametric surface Let S = r(T) be a parametric surface described by a vector-valued function r definedon a region Tin the uv-plane. In Section 12.2 we found that the length of the fundamentalvector product &/au x &/au could be interpreted as a local magnification factor for areas.(See Figure 12.6.) A rectangle in T of area Au Au is mapped by r onto a curvilinear paral-lelogram on S with area nearly equal to & &/I /Ix - AuAv.aU a0 This observation suggests the following definition. DEFINITION OF AREA OF A PARAMETRIC SURFACE. The area of S, denoted by a(S), is de$nedby the double integral (12.7) a(S) = //II: x E//dudv. T Area of a parametric surface 425 In other words, to determine the area of S we first compute the fundamental vectorproduct ar/au x &/au and then integrate its length over the region T. When &/au x &/auis expressed in terms of its components, by.means of Equation (12.4), we have (12.8) Written in this form, the integral for surface area resembles the integral for computingthe arc length of a curve.? If S is given explicitly by an equation of the form z = f(x, y) we may use x and y asthe parameters. The fundamental vector product is given by Equation (12.5), so we have In this case the integral for surface area becomes where the region T is now the projection of S on the xy-plane, as illustrated in Figure 12.7. i4 s tI ," := f'(x,y) FIGURE 12.7 A surface S with an explicit representation, z =f(x, y). The region Tis the projection of S on the xy-plane. t Since the integral in (12.7) involves r, the area of a surface will depend on the function used to describethe surface. When we discuss surface integrals we shall prove (in Section 12.8) that under certain generalconditions the area is independent of the parametric representation.~~

~~The result is analogous to Theorem10.1, in which we discussed the invariance of line integrals under a change of parameter. 426 Surface integrals When S lies in a plane parallel to the xy-plane, the function f is constant, so af/lax =af/lay = 0, and Equation (12.9) becomes a(S) = if dx dyT This agrees with the usual formula for areas of plane regions.Equation (12.9) can be written in another form that gives further insight into its geometric significance. At each point of S, let y denote the angle between the normal vectorN = &/ax x &lay and the unit coordinate vector k. (See Figure 12.8.) Since the z-component of N is 1, we have - - lN-k 'OS ' - IIN llkll = (I = and hence @/ax x atjay I\ = l/cos y . Therefore Equation (12.9) becomes (12.10) a(S) = Suppose now that S lies in a plane not perpendicular to the xy-plane. Then y is constantand Equation (12.10) states that the area of S = (area of T)/cos y , or that (12.11) a(T) = a(S) cos y . Equation (12.11) is sometimes referred to as the area cosine principle. It tells us that ifa region S in one plane is projected onto a region Tin another plane, making an angle ywith the first plane, the area of T is cos y times that of S. This formula is obviously true ' A &?xarax ay FIGURE 12.8 The length of% x d' is l/cos y.aY h FIGURE 12.9 The area cosine principle fora rectangle. Area of a parametric surface 421 when S is the rectangle shown in Figure 12.9, because distances in one direction areshortened by the factor cos y while those in a perpendicular direction are unaltered byprojection. Equation (12.11) extends this property to any plane region S having an area. Suppose now that S is given by an implicit representation F(x, y, z) = 0. If S can beprojected in a one-to-one fashion on the xy-plane, the equation F(x, y, z) = 0 defines zas a function of x and y, say z = f (x, y), and the partial derivatives af/lax and af/lay arerelated to those of F by the equations af aFlax af way- - -a x - and - = - -aqaz ay afyaz for those points at which aF/az # 0. Substituting these quotients in (12.9), we find (12.12) a(S) =ss Jcwx)2 + cway)z + (aqaz)2 dx dy T 1 amw EXAMPLE 1. Area of a hemisphere. Consider a hemisphere S of radius a and center atthe origin. We have at our disposal the implicit representation x2 + y2 + z2 = a2, z 2 0;the explicit representation z = da2 -x2 - y2; and the parametric representation (12.13) r(u,v)=acosucosvi+asinucosvj+asinvk. To compute the area of S from the implicit representation we refer to Equation (12.12)with F(x,y,z) =x2 +y2 + z2 - a2. The partial derivatives of F are aF/ax = 2x, aFlay = 2y, aF/az = 22. The hemisphere Sprojects in a one-to-one fashion onto the circular disk D = {(x, y)) x2 + y2 5 a"} inthe xy-plane.~~

~~We cannot apply Equation (12.12) directly because the partial derivativeaF/az is zero on the boundary of D. However, the derivative aF/az is nonzero everywherein the interior of D, so we can consider the smaller concentric disk D(R) of radius R,where R < a. If S(R) denotes the corresponding portion of the upper hemisphere, Equa-tion (12.12) is now applicable and we find area of S(R) = J(W2 + (2Y)2 + (2z)2 dx dy D(R)PI = _a dx dy = aZ D(R) J,1, - y2 dy*ss D(R) The last integral can be easily evaluated by the use of polar coordinates, giving us area of S(R) = a1 -rdr d0=2na(a-da2-R2).Ja2 - r2 I When R + a this approaches the limit 2ra2. 428 Surface integrals We can avoid the limiting process in the foregoing calculation by using the parametric-representation in (12.13). The calculations of Example 2 in Section 12.2 show that II IIE X 5 = Ila cos ur(u, u)II = a2 lcos vi. Therefore we can apply Equation (12.7), taking for the region T the rectangle [0, 2~1 x[0, +T] . We find a(S) = a2 II lcos 01 du du = a2 ,""I .[S~'2cos u du1 du = 2rra2.T EXAMPLE 2. Another theorem of Pappus. One of the theorems of Pappus states that asurface of revolution, obtained by rotating a plane curve of length L about an axis in theplane of the curve, has area 2xLh, where h is the distance from the centroid of the curveto the axis of rotation. We shall use Equation (12.7) to prove this theorem. FIGURE 12.10 Area of a surface of revolution determined by Pappus' theorem. Suppose a curve C, initially in the xz-plane, is rotated about the z-axis. Let its equationin the xz-plane be z = f(x), where a 5 x < b , a 2 0. The surface of revolution S sogenerated can be described by the vector equation r(u,u)=ucosui+usinuj+f(u)k, where (u, u) E [a, b] x [0,27r]. The parameters u and u can be interpreted as the radiusand angle of polar coordinates, as illustrated in Figure 12.10. If a 5 u 5 b, all points(x,;y, z) at a given distance u from the z-axis have the same z-coordinate, f(u), so they Exercises 429 all lie on the surface. The fundamental vector product of this representation is i i k arxa'= cos vau au sin u f'(u) = - uf'(u) cos vi - uf'(u) sin u j + uk, -24 sin v u cosv 0 and hence Therefore Equation (12.7) becomes a(S) = joZ" [j: UJl + [f'(u)12 du] du = 27r s," uvll + [f'(u>12 du. The last integral can be expressed as SC x ds, a line integral with respect to arc length alongthe curve C. As such, it is equal to RL, where 1 is the x-coordinate of the centroid of C andL is the length of C. (See Section 10.8.) Therefore the area of S is 2rrLR. This proves thetheorem of Pappus. 12.6 Exercises 1. Let S be a parallelogram not parallel to any of the coordinate planes. Let S, , S,, and S,denote the areas of the projections of S on the three coordinate planes. Show that the area of Sis Js,2 +s,2 +s,:. 'Ik2. Compute the area of the region cut from the plane x + y + z = a by the cylinderx2 +y2 = a2. 3. Compute the surface area of that portion of the sphere x2 + y2 + z2 = a2 lying within thecylinder x2 + y2 = ay , where a > 0. 4. Compute the area of that portion of the surface z2 = 2xy which lies above the first quadrantof the xy-plane and is cut off by the planes x = 2 and y = 1. 0s.~~

~~A parametric surface S is described by the vector equation r(u,v) = ucosvi + usinvj + u2k, where0 f;u <4andO Make a sketch and indicate thegeometric meanings of the parameters u and v on the surface.(b) Compute the fundamental vector product ar/ au x arl au in terms of u and v.(c) The area of S is ~(65465 - 1)/n, where n is an integer. Compute the value of n. 6. Compute the area of that portion of the conical surface x2 + y2 = z2 which lies above thexy-plane and is cut off by the sphere x2 + y2 + z2 = 2ax. 7. Compute the area of that portion of the conical surface x2 + y2 = z2 which lies between thetwo planes z = 0 and x + 22 = 3. 8. Compute the area of that portion of the paraboloid x2 + z2 = 2uy which is cut off by theplaney = u. 9. Compute the area of the torus described by the vector equation r(u,v) = (a fbcosu)sinvi + (a +bcosu)cosvj +bsinuk, where 0 < b < a and 0 2 u 5 2~, 0 < v 5 2~. [Hint: Use the theorem of Pappus.] 430 Surface integrals 10. A sphere is inscribed in a right circular cylinder. The sphere is sliced by two parallel planesperpendicular to the axis of the cylinder. Show that the portions of the sphere and cylinderlying between these planes have equal surface areas. 11. Let T be the unit disk in the uv-plane, T = {(u, V) 1 u2 + u2 i; 1}, and let 2 ur(u, 21) = u2 + u2 + 1i+ &&j+1:Iz22++. (a) Determine the image of each of the following sets under r: the unit circle u2 + u2 = 1 ;the interval -1 < u 5 1 ; that part of the line u = u lying in T.(b) The surface S = r(T) is a familiar surface. Name and sketch it.(c) Determine the image of the uv-plane under r. Indicate by a sketch in the xyz-space thegeometric meanings of the parameters u and o. 12.7 Surface integrals Surface integrals are, in many respects, analogous to line integrals; the integration takesplace along a surface rather than along a curve. We defined line integrals in terms of aparametric representation for the curve. Similarly, we shall define surface integrals interms of a parametric representation for the surface. Then we shall prove that undercertain general conditions the value of the integral is independent of the representation. DEFINITION OF A SURFACE INTkGRAL.~~

~~Let S = r(T) be a parametric surface describedby a diflerentiable function Y defined on a region T in the uv-plane, and let f be a scalar$eldde$ned and bounded on S. The surface integral off over S is denoted by the symbol Jj f dS[or by ~Jfb,y, 4 dsl, and is dejned by the equation r(T) (12.14) whenever the double integral on the right exists. The following examples illustrate some applications of surface integrals. EXAMPLE 1. Surface area. When f = 1 , Equation (12.14) becomes The double integral on the right is that used earlier in Section 12.5 to define surface area.Thus, the area of S is equal to the surface integral JJ dS. For this reason, the symbol dS is sometimes referred to as an "element of surface area," and the surface integral ;I';'fdSis said to be an integral off with respect to the element of surface area, extended over the surface r(T). Surface integrals 431 EXAMPLE 2. Center of mass. Moment of inertia. If the scalar field f is interpreted asthe density (mass per unit area) of a thin material in the shape of the surface S, the totalmass m of the surface is defined by the equation m = IS f(x,~,z)dS.s Its center of mass is the point (2, J, z) determined by the equations Zm = jj xf(x, Y, z> ds, Im = SI yf(x, Y, z> dS, im = J'i' zf(x, Y, 4 dS .S S s The moment of inertia IL of S about an axis L is defined by the equation 1, = j j d2(x, Y, z)f(x, Y, z> dS,S where 6(x, y, z) denotes the perpendicular distance from a general point (x, y, z) of Sto the line L. To illustrate, let us determine the center of mass of a uniform hemispherical surface ofradius a.~~

~~We use the parametric representation r(u, u) = a cos u cos 2, i + a sin u cos uj + a sin v k , where (u, u) E [0,2x] x [0, &T]. This particular representation was discussed earlier inExample 2 of Section 12.2, where we found that the magnitude of the fundamental vectorproduct is a2 (cos ~1. In this example the density f is constant, say f = c, and the massm is 2na2c, the area of S times c. Because of symmetry, the coordinates f and J of thecenter of mass are 0. The coordinate Z is given by Pm = c SI z d S = c s.i a sin v . u2]cos V] du dvS T soZ=a/2.s R/2= 2na'c asin v cos v dv = na3c = - m, 0 2 EXAMPLE 3. Fluidjow through a surface. We consider a fluid as a collection of pointscalled particles. At each particle (x, y, z) we attach a vector V(x, y, z) which represents thevelocity of that particular particle. This is the velocity field of the flow. The velocity fieldmay or may not change with time. We shall consider only steady-state flows, that is, flowsfor which the velocity V(x, y, z) depends only on the position of the particle and not ontime. We denote by p(x, y, z) the density (mass per unit volume) of the fluid at the point(x, y, z). If the fluid is incompressible the density p will be constant throughout the fluid.For a compressible fluid, such as a gas, the density may vary from point to point. In anycase, the density is a scalar field associated with the flow. The product of the density and 432 Surface integrals the velocity we denote by F; that is, F,(x, y, 4 = pk Y, z>W Y, 4. This is a vector field called theflux density of the flow. The vector F(x, y, z) has the samedirection as the velocity, and its length has the dimensions mass distance mass unit volume * unit time = (unit area)(unit time) .~~

~~In other words, the flux density vector F(x, y, z) tells us how much mass of fluid per unitarea per unit time is flowing in the direction of V(x, y, z) at the point (x, y, z). Let S = r(T) be a simple parametric surface. At each regular point of S let n denote theunit normal having the same direction as the fundamental vector product. That is, let (12.15) It= i3r i?r auxav II II ar,@ * au au The dot product F * n represents the component of the flux density vector in the directionof n. The mass of fluid flowing through S in unit time in the direction of n is defined to bethe surface integral 12.8 Change of parametric representation We turn now to a discussion of the independence of surface integrals under a changeof parametric representation. Suppose a function r maps a region A in the uu-plane ontoa parametric surface r(A). Suppose also that A is the image of a region B in the St-planeunder a one-to-one continuously differentiable mapping G given by (12.16) G(s, t) = U(s, t>i + V(s, t)j if (s, t) E B. Consider the function R defined on B by the equation (12.17) R(s, t) = r[G(s, t)] . (See Figure 12.11.) Two functions r and R so related will be called smoothly equivalent.Smoothly equivalent functions describe the same surface. That is, r(A) and R(B) areidentical as point sets. (This follows at once from the one-to-one nature of G.) The nexttheorem describes the relationship between their fundamental vector products. Change of parametric representation 433 2 t S = r (A) = R(B) / FIGURE 12.11 Two parametric representations of the same surface. THEOREM 12.1. Let r and R be smoothly equivalent functions related bJ1 Equation (12.179,where G = Ui + Vj is a one-to-one continuously mapping of a region B in theSt-plane onto a region A in the uv-plane given by Equation (12.16). Then we have (12.18) where thepartial derivatives &/au and &/au are to be evaluated at the point (U(s, t), V(s, t)) .In other words, the fundamental vector product of R is equal to that of r, times the Jacobiandeterminant of the mapping G. Proof. The derivatives aRlas and aRli3t can be computed by differentiation of Equation(12.17). If we apply the chain rule (Theorem 8.8) to each component of R and rearrangeterms, we find that andaR at- au &aGf -=5-z+-'at u a?,,atwhere the derivatives &/au and &/au are evaluated at (U(s, t), V(s, t)). Now we crossmultiply these two equations and, noting the order of the factors, we obtain This completes the proof. The invariance of surface integrals under smoothly equivalent parametric representa-tions is now an easy consequence of Theorem 12.1. 434 Surface integrals THEOREM 12.2. Let r and R be smoothly equivalent functions, as described in Theorem12.1. If the surface integral JJ f dS exists, the surface integral jj f dS also exists and wehave r(A) R(B) jjfds = jjfds.r(A) R(B) Proof. By the definition of a surface integral we have Now we use the mapping G of Theorem 12.1 to transform this into a double integral overthe region B in the St-plane. The transformation formula for double integrals states that where the derivatives &/ih and &/au on the right are to be evaluated at (U(s, t), V(s, t)).Because of Equation (12.18), the integral over B is equal to This, in turn, is the definition of the surface integral Jj f dS. The proof is now complete.R(B) 12.9 Other notations for surface integrals If S = r(T) is a parametric surface, the fundamental vector product N = &/au x &/auis normal to S at each regular point of the surface. At each such point there are two unitnormals, a unit normal n, which has the same direction as N, and a unit normal n2 whichhas the opposite direction. Thus, N n1 = fiand n2 = -n,. Let it be one of the.two normals n, or n2. Let F be a vector field defined on S and assumethe surface integral jj F * n dS exists. Then we can write s I (12.19) F.ndS= F[r[u, v)] . n(u, v) $ x $ du dv T I I I/s = f F[r[u, v)] - E x $ du du, T where the + sign is used if n = n, and the - sign is used if n = n2. Other notations for surface integrals Suppose now we express F and Y in terms of their components, say 435 andF(x, y, 4 = % y, z>i + QGG y, 4j + R(x, Y, z)k Y(U, 0) = X(u, u)i + Y(z.4, u)j + Z(u, v)k. Then the fundamental vector product of Y is given by If n = n,, Equation (12.19) becomes (12.20) J's F. n dS = SJ' P[v(u, v)] '$$fdu du3 S T + Q[r(u, u)] a(z, du dv +ah 4 R[v(u, v)] a(x, du dv ; T Ta64 24 if n = n2, each double integral on the right must be replaced by its negative. The sum ofthe double integrals on the right is often written more briefly as (12.21) jj" P(x, Y, z) dy A dz +ss [j Q(x, y, 2) dz h dx + 11 R(x, y, Z) dx h dy ,s or even more briefly as (12.22) IS PdyAdz+QdzAdx+RdxAdy.S The integrals which appear in (12.21) and (12.22) are also referred to as surface integrals.Thus, for example, the surface integral JJ P dy A dz is defined by the equation S (12.23) P dy A dz = wu, u>l___a(y? ') du doS T ah 4 This notation is suggested by the formula for changing variables in a double integral.Despite similarity in notation, the integral on the left of (12.23) is not a double integral. First of all, P is a function of three variables. Also, we must take into consideration theorder in which the symbols dy and dz appear in the surface integral, because and hence / am a w, Y)-=-ah, 4 ah 4 SI P dy A dz = - i s Pdz A dy.S S 436 Surface integrals In this notation, formula (12.20) becomes (12.24) II F*ndS = II Pdyhdz+Qdzhdx+RdxAdys s ifn = n,. If n = n2 the integral on the right must be replaced by its negative. This formularesembles the following formula for line integrals: IcF*da= fcPdx + Qdy + Rdz. If the unit normal n is expressed in terms of its direction cosines, say +cosyk, thenF-n=Pcosa+Qcos/3+Rcosy,andwecanwrite I/F.ndS=jJ(P cos a + Q cos /3 + R cos r) dS.S S This equation holds when n is either n, or n2. The direction cosines will depend on thechoice of the normal. If n = n, we can use (12.24) to write (12.25) SI (Pcosa+Qcos@+Rcosy)dS=J/Pdyndz+QdzAdx+Rdxhdy.s s Ifn= n2 we have, instead, (12.26) si (P cos a + Q cos fl + R cos y) dS = -11 P dy A dz + Q dz A dx + R dx A dy .S s 12.10 Exercises 1. Let S denote the hemisphere x2 + y2 + z2 = 1, z 20, and let F(x,y,z) =xi +yj. Letn be the unit outward normal of S. Compute the value of the surface integral JJ F n dS,using: S (a) the vector representation r(u, 0) = sin u cos u i + sin u sin u j + cos u k,(b) the explicit representation z = dm. 2. Show that the moment of inertia of a homogeneous spherical shell about a diameter is equal to$ma2, where m is the mass of the shell and a is its radius. 3. Find the center of mass of that portion of the homogeneous hemispherical surface x2 + y2 +z2 = a2 lying above the first quadrant in the xy-plane.~~

~~Exercises 431 4. Let S denote the plane surface whose boundary is the triangle with vertices at (1, 0, 0), (0, 1, O),and (0, 0, l), and let F(x, y, z) = xi + yj + zk . Let n denote the unit normal to S havinga nonnegative z-component. Evaluate the surface integral JJ F. n dS, using: (a) the vector representation r(u, u) = (u + u)i + (u - u)j"+ (1 - 2u)k,(b) an explicit representation of the form z = f(x, y) . 5. Let S be a parametric surface described by the explicit formula z = f(x, y) , where (x, y) variesover a plane region T, the projection of S in the xy-plane. Let F = Pi + Qj + Rk and letn denote the unit normal to S having a nonnegative z-component. Use the parametric re-presentation r(x, y) = xi + yj -tf(x, y)k and show that where each of P, Q, and R is to be evaluated at (x, y,f(x, y)).6. Let S be as in Exercise 5, and let p be a scalar field. Show that: ( ) sfa Pk y, z> dS = PLx,y,f(x,y)$ + (q + (qfxdy. S T (b)J's p(x,y,z)dy Adz = -ss dx,y,f(x,y)l ;dxdy.s T (clss p(x,y,z)dz lrdx = -ss af dx,y&,yNaydxdy. s T 7. If S is the surface of the sphere x2 + y2 + z2 = u2, compute the value of the surface integral J.i xzdy/tdz+yzdzAdx+x2dxAdy.S Choose a representation for which the fundamental vector product points in the direction of theoutward normal. 8. The cylinder x2 + y 2 = 2x cuts out a portion of a surface S from the upper nappe of the conex2 +y2 =z2. Compute the value of the surface integral (x4 - y4 + y2z2 -z2x2 + 1)dS.S 9. A homogeneous spherical she11 of radius a is cut by one nappe of a right circular cone whosevertex is at the center of the sphere. If the vertex angle of the cone is c(, where 0 < c(< r,determine (in terms of a and a) the center of mass of the portion of the spherical shell that liesinside the cone.~~

~~10. A homogeneous paper rectangle of base 2aa and altitude h is rolled to form a circular cylindricalsurface S of radius a. Calculate the moment of inertia of S about an axis through a diameterof the circular base. 438 Surface integrals 11. Refer to Exercise 10. Calculate the moment of inertia of S about an axis which is in the planeof the base and is tangent to the circular edge of the base. 12. A fluid flow has flux density vector F(x, y, z) = xi - (2x + y)j + zk. Let S denote the hemi-sphere x2 + y2 + z2 = 1, z > 0, and let R denote the unit normal that points out of the sphere. Calculate the mass of fluid flowing through S in unit time in the direction of n.~~

~~13. Solve Exercise 12 if S also includes the planar base of the hemisphere. On the lower base theunit normal is -k . 14. Let S denote the portion of the plane x + y + z = t cut off by the unit sphere x2 + y2 +22 =l. Letq(x,y,z)=l - x2 - y2 - z2 if (x, y, z) is inside this sphere, and let ~(x, y, z)be 0 otherwise. Show that ssdx,y,z)dS = ; (3 - t2)2 if ItI I J5, S 0 i f jtJ > JS. [Hint: Introduce new coordinates (x1, y, , 1z) with the z,-axis normal to the planex + y + z = t . The use the polar coordinates in the x,yl-plane as parameters for S.] 12.11 The theorem of Stokes The rest of this chapter is devoted primarily to two generalizations of the second funda-mental theorem of calculus involving surface integrals. They are known, respectively, asStokes' theorem? and the divergence theorem. This section treats Stokes' theorem. Thedivergence theorem is discussed in Section 12.19. Stoke's theorem is a direct extension of Green's theorem which states that ll(E-;)dxdy=JcPdx+Qdy,S where S is a plane region bounded by a simple closed curve C traversed in the positive(counterclockwise) direction. Stokes' theorem relates a surface integral to a line integraland can be stated as follows. THEOREM 12.3 STOKES' THEOREM. Assume that S is a smooth simple parametric surface,say S = r(T), where T is a region in the uv-plane bounded by a piecewise smooth Jordancurve I'. (See Figure 12.12.) Assume also that r is a one-to-one mapping whose componentshave continuous second-order partial derivatives on some open set containing T u I'. Let Cdenote the image of r under r, and let P, Q, and R be continuously scalarJieldson S. Then we have (12.27) j j- ($f - 2) dy A dz + (E - E) dz A dx + (2 - $j dx A dy x = s Pdx+Qdy+Rdz.c t In honor of G. G. Stokes (1819-1903), an Irish mathematician who made many fundamental contributionsto hydrodynamics and optics. The theorem of Stokes - 439 V T I' 'r X/ n S = r(T) / /c WY FIGURE 12.12 An example of a surface to which Stokes' theorem is applicable. The curve I? is traversed in thepositive (counterclockwise) direction and the curve C is traversedin the direction inherited from r through the mapping function r. Proof. To prove the theorem it suffices to establish the following three formulas, (12.28) A dy + g dz A dxQ : Cl A 3 , A dz + zdx A dy1 Q . 2 0, hdx + gdy r\dzay Addition of these three equations gives the formula (12.27) in Stokes' theorem. Since thethree are similar, we prove only Equation (12.28). The plan of the proof is to express the surface integral as a double integral over T. Thenwe use Green's theorem to express the double integral over T as a line integral over I'.Finally, we show that this line integral is equal to Jc P d.y. We write r(u, v) = X(u, v)i + Y(u, v)j + %(zr, v)k and express the surface integral over S in the form ap--dxAdy+EdzAdx ay 440 Surface integrals Now let p denote the composite function given by PC% 0) = PW(u, u>, Y(u, u), Z(u, u)] .The last integrand can be written as (12.29) --- az~=iPE)-i(PE)*ap 8(X, Y) + ap qz, X) 3Y %u, u) The verification of (12.29) is outlined in Exercise 13 of Section 12.13. Applying Green'stheorem to the double integral over T we obtain where P is traversed in the positive direction. We p,arametrize I? by a function y defined onan interval [a, b] and let (12.30) a(t) = Wt)rl be a corresponding parametrization of C. Then by expressing each line integral in termsof its parametric representation we find that s $du+pEdv=r au s P d x ,c which completes the proof of (12.28).~~

~~12.12 The curl and divergence of a vector field The surface integral which appears in Stokes' theorem can be expressed more simply interms of the curl of a vector field. Let F be a differentiable vector field given by F(x, y, z) = P(x, y, z>i + Q(x, y, z>i + W> y, z)k. The curl of F is another vector field defined by the equation (12.31) curlF= (z-g)i+ ($- E)j+ (z-$)k. The components of curl F are the functions appearing in the surface integral in Stokes'formula (12.27). Therefore, this surface integral can be written as H (curl F) * n dS,S The curl and divergence of a vectorjeld 441 where n is the unit normal vector having the same direction as the fundamental vectorproduct of the surface; that is, at- i% -"auauil= The line integral in Stokes' formula (12.27) can be written as SC F * da, where a is therepresentation of C given by (12.30). Thus, Stokes' theorem takes the simpler form SI (curlF).ndS= F.da.s sc For the special case in which S is a region in the xy-plane and n = k, this formula reducesto Green's theorem, Equation (12.31) defining the curl can be easily remembered by writing it as an expansionof a 3 x 3 determinant, i j k curl F = -a 2 -? - = (g-E)i+ (?$-g)j+ (z-z)k.ax ay aZ This determinant is to be expanded in terms of first-row minors, but each "product" suchas a/ay times R is to be interpreted as a partial derivative aR/ay. We can also write thisformula as a cross product, curlF= V x F , where the symbol V is treated as though it were a vector, V=&i+zj+ik.ay If we form the "dot product" V . F in a purely formal way, again interpreting productssuch as a/ax times P as aP/ax, we find that (12.32) V.F=a~+!%+a~ax ay aZ * Equation (12.32) defines a scalar field called the divergence of F, also written as div F. 442 Surface integrals We have already used the symbol VP, to denote the gradient of a scalar field p, given by This formula can be interpreted as formal multiplication of the symbolic vector V by thescalar field q- . Thus, the gradient, the divergence, and the curl can be represented sym-bolically by the three products Vy, V . F, and C' x F, respectively. Some of the theorems proved earlier can be expressed in terms of the curl. For example,in Theorem 10.9 we proved that a vector fieldf =: (f, . . . ,fn) , continuously differentiableon an open convex set Sin n-space, is a gradient Ion S if and only if the partial derivatives ofthe components off satisfy the relations (12.33) w (4 = DjM4 (j!, k = 1, 2, . . ., n) In the 3-dimensional case Theorem 10.9 can be restated as follows. THEOREM 12.4. Let F = Pi + Qj + Rk be a continuously differentiable vectorJeld on anopen convex set S in 3-space. Then F is a gradient 011 S if and only lj- we have (12.34) c u r l F = O o n S . Proof. In the 3-dimensional case the relations (12.33) are equivalent to the statementthat curl F = 0. 12.13 Exercises In each of Exercises 1 through 4, transform the surface integral jj (curl F) . II dS to a line integral by the use of Stokes' theorem, and then evaluate the line integral! 1. F(x, y, z) = y2i + xy/" + xzk, where S is the hemisphere x2 i- y2 + z2 = 1, z 2 0, and n isthe unit normal with a nonnegative z-component. 2. F(x, y, z) = yi + zj -t xk, where S is the portion of the paraboloid z = 1 -x2 - y2 withz 2 0, and n is the unit normal with a nonnegative z-component. 3. F(x, y, z) = (]J - z)i + yzj - xzk, where S consists of the five faces of the cube 0 < x < 2,0 5 y 5 2, 0 < z < 2 not in the xy-plane. The unit normal II is the outward normal. 4. F(x, y, z) = xzi - yi + x2yk, where S consists of the three faces not in the xz-plane of thetetrahedron bounded by the three coordinate planes and the plane 3x + y + 32 = 6. Thenormal n is the unit normal pointing out of the tetrahedron, In Exercises 5 through 10, use Stokes' theorem to show that the line integrals have the valuesgiven. In each case, explain how to traverse C to arrive at the given answer.~~

~~5. Joy dx + z dy + x dz = - - - - 43, where C is the curve of intersection of the sphere x2 +y2 + z2 = a2 and the plane x + y -t z = 0, 6. Jo (y + z) dx + (z + x) dy + (x + y) dz = 0, where C is the curve of intersection of thecylinder x2 + y2 = 2y and the plane y = z. 7. Jo y2 dx + xy dy + xz dz = 0, where C is the curve of Exercise 6. Further properties of the curl and divergence 443 8. 9. 10. 11. 12. 13. SC (y - z) dx + (z - x) dy + (X - y) ciz = 24~ + b), where C is the intersection of thecylinder x2 + y2=a2andtheplanex/a+z/b=1,a>0,b>0.Jc (y" + z2) dx + (x2 + z") dy + (x2 + yz) dz = 2mb2, where C is the intersection of the hemisphere x2 + y2 + z2 = 2ux, z > 0, and the cylinder x2 + y2 = 2bx, where 0 < b < a.jc (y" - z2) dx + (z2 - x2) dy + (x2 - y2) dz = 9a3/2, where C is the curve cut from theboundaryofthecubeo Ix div F =f'(x) + g'(y) + h'(z) and curlF= 0 . Let F(x, y, z) = xy2z2i + z2 sinyj + x2evk. The Jacobian matrix is Therefore, anddiv F = xy2z2 + z2 cos y curl F = (x2,, - 22 sin y)i + (2xy2z - 2xeYj - 2xyz2k. EXAMPLE 3. The divergence and curl of a gradient. Suppose F is a gradient, sayF = grad p = i3p/ax i + ap/ay j + aq/az k. The Jacobian matrix is (12.36) a2p, a2p, a29,ax2 ayax aZ ax 1 1a29, 3 3axay ay2 aZ ay * Therefore, azp azp 3291- --axaZ ay a2 a22 J a2p a24, 3%divF=s+G+s. The expression on the right is called the Laplacian of a) and is often written more brieflyas V2~. Thus, the divergence of a gradient VP, is the Laplacian of ~ZJ. In symbols, this iswritten (12.37) div (VT) = V2rp.~~

~~Further properties qf the curl and divergence 44.5 When VP = 0, the function 9 is called harmonic. Equation (12.37) shows that the gradientof a harmonic function has zero divergence. When the mixed partial derivatives in matrix(12.36) are continuous, the matrix is symmetric and curl F is zero. In other words, curl (grad q) = 0 for every scalar field p with continuous second-order mixed partial derivatives. Thisexample shows that the condition curl F = 0 is necessary for a continuously differentiablevector field F to be a gradient. In other words, if curl F # 0 on an open set S, then F isnot a gradient on S. We know also, from Theorem 12.4 that if curl F = 0 on an openconvex set S, then F is a gradient on S. A field with zero curl is called irrotational. EXAMPLE 4. A vector jield with zero divergence and zero curl. Let S be the set of all(x, y) # (0, 0) , and let F(x, y) = - 1 i + -&- 'x2 + y2 x2 + y2j if (x, u) E S. From Example 2 in Section 10.16 we know that F is not a gradient on S(although F is a gradient on every rectangle not containing the origin). The Jacobianmatrix is 1 -2XY)p - x2 (x2 + yy (x2 + y2y O DO, Y> =y2 - x2 -2XJ' (x " + yy2 (x2 + yy2 O 0 0- O- and we see at once that div F = 0 and curl F = 0 on S. EXAMPLE 5. The divergence and curl oj' a curl. If F = Pi + Qj + Rk, the curl of Fis a new vector field and we can compute its divergence and curl. The Jacobian matrix ofcurl F is a2P a2R a2P a2R a2P a2R ay ax ay2 aZ ax azay- If we assume that all the mixed partial derivatives are continuous, we find that anddiv (curl F) = 0 (12.38) curl (curl F) = grad (div F) - V2F, 446 Surface integrals where V2F is defined by the equation V2F = (V2P)i + (V2Q)j + (V2R)k. The identity in (12.38) relates all four operators, gradient, curl, divergence, and Laplacian.The verification of (12.38) is requested in Exercise 7 of Section 12.15. The curl and divergence have some general properties in common with ordinaryderivatives. First, they are linear operators. That is, if a and b are constants, we have (12.39) d i v (a F + b G) = a d i v F + bdivG, and (12.40) curl (aF + bG) = a curl F + b curl G. They also have a property analogous to the formula for differentiating a product: (12.41) div (plF) = a, div F -I- Vp - F, and (12.42) curl (qF) = q curl F + VP x F, where 9 is any differentiable scalar field. These properties are immediate consequencesof the definitions of curl and divergence; their proofs are requested in Exercise 6 ofSection 12.15. If we use the symbolic vector v=$i+dj+ikay once more, each of the formulas (12.41) and (12.42) takes a form which resembles moreclosely the usual rule for differentiating a product: andV*(tpF)= cpV*F-t Vpl*F Vx(pF)=pVxF+Vg;xF. In Example 3 the Laplacian of a scalar field, V2p, was defined to be azq/ax2 + azY/ay2 +azv/az2. In Example 5 the Laplacian V2F of a vector field was defined by components. Weget correct formulas for both V2p, and V2F if we interpret V2 as the symbolic operator This formula for V2 also arises by dot multiplication of the symbolic vector V with itself. Exercises 447 Thus, we have V2 = V * V and we can write V"q, = (V * V)p and V2F = (V * V)F. Now consider the formula V * Vq. This can be read as (V . V)rp, which is Vz~y; or asV . (VP), which is div ('17~). In Example 3 we showed that div (VP) = Vzgi, so we have (V * V)pl = v * (VqJ) ;hence we can write V. Vpj for either of these expressions without danger of ambiguity.This is not true, however, when qj is replaced by a vector field F. The expression (V * V)Fis V2F, which has been defined. However, V * (VF) is meaningless because VF is notdefined.~~

~~Therefore the expression V . OF is meaningful only when it is interpreted as(V - V)F. These remarks illustrate that although symbolic formulas sometimes serve as aconvenient notation and memory aid, care is needed in manipulating the symbols. 12.15 Exercises 1. For each of the following vector fields determine the Jacobian matrix and compute the curland divergence.(a) F(x, y, 2) = (x2 + yz)i + (y" + xt)j + (z2 + xy)k.(b) F(x, y, z) = (22 - 3y)i + (3x - z)j + (y - 2x)k.(c) F(x, y, z) = (z + sin y)i - (z - x cos y)j.(d) F(x, y, z) = eZrj + cos xy k.(e) F(x, y, z) = x2 sin y i f y2 sin xzj + xy sin (cos z)k . 2. If r = xi + yj + zk and r = llrll , compute curl [f(r)r], where f'is a differentiable function.3. If r = xi + yj + zk and A is a constant vector, show that curl (A x r) = 2A.4. If r = xi + yj + zk and r = ((r((, find all integers n for which div (r-9) = 0.5. Find a vector field whose curl is xi + yj + zk or prove that no such vector field exists.6. Prove the elementary properties of curl and divergence in I(The converse is also true. That is, if F. curl F = 0 in a suitableregion, a nonzero ,U exists such that ,uFis a gradient. The proof of the converse is not required.) 10. Let F(x, y, z) = y2z2i + z2x2j + x2y2k. Show that curl F is not always zero, but thatF. curl F = 0. Find a scalar field /c such that ,uF is a gradient. 11.~~

~~Let V(x, y) = yci + xcj, where c is a positive constant, and let r(x, y) = xi + yj. Let R bea plane region bounded by a piecewise smooth Jordan curve C. Compute div (V x r) andcurl (V x r), and use Green's theorem to show that P c V x r*da =0, where a describes C. 448 Surface integrals 12. Show that Green's theorem can be expressed as follows: jj(curl v>.kdxdy =$ V.Tds,R c where T is the unit tangent to C and s denotes arc length.13. A plane region R is bounded by a piecewise smooth Jordan curve C. The moments of inertia of R about the x- and y-axes are known to be a and b, respectively. Compute the line integral I c V(r4) . n ds in terms of a and b. Here Y = /lxi + rjll , n denotes the unit outward normal of C, and s denotesarc length. The curve is traversed counterclockwise. 14. Let F be a two-dimensional vector field. State a definition for the vector-valued line integralJo F x du. Your definition should be such that the following formula is a consequence ofGreen's theorem: sF x da =k ' (divF)dxdy,ii c R where R is a plane region bounded by a simple closed curve C. kl2.16 Reconstruction of a vector field from its curl In our study of the gradient we learned how to determine whether or not a given vectorfield is a gradient. We now ask a similar question concerning the curl. Given a vector fieldF, is there a G such that curl G = F? Suppose we write F = Pi + Qj + Rk and G =Li + Mj + Nk. To solve the equation curl G = F we must solve the system of partialdifferential equations axv aiv aL aN aM aL(1 2 . 4 3) --aZ=P, z-ax=Q, %-%=R ayfor the three unknown functions L, M, and N when P, Q, and R are given. It is not always possible to solve such a system. For example, we proved in Section12.14 that the divergence of a curl is always zero. Therefore, for the system (12.43) to havea solution in some open set S it is necessary to have (12.44) ap+aQ+aR=,ax ay a2 everywhere in S. As it turns out, this condition is also sufficient for system (12.43) to havea solution if we suitably restrict the set S in which (12.44) is satisfied.~~

~~We shall prove nowthat condition (12.44) suffices when S is a three-dimensional interval. THEOREM 12.5 Let F be continuously dlrerentiable on an open interval S in 3-space.Then there exists a vector field G such that curl G = F if and only if div F = 0 everywherein S. Reconstruction of a vector field from its curl 449 Proof. The necessity of the condition div F = 0 has already been established, sincethe divergence of a curl is always zero. To establish the sufficiency we must exhibit threefunctions L, M, and N that satisfy the three equations in (12.43). Let us try to make achoice with L = 0. Then the second and third equations in (12.43) become aN Q aM-= a n d -= R .ax ax This means that we must have W, Y, z> = -1; Q(t, Y, z> dt + f(y, 4 and M(x, Y, z> = Jz R(t, Y, z> dt + dy, z>, where each integration is along a line segment in S and the "constants of integration"f(v, z) and g(y, z) are independent of x. Let us try to find a solution with f(y, z) = 0.The first equation in (12.43) requires (12.45)aN aM- - =p *ay aZ For the choice of M and N just described we have (12.46)aN aM- - az = - ; j Q(t, y, z) dt - ,-'R(t, Y, z) dt - 5.ay XII z a3 At this stage we interchange the two operations of partial differentiation and integration,using Theorem 10.8. That is, we write (12.47) and $ /=Q(t, Y, z> dt = JkQ(t, Y, z> dtY x0 x0 (12.48)za 5 ii $0sR(t, y, z) dt = f &R(t, Y, z> dt .20 Then Equation (12.46) becomes (12.49)aN aM- - aZ =ay s 'F&Q@, Y, z> - D&t, y, z)] dt - 2 .aI Using condition (12.44) we may replace the integrand in (12.49) by D,P(t, y, z); Equation(12.49) becomes aN aM ' ag- - - =ay aZ sz.D,P(t, y, z) dt - ff = P(x, y, z) - P(x,, y, z) - z. 450 Surface integrals Therefore (12.45) will be satisfied if we choose g so that ag/az = -P(x,, y, z) . Thus, forexample, we may take dy, z> = -s:, W, 3 Y, u> du .~~

~~This argument leads us to consider the vector field G = Li + Mj + Nk, whereL(x, y, z) = 0 and M(x, Y, z> = j; Nt, Y, z) dt - j; P(x,, Y, u> du, Nx, Y, z) = - j; Q(c Y, z> dt . For this choice of L, M, and N it is easy to verify, with the help of (12.47) and (12.48),that the three equations in (12.43) are satisfied, giving us curl G = F, as required. It should be noted that the foregoing proof not only establishes the existence of a vectorfield G whose curl is F, but also provides a straightforward method for determining G byintegration involving the components of F.~~

~~For a given F, the vector field G that we have constructed is not the only solution of theequation curl G = F. If we add to this G any continuously differentiable gradient Vp7 weobtain another solution because curl (G -I- Vqj) = curl G + curl (Vpl) = curl G = F, since curl (Vy) = 0. Moreover, it is easy to show that all continuously differentiablesolutions must be of the form G + Vg,. Indeed, if H is another solution, then curl H =curl G, so curl (H - G) = 0. By Theorem 10.9 it follows that H - G = Vv for somecontinuously differentiable gradient Vgj; hence H = G + VP,, as asserted. A vector field F for which div F = 0 is sometimes called solenoidal. Theorem 12.5states that a vector field is solenoidal on an open interval S in 3-space if and only if it is thecurl of another vector field on S. The following example shows that this statement is not true for arbitrary open sets. EXAMPLE. A solenoidal vector field that is not a curl. Let D be the portion of 3-spacebetween two concentric spheres with center at the origin and radii a and 6, where0 < a < b . Let V = r/r3, where Y = xi + yj + zk and r = 11~ 11 . It is easy to verify thatdiv V = 0 everywhere in D. In fact, we have the general formula div (Pr) = (n + 3)r", and in this example n = -3. We shall use Stokes' theorem to prove that this V is not acurl in D (although it is a curl on every open three-dimensional interval not containing theorigin). To do this we assume there is a vector field U such that V = curl U in D and obtaina contradiction. By Stokes' theorem we can write (12.50) jj(curl U).ndS = $cU.dol,S Reconstruction of a vectorjeldj7o.m its curl 451 S --L' X/ FIGURE 12.13 The surface S and curve C in Equation (12.50). where S and C are the surface and curve shown in Figure 12.13. To construct S, we takea spherical surface of radius R concentric with the boundaries of D, where a < R < b,and we remove a small "polar cap," as indicated in the ligure. The portion that remainsis the surface S. The curve C is the circular edge shown. Let n denote the unit outer normalof S, so that n = r/r. Since curl U = V = rlr3, we have YY 1(curl V) * n = - . - = - , r3 r r2 On the surface S this dot product has the constant value 1/R2. Therefore we have ss(curl U) . n dS = -$ ssdS = 'y . s S When the polar cap shrinks to a point, the area of S approaches 47iR2 (the area of thewhole sphere) and, therefore, the value of the surface integral in (12.50) approaches 47r. Next we examine the line integral in (12.50). It is easy to prove that for any line integralJo U * da we have the inequality IJ IclJ.da 5 M*(lengthofC), where M is a constant depending on U. (In fact, M can be taken to be the maximum valueof /IUi/ on C.) Therefore, as we let the polar cap shrink to a point, the length of C and thevalue of the line integral both approach zero. Thus we have a contradiction; the surfaceintegral in (12.50) can be made arbitrarily close to 477, and the corresponding line integralto which it is equal can be made arbitrarily close to 0. Therefore a function U whose curlis V cannot exist in the region D. The difficulty here is caused by the geometric structure of the region D. Although thisregion is simply connected (that is, any simple closed curve in D is the edge of a parametric 452 Surface integrals surface lying completely in D) there are closed surfaces in D that are not the completeboundaries of solids lying entirely in D. For example, no sphere about the origin is thecomplete boundary of a solid lying entirely in D. If the region D has the property thatevery closed surface in D is the boundary of a solid lying entirely in D, it can be shown that avector field U exists such that V = curl U in D if and only if div V = 0 everywhere in D.The proof of this statement is difficult and will not be given here. *12.17 Exercises 1. Find a vector field G(x, y, z) whose curl is 2i + j + 3k everywhere in 3-space. What is themost general continuously differentiable vector field with this property? 2. Show that the vector field F(x, y, z) = (y - z)i + (z - x)j + (x - y)k is solenoidal, andfind a vector field G such that F = curl G everywhere in 3-space. 3. Let F(x, y, z) = -zi + xyk. Find a continuously differentiable vector field G of the formG(x, y, z) = L(x, y, z)i + M(x, y, z)j such that F = curl G everywhere, in 3-space. What isthe most general G of this form? 4. If two vector fields Uand Vare both irrotational, show that the vector field U x Vis solenoidal.5. Let r = xi + yj + zk and let Y = IlrlI . Show that n = -3 is the only value of n for which Pr is solenoidal for r # 0. For this n, choose a 3-dimensional interval S not containing theorigin and express re3r as a curl in S. Note: Although re3r is a curl in every such S, it is nota curl on the set of all points different from (0, 0,O). 6. Find the most general continuously differentiable functionfof one real variable such that thevector fieldf(r)r will be solenoidal, where r = xi + yj + zk and r = ljrll . 7.~~

~~Let V denote a vector field that is continuously differentiable on some open interval S in 3-space. Consider the following two statements about V:(i) curl V = 0 and V = curl U for some continuously differentiable vector field U (everywhereon S).(ii) A scalar field v exists such that VP, is continuously differentiable and such that V = grad q and V2v = 0 everywhere on S. (a) Prove that (i) implies (ii). In other words, a vector field that is both irrotational andsolenoidal in S is the gradient of a harmonic function in S.(b) Prove that (ii) implies (i), or give a counterexample. 8. Assume continuous differentiability of all vector fields involved, on an open interval S. SupposeH = F + G, where F is solenoidal and G is irrotational. Then there exists a vector field Usuch that F = curl U and a scalar field p such that G = K'v in S. Show that U and p satisfythe following partial differential equations in S: V2e, = div H, grad (div U) - V2U = curl H. Note: This exercise has widespread applications, because it can be shown that everycontinuously differentiable vector field H on Scan be expressed in the form H = F +G, where F is solenoidal and G is irrotational. 9. Let H(x, y, z) = x2yi + y2zj + z2xk. Find vector fields F and C, where F is a curl and Gis a gradient, such that H = F + G. 10. Let u and L' be scalar fields that are continuously difl'erentiable on an open interval R in 3-space.(a) Show that a vector field F exists such that Gu x Gu = curl F everywhere in R. Extensions of Stokes' theorem 453 (b) Determine whether or not any of the following three vector fields may be used for F inpart (a): (i) T(W); (ii) uVu; (iii) VVU.(c) If u(x, y, z) = X3 - y3 + z2 and v(x, y, z) = x + y + z, evaluate the surface integral /.IVu x Cv.ndS, S where S is the hemisphere x2 + y2 + z2 = 1, z > 0, and n is the unit normal with a non--negative z-component. 12.18 Extensions of Stokes' theorem Stokes' theorem can be extended to more general simple smooth surfaces. If T is amultiply connected region like that shown in Figure 12.14 (with a finite number of holes),the one-to-one image S = r(T) will contain the same number of holes as T. To extend V Z - r cu X/ FIGURE 12.14 Extension of Stokes' theorem to surfaces that are one-to-one imagesof multiply connected regions. Stokes' theorem to such surfaces we use exactly the same type of argument as in the proofof Stokes' theorem, except that we employ Green's theorem for multiply connected regions(Theorem 11.12). In place of the line integral which appears in Equation (12.27) we needa sum of line integrals, with appropriate signs, taken over the images of the curves formingthe boundary of T. For example, if T has two holes, as in Figure 12.14, and if the boundarycurves I', r,, and I'2 are traversed in the directions shown, the identity in Stokes' theoremtakes the form where C, C1, and C, are the images of l?, rl, and P2, respectively, and p, t+, and e2 arethe composite functions p(t) = r[y(t)], &(t) = r[y,(t)], Q2(t) = r[y2(t)]. Here y, yl, andy2 are the functions that describe I?, rl, and rz in the directions shown. The curves C,C,, and C, will be traversed in the directions inherited from r, rI, and rz through themapping function r. 454 Surface integrals Stokes' theorem can also be extended to some (but not all) smooth surfaces that are notsimple. We shall illustrate a few of the possibilities with examples. Consider first the cylinder shown in Figure 12.15. This is the union of two simple smoothparametric surfaces S1 and S,, the images of two adjacent rectangles TI and T,, undermappings rr and r2, respectively. If y1 describes the positively oriented boundary Fl ofTl and yz describes the positively oriented boundary I'2 of T,, the functions p1 and pzdefined by t+(t) = rI[yl(t)l y p2(t) = dYdt)l describe the images C, and C, of rr and rz, respectively. In this example the representa-tions r1 and r2 can be chosen so that they agree on the intersection PI A rz. If we apply FIGURE 12.15 Extension of Stokes' theorem to a cylinder. Stokes' theorem to each piece SI and S, and add the two identities, we obtain (12.51) J J (curl F). n, dS + ss (curl P) . n2 dS = jc/'.LIPl + jc2F*dp,,Sl sz where its and n2 are the normals determined by the fundamental vector products of r1and r2, respectively. Now let r denote the mapping of Tl U T, which agrees with rl on Tl and with r2 onT,, and let n be the corresponding unit normal determined by the fundamental vectorproduct of r. Since the normals IZ~ and IZ~ agree in direction on S, n S,, the unit normaln is the same as n, on S, and the same as IQ on S,. Therefore the sum of the surfaceintegrals on the left of (12.51) is equal to ss (curl F) * n dS .Sl us2 For this example, the representations r1 and r2 can be chosen so that p1 and pz determineopposite directions on each arc of the intersection C1 n C,, as indicated by the arrowsin Figure 12.15. The two line integrals on the right of (12.51) can be replaced by a sumof line integrals along the two circles C; and C,l forming the upper and lower edges ofS, u S,, since the line integrals along each arc of the intersection C, n C, cancel. There-fore, Equation (12.51) can be written as (12.52) (curl F) . n dS =.s F-dp, + s F.&z,Sl "Sz Cl' Cz' Extensions of Stokes' theorem 455 where the line integrals are traversed in the directions inherited from rI and rZ.~~

~~The twocircles C; and C,l are said to form the complete boundary of S, u S,. Equation (12.52)expresses the surface integral of (curl 1") * II over S, U Sn as a line integral over thecomplete boundary of S1 u S,. This equation is the extension of Stokes' theorem for acylinder. Suppose now we apply the same concepts to the surface shown in Figure 12.16. Thissurface is again the union of two smooth simple parametric surfaces S, and S,, the imagesof two adjacent rectangles Tl and T, . This particular surface is called a Miibius band;? a n, FIGURE 12.16 A Miibius band considered as the union of two simple parametricsurfaces. Stokes' theorem does not extend to a Mabius band. model can easily be constructed from a long rectangular strip of paper by giving one end ahalf-twist and then fastening the two ends together. We define pl, Q~, C1, and C, for theMijbius band as we defined them for the cylinder above. The edge of S1 u S, in this case isone simple closed curve C', rather than two. This curve is called the complete boundaryof the Mbbius band. If we apply Stokes' theorem to each piece S1 and S,, as we did for the cylinder, we obtainEquation (12.51). But if we try to consolidate the two surface integrals and the two lineintegrals as we did above, we encounter two difficulties. First, the two normals n, and n2do not agree in direction everywhere on the intersection C, n C,. (See Figure 12.16.)Therefore we cannot define a normal n for the whole surface by taking n = n, on S1 andn = n,, on S,, as we did for the cylinder. This is not serious, however, because we can definen to be n, on S, and on C, n C, , and then define n to be n2 everywhere else. This gives adiscontinuous normal, but the discontinuities so introduced form a set of content zero inthe tlv-plane and do not affect the existence or the value of the surface integral ss (curl F). n dS.SlUSZ A more serious difficulty is encountered when we try to consolidate the line integrals.In this example it is not possible to choose the mappings rI and Ye in such a way that p1and pZ determine opposite directions on each arc of the intersection C, n C,. This isillustrated by the arrows in Figure 12.16; one of these arcs is traced twice in the samedirection. On this arc the corresponding line integrals will not necessarily cancel as they 7 After A. F. Mijbius (1790-1868), a pupil of Gauss. At the age of 26 he was appointed professor ofastronomy at Leipzig, a position he held until his death. He made many contributions to celestial mechanics,but his most important researches were in geometry and in the theory of numbers. 456 Surface in t(egral.7 did for the cylinder. Therefore the sum of the line integrals in (12.51) is not necessarilyequal to the line integral over the complete boundary of S, U &, and Stokes' theoremcannot be extended to the Mbbius band. Note: The cylinder and the Mobius band are examples oforientable and nonorientablesurfaces, respectively. We shall not attempt 1.0 define these terms precisely, but shallmention some of their differences. For an orientable surface S, u S, formed from twosmooth simple parametric surfaces as described above, the mappings r1 and r2 can alwaysbe chosen so that pI and pa determine opposite directions on each arc of the intersectionC, C I C, . Fora nonorientable surface no such choice is possible. For a smooth orientablesurface a unit normal vector can be defined in a 'continuous fashion over the entire surface.For a nonorientable surface no such definition (of a normal is possible. A paper mode1 ofan orientable surface always has two sides that can be distinguished by painting them withtwo different colors. Nonorientable surfaces have only one side. For a rigorous discussionof these and other properties of orientable and. nonorientable surfaces, see any book oncombinatorial topology. Stokes' theorem can be extended to orientable surfaces by aprocedure similar to that outlined above for the cylinder. Another orientable surface is the sphere shown in Figure 12.17. This surface is theunion of two simple parametric surfaces (hemispheres) Sr and S,, which we may consider FIGURE 12.17 Extension of Stokes' theorem to a sphere. images of a circular disk in the xy-plane under mappings y1 and r2, respectively. We giver, pl, ps, C,, C, the same meani,ngs as in the above examples. In this case the curves C1and C, are completely matched by the mapping r (they intersect along the equator), and thesurface S, U S, is said to be closed. Moreover, lrI and r2 can be chosen so that the directionsdetermined by p1 and pz are opposite on C1 and C, , as suggested by the arrows in Figure12.17. (This is why S, U S, is orientable.) If we apply Stokes' theorem to each hemisphereand add the results we obtain Equation (12.51) as before. The normals n, and n2 agree onthe intersection C1 n Cz, and we can consolidate the integrals over S1 and Sz into oneintegral over the whole sphere. The two line integrals on the right of (12.51) cancelcompletely, leaving us with the formula Jl (curlF).ndS = 0.SIUSZ This holds not only for a sphere, but for any orientable closed surface. The divergence theorem (Gauss' theorem) 457 12.19 The divergence theorem (Gauss' theorem) Stokes' theorem expresses a relationship between an integral extended over a surfaceand a line integral taken over the one or more curves forming the boundary of this surface.The divergence theorem expresses a relationship between a triple integral extended over asolid and a surface integral taken over the boundary of this solid. THEOREM 12.6. DIVERGENCE THEOREM. Let V be a solid in 3-space bounded by an orient-able closedsurface S, and let n be the unit outer normal to S. [f'F isa continuously differentiablevector jeld dejned on V, we have (12.53).JJj(divF)dxdydz = jj F.ndS.V s Note: If we express F and n in terms of their components, say andW, y, z) = P(x, y, z)i + Q&y, z1.i + W, y, zN n=cosui+cos/3j+cosyk, then Equation (12.53) can be written as (12.54)ap aQ aRz + - + aYaz dx dy dz = (PCOSCL + Qcosp + Rcosy)dS. V S Proof.~~

~~It suffices to establish the three equations apax dx dy dz = PcosadS,V s Mg dx dy dz = ssQcos[!dS, Vay S sss$f dx dy dz = J'sRcosydS, V s and add the results to obtain (12.54). We begin with the third of these formulas and proveit for solids of a very special type. Assume V is a set of points (x, y, z) satisfying a relation of the form gk Y) I z I.% Y) for (x, y) in T, where T is a connected region in the xy-plane, and f and g are continuous functions on T,with g(x, y) IfTheboundary surface S consists of an upper cap S , given by the explicit formula z = f(x, y) ;a lower part SZ , given by z = g(x, y) ; and (possibly) a portion S, of the cylinder generated 458 Surface integrals by a line moving parallel to the z-axis along the boundary of T. The outer normal to S has anonnegative z-component on S, , has a nonpositive component on S, , and is parallel to thexy-plane on S, . Solids of this type will be called "xy-projectable." (An example is shownin Figure 12.18.) They include all convex solids (for example, solid spheres, ellipsoids,cubes) and many solids that are not convex (for example, solid tori with axes parallel to thez-axis). The idea of the proof is quite simple. We express the triple integral as a double integralextended over the projection T. Then we show that this double integral has the same value FIGURE 12.18 An example of a solid that is xy-projectable. as the surface integral in question. We begin with the formula sss aR-dxdydz = j-1 [/gyI;:,Jgdz] dxdy.aZV T The one-dimensional integral with respect to z may be evaluated by the second fundamentaltheorem of calculus, giving us (12.55) Edxdydz = {Rb, y,f(x, ~11 - R[x, Y, g(x, ~11) dx dyeV T For the surface integral we can write (1 2 . 5 6) jjRcosydS=jjRcosydS+jjRcosydS+ jjRcosydS.x Xl X2 X3 On S, the normal n is parallel to the xy-plane, so cos y = 0 and the integral over S, is zero.On the surface S, we use the representation r(x, y) = xi + yi + f(x, y)k, The divergence theorem (Gauss' theorem) 459 and on S, we use the representation r(x, y) = xi + yj + g(x, y)k.On S, the normal n has the same direction as the vector product &/ax x &/ay , so we canwrite [see Equation (12.25), p. 4361 jj R cos y dS = j j R dx A dy = j j R[x, y, f(x, y)] dx dy .Sl 81 T On S, the normal n has the direction opposite to that of ar/ax x ar/ay so, by Equation(12.26), we have jj R cos y dS = - jj R dx A dy = - jj R[x, y, g(x, y)] dx dy .sz SZ T Therefore Equation (12.56) becomes jj R cos y dS = jj {Rk y,f(x, Y)I - Rh, Y, g(x, y)l)dxdy.s T Comparing this with Equation (12.55) we see that aRa, dx dy dz = Rcos7)dS. V s In the foregoing proof the assumption that V is xy-projectable enabled us to expressthe triple integral over V as a double integral over its projection T in the xy-plane. It isclear that if V is yz-projectable we can use the same type of argument to prove the identity g dx dy dz = P cos 0: dS ; V s and if V is xz-projectable we obtain 3 dx dy dz =ay QcosB'dS. V s Thus we see that the divergence theorem is valid for all solids projectable on all threecoordinate planes. In particular the theorem holds for every convex solid. A solid torus with its axis parallel to the z-axis is xy-projectable but not xz-projectableor yz-projectable. To extend the divergence theorem to such a solid we cut the torus intofour equal parts by planes through its axis parallel to the xz- and yz-planes, respectively,and we apply the divergence theorem to each part. The triple integral over the whole torusis the sum of the four triple integrals over the four parts. When we add the surface integrals overthe four parts we find that the contributions from the faces common to adjacent parts cancel 460 Surface integrals each other, since the outward normals have opposite directions on two such faces. There fore the sum of the surface integrals over the four parts is equal to the surface integral overthe entire torus. This example illustrates how the divergence theorem can be extended tocertain nonconvex solids. 12.20 Applications of the divergence theorem The concepts of curl and divergence of a vector field F = Pi + Qj + Rk were introducedin Section 12.12 by the formulas (12.57) and div F ap aQ aR=z+-+z ay (12.58) curlF= (g-F)i+ (E-g)j+ (g-E)k. To compute div F and curl F from these formulas requires a knowledge of the componentsof F. These components, in turn, depend on the choice of coordinate axes in 3-space. Achange in the position of the coordinate axes would mean a change in the components of Fand, presumably, a corresponding change in the functions div F and curl F. With the helpof Stokes' theorem and the divergence theorem we can obtain formulas for the divergenceand curl that do not involve the components of F.~~

~~These formulas show that the curl anddivergence represent intrinsic properties of the vector field F and do not depend on theparticular choice of coordinate axes. We discuss first the formula for the divergence. THEOREM 12.7. Let V(t) be a solid sphere of radius t > 0 with center at a point a in3-space, and let S(t) denote the boundary of V(t). Let F be a vectorfield that is continuouslydifferentiable on V(t). Then if 1 V(t)1 denotes the volume of V(t), and if n denotes the unitouter normal to S, we have (12.59) div F(a) = lim - F.ndS. Proof. Let P = div F. If E > 0 is given we must find a 6 > 0 such that da> - &//F ndSI IUOI = JJs ~(4 dx dy dz + ssJ [da> - dx)l dx dy dz .V(t) V(t) Applications of the divergence theorem 461 If we apply the divergence theorem to the first triple integral on the right and then trans-pose this term to the left, we obtain the relation Id4 - dx)l dx LAY dz I ; IV01 < E IV01 . V(t) When we divide this inequality by 1 V(t)1 we see that (12.60:) holds with 6 = h. This provesthe theorem. In the foregoing proof we made no special use of the fact that V(t) was a sphere. Thesame theorem holds true if, instead of spheres, we use any set of solids V(t) for whichthe divergence theorem is valid, provided these solids contain the point a and shrink toa as t --f 0. For example, each V(t) could be a cube inscribed in a sphere of radius t abouta; exactly the same proof would apply. Theorem 12.7 can be used to give a physical interpretation of the divergence. Suppose Frepresents the flux density vector of a steady flow. Then the surface integral JJ F - n dS S(t)measures the total mass of fluid flowing through S in unit time in the direction of n. Thequotient ,J;J F. n dS/l V(t)/ represents the mass per unit volume that flows through S in unit time in the direction of n. As t -+ 0, the limit of this quotient is the divergence of F ata. Hence the divergence at a can be interpreted as the time rate of change of mass per unitvolume per unit time at a. In some books on vector analysis, Equation (12.59) is laken as the dejnition of diver-gence. This makes it possible to assign a physical meaning to the divergence immediately.Also, formula (12.59) does not involve the components of F. Therefore it holds true inany system of coordinates. If we choose for V(t) a cube with its edges parallel to thexyz-coordinate axes and center at a, we can use Equation (12.59) to deduce the formulain (12.57) which expresses div F in terms of the components of F. This procedure is out-lined in Exercise 14 of Section 12.21. There is a formula analogous to (12.59) that is sometimes used as an alternative definitionof the curl.~~

~~It states that (12.61) curl F(u) = lim - n x FdS, where V(t) and S(t) have the same meanings as in Theorem 12.7. The surface integralthat appears on the right has a vector-valued integrand. Such integrals can be defined interms of components. The proof of (12.61) is analogous to that of Theorem 12.7. There is another formula involving the curl that can be deduced from (12.61) or derivedindependently.~~

~~It states that (12.62)1 n * curl F(a) = lim - I,* da.t 10 IS(t)1 C(t) In this formula, S(t) is a circular disk of radius t and center at a, and IS(t)1 denotes itsarea. The vector n is a unit normal to S(t), and a is the function that traces out C(t) in a 462 Surface integrals direction that appears counterclockwise when viewed from the tip of n. The vector fieldF is assumed to be continuously differentiable on S(t). A proof of (12.62) can be given bythe same method we used to prove (12.59). We let m(x) = n * curl F(x) and argue asbefore, except that we use surface integrals instead of triple integrals and Stokes' theoreminstead of the divergence theorem. If F is a velocity field, the line integral over C(t) is called the circulation of F along C(t);the limit in (12.62) represents the circulation per unit area at the point a. Thus, II * curl F(a) can be regarded as a "circulation density" of F at point a, with respect to a plane perpen-dicular to n at a. When n takes the successive values i, j, and k, the dot products i * curl F, j * curl F,and k . curl I: are the components of curl F in rectangular coordinates. When Equation(12.61) is taken as the starting point for the definition of curl, the formula in (12.58) for therectangular components of curl F can be deduced from (12.62) in exactly this manner. 12.21 Exercises I. Let S be the surface of the unit cube, 0 5 x < 1, 0 5 y 5 1, 0 I z 5 1, and let n be theunit outer normal to S.~~

~~If F(x, y, z) = x2i + y2j $ z2k, use the divergence theorem to evaluatethe surface integral Jj I,. n dS. Verify the result by evaluating the surface integral directly. 2. The sphere x2 + y2 "+ z2 = 25 is intersected by the plane z = 3. The smaller portion forms asolid Y bounded by a closed surface S,, made up of two parts, a spherical part S, and a~~

planarpart S, . If the unit outer normal of V is cos a i + cos B j + cos y k, compute the value of thesurface integral J's(xz cos u + yz cos B + cos y) dS s if (a) S is the spherical cap St, (b) S is the planar base S, , (c) S is the complete boundary S, .Solve part (c) by use of the results of parts (a) and (b), and also by use of the divergencetheorem. 3.
Let n = cos a i + cos /l j + cos y k be the unit outer normal to a closed surface S whichboundsa homogeneous solid V of the type described in the divergence theorem. Assume that thecenter of mass (2, 8, i) and the volume 1 V(of V are known.
Evaluate the following surfaceintegrals in terms of 1 V(and X, 7, i.
(a> rj(xcosa +ycosp +zcosy)dS.s (b) JJ (xz cos a + 2yz cos /? + 3z2 cos y) dS.s (c) jr 01" cos a + 2xy cos /3 - xz cos y) dS.s (d) Express JJ (x2 + y2)(xi + yj) n dS in terms of the volume I VI and a moment of inertiaof the solid. S In Exercises 4 through 10, af/ an and ag/ an denote directional derivatives of scalar fields f and gin the direction of the unit
outer normal n to a closed surface S which bounds a solid V of the type Exercises 463-.
described in the divergence theorem. That is, af/ an = Vf ' n and ag/ an = Vg . n . In each of theseexercises prove the given statement. You may assume continuity of all derivatives involved. 5.
gdS=O whenever f is harmonic in V. 6.
jif;dS = ~~~fVzgdxdydz + Jj-1 vf. vgdxdj'dz.S v V 8. //f;dS = //g$dS if both f and g are harmonic in V.
9. jif~~~=~~~,Vf,sbdv" if f is harmonic in V. S V 110. V2f(a) = lim - t-0 I v(t)1-gdS, where V(t) is a solid sphere of radius I with center at II, is(t)S(t) is the surface of V(t), and 1 V(t)1 is the volume of V(t). 11. Let V be a convex region in 3-space whose boundary is a closed surface S and let n be theunit outer normal to S. Let F and G be two
continuously differentiable vector fields such that curl F = curl G and div F = div G everywhere in V, and such thatG.n = F.n everywhere on S , Prove that F = G everywhere in V.
[Hint: Let H = F - G, find a scalar field f such thatH = Vf, and use a suitable identity to prove that JJj II Of (I2 dx du dz = 0. From this deducethat H = 0 in V.] V 12. Given a vector field G and two scalar fields f and g, each continuously differentiable on aconvex solid V bounded by a closed surface S. Let n denote the unit outer normal to S. Provethat
there is at most one vector field F satisfying the following three conditions: curl F = G and div F = g in V, F.n =f on S. 13. Let S be a smooth parametric surface with the property that each line emanating from a pointP intersects S once at most. Let Q(S) denote the set of lines emanating from P and passingthrough S. (See Figure 12.19.) The set n(S) is
called the solid angle with vertex P subtendedby S. Let E(a) denote the intersection of n(S) with the surface of the sphere of radius acentered at P. The quotient area of X(a)a2 464 Surface integrals FIGURE 12.19 The solid angle n(S) with vertex P subtended by a surface S. It is measured by the quotient ISZ(S)J =area of Z(a) a2 .
is denoted by] n(S)] and is used as a measure of the solid angle n(S).(a) Prove that this quotient is equal to the surface integral 'Let S denote the smaller portion of the surface of the sphereintercepted by the two planes. Show that 1 n(S)] = 2%. 14. Let k'(t) denote a cube of edge 2t and center at (I, and let S(t) denote the boundary of thecube. Let n be
the unit outer normal of S(t) and let) V(t)1 denote the volume of the cube.For a given vector field F that is continuously differentiable at a, assume that the followinglimit exists : 1 ','t", I W)l4F.ndS, S(t)and use this limit as the definition of the divergence, div F(a). Choose xyz-coordinate axesparallel to the edges of V(t) and let P, Q, and R be the
components of F relative to this coordi-nate system. Prove that div F(a) = &P(u) + &Q(u) + &R(a). [Hint: Express the surface Exercises 465 integral as a sum of six double integrals taken over the faces of the cube. Then show thatl/l k'(t)/ times the sum of the two double integrals over the faces perpendicular to the z-axisapproaches the limit D&a) as t
-f 0. Argue in a similar way for the remaining terms.] 15. A scalar field c, which is never zero has the properties llV~l12 = 4q and div (v,Oq) = 10~. Evaluate the surface integral where S is the surface of a unit sphere with center at the origin, and aplan is the directionalderivative of 'p in the direction of the unit outer normal to S. PART 3 SPECIAL
TOPICS 13SET FUNCTIONS AND ELEMENTARY PROBABILITY 13.1 Historical introduction A gambler's dispute in 1654 led to the creation of a mathematical theory of probability bytwo famous French mathematicians, Blaise Pascal and Pierre de Fermat. AntoineGombaud, Chevalier de M&C, a French nobleman with an interest in gaming and
gamblingquestions, called Pascal's attention to an apparent contradiction concerning a popular dicegame. The game consisted in throwing a pair of dice 24 times; the problem was to decidewhether or not to bet even money on the occurrence of at least one "double six" duringthe 24 throws. A seemingly well-established gambling rule led de M&C to
believe thatbetting on a double six in 24 throws would be profitable, but his own calculations indicatedjust the opposite. This problem and others posed by de MCrC led to an exchange of letters between Pascaland Fermat in which the fundamental principles of probability theory were formulatedfor the first time. Although a few special problems on
games of chance had been solvedby some Italian mathematicians in the 15th and 16th centuries, no general theory wasdeveloped before this famous correspondence.
The Dutch scientist Christian Huygens, a teacher of Leibniz, learned of this corre-spondence and shortly thereafter (in 1657) published the first book on probability; entitledDe Ratiociniis in Ludo Alcae, it was a treatise on problems associated with gambling.Because of the inherent appeal of games of chance, probability theory soon became
popular,and the subject developed rapidly during the 18th century. The major contributors duringthis period were Jakob Bernoulli? (1654-1705) and Abraham de Moivre (1667-1754).
In 1812 Pierre de Laplace (1749-1827) introduced a host of new ideas and mathematicaltechniques in his book, Thkorie Analytique des Probabilitb.
Before Laplace, probabilitytheory was solely concerned with developing a mathematical analysis of games of chance.Laplace applied probabilistic ideas to many scientific and practical problems. The theoryof errors, actuarial mathematics, and statistical mechanics are examples of some of theimportant applications of probability theory developed in
the 19th century. Like so many other branches of mathematics, the development of probability theoryhas been stimulated by the variety of its applications. Conversely, each advance in thetheory has enlarged the scope of its influence. Mathematical statistics is one importantbranch of applied probability; other applications occur in such widely
different fields as t Sometimes referred to as James Bernoulli. 470 Set functions and elementary probability genetics, psychology, economics, and engineering.
Many workers have contributed to thetheory since Laplace's time; among the most important are Chebyshev, Markov, vonMises, and Kolmogorov. One of the difficulties in developing a mathematical theory of probability has been toarrive at a definition of probability that is precise enough for use in mathematics, yetcomprehensive enough to be
applicable to a wide range of phenomena. The search for awidely acceptable definition took nearly three centuries and was marked by muchcontroversy. The matter was finally resolved in the 20th century by treating probabilitytheory on an axiomatic basis.
In 1933 a monograph by a Russian mathematicianA. Kolmogorov outlined an axiomatic approach that forms the basis for the modern theory.(Kolmogorov's monograph is available in English translation as Foundations of ProbabilityTheory, Chelsea, New York, 1950.) Since then the ideas have been refined somewhat andprobability theory is now part of
a more general discipline known as measure theory. This chapter presents the basic notions of modern elementary probability theory alongwith some of its connections to measure theory. Applications are also given, primarily togames of chance such as coin tossing, dice, and card games. This introductory account isintended to demonstrate the logical
structure of the subject as a deductive science and togive the reader a feeling for probabilistic thinking. 13.2 Finitely additive set functions The area of a region, the length of a curve, or the mass of a system of particles is anumber which measures the size or content of a set. All these measures have certainproperties in common. Stated abstractly,
they lead to a general concept called a finitelyadditive set function. Later we shall introduce probability as another example of such afunction. To prepare the way, we discuss first some properties common to all thesefunctions. A functionf: & + R whose domain is a collection JZZ of sets and whose function valuesare real numbers, is called a set
function. If A is a set in class &, the value of the functionat A is denoted by f(A). DEFINITION OF A FINITELY ADDITIVE SET FUNCTION. A set function f: d - + R is said to befinitely additive if (13.1) f(A u B) =fV) +fW whenever A and B are disjoint sets in ~2' such that A U B is also in &. Area, length, and mass are all examples of finitely additive set
functions. This sectiondiscusses further consequences of Equation (13.1). In the usual applications, the sets in ~2 are subsets of a given set S, called a universal set.It is often necessary to perform the operations of union, intersection, and complementationon sets in &. To make certain that & is closed under these operations we restrict & to bea
Boolean algebra, which is defined as follows. Finitely additive measures 471 DEFINITION OF A BOOLEAN ALGEBRA OF SETS. A nonempty class & of subsets of a givenuniversal set S is called a Boolean algebra if for every A and B in d we have AuBE& and A'E22. Here A' = S - A, the complement of A relative to S. A Boolean algebra &? is also closed
under intersections and differences, since we have A n B = (A' u B')' and A-B=AnB'. This implies that the empty set 0 belongs to Se since ,Z = A - A for some A in &.Also, the universal set S is in JZ? since S = 0 ' . Many Boolean algebras can be formed from the subsets of a given universal set S. Thesmallest such algebra is the class d, = { 0 , S} , which
consists of only two special subsets:ia and S. At the other extreme is the class zJca,, which consists of all subsets of S.
EveryBoolean algebra J&' consisting of subsets of S satisfies the inclusion relations s9, c JXI E&I. The property of finite additivity of set functions in Equation (13.1) requires A and B tobe disjoint sets. The next theorem drops this requirement. T H E O R E M 13.1. Let f: JY - + R be a finitely additive set function defined on a Booleanalgebra &' of sets.
Then for all A and B in .& we have (13.2) f(A u 8 =fW +f(B - 4, and (13.3) f(A u B) = f(A) + f(B) - f(A n B) . Proof. The sets A and B - A are disjoint and their union is A U B. Hence, by applying(13.1) to A and B - A we obtain (13.2). To prove (13.3) we first note that A n B' and B are disjoint sets whose union is A U B.Hence by (13.1) we have (13.4) f (A u
B) = f (A n B') + f(B) I Also, A n B' and A n B are disjoint sets whose union is A, so (13.1) gives us (13.5) f(A) = f(A n B') + f(A n B) . Subtracting (13.5) from (13.4) we obtain (13.3). 13.3 Finitely additive measures The set functions which represent area, length, and mass have further properties incommon.
For example, they are all nonnegative set functions. That is, for each A in the class @' under consideration. Set functions and elementary probability DEFINITION OF A FINITELY ADDITIVE MEASURE. A nonnegative set function f: J# + R that isfinitely additive is called a finitely additive measure, or simply a measure. Using Theorem 13.1 we
immediately obtain the following further properties of measures. THEOREM 13.2. Let f: SI + R be ajnitely additive measure dejned on a Boolean algebra&.Then for all sets A and B in ~2 we have (a> f (A U B) I SW + f W .(b) f(B - 4 =fW -fW i f AGB.Cc> f(A) < f(B) if A c B. (Monotone property) (4 f(O) = 0. Proof. Part (a) follows from (13.3), and part (b)
follows from (13.2). Part (c) followsfrom (b), and part (d) follows by taking A = B = 0 in (b). EXAMPLE. The number of elements in a finite set. Let S = {a,, a2, .
-
. , a,} be a setconsisting of n (distinct) elements, and let & denote the class of all subsets of S. For eachA in JZZ, let y(A) denote the number of distinct elements in A (V is the Greek letter "nu").It is easy to verify that this function is finitely additive on &. In fact, if A has k elementsand if B has m elements, then v(A) = k and V(B) = m .
If A and B are disjoint it is clearthat the union A V B is a subset of S with k + m elements, so This particular set function is nonnegative, so v is a measure. 13.4 Exercises 1. Let & denote the class of all subsets of a given universal set and let A and B be arbitrary setsin & . Prove that:(a) A n B' and B are disjoint.(b) A u B = (A n B') u B. (This formula
expresses A u B as the union of disjoint sets.)(c) A n B and A n B' are disjoint.(d) (A n B) u (A n B') 7 A. (This formula expresses A as a union of two disjoint sets.) 2.
Exercise l(b) shows how to express the union of two sets as the union of two disjoint sets.Express in a similar way the union of three sets '4, u A, v A, and, more generally, of n setsA,uA,u~~~uA,. Illustrate with a diagram when II = 3. 3. A study of a set S consisting of 1000 college graduates ten years after graduation revealedthat the "successes"
formed a subset A of 400 members, the Caltech graduates formed a subsetB of 300 members, and the intersection A n B consisted of 200 members.(a) For each of the following properties, use set notation to describe, in terms of unions andintersections of A, B, and their complements A' and B' relative to S, the subsets consisting ofthose persons in S
that have the property: (i) Neither a "success" nor a Caltech graduate.(i) A "success" but not a Caltech graduate.(iii) A "success" or a Caltech graduate, or both. The dejnition of probability for finite sample spaces 413 (iv) Either a "success" or a Caltech graduate, but not both.(v) Belongs to not more than one of A or B. (b) Determine the exact number
of individuals in each of the five subsets of part (a).4. Letfbe a finitely additive set function defined on a class .sfl of sets. Let A,, ...
, A, be n sets in &' such that Ai n Aj = 0 if i # j. (Such a collection is called a disjoint collection of sets.) If the union 6 Al, is in class & for all m < n, use induction to prove thatk=l f (kcl Ak) = kzlf(Ak) . In Exercises 5, 6, and 7, S denotes a finite set consisting of n distinct elements, say S =h.~2,.~~~,GJ. 5. Let A, = {al}, the subset consisting of a, alone.
Show that the class &?r = { 121, A,, A,; S} isthe smallest Boolean algebra containing A,. 6. Let A, = {al}, A, = {u2}, Describe, in a manner similar to that used in Exercise 5, the smallestBoolean algebra gz containing both A, and A,. 7. Do the same as in Exercise 6 for the subsets A, = {al}, Aa = {u2}, and A, = {ua}.8.
If gk denotes the smallest Boolean algebra which contains the k subsets A, = {al}, A, = (4, . .
. , Al, = {Us}, show that gk contains 2"+l subsets of S if k < n and 2n subsets if k = n.9. Letfbe a finitely additive set function defined on the Boolean algebra of all subsets of a given universal set S.
Suppose it is known that fU A B) =f(A)f(W for two particular subsets A and B of S. Iff(S) = 2, prove that f(A " B) = f(A') + f(B') - f(A')f(B') .
10. If A and B are sets, their symmetric difference A n B is the set defined by the equation A A B =(A - B) u (B - A). Prove each of the following properties of the symmetric difference.(a) AAB=BAA.(b)AAA=@.(c) AABr(AAC)u(CAB).(d) A a B is disjoint from each of A and B.(e) (A AB) AC=AA(BAC).(f) Iffis a finitely additive set function defined on the
Boolean algebra d of all subsets of agiven set S, then for all A and B in d we havef(A A B) =f(A) +f(B) - 2f(A n B). 13.5 The definition of probability for finite sample spaces In the language of set functions, probability is a specific kind of measure (to be denotedhere by P) defined on a specific Boolean algebra g of sets. The elements of 3 are subsetsof a
universal set S. In probability theory the universal set S is called a sample space.
Wediscuss the definition of probability first for finite sample spaces and later for infinite samplespaces. DEFINITION OF PROBABILITY FOR FINITE SAMPLE SPACES. Let 9I denote a Boolean algebrawhose elements are subsets of a givenJinite set S. A set jbnction P dejined on 93' is called a 474 Set functions and elementary probability probabilitjv
measure if it satisfies the following three conditions.(a) P isfinitetv additive.(b) P is nonnegative.(c) P(S) = 1. In other words, for finite sample spaces probability is simply a measure which assigns thevalue 1 to the whole space. It is important to realize that a complete description of a probability measure requiresthree things to be specified: the sample
space S, the Boolean algebra 9? formed from certainsubsets of S, and the set function P. The triple (S, 9, P) is often called aprobability space.In most of the elementary applications the Boolean algebra 3' is taken to be the collectionof all subsets of S. EXAMPLE. An illustration of applied probability theory is found in the experiment oftossing a coin
once. For a sample space S we take the set of all conceivable outcomes ofthe experiment. In this case, each outcome is either "heads" or "tails," which we label by thesymbols h and t. Thus, the sample space S is {h, t}, the set consisting of h and t. For theBoolean algebra we take the collection of all subsets of S; there are four, ,@, S, H, and T,where H
= {h} and T = {t}. Next, we assign probabilities to each of these subsets. Forthe subsets ia and S we have no choice in the assignment of values. Property (c) requiresthat P(S) = 1, and, since P is a nonnegative measure, P(D) = 0. However, there is somefreedom in assigning probabilities to the other two subsets, Hand T.
Since H and Taredisjoint sets whose union is S, the additive property requires that pm + P(T) = P(S) = 1. We are free to assign any nonnegative values whatever to P(H) and P(T) so long as theirsum is 1.
If we feel that the coin is unbiased so that there is no apriori reason to preferheads or tails, it seems natural to assign the values P(H) = P(T) = 4. If, however, the coin is "loaded," we may wish to assign different values to these twoprobabilities. For example, the values P(H) = 4 and P(T) = $ are just as acceptable asP(H) = P(T) = 4. In fact, for any real
p in the interval 0 < p 5 1 we may defineP(H) = p and P(T) = 1 - p, and the resulting function P will satisfy all the conditionsfor a probability measure. For a given coin, there is no mathematical way to determine what the probability p"really" is. If we choosep = 4 we can deduce logical consequences on the assumption thatthe coin is fair or unbiased.
The theory for unbiased coins can then be used to test thefairness of an actual coin by performing a large number of experiments with the coin andcomparing the results with the predictions based on the theory. The testing of agreementbetween theory and empirical evidence belongs to that branch of applied probability knownas statistical inference,
and will not be discussed in this book.
The foregoing example is a typical application of the concepts of probability theory.Probability questions often arise in situations referred to as "experiments." We shall not Special terminology peculiar to probability theory 415 attempt to define an experiment; instead, we shall merely mention some familiar examples:tossing one or more coins, rolling
a pair of dice, dealing a bridge hand, drawing a ball froman urn, counting the number of female students at the California Institute of Technology,selecting a number from a telephone directory, recording the radiation count of a Geigercounter. To discuss probability questions that arise in connection with such experiments, ourfirst task is to construct
a sample space S that can be used to represent all conceivableoutcomes of the experiment, as we did for coin tossing. Each element of S should representan outcome of the experiment and each outcome should correspond to one and only oneelement of S. Next, we choose a Boolean algebra L%' of subsets of S (usually all subsets ofS) and then define
a probability measure P on g.
The choice of the set S, the choice of ~23,and the choice of P will depend on the information known about the details of the experi-ment and on the questions we wish to answer. The purpose of probability theory is not todiscuss whether the probability space (S, LAY', P) has been properly chosen. This motivationbelongs to the science or gambling
game from which the experiment emanates, and onlyexperience can suggest whether or not the choices were well made. Probability theory is thestudy oj'logical consequences that can be derived once the probability space is given. Makinga good choice of the probability space is, strictly speaking, not probability theory - it isnot even mathematics;
instead, it is part of the art of applying probability theory to thereal world. We shall elaborate further on these remarks as we deal with specific examplesin the later sections.
If S = {a,, a,, . . . , a,}, and if z% consists of all subsets of S, the probability function Pis completely determined if we know its values on the one-element subsets, or singletons: P(h>>, P(hH, . . . , P({a,l).In fact, every subset A of S is a disjoint union of singletons, and P(A) is determined by theadditive property. For example, when A = {al} U {az} U . * .
U {ok}, the additive property requires that W) = i: fY{aJ).i=l To simplify the notation and the terminology, we write P(a,) instead of P({a,}). Thisnumber is also called the probability C$ the point ai . Therefore, the assignment of the pointprobabilities P(x) for each element x in a finite set S amounts to a complete description ofthe probability function
P. 13.6 Special terminology peculiar to probability theory In discussions involving probability, one often sees phrases from everyday languagesuch as "two events are equally likely," "an event is impossible," or "an event is certainto occur." Expressions of this sort have intuitive appeal and it is both pleasant and helpfulto be able to employ such
colorful language in mathematical discussions. Before we can doso, however, it is necessary to explain the meaning of this language in terms of the funda-mental concepts of our theory. Because of the way probability is used in practice, it is convenient to imagine that each 476 Set functions and elementary probability probability space (S, 2?, P) is
associated with a real or conceptual experiment. Theuniversal set S can then be thought of as the collection of all conceivable outcomes of theexperiment, as in the example of coin tossing discussed in the foregoing section. Eachelement of S is called an outcome or a sample and the subsets of S that occur in the Booleanalgebra 9? are called events.
The reasons for this terminology will become more apparentwhen we treat some examples. Assume we have a probability space (S, 9Y, P) associated with an experiment. Let Abe an event, and suppose the experiment is performed and that its outcome is X. (In otherwords, let x be a point of S.) This outcome x may or may not belong to the set A. Ifit
does, we say that the event A has occurred. Otherwise, we say that the event A has notoccurred, in which case x E A', so the complementary event A' has occurred. An event Ais called impossible if A = 0 , because in this case no outcome of the experiment can bean element of A. The event A is said to be certain if A = S, because then every outcomeis
automatically an element of A. Each event A has a probability P(A) assigned to it by the probability function P. [Theactual value of P(A) or the manner in which P(A) is assigned does not concern us atpresent.] The number P(A) is also called the probability that an outcome of the experimentis one of the elements of A. We also say that P(A) is the
probability that the event A occurswhen the experiment is performed.
The impossible event @ must be assigned probability zero because P is a finitely additivemeasure. However, there may be events with probability zero that are not impossible.In other words, some of the nonempty subsets of S may be assigned probability zero. Thecertain event S must be assigned probability 1 by the very definition of probability,
butthere may be other subsets as well that are assigned probability 1.
In Example 1 of Section13.8 there are nonempty subsets with probability zero and proper subsets of S that haveprobability 1. Two events A and B are said to be equally likely if P(A) = P(B).
The event A is calledmore likely than B if P(A) > P(B), and at least as likely as B if P(A) > P(B). Table 13.1provides a glossary of further everyday language that is often used in probability discus-sions. The letters A and B represent events, and x represents an outcome of an experimentassociated with the sample space S. Each entry in the left-hand
column is a statementabout the events A and B, and the corresponding entry in the right-hand column defines thestatement in terms of set theory. TABLE 13.1. Glossary of Probability Terms Statement Meaning in set theory At least one of A or B occursBoth events A and B occurNeither A nor B occursA occurs and B does not occurExactly one of A or
B occursNot more than one of A or B occursIf A occurs, so does B (A implies B)A and B are mutually exclusiveEvent A or event BEvent A and event B - XEA uBXEA nB XEA' n B'xgAnB'EA' n B'xgAnB' x E (A n B') u (A' n B)x E (A n B)' AsBAnB=aAuBAnB Worked examples 411 13.7 Exercises Let S be a given sample space and let A, B, and C denote arbitrary events
(that is, subsets of Sin the corresponding Boolean algebra g). Each of the statements in Exercises 1 through 12 isdescribed verbally in terms of A, B, C. Express these statements in terms of unions and intersectionsof 11, B, C and their complements. 1. If A occurs, then B does not occur. 7. At least two of A, B, C occur.2. None of the events A, B, C
occurs. 8. Exactly two occur.3. Only A occurs. 9. Not more than two occur.4. At least one of A, B, C occurs. 10.
A and C occur but not B.5. Exactly one of A, B, C occurs. Il. All three events occur.6. Not more than one occurs. 12. Not more than three occur. 13. Let A denote the event of throwing an odd total with two dice, and let B denote the event ofthrowing at least one 6. Give a verbal description of each of the following events: (a) A u B, (d) A' n B,@)A nB, (e)
A' n B',(c) A n B', (f) A' u B. 14. Let A and B denote events. Show that JYA n B) 5 P(A) I f'(A u B) 5 P(A) + P(B). 15. Let A and B denote events and let a = P(A), b = P(B), c = P(A n B) . Compute, in terms ofa, b, and c, the probabilities of the following events: (a) A', (d) A' u B',(b) B', (e) A' u B,Cc> A u B, (f) A n B'. 16. Given three events A, B, C. Prove that
13.8 Worked examples We shall illustrate how some of the concepts of the foregoing sections may be used toanswer specific questions involving probabilities.
EXAMPLE 1. What is the probability that at least one "head" will occur in two throwsof a coin? First Solution. The experiment in this case consists of tossing a coin twice; the set Sof all possible outcomes may be denoted as follows: S = {hh, ht, th, tt}. If we feel that these outcomes are equally likely, we assign the point probabilities P(x) = 4for each x
in S. The event "at least one head occurs" may be described by the subset A = {hh, ht, th}. 478 Set functions and elementary probability The probability of this event is the sum of the point probabilities of its elements. Hence,P(A)=$+f+~=g.
Second Solution. Suppose we use the same sample space but assign the point proba-bilities as follows :t P(hh) = 1) P(h) = P(h) = lJ(tt) = 0. Then the probability of the event "at least one head occurs" is P(hh) + P(ht) + P(th) = I + 0 + 0 = 1 . The fact that we arrived at a different answer from that in the first solution should notalarm the reader. We
began with a different set of premises. Psychological considerationsmight lead us to believe that the assignment of probabilities in the first solution is the morenatural one. Indeed, most people would agree that this is so if the coin is unbiased.
How-ever, if the coin happens to be loaded so that heads always turns up, the assignment ofprobabilities in the second solution is more natural. The foregoing example shows that we cannot expect a unique answer to the questionasked. To answer such a question properly we must specify the choice of sample spaceand the assignment of point
probabilities. Once the sample space and the point proba-bilities are known only one probability for a given event can be logically deduced. Differentchoices of the sample space or point probabilities may lead to different "correct" answersto the same question. Sometimes the assignment of probabilities to the individual outcomes of an experimentis
dictated by the language used to describe the experiment. For example, when an objectis chosen "at random" from a finite set of II elements, this is intended to mean that eachoutcome is equally likely and should be assigned point probability l/n. Similarly, when wetoss a coin or roll a die, if we have no apriori reason to feel that the coin or die is
loaded,we assume that all outcomes are equally likely. This agreement wiI1 be adopted in all theexercises of this chapter. EXAMPLE 2. If one card is drawn at random from each of two decks, what is the proba-bility that at least one is the ace of hearts? Solution. The experiment consists in drawing two cards, a and b, one from each deck.Suppose we
denote a typical outcome as an ordered pair (a, b). The number of possibleoutcomes, that is, the total number of distinct pairs (a, b) in the sample space S is 522.We assign the probability l/522 to each such pair. The event in which we are interestedis the set A of pairs (a, b), where either a or b is the ace of hearts. There are 52 + 51elements in A.
Hence, under these assumptions we deduce that 52+51 1 1P(A) = ~ - - - . 522 = 26 522 t Note that for this assignment of probabilities there are nonempty subsets of S with probability zero andproper subsets with probability 1.
Exercises 419 EXAMPLE 3. If two cards are drawn at random from one deck, what is the probabilitythat one of them is the ace of hearts? Solution. As in Example 2 we use ordered pairs (a, b) as elements of the sample space.In this case the sample space has 52 * 51 elements and the event A under considerationhas 51 + 51 elements. If we assign the
point probabilities l/(52 * 51) to each outcome,we obtain P(A) = g+ = &.
EXAMPLE 4.
What is the probability of throwing 6 or less with three dice? Solution. We denote each outcome of the experiment as a triple of integers (a, b, c)where a, 6, and c may take any values from 1 to 6. Therefore the sample space consistsof 63 elements and we assign the probability 1/63 to each outcome. The event A in questionis the set of all triples
satisfying the inequality 3 < a + b + c 5 6. If A, denotes theset of (a, 6, c) for which a + b + c = n, we have A =A, UA, WA, UA,.
Direct enumeration shows that the sets A,, with n = 3, 4, 5 , and 6 contain 1 , 3 , 6, and 10elements, respectively. For example, the set A, is given by A, = {Cl,27 3), (1, 3,2), (1, 1,4), (1,4, I), (2, 1,3), (2 , 3, 11, (2,2,2), (3 , 1,2), (392, 11, (4, 1, 1)). Therefore A has 20 elements and P(A) = ; = ; . EXAMPLE 5. A die is thrown once. What is the probability
that the number of pointsis either even or a multiple of 3 ?
Solution. We choose the sample space S = {1,2, 3,4, 5,6}, consisting of six elements,to each of which we assign the probability B. The event "even" is the set A = {2,4, 6))the event "a multiple of 3" is B = (3, 6). We.are interested in their union, which is theset A U B = (2, 3,4, 6). Since this set contains four elements we have P(A U B) = 4/6. This
example can be solved in another way, using the formula P(AUB)=P(A)+P(B)-l'(AnB)=$+s-;g. 13.9 Exercises1. Let S be a finite sample space consisting of n elements. Suppose we assign equal probabilities to each of the points in S. Let A be a subset of S consisting of k elements. Prove that P(A) = k/n. For each of Exercises 2 through 8, describe your
choice of sample space and state how you are 480 Set functions and elementary probability assigning the point probabilities. In the questions associated with card games, assume all cardshave the same probability of being dealt. 2.
Five counterfeit coins are mixed with nine authentic coins.(a) A coin is selected at random. Compute the probability that a counterfeit coin is selected.If two coins are selected, compute the probability that:(b) one is good and one is counterfeit.(c) both are counterfeit.(d) both are good. 3. Compute the probabilities of each of the events described in
Exercise 13 of Section 13.7. Assignequal probabilities to each of the 36 elements of the sample space. 4. What is the probability of throwing at least one of 7, 11, or 12 with two dice?5. A poker hand contains four hearts and one spade. The spade is discarded and one card is drawn from the remainder of the deck. Compute the probability of filling the
flush-that is, ofdrawing a fifth heart. 6. In poker, a straight is a five-card sequence, not necessarily all of the same suit. If a poker handcontains four cards in sequence (but not A234 or JQKA) and one extra card not in the sequence,compute the probability of filling the straight. (The extra card is discarded and a new card isdrawn from the remainder of
the deck.) 7. A poker hand has four cards out of a five-card sequence with a gap in the middle (such as5689), and one extra card not in the sequence. The extra card is discarded and a new one isdrawn from the remainder of the deck. Compute the probability of filling the "inside straight." 8. An urn contains A white stones and B black stones. A second
urn contains C white stones andD black stones. One stone is drawn at random from the first urn and transferred to the secondurn. Then a stone is drawn at random from the second urn. Calculate the probability of eachof the following events.(a) The first stone is white.(b) The first stone is black.(c) The second stone is white, given that the transferred
stone was white.(d) The second stone is white, given that the transferred stone was black. 9. Two stones are drawn with replacement from an urn containing four red stones and two whitestones. Calculate the probability of each of the following events.(a) Both stones are white.(b) Both stones are red.(c) Both stones are the same color.(d) At least one
stone is red. 10. Let P, denote the probability that exactly 12 of the events A and B will occur, where n takes thevalues 0, 1, 2. Express each of P,, P,, P, in terms of P(A), P(B), and P(A n B).
Odds. Some gambling games are described in terms of "odds" rather than in terms ofprobabilities.
For example, if we roll a fair die, the probability of the event "rolling a three" is $..Since there are six possible outcomes, one of which is favorable to the event "rolling a three" andfive of which are unfavorable, this is often described by saying that the odds in favor of the eventare 1 to 5, or the odds against it are 5 to 1. In this case the odds are
related to the probability bythe equation 1 1-=1+5*6 In general, if A is an event with probability P(A) and if a and b are two real numbers such that (13.6) P(A) = sb , Some basic principles of combinatorial analysis 481 we say the odds in favor of A are a to b, or the odds against .4 are b to a. Since 1 - a/(a + b) =b/(a + b) , the odds against A are the
same as the odds in favor of the complementary event A'.The following exercises are devoted to further properties of odds and their relation to probabilities. 11. If P(A) = 1, show that (13.6) can be satisfied only when b = 0 and a z 0. If P(A) # 1,show that there are infinitely many choices of a and b satisfying (13.6) but that all have thesame ratio a/b.
12. Compute the odds in favor of each of the events described in Exercise 2.13. Given two events A and B. If the odds against A are 2 to 1 and those in favor of A u B are 3 to 1, show that& IWO 52. Give an example in which P(B) = & and one in which P(B) = 2. 13.10 Some basic principles of combinatorial analysis Many problems in probability theory
and in other branches of mathematics can bereduced to problems on counting the number of elements in a finite set. Systematic methodsfor studying such problems form part of a mathematical discipline known as combinatorialanabsis. In this section we digress briefly to discuss some basic ideas in combinatorialanalysis that are useful in analyzing
some of the more complicated problems of probabilitytheory. If all the elements of a finite set are displayed before us, there is usually no difficulty incounting their total number. More often than not, however, a set is described in a waythat makes it impossible or undesirable to display all its elements. For example, we mightask for the total number of
distinct bridge hands that can be dealt. Each player is dealt13 cards from a 52-card deck. The number of possible distinct hands is the same as thenumber of different subsets of 13 elements that can be formed from a set of 52 elements.Since this number exceeds 635 billion, a direct enumeration of all the possibilities is clearlynot the best way to
attack this problem; however, it can readily be solved by combinatorialanalysis.
This problem is a special case of the more general problem of counting the number ofdistinct subsets of k elements that may be formed from a set of it elements,7 where n 2 k .Let us denote this number byf(n, k).
It has long been known that (13.7) An, k) = 0;: 'where, as usual (3 denotes the binomial coefficient, n0 n.Ik = k!(n - k)!' In the problem of bridge hands we have f(52, 13) = (:i) = 635,013,559,600 differenthands that a player can be dealt. t When we say that a set has n elements, we mean that it has n distinct elements. Such a set is sometimescalled an
n-element set. 482 Set functions and elementary probability There are many methods known for proving (13.7). A straightforward approach is toform each subset of k elements by choosing the elements one at a time.
There are npossibilities for the first choice, n - 1 possibilities for the second choice, and n - (k - 1)possibilities for the kth choice. If we make all possible choices in this manner we obtaina total of In(n - 1). * + k+l)=(n~.k)! subsets of k elements. Of course, these subsets are not all distinct. For example, if k = 3the six subsets ia, b, 4, {b, c, a>, {c, a, bl,
{a,~, bl, {c, b, 4, (6 a, 4 arc all equal. In general, this method of enumeration counts each k-element subset exactlyk! times.7 Therefore we must divide the number n!/(n - k)! by k! to obtainf(n, k). Thisgives usf(n, k) = (;) , as asserted. This line of argument is more or less typical of the combinatorial analysis required inthe later sections. Hence it seems
worthwhile to digress briefly to discuss the fundamentalprinciples on which this analysis is based. We often wish to count the number of elements in the Cartesian product of n finite setsA . . . . A The Cartesian product is denoted by the symbol A, x * * . x A, and isdt&ed by tck equation A, x ..a x A,={(a, ,..., a,)/a,~A~ ,..., ~,EA,}. That is, the Cartesian
product consists of the set of all ordered n-tuples (a,, . . . , a,) wherethe kth component of the n-tuple comes from the kth set A,. An example with n = 2 is shown in Figure 13.1. Here A, = { 1, 2,4, 5) and A, = {I, 3).There are 4 elements in Al, and 2 elements in A,, giving a total of 8 elements in the Cartesianproduct A, x A,. More generally, if A, consists
of k, elements and A, consists of k,elements, then Al x A, consists of k,k, elements. By induction on n, it follows that if A,consists of k, elements, then the Cartesian product A, x . * * x A, consists of k, * * * k,elements. To express this result in the language of set functions, let 9 denote the class of allfinite sets and let Y be the set function defined on 9 as
follows : If A E 9, v(A) is the numberof distinct elements in A. (For the empty set we define V(D) = 0 .) Then it is easy to verifythat Y is a finitely additive set function, so we may write (13.8) if{&,&,..., S,} is a disjoint collection of finite sets (that is, if Si n Sj = QI wheneveri #j). The number of elements in a Cartesian product may be expressed in terms
of Yas follows : v(A, x A, x . . . x A,) = v(A&(A& * . * v(A,). t The reason for this will become clear in Example 3 on p. 484, where we give a more detailed derivationof (13.7).
Some basic principles of combinatorial analysis 483 .- ii?- .__ 2 3 4 4 = &2,4,5) 5- a! FIGURE 13.1 An example illustrating the Cartesian of two sets. The plotted pointsrepresent A, x A,. A similar formula tells us how to count the number of elements in any set T of n-tuplesif we know the number of possible choices for each of the successive
components. Forexample, suppose there are k, possible choices for the first component x1. Let k, be thenumber of possible choices for the second component x2, once x1 is known. Similarly,let k, be the number of possible choices for the rth component x,, once x1, xp , . . . , xreIare known. Then the number of n-tuples that can be formed with these
choices is v(T) = k,k, . . . k, . This formula is often referred to as the principle of sequential counting. It can be provedby induction on n. In many applications the set of choices for x,. may not be easy to describesince it may not be determined until after the choices of the earlier components have beenmade. (This was the case when we used the
principle to count bridge hands.) Fortunately,to apply the principle of sequential counting we do not need to know the actual set ofchoices for x, but only the number of possible choices for x,. The additive property in formula (13.8) and the principle of sequential counting providethe key to the solution of many counting problems. The following
examples show how theycan be applied.
EXAMPLE 1. Sampling with replacement. Given a set S consisting of n elements. Ifk > 1, how many ordered k-tuples can be formed if each component may be an arbitraryelement of S? Note: It may be helpful to think of S as an urn containingn balls labeled 1, 2, . . .
, n.We select a ball and record its label as the first component of our k-tuple. Replacingthe ball in the urn, we again select a ball and use its label as the second component,and so on, until we have made k selections. Since we replace each ball after it is drawn,the same label may appear in different components of our k-tuple. 484 Set functions and
elementary probability Solution. Each k-tuple is an element of the Cartesian product T = S, x ... x s,, where each Si = S.
Conversely, each element of T is one of the k-tuples in question.Hence the number of k-tuples formed in this way is v(T) = Y(&) * * * Y(,&) = nk. EXAMPLE 2. Sampling without replacement. Given a set S consisting of n elements. Ifk 5 n, how many ordered k-tuples can be formed if the components are chosen from Swithout replacement, that is to say,
if no element of S may be used twice in any givenk-tuple ? Solution. Consider any k-tuple (x,, x2, . . . , x1<) formed from the elements of S withoutreplacement. For the first component x1 there are n choices (the IZ elements of S). Whenx1 is chosen there remain n - 1 ways to choose x2. With x2 chosen, there remain n - 2ways to choose x3, and so on,
there being n - k + I ways to choose x,. Therefore, bythe principle of sequential counting, the total number of k-tuples so formed is n!n(n - l)(n - 2) * * * (n - k + 1) = ~(n - k)!' In particular, when k = n this result tells us that n! distinct n-tuples may be formed froma given set S of n elements, with no two components of any n-tuple being equal.
EXAMPLE 3. The number of k-element subsets of an n-element set. If k 5 n, how manydistinct subsets of k elements can be formed from a set S consisting of n elements?
Solution. Let r denote the number of subsets in question and let us denote these subsets bYAl,&,...,&. These sets are distinct but need not be disjoint. We shall compute r in terms of n and kby an indirect method. For this purpose, let Bi denote the collection of k-tuples that canbe formed by choosing the components from the elements of Ai without
replacement.The sets B, , B, , . . . , B,. are disjoint. Moreover, if we apply the result of Example 2 withn = k we have v(Bi) = k! foreachi= I,2 ,..., r.Now let T=B,vB,V...uB,. Then T consists of all k-tuples that can be formed by choosing the components from Swithout replacement. From Example 2 we have Y(T) = n!/(n - k)! Exercises 485 and by additivity
we also have v(T) = zvY(BJ = k! r.i=l Equating the two expressions for V(T) we obtain n!r=k!(n-k)!= This proves formula (,13.7) stated earlier in this section. If we use the result of Example 3 to count the total number of subsets of a set S consistingof n elements, we obtain n Since this sum is also obtained when we expand (1 + 1)" by the binomial
theorem, thenumber of subsets of S is 2". 13.11 Exercises 1. Let A = {1,2,3). Display in roster notation the set of ordered pairs (a, b) obtained bychoosing the first component from A and the second component from the remaining elementsof A. Can this set of pairs be expressed as a Cartesian product? 2. A two-card hand can be dealt from a deck of 52
cards in 52.51 = 2652 ways. Determine thenumber of distinct hands, and explain your reasoning. 3. A senate committee consisting of six Democrats and four Republicans is to choose a chairmanand a vice-chairman. In how many ways can this pair of officers be chosen if the chairmanmust be a Democrat ? 4. An experiment consists of tossing a coin
twice and then rolling a die. Display each outcomeof this experiment as an ordered triple (a, 6, c), where each of a and b is either H (heads)or T (tails) and c is the number of points on the upturned face of the die. For example,(H, H, 3) means that heads came up on both tosses and 3 appeared on the die. Express theset of all possible outcomes as a
Cartesian product and determine the number of possibleoutcomes. 5. In how many ways can a bridge deck of 52 cards be dealt into four hands, each containing 13cards? Explain your reasoning. 6. Two dice, one red and one white, are tossed.
Represent the outcome as an ordered pair (a, b),where a denotes the number of points on the red die, b the number on the white die.
Howmany ordered pairs (a, 6) are possible ? How many are there for which the sum a + b is:(a) even? (b) divisible by 3?
(c) either even or divisible by 3? 7. A poker hand contains five cards dealt from a deck of 52. How many distinct poker handscan be dealt containing:(a) two pairs (for example, 2 kings, 2 aces, and a 3)?(b) a flush (five cards in a given suit)?(c) a straight flush (any five in sequence in a given suit, but not including ten, jack, queen,king, ace)?(d) a royal
flush (ten, jack, queen, king, ace in a single suit)? 486 Set functions and elementary probability 8.
Refer to Exercise 7. Compute the probability for a poker hand to be:(a) a flush.(b) a straight flush.(c) a royal flush. 9. How many committees of 50 senators can be formed that contain:(a) exactly one senator from Alaska?(b) both senators from Alaska? 10. A committee of 50 senators is chosen at random. Compute the probability that both senatorsfrom
Alaska are included. 11. A code group consists of four symbols in a row, each symbol being either a dot or a dash.How many distinct code groups can be formed? 12. How many k-letter words can be formed with an alphabet containing n letters?13. Show that: (a) (It) +j3) +(:) +..,=(y) +(I) +(:) +...=2+'.
(b) (;i' + (;r + + (;y = (;) . 14. Suppose a set of ordered pairs (a, b) is constructed by choosing the first component from aset of k elements, say from the set {a,, . . . , a,}, and the second component from a set of melements, say from {b,, . . . , b,}. There are a total of m pairs with first component (11, namely,(01, bl), . . . 1 (al, b,,). Similarly, there are m
pairs (ai, b,), . .
-
; (ai, b,) with first componentui. Therefore the total number of ordered pairs (a, b) is m + m + + m (k summands).This sum equals km, which proves the principle of sequential counting for sets of orderedpairs.
Use induction to prove the principle for sets of ordered n-tuples. 13.12 Conditional probability An unbiased die is thrown and the result is known to be an even number. What is theprobability that this number is divisible by 3? What is the probability that a child is colorblind, given that it is a boy?
These questions can be put in the following form: Let A andB be events of a sample space S.
What is the probability that A occurs, given that B hasoccurred? This is not necessarily the same as asking for the probability of the eventA n B. In fact, when A = B the question becomes: If A occurs, what is the probabilitythat A occurs? The answer in this case should be 1 and this may or may not be the proba-bility of A n B.
To see how to treat such problems in general, we turn to the questionpertaining to rolling a die. When we ask probability questions about rolling an unbiased die, we ordinarily usefor the sample space the set S = {1,2, 3,4, 5, 6} and assign point probability i to eachelement of S. The event "divisible by 3" is the subset A = (3, 6) and the event "even"
isthe subset B = {2,4, 6}. We want the probability that an element is in A, given that it isin B. Since we are concerned only with outcomes in which the number is even, we disregardthe outcomes 1, 3, 5 and use, instead of S, the set B = {2,4,6} as our sample space. Theevent in which we are now interested is simply the singleton {6}, this being the
only outcomeof the new sample space that is divisible by 3. If all outcomes of B are considered equallylikely, we must assign probability i to each of them; hence, in particular, the probability of(6) is also i. Note that we solved the foregoing problem by employing a very elementary idea. We Conditional probability 487 simply changed the sample space
from S to B and provided a new assignment of probabilities.This example suggests a way to proceed in general. Let (S, J%', P) be a given probability space. Suppose A and B are events and considerthe question: "What is the probability that A occurs, given that B has occurred?"As in the example just treated, we can change the sample space from S to
B and provide anew assignment of probabilities. We are at liberty to do this in any manner consistent withthe definition of probability measures. For B itself we have no choice except to assign theprobability 1. Since we are interested in those elements of A which lie in the new samplespace B, the problem before us is to compute the probability of the
event A n B accordingto the new assignment of probabilities. That is, if P' denotes the probability functionassociated with the new sample space B, then we must compute P'(A n B). We shall show now that if P(B) # 0 we can always define a probability function P'and a Boolean algebra 59' of subsets of B such that (B, .%', P') is a probability space.
Forthe Boolean algebra 3?' we take the collection of all sets T n B where T is in the originalBoolean algebra G?. It is easy to verify that .%'I, so defined, is indeed a Boolean algebra.One way to define a probability function P' on 23" is simply to divide each of the oldprobabilities by P(B). That is, if C E 9?' we let P(C)P'(C) = -f'(B) ' (This is where the
assumption P(B) # 0 comes in.) We are only changing the scale, withall probabilities magnified by the factor l/P(B).
It is easy to check that this definition ofP' gives us a bona jide probability measure. It is obviously nonnegative and it assignsprobability 1 to B. The additive property follows at once from the corresponding additiveproperty for P.
Since each C in 53' is of the form A n B, where A is an event in the original samplespace S, we can rewrite the definition of P' as follows: P'(A n B) =P(A n B) J'(B) ' This discussion suggests that the quotient P(A 17 B)/P(B) provides a reasonable measureof the probability that A occurs, given that B occurs. The following definition is madewith this
motivation in mind: DEFINITION OF CONDITIONAL PROBABILITY. Let (S, ~3, P) be a probability space and letB be an eoent such that P(B) # 0. The conditional probability that an event A will occur,given that B has occurred, is denoted by the symbol P(A 1 B) (read: "the probability of A,given B") and is dejned by the equation P(A 1 B) = '(;(;) B). The
conditional probability P(A [B) is not defined if P(B) = 0. The following examples illustrate the use of the concept of conditional probability. 488 Set functions and elementary probability EXAMPLE 1. Let us consider once more the problem raised earlier in this section: A dieis thrown and the result is known to be an even number.
What is the probability that thisnumber is divisible by 3? As a problem in conditional probabilities, we may take for thesample space the set S = { 1, 2, 3,4, 5, 6) and assign point probabilities Q to each elementof S. The event "even" is the set B = {2,4, 6) and the event "divisible by 3" is the setA = (3, 6). Therefore we have This agrees, of course, with
the earlier solution in which we used B as the sample spaceand assigned probability 4 to each element of B. EXAMPLE 2.
This is an example once used by the Caltech Biology Department to warnagainst the fallacy of superficial statistics. To "prove" statistically that the population ofthe U.S. contains more boys than girls, each student was asked to list the number of boysand girls in his family.
Invariably, the total number of boys exceeded the total number ofgirls. The statistics in this case were biased because all undergraduates at Caltech were males.Therefore, the question considered here is not concerned with the probability that a childis a boy; rather, it is concerned with the conditional probability that a child is a boy, giventhat he
comes from a family with at least one boy. To compute the probabilities in an example of this type consider a sample of 4n families,each with two children. Assume that n families have 2 boys, 2n families one boy and onegirl, and n families 2 girls.
Let the sample space S be the set of all 8n children in thesefamilies and assign the point probability P(X) = 1/(8n) to each x in S. Let A denote theevent "the child is a boy" and B the event "the child comes from a family with at least oneboy." The probability P(A) is obviously t. Similarly, P(B) = $ since 3n of the 4n familieshave at least one boy.
Therefore the probability that a child is a boy, given that he comesfrom a family with at least one boy, is the conditional probability 13.13 Independence An important idea related to conditional probability is the concept of independence ofevents, which may be defined as follows: DEFIWTION OF ISDEPENDENCE. Two events A and B are called
independent (or stochasti-tally independent) if and only if (13.9) P(A n B) = P(A)P(B). If A and B are independent, then P(A 1 B) = P(A) if P(B) # 0. That is, the conditionalprobability of A, given B, is the same as the "absolute" probability of A. This relationexhibits the significance of independence. The knowledge that B has occurred does notinfluence
the probability that A will occur. Independence 489 EXAMPLE 1. One card is drawn from a 52-card deck. Each card has the same probabilityof being selected. Show that the two events "drawing an ace" and "drawing a heart" areindependent. Solution. We choose a sample space S consisting of 52 elements and assign the pointprobability A to each
element. The event A, "drawing an ace," has the probabilityP(A) = & = lLs. The event B, "drawing a heart," has the probability P(B) = 5; = a.The event A n B means "drawing the ace of hearts," which has probability .&.
SincePA n B) = P(A)P(B), Equation (13.9) is satisfied and events A and B are independent. EXAMPLE 2. Three true dice are rolled independently, so that each combination is equallyprobable. Let A be the event that the sum of the digits shown is six and let B be the eventthat all three digits are different. Determine whether or not these two events are
independent. Solution. For a sample space S we take the set of all triples (a, b, c) with a, b, c rangingover the values 1, 2, 3, 4, 5, 6. There are 63 elements in S, and since they are equallyprobable we assign the point probability 1/63 to each element. The event A is the set of alltriples (a, b, c) for which a + b + c = 6. Direct enumeration shows that
there are 10such triples, namely : (1,2,3), (1, 3,2), (1, 1,4), (194, I>, (2, 1,3), (2, 3, 11, (2,2, a, (3, 1,2), (3,L 11, (4,131). The event B consists of all triples (a, b, c) for which n # 6, b # c, and a # c. Thereare 6 . 5 .4 = 120 elements in B. Exactly six of these elements are in set A, so that A n Bhas six elements. Therefore P(A n B) = 563 ' P(A) = g) and P(B)
= $. In this case P(A n B) # P(A)P(B); therefore events A and B are not independent.
Independence for more than two events is defined as follows. A finite collection & ofn events is said to be independent if the events satisfy the multiplicative property (13.10) Pt iii A, = flP(A,:)k=l k=l for every finite subcollection {A,, A,, . . . , A,}, where m may take the values m = 2,3 ,*.*, n, the sets Ai being in &'. When & consists of exactly three
events A, B, and C, the condition of independence in(13. IO) requires that (13.11) P(A n B) = P(A)P(B), P(A n C) = P(A)P(C), P(B n C) = P(B)P(C), 490 and Set functions and elementary probability (13.12) P(A n B n C) = P(A)P(B)P(C). It might be thought that the three equations in (13.11) suffice to imply (13.12) or, in otherwords, that independence of
three events is a consequence of independence in pairs.
Thisis not true, as one can see from the following example: Four tickets labeled a, b, c, and abc, are placed in a box. A ticket is drawn at random,and the sample space is denoted by S = {a, b, c, abc}. Define the events A, B, and C as follows: A = {a, abc}, B = {b, abc}, C = {c, abc}. In other words, the event X means that the ticket drawn contains the
letter x. It is easyto verify that each of the three equations in (13.11) is satisfied so that the events A, B,and C are independent in pairs. However, (13.12) is not satisfied and hence the threeevents are not independent. The calculations are simple and are left as an exercise forthe reader. 13.14 Exercises 1. Let A and B be two events with P(A) # 0, P(B)
0. Show that (13.13) P(A n B) = P(B)P(A 1 B) = P(A)P(B 1 A). Sometimes it is easier to compute the probabilities P(A) and P(B) A) directly by enumeration ofcases than it is to compute P(A n B). When this is the case, Equation (13.13) gives a con-venient way to calculate P(A n B) . The next exercise is an example. 2.
An urn contains seven white and three black balls. A second urn contains five white andfive black balls. A ball is selected at random from the first urn and placed in the second. Thena ball is selected at random from the second. Let A denote the event "black ball on first draw"and B the event "black ball on second draw."(a) Compute the probabilities
P(A) and P(B 1 ,4) directly by enumerating the possibilities.Use Equation (13.13) to compute P(A n B) .(b) Compute P(A n B) directly by enumerating all possible pairs of drawings. 3. (a) Let A,, A,, A, be three events such that P(A, A A,) # 0. Show that (b) Use induction to generalize this result as follows: If A,, A,, . . . , A, are n events (n > 2)such that
P(A, n A, n . . n A,-,) # 0, then P(A, n A, n n A,) = P(A,)P(A, 1 AJP(A, 1 A, n'A,) . P(A, 1 A, n A, n . . n A,-,). 4. A committee of 50 senators is chosen at random. Find the probability that both senatorsfrom Alaska are included, given that at least one is. Exercises~.~ 491 5. 6 . 7.8 . 9 . 10.
11. 12. 1 3 An urn contains five gold and seven blue chips. Two chips are selected at random (withoutreplacement). If the first chip is gold, compute the probability that the second is also gold.A deck of cards is dealt into four hands containing 13 ca.rds each. If one hand has exactlyeight spades, what is the probability that a particular one of the other
hands has (a) at leastone spade? (b) at least two spades? (c) a complete suit?ShowthatP(AuBIC)=P(AIC)+P(BIC)-P(AnBIC).Let A,, A,, . . . , A, be it disjoint events whose union is the entire sample space S. For everyevent Ewe have the equation E=EnS=EnijAi=iJ(EnAi).i-1 i=l This equation states that E can occur only in conjunction with some Ai. Show
that (a) P(E) = 5 P(E n AJ.i=l (b) P(E) = 2 P(E 1 Ai)!'(i=l This formula is useful when the conditional probabilities P(El Ai) are easier to computedirectly than P(E).An unbiased coin is tossed repeatedly. It comes up heads on the first six tosses. What is theprobability that it will come up heads on the seventh toss?Given independent events A and B whose
probabilities are neither 0 nor 1. Prove that A'and B' are independent. Is the same true if either of A or B has probability 0 or 1 ?Given independent events A and B. Prove or disprove in each case that :(a) A' and B' are independent.(b) A v B and A n B are independent.(c) P(A u B) = 1 - P(A')P(B').IfAl,A,,..., A, are independent events, prove that Pi 1; Ai +
fiP(A;) = 1 .i=l i=l If the three events A, B, and C are independent, prove that A U B and C are independent. [Hint: Use the result of Exercise 7 to show that P(A u B 1 C) = P(A u B).] 14. Let A and B be events, neither of which has probability 0. Prove or disprove the followingstatements :(a) If A and B are disjoint, A and B are independent.(b) If A and B

are independent, A and B are disjoint. 15. A die is thrown twice, the sample space S consisting of the 36 possible pairs of outcomes(a, b) each assigned probability &. Let A, B, and C denote the following events: A = {(a, b) 1 a is odd}, B = {(a, b) 1 b is odd}, C = {(a, b) 1 a + b is odd}. (a) Compute P(A), P(B), P(C), P(A n B), P(A n C), P(.B n C), and P(A n B n C).(b) Show that A, B, and C are independent in pairs.(c) Show that A, B, and C are not independent.

492 Set functions and elementary probability 13.15 Compound experiments We turn now to the problem of de Mere mentioned in the introduction -whether or notit is profitable to bet even money on the occurrence of at least one “double six” in 24 throwsof a pair of dice. We treat the problem in a more general form: What is the probability ofthrowing a double six at least once in n throws of a pair of dice? Is this probability morethan one-half or less than one-half when n = 24? Consider first the experiment of tossing a pair of fair dice just once. The outcomes of thisgame can be described by ordered pairs (a, b) in which a and b range over the values 1, 2,3,4, 5,6. The sample space S consists of 36 such pairs. Since the dice are fair we assign theprobability & to each pair in S. Now suppose we roll the dice n times. The succession of the n experiments is onecompound experiment which we wish to describe mathematically. To do this we need anew sample space and a corresponding probability measure. We consider the outcomes ofthe new game as ordered n-tuples (x1, . . . , xJ, where each component xi is one of theoutcomes of the original sample space S. In other words, the sample space for the com-pound experiment is the n-fold Cartesian product S x * * * x S, which we denote by S”.The set S” has 36” elements, and we assign equal probabilities to each element: P(x) = $ i f xESn. We are interested in the event “at least one double six in n throws.” Denote this eventby A. In this case it is easier to compute the probability of the complementary event A’,which means “no double six in n throws.” Each element of A’ is an n-tuple whose com-ponents can be any element of S except (6, 6). Therefore there are 35 possible values foreach component and hence (35)” n-tuples altogether in A’. Since each element of A’ hasprobability (3Le)n, the sum of all the point probabilities in A’ is ($$)a. This gives us P(A) = 1 - P(X) = 1 - ($g)n.

To answer de Mere’s question we must decide whether P(A) is more than one-half or lessthan one-half when n = 24. The inequality P(A) 2 4 is equivalent to 1 - (+g)” 2 4, or($8)” < i.

Taking logarithms we find nlog35 - nlog36 < -1og2, or n 2log 2 log 36 - log 35= 24.6+. Therefore P(A) < 4 when n = 24 and P(A) > 4 when n 2 25. It is not advantageous tobet even money on the occurrence of at least one double six in 24 throws. The foregoing problem suggests a general procedure for dealing with successive experi-ments. If an experiment is repeated two or more times, the result can be considered onecompound experiment. More generally, a compound experiment may be the result ofperforming two or more distinct experiments successively. The individual experiments maybe related to each other or they may be stochastically independent, in the sense that theprobability of the outcome of any one of them is unrelated to the results of the others.For the sake of simplicity, we shall discuss how one can combine two independent experi-ments into one compound experiment. The generalization to more than two experimentswill be evident. Compound experiments 493 To associate a bona$de probability space with a compound experiment we must explainhow to define the new sample space S, the corresponding Boolean algebra L8’ of subsetsof S, and the probability measure P defined on 9. As in the above example, we use theconcept of Cartesian product. Suppose we have two probability spaces, say (S,, :gl, P,) and (S,, g2, PJ. Thesespaces may be thought of as associated with two experiments El and E2. By the compoundexperiment E we mean the one for which the sample space S is the Cartesian productS1 x S,.

An outcome of E is a pair (x, y) in S, with the first component x an outcome ofE, and the second component y an outcome of E2. If S, has n elements and if S, has melements, then S, x S2 has nm elements. For the new Boolean algebra $ we take the collection of all subsets of S. Next wedefine the probability function P. Since S is finite we can define P(x, y) for each point(x, y) in S and then use additivity to define P for subsets of S. The point probabilitiesP(x, y) can be assigned in many ways. However, if the two experiments El and E, arestochastically independent, we define P by the equation (13.14) P(X? Y> = PlW2W for each (x,y) in S.

Motivation for this definition can be given as follows. Consider two special events Aand B in the new space S, andB = {(XI 9 YA (x, 7 YA - . . > (x, 5 yd> . That is, A is the set of all pairs in S, x S, whose first element is x1, and B is the set of allpairs whose second element is yl. The intersection of the two sets A and B is the singleton{(xl,yl)j. If we feel that the first outcome x1 should have no influence on the second out-come y1 it seems reasonable to require events A and B to be independent. This means wewould like to define the new probability function P in such a way that we have (13.15) P(A A B) = P(A)P(B).

If we decide how to assign the probabilities P(A) and P(B), Equation (13.15) will tell ushow to assign the probability P(A n B), that is, the probability P(x,, yi). Event A occursif and only if the outcome of the first experiment is x1. Since P,(x,) is its probability, itseems natural to assign the value P,(x,) to P(A) as well. Similarly, we assign the valueP2(yl) to P(B). Equation (13.15) then gives us All this, of course, is merely motivation for the assignment of probabilities in (13.14).The only way to decide whether or not (13.14) is a permissible assignment of point proba-bilities is to check the fundamental properties of probability measures. Each numberP(x, y) is nonnegative, and the sum of all the point probabilities is equal to 1, since we have 494 Set functions and elementary probability When we say that a compound experiment E is determined by two stochastically inde-pendent experiments El and E2, we mean that the probability space (S, B, P) is defined inthe manner just described, “independence” being reflected in the fact that P(x,,r) is theproduct P,(x)P,(y). It can be shown that the assignment of probabilities in (13.14) impliesthe formula (13.16) P(U x V) = P,(U)P,(V) for every pair of subsets U in gI and V in gz. (See Exercise 12 in Section 13.23 for anoutline of the proof.) We shall deduce some important consequences of this formula. Let A be an event (in the compound experiment E) of the form A = Cl x S,, where C, E 9YI. Each outcome in A is an ordered pair (x, y) where x is restricted to bean outcome of C1 (in the first experiment I&) but y can be any outcome of Sz (in the secondexperiment Ez). If we apply (13.16) we find P(A) = P(C, x S,) = P,(C,)P,(S,) = P1(Cl), since P,(S,) = 1 .

Thus the definition of P assigns the same probability to A that P,assigns to Ct. For this reason, such an event A is said to be determined by the$rst experi-ment E,. Similarly, if B is an event of E of the form where C, E gz, we have P(B) = P(S, x C,) = P,(S,)P,(C,) = P2(C,) and B is said to be determined by the second experiment E,. We shall now show, using(13.16), that two such events A and B are independent. That is, we have (13.17) P(A n B) = P(A)P(B).

First we note that A nB={(x,y)I(x,y)EC1 x S,and(x,y)ES1 x C,) = {(x,y)lx~C~ andyt:C,} = c, x c,. Hence, by (13.16), we have (13.18) P(A n B) = P(C1 x C,) = Pl(C,)P,(C,). SincePI = P(A) andP,(C,) = P(B) we obtain (13.17). Note that Equation (13.18) alsoshows that we can compute the probability P(A n B) as a product of probabilities in the Bernoulli trials 495 individual sample spaces S, and S,; hence no calculations with probabilities in thecompound experiment are needed. The generalization to compound experiments determined by n experiments E, , E, , . . . ,E, is carried out in the same way. The points in the new sample space are n-tuples (Xl,-%, * *. 7 x,) and the point probabilities are defined as the product of the probabilitiesof the separate outcomes, (13.19) P(Xl, -%,... 7 x,) = Pl(Xl)P2(X2> * * * P&n) * When this definition of P is used we say that E is determined by n independent experiments El, E,, . . . , E,. In the special case in which all the experiments are associated with thesame probability space, the compound experiment E is said to be an example of independentrepeated trials under identical conditions.

Such an example is considered in the nextsection. 13.16 Bernoulli trials An important example of a compound experiment was studied extensively by JakobBernoulli and is now known as a Bernoulli sequence of trials. This is a sequence of repeatedtrials executed under the same conditions, each result being stochastically independent ofall the others.

The experiment being repeated has just two possible outcomes, usuallycalled “success” and “failure ;” the probability of success is denoted byp and that of failureby q. Of course, q = 1 - p. The main result associated with Bernoulli sequences is thefollowing theorem: THEOREM 13.3. BERNOULLI’S FORMULA. The probability of exactly k successes in nBernoulli trials is (13.20) 0;2- pkqn-kydenotes the binomial coefficient, ” Iwhere k! (n - k>?’ Proof. Denote “success” by S and “failure” by F and consider a particular sequence ofn results. This may be represented by an n-tuple (Xl, x2, * . * ,x,), where each xi is either an S or an F.

The event A in which we are interested is the collectionof all n-tuples that contain exactly k S’s and n - k F’s. Let us compute the point probabilityof a particular n-tuple in A. The probability of each S isp, and that of each F is q. Hence,by (13.19), the probability of each particular n-tuple in A is the product of k factors equaltop with n - k factors equal to q. That is, P(x,) x*) . .

. , XJ = p” qn-” if (x1,x2,. . . ,x,)EA. 496 Set functions and elementary probability Therefore, to compute P(A) we need only count the number of elements in A and multiplythis number by p’; qn-’. But the number of elements in A is simply the number of waysof putting exactly k S’s into the n possible positions of the n-tuple. This is the same asthe number of subsets of k elements that can be formed from a set consisting of n elements;we have already seen that this number is ($). Therefore, if we add the point probabilitiesfor all points in A we obtain P(A) = ; pkql”-”.0EXAMPLE 1. An unbiased coin is tossed 50 times. Compute the probability of exactly 25 heads. Solution. We interpret this experiment as a sequence of 50 Bernoulli trials, in which“success” means “heads” and “failure” means “tails.” Since the coin is unbiased we assignthe probabilities p = q = -$, and formula (13.20) gives us (T)(#5o for the probability ofexactly k heads in 50 tosses. In particular, when k = 25 we obtain To express this number as a decimal it is best to use logarithms, since tables of logarithmsof factorials are readily available. If we denote the number in question by P, a table ofcommon logarithms (base IO) gives us log P = log 50! - 2log25! - 5Olog2 = 64.483 - 50.381 - 15.052 = -0.950 = 0.05 - 1.00 so P = 0.112.= log 1.12 - log 10 = log 0.112, EXAMPLE 2. What is the probability of at least r successes in n Bernoulli trials? Solution. Let A, denote the event “exactly k successes in n trials.” Then the event Ein which we are interested is the union E=A,uA,,u.*.uA,. Since the A, are disjoint, we find P(E) =sP(A,) =$&q?k=r k=r Since = (p + qjn = 1, The most probable number of successes in n Bernoulli trials-.~-~ ~-- - 491 the probability of the complementary event E’ can be computed as follows: This last sum gives us the probability of at most r - 1 successes in ti trials. 13.17 The most probable number of successes in n Bernoulli trials A pair of fair dice is rolled 28 times. What is the most probable number of sevens?To solve this problem we let f(k) denote the probability of exactly k sevens in 28 tosses.The probability of tossing a seven is 6. Bernoulli’s formula tells us that f(k) = (2k8) ($(;r-” We wish to determine what value (or values) of k in the range k = 0, 1, 2, . . . , 28 makef(k) as large as possible. The next theorem answers this question for any sequence ofBernoulli trials. THEOREM 13.4. Given an integer n 2 1 and a real p, 0 < p < 1, consider the set ofnumbers f(k) = 0”k Pk(l - Prk, for k = O,l,. . . , n. (a) If(n + 1)p is not an integer, the largest value off(k) occursfor exactly one k: k = t(n + IlpI, thegreatest integer < (n + 1)~. (b) If (n + 1)p is an integer, the largest value of,f(k) occurs for exact4 two values of k: k = (n + l)p and k = (n + 1)~ - 1. Proof. To study the behavior off(k) we consider the ratio f(k) k + 1 1 -- p =---r(k) = f(k + 1) n - k 17 fork=O,l,...,n- 1, The function r(k) is strictly increasing so we have 0 < r(0) < r(1) < * * * < r(n -1).

We consider six cases, illustrated in Figure 13.2. In the first three cases we show that f (k)takes its largest value for exactly one k. In the remaining casesf(k) takes its largest valuefor two consecutive values of k. 498 Set functions and elementary probability k k k0 1 2 n 0 1 2 n 0 I s n Case I Case 2 Case 3 0 1k kk s n 0 1 n 0 1 2 nCase 4 Case 5 Case 6

FIGURE 13.2 Calculation of the most probable number of successes in n Bernoullitrials. CASE I. r(O) > 1. In this case r(k) > 1 for every k so we have f(0) > f’(l) > . . * > f(n). Therefore the largest value off(k) occurs only for k = 0. Also, r(0) = (1 - p)/(n,o) > 1 ,so 1 -p > np, (n + 1)p < 1, hence [(n + l)p] = 0. CASE 2. r(n - 1) < 1. In this case r(k) < 1 for every k so f(0) 1, and r(k) # 1 for all k. In this case there is a uniqueinteger s, 0 < s < n, such that r(s - 1) < 1 and r(s) > 1 . The functionf’(k) increases inthe range 0 5 k < s and decreases in the range s < k < n. Thereforef(k) has a uniquemaximum at k = s.

Since r(s - 1) = -(1 - p)/(np - sp + p) < 1 we have s < (n + 1)-.The inequality r(s) > 1 shows that (n + 1)p < s + 1 , hence [(n + l)p] = s. Note that in each of the first three cases the maximum value of f(k) occurs when k = Kn + IlpI; also (n + 1)p is not an integer in any of these cases. CASE 4. r(0) < 1, r(n - 1) > 1, and r(s - 1) = lfor some s, 2 < s < n.

In this casef(k) increases for 0 < k 5 s - 1 and decreases for s < k 5 n. The maximum value off(k) occurs twice, when k = s - 1 and when k = s. The equation r(s - 1) = 1 implies(n + 1)p = s. CASE 5. r(n - 1) = 1. In this case r(k) < 1 for k < n - 2, so J’(k) increases in therange 0 2 k < n - 1, and .f(n - 1) =f(n). Hence the maximum of f(k) occurs twice,when k = n - 1 and when k = n. The equation r(n - 1) = 1 implies (n + 1)p = n.

Exercises 499- ~ ~~ CASE 6. r(0) = 1. In this case r(k) > 1 for k 2 1 , so f’(k) decreases in the range1 5 k 5 n. The maximumf(k) occurs twice, when k = 0 and when k = 1. The equationr(0) = 1 implies (n + I)p = 1. In each of the last three cases the maximum value off(k) occurs for k = (n + 1)~ andfor k = (n + 1)~ - 1 . This completes the proof.

EXAMPLE 1. A pair of fair dice is rolled 28 times.

What is the most probable number ofsevens ? Solution. We apply Theorem 13.4 with n = 28, p = Q, and (n + 1)~ = z>. This isnot an integer so the largest value off(k) occurs for k = [$%I = 4. Note: If the dice are rolled 29 times there are two solututions, k = 4 and k = 5. EXAMPLE 2. Find the smallest n such that if a pair of fair dice is thrown n times theprobability of getting exactly four sevens is at least as large as the probability of getting anyother number of sevens. Solution.

We take p = & in Theorem 13.4. We want the largest value of f(k) to occurwhen k = 4. This requires either [(n + 1)p] = 4, (n + l), -, = 4, or (n + 1)p - 1 = 4.The smallest n satisfying any of these relations is n = 23. 13.18 Exercises 1. A coin is tossed twice, the probability of heads on the first toss beingp, and that on the secondtoss pa. Consider this a compound experiment determined by two stochastically independentexperiments, and let the sample space be (a) Compute the probability of each element of S.(b) Canp, and pz be assigned so that P(H, H) =Q, P(H, T) = P(T, H) = $, P(T, T) = $? (c) Can pr and pz be assigned so that P(H, H) = P(T, T) = .+, P(H, T) = P(T, H) = Q ? (d) Consider the following four events (subsets of S): H, : heads on the first toss,Hz : heads on the second toss,T, : tails on the first toss,T2 : tails on the second toss. Determine which pairs of these four events are independent.

500 Set functions and elementary probability In each of Exercises 2 through 12 describe your sample space, your assignment of probabilities,and the event whose probability you are computing.

2 . 3 .4 . 5 . 6 . 7. 8 . 9 . 10. 11. 12. 13.

1 4 . A student takes a true - false examination consisting of 10 questions. He is completely unpre-pared so he plans to guess each answer.

The guesses are to be made at random. For example,he may toss a fair coin and use the outcome to determine his guess.(a) Compute the probability that he guesses correctly at least five times.(b) Compute the probability that he guesses correctly at least nine times.(c) What is the smallest n such that the probability of guessing at least n correct answers isless than &?Ten fair dice are tossed together. What is the probability that exactly three sixes occur?A fair coin is tossed five times.

What is the probability of getting (a) exactly three heads?(b) at least three heads? (c) at most one head?A man claims to have a divining rod which locates hidden sources of oil.

The Caltech GeologyDepartment conducts the following experiment to test his claim. He is taken into a room.inwhich there are 10 sealed barrels. He is told that five of them contain oil and five containwater. His task is to decide which of the five contain oil and which do not.(a) What is the probability that he locates the five oil barrels correctly just by chance?(b) What is the probability that he locates at least three of the oil barrels correctly by chance?A little old lady from Pasadena claims that by tasting a cup of tea made with milk she can tellwhether the milk or the tea was added first to the cup. The lady’s claim is tested by requiringher to taste and classify 10 pairs of cups of tea, each pair containing one cup of tea made byeach of the two methods under consideration. Letp denote her “true” probability of classifyinga pair of cups correctly. (If she is skillful, p is substantially greater than 4; if not, p I 4 .)Assume the 10 pairs of cups are classified under independent and identical conditions.(a) Compute, in terms of p, the probability that she classifies correctly at least eight of the10 pairs of cups.(b) Evaluate this probability explicitly whenp = 4.(Another problem of Chevalier de Mere.) Determine whether or not it is advantageous tobet even money on at least one 6 appearing in four throws of an unbiased dic. [Hint: Showthat the probability of at least one 6 in n throws is 1 - ($)n .]An urn contains w white balls and b black balls. If k 5 n, compute the probability of drawingk white balls in n drawings, if each ball is replaced before the next one is drawn.Two dice are thrown eight times. Compute the probability that the sum is 11 exactly threetimes.Throw a coin 10 times or 10 coins once and count the number of heads.

Find the probabilityof obtaining at least six heads.After a long series of tests on a certain kind of rocket engine it has been determined that inapproximately 5 ‘A of the trials there will be a malfunction that will cause the rocket to misfire.Compute the probability that in 10 trials there will be at least one failure.A coin is tossed repeatedly.

Compute the probability that the total number of heads will beat least 6 before the total number of tails reaches 5.Exercise 12 may be generalized as follows: Show that the probability of at least nz successesbefore n failures in a sequence of Bernoullian trials is VIZ++--l cc m + n - 1 k 1Pligm+n--lc--l. k=rrr Determine all n with the following property: If a pair of fair dice is thrown n times, the prob-ability of getting exactly ten sevens is at least as large as the probability of getting any othernumber of sevens. Countable and uncountable sets 501 15. A binary slot machine has three identical and independent wheels. When the machine isplayed the possible outcomes are ordered triples (x, y, z), where each of x, y, z can be 0 or 1.On each wheel the probability of 0 is p and the probability of 1 is 1 - p , where 0 < p < 1 .The machine pays $2 if the outcome is (I, 1, 1) or (0, 0,O); it pays $1 for the outcome (1, 1,O);otherwise it pays nothing. Letf(p) denote the probability that the machine pays $1 or morewhen it is played once.(a) CalculatefQ).(b) Define the “payoff” to be the sum &,Yg(~)P(~), where S is the sample space, P(x) is theprobability of outcome x, andg(x) is the number of dollars paid by outcome x. Calculate thevalue of p for which the payoff is smallest. 13.19 Countable and uncountable sets Up to now we have discussed probability theory only for finite sample spaces. We wishnow to extend the theory to infinite sample spaces. For this purpose it is necessary todistinguish between two types of infinite sets, countable and uncountable. This sectiondescribes these two concepts.

To count the members of an n-element set we match the set, element by element, with theset of integers {1,2, . . . , n}. Comparing the sizes of two sets by matching them element byelement takes the place of counting when we deal with infinite sets. The process of “match-ing” can be given a neat mathematical formulation by employing the function concept: DEFINITION. Two sets A and B are said to be in one-to-one correspondence ifafinction fexists with the following properties: (a) The domain off is A and the range off is B.(b) If x and y are distinct elements of A, then f (x) and f (y) are distinct elements of B. That is, for all x and y in A, (13.21) X#Y implies f(x) # f(y). A function satisfying property (13.21) is said to be one-to-one on A.

Two sets A and Bin one-to-one correspondence are also said to be equivalent, and we indicate this bywriting A - B. It is clear that every set A is equivalent to itself, since we may let f (x) = xfor each x in A.

A set can be equivalent to a proper subset of itself. For example, the set P = (1, 2,3 >, consisting of all the positive integers,{i,‘4,6, .

. . is equivalent to the proper subset Q =} consisting of the even integers. In this case a one-to-one function which makes them equivalent is given byf(x) = 2x for x in P.If A - B we can easily show that B N A. In fact, if f is one-to-one on A and if the range offis B, then for each b in B there is exactly one a in A such that f (a) = b . Thereforewe can define an inverse function g on B as follows: If b E B, g(b) = a, where a is theunique element of A such that f(a) = b. This g is one-to-one on B and its range is A;hence BN A. This property of equivalence is known as symmetry: (13.22) A-B implies B-A. 502 Set functions and elementary probability It is also easy to show that equivalence has the following property, known as transitivity: (13.23) A - B a n d B - C implies A - C . A proof of the transitive property is requested in Exercise 2 of Section 13.20.A set S is calledJinite and is said to contain n elements if S-(1,2 ,..., n}. The empty set is also considered to be finite. Sets which are not finite are called inznite sets.A set S is said to be countably injinite if it is equivalent to the set of all positive integers,that is, if (13.24) SN{1)2,3,. . .}. In this case there is a function f which establishes a one-to-one correspondence betweenthe positive integers and the elements of S; hence the set S can be displayed in rosternotation as follows: s = {f(I),f(2),f(3), . . 4. Often we use subscripts and denote f(k) by ar (or by a similar notation) and we writeS = {a,, 4, a,, . . . } . The important thing here is that the correspondence in (13.24)enables us to use the positive integers as “labels” for the elements of S. A set is said to be countable if it is finite or countably infinite. A set which is not countableis called uncountable.? (Examples will be given presently.) Many set operations whenperformed on countable sets produce countable sets. For example, we have the followingproperties : (a) Every subset of a countable set is countable.(b) The intersection of any collection of countable sets is countable.(c) The union of a countable collection of countable sets is countable.(d) The Cartesian product of a finite number of countable sets is countable.Since we shall do very little with countably infinite sets in this book, detailed proofs of these properties will not be given.j: Instead, we shall give a number of examples to showhow these properties may be used to construct new countable sets from given ones. EXAMPLE 1. The set S of all integers (positive, negative, or zero) is countable.

Proof. If n E S, letf(n) = 2n if n is positive, and letf(n) = 2 (nl + 1 if n is negative orzero. The domain off is S and its range is the set of positive integers. Since f is one-to-oneon S, this shows that S is countable. EXAMPLE 2. The set R of all rational numbers is countable. Proof. For each fixed integer n 2 1, let S, denote the set of rational numbers of theform x/n, where x belongs to the set S of Example 1. Each set S, is equivalent to S [take t The words denumerable and nondenumerable are sometimes used as synonyms for countable and un-countable, respectively.$ Proofs are outlined in Exercises 3 through 8 of Section 13.20. Countable and uncountable sets 503 f(t) = nt if t E S,] and hence each S, is countable. Since R is the union of all the S,,property (c) implies that R is countable. Note . If F = {A,, A,, A,, . . .} is a countable collection of sets, the union of all setsin the family 9 is denoted by the symbols l.J A, or A, uA, VA, u*.*.k = l EXAMPLE 3. Let A be a countably infinite set, say A = {a,, a2, a,, . . . > . For each integern 2 1, let 9n denote the family of n-element subsets of A. That is, let 9n = {S 1 S s A and S has n elements}. Then each 9% is countable. Proof. If S is an n-element subset of A, we may write s = iakl, ak2, . . . , s,>? where k, < k, < * . * < k,.

Let f(S) = (ak, , ak,, . . . , ak,). That is, f is the functionwhich associates with S the ordered n-tuple (akl, ak,, . . . , a,*). The domain of f isgm and its range, which we denote by T,, is a subset of the Cartesian productC, = A x A x . . - x A (n factors). Since A is countable, so is C, [by property (d)] andhence T, is also [by property (a)]. But T, N gn becausefis one-to-one. This shows that9n is countable. EXAMPLE 4. The collection of all finite subsets of a countable set is countable.

Proof. The result is obvious if the given set is finite.

Assume, then, that the given set(call it A) is countably infinite, and let 9 denote the class of all finite subsets of A: 9={SjSc AandSisfinite}. Then 9 is the union of all the families & of Example 3; hence, by property (c), 9 iscountable. EXAMPLE 5.

The collection of all subsets of a countably infinite set is uncountable. Proof. Let A denote the given countable set and let & denote the family of all subsetsof A. We shall assume that & is countable and arrive at a contradiction. If ~2’ is countable,then .d N A and hence there exists a one-to-one functionf whose domain is A and whoserange is ~2. Thus for each a in A, the function valuef(a) is a subset of A. This subsetmay or may not contain the element a. We denote by B the set of elements a such thata $f(a). Thus, B={a[a~Abuta$f(a)). 504 Set functions and elementary probability This B, being a subset of A, must belong to the family &’. This means that B = f(b) forsome b in A. Now there are only two possibilities: (i) b E B, or (ii) b # B. If b E B, thenby the definition of B we have b $f(b), which is a contradiction since f(b) = B. There-fore (i) is impossible.

In case (ii), b $ B, which means b $f(b) .

This contradicts the defini-tion of B, so case (ii) is also impossible. Therefore the assumption that J&’ is countableleads to a contradiction and we must conclude that & is uncountable. We give next an example of an uncountable set that is easier to visualize than that inExample 5. EXAMPLE 6. The set of real x satisfying 0 < x <: 1 is uncountable.

Proof. Again, we assume the set is countable and arrive at a contradiction. If the setis countable we may display its elements as follows: {x,, xp, x3, . . .}. Now we shall con-struct a real number y satisfying 0 < y < 1 which is not in this list. For this purpose wewrite each element x, as a decimal: x, = O.a,,, an.2 q3. . . , where each a,,i is one of the integers in the set {O., 1,2, . . .

, 9}. Let y be the real numberwhich has the decimal expansion y = O.YlYZY3~ * * 7where 1 i f an.% # 1,Yn = 2 i f an,%=l. Then no element of the set {x1, xp , x3,.

. .} can be equal to y, because y differs from x1in the first decimal place, differs from xz in the second decimal place, and in general, ydiffers from xlc in the kth decimal place. (A situation like x, = 0.249999 . * * and y =0.250000 * * - cannot occur here because of the way the yn are chosen.) Since this y satisfies0 < y < 1, we have a contradiction, and hence the set of real numbers in the open interval(0, 1) is uncountable. 13.20 Exercises 1. Let P ={1,2,3,. . .} denote the set of positive integers. For each of the following sets,exhibit a one-to-one function f whose domain is P and whose range is the set in question:(a) A = {2,4, 6, . . .}, the set of even positive integers.(b) B = (3, 3’, 33, . . .} , the set of powers of 3.(c) C = (2, 3, 5,7, 11, 13, . . .}, the set of primes. [Note: Part of the proof consists inshowing that C is an infinite set.](d) P x P, the Cartesian product of P with itself.(e) The set of integers of the form 2”3”, where m and n are positive integers. 2. Prove the transitive property of set equivalence: I f A - B a n d B - C , t h e n A - C . [Hint: Iffmakes A equivalent to B and ifg makes B equivalent to C, show that thecomposite function h = g of makes A equivalent to C.] Exercises 505 Exercises 3 through 8 are devoted to providing proofs of the four properties (a), (b), (c), (d)of countable sets listed in Section 13.19. 3. Prove that every subset of a countable set is countable. [Hint: Suppose S is a countablyinfinite set, say S = {x1, x2, xs, . .

.}, and let A be an infinite subset of S. Let k(1) be thesmallest positive integer m such that x, E A. Assuming k(l), k(2), . . . , k(n - 1) have beendefined, let k(n) be the smallest positive integer m > k(n - 1) such that x,; E A. Letf(n) =xktn). Show that f is a one-to-one function whose domain is the set of positive integers andwhose range is A. This proves the result when S is countably infinite. Construct a separateproof for a finite S.] 4. Show that the intersection of any collection of countable sets is countable.

[Hint: Use theresult of Exercise 3.1 5. LetP={1,2,3,. . .} denote the set of positive integers.(a) Prove that the Cartesian product P x P is countable. [Hint: Let Q denote the set ofpositive integers of the form 2”3”, where m and n are positive integers. Then Q = P, so Q iscountable (by Exercise 3). If (m, n) E p x P, let f(m, ?I) = 2”3” and use this function to show that P x P - Q .](b) Deduce from part (a) that the Cartesian product of two countable sets is countable. Thenuse induction to extend the result to n countable sets. 6. Letg = {Br,&,&,.. .} be a countable collection of disjoint sets (& A Bj = ,@ when i #j)a, such that each B, is countable. Show that the union U B, is also countable. [Hint: Let & = {bl,, b,n, bs,n, . . .

} and S = fi B,.

If x E S, thin .Y = b,,, for some unique pairk=l (m, n) and we can define f(x) = (m, n) . Use this f to show that S is equivalent to a subsetof P x P and deduce (by Exercise 5) that S is countable.] 7. Let&={A,,A,,A,,.. .} be a countable collection of sets, and let g = {B,, B,, B,, . . . }be defined as follows: B, = A, and, for n > 1, 12-l B, = A , -u A , .k=l That is, B, consists of those points in A, which are not in any of the earlier sets A,, . . . , A, , .Prove that J!% is a collection of disjoint sets (Bi n Bj = 13 when i #j) and that ~z’i,=&?k.k=l k=l This enables us to express the union of any countable collection of sets as the union of a count-able collection of disjoint sets. 8. If 9 is a countable collection of countable sets, prove that the union of all sets in 9 is count-able. [Hint: Use Exercises 6 and 7.1 9. Show that the following sets are countable:(a) The set of all intervals on the real axis with rational end points.(b) The set of all circles in the plane with rational radii and centers having rational coordinates.(c) Any set of disjoint intervals of positive length. 10. Show that the following sets are uncountable:(a) The set of irrational numbers in the interval (0, 1).(b) The set of all intervals of positive length.(c) The set of all sequences whose terms are the integers 0 and 1. (Recall that a sequence isa function whose domain is the set of positive integers.) 5 0 6 Set functions and elementary probability 13.21 The definition of probability for countably infinite sample spaces This section extends the definition of probability to countably infinite sample spaces.Let S.be a countably infinite set and let 99 be a Boolean algebra of subsets of S.

We definea probability measure P on 98 as we did for the finite case, except that we require countableadditivity as well as finite additivity. That is, for every countably infinite collection vl,Az>-- .} of elements of 97, we require that (13.25) if Ai n A, = 0 whenever i # j . Finitely additive set functions which satisfy (13.25) are said to be countably additive (orcompletely additive). Of course, this property requires assuming also that the countableunion A, V A, U A, U .

* * is in 97 whenever each A, is in 97. Not all Boolean algebrashave this property. Those which do are called Boolean c-algebras.

An example is theBoolean algebra of all subsets of S. DEFINITION OF PROBABILITY FOR COUNTABLY INFINITE SAMPLE SPACES. Let .9# denote aBoolean o-algebra whose elements are subsets of a given countably infinite set S. A setfunction P is called a probability measure on .!Z? if it is nonnegative, countably additive, andsatisfies P(S) = 1 .

When W is the Boolean algebra of all subsets of S, a probability function is completelydetermined by its values on the singletons (called point probabilities).

Every subset A ofS is either finite or countably infinite, and the probability of A is computed by adding thepoint probabilities for all elements in A, P(A) = z.PW. The sum on the right is either a finite sum or an absolutely convergent infinite series.The following example illustrates an experiment with a countably infinite sample space.

EXAMPLE. Toss a coin repeatedly until the first outcome occurs a second time; at thispoint the game ends. For a sample space we take the collection of all possible games that can be played. Thisset can be expressed as the union of two countably infinite sets A and B, where A = (TT, THT, THHT, THHHT, . . .> and B = (HH, HTH, HTTH, HTTTH, . . .>. We denote the elements of set A (in the order listed) as a,, a,, a,, . . . , and those of set Bas b,, b,, b,, ... We can assign arbitrary nonnegative point probabilities P(a,) and P(b,)provided that Miscellaneous exercises on probability 507 For example, suppose the coin has probability p of coming up heads (H) and proba-bility q = 1 - p of coming up tails (T), where 0 < p < 1. Then a natural assignment ofpoint probabilities would be (13.26) WJ = 42p” and P(b,) = pzq*. This is an acceptable assignment of probabilities because we have Now suppose we ask for the probability that the game ends after exactly n + 2 tosses.This is the event {a,} U {b,} , and its probability is P(a,> + P&J = q2p” + p2q”.

The probability that the game ends in at most n + 2 tosses is $$‘(a,) + -Wd = q2 ‘ L/b’ + $ 1 ,yyr = 1 _ qpn+l_ pqn+l.k=O k = O 13.22 Exercises The exercises in this section refer to the example in Section 13.21. 1. Using the point probabilities assigned in Equation (13.26), let&(p) denote the probability thatthe game ends after exactly n + 2 tosses.

Calculate the absolute maximum and minimum valuesoff,(p) on the interval 0 5 p 5 1 for each of the values IZ = 0, 1, 2, 3.

2. Show that each of the following is an acceptable assignment of point probabilities. (4 W,) = Wh) = & for n = 0, 1,2, ... 1W JYa,) = WJ = (n + 2)(n + 3) f o r n=0,1,2 ,.... 3. Calculate the probability that the game ends before the fifth toss, using:(a) the point probabilities in (13.26).(b) the point probabilities in Exercise 2(a).(c) the point probabilities in Exercise 2(b). 4. Calculate the probability that an odd number of tosses is required to terminate the game, using:(a) the point probabilities in (13.26).(b) the point probabilities in Exercise 2(a).(c) the point probabilities in Exercise 2(b). 13.23 Miscellaneous exercises on probability 1. What is the probability of rolling a ten with two unbiased dice?2. Ten men and their wives are seated at random at a banquet. Compute the probability that a particular man sits next to his wife if (a) they are seated at a round table; (b) they are seated in a row. 508 Set functions and elementary probability 3. A box has two drawers. Drawer number 1 contains four gold coins and two silver coins.Drawer number 2 contains three gold coins and three silver coins. A drawer is opened atrandom and a coin selected at random from the open drawer. Compute the probability of eachof the following events:(a) Drawer number 2 was opened and a silver coin was selected.(b) A gold coin was selected from the opened drawer. 4. Two cards are picked in succession from a deck of 52 cards, each card having the same prob-ability of being drawn.(a) What is the probability that at least one is a spade? The two cards are placed in a sack unexamined. One card is drawn from the sack andexamined and found not to be a spade. (Each card has the same probability of being drawn.)(b) What is the probability now of having at least one spade?

The card previously drawn is replaced in the sack and the cards mixed. Again a card isdrawn and examined. No comparison is made to see if it is the same card previously drawn.The card is again replaced in the sack and the cards mixed. This is done a total of three times,including that of part (b), and each time the card examined is not a spade.(c) What is a sample space and a probability function for this experiment? What is the prob-ability that one of the two original cards is a spade? 5. A man has ten pennies, 9 ordinary and 1 with two heads.

He selects a penny at random,tosses it six times, and it always comes up heads. Compute the probability that he selectedthe double-headed penny. 6.

Prove that it is impossible to load a pair of dice so that every outcome from 2 to 12 will havethe same probability of occurrence. 7. A certain Caltech sophomore has an alarm clock which will ring at the appointed hour withprobability 0.7. If it rings, it will wake him in time to attend his mathematics class with proba-bility 0.8. If it doesn’t ring he will wake in time to attend his class with probability 0.3.Compute the probability that he will wake in time to attend his mathematics class. 8.

Three horses A, B, and Care in a horse race. The event “A beats B” will be denoted symboli-cally by writing AB.

The event “A beats B who beats C” will be denoted by ABC, etc. Sup-pose it is known that P(AB) = $, .P(AC) = $, P(BC) = 4,and that P(ABC) = P(ACB), P(BAC) = P(BCA) , P(CAB) = P(CBA) . (a) Compute the probability that A wins.(b) Compute the probability that B wins.(c) Compute the probability that C wins.(d) Are the events AB, AC, and CB independent? 9. The final step in a long computation requires the addition of three integers aI, a2, a3. Assumethat (a) the computations of a,, a2, and as are stochastically independent; (b) in the com-putation of each ai there is a common probability p that it is correct and that the prob-ability of making an error of + 1 is equal to the probability of making an error of -1; (c)no error larger than + 1 or less than - 1 can occur. Remember the possibility of compensatingerrors, and compute the probability that the sum a, + aa + us is correct. 10. The game of “odd man out.” Suppose n persons toss identical coins simultaneously and in-dependently, where n 2 3. Assume there is a probabilityp of obtaining heads with each coin.Compute the probability that in a given toss there will be an “odd man,” that is, a person whosecoin does not have the same outcome as that of any other member of the group.

11. Suppose n persons play the game “odd man out” with fair coins (as described in Exercise 10).For a given integer m compute the probability that it will take exactly m plays to conclude thegame (the mth play is the first time there is an “odd man”). Miscellaneous exercises on probability 509 12. Suppose a compound experiment (S, g’, P) is determined by two stochastically independentexperiments (S, , gI, PI) and (S, , gz, Pz), where S = S, x & and m, y> = P,(x)&(y)for each (x, y) in S. The purpose of this exercise is to establish the formula (13.27) P(U x V) = P,(U)P,(V) for every pair of subsets U in gI and V in gz. The sample spaces S, and S, are assumed tobe finite.(a) Verify that Equation (13.27) is true when U and V are singletons, and also when at leastone of U or V is empty.

Suppose now that U={Ul,U2,...,Uk) a n d V={u,,u,)...) u,}. Then U x V consists of the km pairs (Ui, uj). For each i = 1,2,. . .

, k let Ai denote the setof m pairs in U x V whose first component is ui .(b) Show that the Ai are disjoint sets whose union is U x V.(c) Show that P(A,) = 2 I’&, uj) = Pl(ui)P2(V) .61 (d) From (b) and (c) deduce that P(U X V) =iP(AJ =Pl(U)Pz(V).i=l 14CALCULUS OF PROBABILITIES 14.1 The definition of probability for uncountable sample spaces A line segment is broken into two pieces, with the point of subdivision chosen at random.What is the probability that the two pieces have equal length?

What is the probability thatthe longer segment has exactly twice the length of the shorter?

What is the probabilitythat the longer segment has at least twice the length of the shorter? These are examples ofprobability problems in which the sample space is uncountable since it consists of allpoints on a line segment. This section extends the definition of probability to includeuncountable sample spaces. If we were to use the same procedure as for countable sample spaces we would start withan arbitrary uncountable set S and a Boolean o-algebra 9 of subsets of S and define aprobability measure to be a completely additive nonnegative set function P defined on L4?with P(S) = 1. As it turns out, this procedure leads to certain technical difficulties thatdo not occur when S is countable. To attempt to describe these difficulties would take ustoo far afield. We shall avoid these difficulties by imposing restrictions at the outset on theset S and on the Boolean algebra a. First, we restrict S to be a subset of the real line R, or of n-space R”.

For the Booleanalgebra g we use special subsets of S which, in the language of modern integration theory,are called measurable subsets of S. We shall not attempt to describe the exact meaning of ameasurable set; instead, we shall mention some of the properties possessed by the class ofmeasurable sets. First we consider subsets of R. The measurable subsets have the following properties: 1. If A is measurable, so is R - A, the complement of A.2. If {A,, A,, A,, . . . } is a countable collection of measurable sets, then the union A, U A, U A, U * 9 . is also measurable.3. Every interval (open, closed, half-open, finite, or infinite) is measurable. Thus, the measurable sets of R form a Boolean o-algebra which contains the intervals.A smallest Boolean o-algebra exists which has this property; its members are called Borelsets, after the French mathematician Emile Bore1 (1871-1956). Similarly, in 2-space a smallest Boolean o-algebra exists which contains all Cartesianproducts of pairs of intervals; its members are the two-dimensional Bore1 sets. Bore1 setsin n-space are defined in an analogous fashion. 510 Countability of the set of points with positive probability 511 Henceforth, whenever we use a set S of real numbers as a sample space, or, moregenerally, whenever we use a set S in n-space as a sample space, we shall assume that thisset is a Bore1 set. The Bore1 subsets of S themselves form a Boolean a-algebra. Thesesubsets are extensive enough to include all the events that occur in the ordinary applicationsof probability theory. DEFINITION OF PROBABILITY FOR UNCOUNTABLE SAMPLE SPACES. Let S be a subset of Rn,and let 93’ be the Boolean a-algebra of Bore1 subsets of S. A nonnegative completely additiveset function P dejned on % with P(S) = 1 is called a probability measure. The triple(S, 9, P) is called a probability space.

14.2 Countability of the set of points with positive probability For countable sample spaces the probability of an event A is often computed by addingthe point probabilities P(x) for all x in A.

This process is not fruitful for uncountablesample spaces because, as the next theorem shows, most of the point probabilities are zero.

THEOREM 14.1. Let (S, 97, P) be a probability space and let T denote the set of all x in Sfor which P(x) > 0. Then T is countable. Proof. Foreachn=1,2,3 ,..., let T, denote the following subset of S: < P(x) < 1 ‘n If P(x) > 0 then x E T, for some n. Conversely, if x E T, for some n then x E T. HenceT = T, v T, v.. . . Now T, contains at most n points, because if there were n + 1 ormore points in T, the sum of their point probabilities would exceed 1. Therefore T iscountable, since it is a countable union of finite sets. Theorem 14.1 tells us that positive probabilities can be assigned to at most a countablesubset of S. The remaining points of S will have probability zero. In particular, if all theoutcomes of S are equally likely, then every point in S must be assigned probability zero. Note: Theorem 14.1 can be given a physical interpretation in terms of mass distri-bution which helps to illustrate its meaning. Imagine that we have an amount of mass, withthe total quantity equal to 1. (This corresponds to P(S) = 1.) Suppose we are able todistribute this mass in any way we please along the real line, either by smearing it alongthe line with a uniform or perhaps a varying thickness, or by placing discrete lumps of massat certain points, or both. (We interpret a positive amount of mass as a discrete lump.)We can place all the mass at one point. We can divide the mass equally or unequally indiscrete lumps among two points, among ten points, among a million points, or amonga countably infinite set of points. For example, we can put & at 1, 2 at 2, + at 3, and so on,with mass ($)n at each integer n 2 1 . Or we can smear all themasswithoutanyconcentratedlumps. Or we can smear part of it and distribute the rest in discrete lumps. Theorem 14.1tells us that at most a countable set of points can be assigned discrete lumps of mass. Since most (if not all) the point probabilities for an uncountable sample space will bezero, a knowledge of the point probabilities alone does not suffice to compute the 512 Calculus of probabilities probabilities of arbitrary events. Further information is required; it is best described interms of two new concepts, random variables and distribution functions, to which we turnnext. These concepts make possible the use of integral calculus in many problems withuncountable sample spaces. Integration takes the place of summation in the computationof probabilities. 14.3 Random variables In many experiments we are interested in numbers associated with the outcomes of theexperiment. For example, n coins are tossed simultaneously and we ask for the number ofheads. A pair of dice is rolled and we ask for the sum of the points on the upturned faces.A dart is thrown at a circular target and we ask for its distance from the center.Whenever we associate a real number with each outcome of an experiment we are dealingwith a function whose domain is the set of possible outcomes and whose range is the setof real numbers in question. Such a function is called a random variable. A formaldefinition can be given as follows: DEFINITION OF A RANDOM VARIABLE. Let S denote a sample space. A real-valuedfunctiondejined on S is called a one-dimensional random variable. If the function values are orderedpairs of real numbers (that is, vectors in 2-space), the function is said to be a two-dimensionalrandom variable. More generally, an n-dimensional random variable is simply a functionwhose domain is the given sample space S and whose range is a collection of n-tuples of realnumbers (vectors in n-space). Thus, a random variable is nothing but a vector-valued function defined on a set. Theterm “random” is used merely to remind us that the set in question is a sample space.7 Because of the generality of the above definition, it is possible to have many randomvariables associated with a given experiment. In any particular example the experimentermust decide which random variables will be of interest and importance to him. In general,we try to work with random variables whose function values reflect, as simply as possible,the properties of the outcomes of the experiment which are really essential. Notations. Capital letters such as X, Y, 2 are ordinarily used to denote one-dimensionalrandom variables. A typical outcome of the experiment (that is, a typical element of thesample space) is usually denoted by the Greek letter cc) (omega). Thus, X(w) denotes thatreal number which the random variable X associates with the outcome w. The following are some simple examples of random variables. EXAMPLE 1. An experiment consists of rolling a die and reading the number of pointson the upturned face.

The most “natural” random variable X to consider is the onestamped on the die by the manufacturer, namely: X(w) = w f o r o = 1,2,3,4, 5,6. t The terms “stochastic variable” and “chance variable” are also used as synonyms for “random variable.”The word “stochastic” is derived from a Greek stem meaning “chance” and seems to have been invented byJakob Bernoulli. It is commonly used in the literature of probability theory. Exercises 513 If we are interested in whether the number of points is even or odd, then we can considerinstead the random variable Y, which is defined as follows: Y(0) = 0 if w is even, Y(w) = 1 i f w i s o d d . The values 0 and 1 are not essential-any two distinct real numbers could be used instead.However, 0 and 1 suggest “even” and “odd,” respectively, because they represent theremainder obtained when the outcome co is divided by 2. EXAMPLE 2. A dart is thrown at a circular target. The set of all possible outcomes isthe set of all points w on the target. If we imagine a coordinate system placed on thetarget with the origin at the center, we can assign various random variables to this experi-ment. A natural one is the two-dimensional random variable which assigns to the point oits rectangular coordinates (x, y). Another is that which assigns to (0 its polar coordinates(r, 19). Examples of one-dimensional random variables are those which assign to each wjust one of its coordinates, such as the x-coordinate or the r-coordinate (distance from theorigin). In an experiment of this type we often wish to know the probability that the dartwill land in a particular region of the target, for example, the first quadrant. This event canbe described most simply by the random variable which assigns to each point w its polarcoordinate angle 19, so that X(W) = 8; the event “the dart lands in the first quadrant” is theset of w such that 0 < X(w) < &r.

Abbreviations.

We avoid cumbersome notation by using special abbreviations to describecertain types of events and their probabilities.

For example, if t is a real number, the setof all LC) in the sample space such that X(w) = t is denoted briefly by writing x= t. The probability of this event is written P(X = t) instead of the lengthier P({w 1 X(o) = t}).Symbols such as P(X = a or X = b) and P(a < X 5 b) are defined in a similar fashion.Thus, the event “X = a or X = b” is the union of the two events “X = a” and “X = b” ;the symbol P(X = a or X = b) denotes the probability of this union. The event“a < X < b” is the set of all points w such that X(w) lies in the half-open interval (a, b],and the symbol P(a < X 5 b) denotes the probability of this event. 14.4 Exercises 1. Let X be a one-dimensional random variable.(a) If a < b, show that the two events a < X < b and X < a are disjoint.(b) Determine the union of the two events in part (a).(c) Show that P(a < X < b) = P(X < b) - P(X I a). 2.

Let (X, Y) denote a two-dimensional random variable defined on a sample space S. This meansthat (X, Y) is a function which assigns to each o in S a pair of real numbers (X(w), Y(w)).Of course, each of X and Y is a one-dimensional random variable defined on S. The notation 514 Calculus of probabilities stands for the set of all elements w in S such that X(w) < a and Y(w) 5 b.(a) If a < b and c < d, describe, in terms of elements of S, the meaning of the followingnotation: a 3. Two fair dice are rolled, each outcome being an ordered pair (a, b), where each of a and b isan integer from 1 to 6. Let X be the random variable which assigns the value a + b to theoutcome (a, b).(a) Describe, in roster notation, the events “X = 7,” “X = 11,” “X = 7 or X = 11.”(b) Compute the probabilities of the events in part (a). 4. Consider an experiment in which four coins are tossed simultaneously (or one coin is tossedfour times). For each coin define a random variable which assigns the value 1 to heads and thevalue 0 to tails, and denote these random variables by X1, X, , X, , X, . Assign the probabilitiesP(Xi = 1) = P(Xi = 0) = i for each Xi. Consider a new random variable Y which assignsto each outcome the total number of heads among the four coins. Express Yin terms of X1,X, , X3, X, and compute the probabilities P(Y = 0) , P(Y = 1)) and P(Y < 1). 5.

A small railroad company has facilities for transporting 100 passengers a day between two cities,at a fixed cost (to the company) of $7 per passenger. If more than 100 passengers buy ticketsin any one day the railroad is obligated to provide bus transportation for the excess at a cost of$10 per passenger. Let X be the random variable which counts the number of passengers thatbuy tickets in a given day. The possible values of Xare the integers 0, 1, 2, 3, .

. .

up to a certainunknown maximum. Let Y denote the random variable which describes the total daily cost(in dollars) to the railroad for handling passengers. Express Yin terms of X. 6. A factory production line consists of two work stations A and B. At station A, X units perhour are assembled; they are immediately transported to station B, where they are inspectedat the rate of Y units per hour, where Y < X. The possible values of X and Y are the integers8,9, and 10.

Let Z denote the random variable which counts the number of units that come offthe production line during the first hour of production.(a) Express Z in terms of X and Y, assuming each of X and Y is constant during this hour.(b) Describe, in a similar way, the random variable U which counts the number of units de-livered in the first two consecutive hours of production. Each of X and Y is constant during eachhour, but the constant values during the second hour need not be the same as those during thefirst. 14.5 Distribution functions We turn now to the problem of computing the probabilities of events associated witha given random variable.

Let X be a one-dimensional random variable defined on asample space S, where S is a Bore1 set in n-space for some n > 1 .

Let P be a probabilitymeasure defined on the Bore1 subsets of S. For each w in S, X(o) is a real number, andas OJ runs through the elements of S the numbers A’(U) run through a set of real numbers(the range of X). This set may be finite, countably infnite, or uncountable. For each realnumber t we consider the following special subset of S: /l(t) = {w (X(0) < t). If t is less than all the numbers in the range of X, the set A(t) will be empty; otherwise,A(t) will be a nonempty subset of S.

We assume that for each t the set -4(t) is an event, Distribution functions 515 that is, a Bore1 set. According to the convention discussed at the end of Section 14.3, wedenote this event by the symbol X 5 t.

Suppose we know the probability P(X 5 t) for every real t. We shall find in a momentthat this knowledge enables us to compute the probabilities of many other events ofinterest. This is done by using the probabilities P(X < t) as a basis for constructing anew function F, called the distribution function of X. It is defined as follows: DEFINITION OF A DISTRIBUTION FUNCTION. Let X be a one-dimensional random variable.The function F de$ned for all real t by the equation F(t) = P(X 5 t) is called the distribution function of the random variable X. Note: Sometimes the notation FX is used to emphasize the fact that the distributionfunction is associated with the particular random variable X. The value of the functionat t is then denoted by Fx(t). It is important to realize that the distribution function F is defined over the entire realaxis, even though the range of X may be only a bounded portion of the real axis. In fact,if all numbers X(a) lie in some finite interval [a, b], then for t < a the probability P(X < t) is zero (since for t < a the set X < t is empty) and for t 2 b the probability P(X < t) is 1(because in this case the set X 5 t is the entire sample space). This means that for boundedrandom variables X whose range is within an interval [a, b] we have F(t) = 0 for allt 0 I F(t) I 1 for all 1.(b) P(a < X 5 b) = F(b) - F(a) if a FW I F(b) i f a a we write (14.4) F(t) = F(a) + P(a < X < t) ; if t < a we write (14.5) F(t) = F(a) - P(t < X 5 a). Letting t--j a+ in (14.4) we find lim F(t) = F(a) + lim P(a < X 5 t) ,t-+n+ t+n-t whereas if t + a- in (14.5) we obtain lim F(t) = F(a) - lim P(1 < X 5 a).t+a-t-+a-- Therefore to prove (14.2) and (14.3) we must establish two equations: (14.6) limP(a < X 5 t) = 0t+at Discontinuities of distribution functions 519 and (14.7) limP(t a, is empty. On the other hand, when t + a- the half-open interval (t, a] shrinksto the point a. (The intersection of all intervals (t, a] for t < a is the set {a}.) Therefore,if probability behaves in a continuous fashion, Equations (14.6) and (14.7) must be valid.To convert this argument into a rigorous proof we proceed as follows: For each integer n 2 1, let (14.8) pn=P a 0 is either finite or countable. We denotethis set by T, that is, we let T = it.1 p(t) > 01. Discrete distributions. Probability mass functions 521 The random variable X is said to be discrete if &p(t) = 1.

In other words, X is discrete if a unit probability mass is distributed over the real line byconcentrating a positive mass p(t) at each point t of some finite or countably infinite setT and no mass at the remaining points. The points of T are called the mass points of X. For discrete random variables a knowledge of the probability mass function enablesus to compute probabilities of arbitrary events. In fact, we have the following theorem.

THEOREM 14.6. Let A be a Bore1 subset of the real line R, and let P(X E A) denote theprobability of the set of u such that X(o) E A. Then we have (14.11) P(X E 4 = zEJTP(4 9 where T is the set of mass points of X. Proof. Since A A T E A and T - A G R - A, we have (14.12) .TnTp(x) I P(X E A) and But A n T and T - A are disjoint sets whose union is T, so the second inequality in(14.12) is equivalent to l-zE&T-(x) I 1 - p(X E A) or zsFn:nT-(4 2 P(X E A).

Combining this with the first inequality in (14.12) we obtain (14.11). Note: Since@) = 0 when x $ T, the sum on the right of (14.11) can be written asIseA p(x) without danger of its being misunderstood. When A is the interval (- co, t], the sum in (14.11) gives the value of the distributionfunction F(t). Thus, we have F(t) = P(X 5 t) = z: p(x).rst If a

random variable X is discrete, the corresponding distribution function F is also calleddiscrete. The following examples of discrete distributions occur frequently in practice. EXAMPLE 1. Binomial distribution. Let p be a given real number satisfying 0 5 p 5 1and let q = 1 -p. Suppose a random variable X assumes the values 0, 1, 2, . .
. , n,where n is a fixed positive integer, and suppose the probability P(X = k) is given by theformula P(X = k) = for k=0,1,2 ,..., n. 522 Calculus of probabilities 0.4 0.2 - - (a) The mass function (b) The distribution function FIGURE 14.5 The probability mass function and the distribution function of abinomial distribution with pararneters n = 5 and p = &.
This assignment of probabilities is permissible because the sum of all the point probabilitiesis -$P(X = k) =$(;) p”,“” = (p + q)” = 1.k=O k=O The corresponding distribution function F, is said to be a binomial distribution withparameters n andp. Its values may be computed by the summation formula F,(t) = 2 (JPkPk.OlkSt Binomial distributions arise
naturally from a Bernoulli sequence of trials where p isthe probability of “success” and q the probability of “failure.” In fact, when the randomvariable X counts the number of successes in k trials, P(X = k) is precisely (!J)P”~“~because of Bernoulli’s formula. (See Theorem 13.3 in Section 13.16.) Figure 14.5 showsthe graphs of the probability mass
function and the corresponding distribution functionfor a binomial distribution with parameters n = 5 and p = 4. EXAMPLE 2. Poisson distribution. Let il be a positive real number and let a randomvariable X assume the values 0, 1, 2, 3, . .
. . If the probability P(X = k) is given by theformula p(X = k) = $ for k = 0, 1, 2, . . .) the corresponding distribution function Fx is said to be a Poisson distribution with param-eter A. It is so named in honor of the French mathematician S. D. Poisson (1781-1840).This assignment of probabilities is permissible because xp(X = k) = g-k2 5 = e-“e” = 1.
k=O k=O ’ Exercises 523 The values of the distribution function are computed from the partial sums The Poisson distribution is applicable to many problems involving random eventsoccurring in time, such as traffic accidents, connections to wrong numbers in a telephoneexchange, and chromosome interchanges in cells induced by x-ray radiation.
Some specificapplications are discussed in the books by Feller and Parzen listed at the end of thischapter. 14.8 Exercises 1. A perfectly balanced die is rolled.
For a random variable X we take the function which countsthe number of points on the upturned face. Draw a graph of the corresponding distributionfunction F, .
2. Two dice are rolled. Let X denote the random variable which counts the total number ofpoints on the upturned faces.
Construct a table giving the nonzero values of the probabilitymass function pX and draw a graph of the corresponding distribution function FX . 3. The distribution function F of a random variable X is given by the following formulas: 0 i f t C-2, i f -2Consider a random variable X whose possible values are all rational numbers of the formn nfl - a n d -
n+l , where n = 1,2, 3, . .
. . Ifn P(X=$) =P(X+J) =A,verify that this assignment of probabilities is permissible and sketch the general shape of thegraph of the distribution function F”. 5. The probability mass function p of a random variable X is zero except at points t = 0, 1, 2.At these points it has the values p(0) = 3c3, p(1) = 4c - lOc2, p(2) = 5c - 1, forsomec >O.(a) Determine
the value of c.(b) Computethefollowingprobabilities: P(X < l),P(X < 2),P(l < X < 2),P(O < X < 3).(c) Describe the distribution function F and sketch its graph.(d) Find the largest t such that F(t) < 4.(e) Find the smallest t such that F(t) > t. 524 Calculus of probabilities 6. A random variable X has a binomial distribution with parameters n = 4 andp = +.
(a) Describe the probability mass function p and sketch its graph.(b) Describe the distribution function F and sketch its graph.(c) Compute the probabilities P(l < X 5 2) and. P(l < X 5 2).
7. Assume that if a thumbtack is tossed on a table, it lands either with point up or in a stableposition with point resting on the table. Assume there is a positive probability p that it landswith point up.(a) Suppose two identical tacks are tossed simultaneously. Assuming stochastic independence,show that the probability that both land with point up is
p2.(b) Continuing part (a), let X denote the random variable which counts the number of tackswhich land with point up (the possible values of Xare 0, 1, and 2).
Compute the probabilitiesP(X = 0) and P(X = 1).(c) Draw the graph of the distribution function F’ when p = *. 8. Given a random variable X whose possible values are 1, 2, . .
. , n. Assume that the prob-abilityP(X = k) is proportional to k. Determine the constant of proportionality, the probabilitymass function px , and the distribution function F”. 9. Given a random variable X whose possible values are 0, 1,2, 3, ... Assume that P(X = k)is proportional to CL/k!, where c is a fixed real number. Determine the constant of
proportion-ality and the probability mass function p.
10.
(a) A fair die is rolled. The sample space is S = { 1,2, 3,4, 5,6}. If the number of points onthe upturned face is odd a player receives one dollar; otherwise he must pay one dollar. LetX denote the random variable which measures his financial outcome (number of dollars) oneach play of the game. (The possible values of X are + 1 and - 1.) Describe the
probabilitymass function px and the distribution F, . Sketch their graphs.(b) A fair coin is tossed. The sample space S == {H, T}. If the outcome is heads a playerreceives one dollar; if it is tails he must pay one dollar. Let Y denote the random variablewhich measures his financial outcome (number of dollars) on each play of the game. Show thatthe
mass function py and the distribution FY are identical to those in part (a). This exampleshows that different random variables may have the same probability distribution function.Actually, there are infinitely many random variables having a given probability distributionF. (Why?) Such random variables are said to be identically distributed. Each
theoremconcerning a particular distribution function is applicable to any of an infinite collection ofrandom variables having this distribution. 11.
The number of minutes that one has to wait for a train at a certain subway station is knownto be a random variable X with the following probability mass function: pw = 0 unless t = 3k/lO forsome k =0,1,2 ,..., 10. p(t) = & if t = 0,0.3,0.6,0.9,2.1,2.4, 2.7, 3.0. pw = * i f t = 1.2, 1.5, 1.8. Sketch the graph of the corresponding distribution function F. Let A
be the event that onehas to wait between 0 and 2 minutes (including 0 and 2), and let B be the event that one hasto wait between 1 and 3 minutes (including 1 and 3). Compute the following probabilities:N),W),P(A nB),WI4,%4 UB). 12. (a) If0

0, let p =+forn>1. ShowthatQ, +lasn--ta3andthat n0k pkq’+ k--+ z e-1 a s n+co. This result suggests that for large n and small p, the binomial distribution is approximatelythe same as the Poisson distribution, provided the product np is nearly constant; this constantis the parameter 1 of the Poisson distribution. 14.9 Continuous distributions. Density
functions Let X be a one-dimensional random variable and let F be its distribution function, sothat F(t) = P(X 5 t) for every real t. If the probability P(X = t) is zero for every 1 then,because of Theorem 14.5, F is continuous everywhere on the real axis. In this case F iscalled a continuous distribution and X is called a continuous random variable. If the
deriva-tive F’ exists and is continuous on an interval [a, t] we can use the second fundamentaltheorem of calculus to write (14.13) F(t) - F(a) = s,‘fCu, du, where f is the derivative of F. The difference F(t) - F(a) is, of course, the probabilityP(a < X 5 t), and Equation (14.13) expresses this probability as an integral. Sometimes the distribution function F
can be expressed as an integral of the form(14.13), in which the integrand f is integrable but not necessarily continuous. Wheneveran equation such as (14.13) holds for all intervals [a, t], the integrand f is called a proba-bility density function of the random variable X (or of the distribution F) provided thatfisnonnegative. In other words, we have the
following definition: DEFINITION OF A PROBABILITY DENSITY FUNCTION. Let X be a one-dimensional randomvariable with a continuous distribution function F. A nonnegative function f is called aprobability density of X (or of F) zyf is integrable on every interval [a, t] and if (14.14) F(t) - F(a) = s,“f(u) du. If we let a -+ - co in (14.14) then F(a) + 0
and we obtain the important formula (14.15) F(t) = P(X 5 t) = fm f(u) du , valid for all real t. If we now let t + + co and remember that F(t) + 1 we find that (14.16) I+cO --41 f(u) du = 1. 526 Calculus of probabilities For discrete random variables the sum of all the probabilities P(X = t) is equal to 1.Formula (14.16) is the continuous analog of this
statement There is also a strong analogybetween formulas (14.11) and (14.15). The density functionf plays the same role for con-tinuous distributions that the probability mass functionp plays for discrete distributions -integration takes the place of summation in the computation of probabilities. There isone important difference, however. In the
discrete case p(t) is the probability that X = t,but in the continuous casef(t) is not the probability that X = t . In fact, this probability iszero because F is continuous for every t. Of course, this also means that for a continuousdistribution we have P(ab. (a) The distribution function F. (b) The density functionf. FIGURE 14.6 A uniform distribution over an
interval [a, 61 and the correspondingdensity function. Uniform distribution over an interval 527 This is a continuous distribution whose graph is shown in Figure 14.6(a).The derivative F’(t) exists everywhere except at the points a and b, and we can write where f is the density function, defined as follows: l/P - 4 i f aThe graph off is shown in Figure
14.6(b).The next theorem characterizes uniform distributions in another way. THEOREM 14.7. Let X be a one-dimensional random variable with all its values in a finiteinterval [a, b], and let F be the distribution function of X. Then F is uniform over [a, b] ifand only if (14.17) P(XEZ) = P(XEJ) for every pair of subintervals Z and J of [a, b] having the
same length, in which case we have P(X E Z) = A!-b - a ’where h is the length of I. Proof. Assume first that X has a uniform distribution over [a, b]. If [c, c + h] is anysubinterval of [a, b] of length h we have P(c < X 5 c + h) = F(c + h) - F(c) = ’ ; ’ -a a - ;fl = $ a . - - This shows that P(XE I) = P(XE J) = h/(b - a) for every pair of subintervals Z and J of[a,
b] of length h. To prove the converse, assume that X satisfies (14.17). First we note that F(t) = 0 ift < a and F(t) = 1 if t > b, since X has all its values in [a, b]. Introduce a new function g defined on the half-open interval (0, b - a] by the equation (14.18) g(u) = P(a < X 5 a + 24) i f OUsing additivity and property (14.17) we find g(u + v) = P(a < X I a +
u + v)=P(a 0, v > 0, u + v < b - a. This is known as Cauchy’sfunctionalequation. In a moment we shall prove that every nonnegative solution of Cauchy’sfunctional equation is given by g(u) = & g(b - 4 f o r O f o r O = G 4 + G(y) whenever x, y, and x + y are in (0, 11. If we prove that (14.19) G(x) = xG(I) f o r O. Taking y = x/n and using (14.20) we
obtain if x and mx/n are in (0, 11. In other words, we have (14.21) W-3 = rG(x) for every positive rational number r such that x and rx are in (0, 11.Now take any x in the open interval (0, 1) and let {r,} and {R,} be two sequences of rational numbers in (0, I] such that r,, < x < R, and such that limr, = lim R, = x.n-rcc n-rm Cauchy’s functional equation
and the nonnegative property of G show that G(x + y) 2G(x) so G is monotonic increasing in (0, 11. Therefore W,) I G(x) I W,).Using (14.21) we rewrite this as r,W) 5 G(x) I W(1). Letting n --t CO we find xG(1) 5 G(x) 5 xG(l) , so G(x) = xG(1) , which proves (14.19). 530 Calculus of probabilities Note: Uniform distributions are often used in experiments
whose outcomes are pointsselected at random from an interval [a, b], or in experiments involving an interval [a, b] asa target, where aiming is impossible. The terms “at random” and “aiming is impossible”are usually interpreted to mean that if Z is any subinterval of [a, b] then the probabilityP(X E Z) depends only on the length of Z and not on its
location in [a, b]. Theorem 14.7shows that uniform distributions are the only distributions with this property. We turn now to the probability questions asked at the beginning of this chapter. EXAMPLE. A line segment is broken into two pieces, with the point of subdivision chosenat random. Let X denote the random variable which measures the ratio
of the length of theleft-hand piece to that of the right-hand piece. Determine the probability distributionfunction F, . Solution. Use the interval [0, l] to represent the line segment and let the point ofsubdivision be described by the random variable Y(w) = co for each OJ in (0, 1). Since thepoint of subdivision is chosen at random we assume that Y has a
uniform distributionfunction Fy over [0, 11. Hence Fy(t) = t f o r OIf the segment is broken at co, then w/(1 - w) is the ratio of the length of the left-handpiece to that of the right-hand piece. Therefore X(w) = cu/(l - w). If t <: 0 we have F,(t) = 0 since the ratio X(o) cannot be negative. If t 2 0, theinequality X(o) < t is equivalent to w/(1 - o) 5 t , which is
equivalent to o 2 t/(1 + t).Therefore since 0 < t/(1 + t) < 1 .Now we can calculate various probabilities. For example, the probability that the two pieces have equal length is P(X = 1) = 0. In fact, since F, is a continuous distribution,the probability that X takes any particular value is zero. The probability that the left-hand segment is at least twice as
long as the right-handsegment isP(X>2)= 1 -P(X<2)= 1 -g=&. Similarly, the probability that theright-hand segment is at least twice as long as the left-hand segment is P(X < &) = $. Theprobability that the longer segment is at least twice as long as the shorter segment isP(X>2)+P(X<+)=Q. 14.11 Cauchy’s distribution A random variable X is said to
have a Cauchy distribution F if F(t) = i + 1 arctan tTr Cauchy’s distribution (a) The distribution function F. (b) The density function jI FIGURE 14.7 Cauchy’s distribution function and the corresponding density function. for all real t. This function has a continuous derivative everywhere; a continuous densityfunctionfis given by the formula f(t) = l?7(1
+ t2) * The graphs of F andfare shown in Figures 14.7(a) and (b), respectively.The following experiment leads to a Cauchy distribution. A pointer pivoted at the point (1, 0) on the x-axis is spun and allowed to come to rest. An outcome of the experimentis 8, the angle of inclination from the x-axis made by a line drawn through the pointer;8 is measured
so that -$r < 19 5 +T. Let X be the random variable defined by X(0) = 0,and let Y be the random variable which measures the y-intercept of the line through thepointer.
If 0 is the angle described above, then Y(0) = tan 8. We shall prove that Y has a Cauchy distribution Fy if X has a uniform distribution over[-gr, &T]. If a < t let IX = arctan a and let 8 = arctan t. Then we have Fy(t) - F,(a) = P(a < Y_< t) = P(a < X_< 0) =s ‘fx(u) du = 8--cr.a 7r Sincea+ +asa+--wefind e + %nFy(t) =___ =n A arctan t + i .7r This shows
that Y has a Cauchy distribution, as asserted.
532 Calculus of probabilities 14.12 Exercises 1. A random variable X has a continuous distribution function F, where 0 i f tl. (a) Determine the constant c and describe the density functionf.(b) Compute the probabilities P(X = &), P(X < 4) , P(lXl < 4). 2. Letf(t) = c /sin II for 1 t 1 < a/2 and letf(t) = 0 otherwise. Determine the value of the constantc so
that f will be the density of a continuous distribution function F. Also, describe F andsketch its graph. 3. Solve Exercise 2 iff(t) = c(4t - 2t2) for 0 5 I 5 2, andf(t) = 0 otherwise.4. The time in minutes that a person has to wait for a bus is known to be a random variable with density function f given by the following formulas : f(t) = +2 f o r O t) = 4. 7. The
Lethe Subway Company schedules a northbound train every 30 minutes at a certainstation.
A man enters the station at a random time. Let the random variable X count thenumber of minutes he has to wait for the next train. Assume X has a uniform distribution overthe interval [0, 301. (This is how we interpret the statement that he enters the station at“random time.”)(a) For each k = 5, 10, 15, 20, 25, 30, compute the probability that he has
to wait at leastk minutes for the next train.(b) A competitor, the Styx Subway Company, is allowed to schedule a northbound trainevery 30 minutes at the same station, but at least 5 minutes must elapse between the arrivalsof competitive trains. Assume the passengers come into the station at random times and alwaysboard the first train that arrives.
Show that the Styx Company can arrange its schedule so thatit receives five times as many passengers as its competitor. 8. Let X be a random variable with a uniform distribution F, over the interval [0, 11. Let Y =uX + b, where a > 0. Determine the distribution function F, and sketch its graph. 9. A roulette wheel carries the integers from 0 to 36,
distributed among 37 arcs of equal length.The wheel is spun and allowed to come to rest, and the point on the circumference next to atixed pointer is recorded. Consider this point as a random variable X with a uniform distri-bution. Calculate the probability that X lies in an arc containing (a) the integer 0; (b) aninteger n in the interval 11 < n I; 20 ;
(c) an odd integer. Exponential distributions 5 3 3 10. A random variable is said to have a Cauchy distribution with parameters a and b, where a > 0,if its density function is given by f(’) = ; a2 + ; _ b)2 * Verify that the integral offfrom - co to + co is 1, and determine the distribution function F.11. Letf,(t) = 1 for 0 < t < 1, and letf,(t) = 0 otherwise.
Define a sequence of functions (fn} by the recursion formula fn+~(x) = ~-~J(x - t)f,(t>dt. (a) Prove thatf,+,(x) = Jo& dt.(b) Make a sketch showing the graphs of fi , f2, and f 3. 12. Refer to Exercise 11. Prove that each functionf,, is a probability density. 14.13 Exponential distributions Let jl be a positive constant. A one-dimensional random variable X is
said to have anexponential distribution F with parameter A if 1 - e AtF(t) = f o r t20, 0 f o r tA corresponding density functionf is given by the formulas 1 AecAtf(t) = f o r t>O,o f o r t < O . The graphs of F and f are like those shown in Figure 14.8. A A -------- _--------- 0 0ct (a) The distribution function F. (b) The density functionf.
FIGURE 14.8 An exponential distribution and the corresponding density function. 534 Calculus of probabilities Exponential distributions have a characteristic property which suggests their use incertain problems involving radioactive decay, traffic accidents, and failure of electronicequipment such as vacuum tubes. This property is analogous to that
which characterizesuniform distributions and can be described as follows. Let X denote the observed waiting time until a piece of equipment fails, and let F be thedistribution function of X.
We assume that F(t) = 0 for r 5 0, and for the moment weput no further restrictions on F. If t > 0, then X 5 t is the event “failure occurs in theinterval [0, t].” Hence X > t is the complementary event, “no failure occurs in the interval[O, t].” Suppose that no failure occurs in the interval [0, t]. What is the probability of continuedsurvival in the interval [t,
t + s]? This is a question in conditional probabilities. We wishto determine P(X > t + s 1 X > t), the conditional probability that there is no failure inthe interval [0, t + s], given that there is no failure in the interval [0, t]. From the definition of conditional probability we have (14.22) P(X > t + s (X > t) = ‘[(‘ > ;:,‘;:,‘”’ ’ t)l = ‘a:; ;; ’) . Suppose now that F
is an exponential distribution with parameter 1 > 0. Then F(t) =1 - e-*t for t > 0, and P(X > t) = 1 - P(X 2 t) = e-At. Hence Equation (14.22)becomes e--L(t+s)P(X > t + s 1 x > t) = 7 = e--As = P(X > s). In other words, if the piece of equipment survives in the interval [0, t], then the probabilityof continued survival in the interval [t, t + s] is equal to the
probability of survival in theinterval [0, s] having the same length. That is, the probability of survival depends only onthe length of the time interval and not on the age of the equipment. Expressed in terms ofthe distribution function F, this property states that (14.23)1 - F(t + s) = 1 - F(s) for all1 - F(t) t > 0 and s > 0. The next theorem shows that
exponential distributions are the only probability distribu-tions with this property. THEOREM 14.9. Let F be a probability distribution ,function satisfying the functionalequation (14.23), where F(t) < 1 for t > 0. Then there is a positive constant A > 0 such that F(t) = 1 - e-At for all t > 0.
Proqf.
Let g(t) = -log [l - F(t)] for t > 0. Then 1 - F(t) = e-Ott), so to prove thetheorem it suffices to prove that g(t) = At for some il > 0. Now (g is nonnegative and satisfies Cauchy’s functional equation, g(t + 4 = g(t) + g(s) Normal distributions 535 for all t > 0 and s > 0. Therefore, applying Theorem 14.8 with c = I, we deduce thatg(t) = tg(1) for 0 < t 5 1 .
Let I = g(1). Then ;I = -log [I - F(l)] > 0, and henceg(t) = At for 0 < t < 1 . To prove that g(t) = At for all t > 0, let G(t) = g(t) - At. The function G also satisfiesCauchy’s functional equation, Moreover, G is periodic with period 1 because G(t + 1) =G(t) + G(1) and G(1) = 0. Since G is identically 0 in (0, I] the periodicity shows thatG(t) = 0 for all t > 0. In
other words, g(t) = At for all t > 0, which completes the proof. EXAMPLE 1. Let X be a random variable which measures the lifetime (in hours) of acertain type of vacuum tube. Assume X has an exponential distribution with parameter1 = 0.001. The manufacturer wishes to guarantee these tubes for T hours. Determine Tso that P(X > T) = 0.95.
Solution. The distribution function is given by F(t) = 1 - ePAt for t > 0, where2 = 0.001. Since P(X > T) = 1 - F(T) = e--IT, we choose T to make e-IT = 0.95.Hence T = -(log 0.95)/J = - 1000 log 0.95 = 51.25+ . EXAMPLE 2.
Consider the random variable of Example 1, but with an unspecified valueof 1. The following argument suggests a reasonable procedure for determining il. Startwith an initial number of vacuum tubes at time t = 0, and let g(t) denote the number oftubes still functioning t hours later. The ratio [g(O) - g(t)]/g(O) is the fraction of theoriginal number that
has failed in time t. Since the probability that a particular tube failsin time t is 1 - ecat, it seems reasonable to expect that the equation (14.24) g(O) - g(t) = 1 _ e-“tg(O) should be a good approximation to reality. If we assume (14.24) we obtain g(t) = g(0)e-“t. In other words, under the hypothesis (14.24), the number g(t) obeys an exponential
decaylaw with decay constant 2. The decay constant can be computed in terms of the half life.If t, is the half life then 4 = g(tJ/g(O) = eWAtl, so il = (log 2)/t,.
For example, if thehalf life of a large sample of tubes is known to be 693 hours, we obtain il = (log 2)/693 =0.001. 14.14 Normal distributions Let m and c be fixed real numbers, with o > 0. A random variable X is said to have anormal distribution with mean m and variance o2 if the density function f is given by theformula jyt) _ 1 e-Chn)lol~/2 a&r for
all real t. The corresponding distribution function F is, of course, the integral F(t) = -J oJ27T :a e-[(~-dfd2/2 du . s TABLE 14.1 Values of the standard normal distribution function t 0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.0 0.5000 0.5040 0.5080 0.5120 0.5160 0.5199 0.5239 0.5279 0.5319 0.53590.1 0.5398 0.5438 0.5478 0.5517 0.5557
0.5596 0.5636 0.5675 0.5714 0.57530.2 0.5793 0.5832 0.5871 0.5910 0.5948 0.5987 0.6026 0.6064 0.6103 0.61410.3 0.6179 0.6217 0.6255 0.6293 0.6331 0.6368 0.6406 0.6443 0.6480 0.65170.4 0.6554 0.6591 0.6628 0.6664 0.6700 0.6736 0.6772 0.6808 0.6844 0.6879 0.5 0.6915 0.6950 0.6985 0.7019 0.7054 0.7088 0.7123 0.7157 0.7190
0.72240.6 0.7257 0.7291 0.7324 0.7357 0.7389 0.7422 0.7454 0.7486 0.7517 0.75490.7 0.7580 0.7611 0.7642 0.7673 0.7704 0.7734 0.7764 0.7794 0.7823 0.78520.8 0.7881 0.7910 0.7939 0.7967 0.7995 0.8023 0.8051 0.8078 0.8106 0.81330.9 0.8159 0.8186 0.8212 0.8238 0.8264 0.8289 0.8315 0.8340 0.8365 0.8389 1.0 0.8413 0.8438 0.8461
0.8485 0.8508 0.8531 0.8554 0.8577 0.8599 0.86211.1 0.8643 0.8665 0.8686 0.8708 0.8729 0.8749 0.8770 0.8790 0.8810 0.88301.2 0.8849 0.8869 0.8888 0.8907 0.8925 0.8944 0.8962 0.8980 0.8997 0.90151.3 0.9032 0.9049 0.9066 0.9082 0.9099 0.9115 0.9131 0.9147 0.9162 0.91771.4 0.9192 0.9207 0.9222 0.9236 0.9251 0.9265 0.9279 0.9292
0.9306 0.9319 1.5 0.9332 0.9345 0.9357 0.9370 0.9382 0.9394 0.9406 0.9418 0.9429 0.94411.6 0.9452 0.9463 0.9474 0.9484 0.9495 0.9505 0.9515 0.9525 0.9535 0.95451.7 0.9554 0.9564 0.9573 0.9582 0.9591 0.9599 0.9608 0.9616 0.9625 0.96331.8 0.9641 0.9649 0.9656 0.9664 0.9671 0.9678 0.9686 0.9693 0.9699 0.97061.9 0.9713 0.9719
0.9726 0.9732 0.9738 0.9744 0.9750 0.9756 0.9761 0.9767 2.0 0.9772 0.9778 0.9783 0.9788 0.9793 0.9798 0.9803 0.9808 0.9812 0.98172.1 0.9821 0.9826 0.9830 0.9834 0.9838 0.9842 0.9846 0.9850 0.9854 0.98572.2 0.9861 0.9864 0.9868 0.9871 0.9875 0.9878 0.9881 0.9884 0.9887 0.98902.3 0.9893 0.9896 0.9898 0.9901 0.9904 0.9906 0.9909
0.9911 0.9913 0.99162.4 0.9918 0.9920 0.9922 0.9925 0.9927 0.9929 0.9931 0.9932 0.9934 0.9936 2.5 0.9938 0.9940 0.9941 0.9943 0.9945 0.9946 0.9948 0.9949 0.9951 0.99522.6 0.9953 0.9955 0.9956 0.9957 0.9959 0.9960 0.9961 0.9962 0.9963 0.99642.7 0.9965 0.9966 0.9967 0.9968 0.9969 0.9970 0.9971 0.9972 0.9973 0.99742.8 0.9974
0.9975 0.9976 0.9977 0.9977 0.9978 0.9979 0.9979 0.9980 0.99812.9 0.9981 0.9982 0.9982 0.9983 0.9984 0.9984 0.9985 0.9985 0.9986 0.9986 3.0 0.9987 0.9987 0.9987 0.9988 0.9988 0.9989 0.9989 0.9989 0.9990 0.99903.1 0.9990 0.9991 0.9991 0.9991 0.9992 0.9992 0.9992 0.9992 0.9993 0.99933.2 0.9993 0.9993 0.9994 0.9994 0.9994 0.9994
0.9994 0.9995 0.9995 0.99953.3 0.9995 0.9995 0.9995 0.9996 0.9996 0.9996 0.9996 0.9996 0.9996 0.99973.4 0.9997 0.9997 0.9997 0.9997 0.9997 0.9997 0.9997 0.9997 0.9997 0.9998 3.5 0.9998 0.9998 0.9998 0.9998 0.9998 0.9998 0.9998 0.9998 0.9998 0.99983.6 0.9998 0.9998 0.9999 0.9999 0.9999 0.9999 0.9999 0.9999 0.9999 0.9999 5 3 6
Normal distributions 531 FIGURE 14.9 The standard normal distribu-tion function: m = 0, G = 1. FIGURE 14.10 The density function of anormal distribution with mean m and variance 02. It is clear that this function F is monotonic increasing, continuous everywhere, and tendsto 0 as t 4 -co. Also, it can be shown that F(t) + 1 as t --f + cc. (See
Exercise 7 ofSection 14.16.) The special case m = 0, o = 1 is called the standard normal distribution. In this casethe function F is usually denoted by the letter 0. Thus, cD(t) = -& Irn e-u2‘2 du .s The general case can be reduced to the standard case by introducing the change of variablev = (U - m)/o in the integral for F. This leads to the formula F(t) =
tD t--m .c 1uA four-place table of values of O(t) for values of t spaced at intervals of length 0.01 is given in Table 14.1 for t = 0.00 to t = 3.69. The graph of 0 is shown in Figure 14.9. Thegraph of the densityfis a famous “bell-shaped” curve, shown in Figure 14.10. The top ofthe bell is directly above the mean m. For large values of o the curve tends to
flatten out;for small (5 it has a sharp peak, as in Figure 14.10.
Normal distributions are among the most important of all continuous distributions.Many random variables that occur in nature behave as though their distribution functionsare normal or approximately normal.
Examples include the measurement of the heightof people in a large population, certain measurements on large populations of livingorganisms encountered in biology, and the errors of observation encountered when makinglarge numbers of measurements. In physics, Maxwell’s law of velocities implies that thedistribution function of the velocity in
any given direction of a molecule of mass M in agas at absolute temperature T is normal with mean 0 and variance M/(kT), where k is aconstant (Boltzmann’s constant). The normal distribution is also of theoretical importance because it can be used toapproximate the distributions of many random phenomena. One example is the binomial 538
Calculus of probabilities 0.3 0.2 -/0 / IIII 1 3 4 5 6 7 8 9 10 FIGURE 14.11 The density function of a normal distribution considered as an approxi-mation to the probability mass function of a binomial distribution. distribution with parameters n and p. If X is a random variable having a binomial distri-bution with parameters IZ andp, the probability P(a 5
X 5 b) is given by the sum where 4 = 1 - p. For a large n, laborious computations are needed to evaluate this sum.In practice these computations are avoided by use of the approximate formula (14.25)i b--p+& where the symbol ry means that the two sides of (14.25) are asymptotically equal; that is,the ratio of the left member to the right member
approaches the limit 1 as IZ + cc. Thelimit relation expressed in (14.25) is a special case of the so-called central limit theorem ofthe calculus of probabilities. This theorem (discussed in more detail in Section 14.30)explains the theoretical importance of normal distributions.
Figure 14.11 illustrates approximate formula (14.25) and shows that it can be accurateeven for a relatively small value of n. The dotted lines are the ordinates of the probabilitymass function p of a binomial distribution with parameters n = 10 and p = +. Theseordinates were computed from the formula p(t)=P(X=t)= (:~)[~J~-~-’ f o r t=0,1,2,...,10.
Remarks on more general distributions 539 The ordinates for t = 7, 8, 9, and 10 are not shown because their numerical values aretoo near zero. For example, ~(10) = (i)‘” = 210/1010 = 0.0000001024. The smoothcurve is the graph of the density function f of a normal distribution (with mean m = np = 2and variance 8 = npq = 1.6). To compute the
probability P(a < t < b) from the massfunction p we add the function values p(t) at the mass points in the interval a < t < b.Each value p(t) may be interpreted as the area of a rectangle of height p(t) located overan interval of unit length centered about the mass point t. (An example, centered aboutt = 3, is shown in Figure 14.11.) The approximate
formula in (14.25) is the result ofreplacing the areas of these rectangles by the area of the ordinate set off over the interval[a - 4, b + 41. 14.15 Remarks on more general distributions In the foregoing sections we have discussed examples of discrete and continuous distribu-tions. The values of a discrete distribution are computed by adding the values
of thecorresponding probability mass function. The values of a continuous distribution with adensity are computed by integrating the density function. There are, of course, distribu-tions that are neither discrete nor continuous. Among these are the so-called “mixed”types in which the mass distribution is partly discrete and partly continuous. (An
exampleis shown in Figure 14.3.) A distribution function F is called mixed if it can be expressed as a linear combinationof the form (14.26) F(t) = @l(t) + c&(t), where FI is discrete and F, is continuous. The constants cr and c2 must satisfy the relations 0 < Cl < 1, 0 < c2 < 1, Cl + c2 = 1 . Properties of mixed distributions may be found by studying
those that are discrete orcontinuous and then appealing to the linearity expressed in Equation (14.26). A general kind of integral, known as the Riemann-Stieltjes integral, makes possible asimultaneous treatment of the discrete, continuous, and mixed cases.7 Although thisintegral unifies the theoretical discussion of distribution functions, in any
specific problemthe computation of probabilities must be reduced to ordinary summation and integration.In this introductory account we shall not attempt to describe the Riemann-Stieltjes integral.Consequently, most of the topics we discuss come in pairs, one for the discrete case andone for the continuous case. However, we shall only give complete
details for one case, leavingthe untreated case for the reader to work out. Even the Riemann-Stieltjes integral is inadequate for treating the most general distribu-tion functions. But a more powerful concept, called the Lebesgue-Stieltjes integral,$ doesgive a satisfactory treatment of all cases. The advanced theory of probability cannot beundertaken
without a knowledge of the Lebesgue-Stieltjes integral. t A discussion of the Riemann-Stieltjes integral may be found in Chapter 9 of the author’s MathematicalAnalysis, Addison-Wesley Publishing Company, Reading, Mass. 1957.: See any book on measure theory.
540 Calculus of probabilities 14.16 Exercises1. 2 . 3 . Let X be a random variable which measures the lifetime (in hours) of a certain type of vacuumtube. Assume X has an exponential distribution with parameter A = 0.001 . Determine Tso that P(X > T) is (a) 0.90; (b) 0.99. You may use the approximate formula -log (1 - x) =x + x2/2 in your
calculations.A radioactive material obeys an exponential decay law with half life 2 years. Consider thedecay time X (in years) of a single atom and assume that X is a random variable with anexponential distribution. Calculate the probability that an atom disintegrates (a) in the interval1 5 X 5 2 ; (b) in the interval 2 5 X 5 3 ; (c) in the interval 2 < X 5 3,
given that it hasnot disintegrated in the interval 0 < X < 2 ; (d) in the interval 2 5 X < 3, given that ithas not disintegrated in the interval 1 < X 5 2.The length of time (in minutes) of long distance telephone calls from Caltech is found to be arandom phenomenon with probability density function ((., t/3 f o r t>O,j w = 0 for t 50. Determine the value of
c and calculate the probability that a long distance call will last (a)less than 3 minutes; (b) more than 6 minutes; (c) between 3 and 6 minutes; (d) more than9 minutes. 4 . Given real constants 3, > 0 and c. Let le l(t-o) i f t>c,f 0) = o i f t < c . 5 . 6 . Verify that J”,/ (t) dt = 1, and determine a distribution function F havingfas its density.This is called an
exponential distribution with two parameters, a decay parameter 1 and alocation parameter c.State and prove an extension of Theorem 14.9 for exponential distributions with two parameters1 and c. 7 .
A random variable X has an exponential distribution with two parameters 1 and c. Let Y =aX + b, where a > 0. Prove that Y also has an exponential distribution with two parametersI’ and c’, and determine these parameters in terms of a, 6, c, and ii.In Exercise 16 of Section 11.28 it was shown that Jo” F2 dx = &/2. Use this result to provethat for u > 0
we have 8 . 9 . 10. 11.
A random variable X has a standard normal distribution Q. Prove that (a) Q(--x) = 1 - Q(x); (b)P(IXj < k) = 2@(k) - 1 ; (c)P(IXI > k) = 2(1 - Q(k)).A random variable X has a standard normal distribution @. Use Table 14.1 to calculate eachof the following probabilities: (a) P(X > 0) ; (b) P(1 < X < 2) ; (c) P(lXl < 3) ; (d)mfl > a.A random variable X has a
standard normal distribution a.
Use Table 14.1 to find a number csuch that (a)P(IXJ > c) = 4; (b) P(lX(> c) = 0.98.Assume Xhas a normal distribution function Fwith mean m and variance u2, and let @ denotethe standard normal distribution. Distributions of functions of random cariables 541 (a) Prove thatt - m F (f) = CD - (1u .
(b) Find a value of c such that P(jX - ml > c) = 4.(c) Find a value of c such that P(IX - ml > c) = 0.98. 12. A random variable Xis normally distributed with mean m = 1 and variance u2 = 4. Calculateeach of the following probabilities: (a) P(--3 5 X 5 3) ; (b) P(--5 5 X < 3). 13. An architect is designing a doorway for a public building to be used by people
whose heightsare normally distributed, with mean m = 5 ft. 9 in., and variance o2 where 0 = 3 in.
How lowcan the doorway be so that no more than 1 ’ A of the people bump their heads? 14. If X has a standard normal distribution, prove that the random variable Y = aX + b is alsonormal if a Z 0. Determine the mean and variance of Y.
15. Assume a random variable X has a standard normal distribution, and let Y = X2. (a) Show that Fy(t) = yJtT os eeu212 du if t 2 0. (b) Determine Fy(t) when 1 < 0 and describe the density function,fy . 14.17 Distributions of functions of random variables If p is a real-valued function whose domain includes the range of the random variableX, we can
construct a new random variable Y by the equation which means that Y(W) = m[X(w)] for each w in the sample space. If we know the distribu-tion function F, of X, how do we find the distribution FY of Y? We begin with an impor-tant special case. Suppose that CJJ is continuous and strictly increasing on the whole realaxis and takes on every real
value. In this case 31 has a continuous strictly increasinginverse y such that, for all x and y, Y = v(x) if and only if x = y(v). By the definition of FY we have Fy(t) = P(Y I t) = P[y(X) I t]. Since ~1 is strictly increasing and continuous, the events “p?(X) < t” and “X < y(t)” areidentical. Therefore P[p(X) 5 t] = P[X 5 y(t)] = F,[y(t)] . Hence the distributions
FYand Fx are related by the equation (14.27) f’=(t) = FxMtN.
When the distribution F, and the function y have derivatives we can differentiate bothsides of (14.27), using the chain rule on the right, to obtain F;(t) = Edy(Ol * y’(t). 542 Calculus of probabilities This gives us the following equation relating the densities: fu(t> = fxMt)l * v’(t). EXAMPLE 1. Y = aX + b , a > 0. In this case we have 9(x> = ax + b, Y(Y) =
Y--b) y’(y) = 1.a a Since 9 is continuous and strictly increasing we may write EXAMPLE 2. Y= x2. In this case q(x) = x2 and the foregoing discussion is not directlyapplicable because v is not strictly increasing. However, we can use the same method ofreasoning to determine Fy andf, . By the definition of Fy we have FY(t) = P(X2 5 t).
If t < 0 the event “X2 5 t” is empty and hence P(X2 < t) = 0. Therefore ISy(t) = 0 fort < O . I f t > O w e h a v e P(X” < t) = P(& 2 X 5 Jr) = F&f) - F,(&) + P(X = -4;). For a continuous distribution F, we have P(X = -4,) = 0 and we obtain the followingrelation between Fy and F,: I 0 i f t O . For all t < 0 and for those t > 0 such that F, is differentiable at
4; and at--& we havethe following equation relating the densities: J u(t) = r 0 i f t + f,(4) 2& if f> o Further problems of this type will be discussed in Section 14.23 with the help of two-dimensional random variables. 14.18 Exercises 1. Assume X has a uniform distribution on the interval [0, 11. Determine the distribution functionF, and a probability
densityf, of the random variable Y if:(a) Y = 3X + 1, (4 Y = log IX/,(b) Y = -3X + 1, (e) Y = log X2,(c) Y = x2, (f) Y = 9. Distributions of two-dimensional random variables 543 2. Let X be a random variable with a continuous distribution function F, . If e, is continuous andstrictly increasing on the whole real axis and if q(x) --f a as x --+ --00 and p(x) -+ b as
x --f $ ~0 ,determine the distribution function FY of the random variable Y = v(X). Also, compute adensity fy , assuming that F, and a, are differentiable. 3. Assume X has a standard normal distribution. Determine a probability density function of therandom variable Y when(a) Y=X2, (c) Y = 9,(b) Y = IXl’m, (d) Y = arctan X. 14.19 Distributions of two-
dimensional random variables The concept of a distribution may be generalized to n-dimensional random variablesin a straightforward way. The treatment of the case n = 2 will indicate how the extensiontakes place. If X and Y are two one-dimensional random variables defined on a common samplespace S, (X, Y) will denote the two-dimensional
random variable whose value at a typicalpoint (u of S is given by the pair of real numbers (X(w), Y(w)). The notation is an abbreviation for the set of all elements w in S such that X(o) 5 a and Y(w) 5 b ;the probability of this event is denoted by P(XThe set of points (x, J) such that x 5 a and y 5 b is the Cartesian product A x B ofthe two one-dimensional
infinite intervals A = {x 1 x 2 a} and B = {y 1 y 5 b}. Theset A x B is represented geometrically by the infinite rectangular region shown in Figure14.12.
The number P(X < a, Y < 6) represents the probability that a point (X(o), Y(w))lies in this region. These probabilities are the two-dimensional analogs of the one-dimen-sional probabilities P(X 5 a), and are used to define two-dimensional probability distribu-tions. DEFINITION. The distribution function of the two-dimensional random variable (X, Y)is
the real-valued,function F defined for all real a and b b v the equation F(a,b)=P(XIt is also known as the joint distribution of the two one-dimensional random variables X and Y.
To compute the probability that (X, Y) lies in a rectangle we use the following theorem,a generalization of Theorem 14.2(b). 544 Calculus qf probabilities FIGURE 14.12 An infinite rectangularregion A x B, where A = {x 1 x < a} and B = {y I y I b}. FIGURE 14.13 The event “X < b, Y 5 d” ex-pressed as the union of four disjoint events. THEOREM
14.10. Let F be the distribution function of a two-dimensional random variable(X, Y). Then if a < b and c < d we have (14.28) P(a < X < b, c < Y 5 d) = F(b, d) - F(a, d) - F(b, c) + F(a, c). Proof.
The two events “X < a, c < Y 5 d” and “X 5 a, Y < c” are disjoint, andtheir union is “X 5 a, Y < d .” Adding probabilities we obtain P(X 5 a, c < Y < d) +P(X < a, Y < c) = P(X 5 a, Y 2 d); hence P(X < a, c < Y 5 d) = F(a, d) - F(u, c). Similarly, we have P(a < X 5 b, Y I c) = F(b, c) - F(a, c). Now the four events “X 5 a, Y < c,” “X 5 a, c < Y 5 d,” “aThere are,
of course, corresponding formulas for the rec-tangles, [a, b] x [c, d], (a, b) x (c, d), [a, b) x [c, d), and so forth. Note: The analogy with mass may be extended to the two-dimensional case. Here thetotal mass 1 is distributed over a plane.
The probability P(a < X 5 b, c < Y < d)represents the total amount of mass located in the rectangle (a, b] x (c, d]. The numberF(a, b) represents the amount in the infinite rectangular region X < a, Y < b . As inthe one-dimensional case, the two most important types of distributions are those known asdiscrete and continuous. In the discrete case the
entire mass is located in lumps concen-trated at a finite or countably infinite number of points. In the continuous case the massis smeared all over the plane with a uniform or varying thickness. 14.20 Two-dimensional discrete distributions If a random variable (X, Y) is given we define a new function p, called the probabilitymass function of (X, Y), such
that p(x,y)=P(X=x, Y=y) for every pair of real numbers (x, y). Let T denote the set of (x, y) for whichp(x, y) > 0.It can be shown that T is either finite or countably infinite. If the sum of the p(x, y) forall (x, y) in T is equal to 1, that is, if (14.29) (r;ETP(X, Y) = 12 the random variable (X, Y) is said to be discrete (or jointly discrete). The points (x, y)in Tare
called the mass points of (X, Y).
Suppose that x1, x2, x3, . . . and yl, yz, y3, . . . are among the possible values of X andY, respectively, and let pij =fyx=xi, Y=yj). If each pij is positive and if the sum of all the pij is 1, then the probability of an event“(X, Y) E E” is the sum of all the pij taken over all xi and yj for which (xi, yj) E E. Weindicate this by writing J’KX, Y)EEI = 12 pij.h;,i& In
particular, since P(X 5 x, Y 5 u) = F(x, J), the joint distribution F (which is alsocalled discrete) is given by the double sum The numbers pij can also be used to reconstruct the probability mass functions pxand py of the one-dimensional random variables X and Y. In fact, if Eii denotes the 546 Calculus of probabilities event “X = .yi, Y = yj,” the events Eil,
Eiz, Ei3, . . . are disjoint and their union is theevent “X = xi .” Hence, by countable additivity, we obtain (14.31) P(X = Xi) = $P(E,,) = $ pii.j=l i=l Similarly, we have (14.32) P(Y = yj) = &YE,,) = 5 Pij fi=l Therefore, the corresponding one-dimensional distributions FX and FY can be computedfrom the formulas and For finite sample spaces, of course,
the infinite series are actually finite sums. 14.21 Two-dimensional continuous distributions. Density functions As might be expected, continuous distributions are those that are continuous over thewhole plane. For the majority of continuous distributions F that occur in practice thereexists a nonnegative function f (called the probability density of F)
such that the proba-bilities of most events of interest can be computed by double integration of the density.That is, the probability of an event “(X, Y) E Q” is given by the integral formula (14.33) f‘[(X, y> E Ql = j jr.Q When such anfexists it is also called a probability density of the random variable (X, Y),or a joint density of X and Y. We shall not
attempt to describe the class of regions Q forwhich (14.33) is to hold, except to mention that this class should be extensive enough toinclude all regions that arise in the ordinary applications of probability. For example, if ajoint density exists we always have (14.34) P(a < X S b, c < Y I d) = jjf(x, y) dx dy,R where R = [a, b] x [c, d]. The integrand f is
usually sufficiently well behaved for thedouble integral to be evaluated by iterated one-dimensional integration, in which case(14.34) becomes P(a < x 5 by C < y I 4 = j; [jabfk Y> dx] dy = jab [j;f(x, y) dy] dx.
In all the examples we shall consider, this formula is also valid in the limiting cases in Two-dimensional continuous distributions. Density functions 547 which a and c are replaced by -co and in which b and d are replaced by + co. Thus we have (14.35) F(b, 4 = jTa [j:, fh Y) dx] dy = j_“, [j;;, Rx, Y) dy] dx for all b and d, and (14.36) j’,” [j:: f(x. Y) dx] dy =
j:I [j;I f(x, y) dy] dx = 1, Equations (14.35) and (14.36) are the continuous analogs of (14.30) and (14.29),respectively.
If a density exists it is not unique, since the integrand in (14.33) can be changed at afinite number of points without affecting the value of the integral.
However, there is atmost one continuous density function. In fact, at points of continuity off we have theformulas fk Y> = D,,,F(x, Y> = &,,WG Y) 3 obtained by differentiation of the integrals in (14.35).As in the discrete case, the joint density f can be used to recover the one-dimensional densities fX and fy. The formulas analogous to (14.31) and
(14.32) are fx(x> = j_+,m Rx, Y> dy and h.4~) = jTIf(x, Y> dx. The corresponding distributions F,(t) and Fy(t) are obtained, of course, by integrating therespective densitiesf, and fu from - co to t. The random variables X and Y are called independent if the joint distribution F(x, y) canbe factored as follows, F(x, Y) = FxCW,(y) for all (x, y). Some
consequences of independence are discussed in the next set of exercises. EXAMPLE. Consider the functionfthat has the constant value 1 over the square R =[0, l] x [0, 11, and the value 0 at all other points of the plane. A random variable (X, Y)having this density function is said to be uniformly distributed over R.
The correspondingdistributions function F is given by the following formulas: IXY if (x,y)~R, X i f O 1, F(x,Y) = ‘Y i f O 1, 1 i f x21 and y2 1, (0 otherwise. The graph of F over R is part of the saddle-shaped surface z = xy .
At all points (x, y) noton the boundary of R the mixed partial derivatives D,,,F(x, y) and D,,,F(x, y) exist and 548 Calculus of probabilities are equal to f(x, u). This distribution is the product of two one-dimensional uniformdistributions Fx and Fy . Hence X and Y are independent. 14.22 Exercises 1. Let X and Y be one-dimensional random variables with
distribution functions F, and Fr,and let F be the joint distribution of X and Y.(a) Prove that X and Y are independent if and only if we have P(a < X I; b, c < Y < d) = P(a < X I b)P(c < Y 2 d) for all a, b, c, d, with a < b and c < d.(b) Consider the discrete case. Assume x1, x2, . . . and y, , yz, . . . are the mass points of Xand Y, respectively. Let ai = P(X = xi)
and bj = P(Y = vj). If pii = P(X = xi, Y = YJ,show that X and Y are independent if pij = aibj for all i and j.(c) Let X and Y have continuous distributions with corresponding densities fx and fy andlet f denote the density of the joint distribution.
Assume the continuity of all three densities.Show that the condition of independence is equivalent to the statement f(x, y) =fx(x)fr~)for all (x, y). [Hint: Expressfas a derivative of the joint distribution F.] 2. Refer to Exercise 1. Suppose that P(X = x1, Y = yJ = P(X = x2, Y = yz) = p/2 and thatP(X=x,, Y=J$) =P(X=x,, Y =yl) =q/2, where p and q are
nonnegative with sum 1.(a) Determine the one-dimensional probabilities P(X = xi) and P(Y = yi) for i = 1, 2 andj = 1,2.(b) For what value (or values) of p will X and Y be independent? 3. If a < band c < d, definefas follows: fk y> =I (b - kid - clif (x, y> E [a, bl x [c, 4, 0 otherwise. (a) Verify that this is the density of a continuous distribution F and
determine F.(b) Determine the one-dimensional distributions F, and Fy .(c) Determine whether or not X and Y are independent. 4.
If P(Y 5 6) # 0, the conditional probability that X < a, given that Y < b , is denoted byP(X < a 1 Y 5 6) , and is defined by the equation P(X 5 a 1 Y < b) =P(X < a, Y I b) P(YSb) . IfP(Y15. Letf(x) = l/x for x # 0.(a) Show that the constant polynomial P that best approximates f over the interval [l, n]relative to the square norm is P(x) = (log n)/(n - 1).
Compute IIP--f/l2 for this P.(b) Find the linear polynomial P that best approximates f over the interval [l , n] relative tothe square norm. Compute IIP--fjlz for this P when n = 2.Let f(x) = eZ .(a) Show that the constant polynomial P that best approximates f over the interval [0, n]relative to the square norm is P(x) = (en - 1)/n.
Compute 11 P--f II2 for this P.(b) Find the linear polynomial P that best approximates f over the interval [0, l] relative tothe square norm. Compute IIP - f II2 for this P.LetP,,P,,... , P, be n + 1 polynomials orthonormal on [a, b] relative to the inner productin Exercise 11. Assume also that Pk has degree k.(a) Prove that any three consecutive polynomials
in this set are connected by a recurrencerelation of the form Pk+dx) = (akx + bk)pk(x) + ~kp~--~(x) for 1 I k < n - 1, where akr bk, ck are constants.(b) Determine this recurrence relation explicitly when the polynomials are the orthonormalLegendre polynomials.Refer to Exercise 14, and let pk denote the coefficient of xk in P&c).(a) Show that ak =
pk+Jpk.(b) Use the recurrence relation in Exercise 14 to derive the formula m c P,(x)P,(y) = p$pm+l(4p&> --p,(x)p,+l(y) 7k=O m x - Y valid for x # y . Discuss also the limiting case x = y . 15.6 Interpolating polynomials We turn now to approximation by interpolation polynomials. The values of a functionfare known at n + 1 distinct points x0, x1, .
. .
, x, and we seek a polynomial P of degreefn that satisfies the conditions (15.4) P(xo> =f(xo), %> =fW, * * * > p&J =fbJ* First we prove that if such a polynomial exists it is unique. Then we prove it exists byexplicit construction. This polynomial minimizes the distance from f to P, measured inthe interpolation seminorm for this n, Ilf - PII =k~Mx,, - JY%)l
* Since this distance is 0 if P satisfies (15.4), the interpolating polynomial P is the bestapproximation relative to this seminorm. THEOREM 15.1. UNIQUENESS THEOREM. Given n + 1 distinct points x0, x1, . . . , x,, letP and Q be two polynomials of degree . , n . Then P(x) = Q(x) for all x. 580 Introduction to numerical analysis Proof. Let R(x) = P(x) -
Q(x). The function R is a polynomial of degree . . , x,. The only polynomial with thisproperty is the zero polynomial. Therefore R(x) = 0 for all X, so P(x) = Q(X) for all x. The interpciating polynomial P can be constructed in many ways.
We describe first amethod of Lagrange. Let A(x) be the polynomial given by the equation (15.5) A(x)=(x-xx,)(x-xx,)~ (x-xx,)=fi(x-x,).i=O This polynomial has a simple zero at each of the points xj. Let F&(X) denote the poly-nomial of degree n obtained from A(x) by deleting the factor x - x, . That is, let (15.6) AR(X) = fi(x - Xj).j=OI#k The polynomial Ak(x)
has a simple zero at each point xj # x,. At the point X, itself wehave (15.7) Ak(xk) = -f,--cxk - xi> ‘j=O5+k This is nonzero since no factor in the product is zero. Therefore the polynomialAk(x)/Ak(xk) has the value 1 when x = xk and the value 0 when x = xj for xi # x,. Nowlet When x = xj, each term in this sum vanishes except the jth term, which has the
valuef(xj). Therefore P(xJ =f(xJ for each j. Since each term of this sum is a polynomial ofdegree in, the sum itself is a polynomial of degree In. Thus, we have found a polynomialsatisfying the required conditions. These results can be summarized by the followingtheorem : THEOREM 15.2. Given n + 1 distinct points x0, x1, . . . , x, and n + 1 real numbers
f(xo>7fW ’ * . , f (x,), not necessarily distinct, there exists one and only one polynomial P ofdegree In such that P(xJ = f (xi) for each j = 0, 1, 2, . . . , n , This polynomial is given bythe formula (15.8) where A,(x) is the polynomial defined by (15.6). Interpolating polynomials 581 Formula (15.8) for P(x) is called Lagrange’s interpolation formula.
We can write it inthe form P(x) =k~o f~~kwk(x) 9 where Lk(x) is a polynomial of degree IZ given by (15.9)Arc(x)Lk(X) = . .4SxJ Thus, for each fixed x, P(x) is a linear combination of the prescribed valuesf(x,),f(x,),. . . ,f(xJ. The multipliers Lk(x) depend only on the points x,, , x1, . . .
, x, and not on theprescribed values. They are called Lagrange interpolation coejficients. If we use theformulas in (15.6) and (15.7) we can write Equation (15.9) in the form (15.10) Lk(X) = fl= .j=o Xk - xji#k This product formula provides an efficient method for evaluating the number Lk(x) fora given x. Note: The Lagrange coefficients &(x) are often
expressed in the form A,(x)Lk(x) = I A (xk) ’ where A’ is the derivative of the polynomial in (15.5). To prove this formula it suffices toshow that A’(xJ = A,(x,) . Differentiating the relation A(x) = (x - %cb’b(--d we obtain A’(x) = (x - x,)Ak(x) + Ak(x). When x = xk this gives us A’(&) = Ak(xk) . EXAMPLE.
Determine the polynomial of degree 13 that takes the values yo, y, , yz , y3at the points -2, - 1, 1, 2, respectively. Solution. We take x0 = -2, x1 = -1 , x2 = 1, xg = 2. The polynomials Lk(x) in(15.10) are given by the formulas Lo(x) = (x + 1)(x - 1)(x - 2)=(- 2 + l)(- 2 - l)(- 2 2) - + - - - ; (x 1)(x 1)(x 2), (x + 2)(x 1)(x - L1(x) 2)= = (- 1 + 2)(l - 1)(l - 2)f(x+2)(x-
1)(x 2),6 L,(x) = (x + 2)(x + w - 2)(1 + 2)(1 + l)(l - 2) = 4,+2)(x+1)(x 2),6 L3(x) = (x + 2)(x + 1)(x - 1)(2 + 2)(2 + 1x2 - 1) = ;1,(x + 2)(x + 1)(x - 1). 582 Introduction to numerical analysis Therefore the required polynomial is P(x) = YoLo(X) + Y&,(X) + Y&(X) + Y&,(X) = - ; (x + 1)(x - 1)(x - 2) + 2 (x + 2)(x - 1)(x - 2) - f (x + 2)(x + 1)(x - 2) + ; (x + 2)(x + 1)
(x - 1). To compute the value of P(x) for a specific x it is usually better to leave the polynomialin this form rather than to rewrite it in increasing powers of x. For example, if y0 = -5,y,=l,y,=l,andy,=7,thevalueofP(x)forx=+isgivenby 15.7 Equally spaced interpolation points In the foregoing discussion the interpolation points x,, , x1, . . . , x, were assumed
to bedistinct but otherwise arbitrary. Now we assume they are equally spaced and show thatthe Lagrange coefficients &(x) can be considerably simplified. Suppose x,, < x1 . , IZ. Since xk - xj = (k - j)h , Equation (15.10) becomes (15.11) wherex - x0 t=- h * In the last term on the right of (15.11) the product of the factors independent of t is (15.12) (
l)n-k = k! (n - k)!(l)“-” n=--0n! k’ where ($) is the binomial coefficient. Since x = x0 + th , Equation (15.11) now becomes (15.13) Lk(xo + th) =‘+(i) c(t--j).3 Oi#k Error analysis in polynomial interpolation 583 For each fixed n, the right member of (15.13) is a function of k and t that can be tabulated.Extensive tables of the Lagrangian coefficients for
equally spaced interpolation points havebeen prepared by the National Bureau of Standards. (See Reference 13 in the bibliographyat the end of this chapter.) If x and h are chosen so that the number t = (x - x,)/h is onefor which the Lagrangian coefficients L,(x, + th) are tabulated, the actual calculation ofP(x,, + th) is reduced to a multiplication of
thef(x,) by the tabulated L,(x, + th), followedby addition. 15.8 Error analysis in polynomial interpolation Let f be a function defined on an interval [a, b] containing the n + 1 distinct points x0, Xl, ***,x,, and let P be the interpolation polynomial of degree In which agreeswith f at these points. If we alter the values off at points other than the
interpolationpoints we do not alter the polynomial P. This shows that the function f and the poly-nomial P may differ considerably at points other than the interpolation points.
If thegiven function f has certain qualities of “smoothness” throughout the interval [a, b] wecan expect that the interpolating polynomial P will be a good approximation to f at pointsother than the xk. The next theorem gives a useful expression that enables us to study theerror in polynomial interpolation when the given function has a derivative of
order n + 1throughout [a, b]. THEOREM 15.3 . Let x0, x1, . .
. , x, be n + 1 distinct points in the domain of a function f,and let P be the interpolation polynomial of degree In that agrees with fat these points.Choose a point x in the domain off and let [c(, p] be any closed interval containing the pointsx0, Xl,. * f , x,, and x. lff has a derivative of order n + 1 in the interval [LX, /I] there is atleast one point c in the
open interval (CI, 1) such that (15.14) where f(x) - P(x) = (of“+”, A(x) = (x - x0)(x - x1) * * . (x - x,). Note: Point c depends on both x and n. Proof. If x is one of the interpolation points x,, then A(x,) = 0 and Equation (15.14)is trivially satisfied for any choice of c in (CC, @). Suppose, then, that x is not one of theinterpolation points. Keep x fixed and
define a new function F on [a, p] by the equation (15.15) F(t) = &)[f(t) - P(t)1 - AW[f(x) - JWl. The right-hand side of this equation, as a function of t, has a derivative of order n + 1 ;hence the same is true of the left-hand side. Since P(t) is a polynomial in t of degree In,its (n + 1)st derivative is identically zero. The polynomial A(t) has degree n + 1,
theterm of highest degree being P+l, and we have A(‘+l)(t) = (n + l)! . Therefore, if wedifferentiate Equation (15.15) n + 1 times with respect to t we obtain the formula (15.16) F(n+l)(t) = A(x)- (n+l)(t) - (n + l)! [f(x) - P(x)]. 584 Introduction to numerical analysis From the definition in Equation (15.15) we see that F has the value zero at the n +
1interpolation points x,, , x1, . . .
, x, and also at the point x. Therefore F(t) = 0 at n + 2distinct points in the interval [M, /3].
These points determine n + 1 adjacent subintervalsof [a, p] and the function F vanishes at both endpoints of each of these subintervals. ByRolle’s theorem, the derivative F’(t) must be zero for at least one t interior to each sub-interval. If we choose exactly one such t from each subinterval we obtain n + 1 distinctpoints in the open interval (c(, j3) at
which F’(t) = 0. These points, in turn, determine nsubintervals at whose endpoints we have F’(t) = 0.
Applying Rolle’s theorem to F’ wefind that the second derivative F”(t) is zero for at least n distinct points in (a, p).
Afterapplying Rolle’s theorem n + 1 times in this manner we finally find that there is at leastone point c in (a, ,Q at which FCn+l) (c) = 0. Substituting this value of c in Equation(15.16) we obtain (n + l)! [f(x) - P(x)] = A(x)f(“+l)(c), which is the same as (15.14). This completes the proof. It should be noted that, as with approximation by Taylor
polynomials, the error terminvolves the (n + 1)st derivativef (nfl)(~) evaluated at an unknown point c. If the extremevalues of ffn+l)in [K, p] are known, useful upper and lower bounds for the error can beobtained. Suppose now that the interpolation points are equally spaced and that x0 < x1 In each case find the polynomial P of lowest possible
degree satisfying the given conditions.(a) P(-1) = 0, P(0) = 2, P(2) = 7.(b) P(1) = 1, P(2) = 0, P(3) = 0, P(4) = 1.(c) P(1) = 1, P(2) = 2, P(3) = 3, P(0) = 1.(d) P(0) = -2, P(1) =0, P(-1) = -2, P(2) = 16.(e) P (--2) = 11 , P(--1) = --11 , P (0) = --5 , P(1) = --1 . 2. Letf(x) = cos (CC/~). Find the polynomial of smallest possible degree that takes the samevalues
as f at the points -2, -$, 0, 8, 2.
3.
Let P be a polynomial of degree s;n and let A(x) = (X - x&(x - x,) . (x - A$, wherex0,x1,. * . , x, are n + 1 distinct points.(a) Show that for any polynomial B the polynomial Q given by Q(x) = P(x) + A(x)B(x)agrees with P at the points x0, x1, . . . , x,,. 586 Introduction to numerical analysis (b) Prove also the converse. That is, if Q is any polynomial that
agrees with P at the pointsxo,xl,.*~ ,&, then Q(x) = P(x) + A(x)B(x) for some polynomial B. 4. (a) Find the polynomial Q of lowest possible degree that satisfies the conditions Q (--2) = --5 , Q (--1) = --1 , Q(l) = 1, Q’(0) = --1 . [Hint: First find a polynomial P that takes the prescribed values at -2, --1, 1,and then use Exercise 3 to determine Q.] (b) Find
the polynomial Q of lowest possible degree that satisfies the conditions in part (a)with Q’(0) = -3 instead of Q’(0) = -1. 5. Letf(x) = lo&x for x > 0. Compute P(32), where P is the polynomial of lowest possibledegree that agrees with f at the points :(a) x = 1,64. (c) x = 4, 16,64.(b) x = 1, 16,256. (d) x = 1,4, 16,64,256.In each case compute the
differencef(32) - P(32). These examples show that the accuracy inpolynomial interpolation is not necessarily improved by increasing the number of interpolationpoints. 6. The Lagrange interpolation coefficients &(x) given by Equation (15.10) depend not only onx but also on the interpolation points x0, x1, . . . , x, . We can indicate this dependence
bywriting L,(x) = L,(x; X), where X denotes the vector in (n + I)-space given by X =(x0,x1,... ,&J. For a given real number b, let b denote the vector in (n + 1) space all ofwhose components are equal to b. If a # 0, show that L&x + b ; ax + b) = &(x; X).
This is called the invariance property of the Lagrange interpolation coefficients. The nextexercise shows how this property can be used to help simplify calculations in practice. 7.
Let P denote the polynomial of degree <4 that has the values P(2.4) = 72, P(2.5) = 30, P(2.7) = 18, P(2.8) = 24, P(3.0) = 180. (a) Introduce new interpolation points uj related to the given points x by the equation10 = 10xj - 24. The ui are integers. For each k = 0, 1,2, 3,4, determine the Lagrangeinterpolation coefficients &(x) in terms of u, where u =
10x - 24.(b) Use the invariance property of Exercise 6 to compute P(2.6).
8. A table of the functionf(x) = log x contains entries for x = 1 to x = 10 at intervals of 0.001.Values intermediate to each pair of consecutive entries are to be computed by linear inter-polation. Assume the entries in the table are exact.(a) Show that the error in linear interpolation will not exceed & in the sixth decimal place.(b) For what values of x will
linear interpolation be satisfactory for a seven-place table?(c) What should be the spacing of the entries in the interval 1 5 x 2 2 so that linear inter-polation will be satisfactory in a seven-place table? In Exercises 9 through 15, x0, x1, .
. .

, x, are distinct points and 44 = fi (x - Xj),GO AkW = fi (x - $1,60ifk Mx)Lk(X) = .A&k) Exercises 587 9. Derive the formula A'(x) = zz==, A,(x) by use of (a) logarithmic differentiation; (b)Lagrange's interpolation formula. 10. Prove each of the following formulas: (a) 2 &Ax) = 1 and R A&dc y - = ok=O A CxJ for all x.k=O Cb) IL&= 0. [Hint: Use part (a)
with suitable values of x.1 kc=0 11. Let P be any polynomial of degree <,n. Show that the coefficient of xn is equal to n m/Jck=O AI * 12. (a) Determine a and b so that the polynomial pk(x) = {a + b(x - &)}&(x)2will have the following properties : Pk(XJ = 0 for all i, p;(b) Determine c and d so that the polynomial Qk(X> = {C + 4x - xk)}Lk(x)2 will have
the following properties: Q&k) = 1 9 Q&i) = 0 f o r i#k, and Q; = 0 for all i . 6) Let f&d = cEso f(xk)Qk(x) + zE=t=, f'(Xk)Pk(X)T where f is a given function that isdifferentiable at x0, x1, . . . , x,.
Prove that ff(xJ = f(Xi) and H'(xJ = f'(xJ for all i.
Prove also that there is at most one polynomial H(x) of degree 12n + 1 with this property.13. (a) Let P and Q be two polynomials of degree In satisfymg the n + 1 conditions P(xo) = Q(xo) , P'W = Q'W, P"(x2) = Q"(x~), . . . , Pcn)(x,) = Qtn'(x)n * Prove that P(x) = Q(x) for all x.(b) Let B,(x) = 1, and for n 2 1 define B,(x) =x(x - n)+ l tt! Show that B;(x) =
B, ,(x - 1) for n 2 1 and deduce that B,(O) = B;(l) = B;(2) = . . . = B, l'(n - 1) = 0 and B:)(n) = 1. 588 Introduction to numerical analysis 14.
15. 16. (c) Show that the one and only polynomial of degree In satisfying the conditions P(O) = co, P'(1) = Cl, p"(2) = c2, .
. . , P'"'(n) = c,is given by P(x) = 2 CJJ?k(X) .k=O (d)Ifx,=x,+khfork=0,1,2,... , n, where h > 0, generalize the results in (b) and (c).Assumex,,x,,..., x, are integers satisfying x0 < x1 < . . * < x,.(a) Prove that IA' 2 k! (n - k)! and deduce that n c1 2" k=olA'ol (b) A"(x) A,(x) nA (xl?=A'0 cj=O&+i:& g$&. j=Oi#k i#k i#k A"(x,) n 1(cl) =2 A cxk) cj=. xk - xj *
j+k Let P,(x) be the polynomial of degree In that agrees with the functionf(x) = ea" at the n + 1integers x =O,1,... ,n. Since this polynomial depends on u we denote it by P,(x; a).Prove that the limit limP,(x; a) - 1 a 0 a exists and is a polynomial in x. Determine this polynomial explicitly. 15.10 Newton's interpolation formula Let P, denote the
interpolation polynomial of degree =Lz5(x)f(xk) 3 where Lk(x) is a polynomial of degree n (the Lagrange interpolation coefficient) given bythe product formula (15.19) Lk(x)=fix, for k=0,1,2 ,..., n.GO Xk - xji+k Newton's interpolation formula 589 Suppose we adjoin a new interpolation point x,+~ to the given points x,, , x1, . . . , x, . Todetermine the
corresponding polynomial P,+l by Lagran,ge's formula it is necessary tocompute a new interpolation coefficient Lnfl and to recompute all the earlier coefficients Lo, L,, *. ., L,, each of which is now a polynomial of (degree n + 1. In practice thisinvolves considerable labor. Therefore it is desirable to have another formula for deter-mining P, that
provides an easier transition from P, to P,+l. One such formula wasdiscovered by Newton; we shall derive it from the following theorem. THEOREM 15.4. Given n + 2 distinct points x,, , x1, . . . , x,, x,+~. Let P, be the poly-nomial of degree In that agrees with a given function f at x0, . . . , x,, and let P,+l be thepolynomial of degree In + 1 that agrees
with f at x,, , x1, . . . , x, , x,+~. Then there is aconstant c,+~, uniquely determ&ied by f and by the interpolation points x,, , . . .
. , x,+~, suchthat (15.20) P,+,(x) = P,(x) + %+1(X - x3 * * * (x - xn) * Proof. Let Q(x) = P,(x) + c(x - x,,) . . . (x - x,) , whlere c is an unspecified constant.Then Q is a polynomial of degree In + 1 that agrees with P, and hence with f at each ofthen+lpointsx,,...,x,. Now we choose c to make Q agree with f also at x,+~.
Thisrequires f (x,+1) = P,(x,+1) + 4x,+1 - x0) * * * (x,+1 - xn) * Since the coefficient of c is nonzero, this equation has a unique solution which we callc,+~. Taking c = c,+~ we see that Q = P,+l. The next theorem expresses P,(x) in terms of the numbers cl, .
. . , c,. THEOREM 15.5. NEWTON'SINTERPOLATIONFORMULA. ZfxO,...,x, aredistinct, wehave (15.21) Proof. We define P,,(x) = f (x, J and take n = 0 in (15.20) to obtain PI(X) = f (xcl) + c1(x - xll) * Now take n = 1 in (15.20) to get Pz(x> = PI(X) + cz(x - x&t - Xl> = f (x0) + c1(x - x0) + 4x - x0)(x - x1) * By induction, we obtain (15.21).
The property of Newton's formula expressed in Equation (15.20) enables us to calculatePnil simply by adding one new term to P, . This property is not possessed by Lagrange'sformula. The usefulness of Newton's formula depends, of course, on the ease with which thecoefficients cr, cZ, . . . , c, can be computed. The next thelorem shows that c, is a
linearcombination of the function values f (x,) , . . . , f (x,). 590 Introduction to numerical analysis~ THEOREM 15.6. The coeficients in Newton's interpolation formula are given by (15.22) c = n f(Xk)n c where Ak(xk) = c (xk - xj>.k=O A,kJ 'i+k Proof. By Lagrange's formula we have P,(x) =k~oww(~k~ f where Lk(x) is the polynomial of degree IZ given
by (15.19). Since the coefficient of xn inLk(x) is 1/,4,(x,), the coefficient of x" in P,(x) is the sum appearing in (15.22). On the otherhand, Newton's formula shows that the coefficient of xn in P,(x) is equal to c,. Thiscompletes the proof. Equation (15.22) provides a straightforward way for calculating the coefficients inNewton's formula.
The numbers A,(x,) also occur as factors in the denominator of theLagrange interpolation coefficient Lk(x). The next section describes an alternate methodfor computing the coefficients when the interpolation points are equally spaced. 15.11 Equally spaced interpolation points. The forward difference operator In the case of equally spaced
interpolation points with xk = x0 + kh for k = 0, 1, . . . , nwe can use Equation (15.12) to obtain -=fi'=Lfl' = 1 (l)" " n Ak(xk) i=o Xk - xj 0j*k h";;;k j n!h" k' In this case the formula for c, in Theorem 15.6 becomes (15.23) c n = -$ & l,n k(;)f(xk), The sum on the right can be calculated in another way in terms of a linear operator Acalled the forward
dtyerence operator. DEFINITION.
Let h be a fixed real number and let f be a given function. The functionAf dejmed by the equation AfCx) = f = U (x + 2h) - f (x + h)) - {j (x + h) -f(x)) =f(x + 2h) - 2f(x + h) + f(x).In general, we have (15.24) A"f(x) =$ l)'+k(;)f(x + kh).k=O This is easily proved by induction on it, using the law of Pascal's triangle for binomialcoefficients : (kii 1) + (i) =
("Y).Now suppose f is defined at n + 1 equally spaced points xk = x0 + kh for k = 0, 1, . . . , n . Then from (15.23) and (15.24) we obtain the formula cn = j& A"f(xo) . This provides a rapid method for calculating the coefficients in Newton's interpolationformula. The diagram in Table 15.1, called a diflerence table, shows how the successivedifferences
can be systematically calculated from a tabulation of the values offat equallyspaced points. In the table we have written fk forf(x,). Newton's interpolation formula (15.21) now becomes (15.25) P,(x) =f(x,> +ZAS kg(x - Xi). TABLE 15.1 x x0 Xl x2 X3 Tf(X) Af (x> A"f (xl A3fW hL fi - f o = Af(xo) fl; L Af (x1) - Af(x,) = A2f(;lco) fi - fi = AfCx3' 'xA"f(xJ
@f(x,,> = @fCxo> hf LAfCx,, - Af(xJ = Aafd Lfs -fi = AfCxd' f3f 592 If we write Introduction to numerical analysis k l k l x0 jh)=h"E(T j) =h'TJ(t j), where t = (x - x,)/h, Equation (15.25) becomes (15.26) P,(x) = f(x0> +gl '* 'E (t - j) .
15.12 Factorial polynomials The product t(t - 1) . . * (t - k + 1) which appears in the sum in (15.26) is a polynomialin t of degree k called a factorial polynomial, or the factorial kth power of t. It is denotedby the symbol t ck). Thus, by definition, k - l t(k) = TJ 0 - j) . We also define t(O) = 1. If we consider the forward difference operator A with h = 1,that
is, Af (x) = f (x + 1) - f(x), we find that At(n) = nt(-l) f o r n>l. This is analogous to the differentiation formula Dt" = ntn-l for ordinary powers. Thus,the factorial power ten) is related to differences much in the same way that the ordinarypower tn is related to derivatives. With the use of factorial polynomials, Newton's interpolation formula (15.26)
becomes P,(x, + th) =cA+ tfk).k=O ' Expressed in this form, Newton's formula resembles the Taylor formula for the poly-nomial of degree In that agrees with f and its first n derivatives at x0. If we write t0+k' t(t - 1) * * . (t - k + 1)= = k k ! k !, Newton's formula takes the form P&o + th) =z(;) W(xo>.k=O Further properties of factorial polynomials are
developed in the following exercises. Exercises 593 15.13 Exercises 1. Let Afn. 2. Let Af 0, show that Alcf(x) = (ah ' 1)"~~.(b) Ifg(x) = (1 + c@@, where u > 0, show that A"g(x) = @g(x).(c) Show that the polynomial P,, of degree n that takes: the values P,(k) = (1 + a)" for k =0,1,2)...) n is given by P,(x) = -$$x?k=O 4. Let xtn) be the factorial nth
power of x. Since xt")IS a polynomial in x of degree n with the value0 when x = 0, we can write X(n' =2 sk,nxk *k=l The numbers Sk8, are called Stirling numbers of thefirst kind. From the definition of x(") it isclear that S,., = 1 for n 2 0.(a) Show that S, , It = -n(n - 1)/2 and that S, n = (-l)+l(n - l)! for n 2 1.(b) Prove that Sk,i+1 = Sk-l,n - nSk lE. Use
this relation to verify the entries in Table 15.2,a table of Stirling numbers of the first'kind, and construct the next three rows of the table. TABLE 15.2 n S1.n S2.n S3.n S4.5% S 5.m s h6.n 1 12 - 1 13 2 - 3 14 - 6 1 1 - 6 :15 24 - 5 0 3 5 - 1 0 16 -120 274 -225 85 - 1 5 17 720 -1764 1624 -73s 1 7 5 -21 1 (c) Express the polynomial xt4) + 3~'~) + 2x(l) + 1
as a. linear combination of powers of x.5. (a) Prove that x = $1)) x2 = x(1' + x(2' , x3 =: $1) + 3x'2' + x(3), 594 Introduction to numerical analysis and that, in general, xn =c' A"f(o) (k)xx ' k=l where f(x) = xn and Af(x) =f(x + 1) -f(x). The numbers Tk,n = AkCf(0)/k! are calledStirling numbers of the second kind.(b) Prove that Akxnfl = (x + k) Akxn + k
Ak-lx" and use this to deduce that Tk,,+l = Tk-l,n + kT,,, .(c) Use the recursion formula in part (b) to verify the entries in Table 15.3, a table of Stirlingnumbers of the second kind, and construct the next three rows of the table. TABLE 15.3 n Tl.,, Tzen T3.n T4.n Tb. n T6.n T,,n 1 1 12 1 1 13 1 3 14 1 7 6 15 1 15 25 10 16 1 31 90 65 15 17 1 63 3 0 1 350 140
21 1 (d) Express the polynomial x4 + 3x3 + 2x - 1 as a linear combination of factorial poly-nomials. 6. (a) If p is a positive integer and if a and b are integers with a < b , prove that b l ck(P) = b(P+l)_ _&'+l) k=ap+l - This formula is analogous to the integration formula for Jl xp dx. It should be noted, however,that the upper limit in the sum is b - 1, not
b.(b) Verify that k(k + 3) = 4k w + kc2). The use part (a) to show that nc k(k + 3) = 4 (n +21'(2) + (n +31P = n(n + 1;'" + 5). k=l (c) If f (k) is a polynomial in k of degree r, prove that kzlf(k) is a polynomial in n of degree r + 1. A minimum problem relative to the max norm 595 7. Use the method suggested in Exercise 6 to express each of the following
sums as a polynomialin n. (a) i (4P -t 7k + 6).k=l Cc) kzlk(k + l)(k -t -2). (b) 3 k2(k + 1). (d) 5 k4.k=l k=l 8. Let A denote the linear operator defined by the equation A(f) = a, A"f + al Anplf + . . . + unpIL Af + a,f, where a,,, a,, . . .
, a, are constants. This is called a constant coeficient difference operator.It is analogous to the constant coefficient derivative operator described in Section 6.7. With eachsuch A we can associate the characteristic polynomial pA defined by PA(r) = a# + ulrn -l + . . . + a, ,r + a,. Conversely, with every polynomialp we can associate an operator A
having this polynomial as itscharacteristic polynomial. If A and B are constant coefficient difference operators and if A is areal number, define A + B, AB, and iA by the same formulas used in Section 6.7 for derivativeoperators. Then prove that Theorem 6.6 is valid for constant coefficient difference operators. 15.14 A minimum problem relative to the
max norm We consider a problem that arises naturally from the theolry of polynomial interpolation.In Theorem 15.3 we derived the error formula (15.27) .f(x> - P(x) = *f(n+lyc)) whereA(x) = (x - x0)(x - x1). * * (x - xJ. Here P is the unique polynomial of degree < n that agrees with f at n + 1 distinct points x0, Xl, . . * ? x, in [a, b].
The function f is assumed to have a derivative of order n + 1on [a, b], and c is an unknown point lying somewhere in [or, b]. To estimate the error in(15.27) we need bounds for the (n + 1)st derivative f (+l) and for the product A(x). SinceA is a polynomial, its absolute value has a maximum somewhere in the interval [a, b].This maximum will depend on
the choice of the points x0, X~ , . . . , x, , and it is natural totry to choose these points so the maximum will be as small as possible. We can denote this maximum by 1) A I/, where 11 A II is the max norm, given by IMII = max IA(x>l.aThe next five polynomials are T,(x) = 2x2 - 1 T3(x) = 4x3 -) 3X) T*(x) = 8x4 - 8x2 + 1, T5(x) = 1 6x5 - 20x3 + 5x, To(x)
=32xfi 48x4+ 18x2 -1 . The recursion formula shows that all the coefficients of T,(x) are integers; moreover,the coefficient of x" is 2n-1. The next theorem shows that T,(x) has exactly n first order zeros and that they all lie inthe interval [- 1, 11.
THEOREM 15.9. If n 2 1 the polynomial T,(x) has zeros at the n points X, = ~0s~~~ ' 'IT,2 n k = 0, 1,2, . .
. , n - 1. Hence T,(x) has the factorization !I -1 T,(x) = 27x - x&(x - x1) * * * (x - x, ,) = 2+l n x - cos(2k + 1)~ k=O 2n Proof. We use the formula T,(x) = cos no. Since cos n0 = 0 only if n0 is an oddmultiple of 7r/2, we have T,(x) = 0 for x in [- 1, l] only if n arccos x = (2k + l)n/2 forsome integer k. Therefore the zeros of T, in the interval [- 1, l] are to be
found among thenumbers (15.29) x,=cos2= t k=O,fl,:&2 ,....n 2' The values k = 0, 1, 2, .
.
. , n - 1 give n distinct zeros x0, x1, . . . , x, ~, all lying in theopen interval (-1, 1). Since a polynomial of degree n cannot have more than n zeros,these must be all the zeros of T,, . The remaining xk in (15.29) are repetitions of these n. THEOREM 15.10. In the interval [- 1, l] the extreme values of T,(x) are + 1 and - 1,taken alternately at the n + 1 points
(15.30) t, = cos !cJ) for k = 0,1,2,. .
. , n.n Proof. By Rolle's theorem, the relative maxima and minima of T, must occur betweensuccessive zeros; there are n - 1 such points in the open interval (-1, 1). From the cosineformula for T, we see that the extreme values, f 1, are taken at the n - 1 interior points 598 Introduction to numerical analysis $ Tz 64 $ TX (4 $ T, (4 $ T.5 (x) FIGURE 15.2
Graphs of Chebyshev polynomials over the interval [- 1, 11.
cos(kn/n), k= 1, 2, . . .
, n - 1, and also at the two endpoints x = 1 and x = - 1.Therefore in the closed interval [- 1, l] the extreme values + 1 and - 1 are taken alter-natelyatthen+lpointst,,t,,...,t,givenbyt,=cos(kn/n)fork=0,1,2,...,n. Figure 15.2 shows the graphs of Tz , . . . , T5 over the interval [- 1, 11. 15.16 A minimal property of Chebyshev polynomials We return now to the
problem of finding a polynomial of a specified degree for which themax norm is as small as possible. The problem is solved by the following theorem. THEOREM 15.11. h?tp,(X) = Xn -i - -* * be any polynomial of degree n 2 1 with leadingcoeficient 1, and let lIPnIl = max IP,(x>l . -11zAssume, then, thatp, satisfies (15.32) and consider the difference 44 =
~k4 - P,(X).
At the points t, given by (15.30) we have r(tJ = (2 - PJfk) = (-1)"[& - (VP?&) 1 * Application to the error formula for interpolation 599 Because of (15.32) the factor in square brackets is positive. Therefore r(tJ has alternatingsignsatthen+ lpointst,,t, ,..., t,. Since r is continuous it must vanish at least oncebetween consecutive sign changes.
Therefore r has at least n distinct zeros. But since r is apolynomial of degree 5 n - 1, this means that r is identically zero. Therefore P, = Fm,,so IIP, I I = IlrT,II = 1/2'+1, contradicting (15.32). Thiis proves that we must have llpn 11 2l/2" 1 = I(Fn 11 . Although Theorem 15.11 refers to the interval [- 1, l] and to a polynomial with leadingcoefficient 1, it
can be used to deduce a corresponding result for an arbitrary interval [a, b]and an arbitrary polynomial. THEOREM 15.12. Let qn(x) = c,xn + * * * be any polynomial of degree n 2 1, and let Then we have the inequality (15.33) Moreover, equality holds in (15.33) if (b - a)" T4n(x> = cn 22n -1 n (-- Proof. Consider the transformation t=2x-a-bbb-a This
maps the interval a 5 x 5 b in a one-to-one fashion onto the interval - 1 2 t 5 1.Since b-a2 t+b+ax=- -2 we have hence P + terms of lower degree, p,(t) 7 where p,(t) is a polynomial in t of degree n with leading coefficient 1. Applying Theorem15.11 top% we obtain Theorem 15.12. 15.17 Application to the error formula for interpolation We return now to
the error formula (15.27) for polynomial interpolation. If we choosethe interpolation points x0, x1, .
. . , x, to be the n + 1 zeros of the Chebyshev polynomial 600 Introduction to numerical analysis r, , we can write (15.27) in the form f(x) - P(x) = 2n2y)! f'"+l'(c). The points x0, x1,. . . , x, all lie in the open interval (-1, 1) and are given by XkECOS (S;) for k=0,1,2 ,..., n. If x is in the interval [- 1, l] we have Ir,+,(x)l 5 1 and the error is estimated by
theinequality If(x) - P(x)1 I 2"(n1+ 1), P+%)l. If the interpolation takes place in an interval [a, b] with the points b - ax +b+ay,=- - 2k 2 as interpolation points, the product ml = (x - YONX - Yl) * * - (x - y,> satisfies the inequality IA(x)/ < (b - a)n+1/22n+1 for all x in [a, b]. The correspondingestimate forf(x) - P(x) is If(x) - W)l I 2!f+ina;;), 1 f (n+l'(C)I .
15.18 Exercises In this set of exercises T, denotes the Chebyshev polynomial of degree n. 1. Prove that T,(-x) = (-l)nT,(x). This shows that T, is an even function when n is evenand an odd function when n is odd. 2. (a) Prove that in the open interval (-1, 1) the derivative TA is given by the formula n sin n6T;(x) = sin 9 where 8 = arccos x. (b) Compute
T;(l) and TL(-1). s 23. Prove that T, ,(x) - T, ,(O) Ii f n22. 0 n - l 4.
(a) Prove that 2T,(x)T,(x) = T,+,(x) + T,, n(x).(b) Prove that Tm,(x) = T,[T,(x)] = T,[T,(x)] . 5. If x = cos 0, prove that sin 6 sin n0 is a polynomial in x, and determine its degree.6. The Chebyshev polynomial T,, satisfies the differential equation (1 - x2)y" - xy' + n2y = 0 Exercises 601 7.
8 . 9. 10. over the entire real axis. Prove this by each of the following methods:(a) Differentiate the relation T(x) sin 19 = n sin n0 obtained in Exercise 2(a).(b) Introduce the change of variable x = cos 8 in the differential equation d2 (cos ne> do2 = 9cos no. Determine, in terms of Chebyshev polynomials, a polynomial Q(x) of degree Prove each of the
followingstatements :(a) If b - a < 4, for every E > 0 there exists a primary polynomial P with lIPI\ < E.(b) If for every E > 0 there exists a primary polynomial P with lIPI/ < E, then b - a < 4. In other words, primary polynomials with arbitrarily small norm exist if, and only if, theinterval [a, b] has length less than 4.The Chebyshev polynomials satisfy the
following orthogonality relations: 0 i f n#m, s l Tn(x)TmW-l Ji ,z dx = I 1 i f n=m=O, 2i f n=m>O. Prove this by each of the following methods: (a) From the differential equation in Exercise 6 deduce that T,(x) 2 Jl - ~2 T;(X)( ) TTJx)Tm(x)+ n2 -+==- = 0.
dl -x2 Write a corresponding formula with n and m interchanged, subtract the two equations, andintegrate from -1 to 1.(b) Use the orthogonality relations 0 if n #m, n >0, m >0, s Icos m0cosn8d0 = i vi f n=m=O, 0 x 2i f n=m>O, and introduce the change of variable x = cos 0.11. Prove that for -1 < x < 1 we have T,(x) 2%! dn Jr2= (-1)" W! dx" (1 -
x2)* 1'2. 12. Lety,,y,, . . . ,y, be n real numbers, and let (2k - 1)~Xk = cos 2nfor k=1,2 ,..., n. 602 Introduction to numerical analysis Let P be the polynomial of degree In - 1 that takes the value ylc at xk for 1 5 k 5 n. If x isnot one of the xk show that P(x) = ; 2 (-l)'c lyk Jl - xi s .k=l k 13. Let P be a polynomial of degree In - 1 such that J1 -x2 IPcd I; 1 for
-1 < x 5 1. Prove that /P/l I n , where llP/l is the maximum of IPI on the interval [- 1 , 11.
[Hinr: Use Exercise 12. Consider three cases: x1 < x < 1 ; - 1 < x < x, ;x, 5; x 5 Xl ,* in the first two use Exercise 15(a) of Section 15.9. In the thirdcase note that Jl - x2 2 sin (7r/2n) > l/n.] In Exercises 14 through 18, V,(x) = Ti+,(x)/(n + 1) for n = 0, 1,2, ... 14. (a) Prove that V,(x) = 2xU, r(x) - U,,(x) for n > 2.(b) Determine the explicit form of the
polynomials U,, (/1, . . . , U,.(c) ProvethatIU,(x)l Prove that Ji z U,(x) = (-l)n :,I: +' $ -$ (1 - Xa)n+l/z. 18. Lety,,y2,. . . , y, be n real numbers and let krr -xk = 'OS n + 1for k=1,2 ,..., n. Let P be the polynomial of degree I;n - 1 that takes the value yk at xk for 1 5 k 5 n. If x isnot one of the xk show that P(x) = j& 2 (l)k I(1 Udx) -$yk - * kc1x - xk 15.19
Approximate integration. The trapezoidal rule Many problems in both pure and applied mathematics lead to new functions whoseproperties have not been studied or whose values have not been tabulated. To satisfycertain practical needs of applied science it often becomes necessary to obtain quantitative Approximate integration. The trapezoidal rule
603 a a+h u + 2h . u . a+nh=b FIGURE 15.3 The trapezoidal rule obtained by piecewise linear interpolation.
information about such functions, either in graphical or numerical form. Many of thesefunctions occur as integrals of the type where the integrandfis given by an explicit analytic formula or is known in part by tabulardata. The remainder of this chapter describes some of the most elementary methods forfinding numerical approximations to such
integrals. The basic idea is very simple. Weapproximate the integrand f by another function P whose integral is easily computed, andthen we use the integral of P as an approximation to the integral off. Iffis nonnegative the integral Ji f(x) dx represents the area of the ordinate set off over[a, b]. This geometric interpretation of the integral immediately
suggests certain proce-dures for approximate integration. Figure 15.3 shows an example of a function f withknown values at n + 1 equally spaced points a, a + h , a + 215, . . . , a + nh = b, whereh = (b - a)/n. Letx,=a+kh. Foreachk=0,1,2,...,n lthegraphoffoverthe interval [x~ , xkfl] has been approximated by a linear function that agrees with f at
theendpoints x, and xk+r. Let P denote the corresponding piecewise linear interpolatingfunction defined over the full interval [a, b]. Then we have (15.34) jyx) = %+1 - xh f(Xk) + yvcxk+J Integrating over the interval [xk , xk+J we find that s %+Ip(x) dx = h fcXk) + f(Xk+l) - *lc 2 * When f is positive this is the area of the trapezoid determined by the
graph of P over Lxk, xk+ll. Th e formula holds, of course, even iff is not positive everywhere. Adding the 604 Introduction to numerical analysis integrals over all subintervals [A~, s,+J we obtain (15.35) P (x) d x = ;h '2 D (Xk) + f(%+dlk=O =k(f(a)+gf(a+kh)+f(b)). To use this sum as an approximation to the integral Jif(x) dx we need an estimate
forthe error, j"zf(x) dx - ji P(x) dx. Ifs has a continuous second derivative on [a, b] thiserror is given by the following the&em. THEOREM 15.13. TRAPEZOIDAL RULE. Assume f has a continuous second derivative f"on [a, b]. Zf n is a positive integer, let h = (b - a)/n . Then we have (15.36) /'f(x) dx = e(f(a) + 2zf(a + kh) + f(b)) - (*f"(c).n k=l for some c in
[a, b]. Note: Equation (15.36) is known as the trapezoidalrule, The term -f"(c)@ - a)3/12n2represents the error in approximating Ji f(x) dx by Ji P(x) dx. Once the maximum valueoff" on [a, b] is known we can approximate the integral off to any desired degree ofaccuracy by taking n sufficiently large. Note that no knowledge of the interpolatingfunction
Pis required to use this formula. It is only necessary to know the values off at thepoints a, u + h, . . . , u + nh, and to have an estimate for If"(c)j. ProoJ Let P be the interpolating function given by (15.34), where xK = a + kh . In eachsubinterval [xk, xktl] we apply the error estimate for linear interpolation given by Theorem15.3 and we find (15.37) f(x) -
P(x) = (x - Xk)(X - x,+$+ for some c, in (xk, xk+l). Let M2 and rn, denote the maximum and minimum, respectively,off" on [a, b], and let B(x) = (x - Xk)(Xk+l - x)/2. Then B(x) 2 0 in the interval [xk, x,,,], and from (15.37) we obtain the inequalities m2Wx) I; P(x) - f(x) S M2W) in this interval. Integrating, we have (15.38) m2 Jr+'B(x) dx 5 j+;+" [P(x) - f(x)]
dx < M, j z+' B(x) dx . The integral of B is given by B(x) dx = 1s W h x)dx=Is h 2(x - xk)(xktl - t=h 2 0t(h - t) dt = $. Simpson's rule Therefore the inequalities (15.38) give us 605 m < 12 xk+h2- sh3 z,t[P(x) - fW1 dx I M, - Adding these inequalities for k = 0, 1, 2, . . . , n - 1 and dividing by n, we obtain m <122 s b[P(x)-f(x)]dxIM2.nh3 a Since the functionf"
is continuous on [a, b], it assumes every value between its minimumm2 and its maximum M, somewhere in [a, b]. In particular, we have f"(c) = -$ Jb [P(x) - ./(x)1 dxafor some c in [a, b]. In other words, sI/(x) dx = 1'a P(x) dx - gf"(c).
Using (15.35) and the relation h = (b - a)/n we obtain (15.36). To derive the trapezoidal rule we used a linear polynomial to interpolate between eachadjacent pair of values off. More accurate formulas can be obtained by interpolating withpolynomials of higher degree. In the next section we consider an important special casethat is remarkable for its
simplicity and accuracy. 15.20 Simpson's rule The solid curve in Figure 15.4 is the graph of a function f over an interval [a, b]. Themid-point of the interval, (a + b)/2, is denoted by m. The dotted curve is the graph of aquadratic polynomial P that agrees with f at the three points a, m, and 6. If we use theintegral 1: P(x) dx as an approximation to jlf(~)
dx we are led to an approximate inte-gration formula known as Simpson's rule. Instead of determining P explicitly, we introduce a linear transformation that carries theinterval [a, b] onto the interval [0,2]. If we write x - at= -3 or x = a + (m - a)t,m - u we see that t takes the values 0, 1, 2 when x takes the values a, m, b. Now let p)(t) = P[u + (m - u)t].
Then v(t) is a quadratic polynomial in t that takes the values P(u), P(m), P(b) at the pointst = 0, 1, 2, respectively. Also, we have lpl(fW= c P[u + (m - a)t] dt := & bP(x) dx ;s 606 Introduction to numerical analysis . a m b FIGURE 15.4 Interpolation by a quadratic polynomial P. hence f b P(x) dx = (m - a)s 2p(t) dt = -b - a 2 a 0 s2 0v(t) dt. Now we use
Newton's interpolation formula to construct 9. We have A2do)$40 = P(O) + r AT(O) + t(t - 1) - 2! ' where A&t) = y(t + 1) - y(t). Integrating from 0 to 2 we obtain io2 v,(t) dt = 2940) + 2 AM(O) + ; A"q(O). Since AF'(O) = ~(1) - P(O) and A2v(0) = ~(2) - 29?(l) + p(O), the integral is equal to . o2 P(t) dt = %dO) + 4~(1) + q(2)]= $[P(a> + 4P(m) + P(b)].1
Using (15.39) and the fact that P agrees with f at a, m, b, we obtain (15.40) s b b - aP(x) dx = -6 [f(a) + V(m) + f(b)].a Therefore, we may write sabf(x) dx = + [f(a) + 4f(m> + f(b)] + R , where R = jif(x) dx - J; P(X) dx. Simpson's rule 601 IfSis a quadratic polynomial, P is identical tofand the error R is zero. It is a remark-able fact that we also have R = 0
when f is a cubic polynomial. To prove this property weuse the error estimate for Lagrange interpolation given by Theorem 15.3, and we write (15.41) f(x) - P(x) = (x - a)(x - m)(x -- b)'- , where c E (a, b).
When f is a cubic polynomial the third derivative f"' is constant, sayf"'(x) = C, and the foregoing formula becomes f(x) - P(x) = 2 (x - a)(x - m)(x - b) = : (t + h)t(t - h), where t = x - m and h = (b - a)/2. Therefore R =s b [f(x) - P(x)] dx = : -;,(t3 -- h't) dt = 0,a s since the last integrand is an odd function. This property is illustrated in Figure 15.5. Thedotted
curve is the graph of a cubic polynomialf that agrees with P at a, m, b. In thiscase R = Jz [f(x) - P(x)] dx = A, - A,, where A, and A, are the areas of the two shadedregions.
Since R = 0 the two regions have equal areas. We have just seen that Equation (15.40) is valid if P is a polynomial of degree 5 3 thatagrees with f at a, m, and 6. By choosing this polynomial carefully we can considerablyimprove the error estimate in (15.41). We have already imposed three conditions on P,namely, P(a) =f(u), P(m) =f(m), P(b) =f(b) .
Now we impose a fourth condition,P'(m) =f'(m). This will give P andfthe same slope at (m,f(m)), and we can hope thatthis will improve the approximation off by P throughout [a, b]. FIGURE 15.5 The two shaded regions have equal areas, for every cubic interpolatingpolynomial. 608 Introduction to numerical analysis To show that such a P can always
be chosen, we let Q be the quadratic polynomial thatagrees with f at a, m, b, and we let P(x) = Q(x) + A(x - a)(x - m)(x - b), where A is a constant to be determined. For any choice of A, this cubic polynomial Pagrees with Q and hence with f at a, m, b. Now we choose A to make P'(m) = f '(m).Differentiating the formula for P(x) and putting x = m we
obtain P'(m) = Q'(m) + A(m - a)(m - b), Therefore if we take A = [f'(m) - Q'(m)]/[(m - a)(m - b)] we also satisfy the conditionP'(m) = f'(m). Next we show that for this choice of P we have (15.42) f(x) - P(x) = (x - a)(x - m)"(x - b)fq for some z in (a, b), provided that the fourth derivative ft4) exists in [a, b]. To prove(15.42) we argue as in the proof of
Theorem 15.3. First we note that (15.42) is triviallysatisfied for any choice of z if x = a, m, or 6.
Therefore, assume x # a, x # m , x # b ,keep x fixed, and introduce a new function F defined on [a, b] by the equation whereF(t) = AWW) - P(t)1 - AWfW - P(x)l, A(t) = (t - a)(t - m)"(t - b). Note that F(t) = 0 for t = a, ~2, b, and x. By Rolle's theorem, F'(t) vanishes in each ofthe three open intervals determined by these four points. In addition, F'(m) = 0
becauseA'(m) = 0 and f'(m) = P'(m). Therefore F'(t) = 0 for at least four distinct points in(a, 6). By Rolle's theorem F"(t) = 0 for at least three points, F"'(t) = 0 for at least twopoints, and Fc4)(t) = 0 for at least one point, say for t = z. From the definition of F wefind Fc4)(t) = A(x)[ft4)(t) - P'"'(t)] - At4)(t)[f(x) - P(x)] = A(x)fc4)(t) - 4! [f(x) - P(x)]. When we
substitute t = z in this equation we obtain (15.42).Now it is a simple matter to prove Simpson's rule in the following form. THEOREM 15.14. SIMPSON'S RULE. Assume f has a continuous fourth derivative on[a, b], and let m = (a + b)/2.
Then \+'e have (15.43s)(x) dx = y [f(a) + 4f(m) +f(b>l - 'of for some c in [a, b]. Simpson's rule 609 Proof. Let M4 and m4 denote, respectively, the maximum and minimum values off(") on [a, b], and let B(x) = -(x - a)(x - m)"(x - 6)/4!. Since B(x) 2 0 for each xin [a, b], Equation (15.42) leads to the inequalities Integrating, we findm4 Nx) I P(x) -f(x) I W
B(x). (15.44) m4 j," B(x) dx 5 jab [P(x) - f(x)] dx 5 MM, jb" B(x) dx .
To evaluate the integral sl B(x) dx we let h = (b - a)/2 and we have s b aB(x) dx = - $ f'(x - a)(x - m)"(x - b) dx = - 1 . a .s4r :(t + h)t'(t - h) dt 2 h= -s4! 0 tz(t2 _ h2) dt = I 4h5 = (b -4! 15 2880 ' Therefore the inequalities in (15.44) give us 2880 m4 s (b - a)5 sb P'(x) - f(x)1 dx I Ma. a But sincefc4) is continuous on [a, b], it assumes every value between its
minimum m4 andits maximum M4 somewhere in [a, b]. Therefore f""(c) = s 1' V'(x) - f(x)1 dxn for some c in [a, b]. Since ji P(x) dx = Q(b - a)[f(a) + Y(m) +f(b)] , this equationgives us (15.43). Simpson's rule is of special interest because its accuracy is greater than might be expectedfrom a knowledge of the functionfat only three points. If the values
offare known at anodd number of equally spaced points, say at a, a + h, . . . , u + 2nh, it is usually simplerto apply Simpson's rule successively to each of the intervals [a, a + 2h], [a + 2h, a + 4/z], . . .) rather than to use an interpolating polynomial of degree <2n over the full interval[a, a + 2nh].
Applying Simpson's rule in this manner, we obtain the following extensionof Theorem 15.14. THEOREM 15.15. EXTENDED SIMPSON'S RULE. Assume f has a continuous fourth derivativein [a, b].
Let h = (b - u)/(2n) and let fk =f(u + kh) for k = 1, 2, . . .
, 2n - 1. Thenwe have for some E in [a, 61. The proof of this theorem is requested in Exercise 9 of the next section. 610 Introduction to numerical analysis 15.21 Exercises 1. (a) Apply the trapezoidal rule with n = 10 to estimate the value of the integral s 2 dx log2= 1X' Obtain upper and lower bounds for the error. (See Exercise 10(b) to compare the
accuracywith that obtained from Simpson's rule.)(b) What is the smallest value of n that would ensure six place accuracy in the calculation oflog 2 by this method?
2. (a) Show that there is a positive number c in the interval [0, l] such that the formula s;,f(x) dx =f +./X c) is exact for all polynomials of degree 13.(b) Generalize the result of part (a) for an arbitrary interval. That is, show that constants c1and c2 exist in [a, b] such that the formula sabf(~) dx = 7 [fProve that 12c b'+l -(f+1 x;Kc(n> =r + l f o r
r=O,l,..., n.k=O This is a system of IZ + 1 linear equations that can be used to determine the weights. It canbe shown that this system always has a unique solution. It can also be shown that for a suitablechoice of interpolation points it is possible to make all the weights equal. When the weightsare all equal the integration formula is called a
Chebyshev integration formula. Exercises 2and 3 give examples of Chebyshev integration formulas.
The next exercise shows that for aproper choice of interpolation points the resulting integration formula is exact for all poly-nomials of degree <2n + 1 . 7. In this exercise you may use properties of the Legendre polynomials stated in Sections 6.19and6.20. Letx,,x, ,..., x, be the zeros of the Legendre polynomial P,+l(x). These zerosare distinct and they
all lie in the interval [-1, 11. Letf(x) be any polynomial in x of degree12n + 1 . Dividef(x) by P,+l(x) and write f(x) = f'n+,WQW - t R(x), where the polynomials Q and R have degree In.(a) Show that the polynomial R agrees with f at the zeros of Pn+l and that s1, f(x) dx = j' R(x) dx. 1 (b) Show that n + 1 weights IVo(n), . . . , W,(n) exist (independent off)
such that s;,f(x)dx =kzo w,(n)f(xk). This gives an integration formula with n + 1 interpolation points that is exact for all poly-nomials of degree <2n + 1.(c) Take n = 2 and show that the formula in part (b) becomes sf,f(x)dx =$f(- &> ++f(O) +$f(& This is exact for all polynomials of degree 15.(d) Introduce a suitable linear transformation and rewrite
the formula in part (c) for anarbitrary interval [a, b]. 8. This exercise describes a method of Peano for deriving the error formula in Simpson's rule.(a) Use integration by parts repeatedly to deduce the relation J u(t>d"(t) dt = u(t)/(t) - u'(tb'(t) + u"(t>u(t) - j&t) dt, whereg(r) = u"'(t)u(t).
612 Introduction to numerical analysis (b) Assume p has a continuous fourth derivative in the interval [- 1, 11. Take u(t) = t(1 - t)2/6, u(f) = P(t) + d -t), and use part (a) to show that j:, p(t) dt = i[d -1) + 4p(O) + p(l)] - jig(t)&.
Then show that Jtg(t) dt = p(*)(c)/90 for some c in [- 1, 11.(c) Introduce a suitable linear transformation to deduce Theorem 15.14 from the result ofpart (b). 9. (a) Let a,, a2, . .
. , a, be nonnegative numbers whose sum is 1. Assume IJJ is continuous onan interval [a, b]. If cl, c2, . . .
, c, are any n points in [a, b] (not necessarily distinct), provethat there is at least one point c in [a, b] such that [Hint: Let M and m denote the maximum and minimum of q~ on [a, b] and usethe inequality m I &ck) I M.] (b) Use part (a) and Theorem 15.14 to derive the extended form of Simpson's rule given inTheorem 15.15. 10. Compute log 2 from
the formula log 2 = J," x l dx by using the extension of Simpson's rulewith (a) n = 2 ; (b) n = 5. Give upper and lower bounds for the error in each case. 11. (a) Let p(t) be a polynomial in t of degree 53. Express p)(t) by Newton's interpolationformula and integrate to deduce the formula 3 v(f)dt = $[P(O) + 3N) + 3&2) + p(3)]. (b) Let P be the
interpolation polynomial of degree 53 that agrees with f at the points a,a +h,a +2h,a +3h,whereh >O.
Usepart(a)toprovethat s a ! 3h aP(x) dx = ; [f (a> + 3f (a + h) + 3f(a + 2h) + f(a + 3h)3. (c) Assume f has a continuous fourth derivative in [a, b], and let h = (b - a)/3. Prove that saiif(x)dx = 7 [f(a) + 3f(a + h) + 3f(a + 2h) +f(b)] - for some c in [a, b]. This approximation integration formula is called Cotes' rule.(d) Use Cotes' rule to compute log 2 = Jf x l
dx and give upper and lower bounds for theerror. 12. (a) Use the vector equation r(t) = a sin t i + b cos t j, where 0 < b < a, to show that thecircumference L of an ellipse is given by the integral L =4a j;" ,/mdt, where k = da2 - b2/a. The Euler summation formula 613 (b) Show that Simpson's rule gives the formula for some c in [0, p/2], wheref(t) = Jl -
k2 sin2 1. 15.22 The Euler summation formula Let n be a positive integer. When the trapezoidal formula (Theorem 15.13) is applied tothe interval [0, n] it becomes sb'j(x) dx = f f(k) + *(f(n) -S(O)) -'F b=O for some c in [0, n]. If f is a quadratic polynomial thenf" is constant and hence f "(c) =f"(0). In this case the formula can be rewritten in the form
(15.45) It is exact when f is any polynomial of degree 12.Euler discovered a remarkable extension of this formula that is exact for any function with a continuous first derivative. It can be used to approximate integrals by sums, 'or, asis more often the case, to evaluate or estimate sums in terms of integrals. For this reasonit is usually referred to as a
"summation" formula rat'her than an integration formula.
Itcan be stated as follows. THEOREM 15.16. EULER'S SUMMATION FORMULA. Assume f has a continuous derivative on[0, n]. Then we have (15.46) dx + '(') ; f(n) + j+" (x - [x] - Qf'(x) dx,0 where [x] denotes the greatest integer Ix. Proof. Integration by parts gives us (15.47) s,"(x - &)f'(x) dx = (n - t)f(n) + &f(O) -/: f(x) dx. Now we consider the
integral j'," [x]f '(x) dx and write it as a sum of integrals in each ofwhich [x] has a fixed value. Thus, we have /," [x]f'(x) dx =;z; j ;+l [x]f'(x) dx =z;r j -Tdlf'(x) dx n-1 n l 11 l = .Xor(f(r + 1) - f(r)> =rz rf(r + 1) -rz rf(r> n-1 n 1 n 1 = - zof(r + 1) +TzCr + l>f(r + 1) -Tzyf(r) = -kzl.f@) + nf(n> = -k$o f(k) +fW + nf(n>. 614 Introduction to numerical analysis
Subtracting this from Equation (15.47) we obtain s"(x - [x] - +)f'(x) dx =$ f(k) -f(o) ; j(n) - fa'f(x) dx , 0k=O which is equivalent to (15.46). The last integral on the right of (15.46) can be written as lb" (x - [xl - f>f'(x> dx = sb" dx)f'(x) dx, where IJJ~ is the function defined by VI(X) =i x - [x] - + if x is not an integer, 0 if x is an integer. The graph of (pl is
shown in Figure 15.6(a). We note that vr(x + 1) = QI~(x), which meansthat e)l is periodic with period 1.
Also, if Of"(x) dx = -Jon gdx).f"(x) dx , since p&n) = ~~(0) = 0.
Using this in (15.46) we obtain (15.48). Note: Although Theorems 15.16 and 15.17 refer to the interval [0, n], both formulasare valid when 0 is replaced throughout by 1 or by any positive integer < n. To illustrate the use of Euler's summation formula we derive the following formula forl o g n ! . THEOREM 15.18. For any positive integer n we have
(15.49) log n! = (n + $> log n - n + C + E(n), where 0 < E(n) < 1/(8n) and C = 1 + j," t "p2(t) dt . Proof. We take f(x) = log x and apply Theorem 15.17 to the interval [l, n]. Thisgives us log x dx + 4 log n +sI n!??$d,x. Using the relation j log t dt = t log t - t we rewrite this equation as follows, (15.50) logn! =(n + $)logn - n + 1 +r"%&&. .1 t2 Since
Iy2(t)l < + the improper integral s; te2y2(t) dt converges absolutely and we canwrite s " =1 t2 s "pj20df 1 t2 s "j7"(t>dt. 12 t2Therefore Equation (15.50) becomes logn!=(n+~)logn n+C-sm 92(0& , 12 t2 616._ Introduction to numerical analysis where C = 1 + j," t "y2(t) dt . Since we have -$ < y2(t) < 0 except when t is an integeror
half an integer, we obtain o< f "%@dt< ..$It t2 This proves (15.49), with E(n) = -jz t 2p,(t) dt. From Theorem 15.18 we can derive Stirling's formula for estimating n!.
THEOREM 15.19.
STIRLING'S FORMULA. If n is a positive integer we have Proof. Using Equation (15.49) and the inequalities for E(n) we obtain exp((n+$)logn-n+C) 0 we obtain A4rr< We note that I, = 7~12 and I1 = 1. For 0 5 t 5 n/2we have 0 < sin t 5 1 ; hence 0 < sinn+l t 5 sin?' t. This shows that the sequence {I,}is monotonic decreasing. Therefore we can write
(15.53)1 I2J2, 1 Now we shall evaluate each member of this inequality; this will lead at once to Wallis'inequality. Integration of the identity 2 (cos t sin"+l t) = (n + 1) sin" t - (n + 2) sinn+2 t over the interval [0,77/2] gives us or (15.54) 0 = (n + 1>4 - (n + 2)1n+2, Using this recursion formula with n replaced by 2k - 2 we find 2k(2k - 1) (2k)2 '
Multiplying these equations for k = 1, 2, . . . , n, we find fi 2k(2k - 1) (2n)!= k=l (2k)2 22"(n !)" * The product on the left telescopes to Izn/Io. Since I, = 7r/2 we obtain (15.55) I,, = o! .4.22n(n!)2 2 618 Introduction to numerical analysis In a similar way we apply the recursion formula (15.54) with n replaced by 2k - 1 andmultiply the resulting equations
for k = 1, 2, . . . , n to obtain l rn I,,+1 _(2kY n 22"(n !)" 1 77 1= . . k=l I,, , k=l 2k(2k + 1) = (2n + l)! 2n + 1 2 I,, . The product on the left telescopes to 12n+1/Z1 = Z2n+l, so we get (15.56) ~2J2n+l = =2(2n + 1) ' Since Z2n.t 1 = 2nZ2, ,/(2n + 1)) Equation (15.56) implies ~2J2n -1 = :. We use this in (15.53), together with the two relations (15.55)
and (15.56). Then we multiplyby .rr2/4 to obtain Wallis' inequality (15.52). 15.23 Exercises 1. Iffis a polynomial of degree < 2, show that Euler's summation formula (15.48) reduces to thetrapezoidal formula, as expressed in Equation (15.45). 2. Euler's constant C is defined by the limit formula c =JLn(2; -1ogn). (See Section 10.17 in Volume I.) Use
Euler's summation formula to prove that n 1c -=1ogn+c+; F,k k=l where 0 < E(n) 5 4. Also, show that C=l s * t - [t]1 t2dt. 3.
(a) Ifs > 0, s # 1, use Euler's summation formula to prove that n 1cnl-" jp - rs + C(s) + ss cc Gdt,kc1 n where C(s) = 1 + -& - ss1 * f$d;. Exercises 619 (b) If s > 1, show that C(S) = 4(s), where 5 is the Riemann zeta function defined for s > 1by the series m 15(s) = c jp * k=l The series for l(s) diverges for s I 1 . However, since the formula for C(S) in
part (a) is meaning ful for 0 < s < 1, it can be used to extend the definition of 4(s) to the open interval 0 < s < 1 .Thus, for s > 0 and s # 1 we have the formula C(s) = 1 + -& - ss m t - [t] 1tS+1dt- This is a theorem ifs > 1, and a definition if 0 < s < 1 . In Exercises 4 through 6, qZ is the function introduced in Section 15.22. 4. (a) Use Euler's summation
formula to prove that 11c log2 k = (n + +) log'n - 2 n l o g n +2n - 2 + 2k=l sn log x - 1 1R(X) 7 dx . (b) Use part (a) to deduce that for n > e we have 2 log" k = (n + a) log2 n - 2nlogn + 2n + A - E(n),k=l log nwhere A is a constant and 0 < E(n) < z . 5. (a) Use Euler's summation formula to prove that n logk 1c 1 logn- =210g2n + 2n - s "2logx - 3 k
X3~)a(4 dx . k=l1 (b) Use part (a) to deduce that for n > ex we have n logk 1c 1 logn- =slog2n +~n + A - E (n) , kk=l log nwhere A is a constant and 0 < E(n) < q . 6. (a) If n > 2 use Euler's summation formula to prove that Rc 1 k=2k l o g k 1 1= log (log n) + -+- - log (log 2) - s ?a%(X) 2 + 310gx + 2log2xdx.
2n log n 4log2 2 (x log x)3 620 Introduction to numerical analysis (b) Use part (a) to deduce that for n > 2 we have 12c 1 1 -=log(logn) +A + k=2klogk - - E(n),2n log n 1where A is a constant and 0 < E(n) < ~.
4n2 log n7. (a) If a > 0 andp > 0, use Euler's summation formula to prove that mc 00e-ak' = s~l(x)x~ le "" dx , k=O where I' is the gamma function.(b) Use part (a) to deduce that mc e-ak' = where 0 < 0 < 1.kc0 8. Deduce the following limit relations with the aid of Stirling's formula and/or Wallis' inequality. (c) lim (-l)n-4 ( )n n=-n-a, ;: s nl29. Let Z,
= sinn t dt, where n is a nonnegative integer. In Section 15.22 it was shown that the 0sequence {Zn} satisfies the recursion formula n+l Letf(n) =&/Gr I is the gamma function. (a) Use the functional equation I(s + 1) = sl?(s) to show that f(n + 2) = s2 f(n). (b) Use part (a) to deduce that Suggested references 6 2 1 SUGGESTED REFERENCES This
small list contains only a few books suggested for further reading on the general principlesof numerical analysis. All these books contain further references to works of a more special nature.The list of tables given in Todd's survey (Reference 9 below) is especially recommended. B o o k s1. A.
D. Booth, Numerical Methods, Academic Press, New York, 1958; 3rd ed., Plenum Press, New York, 1966.2. P. J. Davis, Interpolation and Approximation, Blaisdell, Waltham, Mass., 1963.3. D. R. Hartree, Numerical Analysis, Oxford Univ. Press (Clarendon), London and New York, 1958.4. F. B. Hildebrand, Zntroduction to Numerical Analysis, McGraw-
Hill,'New York, 1956.5. A. S. Householder, Principles of Numerical Analysis, McGraw-Hill, New York, 1953.6. W. E. Milne, Numerical Calculus, Princeton Univ. Press, Princeton, N.J., 1950.7. J. B. Scarborough, Numerical Mathematical Analysis, Johns Hopkins Press, Baltimore, Md., 1958, 6th ed., 1966.8. J. Todd, Introduction to the Constructive
Theory of Functions, Academic Press, New York, 1963.9. J. Todd (ed.), Survey of Numerical Analysis, McGraw-Hill, New York, 1962. Tables10. L. J. Comrie (ed.), Chambers' Six-figure Muthematical Tables, W. & R. Chambers, London and Edinburgh, 1949.11. L. J. Comrie, "Interpolation and Allied Tables," 2nd rev. reprint from Nautical Almanac for
1937, H.M. Stationery Office, London, 1948.12. A. J.
Fletcher, J. C.
P. Miller, and L.
Rosenhead, Index of Mathematical Tables, McGraw- Hill, New York, 1946.13. Tables of Lagrangian Interpolation Coefficients, Natl.
Bur. Standards Columbia Press Series, Vol. 4., Columbia Univ. Press, New York, 1944. ANSWERS TO EXERCISES 1.5 Exercises (page 7) 1.
Yes2 . Yes3 . Yes4 . Yes5 . No6 . Yes7. Yes 31. (a) No 8 .9 . 10.11.12.13. (b) ;:* Yes 15. YesYes 16. YesYes 17. YesNo 18. YesYes 19. YesYes 20. YesNo 21. Yes Cc) No (4 No 1.10 Exercises (page 13) 1. Yes; 2 5. Yes; 1 9 . Yes; 12 . Yes; 2 6 . No 10. Yes; 13. Yes; 2 7 . No 11.
Yes; n4 . Yes; 2 8 . No 12. Yes; n 17.
Yes; dim = 1 + &PZ if n is even, !&z + 1) if n is odd18. Yes; dim = & if n iseven, i(n + 1) if n is odd19. Yes; k + 1 22. Yes23. No24.
Yes25. No26. Yes27, Yes28. Yes 13. Yes; n14. Yes; n15. Yes; n16. Yes; n 20. No21. (a) dim = 3 (b) dim = 3 (c) dim = 2 (d) dim = 223.
(a) If a # 0 and b # 0, set is independent, dim = 3 ; if one of a or b is zero, set is de- pendent, dim = 2 (b) Independent, dim = 2 (c) If a # 0, independent, dim = 3 ;if a = 0, dependent, dim = 2 (d) Independent; dim = 3 (e) Dependent; dim = 2(f) Independent; dim = 2 (g) Independent; dim = 2 (h) Dependent; dim = 2(i) Independent; dim = 2 (j)
Independent; dim = 2 1.13 Exercises (page 20) 1.
(a) No @I No Cc) No (4 No (e) Y e s ~8. (a) &de2 + 1 (b) &T(X) =b(x - y), barbitrary 10. (b)(n + 1)(2n + 1) n+l 6na+ -2 b(C) g(r)=a(r -%$), aarbi t rary 11. (c) 43 (d) g(t) = a(1 - St), a arbitrary12. (a) No (b) No Cc> No Cd) No13. (c) 1 (d) e2 - 114. (c) n!/2n+' 622 Answers to exercises 623 1.17 Exercises (page 30) 1. (a) and (b) id3 (1, 1, l), gJ6 (1, -2, 1) 2
. (a) +Jz(l,l,O,O), $J6(1,1,2,0), +J3(1, 1,1,3) (b) 6$ (1, 1, 0, l), L (1,JG - 2 , 6 , 1) 6. 8 - 4 log2 3 9 . 77 - 2 sin x7 . e2 - 1 10. 2 -ax *(e - e 9 +3 8. ; x ; 1 - 7em2 2.4 Exercises (page 35) Chapter 2 1. Linear; nullity 0 , rank 2 13. Nonlinear2 . Linear; nullity 0 , rank 2 14. Linear; nullity 0, rank 23 . Linear; nullity 1 , rank 1 15. Nonlinear4 . Linear; nullity 1 ,
rank 1 16. Linear; nullity 0, rank 35. Nonlinear 17.
Linear; nullity 1, rank 26 . Nonlinear 18. Linear; nullity 0, rank 37 . Nonlinear 19. Nonlinear8. Nonlinear 20. Nonlinear9 .
Linear; nullity 0 , rank 2 21. Nonlinear 10. Linear ; nullity 0, rank 2 22. Nonlinear11. Linear ; nullity 0 , rank 2 23. Linear ; nullity 1, rank 212. Linear ; nullity 0 , rank 2 24. Linear; nullity 0, rank n + 125. Linear; nullity 1, rank infinite 26. Linear; nullity infinite, rank 227. Linear; nullity 2 , rank infinite28. N(T) is the set of constant sequences; T(V) is
the set of sequences with limit 029. (d) { 1, cos x, sin x} is a basis for T(V); dim T(V) = 3 (e) N(T) = S (0 If T(f) = cf with c # 0, thencET(V)so wehavef(x) =c, +c,cosx+c,sinx; ifcr =0 , thenc = TT andf(x) = cr cos x + c2 sin x, where cl, c2 are not both zero but otherwise arbitrary;if cr # 0, then c = 2n andf(x) = cl, where cr is nonzero but otherwise
arbitrary. 2.8 Exercises (page 42) 3. Yes; x =v, y =u 10. Yes; x=u-1, y = v - 14.
Yes; x=u, y = -v 11. Yes; x=&(v+u), y=gv-u)5. No 12. Yes; x =Q(v +u), y =Q(Zv -U)6. No 1 3 . Y e s ; x=w, y = v , z=u7 . N o 14. No8 . Y e s ; x =logu, y =logv 1 5 .
Y e s ; x =u, y E&V, z =#w 9. No 1 6 . Y e s ; x=v, y=u, Z=W U V1 7 . Y e s ; x=u-1, y=v-1, z=w+l1 8 . Y e s ; x=u-1, y=v-2, Z=W-31 9 . Y e s ; x=u, y - v - u , Z=W v 20.
Yes; x=i(u-vu+), y=gv-w+u), z=gw-u+v)25. (S + T)2 = S2 + ST + TS + T2 ; (s + T)3 = S3 -I- TS2 + STS + S2T + ST2 + TST + T2S + Z-3 624 Answers to exercises 26. (a) (ST)(x,y, z) = (x + y + z, x + y, x) ; (TS)(x, y, z) = (z, z + y, z + y + x) ;(ST - TWX, y, z) = (x + y, x - z, -y - z) ; SYX, y, z) = (x, y, z) ;T2(x,y,z) = (x ,2x +y,3x +2y +z);(ST)2(x, y, 2) = (3x
+ 2y + z, 2x + 2y + z, x + y + z) ;(T.!q2(x,y,z) = (x +y +z,x +2y +2z,x +2y +3zj;(ST - TS)2 = (2x + y - z, x + 2y + z, -x + 2y + 22)(b) S -'(u, v, w) = (w, v, U) ; T-l(u, v, w) = (u, v - u, w - v) ;(ST)-l(u,v, w) = (w, v - w, I.4 - v); (TS) -'(u, v, w) = (w - v, v - u, u)cc> CT- Nx,y,d =(%x,x +y>; (T-02(x,y,d =(0,0,x);(T - Z)n(x,y, z) = (0, 0,O) if n 2 3 28.
(a) Q(x) = 3 - 2x + 12x2; 7"(x) = 3x - 2x2 + 12x3 ; (Dr)p(x) = 3 - 4x + 36x2 ;(T@(x) = -2x + 24x2 ; (DT - TD)p(x) = 3 - 2x + 12x2;(T2D2 - D2T2)p(x) = 8 - 192x (b) p(x) = ax, a an arbitrary scalar(c) p(x) = ax2 + 6, a and b arbitrary scalars (d) Allp in V 3 1 . (a) Rp(x) =2; Sp(x) =3 -x+x2; Tp(x) = 2 x +3x2 -x3 +x4;(ST)p(x) = 2 + 3x - x2 + x3 ; (z-S)p(x) =
3x - x2 + x3 ; (TS)2p(x) = 3x - x2 + x3 ;(zv)p(x) = -x2 + x3 ; (ST)p(x) = 2 + 3x - x2 + x3 ; (TRS)p(x) = 3x ;(RST)p(x) = 2 (b) NW = {p 1 p(O) = 01; R(V) = {p) p is constant} ; N(S) ={p 1 p is constant} ; S(V) = V; N(T) = (0); T(V) = {p Ip(0) = 0} (c) T-1 = s(d) (Z-S)n = Z - R ; SnTn = Z 32.
T is not one-to-one on V because it maps all constant sequences onto the same sequence 2.12 Exercises (page 50) 1. (a) The identity matrix Z = (a&, where aj, = 1 ifj = k, and fij, = 0 ifj # k(b) The zero matrix 0 = (uik) where each entry ujk = 0(c) The matrix (cajr), where (Sjk) is the identity matrix of part (a) 2. (4 3. (a) -5 + 7j, 9i - 12j Cb) [: J [: J (4 [i -
g* [: I] - 1 - 1 i 5. (a) 3 i + 4j + 4k ; nullity 0 , rank 3 (b) 1 - 3 3 - 1 - 5 5. 6. Answers to exercises 625 7. (a) T(4i - j + k) = (0, -2) ; nullity 1, rank 2 (c) [i i -iI (d) e,=j, e2=k, e,=i, w,=(~,l), wz=(l, l) 1 - 1 8.
(a) (5, 0, -1); nullity 0, rank 2 (b) 0 0 [I1 1 (4 e, =i, e2 = i + j , Wl = Cl,09 11, w, = (O,O, 21, w 3 = (0, 1,O) 1 1 9. (a) (-1, -3, -1); nullity 0, rank 2 (b) 0 0 [I1 1 (4 e, =i, e2=j i, wl=(l,O,l), w2=(0,1,0), w,=(O,O,l) 1 210. (a) e, - e2 ; nullity 0, rank 2 (b) 5 4 [I(c) a = 5, b = 4 11* [y -32 [-:, -;I 1 2 . pj.
[Z] 13. [% A -g, i & j] 14* [:, :I. [i :] 15* c -:I> [-:, -:I 16.[: g qy;] 17. i -;I. 6 !:I 18. [: -;I, [-; ";] 626 Answers to exercises 19. (a)19. (a) 0 0 0 04 0(cl0 0 19 i 00 00 0 0 0'0 0 0 00 1 0 0 (00 0 4 0 p 0 0 9 I r 0 O - 8 0 0 0 - 4 8(4 (e) 0 0 0 0 p 0 0 0 1020. Choose (2, x2, x, 1) as a basis for V, and (x2, x) as a basis for W. Then the matrix of TD is 0 012 0 00 6 I0
2.16 Exercises (page 57) 0A C = 0 1 II6.
An = [0 1 17. A " = 8. An = Answers to exercises 627 where b and c are arbitrary, and a is any solution of the equation a2 = -bc where a is arbitrary -:I. a n d [I -11, where b and c are arbitrary and a is any solution of the equation a2 = 1 - bc 14. (b) (A + B)* = A* + AB + BA + 82; (A + @(A - B) = A* + IL4 - AB - B*(c) For those which commute 2.20
Exercises (page 67) 1. (x9 y, z> = (E, -;, $)2. No solution 4 .
(x,y,z> =u, -1,O) +t(3,4,1) 3. b,y, z) = (1, -1,O) + t(3,4, 1)5. (x,y,z,~)=(l,l,0,0)+t(l,14,5,0)6 . by, z, u> = (1,&O, -4) + t(2,7,3,0) 7.
(x,y,z,u,u) =t,(l,l,O,O,O) $I,(1,0,3, -3,l)8. (x,y,z,~)=(1,1,1, 1)+t~(1,3,7,0)+t~(4,9,0,7)9. (x, y, z> = (9, $, 0) + t(5, 1, -3) 10.
(4 (x, y, z, 4 = (1,6,3,0) + t,(4, 11,7,0) + t,(o, o, o, I)(b) (X,)', Z, U) = (&, 4,+f, 0) + t(4, -11, 7, 22) - 1 2 1 12.
[5 - 8 -3 5 - 6 14A# 13. [-1 Q ?0; -$ 1 1 Q i 0 h 0 - 1 1 0 0 0 0 0 0 1 16.
- 3 0 1 0 0 0 0 0 0 9 0 - 3 0 0 1 - 0 0 0 -1 0 0 0 s 1 0 14. 15. 14 8 3 8 5 2 3 2 1 1' 1 -2 0'1 0 l - 2 1 0 0 l - 2 0 0 0 1 628 Answers to exercises 2.21 Miscellaneous exercises on matrices (page 68) 2 1 3 .
P = [5 - 114- T,"o :]s [:, ;]7 and [: 1 -"J where b and c are arbitrary and a is any solution of the quadratic equation a2 - a + bc = 0 10. (a)[: -: :]s [: -:]s [: -: :I. [: -:]s [: I:]3 [I: -:]p [: I:]9 [I: -:I Chapter 3 3.6 Exercises (page 79) 1. (a) 6 (b) 76 (c) a3 - 4a2. (a) 1 (b) 1 Cc) 13.
(b) (b - a)(c - a)(c - b)(a + b + c) and (b - a)(c - a)(c - b)(ab + ac + bc)4. (a) 8 (b) (b - a)(c - a)(d - a)(c - b)(d - b)(d - c) (4 (b - a)(c - a)@ - a"'"(7 b!t6; b)(d - c)(a(d) a(a2 - 4)(a2 - 16) e fi fi' fi fi fi f3 fl fi f3 7. F' = 81 g2 g3 + 8; g:l gj, + gl gz g3 h, h, h, h h, h, h; h;; h; fi h f3 fl b f3 8. (b) If F = fi fi fi then F' = f; fi fi f;' f; fif; f; ff p -2 +b+c+d) 10 det A = 16, det
(A l) = k , 3.11 Exercises (page 85) 6 . det A = (det B)(det D)7. (a) Independent (b) Independent (c) Dependent 3.17 Exercises (page 94) Answers to exercises 629 3 . (a) A=2, I = -3 (b) A = 0, I = f3 (c) il=3, 1=& i5 . (a) x=0, y=l, 2=2 (b) x=1, JJ=~, ,z= - 1 -x1 yz -y1 z2 -%]=O; detk i i! :]O x3 - Xl y3 -y1 z3 - z1 (x - XlP + 0, - y1Y (x - x1) fjJ - yl) (4 det
(x2 - x112 + (y2 - ylj2 (x2 - xl) cy2 - yl) I = 0 ; (x3 - a2 + cy3 -y1Y (x3 - Xl) (y3 -yJ x2 + y2 x y 1 det4 +Yi Xl Yl 1I 1 =4 +u: 0x2 Y2 1 4 +Yi x3 y3 1 Chapter 4 4.4 Exercises (page 101) 5. Eigenfunctions: f(t) = Ct", where C # 06 .
The nonzero constant polynomials7 . Eigenfunctions : f(t) = Get'" , where C # 08 . Eigenfunctions: f(t) = Ce~~t2/~, where C # 0 10. Eigenvectors belonging to 1 = 0 are all constant sequences with limit a # 0.
Eigenvectorsbelonging to 1 = -1 are aI1 nonconstant sequences with limit a = 0 4.8 Exercises (page 107) Eigenvalue 1. (a) 1,1(b) 131(c) 1,lCd) 2 0 2. 1+ Jab1 - Jab Eigenvectors dim E(1) (a, 6) Z (O,O)t(1, O), t # 0t(0, l), t # 0t(1,11 t #Ot(1, -l), t # 0 t&I; Jb), r z 0t(J; -J&t #O : 630 Answers to exercises 3. If the field of scalars is the set of real numbers R,
then real eigenvalues exist only whensin tl = 0, in which case there are two equal eigenvalues, I, = I, = cos 0, where cos 8 = 1 or -1. In this case every nonzero vector is an eigenvector, so dim E(I,) = dim E(&) = 2. If the field of scalars is the set of complex numbers C, then the eigenvalues are A, =cos f3 + i sin 0, I, = cos f3 - i sin 8. If sin 0 = 0 these
are real and equal. If sin 8 # 0 they aredistinct complex conjugates; the eigenvectors belonging to I, are t(i, l), r # 0 ; those be-longing to 1, are r(l) i), I # 0 ; dim E(1,) = dim E(1,) = 1 . a b4. [I2 where b and c are arbitrary and a is any solution of the equation a2 = 1 - bc . c - aa b 5. Let A = [1 , and let A = (a - d)2 + 4bc. The eigenvalues are real and
distinct if A > 0,c d real and equal if A = 0, complex conjugates if A < 0.6. a=b=c=d=e=f=l. Eigenvalue Eigenvectors dim E(I) 7 . (a) l,l,l t(O,0, l), t # 0 1(b) 1 t(1, -l,O), t # 0 1 2 t(3, 3, -l), t # 0 12 1 t(1, 1, 6), t # 0 1 (cl 1 t(3, -1,3), t # 0 1222 t(2,2, -l), t # 0 1 8. 1, 1, -1, -1 for each matrix 4.10 Exercises (page 112) 2. (a) Eigenvalues 1, 3; C = , where cd
0 (b) Eigenvalues 6, - 1; C = , where ab Z 0 (c) Eigenvalues 3, 3; if a nonsingular C exists then C lAC = 31, so AC = 3C, A = 31(d) Eigenvalues 1, 1; if a nonsingular C exists then @AC = I, so AC = C, A = Z 3. C=A lB.4 . (a) Eigenvalues 1 , 1 , - 1 ; eigenvectors (1, 0 , l), (0, 1 , 0), (1, 0 , - 1) ; 1 c = o i0 1 11 0 0 I -1 (b) Eigenvalues 2, 2, 1; eigenvectors
(1, 0, -I), (0, 1, -l), (1, -1, 1); c = i1 0 1 0 l -1 -1 - 1 1I Answers to exercises 631 5. (a) Eigenvalues 2, 2; eigenvectors t(1, 0), t # 0. If C = , b #O, then C -'AC = [2 0 1 21(b) Eigenvalues 3, 3; eigenvectors t(1, l), t # 0. If C = , b#O, t h e n @AC = [3 '1 3 1 6 . Eigenvalues 1, 1, 1; eigenvectors t(1, - 1 , -l), I # 0 5.5 Exercises (page 118) Chapter 5 3 . (b) Tn is
Hermitian if n is even, skew-Hermitian if n is odd7. (a) Symmetric (b) Neither (c) Symmetric (d) Symmetric9.
(4 Q(x + ty> = Q(x) + tfQQ> + fUW,y) + U(y), x> 5.11 Exercises (page 124) 1 . 4 . 5 .
6. 7. (a) Symmetric and Hermitian(b) None of the four types(c) Skew-symmetric(d) Skew-symmetric and skew-Hermitian Cb) [;;; -1: -;] Eigenvalues A, = 0, 1, = 25; orthonormal eigenvectors ICY = $(4, -3)) u2 = 5(3,4). c=f4 3 [I 3 4 Eigenvalues ii, = 2i, AZ = -2i; orthonormal eigenvectors u, = L(1, -i),J2 u2 =L(l,i).1 J2 C=$ yi i[1Eigenvalues 1, = 1, 1, =
3, 1, = -4; orthonormal eigenvectors u1=+1,0,3),Ju2 = + (3,2, -l), JICY = Lz (3, -5, -1). J 632 Answers to exercises 8. Eigenvalues A, = 1, I, = 6, 1, = -4; orthonormal eigenvectors I41 = ; 4 = (O,l,O>, u2 =L(2,0,1),J5 I+ = $ (1, 0, -2) 0 2 1 (d) C = iJ5 [$ 00 1 0 1 -2 3 (a) A = [2 2 4 0 2 1 (b) I,=A2= 1, A,=8 4 2 3 (4 4 = 5 (l,O, -l), u2 =A(1,4, -l), u,
=k(2,1,2)3J2 3 - 1 (d) C = L 242 3J2 [0 4 -3 - 1 Ji 12JZ Ellipse; center at (0,O)Hyperbola; center at (-$, -$)Parabola; vertex at (&, -)8)Ellipse; center at (0,O)Ellipse; center at (6, -4)Parabola*, vertex at (&, s$) 5.20 Exercises (page 141) 8.
a = fi$ Chapter 66.3 Exercises (page 144) 1. y =e32 -$5 6. (b) y = e*+ - e r'/32. y = %x2 + $5 7. y = cle2@ + c2e-223. y = 4 cos x - 2 COG x 8. y = cl cos 2x + c2 sin 2x4. Four times the initial amount 9. y = @(cl cos 2x + c2 sin 2x)5 . f(x) = Cxn, o r f(x) = Cxlln 10. y = e+(c, + c2x) 14. Ellipse; center at (0,O)15. Parabola; vertex at (2, Q)16. Ellipse;
center at (- 1, +)17. Hyperbola; center at (0,O)18.
Hyperbola; center at (-1,2)19. - 1 4 13. (a), (b), and (4 6 3 4 Answers to exercises 11. k = $T? ; fk(x) = C sin mx (n = 1,2, 3, . . .)13. (a) y" -y = 0 (b) y" - 4y' + 4y = 0 (c) y" + y' + & = 0(d) y" + 4y = 0(e) y" - y = 0 14. y = JJZ, y" = -12y = -446 6.9 Exercises (page 154) 1.
y = Cl + c'p + cp 3. y = Cl + (c2 + c3x)e-222. y = cl + c2ex + c3e@ 4. y = (cl + c2x + c3x2)ez5.
y = (Cl + c2x + c3x2 + c,x3)P6. y = cle2x + c2e-2x + c, cos 2x + c, sin 2x 7. y = ediz(cl cos JZx + c2 sin JZx) + e -'/22(c3 cosJZx + c, sin JZx) 8 . y = clee + e -"/2(c2 cos &/% + c3 sin iJ3x)9. y = e -"[(cl + c2x) cos x + (c3 + c4x) sin x] 10. y = (cl + c2x) cos x + (c3 + c4x) sin x 11. y = cl + c2x + (c3 + cg) cos JZx + (cc + cSx) sin JZx12. y = c1 + c2x + (c3
+ cqx) cos 2x + (cg + csx) sin 2x 13. f(x) = k2 (em" - cos mx - sin mx) 15. (4 y (4) - 5y" + 4y = 0 (b) y"' + 6~" + 12~' + sy = 0 Cc) y (4) _ Zy" + y" = 0(4) _ zy" + y" = 0(4 y (4 y (3 - ~~(4) + yflf = 0 (f) yt4' + 8~"' + 33~" + 68~' + 52~ = o (g) Y (4) - 2y" + y = 0 (h) yf6' + 4~" = 0 6.15 Exercises (page 166) 1. y1 = -2x - x2 - ix" 5.

y, = ix2F + e2z2. y1 = $xe2x 6. y, = hxe"3. y1 = (x - $)e2 7. y1 = x cash x4. y, = 6 sin x 8. y1 = 2L,x4e r9. 50yl = (11 - 5x)e" sin 2x + (2 - 10x)e5 cos 2x 10. y, = -(fX + $x2 + & x3)e " xmeoLx12. Y l =(m) P A (u) 15. (b) 20 (c) 3D2 (d) nD'+l 16. y = Ae" + Be- + i exs e+; dx - 1 e++ s;dx 17. y = (A + ix) sin 2x + (B + 1 log lcos 2x1) cos 2x18.
y = Ae" + Be+ + i set x 19. y = (A + Bx)F + eez - xe"s ee2 dx + e"s xeez dx Answers to exercises 635 s eP20. y = - $ log 1x1 + f e" x dx - i ez5 s e-2x - dxX + A egxsT dx + Ae" + Bezx + Ce4" 6.16 Miscellaneous exercises on linear differential equations (page 167) 1. u(x) = 6(e4" - eP)/5 ; v(x) = e" - e-5r2 . u(x) = $2z -a sin 5x ; v(x) = $e-2x * sin 3x3. u(x)
= ewx2; Q(x) = 4x2 + 25. y = (A + Bx3)e" + (x2 - 2x + 2)e2"fj. y = A& J" e 4X x3/3 & + B& 7. y = Ax% + Bx ?i8. y = Ac5 + Bx2e+ - x9. y = A(x2 - 2) + B/x 10. y = x ~[A + B(x - 1)3 + ix" + ix" - ;x + 4 - (x - 1)3 log Ix - 11111. a = 1, -1; y = [Acg(") + Bcpg(")]/x 6.21 Exercises (page 177) 2. f(x) = Ul(X) (a = 1)3. (a) A = (a - b)/2, B = (a + b)/2 (b) -$I' -)$] -
u(cc+l)y=O, where cc = 1 or -2, and x = (t + 1)/2 4. z+(x) = 1 -I - $ (1p2m4~ - 2) Cm - 2m + 2)x2ot for all x; V&=1(2m) ! m u&x) = x +c (l)m2m(a - l)(a - 3) . . . (a - 2m + 1) X2m+l for all x TfZ=l(2m + l)! cc 5. Ul(X) = 1 +c C 1)"m=l (3m + 2)(3m - 1) . . . 8 5 x3m for all x; u2(x) =~ ~(l +zzx3n) forallx ZO 11. (b) f(x) = BP,,(x) + +=2(x> + &p4(x) 6.24
Exercises (page 188) + sin x 636 Answers to exercises 9.
(a) y = x~[c,J5~(~x?~) + c.&&x~)](b) y = ~'~[c,J?,,(+x*) + cp ,,(+x",l(c) y = x!*[cJ,(2 axlfm'*) + c*J,(2axl+"'*)], where do = l/(m + 2), provided thatl/(m + 2) is not an integer; otherwise replace the appropriate J by K(d) y = x?~[cJ,(~x*) + c2S,($x2)], where t(= &!/8 10. y =g, satisfies x2/' + (1 - 2c)xy' + (a*b*~*~ + C* - C&~)]J = 0(a) y = x ~[cIJ,
(2x~) + c2K5(2xx)](b) y = x x k&(x) + c2Lm(x)](c) y = x ~[cJ,($x~) + c,K,($xG)](d) y = x[c,J&x'm) + c&&x";)] 11.
a =2, c = o 12. y=c* onx"; n=O @!I2y = J&2x5*) if x > 0 13. b = (p.
-aJ/a,, c = qo/ao14. y = x*4 co15. t = 1: y = c(1P (2;)! ! (2x)n Tk=l 16. Q(X) = cos x ; ul(x) = 3 - % cos x - + cos 2x 7.4 Exercises (page 195)Chapter 7 k 1 3. (b) (Pk)' = 2 ~v"P~ l mnko 7.12 Exercises (page 205) 1. (a) A l =21 A, A" = nA - (n - l)Z (b) etA = et(l - t)Z + tetA = et 2. (a) A l = $Z - $A, An = (2n - 1)A - (2n - 2)Z (b) eta = (2et - e2t)Z + (ePt -
et)A =et 0 [1$t - et e2t 3. (a) A l = A, An =1 + (lYz + 1 - (l)nA 2 (b) eta = (cosht)Z + (sinht)A = [:~ ~:2~~] Answers to exercises 631 1 + 1 4. (a) A ' = A , A" (l)n= (1)"Z2 +2 A e& 0(b) etA = (cash t)Z + (sinh t)A = [1o et 5. (b) etA = Ft['I;n:, :::I 7. eAtt) = Z + (e - l)A(t) ; (eA@))' = (e - l)A'(t) = 0 e 0 1ed(t = [1 ; A'(')&(t) = 0 0 [0 0 18. (a) An = 0 if n
2 3 9 . (a) An = A if n 2 1 etA = 0 et 0 ie2t 0 0 @I 0 tet et111. eta = Z + tA + &t2A2 13. eAcB = r; '"', ')1; eBcA = 7.15 Exercises @age 211) 1. eta = &(3et - e3t)Z + i(e"" - et)A 2 . eta = (cash $ t)Z + L (sinh $ t)AJ5 3. etA = &et((t2 - 2t + 2)Z + (2t2 + 2t)A + t2A2}4. etA = (se t - 3e 2t + @)I + ($e t - ] + *@)A + (&e t - e 2t + $@)A2 5 .
eta = (4et - 3e2t + 2te2t)Z + (4e2t - 3te2t - 4et)A + (et - e2t + te2t)A26. &A = (&t _ &2t + &St _ @)I + (13aet + L2ae2t - j'$t + l&&t)A + (set _ 4e2t + $3 - e4t)A2 + (set + $,ezt _ &y3t + *&)A3 7. (b) etA = $e"t{(6 - 6A.t + 3A2t2 - L3t3)Z + (6t - 61t2 + 3a2t3)A + (3t2 - 3;it3)A2 + t3A3} 638 Answers to exercises yl = cl cash JS t +Cl + 2c, 2c, -
c2psinh Jjt, ya =c,coshJjt + - 8 . Jj Jsmh JS t 9. .yl = d(COs 3t - sin 3t), yz = et(cos 3t - 3 sin 3tj10. y1 = ezt + 4tezt, yz = 2et + ezt + 4tezt, y3 = 2et + 4e2t 7.17 Exercises (page 215) 2. (Cj y1. = (b - l)@ + 2(C + 1 - bjxe" + 1, y, = ce" + 2(c + 1 - b)xcz4. y1 = *ct - +& + $$t, y2 = $ct - $c4t + gc2t 5. (a) B, = B, B, = AB, B, =A A2B,. . . , B, =lA"B *RZ!
(b) B = m! (A l)"+lC 6. (a) I + tA + i t2A2 + it3A3 B, where B = 6A 4C = - 'This gives the particular solution y, = y2 = & - &t - &.t2 - its I(b) y?, = y2 = A - $zt - &t2 - at3 + iT%e131 4t 7 . .E=B, F=;(AB+C) 8. I(a) yI = COS 2t - 8 sin 2, y2 = 4 sin 2t I:bj y].
= 2 cash 2t + 8 sinh 2t - cos 2t - & sin 2t, y2 = cash 2t + & sinh 2t - 4 sin 2t9. x(X) = c2s + 2" - e5, y2(x) = 2e2x - e3= + 38 10. JVl(X) = &se2 - 3Lae2a: + (cl - $$+cdx + (+* - c1 - c2)xeGs, JV2(Xj = 2%e3: + 3%3e2z + (c2 - $+j$)e 45 + (cl + c2 - $&)xe 4z 11.
y,(x) = eC4"(2 cos x + sin x) + g+ex - g, y2(xj = eW4"(sin x + 3 cos x) - Aez + I6712. n(x) = eez(x2 + 2x + 3) + x2 - 3x + 3, y2(x) = e "(2x - 2) + x, y3(x) = 2e " +x - 1 7.20 Exercises (page 221) 4. (c:) Y(x) = eze'AzzABm 5. If A(x) = zx"A,, then Y(x) = B + XC + z xkBk,k=O k=2 where (k + 2j(k + lj&,, = $ A,B, , f o r k 20.7=0 7.24 Exercises (page 230) 1.
(a) Y(x) = e"n c Xk(b) Y,(x) = 2e" - F if n is odd; kc0 ' Y,(x) = 2 g if n is evenk=O Answers to exercises 639 2 . Y,(x) = ; + ;; + so + g(; x4 x7 Xl03. Y,(x) = x + 2 + j j + F. x3 4x' 8x9 184~~~ 4x134 .
Y3(x) = 5 + 63 + 405 + _ _ 51975 + 12285 3 x4 55 . (a) YA(x) = 1 + x + x2 + 7 + - 6 + g + i (b) M=2; c=& (c) Y(x) = 1 + x + x3 + y + T + y + . . . 4x13 Xl56.
(a) Y,(x) = x + z + $s + g + g5 + g + _12285 + ~59535 (d) Y(x) = tan x = x + I + g + g + gi + . . for 1x1 < ; 8.
Y,(x) = 2 + x2 + x3 + F; + r" ;1 0z3cx) = 3x2 + 'G + '!! + g + 'g x4 x6 2x7 x99 . Y,(x) = 5 + x + r2 + 6 + 63 + 72 ; x3 6 x8 9 7x11Z,(x) = 1 + 7 + x5 + g + g + g + F4 10. (d) Y,(x) = 0; lim Y,(x) = 0n CC (4 Y,(x) =t x2 if x 2 0 x2 if x 2 0 -X2; lim Y, (x) = if x < 0 n+a -x2 i f x50 (f) Y,(x) = g ; lim m(x) = 07k CC x2 if x 2 0 x2 if x 2 0(g) Y,(x) = -x2; lim Y,(x) =
if x 10 ~l m -x2 i f x50 Chapter 8 8.3 Exercises (page 245) 2 . All open except (d), (e), (h), and (j)3. All open except (d)5. (e) One example is the collection of all 2-balls B(0; l/k), where k = 1, 2, 3, . . .6. (a) Both (b) Both (c) Closed (4 Open (e) Closed (f) Neither (g) Closed (h) Neither (i) Closed (j) Closed (k) Neither (1) Closed8. (c) One example is the
collection of all sets of the form Sk = {x 1 IIxII I 1 - l/k} for k = 1,2, 3, ... Their union is the open ball B(0; 1)10. No 640 Answers to exercises 8.5 Exercises (page 251) 1. (a> All (x,y)(b) AH (x, y) z (0, 0)(c) All (x, y) withy # 0(d) All(x,y)withy f Oand: #f+k~(k =0,1,2,...)(e) All (x, y) with x # 0(0 All z (0, 0)(g) All (x, y) with xy # 1U d All (x, y) # (0, 0)(0
All k y) z (0, 0)(j) All(x,y)withy #OandO 5x I;yory Sx $O 5 . lim f[x, y) does not exist if x # 0 6. Ti" m2)/(1 + m2) ; NO 7. y = 4x2; f not continuous at (0,O)8. f(O,O) = 1 8.9 Exercises (page 255) 1. f'(x;y) = a*y 2. (a) f'(x;~) = 4 11~11~ x -Y(b) All points on the line 2x + 3y = &(c) All points on the plane x + 2y + 32 = 0 3. f'(x;y> =x T(y) +y T(x) af4. =&
=2x +y3cos(xy); af - = 2y sin (xy) + xy2 cos (xy)aY 5.af7& = x/(x2 + y2)" ; afI& = y/(x2 + y2)% 6.af 2,z = y2/(x2 + y2)s ; af = -xy/(x2 + y2)3f' aY af7* zi af= -2y/(x-y)"; =2x/(x-y)2 aY 8. &f(x) = uk, where a = (a,, . . . , a,) 9. &f(x) = 2 2 a,,x,5=1 af af10. =4x3-8~~2; =4~3-8~2~ax aY Answers to exercises 641 ,4 K L. af l* ax = x2 + y2 ' ay x2 + y2 15
af1 +y2 . af 1 +x2 * is= 1 +x2+y2+.?y2 q=1 +x2+y2+x2yZ 16 af a!. z& = y2x+1 ; gj = 2yxyB log x 17. af 1 af J;;ax = -2Jxo, -x); G=2yJy x 18. n = -Q19 . a=b=l22. (b) One example is f (x) = x . y , where y is a fixed nonzero vector 8.14 Exercises (page 262) 1. (a) (2x + y3 cos xy)i + (2y sin xy + xy2 cos xy)j(b) 8 cos yi - eZ sin yj(c) 2xy3z4i + 3x2y2z4i +
4x2y3z3k(d) 2xi - 2yj + 4zk 2x 62(4 x2 + 2y2 - 3z2 i + 4Y .x2 + 2y2 - 322 rx2+2y2 322 k (f) yzx~' % + z~%P" log xj + y*xY' log x log yk 2 .
(a) -2/&i (b) ll&3 . (1, 0), in the direction of i; (1, 0), in the direction of -i4. 2i + 2j ; L545. (a, b, c) = (6,24, -8) or (6, -24, 8)6 . The set of points (x, y) on the line 5x - 3y = 6 ; Vf (a) = 5i - 3j8. (c) Yes (4 fk y, z) = SW + y2 + z2>11. (b) implies (a) and (c); (d) implies (a), (b), and (c); (f) implies (a) 8.17 Exercises (page 268) 1. (b)w ?f ?f af afF"(t) = Z2
[X'(t)12 + 2 axy X'(t) Y'(t) + [Y'(t)12 + z X"(t) + q Y"(t) aY2 2. (a) F'(t) = 4t3 + 2t ; F"(t) = 12t2 + 2(b) F'(t) = (2 cos2t - l)ecostsin~ cos(cos t sin2t) + (3 sin3t - 2 sin t)ecos t sin t sin(cos t sin2t) ;F"(t) = (5 co@ t - 3 cos4 t - 4 cos3 t - cos2 t - 4 cos t)ecost sin c cos (cos t sin2 I) + (14 sin3 t - 12 sin5 t - 4 sin t + 7 cos t - 9 cos3 t)ecostsint sin (cos t sin2 t)
(c) F'(t) =2e2t exp (e2t) 2e-2t exp (eC2t)1 + exp (e2t) +1 + exp (e-2t) ' where exp (u) = e" ; F"(t) =4[1 + e2t + exp (e2t)]e2t exp (e2t) 4[1 + e-2t + exp (e-2t)]e-2t exp (ew2') -[l + exp (e2t)]2 [l + exp (e-2t)]2 642 Answers to exercises 3. (a) -3(b) x2 -y2(cl 0 4. (a) (1 + 3x2 + 3~4(xi + rj) - (x2 + y2)%, or any scalar multiple thereof(b) cos 0 = -[l + (1 +
3(x2 +~~))~] !d; cos 0 + -iJ2 as (x, y, z) -+ (0, 0,O) 5.
U(x, y) = 4 log (x2 + r2) ; V(x, v) = arctan Q/x)6. (b) No8 . x/x0 + y/y0 + z/z0 = 39 . x+y+2z=4, x y z= 1 10. c = *J3 8.22 Exercises (page 275) af1.
(b) ~ = -2x sin (x2 + y2) cos [cos (X2 + y2)lesin[cos W + ya)1 af laf laf. aF2 .
2&= I af 1 afzzG++z2 q= z;i;+zz aF afaX afar aF afaX afaY3 . (4 x=zb+qx; ;i;=zz+%z a2F a2faXaX(C) ZZ 3XGTt + azfayay af a2x af av+t$cGz+zas++mi aF af af aF af af a2F ay ay ,?f4 . (a) x=g+tG; ~=z+s%; B=z2+2tmy+t=p; a2F azf a!f_ _ p?fa2F ay = - = ax2 + 2saxay +' ay2 ; asat azf ?f afat2 ax2 + (' + ')axay + 'Cp + 5 ay S a?f af 1 af =st - + -
asat ax2 t3 ay2 ax t2ay a2F 1 azf 1 azf 1 i?f a2F 1 8f 1 azf 1 a*f _ _ =5Z2+ZZFy+;i~; af2=4ax2 2axay+4vas2 azv5* p = co@ (j Ef + ~0s /j sin e azf azf ax2 ax av + ay ax+ sin2 0 Ef . ay2 3 azq r cos 8 sin e Ef + r cos2 e ay w ZFe= - ax2 ax ay r sin2 e ayx+ r cos e sin 8 Ef aY2 -sin 8 af + cos 8 af ;ax aY af af af aF9. (a) g = 2s~ + 2s + 2t z ; z = aY2t af _
2t af + zs afax ay aZ aF af af . aF Jf af aF af af(b) ,r=z+2rq, ~=~+2sq; ,t=;r++tay 13. (a) f(x, y, z) = xi + yj + zk, plus any constant vector(b) f(x, y, z) = P(x)i + Qf$)j + R(z)k, where P, Q, R are any three functions satisfyingP'=p,Q'=q,R'=r 14. (a) @I w 15. (a) (b) (4 &X+z?J 2F+2'm"(x, y) = 2 cos (y + 2x) cos 0, + 2x)1 ; &du,u,w) = [iu "," ';'I,,(u, v, w)
= cu+2va+3ws+4V 2uz~ I + sin (2~ - u2 + 2u + 4u2 + 6w3)j - 3 0 9Dh(1, -1, 1) = 0 - 6 c o s 9 18cos9IU2W2 2uvw2 2uv2w 2x 1 1Df(x, y, z) = 2 1 2z ; Mu, u, w) =[1 0 w2 cos u 2w sin v 2uc' u2c* 0 I h(u, v, w) = (u2v4w4 + w2 sin v + u2cv)i + (2uu2w2 + w2 sin v + u4c2V 2u w2 + u2 0Dh(u, 0, w) = 4u3 w2 + 2u4 I0 644 Answers to exercises 8.24
Miscellaneous exercises (page 281) 1. One example: f(x, y) = 3x when x = y,f(x, y) = 0 otherwise2. D,f(O, 0) = 0 ; D,f(O, 0) = -1; D,,,f(O, 0) = 0 ; D, 2f(0, 0) does not exist3 . (a) Ifa =(a1,a2j,thenf'(O;aj =a$,2ifa, #O,andf'(O;a) =Oifa, = 0 (b) Not continuous at the origin 4.
af 1 af 1z = e+U &dyW ; _ = z e XY J4y ?4aY 5. ?f ?f ?fF""(t) = ax3 W'(t)13 + 3 may W'(t)12 Y'(f) + 3 ax ay2 X'(f)[Y'(r)]2 w ?f+ - [y'(t)13 + 3 g-2 X'(t)xyt) + 3 ?faY3 - W"(t) Y'(t) + x'(t)"(t) Y'(t) + x'(t) Y"(t)]ax ay ?f af af+ 3 ay2 Y'(f) Y"(t) + ?& x,,,(t) + 5 Y(t), assuming the mixed partial derivatives are independent of the order of differentiation6. 8 7 (a) ag =?fv + af, %
af af a28 ?f wau ax aY = ; au ax" -3"; aj =uvQ + (U2 -V2)my - .va"r + afay2 ax (b) a=&, b= 4 10. (a) v'(t) = A'(t) J,Btt) f [A(t), y] dy + B'(t) j$(t) f [x, B(t)] dx(b) a)'(t) = 2tet2(2et2 - ea - eC) 13. A sphere with center at the origin and radius ,/i14. f(x) = x2 Chapter 9 9.3 Exercises (page 286) 1. f(x, yj = sin (x - $yV,2. f(x, y) = cx+5r'2 - 13. (a) u(x, yj =
x2y2Py (b) v(x, yj = 2 + log ;I I 5.
A = B = C = 1, D = -3; f(x, y) = q5(3x + y) + v2(x - yj6 .
G(x,y) = x - y 9.8 Exercises (page 302) I. ax/au = (1 + xuy(x - yj ; aY/au = (1 - yv)/(x-y) ; aY/av = (1 + yuj/~ - xj2.
ax/a, = -(I + xuji(~ + 4 ; av/au = (1 - yvj/(l + yu); av/ay = (1 - xj/(l + uj aX a(F, G>3 . x= -Ia(F, G)a(F, G) . aY a(F, Gj Ia(F, Gj aY a(F, G) acY, vj acx, yj 3 ZG = -acu, xj a(x,y> ; Z = -ah XI Iacxl Y) 4. T = f -k (24i - 4J? j + 3J? k)J Answers to exercises 645 -5 .6 .8 .9 .
10. 11. 12.
2i + i + J? k , or any nonzero scalar multiple thereofax/au = 0, ax/au = 42n=2aflax = -1/~2~ + 2z + 1); aflay = -2~ ~ + z)i(2y + 2z + 1);ay(axay) = 2/(2y + 2z + 1)sazz/caxay) = [sin (X +Y)COS~ (,Y + Z) + sin (y + Z)COS~ (X +y)l/~ ~ ~s(y + Z) af D,F + 2xD,F af D,F + 2yD,F = ax = D,F + 2zD,F ' ay D,F + 2zD,F D,F=f'[x +g(y)]; D,F=f'[x +g(y)lg'(y);
Dl~F=f"[x+g(~)];D,,,F = f v [x + g(y)]g'(y) ; D,,,F = f N [x + g(y)] [g'iy)12 + f 'Ix + g(J)&/(y) 9.13 Exercises (page 313) 1.2 .3.4 .5.6 .7 .8. 9 .10.11.12.13. 14.15.17.18. 19.
21. 22. Absolute minimum at (0, 1)Saddle point at (0, 1)Saddle point at (0,O)Absolute minimum at each point of the line y = x + 1Saddle point at (1, 1)Absolute minimum at (1,O)Saddle point at (0,O)Saddle points at (0,6) and at (x, 0), all x; relative minima at (0, y), 0 < y < 6 ; relativemaxima at (2,3) and at (0, y) for y < 0 and y > 6Saddle point at (0,O);
relative minimum at (1, 1)Saddle points at (nn + a/2, 0) , where n is any integerAbsolute minimum at (0,O); saddle point at (-$, -3)Absolute minimum at (-26, -","); absolute maximum at (1, 3)Absolute maximum at (p/3, 7~13); absolute minimum at (2~/3,2a/3); relative maximum at(n, n); relative minimum at (0,O); saddle points at (0, r) and (v, 0)Saddle
point at (1, 1)Absolute maximum at each point of the circle x2 + y2 = 1 ; absolute minimum at (0,O)(c) Relative maximum at (2, 2); no relative minima; saddle points at (0, 3), (3,0), and (3, 3)Relative maximum & at (4, 4) and (4, -4); relative minimum -& at (4, -4) and (4, $);saddle points at (0, 0), (fl, 0), and (0, *l); absolute maximum 1 at (1, -1) and
(1, 1);absolute minimum - 1 at (1, 1) and (- 1, - 1)(a) a = 1, b = -f(b) (I = 6log2 - 3a/2, b = v - 3log2 1 7ZLet x* = - n cxj, y* = ;$Yj, i=l i=l ui =xi - x * . T h e n a = ($Y~LQ)/($u~),i=l 61 andb =y* - ax * uj =xi -x*, vi = yi - y* , and leti=l i=l , where the sums are for i = 1,2, . . .
, n. Then 646 Answers to exercises 1 2 'izi 2 uivi (I = . A >izi 2 v, , b =z c~izi c~" , c=z* - a x * -by* 25. Eigenvalues 4 , 16, 16; relative minimum at (1, 1, 1) 9.15 Exercises (page 318) 1. Maximum value is &; no minimum2. Maximum is 2; minimum is 1 3. (a) Maximum is J -a2 + b2 J at (b(a2 + b2) 'h, a(a2 + b2) p") ; minimum is - ~a2 + b2 at a b a b (-
b(,a2 + b2) N, -a(a2 + b2) 'h) (b) Minimum is a2b2/(a2 + b2) atab2 a2b -a2 + b2 ' a2 + b2 ; no maximum 4. 5 .6 .7 .8 . Maximum is I + $/2 at the points (PDT + n/8, n VT - ~$3)) where n is any integer; minimumis 1 - &/2 at (na + 5x18, TZTT + 3a/8), where n is any integerMaximum is 3 at (4, -f, $); minimum is -3 at (4, 3, -$)(O,O, 1) and (O,O, -1) 9.
10.11. 12. 13.14. aabbcc a b C (a + b + ~)a+~+~ at a + b + c ' a + b + c ' a + b + c abet/?/2 5logr +3logJ5 m2 = A + c - &A - C)' + 4B22(AC - B2) (4 f &NJ2Angle is r/3 ; width across the bottom is c/3 ; maximum area is c2/(4J3) 10.5 Exercises (page 328) 1. -+$2. -2pa23. &4. Q 5. 06. 407. z.. Chapter 10 8 . +9. 360 - 3 x -10.
-27711. 012. (a) - 2 JZ r (b) -r Answers to exercises 647 256a3/1 52p2a3(1 + 25r2) [(2 + '02)s - 2&/3moment of inertia = 4a42~13 600 - 3642 - 49 log (9 - 4J2) 64[6]2 + log (3 + 2$)] 6ab2 6?rab2x = 3a2 + drr2b2 ; J = - 3a2 + &2b2 10.9 Exercises (page 331) 1. +a2 .
2a3 3 . a = (3c/2)%4 . 05. 8&in 0 - cos 0) 6 . au314 7 . -JZ 8.9 . 10.12.13. 14. 15.
16. 1% = (a2 + b2)x[na4 + (47~ ~ - n/2)a2b2 + 32a5b4/5]Z, = (a2 + b2)'h[?ra4 + (4 r3 + v/2)a2b2 + 32P5b4/5] 10.13 Exercises (page 336) 1. All except (f) are connected6.
(a) Not conservative (b) (2e2= - 5eR - 5n - 3)/107. (b) 38. $ 10. 4b2 - 8ab + 4 ; minimum occurs when b = T 10.18 Exercises (page 345) 1 . &x, y) = $(x2 + y2> + c2 . y(x, y) = x"y + c 3 . y(x,y)=x2cu+xy-y2+C4. y(x,y> =xsiny +ycosx +(x2 +y2)/2 + C5 . 9(x, y) = x sin (xy) + C6. p(x, y, z) = (x2 + y2 + z2)/2 + C7. q(x, y, z) = x2/2 - y/2 + xz - yz + c8. f is
not a gradient9.
f is not a gradient 10. f is not a gradient1 1 . q(x,y,z)=y2sinx+xz3-4y+2z+C12. q(x,y,z) =x + 2x2y - xaz2 + 2y - za + c arnfl1 3 . (b) q(x,y) =n+l + C i f n # - 1 ; v,(x,y) =alogr + C i f n = - 1 VP+215. p(x) = - P+2+ Cifp # - 2 ; q(x) = logr + Cifp = -2 16. P(X) =g(r) + C 10.20 Exercises (page 349) 1 . x2/2+2xy+y2/2=c2. x"y = c 648 Answers to exercises
3. $13 - xy -y/2 + (sin 2y)/4 = C4. cos 2x sin 3y = C5. Sy + 4x2y2 - 12eY + 12ye' = C6. j Q(X)e~P(z)dz dx _ ye!P(z)dz = C 8. (a) x + y = C'(b) r'/$ - 3 log 1x1 = C 9 . (a) 6(xy)!' - (y/x)% = C ; (x5y) " is an integrating factor(b) x + eB2 sin y = C ; eP cos y is an integrating factor 10. gy* + zX+y" = c, 10x3~~ + xsy5 = C, respectively; x2ys is a common
integrating factor Chapter 11 11.9 Exercises (page 362) 1. 42. 13. 245 - v4. ?r2/45. 26. 271 7. 68. tm3(eta - et) + r2 - r1 10. 4 11.12.
37 - J2) 13. (log 2)/6 11.15 Exercises (page 371) 1.2.3.4. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 5. 44 - 0 6. 67. B#8. (a) $ (b) 2 (c) 320~ - 3n/2Q + cos 1 + sin 1 - cos 2 - 2 sin 2e - e l$log2 Answers to exercises 649 19. s,'[s," " (x2 + y2) dy] dx = + 2 0 . y = 0 , y = x t a n c , x2 +y2 =a2, x2 +y2 =b2 (b) 4es + 2e/32 2 . m=2; n = l 11.18 Exercises (page 377)
1.2 .3 .4 . 5 . 6 .
7 .8 .9 . 10. 11. 12.13.14.15.16.19.20.21. 22.23.
_2a210ga a2+1 _ x = 4(a log a - a + 1) 'a(log a)2 ' = 2(a log a - a + 1) -1 jj =j7 =+ ia= jj = 256/(31%)3 - v log 3 P = 8 IIA3lI) 7 = Q IlA'bll ; assuming the x- and y-axes are chosen along sides AB and AD,respectively Z, = &b3(a - c) , Z, = &b(a3 - c3)Z, = Iv = (1 - 5a/16)r4z, = zg = QZ, = &[(4a - l)e4= - 11,z, = & , z, = $&~ I, = &[(a3 - 3a2 + 6a - 6)e2a +
61 fh[JZ + log (1 + JZ,]h2 + &r2 11.22 Exercises (page 385) 1. (a) - 4(b) 4td 8td) 4~(4 3n12 650 Answers to exercises 2. 03.
n=34. -779. &'(x,y) = *[P2(x,y) + Q2(x,y)l'"' - *11.25 Exercises (page 391) 1. (b) 02. 0, 27r, -2573. As many as three4. As many as seven5 . (a) - 36. 2a 11.28 Exercises (page 399) cos 0, r sin 0)r dr] d0 2. I:;, [s,""" f(r cos 0, r sin 0)r dr] d0 where g(e) = l/(cos 0 + sin 0) 5. rl4s [I0 t&nesecef(r~os8,rsinO)rdr]dB + ~T~'4[~~ef(r~ ~ ~0,rsinB)rdr]de0 + j,",,4
[jot"" e Se' e f (r cos 0, r sin O)r dr] de 7. +a"[$ + log (1 + J2)] 10. 'f (r cos 0, r sin 0)r dr] d0 + jiy [/is" e f (r cos 0, r sin O)r dr]dO 11 . sr;'," [/f"'"f(r)rdr] de 12. I [I:I2 g:o, f(r cos EJ, r sin O)r dr1 de, where g(0) = I/(COS e + sin e) 13. JT [/~~~,,of(rcos 0,rsin O)rdr] d0 1 4 .
7741315. (a) u=~x y, V= 5x+Y (b) 6017. (a) 1 + 2~ Answers to exercises 651 18. (a) 4(u2 + u2) Cc) 0 19. ep [(p2 + r2j1 -P - ~~(l ~)] if p # 1 ; 57 log (1 + 9) if p = 1 . I@, r) tends to a finite limit whenp > 1 11.34 Exercises (page 413) 10.11.12.13.14.15. 18.19.20.22.23. 16n/3.lg &xa%(3a2 + 2h2)$,a3+(b3 - a3)$rR3(a2 + b2 + c2) 'h $T(SJS -
4)329Qx(b5 - a5)On the axis at distance Qh from the baseOn the axis at distance #h from the base 3 b4 - u424. On the axis of symmetry at distance 8 . b33 from the "cutting plane" of the hemispheres 25. P = J = Z = -&h (assuming the specified corner is at the origin)26. &M(u2 + 4h2)2 7 . $14R228. $W12 2 9 .
2% Chapter 12 12.4 Exercises (page 424) 1. b2b3 - a3b2j(x - x0) + (a3bl - alb3)(y - yo) + (alb2 - a,b&z - q) = 0 ; ar ar; x z = (@, - u,b,ji + (u,b, - a,b,jj + (a,b, - u,b,jk 2. x2/a2 + y2/b2 = z ; FU x z = -2bu2 cos vi - 2~1.4~ sin vj + ubuk 652 Answers to exercises sin u cos v sin u sin v cos llx - = abc sin u i+a b j+ck ar ar4 .
z = f(,/x2 + u2) ; z x a~ = -uf'(u) cos vi - f(u) sin vj + uk Zi ar5 . $'g=l; -&x;= b sin vj + a cos vk 6 . (Jx" + y2 - a)2 + z2 = b2 ; ar ara; x s = b(a + b cos u)(cos u sin vi + cos u cos vj + sin uk) sin2 u CO82 u sinh2 v s7. label cash u [(7 + 7 1cosh2 v + - c2 18.
$28~ ~ + 49 . Iu - VI J36u2v2 + 9(u + v)~ + 4 10. Ju" + z42 12.6 Exercises (page 429) 23 ' ;;y4,4: 4 a2 5. (a) A circular paraboloid(b) -2u2 cos vi - 2u2 sin vj + uk(c) n = 6 6 . ,/i =a2147 .
27rJ68 .
2ra2(3$ - 1)/39 . 4n2ab 11. (a) A unit circle in the xy-plane; a unit semicircle in the xz-plane, with z I; 0 ; a unitsemicircle in the plane x = y with z 5 0(b) The hemisphere x2 + y2 + z2 = 1, z 5 0(c) The sphere x2 + y2 + z2 = 1 except for the North Pole; the line joining the NorthPole and (x, y, I) intersects the xy-plane at (u, v, 0) 12.10 Exercises (page
436) 1.3 .4 .7 .8.9 . 10.11. 4?r/3f=ji=f=@ 90?TJZOn the axis of the cone, at a distance $a(1 - cos c~)/[l - cos (u/2)] from the center of thespherevu% + &ah3 12 . 2n/33?ra3h + +ah3 13.
-813 Answers to exercises 653 12.13 Exercises (page 442) 1. 02 . -77 3 . - 44 .
+ 12.15 Exercises (page 447) 1.
(a) div F(x, y, z) = 2x + 2y + 22 ; curl F(x, y, z) = 0(b) div F(x, y, z) = 0 ; curl F(x, y, z) = 2i + 4j + 6k(c) div F(x, y, z) = -x sin y ; curl F(x, y, z) = i + j(d) div P(x, y, z) = yeZv - x sin (xy) - 2xz sin (xzz) ; curl F(x, y, z) = z2 sin (xz2)j - [xFy + y sin (xy)]k(e) div F(x, y, z) = 2x sin y -I- 2y sin (xz) - xy sin z cos (cos z) ; curl F(x, y, z) = [x sin (cos z) - xy2 cos (xz)]i
-y sin (cos z)j + [y22 cos (xz) -x2 cos y]k 2. 04 . n = -35 . No such vector field 10. One such field is ~(x, y, z) = (xyzr211. div(V x r) =O; curl (V X r) = (c + 1)V13. 16(u + b) k12.17 Exercises (page 452) 1. (3x - 2z)j - xk is one such field2 . (x2/2 - xy - yz + z2/2)j + (x2/2 - xz)k is one such field3 . (x2y/2 + z2/2)j + Vf(x, y) for somefindependent of z 5. G(x, y,
z) = " * xzr(x2 -t y2)' - r(x2 + y2)j satisfies curl G = r4r at all points not on the z-axis 6. f(r) = CrF39 . F(x, y, z) = -i(z3i + 2j + yk) , G(x, y, z) = 6V(xay f r'z + z3x) 10. (c) 3?r/2 12.21 Exercises (page 462) 1. 3 3 . G 4 3 IV2 . (a) 14477 (b) 9 IV2 (b) -1677(c) 128~ 15. 87 r Chapter 13 13.4 ] (page 472) 2. A, u A, u A, = (A, n A; n A;) v (A, n
A$ u A3 ; kcl 4 = "ij' (4 n fi *? u *nk=l 1=k+l 3. (9 (ii) (iii) (iv) (4 (a) A' n B' A nB' AuB (A n B') u (A' n B) A' u B' (b) 500 200 500 300I 800 654 Answers to exercises 6.
a2 ={O,A,,A,,A,,uAz,A;,A;,A;,nA;,S)7. ~3~ ={~zI,A~,A,,A,,A,,UA,,A,,uA,,A,,uA,,A,,uA,,uA,,A;,A;,A;,A;,nA4,,A;,nAj,A;,nA;,A;,nA;,nA;,S} (ifn>3) 13.7 Exercises (page 477) 1. A E B'2. x~A'nB'nc'3 . x~AnB'nc'4. XEAUBUC5. XE(A n B' n C') u (A' n B n C') u (A' n B' n C)6. xE(A'nB')u(B'nC')u(A'nC')7. XE(A nB) u (A n C) u(B n C)8
xE(AnBnC')U(AnB'nC)U(A'nBnC)9. xE(A nBnC)' 10. XEA nC nB'11. xeAnBnC1 2 . XEAUBUC15. (a) 1 - a (d) 1 - c (b) 1 - b (e) 1 - a+cCc) a+b-c (f) a - c 13.9 Exercises (page 479) 2. (a) A 5. &(b) 3 6.
&(4 tl 7 . 27 *(4 $I 8. (a) A/(A + B) 3. (a) $g (b) B/(A + B)0) % (c) (C + l)/(C + D + 1)(cl * (d) C/(C -t- D + 1)(4 2ii 9. (a) 8(4 QQ @I it(f) s Cc) # 4. 4 (4 010. P,=l P(A) P(B)+P(A nB), P, = P(A nB), P, = P(A) + P(B) - 2P(A n B), Pz =P(A n B)12. (a) 5 to 9 (b) 45 to 46(c) 10 to 81(d) 36 to 55 13.11 Exercises (page 485) ; 3: ;(31;62), (1,3), (2, l), (2,3), (3, 11, (332)) 544. {H, T}
x {H, T} x {1,2, 3,4, 5,6} ; 24 outcomes5. 52!/(13!)46.
36 (4 18(b) 12(4 24 Answers to exercises 6 5 5 7 . (a) 13 . 12, 11 . 72 = 123552 (not including triplets or quadruplets)(b) 5148(c) 36 (not including IOJQKA)(4 4 8. (4 4(:3)/(y) (b) 36/(y) Cc) 4/(y) 2.98! 98!g. Cal (49 (b) ___48!. 50! lo. (zy(:"o")11. 1612. nk 13.14 Exercises (page 490) 2. (a) P(A) = &; P(B 1 A) = & ; P(A n B) = 3% 9850) 26! .34!6. (a) 1 -m=l
. . (b) 1 _ (3 + C)39 (>( ) C3(256) (39 >(39 1 3 5 1 3 1 9. 41 5 . (a) P(A) = P(B) = P(C) = 4 ; P(A n B) = P(A n C) = P(B n c) = & ; P(A n B n C) = 0 13.18 Exercises (page 499) 1. (a) PW, HI =plp2 ; PW, T) =pl(l - p2) ; P(T, ff) = (1 - pl)pz ;P(T, n = (1 -p&l -pz)(b) Yes(4 No(d) HI and Hz, HI and T2 , Hz and TI, T, and T2 2.
(a) #(b) i+&(4 6 656 Answers to exercises 3. 10 5' 390625(1 - 3 610 2519424 4. (a) +j(b) 4(4 & 5 . (a) (5!)2/10! = 2$y(b) !I 6.
(a) 36p'O - Sop9 + 4Sp8(b) i& 7. It is advantageous to bet even money 15. (a> f(p) = (1 -pj2 +p3 W (6.
-4)/3 13.20 Exercises (page 504) 8 175 9938999 9 . 03 iiF= 1377495072 10 . 1835 1 2 11. 1 - (19/20)10 = 0.401312. +s 14.
59 In I; 65 and f(k) = 2kf(k) = 3"f(k) = pk , wherep, is the kth prime 22one such function is f(k) = (g(k), h(k)), where g(k) =m2(k) + 3m(k) -k+2, h(k)=k m2(k) + m(k) 2 2 ' m(k) =&CT l 2 1where [x] denotes the greatest integer IX(e) f(k) = 2g(k)3h(k), whereg(k) and h(k) are as defined in part (d) 13.22 Exercises (page 507) 1 . n=O: m a x = l , min =
4n=l: max = $, min = 0n =2: m a x =&, min = 0n =3: max = A, min = 0 3. (a) 1 - qp3 - pq3 4 . (a) 3pqKpq + 2)0) % (b) 4Cc) % (c) 2 log 2 - 1 13.23 Miscellaneous exercises on probability (page 507) 1. T'12. (a) lig W i%3. (4 t @I I - 4. (a) $4(b) %f(cl 1i:5 5 . -gi$ 7 .
0.65 8. (a) 3(b) 6(4 8Cd) No 9 . p3 + 1 -pa 6p (12 10. np(1 - p)n l + npn l(1 - p) Answers to exercises 657 Chapter 14 14.4 Exercises (page 513) 1. (b) Xlb2. (a) (0 I X(w) E (a, bl, Y(w) E Cc, 41 (4 XSa, Y5d(d) P(a < X 5 b, c < Y I d) = P(X I 6, Y I; d) - P(X 5 a, Y 5 d) - P(XS b, Y I; c)+ P(X I a, Y .S c) 3. (a) ((1, 6), (2,5>, (3,4), (4,3), (5,2), (6, l>>, {(5,6),
(6,5>>. {(1,6), (2,% (3,4), (4, 3),(5,2), (6, l), (596)s (69%)(b) P(X = 7) = $; P(X = 11) = & ; P(X = 7 or X = 11) = $ 4 . Y = Xl + x, + x, + x, ; P(Y=O)=&; P(Y=l)=i; P(YlOO6. (a) Z = Y - 1 (b) U = Y, + Y, - 1 14.8 Exercises (page 523) 2 . t 1 2 3 4 5 6 7 8 9 1 0 1 1 1 2 3.
(b) p(-2) = 4, p(0) =p(2) = t(4 0. t,;, t, 190, a, a 5. (a) c = *(b) 6,6,8, !(c) F(t)=Ofort n 9. p(k) = c -'z , k =0,1,2,3 ,...; c 20 p(t) = 0 for t # 0, 1,2, 3, . , .10. (a) px(t) = 4 at t = -1 and t = +l ; px(t) = 0 elsewhere F,(t) = 0 for t < -1 ; F,(t) = g for - 1 I t < 1 ; F,(r) = 1 for t 2 11 1 . P(A)=Q; P(B)=Q; P(Ar,B)=Q; P(BIA)=&; JJ(AUB)=l 658 Answers to exercises
14.12 Exercises (page 532) 1. (a) c = 1 ; f(t) = 1 if 0 5 t 5 1 ; f(t) = 0 otherwise(b) O,Q,h 2 . c=$; F(t)=Oift < -r/2; F(t)=&costif 7r/2 a/2 3 . c=;; F(l) = 0 if t < 0 ; F(t) = $(3t2 - t3) if 0 5 t 5 2; F(t) = 1 i f t >24 . (a> S (b) iCc) i 5 . (a) f(t)=Oift<$; f(t)=& liff4 (b) F(t)=Oift <$; F(t) = 4t2 - t if & 5 t < 4 ; F(t) = -4t2 + 7t - 2 if 4 5 t < 2;F(t) = 1 if t 2 2 cc>
1,1,6,O,i%6. (a) 0,9,$,* (b) t = 1 7 . (a> 1k 5 1 0 1 5 20 25 30 (b) Let each Styx train arrive 5 minutes before a Lethe train8 . Fy (t) = 0 if I < b ; F,(t) = (t - b)/a if b < I < b + a ; Fy(t) = 1 if t > b + a9. (a> 3% (b) ++(4 $8 t - b10. F(t) = i + 1 arctan a ?T 14.16 Exercises (page 540) 1. (a) 105(b) 10.05 2 . (a) *(JZ - 1) 0) a(2 - $9(cl m - 42)(4 A(4 - 4% 3 . (a) 1
- e l(b) em2Ez F3 - O/e2 4 . F(t) = 0 if t < c ; F(t) = 1 - I+(~ ~) if I ;r c6 . 1'=t/a, c'=b+ac9 .
(a) 0.5000 (b) 0.1359(c) 0.9974(d) 0.0456 10.
(a) 0.675(b) 0.025 11. (a) 0.6750(b) 0.025~ Answers to exercises 659 12. (a) 0.8185(b) 0.8400 13. 75.98 inches14. mean = b, variance = a215. Fy(t) = Cl if t < 0 ; fy(t) = 0 if t < 0 ; fy(t) = c t/2/J2nt if t > 0 14.18 Exercises (page 542) (a) Fy(t) = 0 if t < 1 ; F,(t)=(t-1)/3ifl 4;f,(t)=+if1 < t < 4 ; f=(t) = 0 otherwise (b) Fy(t) = 0 if t < -2 ; F=(t) = (t + 2)/3 if
-2 I t I 1 ; Fy(t) = 1 if t > 1 ;fu(t) = 4 if -2 < t 5 1 ; fy(t) = 0 otherwise (c) F,(t)=Oift 1 ; ,fv(t) = (2t) !" if0 5 t 5 1 ; fy(t) = 0 otherwise (d) Fy(t) = et if t 5 0 ; Fy(t) = 1 if t > 0 ; fu(t) = et if t 5 0 ; fu(t) = 0 if t > 0(e) Fu(t) = et/2 if t < 0 ; Fy(t) = 1 if t > 0 ; fu(t) = &t/2 if t 5 0 ; fy(t) = 0 if t > 0(f) F=(t) = 0 if t < 1 ; Fy(t) =logt if 1 5 t se; Fy(t) = 1 if t > e; &(t)
= l/t if 1 < t < e ; fu(t) = 0 otherwiseLet w be the inverse of p, defined on the open interval (a, b).Then Fy(t) = 0 if t < (I; FY(t) = F,[y(t)l if a < t < b ;fx[y(t>]y'(t) if a < t < b ; fu(t) = 0 otherwise Fy(t) = 1 if t 2 b ; f&(t)= (a) fu(t) = 0 if t < 0 ; ,fu(t) = (2xt) -%e t/2 if t > 0(b) &(t) = 0 if t < 0; ,fu(t) = 4t(2v) b6e t"/2 if t 2 0(C) &(t) = 0 if t < 0 ; ,fu(t) =
(257t2) -lmc (10gt)2/2 if t > 0 (4 fy(t> = (277 'm sec2 t c (tan2 t'/2 if It) < 77/2 ; ,fu(t) = 0 if If (2 n/2 14.22 Exercises (page 548) 2. (a) P(X = x1) = P(X = x2) = P(Y = yl) = p(y = 4'2) = &cp + q)(b) p = q = 4 3.
(a) F(x ,y) = (z)($z) ifa Ix Ib and c ly Id, x aF(x,y) = bifa I x Ib a n d y >d, F(x,y) ='ez i f x > b a n d c b and y > d, F(x, y) = 0 otherwise(b) F~(x) = (X - a)/(b - a> if a 5 x 5 b ; F,(x) = 0 if x < a ; F,(x) = 1 if x > b ; Fy(y)=0, c)l(d c)ifcjyId; F,(y)=Oifyd(c) X and Y are independent 5 . P(Y=O)=i; P(Y=l)=P(Y=2)=)7 . +,g8. ; 9.
(b) fX(x) =$if(x)EQ; f&y) =Oif(x,y)$Q;f-(x) = 1 - 1x1 if 1x1 < 1 ; ,fX(x) = 0 if 1x1 > 1 ;f-(y) = 1 - lyl if lyl I 1 ; fY(y) = 0 if lyl > 1.X and Y are not independent 10. g(u, u) =f(u + a, u + b) 14.24 Exercises (page 553) 1. (b) fv(o) = 1 + v if -1 5 v < 0 ; f&u) = 1 - v if 0 5 o 5 1 ; f&u) = 0 if 1~1 > 1(c) U and V are not independent 660 Answers to exercises 2 .
(b) fv(t) = 2 - 2t if 0 I. t < 1 ; fv(t) = 0 otherwise(c) U and V are independent 3 . (b) g(u. v) = UP if II > 0, 0 < u < 1 ; g(u, v) = 0 otherwise(Cl fv(u) = ue " if u > 0 ; &(u) = 0 if u < 0(d) f"(u) = 1 if 0 < u < 1 ; f"(u) = 0 otherwise 4 . (b) g(u, v) = (v/2~)e (1+U*)~*~2 if u 2 0(4 f&4 = [a(1 + UT' 5. fi(t) = n xemt* 6. (b) f&) = 0 if u < 0 ; f&u) = (u/u2) exp
F&t)=Oiff < O ; FU(t)=l - e x p 14.27 Exercises (page 560) 1 . E(X)=;, Var(X)=g7. (a) E(X) = Var (X) = A (b) None(c) E(X) = l/A, Var (X) = l/A2(d) E(X) = m, Var (X) = a2 8. (a) C(r) = (r - 1)/2(b) Fx(t) = 4 ItI' 'if t < -1 ; F'(t) = 4 if -1 5 t 5: 1 ; F,(t) = 1 ] > 1(c) P(X < 5) = 1 - 5? l/2 ; P(5 < X < 10) = (51~' - l(W)/2(d) X has a finite expectation
when r > 2 ; E(X) = Cl (e) Variance is finite for r > 3 ; Var (X) = (r - l)/(r - 3)9. E(X) = E(Y) = A ; E(Z) = -1767/50653 ; E(X + Y + Z) = 4505/50653 10 . E (X) + a, asn+ co12 . (a) (2/m)% (b) ex 14.31 Exercises (page 568) 4 . 2 5 15 . 06 . Chebyshev's inequality gives 4; tables give 0.00278 . (b) 0.68269 . (b) 0.0796 10. (a) 0.0090(b) 0.0179 Chapter
1515.5 Exercises (page 577) 2. (a) No(b) Yes(c) Yes(4 No(e) Yes(0 No(g, Y e s 3.
(a) Neither(b) Seminorm(c) Seminorm(d) Neither(e) Seminorm(f) N o r m(g) Neither(h) Neither Answers to exercises 6 6 1 6. (4, W, 6)8. (b) The polynomial in (a) plus g(Sx3 - 3x) JL1 (St3 - 3t)f(t) dt + &(35x4 - 30x2 f 3) x 51, (3V - 30t2 + 3)f(t) dt9. a = -3e/4 + 33/(4e), b = 3/e, c = +f(e - 7/e) l&&l + 14x2 - 7x4):f: (b) Ilf - P1l2 = &@ n - l log2 n12. (a)
IlP -fl12 = - - II n - l 1 2 6(n + 1)(n - (n 1% n 4(n3 - 1) log n 6(n + 1)(n _1)4 -(n - IIP -flj" = 36 log 2 - 28 log2 2 - 2; = 0.0007 when n = 2, 1 3 (a) I/P -Sir2 = k [(n - 2)e2n + 4en - n - 21 (b) P(x) = (18 - 6e)x + 4e - 10; I/P -fjl" = 20e - Ge2 - '?$ = 0.0038 14. (b) Vk+l(x) = J(2k ; :"1" + 3)xPk(x) - &JS PkVI(x), where qk = pk/IIpkll 15. (b) ,$ f';(x) =pjf P',
(xP~+,(x) - P,+,(x)P;(~)]k=O 15.9 Exercises (page 585) 1. (a) P(x) = $(x2 + 13x + 12)(b) P(x) = 4(x" - 5x + 6)(c) P(x) = -+(x3 - 6x2 + 5x - 6)(d) P(x) = 2x3 + x2 - x - 2(e) P(x) = -5x3 - x2 + 10x - 5 2.
P(x) = &(9x" - 196x2 + 640)4. (4 (b)5.
(a> (b) 197. (4 (b)8. 0) (412. (a) @I13. (d) 16. x Q(x) = 2x3 + 3x2 - x - 3Q(x) = 4x3 + 7x2 - 3x - 7P(32) = g; f(32) - P(32) = gP(32) = $+ ; f(32) - P(32) = -zP(32) = g; f(32) - P(32) = -&P(32) = -$& ; f(32) - P(32) = ++&+L,(x) = &(u - l)(u - 3)(u - 4)(u - 6) ;L,(x) = &(u - l)(u k(x) = -&u(u - 3)(u - 4)(u - 6) ; -4)(u - 6) ; L3(x) = +(u - l)(u - 3)(u - 6) ; L4(x) = &u(u -
l)(u - 3)(u - 4)P(2.6) = 20x 2 1.581h 5 0.0006a=O, b=lc = 1, d = -2L,(xk)Let B,(X) = 1 and let B,(X) = (x - x0)(x - x0 - nh)n l/n! for n 2 1 ; the one andonly polynomial P of degree In satisfying the conditions P(x,) = co, P'(xl) = cl,P"(x2) =c2,..., P(")(x,) = c, is given by P(x) = c,B,(x) + . .
+ C,&(X) 662 Answers to exercises 15.13 Exercises (page 593) 4 . (b) 8 -5040 13068 -13132 6769 -1960 322 - 2 8 1 ~ ~ ________ 9 40320 - 109584 118124 -67284 22449 - 4536 546 - 3 6 1 ________ ______ ~ 1 0 -3628800 1026576 -1172700 723680 -269325 63273 -9450 870 - 4 5 1 (c) 1 + 2x + 2x2 - 3x3 + x4 5 . (c) 8 1 1 2 7 ~ 9 1 255~~ 1 0 1 5 1 1 . - 966
1050 3025 6951 9330 1 34105 1 42525 266 28 1 2646 462 36 1 22827 5880 750 45 (d) - 1 + 6 x1(1 + 16,#) + gx(3) + $4?7. (a) 4n3 + J -tn2 + S?n (b) in4 + in3 + $,"z + ;n(c) &z" + $2" + +l" + ;n(d) +" + +n4 + +n3 - &I 15.18 Exercises (page 600) 2 . (b) T,(l) =n2, T;(1) = (l)n $22 1 -x25. sin 8 sin n0 = n TX 4 ; degree = 12 + 1 7. Q(x) = x"+l - 2
nT,+1(x)8. Q(x) = -++ + fjx3 _ &" + iZ556x _ ,+, 14. (b) U,,(x) = 1, o;(x) = 2 x , U,(x) = 4x2 - 1 , U,(x) = 8x3 - 4x,U,(x) = 16x4 - 12x2 + 1, U,(x) = 32x5 - 32x3 + 6x 15.21 Exercises (page 610) 1 . (a) 0.693773 - where 0.000208 5 E 5 0.001667.
This4, gives the inequalities0.6921 < log2 < 0.6936 (b) n = 5782. (a) c = $13 0)a+b b ad3c , = - + - - a+b b a@ 2 2 3, c2=2 - 2 3 3. (a) c = JS/2 (b)a+b b u@ a+b b u@ c l = - + - -) c,= - 2 2 2 2 2 2 4 . u=2+$, b=2 $5. c = & 1 Answers to exercises 663 10. (a) log 2 = 0.693254 - E, where 0.000016 5 E < 0.000521 ; this leads to the inequalities0.69273
< log 2 < 0.69324 (b) log 2 = 0.69315023 - E , where 0.00000041 5 E 5 0.00001334 ; this leads to the in equalities 0.693136 < log 2 < 0.693149 11. (d) log 2 = 0.693750 - E, where 0.000115 < E 5 0.003704 ; this leads to the inequalities0.69004 < log 2 < 0.69364 INDEX ABEL, NIELS HENRIK, 162Abel's formula for Wronskian determinants, 1 6
2Absolute maximum, 304Absolute minimum, 304ADAMS, JOHN COUCH, 571Adjoint of a matrix, 122Angles in a real Euclidean space, 18Annihilators, 151 method of, 163table of, 166 Approximate integration :by Cotes' rule, 612 (Exercise 11)by Euler's summation formula, 613, 615by Simpson's rule, 608, 609by the trapezoidal rule, 604
Approximations :by interpolation polynomials, 579by Legendre polynomials, 29by Taylor polynomials, 576by trigonometric polynomials, 29in a Euclidean space, 28successive, 223 ARCHIMEDES, 571Area : as a line integral, 383of a plane region, 368of a surface, 424 Area cosine principle, 426At random, 478Attraction, gravitational, 335Average
density, 374Average of a function, 374Average rate of change, 253Axioms : for a determinant function, 73for a linear space, 3for an inner product, 14for probability, 474, 506, 511 Ball in n-space, 244Basis, 12Bayes' formulas, 549 (Exercise 6)BERNOULLI, DANIEL, 182BERNOULLI, JAKOB, 469, 495Bernoulli trials, 495BESSEL, FRIEDRICH
WILHELM, 182Bessel differential equation, 182Bessel functions : graphs of, 186of first kind, 186of second kind, 188 Binomial coefficients, 481Binomial distribution, 521 approximation by normal, 538Boolean algebra of sets, 471BOREL, SMILE, 510Bore1 set, 510Boundary of a set, 245Boundary point, 245Bounded function, 357Boundedness of
continuous functions, 319Bridge hands, 481 CAQUh, J., 223Cartesian product, 244,CAUCHY, AUGUSTIN LOUIS, 16, 142, 191, 530Cauchy distribution, 530Cauchy functional equation, 528Cauchy-Schwarz inequality, 16CAYLEY, ARTHUR, 202Cayley-Hamilton theorem, 203Center of mass, 373, 374,431Central limit property, 567Central limit theorem,
566Centroid, 374Chain rule for derivatives : of matrix functions, 194of scalar fields, 264, 273of vector fields, 272 665 666 Index Change of variable:in a double integral, 394in a line integral, 327in an n-fold integral, 408in a surface integral, 433 Characteristic equation, 150Characteristic polynomial, 103, 149CHEBYSHEV, PAFNUTI LIWOWICH, 563,
596, 611Chebyshev inequality, 563Chebyshev integration formulas, 611Chebyshev polynomials, 596Christoffel numbers, 610 (Exercise 6)Circulation, 330, 462Clockwise (negative) direction, 389Closed curve, 334, 379Closed set, 246Closure axioms, 3Coefficient matrix, 58Cofactor, 87Cofactor matrix, 92Coin-tossing games, 506, 558Column matrix
(column vector), 51Combinatorial analysis, 481Complement of a set, 246Complex Euclidean space, 15Complex linear space, 4Components, 13Composite functions, 37, 194, 250 continuity of, 250differentiation of, 194, 264, 272 Composition of transformations, 37Compound experiments, 492Conditional probability, 486Connected set, 332Conservative
field, 329Constant-coefficient derivative operators. 148Constant-coefficient difference operators, 595 (Exercise 8)Content zero, 364, 406Continuity of scalar and vector fields, 247Continuous distribution, 525Continuous random variable, 525Continuously differentiable scalar field, 261Contraction constant, 235Contraction operator, 234Convex
combination, 377Convex set, 344Coordinates : cylindrical, 409polar, 274, 397spherical, 293, 410, 414 COTES, ROGER , 612 (Exercise 11)Cotes' rule, 612 (Exercise 11)Countable set, 502Countably additive set function, 506Countably infinite set, 502 Counterclockwise (positive) direction, 389CRAMER, GABRIEL, 93Cramer's rule, 93CRAMfR,
HARALD, 520Curl of a vector field, 441Curve : closed, 334, 379in n-space, 324Jordan, 379piecewise smooth, 324 Cylindrical coordinates, 409 D'ALEMBERT, JEAN, 288Del operator V, 259DE MOIVRE, ABRAHAM, 469,567Density function, 525 for Cauchy distributions, 531for exponential distributions, 533for normal distributions, 535for uniform
distributions, 527 Dependent sets, 9Derivative : and continuity, 260directional, 254of a matrix function, 193of a scalar field with respect to a vector, 253of a vector field with respect to a vector, 270partial, 254total, 258, 270 Determinants, 71axiomatic definition of, 73differentiation of, 80 (Exercise 6)expansion formulas for, 86 Diagonal matrix, 48,
96Diagonal quadratic form, 127Diagonalizing matrix, 112Difference of two sets, 246Difference operator A, 590Differentiable scalar field, 258Differentiable vector field, 270Differential equations, 142 first-order linear, 143homogeneous, 146nth-order linear, 145partial, 283power-series solutions of, 169, 220systems of, 191 Dimension of a linear space,
12Dimensionality theorem, 147DIRAC , PAUL A.
M . , 108 (Exercise 8)Directional derivative, 254Disconnected set, 333Discontinuity, 366, 517Discrete distribution, 521Discrete random variable, 521 Index 667 Discriminant of a quadratic form, 134Disk, 244Distance in a Euclidean space, 26Distribution function: binomial, 521Cauchy, 530continuous, 525discrete, 521exponential, 533joint, 543mixed,
539normal, 535normal bivariate, 555 (Exercise 9)of a function of a random variable, 541, 550of a one-dimensional random variable, 515of a two-dimensional random variable, 543Poisson, 522standard normal, 537uniform, 526 Divergence of a vector field, 441, 460Divergence theorem, 457Dot product, 14Double integral(s) : applications of,
374evaluation by repeated integration, 358, 367existence of, 358, 363in polar coordinates, 397of a bounded function, 357of a step function, 355transformation formula for, 394 Doubly connected region, 391 (Exercise 3) Eigenfunction, 99Eigenspace, 98Eigenvalue, 97Eigenvector, 97Element of a linear space, 3Entries of a matrix, 45, 51Equipotential
lines and surfaces, 335Equivalence of sets, 501Error : in Lagrange's interpolation formula, 583in linear interpolation, 584in Taylor polynomial approximation, 573in Taylor's formula, 258, 270 Euclidean space, 15EULER, LEONARD, 142, 182, 613Euler's constant, 618 (Exercise 2)Euler's gamma function, 184, 413,620 (Exer- cises 7, 9)Euler's
summation formula, 613, 615Euler's theorem for homogeneous functions, 287Events, 476Exact differential equation, 347 Existence theorem(s) :for determinants, 90for ftrst order linear differential equations, 1 4 3for first order linear systems of differential equations, 213, 219, 220for first-order nonlinear systems of differ- ential equations, 229for
implicit functions, 237for integral equations, 239for interpolation polynomials, 580for linear nth order differential equations, 1 4 7forlinearsecond orderdifferentialequations, 1 4 3for potential functions, 339for Taylor polynomials, 572for vector fields with prescribed curl, 448 Expectation (expected value) :of a function of a random variable, 559of a
random variable, 556 Exponential distribution, 533Exponential matrix, 197Exterior: of a Jordan curve, 380of a set, 245 Exterior point, 245Extreme-value theorem for continuous scalar fields, 321Extremum, 304 second-derivative test for, 312with constraints, 314 Factorial polynomials (factorial nth powers),592 FELLER, WILLIAM, 568FERMAT,
PIERRE DE, 469Finite additivity, 470Finite sample space, 473Finite set, 502Finitely additive measure, 472Finitely additive set function, 470Fixed point of an operator, 233Fixed-point theoremforcontractionoperators, 235applications of, 237 Flow integral, 330Flux density, 432, 461Forward difference operator A, 590FOURIER, JOSEPH, 29Fourier
coefficients, 29FRiCHET, RENk MAURICE, 258FROBENIUS, GEORG, 181Frobenius' method, 180FUCHS, LAZARUS, 223 668 Index Function space, 5Functional equation : for the gamma function, 185of Cauchy, 528 Functions of random variables, 541, 550Fundamental theorems for line integrals: first, 338second, 334 Fundamental vector product of
a surface, 420 GALLE, JOHANN, 571Gamma function, 184, 413, 620 (Exercises 7,9)GAUSS, KARL FRIEDRICH, 61, 78, 457Gauss-Jordan process : for calculating determinants, 78for solving linear equations, 61 Gauss' theorem (divergence theorem), 457General solution : of a linear differential equation, 148, 156of a system of linear equations, 60
GOMBAUD , ANTOINE (Chevalier de Mere), 469Gradient, 259GRAM, J0RGEN PEDERSEN, 22Gram-Schmidt process, 22GREEN, GEORGE, 379Green's formula, 387 (Exercise 8)Green's theorem : for multiply connected regions, 387for simply connected regions, 380 HADAMARD, JACQUES, 69Hadamard matrices, 69 (Exercise 10)HAMILTON,
WILLIAM ROWAN, 202Harmonic function, 445Heat equation, 292 (Exercise 1)HERMITE, CHARLES, 15, 114, 122, 178Hermite differential equation, 178 (Exercise 4)Hermitian matrix, 122Hermitian operator, 115Hermitian symmetry, 15, 114HERSCHEL, WILLIAM, 571HESSE, LUDWIG OTTO, 308Hessian matrix, 308Homogeneous function,
287Homogeneous linear differential equation, 146Homogeneous system: of differential equations, 200of linear equations, 59 HUYGENS, CHRISTIAN, 469 Identity matrix, 54Identity transformation, 32 Implicit differentiation, 294Implicit function theorem, 237Independence : in a linear space, 9of eigenvectors, 100of events, 488of nonzero orthogonal
elements, 18of parametrization, 327, 434of path, 333of random variables, 547 Independent trials, 495Indicial equation, 183Inequality: Cauchy-Schwarz, 16Chebyshev, 563triangle, 17Wallis', 616 Infinite set, 502Initial conditions, 143, 147Initial-value problem, 193Inner product, 14Integral : double, 357line, 324n-fold, 406of a matrix function,
193surface, 430triple, 406 Integral equation, 239Integrating factor, 349Interior (inner region) of a Jordan curve, 380Interior of a set, 244Interior point, 244Interpolation : by Lagrange's formula, 580by Newton's formula, 589 Interval in n-space, 405Invariance property : of Lagrange interpolation coefficients, 586(Exercise 6) of line integrals under
change of parame-trization, 327 of line integrals under deformation of thepath, 388 of surface integrals under change of parame-trization, 433 Invariant subspace, 99Inverse function, 39Inverse matrix, 66Inverse transformation, 39, 41Invertible function, 40Invertible transformation, 40, 41Irrotational vector field, 445Isometry, 129Isomorphism,
53Isothermals, 266 Index 669 Iterated integrals, 359, 367, 406Iterated limits, 251 (Exercise 2) JACOBI, CARL GUSTAV JACOB, 271Jacobian determinant, 297, 394, 408, 420Jacobian matrix, 271Joint probability density, 546Joint probability distribution, 543Jointly distributed random variables, 543JORDAN, CAMILLE, 61, 78, 379Jordan curve,
379Jordan curve theorem, 380Jump discontinuity, 518 Kernel :of a linear transformation, 33of an integral equation, 239 Kinetic energy, 329KOLMOGOROV, ANDRE1 NIKOLAEVICH, 470 LAGRANGE, JOSEPH LOUIS, 142, 315, 580Lagrange interpolation coefficient, 209, 581Lagrange interpolation formula, 580Lagrange multipliers, 3 15LAPLACE,
PIERRE SIMON, 142Laplace's equation, 283, 292Laplacian, 292, 444Law of large numbers, 565, 566Least-square approximation, 573LEGENDRE, ADRIEN MARIE, 25, 171Legendre polynomials, 25, 174 graphs of, 175normalized, 26Rodrigues' formula for, 176zeros of, 180 (Exercise 14) Legendre's differential equation, 171LEIBNIZ, GOTTFRIED
WILHELM, 142Level sets (curves, surfaces), 266LEVERRIER, JEAN JOSEPH, 571LINDEBERG, JARE W., 567Lindeberg's condition, 568Line integrals : applications of, 330first fundamenta1 theorem for, 338independence of parametrization, 327independence of path, 333notations for, 324second fundamental theorem for, 334 Linear combination,
8Linear differential equation, 145Linear differential operator, 146Linear interpolation, 584 Linear operator (transformation), 31Linear space (vector space), 3Linear subspace, 8Linear system, 58, 192Linear transformation, 31LIOUVILLE, JOSEPH, 143, 191,223LIPSCHITZ, RUDOLF, 229Lipschitz condition, 229Lorentz transformation, 69 (Exercise
6)Lower integral, 358LYAPUNOV, ALEXSANDR MIKHAILOVICH, 567 Mapping, 31, 393, 550MARKOV, ANDRE1 ANDREEVICH, 470Mass, 330, 373, 414, 431Mass interpretation of probability, 511, 5 4 5Matrix : adjoint, 122adjugate, 92algebraic operations on, 52, 54block-diagonal, 84cofactor, 92conjugate, 122definition of, 46diagonal, 48,
96exponential, 197Hermitian, 122minor of, 87nonsingular, 65orthogonal, 123representation, 46self-adjoint, 122series, 194singular, 67skew-Hermitian, 122skew-symmetric, 124symmetric, 124trace of, 106transpose of, 91unitary, 123 Max norm, 575Maximum of a function : absolute, 304existence of, 321relative, 304tests for, 311, 312 Mean :of a
function, 374of a random variable, 535, 556 Mean-value theorem for scalar fields, 254Measurable sets, 510Measure : finitely additive, 472probability, 474, 506 670 Index Minimum of a function :absolute, 304existence of, 321relative, 304tests for, 311, 312 Minor of a matrix, 87MISES, RICHARD VON, 470MSBIUS, AUGUSTUS FERDINAND,
455Mobius band, 455Moment of inertia, 331, 374, 414, 431Multiple integrals: applications of, 373, 414, 431definition of, 357, 406evaluation of, 358, 406existence of, 363transformation formulas for, 394, 408 Multiply connected region, 384 NAPIER, JOHN, 572Negative definite quadratic form, 310Negative (clockwise) direction, 389Neighborhood,
244Neptune, 571NEWTON, ISAAC, 142, 335, 571, 589Newton's interpolation formula, 589Newtonian potential, 335n-fold integral, 405Nonlinear differential equation, 228Nonorientable surface, 456Nonsingular matrix, 65Norm : in a Euclidean space, 17in a linear space, 233max, 575of a matrix, 195square, 575 Normal approximation to binomial dis-
tribution, 538 Normal bivariate distribution, 555 (Exercise 9)Normal derivative, 386Normal distribution, 535Normal vector to a surface, 267, 423Norrned linear space, 233, 574Null space, 33Nullity, 34Numerical integration : by Chebshev formulas, 611by Cotes' rule, 612 (Exercise 11)by Euler's summation formula, 613, 615by Simpson's rule, 608,
609by the trapezoidal rule, 604 Odd man out, 508 (Exercises 10, 11)Odds, 480 One-element set (singleton), 475One-to-one correspondence, 501One-to-one function, 40Open ball, 244Open set, 244Operator: contraction, 235curl, 441derivative, 32divergence, 441forward difference, 590gradient, 259identity, 32integration, 32Laplacian, 292linear, 31
Ordered basis, 13Ordinate set, 360, 368Orientable surface, 456Orientation of Jordan curve, 389Orthogonal basis, 19Orthogonal complement, 27Orthogonal decomposition theorem, 27Orthogonal elements, 18Orthogonal functions, 18Orthogonal matrix, 123Orthogonal operator, 138Orthogonal projection, 28Orthogonal trajectory,
267Orthogonalization theorem, 22Orthonormal set, 18Outcome, 476 PAPPUS OF ALEXANDRIA, 376, 426Pappus' theorems, 376, 428Parallel-axis theorem, 378 (Exercise 17)Parametric surface, 420PARSEVAL, MARK-ANTOINE, 20Parseval's formula, 20Partial derivative, 254, 255Partial differential equation, 283Partition of a rectangle, 354PASCAL,
BLAISE, 469Path in n-space, 323PAULI, W O L F G A N G , 107 (Exercise 4)Pauli spin matrices, 107 (Exercise 4)PEANO, GUISEPPE, 223, 611Petersburg problem, 562 (Exercise 10)PICARD, C H A R L E S SMILE, 223Picard's method, 223, 227Piecewise linear interpolation, 603Piecewise smooth path, 324POISSON, SIMiON DENIS, 522Poisson
distribution, 522Polar coordinates, 274, 397 Index 671 Polar moment of inertia, 375Polynomials : Chebyshev, 596interpolation, 579, 589Legendre, 26, 176, 577Taylor, 572 Positive definite quadratic form, 310Positive (counterclockwise) direction, 390Potential energy, 336Potential function, 335 construction of, 338, 342existence of, 339Newtonian,
335on convex sets, 344 Power series solutions :of linear differential equations, 169of homogeneous linear systems, 220 Probability :conditional, 487definition of, 474, 506, 511 Probability density function, 525, 546Probability distribution, 515, 543Probability mass function, 520, 545Probability measure, 474Probability space, 474Projections, 24,
28PTOMELY, CLAUDIUS, 57 1PUTZER, E. J., 202Putzer's method, 205Pythagorean formula, 10, 20, 27 Quadratic form, 127, 310extreme values of, 137negative definite, 310positive definite, 310reduction to diagonal form, 129 Radius of gyration, 557Random variable(s), 512 continuous, 525discrete, 521, 545functions of, 541independent, 547jointly
distributed, 543 Range of a linear transformation, 32Rank, 34Rate of change, 253Rational function, 249Real linear space, 4Regular point of a surface, 421Regular 'singular point, 18iRelative maximum. 304Relative minimum,' 304Repeated experiments, 492 Repeated integrals, 359, 406Repeated trials, 495RICATTI, VINCENZO, 142Ricatti's differential
equation, 142RIEMANN, GEORG FRIEDRICH BERNHARD, 539, 619Riemann zeta function, 619 (Exercise 3)RODRIGUES, OLINDE, 176Rodrigues' formula, 176Rotation, 129Roulette, 558 Saddle point, 304Sample, 476Sample space, 473Sampling, 483, 484Saturn, 571Scalar, 4Scalar field, 243 total derivative of, 258SCHMIDT, ERHARD, 22SCHWARZ,
HERMANN AMANDUS, 16Second-derivative test for extrema, 311, 312Self-adjoint matrix, 122Seminorm, 574 interpolation, 575Taylor, 574 Sequential counting, principle of, 483Series of matrices, 194Series solutions : of homogeneous linear systems, 220of linear differential equations, 169 Set function, 470Similar matrices, 111Simple closed curve,
379Simple parametric surface, 420Simply connected set, 384SIMPSON, THOMAS, 608Simpson's rule, 608, 609Singleton (one-element set), 475Singular point: of a differential equation, 146, 181of a mapping, 396of a surface, 421 Skew-Hermitian matrix, 122Skew-Hermitian operator, 115Skew-symmetric matrix, 124Skew-symmetric operator,
115Smooth path, 323Smooth surface, 421Solenoidal vector field, 450Solid angle, 463 (Exercise 13)Sphere : area of, 427volume of (in n-space), 411 672 Index Spherical coordinates, 293,410, 414Standard deviation, 557Standardized normal distribution, 537 table of values of, 536Standardized random variable, 567Stationary point, 304Step function,
354,406 integral of, 355, 406STIRLING, JAMES, 572, 593Stirling formula for n factorial, 616Stirling numbers : of first kind, 593 (Exercise 4)of second kind, 594 (Exercise 5) Stochastic, 512Stochastic independence, 488Stochastic variable, 512STOKES, GEORGE GABRIEL, 438Stokes' theorem, 438, 455Strong law of large numbers, 566Sturm-Liouville
operator, 116Subspace, 8Summation formula of Euler, 613, 615Surface : area of, 424closed, 456equipotential, 335nonorientable, 456of revolution, 428orientable, 456parametric, 420smooth, 421 Surface integrals, 430applications of, 431independence of parametrization, 434notations for, 434 Symmetric difference, 473 (Exercise 10)Symmetric
matrix, 124Symmetric operator, 115Systems : of differential equations, 197of linear algebraic equations, 58 Tangent plane to a surface, 268, 424TAYLOR, BROOK, 571Taylor's formula : for scalar fields, 258, 308for vector fields, 270 Taylor polynomials, 571Torus, 376Total derivative : of a scalar field, 258of a vector field, 270 Trace of a matrix,
106Trajectory, orthogonal, 267 Transformations:linear, 31of double integrals, 394of n-fold integrals, 408 Trapezoidal rule, 604Triangle inequality, 17Trigonometric polynomial, 29Triple integrals : applications of, 414in cylindrical coordinates, 410in spherical coordinates, 411 Uncountable set, 502Uniform continuity, 321Uniform distribution : over an
interval, 526over a plane region, 549 (Exercise 9)over a square, 547 Uniqueness theorem for determinants, 79Uniqueness theorems for differential equa- tions :first-order linear systems, 226first-order nonlinear systems, 229linear equations of order IZ, 147matrix differential equations, 198 Unit sphere in n-space, 135volume of, 411 Unitary matrix,
123Unitary operator, 138Unitary space, 15Upper integral, 358Uranus, 571 Variance, 556binomial distribution, 557Cauchy distribution, 561 (Exercise 7)exponential distribution, 561 (Exercise 7)normal distribution, 558Poisson distribution, 561 (Exercise 7)uniform distribution, 557 Variation of parameters, 157VEBLEN, OSWALD, 380Vector equation
of a surface, 418Vector field, 243 total derivative of, 270Vector space, 4Volume : of an ordinate set, 360, 369of an n-dimensional sphere, 411 WALLIS, JOHN, 616Wallis' inequality, 616 Index 613 Wave equation, 288 YOUNG, w.
H., 258Weak law of large numbers, 565Winding number, 389Work : Zeros : and energy principle, 329 of Bessel functions, 188 (Exercise 3)as a line integral, 328 of Chebyshev polynomials, 597 WRONSKI, J. M. HO&N~$ 95 (Exercise 8), 159 of Legendre polynomials, 177Wronskian determinant, 161 Zeta function of Riemann, 619 (Exercise
3)Wronskian matrix, 95 (Exercise 8), 159 ZUCKERMAN, HERBERT S., 559