Mathematical Tools for Physicists

Edited by

George L. Trigg

WILEY-VCH Verlag GmbH & Co. KGaA

Mathematical Tools for Physicists

Edited by
George L. Trigg

Mathematical Tools for Physicists

Edited by

George L. Trigg

WILEY-VCH Verlag GmbH & Co. KGaA

Editor:

Dr. George L. Trigg
New Paltz, New York, USA

Editorial Advisor:

Professor Stuart P. Smith
California State University, Hayward (CSUH), USA

This edition is based on Review Articles originally written for the ''Encyclopedia of Applied Physics'', edited by George L. Trigg, 25 Volumes, VCH Publishers/Wiley-VCH, New York/Weinheim, 1991–2000.

Cover Picture

Sound pressure plot of the acoustic waves in water around an aluminum cylinder. Arrow and deformation plot shows the deformation of the cylinder. Model computed by using the FEMLAB® Structural Mechanics Module. © COMSOL AB, Stockholm, Sweden. Printed with kind permission of COMSOL AB, www.comsol.com.

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data: A catalogue record for this book is available from the British Library.

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the internet at http://dnb.ddb.de.

Printed in the Federal Republic of Germany. Printed on acid-free paper.

Composition: Laserwords Private Ltd., Chennai, India
Printing: betz-druck gmbh, Darmstadt
Bookbinding: J. Schäffer GmbH, Grünstadt
ISBN-13: 978-3-527-40548-0
ISBN-10: 3-527-40548-8

Contents

Preface

Mathematics is a central structure in our knowledge. The rigor of mathematical proof places the subject in a very special position with enormous prestige. For the potential user of mathematics this has both advantages as well as disadvantages. On the one hand, one can use mathematics with confidence that in general the concepts, definitions, procedures, and theorems have been thoroughly examined and tested, but the sheer amount of mathematics is often very intimidating to the non-expert. Since the results of mathematics once proved stay in the structure forever, the subject just gets larger and larger, and we do not have the luxury of discarding older theories as obsolete. So the quadratic formula and the Pythagorean theorem are still useful and valid even though they are thousands of years old. Euclid's Elements is still used as a text in some classrooms today, and it continues to inspire readers as it did in the past although it treats the mathematics from the time of Plato over 2300 years ago.

Despite the prestige of mathematical proof, most mathematics that we use today arose without proof. The history of the development of calculus is a good example. Neither Newton nor Leibniz gave definitions of limits, derivatives, or integrals that would meet current standards. Even the real number system was not rigorously treated until the second half of the nineteenth century. In the past, as in modern times, large parts of mathematics were initiated and developed by scientists and engineers. The distinction between mathematicians and scientists was often rather vague. Consider for example, Newton, Euler, Lagrange, Gauss, Fourier, and Riemann. Although these men did important work in mathematics, they were also deeply involved in the sciences of their times. Toward the end of the nineteenth century a splitting occurred between mathematics and the sciences. Some see it in the development of non-Euclidean geometry and especially axiomatic methods reminiscent of Euclid.

At this time mathematics appeared to be taking its own path independent of the sciences. Here are two cases that participated in this division. In the late nineteenth century Oliver Heaviside developed the Operational Calculus to solve problems in electrical engineering. Although this calculus gave solutions in agreement with experiment, the mathematicians of Heaviside's time could not justify or condone his procedures. Physicists also found the Dirac delta function and Green's functions extremely useful and developed an appropriate calculus for their use, but the underlying mathematical theory was not available. It was not until the early 1950's that Laurent Schwartz was able to give a rigorous mathematical foundation for these methods with

his Theory of Distributions. Also, early in the twentieth century the relatively new subject of Group Theory was seen as being of use in applications to chemistry and physics, but the few texts available at the time were written in a rather abstract and rigorous mathematical style that was not easily accessible to most non-mathematicians. The subject was quickly labeled the "Gruppenpest" and ignored by many researchers. Needless to say, today group theory with its applications to symmetry is a fundamental tool in science.

With the complexity of each field in science and engineering growing so rapidly, a researcher in these fields has little time to study mathematics for its own sake. Each field has more material than can possibly be covered in a typical undergraduate program, and even graduate students must quickly pick a sub-area of specialization. Often, however, there is a sense that if we just knew more mathematics of the right sort, we could get a better grasp of the subject at hand. So, if we are still in school, we may take a mathematics course, or if not in school, we may look at some mathematical texts. Here some questions arise: which course should we take, do we have the correct prerequisites, what if our mathematics instructor has no knowledge of our field or any applications that we are interested in, are we really in the right course? Furthermore, most texts in mathematics are intended for classroom use. They are generally very proof oriented, and although many now include some historical remarks and have a more user friendly tone, they may not get to the point fast enough for the reader outside of a classroom.

This book is intended to help students and researchers with this problem. The eighteen articles included here cover a very wide range of topics in mathematics in a compact, user oriented way. These articles originally appeared in the Encyclopedia of Applied Physics, a magnificent twenty-three volume set edited by George L. Trigg, with associate editors Eduardo S. Vera and Walter Greulich and managing editor Edmund H. Immergut. The full Encyclopedia was published in the 1990's by VCH, a subsidiary of John Wiley & Sons, New York. Each article in this volume covers a part of mathematics especially relevant to applications in science and engineering. The articles are designed to give a good overview of the subject in a relatively short space with indications of applications in applied physics. Suggestions for further reading are provided with extensive bibliographies and glossaries. Most importantly, these articles are accessible. Each article seeks to give a quick review of a large area within mathematics without lapsing into vagueness or overspecialization.

Of course not all of mathematics can be covered in this volume: choices must be made in order to keep the size of the work within bounds. We can only proceed based on those areas that have been most useful in the past. It is certainly possible that your favorite question is not discussed here, and certainly the future will bring new mathematics and applications to prominence, but we sincerely expect that the articles in this volume will be valuable to most readers.

Stuart P. Smith
CSUH – January 2005

List of Contributors

K. Binder
Institut für Physik,
Johannes-Gutenberg-Universität Mainz,
Mainz,
Germany

Ronald N. Bracewell
Electrical Engineering Department,
Stanford University,
Stanford, California,
USA

Samuel L. Braunstein
SEECS,
University of Wales,
Bangor,
UK

Christina C. Christara
Computer Science Department,
University of Toronto,
Toronto, Ontario,
Canada

A. J. Coleman
Department of Math and Statistics,
Queens University,
Kingston, Ontario,
Canada

G. W. F. Drake
Department of Physics,
University of Windsor,
Windsor, Ontario,
Canada

David J. Foulis
University of Massachusetts,
Amherst, Massachusetts,
USA

Roberto Giuntini
Università di Firenze,
Florence,
Italy

Richard J. Greechie
Louisiana Tech University,
Ruston, Louisiana,
USA

M. Hamermesh
School of Physics and Astronomy,
University of Minnesota,
Minneapolis, Minnesota,
USA

Charlie Harper
Department of Physics,
California State University,
Hayward, California,
USA

Kenneth Hartt
Physics Department,
University of Rhode Island,
Kingston, Rhode Island,
USA

Kenneth R. Jackson
Computer Science Department,
University of Toronto,
Toronto, Ontario,
Canada

V. Alan Kostelecký
Physics Department,
Indiana University,
Bloomington, Indiana,
USA

Melvin Lax
Physics Department,
City College of the City
University of New York,
New York,
USA

and

Bell Laboratories,
Lucent Technologies,
Murray Hill, New Jersey,
USA

Maria Louisa Dalla Chiara
Università di Firenze,
Florence,
Italy

Paul Meakin
Department of Physics,
University of Oslo,
Oslo,
Norway

James Murdock
Iowa State University,
Ames,
USA

Kazuo Ohtaka
Laboratory of Applied Physics,
Faculty of Engineering,
Chiba University, Chiba-shi,
Japan

Gino Segrè
Department of Physics,
University of Pennsylvania,
Philadelphia, Pennsylvania,
USA

S. P. Smith
Department of Mathematics and
Computer Science,
California State University,
Hayward, California,
USA

Algebraic Methods

A. J. Coleman
Department of Math and Statistics, Queens University, Kingston, Ontario, Canada

Introduction

The use of mathematics by physicists, and in particular of algebra, has increased in a remarkable degree during the last 50 years, both in the amount of space occupied in journal articles and in the type and abstractness of the methods employed.

Following N. Bourbaki, it is now conventional to characterize as algebraic structures those parts of mathematics that employ operations, such as addition, which act on a finite set of objects to produce a unique corresponding object. Such operations are contrasted with ideas like limit in calculus or closure in topology, which associate a number or other mathematical object to an infinite set or sequence. Thus, whereas the passage from (2, 3) to $2 + 3 = 5$ is an algebraic operation, to go from the infinite sequence $n \to 1/n$ (where n is any positive integer) to the limit 0 is a topological operation. The present section is concerned chiefly with algebra.

In this brief article it is impossible to describe all the many algebraic structures which occur in the literature of applied physics. Therefore we have selected those which are absolutely essential for understanding the contemporary literature under the following rubrics: Groups; Fields; Linear Algebra; Rings; Algebras and Modules. As to style, we have attempted to steer a course between that which physicists would have liked 20 years ago and the austerity of contemporary pure mathematicians with which all physicists will be happy 20 years from now. This should leave all readers equally unhappy! Our definitions are seldom painstakingly detailed but rather highlight the essential ideas leaving the reader to use common sense to fill them out. We shall assume that the reader is familiar with elementary properties of vectors and matrices. Recall that a square matrix A is invertible or nonsingular if there is a matrix B such that $AB = BA = I$, where I is the identity matrix. In this case A and B are inverses of each other and we denote B by A^{-1}. Although, logically, rings should be discussed before fields, teaching experience suggests that the reverse order is pedagogically sounder.

NOTATION: We shall adopt the following widely used symbolism: $\mathbb{N} :=$ the natural numbers, $\{1, 2, 3, \ldots\}$; $\mathbb{Z} :=$ the positive and negative integers and zero; $\mathbb{R} :=$ the real numbers; $\mathbb{C} :=$ the complex numbers; $i := \sqrt{-1}$; $\mathbb{Q} :=$ the rational numbers. We shall employ Einstein's summation convention in the restricted form that in any monomial an index which is repeated as a subscript and as a superscript will be interpreted as summed over its range unless the contrary is explicitly stated.

1 Groups

A group is a set, G, say, together with a binary operation which we temporarily denote by "$*$", which satisfies certain definite rules. A binary operation is one which combines two elements of G to

obtain an element of G. Perhaps our first encounter with a group occurs when as babies we push our blocks around on the floor using the translation group in two dimensions! Later in grade school we learn the properties of the integers. Under addition the integers $\mathbb{Z}$ exemplify the axioms of a group:

(i) A group $(G,*)$ is a set, G, together with an operation, $*$, which to any two elements x and y of G associates an element $z = x * y$ of G. For example, in $(\mathbb{Z}, +)$, $2 + 3 = 5$, $5 + (-3) = 2$. This property is described by saying that G is closed under the binary operation $*$.
However, for the structure $(G,*)$ to be dignified with the title "group," it must satisfy the additional properties:

(ii) The operation is associative, that is for any x, y, z in G, $(x * y) * z = x * (y * z)$.

(iii) There is a unique neutral or identity element, n, such that $x * n = n * x = x$ for all x in G.

(iv) For any element x in G there is a unique element y in G such that $x * y = n$. In this case, x and y are said to be inverses of each other.

Thus while $(\mathbb{N}, +)$ satisfies (i) and (ii) it is not a group because (iii) and (iv) fail. However, $(\mathbb{Z}, +)$ is a group when we take $n := 0$.

If G has a finite number of elements, the group is a finite group and the number of elements is called the order of the group. If $x * y = y * x$ for all $x, y \in G$, the group is Abelian or commutative.

The set of symmetries of any mathematical or physical structure constitutes a group under composition of symmetries. Such groups play a major role in physics for analyzing the properties of space-time, understanding crystal structure, and classifying the energy levels of atoms, molecules, and nuclei. Indeed, the role of groups is so important in physics that an article of the *Encyclopedia* is devoted to them. We therefore shall not explicitly pursue the detailed properties of groups further, even though they will occur as substructures in rings and fields.

2 Fields

Whereas a group consists of a set together with a single binary operation, a field consists of a set together with two binary operations linked together by a distributive law. The two operations are usually called addition and multiplication. The familiar fields are the real numbers, $\mathbb{R}$; the complex numbers, $\mathbb{C}$; and the rational numbers, $\mathbb{Q}$. We shall use the symbol $\mathbb{F}$ for an arbitrary field. Strictly speaking, we should employ a notation such as $(\mathbb{F}, +, \times)$ to denote a field; however, the relevant operations are generally obvious from context in which case it is sufficient to use $\mathbb{F}$ alone.

$(\mathbb{F}, +, \times)$ is a field if:

(i) $(\mathbb{F}, +)$ is a commutative or Abelian group. That is, $x + y = y + x$ for any x and y in $\mathbb{F}$.

(ii) The elements of $\mathbb{F}$ other than zero form a group under multiplication.

(iii) Multiplication distributes over addition. That is, if $a, b, c,$ belong to $\mathbb{F}$ then $a \times (b + c) = a \times b + a \times c$, and $(b + c) \times a = b \times a + c \times a$.

These properties are, of course, familiar for the reals, complexes, and rationals, but there are fields, such as the quaternions, for which multiplication is not commutative. There are also fields with only a finite number of elements.

A field always has at least two elements, 0 and 1.

2.1 The Characteristic of $\mathbb{F}$

Since a field is closed under addition, $\mathbb{F}$ contains $1+1$, which cannot equal 1 since this would imply that $1=0$, which we excluded. But $1+1$ might equal 0 in which case $(1+1)+1=1$ and one can easily check that $\mathbb{F}=\{0,1\}$ can serve as the set of a field of two elements. This is the smallest possible field and is both famous and useful since it plays a key role in the design of electric circuits, such as those which occur in computers.

More generally, if p is a prime number, we can obtain a field containing p elements in which the sums of j 1's are numbers which are distinct if $0 \leq j < p$ and equal to 0 if $j=p$. When this occurs in any field $\mathbb{F}$ we say that p is the characteristic of $\mathbb{F}$ and that $\mathbb{F}$ has finite characteristic. When there is no such p we say that $\mathbb{F}$ has characteristic zero. A field of characteristic zero has an infinite number of elements. If $\mathbb{F}$ has only a finite number of elements it will contain p^n elements, where p is a prime and n is a positive integer. If $n>1$, $\mathbb{F}$ will contain a subfield of the above type with p elements. The fields with p^n elements are called Galois fields. They are important in coding and communication theory. A finite field is necessarily commutative.

2.2 Algebraically Closed Fields

We know that the square of a real number is positive, so there is no real number x such that $x^2=-1$. In other words, in $\mathbb{R}$ there is no element x satisfying the equation $x^2+1=0$. If $\mathbb{F}$ has the property that for every equation of the form $a_j x^j = 0$, $0 \leq j \leq n$, where the a_j belong to $\mathbb{F}$, there is an element of $\mathbb{F}$ which satisfies the equation, we say that $\mathbb{F}$ is algebraically closed. Otherwise, it is not algebraically closed. Clearly $\mathbb{R}$ is not algebraically closed. If we assume that there is a "number" i such that $i^2+1=0$, then, as we know, the field containing $\mathbb{R}$ and i is the complex numbers, $\mathbb{C}$. It was proved by Gauss that $\mathbb{C}$ is algebraically closed.

Notice that if σ is a 1:1 map of $\mathbb{C}$ onto itself, such that $\sigma(x+iy)=x-iy$ for all $x, y \in \mathbb{R}$, then σ preserves all the properties of a field and is therefore an automorphism of $\mathbb{C}$. Recall that an isomorphism of two algebraic structures is a bijective (or one-to-one) correspondence between their sets, which preserves all the relations among their elements, and that an automorphism is an isomorphism of an algebraic structure onto itself. Note that $\sigma(x)=x$ if $x \in \mathbb{R}$ and that $\sigma(i)=-i$, which is the root other than i of the equation $x^2+1=0$. An automorphism of $\mathbb{C}$ must send 0 into 0 and thus must either leave i fixed (and so everything in $\mathbb{C}$ is fixed) or, like σ, send i to $-i$. The set consisting of σ and the identity map is a group of order two under composition of mappings. It is the Galois group of $\mathbb{C}$ over $\mathbb{R}$. Alternatively, it is also called the Galois group of the equation $x^2+1=0$ with respect to the reals. For more detail about fields, their algebraic extensions, and their Galois groups, the reader is referred to Jacobson (1964) or any of the multitude of algebra texts at the same level.

Are there fields containing $\mathbb{R}$ other than $\mathbb{C}$ which are at most finite dimensional over $\mathbb{R}$? The answer was given by Frobenius. There is one and only one, the quaternions, but in this field multiplication is not commutative. We shall see below that the quaternions can be realized

as linear combinations with real coefficients of the Pauli matrices and the 2×2 identity matrix. The significance of the field of quaternions is dramatized by the observation that if it did not exist there would be no spin in physics, therefore no sigma and pi bonds in chemistry, and therefore no life on planet earth if, indeed, there were any stars or planets!

2.3 Rational Functions

If we adjoin a symbol x to any field $\mathbb{F}$ and form all possible sums, differences, products, and quotients involving x and the elements of $\mathbb{F}$, the result is a set which is closed under any finite sequence of these operations and forms a field, denoted by $\mathbb{F}(x)$, which we might describe as the field of rational functions in x over $\mathbb{F}$. There is a subset, denoted by $\mathbb{F}[x]$, of polynomials of the form $a_j x^j$ where j is summed from 0 to some $n \in \mathbb{N}$, where n is arbitrary and the $a_j \in \mathbb{F}$. If a_n is not zero we say that the polynomial has degree n. As we shall remark below, the polynomials constitute a ring. As usual $x^0 := 1$, by definition, so when $n = 0$ the preceding polynomial reduces to a_0. Thus $\mathbb{F}$ is contained in $\mathbb{F}[x]$. The field $\mathbb{F}(x)$ consists of all possible quotients of elements of $\mathbb{F}[x]$.

Suppose that the rational function $R(x) = P(x)/Q(x)$, where P and Q are polynomials. Suppose further that $Q(x) = Q_1(x)Q_2(x)$, where Q_1 and Q_2 are polynomials with no common factor. Since we could have used long division to ensure that R is the sum of a polynomial and a rational function, the numerator of which has degree strictly less than the degree of Q, we may assume that $\deg(P)$ – the degree of P – is less than $\deg(Q)$. It is relatively easy to show that it is then possible to find polynomials P_1 and P_2 with $\deg(P_i) < \deg(Q_i)$ such that

$$\frac{P}{Q} = \frac{P_1}{Q_1} + \frac{P_2}{Q_2}.$$

This is the fundamental theorem of the so-called method of partial fractions, by repeated application of which it follows that any rational function can be expressed as a sum of a polynomial and rational functions whose denominators have no nontrivial factors.

In particular, if $\mathbb{F}$ is algebraically closed (e.g., $\mathbb{F} = \mathbb{C}$), then Q is a product of factors such as $(x - a)^m$, where $a \in \mathbb{F}$. A summand in $R(x)$ of the form $g(x)/(x - a)^m$ with $\deg(g) < m$ can, by Taylor's theorem applied to $g(x)$, be expressed as the sum $c_j(x - a)^{-j}$, where $1 \leq j < m$ and $c_j = g^{(m-j)}(a)/(m - j)!$. Here $g^{(k)}$ is the kth order derivative of g.

The method of partial fractions is quite useful for finding integrals of rational functions. Books on calculus explain helpful tricks for obtaining the partial fraction decomposition of a given rational function.

3 Linear Spaces

The theory of linear space with its related concepts of linear transformation, eigenvector, matrix, determinant, and Jordan canonical form is certainly one of the most important and most useful part of mathematics. The abstract concept of linear space is frequently approached by a long discussion of the problem of solving systems of linear equations. We take a direct approach defining a linear space as consisting of a field $\mathbb{F}$ whose elements are called scalars, a set V, called vectors, and two operations called vector addition and scalar multiplication together

with a set of rules governing the relation among these various entities. The vectors under addition form an additive Abelian group $(V, +)$. Under multiplication by scalars the set V is closed. Thus, $v \in V$ and $a \in \mathbb{F}$ imply that $av \in V$. Another important property is that multiplication by scalars distributes over addition of vectors. That is $a(v_1 + v_2) = av_1 + av_2$ for all $a \in \mathbb{F}$ and $v_i \in V$.

3.1 Independence of Vectors

This seems to be the most difficult idea in teaching elementary courses in linear algebra – possibly, the only difficult idea! Two nonzero vectors v_1 and v_2 are linearly dependent if there are scalars a^1 and a^2, not both zero, such that $a^1v_1 + a^2v_2 = 0$, where, of course, by 0 we mean the zero vector. It is clear that neither a^1 nor a^2 is zero, and thus each vector is a scalar multiple of the other. More generally, if, given n vectors v_i, $1 \leq i \leq n$, there exist scalars a^i such that $a^iv_i = 0$, where all $v_i \neq 0$ and not all $a^i = 0$; then we say that the n vectors are linearly dependent. If no such relation holds, the vectors are linearly independent. For example, for $n \in \mathbb{N}$ there are no numbers a_n other than 0 such that $\Sigma_n a_n \times \cos(n\vartheta) = 0$ for all ϑ. Thus the functions $\vartheta \to \cos(n\vartheta)$ are linearly independent.

If n vectors v_i are such that any vector v can be written as $v = a^iv_i$ for some choice of scalars a^i, we say that the set $\{v_i\}$ spans V. If the v_i are also linearly independent then the coefficients a^i are unique. We then say that $B = \{v_i\}$ is a basis of V, that the linear space V has dimension n, and that a^i are the components of v with respect to B. A basic theorem assures us that the dimension depends only on the space V and not on the choice of basis. If a linear space does not have a finite basis it is infinite dimensional.

3.2 Change of Basis

How do the components of a given vector change if the basis is changed? This was a key question which led to the theory of invariants in the mid-19th century and opened up the development of much of contemporary algebra. It also led to the emergence of the tensor calculus which was essential for Einstein's exposition of General Relativity Theory.

Suppose V is a linear space and that $B = \{v_i\}$ and $B' = \{v'_i\}$ are two different bases for V. Then there is a matrix, P^j_i called the transition matrix from B to B' such that $v'_i = P^j_i v_j$. Thus if the vector $x = x^jv_j = x'^iv'_i = x'^iP^j_iv_j$, it follows that $x^j = P^j_i x'^i$. If in the usual matrix notation we regard x^j as the jth component of a column vector $\mathbf{x}$, and P^j_i as the element in the jth row and ith column of the transition matrix, $\mathbf{P}$, from the old base B to the new base B', the preceding equation takes the form

$$\mathbf{x} = \mathbf{P}\mathbf{x}' \quad \text{or} \quad \mathbf{x}' = \mathbf{P}^{-1}\mathbf{x}.$$

There is an even more convenient notation. Define $P(B', B) := \mathbf{P}$; then the preceding equations imply that $\mathbf{P}^{-1} = P(B, B')$. Subsequently we shall need the formulas

$$\mathbf{x} = P(B', B)\mathbf{x}' \quad \text{and} \quad \mathbf{x}' = P(B, B')\mathbf{x}.$$

To understand tensor notation it will prove important to note that, whereas $\mathbf{P}$ sends the old to the new basis, it sends the new coordinates to the old ones. This observation underlies duality in homological algebra and the distinction between

covariant and contravariant tensors, which we define below.

3.3 Linear Maps and Their Associated Matrices

Suppose that V and U are linear spaces of dimension n and m with bases $B_v = \{v_i\}$ and $B_u = \{u_j\}$, respectively. A transformation, function, or map from V to U sends each vector $x \in V$ to a vector, say, y of U, which we denote by $Ax := y$. If A has the property that for any two vectors x and $x' \in V$ and arbitrary scalars a and $a' \in \mathbb{F}$, $A(ax + a'x') = aAx + a'Ax'$, we say that A is linear. The condition that a map be linear is very restrictive. Nonetheless, linear maps play a big role in the application of mathematics to physics (as well as statistics, economics, biology, etc.) for the same reason that the derivative is important in analysis. For example, if $f(x)$ is a real-valued function of the real variable x, such that $f(0) = 0$, then $f'(0)x$ is the linear function of x which is the best possible linear approximation to $f(x)$ near 0.

A linear map or transformation, $A: V \to U$, can be completely described by an $m \times n$ matrix A_i^j, such that $Av_i = A_i^j u_j$, which we describe as the matrix associated to the linear transformation or map A with respect to the bases B_u and B_v. It has m rows and n columns. If we denote this matrix by $A(u,v)$ then if the bases in V and U are changed to B'_v and B'_u, respectively,

$$A(u', v') = P(B'_u, B_u)A(u, v)P(B_{v'}, B'_v),$$

where we use the notation of Sec. 3.2. In terms of coordinates, if $\mathbf{y} = \mathbf{Ax}$, then $y^j = A_i^j x^i$, where, as follows from the context, $1 \le j \le m$ and $1 \le i \le n$.

In the particular case that $V = U$ of dimension n, with $B_u = B_v$, $A(u, u)$ is an $n \times n$ matrix which we denote by $A(u)$. We deduce that for a change of basis

$$A(u') = P(B'_u, B_u)A(u)P(B_u, B'_u).$$

We thus associate to any linear transformation a matrix which is unique, *once bases are chosen* for the domain and codomain of the transformation. But conversely, if the bases are given, then there is a unique linear transformation associated with a given matrix of the appropriate shape. Thus there is a bijection (i.e., a one-to-one correspondence) between $m \times n$ matrices with entries in $\mathbb{F}$ and linear maps from a linear space of dimension n into one of dimension m. We have found how the bijection changes when the bases are altered. It is this bijection which gives meaning to the familiar addition and multiplication of matrices.

A linear map between two spaces over the same field $\mathbb{F}$ has the property of preserving the linear structure and is said to be a homomorphism (i.e., a structure-preserving map), so it is common to denote by $\mathrm{Hom}(V_1, V_2)$ the set of all linear maps between linear spaces V_1 and V_2 where both have the same field. If $A \in \mathrm{Hom}(V_1, V_2)$ then V_1 is the domain of A and V_2 is the codomain of A. The kernel of A, frequently denoted by $\ker(A)$, is the set of all elements in the domain which are mapped onto 0 by A. The range of A consists of all elements of the codomain of the form Ax for some x in the domain. Of course these last four definitions are valid for any function, not merely linear maps. However, when A is linear it can be easily proved that both the kernel and the range are linear subspaces of their ambient spaces. This is probably the secret of the power and relative simplicity of the theory of linear spaces. When $V_1 = V_2 = V$, we denote $\mathrm{Hom}(V,V)$ by $\mathrm{Hom}(V)$.

If G is a map from V_1 to V_2 and F one from V_2 to V_3, we denote the composition

of these two maps by FG and, having fixed bases in the spaces, we define the matrix corresponding to FG as the product of the matrices corresponding to F and G. This "explains" the usual rule for matrices that $(FG)_i^j = F_k^j G_i^k$, where i, k, and j range from 1 to the dimensions of V_1, V_2, and V_3, respectively.

The composition (or product) of two maps can be well-defined if the range of the first is in the domain of the second. The sum of two maps is only meaningful if the codomain is an additive group in order for the sum of Fx and Gx to be meaningful. In this case it is possible to let $F + G$ denote the map such that $(F + G)x = Fx + Gx$ for all x in the intersection of the domains of F and G. When the domain and codomain are fixed linear spaces over the same field $\mathbb{F}$ we can do even better and give $\mathrm{Hom}(V_1, V_2)$ the structure of a linear space over $\mathbb{F}$. This implies that the set of all $m \times n$ matrices with entries from $\mathbb{F}$ is a linear space of dimension mn over $\mathbb{F}$.

The dimension of the range of a linear operator is called the rank of the operator and also the rank of any matrix associated with the operator by a particular choice of bases. The dimension of the kernel of a linear transformation is called the nullity of the transformation and of its associated matrices. It follows from this definition that the various matrices obtained from one another by a change of basis all have the same rank and nullity. The rank of a product of operators or matrices is not greater than the minimum rank of its factors.

3.4 Determinants

If $\mathbb{F}$ is a commutative field, to any square matrix, it is possible to assign a number in $\mathbb{F}$ which is expressible as a polynomial in the elements of the matrix and which vanishes only if the matrix is not invertible. To two square matrices which are related as in Sec. 3.3 by a change of basis, we assign the same number, and therefore it is meaningful to also assign this number to the associated linear transformation belonging to $\mathrm{Hom}(V)$. The function, det, from $\mathrm{Hom}(V)$ into $\mathbb{F}$, has the following properties: (i) $\det(AB) = \det(A)\det(B)$; (ii) $\det(fI) = f^n$, where n is the dimension of V, I is the identity map, and f is any element of $\mathbb{F}$. The usual definition of the determinant follows from these properties (MacDuffee, 1943). In particular since, for a fixed basis, the equation $Ax = y$ is equivalent to the system of equations $A_i^j x^i = y^j$, Cramèr's rule implies

$$\det(A)x^i = \det(Y^i),$$

where Y^i is the matrix obtained from (A_i^j) by replacing its ith column by the column vector (y^k) where i, j, k run from 1 to n. Thus if $\det(A) \neq 0$ there is a unique x for every y so A is invertible; whereas if $\det(A) = 0$, there is an x only for particular y satisfying the n conditions $\det(Y^k) = 0$. Thus for a finite dimensional linear space V, $A \in \mathrm{Hom}(V)$ is invertible if and only if $\det(A) \neq 0$.

The theory of $\mathrm{Hom}(V_1, V_2)$ is really equivalent to the theory of systems of linear equations in several variables. This topic occurs in articles of this book devoted to NUMERICAL METHODS and to MATHEMATICAL MODELING and in at least one hundred elementary textbooks; so we shall not pursue it here.

3.5 Eigenvectors and Eigenvalues

If $A \in \mathrm{Hom}(V)$ then for any $x \in V$, $Ax \in V$. In general we shall not expect Ax to

equal x or indeed, even, that Ax be parallel to x. However, in the latter case Ax would be a multiple of x, say, λx. The equation $Ax = \lambda x$ is equivalent to $(\lambda I - A)x = 0$. By the preceding section, if the determinant of $\lambda I - A$ is different from zero, the only possible solution of this equation is $x = 0$, which is of no great interest. When there is a nontrivial solution of this equation it will be somewhat unusual and is called an eigenvector of A and can occur only for special values of λ. Such a value of λ is the eigenvalue of A corresponding to the particular eigenvector x. The eigenvalue, λ, will satisfy the nth degree algebraic equation

$$f(z; A) := \det(zI - A) = 0.$$

The nth degree polynomial $f(z;A)$ is called the characteristic function of A, and the preceding equation is the characteristic equation of A. Any eigenvalue of A satisfies its characteristic equation. For each zero of the characteristic equation there is at least one nontrivial eigenvector.

There is a one-to-one correspondence between the operators in $\mathrm{Hom}(V)$ and the set of $n \times n$ matrices over $\mathbb{F}$, and this set spans a linear space over $\mathbb{F}$ of dimension n^2. If we interpret A^0 as the identity operator, I, it follows that the operators A^k for $0 \le k \le n^2$ are linearly dependent. That is, there are $c_j \in \mathbb{F}$ such that $c_j A^j = 0$, where not all c_j are zero. Thus there exists at least one polynomial, $p(z)$, such that $p(A) = 0$. From the algorithm for long division it easily follows that there is a unique monic polynomial (i.e., a polynomial with highest coefficient 1) of minimal degree with this property. We shall denote this so-called minimal polynomial of A by $m(z;A)$. A famous theorem of Hamilton asserts that A satisfies its characteristic equation. That is, $f(A;A) = 0$. Since $\deg(f) = n$, $\deg[m(z;A)] \le n$. Since $m(z;A)$ divides any polynomial $p(z)$ such that $p(A) = 0$, it follows that $m(z;A)$ divides $f(z;A)$.

The form of $m(z;A)$ provides information about A.

(i) $m(z : A) = z^p$ implies that $A^p = 0$ but that $A^{p-1} \neq 0$. Such an operator is called nilpotent, with nilpotency index p.

(ii) $m(z;A) = (z-1)^p$ implies that $A - I$ is nilpotent with index p. Thus in this case $A = I + N$, where N is nilpotent. An operator of this form is called unipotent.

(iii) Suppose the minimal polynomial of A has no multiple zeros, which is equivalent to saying that m and its derivative have no common factors. Then there is a basis of V consisting of eigenvectors of A. Equivalently, among the matrices associated with A there is one which is diagonal. In this case we say that A and its associated matrices are diagonalizable or semisimple.

(iv) If $m(z;A) = (z-\lambda)^p$, then, of course, $p \le n$. A basis can be chosen so that the matrix corresponding to A has zero entries except along the diagonal where there are so-called Jordan blocks, which in case $n = 4$, for example, would be

$$\begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{bmatrix}.$$

That is $n_i \times n_i$ matrices with λ on the diagonal and 1's on the first superdiagonal, $\Sigma n_i = n$, $1 \le n_i \le p$, and for at least one value of i, $n_i = p$.

In the preceding we have assumed that the entries of the matrix A could

be arbitrary. However, if they are real and nonnegative the remarkable Perron-Frobenius theorem (Senata, 1973) about the eigenvalues and eigenvectors of A gives information which is useful in many contexts; we thus state it here. A matrix $M = (m_{ij})$ is connected or indecomposable if for any two indices i and j there is a sequence $r_k, 1 \leq k \leq s$, such that the continued product $m_{ir_1} m_{r_1 r_2} m_{r_2 r_3} \ldots m_{r_s j} \neq 0$. We write $M > 0$ if all $m_{ij} > 0$, and $M \geq 0$ if all $m_{ij} \geq 0$. Then, if $M \geq 0$ is a real connected matrix, it has a largest simple positive eigenvalue, $r(M) = r$, and an associated column vector $x > 0$, such that $Mx = rx$ where $r > 0$; any other eigenvalue λ of M has absolute value less than or equal to r. Further, if $N \geq 0$ is another real matrix of the same dimension, such that $M - N \geq 0$, then $r(N) \leq r(M)$ with equality only if $N = M$. This theorem can be used to quickly give the basic classification of Kac-Moody algebras.

3.6
Canonical Form of Matrices

In Sec. 3.3 we noticed that distinct matrices were associated with the same linear operator, so there is a sense in which such matrices are "equivalent." Recall that by an equivalence relation a set is partitioned into distinct mutually exclusive subsets which exhaust the given set. One method of partitioning a set is into the orbits of a group which acts on the set. Thus if g belongs to a group G which is acting on a set S and we denote by gs the element of S into which g sends s, the orbit of s is the set $M_s = \{gs | g \in G\}$. It follows that $x \in M_s$ implies that $M_x = M_s$. Given an equivalence relation on a set of matrices, the problem considered in this section is that of choosing a canonical or "simplest" matrix in each equivalence class. There are different canonical forms depending on the types of matrices we consider and the different group actions contemplated.

Possibly the basic and most general situation is that considered by H. J. S. Smith in 1861. It is that of Sec. 3.3 where the equation $A(u', v') = PA(u, v)Q$ occurs in slightly different notation. There P and Q are arbitrary invertible $m \times m$ and $n \times n$ matrices, respectively. By choosing B'_u so that the last elements of the basis span the kernel of A and the first ones span a subspace which is complementary to the kernel, while the first elements of B_u span the range of A, one can arrive at Smith's canonical matrix which has 1's in the (i, i) positions for $1 \leq i \leq r$ where r is the rank of A, and zero everywhere else. It would be difficult to demand anything "simpler." It follows that with this meaning of equivalence there are $p + 1$ equivalence classes of $m \times n$ matrices where p is the minimum of $\{m, n\}$.

At first one is surprised that there are so few classes. However, on second thought, one notices that we have been acting on a space of matrices of dimension mn by the group $Gl(n, \mathbb{F}) \times Gl(m, \mathbb{F})$ $(= G$, say), which has $n^2 + m^2 \geq 2mn$ parameters; there is plenty of redundancy unless one of m and n is 1 and the other is 1 or 2.

If we consider an action on the set of $n \times n$ matrices by a smaller group we shall expect more equivalence classes. For $(P, Q) \in G$, subgroups of G can be defined by imposing restrictions on P and Q.

Recall the following definitions. If A is a square matrix the transpose of A, denoted by A^t, is obtained from A by interchanging rows and columns or by reflecting across the main diagonal. The operation of taking the transpose is an involution, that is $(A^t)^t = A$. If $A^t = A$, then we say A is symmetric. If $A^t = -A$, then A is antisymmetric or skew-symmetric.

An important property of transposition is $(AB)^t = B^t A^t$. It is worth noting that once the basis of V has been fixed, the mapping defined by transposition of matrices can be transferred to the associated linear transformations, thus defining an involution on $\mathrm{Hom}(V)$.

If σ is an automorphism of $\mathbb{F}$, we can define an operation on the matrix A by replacing each of its elements by its conjugate under the automorphism, and denote the new matrix by A^σ. If the field is commutative $(AB)^\sigma = A^\sigma B^\sigma$. In particular, when $\mathbb{F} = \mathbb{C}$, complex conjugation is an automorphism of period two. We follow a common custom and denote the complex conjugate of A by $\overline{A}$, so $\overline{AB} = \overline{A}\,\overline{B}$.

The Hermitian conjugate of A is denoted by $A^* = \overline{A}^t$ and satisfies $(AB) = B^* A^*$. A matrix A is Hermitian if $A^* = A$ and anti-Hermitian if $A^* = -A$.

The approach of this section is based on that of Turnbull and Aitken (1932), a superb book which goes far beyond our brief summary. They distinguish five subgroups of G.

(i) The Collinearity Group is characterized by $PQ = I$. It arises in Sec. 3.3 when $v = u$ and $v' = u'$. Under the action of this group, a square matrix, A, can be reduced to Jordan canonical form, that is to a sum of diagonal blocks, each of which has the form $\lambda I + N$, where λ is an eigenvalue of A and N is a nilpotent matrix, all of whose entries are zero except for 1's along the first superdiagonal. A particular eigenvalue occurs on the diagonal of the canonical form as many times as its multiplicity in the characteristic equation. For any eigenvalue the dimension of the largest Jordan block is equal to the multiplicity of the eigenvalue in the minimal polynomial $m(z;A)$. Thus if the zeros of $m(z;A)$ are simple, A is diagonalizable.

(ii) The Congruent Subgroup is defined by the condition $P^t = Q$. Under this group, symmetry or antisymmetry of A is invariant. A symmetric matrix can be diagonalized. If $\mathbb{F}$ is closed under taking square-roots, we can choose as the canonical element of an equivalence class a diagonal matrix which has only 0's or 1's on the diagonal. If $\mathbb{F} = \mathbb{R}$, the diagonal could also contain -1. In the real case, Sylvester's Law of Inertia asserts that the number of 1's and the number of -1's are invariants. A nonsingular antisymmetric matrix has even rank r and there is a canonical form under the congruent group which contains zeros everywhere except for $r/2$ blocks of 2×2 antisymmetric matrices down the diagonal; each has 1 and -1 off the diagonal and 0 on the diagonal.

(iii) The Conjunctive Subgroup is defined by the condition $P = Q^*$. It changes Hermitian matrices into Hermitian matrices. For real matrices, the conjunctive and the congruent transformations are the same. For any $\mathbb{F}$, one may choose a diagonal matrix as canonical. If $\mathbb{F} = \mathbb{C}$, the diagonal can consist of 1's and 0's.

(iv) The Orthogonal Group is defined by $PQ = I$ and $P = Q^t$ and is thus a subgroup of the groups (i) and (ii). It will preserve symmetry or antisymmetry of a matrix. A symmetric matrix will be equivalent to a diagonal matrix whose diagonal elements are eigenvalues of the original matrix. An antisymmetric matrix will be equivalent to one with zeros everywhere except for 2×2 blocks on the diagonal, the determinants of these blocks

being equal to the negatives of the squares of eigenvalues of the original matrix.

(v) The Unitary Subgroup is defined by $PQ = I$ and $P = Q^*$, and is thus a subgroup of (i) and (iii). It preserves the property of a matrix being Hermitian or anti-Hermitian. If $\mathbb{F} = \mathbb{R}$, groups (v) and (iv) are the same. Under this group, a Hermitian matrix is equivalent to a diagonal matrix whose nonzero elements are eigenvalues of the original matrix. An anti-Hermitian matrix is equivalent to one with 2-dimensional blocks on the diagonal whose determinants are the negatives of the squares of eigenvalues of the original matrix.

3.7 Dual Space

We have already noted that $\mathrm{Hom}(V,U)$, where V and U are linear spaces of dimension n and m, respectively, over a common field $\mathbb{F}$, can be given a structure of a linear space of dimension nm over $\mathbb{F}$. We can, of course consider $\mathbb{F}$ as a linear space of dimension 1 over F. Thus, $\mathrm{Hom}(V,\mathbb{F})$ is a linear space of dimension n over $\mathbb{F}$ and therefore isomorphic to $\mathbb{F}^n$ and hence also to V. It is called the dual space of V and usually denoted by V^*. This use of the asterisk can be distinguished from its use to indicate Hermitian conjugation by the context. The elements of V^* are linear functions on V with values in $\mathbb{F}$. We shall denote them by lower case Greek letters. Recall that the Kronecker symbol δ^i_j takes the value 1 if $i = j$ and 0 otherwise.

If $\alpha \in V^*$ and $x = x^j v_j$ is an arbitrary vector in V expressed in terms of the basis B_v, then $\alpha(x) = x^j \alpha(v_j) = a_j x^j$, where $a_j = \alpha(v_j)$. It is possible to define various bases for V^*. The basis which is said to be dual to B_v, and may be denoted by B^*_v, is defined as follows. Recall that a linear function on V is completely determined by the values it assumes for the elements of a basis of V.

Let α^i be a linear function such that $\alpha^i(v_j) = \delta^i_j$ for all j, $1 \le j \le n$. Then $\alpha^i(x) = x^i$. Thus α^i is the ith coordinate function. It easily follows that α^i are linearly independent and that $\alpha = a_j \alpha^j$, where $a_j = \alpha(v_j)$. Thus any element of V^* is a linear combination of the n elements α^j, $1 \le j \le n$, so that $B^*_v = \{\alpha^j\}$ is a basis for V. Just as the x^i are coordinates of an arbitrary element of V with respect to B_v, so a_i are coordinates of an arbitrary element of V^*. Since $a_i = \alpha(v_i)$, when the basis of V is changed, a_i changes by the same transformation as, or cogrediently with, the basis. As we noted at the end of Sec. 3.2, the x^i transform contragrediently to the basis. This distinction reflects the fact that the definition of the linear function $\alpha: x \to \alpha(x)$ is independent of the coordinate system used to describe it. A geometrical or physical entity which is described by a sequence of n numbers which transform like (a_i) is called a covariant vector. Similarly, an entity described by a sequence of n numbers which transform like (x^i) when the basis is changed is called a contravariant vector.

3.8 Tensors

Possibly it was algebraic geometers in the middle of the nineteenth century who first focused attention on the behavior of the coordinates of geometrical objects when the frame of reference is changed. But the first time this issue really impinged on physics was with the advent of Einstein's General Relativity Theory (GRT). The basic metric of GRT, $g_{ij}dx^i dx^j$, is clearly independent of the coordinate system but

since dx^i is a contravariant vector, g_{ij} will have to vary covariantly in both subscripts i and j. Then the n^2 symbols g_{ij} must be describing something (in fact, according to Einstein, the gravitational field!) which is a doubly covariant tensor.

The curvature of space-time, which allegedly explains black holes and how planets circle around the sun, is described by the Riemann-Christoffel tensor, R^i_{jkl}, which is contravariant in the index i and covariant in the other three.

The great advantage of the indicial notation, as it evolved in the writings of Eddington, Weyl, Synge, and other mathematical physicists between 1920 and 1940, is that it immediately indicates the behavior of the tensor when the underlying basis, or frame of reference, is changed. Thus if a_{ij} is a double covariant tensor and b^i is a contravariant vector (or first order tensor), then $a_{ij}b^k$ is a third order tensor covariant in two indices and contravariant in one. If we now contract on the indices j and k, we see immediately that $c_i = a_{ij}b^j$ is a covariant vector.

An algebraist would say that a_{ij} are the components of an element of $V^* \otimes V^*$, the tensor product of the dual space of V with itself. Similarly, $a^i b^j_k$ are the components of an element in the tensor product $V \otimes V \otimes V^*$. In general, the tensor product (see Sec. 4) of two linear spaces of dimension n and m is a linear space of dimension nm. In particular, $V^* \otimes U$ is isomorphic to $\mathrm{Hom}(V, U)$ and is spanned by a basis consisting of elements noted as $\alpha^i \otimes u_j$, where $1 \leq i \leq n$ and $1 \leq j \leq m$.

4 Creating Algebraic Structures

What experimental apparatus is for the physicist, the Cartesian product and quotient structures are for the algebraist. These are the principal tools with which he makes new mathematical structures.

If A and B are two sets, the Cartesian product of A and B is denoted by $A \times B$ and defined as the set $\{(x, y) | x \in A, y \in B\}$. Thus it is a new set consisting of ordered pairs with the first element of the pair belonging to A and the second to B. If $A \neq B$, $A \times B \neq B \times A$, since by definition two ordered pairs (a, b) and (c, d) are equal only if $a = c$ and $b = d$.

Things become more interesting when A and B have some algebraic structure which can be used to impose structure on the Cartesian product. For example, suppose that $A = B = \mathbb{Z}$. We define the addition of pairs $\in \mathbb{Z} \times \mathbb{Z}$ by $(x, y) + (u, v) := (x + u, y + v)$. Notice that the plus signs on the right and left have quite different meanings. One acts on pairs of integers; the others on integers. If we think of $+3$ as a translation by 3 units along the number line, we can call $(\mathbb{Z}, +)$ a translation group in one dimension. We could then think of $(\mathbb{Z} \times \mathbb{Z}, +)$ as the translation group of a two-dimensional lattice. Another familiar example is the idea due to Gauss of imposing the structure of the complex numbers on $\mathbb{R} \times \mathbb{R}$.

The direct sum of two vector spaces provides us with another important example of this construction. Suppose X and V are two linear spaces over the same field $\mathbb{F}$ with bases $\{e_i\}$, $1 \leq i \leq n$, and $\{f_j\}$, $1 \leq j \leq m$ respectively. For $x, y \in X$, $u, v \in V$, and $\alpha \in \mathbb{F}$, define (i) $(x, u) + (y, v) := (x + y, u + v)$; (ii) $\alpha(x, u) := (\alpha x, \alpha u)$. By these definitions we have imposed on $X \times V$ the structure of a linear space for which the $n + m$ elements $\{(e_i, 0), (0, f_j)\}$ form a basis. This new linear space is called the direct sum of the linear spaces X and V, and has dimension $n + m$, and is denoted by $X \oplus V$.

An apparently minor variation on the preceding definition leads us to an important but quite different object – the tensor product of X and V which is denoted by $X \otimes V$. This is a linear space which has a basis of elements belonging to the Cartesian product $X \times V$, but addition and scalar multiplication are different. (i) $(x_1 + x_2, v_1 + v_2) = (x_1, v_1) + (x_1, v_2) + (x_2, v_1) + (x_2, v_2)$; (ii) $\alpha(x, v) = (\alpha x, v) = (x, \alpha v)$. These conditions imply that $X \otimes V$ is a vector space over $\mathbb{F}$ of dimension mn, with $\{(e_i, f_j)\}$ as a basis.

Recall that an equivalence relation ρ on a set S partitions S into mutually exhaustive subsets which we call equivalence classes. A binary relation ρ on a set is an equivalence relation if it has the following three properties: (i) reflexive, $x\rho x$ for all $x \in S$, (ii) symmetric, $x\rho y$ implies $y\rho x$, (iii) transitive, $x\rho y$ and $y\rho z$ imply $x\rho z$. A subset of the partition of S contains exactly all the elements of S which are related by ρ to any one member of the subset. For example, the nails in a hardware store can be partitioned by length. Thus $x\rho y$ means $\text{length}(x) = \text{length}(y)$.

Now consider the equivalence relation ρ on $\mathbb{Z} \times \mathbb{Z}$ such that $(a, b)\rho(u, v)$ if and only if $av = bu$. We have used only properties of the integers to partition $\mathbb{Z} \times \mathbb{Z}$ into equivalence classes. But the condition we used is identical with the equality of the rational numbers a/b and u/v. We have thus established a bijection, or one-to-one correspondence, between $\mathbb{Q}$ and the equivalence classes of $\mathbb{Z} \times \mathbb{Z}$ under the relation ρ.

For any set S with equivalence relation ρ, the new set whose elements are equivalence classes of S is denoted by S/ρ and called the quotient set of S by the relation ρ. Starting from $\mathbb{Z}$ we have just created the rationals $\mathbb{Q}$ as $(\mathbb{Z} \times \mathbb{Z})/\rho$. The notion of quotient structure frequently arises in physics when we have a group G acting on some set S.

Suppose that G acts transitively on S, that is if Q is a fixed point, and P is any point, there is at least one transformation in G which sends Q to P. The set of all transformations which leave Q fixed is a subgroup H of G – the so-called stabilizer of Q. For any two elements f and g of G we shall say that they are in the relation ρ, that is $f\rho g$, if $fg^{-1} \in H$. We easily prove that ρ is an equivalence relation and that the points of S are in one-to-one correspondence with the elements of G/ρ. Thus the physics of S can be transferred to G/ρ and the symmetries of the physical situation may become more transparent.

When the relation is defined by a subgroup H as above, G/ρ is usually denoted by G/H. Suppose we denote the equivalence class containing g by $\pi(g)$, if g is any element of G. That is π is a mapping from G to G/H, the so-called canonical map. We could ask whether it is possible to impose on G/H a structure of a group in such a way that for any $f, g \in G$, $\pi(fg) = \pi(f)\pi(g)$. The answer is yes – if and only if H is a normal subgroup of G. A normal subgroup is not only a group but has the additional property that for all $g \in G$, $gHg^{-1} = H$. Further $H = \{g \in G | \pi(g) = e\}$ where e is the neutral element of the new group G/H. When the subgroup H is not normal, G/H is not a group but is called a homogeneous space on which G acts transitively.

We shall meet below other examples of the use of quotienting as a method of creating new structures.

5
Rings

A ring like a field consists of a set, R, together with two binary operations which are usually called addition and multiplication. $(R, +, \times)$ is a ring if

(i) $(R, +)$ is a commutative additive group with zero;
(ii) $(R, \times)$ is closed under multiplication and may or may not have a unit;
(iii) multiplication distributes over addition, i.e., $a(x + y) = ax + ay$ for all a, x, and y in R.

We do not require that nonzero elements of R have reciprocals in R, nor that multiplication be commutative or associative, but we do not exclude these properties. Thus a field is a ring but not all rings are fields.

5.1
Examples of Rings

We now list five rings and one "almost ring" which occur frequently in the physics literature.

(a) *The Integers* $\mathbb{Z}$. Perhaps it was this example which led to the emergence of the concept of ring. The integers form a group under addition and are therefore closed under addition and subtraction. They are also closed under multiplication, which distributes over addition. However, the solution, x, of the equation $mx = n$, where $m, n \in \mathbb{Z}$, is not, in general, an element of $\mathbb{Z}$. In contrast with some other rings there are no divisors of zero in the integers. That is you cannot find two integers, neither of which is zero, whose product is zero.

(b) *Square Matrices.* Suppose $A = \left(a^i_j\right)$, $B = \left(b^i_j\right)$, and $C = (c^i_j)$ are $n \times n$ matrices with entries in a field $\mathbb{F}$; then we define $A + B$ and AB or $A \times B$ to be $n \times n$ matrices whose entries in the ith row and jth column are, respectively, $a^i_j + b^i_j$ and $a^i_k b^k_j$. (Recall the summation convention in the Introduction.) Here $1 \le i, j, k \le n$. Let $M_n(\mathbb{F}) = M_n$ denote the set of all $n \times n$ matrices with entries in the commutative field $\mathbb{F}$. Then one can verify that $(M_n, +, \times)$ is an associative ring which is noncommutative if $n \ge 2$. The zero element of the ring is the matrix all of whose entries are 0, whereas the unit or identity for multiplication is the matrix (δ^i_j) which has 1 on the diagonal and 0 elsewhere. Notice that if $n = 2$,

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 3 & 7 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix};$$

thus the ring of square matrices possesses zero divisors.

(c) *Quaternions* were invented by Sir William Rowan Hamilton in order to give a convenient description of rotations in 3-space.
In this section we shall use j, k, s, t as indices with their ranges restricted as follows: $1 \le j, k \le 3$, and $0 \le s, t \le 3$. The quaternions, H, form an associative ring with multiplicative identity, $I = e_0$, and contain three elements e_j satisfying the conditions $e_j e_k + e_k e_j = -2\delta_{jk} e_0$, so $e_j^2 = -e_0$. Further, $e_j e_k = e_m$ where (j, k, m) is an even permutation of $(1, 2, 3)$. As a ring, H will contain

$e_0 + e_0 := 2e_0$, etc., so that H contains $\mathbb{Z}e_0$. More generally if R denotes any commutative ring, we could assume that H contains Re_0 and note this explicitly by denoting the quaternions as $H(R)$. Hamilton considered only the possibility that $R = \mathbb{R}$, the real numbers, since his concern was rotations in the 3-dimensional space of Newtonian physics – not some esoteric space of super string theory! Over R we can define H by

$$H = \{x^s e_s | x^s \in R\}.$$

Then it follows that H is closed under addition and multiplication. If we demand that the associative and distributive properties hold, we obtain a noncommutative associative ring. That it is consistent to demand the preceding properties follows from the fact that they are satisfied by 2×2 matrices with entries in R if we represent e_0 by the identity matrix and e_j by $-i\sigma_j$, where σ_j are the three Pauli matrices:

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix},$$

$$\sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Thus, if we set $E_0 = I$ and $E_j = -i\sigma_j$, we find that

$$X = x^s E_S = \begin{bmatrix} x^0 - ix^3 & -x^2 - ix^1 \\ x^2 - ix^1 & x^0 + ix^3 \end{bmatrix},$$

and that if $x^0 = 0$, then $\det(X) = \delta_{jk} x^j x^k$, which equals the square of the Euclidean length of the vector with components x^j.

If T is any invertible 2×2 matrix and $Y = TXT^{-1}$, then trace of $Y = \mathrm{tr}(Y) = \mathrm{tr}(X)$ and the determinant $\det(Y) = \det(X)$. Since $\mathrm{tr}(X) = x^0$, it follows that $x^0 = 0$ implies $y^0 = 0$. Further, $\delta_{jk} x^j x^k = \delta_{jk} y^j y^k$, that is, Euclidean distance is preserved so the transformation from (x^1, x^2, x^3) to (y^1, y^2, y^3) is orthogonal. In particular if $T = \exp(\vartheta\sigma_3) = \cos\vartheta\, I + \sin\vartheta\,\sigma_3$, this transformation is a rotation though an angle 2ϑ about the x^3 axis.

If R is a finite commutative ring with m elements, $H(R)$ would be a noncommutative ring with m^4 elements.

When is $H(R)$ a field? Define X' by $X = x^0 I + X'$ and $\overline{X}$ by $\overline{X} = x^0 I - X'$. It then follows that $X\overline{X} = \delta_{st} x^s x^t I$. If $R = \mathbb{R}$, this vanishes only if $X = 0$. Thus $\overline{X}$ divided by $\delta_{st} x^s x^t$ is the reciprocal of X. It is not difficult to verify that $H(\mathbb{R})$ satisfies the requirements of an anticommutative or skew field. This is the field discovered by Hamilton to which we alluded in Sec. 2.2.

(d) *Boolean "Ring".* In studying what he called "The Laws of Thought", George Boole was led to introduce an algebraic structure on the subsets of any fixed set in which union, $\cup$, and intersection, $\cap$, are analogs of addition and multiplication, respectively. The original set acts as the identity for multiplication, and the empty set serves as the zero for addition. The reader can verify that most of the properties of a commutative ring are satisfied by Boole's structure, but a given subset does not have an additive inverse so that $\mathscr{P}(S)$, the set of subsets of S, is not an additive group under the binary operation $\cup$.

(e) *Lie Rings.* Let $(L, +, \circ)$ be a set L together with two binary operators such that $(L, +)$ is an additive commutative group such that the operation $\circ$ distributes over addition, so that

$$x \circ (y + z) = x \circ y + x \circ z.$$

However, the Lie product is neither commutative nor associative but satisfies the properties:

$$x \circ y + y \circ x = 0$$

and

$$x \circ (y \circ z) + y \circ (z \circ x) + z \circ (x \circ y) = 0.$$

Because of the first of these conditions, we say that the Lie product is anticommutative. The second, which replaces the associativity property of the familiar rings, is referred to as the Jacobi identity. Lie groups are discussed in other articles of this work so we do not go into details here. We merely remark that the elements of a finite dimensional Lie group can be parametrized by continuous real variables and that Sophus Lie associated to such groups what he called an infinitesimal group which is a particular case of a Lie ring. Associativity of multiplication in the group implies the validity of the Jacobi identity in the corresponding Lie ring. The Jacobi identity can be rewritten in the form

$$z \circ (x \circ y) = (z \circ x) \circ y + x \circ (z \circ y),$$

which is the same as

$$D(x \circ y) = (Dx) \circ y + x \circ (Dy),$$

if we set $Dw = z \circ w$, for fixed z and all $w \in L$. This last equation reminds us of the product rule in calculus, so we say that the linear map $D: w \to z \circ w$ is a derivation of L. The concept of Lie ring, which apparently (Witt, 1937) was first defined by Wilhelm Magnus in 1936, is a generalization of the concept of Lie algebra introduced under the name "infinitesimal group" by Lie and Killing independently before 1880.

(f) *Grassmann Ring.* As a final example of the concept of ring we briefly describe an algebraic structure invented by Hermann Grassmann about 1840 which is basic to the theory of fermions as well as the geometry of many dimensions. Given a field, $\mathbb{F}$, and a finite vector space $(V, \mathbb{F}, +)$ of dimension n, it is possible to define a new vector space, $V^\wedge$, of dimension 2^n over $\mathbb{F}$ and a binary operation, denoted by $\wedge$, called the wedge or Grassmann product, which distributes over addition. $(V^\wedge, \mathbb{F}, +, \wedge)$ will be the Grassmann or exterior algebra. In order to define the product $\wedge$, which is the same as that for fermion creation operators in second quantization, we proceed by induction on the grade of homogeneous elements of the algebra. Recall that in the ring $\mathbb{F}[x, y]$ of all polynomials in x and y there are special subspaces such as $ax + by$, or $ax^2 + bxy + cy^2$, or $ax^3 + bx^2y + cxy^2 + dy^3$, of homogeneous elements of dimension 2, 3, 4, respectively. Any polynomial can be expressed as a sum of homogeneous polynomials, and the summands are unique. It turns out, analogously, that if $\dim(V) = n$, $V^\wedge$ contains $n + 1$ subspaces V^p, $0 \le p \le n$, such that any element x of $V^\wedge$ can be expressed in precisely one way as $x = \Sigma_0^n x^p$, where $x^p \in V^p$. An element of V^p is said to be homogeneous of grade p. If x and y are homogeneous of grades p and q, respectively, then $x \wedge y = (-1)^{pq} y \wedge x$ is of grade $p + q$. In particular, $V^0 = \mathbb{F}$ and $V^1 = V$ by definition. It follows that if x and y are of grade 1, that is belong to V, $x \wedge y = -y \wedge x$. So if $\mathbb{F}$ has characteristic other than 2 it follows that $x \in V$ implies that $x \wedge x = 0$.
If $\{v_i\}$ is a basis of V, the $n(n-1)/2$ elements $v_i \wedge v_j$ for $i < j$ are linearly

independent and span the subspace V^2 of $V^\wedge$. Similarly V^3 is spanned by $v_i \wedge v_j \wedge v_k := (v_i \wedge v_j) \wedge v_k = v_i \wedge (v_j \wedge v_k)$ with $i < j < k$ between 1 and n. The $\dim(V^3) = n(n-1) \quad (n-2)/6$. Proceeding in this way we define all the $n+1$ homogeneous subspaces. As is known, the sum of the coefficients of the nth power of a binomial is $(1+1)^n = 2^n$, so $V^\wedge$ has dimension 2^n. The preceding terse abstract definition does not immediately suggest that the Grassmann ring is significant for fermion physics. However, this becomes plausible when one realizes that the above basis elements of grade p correspond to the Slater determinants for a system of p electrons which can be formed from a basis set of n linearly independent spin-orbitals.

5.2 Polynomial Rings

For everyday applications there is little doubt that the integers $\mathbb{Z}$ constitute the most important ring which is not also a field. Perhaps the next most important is the ring of polynomials involving one or more variables. Suppose R is any ring; then we consider all expressions of the form $P(x) = a_s x^s$, where $0 \le s \le n$, x^s denotes the sth power of the variable x, and $a_s \in R$. If $a_n \neq 0$ we say that $P(x)$ is a polynomial of degree n in x. The set of all such polynomials of arbitrary finite degree will be denoted by $R[x]$. (Note the square bracket which distinguishes the ring from the field $R(x)$ of rational functions.) Assume that the powers of x commute with the elements of R and define addition and multiplication in the obvious way. Then $(R[x], +, \times)$ is a ring which is commutative if and only if R is commutative. For example, if $R = \mathbb{Z}$, $R[x]$ is the ring of all polynomials with integer coefficients. If R is the ring of 2×2 matrices with complex entries, $R[x]$ consists of all 2×2 matrices whose entries are polynomials in x with complex coefficients. In this case the variable is often called λ. The theory of this particular ring is discussed by Turnbull and Aitken (1932), for example, under the title λ-matrices.

An obvious extension of the preceding is to adjoin two or more variables to R. Thus $R[x, y]$ denotes the set of polynomials in x and y with coefficients in R. A term such as $3x^2y^5$, formed by multiplication without addition, is called a monomial. The sum of the powers of x and y is called the degree of the monomial. Thus the degree of the preceding monomial is $2 + 5 = 7$. Clearly there are 8 different monomials of degree 7 in two variables. Any sum of these with coefficients in R is a homogeneous polynomial in x and y of degree 7. When R is a field we see that the homogeneous polynomials of degree 7 form a linear space of dimension 8.

More generally, it is of considerable interest to determine how many distinct monomials of degree n can be obtained from r variables. It is not difficult to see that the possible such monomials occur as the coefficients of t^n in the expansion of the r-fold product $\Pi(1 - x_i t)^{-1}$, where $1 \le i \le r$ and x_i are distinct variables. Setting all $x_i = 1$, we see that the required number is the binomial coefficient $\binom{r+n-1}{n}$.

This is an opportune point at which to explain the concept of a graded ring which appeared in Sec. 5.1(f) and has recently entered quantum physics in connection with super-symmetry. It is clear that any element of $R[x, y]$ is a unique sum of homogeneous terms and that the product of two homogeneous terms of degree p and