Infosys Science Foundation Series in Mathematical Sciences

Ramji Lal

Algebra 2

Linear Algebra, Galois Theory, Representation Theory, Group Extensions and Schur Multiplier

Springer

Infosys Science Foundation Series

Infosys Science Foundation Series in Mathematical Sciences

The *Infosys Science Foundation Series in Mathematical Sciences* is a sub-series of The *Infosys Science Foundation Series*. This sub-series focuses on high quality content in the domain of mathematical sciences and various disciplines of mathematics, statistics, bio-mathematics, financial mathematics, applied mathematics, operations research, applies statistics and computer science. All content published in the sub-series are written, edited, or vetted by the laureates or jury members of the Infosys Prize. With the Series, Springer and the Infosys Science Foundation hope to provide readers with monographs, handbooks, professional books and textbooks of the highest academic quality on current topics in relevant disciplines. Literature in this sub-series will appeal to a wide audience of researchers, students, educators, and professionals across mathematics, applied mathematics, statistics and computer science disciplines.

More information about this series at http://www.springer.com/series/13817

Ramji Lal

Algebra 2

Linear Algebra, Galois Theory, Representation Theory, Group Extensions and Schur Multiplier

Springer

Ramji Lal
Harish Chandra Research Institute (HRI)
Allahabad, Uttar Pradesh
India

ISSN 2363-6149 ISSN 2363-6157 (electronic)
Infosys Science Foundation Series
ISSN 2364-4036 ISSN 2364-4044 (electronic)
Infosys Science Foundation Series in Mathematical Sciences
ISBN 978-981-10-4255-3 ISBN 978-981-10-4256-0 (eBook)
DOI 10.1007/978-981-10-4256-0

Library of Congress Control Number: 2017935547

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Dedicated to the memory of
my mother
(Late) Smt Murti Devi,
my father
(Late) Sri Sankatha Prasad Lal, and
my father like brother
(Late) Sri Gopal Lal

Preface

Algebra has played a central and decisive role in all branches of mathematics and, in turn, in all branches of science and engineering. It is not possible for a lecturer to cover, physically in a classroom, the amount of algebra which a graduate student (irrespective of the branch of science, engineering, or mathematics in which he prefers to specialize) needs to master. In addition, there are a variety of students in a class. Some of them grasp the material very fast and do not need much of assistance. At the same time, there are serious students who can do equally well by putting a little more effort. They need some more illustrations and also more exercises to develop their skill and confidence in the subject by solving problems on their own. Again, it is not possible for a lecturer to do sufficiently many illustrations and exercises in the classroom for the aforesaid purpose. This is one of the considerations which prompted me to write a series of three volumes on the subject starting from the undergraduate level to the advance postgraduate level. Each volume is sufficiently rich with illustrations and examples together with numerous exercises. These volumes also cater for the need of the talented students with difficult, challenging, and motivating exercises which were responsible for the further developments in mathematics. Occasionally, the exercises demonstrating the applications in different disciplines are also included. The books may also act as a guide to teachers giving the courses. The researchers working in the field may also find it useful.

The first volume consists of 11 chapters, which starts with language of mathematics (logic and set theory) and centers around the introduction to basic algebraic structures, viz., groups, rings, polynomial rings, and fields together with fundamentals in arithmetic. This volume serves as a basic text for the first-year course in algebra at the undergraduate level. Since this is the first introduction to the abstract-algebraic structures, we proceed rather leisurely in this volume as compared with the other volumes.

The present (second) volume contains 10 chapters which includes the fundamentals of linear algebra, structure theory of fields and the Galois theory, representation theory of groups, and the theory of group extensions. It is needless to say that linear algebra is the most applicable branch of mathematics, and it is essential

for students of any discipline to develop expertise in the same. As such, linear algebra is an integral part of the syllabus at the undergraduate level. Indeed, a very significant and essential part (Chaps. 1–5) of linear algebra covered in this volume does not require any background material from Volume 1 of the book except some amount of set theory. General linear algebra over rings, Galois theory, representation theory of groups, and the theory of group extensions follow linear algebra, and indeed these are parts of the syllabus for the second- and the third-year students of most of the universities. As such, this volume together with the first volume may serve as a basic text for the first-, second-, and third-year courses in algebra.

The third volume of the book contains 10 chapters, and it can act as a text for graduate and advance graduate students specializing in mathematics. This includes commutative algebra, basics in algebraic geometry, semi-simple Lie algebras, advance representation theory, and Chevalley groups. The table of contents gives an idea of the subject matter covered in the book.

There is no prerequisite essential for the book except, occasionally, in some illustrations and exercises, some amount of calculus, geometry, or topology may be needed. An attempt to follow the logical ordering has been made throughout the book.

My teacher (Late) Prof. B.L. Sharma, my colleague at the University of Allahabad, my friend Dr. H.S. Tripathi, my students Prof. R.P. Shukla, Prof. Shivdatt, Dr. Brajesh Kumar Sharma, Mr. Swapnil Srivastava, Dr. Akhilesh Yadav, Dr. Vivek Jain, Dr. Vipul Kakkar, and above all, the mathematics students of the University of Allahabad had always been the motivating force for me to write a book. Without their continuous insistence, it would have not come in the present form. I wish to express my warmest thanks to all of them.

Harish-Chandra Research Institute (HRI), Allahabad, has always been a great source for me to learn more and more mathematics. I wish to express my deep sense of appreciation and thanks to HRI for providing me all infrastructural facilities to write these volumes.

Last but not least, I wish to express my thanks to my wife Veena Srivastava who had always been helpful in this endeavor.

In spite of all care, some mistakes and misprints might have crept in and escaped my attention. I shall be grateful to any such attention. Criticisms and suggestions for the improvement of the book will be appreciated and gratefully acknowledged.

Allahabad, India Ramji Lal
April 2017

Contents

About the Author

Ramji Lal is Adjunct Professor at the Harish-Chandra Research Institute (HRI), Allahabad, Uttar Pradesh. He started his research career at the Tata Institute of Fundamental Research (TIFR), Mumbai, and served at the University of Allahabad in different capacities for over 43 years: as a Professor, Head of the Department, and the Coordinator of the DSA Program. He was associated with HRI, where he initiated a postgraduate (PG) program in mathematics and coordinated the Nurture Program of National Board for Higher Mathematics (NBHM) from 1996 to 2000. After his retirement from the University of Allahabad, he was Advisor cum Adjunct Professor at the Indian Institute of Information Technology (IIIT), Allahabad, for over 3 years. His areas of interest include group theory, algebraic K-theory, and representation theory.

Notations from Algebra 1

$\langle a \rangle$	Cyclic subgroup generated by a, p. 122
a/b	a divides b, p. 57
$a \sim b$	a is an associate of b, p. 57
A^t	The transpose of a matrix A, p. 200
$A^\star$	The hermitian conjugate of a matrix A, p. 215
$Aut(G)$	The automorphism group of G, p. 105
A_n	The alternating group of degree n, p. 175
$B(n, \mathbb{R})$	Borel subgroup, p. 187
$C_G(H)$	The centralizer of H in G, p. 159
$\mathbb{C}$	The field of complex numbers, p. 78
D_n	The dihedral group of order $2n$, p. 90
det	Determinant map, p. 191
$End(G)$	Semigroup of endomorphisms of G, p. 105
$f(A)$	Image of A under the map f, p. 34
$f^{-1}(B)$	Inverse image of B under the map f, p. 34
$f\|Y$	Restriction of the map f to Y, p. 30
E_{ij}^{λ}	Transvections, p. 200
$Fit(G)$	Fitting subgroup, p. 353
$g.c.d.$	Greatest common divisor, p. 58
$g.l.b.$	Greatest lower bound, or inf, p. 40
$G/^l H (G/^r H)$	The set of left(right) cosets of G mod H, p. 135
G/H	The quotient group of G modulo H, p. 151
$[G : H]$	The index of H in G, p. 135
$\|G\|$	Order of G, p. 331
$G' = [G, G]$	Commutator subgroup of G, p. 403
G^n	nth term of the derived series of G, p. 345
$GL(n, \mathbb{R})$	General linear group, p. 186
I_X	Identity map on X, p. 30
i_Y	Inclusion map from Y, p. 30
$Inn(G)$	The group of inner automorphisms, p. 407

$ker\, f$	The kernel of the map f, p. 35
$L_n(G)$	nth term of the lower central series of G, p. 281
l.c.m.	Least common multiple, p. 58
l.u.b.	Least upper bound, or sup, p. 40
$M_n(R)$	The ring of $n \times n$ matrices with entries in R, p. 350
$\mathbb{N}$	Natural number system, p. 21
$N_G(H)$	Normalizer of H in G, p. 159
$O(n)$	Orthogonal group, p. 197
$O(1, n)$	Lorentz orthogonal group, p. 201
$PSO(1, n)$	Positive special Lorentz orthogonal group, p. 201
$\mathbb{Q}$	The field of rational numbers, p. 74
Q_8	The quaternion group, p. 88
$\mathbb{R}$	The field of real numbers, p. 75
$R(G)$	Radical of G, p. 346
S_n	Symmetric group of degree n, p. 88
$Sym(X)$	Symmetric group on X, p. 88
S^3	The group of unit quaternions, p. 92
$\langle S \rangle$	Subgroup generated by a subset S, p. 116
$SL(n, \mathbb{R})$	Special linear group, p. 196
$SO(n)$	Special orthogonal group, p. 197
$SO(1, n)$	Special Lorentz orthogonal group, p. 201
$SP(2n, \mathbb{R})$	Symplectic group, p. 202
$SU(n)$	Special unitary group, p. 202
$U(n)$	Unitary group, p. 202
U_m	Group of prime residue classes modulo m, p. 100
V_4	Kleins four group, p. 102
X/R	The quotient set of X modulo R, p. 36
R_x	Equivalence class modulo R determined by x, p. 27
X^+	Successor of X, p. 20
X^Y	The set of maps from Y to X, p. 34
$\subset$	Proper subset, p. 14
$\wp(X)$	Power set of X, p. 19
$\prod_{k=1}^{n} G_k$	Direct product of groups $G_k, 1 \leq k \leq n$, p. 142
$\trianglelefteq$	Normal subgroup, p. 147
$\trianglelefteq\trianglelefteq$	Subnormal subgroup, p. 332
$Z(G)$	Center of G, p. 108
$\mathbb{Z}_m$	The ring of residue classes modulo m, p. 256
$p(n)$	The number of partition of n, p. 172
$H \prec K$	Semidirect product of H with K, p. 204
$\sqrt{A}$	Radical of an ideal A, p. 286
$R(G)$	Semigroup ring of a ring R over a semigroup G, p. 238
$R[X]$	Polynomial ring over the ring R in one variable, p. 240
$R[X_1, X_2, \cdots, X_n]$	Polynomial ring in several variables, p. 247
μ	The Mobius function, p. 256

σ	Sum of divisor function, p. 256
$\left(\frac{a}{p}\right)$	Legendre symbol, p. 280
$Stab(G, X)$	Stabilizer of an action of G on X, p. 295
G_x	Isotropy subgroup of an action of G at x, p. 295
X^G	Fixed point set of an action of G on X, p. 296
$Z_n(G)$	nth term of the upper central series of G, p. 351
$\Phi(G)$	The Frattini subgroup of G, p. 355

Notations from Algebra 2

$B^2_\sigma(K,H)$	Group of 2 co-boundaries with given σ, p. 385
$C(A)$	Column space of A, p. 42
$Ch(G, K)$	Set of characters from G to K, p. 278
$Ch(G)$	Character ring of G, p. 350
$dim(V)$	Dimension of V, p. 18
EXT	Category of Schreier group extensions, p. 368
$E(H, K)$	The set of equivalence classes of extensions of H by K, p. 376
$E_1 \uplus E_2$	Baer sum of extensions, p. 388
$EXT_\psi(H, \mathrm{K})$	Set of equivalence classes of extensions associated to abstract kernel ψ, p. 384
$E(V)$	Exterior algebra of V, p. 257
$FACS$	Category of factor systems, p. 375
$F(X)$	The fixed field of a set of automorphism of a field, p. 275
$G(L/K)$	The Galois group of the field extension L of K, p. 275
$G \wedge G$	Non-abelian exterior square of a group G, p. 413
$\bar{K}$	Algebraic closure of K, p. 289
$H^2_\sigma(K,H)$	Second cohomology with given σ, p. 385
$\mathrm{K}_0(R)$	Grothendieck group of the ring R, p. 257
$\mathrm{K}_1(R)$	Whitehead group of the ring R, p. 260
K_S^L	Separable closure of K in L, p. 295
L/K	Field extension L of K, p. 262
$m_T(X)$	Minimum polynomial of linear transformation T, p. 212
$min_K(\alpha)(\mathrm{X})$	Minimum polynomial of α over the field K, p. 265
$M(V)$	Group of rigid motion on V, p. 122
$M \otimes_R N$	Tensor product of R-modules M and N, p. 250
$N_{L/K}$	Norm map from L to K, p. 279
$N(A)$	Null space of A, p. 41
$Obs(\psi)$	Obstruction of the abstract kernel ψ, p. 393
$R(A)$	Row space of A, p. 42
$St(R)$	Steinberg group, p. 422

$Sym^r(V)$	rth symmetric power of V, p. 345
$T_{L/K}$	Trace map from L to K, p. 314
$T(V)$	Tensor algebra of V, p. 257
T_S	Semi-simple part of T, p. 219
T_n	Nilpotent part of T, p. 220
$Z^2_\sigma(K,H)$	Group of 2 co-cycles with given σ, p. 385
$\bigwedge^r V$	rth exterior power of V, p. 255
Ψ_E	Abstract kernel associated to the extension E, p. 377
$\rho \oplus \eta$	Direct sum of representations ρ and η, p. 345
$\rho \otimes \eta$	Tensor product of representations ρ and η, p. 345
$Sym^r\rho$	rth symmetric power of the representation ρ, p. 345
$SF(L/K)$	Set of all intermediary fields of L/K, p. 275
$\bigwedge^r \rho$	rth exterior power of the representation ρ, p. 345
χ_ρ	Character afforded by the representation ρ, p. 350
$\phi_n(X)$	nth cyclotomic polynomial, p. 311
$\phi_A(X)$	Characteristic polynomial of A, p. 149

Chapter 1
Vector Spaces

This chapter is devoted to the structure theory of vector spaces over arbitrary fields. In essence, a vector space is a structure in which we can perform all basic operations of vector algebra, can talk of lines, planes, and linear equations. The basic motivating examples on which we shall dwell are the Euclidean 3-space $\mathbb{R}^3$ over $\mathbb{R}$ in which we live, the Minkowski Space $\mathbb{R}^4$ of events (in which the first three coordinates represent the place and the fourth coordinate represents the time of the occurrence of the event), and also the space of matrices.

1.1 Concept of a Field

Rings and fields have been introduced and studied in Algebra 1. However, to make the linear algebra part (Chaps. 1–5) of this volume independent of Algebra 1, we recall, quickly, the concept of a field and its basic properties. Field is an algebraic structure in which we can perform all arithmetical operations, viz., addition, subtraction, multiplication, and division by nonzero members. The basic motivating examples are the structure $\mathbb{Q}$ of rational numbers, the structure $\mathbb{R}$ of real numbers, and the structure $\mathbb{C}$ of complex numbers with usual operations. The precise definition of a field is as follows:

Definition 1.1.1 A **Field** is a triple $(F, +, \cdot)$, where F is a set, $+$ and $\cdot$ are two internal binary operations, called the addition and the multiplication on F, such that the following hold:

1. $(F, +)$ is an abelian Group in the following sense:
(i) The operation $+$ is associative in the sense that
$(a + b) + c = a + (b + c)$ *for all* $a, b, c \in F$.
(ii) The operation $+$ is commutative in the sense that
$(a + b) = (b + a)$ *for all* $a, b \in F$.

R. Lal, *Algebra 2*, Infosys Science Foundation Series in Mathematical Sciences,
DOI 10.1007/978-981-10-4256-0_1

(iii) There is a unique element $0 \in F$, called the zero of F, such that
$a + 0 = a = 0 + a$ *for all* $a \in F$.
(iv) For all $a \in F$, there is a unique element $-a \in F$, called the negative of a, such that
$a + (-a) = 0 = -a + a$.
2. (i) The operation $\cdot$ is associative in the sense that
$(a \cdot b) \cdot c = a \cdot (b \cdot c)$ *for all* $a, b, c \in F$.
(ii) The operation $\cdot$ is commutative in the sense that
$(a \cdot b) = (b \cdot a)$ *for all* $a, b \in F$.
3. The operation $\cdot$ distributes over $+$ in the sense that
(i) $a \cdot (b + c) = a \cdot b + a \cdot c$, and
(ii) $(a + b) \cdot c = a \cdot c + b \cdot c$ *for all* $a, b, c \in F$.
4. (i) There is a unique element $1 \in F - \{0\}$, called the one of F, such that
$1 \cdot a = a = a \cdot 1$ *for all* $a \in F$.
(ii) For all $a \in F - \{0\}$, there is a unique element $a^{-1} \in F$, called the multiplicative inverse of a, such that
$a \cdot a^{-1} = 1 = a^{-1} \cdot a$.

Before having some examples, let us observe some simple facts:

Proposition 1.1.2 *Let* $(F, +, \cdot)$ *be a field.*

(i) *The cancellation law holds for the addition* $+$ *in* F *in the sense that* $(a + b = a + c)$ *implies* $b = c$. *In turn,* $(b + a = c + a)$ *implies* $b = c$.
(ii) $a \cdot 0 = 0 = 0 \cdot a$ *for all* $a \in F$.
(iii) $a \cdot (-b) = -(a \cdot b) = (-a) \cdot b$ *for all* $a, b \in F$.
(iv) *The restricted cancellation for the multiplication in* F *holds in the sense that* $(a \neq 0$ *and* $a \cdot b = a \cdot c)$ *implies* $b = c$. *In turn,* $(a \neq 0$ *and* $b \cdot a = c \cdot a)$ *implies* $b = c$.
(v) $(a \cdot b = 0)$ *implies that* $(a = 0$ *or* $b = 0)$.

Proof (i) Suppose that $a + b = a + c$. Then $b = 0 + b = (-a + a) + b = -a + (a + b) = -a + (a + c) = (-a + a) + c = 0 + c = c$.
(ii) $0 + a \cdot 0 = a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0$. Using the cancellation for $+$, we get that $0 = a \cdot 0$. Similarly, $0 = 0 \cdot a$.
(iii) $0 = a \cdot 0 = a \cdot (b + (-b)) = a \cdot b + a \cdot (-b)$. It follows that $a \cdot (-b) = -(a \cdot b)$. Similarly, the other part follows.
(iv) Suppose that $a \neq 0$ and $a \cdot b = a \cdot c$. Then $b = 1 \cdot b = (a^{-1} \cdot a) \cdot b = a^{-1} \cdot (a \cdot b) = a^{-1} \cdot (a \cdot c) = (a^{-1} \cdot a) \cdot c = 1 \cdot c = c$. Similarly, the other part follows.
(v) Suppose that $(a \cdot b = 0)$. If $a = 0$, there is nothing to do. Suppose that $a \neq 0$. Then $a \cdot b = 0 = a \cdot 0$. From (iv), it follows that $b = 0$. ♯

Integral Multiples and the Integral Powers of Elements of a Field
Let $a \in F$. For each natural number n, we define the multiple na inductively as follows: Define $1a = a$. Assuming that na is defined, define $(n + 1)a = na + a$.

Thus, for a natural number n, $na = \underbrace{a + a + \cdots + a}_{n\,times}$. We define $0a = 0$. Further, if $m = -n$ is a negative integer, then we define $ma = n(-a)$. Thus, for a negative integer $m = -n$, $ma = \underbrace{-a + (-a) + \cdots + (-a)}_{n\,times}$. This defines the integral multiple na for each integer n. Similarly, we define all integral powers of a nonzero element a of F as follows: Define $a^1 = a$. Assuming that a^n has already been defined, define $a^{n+1} = a^n \cdot a$. This defines all positive integral powers of a. Define $a^0 = 1$, and for negative integer $n = -m$, define $a^n = (a^{-1})^m$. The following law of exponents follow immediately by the induction.

(i) $(n + m)a = na + ma$ *for all* $n, m \in \mathbb{Z}$.
(ii) $(nm)a = n(ma)$ *for all* $n, m \in \mathbb{Z}$.
(iii) $a^{n+m} = a^n \cdot a^m$ *for all* $a \in F - \{0\}$, *and* $n, m \in \mathbb{Z}$.
(iv) $a^{nm} = (a^n)^m$ *for all* $a \in F - \{0\}$, *and* $n, m \in \mathbb{Z}$.

Examples of Fields

Example 1.1.3 The rational number system $\mathbb{Q}$, the real number system $\mathbb{R}$, and the complex number system $\mathbb{C}$ with usual addition and multiplications are basic examples of a field.

Example 1.1.4 Consider $F = \mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$. The addition and multiplication in $\mathbb{R}$ induce the corresponding operations in $\mathbb{Q}(\sqrt{2})$. We claim that $\mathbb{Q}(\sqrt{2})$ is a field with respect to the induced operations. All the defining properties of a field are consequences of the corresponding properties in $\mathbb{R}$ except, perhaps, 4(ii) which we verify. Let $a, b \in \mathbb{Q}$ such that $a + b\sqrt{2} \neq 0$. We claim that $a^2 - 2b^2 \neq 0$. Suppose not. Then $a^2 - 2b^2 = 0$. In turn, $b = 0$ (and so also $a = 0$), otherwise, $(\frac{a}{b})^2 = 2$, a contradiction to the fact that $\sqrt{2}$ is not a rational number. Thus, then $\frac{1}{a+b\sqrt{2}} = \frac{a-b\sqrt{2}}{a^2-2b^2} = \frac{a}{a^2-2b^2} + \frac{-b}{a^2-2b^2}\sqrt{2}$ is in $\mathbb{Q}(\sqrt{2})$.

Remark 1.1.5 There is nothing special about 2 in the above example, indeed, we can take any prime, or for that matter any rational number in place of 2 which is not a square of a rational number.

So far all the examples of fields are infinite. Now, we give an example of a finite field.

Let p be a positive prime integer. Consider the set $\mathbb{Z}_p = \{\overline{1}, \overline{2}, \ldots, \overline{p-1}\}$ of residue classes modulo a prime p. Clearly, $\overline{a} = \overline{r}$, where r is the remainder obtained when a is divided by p. The usual addition $\oplus$ modulo p, and the multiplication $\star$ modulo p are given by

$$\overline{i} \oplus \overline{j} = \overline{i+j},\ i, j \in \mathbb{Z},$$

and

$$\overline{i} \star \overline{j} = \overline{i \cdot j},\ i, j \in \mathbb{Z}$$

For example, in $\mathbb{Z}_{11}$, $\overline{6} \oplus \overline{7} = \overline{13} = \overline{2}$. Similarly, the product $\overline{6} \star \overline{7} = \overline{42} = \overline{9}$. We have the following proposition.

Proposition 1.1.6 *For any prime p, the triple* $(Z_p, \oplus, \star)$ *introduced above is a field containing p elements.*

Proof Clearly, $\overline{1}$ is the identity with respect to $\star$. We verify only the postulate 4(ii) in the definition of a field. The rest of the postulates are almost evident, and can be verified easily. In fact, we give an algorithm (using Euclidean Algorithm) to find the multiplicative inverse of a nonzero element $\overline{i} \in Z_p$. Let $\overline{i} \in Z_p - \{\overline{0}\}$. Then p does not divide i. Since p is prime, the greatest common divisor of i and p is 1. Using the Euclidean algorithm, we can find integers b and c such that

$$1 = i \cdot b + p \cdot c.$$

Thus, $\overline{1} = \overline{i \cdot b} = \overline{i} \star \overline{b}$. It follows that $\overline{b}$ is the inverse of $\overline{i}$ with respect to $\star$. ♯

The above proof is algorithmic and gives an algorithm to find the multiplicative inverse of nonzero elements in Z_p.

Definition 1.1.7 Let $(F, +, \cdot)$ be a field. A subset L of F is called a *subfield* of F if the following hold:

(i) $0 \in L$.
(ii) If $a, b \in L$, then $a + b \in L$ *and* $a \cdot b \in L$.
(iii) $1 \in L$.
(iv) For all $a \in L$, $-a \in L$.
(v) For all $a \in L - \{0\}$, $a^{-1} \in L$.

Thus, a subfield L of a field F is also a field at its own right with respect to the induced operations. The field F is a subfield of itself. This subfield is called the improper subfield of F. Other subfields are called **proper subfields**. The set $\mathbb{Q}$ of rational numbers, the set $\mathbb{Q}(\sqrt{2})$ described in Example 1.1.4, are proper subfields of the field $\mathbb{R}$ of real numbers. The field $\mathbb{R}$ of real numbers is a subfield of the field $\mathbb{C}$ of complex numbers.

Proposition 1.1.8 *The field* $\mathbb{Q}$ *of rational numbers, and the field* $\mathbb{Z}_p$ *have no proper subfields.*

Proof We first show that $\mathbb{Q}$ has no proper subfields. Let L be a subfield of $\mathbb{Q}$. Then by the Definition 1.1.7(iii), $1 \in L$. Again, by (ii), $n = \underbrace{1 + 1 + \cdots + 1}_{n}$ belongs to L for all natural numbers n. Thus, by (iv), all integers are in L. By (v), $\frac{1}{n} \in L$ for all nonzero integers n. By (ii), $\frac{m}{n} \in L$ for all integers $m, n; n \neq 0$. This shows that $L = \mathbb{Q}$.

Next, let L be a subfield of $\mathbb{Z}_p$. Then by the Definition 1.1.7(iii), $\overline{1} \in L$. By (ii), $\overline{i} = \underbrace{\overline{1} \oplus \overline{1} \oplus \cdots \oplus \overline{1}}_{i}$ belongs to L for all $i \in \mathbb{Z}_p$. This shows that $L = \mathbb{Z}_p$. ♯

We shall see that, essentially, these are the only fields which have no proper subfields. Such fields are called **prime fields**.

Homomorphisms and Isomorphisms Between Fields

Definition 1.1.9 Let F_1 and F_2 be fields. A map f from F_1 to F_2 is called a *fieldhomomorphism* if the following conditions hold:

(i) $f(a + b) = f(a) + f(b)$ for all $a, b \in F_1$ (note that + in the LHS is the addition of F_1, and that in RHS is the addition of F_2).
(ii) $f(a \cdot b) = f(a) \cdot f(b)$ for all $a, b \in F_1$ (again $\cdot$ in the LHS is the multiplication of F_1, and that in RHS is the multiplication of F_2).
(iii) $f(1) = 1$, where 1 in the LHS denotes the multiplicative identity of F_1, and 1 in RHS denotes the multiplicative identity of F_2.
A bijective homomorphism is called an **isomorphism**. A field F_1 is said to be isomorphic a field F_2 if there is an isomorphism from F_1 to F_2.

We do not distinguish isomorphic fields.

Proposition 1.1.10 *Let f be a homomorphism from a field F_1 to a field F_2. Then, the following hold.*

(i) $f(0) = 0$, *where 0 in the LHS is the zero of F_1, and 0 in the RHS is the zero of F_2.*
(ii) $f(-a) = -f(a)$ *for all* $a \in F_1$.
(iii) $f(na) = nf(a)$ *for all* $a \in F_1$, *and for all integer n.*
(iv) $f(a^n) = (f(a))^n$ *for all* $a \in F_1 - \{0\}$, *and for all integer n.*
(v) f is injective, and the image of F_1 under f is a subfield of F_2 which is isomorphic to F_1.

Proof (i) $0 + f(0) = f(0) = f(0 + 0) = f(0) + f(0)$. Using cancellation law for addition in F_2, we get that $f(0) = 0$.
(ii) $0 = f(0) = f(a + (-a)) = f(a) + f(-a)$. This shows that $f(-a) = -f(a)$.
(iii) Suppose that $n = 0$. Then $0f(a) = 0 = f(0) = f(0a)$. Clearly, $f(1a) = f(a) = 1f(a)$. Assume that $f(na) = nf(a)$ for a natural number n. Then $f(n + 1)a = f(na + a) = f(na) + f(a) = nf(a) + f(a) = (n + 1)f(a)$. By induction, it follows that $f(na) = nf(a)$ for all $a \in F_1$, and for all natural number n. Suppose that $n = -m$ is a negative integer. Then, $f(na) = f((-m)a) = f(-(ma)) = -f(ma) = -(mf(a)) = -(m)f(a) = nf(a)$.
(iv) Replacing na by a^n, imitate the proof of (iii).
(v) Suppose that $a \neq b$. Then $(a - b) \neq 0$. Now, $1 = f(1) = f((a - b)(a - b)^{-1}) = f(a - b)f((a - b)^{-1})$. Since $1 \neq 0$, it follows that $(f(a) - f(b)) = f(a - b) \neq 0$. This shows that $f(a) \neq f(b)$. Thus, f is injective, and it can be realized as a bijective map from F_1 to $f(F_1)$. It is sufficient, therefore, to show that $f(F_1)$ is a subfield of F_2. Clearly, $0 = f(0)$, and $1 = f(1)$ belong to $f(F_1)$. Let

$f(a),\ f(b) \in f(F_1)$, where $a, b \in F_1$. Then $(f(a) + f(b)) = f(a+b) \in f(F_1)$, and also $(f(a)f(b)) = f(ab) \in f(F_1)$. Finally, if $f(a) \neq 0$, then $a \in F_1 - \{0\}$. But, then $(f(a))^{-1} = f(a^{-1}) \in F_1$. ♯

Characteristic of a Field

Let F be a field. Consider the multiplicative identity 1 of F. There are two cases:

(i) Distinct integral multiples of 1 are distinct, or equivalently, $n1 = m1$ *implies that* $n = m$. This is equivalent to say that $n1 = 0$ if and only if $n = 0$. In this case we say that F is of **characteristic** 0. Thus, for example, the field $\mathbb{R}$ of real numbers, the field $\mathbb{Q}$ of rational numbers, and the field $\mathbb{C}$ of complex numbers are the fields of characteristic 0.

(ii) Not all integral multiples of 1 are distinct. In this case there exists a pair $n,\ m$ of distinct integers such that $n1 = m1$. But, then, $(n-m)1 = 0 = (m-n)1$. In turn, there is a natural number l such that $l1 = 0$. In this case, the smallest natural number l such that $l1 = 0$ is called the **characteristic** of F. Thus, the characteristic of $\mathbb{Z}_p$ is p.

Proposition 1.1.11 *The characteristic of a field is either 0 or a prime number p. A field of characteristic 0 contains a subfield isomorphic to the field* $\mathbb{Q}$ *of rational numbers, and a field of characteristic p contains a subfield isomorphic to the field* $\mathbb{Z}_p$.

Proof Suppose that F is a field of characteristic 0. Then $n1 = m1$ implies that $n = m$. Also $(m1 \neq 0)$ *if and only if* $(m \neq 0)$. Suppose that $(\frac{m}{n} = \frac{r}{s})$. Then $(m1)(s1) = ms1 = nr1 = (n1)(r1)$. In turn, $((m1)(n1)^{-1} = (r1)(s1)^{-1})$. Thus, we have a map f from $\mathbb{Q}$ to F given by $f(\frac{m}{n}) = (m1)(n1)^{-1}$. Next, suppose that $((m1)(n1)^{-1} = (r1)(s1)^{-1})$. Then $ms1 = (m1)(s1) = (n1)(r1) = nr1$. This means that $ms = nr$, or equivalently, $(\frac{m}{n} = \frac{r}{s})$. This shows that f is an injective map. It is also straight forward to verify that f is a field homomorphism. Thus, $L = \{(m1)(n1)^{-1} \mid m \in \mathbb{Z},\ n \in \mathbb{Z} - \{0\}\}$ is a subfield of F which is isomorphic to $\mathbb{Q}$.

Next, suppose that the characteristic of F is $l \neq 0$. Then l is the smallest natural number such that $l1 = 0$. We show that l is a prime p. Suppose not. Then $l = l_1 l_2,\ 1 < l_1 < l,\ 1 < l_2 < l$. But, then $0 = l1 = (l_1 l_2)1 = (l_1 1)(l_2 1)$. In turn, $l_1 1 = 0$ or $l_2 1 = 0$. This is a contradiction to the choice of l. Thus, the characteristic of F is a prime p. Suppose that $\bar{i} = \bar{j}$. Then p divides $i - j$. In turn, $(i - j)1 = 0$, and so $i1 = j1$. Thus, we have a map f from $\mathbb{Z}_p$ to F defined by $f(\bar{i}) = i1$. Clearly, this is an injective field homomorphism. ♯

Exercises

1.1.1 Show that $\mathbb{Q}(\omega) = \{a + b\omega \mid a, b \in \mathbb{Q}\}$, where ω a primitive cube root of 1, is a subfield of the field $\mathbb{C}$ of complex numbers.

1.1.2 Show that $\sqrt{\sqrt{2}}$ is not a member of $\mathbb{Q}(\sqrt{2})$. Use the method of Example 1.1.4 to show that $\mathbb{Q}(\sqrt{2})(\sqrt{\sqrt{2}}) = \{a + b\sqrt{2} + (c + d\sqrt{2})(\sqrt{\sqrt{2}}) \mid a, b, c, d \in \mathbb{Q}\}$ is a field with respect to the addition and multiplication induced by those in $\mathbb{R}$. Generalize the assertion.

1.1.3 Show that $\mathbb{Q}(\sqrt{2})(\sqrt{3}) = \{a + b\sqrt{2} + (c + d\sqrt{2})(\sqrt{3}) \mid a, b, c, d \in Q\}$ is a field with respect to the addition and multiplication induced by those in $\mathbb{R}$.

1.1.4 Show that $\mathbb{Q}(2^{\frac{1}{3}}) = \{a + b2^{\frac{1}{3}} + c2^{\frac{2}{3}} \mid a, b, c \in \mathbb{Q}\}$ is also a field with respect to the addition and multiplication induced by those in $\mathbb{R}$. Express $\frac{1}{1+2^{\frac{1}{3}}}$ as $a + b2^{\frac{1}{3}} + c2^{\frac{2}{3}}$, $a, b, c \in \mathbb{Q}$.

1.1.5 Show that $F = \{0, 1, \alpha, \alpha^2\}$ is a field of characteristic 2 with respect to the addition + and multiplication $\cdot$ given by the following tables:

$+$	0	1	α	α^2
0	0	1	α	α^2
1	1	0	α^2	α
α	α	α^2	0	1
α^2	α^2	α	1	0

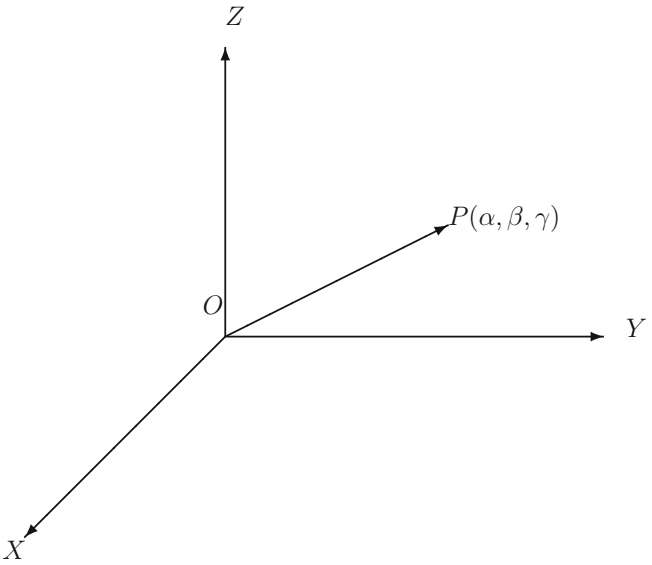
$\cdot$	0	1	α	α^2
0	0	0	0	0
1	0	1	α	α^2
α	0	α	α^2	1
α^2	0	α^2	1	α

1.1.6 Find the multiplicative inverse of $\overline{20}$ in $\mathbb{Z}_{257}$, and also find the solution of $\overline{10}x \oplus \overline{2} = 3$.

1.1.7 Write a program in C++ language to check if a natural number n is prime, and if so to find the multiplicative inverse of a nonzero element $\overline{m}$ in $\mathbb{Z}_n$. Find the output with $n = 2^{2^4} + 1$, and $m = 641$.

1.2 Concept of a Vector Space (Linear Space)

Consider the space (called the Euclidean 3-space) in which we live. If we fix a point (place) in the three space as origin together with three mutually perpendicular lines (directions) passing through the origin as the axes of reference, and also a segment of line as a unit of length, then any point in the 3-space determines, and it is determined uniquely by an ordered triple (α, β, γ) of real numbers.



Thus, with the given choice of the origin and the axes as above, the space in which we live can be represented faithfully by

$$\mathbb{R}^3 = \{\overline{x} = (x_1, x_2, x_3) \mid x_1, x_2, x_3 \in \mathbb{R}\},$$

and it is called the **Euclidean 3-space**. The members of $\mathbb{R}^3$ are called the usual **3-vectors**. It is also evident that the physical quantities which have magnitudes as well as directions (e.g., force, velocity, or displacement) can be represented by vectors. More generally, for a fixed natural number n,

$$\mathbb{R}^n = \{\overline{x} = (x_1, x_2, \ldots, x_n) \mid x_1, x_2, \ldots, x_n \in R\}$$

is called the **Euclidean n-space**, and the members of the Euclidean n-space are called the **Euclidean n-vectors**. We term $x_1, x_2, \ldots, x_n$ as components, or coordinates of the vector $\overline{x} = (x_1, x_2, \ldots, x_n)$. Thus, $\mathbb{R}^2$ represents the Euclidean plane, and $\mathbb{R}^4$ represents the **Minkowski space** of events in which the first three coordinates represent the place, and the fourth coordinate represents the time of the occurrence of the event. $\mathbb{R}^1$ is identified with $\mathbb{R}$. By convention, $\mathbb{R}^0 = \{0\}$ is a single point. We have the addition + in $\mathbb{R}^n$, called the addition of vectors, and it is defined by

$$\overline{x} + \overline{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n),$$

where $\overline{x} = (x_1, x_2, \ldots, x_n)$ and $\overline{y} = (y_1, y_2, \ldots, y_n)$. We have also the external multiplication $\cdot$ by the members of $\mathbb{R}$, called the multiplication by scalars, and it is given by

$$\alpha \cdot \overline{x} = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n), \alpha \in \mathbb{R}.$$

Remark 1.2.1 The addition + of vectors in 3-space $\mathbb{R}^3$ is the usual addition of vectors, which obeys the parallelogram law of addition.

The **Euclidean 3-space** $(\mathbb{R}^3,\ +,\ \cdot)$ introduced above is a **Vector Space** in the sense of the following definition:

Definition 1.2.2 A **Vector Space** (also called a **Linear Space**) over a field F (called the field of **Scalars**) is a triple $(V,\ +,\ \cdot)$, where V is a set, $+$ is an internal binary operation on V, called the addition of vectors, and $\cdot : F \times V \to V$ is an external multiplication, called the multiplication by scalars, such that the following hold:
A. $(\mathbf{V},\ +)$ is an abelian group in the sense that:

1. + is associative, i.e.,
$$(x + y) + z \ = \ x + (y + z)$$
for all x, y, z in V.
2. + is commutative, i.e.,
$$x + y \ = \ y + x$$
for all x, y in V.
3. We have a unique vector 0 in V, called the null vector, and it is such that
$$x + 0 \ = \ x \ = \ 0 + x$$
for all x in V.
4. For each x in V, we have a unique vector $-x$ in V, called the negative of x, and it is such that
$$x + (-x) = 0 = (-x) + x.$$

B. The external multiplication $\cdot$ by scalars satisfies the following conditions:

1. It distributes over the vector addition + in the sense that
$$\alpha \cdot (x + y) = \alpha \cdot x + \alpha \cdot y$$
for all $\alpha \in F$ and $x,\ y$ in V.
2. It distributes over the addition of scalars also in the sense that
$$(\alpha + \beta) \cdot x = \alpha \cdot x + \beta \cdot x$$
for all $\alpha, \beta \in F$ and x in V.
3. $(\alpha\beta) \cdot x = \alpha \cdot (\beta \cdot x)$ for all $\alpha, \beta \in F$ and x in V.
4. $1 \cdot x = x$ for all x in V.

Example 1.2.3 Let F be a field, and n be a natural number. Consider the set
$$V \ = \ F^n \ = \ \{\overline{x} \ = \ (x_1, x_2, \ldots, x_n) \mid x_1, x_2, \ldots, x_n \ \in \ F\}$$

of row vectors with n columns, and with entries in F. We have the addition + in F^n defined by

$$\bar{x} + \bar{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n),$$

where $\bar{x} = (x_1, x_2, \ldots, x_n)$ and $\bar{y} = (y_1, y_2, \ldots, y_n)$. We have also the external multiplication $\cdot$ by the members of F defined by

$$\alpha \cdot \bar{x} = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n), \alpha \in F.$$

The field properties of F ensures that the triple $(F^n, +\cdot)$ is a vector space over F. The zero of the vector space is the zero row $\bar{0} = (0, 0, \ldots, 0)$, and the negative of $\bar{x} = (x_1, x_2, \ldots, x_n)$ is $-\bar{x} = (-x_1, -x_2, \ldots, -x_n)$. We can also treat the members of F^n as column vectors.

Example 1.2.4 Let L be a subfield of a field F. Consider $(F, +, \cdot)$, where $+$ is the addition of the field F, and $\cdot$ is the restriction of the multiplication in F to $L \times F$. Then it is evident that $(F, +, \cdot)$ is a vector space over L. Thus, every field can be considered as vector spaces over its subfields.

Example 1.2.5 Let $C[0, 1]$ denote the set of all real valued continuous functions on the closed interval $[0, 1]$. Since sum of any two continuous functions is a continuous function, we have an addition on $C[0, 1]$ with respect to which it is an abelian group. Define the external multiplication $\cdot$ by $(a \cdot f)(x) = a \cdot f(x)$. Then $C[0, 1]$ is a vector space over the field $\mathbb{R}$ of reals. Note that the set $D[0, 1]$ of differentiable functions is also a vector space over the field $\mathbb{R}$ of reals with respect to the addition of functions, and multiplication by scalars as defined above.

Example 1.2.6 Let $P_n(F)$ denote the set of all polynomials of degree at most n over a field F. Then $P_n(F)$ is an abelian group with respect to the addition of polynomials. Further, if $a \in F$ and $f(X) \in P_n(F)$, then $af(X) \in P_n(F)$. Thus, $P_n(F)$ is also a vector space over F.

Proposition 1.2.7 *Let V be a vector space over a field F. Then the following hold:*

(i) The cancellation law holds in $(V, +)$ in the sense that $(x + y = x + z)$ implies $y = z$ (In turn, $(y + x = z + x)$ implies $y = z$).
(ii) $0 \cdot x = 0$, where 0 in the left side is the 0 of F, 0 on right side is that of V, and $x \in V$.
(iii) $\alpha \cdot 0 = 0$, where both 0 are that of V, and $\alpha \in F$.
(iv) $(-\alpha) \cdot x = -(\alpha \cdot x)$ for all $\alpha \in F$, and $x \in V$. In particular, $(-1) \cdot x = -x$.
(v) $(\alpha \cdot x = 0)$ implies that $(\alpha = 0$ or $x = 0)$.

Proof (i) Suppose that $(x + y = x + z)$. Then $y = 0 + y = (-x + x) + y = -x + (x + y) = -x + (x + z) = (-x + x) + z = 0 + z = z$.
(ii) $0 + 0 \cdot x = 0 \cdot x = (0 + 0) \cdot x = 0 \cdot x + 0 \cdot x$. By the cancellation in $(V, +)$,

$0 = 0 \cdot x$.
(iii) $0 + \alpha \cdot 0 = \alpha \cdot 0 = \alpha \cdot (0 + 0) = \alpha \cdot 0 + \alpha \cdot 0$. By the cancellation in $(V, +)$, $0 = \alpha \cdot 0$.
(iv) $0 = 0 \cdot x = (-\alpha + \alpha) \cdot x = (-\alpha) \cdot x + \alpha \cdot x$. This shows that $(-\alpha) \cdot x = -(\alpha \cdot x)$
(v) Suppose that $(\alpha \cdot x = 0)$, and $\alpha \neq 0$. Then, $x = 1 \cdot x = (\alpha^{-1}\alpha) \cdot x = \alpha^{-1} \cdot (\alpha \cdot x) = \alpha^{-1} \cdot 0 = 0$. ♯

1.3 Subspaces

Definition 1.3.1 Let V be a vector space over a field F. A subset W of V is called a **subspace**, or a **linear subspace** of V if

(i) $0 \in W$.
(ii) $x + y \in W$ for all $x, y \in W$.
(iii) $\alpha \cdot x \in W$ for all $\alpha \in F$ and $x \in W$.

Thus, a subspace is also a vector space over the same field at its own right.

Proposition 1.3.2 *Let V be a vector space over a field F. Then a nonempty subset W of V is a subspace if and only if $ax + by \in W$ for all $a, b \in F$, and $x, y \in V$.*

Proof Suppose that W is a subspace of V. Let $a, b \in F$, and $x, y \in V$. From the Definition 1.3.1(i), $ax, by \in W$. In turn, by Definition 1.3.1(ii), $ax + by \in W$. Conversely, suppose that W is a nonempty subset of V such that $ax + by \in W$ *for all* $a, b \in F$, and for all $x, y \in W$. Let $x, y \in W$. Then $x + y = 1x + 1y$ belongs to W. Further, since W is nonempty, there is an element $x \in W$, and then $0 = 0x + 0x$ belongs to W. Also for $x \in W$, and $a \in F$, $ax = ax + 0x \in W$. This shows that W is a subspace of V. ♯

Example 1.3.3 Let V be a vector space over a field F. Then V is clearly a subspace of V, and it is called an **improper** subspace of V. The singleton $\{0\}$ is also a subspace of V, and it is called the **trivial** subspace of V. Other subspaces of V are called **Proper** subspaces of V.

Example 1.3.4 (**Subspaces of $\mathbb{R}^2$ over $\mathbb{R}$**) Let W be a nontrivial subspace of $\mathbb{R}^2$. Then there is a nonzero element $(l, m) \in W$. Since W is a subspace, $\alpha \cdot (l, m) = (\alpha l, \alpha m) \in W$ for all $\alpha \in \mathbb{R}$. Thus, $W_{lm} = \{(\alpha l, \alpha m) \mid \alpha \in \mathbb{R}\} \subseteq W$. W_{lm} is easily seen to be a subspace of $\mathbb{R}^2$. Indeed, W_{lm} is the line in the plane $\mathbb{R}^2$ passing through origin and the point (l, m). Note that all lines in $\mathbb{R}^2$ are of this type. Suppose that $W \neq W_{lm}$. Then there is a nonzero element (p, q) in $W - W_{lm}$. We claim that $ql - pm \neq 0$. Suppose that $ql - pm = 0$. Since $(l, m) \neq (0, 0)$, $l \neq 0$ *or* $m \neq 0$. Suppose that $l \neq 0$. Then, $(p, q) = (\frac{p}{l} l, \frac{p}{l} m)$ turns out to be in W_{lm}, a contradiction to the choice of (p, q). Similarly, if $m \neq 0$, then $(p, q) = (\frac{q}{m} l, \frac{q}{m} m)$, a contradiction. Now, let (a, b) be an arbitrary member of $\mathbb{R}^2$. Since $ql - pm \neq 0$, we can solve the

pair of equations $\alpha l + \beta p = a$ and $\alpha m + \beta q = b$. In other words, $(a, b) = \alpha(l, m) + \beta(p, q)$ belongs to W, and so $W = \mathbb{R}^2$. This shows that only proper subspaces of $\mathbb{R}^2$ are the lines passing through origin.

Example 1.3.5 (**Subspaces of $\mathbb{R}^3$ over $\mathbb{R}$**) As in the above example, lines and planes passing through origin are proper subspaces of $\mathbb{R}^3$ *over* $\mathbb{R}$. Indeed, they are the only proper subspaces.

Proposition 1.3.6 *Intersection of a family of subspaces is a subspace.*

Proof Let $\{W_\alpha \mid \alpha \in \Lambda\}$ be a family of subspaces of a vector space V over F. Then $0 \in W_\alpha$ *for all* α, and so 0 belongs to the intersection of the family. Thus, the intersection of the given family is nonempty. Let $x, y \in \bigcap_{\alpha\in\Lambda} W_\alpha$, and $a, b \in F$. Then $x, y \in W_\alpha$ *for all* α. Since each W_α is a subspace, $ax + by \in W_\alpha$ *for all* α. Hence $ax + by$ belongs to the intersection. This shows that the intersection of the family is a subspace. ♯

Proposition 1.3.7 *Union of subspaces need not be a subspace. Indeed, the union $W_1 \bigcup W_2$ of two subspaces is a subspace if and only if $W_1 \subseteq W_2$ or $W_2 \subseteq W_1$.*

Proof If $W_1 \subseteq W_2$, then $W_1 \bigcup W_2 = W_2$ a subspace. Similarly, if $W_2 \subseteq W_1$, then also the union is a subspace. Conversely, suppose that $W_1 \bigcup W_2$ is a subspace and W_1 is not a subset of W_2. Then there is an element $x \in W_1$ which is not in W_2. Let $y \in W_2$. Then, since $W_1 \bigcup W_2$ is a subspace, $x + y \in W_1 \bigcup W_2$. Now $x + y$ does not belong to W_2, for otherwise $x = (x + y) - y$ will be in W_2, a contradiction to the supposition. Hence $x + y \in W_1$. Since $x \in W_1$ and W_1 is subspace, $y = -x + (x + y)$ belongs to W_1. This shows that $W_2 \subseteq W_1$. ♯

Proposition 1.3.8 *Let W_1 and W_2 be subspaces of a vector space V over a field F. Then $W_1 + W_2 = \{x + y \mid x \in W_1, y \in W_2\}$ is also a subspace (called the* ***sum*** *of W_1 and W_2) which is the smallest subspace containing $W_1 \bigcup W_2$.*

Proof Since $0 \in W_2$, $x \in W_1$ implies that $x = x + 0 \in W_1 + W_2$. Thus, $W_1 \subseteq W_1 + W_2$. Similarly, $W_2 \subseteq W_1 + W_2$. Also, if L is a subspace containing $W_1 \bigcup W_2$, then $x + y \in L$ *for all* $x \in W_1$, and $y \in W_2$. Therefore, it is sufficient to show that $W_1 + W_2$ is a subspace. Clearly, $W_1 + W_2 \neq \emptyset$. Let $x + y$ and $u + v$ belong to $W_1 + W_2$, where $x, u \in W_1$, and $y, v \in W_2$. Since W_1 and W_2 are subspaces, $\alpha x + \beta u \in W_1$, and $\alpha y + \beta v \in W_2$. But, then $\alpha(x + y) + \beta(u + v) = (\alpha x + \beta u) + (\alpha y + \beta v)$ belongs to $W_1 + W_2$. ♯

Definition 1.3.9 A family $\{W_\alpha \mid \alpha \in \Lambda\}$ of subspaces of a vector space V over a field F is called a **chain** if for any given pair $\alpha, \beta \in \Lambda$, $W_\alpha \subseteq W_\beta$, *or* $W_\beta \subseteq W_\alpha$.

Proposition 1.3.10 *Union of a chain of subspaces is a subspace.*

Proof Let $\{W_\alpha \mid \alpha \in \Lambda\}$ be a chain of subspaces of a vector space V over a field F. Clearly, $0 \in \bigcup_{\alpha\in\Lambda} W_\alpha$. Let $x, y \in \bigcup_{\alpha\in\Lambda} W_\alpha$, and $\alpha, \beta \in F$. Then $x \in W_\alpha$, and $y \in W_\beta$ for some $\alpha, \beta \in F$. Since the family is a chain, $W_\alpha \subseteq W_\beta$, *or* $W_\beta \subseteq W_\alpha$.

This means that $x, y \in W_\alpha$, *or* $x, y \in W_\beta$. Since W_α and W_β are subspaces, $\alpha x + \beta y$ belongs to W_α or to W_β. It follows that $\alpha x + \beta y \in \bigcup_{\alpha\in\Lambda} W_\alpha$. This shows that $\bigcup_{\alpha\in\Lambda} W_\alpha$ is a subspace. ♯

Subspace Generated (Spanned) by a Subset

Definition 1.3.11 A subset S of a vector space V over a field F need not be a subspace, for example, it may not contain 0. The intersection of all subspaces of V containing S is the smallest subspace of V containing S. This subspace is called the **subspace generated (spanned) by S**, and it is denoted by $< S >$. If $< S > = V$, then we say that **S generates V**, or S is a **set of generators** of V. A vector space V is said to be **finitely generated** if it has a finite set of generators.

Clearly, $< \emptyset > = \{0\}$.

Remark 1.3.12 The subspace $< S >$ of V generated by S is completely characterized by the following 3 properties:

(i) $< S >$ is a subspace.
(ii) $< S >$ contains S.
(iii) If W is a subspace containing S, then $< S > \subseteq W$.

Definition 1.3.13 Let S be a nonempty subset of a vector space V over a field F. An element $x \in V$ is called a **linear combination** of members of S if

$$x = a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

for some $a_1, a_2, \ldots, a_n \in F$ and $x_1, x_2, \ldots, x_n \in V$. We also say that x depends linearly on S.

Remark 1.3.14 If S is a nonempty set, then 0 is always a linear combination of the members of S, for $0 = 0x$. All the members of S are linear combination of members of S, for any $x \in S$ is $1x$. Further, if x is a linear combination of members of S, and $S \subseteq T$, then x is also a linear combination of members of T. A Linear combination of linear combinations of members of S is again a linear combination of members of S.

Proposition 1.3.15 *Let S be a nonempty subset of a vector space V over a field F. Then $< S >$ is the set of all linear combinations of members of S.*

Proof Let W denote the set of all linear combinations of members of S. Since members of S are also linear combinations of members of S, it follows that $S \subseteq W$. Thus, W is nonempty set. Let $x = a_1x_1 + a_2x_2 + \cdots a_nx_n$ and $y = b_1y_1 + b_2y_2 + \cdots b_my_m$ be members of W, and $a, b \in F$. Then

$$ax + by = a_1x_1 + a_2x_2 + \cdots a_nx_n + b_1y_1 + b_2y_2 + \cdots b_my_m,$$

being a linear combination of members of S, is again a member of W, and so W is a subspace of V. Let L be a subspace of V containing S. It follows, by induction on r,

that any linear combination $a_1x_1 + a_2x_2 + \cdots + a_rx_r$ belongs to L. Thus, W is the smallest subspace of V containing S. ♯

In particular, S is a set of generators of a vector space V over a field F if and only if every element of V is a linear combination of members of S.

Example 1.3.16 The set $E = \{\overline{e_1}, \overline{e_2}, \cdots, \overline{e_n}\}$, where

$$\overline{e_i} = (0, 0, \ldots, 0, \overbrace{1}^{i}, 0, \ldots, 0),$$

is a set of generators of the vector space F^n. Indeed, any member $\overline{x} = (x_1, x_2, \ldots, x_n)$ of F^n is the linear combination $\overline{x} = x_1e_1 + x_2e_2 + \cdots + x_ne_n$ of members of E. The subset $S = \{\overline{e_1} + \overline{e_2}, \overline{e_2} + \overline{e_3}, \overline{e_3} + \overline{e_1}\}$ is also a set of generators of F^3, for $\overline{x} = (x_1, x_2, x_3) = \alpha_1(\overline{e_1} + \overline{e_2}) + \alpha_2(\overline{e_2} + \overline{e_3}) + \alpha_3(\overline{e_3} + \overline{e_1})$, where $\alpha_1 = \frac{x_1+x_2-x_3}{2}$, $\alpha_2 = \frac{x_2+x_3-x_1}{2}$, $\alpha_3 = \frac{x_1+x_3-x_2}{2}$ (verify).

Example 1.3.17 Consider The subset $S = \{\overline{e_1} - \overline{e_2}, \overline{e_2} - \overline{e_3}, \overline{e_3} - \overline{e_1}\}$ of $\mathbb{R}^3$. It is easy to verify that $\overline{x} = (x_1, x_2, x_3)$ is a linear combination of $S = \{\overline{e_1} - \overline{e_2}, \overline{e_2} - \overline{e_3}, \overline{e_3} - \overline{e_1}\}$ if and only if $x_1 + x_2 + x_3 = 0$. Thus, the subspace $< S >$ of $\mathbb{R}^3$ generated by S is the plane $\{\overline{x} = (x_1, x_2, x_3) \mid x_1 + x_2 + x_3 = 0\}$.

Linear Independence

Definition 1.3.18 A subset S of a vector space V over a field F is called **linearly independent** if given any finite subset $\{x_1, x_2, \ldots x_n\}$ of S, $x_i \neq x_j$ for $i \neq j$,

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0 \textit{ implies that } a_i = 0 \textit{ for all } i.$$

A subset S which is not linearly independent is called a **linearly dependent** subset.

Thus, a subset S of a vector space V over a field F is linearly dependent if there is a subset $\{x_1, x_2, \ldots, x_n\}$ of distinct members of S, and $a_1, a_2, \ldots, a_n$ not all zero in F such that

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0.$$

Vacuously, the empty set $\emptyset$ is linearly independent. The observations in the following proposition are easy but crucial, and they will be used often.

Proposition 1.3.19 *Let V be a vector space over a field F. Then,*

(i) any subset of V containing 0 is linearly dependent,
(ii) every subset of a linearly independent subset of V is linearly independent,
(iii) every subset containing a linearly dependent set is linearly dependent,
(iv) if S is a subset of V, and $x \in < S > - S$, then $S \bigcup \{x\}$ is linearly dependent, and
(v) if S is linearly independent, and $x \notin < S >$, then $S \bigcup \{x\}$ is linearly independent.

Proof (i) If $0 \in S$, then $1 \cdot 0 = 0$ but $1 \neq 0$. It follows from the definition that S is linearly dependent.
The assertions (i) and (iii) are immediate from the definition itself.
(iv) Suppose that $x \notin S$, and $x \in < S >$. Then there are distinct members $x_1, x_2, \ldots, x_n \in S$, and $a_1, a_2, \ldots, a_n \in F$ such that

$$x = a_1x_1 + a_2x_2 + \cdots + a_nx_n.$$

But, then

$$-1x + a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0.$$

Since $1 \neq 0$, it follows that $S \bigcup \{x\}$ is linearly dependent.
(v) Suppose that S is linearly independent, and $x \notin < S >$. Suppose that $x_1, x_2, \ldots, x_n \in S$ are distinct members of $S \bigcup \{x\}$ such that

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0.$$

If $x_i \neq x$ *for all* i, then since S is linearly independent, $a_i = 0$ *for all* i. Suppose that $x_i = x$ for some i. Without any loss, we may suppose that $x_1 = x$. Then $a_1 = 0$, otherwise,

$$x = (-a_1)^{-1}a_2x_2 + (-a_1)^{-1}a_3x_3 + \cdots + (-a_1)^{-1}a_nx_n.$$

belongs to $< S >$, a contradiction to the supposition. Thus, $a_1 = 0$. Hence

$$a_2x_2 + a_3x_3 + \cdots + a_nx_n = 0.$$

Since S is linearly independent, $a_i = 0$ for all i. ♯

Proposition 1.3.20 *A subset S of a vector space V over a field F is linearly independent if and only if given distinct members $x_1, x_2, \ldots, x_n \in S$, and $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n \in F$,*

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b_1x_1 + b_2x_2 + \cdots + b_nx_n.$$

implies that $a_i = b_i$ for all i.

Proof Suppose that S is linearly independent, and

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b_1x_1 + a_2x_2 + \cdots + b_nx_n,$$

where $x_1, x_2, \ldots, x_n$ are distinct members of S. Then

$$(a_1 - b_1)x_1 + (a_2 - b_2)x_2 + \cdots + (a_n - b_n)x_n = 0.$$

Since S is linearly independent, $a_i - b_i = 0$ *for all* i, and so $a_i = b_i$ *for all* i. Conversely, suppose that the condition is satisfied, and

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0,$$

where $x_1, x_2, \ldots, x_n$ are distinct members of S. Then,

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0x_1 + 0x_2 + \cdots + 0x_n.$$

From the given condition $a_i = 0$ *for all* i. This shows that S is linearly independent. ♯

Example 1.3.21 The set $E = \{\overline{e_1}, \overline{e_2}, \cdots, \overline{e_n}\}$ described in Example 1.3.16 is linearly independent subset of F^n, for $(x_1, x_2, \ldots, x_n) = x_1e_1 + x_2e_2 + \cdots + x_ne_n = \overline{0} = (0, 0, \ldots, 0)$ implies that each $x_i = 0$. Also, the subset $S = \{\overline{e_1} + \overline{e_2}, \overline{e_2} + \overline{e_3}, \overline{e_3} + \overline{e_1}\}$ of F^3 is linearly independent, for $a_1(\overline{e_1} + \overline{e_2}) + a_2(\overline{e_2} + \overline{e_3}) + a_3(\overline{e_3} + \overline{e_1}) = \overline{0} = (0, 0, 0)$ implies that $a_1 + a_3 = 0 = a_1 + a_2 = a_2 + a_3$. But, then $a_1 = a_2 = a_3 = 0$. However, the subset $S = \{\overline{e_1} - \overline{e_2}, \overline{e_2} - \overline{e_3}, \overline{e_3} - \overline{e_1}\}$ of F^3 is linearly dependent, for $1(\overline{e_1} - \overline{e_2}) + 1(\overline{e_2} - \overline{e_3}) + 1(\overline{e_3} - \overline{e_1} = \overline{0}$.

1.4 Basis and Dimension

Definition 1.4.1 A subset S of a vector space V over a field F is said to be a **minimal set of generators** or **irreducible set of generators** if
(i) S generates V, i.e., $< S > = V$, and
(ii) no proper subset of S generates V.
More precisely, $< S > = V$, and $< S - \{x\} > \neq V$ for all $x \in S$.

Definition 1.4.2 A subset B of a vector space V over a field F is said to be a **maximal linearly independent set** if
(i) B is linearly independent, and
(ii) $B \subset S$ implies that S is linearly dependent.
More precisely, a linearly independent subset B is maximal linearly independent if for all $x \notin B$, $B \bigcup \{x\}$ is linearly dependent.

The following two propositions says that maximal linearly independent sets and minimal sets of generators are same.

Proposition 1.4.3 *Every minimal set of generators is also a maximal linearly independent set.*

Proof Let S be a minimal set of generators of a vector space V over a field F. Suppose that S is not linearly independent. Then there exists a set $\{x_1, x_2, \ldots, x_n\}$ of distinct members of S, and $a_1, a_2, \ldots, a_n$ not all 0 in F such that

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0.$$

Since the addition is commutative, without any loss, we may assume that $a_1 \neq 0$. But, then

$$x_1 = (-a_1)^{-1}a_2x_2 + (-a_1)^{-1}a_3x_3 + \cdots + (-a_1)^{-1}a_nx_n.$$

This shows that x_1 is a linear combination of members of $S - \{x_1\}$, or equivalently, $x_1 \in < S - \{x_1\} >$. Thus, $S \subseteq < S - \{x_1\} >$. Since $< S >$ is the smallest subspace containing S, $V = < S > \subseteq < S - \{x_1\} >$. It follows that $< S - \{x_1\} > = V$. This is a contradiction to the supposition that S is a minimal set of generators of V. Thus, S is linearly independent. Next, suppose that $x \notin S$. Since S is also a set of generators, it follows from the Proposition 1.3.19(iv) that $S \bigcup \{x\}$ is linearly dependent. This completes the proof of the fact that S is maximal linearly independent. ♯

Conversely, have the following proposition:

Proposition 1.4.4 *A maximal linearly independent subset is also a minimal set of generators.*

Proof Let B be a maximal linearly independent subset of a vector space V over a field F. Let $x \in V$. If $x \in B$, then $x \in < B >$. Suppose that $x \notin B$. Since B is maximal linearly independent subset of V, $B \bigcup \{x\}$ is linearly dependent. Hence there exists a set $\{x_1, x_2, \ldots, x_n\}$ of distinct members of $B \bigcup \{x\}$, and $a_1, a_2, \ldots, a_n$ not all 0 in F such that

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0.$$

One of the x_i is x and corresponding $a_i \neq 0$, otherwise B will turn out to be linearly dependent, a contradiction to the supposition that B is linearly independent. We may assume, without loss of generality, that $x_1 = x$ and $a_1 \neq 0$. But, then

$$x = x_1 = (-a_1)^{-1}a_2x_2 + (-a_1)^{-1}a_3x_3 + \cdots + (-a_1)^{-1}a_nx_n.$$

Hence $x \in < B >$. This shows that B is a set of generators of V. Finally, $x \notin < B - \{x\} >$, otherwise, from Proposition 1.3.19(iv), B will turn out to be linearly dependent. This shows that B is a minimal set of generators. ♯

Most of the implications in the following theorem are already established.

Theorem 1.4.5 *Let B be a subset of a vector space V over a field F. Then the following conditions are equivalent:*
1. B is maximal linearly independent subset of V.
2. B is a minimal set of generators of V.
3. B is linearly independent as well as a set of generators of V.
4. Every nonzero element $x \in V$ can be expressed uniquely (upto order) as

$$x = a_1x_1 + a_2x_2 + \cdots + a_nx_n,$$

where $x_1, x_2, \ldots, x_n$ are distinct members of B, and $a_1, a_2, \ldots, a_n$ are all nonzero members of F.

Proof The equivalence of 1 and 2 follows from the Proposition 1.4.3 and the Proposition 1.4.4. The implication 2 $\Rightarrow$ 3 follows from the Proposition 1.4.3.
(3 $\Rightarrow$ 4). Assume 3. Since B is a set of generators and also linearly independent, 4 follows from the Proposition 1.3.20.
(4 $\Rightarrow$ 1). Assume 4. It follows again from the Proposition 1.3.20 that B is linearly independent. Suppose that $x \notin B$. By (4), x is a linear combination of members of B, and so $B \bigcup \{x\}$ is linearly dependent. This shows that B is maximal linearly independent subset. ♯

Definition 1.4.6 A subset B of a vector space V over a field F is called a **basis** of V if it satisfies any one, and hence all, of the conditions in the Theorem 1.4.5.

Example 1.4.7 The set $E = \{\overline{e_1}, \overline{e_2}, \cdots, \overline{e_n}\}$ described in Example 1.3.16 is linearly independent (Example 1.3.21) subset as well as a set of generators of $\mathbf{F^n}$ (Example 1.3.16), and hence it is a basis of F^n. This basis is called the **standard basis** of F^n. Similarly, $S = \{\overline{e_1} + \overline{e_2}, \overline{e_2} + \overline{e_3}, \overline{e_3} + \overline{e_1}\}$ is another basis of F^3.

Proposition 1.4.8 *Let V be a finitely generated vector space over a field F. Then V has a finite basis. Indeed, any finite set of generators contains a basis.*

Proof Let S be a finite set of generators of V. It may be a minimal set of generators and so a basis. If not, $< S - \{x_1\} > = V$ for some $x_1 \in S$. $S - \{x_1\}$ may be a minimal set of generators and so a basis. If not, then $< S - \{x_1, x_2\} > = V$ for some $x_2 \in S - \{x_1\}$. $S - \{x_1, x_2\}$ may be a minimal set of generators and so a basis. If not, proceed. This process stops after finitely many steps giving us a basis contained in S, for S is finite. ♯

Theorem 1.4.9 *Let V be a finitely generated vector space over a field F. Then every basis of V is finite, and any two bases of V contain the same number of elements.*

Proof From the above proposition, V has a finite basis

$$B_1 = \{x_1, x_2, \cdot, x_n\}(\text{say}).$$

Let B_2 be another basis of V. If $B_1 - B_2 = \emptyset$, then $B_1 \subseteq B_2$. Since B_1 and B_2 are both maximal linearly independent sets (being bases), $B_1 = B_2$, and we are done. Suppose that $B_1 \neq B_2$. Then $B_2 - B_1 \neq \emptyset$, otherwise $B_2 \subseteq B_1$, and again $B_2 = B_1$. Let $y_1 \in B_2 - B_1$. Since B_1, being a basis, is maximal linearly independent, $B_1 \bigcup \{y_1\}$ is linearly dependent. Thus, there exist $a_1, a_2, \ldots, a_n, b_1$ not all 0 in the field F such that

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n + b_1y_1 = 0.$$

Indeed, $b_1 \neq 0$, otherwise

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n \ = \ 0,$$

and then all $a_i \ = \ 0$. Further, since $y_1 \neq 0, \ b_1y_1 \neq 0$. Hence $a_i \neq 0$ for some i. We may assume that $a_1 \neq 0$. But, then

$$x_1 \ = \ (-a_1)^{-1}a_2x_2 \ + \ (-a_1)^{-1}a_3x_3 + \cdots \ + \ (-a_1)^{-1}a_nx_n \ + \ (-a_1)^{-1}b_1y_1.$$

Hence, $x_1 \in < (B_1 - \{x_1\}) \bigcup \{y_1\} >$. This shows that $B_1 \subseteq < (B_1 - \{x_1\}) \bigcup \{y_1\} >$, and so $(B_1 - \{x_1\}) \bigcup \{y_1\}$ generates V. We also show that $(B_1 - \{x_1\}) \bigcup \{y_1\}$ is linearly independent. Suppose that

$$a_2x_2 + a_3x_3 + \cdots + a_nx_n + b_1y_1 \ = \ 0.$$

If $b_1 \ = \ 0$, then

$$a_2x_2 + a_3x_3 + \cdots + a_nx_n + \ = \ 0.$$

Since B_1 is linearly independent, $a_i \ = \ 0$ for all $i \geq 2$. Suppose that $b_1 \neq 0$. Then

$$y_1 \ = \ (-b_1)^{-1}a_2x_2 \ + \ (-b_1)^{-1}a_3x_3 + \cdots \ + \ (-b_1)^{-1}a_nx_n,$$

and so $y_1 \in < B_1 - \{x_1\} >$. Since $(B_1 - \{x_1\}) \bigcup \{y_1\}$ is already seen to be a set of generators of V, $< B_1 - \{x_1\} > \ = \ V$. This is a contradiction to the supposition that B_1 is a basis (minimal set of generators). This shows that $(B_1 - \{x_1\}) \bigcup \{y_1\}$ is also a basis containing n elements. If $(B_1 - \{x_1\}) \bigcup \{y_1\} \ - \ B_2 \ = \ \emptyset$, then as before $(B_1 - \{x_1\}) \bigcup \{y_1\} \ = \ B_2$, and so B_2 contains n elements. If not, as before, $B_2 \ - \ ((B_1 - \{x_1\}) \bigcup \{y_1\})$ is nonempty, and then proceed as above. The process stops after finitely many steps, at most at the *nth* step, showing that B_2 is finite, and contains exactly n elements. ♯

Definition 1.4.10 The number of elements in a basis of a finitely generated vector space V over a field F is called the **dimension** of V, and it is denoted by $dim(V)$.

It follows from Example 1.4.7 that the dimension of F^n is n. The dimension of the plane $W \ = \ \{\overline{x} \ = \ (x_1, x_2, x_3) \ | \ x_1 + x_2 + x_3 \ = \ 0\}$ is 2, for $\{\overline{e_1} \ - \ \overline{e_2}, \ \overline{e_2} \ - \ \overline{e_3}\}$ is a basis of W (verify). To determine the dimension of a vector space, one needs to determine a basis of the vector space, and then count the number of elements in the basis. In the next chapter we shall have an algorithm to find a basis, and so also the dimensions of the subspaces of F^n, which are generated by finite sets of elements of F^n.

Proposition 1.4.11 *Every set of generators of a finite dimensional vector space contains a basis.*

Proof Let $B = \{x_1, x_2, \ldots, x_n\}$ be a finite basis of a finite dimensional vector space V over a field F. Let S be a set of generators of V. Then each member x_i is a linear combination of a finite subset A_i (say) of S. In turn, each member of B is a linear combination of the finite subset $A = A_1 \bigcup A_2 \bigcup, \ldots, \bigcup A_n$ of S. Since B generates V, A also generates V. Since A is finite, we can reduce A to a minimal set of generators, a basis of V. ♯

Proposition 1.4.12 *Every linearly independent subset of a finite dimensional vector space V over a field F can be enlarged to a basis of V.*

Proof Let $B = \{x_1, x_2, \ldots, x_n\}$ be a finite basis of a finite dimensional vector space V over a field F, and S a linearly independent subset of V. If $B \subseteq < S >$, then $< S > = V$, and so S is a basis, and there is nothing to do. If not, then some $x_i \in B - < S >$. We may assume that $x_1 \in B - < S >$. Then by the Proposition 1.3.19(v), $S \bigcup \{x_1\}$ is linearly independent. If $B \subseteq < S \bigcup \{x_1\} >$, then $V = < B > \subseteq < S \bigcup \{x_1\} >$, and so $S \bigcup \{x_1\}$ turns out to be a basis. If not, proceed. This process stops at most at the *nth* step enlarging S to a basis. ♯

Corollary 1.4.13 *If $dim V = n$, then*

(i) every linearly independent subset contains at most n elements, and
(ii) any set of generators contain at least n elements. ♯

Proposition 1.4.14 *Let F be a finite field containing q elements, and V a vector space over F of dimension n. Then V contains exactly q^n elements.*

Proof Since $dim V = n$, there is a basis $\{x_1, x_2, \ldots, x_n\}$ of V containing n elements. Hence every element v of V can be expressed uniquely as

$$v = \alpha_1 x_1 + \alpha_2 x_2 \cdots + \alpha_n x_n,$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n$ belong to F. This says that we have a bijective map η from F^n to V defined by

$$\eta(\alpha_1, \alpha_2, \ldots, \alpha_n) = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n.$$

Since F contains q elements, F^n and hence V contains q^n elements. ♯

Corollary 1.4.15 *Let F be a finite field of characteristic p (note that p is prime). Then F contains p^n elements for some $n \in \mathbb{N}$.*

Proof Since F is finite, its characteristic is some prime $p \neq 0$. By proposition1.1.1, F has a subfield isomorphic to the field $\mathbb{Z}_p$ of prime residue classes modulo p. Thus, F is a vector space over a field containing p elements. Since it is finite, its dimension is finite n (say). From the previous proposition, the result follows. ♯

Corollary 1.4.16 *Let L be a field containing p^n elements, where p is a prime, and $n \geq 1$. Let F be a subfield of L. Then F contains p^m elements for some divisor m of n.*

Proof Since F is a subfield of L, $charL = charF = p$. Thus, F contains p^m elements for some m. Since L is a vector space over F, it follows that L contains $(p^m)^r = p^{mr}$ elements for some r. Hence $n = mr$. ♯

Remark 1.4.17 We shall see in a latter chapter that for every prime p, and for all $n \geq 1$, there is a unique (upto isomorphism) field of order p^n. Further, corresponding to any divisor m of n, there is a unique subfield of order p^m.

Definition 1.4.18 Let V be a vector space over a field F. An ordered n-tuple $(x_1, x_2, \ldots, x_n)$ is called an **ordered basis** of V if the set $\{x_1, x_2, \ldots, x_n\}$ is a basis of V. Thus, to every basis there are exactly $n!$ distinct ordered bases which give rise to the same basis.

Proposition 1.4.19 *Let V be a vector space of dimension n over a finite field F containing q elements. Then the number of ordered bases of V is*

$$(q^n - 1)(q^n - q)(q^n - q^2) \cdots (q^n - q^{n-1}),$$

and the number of bases of V is

$$\frac{(q^n - 1)(q^n - q) \cdots (q^n - q^{n-1})}{n!}.$$

Proof We find the number of ordered n-tuples $(x_1, x_2, \ldots, x_n)$ such that the set $\{x_1, x_2, \ldots, x_n\}$ is a basis. Since x_1 can be any nonzero element of the vector space, the number of ways in which x_1 can be selected is $q^n - 1$. Having chosen x_1, we have to select x_2 such that $\{x_1, x_2\}$ is linearly independent. Clearly, $\{x_1, x_2\}$ is linearly independent if and only if $x_2 \neq \alpha x_1$ for all $\alpha \in F$. Thus, the number of ways in which the ordered pair (x_1, x_2) can be chosen so that the set $\{x_1, x_2\}$ is linearly independent is $(q^n - 1)(q^n - q)$. Again, having chosen (x_1, x_2), we have to find x_3 so that $\{x_1, x_2, x_3\}$ is linearly independent. Now, $\{x_1, x_2, x_3\}$ is linearly independent if and only if $x_3 \neq \alpha_1 x_1 + \alpha_2 x_2$ for every pair $\alpha_1, \alpha_2 \in F$. Thus, the number of choices for x_3 is $q^n - q^2$. Hence the number of choices for the ordered triple (x_1, x_2, x_3) so that the set $\{x_1, x_2, x_3\}$ is linearly independent is $(q^n - 1)(q^n - q)(q^n - q^2)$. Proceeding inductively, the number of choices for the ordered n-tuple $(x_1, x_2, \ldots x_n)$ so that that $\{x_1, x_2, \ldots x_n\}$ is linearly independent (and so a basis) is

$$(q^n - 1)(q^n - q)(q^n - q^2) \cdots (q^n - q^{n-1}).$$

In turn, the number of bases of V is

$$\frac{(q^n - 1)(q^n - q) \cdots (q^n - q^{n-1})}{n!}.$$ ♯

Remark 1.4.20 If W is a subspace of a vector space V, then since a basis of W is a linearly independent subset of V, $dim W \leq dim V$. Further, if n is the dimension of V, and $m \leq n$, then there is a subspace of dimension m, for if $\{x_1, x_2, \ldots, x_n\}$ is a basis of V, then the subspace $< \{x_1, x_2, \ldots, x_m\} >$ is a subspace of dimension m.

Proposition 1.4.21 *Let V be a vector space of dimension n over a field F containing q elements. Then the number b_r of r dimensional subspaces, $r \geq 1$ is*

$$b_r = \frac{(q^n - 1)(q^n - q) \cdots (q^n - q^{r-1})}{(q^r - 1)(q^r - q) \cdots (q^r - q^{r-1})}.$$

The total number of subspaces is

$$1 + b_1 + b_2 + \cdots + b_n,$$

where b_r is given above.

Proof Let $0 < r < n$. Any subspace of dimension r is determined by a linearly independent subset $\{x_1, x_2, \ldots, x_r\}$ of V. From the proof of the above proposition, it follows that the number of linearly independent subsets containing r elements is

$$\frac{(q^n - 1)(q^n - q) \cdots (q^n - q^{r-1})}{r!}.$$

Further, a linearly independent subset $\{y_1, y_2, \ldots, y_r\}$ determine the same subspace as the set $\{x_1, x_2, \ldots, x_r\}$ if and only if $\{y_1, y_2, \ldots, y_r\}$ is a basis of the the subspace $< \{x_1, x_2, \ldots, x_r\} >$ which is of dimension r. The number of bases of a vector space of dimension r is

$$\frac{(q^r - 1)(q^r - q) \cdots (q^r - q^{r-1})}{r!}.$$

Thus, the number of r dimensional subspaces of V is b_r given in the statement of the proposition. ♯

Proposition 1.4.22 *Let V be a vector space of finite dimension over a field F. Let W_1 and W_2 be subspaces of V. Then $W_1 + W_2 = \{x + y \mid x \in W_1, y \in W_2\}$ is a subspace, and*

$$dim(W_1 + W_2) = dim W_1 + dim W_2 - dim(W_1 \bigcap W_2).$$

Proof We have already seen that $W_1 + W_2$ is a subspace. Let $\{x_1, x_2, \ldots, x_r\}$ be a basis of $W_1 \bigcap W_2$. Then it is a linearly independent subset of W_1 as well as of W_2. Thus, it can be enlarged to a basis

$$\{x_1, x_2, \ldots, x_r, y_1, y_2, \ldots, y_m\}$$

of W_1, and to a basis

$$\{x_1, x_2, \ldots, x_r, z_1, z_2, \ldots z_n\}$$

of W_2, where $\dim W_1 = r + m$, and $\dim W_2 = r + n$. We show that

$$S = \{x_1, x_2, \ldots, x_r, y_1, y_2, \ldots, y_m, z_1, z_2, \ldots, z_n\}$$

is a basis of $W_1 + W_2$. Clearly, $W_1 \subseteq < S >$, and $W_2 \subseteq < S >$. Hence $W_1 + W_2 \subseteq < S >$. Also $S \subseteq W_1 \bigcup W_2$. Hence, $< S > \subseteq W_1 + W_2$. Now, we show that S is linearly independent. Suppose that

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_r x_r + \beta_1 y_1 + \beta_2 y_2 + \cdots + \beta_m y_m + \delta_1 z_1 + \delta_2 z_2 + \cdots + \delta_n z_n = 0.$$

Then,

$$\alpha_1 x_1 + \cdots \alpha_r x_r + \beta_1 y_1 + \cdots + \beta_m y_m = -\delta_1 z_1 - \cdots - \delta_n z_n$$

belongs to $W_1 \bigcap W_2$. Since $\{x_1, x_2, \ldots, x_r\}$ is a basis of $W_1 \bigcap W_2$,

$$-\delta_1 z_1 - \delta_2 z_2 - \cdots - \delta_n z_n = \gamma_1 x_1 + \gamma_2 x_2 + \cdots \gamma_r x_r$$

for some $\gamma_1, \gamma_2, \ldots, \gamma_r$ in F. Since $\{x_1, x_2, \ldots, x_r, z_1, z_2, \ldots, z_n\}$ is linearly independent (being a basis of W_2), it follows that $\delta_1, \delta_2, \ldots, \delta_n$ are all zero. Similarly, $\beta_1, \beta_2, \ldots \beta_m$ are all zero. But, then $\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_r x_r = 0$. Since $\{x_1, x_2, \ldots, x_r\}$ is linearly independent(being a basis of $W_1 \bigcap W_2$), we see that $\alpha_1, \alpha_2, \ldots, \alpha_r$ are all zero. This shows that S is also linearly independent and so a basis of $W_1 + W_2$. In turn, $dim(W_1 + W_2) = r + m + n = (r + n) + (r + m) - r = dim W_1 + dim W_2 - dim(W_1 \bigcap W_2)$. ♯

Example 1.4.23 Let V be a vector space of dimension n. Let W_1 and W_2 be distinct subspaces of dimension $n - 1$ each. Then $W_1 + W_2$ is a subspace of V containing W_1 properly, and so it is V. Thus, $dim(W_1 + W_2) = n = dim W_1 + dim W_2 - dim(W_1 \bigcap W_2$. Hence, from the above proposition $dim(W_1 \bigcap W_2) = n - 1 + n - 1 - n = n - 2$.

1.5 Direct Sum of Vector Spaces, Quotient of a Vector Space

Let $V_1, V_2, \ldots, V_r$ be vector spaces over a field F. Consider the Cartesian product $V = V_1 \times V_2 \times \cdots \times V_r$. Define addition using the coordinate-wise addition, and the external multiplication $\cdot$ by $\alpha \cdot (x_1, x_2, \ldots, x_r) = (\alpha x_1, \alpha x_2, \ldots, \alpha x_r)$. It is straight forward to verify that V is a vector space over F with respect to these operations. This vector space is called the **external direct sum** of $V_1, V_2, \ldots, V_r$.

A vector space V over a field F is said to be **internal direct sum** of its subspaces $V_1, V_2, \ldots, V_r$ if every element x of V has a unique representation as

$$x = x_1 + x_2 + \cdots + x_r,$$

where $x_i \in V_i$ *for all* i. The notation

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_r$$

stands to assert that V is direct sum of its subspaces $V_1, V_2, \ldots, V_r$.

Proposition 1.5.1 *Let $V_1, V_2, \ldots, V_r$ be subspaces of a vector space V over a field F. Then the following conditions are equivalent.*

(1) $V = V_1 \oplus V_2 \oplus \cdots \oplus V_r$.
(2) $V = V_1 + V_2 + \cdots + V_r$, *and* $V_i \bigcap V^i = \{0\}$ *for all* i, *where* $V^i = (V_1 + V_2 + \cdots + V_{i-1} + V_{i+1} + V_{i+2} + \cdots + V_r)$.

Proof $1 \Longrightarrow 2$. Assume 1. Since every element x of V has a unique representation as $x = x_1 + x_2 + \cdots + x_r$, $V = V_1 + V_2 + \cdots + V_r$. Let $x \in V_i \bigcap V^i$. Then $x = x_1 + x_2 + \cdots + x_{i-1} + x_{i+1} + x_{i+2} + \cdots x_r$, where $x_j, j \neq i$ belong to V_j. Thus,

$$0 = x_1 + x_2 + \cdots + x_{i-1} - x + x_{i+1} + x_{i+2} + \cdots + x_r = 0 + 0 + \cdots + 0.$$

From the uniqueness of the representation of 0, it follows that x_j is zero for all j, and x is also 0. Hence $x = 0$.

$2 \Longrightarrow 1$. Assume 2. Clearly, every element x of V has a representation $x = x_1 + x_2 + \cdots + x_r$, where $x_i \in V_i$ for all i. Now, we prove the uniqueness of the representation. If

$$x = x_1 + x_2 + \cdots + x_r = y_1 + y_2 + \cdots + y_r,$$

where $x_i, y_i \in V_i$ for all i. Then $x_i - y_i \in V_i \bigcap V^i = \{0\}$. This shows that $x_i = y_i$ *for all* i. ♯

Remark 1.5.2 If V is direct sum of $V_1, V_2, \ldots, V_r$, then V^i as defined in the above proposition is direct sum of $V_1, V_2, \ldots, V_{i-1}, V_{i+1}, V_{i+2}, \ldots, V_r$.

Proposition 1.5.3 *Let $W_1, W_2, \ldots, W_r$ be subspaces of a vector space V such that $V = W_1 + W_2 + \cdots + W_r$. Then V is direct sum of $W_1, W_2, \ldots, W_r$ if and only if $dim V = dim W_1 + dim W_2 + \cdots + dim W_r$.*

Proof Suppose that V is direct sum of $W_1, W_2, \ldots, W_r$. Then, we show that $dim V = dim W_1 + dim W_2 + \cdots + dim W_r$. The proof is by induction on r. If $r = 1$, then there is nothing to do. Assume that the result is true for r. Since $W_1 \bigcap W^1 = \{0\}$, it follows from the Proposition 1.4.22 that $dim V = dim(W_1 + W^1) = dim W_1 + dim W^1 - dim\{0\} = dim W_1 + dim W_2$. Further, it is clear that W^1 is direct sum of $W_2, W_3, \ldots, W_{r+1}$, and so by the induction assumption $dim W^1 = dim W_2 + dim W_3 + \cdots + dim W_{r+1}$. Hence $dim V = dim W_1 + dim W_2 + \cdots + dim W_r$. Conversely, suppose that $dim V = dim W_1 + dim W_2 + \cdots + dim W_r$. Clearly, then $W_i \bigcap W^i = \{0\}$ for all i. The result follows. ♯

Example 1.5.4 The plane $W = \{(x, y, z) \mid lx + my + nz = 0\}$, $(l, m, n) \neq (0, 0, 0)$, is a subspace of $\mathbb{R}^3$ of dimension 2 (verify). Let $(a, b, c) \notin W$. The line $L = \{(a\alpha, b\alpha, c\alpha) \mid \alpha \in \mathbb{R}\}$ is also a subspace of $\mathbb{R}^3$ of dimension 1 such that $W + L = \mathbb{R}^3$. Since $dim W + dim L = 3, \mathbb{R}^3 = W \oplus L$.

Let V be a vector space, and W a subspace of V. Let $x \in V$. The subset $x + W = \{x + w \mid w \in W\}$ is called the **coset** of V modulo W determined by x. This is also called a Plane in V passing through x and parallel to W. The set $\{x + W \mid x \in V\}$ of cosets of V modulo W is denoted by V/W.

Proposition 1.5.5 *Let W be a subspace of a vector space V over a field F. Then the following hold:*

(i) $x \in x + W$ *for all* $x \in V$.
(ii) $(x + W = y + W)$ *if and only if* $(x - y) \in W$.
(iii) $(x + W \neq y + W)$ *if and only if* $(x + W) \bigcap (y + W) = \emptyset$.

In particular, V/W is a partition of the vector space V.

Proof (i) Since $0 \in W$, $x = x + 0 \in x + W$ for all $x \in V$.
(ii) Suppose that $(x + W = y + W)$. Then $x \in y + W$. Hence $x = y + w$ for some $w \in W$. In turn, $x - y = w$ belongs to W. Conversely, suppose that $x - y \in W$. Then $x + w = y + (x - y) + w$ belongs to $y + W$ for all $w \in W$. This shows that $x + W \subseteq y + W$. Similarly, since $y - x$ also belongs to W, it follows that $y + W \subseteq x + W$. Thus, $x + W = y + W$.
(iii) Suppose that $(x + W) \bigcap (y + W) \neq \emptyset$. Let $z \in (x + W) \bigcap (y + W)$. Then $z = x + u = y + v$ for some $u, v \in W$. But, then $x - y = v - u$ belongs to W. It follows from (ii) that $(x + W = y + W)$. ♯

Corollary 1.5.6 *Let W be a subspace of a vector space V over a field F. Then the following hold:*

(i) If $x + W = x' + W$ *and* $y + W = y' + W$, *then* $(x + y) + W = (x' + y') + W$.
(ii) If $x + W = x' + W$, *then* $ax + W = ax' + W$ *for all* $a \in F$.

In turn, we have an internal binary operation $+$ on V/W, and an external multiplication $\cdot$ on V/W by scalars given by

$$(x + W) + (y + W) = (x + y) + W,$$

and

$$a \cdot (x + W) = a \cdot x + W.$$

Further, V/W is a vector spae with respect to these operations.

Proof (i) Suppose that $x + W = x' + W$, and $y + W = y' + W$. Then $x - x'$ and $y - y'$ belong to W. Since W is a subspace $(x + y) - (x' + y')$ belongs to

W. This shows that $(x + y) + W = (x' + y') + W$.
(ii) Suppose that $x + W = x' + W$. Then, $x - x' \in W$. Since W is a subspace $(ax - ax') = a(x - x')$ belongs to W. This shows that $ax + W = ax' + W$ for all $a \in F$.
(i) and (ii) ensure that we have the addition $+$ on V/W, and the multiplication $\cdot$ by scalars as described in the corollary. The verification of the fact that V/W is a vector space with respect to the operations described is straight forward. The zero of the vector space is the coset $0 + W = W$, and the negative of of $x + W$ is $-x + W$. ♯

Definition 1.5.7 The vector space V/W described above is called the **quotient space** of V modulo W.

Proposition 1.5.8 *Let V be a finite dimensional vector space over a field F, and W a subspace of V. Then*

$$dim V/W = dim V - dim W$$

Proof Let $\{x_1, x_2, \ldots, x_r\}$ be a basis of W, and $\{y_1 + W,\ y_2 + W,\ \ldots,\ y_s + W\}$ be a basis of V/W. We show that $\{x_1, x_2, \ldots, x_r, y_1, y_2, \ldots, y_s\}$ is a basis of V. Let $x \in V$. Then $x + W \in V/W$. Since $\{y_1 + W,\ y_2 + W,\ \ldots,\ y_s + W\}$ is a basis of V/W,

$$x + W = \alpha_1(y_1 + W) + \alpha_2(y_2 + W) + \cdots + \alpha_s(y_s + W) = (\alpha_1 y_1 + \alpha_2 y_2 + \cdots + \alpha_s y_s) + W$$

for some $\alpha_1, \alpha_2, \ldots, \alpha_s$ in F. Hence,

$$x - (\alpha_1 y_1 + \alpha_2 y_2 + \cdots + \alpha_s y_s)$$

belongs to W for some $\alpha_1, \alpha_2, \ldots, \alpha_s$ in F. Since $\{x_1, x_2, \ldots, x_r\}$ is a basis of W,

$$x - (\alpha_1 y_1 + \alpha_2 y_2 + \cdots + \alpha_s y_s) = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_r x_r$$

for some $\beta_1, \beta_2, \ldots, \beta_r$ in F. Thus,

$$x = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_r x_r + \alpha_1 y_1 + \alpha_2 y_2 + \cdots + \alpha_s y_s$$

for some $\beta_i, \alpha_j \in F$. This shows that $\{x_1, x_2, \ldots, x_r, y_1, y_2, \ldots, y_s\}$ generates V. Next, suppose that

$$\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_r x_r + \alpha_1 y_1 + \alpha_2 y_2 + \cdots + \alpha_s y_s = 0.$$

Since $\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_r x_r$ belongs to W,

$$\alpha_1(y_1 + W) + \alpha_2(y_2 + W) + \cdots + \alpha_s(y_s + W) = (\alpha_1 y_1 + \alpha_2 y_2 + \cdots + \alpha_s y_s) + W = W.$$

Since W is the zero of V/W, and $\{y_1 + W, y_2 + W, \cdots, y_s + W\}$ (being a basis of V/W) is linearly independent, $\alpha_j = 0$ *for all* j. But, then $\beta_1 x_1 + \beta_2 x_2 + \cdots +$

$\beta_r x_r = 0$. Since $\{x_1, x_2, \ldots, x_r\}$ (being a basis of W) is linearly independent, $\beta_i = 0$ *for all* i. This shows that $\{x_1, x_2, \ldots, x_r, y_1, y_2, \ldots, y_s\}$ is linearly independent, and so it is a basis also. Thus,

$$dim V = r + s = dim W + dim V/W. \qquad \sharp$$

In this book, we shall be mainly interested in finite dimensional vector spaces. However, a vector space need not be finite dimensional, and to develop the analogues theory for infinitely generated vector spaces, one uses some equivalents of axiom of choice (for example, Zornes Lemma).

Proposition 1.5.9 *Union of a chain of linearly independent subsets is linearly independent.*

Proof Let $\{S_\alpha \mid \alpha \in \Lambda\}$ be a chain of linearly independent subsets. Then $0 \notin S_\alpha$ *for all* α, and hence $0 \notin \bigcup_{\alpha\in\Lambda} S_\alpha$. Let $x_1, x_2, \ldots, x_n$ be distinct elements $\bigcup_{\alpha\in\Lambda} S_\alpha$. Suppose that $x_i \in S_{\alpha_i}$. Since $\{S_\alpha \mid \alpha \in \Lambda\}$ is a chain, there exists α_r such that $S_{\alpha_i} \subseteq S_{\alpha_r}$ *for all* i. Thus $x_1, x_2, \ldots, x_n$ all belong to S_{α_r}. Since S_{α_r} is linearly independent,

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0 \text{ implies that } a_i = 0 \text{ for all } i.$$

This shows that the union is linearly independent. $\sharp$

Proposition 1.5.10 *Every linearly independent subset can be embedded in to a basis.*

Proof All that we need to show that every linearly independent subset can be embedded in a maximal linearly independent subset. Let S be a linearly independent subset of a vector space V over a field F. Let X be the set of all linearly independent subsets which contain S. Then $X \neq \emptyset$, for $S \in X$. Thus, $(X, \subseteq)$ is a nonempty partially ordered set. From the previous proposition, it follows that every chain in X has an upper bound. By the Zorn's Lemma, X has a maximal element T (say). Clearly, T is also maximal linearly independent subset. $\sharp$

Proposition 1.5.11 *Every set S of generators contains a basis.*

Proof Let S be a set of generators of V. If $S = \{0\}$, or $S = \emptyset$, then $V = \{0\}$, and then $\emptyset \subseteq S$ is a basis of V. Suppose that $S \neq \{0\}$. Let X be the set of all linearly independent subsets of S. If x is a nonzero element in S, then $\{x\} \in X$, and so $X \neq \emptyset$. Thus, $(X, \subseteq)$ is a nonempty partially ordered set. Since union of a chain of linearly independent subsets is a linearly independent subset, every chain in X has an upper bound. By the Zorn's lemma, X has a maximal element B (say). $S \subseteq < B >$, for if not, there exists an element $x \in S - < B >$. Consider the subset $B' = B \bigcup \{x\}$. Suppose that

$$\alpha x + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0,$$

where $x_1, x_2, \ldots x_r$ are all in B, and $x_i \neq x_j$ for $i \neq j$. But, then $\alpha x \in < B >$. Since $x \notin < B >$, $\alpha = 0$. Since B is linearly independent, $\alpha_i = 0$ *for all* i. This shows that B' is linearly independent. This is a contradiction to the supposition that B is maximal linearly independent subset of S. Thus, $S \subseteq < B >$, and so $< B > = V$. This shows that B is a basis of V contained in S. ♯

Exercises

1.5.1 Test the following for being a subspace of F^3. Find the subspace generated by those which are not subspaces.

(i) $W = \{\overline{x} = (x_1, x_2, x_3) \mid x_1 + x_2 = x_3\}$.
(ii) $W = \{\overline{x} = (x_1, x_2, x_3) \mid x_1 + 2x_2 + x_3 = 0 = 2x_1 + x_2 + 3x_3\}$.
(iii) $W = \{\overline{x} = (x_1, x_2, x_3) \mid x_1 + 2x_2 + x_3 = 1\}$.
(iv) $W = \{\overline{x} = (x_1, x_2, x_3) \mid x_1 + 2x_2 = x_3^2\}$.
(v) $W = \{\overline{x} = (x_1, x_2, x_3) \mid x_1^2 + x_2^2 + x_3^2 = 1\}$.
(vi) $W = \{(x, sinx, cosx) \mid x \in \mathbb{R}\}$.
(v) $W = \{\overline{x} = (x_1, x_2, x_3) \mid x_1^2 + x_2^2 + x_3^2 = 1 = x_1 + 2x_2 + x_3\}$.

1.5.2 Which of the following subsets are linearly independent? and Why?

(i) The subset $\{(1, 1, 0), (0, 1, 1), (1, 0, 1)\}$ of F^3.
(ii) The subset $\{(1, 1, 1, 0), (2, 3, 4, 0), (4, 9, 16, 0), (2, 3, \pi, 0)\}$ of F^4.
(iii) The subset $\{(1, 1, 1, 0), (1, 3, 4, 0), (1, 9, 16, 0), \}$ of F^4.
(iv) The sphere $S^2 = \{(x, y, z) | x^2 + y^2 + z^2 = 1\}$ in F^3.
(v) The subset $\{(x, x^2) | x \in R\}$ of $\mathbb{R}^2$.

1.5.3 Let V be a finite dimensional vector space, and W a subspace such that $dimW = dimV$. Show that $W = V$. Is this result true for infinite dimensional vector spaces? Support.

1.5.4 Show that a non trivial proper subspace of $\mathbb{R}^3$ is either a line passing through origin or a plane passing through origin.

1.5.5 Show that the intersection of two distinct planes passing through origin is a line passing through origin.

1.5.6 Let W_1 and W_2 be two distinct subspaces of dimension n − 1 of a vector space W of dimension n. Show that the dimension of $W_1 \bigcap W_2$ is n − 2.

1.5.7 A subspace W of dimension $n - 1$ of a vector space V of dimension n is called a hyperplane in V. Show that for every hyperplane W of dimension $n - 1$ of the vector space F^n, *there exists* $\overline{a} = (a_1, a_2, \ldots, a_n) \in F^n - \{\overline{0}\}$ such that

$$W = \{\overline{x} = (x_1, x_2, \ldots, x_n) \mid a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0\}.$$

1.5.8 Show that a subspace (also called a plane) of dimension r of a vector space V of dimension n is an intersection of $n - r$ distinct hyperplanes.

1.5.9* What are the results in the above section which remain true even when a field is replaced by the set $\mathbb{Z}$ of integers with usual addition and multiplication? What are the best possible modifications in the results which are not true for $\mathbb{Z}$ so that it holds for $\mathbb{Z}$ also?

1.5.10 Show that $\{e_1 + e_2, e_2 + e_3, \ldots, e_{n-1} + e_n, e_n\}$, where $\{e_1, e_2, \ldots, e_n\}$ is the standard basis of the vector space F^n, is another basis of F^n.

1.5.11 Characterize a vector space with unique basis. Can we have a vector space with exactly 2 bases? Support.

1.5.12 Find the number of bases of $\mathbb{Z}_p^3$ over $\mathbb{Z}_p$.

1.5.13 Find the number of subspaces of $\mathbb{Z}_p^3$ over $\mathbb{Z}_p$.

1.5.14 Show that a vector space V has no proper subspace if and only if it is of dimension 1.

1.5.15 Embed $(1, 1, 0, 2)$ into a basis of $\mathbb{R}^4$.

1.5.16 Show that $\{(1, 2, 1), (3, 1, 2), (1, 1, 1)\}$ forms a basis of $\mathbb{R}^3$ over R. Express $(3, 5, 2)$ as linear combination of members of the above basis.

1.5.17 Can we have a nontrivial finite vector space over a field of characteristic 0, or over any infinite field? Support.

1.5.18 Let F be a field of order 32. Find the number of subfields of F.

1.5.19 Let $P_n(F)$ denote the set of polynomials of degree at most n over the field F. Show that $P_n(F)$ is a vector space over F with respect to the usual addition of polynomials and multiplication by scalars. Find a basis of this vector space and so also the dimension.

1.5.20 Let W_1 be a subspace of dimension $n - 1$ of a vector space V of dimension n. Let W_2 be a subspace of dimension r such that W_2 is not contained in W_1. Find the dimension of $W_1 \bigcap W_2$.

1.5.21 Let V be a vector space over infinite field. Can we express V as finite union of proper subspaces? Support.

1.5.22 Show that $\{1, (X - 1), (X - 1)^2, \ldots, (X - 1)^n\}$ is a basis of the vector space $P_n(F)$ of polynomials of degree at most n.

1.5.23 Let $W = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \mid x_1 + 2x_2 + 3x_3 + \cdots + nx_n = 0\}$ Show that W is a subspace of $\mathbb{R}^n$. Find the dimension of W.

1.5.24 Find a vector space having exactly 3 bases. Is such a vector space unique? Support.

1.5.25 Let V be a vector space over a field F and W a subspace. Show that there is a subspace W' such that $V = W \oplus W'$.

1.5.26 Let W be a nontrivial proper subspace of the Euclidean vector space $\mathbb{R}^3$ of dimension 3. Show that W is a line passing through origin, or it is a plane passing through origin.

1.5.27 Let W_1 be a line passing through origin, and W_2 a plane passing through origin which does not contain the line W_1. Show that $\mathbb{R}^3 = W_1 \oplus W_2$.

1.5.28 Let $(l, m, n) \neq (0, 0, 0)$. Show that $W = \{(\lambda l, \lambda m, \lambda n) \mid \lambda \in R\}$ is a subspace of $\mathbb{R}^3$, and the quotient space $\mathbb{R}^3/W$ is the set of all lines having the direction ratio (l, m, n).

1.5.29 Show that $\mathbb{R}$ considered as a vector space over the field $\mathbb{Q}$ of rational numbers is infinite dimensional.

1.5.30 Show that in the real vector space $C[0, 1]$ of all real valued continuous functions on $[0, 1]$, the set $\{Sin\ x, Sin\ 2x, \ldots, Sin\ nx\}$ is linearly independent subset for all $n \geq 1$. Show also that the set of all Legendre polynomials also form a linearly independent set. Deduce that $C[0, 1]$ is infinite dimensional.

Chapter 2
Matrices and Linear Equations

Matrices play a pivotal role in mathematics, and in turn, in all branches of science, social science, and engineering. This chapter is devoted to the interplay between matrices and systems of linear equations.

2.1 Matrices and Their Algebra

By definition, a $m \times n$ matrix A with entries in a field F is an arrangement

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

of m rows and n columns of elements of F. In short A is denoted by $[a_{ij}]$, where a_{ij} is the entry in the ith row and jth column of A. The ith row

$$(a_{i1}, a_{i2}, \ldots, a_{in})$$

of the matrix A is a vector in F^n, called the ith row vector of A, and it will be denoted by $R_i(A)$. Thus, the matrix A can also be expressed as a column

R. Lal, *Algebra 2*, Infosys Science Foundation Series in Mathematical Sciences,
DOI 10.1007/978-981-10-4256-0_2

$$\begin{bmatrix} R_1(A) \\ R_2(A) \\ . \\ . \\ . \\ R_m(A) \end{bmatrix}$$

of m rows with entries in F^n.
Similarly, if we treat the members of F^m as column vectors, then the jth column

$$\begin{bmatrix} a_{1j} \\ a_{2j} \\ . \\ . \\ . \\ a_{mj} \end{bmatrix}$$

of the matrix A is a column vector in F^m, called the jth column vector of A, and it will be denoted by $C_j(A)$. As such, the matrix A can also be expressed as a row

$$A = [C_1(A),\ C_2(A),\ \ldots\ ,\ C_m(A)].$$

Thus,

$$A = \begin{bmatrix} 2 & 0 & 1+i & 1-i & 3 \\ 4 & 1+2i & 0 & 1 & i \\ 0 & 8 & 1 & 2 & i \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}$$

is a 4×5 matrix with entries in the field $\mathbb{C}$ of complex numbers.

A matrix A is called a square matrix if the number of rows and columns are same. The matrix

$$A = \begin{bmatrix} 2 & 0 & 1 \\ 4 & 1 & 0 \\ 0 & 8 & 1 \end{bmatrix}$$

is a square 3×3 matrix with entries in the field $\mathbb{R}$ of real numbers.
The set of all $m \times n$ matrices with entries in a field F is denoted by $M_{mn}(F)$. The set of all square $n \times n$ matrices is denoted by $M_n(F)$. We have a binary operation $+$ on $M_{mn}(F)$, called the matrix addition, and which is defined by

$$[a_{ij}] + [b_{ij}] = [c_{ij}],$$

where $c_{ij} = a_{ij} + b_{ij}$.

For example,

$$\begin{bmatrix} 2 & 0 & 1 \\ 4 & 1 & 0 \\ 0 & 8 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 2 \\ 3 & 1 & 0 \\ 5 & 8 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 3 \\ 7 & 2 & 0 \\ 5 & 16 & 2 \end{bmatrix}$$

The $m \times n$ matrix $0_{m\times n}$ all of whose entries are 0 is called the **zero** $m \times n$ matrix. Clearly, the matrix $0_{m\times n}$ is described by the property that for any $m \times n$ matrix A, $A + 0_{m\times n} = A = 0_{m\times n} + A$. Further, if $A = [a_{ij}]$ is a $m \times n$ matrix, then the matrix $-A = [-a_{ij}]$ all of whose entries are the negatives of the corresponding entries of A is called the negative of A, and it is described by the property that $A + (-A) = 0_{m\times n} = (-A) + A$. The proof of the following proposition is an immediate consequence of the corresponding properties of the addition + in F.

Proposition 2.1.1 *The set $M_{mn}(F)$ of $m \times n$ matrices with entries in F is an abelian group with respect to the matrix addition in the sense that it satisfies the following properties:*
(i) The matrix addition + is associative in the sense that
$(A + B) + C = A + (B + C)$
for all A, B and C in $M_{mn}(F)$.
(ii) The matrix addition + is commutative in the sense that
$(A + B) = (B + A)$
for all A, B in $M_{mn}(F)$.
(iii) There is a unique matrix $0_{m\times n}$ in $M_{mn}(F)$ such that $A + 0_{m\times n} = A = 0_{m\times n} + A$ for all A in $M_{mn}(F)$.
(iv) For each matrix A in $M_{mn}(F)$, there is a unique matrix $-A$ in $M_{mn}(F)$ such that $A + (-A) = 0_{mn} = (-A) + A$. ♯

We have an external multiplication $\cdot$ on $M_{mn}(F)$ by scalars in F defined by $a \cdot [a_{ij}] = [b_{ij}]$, where $b_{ij} = a \cdot a_{ij}$. Thus, for example,

$$2 \cdot \begin{bmatrix} 2 & 0 & 1 \\ 4 & 1 & 0 \\ 0 & 8 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 2 \\ 8 & 2 & 0 \\ 0 & 16 & 2 \end{bmatrix}$$

It can be further observed that the triple $(M_{mn}(F), +, \cdot)$ is a vector space over F. Indeed, $(M_{mn}(F), +, \cdot)$ can be identified with the triple $(F^{mn}, +, \cdot)$ under the correspondence $A \longleftrightarrow (R_1(A), R_2(A), \ldots, R_m(A))$ which respects all the operations. Let e_{ij} denote the matrix in which ith row jth column entry is 1 and the rest of the entries are 0. For example, the 3×3 matrix e_{23} is given by

$$e_{23} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

It follows that the set $\{e_{ij} \mid 1 \leq i \leq m,\ 1 \leq j \leq n\}$ corresponds to the standard basis of F^{mn} under the above correspondence. Clearly,

$$[a_{ij}] = \Sigma_{i,j} a_{ij} e_{ij},$$

and

$$\Sigma_{i,j} a_{ij} e_{ij} = 0_{mn}$$

if and only if $a_{ij} = 0$ *for all* i, j. Thus, $\{e_{ij} \mid 1 \leq i \leq m, 1 \leq j \leq n\}$ is a basis, called the standard basis, of the vector space $M_{mn}(F)$. Thus, the dimension of $M_{mn}(F)$ is $m \cdot n$. In particular, $M_n(F)$ is of dimension n^2.

Apart from the above operations, we have an external operation $\cdot$ from $M_{mn}(F) \times M_{np}(F)$ to $M_{mp}(F)$, called the **matrix multiplication**, defined as follows: Let $A = [a_{ij}]$, $1 \leq i \leq m$, $1 \leq j \leq n$, and $B = [b_{jk}]$, $1 \leq j \leq n$, $1 \leq k \leq p$. Then $A \cdot B = [c_{ik}]$, where $c_{ik} = \Sigma_j a_{ij} b_{jk}$. Thus, for example,

$$\begin{bmatrix} 2 & 0 & 1 \\ 4 & 1 & 0 \\ 0 & 8 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & 2 \\ 3 & 1 & 0 \\ 5 & 8 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 10 & 5 \\ 3 & 5 & 8 \\ 29 & 16 & 1 \end{bmatrix}$$

It can be observed easily that the matrix multiplication is distributive over addition from left as well as from right in the sense that $(A + B) \cdot C = A \cdot C + B \cdot C$ and $A \cdot (B + C) = A \cdot B + A \cdot C$. Evidently, $A \cdot 0_{n \times p} = 0_{m \times p}$, and $0_{p \times m} \cdot A = 0_{p \times n}$. Again, since $\Sigma_k (\Sigma_j a_{ij} b_{jk}) c_{kl} = \Sigma_j a_{ij} (\Sigma_k b_{jk} c_{kl})$, it follows that the matrix multiplication is associative in the sense that $(A \cdot B) \cdot C = A \cdot (B \cdot C)$ whenever the products are defined. In particular, we have a multiplication $\cdot$ in $M_n(F)$. Note that matrix multiplication is not commutative, for example,

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

where as

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Thus, the set $M_n(F)$ of $n \times n$ matrices with entries in F together with matrix addition $+$, the multiplication by scalars, and the matrix multiplication $\cdot$ is an algebra in the sense of the following definition.

Definition 2.1.2 A vector space V over a field F together with an internal multiplication $\cdot$ on V is called an **algebra** over F if the following conditions hold:

1. The internal multiplication $\cdot$ is associative, i.e., $(x \cdot y) \cdot z = x \cdot (y \cdot z)$ for all $x, y, z \in V$.
2. $\cdot$ distributes over addition $+$, i.e., $(x + y) \cdot z = x \cdot z + y \cdot z$, and also $x \cdot (y + z) = x \cdot y + x \cdot z$ for all $x, y, z \in V$.
3. $\alpha(x \cdot y) = (\alpha x) \cdot y = x \cdot (\alpha y)$ for all $\alpha \in F$, and $x, y \in V$.

Let A be a $n \times m$ matrix. The $m \times n$ matrix A^t obtained by interchanging rows and columns of A is called the **transpose** of A. More precisely, if $A = [a_{ij}]$ is a $n \times m$

matrix, then the $m \times n$ matrix $A^t = [b_{ji}]$, where $b_{ji} = a_{ij}$ is called the **transpose** of A. Let $A = [a_{ij}]$ be a $n \times m$ matrix with entries in the field $\mathbb{C}$ of complex numbers. The matrix $\overline{A} = [b_{ij}]$, where $b_{ij} = \overline{a_{ij}}$ (the complex conjugate of a_{ij}) is called the **conjugate** of the matrix A. The matrix $A^\star = \overline{A}^t$ is called the **tranjugate**, also called the **hermitian conjugate** of A

Thus, for example

$$\begin{bmatrix} 2 & 0 & 1 \\ 4 & 1 & 0 \\ 0 & 8 & 1 \end{bmatrix}^t = \begin{bmatrix} 2 & 4 & 0 \\ 0 & 1 & 8 \\ 1 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 2+i & i & 1+i \\ 4+i & i & 0 \\ 1-i & 8 & 1+i \end{bmatrix}^\star = \begin{bmatrix} 2-i & 4-i & 1+i \\ -i & -i & 8 \\ 1-i & 0 & 1-i \end{bmatrix}$$

Proposition 2.1.3 *Let A, B be matrices with entries in a field F. Then*

(i) $(A + B)^t = A^t + B^t$
(ii) $(A^t)^t = A$.
(iii) $(a \cdot A)^t = a \cdot A^t$
(iv) $(A \cdot B)^t = B^t \cdot A^t$

provided the relevant sums and the products are defined.
Further, if A, B are matrices with entries in the field $\mathbb{C}$ of complex numbers, then

(v) $(A + B)^\star = A^\star + B^\star$
(vi) $(A^\star)^\star = A$.
(vii) $(a \cdot A)^\star = \overline{a} \cdot A^\star$
(viii) $(A \cdot B)^\star = B^\star \cdot A^\star$

provided the relevant sums and the products are defined.

Proof The identities (i), (ii), and (iii) are evident from the definition. We prove the (iv). Suppose that $A = [a_{ij}]$ is a $n \times m$ matrix, and $B = [b_{jk}]$ is a $m \times p$ matrix. Then, by the definition, $A \cdot B = [c_{ik}]$, where $c_{ik} = \Sigma_j a_{ij} b_{jk} = \Sigma_j v_{kj} u_{ji}$ where $v_{kj} = b_{jk}$ and $u_{ji} = a_{ij}$. By the definition $B^t = [v_{kj}]$, $A^t = [u_{ji}]$, and $(A \cdot B)^t = [w_{ki}]$, where $w_{ki} = c_{ik}$. This shows that the k_{th} row j_{th} column entry of both sides are same. This proves the result. The proofs of the rest of the identities are similar. ♯

2.2 Types of Matrices

1. **Identity matrix**. The $n \times n$ matrix all of whose diagonal entries are 1 and off diagonal entries are 0 is called the **identity** matrix of order n, and it is denoted by I_n. For example,

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

It can be checked that $I_n \cdot A = A = A \cdot I_m$ for every $n \times m$ matrix A. Indeed, if C is a $n \times m$ matrix such that $C \cdot A = A$ for every $n \times m$ matrix A, then $C = I_n$.

2. **Diagonal matrix**. A matrix $A = [a_{ij}]$ is called a **diagonal matrix** if all off diagonal entries are 0. Thus, $[a_{ij}]$ is a diagonal matrix if $a_{ij} = 0$ for all $i \neq j$. The diagonal matrix whose ith row ith column entry is α_i is denoted by $Diag(\alpha_1, \alpha_2, \ldots, \alpha_n)$. For example,

$$Diag(1, 2, 3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

The effect of multiplying the diagonal matrix $diag(\alpha_1, \alpha_2, \ldots, \alpha_n)$ to a $n \times m$ matrix A from left is to multiply the ith row by α_i. Thus $diag(\alpha_1, \alpha_2, \ldots, \alpha_n) \cdot [a_{ij}] = [b_{ij}]$, where $b_{ij} = \alpha_i a_{ij}$. Similarly, the effect of multiplying this matrix to a $m \times n$ matrix A from right is the same as multiplying the ith column by α_i. In particular, $diag(\alpha_1, \alpha_2, \ldots, \alpha_n) \cdot diag(\beta_1, \beta_2, \ldots \beta_n) = diag(\alpha_1\beta_1, \alpha_2\beta_2, \ldots, \alpha_n\beta_n)$.

3. **Scalar matrix**. A $n \times n$ diagonal matrix all of whose diagonal entries are same is called a **scalar** matrix. Thus, a scalar matrix is of the form αI_n, and effect of multiplying this matrix to a matrix A is αA.

4. **Symmetric matrix**. A matrix A is called a **symmetric matrix** if $A^t = A$. Thus, a diagonal matrix is a symmetric matrix. The matrix

$$\begin{bmatrix} 1 & 3 & 2 \\ 3 & 2 & 0 \\ 2 & 0 & 3 \end{bmatrix}$$

is a symmetric matrix. It follows from the Proposition 2.1.3 that sum of two symmetric matrices are symmetric, scalar multiple of a symmetric matrix is a symmetric matrix. Thus, the set $S_n(F)$ of all $n \times n$ symmetric matrices forms a subspace of $M_n(F)$. For all matrices A, AA^t is a symmetric matrix. For a square matrix A, $A + A^t$ is a symmetric matrix. Product of two symmetric matrices is symmetric if and only if they commute.

5. **Skew symmetric matrix**. A matrix A is called a **skew symmetric matrix** if $A^t = -A$. For example, the matrix

$$\begin{bmatrix} 0 & 3 & 2 \\ -3 & 0 & 0 \\ -2 & 0 & 0 \end{bmatrix}$$

is a skew symmetric matrix. It follows from the Proposition 2.1.3 that sum of two skew symmetric matrices are skew symmetric, scalar multiple of a skew symmetric matrix is a skew symmetric matrix. Thus, the set $SS_n(F)$ of all $n \times n$ skew symmetric

matrices forms a subspace of $M_n(F)$. $A - A^t$ is skew symmetric for all square matrices A. Product of two skew symmetric matrices is skew symmetric if and only if they anti commute in the sense that $A \cdot B = -B \cdot A$. Also observe that the diagonal entries of a skew symmetric matrices are 0.

Every square matrix A with entries in a field F can be uniquely represented as sum $A = \frac{A+A^t}{2} + \frac{A-A^t}{2}$ of a symmetric matrix $\frac{A+A^t}{2}$ and a skew symmetric matrix $\frac{A-A^t}{2}$ (prove the uniqueness of the representation).

6. **Hermitian matrix**. A matrix A with entries in the field $\mathbb{C}$ of complex numbers is called a **hermitian matrix** (also termed as **self adjoint**) if $A^\star = A$. Thus, a matrix A with real entries is Hermitian if and only if it is symmetric. The matrix

$$\begin{bmatrix} 1 & 3+i & 2 \\ 3-i & 2 & i \\ 2 & -i & 3 \end{bmatrix}$$

is a Hermitian matrix. Evidently, all diagonal entries of Hermitian matrices are real. It follows from the Proposition 2.1.3 that sum of two Hermitian matrices are Hermitian. However, only real scalar multiple of a Hermitian matrix is a Hermitian matrix. For all matrices A, $AA^\star$ is a Hermitian matrix. For a square matrix A, $A + A^\star$ is also a Hermitian matrix. Product of two Hermitian matrices is Hermitian if and only if they commute.

7. **Skew-Hermitian matrix**. A matrix A with entries in the field $\mathbb{C}$ of complex numbers is called a **skew-Hermitian matrix** if $A^\star = -A$. Thus, a matrix A with real entries is skew-Hermitian if and only if it is skew symmetric. The matrix

$$\begin{bmatrix} i & 3i-1 & 2 \\ 3i+1 & 2i & -1 \\ 2 & 1 & 3i \end{bmatrix}$$

is a skew-Hermitian matrix. Evidently, all diagonal entries of skew-Hermitian matrices are purely imaginary. It follows from the Proposition 2.1.3 that sums of two skew-Hermitian matrices are skew-Hermitian. However, only real scalar multiple of a skew-Hermitian matrix is a skew-Hermitian matrix. Observe that a matrix A is skew-Hermitian if and only if iA is a Hermitian matrix. For all matrices A, $iAA^\star$ is a skew-Hermitian matrix. For a square matrix A, $A - A^\star$ is also a skew-Hermitian matrix. Product of two skew-Hermitian matrices is skew-Hermitian if and only if they anticommute in the sense that $AB = -BA$.

Every square matrix A with entries in the field $\mathbb{C}$ of complex numbers can be uniquely represented as sum $A = \frac{A+A^\star}{2} + \frac{A-A^\star}{2}$ of a Hermitian matrix $\frac{A+A^\star}{2}$, and a skew-Hermitian matrix $\frac{A-A^\star}{2}$ (prove the uniqueness of the representation). In turn, it follows that every square matrix A with entries in the field $\mathbb{C}$ of complex numbers can be uniquely represented as $A = B + iC$, where B and C are Hermitian matrices.

8. **Nonsingular matrices**. A $n \times n$ matrix A is called a **nonsingular matrix** (also called an **invertible matrix**) if there is a $n \times n$ matrix B such that $A \cdot B = I_n = B \cdot A$.

Note that such a B, if exists, will be unique, for if B_1 and B_2 are such matrices, then $B_1 = B_1 \cdot I_n = B_1 \cdot (A \cdot B_2) = (B_1 \cdot A) \cdot B_2 = I_n \cdot B_2 = B_2$. If A is an invertible matrix, then the unique B such that $A \cdot B = I_n = B \cdot A$ is called the **Inverse** of A, and it is denoted by A^{-1}. Following are some simple observations:
(i) The identity matrix I_n is invertible and $I_n^{-1} = I_n$.
(ii) Consider a diagonal matrix $diag(\alpha_1, \alpha_2, \ldots, \alpha_n)$. As already observed in 2, $diag(\alpha_1, \alpha_2, \ldots, \alpha_n) \cdot [a_{ij}] = [b_{ij}]$, where $b_{ij} = \alpha_i a_{ij}$. Thus, $diag(\alpha_1, \alpha_2, \ldots, \alpha_n) \cdot [a_{ij}] = I_n$ if and only if $\alpha_i a_{ij} = 1$ for $j = i$, and 0 other wise. This is so if and only if $\alpha_i \neq 0$, $a_{ii} = \alpha_i^{-1}$ for each i, and $a_{ij} = 0$ for all $i \neq j$. This shows that $Diag(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is invertible if and only if each $\alpha_i \neq 0$, and then its inverse is $Diag(\alpha_1^{-1}, \alpha_2^{-1}, \ldots, \alpha_n^{-1})$.
(iii) Let A and B be invertible $n \times n$ matrices. Then, $(AB)(B^{-1}A^{-1}) = I_n = (B^{-1}A^{-1})(AB)$. This shows that AB is also invertible and $(AB)^{-1} = B^{-1}A^{-1}$.

In due course, we shall describe an algorithms to check if a matrix is invertible, and then to find its inverse.
9. **Triangular matrices**. A square matrix A is said to be an **upper (lower) triangular** matrix if all its below (above) diagonal entries are 0. More precisely, a $n \times n$ matrix $A = [a_{ij}]$ is called an **upper (lower) triangular** matrix if $a_{ij} = 0 \, for \, all \, i > j (i < j)$. It is called **strictly upper (lower)** triangular if in addition to that all the diagonal entries are also 0. For example,

$$\begin{bmatrix} 1 & 4 & 6 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

is an upper triangular matrix.

Clearly, the sum of any two **upper (lower) triangular** matrices is an **upper (lower) triangular** matrix. Also a scalar multiple of an **upper (lower) triangular** matrix is a **upper (lower) triangular** matrix. Thus, the set $T_+(n, F)(T_-(n, F))$ of upper (lower) triangular matrices forms a subspace of $M_n(F)$.

Further, $T_+(n, F)(T_-(n, F))$ is closed under matrix multiplication: For, let $A = [a_{ij}]$ and $B = [b_{jk}]$ be upper triangular matrices. Then $a_{ij} = 0 = b_{jk} \, for \, all \, i > j > k$. Let $A \cdot B = [c_{ik}]$. Then $c_{ik} = \Sigma_j a_{ij} b_{jk} = 0 \, for \, all \, i > k$.

Next, let $A = [a_{ij}] \in T_+(n, F)$ be a nonsingular matrix. Then there is a matrix $B = [b_{ij}]$ such that $B \cdot A = I_n$. Equating the first row first column entry from both side we get $b_{11}a_{11} = 1$. But then $a_{11} \neq 0$ and $b_{11} = a_{11}^{-1}$. Equating second row first column entry, we obtain that $b_{21}a_{11} = 0$. Hence $b_{21} = 0$. Similarly, equating ith row 1_{st} column entry we obtain that $b_{i1}a_{11} = 0$, and so $b_{i1} = 0 \, for \, all \, i > 1$. Equating the 1_{st} row 2_{nd} column entry, we get that $b_{11}a_{12} + b_{12}a_{22} = 0$, and equating the 2_{nd} row 2_{nd} column entry, we get $b_{22}a_{22} = 1$. Thus $a_{22} \neq 0$, $b_{22} = a_{22}^{-1}$, and $b_{12} = a_{22}^{-1}a_{11}^{-1}a_{12}$. Proceeding in this way we obtain that all the diagonal entries a_{ii} of A are nonzero, and then we can solve b_{ij} to get the inverse of A. We also observe that the inverse of A is also a member of $T_+(n, F)$. For example, consider the upper triangular matrix

$$\begin{bmatrix} 2 & 4 & 6 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

all of whose diagonal entries are nonzero. We find its inverse. Suppose that

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} 2 & 4 & 6 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Then we have the following equations:
$2a_{11} = 1,\ 4a_{11} + 2a_{12} = 0,\ 6a_{11} + 3a_{13} = 0,$
$2a_{21} = 0,\ 4a_{21} + 2a_{22} = 1,\ 6a_{21} + 3a_{23} = 0,$
$2a_{31} = 0,\ 4a_{31} + 2a_{32} = 0,\ 3a_{33} = 1$
Solving, we get that $a_{11} = \frac{1}{2},\ a_{12} = -1 = a_{13},\ a_{21} = a_{31} = a_{32} = 0,\ a_{23} = 0,\ a_{22} = \frac{1}{2},\ a_{33} = \frac{1}{3}$. Thus, the inverse of the said matrix is

$$\begin{bmatrix} \frac{1}{2} & -1 & -1 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

Block multiplication
We can multiply two matrices by using suitable blocks of their submatrices. More explicitly, let A be a $m \times n$ matrix, and B a $n \times p$ matrix. Suppose that $m = m_1 + m_2 + \cdots + m_r,\ n = n_1 + n_2 + \cdots n_s$, and $p = p_1 + p_2 + \cdots + p_t$, where m_i, n_j and p_k are positive integers. Then A and B can be expressed uniquely as

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1s} \\ A_{21} & A_{22} & \cdots & A_{2s} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ A_{r1} & A_{r2} & \cdots & A_{rs} \end{bmatrix},$$

where A_{ij} is a $m_i \times n_j$ matrix and

$$B = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1t} \\ B_{21} & B_{22} & \cdots & B_{2t} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ B_{s1} & B_{s2} & \cdots & B_{st} \end{bmatrix},$$

where B_{jk} is $n_j \times p_k$ matrix. Further, then

$$A \cdot B = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1t} \\ C_{12} & C_{22} & \cdots & C_{2t} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ C_{r1} & C_{r2} & \cdots & C_{rt} \end{bmatrix},$$

where $C_{ik} = \Sigma_{j=1}^{s} A_{ij} B_{jk}$.

2.3 System of Linear Equations

A system of m linear equations in n unknowns $x_1, x_2, \ldots, x_n$ over a field F is given by

$$\begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m = b_2 \\ . \quad . \quad . \quad \ldots \quad . \quad . \quad . \\ . \quad . \quad . \quad \ldots \quad . \quad . \quad . \\ . \quad . \quad . \quad \ldots \quad . \quad . \quad . \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m = b_n \end{pmatrix}, \tag{2.1}$$

where $a_{ij} \in F$.

Example 2.3.1 Following is a system of two linear equations in three unknowns over the field of real numbers:

$$3x_1 + 2x_2 + x_3 = 1.$$

$$x_1 + x_2 + x_3 = 2.$$

We say that a *n-tuple* $(a_1, a_2, \ldots, a_n)$ in F^n is a solution of the system (2.1) of linear equations if $x_1 = a_1, x_2 = a_2, \ldots, x_n = a_n$ satisfies all the equations in the system (2.1). Thus, $(-2, 3, 1)$ is a solution of the system of linear equations in the above example. $(-3, 5, 0)$ is also a solution to the above system. Indeed, there are infinitely many solutions which can be parametrized in terms of x_3 as $(x_3 - 3, 5 - 2x_2, x_3)$. Clearly, this represents a line.

Example 2.3.2 The system

$$x_1 + 2x_2 + 3x_3 = 1.$$

$$x_1 + x_2 + 3x_3 = 2.$$

$$4x_1 + 6x_2 + 12x_3 = 5.$$

of linear equations has no solution (why?).

where as

Example 2.3.3 The system

$$x_1 + 2x_2 = 1.$$

$$2x_1 + 2x_2 = a.$$

has a unique solution for all a (why?).

Definition 2.3.4 A system of linear equations is said to be **consistent** if it has a solution. It is said to be **inconsistent** otherwise.

The Example 2.3.1 is consistent having infinitely many solutions, the Example 2.3.2 is inconsistent, whereas Example 2.3.3 is consistent with unique solution.

Most of the problems in real life, in engineering, in industries, in social life, and in medical science can be modeled in terms of systems of linear equations. As such, describing and interpreting the solutions of a system of linear equations is one of the main themes of linear algebra. In the following few sections we shall concentrate on this.

The system (2.1) of m linear equations in n unknowns can be expressed in a single matrix equation

$$A\overline{x}^t = \overline{b}^t \tag{2.2}$$

where $A = [a_{ij}]$ is the $m \times n$ matrix whose i_{th} row j_{th} column entry is a_{ij}, $\overline{x} = [x_1, x_2, \ldots, x_n] \in F^n$ is the $1 \times n$ row matrix of unknowns, and $\overline{b} = [b_1, b_2, \ldots, b_m] \in F^m$ is the $1 \times m$ matrix.

Thus, the system of linear equations in Example 2.3.1 can be expressed as

$$\begin{bmatrix} 3 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

The matrix A in (2.2) is called the **coefficient matrix** of the system (2.1) of linear equations, and the $m \times (n+1)$ matrix $A^+ = [A\ \overline{b}^t]$ whose first n columns are those of A, and the last $(n+1)_{th}$ column is $\overline{b}^t$, is called the **augmented matrix** of the system of linear equations.

Thus, the coefficient matrix of the Example 2.3.2 is

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 3 \\ 4 & 6 & 12 \end{bmatrix},$$

and the augmented matrix of the example is

$$\begin{bmatrix} 2\,4\,6\,1 \\ 0\,2\,0\,2 \\ 0\,0\,3\,5 \end{bmatrix}$$

Definition 2.3.5 A system of linear equations given by the matrix equation

$$A\overline{x}^t = \overline{0}^t \ldots . \tag{2.3}$$

is called a **homogeneous system** of linear equations. It is also called the **homogeneous part** of the system of linear equations given by

$$A\overline{x}^t = \overline{b}^t.$$

Proposition 2.3.6 *A homogeneous system of linear equations given by the matrix equation*

$$A\overline{x}^t = \overline{0}^t.$$

is always consistent, and the set of solutions of the homogeneous system is a subspace of F^n.

Proof Let $N(A)$ denote the set of all solutions of $A\overline{x}^t = \overline{0}^t$. Since $A\overline{0}^t = \overline{0}^t$, it follows that $\overline{0} \in N(A)$. Let $\overline{u}, \overline{v} \in N(A)$, and $a, b \in F$. Then $A(a\overline{u} + b\overline{v})^t = aA\overline{u}^t + bA\overline{v}^t = \overline{0}^t$. This shows that $a\overline{u} + b\overline{v} \in N(A)$. It follows that $N(A)$ is a subspace of F^n. ♯

Definition 2.3.7 The subspace $N(A)$ described in the above proposition is called the **solution space** of the system (2.3) of linear equations, and it is also called the **null space** of the matrix A. The dimension of the null space $N(A)$ is called the **nullity** of A, and it is denoted by $n(A)$. If $\{\overline{u_1}, \overline{u_2}, \ldots, \overline{u_{n(A)}}\}$ is a basis of $N(A)$, then any solution of (2.3) is uniquely expressed as $c_1\overline{u_1} + c_2\overline{u_2} + \cdots + c_{n(A)}\overline{u_{n(A)}}$, where $c_1, c_2, \ldots, c_{n(A)}$ are constants in F. As such $c_1\overline{u_1} + c_2\overline{u_2} + \cdots + c_{n(A)}\overline{u_{n(A)}}$ is called a general solution of the homogeneous system (2.3).

A little later, we shall give an algorithm to find $N(A)$, indeed a basis of $N(A)$, and so also a general solution of the system (2.3) of linear equations.

Proposition 2.3.8 *Suppose that the system of linear equations given by the matrix equation*

$$A\overline{x}^t = \overline{b}^t.$$

is consistent, and $\overline{a} = [a_1, a_2, \ldots, a_n]$ *is a solution of the above equation. Then the coset* $\overline{a} + N(A) = \{\overline{a} + \overline{u} \mid \overline{u} \in N(A)\}$ *is the complete set of all solutions of the system of linear equations. In turn, if* $\{\overline{u_1}, \overline{u_2}, \ldots, \overline{u_{n(A)}}\}$ *is a basis of* $N(A)$, *then* $\overline{a} + c_1\overline{u_1} + c_2\overline{u_2} + \cdots + c_{n(A)}\overline{u_{n(A)}}$ *is a general solution of the system of linear equations, where* $c_1, c_2, \ldots, c_{n(A)}$ *are arbitrary constants.*

Proof Since $\overline{a}$ is a solution of $A\overline{x}^t = \overline{b}^t$, $A\overline{a}^t = \overline{b}^t$. If $\overline{u} \in N(A)$, then $A\overline{u}^t = \overline{0}^t$. But, then $A(\overline{a} + \overline{u})^t = (A\overline{a}^t + A\overline{u}^t) = (\overline{b}^t + \overline{0}^t) = \overline{b}^t$. This shows that $\overline{a} + \overline{u}$ is also a solution of $A\overline{x}^t = \overline{b}^t$. Conversely, let $\overline{c}$ be a solution of $A\overline{x}^t = \overline{b}^t$. Then $A\overline{c}^t = \overline{b}^t$. Hence $A(\overline{c} - \overline{a})^t = (A\overline{c}^t - A\overline{a}^t) = \overline{0}$. It follows that $(\overline{c} - \overline{a}) \in N(A)$. This shows that $\overline{c} \in \overline{a} + N(A)$. The rest is an immediate observation. ♯

Definition 2.3.9 The subspace $R(A)$ of F^n generated by the set $\{R_1(A), R_2(A), \ldots, R_m(A)\}$ of the rows of A is called the **row space** of A, and the dimension of $R(A)$ is called the **row rank** of A. Thus, the maximum number of linearly independent rows of a matrix is the row rank of A. Similarly, the subspace $C(A)$ of F^m (the elements of F^m treated as columns) is called the **column space** of A, and the dimension of $C(A)$ is called the **column rank** of A. Again, it follows that the maximum number of linearly independent columns of A is the column rank of A. We shall see, in due course, that row rank is same as column rank, and it is called the **rank**of A. The rank of A is denoted by $r(A)$.

Proposition 2.3.10 *The system of linear equations given by the matrix equation*

$$A\overline{x}^t = \overline{b}^t.$$

is consistent if and only if the column rank of A is same as that of the augmented matrix A^+.

Proof The system of linear equations given by the matrix equation $A\overline{x}^t = \overline{b}^t$ is also expressible as

$$x_1C_1(A) + x_2C_2(A) + \cdots + x_nC_n(A) = \overline{b}^t,$$

where $\overline{x} = [x_1, x_2, \ldots, x_n]$, and $C_i(A)$ denotes the i_{th} column of A. Thus, the equation has a solution if and only if $\overline{b}^t$ is a linear combination of the columns of A. This is equivalent to say that the column space $C(A)$ of A is same as the column space $C(A^+)$ of the augmented matrix A^+. Since $C(A) \subseteq C(A^+)$, this is equivalent to the fact that column rank of A is same as that of A^+. ♯

We shall look at an algorithm to find the rank of a matrix, and also an algorithm to find a general solution of $A\overline{x}^t = \overline{b}^t$ provided it is consistent.

2.4 Gauss Elimination, Elementary Operations, Rank, and Nullity

Definition 2.4.1 Two systems of m linear equations in n unknowns are said to be equivalent if they have same set of solutions.

Example 2.4.2 The system

$$x_1 + 2x_2 = 1$$

$$2x_1 + 2x_2 = a$$

of two linear equations in two unknowns is equivalent to the system

$$x_1 + 2x_2 = 1$$

$$3x_1 + 4x_2 = a + 1,$$

for they have same set of solutions, whereas the system is not equivalent to

$$x_1 + 2x_2 = 1$$

$$2x_1 + 3x_2 = a$$

In what follows, we shall introduce an algorithm called the **Gaussian elimination** to reduce a system of linear equations into an equivalent system of linear equations from which the solution will become apparent.

Definition 2.4.3 Following operations on a system of linear equations are called the **elementary operations** on the system of linear equations, and the corresponding operations on coefficient and augmented matrices are called the **Elementary row operations** on the matrices:

1. Interchange any two equations in the system.
2. Multiply an equation in the system by a nonzero member of the field.
3. Add a nonzero multiple of an equation in the system to another equation in the system.

In turn, the corresponding elementary **row operations** on matrices are:

1. Interchange any two row of the matrix.
2. Multiply a row of the matrix by a nonzero element of the field.
3. Add a nonzero multiple of a row of the matrix to another row.

The following proposition is an immediate observation.

Proposition 2.4.4 *Any two system of linear equations which differ by a finite sequence of elementary operations are equivalent.* ♯

We shall first discuss an algorithm to find the space of solutions of a homogeneous system of linear equations given by the matrix equation $A\overline{x}^t = \overline{0}^t$. More precisely, we derive an algorithm to find a basis of the null space $N(A)$ of A so that every solution of the system is unique linear combination of the basis members.

Proposition 2.4.5 *The null space $N(A)$, and so also the nullity $n(A)$ of a matrix A remain invariant under the elementary row operations.*

Proof Follows from the Proposition 2.4.4. ♯

Proposition 2.4.6 *The row space $R(A)$ and so also the row rank of a matrix A remain invariant under the elementary row operations.*

Proof Interchange of any two rows of a matrix will not change the row space as the set of rows will not change. Since the subspace of F^n generated by the set $\{R_1(A), R_2(A), \ldots, R_m(A)\}$ of rows of A is the same as the subspace of F^n generated by $\{R_1(A), R_2(A), \ldots, aR_j(A), \ldots, R_m(A)\}$ for each nonzero $a \in F$ and $j \leq m$, it follows that the row space of a matrix remains the same if we multiply a row of the matrix by a nonzero member of the field. Finally, since the subspace of F^n generated by the set $\{R_1(A), R_2(A), \ldots, R_m(A)\}$ of rows of A is the same as the subspace of F^n generated by $\{R_1(A), R_2(A), \ldots, R_k(A) + aR_j(A), \ldots, R_m(A)\}$ for each nonzero $a \in F$ and $j \neq k$, it follows that the row space of a matrix remains the same if we add a nonzero multiple of a row to another row. ♯

The column space of a matrix, in general, is not invariant under elementary row operations. However,

Proposition 2.4.7 *The column rank of a matrix remains invariant under elementary row operations.*

Proof Let A be a matrix, and A' a matrix obtained by applying any of the elementary row operations on A. Then evidently,

$$x_1 C_{i_1}(A) + x_2 C_{i_2}(A) + \cdots + x_r C_{i_r}(A) = \overline{0}^t$$

if and only if

$$x_1 C_{i_1}(A') + x_2 C_{i_2}(A') + \cdots + x_r C_{i_r}(A') = \overline{0}^t$$

This means that the maximum number of linearly independent columns of A is same as that of A'. Thus, the column rank of A is same as that of A'. ♯

We shall describe an algorithm to transform a matrix in to a special form, called a **reduced row echelon form**, of the matrix by using elementary row operations, and from which a basis for the null space of the matrix, and also a basis of the row space of the matrix can be easily obtained.

Definition 2.4.8 A $m \times n$ matrix $A = [a_{ij}]$ is said to be a matrix in **reduced row (column) echelon form**, or it is said to be a **reduced row echelon matrix** if the following hold:

(i) The first nonzero entry in each row (column) is 1. This entry is called a **pivot** entry, and the corresponding columns (rows) are called **pivot column (row)** of the matrix. The columns (rows) which are not pivot columns (rows) are called the **free** columns (rows). The unknown variable corresponding to pivot columns are called **pivot variables**, and those corresponding to free columns are called **free Variables**.

(ii) The pivot entry in any row (column) is towards right (bottom) side to the pivot entry in the previous row (column).
(iii) All of the rest of the entries in a pivot column (row) are 0.
(iv) All the zero rows (columns) are towards bottom (right).

Example 2.4.9 The matrix

$$A = \begin{bmatrix} 1 & 2 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

is in reduced row echelon form. The 1st row 1st column, the 2nd row 3rd column, and the 3rd row 4th column entries are pivot entries, 2nd and 5th columns are free columns. x_1, x_3 and x_4 are pivot variables. x_2 and x_5 are free variables.

Proposition 2.4.10 *Let A be a $m \times n$ matrix with entries in a field F and which is in reduced row echelon form. Suppose that the columns $C_{i_1}(A), C_{i_2}(A), \ldots, C_{i_r}(A)$ with $i_1 < i_2 < \cdots < i_r$ are pivot columns and the columns $C_{j_1}(A), C_{j_2}(A), \ldots, C_{j_s}(A)$ with $j_1 < j_2 < \cdots < j_s$ are free columns. Then,*

(i) the first r rows $R_1(A), R_2(A), \ldots, R_r(A)$ are nonzero rows, and they form a basis of the row space $R(A)$ of A,
(ii) the number of pivots is the row rank of A,
(iii) the pivot columns form a basis of the column space of A,
(iv) row rank of A is the same as the column rank of A. Indeed, it is the number of pivots.

Proof (i) Since each nonzero row contains a unique pivot entry, and the zero rows are towards the bottom, it follows that $R_1(A), R_2(A), \ldots, R_r(A)$ are precisely the nonzero rows of the matrix. Since the pivot entries 1 in $R_1(A), R_2(A), \ldots, R_r(A)$ appear in different columns $i_1, i_2, \ldots, i_r$, it follows that the set $\{R_1(A), R_2(A), \ldots, R_r(A)\}$ of nonzero row of A is linearly independent. As such, it forms a basis of the row space $R(A)$ of A.
(ii) Follows from (i).
(iii) Clearly, the set $\{C_{i_1}(A), C_{i_2}(A), \ldots, C_{i_r}(A)\}$ of pivot columns form a linearly independent set, for the k_{th} row entry in the pivot column $C_{i_k}(A)$ is 1 and the rest of the entries in this column are 0. It is also evident that all the free columns are linear linear combinations of the pivot columns. Indeed,

$$C_{j_l}(A) = a_{1j_l} C_{i_1}(A) + a_{2j_l} C_{i_2}(A) + \cdots + a_{rj_l} C_{i_r}(A).$$

(iv) Follows from (iii). ♯

Proposition 2.4.11 *Consider the homogeneous system of linear equations given by the matrix equation*

$$A\overline{x}^t = \overline{o}^t,$$

where A is a reduced row echelon $m \times n$ matrix with entries in a field F. Suppose that the columns $C_{i_1}(A), C_{i_2}(A), \ldots, C_{i_r}(A)$ with $i_1 < i_2 < \cdots < i_r$ are pivot columns, and the columns $C_{j_1}(A), C_{j_2}(A), \ldots, C_{j_s}(A)$ with $j_1 < j_2 < \cdots < j_s$ are free columns. Then the pivot variables $x_{i_1}, x_{i_2}, \ldots, x_{i_r}$ in the homogeneous system of linear equations are uniquely expressible in terms of free variables $x_{j_1}, x_{j_2}, \ldots, x_{j_s}$ as

$$x_{i_t} = -\sum_{k=1}^{s} a_{tj_k} x_{j_k}.$$

The set $\{\overline{u^1}, \overline{u^2}, \ldots, \overline{u^s}\}$ is a basis for the space $N(A)$ of solutions of the homogeneous system, where $\overline{u^k} = (u_1^k, u_2^k, \ldots, u_n^k)$ is the unique solution of the homogeneous system corresponding to the choice $x_{j_l} = 0$, $l \neq k$, and $x_{j_k} = 1$ of the free variables. Indeed, $u_{j_l}^k = 0$ for $l \neq k$, $u_{j_k}^k = 1$, and $u_{i_t}^k = -a_{tj_k}$. The nullity $n(A) = s$, the number of free variables.

Proof Under the assumption, for all $t \leq r$, $a_{ti_t} = 1$ and $a_{li_t} = 0$ for $l \neq t$. The corresponding homogeneous system of linear equations is given by

$$a_{1i_1}x_{i_1} + a_{1j_1}x_{j_1} + a_{1j_2}x_{j_2} + \cdots + a_{1j_s}x_{j_s} = 0.$$

$$a_{2i_2}x_{i_2} + a_{2j_1}x_{j_1} + a_{2j_2}x_{j_2} + \cdots + a_{2j_s}x_{j_s} = 0.$$

.....

.....

$$a_{ri_1}x_{i_1} + a_{rj_1}x_{j_1} + a_{rj_2}x_{j_2} + \cdots + a_{rj_s}x_{j_s} = 0.$$

the rest of the equations, if any, are the identities

$$0x_1 + 0x_2 + \cdots + 0x_n = 0.$$

Evidently, each pivot variable is uniquely expressible in terms of free variable as described in the proposition. Further, the set $S = \{\overline{u^1}, \overline{u^2}, \ldots, \overline{u^s}\}$ of solutions is a basis of the space $N(A)$ of solutions, for any solution with values $\alpha_1, \alpha_2, \ldots, \alpha_s$ to the free variables $x_{j_1}, x_{j_2}, \ldots, x_{j_s}$ is uniquely expressible as linear combination $\alpha_1\overline{u^1} + \alpha_2\overline{u^2} + \cdots, \alpha_s\overline{u^s}$. The rest is evident. ♯

Proposition 2.4.12 *Consider the system of linear equations given by the matrix equation*

$$A\overline{x}^t = \overline{b}^t,$$

where A is a reduced row echelon $m \times n$ matrix with entries in a field F. Suppose that the columns $C_{i_1}(A), C_{i_2}(A), \ldots, C_{i_r}(A)$ with $i_1 < i_2 < \cdots < i_r$ are pivot columns and the columns

$C_{j_1}(A), C_{j_2}(A), \ldots, C_{j_s}(A)$ with $j_1 < j_2 < \cdots < j_s$ are free columns. Then the system of linear equations is consistent if and only if $b_k = 0$ for all $k \geq r+1$, or equivalently, rank(A) = rank(A^+). Further, then $\overline{v} = (v_1, v_2, \ldots, v_n)$, where $v_{i_t} = -a_{tj_1} + b_t$, $1 \leq t \leq r$, $v_{j_1} = 1$ and $v_{j_l} = 0, 2 \leq l \leq s$, is a particular solution of the above nonhomogeneous system. Finally, a general solution $\overline{x}$ of the system of linear equation is given by

$$\overline{x} = \overline{v} + c_1\overline{u^1} + c_2\overline{u^2} + \cdots + c_s\overline{u^s},$$

where $c_1, c_2, \ldots, c_s$ are arbitrary constants.

Proof From the previous proposition, a general solution of the homogeneous part of the above nonhomogeneous system of linear equations is given by

$$c_1\overline{u^1} + c_2\overline{u^2} + \cdots + c_s\overline{u^s}.$$

Further, the system of linear equations is given by

$$a_{1i_1}x_{i_1} + a_{1j_1}x_{j_1} + a_{1j_2}x_{j_2} + \cdots + a_{1j_s}x_{j_s} = b_1.$$

$$a_{2i_2}x_{i_2} + a_{2j_1}x_{j_1} + a_{2j_2}x_{j_2} + \cdots + a_{2j_s}x_{j_s} = b_2.$$

.....

.....

$$a_{ri_1}x_{i_1} + a_{rj_1}x_{j_1} + a_{rj_2}x_{j_2} + \cdots + a_{rj_s}x_{j_s} = b_r.$$

The rest of the equations, if any, are the identities

$$0x_1 + 0x_2 + \cdots + 0x_n = b_k,\ k \geq r+1.$$

Clearly, the system is inconsistent if $b_k \neq 0$ for any $k \geq r+1$. Now, suppose that $b_k = 0$ *for all* $k \geq r+1$. Putting the free variable $x_{j_1} = 1$, and $x_{j_k} = 0$ for $2 \leq k \leq s$, we get a particular solution $\overline{v} = (v_1, v_2, \ldots, v_n)$, where $v_{i_t} = -a_{tj_1} + b_t$, $1 \leq t \leq r$, $v_{j_1} = 1$, and $v_{j_l} = 0, 2 \leq l \leq s$ of the system. From the Proposition 2.3.8, we get a general solution

$$\overline{x} = \overline{v} + c_1\overline{u^1} + c_2\overline{u^2} + \cdots + c_s\overline{u^s}$$

of the system, where $c_1, c_2, \ldots, c_s$ are arbitrary constants. ♯

Example 2.4.13 Consider the system of linear equations given by the matrix equation

$$A\overline{x}^t = \overline{b}^t,$$

where A is the matrix given by

$$A = \begin{bmatrix} 1\,0\,1\,0\,2 \\ 0\,1\,1\,0\,1 \\ 0\,0\,0\,1\,1 \\ 0\,0\,0\,0\,0 \end{bmatrix}$$

The corresponding system of linear equations is given by

$$x_1 + 0x_2 + x_3 + 0x_4 + 2x_5 = b_1.$$

$$0x_1 + x_2 + x_3 + 0x_4 + x_5 = b_2.$$

$$0x_1 + 0x_2 + 0x_3 + x_4 + x_5 = b_3.$$

$$0x_1 + 0x_2 + 0x_3 + 0x_4 + 0x_5 = b_4.$$

The matrix A is in reduced row echelon form with the pivot columns C_1, C_2, C_4, and the free columns C_3 and C_5. The nonzero rows R_1, R_2, R_3 form a basis of row space, and the pivot columns C_1, C_2, C_4 of A form a basis of the column space of A. Row rank = 3 = Column rank of A. For the system to be consistent $b_4 = 0$. Assuming that $b_4 = 0$, we find a general solution of the system. We first find a basis of the solution space $N(A)$ of the homogeneous part $A\overline{x}^t = \overline{0}^t$ of the given system of linear equations. x_3 and x_5 are free variables. Putting $x_3 = 1$ and $x_4 = 0$, we get a solution $\overline{u^1} = (-1, -1, 1, 0, 0)$ of the homogeneous part of the system. Further putting $x_3 = 0$ and $x_5 = 1$, we get a solution $\overline{u^2} = (-2, -1, 0, -1, 1)$ of the homogeneous part of the system. The set $\{\overline{u^1}, \overline{u^2}\}$ is a basis of the space $N(A)$ of solutions of the homogeneous part. Nullity of A is 2. Finally, putting $x_3 = 1$ and $x_4 = 0$, we get a particular solution $\overline{v} = (-1 + b_1, -1 + b_2, 1, b_3, 0)$ of the given nonhomogeneous system of linear equations. In turn, a general solution of the given nonhomogeneous system of linear equations is $\overline{v} + c_1\overline{u^1} + c_2\overline{u^2}$.

Observe that a square matrix in reduced row echelon form has no zero rows if and only if all the rows, and so also all columns have pivots, or equivalently, it is the identity matrix. Since a matrix with a zero row is singular, we have the following:

Proposition 2.4.14 *A square matrix in reduced row echelon form is nonsingular if and only if it is the identity matrix.* ♯

Elementary operations on a system of linear equations, or equivalently, elementary row operations on the coefficient and augmented matrices, transform the system into equivalent system of linear equations. Further, if the coefficient matrix of the system of linear equations is in reduced row echelon form, then as observed above, a general solution of the system is easily obtained. As such, it is prompting to discover, if possible, an algorithm to reduce an arbitrary matrix

in to a matrix in reduced row echelon form by using elementary row operations. The following theorem gives an algorithm.

Theorem 2.4.15 *Using elementary row operations, every matrix can be reduced to a matrix in reduced row echelon form.*

Proof Let A be a $m \times n$ matrix. If A is the zero matrix, then it is already in reduced row echelon form. Suppose that A is nonzero matrix. Let j_1 be the least number such that the column $C_{j_1}(A)$ is a nonzero column. Further, let i_1 be the smallest number such that $a_{i_1 j_1} \neq 0$. Interchanging the i_1th row and the first row, we may assume that $a_{1j_1} \neq 0$, and $a_{ik} = 0$ for all $k < j_1$. Multiplying the first row by $a_{1j_1}^{-1}$, we may assume that $a_{1j_1} = 1$, and $a_{ik} = 0$ for all $k < j_1$. Next, adding $-a_{ij_1}$ times the first row to the ith row for each $i \geq 2$, we reduce A to a matrix $[a_{ij}]$, where $a_{1j_1} = 1$, $a_{ij_1} = 0$ for all $i \geq 2$, and $a_{ik} = 0$ for all $k \leq j_1 - 1$. If in this reduced matrix $a_{ij} = 0$ for all $i \geq 2$, then it is already in reduced row echelon form. If not, let j_2 be the smallest number such that $a_{ij_2} \neq 0$ for some $i \geq 2$. Further, let i_2 be the smallest number greater than 2 such that $a_{i_2 j_2} \neq 0$. Note that $j_2 > j_1$. Interchanging the i_2th row and the second row, we may assume that $a_{2j_2} \neq 0$. Then multiplying the second row by $a_{2j_2}^{-1}$, we may assume that $a_{2j_2} = 1$. In turn, adding $-a_{ij_2}$ times the second row to the ith row for each $i \neq 2$, A may have been reduced to a matrix in reduced row echelon form. If not, proceed as before. This process reduces A in to reduced row echelon form after finitely many steps (if worst comes, at the n_{th} step). ♯

Corollary 2.4.16 *Row rank of a matrix is the same as the column rank of the matrix.*

Proof From the Proposition 2.4.6, and the Proposition 2.4.7, row rank and column rank of a matrix are invariant under elementary row operations. From the Proposition 2.4.10(iv), row rank of a matrix in reduced row echelon form is same as its column rank (equal to the number of pivots). Combining this with the Theorem 2.4.15, the result follows. ♯

Definition 2.4.17 Row rank of a matrix A, or equivalently, the column rank of a matrix is called the **rank** of the matrix. The rank of a matrix A is denoted by $r(A)$.

Corollary 2.4.18 *Let A be a $m \times n$ matrix. Then $r(A) + n(A) = n$.*

Proof Since the rank and the nullity remain invariant under elementary row operations, using Theorem 2.4.15, it is sufficient to prove the result for matrices in reduced row echelon form. For a matrix A in reduced row echelon form, $r(A)$ is the number of pivot columns and $n(A)$ is the number of free columns. Clearly, a column is either a pivot column or a free column. ♯

Example 2.4.19 Consider the system of linear equations

$$2x_3 + 3x_4 + 8x_5 = 1.$$

$$2x_1 + 4x_2 + x_3 + 5x_5 = 0.$$

$$x_1 + 2x_2 + x_3 + x_4 + 5x_5 = 2.$$

$$5x_1 + 10x_2 + 6x_3 + 6x_4 + 28x_5 = a.$$

The corresponding coefficient matrix A is

$$A = \begin{bmatrix} 0 & 0 & 2 & 3 & 8 \\ 2 & 4 & 1 & 0 & 5 \\ 1 & 2 & 1 & 1 & 5 \\ 5 & 10 & 6 & 6 & 28 \end{bmatrix},$$

and the augmented matrix A^+ is

$$A^+ = \begin{bmatrix} 0 & 0 & 2 & 3 & 8 & 1 \\ 2 & 4 & 1 & 0 & 5 & 0 \\ 1 & 2 & 1 & 1 & 5 & 2 \\ 5 & 10 & 6 & 6 & 28 & a \end{bmatrix}.$$

We discuss the consistency of the above system of linear equations, and if consistent, we determine a general solution. For the purpose, we reduce the coefficient matrix A, and also the augmented matrix A^+ to reduced row echelon forms simultaneously by using the algorithm described in the above theorem. The 1st column of A is nonzero, and the smallest number i for which $a_{i1} \neq 0$ is 2. Thus, interchanging the 1st and the 2nd rows of A, and of A^+, A is transformed to

$$\begin{bmatrix} 2 & 4 & 1 & 0 & 5 \\ 0 & 0 & 2 & 3 & 8 \\ 1 & 2 & 1 & 1 & 5 \\ 5 & 10 & 6 & 6 & 28 \end{bmatrix},$$

and A^+ is transformed to

$$\begin{bmatrix} 2 & 4 & 1 & 0 & 5 & 0 \\ 0 & 0 & 2 & 3 & 8 & 1 \\ 1 & 2 & 1 & 1 & 5 & 2 \\ 5 & 10 & 6 & 6 & 28 & a \end{bmatrix}.$$

Now, multiplying the 1st row by $\frac{1}{2}$, the matrices are transformed to

$$\begin{bmatrix} 1 & 2 & \frac{1}{2} & 0 & \frac{5}{2} \\ 0 & 0 & 2 & 3 & 8 \\ 1 & 2 & 1 & 1 & 5 \\ 5 & 10 & 6 & 6 & 28 \end{bmatrix},$$

and to

$$\begin{bmatrix} 1 & 2 & \frac{1}{2} & 0 & \frac{5}{2} & 0 \\ 0 & 0 & 2 & 3 & 8 & 1 \\ 1 & 2 & 1 & 1 & 5 & 2 \\ 5 & 10 & 6 & 6 & 28 & a \end{bmatrix}.$$

Further, adding -1 times the 1st row to the 3rd row, and adding -5 times the 1st row to the 4th row, the matrices are transformed to

$$\begin{bmatrix} 1 & 2 & \frac{1}{2} & 0 & \frac{5}{2} \\ 0 & 0 & 2 & 3 & 8 \\ 0 & 0 & \frac{1}{2} & 1 & \frac{5}{2} \\ 0 & 0 & \frac{7}{2} & 6 & \frac{31}{2} \end{bmatrix},$$

and to

$$\begin{bmatrix} 1 & 2 & \frac{1}{2} & 0 & \frac{5}{2} & 0 \\ 0 & 0 & 2 & 3 & 8 & 1 \\ 0 & 0 & \frac{1}{2} & 1 & \frac{5}{2} & 2 \\ 0 & 0 & \frac{7}{2} & 6 & \frac{31}{2} & a \end{bmatrix}.$$

Here, in this transformed matrix, $a_{i2} = 0$ *for all* $i \geq 2$. Thus, the 2nd column is a free column. We look at the 3rd column. The 2nd row 3rd column entry $a_{23} = 2 \neq 0$. We divide the 2nd row by 2 to get the pivot entry 1 in 2nd row 3rd column. The matrices, thus, reduce to

$$\begin{bmatrix} 1 & 2 & \frac{1}{2} & 0 & \frac{5}{2} \\ 0 & 0 & 1 & \frac{3}{2} & 4 \\ 0 & 0 & \frac{1}{2} & 1 & \frac{5}{2} \\ 0 & 0 & \frac{7}{2} & 6 & \frac{31}{2} \end{bmatrix},$$

and to

$$\begin{bmatrix} 1 & 2 & \frac{1}{2} & 0 & \frac{5}{2} & 0 \\ 0 & 0 & 1 & \frac{3}{2} & 4 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 1 & \frac{5}{2} & 2 \\ 0 & 0 & \frac{7}{2} & 6 & \frac{31}{2} & a \end{bmatrix}.$$

In turn, to make all other entries in this pivot column 0, we add $-\frac{1}{2}$ times the 2nd row to the 1st row, $-\frac{1}{2}$ times the 2nd row to the 3rd row, and $-\frac{7}{2}$ times the 2nd row to the 4th row. The matrices reduce to

$$\begin{bmatrix} 1 & 2 & 0 & -\frac{3}{4} & \frac{1}{2} \\ 0 & 0 & 1 & \frac{3}{2} & 4 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{3}{4} & \frac{3}{2} \end{bmatrix},$$

and to

$$\begin{bmatrix} 1 & 2 & 0 & -\frac{3}{4} & \frac{1}{2} & -\frac{1}{4} \\ 0 & 0 & 1 & \frac{3}{2} & 4 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{7}{4} \\ 0 & 0 & 0 & \frac{3}{4} & \frac{3}{2} & a - \frac{7}{4} \end{bmatrix}.$$

The 3rd row 4th column entry $a_{34} = \frac{1}{4} \neq 0$. We multiply the 3rd row by 4 to get the pivot entry 1 in 3rd row 4th column. Thus, the matrices further reduce to

$$\begin{bmatrix} 1 & 2 & 0 & -\frac{3}{4} & \frac{1}{2} \\ 0 & 0 & 1 & \frac{3}{2} & 4 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & \frac{3}{4} & \frac{3}{2} \end{bmatrix},$$

and to

$$\begin{bmatrix} 1 & 2 & 0 & -\frac{3}{4} & \frac{1}{2} & -\frac{1}{4} \\ 0 & 0 & 1 & \frac{3}{2} & 4 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 2 & 7 \\ 0 & 0 & 0 & \frac{3}{4} & \frac{3}{2} & a - \frac{7}{4} \end{bmatrix}.$$

In turn, we add $\frac{3}{4}$ times the 3rd row to the 1st row, $-\frac{3}{2}$ times the 3rd row to the 2nd row, and the $\frac{-3}{4}$ times the 3rd row to the 4th row to make the rest of the entries in this pivot column 0. The coefficient matrix A reduces to the following matrix

$$\begin{bmatrix} 1 & 2 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

which is in reduced row echelon form, and the augmented matrix A^+ gets transformed to

$$\begin{bmatrix} 1 & 2 & 0 & 0 & 2 & 5 \\ 0 & 0 & 1 & 0 & 1 & -10 \\ 0 & 0 & 0 & 1 & 2 & 7 \\ 0 & 0 & 0 & 0 & 0 & a - 7 \end{bmatrix}.$$

Thus, the given system of linear equations is equivalent to a system of linear equations whose coefficient matrix is

$$\begin{bmatrix} 1 & 2 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and the augmented matrix is

$$\begin{bmatrix} 1\,2\,0\,0\,2 & 5 \\ 0\,0\,1\,0\,1 & -10 \\ 0\,0\,0\,1\,2 & 7 \\ 0\,0\,0\,0\,0 & a-7 \end{bmatrix}.$$

In turn, using the discussions and the results above, we have the following: (i) A basis of the row space of A is $\{(1, 2, 0, 0, 2), (0, 0, 1, 0, 1), (0, 0, 0, 1, 2)\}$. The rank $r(A) = 3$.
(ii) Putting the free variable $x_2 = 1$, and the free variable $x_5 = 0$, we get a solution $(-2, 1, 0, 0, 0)$ of the homogeneous part of the system. Further, putting the free variable $x_2 = 0$, and the free variable $x_5 = 1$, we get another solution $(-2, 0, -1, -2, 1)$ of the homogeneous part of the system. The set $\{(-2, 1, 0, 0, 0), (-2, 0, -1, -2, 1)\}$ is a basis of the solution space $N(A)$ of the homogeneous part. A general solution of the homogeneous part of the system is

$$c_1(-2, 1, 0, 0, 0) + c_2(-2, 0, -1, -2, 1),$$

where c_1, c_2 are arbitrary constants.
(iii) The nonhomogeneous system is consistent if and only if $3 = r(A) = r(A^+)$, or equivalently, $a = 7$. Then, giving the value $x_2 = 1$, and $x_5 = 0$ of the free variables, in the nonhomogeneous system, we get a particular solution $(3, 1, -10, 7, 0)$. Thus, a general solution of the nonhomogeneous part is

$$(3, 1, -10, 7, 0) + c_1(-2, 1, 0, 0, 0) + c_2(-2, 0, -1, -2, 1),$$

where c_1, c_2 are arbitrary constants.

Definition 2.4.20 A square matrix E obtained by applying elementary row operations on identity matrix is called an **elementary matrix**.

Example 2.4.21 The matrix

$$\begin{bmatrix} 1\,0\,0\,0 \\ 0\,1\,0\,0 \\ 0\,0\,3\,0 \\ 0\,0\,0\,1 \end{bmatrix}$$

is an elementary matrix which is obtained by multiplying the 3rd row of the identity matrix I_4 by 3. The matrix

$$\begin{bmatrix} 0\,0\,1\,0 \\ 0\,1\,0\,0 \\ 1\,0\,0\,0 \\ 0\,0\,0\,1 \end{bmatrix}$$

is an elementary matrix which is obtained by interchanging the 1st row and the 3rd row of the identity matrix I_4. Again, the matrix

$$\begin{bmatrix} 1\,0\,3\,0 \\ 0\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,1 \end{bmatrix}$$

is also an elementary matrix which is obtained by adding 3 times the 3rd row of the identity matrix I_4 to its 1st row.

τ_{ij} denotes the elementary matrix which is obtained by interchanging the i_{th} row and the j_{th} row of identity matrix. Thus,

$$\begin{bmatrix} 0\,0\,1\,0 \\ 0\,1\,0\,0 \\ 1\,0\,0\,0 \\ 0\,0\,0\,1 \end{bmatrix} = \tau_{13}.$$

The elementary matrix which is obtained by adding the λ times the j_{th} row of the identity matrix to its i_{th} row is denoted by E_{ij}^{λ}. Indeed, E_{ij}^{λ} is the matrix all of whose diagonal entries are 1, the i_{th} row j_{th} column entry is λ, and the rest of entries are 0. Thus,

$$\begin{bmatrix} 1\,0\,3\,0 \\ 0\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,1 \end{bmatrix} = E_{13}^{3}$$

The matrices E_{ij}^{λ} are called the **transvections**.

It can be easily observed that the effect of multiplying an elementary matrix E from left (right) to a matrix A is applying the elementary row (column) operation on A which was used to get the matrix E from the identity matrix. Thus, $\tau_{ij}A$ is the matrix obtained by interchanging i_{th} row and j_{th} row of A, and $E_{ij}^{\lambda}A$ is the matrix obtained by adding λ times the j_{th} row of A to its i_{th} row. It is straightforward, in particular, to verify the following relations, called the **Steinberg relations**, among the transvections in $M_n(F)$.

(i) $E_{ij}^{\lambda} \cdot E_{ij}^{\mu} = E_{ij}^{\lambda+\mu}$, $i \neq j$. In particular, $E_{ij}^{\lambda} \cdot E_{ij}^{-\lambda} = E_{ij}^{0} = I_n$. Thus, E_{ij}^{λ} is invertible, and its inverse is $E_{ij}^{-\lambda}$.

(ii) For $i \neq l, j \neq k$, E_{ij}^{λ} and E_{kl}^{μ} commute.

(iii) For $i \neq l$, $(E_{ij}^{\lambda}E_{jl}^{\mu}E_{ij}^{-\lambda}E_{jl}^{-\mu}) = E_{il}^{\lambda\mu}$.

(iv) For $j \neq k$, $(E_{ij}^{\lambda}E_{ki}^{\mu}E_{ij}^{-\lambda}E_{ki}^{-\mu}) = E_{jk}^{-\mu\lambda}$.

Proposition 2.4.22 *Let A be a $m \times n$ matrix. Then, we can find a nonsingular $m \times m$ matrix P such that PA is a matrix in reduced row echelon form. In particular, a square matrix A is nonsingular if and only if its reduced row echelon form PA is the identity matrix.*

Proof Applying an elementary row operation on A is equivalent to multiply A from left by an elementary matrix. Since every matrix can be reduced to a matrix in reduced row echelon form (Theorem 2.4.15), multiplying A successively by elementary matrices from left we arrive at matrix in reduced row echelon form. Since elementary matrices are nonsingular, and product of nonsingular matrices are nonsingular, we get a nonsingular matrix P such that PA is a matrix in reduced row echelon form. Since P is nonsingular, A is nonsingular if and only if PA is nonsingular. From the Proposition 2.4.14, A is nonsingular if and only if PA is the identity matrix. ♯

The above discussion and the results give an algorithm to determine a nonsingular matrix P such that PA is a reduced row echelon matrix. In particular, it gives an algorithm to check if a square matrix A is invertible, and if so, to find the inverse of A. We further illustrate the algorithm by means of examples.

Example 2.4.23 Consider the matrix

$$A = \begin{bmatrix} 0 & 0 & 3 & 1 & 2 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 2 & 1 & -1 & 1 \\ 0 & 1 & 1 & -\frac{1}{3} & 0 \end{bmatrix}.$$

Using the elementary row operations, we transform the matrix A in to a matrix in reduced row echelon form, and simultaneously find a nonsingular matrix P such that PA is a matrix in reduced row echelon form. We start with the pair

$$\left[I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, A = \begin{bmatrix} 0 & 0 & 3 & 1 & 2 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 2 & 1 & -1 & 1 \\ 0 & 1 & 1 & -\frac{1}{3} & 0 \end{bmatrix} \right].$$

There is no nonzero entry in the first column of A, and so no pivot will appear in the first column. We leave and move to the second column. The first nonzero entry in the second column of A is 1, and it is in the second row. We interchange the first row R_1 and the second row R_2 in the pair of matrices. The pair, thus, gets transformed to the pair $(E_1,\ A_1)$ given by

$$\left[E_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, A_1 = \begin{bmatrix} 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 3 & 1 & 2 \\ 0 & 2 & 1 & -1 & 1 \\ 0 & 1 & 1 & -\frac{1}{3} & 0 \end{bmatrix} \right]$$

(note that $E_1A = A_1$). The entry 1 in the first row and second column of A_1 is the pivot entry. To make the rest of the entries in this pivot column 0, we replace R_3 by $R_3 - 2R_1$, and then R_4 by $R_4 - R_1$. In turn, the pair $(E_1,\ A_1)$ gets transformed to the pair $(E_2,\ A_2)$ given by

$$\left[\begin{array}{l} E_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 3 & 1 & 2 \\ 0 & 0 & -3 & -1 & 1 \\ 0 & 0 & -1 & -\frac{1}{3} & 0 \end{bmatrix} \end{array}\right]$$

(Again note that $E_2A_1 = E_2E_1A = A_2$). The second row third column entry is 3 which is nonzero. We replace R_2 by $\frac{R_2}{3}$ to make it a pivot entry 1, and in turn, we replace R_1 by $R_1 - 2R_2$, R_3 by $R_3 + 3R_2$, and R_4 by $R_4 + R_2$ to make all the rest of the entries in this pivot column 0. Thus, the pair $(E_2,\ A_2)$ is transformed to the pair $(E_3,\ A_3)$ given by

$$\left[\begin{array}{l} E_3 = \begin{bmatrix} -\frac{2}{3} & 1 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ \frac{1}{3} & -1 & 0 & 1 \end{bmatrix}, A_3 = \begin{bmatrix} 0 & 1 & 0 & -\frac{2}{3} & -\frac{4}{3} \\ 0 & 0 & 1 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & \frac{2}{3} \end{bmatrix} \end{array}\right]$$

(Again, note that $E_3A_2 = A_3$). Since the 3rd row 4th column, and 4th row 4th column entries are 0, there is no pivot in the 4th column, it is a free column. We go to the 5th column. The 3rd row 5th column entry is 3 which is nonzero. We replace R_3 by $\frac{1}{3}R_3$ to make the 3rd row 5th column entry a pivot entry 1, and then replace R_1 by $R_1 + \frac{4}{3}R_3$, R_2 by $R_2 - \frac{2}{3}R_3$, and R_4 by $R_4 - \frac{2}{3}R_3$. Thus, the pair $(E_3,\ A_3)$ is transformed to the pair $(E_4,\ A_4)$ given by

$$\left[\begin{array}{l} E_4 = \begin{bmatrix} -\frac{2}{9} & \frac{1}{9} & \frac{4}{9} & 0 \\ \frac{1}{9} & \frac{4}{9} & -\frac{2}{9} & 0 \\ \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{9} & \frac{5}{9} & -\frac{2}{9} & 1 \end{bmatrix}, A_4 = \begin{bmatrix} 0 & 1 & 0 & -\frac{2}{3} & 0 \\ 0 & 0 & 1 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}\right],$$

where A_4 is in reduced row echelon form, and $P = E_4$ is an invertible matrix such that $PA = A_4$ is in reduced row echelon form.

Example 2.4.24 Consider the matrix A given by

$$\begin{bmatrix} 0 & 1 & 3 \\ 1 & 0 & 2 \\ 0 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

We apply the following elementary row operations in succession.

(i) Interchange R_1 and R_2,

(ii) replace R_4 by $R_4 - R_1$, R_3 by $R_3 - 2R_2$ and R_4 by $R_4 - R_2$,

(iii) replace R_3 by $-\frac{1}{3}R_3$, R_1 by $R_1 - 2R_3$, R_2 by $R_2 - 2R_3$ and R_4 by $R_4 + 3R_3$.

on A, and also on I_4. Then, A reduces to the reduced row echelon form

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

and I_4 reduces to

$$P = \begin{bmatrix} -\frac{4}{3} & 1 & \frac{2}{3} & 0 \\ -\frac{1}{3} & 0 & \frac{2}{3} & 0 \\ \frac{2}{3} & 0 & -\frac{1}{3} & 0 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

Thus, PA is in the row echelon form given above.

Example 2.4.25 Consider the 3×3 matrix A given by

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{bmatrix}$$

If we use the method of the above example, then A reduces to the identity matrix I_3, and I_3 reduces to

$$P = \begin{bmatrix} 3 & -\frac{5}{2} & \frac{1}{2} \\ -3 & 4 & -1 \\ 1 & -\frac{3}{2} & \frac{1}{2} \end{bmatrix}$$

Thus, A is invertible, and $PA = I_3$. Hence P is the inverse of A.

2.5 LU Factorization

If the coefficient matrix of a system of linear equations is upper triangular square matrix U with nonzero diagonal entries, then the solution is easily obtained by inspection. For example, if a system of linear equations is given by the matrix equation $U\overline{x}^t = \overline{b}^t$, where

$$U = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 3 \\ 0 & 0 & 9 \end{bmatrix},$$

and $\overline{b} = (b_1, b_2, b_3)$, then, evidently, the solution is $(18b_1 - \frac{b_2}{2} + \frac{4b_3}{9}, \frac{3b_2 - b_3}{2}, \frac{b_3}{9})$. Similarly, it is also easy to solve a system of linear equations whose coefficient matrix is lower triangular square matrix with nonzero diagonal entries. Further, suppose the coefficient matrix A is invertible, and it is expressed as $A = LU$, where L is a lower triangular matrix, and U is an upper triangular matrix. Then, we first find the solution $\overline{v}$ of $U\overline{y}^t = \overline{b}^t$, and then the solution $\overline{u}$ of $U\overline{x}^t = \overline{v}^t$. Clearly, $\overline{u}$ is the solution of $A\overline{x}^t = \overline{b}^t$.

The above discussion prompts us to look at the problem of factorizing an invertible matrix A as a product LU of a lower triangular matrix L and an upper triangular matrix U. This, in general, is not possible.

Example 2.5.1 Suppose that

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} a & 0 \\ b & c \end{bmatrix} \cdot \begin{bmatrix} u & v \\ 0 & w \end{bmatrix}.$$

Then $au = 0$, $av = 1$, $bu = 1$. This, however, is impossible. This shows that the invertible matrix

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

cannot be expressed as product of a lower triangular and an upper triangular matrix. Observe that the matrix

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 3 \\ 0 & 2 & 9 \end{bmatrix}$$

is also not expressible as product of a lower triangular and an upper triangular matrix.

The reason behind the impossibility of expressing the above matrices as product of lower and upper triangular matrices is while reducing these matrices in to reduced row echelon forms, we are forced either to interchange rows, or to add a nonzero multiple of a k_{th} row to l_{th} row for some $k > l$. Equivalently, we need to multiply from left by a corresponding elementary matrix τ_{ij}, or by a corresponding matrix E_{kl}^{λ}. Obviously, these matrices are not lower triangular matrices. Indeed, if, while reducing A in to reduced row echelon form, elementary row operations of the above type are not needed, then we can find a lower triangular matrix P with diagonal entries 1 so that PA is upper triangular. In turn, $A = LU$, where $L = P^{-1}$. We illustrate it by means of examples.

Example 2.5.2 Consider the matrix A given by

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{bmatrix},$$

and the system of linear equations given by the matrix equation

$$A\vec{x}^t = [1, 2, 3]^t.$$

Adding -1 times the 1st row of A to the 2nd row, and then adding -1 times the 1st row to the 3rd row, or equivalently, multiplying the matrix $E_{13}^{-1}E_{12}^{-1}$ to A from left, we obtain that

$$E_{13}^{-1}E_{12}^{-1}A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 3 & 8 \end{bmatrix}.$$

Again, adding -3 times the 2_{nd} row of the above matrix to its 3_{rd} row, or equivalently, multiplying E_{23}^{-3} to $E_{13}^{-1}E_{12}^{-1}A$ from left, we obtain that $E_{23}^{-3}E_{13}^{-1}E_{12}^{-1}A$ is the upper triangular matrix U given by

$$U = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix}.$$

Thus, $A = LU$, where $L = E_{12}^{1}E_{13}^{1}E_{23}^{3}$ is the lower triangular matrix given by

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

Now, to find solution of $A\overline{x}^t = [1, 2, 3]^t$, we first find the solution of $L\overline{y}^t = [1, 2, 3]^t$. Equating the corresponding entries of both sides, $y_1 = 1, y_1 + y_2 = 2$, *and* $y_1 + 3y_2 + y_3 = 3$. This gives the solution $[1, 1, -1]^t$ of $L\overline{y}^t = [1, 2, 3]^t$. Finally, we find the solution of $U\overline{x}^t = [1, 1, -1]^t$ to get the solution of the original equation $A\overline{x}^t = [1, 2, 3]^t$. Equating the entries of both sides in the equation $U\overline{x}^t = [1, 1, -1]^t$, we get that $2x_3 = -1, x_2 + 2x_3 = 1$, *and* $x_1 + x_2 + x_3 = 1$. Evidently, $x_3 = \frac{-1}{2}, x_2 = 2$, *and* $x_1 = \frac{-1}{2}$.

2.6 Equivalence of Matrices, Normal Form

Definition 2.6.1 Two $m \times n$ matrices A and B with entries in a field F are said to be equivalent if there exists a nonsingular $m \times m$ matrix P, and a nonsingular $n \times n$ matrix Q such that $A = PBQ$.

Clearly, the relation of being equivalent to is an equivalence relation on $M_{mn}(F)$. We determine a unique representative of each equivalence class of equivalent matrices.

Definition 2.6.2 A $m \times n$ matrix A is said to be in **normal form** if there is $r \leq min(m, n)$ such that

$$A = \begin{bmatrix} I_r & O_{r\,n-r} \\ O_{m-r\,r} & O_{m-r\,n-r} \end{bmatrix},$$

where $O_{m\,n}$ denote the zero $m \times n$ matrix.

Theorem 2.6.3 *Every $m \times n$ matrix is equivalent to a unique matrix in normal form.*

Proof Applying an elementary row operation on a matrix A is equivalent to multiply A from left by an elementary matrix, and applying an elementary column operation is equivalent to multiply matrix A from right by an elementary matrix. Since all elementary matrices are nonsingular, and product of nonsingular matrices are nonsingular, it is sufficient to show that every matrix can be reduced to a matrix in normal form with the help of elementary row, and elementary column operations. The proof of this fact is by the induction on $max(m, n)$, where m is the number of rows and n the number of columns. If $max(m, n) = 1$, then $m = 1 = n$, and $A = [a_{11}]$ is 1×1 matrix. If $A = [0]$, then it is already in normal form. If $a_{11} \neq 0$, then multiplying the row by a_{11}^{-1}, we reduce it to the normal form [1]. Assume that the result is true for all $r \times s$ matrices with $max(r, s) < max(m, n)$. Let $A = [a_{ij}]$ be a $m \times n$ times matrix. If $A = O_{m\,n}$, then it is already in normal form, and there is nothing to do. Suppose that $A \neq O_{m\,n}$. Suppose that $a_{kl} \neq 0$. Interchanging 1_{st} row and kth row, and then interchanging 1_{st} column and the lth column, we may suppose that $a_{11} \neq 0$, and then multiplying the 1_{st} row by a_{11}^{-1}, we may further suppose that $a_{11} = 1$. After this we add $-a_{1j}$ times the first column to the jth column, and then $-a_{i1}$ times the first row to the ith row for all $i \neq 1 \neq j$. This reduces the matrix A into the form

$$\begin{bmatrix} I_1 & O_{1\,n-1} \\ O_{m-1\,1} & B \end{bmatrix},$$

where B is $m - 1 \times n - 1$ matrix. This also gives us a nonsingular $m \times m$ matrix C, and a $n \times n$ nonsingular matrix D such that

$$CAD = \begin{bmatrix} I_1 & O_{1\,n-1} \\ O_{m-1\,1} & B \end{bmatrix}.$$

By the induction hypothesis there is a $m - 1 \times m - 1$ nonsingular matrix C', and there is a nonsingular $n - 1 \times n - 1$ matrix D' such that

$$C'BD' = \begin{bmatrix} I_{r-1} & O_{r-1\,n-r} \\ O_{m-r\,r-1} & O_{m-r\,n-r} \end{bmatrix}$$

Take

$$C'' = \begin{bmatrix} I_1 & O_{1\,n-1} \\ O_{m-1\,1} & C' \end{bmatrix},$$

and

$$D'' = \begin{bmatrix} I_1 & O_{1\,n-1} \\ O_{m-1\,1} & D' \end{bmatrix}.$$

Then C'' and D'' are nonsingular. In fact,

$$(C'')^{-1} = \begin{bmatrix} I_1 & O_{1\,n-1} \\ O_{m-1\,1} & (C')^{-1} \end{bmatrix}$$

(Use block multiplication to show this). Again, using block multiplication, we find that

$$C'' \cdot \begin{bmatrix} I_1 & O_{1\,n-1} \\ O_{m-1\,1} & B \end{bmatrix} \cdot D'' = \begin{bmatrix} I_1 & O_{1\,n-1} \\ O_{m-1\,1} & C'BD' \end{bmatrix} = \begin{bmatrix} I_r & O_{r\,n-r} \\ O_{m-r\,r} & O_{m-r\,n-r} \end{bmatrix}$$

Take $P = C \cdot C''$, and $Q = D \cdot D''$. Then P is nonsingular $m \times m$ matrix, and Q a nonsingular $n \times n$ matrix such that

$$PAQ = \begin{bmatrix} I_r & O_{r\,n-r} \\ O_{m-r\,r} & O_{m-r\,n-r} \end{bmatrix}$$

is in normal form. Finally,

$$\begin{bmatrix} I_r & O_{r\,n-r} \\ O_{m-r\,r} & O_{m-r\,n-r} \end{bmatrix}$$

is equivalent to

$$\begin{bmatrix} I_s & O_{s\,n-s} \\ O_{m-s\,s} & O_{m-s\,n-s} \end{bmatrix}$$

if and only if $r = s$, for one can be obtained from the other using elementary operations if and only if $r = s$. ♯

Corollary 2.6.4 *There are* $min(m, n) + 1$ *equivalence classes of equivalent matrices in* $M_{mn}(F)$.

Proof There are $min(m, n) + 1$ matrices in $M_{mn}(F)$ which are in normal form. ♯

Corollary 2.6.5 *Two matrices A and B are equivalent if and only if they have same rank.*

Proof Since under elementary operations rank of the matrices do not change and rank of the matrix

$$\begin{bmatrix} I_r & O_{r\,n-r} \\ O_{m-r\,r} & O_{m-r\,n-r} \end{bmatrix}$$

is r, the result follows. ♯

Corollary 2.6.6 *All nonsingular matrices in* $M_n(F)$ *are equivalent to* I_n. *The group* $GL(n, F)$ *is a single complete equivalence class of equivalent matrices.* ♯

Proof Let A be a $n \times n$ matrix which is nonsingular. Then there are nonsingular matrices P and Q such that PAQ is in normal form. Clearly, then PAQ is also nonsingular. The result follows if we observe that a matrix in normal form is nonsingular if and only if it is the identity matrix. ♯

Corollary 2.6.7 *The group* $GL(n, F)$ *is generated by elementary matrices. Indeed, every element of* $GL(n, F)$ *is product of elementary matrices.*

Proof All elementary matrices are nonsingular, and so they belong to $GL(n, F)$. Further, given any matrix $A \in GL(n, F)$, there are nonsingular matrices P and Q which are product of elementary matrices such that $PAQ = I_n$. But, then $A = P^{-1}Q^{-1}$. Since inverse of an elementary matrix is an elementary matrix, P^{-1} and Q^{-1} are product of elementary matrices. This shows that A is product of elementary matrices. ♯

Remark 2.6.8 The matrices $\{E_{ij}^{\lambda} \mid i \neq j, \lambda \in F^{\star}\}$ do not generate $GL(n, F)$ (verify).

Remark 2.6.9 The proof of the Theorem 2.6.3 gives us a method by which

(i) we can reduce a matrix A into normal form,
(ii) we can find nonsingular matrices P and Q such that PAQ is in normal form, and
(iii) we can determine whether A is nonsingular, and then we can find its inverse also.

Following two examples illustrates the algorithm.

Example 2.6.10 Let A be a $m \times n$ matrix. To find nonsingular matrices P and Q such that PAQ is in normal form, we proceed as follows: We start with a row with three columns. The first column I_m, the second A, and the third column I_n. Then we try to reduce the matrix A in to normal form by successive elementary row and elementary column operations. Whenever we perform a row operation on A, apply the same operation to the matrix in the first column, and keep the matrix in the third column as it is, and if we perform a column operation on A, then we perform the same operation on the matrix in the third column, and keep the matrix in the first column as it is. Then as the matrix A reduces to a matrix in normal form, the matrix in the first column reduces to the required matrix P, and the matrix in the third column reduces to the required matrix Q. Consider, for example, the matrix

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}.$$

Let R_i denote the ith row, and C_j denote the jth column. We start with a row

$$\left[\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right].$$

Replacing R_2 by $R_2 - 2R_1$, and R_3 by $R_3 - R_1$, we transform the above row to the row

$$\left[\begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & -2 & -1 \\ 0 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right].$$

Next, replacing C_2 by $C_2 - C_1$, and C_3 by $C_3 - C_1$, we get the transformed row as

$$\left[\begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & -1 \\ 0 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right].$$

Interchanging R_2 and R_4, and then replacing R_4 by $R_4 + 2R_2$, it reduces to

$$\left[\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ -2 & 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & -1 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right].$$

Replacing C_3 by $C_3 - 2C_2$, we transform it to

$$\left[\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ -2 & 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}\right].$$

Finally, replacing R_3 by $-R_3$, and then R_4 by $R_4 - 3R_3$, we transform it to

$$\left[\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ -5 & 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}\right].$$

Thus, A reduces to the normal form

$$\begin{bmatrix} I_3 \\ O_{1\,3} \end{bmatrix}.$$

Further, the required nonsingular matrices P and Q are given by

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ -5 & 1 & 3 & 2 \end{bmatrix},$$

and

$$Q = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}.$$

2.7 Congruent Reduction of Symmetric Matrices

Definition 2.7.1 A square matrix A is said to be **congruent** to a matrix B if there is an invertible matrix P such that $PAP^t = B$.

Observe that if A is symmetric, then PAP^t is also symmetric.

Theorem 2.7.2 *Every symmetric matrix A with entries in a field F of characteristic different from 2 is congruent to a diagonal matrix.*

Proof The proof is algorithmic. Let us recall that applying an elementary row operation on A is equivalent to multiply from left the corresponding elementary matrix E, and applying the same type of elementary column operation on A is equivalent to multiply the matrix A from right by the elementary matrix E^t (note that if we apply an elementary row operation on the identity matrix and take its transpose, then it is the same as apply the same elementary column operation on the identity matrix). Thus, it is sufficient to show that a symmetric matrix with entries in a field F of characteristic different from 2 can be reduced to a diagonal matrix by applying successively elementary row followed by the same type of elementary column operations. Let A be a symmetric matrix with entries in F, where characteristic of F if different from 2. If $A = 0$, then there is nothing to do. Suppose that $A \neq 0$. We may suppose that $a_{11} \neq 0$, for if not, suppose that $a_{ij} = a_{ji} \neq 0$, then adding the ith row to the first row, and then adding the ith column to the first column the first row first column entry becomes $2a_{ij} \neq 0$ (note that the characteristic $F \neq 2$). Then, for each $i \neq 1$, adding $-a_{i1}a_{11}^{-1}$ times the first row to the ith row, and $-a_{i1}a_{11}^{-1}$ times the first column to the ith column, we reduce the matrix to a symmetric matrix matrix in which all entries in the first row (and so also in the first column) except a_{11} is 0. Now, if $a_{ij} = 0$ for all $i, j \geq 2$, we have reduced it to a diagonal matrix. If not, using the previous argument, we may take $a_{22} \neq 0$, and then for $i \neq 2$ reduce all the entries $a_{i2} = a_{2i} = 0$. Proceeding inductively we reduce the matrix A to a diagonal matrix. ♯

Taking $Q = P^{-1}$, we get the following corollary.

Corollary 2.7.3 *Every symmetric matrix A with entries in a field of characteristic different from 2 can be decomposed as $A = QDQ^t$, where Q is an invertible matrix, and D is a diagonal matrix.* ♯

Remark 2.7.4 The theorem does not hold over a field of characteristic 2. Consider the matrix

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

Suppose that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} p & 0 \\ 0 & q \end{bmatrix}.$$

Equating the corresponding entries $p = ba + ab$, $q = dc + cd$, $da + cb = 0 = bc + ad$. Since the field is of the characteristic 2, $p = 0 = q$. In turn,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

But, then

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is singular.

We illustrate the algorithm of congruent reduction by means of an example.

Example 2.7.5 Let A be a symmetric $n \times n$ matrix. To find a nonsingular matrix P such that P^tAP is a diagonal matrix, we proceed as follows: We start with a row with 3 columns, the first column I_n, the second column A, and the third column I_n. We reduce the matrix A in to a diagonal form by successive elementary row and corresponding elementary column operations as described in the above theorem. Whenever we apply an elementary row operation on A, we apply the same operation on the matrix in the first column, and keep the matrix in third column as it is, and whenever we apply elementary column operation we apply the same operation on the matrix in the third column, and keep the first column as it is. In this process as soon as A reduces to a diagonal matrix, the first column reduces to P, and the third column, then will be P^t. Further, PAP^t is a diagonal matrix. Consider, for example, the matrix

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}$$

and the triple

$$\begin{bmatrix} I_3 & A & I_3 \end{bmatrix}$$

If we apply the following elementary operations

1. $R_1 \longrightarrow R_1 + R_2$,
2. $C_1 \longrightarrow C_1 + C_2$,
3. $R_2 \longrightarrow R_2 - \frac{1}{2}R_1$,
4. $C_2 \longrightarrow C_2 - \frac{1}{2}C_1$,
5. $R_3 \longrightarrow R_3 - \frac{3}{2}R_1$,
6. $C_3 \longrightarrow C_3 - \frac{3}{2}C_1$,

7. $R_3 \longrightarrow R_3 - R_2$,
8. $C_3 \longrightarrow C_3 - C_2$,

successively, on the triple

$$\left(I_3\ A\ I_3\right),$$

then the triple of matrices reduce to the triple

$$\left[\begin{bmatrix} 1 & 1 & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ -1 & -2 & -1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & -1 \\ 1 & \frac{1}{2} & -2 \\ 0 & 0 & 1 \end{bmatrix}\right].$$

Thus, A is congruent to $diag(2, -\frac{1}{2}, -4)$, and P is the matrix

$$\begin{bmatrix} 1 & 1 & -0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ -1 & -2 & 1 \end{bmatrix}.$$

Further, take $L = P^{-1}$ and $D = diag(2, -\frac{1}{2}, -4)$, then $A = LDL^t$. Note that L is not a lower triangular matrix. However, if we consider the matrix

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}$$

with the triple

$$\left(I_3\ A\ I_3\right)$$

of matrices and apply the following elementary operations on each member of the triple to reduce A to a diagonal matrix.

1. $R_2 \longrightarrow R_2 - R_1$ and $R_3 \longrightarrow R_3 - 2R_1$,
2. $C_2 \longrightarrow C_2 - C_1$ and $C_3 \longrightarrow C_3 - 2C_1$,
3. $R_3 \longrightarrow R_3 - R_2$,
4. $C_3 \longrightarrow C_3 - C_2$.

Then the triple of matrices reduce to the triple

$$\left(\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -3 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}\right)$$

Thus, A is congruent to $diag(1, -1, -3)$ and P is the matrix

$$\begin{bmatrix} 1 & 0 & -0 \\ -1 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix}$$

Further, take $L = P^{-1}$ and $D = diag(2, -\frac{1}{2}, -4)$, then $A = LDL^t$. Note that in this case P and L are lower triangular matrices.

Example 2.7.6 Consider the symmetric matrix

$$A = \begin{bmatrix} 3 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 3 \end{bmatrix}$$

with the triple

$$\begin{pmatrix} I_3 & A & I_3 \end{pmatrix}$$

of matrices and apply the following elementary operations on each member of the triple to reduce A to a diagonal matrix.

1. $R_3 \longrightarrow R_3 + \frac{1}{3}R_1$ and
2. $C_3 \longrightarrow C_3 + \frac{1}{3}C_1$.

Then the triple of matrices reduce to the triple

$$\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{3} & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{8}{3} \end{bmatrix} \begin{bmatrix} 1 & 0 & \frac{1}{3} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$$

Here again, P is a lower triangular matrix and the diagonal matrix D has all diagonal entries positive. As such, if we take $L = P^{-1}\sqrt{D}$, where $\sqrt{D} = Diag(\sqrt{3}, 1, \sqrt{\frac{8}{3}}$, then $A = LL^t$. Later we shall describe those symmetric matrices which can be expressed as LL^t, where L is a lower triangular matrix.

Exercises

2.7.1 Give two bases of the vector space $Mnm(F)$ of $n \times m$ matrices with entries in a field F over the field F.

2.7.2 Find a basis, and so also the dimension of the vector space $S_n(F)$ of $n \times n$ symmetric matrices with entries in a field F.

2.7.3 Let F be a field of characteristic different from 2. Find a basis, and so also the dimension of the vector space $SS_n(F)$ of $n \times n$ skew symmetric matrices with entries in a field F. Do the same for fields of characteristic 2. Are they same?

2.7.4 Let A be a $n \times m$ matrix. Consider the subset $W = \{B \in M_{mp} \mid AB = 0_{np}\}$ of M_{mp}. Show that W is a subspace of M_{mp}. Further, show that the dimension of W is $pn(A)$, where $n(A)$ denotes the nullity of A.

2.7.5 Show that every square matrix A with entries in a field F of characteristic different from 2 is uniquely expressible as sum of a symmetric matrix, and a skew symmetric matrix. Deduce that vector space $M_n(F)$ is direct $S_n(F) \oplus SS_n(F)$.
Hint. $A = \frac{A+A^t}{2} + \frac{A-A^t}{2}$.

2.7.6 Find a basis, and so also the dimension of the vector space $UT_n(F)$ of upper triangular matrices over F.

2.7.7 The sum of the diagonal entries of a square matrix A is called the **Trace** of A, and it is denoted by $Tr(A)$. Let $sl(n, F)$ denote the set of $n \times n$ matrices with trace 0. Show that $sl(n, F)$ is a vector space with respect the addition of matrices and multiplication by scalars. Find a basis of $sl(n, F)$, and so also its dimension.

2.7.8 Let A and B be square $n \times n$ matrices. Show that $Tr(AB - BA) = 0$. Deduce that $AB - BA$ is never identity matrix. Show by means of an example that it may be a nonsingular diagonal matrix.

2.7.9 Show by means of an example that AA^t need not be same as A^tA.

2.7.10 Consider the co-diagonal $n \times n$ matrix $\Gamma_n = [a_{ij}]$, where $a_{ij} = 1$ if $i + j = n + 1$, and $a_{ij} = 0$, otherwise. Show that Γ_n is symmetric and $\Gamma_n^2 = I_n$. What is the matrix $\Gamma_n A \Gamma_n$.

2.7.11 Describe all 2×2 matrices A such that $A^2 = 0_2$.

2.7.12 Let A be a strictly upper (lower) triangular $n \times n$ matrix. Show that $A^n = 0_n$.

2.7.13 Let A be a square $n \times n$ matrix which is nilpotent in the sense that $A^m = 0_n$ for some m. Show that $I_n + A$ is invertible. Show that

$$I_n + A + A^2 + \cdots + A^{m-1}$$

is the inverse of A. Is the converse of this statement true? Support.

2.7.14 Let $A = [a_{ij}]$ be a square $n \times n$ matrix which commutes with e_{12}. Show that $a_{12} = 0 = a_{21}$, and $a_{11} = a_{22}$. Show that a matrix commutes with all e_{ij} if and only if it is a scalar matrix. Show also that the matrices which commute with all transvections are precisely scalar matrices. Deduce that the center $Z(GL(n, F))$ is precisely $\{aI_n \mid a \in F^\star\}$.

2.7.15 Find a basis, and so also the dimension of the subspaces of $\mathbb{R}^4$ generated by the following subsets:

(i) $\{(1, 0, 2, 1), (2, 1, 3, 2), (7, 4, 9, 5), (1, 5, 6, 1)\}$,

(ii) $\{(1, 1, 1, 1), (1, 0, 2, 3), (1, 0, 4, 9), (1, 0, 8, 27)\}$.

2.7.16 Reduce the following matrices in to reduced row echelon form. Find the bases of their row spaces, column spaces, and Null spaces. Find their rank, and the nullities. Further, for each of the matrices A, find an invertible matrix P such that PA is a reduced row echelon form of A.

$$\begin{bmatrix} 0 & 0 & 3 & -3 & -3 \\ 2 & 4 & 3 & 3 & 1 \\ 2 & 4 & 3 & 3 & 3 \\ 1 & 2 & 2 & 1 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 1 & 0 & -1 & 0 \\ -5 & 1 & 3 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 7 & 11 \\ 3 & 7 & 14 & 25 \\ 4 & 11 & 25 & 50 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}$$

2.7.17 Check if the following systems of linear equations are consistent, and if so find their general solutions.

1.
$$\begin{aligned} x_1 + 3x_2 + 4x_3 &= 1. \\ 2x_1 - x_2 + x_3 &= 2. \\ 4x_1 + x_2 - x_3 &= 0. \\ 8x_1 - 3x_2 + x_3 &= 3. \end{aligned}$$

2.
$$\begin{aligned} x_1 + 2x_2 + x_3 + 2x_4 + x_5 &= 2. \\ 2x_1 + 4x_2 + 3x_3 + 3x_4 + x_5 &= 8. \\ 2x_1 + 4x_2 + 4x_3 + 2x_4 + 2x_5 &= 8. \\ x_1 + 2x_2 + 2x_3 + x_4 + 2x_5 &= 2. \end{aligned}$$

3.
$$\begin{aligned} 4x_1 - 15x_2 - 2x_3 - 32x_4 &= -40. \\ x_1 - 2x_2 - 3x_4 &= -4. \\ -3x_1 + 16x_2 + 3x_3 + 38x_4 &= 46. \\ x_1 - 6x_2 - x_3 - 14x_4 &= -17. \end{aligned}$$

2.7.18 Find the value of a, if possible, for which the following system of linear equations is consistent.

$$\begin{aligned} 4x_1 - 15x_2 - 2x_3 - 32x_4 &= -40. \\ x_1 - 2x_2 - 3x_4 &= -4. \\ -3x_1 + 16x_2 + 3x_3 + 38x_4 &= 46. \\ x_1 - 6x_2 - x_3 - 14x_4 &= a. \end{aligned}$$

2.7.19 Check if the matrices in exercise 16 have LU decompositions and if so find their LU decompositions.

2.7.20 Express each of the following symmetric matrices as PDP^t, where P is a nonsingular matrix, and D a diagonal matrix. Which of the matrices are expressible as LDL^t, where L is a lower triangular matrix. Also express them, if possible, as LL^t, where L is a lower triangular matrix.

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 3 \\ 1 & 2 & 1 & 3 \\ 1 & 1 & 3 & 2 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 6 & 7 & 8 \\ 3 & 7 & 11 & 12 \\ 4 & 8 & 12 & 16 \end{bmatrix}$$

2.7.21 Find the maximum number of arithmetic operations needed to reduce a 3×3 matrix into reduced row echelon form. Generalize it to $n \times n$ matrices.

2.7.22 Write a program in C-Language to check if a system of linear equations is consistent, and if so to find a general solution.

2.7.23 Write a program in C-Language to check if a matrix A admits LU decomposition, and if so to find it.

2.7.24 Write a program in C-Language to check if a symmetric matrix A admits LL^t decomposition, and if so to find it.

Chapter 3
Linear Transformations

This chapter centers around the study of linear transformations and their matrix representations.

3.1 Definition and Examples

Definition 3.1.1 Let V_1 and V_2 be vector spaces over a field F. A map T from V_1 to V_2 is called a **linear transformation** or a **homomorphism** if

$$T(ax + by) = aT(x) + bT(y)$$

for all $a, b \in F$, and $x, y \in V_1$.
A bijective linear transformation is called an **isomorphism**.

Proposition 3.1.2 *Let T be a linear transformation from a vector space V_1 to a vector space V_2. Then the following hold:*

(i) $T(0) = 0$ *and*
(ii) $T(-x) = -T(x)$ *for all* $x \in V_1$.
(iii) If W_1 *is a subspace of* V_1*, then* $T(W_1)$ *is a subspace of* V_2.
(iv) If W_2 *is a subspace of* V_2*, then the inverse image* $T^{-1}(W_2)$ *of* W_2 *under* T *is a subspace of* V_1.

Proof (i) Since T is a linear transformation, $T(0) = T(0 \cdot x + 0 \cdot y) = 0 \cdot T(x) + 0 \cdot T(y) = 0$.
(ii) $T(-x) = T(-1 \cdot x + 0 \cdot x) = (-1) \cdot T(x) + 0 \cdot T(x) = -T(x)$.
(iii) Let W_1 be a subspace of V_1. Then $0 \in W_1$, and so $0 = T(0) \in W_2$. Let $T(x), T(y) \in T(W_1)$, where $x, y \in W_1$. Since W_1 is a subspace,

R. Lal, *Algebra 2*, Infosys Science Foundation Series in Mathematical Sciences,
DOI 10.1007/978-981-10-4256-0_3

$ax + by \in W_1$ *for all* $a, b \in F$. It follows that $aT(x) + bT(y) = T(ax + by)$ belongs to $T(W_1)$. This shows that $T(W_1)$ is a subspace of V_2.
(iv) Since $0 = T(0) \in W_2$, it follows that $0 \in T^{-1}(W_2)$. Let $x, y \in T^{-1}(W_2)$. Then $T(x), T(y) \in W_2$. Since W_2 is a subspace, $T(ax + by) = aT(x) + bT(y) \in W_2$ *for all* $a, b \in F$. It follows that $ax + by \in T^{-1}(W_2)$. This shows that $T^{-1}(W_2)$ is a subspace of W_1. ♯

The inverse image $T^{-1}(\{0\})$ of the trivial subspace $\{0\}$ under a linear transformation T is a subspace, called the **null space** of T, and it is denoted by $N(T)$. $N(T)$ is also called the **kernel of T**, and then it is denoted by $ker\ T$.

Proposition 3.1.3 *A linear transformation T is injective if and only if* $N(T) = \{0\}$.

Proof Suppose that T is injective, and $x \in N(T)$. Then $T(x) = 0 = T(0)$. Since T is assumed to be injective, $x = 0$. Hence $N(T) = \{0\}$. Suppose, conversely, that $N(T) = \{0\}$. Suppose that $T(x) = T(y)$. Since T is a linear transformation, $T(x - y) = T(x) - T(y) = 0$. Hence $x - y \in N(T) = \{0\}$, and so $x = y$. Thus, T is injective. ♯

Example 3.1.4 Let $\overline{a} = (a_1, a_2, a_3)$ be a vector in the Euclidean vector space $\mathbb{R}^3$ over $\mathbb{R}$. Define a map T from $\mathbb{R}^3$ to itself by $T(\overline{r}) = \overline{r} \times \overline{a}$, where $\times$ is the vector product in $\mathbb{R}^3$. Then T is a linear transformation (follows from the property of vector product). The null space of T is given by $N(T) = \{\overline{x} \mid \overline{x} \times \overline{a} = \overline{0}\} = \{\alpha\overline{a} \mid \alpha \in \mathbb{R}\}$ provided that $\overline{a} \neq \overline{0}$. What is the image of T ?

Example 3.1.5 Let F be a field. Let $V = F^n$ and $W = F^m$ be standard vector spaces over the field F. Let T be a linear transformation from F^n to F^m. Let $\{\overline{e_1}, \overline{e_2}, \cdots, \overline{e_n}\}$ denote the standard basis of F^n. Let $T(\overline{e_i}) = \overline{r_i} = (a_{i1}, a_{i2}, \cdots, a_{im})$. Then T determines a matrix $A = M(T)$ whose i_{th} row is $\overline{r_i}$. It is evident that $T(\overline{x}) = \overline{x}M(T)$, where $\overline{x}$ is treated as a $1 \times n$ matrix. Thus, a linear transformation from F^n to F^m is precisely multiplication by a $n \times m$ matrix from right (note that elements of F^n are considered as $1 \times n$ matrices). The null space $N(T)$ is precisely the null space $N(M(T))$ of the corresponding matrix $M(T)$.

Example 3.1.6 Let V be a vector space over a field F, and W a subspace of V. Consider the quotient space V/W. The **quotient map** ν from V to V/W given by $\nu(x) = x + W$ is a linear transformation (follows from the definition of operations on V/W). Clearly, ν is surjective, and the null space $N(\nu)$ is given by $N(\nu) = \{x \in V \mid x + W = \nu(x) = W\}$ (note that the zero of V/W is the coset W). Since $x + W = W$ if and only if $x \in W$, it follows that $N(\nu) = W$. In turn, it also follows that every subspace of a vector space is null space of a linear transformation.

Example 3.1.7 Let $\wp_n$ denote the vector space of polynomials over the field $\mathbb{R}$ of real numbers of degrees at most n. Let D denote the derivative. Thus,

$$D(a_0 + a_1X + a_2X^2 + \cdots + a_nX^n) = a_1 + 2a_2X + 3a_3X^2 + \cdots + na_nX^{n-1}.$$

Then D is a linear transformation (verify). The null space of D is space of constant polynomials. Find its rank and nullity. Note that D is nilpotent. Indeed, D^{n+1} is the

zero linear transformation. Further, $I + D$ is an isomorphism. In fact, $I - D + D^2 - D^3 + \cdots + (-1)^n D^n$ is the inverse of $I + D$ (verify).

Example 3.1.8 Let $C^\infty(\mathbb{R})$ denote the vector space of real-valued functions on $\mathbb{R}$ which are r-times continuously differentiable functions for all r. The Differential operator $D^2 - 3D + 2$ from $C^\infty(R)$ to itself given by

$$(D^2 - 3D + 2)(f(X)) = \frac{d^2 f(X)}{dX^2} - 3\frac{df(X)}{dX} + 2f(X).$$

is a linear transformation. The null space of this differential operator is precisely $\{\alpha e^x + \beta e^{2x} \mid \alpha, \beta \in \mathbb{R}\}$ which is of dimension 2.

3.2 Isomorphism Theorems

Theorem 3.2.1 (Fundamental Theorem of Homomorphism). *Let T be a linear transformation from a vector space V over a field F to a vector space V' over the same field F. Let W be a subspace of V. Then there exists a linear transformation $\overline{T}$ from V/W to V' such that $\overline{T}ov = T$ if and only if $W \subseteq N(T)$. Also if such a linear transformation exists, it is unique. Further, then $\overline{T}$ will be injective if and only if $W = N(T)$. Finally, $\overline{T}$ is an isomorphism if and only if T is surjective and $W = N(T)$.*

Proof Suppose that there is a linear transformation $\overline{T}$ from V/W to V' such that $\overline{T}ov = T$. Let $x \in W$. Then $x + W = W$ the zero of V/W. Now,

$$T(x) = \overline{T}(v(x)) = \overline{T}(x + W) = \overline{T}(W) = 0,$$

for $\overline{T}$ is a linear transformation. Thus, $x \in N(T)$. This shows that $W \subseteq N(T)$. Conversely, suppose that $W \subseteq N(T)$ and $x + W = y + W$. Then $x - y \in W$. Since $W \subseteq N(T)$, $T(x - y) = 0$. Since T is a linear transformation, $T(x) = T(y)$. Thus, we have a map $\overline{T}$ from V/W to V' defined by $\overline{T}(x + W) = T(x)$. It is easily observed that $\overline{T}$ is a linear transformation such that $\overline{T}ov = T$. Further, if T' is a linear transformation such that $T'ov = T$. Then, $T'(x + W) = T'(v(x)) = (T'ov)(x) = T(x) = \overline{T}(x + W)$. This shows that $T' = \overline{T}$.

Next, suppose that such a $\overline{T}$ exists, and it is injective. Then already $W \subseteq N(T)$. Let $x \in N(T)$. Then $\overline{T}(x + W) = T(x) = 0 = \overline{T}(W)$. Since $\overline{T}$ is supposed to be injective, $x + W = W$, and so $x \in W$. This shows that $W = N(T)$.

Conversely, suppose that $N(T) = W$. Suppose that $\overline{T}(x + W) = \overline{T}(y + W)$. Then $T(x) = T(y)$. Hence $x - y \in N(T) = W$. This means that $x + W = y + W$. This shows that $\overline{T}$ is injective. Finally, since $\overline{T}ov = T$, $\overline{T}$ is surjective if and only if T is surjective. ♯

Corollary 3.2.2 *Let T from V to V' be a surjective linear transformation. Then $V/N(T) \approx V'$.* ♯

Theorem 3.2.3 (Noether Isomorphism Theorem). *Let V_1 and V_2 be vector subspaces of a vector space V over a field F. Then $V_1 + V_2/V_2$ is isomorphic to $V_1/V_1 \bigcap V_2$.*

Proof Define a map η from V_1 to $V_1 + V_2/V_2$ by $\eta(x) = x + V_2$. Clearly, η is a linear transformation. Any element of $V_1 + V_2/V_2$ is of the form $(x + y) + V_2$, where $x \in V_1$ and $y \in V_2$. But, then $(x + y) + V_2 = x + V_2 = \eta(x)$. This shows that η is surjective. Further,

$$N(\eta) = \{x \in V_1 \mid x + V_2 = V_2\} = \{x \in V_1 \mid x \in V_2\} = V_1 \bigcap V_2.$$

The result follows from the fundamental theorem of homomorphism. ♯

Proposition 3.2.4 *Let V and V' be vector spaces over a field F. Then,*

(i) a linear transformation T from V to V' is surjective if and only if it takes a set of generators to a set of generators.
(ii) a linear transformation T from V to V' is injective if and only if it takes a linearly independent set to a linearly independent set.
(iii) a linear transformation T from V to V' is an isomorphism if and only if it takes a basis to a basis.

Proof (i) Let T be a linear transformation. Since $T(< S >)$ is a subspace containing $T(S)$, it follows that $< T(S) > \subseteq T(< S >)$. Further, since image of linear combination of members S is a linear combination of members of $T(S)$, it follows that $T(< S >) \subseteq < T(S) >$. Thus $< T(S) > = T(< S >)$. The result follows.

(ii) Let T be an injective linear transformation from V to V'. Let S be a linearly independent subset of V. Let $y_1 = T(x_1), y_2 = T(x_2), \ldots, y_r = T(x_r)$ be distinct elements of $T(S)$. Since T is injective, $x_1, x_2, \ldots, x_r$ are distinct elements of S. Suppose that

$$\alpha_1 T(x_1) + \alpha_2 T(x_2) + \cdots + \alpha_r T(x_r) = 0.$$

Since T is a linear transformation,

$$T(\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_r x_r) = 0.$$

Since T is injective,

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_r x_r = 0.$$

Since S is linearly independent, $\alpha_i = 0$ *for all* i. This shows that $T(S)$ is linearly independent. Conversely, suppose that T takes a linearly independent subset to a linearly independent subset. Let $x \in V, x \neq 0$. Then $\{x\}$ is linearly independent, and so $\{T(x)\}$ is linearly independent. Thus, $T(x) \neq 0$. This shows that T is injective.

(iii). Follows from (i) and (ii). ♯

Corollary 3.2.5 *Let V be a finite dimensional vector space over a field F, and $S = \{x_1, x_2, \ldots, x_r\}$ an ordered basis of V. Then a linear transformation T from V to W is an isomorphism if and only if $\{T(x_1), T(x_2), \ldots, T(x_r)\}$ is an ordered basis of W.* ♯

Proposition 3.2.6 *Let V and W be vector spaces over a field F. Let S be a basis of V. Then any map f from S to W has a unique extension to a linear transformation T_f from V to W. More precisely, we have a bijective map η from the set $Hom_F(V, W)$ of all linear transformations from V to W to the set $Map(S, W)$ of all maps from S to W given by $\eta(T) = T/S$.*

Proof Let f be a map from S to W. Since S is a basis of V, every nonzero element $x \in V$ has a unique representation as

$$x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_r x_r,$$

where $x_1, x_2, \ldots, x_r$ are distinct elements of S, and all α_i are nonzero. Thus, we have a map T defined by

$$T(\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_r x_r) = \alpha_1 f(x_1) + \alpha_2 f(x_2) + \cdots + \alpha_r f(x_r).$$

Clearly, T extends f, and it is a linear transformation. If T' is also a linear transformation which extends f, then

$$\begin{aligned} T'(\alpha_1 x_1 + \cdots + \alpha_r x_r) &= \alpha_1 T'(x_1) + \cdots + \alpha_r T'(x_r) = \\ \alpha_1 f(x_1) + \cdots + \alpha_r f(x_r) &= T(\alpha_1 x_1 + \cdots + \alpha_r x_r). \end{aligned}$$

Hence $T = T'$. ♯

Remark 3.2.7 The above proposition says that two linear transformations are same if and only if they agree on a basis.

Corollary 3.2.8 *Any two finite dimensional vector spaces are isomorphic if and only if they are of same dimension. In particular, any n dimensional vector space over F is isomorphic to the standard vector space F^n.*

Proof Suppose that V and W are isomorphic, and T is an isomorphism from V to W. Then T, by the Corollary 3.2.5, takes an ordered basis to an ordered basis. Hence $dim V = dim W$. Conversely, suppose that $dim V = dim W$. Then there is a bijective map from a basis of V to a basis of W which can be extended to a linear transformation taking a basis to a basis. Thus, this extended linear transformation is an isomorphism. ♯

Corollary 3.2.9 *Let V and W be vector spaces of dimensions n and m, respectively, over a field F containing q elements. Then the number of linear transformations from V to W is $(q^m)^n$.*

Proof Since dimension of W is m, and F contains q elements, W contains q^m elements. A basis of V contains n elements. It follows from above results that there are as many linear transformation from V to W as many maps from a basis of V to W. ♯

Proposition 3.2.10 *Let V and W be finite-dimensional vector spaces of same dimension (in particular V may be same as W). Let T be a linear transformation from V to W. Then,*

(i) T is an isomorphism if and only if it is injective.
(ii) T is an isomorphism if and only if it is surjective.

Proof (i) If T is an isomorphism, then it is bijective, and so it is injective also. Conversely, suppose that T is injective. Let $\{x_1, x_2, \ldots, x_n\}$ be an ordered basis of V. Then it is linearly independent also. By the Proposition 3.2.4(ii), $\{T(x_1), T(x_2), \ldots, T(x_n)\}$ is an ordered linearly independent subset of W. Since $n = DimV = DimW$, $\{T(x_1), T(x_2), \ldots, T(x_n)\}$ is a basis of W. By the Proposition 3.2.4(iii), it follows that T is an isomorphism.

(ii) Again, if T is an isomorphism, then it is bijective, and so it is surjective. Suppose that T is surjective, and $\{x_1, x_2, \ldots, x_n\}$ is an ordered basis. Since T is surjective, $\{T(x_1), T(x_2), \ldots, T(x_n)\}$ is an ordered set of generators. Since $n = DimV = DimW$, $\{T(x_1), T(x_2), \ldots, T(x_n)\}$ is an ordered basis. By the Proposition 3.2.4(iii), T is an isomorphism. ♯

Let V be a vector space over a field F. An isomorphism from V to itself is called an automorphism of V. The set of all automorphisms of V is denoted by $GL(V)$. Some times it is also denoted by $Aut(V)$. $GL(V)$ is a group with respect to the composition of maps. This group is called the **general linear group** on V.

Proposition 3.2.11 *Let V be a vector space of dimension n over a field F, and $B(V)$ the set of all ordered bases of V. Let $\{x_1, x_2, \ldots, x_n\}$ be a fixed member of $B(V)$. Then we have a bijective map η from $GL(V)$ to $B(V)$ defined by*

$$\eta(T) = \{T(x_1), T(x_2), \ldots, T(x_n)\}.$$

Proof Since an isomorphism takes an ordered basis to an ordered basis, η is indeed a map from $GL(V)$ to $B(V)$. Since a linear transformation is uniquely determined by its effect on an ordered basis, η is injective. Also given an ordered basis $\{y_1, y_2, \ldots, y_n\}$ of V, the map $x_1 \rightsquigarrow y_1, x_2 \rightsquigarrow y_2, \ldots, x_n \rightsquigarrow y_n$ can be extended to automorphism T of V such that $\eta(T) = \{y_1, y_2, \ldots, y_n\}$. ♯

The following corollary follows from the above proposition and the Proposition 1.4.19.

Corollary 3.2.12 *Let V be a vector space of dimension n over a field F containing q elements. Then the group $GL(V)$ is finite of order*

$$(q^n - 1)(q^n - q) \cdots (q^n - q^{n-1}).$$ ♯

3.3 Space of Linear Transformations, Dual Spaces

Let V and W be vector spaces over a field F. Let $Hom_F(V, W)$ denote the set of all linear transformations from V to W. Let $f, g \in Hom_F(V, W)$. Define $f + g$ by $(f + g)(x) = f(x) + g(x)$. It can be easily verified that $f + g \in Hom_F(V, W)$. This defines an operation $+$ on $Hom_F(V, W)$. It is easily seen that $Hom_F(V, W)$ is an abelian group with respect to the addition $+$. Let $f \in Hom_F(V, W)$, and $\alpha \in F$. We define αf by $(\alpha f)(x) = \alpha \cdot f(x)$. Then, αf also belongs to $Hom_F(V, W)$. This defines a multiplication by scalars on $Hom_F(V, W)$. Indeed, $Hom_F(V, W)$ is a vector space over F under these operations (verify). In particular, $End(V)$ is also a vector space over F. In fact, $End(V)$ is an algebra, the internal multiplication being the composition of maps.

Theorem 3.3.1 *Let V and W be finite-dimensional vector spaces over a field F. Then*

$$dim Hom_F(V, W) = dim V \cdot dim W.$$

Proof Suppose that $dim V = n$ and $dim W = m$. Let $\{x_1, x_2, \ldots, x_n\}$ be an ordered basis of V, and $\{y_1, y_2, \ldots, y_m\}$ an ordered basis of W. Fix a pair (i, j), $1 \leq i \leq n$, $1 \leq j \leq m$. Since every map from a basis of a vector space to a vector space can be extended uniquely to a linear transformation, we have a unique $T_{ij} \in Hom_F(V, W)$ whose restriction to the basis $\{x_1, x_2, \ldots, x_n\}$ is given by $T_{ij}(x_i) = y_j$, and for $k \neq i$, $T_{ij}(x_k) = 0$. We show that $B = \{T_{ij} \mid 1 \leq i \leq n, 1 \leq j \leq m\}$ is a basis of $Hom_F(V, W)$. Let $T \in Hom_F(V, W)$. Then $T(x_i) \in W$. Since $\{y_1, y_2, \ldots y_m\}$ is a basis of W,

$$T(x_i) = \Sigma_{j=1}^{m} \alpha_{ji} y_j$$

for unique $\alpha_{ji} \in F$, $1 \leq j \leq m$, $1 \leq i \leq n$. It follows that the linear transformations T and $\Sigma_{i,j} \alpha_{ji} T_{ij}$ agree on each x_i, and so they agree on a basis. This means that $T = \Sigma_{i,j} \alpha_{ji} T_{ij}$. Thus, B is a set of generators for $Hom_F(V, W)$.

Next, we show that B is linearly independent. Suppose that

$$\Sigma_{i,j} \alpha_{ji} T_{ij} = 0.$$

Then

$$(\Sigma_{i,j} \alpha_{ji} T_{ij})(x_k) = 0 \, for \, all \, k.$$

Hence,

$$\Sigma_{i,j} \alpha_{ji} T_{ij}(x_k) = 0 \, for \, all \, k.$$

Thus,

$$\Sigma_{j=1}^{m}\alpha_{jk}y_j \;=\; 0 \,for\, all\; k.$$

Since $\{y_1, y_2, \ldots, y_m\}$ is linearly independent, $\alpha_{jk} \;=\; 0 \,for\, all\; j, k$. This shows that B is linearly independent, and so it is a basis of $Hom_F(V, W)$. Further, it follows that

$$dim Hom_F(V, W) \;=\; n \cdot m \;=\; dim V \cdot dim W. \qquad \sharp$$

Definition 3.3.2 Let V be a vector space over a field F. Treat F as a vector space over F. The members of $Hom_F(V, F)$ are called the **linear functionals** on V. The vector space $Hom_F(V, F)$, denoted by $V^\star$, is called the **dual space** of V.

If V is finite dimensional, then $dim V^\star \;=\; dim Hom(V, F) \;=\; dim V \cdot dim F \;=\; dim V$. Thus V and $V^\star$ have same dimensions, and so they are isomorphic as vector spaces. This is not true for infinite-dimensional spaces. For example, the vector space $\mathbb{R}$ of real numbers over $\mathbb{Q}$ is of infinite dimension, and it has a basis whose cardinality is the same as the cardinality of $\mathbb{R}$. Thus, the cardinality of $\mathbb{R}^\star$ is the same as that of the set $\mathbb{Q}^{\mathbb{R}}$ of all maps from $\mathbb{R}$ to $\mathbb{Q}$. Clearly, there is no bijective map from $\mathbb{R}$ to $\mathbb{Q}^{\mathbb{R}}$, and so $\mathbb{R}$ and $\mathbb{R}^\star$ are not isomorphic as vector spaces over $\mathbb{Q}$.

Definition 3.3.3 Let V and W be vector spaces over a field F. Let T be a linear transformation from V to W. Define a map T^t from $W^\star$ to $V^\star$ by $T^t(f) \;=\; foT$. Then T^t is a linear transformation(verify), and it is called the **Transpose** of T.

Proposition 3.3.4 *Let V_1, V_2 and V_3 be vector spaces over a field F. Let $T_1 : V_1 \longrightarrow V_2$ and $T_2 : V_2 \longrightarrow V_3$ be linear transformations. Then,*

$$(T_2 o T_1)^t \;=\; T_1^t o T_2^t.$$

Proof $(T_2 o T_1)^t$ is a linear transformation from $V_3^\star$ to $V_1^\star$ given by

$$(T_2 o T_1)^t(f) \;=\; fo(T_2 o T_1) \;= (f o T_2) o T_1 \;=\; T_2^t(f) o T_1 \;=\; T_1^t(T_2^t(f)) \;= \\ (T_1^t o T_2^t)(f)$$

for all $f \in V_3^\star$. $\sharp$.

Let V be a vector space over a field F. The dual $(V^\star)^\star$ of $V^\star$ is called the **double dual** of V, and it is denoted by $V^{\star\star}$. Let $x \in V$. Define a map $x^{\star\star}$ from $V^\star$ to F by

$$x^{\star\star}(f) \;=\; f(x).$$

It can be checked easily that $x^{\star\star}$ is a linear functional on $V^\star$. Thus $x^{\star\star} \in V^{\star\star}$. This gives us a map $x \rightsquigarrow x^{\star\star}$ from V to $V^{\star\star}$.

Proposition 3.3.5 *Let V be a vector space over a field F. Then the map $x \rightsquigarrow x^{\star\star}$ from V to $V^{\star\star}$ is an injective homomorphism. If V is finite dimensional, then it is also an isomorphism.*

Proof

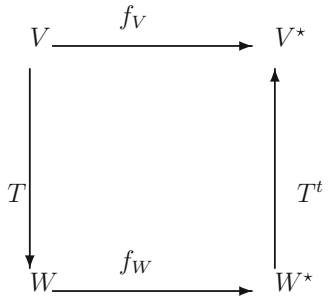
$$(\alpha x + \beta y)^{\star}(f) \;=\; f(\alpha x + \beta y) \;=\; \alpha f(x) + \beta f(y) \;=\; \alpha x^{\star\star}(f) + \beta y^{\star\star}(f)$$

for all $f \in V^{\star}$. Thus, $(\alpha x+\beta y)^{\star\star} = \alpha x^{\star\star}+\beta y^{\star\star}$, and so the map $x \rightsquigarrow x^{\star\star}$ is a linear transformation. To show that this is injective, it is sufficient to show that $x^{\star\star} = 0$ implies that $x = 0$. Suppose that $x \neq 0$. Then $\{x\}$, being linearly independent, can be enlarged to a basis of V. We have a map from this basis to F which is 1 on x and zero at all other members of the basis. This can be extended to a linear functional f of V which is 1 at x. Thus, $x^{\star\star}(f) = f(x) = 1 \neq 0$.

Finally, if V is finite dimensional, then $DimV = DimV^{\star} = DimV^{\star\star}$. Hence any injective linear transformation, in particular $x \rightsquigarrow x^{\star\star}$, is an isomorphism. ♯

Remark 3.3.6 A vector space is said to be **reflexive** if the map $x \rightsquigarrow x^{\star\star}$ is an isomorphism from V to $V^{\star\star}$. Thus, every finite-dimensional vector space is reflexive. Clearly, the vector space $\mathbb{R}$ over $\mathbb{Q}$ is not reflexive.

We know that every finite-dimensional vector space is isomorphic to its dual(being of same dimension). The question is whether we have a natural isomorphism. More precisely, do we have isomorphisms f_V from V to $V^{\star}$ for all vector spaces V such that given a linear transformation T from V to W the following diagram commutes.



The answer to this question is in negative. Suppose that we have a family,

$$\{f_V : V \longrightarrow V^{\star} \mid V \textit{ is a vector space over } F\}.$$

of isomorphisms. Let V be a one-dimensional vector space with $\{x\}$ as a basis, $x \neq 0$. Define a linear functional $x^{\star} \in V^{\star}$ by $x^{\star}(\alpha x) = \alpha$. Then $x^{\star} \neq 0$. Since $V^{\star}$ is also one dimensional, $\{x^{\star}\}$ is a basis of $V^{\star}$. Hence there is a $\lambda \in F$ such that $f_V(x) = \lambda x^{\star}$. Take a $\mu \in F$ such that $\mu^2 \neq 1$. Define a linear transformation T on V by $T(x) = \mu x$. Then,

$$(T^t o f_V o T)(x) = T^t(f_V(T(x))) = T^t(\lambda\mu x^\star) = \lambda\mu x^\star o T.$$

Also, $f_V(x) = \lambda x^\star$. Now

$$(\lambda\mu x^\star o T)(x) = \lambda\mu x^\star(T(x)) = \lambda\mu x^\star(\mu x) = \lambda\mu^2 x^\star(x) = \lambda\mu^2$$

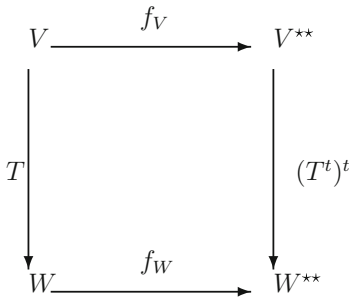
and $f_V(x)(x) = \lambda x^\star(x) = \lambda$. Since $\mu^2 \neq 1$, $f_V \neq T^t o f_V o T$. Thus, the above diagram is not commutative.

However, the following result says that every finite dimensional-vector space is naturally isomorphic to its double dual.

Proposition 3.3.7 *The family*

$$\{f_V : V \longrightarrow V^{\star\star} \mid V \text{ is finite dimensional}, \ f_V(x) = x^{\star\star}\}$$

defines natural isomorphisms from finite-dimensional vector spaces to its double duals in the sense that given any linear transformation $T : V \longrightarrow W$, the following diagram is commutative.



Proof We have already seen that f_V defined above is an isomorphism. Let $x \in V$. Then

$$(f_W o T)(x) = f_W(T(x)) = T(x)^{\star\star}.$$

Also,

$$((T^t)^t o f_V)(x) = (T^t)^t(x^{\star\star}) = x^{\star\star} o T^t.$$

Now,

$$T(x)^{\star\star}(g) = g(T(x))$$

for all $x \in V$. Further,

$$(x^{\star\star} o T^t)(g) = x^{\star\star}(T^t(g)) = x^{\star\star}(goT) = (goT)(x) = g(T(x))$$

for all $x \in V$. This shows that $T(x)^{\star\star} = x^{\star\star} o T^t$ for all $x \in V$. Hence $f_W o T = (T^t)^t o f_V$. ♯

Definition 3.3.8 Let V be a vector space over a field F. Let $\{e_1, e_2, \ldots, e_n\}$ be a basis of V. Consider F as a vector space over F with $\{1\}$ as a basis. Fix i, $1 \leq i \leq n$. Define a linear functional $e_i^\star$ on V by

$$e_i^\star(e_j) = \begin{cases} 1 \text{ if } j = i \\ 0 \text{ otherwise} \end{cases}$$

Then as in Theorem 3.3.1, $\{e_1^\star, e_2^\star, \ldots, e_n^\star\}$ is a basis of $V^\star$ called a **dual basis** which is dual to $\{e_1, e_2, \ldots, e_n\}$.

Let V and W be vector spaces over a field F. Let $\{x_1, x_2, \ldots, x_n\}$ be a basis of V, and $\{y_1, y_2, \ldots, y_m\}$ be a basis of W. Let $\{x_1^\star, x_2^\star, \ldots, x_n^\star\}$ and $\{y_1^\star, y_2^\star, \ldots, y_m^\star\}$ be corresponding dual bases. Let T be a linear transformation from V to W. Suppose that

$$T(x_i) = \Sigma_{j=1}^m \alpha_{ji} y_j.$$

Now, $T^t(y_k^\star) = y_k^\star o T$, and
$(y_k^\star o T)(x_i) = y_k^\star(T(x_i)) = y_k^\star(\Sigma_{j=1}^m \alpha_{ji} y_j)$
$= \Sigma_{j=1}^m \alpha_{ji} y_k^\star(y_j) = \alpha_{ki} = \alpha_{ki} x_i^\star(x_i)$.
Thus,

$$T^t(y_k^\star) = \Sigma_{i=1}^n \alpha_{ki} x_i^\star.$$

This gives us an expression for T^t in terms of dual bases provided we know the expression for T in terms of the given bases.

3.4 Rank and Nullity

Definition 3.4.1 Let V and W be vector spaces of finite dimensions over a field F. Let T be a linear transformation from V to W. The dimension of the image $T(V)$ is called the **rank** of T, and it is denoted by $r(T)$. The dimension of the null space $N(T)$ of T is called the **nullity** of T, and it is denoted by $n(T)$.

Thus, T is injective if and only if $n(T) = 0$, and T is surjective if and only if $r(T) = dim(W)$.

Theorem 3.4.2 *Let V and W be vector spaces of finite dimensions over a field F. Let T be a linear transformation from V to W. Then,*

$$r(T) + n(T) = dim(V).$$

Proof From the fundamental theorem of homomorphism $T(V)$ is isomorphic to $V/N(T)$. Also $dim(V/N(T)) = dim(V) - dimN(T)$. Hence

$$r(T) = dim(T(V)) = dimV/N(T) = dimV - dim(N(T)) = dimV - n(T).$$ ♯

Proposition 3.4.3 *Let $T_1 : V_1 \longrightarrow V_2$ and $T_2 : V_2 \longrightarrow V_3$ be linear transformations between finite-dimensional spaces over a field F. Then*

$$r(T_2oT_1) \leq min(r(T_2), r(T_1)).$$

Proof Since $T_2(T_1(V_1))$ is a subspace of $T_2(V_2)$,

$$r(T_2oT_1) = dim(T_2(T_1(V_1))) \leq dimT_2(V_2) = r(T_2).$$

Next, it follows from the above proposition that

$$dimT_2(T_1(V_1)) = dimT_1(V_1) - n(T_2/T_1(V_1)) \leq dimT_1(V_1) = r(T_1).$$ ♯

Corollary 3.4.4 *Under the hypothesis of the above proposition, if T_1 is an isomorphism, then $r(T_2oT_1) = r(T_2)$, and if T_2 is an isomorphism, then $r(T_2oT_1) = r(T_1)$.*

Proof Suppose that T_1 is an isomorphism. Then from the previous proposition, it follows that

$$r(T_2) = r(T_2oT_1oT_1^{-1}) \leq r(T_2oT_1) \leq r(T_2).$$

Thus, $r(T_2) = r(T_2oT_1)$. The rest of the assertion follows similarly. ♯

Corollary 3.4.5 *Let $T : V \longrightarrow W$ be a linear transformation between finite-dimensional vector spaces over a field F. Then*

$$r(T) = r((T^t)^t).$$

Proof From Proposition 3.3.7, we have $f_WoT = (T^t)^tof_V$. Since f_V and f_W are isomorphisms, from the previous proposition, it follows that

$$r(T) = r(f_WoT) = r((T^t)^tof_V) = r((T^t)^t).$$ ♯

Corollary 3.4.6 *Let $T : V \longrightarrow W$ be a linear transformation between finite-dimensional vector spaces over a field F. Then*

$$r(T) = r(T^t).$$

Proof Let $r = r(T)$. Let $\{y_1 = T(x_1), y_2 = T(x_2), \cdots, y_r = T(x_r)\}$ be a basis of $T(V)$. Enlarge the linearly independent subset $\{y_1, y_2, \ldots, y_r\}$ to a basis $\{y_1, y_2, \ldots, y_r, y_{r+1}, \ldots, y_m\}$ of W. Consider the corresponding dual basis $\{y_1^\star, y_2^\star, \ldots, y_r^\star, y_{r+1}^\star, \ldots, y_m^\star\}$ of W. Then $y_s^\star(y_i) = 0$ for all $i \leq r$ and $s \geq r+1$. This means that $y_s^\star(T(V)) = \{0\}$ for all $s \geq r+1$. Thus, $T^t(y_s^\star) = y_s^\star oT = 0$ for all $s \geq r+1$. This shows that $T^t(W)$ is generated by $\{T^t(y_1^\star), T^t(y_2^\star), \ldots, T^t(y_r^\star)\}$. It follows that $r(T^t) \leq r(T)$. In turn, $r((T^t)^t) \leq r(T^t) \leq r(T)$. Already by Corollary 3.4.5, $r(T) = r((T^t)^t)$. Hence $r(T) = r(T^t)$. ♯

3.5 Matrix Representations of Linear Transformations

Let V_1 and V_2 be vector spaces of dimensions m and n respectively over a field F. Let $\{x_1, x_2, \ldots x_m\}$ be a basis of V_1, and $\{y_1, y_2, \ldots, y_n\}$ a basis of V_2. We have a map $M_{x_1,x_2,\ldots,x_m}^{y_1,y_2,\ldots,y_n}$ from $Hom_F(V_1, V_2)$ to $M_{nm}(F)$ defined by

$$M_{x_1,x_2,\ldots,x_m}^{y_1,y_2,\ldots,y_n}(T) = [a_{ij}],$$

where

$$T(x_j) = \Sigma_{i=1}^n a_{ij} y_i.$$

This map $M_{x_1,x_2,\ldots,x_m}^{y_1,y_2,\ldots,y_n}$ is called the **matrix representation map** of linear transformations with respect to bases $\{x_1, x_2, \ldots, x_m\}$ and $\{y_1, y_2, \ldots y_n\}$ of V_1 and V_2, respectively.

Suppose that

$$M_{x_1,x_2,\ldots,x_m}^{y_1,y_2,\ldots,y_n}(T) = M_{x_1,x_2,\ldots,x_m}^{y_1,y_2,\ldots,y_n}(T') = [a_{ij}].$$

Then

$$T(x_j) = \Sigma_{i=1}^n a_{ij} y_i = T'(x_j)$$

for each j. But, then the effect of T and T' are same on the basis $\{x_1, x_2, \ldots, x_m\}$. This means that $T = T'$. Hence $M_{x_1,x_2,\ldots,x_m}^{y_1,y_2,\ldots,y_n}$ is an injective map. Further, given any $[a_{ij}] \in M_{nm}(F)$, there is a unique linear transformation T from V_1 to V_2 whose effect on the basis $\{x_1, x_2, \ldots, x_m\}$ is given by

$$T(x_j) = \Sigma_{i=1}^n a_{ij} y_i.$$

Clearly, $M_{x_1,x_2,\ldots,x_m}^{y_1,y_2,\ldots,y_n}(T) = [a_{ij}]$. Thus, $M_{x_1,x_2,\ldots,x_m}^{y_1,y_2,\ldots,y_n}$ is a bijective map. Let T_1, T_2 be members of $Hom_F(V_1, V_2)$, and $a, b \in F$. Suppose that $M_{x_1,x_2,\ldots,x_m}^{y_1,y_2,\cdots y_n}(T_1) = [a_{ij}]$ and $M_{x_1,x_2,\ldots,x_m}^{y_1,y_2,\ldots,y_n}(T_2) = [b_{ij}]$. Then $T_1(x_j) = \Sigma_{i=1}^n a_{ij} y_i$ and $T_2(x_j) = \Sigma_{i=1}^n b_{ij} y_i$. But, then

$$(aT_1 + bT_2)(x_j) = \Sigma_{i=1}^{n}(aa_{ij} + bb_{ij})y_i.$$

Hence

$$M_{x_1,x_2,\ldots,x_m}^{y_1,y_2,\ldots,y_n}(aT_1 + bT_2) = a[a_{ij}] + b[b_{ij}] = aM_{x_1,x_2,\ldots,x_m}^{y_1,y_2,\ldots,y_n}(T_1) + bM_{x_1,x_2,\ldots,x_m}^{y_1,y_2,\ldots,y_n}(T_2).$$

This proves the following proposition.

Proposition 3.5.1 *The matrix representation map $M_{x_1,x_2,\ldots,x_m}^{y_1,y_2,\ldots,y_n}$ from $Hom_F(V_1, V_2)$ to $M_{nm}(F)$ with respect to bases $\{x_1, x_2, \ldots, x_m\}$ of V_1 and $\{y_1, y_2, \ldots, y_n\}$ of V_2 is a vector space isomorphism.* ♯

Next, let V_1 be a vector space with a basis $\{x_1, x_2, \ldots, x_m\}$, V_2 a vector space with a basis $\{y_1, y_2, \ldots, y_n\}$, and V_3 a vector space with a basis $\{z_1, z_2, \ldots, z_p\}$ all over the same field F. Let $T_1 : V_1 \longrightarrow V_2$ and $T_2 : V_2 \longrightarrow V_3$ be linear transformations given by $T_1(x_j) = \Sigma_{i=1}^{n} a_{ij} y_i$ and $T_2(y_i) = \Sigma_{k=1}^{p} b_{ki} z_k$. Then

$$(T_2 o T_1)(x_j) = T_2(\Sigma_{i=1}^{n} a_{ij} y_i) = \Sigma_{i=1}^{n} a_{ij} T_2(y_i) = \Sigma_{i=1}^{n} a_{ij} \Sigma_{k=1}^{p} b_{ki} z_k = \Sigma_{k=1}^{p}(\Sigma_{i=1}^{n} b_{ki} a_{ij}) z_k = \Sigma_{k=1}^{p} c_{kj} z_k,$$

where $c_{kj} = \Sigma_{i=1}^{n} b_{ki} a_{ij}$. Thus,

$$M_{x_1,x_2,\ldots,x_m}^{z_i,z_2,\ldots,z_k}(T_2 o T_1) = [c_{ki}] = [b_{ki}][a_{ij}] = M_{y_1,y_2,\cdots,y_n}^{z_1,z_2,\ldots,z_k}(T_2) M_{x_1,x_2,\ldots,x_m}^{y_1,y_2,\ldots,y_n}(T_1).$$

This shows that the matrix representation map with respect to fixed choice of bases preserves product also. In particular, we have the following proposition.

Proposition 3.5.2 *Let V be a vector space of dimension n over a field F with a basis $\{x_1, x_2, \ldots, x_n\}$. Then $M_{x_1,x_2,\ldots,x_n}^{x_1,x_2,\ldots,x_n}$ is an isomorphism from the algebra $End_F(V)$ of endomorphisms of V to the algebra $M_n(F)$ of $n \times n$ matrices with entries in F.* ♯

Corollary 3.5.3 *Let $T \in End_F(V)$ and $\{x_1, x_2, \ldots, x_n\}$ a basis of V. Then $M_{x_1,x_2,\ldots,x_n}^{x_1,x_2,\ldots,x_n}$ induces an isomorphism from $GL(V)$ to $GL(n, F)$.*

Proof $GL(V)$ is the group of units of $End_F(V)$ and $GL(n, F)$ is that of $M_n(F)$. The result follows from the above proposition if we observe that an isomorphism between algebras induces isomorphisms between their group of units. ♯

The following corollary is consequence of the above corollary and the Corollary 3.2.12.

Corollary 3.5.4 *Let F_q denote a finite field with q elements. Then the order of the group (the number of nonsingular $n \times n$ matrices) $GL(n, F_q)$ is*

$$(q^n - 1)(q^n - q) \cdots (q^n - q^{n-1}).$$ ♯

Corollary 3.5.5 *Let V_1 and V_2 vector spaces over a field F of same dimension n. Let $\{x_1, x_2, \ldots, x_n\}$ be a basis of V_1 and $\{y_1, y_2, \ldots, y_n\}$ that of V_2. Then a linear transformation T from V_1 to V_2 is an isomorphism if and only if $M_{x_1,x_2,\ldots,x_n}^{y_1,y_2,\ldots,y_n}(T)$ is invertible $n \times n$ matrix. Also if T^{-1} exists, then*

$$(M_{x_1,x_2,\ldots,x_n}^{y_1,y_2,\cdots y_n}(T))^{-1} = M_{y_1,y_2,\ldots,y_n}^{x_1,x_2,\ldots,x_n}(T^{-1}).$$

Proof Clearly,

$$M_{x_1,x_2,\ldots,x_n}^{x_1,x_2,\ldots,x_n}(I_{V_1}) = I_n = M_{y_1,y_2,\ldots,y_n}^{y_1,y_2,\ldots,y_n}(I_{V_2}).$$

Further, since

$$M_{x_1,x_2,\ldots,x_n}^{x_1,x_2,\ldots,x_n}(T'oT) = M_{y_1,y_2,\ldots,y_n}^{x_1,x_2,\ldots,x_n}(T') \cdot M_{x_1,x_2,\ldots,x_n}^{y_1,y_2,\ldots,y_n}(T),$$

and

$$M_{y_1,y_2,\ldots,y_n}^{y_1,y_2,\ldots,y_n}(ToT') = M_{x_1,x_2,\ldots,x_n}^{y_1,y_2,\ldots,y_n}(T) \cdot M_{y_1,y_2,\ldots,y_n}^{x_1,x_2,\ldots,x_n}(T')$$

for all linear transformations T from V_1 to V_2, and all linear transformations T' from V_2 to V_1, the result follows. Evidently, if T^{-1} exists, then

$$(M_{x_1,x_2,\ldots,x_n}^{y_1,y_2,\cdots y_n}(T))^{-1} = M_{y_1,y_2,\ldots,y_n}^{x_1,x_2,\ldots,x_n}(T^{-1}). \quad \sharp$$

In particular, we have the following corollary.

Corollary 3.5.6 *Let V be a vector space of dimension n. Let $\{x_1, x_2, \ldots, x_n\}$ and $\{y_1, y_2, \ldots, y_n\}$ be bases of V. Then $M_{x_1,x_2,\ldots,x_n}^{y_1,y_2,\ldots,y_n}(I_V)$ is invertible, and its inverse is $M_{y_1,y_2,\ldots,y_n}^{x_1,x_2,\ldots,x_n}(I_V)$.* $\sharp$

The matrix $M_{x_1,x_2,\ldots,x_n}^{y_1,y_2,\ldots,y_n}(I_V)$ is called the **matrix of transformation** from the basis $\{x_1, x_2, \ldots, x_n\}$ to the basis $\{y_1, y_2, \ldots, y_n\}$. Thus, $M_{x_1,x_2,\ldots,x_n}^{y_1,y_2,\ldots,y_n}(I_V) = [a_{ij}]$, where $x_j = \Sigma_{i=1}^{n} a_{ij} y_i$.

Example 3.5.7 Define a map $T : \mathbb{R}^3 \longrightarrow R^2$ by $T((a, b, c)) = (a+b+c, a-b+c)$. Then T is a linear transformation (verify). Let $x_1 = (1, 1, 1)$, $x_2 = (1, 2, 1)$, $x_3 = (1, 2, 0)$, $y_1 = (1, 2)$, and $y_2 = (1, 1)$. Suppose that $a_1x_1 + a_2x_2 + a_3x_3 = (0, 0, 0)$. Then $a_1+a_2+a_3 = 0$, $a_1+2a_2+2a_3 = 0$, and $a_1+a_2 = 0$. Solving we get that $a_1 = a_2 = a_3 = 0$. This shows that $\{x_1, x_2, x_3\}$ is linearly independent. Since dimension of R^3 is 3, it follows that $\{x_1, x_2, x_3\}$ is a basis of R^3. Similarly, $\{y_1, y_2\}$ is a basis of $\mathbb{R}^2$. Now, suppose that $M_{x_1,x_2,x_3}^{y_1,y_2}(T) = [a_{ij}]$. Then $(3, 1) = T((1, 1, 1)) = T(x_1) = a_{11}y_1 + a_{21}y_2 = (a_{11} + a_{21}, 2a_{11} + a_{21})$. This shows

that $a_{11}+a_{21} = 3$ and $2a_{11}+a_{21} = 1$. Solving, we get that $a_{11} = -2$, $a_{21} = 5$. Similarly, looking at $T(x_2)$ and $T(x_3)$, we find that $a_{12} = -4$, $a_{22} = 8$, $a_{13} = -4$, and $a_{23} = 7$. Thus,

$$M_{x_1,x_2,x_3}^{y_1,y_2}(T) = \begin{bmatrix} -2 & -4 & -4 \\ 5 & 8 & 7 \end{bmatrix}.$$

3.6 Effect of Change of Bases on Matrix Representation

Proposition 3.6.1 *Matrix representations of a linear transformation with respect to different pair of bases are equivalent to each other. Conversely, if A and B are $m \times n$ matrices which are equivalent, then they represent same linear transformation with respect to a suitable pair of bases.*

Proof Let V_1 and V_2 be vector spaces over a field F of dimensions m and n, respectively. Let T from V_1 to V_2 be a linear transformation. Let $\{x_1, x_2, \ldots, x_m\}$ and $\{x_1', x_2', \ldots, x_m'\}$ be bases of V_1, and $\{y_1, y_2, \ldots, y_n\}$ and $\{y_1', y_2', \ldots, y_n'\}$ be those of V_2. Since $T = I_{V_2} o T o I_{V_1}$, and matrix representation preserves product, we have

$$M_{x_1,x_2,\ldots,x_m}^{y_1,y_2,\ldots,y_n}(T) = M_{y_1',y_2',\ldots,y_n'}^{y_1,y_2,\ldots,y_n}(I_{V_2}) M_{x_1',x_2',\ldots,x_m'}^{y_1',y_2',\ldots,y_n'}(T) M_{x_1,x_2,\ldots,x_m}^{x_1',x_2',\ldots,x_m'}(I_{V_1}).$$

By the Corollary 3.5.6, $M_{y_1',y_2',\ldots,y_n'}^{y_1,y_2,\cdots y_n}(I_{V_2})$ and $M_{x_1,x_2,\cdots x_m}^{x_1',x_2',\ldots,x_m'}(I_{V_1})$ are nonsingular. This shows that the two matrix representations are equivalent. Conversely, let $A = [a_{ij}]$ and $B = [b_{ij}]$ be $n \times m$ matrices which are equivalent. Let P be $n \times n$ nonsingular matrix, and Q a $m \times m$ nonsingular matrix such that $A = PBQ$. Let $\{x_1, x_2, \cdots x_m\}$ be a basis of V_1 and $\{y_1, y_2, \ldots, y_n\}$ a basis of V_2. Define T from V_1 to V_2 by

$$T(x_j) = \Sigma_{i=1}^{n} a_{ij} y_i.$$

Let $P = [\mu_{ki}]$ and $Q = [\nu_{lj}]$. Take $y_i' = \Sigma_{k=1}^{n} \mu_{ki} y_k$, and $x_j' = \Sigma_{l=1}^{m} \nu_{lj} x_l$. Since P and Q are invertible $\{y_1', y_2', \ldots, y_n'\}$ is a basis of V_2, and $\{x_1', x_2', \ldots, x_m'\}$ is a basis of V_1. Also $M_{y_1,y_2,\cdots y_n}^{y_1',y_2',\ldots,y_n'}(I_{V_2}) = P$ and $M_{x_1,x_2,\ldots,x_n}^{x_1',x_2',\ldots,x_m'}(I_{V_1}) = Q$. But, then $M_{x_1',x_2',\ldots,x_m'}^{y_1',y_2',\ldots,y_n'}(T) = PAQ = B$. ♯

Since every matrix is equivalent to a matrix in normal form, we have the following corollary.

Corollary 3.6.2 *Let T be a linear transformation from V_1 to V_2. Then there exists a basis $\{x_1, x_2, \ldots, x_m\}$ of V_1, $\{y_1, y_2, \ldots, y_n\}$ of V_2, and $r \leq (min(m, n)$ such that $T(x_i) = y_i$ for all $i \leq r$, and $T(x_i) = 0$ for all $i > r$.* ♯

Recall that two square $n \times n$ matrices A and B are said to be **similar** if there is a nonsingular $n \times n$ matrix P such that $PAP^{-1} = B$.

Corollary 3.6.3 *Let V be a vector space of dimension n over a field F. Let $\{x_1, x_2, \ldots, x_n\}$ and $\{x_1', x_2', \ldots, x_n'\}$ be bases of V. Let T be an endomorphism of V. Then $M^{x_1,x_2,\ldots,x_n}_{x_1,x_2,\ldots,x_n}(T)$ is similar to $M^{x_1',x_2',\ldots,x_n'}_{x_1',x_2',\ldots,x_n'}(T)$. Conversely, if A and B are $n \times n$ similar matrices, then there are bases $\{x_1, x_2, \ldots, x_n\}$ and $\{x_1', x_2', \ldots, x_n'\}$, and a linear transformation T such that $M^{x_1,x_2,\ldots,x_n}_{x_1,x_2,\ldots,x_n}(T) = A$ and $M^{x_1',x_2',\ldots,x_n'}_{x_1',x_2',\cdots x_n'}(T) = B$.*

Proof The result follows from the Corollary 3.5.5 (look at its proof), if we observe that the inverse of $M^{x_1',x_2',\ldots,x_n'}_{x_1,x_2,\ldots,x_n}(I_V)$ is $M^{x_1,x_2,\ldots,x_n}_{x_1',x_2',\ldots,x_n'}(I_V)$. ♯

Example 3.6.4 Consider the usual vector space $\mathbb{R}^3$ over $\mathbb{R}$. Consider the bases $\{x_1, x_2, x_3\}$ and $\{y_1, y_2, y_3\}$ of $\mathbb{R}^3$, where $x_1 = (1, 1, 1)$, $x_2 = (1, 2, 4)$, $x_3 = (1, 3, 9)$, $y_1 = (1, 2, 4)$, $y_2 = (1, 3, 9)$, and $y_3 = (1, 4, 16)$ (verify that these are bases). Let T be a linear transformation such that

$$M^{x_1,x_2,x_3}_{x_1,x_2,x_3}(T) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix}.$$

Suppose that $M^{y_1,y_2,y_3}_{x_1,x_2,x_3}(I_{R^3}) = [a_{ij}]$. Then, $x_1 = I_{\mathbb{R}^3}(x_1) = a_{11}y_1 + a_{21}y_2 + a_{31}y_3$. Hence $1 = a_{11} + a_{21} + a_{31}$, $1 = 2a_{11} + 3a_{21} + 4a_{31}$, $1 = 4a_{11} + 9a_{21} + 16a_{31}$. Solving $a_{11} = 3$, $a_{21} = -3$, $a_{31} = 1$. Similarly, looking at the representations of x_2 and x_3 in terms of y_1, y_2 and y_3, we find that $a_{12} = 1$, $a_{22} = 0 = a_{32} = a_{13} = a_{33}$, and $a_{23} = 1$. Thus,

$$M^{y_1,y_2,y_3}_{x_1,x_2,x_3}(I_{R^3}) = \begin{bmatrix} 3 & 1 & 0 \\ -3 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

Similarly,

$$M^{x_1,x_2,x_3}_{y_1,y_2,y_3}(I_{R^3}) = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & -3 \\ 0 & 1 & 3 \end{bmatrix}.$$

It follows that the above two matrices are similar. Further,

$$M^{y_1,y_2,y_3}_{y_1,y_2,y_3}(T) = M^{y_1,y_2,y_3}_{x_1,x_2,x_3}(I_{R^3}) M^{x_1,x_2,x_3}_{x_1,x_2,x_3}(T) M^{x_1,x_2,x_3}_{y_1,y_2,y_3}(I_{R^3}).$$

Substituting the values and multiplying we obtain that

$$M^{y_1,y_2,y_3}_{y_1,y_2,y_3}(T) = \begin{bmatrix} 5 & 1 & -9 \\ -2 & 3 & 13 \\ 1 & 0 & -2 \end{bmatrix}.$$

Proposition 3.6.5 *Let V_1 and V_2 be vector spaces over a field F. Let $\{x_1, x_2, \ldots, x_n\}$ be a basis of V_1 and $\{y_1, y_2, \ldots y_m\}$ a basis of V_2. Let $\{x_1^\star, x_2^\star, \ldots, x_n^\star\}$ and $\{y_1^\star, y_2^\star, \ldots, y_m^\star\}$ be corresponding dual bases of $V_1^\star$ and $V_2^\star$ respectively. Let T be a linear transformation from V_1 to V_2. Then*

$$(M_{x_1,x_2,\ldots,x_n}^{y_1,y_2,\ldots,y_m}(T))^t = M_{y_1^\star,y_2^\star,\ldots,y_m^\star}^{x_1^\star,x_2^\star,\ldots,x_n^\star}(T^t).$$

Proof Let

$$M_{x_1,x_2,\ldots,x_n}^{y_1,y_2,\ldots,y_m}(T) = [a_{ij}].$$

Then

$$T(x_j) = \Sigma_{i=1}^m a_{ij} y_i.$$

Suppose that

$$T^t(y_k^\star) = \Sigma_{j=1}^n b_{jk} x_j^\star.$$

By the definition $T^t(y_k^\star) = y_k^\star o T$. Hence

$$(y_k^\star o T)(x_l) = \Sigma_{j=1}^n b_{jk} x_j^\star(x_l)$$

i.e.,

$$y_k^\star(T(x_l)) = b_{lk}.$$

Now

$$y_k^\star(T(x_l)) = y_k^\star(\Sigma_{i=1}^m a_{il} y_i) = \Sigma_{i=1}^m a_{il} y_k^\star(y_i) = a_{kl}.$$

This shows that $[a_{ij}]^t = [b_{ji}]$. ♯

Exercises

3.6.1 Let F be a field. Show that the vector space F^n is isomorphic to the vector space F^m if and only if $n = m$.

3.6.2 Define a map T from $\mathbb{R}^3$ to $\mathbb{R}^3$ by $T((x, y, z)) = (x - y, y - z, z - x)$. Show that T is a linear transformation. Find its matrix representation with respect to the standard bases. Also find its matrix representation with respect to the basis $\{(1, 1, 1), (1, 2, 3), (1, 4, 9)\}$ of the domain, and the basis $\{(1, 1, 0), (0, 1, 1), (1, 0, 1)\}$ of the range. Show that $T(\mathbb{R}^3)$ is the subspace of $\mathbb{R}^3$ represented by the plane $x + y + z = 0$. What is $N(T)$? Find the rank and the nullity of T.

3.6.3 Let $\overline{a} = (a_1, a_2, a_3)$ be a fixed vector in $\mathbb{R}^3$. Define a map T from $\mathbb{R}^3$ to $\mathbb{R}$ by

$$T(\overline{r}) = \overline{r} \cdot \overline{a} \text{ (scalar product)}$$

Show that T is a linear transformation. Interpret the kernel of T if $\overline{a} \neq \overline{0}$. Find its matrix representation with respect to the standard bases. What is the rank, and what is the nullity of T.

3.6.4 Consider the subspace $W = \{(la, ma, na) \mid (l, m, n) \neq \overline{0} \; and \; a \in \mathbb{R}\}$ of $\mathbb{R}^3$. Show that the quotient space $\mathbb{R}^3/W$ is isomorphic to the subspace represented by the plane $lx + my + nz = 0$.

3.6.5 Let f be a linear functional on $\mathbb{R}^3$. Show that there is a vector $\overline{a}$ in $\mathbb{R}^3$ such that $f(\overline{r}) = \overline{r} \cdot \overline{a}$ (the scalar product).

3.6.6 Determine a linear transformation from $\mathbb{R}^3$ to $\mathbb{R}^3$ whose kernel is $lx + my + nz = 0$.

3.6.7 Fine the number of linear transformations on a vector space V of dimension n over a field F_q containing q elements.

3.6.8 Let V be a vector space of dimension n over a field F. Let $\{x_1, x_2, \ldots, x_n\}$ be a basis of V. Let T_1 and T_2 be linear transformations on V. Show that $T_1 o T_2 - T_2 o T_1$ is also a linear transformation on V. Show that $(x_i^{\star} o(T_1 o T_2 - T_2 o T_1))(x_i) = 0 for all \, i$. Deduce that $T_1 o T_2 - T_2 o T_1$ can never be the identity map.

3.6.9 Let T be a linear transformation on V. Let us call T a **nilpotent** endomorphism if $T^m = 0$ for some m. Suppose that T is nilpotent. Show that $I_V + T$ is an automorphism of V. Find the inverse of $I_V + T$ if $T^m = 0$.

3.6.10 Let $T \in End(V) = Hom_F(V, V)$. Let $f(X) \in F[X]$. Define an element $f(T) \in End(V)$ by

$$f(T) = a_0 I + a_1 T + \cdots + a_m T^m,$$

where $f(X) = a_0 + a_1 X + a_2 X^2 + \cdots a_m X^m$. Suppose that V is finite dimensional. Show that there is a nonzero polynomial $f(X) \in F[X]$ such that $f(T) = 0$.
Hint. If $dim V = n$, then $dim End V = n^2$, and so $\{I_V, T, T^2, \ldots, T^{n^2}\}$ is linearly dependent.

3.6.11 Show that $End(V)$ is a $F[X]$ module with respect to the external operation $\cdot$ given by $f(X) \cdot v = f(T)(v)$.

3.6.12 Let $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be a linear transformation which preserves angle between vectors in the sense that if P and Q are points in $\mathbb{R}^2$, then the angle between $\overline{OP}$ and $\overline{OQ}$, where O is origin, is the same as the angle between $\overline{OT(P)}$ and $\overline{OT(Q)}$. Show

that either T is a reflection about a line passing through origin, or it is a rotation in the plane through an angle α
Hint. Suppose that there is a point P different from the origin such that $T(P) = P$. Then show that $T = I$, or it is a reflection about the line passing through O and P. Next, if T fixes no point other that O, then show that $T((1, 0)) = (cos\alpha, sin\alpha)$ for some α, and then show that $T((x, y)) = (xcos\alpha + ysin\alpha, -xsin\alpha + ycos\alpha)$.

3.6.13 Show that any angle preserving endomorphism of $\mathbb{R}^3$ is either a rotation about a fixed axis, or a reflection about a plane passing through origin.

3.6.14 Let $\wp_n$ denote the vector space of polynomials over the field $\mathbb{R}$ real numbers of degrees at most n. Define a map T from $\wp_n$ to itself by

$$T(a_0 + a_1X + a_2X^2 + \cdots + a_nX^n) = a_1 + 2a_2X + 3a_3X^2 + \cdots + na_nX^{n-1}.$$

Show that T is a linear transformation. Find its rank and nullity. Is T invertible?

3.6.15 Let $C^\infty(\mathbb{R})$ denote the vector space of real-valued functions on $\mathbb{R}$ which are r-times continuously differentiable functions for all r. Define a linear transformation $D^2 - 2D + 1$ from $C^\infty(R)$ to itself by

$$(D^2 - 2D + 1)(f(X)) = \frac{d^2f(X)}{dX^2} - 2\frac{df(X)}{dX} + f(X).$$

Find the nullity of $D^2 - 2D + 1$, and also a basis of the kernel.

3.6.16 Let V be a vector space of dimension m over a field F_q of order q, and W a vector space of dimension n over F_q. Suppose that $m \leq n$. Find the number of injective linear transformations from V to W.

3.6.17 Suppose that $m \geq n$ in the above exercise. Find the number of surjective linear transformations from V to W.

3.6.18 Let V be a vector space of dimension n over a field F. Let $\{e_1, e_2, \ldots, e_n\}$ be an ordered basis of V. Let p be a permutation in S_n. Then we have a map T_p from the ordered basis $\{e_1, e_2, \ldots, e_n\}$ to V given by $T_p(e_i) = e_{p(i)}$. Show that $p \rightsquigarrow T_p$ defines an injective homomorphism from S_n to the group $GL(V)$ of all automorphisms of V. Deduce that every group of order n is isomorphic to a subgroup of $GL(V)$.

3.6.19 Let V be a finite-dimensional vector space over a field F. Let $T, S \in End(V)$ such that $SoT = I_V$ ($ToS = I_v$). Show that $ToS = I_V$ ($SoT = I_V$).

3.6.20 Show by means of an example that the above result is not true for infinite-dimensional spaces.
Hint. Let V be the vector space of all real-valued continuous functions on $[1, \infty)$ over the field $\mathbb{R}$ of real numbers. Consider the map T given by

$$T(f)(x) = \int_1^x f(y)dy.$$

Use the fundamental theorem of integral calculus.

3.6.21 Let V be a vector space of dimension n over a field F, and T be an element of the center of End_V. Then $T = \alpha I_V$ for some $\alpha \in F$.

Proof Let $x \in V, x \neq 0$. We show that there is a $\lambda_x \in F$ such that $T(x) = \lambda_x x$. Suppose not. Then $\{x, T(x)\}$ is linearly independent, and so it can be embedded in a basis $\{x, T(x), x_3, \cdots x_n\}$ of V. Define a linear transformation S by $S(x) = x, S(T(x)) = 0$, and $S(x_i) = 0$ *for all* $i \geq 3$. Then $ST(x) = 0$, where as $TS(x) = T(x) \neq 0$. Thus, *for all* $x \in V$ *there exists a* $\lambda_x \in F$ such that $T(x) = \lambda_x x$. Now, we show that $\lambda_x = \lambda_y, x \neq 0 \neq y$. Suppose that $\lambda_x \neq \lambda_y$. Then $\lambda_{x-y}(x-y) = T(x-y) = \lambda_x x - \lambda_y y$. But, then $(\lambda_x - \lambda_{x-y})x = (\lambda_y - \lambda_{x-y})y$. Since $\lambda_x \neq \lambda_y$, $\{x, y\}$ is linearly independent. Hence $\lambda_x = \lambda_{x-y} = \lambda_y$. This shows that $T = \lambda I_V$ for some λ. ♯

3.6.22 Let V be a vector space of dimension n over a field F. Let T be a linear transformation on V which commutes with each element of the group $GL(V)$. Then $T = \alpha I_V$ for some $\alpha \in F$. In particular, $Z(GL(V)) = \{\alpha I_V \mid \alpha \in F^\star\}$.

Proof We again show that for each $x \in V$, there is an element $\lambda_x \in F$ such that $T(x) = \lambda_x x$. The rest will follow as in the above exercise. Suppose that there is a $x \in V, x \neq 0$ for which there is no $\lambda \in F$ such that $T(x) = \lambda x$. Then $\{x, T(x)\}$ is linearly independent, and so it can be embedded in a basis $\{x, T(x), x_3, \ldots x_n\}$ of V. Let S be a linear transformation defined by $S(x) = x$, $S(T(x)) = -T(x)$, and $S(x_i) = x_i$ *for all* i. Then, since S takes a basis to a basis, it is an element of $GL(V)$. Further, $TS(x) = T(x)$ where as $ST(x) = -T(x)$. Since $T(x) \neq 0$, $TS \neq ST$. ♯

3.6.23 Let V be a finite -dimensional vector space over a field F. Let Γ be a nontrivial subspace of $End_F(V)$ such that ToS and SoT belong to Γ for all $S \in End(V)$ and $T \in \Gamma$. Then $\Gamma = End(V)$ (in the language of ring theory, this is expressed by saying that the ring $End(V)$ has no proper two-sided ideals).

Proof Let T be a nonzero linear transformation in Γ. Let $\{x_1, x_2, \ldots, x_n\}$ be a basis of V. Since $T \neq 0$, $T(x_i) \neq 0$ for some i. Without any loss of generality, we may assume that $T(x_1) \neq 0$. Take any i. There is a linear transformation S such that $S(T(x_1)) = x_i$ (embedded $\{T(x_1)\}$ in to a basis). There is also a linear transformation S' such that $S'(x_1) = x_1$ and $S'(x_j) = 0$ *for all* $j \geq 1$. Thus, $STS'(x_1) = x_i$, and $STS'(x_j) = 0$ *for all* $j \geq 2$. Hence $STS' = T_{1i}$. By the hypothesis, $T_{1i} \in \Gamma$. Also $T_{ji} = T_{1i}T_{j1} \in \Gamma$ *for all* j. Thus, Γ is a subspace of $End_F(V)$ containing all T_{ji}. Since $\{T_{ji}, 1 \leq i \leq n, 1 \leq j \leq n\}$ is a basis of the vector space $End_F(V), \Gamma = End_F(V)$. ♯

3.6.24 Let V be a finite-dimensional vector space over F. Show that $GL(V)$ generates $End_F(V)$ as a vector space.

3.6.25 Let $T \in End_F(V)$ be such that $T^2 = T$, where V is finite dimensional (such a transformation is called idempotent). Show that every element x of V can be uniquely expressed as $x = y + z$, where $T(y) = 0$ and $T(z) = z$.

3.6.26 Let $T \in End_F(V)$ be nilpotent, and $f(X) \in F[X]$. Show that $f(T)$ is an automorphism of V if and only if $f(X)$ has a nonzero constant term.

3.6.27 Let $T \in End_F(V)$, and $dim(V) = n$. Show that there is a monic polynomial $f(X)$ such that $f(T) = 0$. Smallest degree such a monic polynomial is called the **minimum polynomial** of T. Show that T is an isomorphism if and only if the minimum polynomial of T has nonzero constant term. Deduce that $T^{-1} = g(T)$ for some polynomial $g(X)$.

3.6.28 Show that $g(T) = 0$ if and only if the minimum polynomial of T divides $g(X)$.

3.6.29 Let V be a finite-dimensional vector space over a field F. Let T be a nonzero element of $End_F(V)$. Show that there is a $S \in End_F(V)$ such that $ST \neq 0$ and $(ST)^2 = ST$.

3.6.30 Let $T \in End_F(V) - GL(V) - \{0\}$. Show that there is a $S \in End_F(V)$ such that $ST = 0$ but $TS \neq 0$.

3.6.31 Find the minimum polynomial of the linear transformation in Exercise 3.6.2.

3.6.32 Show that $Z(GL(V))$ is isomorphic to the multiplicative group $F^\star$.

3.6.33 Find the order of a Sylow p-subgroup of $GL(V)$, where V is a vector space of dimension n over $\mathbb{Z}_p$. Find also a Sylow p-subgroup.

3.6.34 Let V and W be vector spaces of dimensions n and m, respectively, over a finite field F_q of order q. Let $r \leq min(n, m)$. Find the number of linear transformations of rank r.

3.6.35 Let T_1 and T_2 be endomorphism of a vector space V of dimension n over a field F. Show that

$$r(T_1 o T_2) \geq r(T_1) + r(T_2) - n.$$

3.6.36 Let V be a vector space of dimension 2 over $\mathbb{Z}_2$. Show that $GL(V)$ is isomorphic to S_3.

3.6.37 Let $T \in End_F(V)$, where V is a vector space of dimension n. Suppose that $T^r = 0$. Show that $T^n = 0$.

3.6.38 Show that every linear functional f on V defines a linear transformation T_f from $End_F(V)$ to $V^\star$ by $T_f(T) = foT$. Let V be a vector space over F, and $\{x_1, x_2, \ldots, x_n\}$ be a basis of V. Consider $p = p_1 + p_2 + \cdots + p_n$, where p_i is

the ith projection linear functional on V with respect to the given basis. Show that p is independent of the choice of basis. Show, further, that the linear functional Tr on $End_F(V)$ defined by $Tr(T) = p_1(T(x_1)) + p_2(T(x_2)) + \cdots + p_n(T(x_n))$ is also independent of the choice of basis of V. This is called the **trace form** on $End_F(V)$. Show that $TS - ST \in KerTr$ for all $S, T \in End_F(V)$. Does $Tr(T') = o$ implies that $T' = TS - ST$ for some $S, T \in End_F(V)$?

3.6.39 Let f be a linear functional on $End_F(V)$, where V is as in the above exercise. Suppose that $f(AB - BA) = 0$ for all $A, B \in End_F(V)$, and $f(I_V) = n$. Show that $f = Tr$.

3.6.40 Show that $AB - BA$ can never be an automorphism of V, where V is finite dimensional. Deduce that $I_V + AB - BA$ can never be nilpotent.

3.6.41 Show that the subgroup of the additive group $End_F(V)$ generated by $GL(V)$ is $End_F(V)$.

3.6.42 Let T be a linear transformation from V to V. Show that the following conditions are equivalent:

(i) $N(T) \bigcap imageT = \{0\}$
(ii) $T^2(x) = 0\, implies\, that\, T(x) = 0.$

3.6.43 Let T be a linear transformation on V such that the rank of T is same as rank of T^2. Show that $KerT \bigcap imageT = \{0\}$.

Chapter 4
Inner Product Spaces

In the vector space theory, we have talked about points, lines, and planes as translates of subspaces of a vector space. In this chapter, we shall talk about the concepts of angle between lines (planes), distance between points, shortest distances between planes, area, volumes of parallelepiped, etc. We also discuss rigid motions in an inner product space. For this purpose, we enrich the structure of vector space by putting the concept of inner product. We have to consider vector spaces over particular types of fields. All fields F in this chapter will either be the field $\mathbb{R}$ of real numbers or the field $\mathbb{C}$ of complex numbers. We have a field automorphism $\alpha \mapsto \overline{\alpha}$ from $\mathbb{C}$ to itself called the complex conjugation. The restriction of the complex conjugation to $\mathbb{R}$ is the identity map.

4.1 Definition, Examples, and Basic Properties

Definition 4.1.1 Let F be the field $\mathbb{R}$ of real numbers or the field $\mathbb{C}$ of complex numbers, and V a vector space over F. A map $<>$ from $V \times V$ to F (the image of (x, y) under $<>$ is denoted by $< x, y >$) is called an **inner product** (**real** if $F = \mathbb{R}$, and **complex inner product** if $F = \mathbb{C}$) on V if the following hold:

1. $< \alpha x + \beta y, z > = \alpha < x, z > + \beta < y, z >$ *for all* $\alpha, \beta \in F$, and $x, y \in V$.
2. $< x, y > = \overline{< y, x >}$ *for all* $x, y \in V$.
   In particular, $< x, x > = \overline{< x, x >}$ *for all* $x \in V$, and so $< x, x >$ is a real number for all $x \in V$.
3. $< x, x > \geq 0$, and $< x, x > = 0$ *if and only if* $x = 0$.

A vector space V together with an inner product $<>$ on V is called an **inner product space**.

R. Lal, *Algebra 2*, Infosys Science Foundation Series in Mathematical Sciences,
DOI 10.1007/978-981-10-4256-0_4

Putting $\alpha = 1 = \beta$ in 1, we obtain that

$$< x + y, z > = < x, z > + < y, z >,$$

and putting $\beta = 0$ in 1, we obtain that

$$< \alpha x, y > = \alpha < x, y > .$$

Next, using 2 and 1, we get

$$< x, \alpha y + \beta z > = \overline{< \alpha y + \beta z, x >} = \overline{\alpha < y, x > + \beta < z, x >}$$
$$= \overline{\alpha} \overline{< y, x >} + \overline{\beta} \overline{< z, x >} = \overline{\alpha} < x, y > + \overline{\beta} < x, z > .$$

Thus,

$$< x, \alpha y + \beta z > = \overline{\alpha} < x, y > + \overline{\beta} < x, z >$$

for all $x, y, z \in V$, and $\alpha, \beta \in F$.

Putting $\alpha = 1 = \beta$ in the above equation, we obtain

$$< x, y + z > = < x, y > + < x, z >,$$

and putting $\beta = 0$, we get

$$< x, \alpha y > = \overline{\alpha} < x, y > .$$

Further,

$$< 0, y > = < 0 \cdot 0, y > = 0 \cdot < 0, y > = 0,$$

and similarly,

$$< x, 0 > = \overline{< 0, x >} = 0$$

for all $x, y \in V$.

Example 4.1.2 Let $V = \mathbb{R}^n$ be the Euclidean vector space of dimension n over $\mathbb{R}$. Define

$$< \overline{x}, \overline{y} > = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n,$$

where $\overline{x} = (x_1, x_2, \ldots, x_n)$ and $\overline{y} = (y_1, y_2, \ldots, y_n)$. This gives an inner product on $\mathbb{R}^n$ (verify). This inner product is called the usual **standard Euclidean** inner product on $\mathbb{R}^n$.

Example 4.1.3 We have another inner product $< >$ on $\mathbb{R}^3$ given by

$$< \overline{x}, \overline{y} > = x_1 y_1 + x_2 y_2 + 2x_3 y_3 + x_2 y_3 + x_3 y_2,$$

where $\overline{x} = (x_1, x_2, x_3)$, and $\overline{y} = (y_1, y_2, y_3)$ (verify).

Example 4.1.4 Let $V = \mathbb{C}^n$ the complex vector space of dimension n. Define $< >$ on $\mathbb{C}$ by

$$< \overline{x},\ \overline{y} > = x_1\overline{y_1} + x_2\overline{y_2} + \cdots + x_n\overline{y_n}.$$

Then it is a complex inner product (verify). This inner product space is called the **standard unitary space**.

Example 4.1.5 Let $V = \mathbb{C}^2$. Define $< >$ by

$$< x,\ y > = 4(x_1\overline{y_1} + x_2\overline{y_2}) + i(x_1\overline{y_2} - x_2\overline{y_1}).$$

Then it is a complex inner product (verify).

Example 4.1.6 Let V denote the complex vector space of all complex-valued continuous functions on $[0, 1]$. Define $< >$ by

$$< f,\ g > = \int_0^1 f(x)\overline{g(x)}dx.$$

Then $< >$ is a complex inner product(verify).

Example 4.1.7 Let l^2 denote the set of all real sequences $\{a_n\}$ such that $\Sigma_{n=1}^{\infty} \mid a_n \mid^2 < \infty$. Then it is a vector space over $\mathbb{R}$ with respect to the usual addition of sequences and multiplication by scalars. We have an inner product on l^2 given by

$$< \{a_n\},\ \{b_n\} > = \Sigma_{n=1}^{\infty} a_n b_n.$$

Let $< >$ be a real (complex) inner product on $\mathbb{R}^n$ ($\mathbb{C}^n$). Let $E = \{\overline{e_1}, \overline{e_2}, \ldots, \overline{e_n}\}$ denote the standard basis of the vector space $\mathbb{R}^n$ ($\mathbb{C}^n$). The inner product $< >$ determines a matrix $A = [a_{ij}]$, where $a_{ij} = < \overline{e_i},\ \overline{e_j} >$. Since $a_{ij} = < \overline{e_i},\ \overline{e_j} > = < \overline{e_j},\ \overline{e_i} > (\overline{< \overline{e_j},\ \overline{e_i} >}) = a_{ji}(\overline{a_{ji}})$, A turns out to be symmetric (Hermitian matrix). The matrix A, in turn, determines the inner product. Indeed, if $\overline{x} = [x_1, x_2, \ldots, x_n] = x_1\overline{e_1} + x_2\overline{e_2} + \cdots + x_n\overline{e_n}$, and $\overline{y} = [y_1, y_2, \ldots, y_n] = y_1\overline{e_1} + y_2\overline{e_2} + \cdots + y_n\overline{e_n}$, then $< \overline{x}, \overline{y} > = \overline{x}A\overline{y}^t (\overline{x}A\overline{y}^{\star})$. Not all symmetric (Hermitian) matrices determine inner products in the manner described above. For example, consider the real symmetric matrix A given by

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}.$$

Then $[1, -1]A[1, -1]^t = 0$, where as $[1, -1] \neq [0, 0]$. As such, A will not determine inner product (observe that A is invertible also). Indeed, as we shall see, a matrix A determines an inner product, as described, if and only if there is a non singular matrix B such that $A = BB^t (BB^{\star})$.

Definition 4.1.8 Let $(V, <>)$ be an inner product space. Let $x \in V$. Then $< x,\ x >$ is a non negative real number. Its non negative square root is called the **length** of the vector x, and it is denoted by $|| x ||$. Thus, $|| x || = +\sqrt{< x,\ x >}$.

Clearly,

$$|| x || = 0 \textit{ if and only if } x = 0.$$

Also, since $< \alpha x,\ \alpha x > = \alpha\overline{\alpha} < x,\ x > = | \alpha |^2 || x ||^2$, we have

$$|| \alpha x || = | \alpha | || x ||$$

for all $\alpha \in F$ and $x \in V$.

Theorem 4.1.9 (Cauchy–Schwarz inequality) *Let $(V, <,\ >)$ be an inner product space. Then $|< x,\ y >| \leq || x || || y ||$ for all $x, y \in V$. The equality holds if and only if $\{x, y\}$ is linearly dependent.*

Proof If $y = 0$, then both side is 0, and the equality holds. Assume that $y \neq 0$. Then $< x - \alpha y,\ x - \alpha y > \geq 0$ for all $\alpha \in F$. Thus,

$$< x,\ x > - \overline{-\alpha} < x,\ y > - \alpha < y, x > + \alpha\overline{\alpha} < y,\ y > \geq 0.$$

Putting $\alpha = \frac{<x,\ y>}{<y,\ y>}$ in the above equation, and noting that $\overline{< x,\ y >} = < y,\ x >$, we obtain that

$$< x,\ x > - \frac{\overline{< x,\ y >} < x,\ y >}{< y,\ y >} - \frac{< x,\ y > \overline{< x,\ y >}}{< y,\ y >} + \frac{< x,\ y > \overline{< x,\ y >}}{< y,\ y >^2} < y,\ y > \geq 0,$$

or

$$< x,\ y > \overline{< x,\ y >} \leq < x,\ x >< y,\ y >,$$

or

$$|< x,\ y >|^2 \leq || x ||^2 || y ||^2 .$$

Taking square root, we obtain that

$$|< x,\ y >| \leq || x || || y || .$$

If $\{x, y\}$ is linearly independent, then $x - \alpha y \neq 0$ for all $\alpha \in F$. Hence the inequality becomes strict inequality, and so in this case

$$|< x,\ y >| < || x || || y || .$$

Conversely, if $\{x, y\}$ is linearly dependent, then $x = \alpha y$ for some $\alpha \in F$. But, then

$$|< x,\ y >| = |< \alpha y,\ y >| = |\ \alpha\ |\ |< y,\ y >| = |\ \alpha\ |\ ||\ y\ ||^2 = |\ \alpha\ |\ ||\ y\ ||\ ||\ y\ ||$$
$$= ||\ x\ ||\ ||\ y\ ||\ .$$

♯

Applying the Cauchy–Schwarz inequality for Examples 4.1.4, 4.1.6, and 4.1.7 respectively, we obtain the following corollaries.

Corollary 4.1.10 (Cauchy inequality). *Let $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n$ be complex numbers. Then*

$$|\ \Sigma_{i=1}^{n} x_i \overline{y_i}\ | \leq \sqrt{\Sigma_{i=1}^{n}\ |\ x_i\ |^2}\sqrt{\Sigma_{i=1}^{n}\ |\ y_i\ |^2}.$$

In particular, the inequality holds for real numbers also. ♯

Corollary 4.1.11 *Let f and g be complex valued continuous functions on $[0, 1]$. Then*

$$|\ \int_0^1 f(x)\overline{g(x)}dx\ | \leq \sqrt{\int_0^1\ |\ f(x)\ |^2\ dx}\ \sqrt{\int_0^1\ |\ g(x)\ |^2\ dx}.$$

♯

Corollary 4.1.12 *If $\{x_n\}$ and $\{y_n\}$ is sequence of real numbers such that $\Sigma_{n=1}^{\infty}\ |\ x_n\ |^2 < \infty$ and $\Sigma_{n=1}^{\infty}\ |\ y_n\ |^2 < \infty$, then*

$$|\ \Sigma_{n=1}^{\infty} x_n y_n\ | \leq \sqrt{\Sigma_{n=1}^{\infty}\ |\ x_n\ |^2}\sqrt{\Sigma_{n=1}^{\infty}\ |\ y_n\ |^2}.$$

♯

Find the inequalities coming out of the Examples 4.1.3 and 4.1.5.

Proposition 4.1.13 (Triangle inequality). *Let $(V, <>)$ be an inner product space. Then*

$$||\ x\ +\ y\ || \leq ||\ x\ ||\ +\ ||\ y\ ||$$

for all $x, y \in V$. Equality holds if and only if $\{x, y\}$ is linearly dependent.

Proof
$(||\ x\ +\ y\ ||)^2$
$= |< x\ +\ y,\ x\ +\ y >|$
$= |< x,\ x > +\ < x,\ y > +\ < y,\ x > +\ < y,\ y >|$
$\leq |< x,\ x >| +\ |< x,\ y >| +\ |< y,\ x >| +\ |< y,\ y >|$
$\leq ||\ x\ ||^2 +2\ ||\ x\ ||\ ||\ y\ ||\ +\ ||\ y\ ||^2$ (by Cauchy–Schwarz)
$= (||\ x\ ||\ +\ ||\ y\ ||)^2.$
Taking the square root, we get

$$|| x + y || \leq || x || + || y || .$$

Further, it is clear from the above that equality holds if and only if

$$< x, y > + < y, x > = 2 || x || \cdot || y || .$$

This is so if and only if $|< x, y >| = || x || \cdot || y ||$ (Cauchy–Schwarz inequality). Again, from the second part of Cauchy–Schwarz, it follows that the equality holds if and only if $\{x, y\}$ is linearly dependent. ♯

If we apply the above proposition to Examples 4.1.4, 4.1.6, and 4.1.7, we get the following corollaries:

Corollary 4.1.14 *If $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n$ are complex numbers (or in particular real numbers), then*

$$\sqrt{\Sigma_{i=1}^{n} | x_i + y_i |^2} \leq \sqrt{\Sigma_{i=1}^{n} | x_i |^2} + \sqrt{\Sigma_{i=1}^{n} | y_i |^2}. \quad ♯$$

Corollary 4.1.15 *If f and g are two complex-valued continuous functions on* $[0, 1]$, *then*

$$\sqrt{\int_0^1 | f(x) + g(x) |^2 \, dx} \leq \sqrt{\int_0^1 | f(x) |^2 \, dx} + \sqrt{\int_0^1 | g(x) |^2 \, dx}. \quad ♯$$

Corollary 4.1.16 *If $\{a_n\}$ and $\{b_n\}$ are sequences in l^2, then*

$$\sqrt{\Sigma_{n=1}^{\infty} | a_n + b_n |^2} \leq \sqrt{\Sigma_{n=1}^{\infty} | a_n |^2} + \sqrt{\Sigma_{n=1} \infty | b_n |^2}. \quad ♯$$

Notion of Distance in an Inner Product Space

We first introduce the notion of distance on a set by abstracting the fundamental properties of distance.

Definition 4.1.17 Let X be a set. A map d from $X \times X$ to the set $R^+ \bigcup \{0\}$ of non negative real numbers (the image of (x, y) under d is denoted by $d(x, y)$ instead of $d((x, y))$) is called a **distance** or a **metric** on X if

1. $d(x, y) = 0$ *if and only if* $x = y$.
2. $d(x, y) = d(y, x)$.
3. (Triangle inequality) $d(x, y) \leq d(x, z) + d(y, z)$
   for all $x, y, z \in X$. The pair (X, d) is called a **metric space**.

Proposition 4.1.18 *Let $(V, <>)$ be an inner product space. Then the inner product $<>$ induces a metric d on V defined by*

$$d(x, y) = || x - y ||$$

Proof $d(x, y) = || x - y || \geq 0$, and $d(x, y) = o$ *if and only if* $|| x - y || = 0$. This means that $d(x, y) = 0$ if and only if $x - y = 0$, or equivalently, $x = y$. Since

$$|| x - y || = || -1(y - x) || = | -1 | || y - x || = || y - x ||,$$

it follows that $d(x, y) = d(y, x)$.
Also

$$d(x, y) = || x - y || = || x - z + z - y || \leq || x - z || + || z - y ||$$
$$= d(x, z) + d(z, y).$$ ♯

Remark 4.1.19 Let (V, d) be a metric space. It may not be possible to define a vector space structure on V, and an inner product on V so that the induced metric is d. For example, take a nonempty set V, and define a metric d' on V by $d'(x, y) = 0$ if $x = y$ and 1 otherwise (verify that d' is indeed a metric called a discrete metric). Given any inner product space structure on V, and $x \neq 0$ in V, and d the induced metric, then $d(\frac{2}{||x||}x, 0) = 2$, and so d' can not be induced by an inner product.

Notion of Angle, Orthogonality

Let $(V, <>)$ be an inner product space. Then by the Cauchy–Schwarz inequality $|< x, y >| \leq || x || || y ||$ for all $x, y \in V$. If $x \neq 0 \neq y$, then

$$\frac{|< x, y >|}{|| x || || y ||} \leq 1.$$

If it is a real inner product space, then

$$-1 \leq \frac{< x, y >}{|| x || || y ||} \leq 1.$$

Thus, there is unique θ, $0 \leq \theta \leq \pi$ such that

$$cos\theta = \frac{< x, y >}{|| x || || y ||}.$$

This θ is called the angle between x and y. In case it is a complex inner product space, the above argument implies that there is a unique θ, $0 \leq \theta < 2\pi$ such that

$$cos\theta + isin\theta = \frac{< x, y >}{|| x || || y ||}.$$

This θ may be termed as angle between x and y.

Any two vector x and y in an inner product space is said to be **orthogonal** if $< x, y >= 0$. This definition extends to the null vector 0 also. Thus, the null vector 0 is orthogonal to each vector. The notation $x \perp y$ is used to say that x and y are orthogonal. Thus,

$$x \perp y \text{ if and only if } < x, y >= 0.$$

A vector x is called a **unit** vector if $|| x || = 1$.

Proposition 4.1.20 (Pythagoras Theorem). *Let $(V, <>)$ be a real inner product space, and $x, y \in V$. Then $x \perp y$ if and only if*

$$|| x - y ||^2 = || x ||^2 + || y ||^2,$$

or equivalently,

$$|| x + y ||^2 = || x ||^2 + || y ||^2 .$$

Proof

$$|| x - y ||^2 = < x - y, x - y > = < x, x > - < x, y > - < y, x > + < y, y >$$
$$= || x ||^2 + || y ||^2 - 2 < x, y > .$$

The result follows. ♯

Remark 4.1.21 In complex inner product space also 'if $x \perp y$, then $|| x - y ||^2 = || x ||^2 + || y ||^2$(verify). But the converse is not true. Consider, for example, the unitary space $\mathbb{C}^2$. The vectors $x = (0, 1)$ and $(1, i)$ are not orthogonal but still $|| x - y ||^2 = || x ||^2 + || y ||^2 = 3$.

In a real inner product space $(V, <, >)$, we have

$$|| x - y ||^2 = || x ||^2 + || y ||^2 - 2 || x || || y || cos\theta,$$

and

$$|| x + y ||^2 = || x ||^2 + || y ||^2 + 2 || x || || y || \ cos\theta.$$

These equations give formula for the diagonals of parallelograms in terms of sides.

Proposition 4.1.22 (Parallelogram Law). *In an inner product space $(V, <>)$ we have*

$$|| x - y ||^2 + || x + y ||^2 = 2 || x ||^2 + 2 || y ||^2$$

for all $x, y \in V$.

Proof Adding equations

$$|| x - y ||^2 = || x ||^2 + || y ||^2 - < x, y > - < y, x >,$$

and

$$|| x + y ||^2 = || x ||^2 + || y ||^2 + < x, y > + < y, x >$$

we get the result. ♯

The geometrical meaning of the above proposition is that the sum of the areas of the squares formed on the diagonals of a parallelogram is the sum of the areas of squares formed on the sides of the parallelogram.

The following identities, termed as the polarization identities, relate the norm with the inner product.

Proposition 4.1.23 (Polarization identities)

1. *If* $(V, < >)$ *is a real inner product space, then*

$$< x, y > = \frac{1}{4}[|| x + y ||^2 - || x - y ||^2]$$

for all $x, y \in V$.

2. *If* $(V, < >)$ *is a complex inner product space, then*

$$< x, y > = \frac{1}{2}[|| x + y ||^2 + i || x + iy ||^2 - (1 + i)(|| x ||^2 + || y ||^2]$$

for all $x, y \in V$.

Proof 1. Let $(V, < >)$ be a real inner product space. Then

$$|| x + y ||^2 = < x + y, x + y > = || x ||^2 + || y ||^2 + 2 < x, y > \ \ldots, \tag{4.1}$$

and

$$|| x - y ||^2 = < x - y, x - y > = || x ||^2 + || y ||^2 - 2 < x, y > \ \ldots, \tag{4.2}$$

for all $x, y \in V$. Subtracting the second equation from the first equation, we get the desired identity.

2. Let $(V, < >)$ be a complex inner product space. Then

$$|| x + y ||^2 = < x + y, x + y > = || x ||^2 + || y ||^2 + < x, y > + < y, x > \ \ldots, \tag{4.3}$$

and

$$|| x + iy ||^2 = < x + iy, x + iy > = || x ||^2 + || y ||^2 - i < x, y > + i < y, x > \ \ldots, \tag{4.4}$$

for all $x, y \in V$. Adding the i times the Eq. 4.4 to the Eq. 4.3, we get the desired result. ♯

Let $(V, <>)$ be an inner product space. A subset S of V is called an **orthonormal set** if

(i) $|| x || = 1$ *for all* $x \in S$ and
(ii) $< x, y > = 0$ *for all* $x, y \in S, x \neq y$.

Proposition 4.1.24 *An orthonormal set in an inner product space is always linearly independent.*

Proof Let S be an orthonormal set and

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0,$$

where $x_1, x_2, \ldots, x_n$ are distinct elements of S. Then

$$\alpha_m x_m = < \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n, x_m > = 0.$$

Hence $\alpha_m = 0$ *for all* m. ♯

Proposition 4.1.25 (Bessels inequality). *Let* $(V, <>)$ *be an inner product space, and* $\{x_1, x_2, \ldots, x_r\}$ *an orthonormal set,* $x_i \neq x_j$ *for* $i \neq j$. *Let* $x \in V$. *Then*

$$\Sigma_{i=1}^{r} |< x, x_i >|^2 \leq || x ||^2 .$$

Proof We have

$$< x - \Sigma_{i=1}^{r} < x, x_i > x_i, x - \Sigma_{i=1}^{r} < x, x_i > x_i > \geq 0.$$

Since $\{x_1, x_2, \ldots, x_r\}$ is an orthonormal set, expanding

$$< x, x > - \Sigma_{i=1}^{r} < x, x_i > \overline{< x, x_i >} \geq 0,$$

or

$$|| x ||^2 - \Sigma_{i=1}^{r} |< x, x_i >|^2 \geq 0.$$

Hence

$$\Sigma_{i=1}^{r} |< x, x_i >|^2 \leq || x ||^2 .$$ ♯

Definition 4.1.26 An orthonormal set which is also a basis is called an **orthonormal basis**.

Corollary 4.1.27 *Let $(V, <>)$ be an inner product space. An orthonormal set $\{x_1, x_2, \ldots, x_n\}$ is an orthonormal basis if and only if*

$$|| x ||^2 = \Sigma_{i=1}^n |< x, x_i >|^2 .$$

Proof Suppose that $\{x_1, x_2, \ldots, x_n\}$ is an orthonormal basis. Let $x \in V$. Then

$$x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$$

for some $\alpha_1, \alpha_2, \ldots, \alpha_n$ in F. But, then $< x, x_i > = \alpha_i$ *for all* i. Hence $x = \Sigma_{i=1}^n < x, x_i > x_i$, and so

$$< x - \Sigma_{i=1}^n < x, x_i > x_i, x - \Sigma_{i=1}^n < x, x_i > x_i >= 0.$$

Expanding, we get

$$|| x ||^2 = \Sigma_{i=1}^n |< x, x_i >|^2 .$$

Conversely, suppose that $|| x ||^2 = \Sigma_{i=1}^n |< x, x_i >|^2$. Then

$$|| x - \Sigma_{i=1}^n < x, x_i > x_i ||^2 = 0,$$

and so

$$x = \Sigma_{i=1}^n < x, x_i > x_i$$

for all $x \in V$. This shows that $\{x_1, x_2, \ldots, x_n\}$ is a set of generators. Since an orthonormal set is already linearly independent, it is a basis. ♯

4.2 Gram–Schmidt Process

The proof of the following theorem gives an algorithm by which we can find an orthonormal basis of an inner product space starting from a set of generators for V. The process is called the **Gram–Schmidt process**.

Theorem 4.2.1 (Gram–Schmidt). *Let $(V, <>)$ be an inner product space. Let $\{x_1, x_2, \ldots, x_r\}$ be a finite subset of V. Then there exists an orthonormal set $\{y_1, y_2, \ldots, y_s\}$, $s \leq r$ such that the subspace generated by $\{x_1, x_2, \ldots, x_r\}$ is the same as that generated by $\{y_1, y_2, \ldots, y_s\}$.*

Proof The proof is by the induction on r. If $r = 0$, then the subset is empty set, and since an empty set is (vacuously) orthonormal, there is nothing to do. Suppose that $r = 1$. If $x_1 = 0$, then $< \{x_1\} > = \{0\}$, and empty set is again an orthonormal set which generates $\{0\}$. If $x_1 \neq 0$, take $y_1 = \frac{x_1}{||x_1||}$. Then $\{y_1\}$ is an orthonormal set which generates $< \{x_1\} >$. Assume the result for r. Consider a subset $\{x_1, x_2, \ldots x_r, x_{r+1}\}$ of V. By our induction assumption there is an orthonormal subset $\{y_1, y_2, \ldots, y_s\}$, $s \leq$

r of V such that $< \{y_1, y_2, \ldots, y_s\} > = < \{x_1, x_2 \ldots, x_r\} >$. If x_{r+1} belongs to this subspace, then there is nothing to do. Suppose that $x_{r+1} \notin < \{x_1, x_2, \ldots, x_r\} = < \{y_1, y_2, \ldots, y_s\} >$. Then

$$x_{r+1} - \Sigma_{i=1}^{s} < x_{r+1}, y_i > y_i \neq 0.$$

Take

$$y_{s+1} = \frac{x_{r+1} - \Sigma_{i=1}^{s} < x_{r+1}, y_i > y_i}{|| x_{r+1} - \Sigma_{i=1}^{s} < x_{r+1}, y_i > y_i ||}.$$

Clearly, y_{s+1} is a unit vector which is orthogonal to y_i for each $i \leq s$. Evidently, $\{y_1, y_2, \ldots y_{s+1}\}$ is an orthonormal set. Since y_{s+1} is linear combination of $\{x_{r+1}, y_1, y_2, \ldots, y_s\}$, $< \{y_1, y_2, \ldots, y_{s+1}\} >$ is contained in $< \{y_1, y_2, \ldots, y_s, x_{r+1}\} > = < \{x_1, x_2, \ldots, x_{r+1}\} >$. Also x_{r+1} is linear combination of $\{y_1, y_2, \ldots, y_{s+1}\}$, and so $< x_1, x_2, \ldots, x_{r+1}\} >$ is contained in $< \{y_1, y_2, \ldots, y_{s+1}\} >$. Thus, the result holds for $r + 1$ also. ♯

Corollary 4.2.2 *Every finite dimensional inner product space admits an orthonormal basis.*

Proof Let $(V, <>)$ be a finite-dimensional inner product space. Let $\{x_1, x_2, \ldots, x_n\}$ be a basis of V. Then from the above theorem, there exists an orthonormal set $\{y_1, y_2, \ldots, y_m\}$, $m \leq n$ which generates V. Since an orthonormal set is linearly independent, it is a basis, and in turn $n = m$. ♯

Proposition 4.2.3 *Every orthonormal set of a finite-dimensional inner product space can be enlarged to an orthonormal basis.*

Proof Let $\{x_1, x_2, \ldots, x_m\}$ be an orthonormal set of an inner product space $(V, <>)$ of dimension n. Since an orthonormal set is linearly independent, $m \leq n$. If $m = n$, it is already a basis, and so an orthonormal basis. Suppose that $m < n$. Then $< \{x_1, x_2, \ldots, x_m\} > \neq V$. Let y_{m+1} be a member of $V - < \{x_1, x_2, \ldots, x_m\} >$. Then $y_{m+1} - \Sigma_{i=1}^{m} < y_{m+1}, x_i > x_i \neq 0$. Take

$$x_{m+1} = \frac{y_{m+1} - \Sigma_{i=1}^{m} < y_{m+1}, x_i > x_i}{|| y_{m+1} - \Sigma_{i=1}^{m} < y_{m+1}, x_i > x_i ||}.$$

Then $\{x_1, x_2, \ldots, x_{m+1}\}$ is an orthonormal set. If $m + 1 = n$, then this is an orthonormal basis. If not proceed as above. At the $(n - m)$th step, we shall arrive at an orthonormal basis containing $\{x_1, x_2, \ldots x_m\}$. ♯

Example 4.2.4 This example is to illustrate the Gram–Schmidt process. Consider the usual Euclidean inner product space $\mathbb{R}^3$. Consider the subset $\{(1, 1, 1), (0, 1, 1), (2, 1, 1)\}$ of $\mathbb{R}^3$. We determine an orthonormal set which generates the same space as $< \{(1, 1, 1), (0, 1, 1), (2, 1, 1)\} >$. Take $x_1 = \frac{(1,1,1)}{||(1,1,1)||} = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$, and $x_2 = \frac{y_2}{||y_2||}$, where

$$y_2 = (0,1,1) - <(0,1,1),\ (\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}})>(\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}).$$

Thus, $x_2 = \sqrt{\frac{3}{2}}(-\frac{2}{3},\frac{1}{3},\frac{1}{3})$. Since

$$(2,1,1) - <(2,1,1),\ x_1>x_1 - <(2,1,1),\ x_2>x_2 = 0,$$

it follows that the orthonormal set $\{x_1, x_2\}$ generates the same space as $\{(1,1,1), (0,1,1), (2,1,1)\}$. Note that $\{(1,1,1), (0,1,1), (2,1,1)\}$ does not generate $\mathbb{R}^3$.

Let

$$A = \begin{bmatrix} \overline{r_1} \\ \overline{r_2} \\ . \\ . \\ . \\ \overline{r_n} \end{bmatrix}$$

be a $n \times m$ matrix with rows $\{\overline{r_1}, \overline{r_2}, \ldots, \overline{r_n}\}$. Then $AA^t = [u_{ij}]$, where $u_{ij} = \overline{r_i}\,\overline{r_j}^t = <\overline{r_i}, \overline{r_j}>$ ($AA^\star = [u_{ij}]$, where $u_{ij} = \overline{r_i}\,\overline{r_j}^\star = <\overline{r_i}, \overline{r_j}>$). Thus, to say that the rows of A form an orthonormal set is to say that $AA^t = I_n$ ($AA^\star = I_n$). Dually, to say that the columns of A form an orthonormal set is to say that $A^tA = I_m$ ($A^\star A = I_m$). In particular, we have the following definition.

Definition 4.2.5 A square $n \times n$ matrix A with entries in the field $\mathbb{R}$ ($\mathbb{C}$) of real (complex) numbers is called an **orthogonal** (**unitary**) matrix if the rows of A form an orthonormal basis of $\mathbb{R}^n$ ($\mathbb{C}^n$), or equivalently, $AA^t = I_n = A^tA$ ($AA^\star = I_n = A^\star A$). Alternatively, A is orthogonal if and only if $A^t = A^{-1}$.

Example 4.2.6 The 2×2 matrices

$$\begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix},$$

and

$$\begin{bmatrix} cos\theta & sin\theta \\ sin\theta & -cos\theta \end{bmatrix}$$

are orthogonal 2×2 matrices. Indeed, any 2×2 orthogonal matrix is one of the above two types (prove it). Observe that the linear transformation

$$[x,y] \mapsto [x,y] \cdot \begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix} = [xcos\theta - ysin\theta,\ xsin\theta + ycos\theta]$$

determined by the matrix

$$\begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix}$$

represents rotation of the x, y plane through an angle θ. Interpret the linear transformation determined by the second matrix. Indeed, it will represent the reflexion about a line in the plane (find it).

Example 4.2.7 The 3×3 matrix

$$\begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix}$$

is an orthogonal matrix.

If $AA^t = I_n = BB^t$, then $AB(AB)^t = ABB^tA^t = I_n$. Thus, product of two $n \times n$ orthogonal matrices are orthogonal. Also, since $A^t = A^{-1}$, the inverse of an orthogonal matrix is orthogonal. The identity matrix is obviously an orthogonal matrix. This shows that the set $O(n)$ of orthogonal $n \times n$ matrices form a group under matrix multiplication. This group $O(n)$ is called the **orthogonal** group of $n \times n$ matrices. A subgroup of $O(n)$ is called an **orthogonal** group. Similarly, the set $U(n)$ of all $n \times n$ unitary matrices form a group, called the **unitary** group.

The proof of the following proposition is algorithmic, and it is essentially the Gram–Schmidt process.

Proposition 4.2.8 *Let A be a $n \times m$ matrix with entries in the field $\mathbb{R}$ ($\mathbb{C}$) of real (complex) numbers which is of rank n. Then, we can find a lower triangular $n \times n$ matrix P with positive diagonal entries such that $PA(PA)^t = I_n$ ($PA(PA)^\star = I_n$). Also, if A is of rank m, then we can find an upper triangular $m \times m$ matrix Q with positive diagonals such that $(AQ)^tAQ = I_m$ ($(AQ)^\star AQ = I_m$).*

Proof Since the rank of A is n, the rows $\{\overline{r_1}, \overline{r_2}, \ldots, \overline{r_n}\}$ of A form a linearly independent subset of $\mathbb{R}^m$. Using the Gram Schmidt process, we transform the rows of A to get a $n \times m$ matrix B with the orthonormal rows $\{\overline{s_1}, \overline{s_2}, \ldots, \overline{s_n}\}$. Clearly, then $BB^t = I_n$. Further, while transforming the rows of A to orthonormal rows, we use only, the following types of elementary row operations in succession:

(i) Multiply a row by a nonzero number, for example, $\overline{s_1} = \frac{1}{|\overline{r_1}|}\overline{r_1}$.
(ii) Add certain linear combinations of rows preceding to j_{th} row to the j_{th} row, and then multiply it by a suitable positive number, for example, $\overline{s_2} = \frac{\overline{r_2} - <\overline{r_2}, \overline{s_1}>\overline{s_1}}{|\overline{r_2} - <\overline{r_2}, \overline{s_1}>\overline{s_1}|}$.

We further observe that the corresponding elementary matrices are lower triangular with positive diagonal entries. Thus, $B = PA$ for some lower $n \times n$ triangular matrix P with positive diagonal entries. Finally, let A be a $n \times m$ matrix of rank m. Then A^t is a $m \times n$ matrix of rank m. Applying the above result for A^t, we get a lower triangular $m \times m$ matrix P with positive diagonal entries such that $PA^t(PA^t)^t = I_m$. Take $Q = P^t$. Then Q is an upper triangular $m \times m$ matrix with positive diagonal entries such that $(AQ)^tAQ = I_m$. ♯

Corollary 4.2.9 *Let A be a $n \times m$ matrix with entries in the field $\mathbb{R}$ ($\mathbb{C}$) of real (complex) numbers which is of rank n. Then, we can find a lower triangular $n \times n$ matrix L with positive diagonal entries and a $n \times m$ matrix Q with $QQ^t = I_n$ ($QQ^\star = I_n$) such that $A = LQ$. Also, we can find an upper triangular $m \times m$ matrix U with positive diagonals and a $n \times m$ matrix Q with $QQ^t = I_n$ ($QQ^\star = I_n$) such that $A = QU$.*

Proof Follows from the above proposition if we take $L = P^{-1}$, where $PA = Q$. ♯

Corollary 4.2.10 *Let A be $n \times n$ invertible matrix. Then we can find a lower triangular $n \times n$ matrix P with positive diagonal entries such that PA is an orthogonal matrix.* ♯

Corollary 4.2.11 *Every invertible matrix can be decomposed as product of a lower triangular matrix with positive diagonal entries and an orthogonal matrix. It can also be decomposed as product of an orthogonal matrix with an upper triangular matrix with positive diagonal entries.* ♯

Now, we illustrate the algorithms described above by means of examples.

Example 4.2.12 The set $S = \{\overline{r_1} = (1, 1, 0), \overline{r_2} = (0, 1, 1), \overline{r_3} = (1, 0, 1)\}$ is a basis of R^3, and therefore, the corresponding matrix

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

is invertible. We transform S in to an orthonormal basis using Gram–Schmidt process, and using the corresponding elementary row operations on A, we transform it to an orthogonal matrix O. Further by applying the same elementary operations on the the identity matrix I_3 we obtain the lower triangular matrix P with positive diagonal entries such that $PA = O$. First $\overline{r_1}$ is replaced by $\overline{s_1} = \frac{\overline{r_1}}{|\overline{r_1}|} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)$ and correspondingly, we multiply the first row of A by $\frac{1}{\sqrt{2}}$ to transform it to

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

Further, we replace $\overline{r_2}$ by $\overline{s_2} = \frac{\overline{r_2} - <\overline{r_2}, \overline{s_1}>\overline{s_1}}{|\overline{r_2} - <\overline{r_2}, \overline{s_1}>\overline{s_1}|} = (-\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \sqrt{\frac{2}{3}})$, and we apply the corresponding elementary operations on the matrix to transform it to

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \sqrt{\frac{2}{3}} \\ 1 & 0 & 1 \end{bmatrix}$$

Finally, we replace $\overline{r_3}$ by $\overline{s_3} = \frac{\overline{r_3} - <\overline{r_3}, \overline{s_1}>\overline{s_1} - <\overline{r_3}, \overline{s_2}>\overline{s_2}}{|\overline{r_3} - <\overline{r_3}, \overline{s_1}>\overline{s_1} - <\overline{r_3}, \overline{s_2}>\overline{s_2}|} = (\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$, and we apply the corresponding elementary operations on the matrix to transform it to the orthogonal matrix

$$O = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \sqrt{\frac{2}{3}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

To get the the corresponding lower triangular matrix P so that $PA = O$, we apply the same elementary row operations in succession on the identity matrix I_3 to get

$$P = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 \\ -\frac{1}{\sqrt{6}} & \sqrt{\frac{2}{3}} & 0 \\ -\frac{1}{2\sqrt{3}} & -\frac{1}{2\sqrt{3}} & \frac{\sqrt{3}}{2} \end{bmatrix}$$

Lastly, to find a lower triangular matrix $L = P^{-1}$ so that $A = LO$, we apply the inverses of the same elementary row operations on the identity matrix in reverse order to get

$$L = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{3}} \end{bmatrix}$$

4.3 Orthogonal Projection, Shortest Distance

Let (X, d) be a metric space, and A a subset of X. Then

$$d(x, A) = g.l.b.\{d(x, a) \mid a \in A\} \text{ (the greatest lower bound of the set } \{d(x, a) \mid a \in A\})$$

is called the **shortest distance** (or simply **distance**) between the point x and the set A. More precisely, $d(x, A)$ is characterized by the following two properties:

(i) $d(x, A) \leq d(x, a)$ *for all* $a \in A$
(ii) For all real $\alpha > d(x, A)$, there is an element $a \in A$ such that $\alpha > d(x, a)$.

Proposition 4.3.1 *Let $(V, <>)$ be a finite dimensional inner product space. Let W be a subspace of V and $\{x_1, x_2, \ldots, x_r\}$ an orthonormal basis of W. Let $x \in V$. Then the shortest distance between x and W is*

$$\sqrt{|| x ||^2 - \Sigma_{i=1}^{r} |< x, x_i >|^2}.$$

Proof Since $\Sigma_{i=1}^{r} < x,\ x_i > x_i \in W$, and

$$|| x - \Sigma_{i=1}^{r} < x,\ x_i > x_i ||^2 = || x ||^2 - \Sigma_{i=1}^{r} |< x,\ x_i >|^2,$$

it follows that the shortest distance between x and W is at least

$$\sqrt{|| x ||^2 - \Sigma_{i=1}^{r} |< x,\ x_i >|^2}.$$

Further, given any $y = \Sigma_{i=1}^{r} \alpha_i x_i$ in W,

$$|| x - y ||^2 = || x ||^2 + \Sigma_{i=1}^{r} | \alpha_i |^2 - \Sigma_{i=1}^{r} \overline{\alpha_i} < x,\ x_i > - \Sigma_{i=1}^{r} \alpha_i \overline{< x,\ x_i >}.$$

Since $a + \overline{a} \leq 2 | a |$ for all complex number a, using Cauchy inequality, we get

$$\begin{aligned} &\Sigma_{i=1}^{r} \overline{\alpha_i} < x,\ x_i > + \Sigma_{i=1}^{r} \alpha_i \overline{< x,\ x_i >} \leq 2 | \Sigma_{i=1}^{r} \alpha_i \overline{< x,\ x_i >} | \\ &\leq 2 \sqrt{\Sigma_{i=1}^{r} | \alpha_i |^2} \sqrt{\Sigma_{i=1}^{r} |< x,\ x_i >|^2}. \end{aligned}$$

Hence

$$|| x - y ||^2 \geq || x ||^2 - \Sigma_{i=1}^{r} |< x,\ x_i >|^2 .$$

Thus, $\sqrt{|| x ||^2 - \Sigma_{i=1}^{r} |< x,\ x_i >|^2}$ is the shortest distance between x and W. ♯

Proposition 4.3.2 *Under the hypothesis of the Proposition 4.3.1,* $\Sigma_{i=1}^{r} < x,\ x_i > x_i$ *is the unique point of* W *which is at the shortest distance from* x.

Proof It follows from the proof of the Proposition 4.3.1 that the said point is at the shortest distance from W. Suppose that $\Sigma_{i=1}^{r} \alpha_i x_i$ is a point of W which is also at the shortest distance from x. Then

$$|| x - \Sigma_{i=1}^{r} \alpha_i x_i ||^2 = || x ||^2 - \Sigma_{i=1}^{r} |< x,\ x_i >|^2,$$

and so

$$\Sigma_{i=1}^{r} | \alpha_i |^2 + \Sigma_{i=1}^{r} |< x,\ x_i >|^2 = \Sigma_{i=1}^{r} \overline{\alpha_i} < x,\ x_i > + \Sigma_{i=1}^{r} \alpha_i \overline{< x,\ x_i >} \ldots . \tag{4.1}$$

Consider the usual standard complex inner product on $\mathbb{C}^r$, and points $u = (\alpha_1, \alpha_2, \ldots, \alpha_r)$ and $v = (< x,\ x_1 >, < x,\ x_2 >, \ldots, < x,\ x_r >)$ of C^r, then the above equation means that

$$|| u ||^2 + || v ||^2 = < u,\ v > + < v,\ u > .$$

Hence

$$< u,\ u > \ + \ < v,\ v > = < u,\ v > \ + \ < v,\ u > .$$

This shows that $< u - v,\ u - v > = 0$, and so $u = v$. Thus, $\alpha_i = < x,\ x_i >$ for all i. ♯

Let V be a vector space. The translates of one dimensional subspaces are called **lines** or **affine lines** in V. Thus, a line in V is of the form $\{a + \lambda b \mid \lambda \in F\}$. This line is the line passing through a and parallel to b (or to the subspace $\{\lambda b \mid \lambda \in F\}$). Translates of subspaces of dimension greater than 1 are called **planes** or **affine planes**. The subset $a + W$, where W is a subspace of dimension $r > 1$ is called a plane of dimension r passing through a and parallel to W. If $dim W = dim V - 1$, then it is said to be a hyperplane.

Corollary 4.3.3 *Let $(V, < >)$ be a finite dimensional inner product space. Let W be a subspace of V with $\{x_1, x_2, \ldots, x_r\}$ an orthonormal basis. Then the distance of a point $a \in V$ from the affine plane $b + W$ (or affine line if $r = 1$) is same as that of $a - b$ from W, and it is*

$$\sqrt{|| \ a - b \ ||^2 \ - \ \Sigma_{i=1}^{r} \ | < a - b,\ x_i > |^2}.$$

The line of shortest distance from a to $b + W$ is the same as perpendicular from a to $b + W$, and it is

$$\{a + \lambda((a - b) - \Sigma_{i=1}^{r} < a - b,\ x_i > x_i) \mid \lambda \in F\}.$$

The foot of perpendicular from a to $b + W$ is $b + \Sigma_{i=1}^{r} < a - b,\ x_i > x_i$.

Proof The shortest distance of a from $b + W$ is $g.l.b\{|| \ a - (b + w) \ || \mid w \in W\}$ which is the same as $g.l.b\{|| \ a - b - w \ || \mid w \in W\}$. Thus, the shortest distance from a to $b + W$ is same as the shortest distance between $a - b$ and W. From the above proposition, it is

$$\sqrt{|| \ a - b \ ||^2 \ - \ \Sigma_{i=1}^{r} \ | < a - b,\ x_i > |^2}.$$

Since $a - b - \Sigma_{i=1}^{r} < a - b,\ x_i > x_i$ is orthogonal to each x_i, it is also orthogonal to each member of W. Thus, the line joining a and $b + \Sigma_{i=1}^{r} < a - b,\ x_i > x_i$ is orthogonal to W. Hence, the line passing through a and perpendicular to $b + W$ is given by

$$\{a + \lambda(a - b - \Sigma_{i=1}^{r} < a - b,\ x_i > x_i) \mid \lambda \in F\}.$$

Clearly, $b + \Sigma_{i=1}^{r} < a - b,\ x_i > x_i$ is the foot of perpendicular from a on to $b + W$. ♯

Proposition 4.3.4 *Let $(V, < >)$ be an inner product space. Let W be a subspace of V. Then $W^{\perp} = \{v \in V \mid < x,\ v > = 0$ for all $x \in W\}$ is a subspace of V.*

Proof Clearly $0 \in W^{\perp}$, and so $W^{\perp} \neq \emptyset$. Let $y, z \in W^{\perp}$. Then $< x,\ y >\ = 0 = < x,\ z >$ *for all* $x \in W$. But, then $< x,\ \alpha y + \beta z >\ = \overline{\alpha} < x,\ y >\ + \overline{\beta} < x,\ z >\ = 0$. This shows that $\alpha y + \beta z \in W^{\perp}$. It follows that $W^{\perp}$ is a subspace. ♯

Definition 4.3.5 The subspace $W^{\perp}$ defined in the above proposition is called the **orthogonal complement** of W.

Proposition 4.3.6 *Let V be a finite dimensional inner product space, and W a subspace of V. Then V is the direct sum of W and $W^{\perp}$.*

Proof Let $\{x_1, x_2, \ldots, x_r\}$ be an orthonormal basis of W. Then this being an orthonormal set can be enlarged to an orthonormal basis $\{x_1, x_2, \ldots, x_r, x_{r+1}, \ldots, x_n\}$ of V, where $n = dimV$. An element $x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_r x_r + \alpha_{r+1} x_{r+1} + \cdots \alpha_n x_n$ is orthogonal to W if and only if it is orthogonal to each $x_i,\ i \leq r$. This is so if and only if $\alpha_i = < x,\ x_i >\ = 0\ \forall i \leq r$. This shows that $W^{\perp} = < \{x_{r+1}, x_{r+2}, \ldots, x_n\} >$. Since every element $x \in V$ can be uniquely expressed as $x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_r x_r + \alpha_{r+1} x_{r+1} + \alpha_{r+2} x_{r+2} + \cdots + \alpha_n x_n$, it follows that every element x of V has a unique representation as $x = y + z$, where $y \in W$ and $z \in W^{\perp}$. ♯

Remark 4.3.7 Let $(V, <,\ >)$ be a finite-dimensional inner product space, and W a subspace of V. Suppose that $x = y + z$, where $y \in W$ and $z \in W^{\perp}$. Then y is called the component of x along W, and z is called the component of x orthogonal to W. Clearly, y is the foot of perpendicular from x to W.

Definition 4.3.8 Each subspace W of an inner product space V defines the map P_W from V to V given by $P_W(x) = y$, where y is the foot of perpendicular from x to W. The map P_W is a linear transformation called the **orthogonal projection** of V on to W.

Example 4.3.9 Consider the subspace $W = \{\overline{x} = (x_1, x_2, x_3) \mid x_1 + x_2 + x_3 = 0\}$. The subset $S = \{(1, -1, 0),\ (0, 1, -1)\}$ is a basis of W (verify). Using Gram Schmidt process, we get an orthonormal basis $\{(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0),\ (\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\sqrt{\frac{2}{3}})\}$ of W. There fore, the foot of the perpendicular from a point $\overline{x} = (x_1, x_2, x_3)$ on to W is $< \overline{x}, (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0) > (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0) + < \overline{x}, (\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\sqrt{\frac{2}{3}}) > (\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\sqrt{\frac{2}{3}}) = (\frac{2x_1 - x_2 - x_3}{3}, \frac{-x_1 + 2x_2 - x_3}{3}, \frac{-x_1 - x_2 + 2x_3}{3})$. The matrix of the orthogonal projection P_W is given by

$$P_W = \frac{1}{3}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

Remark 4.3.10 The result of the above proposition is not true for infinite-dimensional inner product space. For example, consider the vector space $C[0, 1]$ of real-valued continuous functions from the closed interval $[0, 1]$ with inner product given by

$$< f,\ g >\ = \int_0^1 f(x)g(x)dx.$$

Let W be the subspace of polynomial functions. Then W is a proper subspace, and $W^{\perp} = \{0\}$ (prove it). Thus, $C[0, 1]$ is not direct sum of W and $W^{\perp}$.

Let $(V, <>)$ be an inner product space. Let $y \in V$. It follows from the definition of the inner product that the map f_y from V to F defined by $f_y(x) = < x, y >$ is a linear functional on V, and the map f^y defined by $f^y(x) = < y, x >$ is anti-linear functional in the sense that $f^y(\alpha x + \beta z) = \overline{\alpha} f^y(x) + \overline{\beta} f^y(z)$. Observe that the set of all anti-linear functionals also form a vector space anti isomorphic to the dual space $V^{\star}$ of V.

Proposition 4.3.11 *Let $(V, <>)$ be a finite-dimensional inner product space. Then the map $y \rightsquigarrow f_y$ is an anti-linear isomorphism from V to $V^{\star}$. Also the map $f \rightsquigarrow f^y$ is an isomorphism from V to the space of all anti-linear functionals on V.*

Proof That the map $y \rightsquigarrow f_y$ is an anti-linear transformation follows from the definition of inner product. We show that it is injective. Suppose that $f_y = f_z$. Then $f_y(x) = f_z(x)$ for all x. Hence $< x, y > = < x, z >$ for all x. This means that $< x, y - z > = 0$ for all x. In particular, $< y - z, y - z > = 0$. This implies that $y = z$. Next, it is easy to observe that an anti-linear transformation takes a subspace to a subspace, and an injective anti-linear transformation takes a linearly independent subset to a linearly independent subset. Thus, the image of V under the injective anti-linear transformation $y \rightsquigarrow f_y$ is a subspace of $V^{\star}$ of dimension equal to the dimension of V. Since $dim V = dim V^{\star}$, it follows that $y \rightsquigarrow f_y$ is also surjective. The rest of the proposition follows similarly. ♯

The following corollary is a restatement of the bijectivity of the map $y \rightsquigarrow f_y$.

Corollary 4.3.12 *Let $(V, <>)$ be a finite-dimensional inner product space, and f a linear functional on V. Then there is a unique $y \in V$ such that $f(x) = < x, y >$ for all $x \in V$.* ♯

Remark 4.3.13 The result of the Proposition 4.3.11, and the Corollary 4.3.12 is not true, in general, for an infinite dimensional inner product spaces. Consider, for example, the space $P[0, 1]$ of polynomial functions on $[0, 1]$. We have an inner product on this space defined by

$$< f, g > = \int_0^1 f(x)g(x)dx.$$

Consider the linear functional ϕ on $P[0, 1]$ defined by $\phi(f) = f(1)$. Check that there is no g in $P[0, 1]$ such that

$$\int_0^1 f(x)g(x)dx = f(1)$$

for all $f \in P[0, 1]$.

Adjoint of a Linear Transformation

Let $(V, <>)$ be a finite dimensional inner product space. Let T from V to V be a linear transformation. The map $x \rightsquigarrow < T(x),\ y >$ is a linear functional on V(verify) for each $y \in V$. Hence from the Corollary 4.3.12, there is a unique element in V which we denote by $T^\star(y)$ such that

$$< T(x),\ y > = < x,\ T^\star(y) >$$

for each $x \in V$. This defines a map $T^\star$ from V to V given by the equation

$$< T(x),\ y > = < x,\ T^\star(y) > .$$

Using the defining property of an inner product, we see that

$$< x,\ T^\star(\alpha y + \beta z) > = < T(x),\ \alpha y + \beta z > = \overline{\alpha} < T(x),\ y > + \overline{\beta} < T(x),\ z >$$
$$= \overline{\alpha} < x,\ T^\star(y) > + \overline{\beta} < x,\ T^\star(z) > = < x,\ \alpha T^\star(y) + \beta T^\star(z) >$$

for each $x \in V$. Thus,

$$< x,\ T^\star(\alpha y + \beta z) - (\alpha T^\star(y) + \beta T^\star(z)) > = 0$$

for all $x \in V$. Putting $x = T^\star(\alpha y + \beta z) - \alpha T^\star(y) - \beta T^\star(z)$, we get that

$$T^\star(\alpha y + \beta z) = \alpha T^\star(y) + \beta T^\star(z)$$

for all $y, z \in V$, and $\alpha, \beta \in F$. This shows that $T^\star$ is a linear transformation.

Definition 4.3.14 The linear transformation defined above is called the **adjoint** of T.

Proposition 4.3.15 *Let $(V, <>)$ be a finite dimensional inner product space, and T a linear transformation from from V to V. Let $B = \{x_1, x_2, \ldots, x_n\}$ be an orthonormal basis of V. Let $M(T)$ denote the matrix of T with respect to the orthonormal basis B. Then $M(T^\star) = (M(T))^\star$ (the tranjugate of the matrix $M(T)$). Further, if $M(T_2) = (M(T_1))^\star$, then $T_2 = T_1^\star$.*

Proof Suppose that $M(T) = [a_{ij}]$ and $M(T^\star) = [b_{jl}]$. Then $T(x_i) = \sum_{k=1}^n a_{ki} x_k$ and $T^\star(x_j) = \sum_{l=1}^n b_{lj} x_l$. Thus, $a_{ji} = < \sum_{k=1}^n a_{ki} x_k,\ x_j > = < T(x_i),\ x_j > = < x_i,\ T^\star(x_j) > = < x_i,\ \sum_{l=1}^n b_{lj} x_l > = \overline{b_{ij}}$. This shows that $M(T^\star) = (M(T))^\star$. Suppose that $M(T_2) = (M(T_1))^\star$. Then $M(T_2) = M(T_1^\star)$. Since the matrix representation map with respect to a fixed basis is bijective, the result follows. ♯

Proposition 4.3.16 *Let $(V, <>)$ be a finite-dimensional inner product space. The map η defined by $\eta(T) = T^\star$ from $EndV$ to $EndV$ is an anti isomorphism of algebras which is an involution in the sense that $\eta^2 = I_V$.*

Proof Since $< x,\ (\alpha T_1 + \beta T_2)^\star(y) > = < (\alpha T_1 + \beta T_2)(x),\ y > = \alpha < T_1(x),\ y > + \beta < T_2(x),\ y > = \alpha < x,\ T_1^\star(y) > + \beta < x,\ T_2^\star(y) > = < x,\ (\overline{\alpha} T_1^\star + \overline{\beta} T_2^\star)(y) >$ for all $x \in V$, it follows that $(\alpha T_1^\star + \beta T_2^\star)(y) = \overline{\alpha} T_1^\star(y) + \overline{\beta} T_2^\star(y)$ for each $y \in V$. Hence $(\alpha T_1 + \beta T_2)^\star = \overline{\alpha} T_1^\star + \overline{\beta} T_2^\star$. Further, $< x,\ (T_1 o T_2)^\star(y) > = < T_1 o T_2(x),\ y > = < T_2(x), T_1^\star(y) > = < x,\ T_2^\star T_1^\star(y) >$ for all $x, y \in V$. Hence $(T_1 o T_2)^\star = T_2^\star o T_1^\star$. Also $< x,\ T(y) > = \overline{< T(y),\ x >} = \overline{< y,\ T^\star(x) >} = < T^\star(x),\ y > = < x,\ (T^\star)^\star(y) >$ for all $x.y \in V$. Hence $T = (T^\star)^\star$. It is clear that $\eta^2 = I_V$, and so η is bijective. ♯

Definition 4.3.17 Let $(V, < >)$ be a complex inner product space. A linear transformation T from V to V is called a

(i) **self adjoint** or **Hermitian** linear transformation if $T^\star = T$, i.e., $< T(x),\ y > = < x,\ T(y) >$ *for all* $x, y \in V$.
(ii) **skew Hermitian** linear transformation if $T^\star = -T$, i.e., $< T(x),\ y > = - < x,\ T(y) >$ *for all* $x, y \in V$.
(iii) **unitary** linear transformation if $T^\star = T^{-1}$, or equivalently $< T(x),\ T(y) > = < x,\ y >$ for all $x, y \in V$.
(iv) **normal** linear transformation if $T^\star T = T T^\star$, or equivalently, $< T(x),\ T(y) > = < T^\star(x),\ T^\star(y) >$ for all $x, y \in V$.

It is clear from the definition that every Hermitian linear transformation as well as every skew Hermitian linear transformation is normal. Also every unitary linear transformation is normal.

The following corollary is immediate from Proposition 4.3.15.

Corollary 4.3.18 *A linear transformation T on an inner product space V is Hermitian (skew-Hermitian, unitary or normal) if and only if the matrix representation $M(T)$ of T relative to an orthonormal basis is Hermitian (skew-Hermitian, unitary, or normal respectively).* ♯

Example 4.3.19 Consider the usual complex inner product space $\mathbb{C}^2$. Define a map T from $\mathbb{C}^2$ to itself by $T((x, y)) = (x + iy,\ -ix + y)$, $x, y \in \mathbb{C}$. Then T is Hermitian. Indeed, the matrix of T relative to the standard orthonormal basis $\{\overline{e_1}, \overline{e_2}\}$ is

$$\begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$$

which is clearly a Hermitian matrix. T is not unitary as the matrix is not unitary. The linear transformation $U = \frac{T}{\sqrt{2}}$ is unitary (check for the corresponding matrix). The linear transformation $(x, y) \rightsquigarrow (ix, y)$ from $\mathbb{C}^2$ to itself is unitary (check), but it is not Hermitian (verify). The linear transformation $(x, y) \rightsquigarrow (ix, 2y)$ is normal, but it is neither Hermitian nor unitary (check).

Let $(V, < >)$ be a real inner product space. A linear transformation T from V to V is called a

(i) **real symmetric** (**real skew symmetric**) if $T^\star = T(T^\star = -T)$.
(ii) **orthogonal** if $T^\star = T^{-1}$.

Example 4.3.20 The linear transformation T_θ from $\mathbb{R}^2$ to $\mathbb{R}^2$ defined by $T_\theta((x, y)) = (xcos\theta + ysin\theta, -xsin\theta + ycos\theta)$ is an orthogonal linear transformation(verify).

Let $H(V)(SH(V))$ denote the set of all Hermitian (skew-Hermitian) linear transformation on a complex inner product spaces $(V, <>)$. If $T \in H(V) \bigcap SH(V)$, then $T = T^\star = -T$. This is equivalent to say that $T = \{0\}$. Thus $H(V) \bigcap SH(V) = \{0\}$. If T_1 and T_2 are in $H(V)$ $(SH(V))$, then $(T_1 \pm T_2)^\star = T_1^\star \pm T_2^\star = T_1 \pm T_2(-(T_1 \pm T_2))$. This shows that $H(V)$ and $SH(V)$ are subgroups of $End(V)$, and their intersection is zero. Further, if T is any Hermitian (skew) linear transformation, and α a complex number, then αT is Hermitian (skew-Hermitian) if and only if α is purely real (imaginary). Thus, $H(V)$ and $SH(V)$ are not subspaces. Also T is Hermitian if and only if iT is skew-Hermitian. The map $T \rightsquigarrow iT$ defines an isomorphism from the group $H(V)$ to $SH(V)$. Given $T_1, T_2 \in H(V)$, $T_1T_2 \in H(V)$ if and only if $T_1T_2 = (T_1T_2)^\star = T_2^\star T_1^\star = T_2T_1$. Similarly, given $T_1, T_2 \in SH(V)$, $T_1T_2 \in SH(V)$ if and only if $T_1T_2 = -T_2T_1$. Let T be any endomorphism of V. Then $(\frac{T+T^\star}{2})$ is Hermitian for

$$(\tfrac{T+T^\star}{2})^\star = \tfrac{T^\star+(T^\star)^\star}{2} = \tfrac{T+T^\star}{2}.$$

To summarize, we have proved the following proposition.

Proposition 4.3.21 *Let $(V, <>)$ be a finite-dimensional complex inner product space. Then the group $EndV$ is direct sum of its subgroups $H(V)$ and $SH(V)$. The subgroups $H(V)$ and $SH(V)$ are not subspaces. They are isomorphic as groups under the map $T \rightsquigarrow iT$. $H(V)$ and $SH(V)$ are not closed under product. Indeed, $T_1, T_2 \in H(V)$ implies that $T_1T_2 \in H(V)$ if and only if $T_1T_2 = T_2T_1$. Also $T_1, T_2 \in SH(V)$ implies that $T_1T_2 \in SH(V)$ if and only if $T_1T_2 = -T_2T_1$.* ♯

The following two propositions follow immediately from the corresponding result in matrices provided we observe that the matrix representation map M relative to an orthonormal basis is an isomorphism from $End(V)$ to $M_n(R)$ which maps $S(V)$ to $S_n(\mathbb{R})$ and $SS(V)$ to $SS_n(\mathbb{R})$.

Proposition 4.3.22 *Let $(V, <>)$ be a real inner product space. Then the set $S(V)(SS(V))$ of symmetric (skew symmetric) linear transformations forms subspaces of $End(V)$, and $End(V)$ is direct sum of these subspaces. Further, product of any two symmetric (skew-symmetric) linear transformations A and B is symmetric (skew-symmetric) if and only if $AB = BA$ $(AB = -BA)$.* ♯

Proposition 4.3.23 *Let $(V, <>)$ be a real inner product space of dimension n. Then the dimension of $S(V)$ is $\frac{n^2+n}{2}$ and that of $SS(V)$ is $\frac{n^2-n}{2}$.* ♯

4.4 Isometries and Rigid Motions

Definition 4.4.1 Let (X, d) be a metric space. A bijective map f from X to itself is called an **isometry** of (X, d) if $d(f(x), f(y)) = d(x, y)$ *for all* $x, y \in X$.

The set of all isometries of (X, d) is denoted by $Iso(X)$, and it is clearly a group under composition of maps. This group is a subgroup of $Sym(X)$. If $(V, <>)$ is an inner product space, then it is already equipped with a metric induced by the inner product. We shall try to describe the isometries of V with respect to the induced metric, and also its group $Iso(V)$ of isometries.

Theorem 4.4.2 *Let* $(V, <>)$ *be a finite dimensional complex inner product space. Let* T *be a linear transformation from* V *to* V. *Then the following conditions are equivalent.*

1. T *is an isometry of* V, *i.e.,* $|| T(x) - T(y) || = || x - y ||$ *for all* $x, y \in V$.
2. $|| T(x) || = || x ||$ *for all* $x \in V$.
3. T *is a unitary linear transformation.*
4. $T^\star = T^{-1}$.

Proof $1 \Longrightarrow 2$. Assume 1. Then $|| T(x) || = || T(x) - T(0) ||$
$= || x - 0 || = || x ||$ *for all* $x \in V$.

$2 \Longrightarrow 3$. Assume 2. We have the following polarization identity (Proposition 4.1.23)

$$< x,\ y > = \frac{1}{2}[|| x + y ||^2 + i\ || x + iy ||^2 - (1 + i)(|| x ||^2 + || y ||^2)]$$

for all $x, y \in V$. Thus,
$< T(x), T(y) > =$
$\frac{1}{2}[|| T(x) + T(y) ||^2 + i\ || T(x) + iT(y) ||^2 - (1 + i)(|| T(x) ||^2 + || T(y) ||^2)]$
for all $x, y \in V$. Since T is a linear transformation, we have
$< T(x), T(y) > =$
$\frac{1}{2}[|| T(x + y) ||^2 + i\ || T(x + iy) ||^2 - (1 + i)(|| T(x) ||^2 + || T(y) ||^2)]$
for all $x, y \in V$. Using 2, we see that
$< T(x), T(y) > =$
$\frac{1}{2}[|| x + y ||^2 + i\ || x + iy ||^2 - (1 + i)(|| x ||^2 + || y ||^2)] = < x, y >$
for all $x, y \in V$. This shows that T is a unitary linear transformation.

$3 \Longrightarrow 4$. Assume 3. Then
$< x, y > = < T(x), T(y) > = < x, T^\star(T(y)) >$
for all $x, y \in V$. Hence $< x, y - T^\star T(y) > = 0$ for all $x, y \in V$. Putting $x = y - T^\star(T(y))$, we get that $T^\star(T(y)) = y$ for all $y \in V$. This shows that $T^\star T = I_V$. Since V is finite dimensional $TT^\star = I_V$. This means that $T^\star = T^{-1}$.

$4 \Longrightarrow 1$. Assume 4. Then $T^\star T = I_V$, and so $< T(x), T(y) >$
$= < x, T^\star T(y) > = < x, y >$ for all $x, y \in V$. In particular, $< T(x), T(x) > =$
$< x, x >$ for all $x \in V$. This means that $|| T(x) || = || x ||$ for all $x \in V$. Since

T is a linear transformation $|| T(x) - T(y) || = || T(x - y) || = || x - y ||$ for all $x, y \in V$. ♯

Corollary 4.4.3 $GL(V) \bigcap Iso(V) = U(V)$. ♯

Proposition 4.4.4 *Let $(V_1, <>_1)$ and $(V_2, <>_2)$ be complex inner product spaces. Let f be a vector space isomorphism from V_1 to V_2 which preserves inner product. Then f induces an isomorphism $\eta(f)$ from $U(V_1)$ to $U(V_2)$ defined by $\eta(f)(T) = foTof^{-1}$.*

Proof It is clear that $\eta(f)$ defined above is an isomorphism from $GL(V_1)$ to $GL(V_2)$. It is sufficient, therefore, to prove that if f preserves inner product then $\eta(f)$ takes $U(V_1)$ onto $U(V_2)$. Suppose that f preserves inner product. Then $< x, y > = < f(f^{-1}(x)), f(f^{-1}(y)) > = < f^{-1}(x), f^{-1}(y) >$. This shows that f preserves inner product if and only if f^{-1} preserves inner product. It is also immediate that composition of inner product preserving maps are inner product preserving. Hence, if f is inner product preserving, and g an isomorphism, then fog (gof) is inner product preserving if and only if $g = f^{-1}o(fog)$ is inner product preserving. We know that $T \in U(V_1)$ if and only if T is an isomorphism which is inner product preserving. It follows that $T \in U(V_1)$ if and only if $foTof^{-1}$ is inner product preserving. Thus, $\eta(f)$ induces an isomorphism from $U(V_1)$ to $U(V_2)$. ♯

Proposition 4.4.5 *Any two n-dimensional complex inner product spaces are isomorphic as inner product spaces, i.e., there is an isomorphism between them which preserve inner product.*

Proof Let $(V_1, <>_1)$ and $(V_2, <>)$ be two complex inner product spaces each of dimension n. Let $\{x_1, x_2, \ldots, x_n\}$ be an orthonormal basis of V_1, and $\{y_1, y_2, \ldots, y_n\}$ be that of V_2. Then there is an isomorphism f from V_1 to V_2 which takes x_i to y_i. But, then

$$< \Sigma_{i=1}^n \alpha_i x_i,\ \Sigma_{i=1}^n \beta_i x_i > = \Sigma_{i=1}^n \alpha_i \overline{\beta_i} = < \Sigma_{i=1}^n \alpha_i y_i,\ \Sigma_{i=1}^n \beta_i y_i > .$$

This shows that f preserves inner product. ♯

Corollary 4.4.6 *Every n-dimensional complex inner product space is isomorphic as inner product space to the standard complex inner product space $\mathbb{C}^n$.* ♯

Thus, if V is a n-dimensional complex inner product space, then the group $U(V)$ is isomorphic to $U(\mathbb{C}^n)$. The group $U(C^n)$ is denoted by $U(n)$, and it is called the **unitary** group on n-dimensional inner product space.

Proposition 4.4.7 *Let $(V, <>)$ be a complex inner product space, and H a linear transformation from V to itself. Then H is a Hermitian linear transformation if and only if $< H(x),\ x >$ is real for all $x \in V$.*

Proof Suppose that H is Hermitian. Then

$$< H(x),\ x > = < x,\ H^{\star}(x) > = < x,\ H(x) > = \overline{< H(x),\ x >}.$$

Hence $< H(x),\ x >$ is real for all $x \in V$. Conversely, suppose that $< H(x),\ x >$ is real for all $x \in V$. Then $< H(x+y),\ (x+y) >$ is real for all $x, y \in V$. Expanding, and noting that $< H(x),\ x >$ and $< H(y),\ y >$ are real, we find that

$$< H(x),\ y > \ + \ < H(y),\ x > \ \text{is real for all } x, y \in V.$$

Again expanding $< H(x+iy),\ (x+iy) >$, we get that

$$< H(x),\ y > \ - \ < H(y),\ x > \ \text{is purely imaginary for all } x, y \in V.$$

It is an elementary fact that if z_1 and z_2 are two complex numbers such that $z_1 + z_2$ is real and $z_1 - z_2$ is purely imaginary, then $z_1 = \overline{z_2}$. This shows that $< H(x),\ y > = \overline{< H(y),\ x >} = < x,\ H(y) >$ for all $x, y \in V$. This shows that H is Hermitian. ♯

Corollary 4.4.8 *Let H be a Hermitian linear transformation from a finite dimensional complex inner product space $(V, < >)$ to itself. Then $I - iH$ and $I + iH$ are isomorphisms. Also $(I + iH)(I - iH)^{-1}$ is a unitary linear transformation.*

Proof Suppose that $(I - iH)(x) = 0$. Then $x = iH(x)$, and so $< x,\ x > = i < H(x),\ x >$ is real. Since $< x,\ x >$ is real, and from the above proposition $< H(x),\ x >$ is also real, it follows that $< x,\ x > = 0$, and so $x = 0$. This shows that $I - iH$ is an injective linear transformation from V to itself. Since V is finite dimensional, it follows that $I - iH$ is an isomorphism. Similarly, $I + iH$ is also an isomorphism. Further, since $(T_1 o T_2)^{\star} = T_2^{\star} o T_1^{\star}$, and $(T^{-1})^{\star} = (T^{\star})^{-1}$, we have

$$\begin{aligned}((I + iH)(I - iH)^{-1})^{\star} &= ((I - iH)^{\star})^{-1}(I + iH)^{\star} \\ &= (I + iH)^{-1}(I - iH) = (I - iH)(I + iH)^{-1} = ((I + iH)(I - iH)^{-1})^{-1}.\end{aligned}$$

(Note that $(I - iH)$ and $(I + iH)$ commute, and so $I - iH$ and $(I + iH)^{-1}$ also commute.) This shows that $(I + iH)(I - iH)^{-1}$ is unitary. ♯

Rigid Motion

Definition 4.4.9 Let $(V, < >)$ be a finite-dimensional real inner product space. Let d be the metric induced by the inner product. An isometry of (V, d) is also called a **rigid motion** on V. Thus, T is a rigid motion if

$$|| T(x) - T(y) || = || x - y ||$$

for all $x, y \in V$.

The group $Iso(V)$ of all rigid motions is called the **group of rigid motions** on V, and it is denoted by $M(V)$.

As in case of complex inner product space, an inner product preserving isomorphism from a real inner product space V_1 to a real inner product space V_2 induces an isomorphism from $M(V_1)$ to $M(V_2)$. Thus, the group of motion on an n-dimensional real inner product space V is isomorphic to $M(\mathbb{R}^n)$. This group is called the group of Euclidean motions.

Theorem 4.4.10 *Let $(V, <>)$ be a finite dimensional real inner product space. Let T be a map from V to V. Then the following conditions are equivalent.*

1. *T is a rigid motion which fixes origin 0.*
2. *T preserves inner product, i.e., $< T(x),\ T(y) > = < x,\ y >$ for all $x, y \in V$. This is equivalent to say that T preserves angle between vectors.*
3. *T is an orthogonal linear transformation.*
4. *T is a linear transformation such that $T^\star = T^{-1}$.*
5. *T is a linear transformation which preserves lengths of vectors, i.e., $|| T(x) || = || x ||$ for all $x \in V$.*

Proof $1 \Longrightarrow 2$. Assume 1. Then

$$|| T(x) || = || T(x) - 0 || = || T(x) - T(0) || = || x - 0 || = || x ||$$

for all $x \in V$. Also, then
$|| x ||^2 + || y ||^2 - 2 < T(x),\ T(y) >$
$= || T(x) ||^2 + || T(y) ||^2 - 2 < T(x),\ T(y) >$
$= || T(x) - T(y) ||^2$
$= || x - y ||^2$
$= || x ||^2 + || y ||^2 - 2 < x,\ y >$
for all $x, y \in V$. Hence $< T(x),\ T(y) > = < x,\ y >$ for all $x, y \in V$.

$2 \Longrightarrow 3$. Assume 2. It is sufficient to prove that T is a linear transformation. Using the fact that T preserves the inner product, we see that

$$\begin{aligned} &< T(x+y) - T(x) - T(y),\ T(x+y) - T(x) - T(y) > \\ &= < x + y - x - y,\ x + y - x - y > = 0 \end{aligned}$$

for all $x, y \in V$. This shows that $T(x + y) = T(x) + T(y)$ for all $x, y \in V$. Similarly, it can be shown that $T(\alpha x) = \alpha T(x)$ for all $\alpha \in \mathbb{R}$ and $x \in V$.

The proof of $3 \Longrightarrow 4$ is similar to the proof of $3 \Longrightarrow 4$ in the Theorem 4.4.2.

$4 \Longrightarrow 5$. Assume 4. Then $T^\star = T^{-1}$. Hence

$$|| T(x) ||^2 = < T(x),\ T(x) > = < x,\ T^\star T(x) > = < x,\ x > = || x ||^2$$

for all $x \in V$.

$5 \Longrightarrow 1$. Assume 5. Since T is a linear transformation, $T(0) = 0$ and

$$|| T(x) - T(y) || = || T(x - y) || = || x - y ||$$

for all $x, y \in V$. ♯

Remark 4.4.11 The analogue of $1 \Longrightarrow 2$ is not valid in complex case. For example the map T from the usual complex inner product space $\mathbb{C}^n$ to itself defined by

$$T((z_1, z_2, \ldots, z_n)) = (\overline{z_1}, z_2, \ldots, z_n)$$

preserves distance, fixes origin, but does not preserve inner product.

Remark 4.4.12 A length preserving map from a real inner product space to itself need not be an orthogonal transformation. Indeed, the translation map $L_{\overline{a}}$ from the usual inner product space $\mathbb{R}^n$ to itself defined by

$$L_{\overline{a}}(\overline{x}) = \overline{x} + \overline{a},$$

$\overline{a} \neq \overline{0}$ preserves length but it is not a linear transformation.

Let $(V, <>)$ be a real inner product space and $a \in V$. The map L_a from V to V defined by $L_a(x) = x + a$ is a rigid motion. This is called the translation by a. The set $\Im(V)$ of all translations on V is a subgroup of the group $M(V)$ of rigid motions which is isomorphic to $(V, +)$ (the map $a \rightsquigarrow L_a$ is an isomorphism). Also $O(V)$ the group of orthogonal linear transformations is a subgroup of $M(V)$.

Theorem 4.4.13 *Every rigid motion of a finite dimensional real inner product space $(V, <>)$ is uniquely expressible as a product of a translation and an orthogonal linear transformation.*

Proof Let $\phi \in M(V)$. Let $\phi(0) = a$. Then the map T from V to V defined by $T(x) = \phi(x) - a$ is also a rigid motion such that $T(0) = 0$. From the previous result, T is an orthogonal linear transformation and $\phi = L_a o T$. Further, suppose that $L_a o T = L_b o T'$, where T and T' are orthogonal linear transformations. Then $a = L_a o T(0) = L_b o T(0) = b$. This shows that $a = b$ and $T = T'$. ♯

Recall that a group G is said to be (internal) semi-direct product of a normal subgroup H by K if

(i) $G = HK$, and
(ii) $H \bigcap K = \{e\}$.

In this case every element g of G has a unique representation as $g = hk$, where $k \in K$ and $h \in H$.

Corollary 4.4.14 *$M(V)$ is semi-direct product of $\Im(V)$ by $O(V)$.*

Proof It follows from the above result that $M(V) = \Im(V)O(V)$. Also if $L_a \in O(V)$, then $a = L_a(0) = 0$, and so $L_a = I_V$. Thus, $\Im(V) \bigcap O(V) = \{I_V\}$.

Now, we show that $\Im(V)$ is a normal subgroup of $M(V)$. Let $L_a oT \in M(V)$ and $L_b \in \Im(V)$, where $T \in O(V)$. Then

$$(L_a oT)^{-1} o L_b o (L_a oT)(x) \;=\; x \;+\; T^{-1}(b) \;=\; L_{T^{-1}(b)}(x).$$

Hence $(L_a oT)^{-1} o L_b o (L_a oT) \;=\; L_{T^{-1}(b)} \in \Im(V)$. ♯

Following corollary follows from the second isomorphism theorem.

Corollary 4.4.15 *$M(V)/\Im(V)$ is isomorphic to $O(V)$.* ♯

Exercises

4.4.1 Define a map $<\,>'$ from $\mathbb{R}^3 \times \mathbb{R}^3$ to $\mathbb{R}$ by

$$<(x_1, x_2, x_3),\ (y_1, y_2, y_3)> = \; x_1y_1 + x_2y_1 + x_1y_2 + 2x_2y_2 + 3x_3y_2 + 3x_2y_3 + x_3y_3.$$

Show that $<\,>'$ is an inner product on $\mathbb{R}^3$. Deduce that
$| x_1y_1 + x_2y_1 + x_1y_2 + 2x_2y_2 + 3x_3y_2 + 3x_2y_3 + x_3y_3 |$
$\leq \sqrt{x_1^2 + 2x_1x_2 + 2x_2^2 + 6x_2x_3 + x_3^2}\sqrt{y_1^2 + 2y_1y_2 + 2y_2^2 + 6y_2y_3 + y_3^2}$ for all real numbers x_i, y_i. Further, find an orthonormal basis of $(\mathbb{R}^3, <\,>')$, and a linear transformation T from $\mathbb{R}^3$ to itself such that

$$<T(x),\ T(y)>' = <x,\ y>$$

for all $x, y \in \mathbb{R}^3$, where $<\,>$ is the standard Euclidean inner product.

4.4.2 Define a map $<\,>'$ from $\mathbb{R}^2 \times \mathbb{R}^2$ to $\mathbb{R}$ by

$$<(x_1, x_2),\ (y_1, y_2)>' = \; x_1y_1 + x_1y_2 + 2x_2y_1 + y_1y_2.$$

Is $<\,>'$ an inner product? Support.

4.4.3 Let $P_3(x)$ denote the vector space of all polynomials of degree at most 3 with coefficients in the field $\mathbb{R}$ of real numbers. Let $f(x)$ and $g(x)$ be elements of $P_3(x)$. Define $<f(x),\ g(x)> = \int_0^1 f(x)g(x)dx$. Show that $<,\ >$ defines an inner product on $P_3(x)$. Find an orthonormal basis of $P_3(x)$.

4.4.4* Let V be the vector space of all real valued smooth functions on $\mathbb{R}$ and

$$T \;=\; D^3 \;-\; 6D^2 \;+\; 11D \;-\; 6I$$

where $D = \frac{d}{dx}$ is the standard differential operator. Show that $ker\,T$ is 3 dimensional inner product space with respect to the inner product

$$<f,\ g> = \int_{-1}^{1} f(x)g(x)dx.$$

Find an orthonormal basis of $kerT$.

4.4.5 Let A be a non singular 3×3 matrix with real entries. Show, by means of an example, that $< \overline{x},\ \overline{y} > = \overline{x}A\overline{y}^t$ need not define an inner product on $\mathbb{R}^3$. Show that it defines an inner product if and only if $A = BB^t$ for some non singular matrix. Such a matrix is called a **positive definite symmetric matrix**

4.4.6 Let $C[0, 1]$ denote the vector space of all continuous functions on the closed interval $[0, 1]$. Show that $< f(x),\ g(x) > = \int_0^1 f(x)g(x)dx$ defines an inner product on $C[0, 1]$. Show that the set $\{1\} \bigcup \{\sqrt{2}sinn\pi x, \sqrt{2}cosn\pi x \mid n \in \mathbb{Z}\}$ forms an orthonormal set.

4.4.7 Find the largest value and also the smallest value of $3x - 4y + 2z$, if exists, on the sphere $x^2 + y^2 + z^2 = 4$, and also on the Ellipsoid $x^2 + 2y^2 + 3z^2 = 1$. Do they exist on a paraboloid, or a hyperboloid?

4.4.8 Consider the standard real inner product space $\mathbb{R}^4$. Use the Gram–Schmidt process to determine an orthonormal basis of the subspace W generated by $\{(1, 1, 1, 1), (1, 2, 2, 2), (1, 2, 3, 3), (1, 0, 0, 0)\}$. What is the dimension of W?

4.4.9 Find a lower triangular matrix P with positive entries, if possible, so that PA is an orthogonal matrix, where

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{bmatrix}$$

Also express A as $A = LO(UO)$, where L is a lower (U an upper) triangular matrix with positive entries, and O is an orthogonal matrix.

4.4.10 Let W be a subspace of the standard Euclidean inner product space $\mathbb{R}^4$ generated by $\{(1, -1, 1, -1), (1, 1, 1, -3), (1, 2, -3, 0)\}$. Find the distance of $(1, 1, 1, 1)$ from W, and also the foot of the perpendicular from $(1, 1, 1, 1)$ to W.

4.4.11 Consider the usual inner product space $\mathbb{R}^3$. Find the shortest distance between the line

$$\frac{x}{1} = \frac{y-2}{2} = \frac{z}{1},$$

and

$$\frac{x-1}{1} = \frac{y-2}{1} = \frac{z-3}{1},$$

and also find the line of shortest distance, if it exists.

4.4.12 In the standard inner product space $\mathbb{R}^4$, find shortest distance between $\{(x, y, z, w) \mid x + y + z + w = 0 = x - y = z - w\}$ and $\{(x, y, z, w) \mid x + z + 1 = 0\}$.

4.4.13 Consider the subspace $W = \{(x, y, z, w) \mid x + 2y + z + w = 0 = x + y - 2z = w\}$ of the standard inner product space $\mathbb{R}^4$. Find an orthonormal basis of W, and also of $W^{\perp}$. Find also the component of $(1, 1, 1, 1)$ along W and perpendicular to W.

4.4.14 Let T be the linear transformation from the standard inner product space $\mathbb{R}^3$ to itself defined by

$$T((x, y, z)) = (x + y, y + z, z + x).$$

Find $T^{\star}$, and also find its matrix representation with respect to the standard basis.

4.4.15 Let W be a subspace of a finite-dimensional inner product space V, and $x \in V$ be such that $< x,\ y > + < y,\ x > \leq < y,\ y >$ for all $y \in W$. Show that $x \in W^{\perp}$.

4.4.16 Let T be a normal linear transformation from V to V such that $T(x) = 0$. Show that $T^{\star}(x) = 0$.

4.4.17 Let A be a bounded subset of a fine-dimensional real inner product space V in the sense that there is a real number M such that $|| x || \leq M$ for all $x \in A$. Show that $\{f \in M(V) \mid f(A) \subseteq A\}$ is a subgroup of $O(V)$.

4.4.18 Let $T \in End(V)$, where V is a complex inner product space. Suppose that $T^{\star}T(x) = 0$. Show that $T(x) = 0$.

4.4.19 Let T be a Hermitian linear transformation on a finite dimensional complex inner product space V. Let $x \in V$ be such that $T^m(x) = 0$ for some $m \geq 1$. Show that $T(x) = 0$.

4.4.20 Let V be a real inner product space of dimension n, and $\{x_1, x_2, \ldots, x_n\}$ an orthonormal basis. Define a map η from S_n to $O(V)$ as follows. Let $p \in S_n$. Then $\eta(p)$ is the unique orthogonal linear transformation whose effect on the orthonormal basis is given by $\eta(p)(x_i) = x_{p(i)}$. Show that η is an injective homomorphism. Deduce that every group of order n is isomorphic to a subgroup of $O(V)$.

4.4.21 Show that every group of order n is isomorphic to a subgroup of U_n.

4.4.22 Show that every finite subgroup of $M(V)$ is a subgroup of $O(V)$.

4.4.23 Find all finite subgroups of $O(2)$.

4.4.24 Show that there is no proper open subspace of an inner product space.

4.4.25 Show that every nonempty open subset of an inner product space generates the space.

4.4.26 Show that every unit sphere $\{x \mid || x || = 1\}$ generates the space.

4.4.27 Show that every subspace of an inner product space is closed.

4.4.28 Show that every linear transformation from a finite dimensional inner product space to an inner product space is continuous.

4.4.29 Show that inner product map is also continuous.

4.4.30 Let W be a subspace of a finite dimensional inner product space V. Show that V/W is a inner product space with respect to the inner product defined by $< x + W,\ y + W > = g.l.b.\{< x + u,\ y + v > |\ u, v \in W\}$.

4.4.31 Let A be a skew-Hermitian transformation. Show that $I + A$ and $I - A$ are isomorphism.

4.4.32 Let V be a finite dimensional real inner product space of dimension n. Define an inner product $< >$ on $End(V)$ by

$$< T,\ T' > = \Sigma_{i,j}\alpha_{ij}\beta_{ij},$$

where $T = \Sigma_{ij}\alpha_{ij}T_{ij}$ and $T' = \Sigma_{ij}\beta_{ij}T_{ij}$. Find the distance between the subspace $S(V)$ and $SS(V)$.

4.4.33 Check, if the following transformations from $\mathbb{R}^3$ to itself are orthogonal transformations.

(i) $T(x_1, x_2, x_3) = (x_2, x_3, x_1)$
(ii) $T(x_1, x_2, x_3) = (\frac{x_1+x_2}{\sqrt{2}}, \frac{x_1-x_2}{\sqrt{2}}, x_3)$
(iii) $T(x_1, x_2, x_3) = (\frac{x_1+x_2}{\sqrt{2}}, \frac{x_1-x_2+x_3}{\sqrt{3}}, \frac{x_1-x_2-2x_3}{\sqrt{6}})$
(iv) $T(x_1, x_2, x_3) = (x_1 + x_2, x_3, x_1)$.

4.4.34 Check, if the following transformations from $\mathbb{R}^3$ to itself are rigid motions.

(i) $T(x_1, x_2, x_3) = (x_2 + 1, x_3 + 2, x_1)$
(ii) $T(x_1, x_2, x_3) = (\frac{x_1+x_2}{\sqrt{2}}, \frac{x_1-x_2}{\sqrt{2}}, x_3)$
(iii) $T(x_1, x_2, x_3) = (\frac{x_1+x_2}{\sqrt{2}} + 1, \frac{x_1-x_2+x_3}{\sqrt{3}}, \frac{x_1-x_2-2x_3}{\sqrt{6}})$
(iv) $T(x_1, x_2, x_3) = (x_1 + x_2, x_3, x_1)$.

4.4.35 Let V be a real inner product space of dimension n and $x \in V, x \neq 0$. Then show that $P_x = \{y \in V |< x,\ y > = 0\}$ is a hyperplane in V. Show that $P_x = P_y$ if and only if $x = \alpha y$ for some $\alpha \neq 0$. Further, show that every hyperplane is of the form P_x for some x. Determine a bijective correspondence between the set of lines and the set of hyperplanes.

4.4.36 Let P_x be the hyperplane determined by an element x in a real inner product space V as described in the above exercise. Let σ_x be a linear transformation on V which fixes the elements of P_x and maps a vector orthogonal to P_x to its negative. Show that σ_x is uniquely defined, and it is given by

$$\sigma_x(y) = y - \frac{2 < y, x >}{< x, x >} x.$$

This linear transformation is called the **reflection** about the hyperplane P_x. Observe that $P_x = P_y$ if and only if $\sigma_x = \sigma_y$. Deduce that every reflection is an orthogonal linear transformation whose square is identity. Further, find the matrix representation of σ_x with respect to the standard basis.

4.4.37 Let Φ be a finite set of generators of a real inner product space V such that $0 \notin \Phi$. Let $\sigma \in GL(V)$ be such that

(i) $\sigma(\Phi) = \Phi$
(ii) σ fixes element wise a hyperplane P
(iii) $\sigma(\alpha) = -\alpha$ for some $\alpha \in \Phi$.

Show that $\sigma = \sigma_\alpha$ and $P = P_\alpha$.

4.4.38$*$ Let Φ be as in the above exercise such that

(i) $\alpha \in \Phi$ *implies that* $-\alpha \in \Phi$
(ii) $\sigma_\alpha(\Phi) = \Phi$ *for all* $\alpha \in \Phi$
(iii) $\frac{2<\beta, \alpha>}{<\alpha, \alpha>} \in \mathbb{Z}$ *for all* $\alpha, \beta \in \Phi$.

Show that $\Phi \bigcap < \alpha > = \{\frac{1}{2}\alpha, -\frac{1}{2}\alpha, \alpha, -\alpha, 2\alpha, -2\alpha\}$ for all $\alpha \in \Phi$.

39$*$ In the above exercise, if in addition, $\Phi \bigcap < \alpha > = \{\alpha, -\alpha\}$, then Φ is called a **root system**. Let Φ be a root system and $\alpha, \beta \in \Phi$. Show that the angle between α and β is one of the following: $0, \frac{\pi}{2}, \frac{\pi}{3}, \frac{2\pi}{3}, \frac{\pi}{4}, \frac{3\pi}{4}, \frac{\pi}{6}, \frac{5\pi}{6}$. Determine the ratio of their lengths in each case.

40$*$ Determine all roots in $\mathbb{R}^2$.

Chapter 5
Determinants and Forms

In this chapter, we introduce the concept of determinant in various ways. The invariant subspaces, the eigen values, the spectral theorem, the geometry of orthogonal transformations, and the geometry of bilinear forms also constitute the subject matter of this paper.

5.1 Determinant of a Matrix

We define determinant $det(A)$ of a $n \times n$ matrix A by the induction on n as follows:

Let $A = [a_{ij}]$ be a $n \times n$ matrix. Let A_{ij} denote the $(n-1) \times (n-1)$ submatrix obtained by deleting the ith row and the jth column of the matrix.

Example 5.1.1 For the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix},$$

$$A_{11} = \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix},$$

and

$$A_{23} = \begin{bmatrix} 1 & 2 \\ 7 & 8 \end{bmatrix}$$

Definition 5.1.2 If $A = [a_{11}]$ is a 1×1 matrix, then define $det(A) = a_{11}$. Assuming that the determinant of all $m \times m$ matrices, $m < n$ has already been defined, define the determinant of a $n \times n$ matrix A by

R. Lal, *Algebra 2*, Infosys Science Foundation Series in Mathematical Sciences,
DOI 10.1007/978-981-10-4256-0_5

$$det(A) = \Sigma_{i=1}^{n}(-1)^{i+1}a_{i1}det(A_{i1}).$$

Thus,

$$det\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = (-1)^{1+1}a_{11}detA_{11} + (-1)^{2+1}a_{21}detA_{21} = a_{11}a_{22} - a_{21}a_{12},$$

and

$$det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$= (-1)^{1+1}a_{11}det\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} + (-1)^{2+1}a_{21}det\begin{bmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{bmatrix} + (-1)^{3+1}a_{31}det\begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix}$$

$$= a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{21}(a_{12}a_{33} - a_{32}a_{13}) + a_{31}(a_{12}a_{23} - a_{22}a_{13}).$$

Theorem 5.1.3 *The determinant map det satisfies the following properties:*
(i) det is linear in each row of the matrix. More precisely,

$$det\begin{bmatrix} \overline{r_1} \\ \overline{r_2} \\ \cdot \\ \cdot \\ a\overline{r_i} + b\overline{r_i'} \\ \cdot \\ \cdot \\ \overline{r_n} \end{bmatrix} = adet\begin{bmatrix} \overline{r_1} \\ \overline{r_2} \\ \cdot \\ \cdot \\ \overline{r_i} \\ \cdot \\ \cdot \\ \overline{r_n} \end{bmatrix} + bdet\begin{bmatrix} \overline{r_1} \\ \overline{r_2} \\ \cdot \\ \cdot \\ \overline{r_i'} \\ \cdot \\ \cdot \\ \overline{r_n} \end{bmatrix}$$

for each i.
(ii) If two distinct rows of a matrix A are same, then $det(A) = 0$.
(iii) $det(I_n) = 1$.

Proof (i) We prove (i) by the induction on n. For $n = 1$, $det([aa_{11}]) = aa_{11} = adet([a_{11}])$, and there is nothing to do. Assume that the result is true for matrices of order less than n. Let A be a matrix of order n. Consider a general term $(-1)^{i+1}a_{i1}det(A_{i1})$ under summation in the definition of $det(A)$. If $k \neq i$, then a_{i1} does not depend on kth row, whereas by the induction hypothesis, $det(A_{i1})$ depends linearly on kth row. However, a_{k1} depends linearly on kth row, whereas $det(A_{k1})$ is independent of kth row. This shows *det* is linear in each row.

(ii) Suppose that the kth row $\overline{r_k}$ is same as the lth row $\overline{r_l}$ of A. Suppose that $k < l$. If $i \neq k$ and $i \neq l$, then two rows of A_{i1} will be same, and so by the induction hypothesis $det(A_{i1}) = 0$. Thus, all terms under summation of the R.H.S. of the expression for $det(A)$ are zero except perhaps $(-1)^{k+1}a_{k1}det(A_{k1})$ and $(-1)^{l+1}a_{l1}det(A_{l1})$. Hence, then, $det(A)$ equals

$$(-1)^{k+1}a_{k1}det(A_{k1}) + (-1)^{l+1}a_{l1}det(A_{l1}).$$

Observe that $a_{k1} = a_{l1}$. Next, A_{k1} can be obtained from A_{l1} by the interchange of $l - k - 1$ consecutive rows. This means that $det(A_{k1}) = (-1)^{l-k-1}det(A_{l1})$. Substituting it we see that the $det(A) = 0$.

(iii) Finally if $A = I_n$, then the only term which appears in the expression for the $det(I_n)$ is $(-1)^{1+1}det(I_{n-1}) = 1$. ♯

Remark 5.1.4 Imitating the proof of the above theorem, one can easily observe that the associations $A \mapsto \Sigma_{i=1}^{n}(-1)^{i+j}a_{ij}det(A_{ij})$, $A \mapsto \Sigma_{j=1}^{n}(-1)^{i+j}a_{ij}det(A_{ij})$ from the set of square matrices to the field F also satisfy the three properties listed in the theorem. We shall show (see Corollary 5.3.16) that an association from M_n to F satisfying the listed 3 properties in the above theorem is unique. As such, it will follow that $det(A) = \Sigma_{i=1}^{n}(-1)^{i+j}a_{ij}det(A_{ij}) = \Sigma_{j=1}^{n}(-1)^{i+j}a_{ij}det(A_{ij})$. Thus, we can expand the determinant from any row, or from any column. In turn, $det(A) = det(A^t)$.

Corollary 5.1.5 $\Sigma_{j=1}^{n}(-1)^{k+j}a_{ij}det(A_{kj}) = 0$ *for all* $i \neq k$.

Proof Suppose that $i \neq k$. Consider the matrix $B = [b_{pq}]$, where $b_{pq} = a_{pq}$ if $p \neq k$ and $b_{kq} = a_{iq}$. In other words B is obtained by replacing the kth row of A by the ith row. Thus, deleting the kth row and jth column of A is same as doing the same thing on B. More precisely, $A_{kj} = B_{kj}$. The expression $\Sigma_{j=1}^{n}(-1)^{k+j}a_{ij}det(A_{kj})$ becomes $\Sigma_{j=1}^{n}(-1)^{k+j}b_{kj}det(B_{kj})$. Since two distinct rows of B are same, it follows, from the above remark, that $\Sigma_{j=1}^{n}(-1)^{k+j}b_{kj}det(B_{kj}) = 0$. The result follows. ♯

Definition 5.1.6 Let A be a square matrix of order n. Then $(-1)^{i+j}det(A_{ij})$ is called the (**i**, **j**) **co-factor** of A. The matrix $A^{cof} = [(-1)^{i+j}det(A_{ij})]$ is called the **co-factor** matrix of A. The transpose $(A^{cof})^t$ of the co-factor matrix is called the **adjoint** of A, and it is denoted by A^{adj}. Thus, $A^{adj} = [b_{ij}]$, where $b_{ij} = (-1)^{j+i}det(A_{ji})$.

Corollary 5.1.7 $A \cdot A^{adj} = det(A)I_n = A^{adj} \cdot A$.

Proof Suppose that $A \cdot A^{adj} = [c_{ij}]$. Then

$$c_{ij} = \Sigma_{k=1}^{n}a_{ik}b_{kj} = \Sigma_{k=1}^{n}a_{ik}(-1)^{j+k}det(A_{jk}).$$

From the above theorem, it follows that $c_{ij} = 0$ if $i \neq j$, and it is $detA$ if $i = j$. This proves that $A \cdot A^{adj} = det(A)I_n$. It also follows that $A^t \cdot (A^t)^{adj} = det(A^t)I_n = det(A)I_n$. Further, it is also clear that $(A^{adj})^t = (A^t)^{adj}$. Taking the transpose of the equality $A \cdot A^{adj} = det(A)I_n$, we see that $(A^{adj}A)^t = A^t(A^{adj})^t = A^t(A^t)^{adj} = det(A^t)I_n = det(A)I_n$. This shows that $A^{adj} \cdot A = det(A)I_n$. ♯

Remark 5.1.8 We shall see a little later that $det(AB) = detAdetB$. Since $A \cdot A^{adj} = det(A)I_n$, it follows that if $det(A) \neq 0$, then $det(A^{adj}) = (det(A))^{n-1}$.

Corollary 5.1.9 *A is invertible if and only if $det(A) \neq 0$, and then $A^{-1} = (det(A))^{-1}A^{adj}$.*

Proof Suppose that A is invertible, and $A^{-1}A = I_n$. Then $1 = det(A^{-1} \cdot A) = detA^{-1} \cdot detA$. Hence $detA \neq 0$. Conversely, if $det(A) \neq 0$, then since $A \cdot A^{adj} = det(A)I_n$, we have $A \cdot (det(A))^{-1}A^{adj} = I_n$. ♯

The above result gives us another method to find the inverse of a matrix. This we illustrate by means of the following example.

Example 5.1.10 Consider the matrix A given by

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

Here

$$(-1)^{1+1}Det(A_{11}) = 2,\ (-1)^{1+2}Det(A_{12}) = 0 = (-1)^{1+3}Det(A_{13}), (-1)^{2+1}Det(A_{21}) = -2,\ (-1)^{2+2}Det(A_{22}) = 2,\ (-1)^{2+3}Det(A_{23}) = 0 = (-1)^{3+1}Det(A_{31}),\ (-1)^{3+2}Det(A_{32}) = -1 \text{ and } (-1)^{3+3}Det(A_{33}) = 1.$$

Thus, the co-factor matrix is

$$\begin{bmatrix} 2 & 0 & 0 \\ -2 & 2 & 0 \\ 0 & -1 & 1 \end{bmatrix}.$$

The adjoint A^{adj}, being the transpose of the co-factor matrix, is

$$\begin{bmatrix} 2 & -2 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Clearly, the $det(A) = 2$, and so the inverse of A is

$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & \frac{1}{2} \end{bmatrix}.$$

Corollary 5.1.11 (Cramer's rule) *Consider a system of n-linear equations in n unknowns given by the matrix equation*

$$A \cdot X = B,$$

where A is invertible matrix. Then $x_i = \Sigma_{j=1}^{n}(-1)^{i+j}b_j\frac{det(A_{ji})}{det(A)}$ for all i.

Proof Since $A^{-1} = (det(A))^{-1}A^{adj}$, it follows that $X = \frac{1}{det(A)}A^{adj} \cdot B$. The result follows if we equate the rows. ♯

5.2 Permutations

This is a brief section on permutations with aim to introduce even and odd permutations. For detail, one may refer to algebra 1. A bijective map p from $\{1, 2, \ldots, n\}$ to itself is called a permutation on $\{1, 2, \ldots, n\}$. The set S_n of all permutations on $\{1, 2, \ldots, n\}$ is a group with respect to the composition of maps (product of permutations). We may represent an element $p \in S_n$ (without any ambiguity) by

$$\begin{pmatrix} 1 & 2 & \ldots. & n \\ p(1) & p(2) & \ldots. & p(n) \end{pmatrix}.$$

Since p is a bijective map, the second row is just the rearrangement (permutation) of $1, 2, \ldots, n$. Thus, any $p \in S_n$ gives a unique permutation described above. Conversely, if we have a rearrangement of $1, 2, \ldots, n$, then it gives rise to a unique bijective map from $\{1, 2, \ldots, n\}$ to itself by putting the rearrangement below $12 \cdots n$ as above. For example, if $n = 4$, the rearrangement 2314 of 1234 gives rise to a bijective map p from $\{1, 2, 3, 4\}$ to itself given by $p(1) = 2$, $p(2) = 3$, $p(3) = 1$ and $p(4) = 4$. In the above notation

$$p = \begin{pmatrix} 1\ 2\ 3\ 4 \\ 2\ 3\ 1\ 4 \end{pmatrix}.$$

Thus, the members of S_n can be viewed as permutations. The product gf of permutations

$$f = \begin{pmatrix} 1 & 2 & \ldots. & n \\ f(1) & f(2) & \ldots. & f(n) \end{pmatrix}$$

and

$$g = \begin{pmatrix} 1 & 2 & \ldots. & n \\ g(1) & g(2) & \ldots. & g(n) \end{pmatrix}$$

is given by

$$gf = \begin{pmatrix} 1 & 2 & \ldots. & n \\ g(f(1)) & g(f(2)) & \ldots. & g(f(n)) \end{pmatrix}.$$

Example 5.2.1 If

$$p = \begin{pmatrix} 1\ 2\ 3\ 4 \\ 3\ 4\ 1\ 2 \end{pmatrix}$$

and

$$q = \begin{pmatrix} 1\ 2\ 3\ 4 \\ 2\ 1\ 3\ 4 \end{pmatrix}$$

then

$$pq = \begin{pmatrix} 1\ 2\ 3\ 4 \\ 4\ 3\ 1\ 2 \end{pmatrix}$$

and

$$qp = \begin{pmatrix} 1\ 2\ 3\ 4 \\ 3\ 4\ 2\ 1 \end{pmatrix}.$$

Thus, $pq \neq qp$.

Cycles and Transpositions

Now, we consider special types of permutations. Consider, for example, the permutation

$$p = \begin{pmatrix} 1\ 2\ 3\ 4\ 5\ 6 \\ 2\ 5\ 3\ 1\ 6\ 4 \end{pmatrix}.$$

p takes 1 to 2, 2 to 5, 5 to 6, 6 to 4 and 4 to 1. The remaining symbol 3 is fixed. We can faithfully represent the permutation p by the row (1 2 5 6 4) with the understanding that each symbol goes to the following symbol, the last symbol is mapped to the first symbol, and the symbol not appearing in the row is kept fixed. Thus, the permutation

$$\begin{pmatrix} 1\ 2\ 3\ 4\ 5\ 6\ 7 \\ 1\ 5\ 2\ 4\ 7\ 6\ 3 \end{pmatrix}$$

can be represented by (2 5 7 3), whereas

$$\begin{pmatrix} 1\ 2\ 3\ 4 \\ 2\ 1\ 4\ 3 \end{pmatrix}$$

cannot be represented in this form.

Definition 5.2.2 A permutation $p \in S_n$ is called a **cycle** of length $r \geq 1$ if there exists a subset $\{i_1, i_2, \ldots, i_r\}$ of $\{1, 2, \ldots, n\}$ containing r distinct elements such that $p(i_1) = i_2, p(i_2) = i_3, \ldots, p(i_{r-1}) = i_r, p(i_r) = i_1$, and $p(j) = j$ for all $j \notin \{i_1, i_2, \ldots, i_r\}$. The cycle p is denoted by $(i_1 i_2 \cdots i_r)$. A cycle of length 2 is called a **transposition**. Thus, a transposition is represented by $(i\ j)$ which interchanges i and j, and keeps the rest of the symbols fixed.

Theorem 5.2.3 *Every nonidentity permutation can be written as product of disjoint cycles. Further, any two representations of a nonidentity permutation as product of disjoint cycles are same up to a rearrangement of the cycles.*

Proof Let p be a nonidentity permutation in S_n. Then *there exists* i_1 such that $p(i_1) \neq i_1$. Clearly, all members of the set $\{i_1, p(i_1), p^2(i_1, \ldots, p^n(i_1)\}$ cannot be distinct. Hence, there exist r, s; $1 \leq r < s \leq n$ such that $p^r(i_1) = p^s(i_1)$. Thus, there exists t, $1 < t \leq n$ such that $p^t(i_1) = i_1$. Let l_1 be the least positive integer such that $p^{l_1}(i_1) = i_1$. Given $m \in \mathbb{Z}$, by the division algorithm, *there exist* q, r such that $m = l_1 q + r$, where $0 \leq r \leq l_1 - 1$. But, then $p^m(i_1) = p^r(i_1)$. It is clear from the above observation that the effect of the permutation p on the symbols in $\{i_1, p(i_1), p^2(i_1), \ldots, p^{l_1-1}(i_1)\}$ is the same as that of the cycle $C_1 = (i_1\ p(i_1)\ p^2(i_2)\ \cdots p^{l_1-1}(i_1))$. If $p = C_1$, then there is nothing to do. If not, there exists $i_2 \notin \{i_1, p(i_1), p^2(i_1), \ldots, p^{l_1-1}(i_1)\}$ such that $p(i_2) \neq i_2$. As before, consider the cycle $C_2 = (i_2\ p(i_2)\ p^2(i_2)\ \cdots p^{l_2-1}(i_2))$, where l_2 is the smallest positive integer such that $p^{l_2}(i_2) = i_2$. Clearly, C_1 and C_2 are disjoint cycles. If $p = C_1C_2$, then there is nothing to do. If not, proceed. This process stops after finitely many steps giving p as product of disjoint cycles, because the symbols are finitely many.

Finally, we prove the uniqueness. Suppose that $p \neq I$ and

$$p = C_1C_2 \cdots C_r = C_1'C_2' \cdots C_s',$$

where C_i and C_j are disjoint for $i \neq j$, and also C_k' and C_l' are disjoint for $k \neq l$. Suppose that $p(t) \neq t$. Then *there exist* i, k such that $C_i(t) \neq t$, and also $C_k'(t) \neq t$. We may assume that $C_1(t) \neq t$ and $C_1'(t) \neq t$. But, then using the arguments of the previous paragraph, we find that $C_1 = C_1'$. Canceling C_1 and C_1', using induction, and the fact that products of nonidentity disjoint cycles can never be identity, we find that $r = s$, and $C_i = C_i'$ for all i. ♯

Remark 5.2.4 The proof of the above theorem is algorithmic, and it gives an algorithm to express a permutation as product of disjoint cycles.

Proposition 5.2.5 *Every cycle is product of transpositions.*

Proof $(i_1 i_2 \cdots i_r) = (i_1 i_r)(i_1 i_{r-1}) \cdots (i_1 i_2)$. ♯

Since every permutation is product of disjoint cycles, we have the following corollary.

Corollary 5.2.6 *Every permutation is product of transpositions.* ♯

Remark 5.2.7 Representation of a permutation as product of transpositions is not unique. For example,

$$(1234) = (14)(13)(12) = (14)(13)(12)(24)(24) = (14)(23)(13).$$

Alternating Map

Let $p \in S_n$. Consider the following rational number:

$$\frac{p(1)-p(2)}{1-2}\frac{p(1)-p(3)}{1-3}\cdots\frac{p(1)-p(n)}{1-n}\cdot\frac{p(2)-p(3)}{2-3}\cdots\frac{p(2)-p(n)}{2-n}\cdots\frac{p(n-1)-p(n)}{(n-1)-n}.$$

The above expression in short is denoted by

$$\prod_{1\le i<j\le n} \frac{p(i)-p(j)}{i-j}.$$

Proposition 5.2.8 $\prod_{1\le i<j\le n} \frac{p(i)-p(j)}{i-j} = \pm 1$ *for all* $p \in S_n$.

Proof Since p is a permutation, for all pair $(k, l), k \neq l$, there is a unique pair $(i, j), i \neq j$ with $p(i) = k$ and $p(j) = l$. If $i < j$, then $k - l$ appears once and only once in the numerator of the expression, and if $j < i$, then $l - k$ appears once and only once in the numerator of the expression. Also $k - l$ or $l - k$ appears once and only once in the denominator according as $k < l$ or $l < k$. This proves the result. ♯

Definition 5.2.9 The map χ from S_n to $\{1, -1\}$ defined by

$$\chi(p) = \prod_{1\le i<j\le n} \frac{p(i)-p(j)}{i-j}$$

is called the **alternating map** of degree n.

Theorem 5.2.10 *The alternating map* $\chi : S_n \longrightarrow \{1, -1\}$ *is a surjective map which takes any transposition to* -1. *Further, it is a homomorphism in the sense that* $\chi(pq) = \chi(p)\chi(q)$ *for all* $p, q \in S_n$.

Proof We first show that χ is a homomorphism.

$$\chi(pq) = \prod_{1\le i<j\le n} \frac{pq(i)-pq(j)}{i-j} = \prod_{1\le i<j\le n} \frac{p(q(i))-p(q(j))}{q(i)-q(j)} \prod_{1\le i<j\le n} \frac{q(i)-q(j)}{i-j} = \prod_{1\le i<j\le n} \frac{p(q(i))-p(q(j))}{q(i)-q(j)} \cdot \chi(q).$$

Since q is a permutation of $1, 2, \ldots n$, it follows that

$$\prod_{1\le i<j\le n} \frac{p(q(i))-p(q(j))}{q(i)-q(j)} = \chi(p).$$

This shows that χ is a homomorphism. Hence $\chi(p)\chi(p^{-1}) = \chi(I) = 1$ for all permutation p. Consider the transposition $\tau = (1, 2)$. Clearly,

$$\chi(\tau) = \frac{2-1}{1-2} \cdot \frac{2-3}{1-3} \cdots \frac{2-n}{1-n} \cdot \frac{1-3}{2-3} \cdot \frac{1-4}{2-4} \cdots \frac{1-n}{2-n} = -1.$$

Consider a general transposition $\sigma = (k, l)$. Take a permutation $p \in S_n$ for which $p(1) = k$, $p(2) = l$. Observe that such a permutation exists. Then $p\tau p^{-1} = \sigma$. Hence $\chi(\sigma) = \chi(p)\chi(\tau)\chi(p^{-1}) = -1$. Thus, χ takes any transposition to -1. ♯

Corollary 5.2.11 *Let $p \in S_n$. Suppose that*

$$p = \sigma_1\sigma_2\cdots\sigma_r = \tau_1\tau_2\cdots\tau_s,$$

where σ_i and τ_j are transpositions. Then $r \equiv s(mod 2)$, i.e., 2 divides $r - s$ (equivalently r and s both are simultaneously even, or both are simultaneously odd).

Proof From the above theorem, it follows that

$$\chi(p) = \chi(\sigma_1)\chi(\sigma_2)\cdots\chi(\sigma_r) = \chi(\tau_1)\chi(\tau_2)\cdots\chi(\tau_s).$$

Since χ takes a transposition to -1, $(-1)^r = (-1)^s$. Hence $r - s$ is even. ♯

Remark 5.2.12 From the above corollary, it follows that if we can write a permutation as a product of even number of transpositions, then we cannot write it as a product of odd number of transpositions, and if we can write it as product of odd number of transpositions, then we cannot write it as a product of even number of transpositions.

Definition 5.2.13 A permutation p is called an **even permutation**, if it can be expressed as product of even number of permutations, or equivalently, $\chi(p) = 1$. It is said to be an **odd permutation** if it can be expressed as product of odd number of transpositions. We also say that $sign(p) = 1$, if p is an even permutation, and $sign(p) = -1$ if p is an odd permutation. Thus, $\chi(p)$ is also written as $sign(p)$. The set A_n of all even permutations is a subgroup of S_n called the **alternating** group.

Example 5.2.14 Consider the permutation p given by

$$p = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 5 & 1 & 3 \end{pmatrix}.$$

Then $p = (124)(35) = (12)(14)(35)$ is product of 3 transpositions. Hence p is an odd permutation and so $sign(p) = \chi(p) = -1$.

Proposition 5.2.15 *Let τ be a transposition in S_n. Then $A_n\tau = \{p\tau \mid p \in A_n\}$ and A_n are disjoint and $S_n = A_n \bigcup A_n\tau$.*

Proof Follows from the fact that $A_n\tau$ is the set of all odd permutations, whereas A_n is the set of all even permutations. ♯

5.3 Alternating Forms, Determinant of an Endomorphism

Let $V_1, V_2, \ldots, V_r$ and W be vector spaces over a field F. A map f from $V_1 \times V_2 \times \cdots \times V_r$ to W is called a **multilinear map** if

$$\begin{aligned} &f(x_1, x_2, \ldots, x_{i-1}, ax_i + bx_i', x_{i+1}, \ldots, x_r) \\ = \; &af(x_1, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_r) + bf(x_1, x_2, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_r) \end{aligned}$$

for all $a, b \in F, x_j \in V_j$, and $x_i' \in V_i$. Thus, a multilinear map is a map which is linear in each coordinate. If $V_i = V$ for each i, then it is called a r-**linear** map on V. If in addition $W = F$, then it is said to be a r-**linear form** on V. A 2-linear form on V is also called a **bilinear** form on V

The vector product on $\mathbb{R}^3$ is a bilinear map from $\mathbb{R}^3 \times \mathbb{R}^3$ to $\mathbb{R}^3$, and the scalar product on $\mathbb{R}^3$, or in general, an inner product on a real vector space V is a bilinear form on V.

Definition 5.3.1 A r-multilinear map f on a vector space V is called r-**alternating** map if $f(x_1, x_2, \ldots, x_r) = 0$ whenever $x_i = x_j$ for some $i \neq j$.

The vector product on $\mathbb{R}^3$ is a $2-$alternating map. The map f from $\mathbb{R}^2 \times \mathbb{R}^2$ to $\mathbb{R}$ given by $f((a_1, a_2), (b_1, b_2)) = a_1b_2 - a_2b_1$ is a 2-alternating form on $\mathbb{R}^2$. The map $(\overline{a}, \overline{b}, \overline{c}) \mapsto (\overline{a} \times \overline{b}) \cdot \overline{c}$ defines 3-alternating form (called the volume form) on $\mathbb{R}^3$.

The sum of two r-alternating maps from V to W is a r-alternating map (verify), and the scalar multiple of a r-alternating map is also a r-alternating map. Thus, the set $A_r(V, W)$ of all r-alternating maps form a vector space with respect the above operations.

Next, let T be a linear transformation from V to W. The map T^r from V^r to W^r defined by $T^r(x_1, x_2, \ldots, x_r) = (T(x_1), T(x_2), \ldots, T(x_r))$ is a linear transformation. If f is a r-alternating map from W to U, and T a linear transformation from V to W, then foT^r defines a r-alternating map from V to U(verify). This defines a linear transformation $A_r(T)$ from $A_r(W, U)$ to $A_r(V, U)$. The following properties can be verified easily.

1. $A_r(I_V) = I_{A_r(V,V)}$.

2. $A_r(T_2oT_1) = A_r(T_1)oA_r(T_2)$, where T_1 is a linear transformation from W_1 to W_2 and T_2 is a linear transformation from W_2 to W_3.

In particular, A_r defines a linear transformation from $End(V)$ to $End(A_r(V))$.

Proposition 5.3.2 *Let f be a r-alternating form on V, and $\{x_1, x_2, \ldots, x_r\}$ a linearly dependent set. Then $f(x_1, x_2, \ldots, x_r) = 0$.*

Proof Under the hypothesis, there is an i such that x_i is a linear combination of the rest of the coordinates. Substituting this linear combination at the ith coordinate, expanding, and using the property of being alternative, we get the result. ♯

Corollary 5.3.3 *Let V be a vector space of dimension n. Then the vector space $A_r(V, W) = \{0\}$ for all $r > n$.* ♯

Proposition 5.3.4 *Let f be a r-alternating map on V. Then*

$$f(x_1, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_j, \ldots, x_r) = \\ -f(x_1, x_2, \ldots, x_{i-1}, x_j, x_{i+1}, \ldots, x_{j-1}, x_i, x_{j+1}, \ldots, x_r).$$

In other words, if we interchange two coordinates, then the value of f changes its sign.

Proof From the definition of alternating map, it follows that

$$f(x_1, x_2, \ldots, x_{i-1}, x_i + x_j, x_{i+1}, \ldots, x_{j-1}, x_i + x_j, x_{j+1}, \ldots, x_r) = 0.$$

Expanding, and observing that

$$f(x_1, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_{j-1}, x_i, x_{j+1}, \ldots, x_r) = 0 = \\ f(x_1, x_2, \ldots, x_{i-1}, x_j, x_{i+1}, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_r)$$

the result follows. ♯

The above result can be restated as follows.

Proposition 5.3.5 *Let f be a r-alternating map on V, and τ a transposition in S_r. Then*

$$f(x_{\tau(1)}, x_{\tau(2)}, \ldots, x_{\tau(r)}) = -f(x_1, x_2, \ldots, x_r).$$ ♯

In general, we have the following proposition.

Proposition 5.3.6 *Let f is a r-alternating map from V to W, and $p \in S_r$. Then*

$$f(x_{p(1)}, x_{p(2)}, \ldots, x_{p(r)}) = sign(p) f(x_1, x_2, \ldots, x_r),$$

where $sign(p) = 1$, if p is even permutation, and it is -1 if p is an odd permutation.

Proof If $p = \tau_1 \tau_2 \cdots \tau_m$, then applying the above result successively, we see that

$$f(x_{p(1)}, x_{p(2)}, \ldots, x_{p(r)}) = (-1)^m f(x_1, x_2, \ldots, x_r).$$

The result follows. ♯

Proposition 5.3.7 *Let f be a r-alternating map on V. Let $\{x_1, x_2, \ldots, x_r\}$ and $\{y_1, y_2, \ldots, y_r\}$ be subsets of V. Suppose that $y_j = \Sigma_{i=1}^r a_{ij} x_i$. Then*

$$f(y_1, y_2, \ldots, y_r) = \Sigma_{p \in S_r} sign(p) a_{1p(1)} a_{2p(2)} \cdots a_{rp(r)} f(x_1, x_2, \ldots, x_r) = \\ \Sigma_{p \in S_r} sign(p) \prod_{i=1}^r a_{ip(i)} f(x_1, x_2, \ldots, x_r).$$

Proof $f(y_1, y_2, \ldots, y_r) = f(\Sigma_{i=1}^r a_{i1} x_i, \Sigma_{i=1}^r a_{i2} x_i, \ldots, \Sigma_{i=1}^r a_{ir} x_i)$. Expanding by multilinearity and keeping in mind that the value of f is 0 whenever two arguments are same, we see that

$$f(y_1, y_2, \ldots, y_r) = \Sigma_{p\in S_r} a_{p(1)1} a_{p(2)2} \cdots a_{p(r)r} f(x_{p(1)}, x_{p(2)}, \ldots, x_{p(r)}).$$

Using the above proposition we find that

$$f(y_1, y_2, \ldots, y_r) = \Sigma_{p\in S_r} sign(p) \prod_{i=1}^{r} a_{p(i)i} f(x_1, x_2, \ldots, x_r). \qquad \sharp$$

Corollary 5.3.8 *Let V be a vector space with a basis $\{x_1, x_2, \ldots, x_n\}$ and f a n-alternating form on V. Then f is uniquely determined by its value $f(x_1, x_2, \ldots, x_n)$ on $(x_1, x_2, \ldots, x_n)$.* $\sharp$

Theorem 5.3.9 *Let V be a vector space of dimension n over a field F. Then the dimension of $A_n(V, F)$ is* 1.

Proof Let $\{u_1, u_2, \ldots, u_n\}$ be a basis of V. Then every n-alternating form f is uniquely determined by its value $f(u_1, u_2, \ldots, u_n)$ on $(u_1, u_2, \ldots, u_n)$. Indeed, given $(x_1, x_2, \ldots, x_n)$ such that $x_j = \Sigma_{i=1}^{n} a_{ij} u_i$, $f(x_1, x_2, \ldots, x_n) = \Sigma_{p\in S_n} sign(p) \prod_{i=1}^{n} a_{p(i)i} f(u_1, u_2, \ldots, u_n)$. This shows that the dimension of $A_n(V, F)$ is at most 1 (any two n-alternating map differ by a scalar multiple). It is sufficient, therefore, to show that $A_n(V, F) \neq \{0\}$. We show that the map f defined by

$$f(x_1, x_2, \ldots, x_n) = \Sigma_{p\in S_n} sign(p) \prod_{i=1}^{n} a_{p(i)i},$$

where $x_j = \Sigma_{i=1}^{n} a_{ij} u_i$ is a nonzero n-alternating form on V. Clearly, $f(u_1, u_2, \ldots, u_n) = 1$, for then the matrix $[a_{ij}]$ is the identity matrix, and so $a_{ii} = 1$ for all i and $a_{ij} = 0$ for $i \neq j$. It is clearly an n-linear map. We show that it is alternating. Suppose that $x_j = x_k$, $j \neq k$. Then $a_{ij} = a_{ik}$ for all i. Let p be a permutation and $\tau = (j\ k)$. Then $a_{p(\tau(j))j} = a_{p(k)j} = a_{p(k)k}$ and $a_{p(\tau(k))k} = a_{p(j)k} = a_{p(j)j}$, and for $l \notin \{j, k\}$, $a_{p(\tau(l))l} = a_{p(l)l}$. It follows that $a_{p(1)1} a_{p(2)2} \cdots a_{p(n)n} = a_{p\tau(1)1} a_{p\tau(2)2} \cdots a_{p\tau(n)n}$, and so $sign(p) a_{p(1)1} a_{p(2)2} \cdots a_{p(n)n} = -sign(p\tau) a_{p\tau(1)1} a_{p\tau(2)2} \cdots a_{p\tau(n)n}$. Now S_n is disjoint union of A_n and $A_n\tau = \{p\tau \mid p \in A_n\}$. Hence,

$$\Sigma_{p\in S_n} sign(p) \prod_{i=1}^{n} a_{p(i)i} = \Sigma_{p\in A_n} sign(p) \prod_{i=1}^{n} a_{p(i)i} + \Sigma_{p\tau | p\in A_n} sign(p\tau) \prod_{i=1}^{n} a_{p(i)i} = 0.$$

This shows that $f(x_1, x_2, \ldots, x_n) = 0$ whenever $x_j = x_k$ for some $j \neq k$. Thus, f is alternating. $\sharp$

Let T be an endomorphism of V, where V is a vector space of dimension n over a field F. Then T induces a linear transformation $A_n(T)$ from $A_n(V, F)$ to itself. Since $A_n(V, F)$ is of dimension 1, it follows that $A_n(T)$ is multiplication by a scalar. This scalar is denoted by $det(T)$, and it is called the **determinant** of T. Thus, $A_n(T)(f) = det(T) \cdot f$. This defines a map det from $End(V)$ to F, and it is called the **determinant** map on $End(V)$. Since $A_n(T_1 o T_2) = A_n(T_2) o A_n(T_1)$, we have the following corollary.

Corollary 5.3.10 $det(T_1 o T_2) = det(T_2) \cdot det(T_1) = det(T_1) \cdot det(T_2)$. $\sharp$

Corollary 5.3.11 *Let T be a linear transformation from a vector space V to itself. Let $\{u_1, u_2, \ldots, u_n\}$ be a basis of V. Suppose that $T(u_j) = \Sigma_{i=1}^n a_{ij}u_i$. Then $det(T) = \Sigma_{p\in S_n} sign(p) \prod_{i=1}^n a_{p(i)i}$.*

Proof Let f be a $n-$ alternating form on V. Then by the definition, $A_n(T)(f) = foT^n$. Now,

$(foT^n)(u_1, u_2, \ldots, u_n) = f(T(u_1), T(u_2), \ldots, T(u_n))$

$= f(\Sigma_{i=1}^n a_{i1}u_i, \Sigma_{i=1}^n a_{i2}u_i, \ldots, \Sigma_{i=1}^n a_{in}u_i)$

$= \Sigma_{p\in S_n} sign(p) \prod_{i=1}^n a_{p(i)i} f(u_1), u_2, \ldots, u_n)$

$= det(T)f(u_1, u_2, \ldots, u_n).$

This shows that

$$det(T) = \Sigma_{p\in S_n} sign(p) \prod_{i=1}^n a_{p(i)i}. \qquad \sharp$$

The following corollary is immediate from the above proposition.

Corollary 5.3.12 $det(I_V) = 1.$ $\sharp$

Corollary 5.3.13 *Let T be a linear transformation from V to V. Then T is invertible if and only if $det(T) \neq 0$.*

Proof Suppose that T is invertible. Then $ToT^{-1} = I_V$. Hence $1 = det(ToT^{-1}) = det(T) \cdot det(T^{-1})$. This shows that $det(T) \neq 0$. Conversely, suppose that $det(T) \neq 0$. Let $\{u_1, u_2, \ldots, u_n\}$ be a basis of V. It is sufficient to show that $\{T(u_1), T(u_2), \ldots, T(u_n)\}$ is linearly independent. Suppose not. Then for any $f \in A_n(V)$, $A_n(T)(f)(u_1, u_2, \ldots, u_n) = f(T(u_1), T(u_2), \ldots, T(u_n)) = 0 = det(T)f(u_1, u_2, \ldots, u_n)$. Since there is a $f \in A_n(V)$ such that $f(u_1, u_2, \ldots, u_n) \neq 0$, it follows that $det(T) = 0$. This is a contradiction. $\sharp$

Corollary 5.3.14 $det(T) = det(T^t)$.

Proof Let T be a linear transformation on V and $\{u_1, u_2, \ldots, u_n\}$ a basis of V. Consider the dual basis $\{u_1^\star, u_2^\star, \ldots, u_n^\star\}$. Suppose that $T(u_j) = \Sigma_{i=1}^n a_{ij}u_i$. Then $T^t(u_j^\star) = \Sigma_{i=1}^n b_{ij}u_i^\star$, where $b_{ij} = a_{ji}$. Thus, (see the above corollary)

$$det(T^t) = \Sigma_{p\in S_n} sign(p) \prod_{i=1}^n b_{p(i)i} = \Sigma_{p\in S_n} sign(p) \prod_{i=1}^n a_{ip(i)}.$$

Since $\prod_{i=1}^n a_{p(i)i} = \prod_{i=1}^n a_{ip^{-1}(i)}$, $sign(p) = sign(p^{-1})$, and $p \rightsquigarrow p^{-1}$ is a bijective map on S_n, we have

$$\Sigma_{p\in S_n} sign(p) \prod_{i=1}^n a_{p(i)i} = \Sigma_{p\in S_n} sign(p) \prod_{i=1}^n a_{ip(i)}.$$

The result follows. $\sharp$

Corollary 5.3.15 *Any map f from the set $M_n(F)$ of $n \times n$ matrices with entries in F to F which is linear on each row and which is 0 on matrices with two rows same is uniquely determined by its value $f(I_n)$ on the identity matrix I_n. In fact, $f(A) = \Sigma_{p \in S_n} sign(p) \prod_{i=1}^{n} a_{p(i)i} f(I_n)$.*

Proof Take $V = F^n$ and realize a $n \times n$ matrix A as an element $(\overline{r_1}, \overline{r_2}, \ldots, \overline{r_n})$ of V^n, where $\overline{r_i} = [a_{i1}, a_{i2}, \ldots, a_{in}]$ is the ith row of A. Then $\overline{r_i} = a_{i1}\overline{e_1} + a_{i2}\overline{e_2} + \cdots + a_{in}\overline{e_n}$. Since f is n-alternating, it follows from the above theorem that $f(A) = f(\overline{r_1}, \overline{r_2}, \ldots, \overline{r_n})$
$= \Sigma_{p \in S_n} sign(p) \prod_{i=1}^{n} a_{p(i)i} f(\overline{e_1}, \overline{e_2}, \ldots, \overline{e_n})$
$= \Sigma_{p \in S_n} sign(p) \prod_{i=1}^{n} a_{p(i)i} f(I_n)$. ♯

In particular, we have the following corollary.

Corollary 5.3.16 *We have the unique map from the set $M_n(F)$ of $n \times n$ matrices with entries in F to F given by $A \mapsto \Sigma_{p \in S_n} sign(p) \prod_{i=1}^{n} a_{p(i)i}$ which is linear on each row and which is 0 on matrices with two rows same and which is 1 on I_n.* ♯

The following corollary follows from Theorem 5.1.3 and the above corollary.

Corollary 5.3.17 $det(A) = \Sigma_{p \in S_n} sign(p) \prod_{i=1}^{n} a_{p(i)i}$

Let $A = [a_{ij}]$ be a $n \times n$ matrix with entries in a field F. Then A defines a linear transformation L_A from the vector space F^n to itself by

$$L_A\left(\begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix}\right) = A \cdot \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix}$$

Let $\{e_1, e_2, \ldots, e_n\}$ be the standard basis of F^n (we write the elements of F^n as columns). Then

$$L_A(e_j) = \Sigma_{i=1}^{n} a_{ij} e_i.$$

In other words $M_{e_1, e_2, \ldots, e_n}^{e_1, e_2, \ldots, e_n}(L_A) = A$. It follows that

$$det(L_A) = \Sigma_{p \in S_n} sign(p) \prod_{i=1}^{n} a_{p(i)i} = det(A).$$

To summarize, we list some of the important properties of determinant of matrices/linear transformations. The first 3 properties are the defining properties, and the rest of them are the consequences which are useful in computations and other discussions.

1. $det(I_n) = 1 = det(I_{F^n})$.
2. Determinant is a multilinear map on rows/columns of matrices.
3. Determinant of a matrix is zero whenever two distinct rows/columns are same.

4. $det(A) = det(A^t) = \Sigma_{p\in S_n} sign(p) \prod_{i=1}^{n} a_{ip(i)}$.

The following property of determinant which is a consequence of 3 and 4 is useful in computing the determinant of a matrix. See the example below.

5. Determinant of a matrix does not change if we add a multiple of a row (column) in another row (column).

6. $det(A \cdot B) = det(A) \cdot det(B)$. This follows from the fact that $L_{A\cdot B} = L_A o L_B$, and $det(L_A o L_B) = det(L_A) \cdot det(L_B) = det(A) \cdot det(B)$.

7. A is invertible if and only if $det(A) \neq 0$. This follows from the facts: (i) A is invertible if and only if L_A is invertible, and (ii) L_A is invertible if and only if $Det(A) = Det(L_A) \neq 0$.

8. Determinant of a upper(lower) triangular matrix is product of their diagonal entries: Let $A = [a_{ij}]$ be an upper triangular matrix. Then $a_{ij} = 0$ for $i > j$. Let $p \in S_n$. If p is a nonidentity permutation, then $p(i) > i$ for at least one i. Hence, the term $sign(p) \prod_{i=1}^{n} a_{p(i)i} = 0$ for every nonidentity permutation p. This shows that $det(A) = \prod_{i=1}^{n} a_{ii}$. In particular, determinant of a diagonal matrix is product of the diagonal entries. $det(aI_n) = a^n . det(E_{ij}^{\lambda}) = 1$.

9. Determinant of the permutation matrix A^p determined by the permutation p is $sign(p)$ (verify).

Example 5.3.18 Consider the $n \times n$ matrix $A_n = [a_{ij}]$, where $a_{ij} = min(i, j)$. For example,

$$A_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix}.$$

Subtracting the first row of A_n from the rest of the following rows of A_n, it reduces to

$$\begin{bmatrix} 1 & 1_{1\times(n-1)} \\ 0_{(n-1)\times 1} & A_{n-1} \end{bmatrix},$$

where $1_{1\times(n-1)}$ is a $1 \times (n-1)$ matrix with each entry 1 and $0_{(n-1)\times 1}$ represents $(n-1) \times 1$ matrix with each entry 0. Thus, $det(A_n = det(A_{n-1})$. Using induction, and the fact that $det(A_1) = 1$, we see that $det(A_n) = 1$ for all n.

Example 5.3.19 **Vandermonde matrix and determinant**. A matrix V_n of the type

$$V_n = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix}$$

is called the **Vandermonde matrix**, and $det(V_n)$ of a Vandermonde matrix is called a Vandermonde determinant. We show, by induction, that $det(V_n) = \prod_{n \geq i > j}(x_i - x_j)$. If $x_i = x_j$ for some $i \neq j$, then two columns of the Vandermonde matrix are same, and so the $det(V_n) = 0$, and then there is nothing to do. Assume that all $x_1, x_2, \ldots, x_n$ are distinct. For $n = 1$, there is nothing to do. For $n = 2$, $det(V_2) = x_2 - x_1$. Thus, the result is true for $n = 1$ and $n = 2$. Assume that the result is true for all $m < n \geq 3$. Let $f(t) = det(V_n(t))$, where the matrix $V_n(t)$ is obtained by replacing $x_n = t$ in the Vandermonde matrix V_n. Clearly, $f(t)$ is a polynomial in t of degree $n - 1$. Each $x_i, i \leq n - 1$ is a root of $f(t)$ for if we replace $t = x_i, i \leq n - 1$, then $det(V_t) = 0$. Thus $f(t) = a(t - x_1)(t - x_2) \cdots (t - x_{n-1})$ for some constant a which is the coefficient of t^{n-1} in $f(t)$. Clearly, the coefficient t^{n-1} is $det(V_{n-1})$. By the induction hypothesis $a = det(V_{n-1}) = \prod_{n \geq i > j}(x_i - x_j)$. Substituting the value of a we find that $det(V_n) = \prod_{n \geq i > j}(x_i - x_j)$.

Determinant as Volume Form

Consider a parallelogram in $\mathbb{R}^2$ with co-terminus edges $\overline{OP}$ and $\overline{OQ}$ with P and Q having position vectors $\overline{r_1} = [a_{11}, a_{12}]$ and $\overline{r_2} = [a_{21}, a_{22}]$ respectively. The area Ω of the parallelogram is given by $\Omega = base \times height = |\overline{r_1}||\overline{r_2}^{\perp}|$, where $\overline{r_2}^{\perp} = \overline{r_2} - < \overline{r_2}, \frac{\overline{r_1}}{|\overline{r_1}|} > \frac{\overline{r_1}}{|\overline{r_1}|}$ is the resolution of $\overline{r_2}$ orthogonal $\overline{r_1}$. Now,

$$(|\overline{r_2}^{\perp}|)^2 = < \overline{r_2} - < \overline{r_2}, \frac{\overline{r_1}}{|\overline{r_1}|} > \frac{\overline{r_1}}{|\overline{r_1}|}, \overline{r_2} - < \overline{r_2}, \frac{\overline{r_1}}{|\overline{r_1}|} > \frac{\overline{r_1}}{|\overline{r_1}|} >$$

$$= |\overline{r_2}|^2 - \frac{< \overline{r_2}, \overline{r_1} >^2}{|\overline{r_1}|^2}.$$

Thus,

$$\Omega^2 = |\overline{r_1}|^2 |\overline{r_2}|^2 - < \overline{r_2}, \overline{r_1} >^2$$

$$= det(\begin{bmatrix} \overline{r_1}\overline{r_1}^t & \overline{r_1}\overline{r_2}^t \\ \overline{r_2}\overline{r_1}^t & \overline{r_2}\overline{r_2}^t \end{bmatrix}) = det(AA^t),$$

where

$$A = \begin{bmatrix} \overline{r_1} \\ \overline{r_2} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Similarly, if we take a parallelogram in $\mathbb{R}^3$ with co-terminus edges $\overline{OP}$ and $\overline{OQ}$ with P and Q having position vectors $\overline{r_1} = [a_{11}, a_{12}, a_{13}]$ and $\overline{r_2} = [a_{21}, a_{22}, a_{23}]$ respectively, then the area Ω of the parallelogram is $base \times height = |\overline{r_1}||\overline{r_2}^{\perp}|$, where $\overline{r_2}^{\perp} = \overline{r_2} - < \overline{r_2}, \frac{\overline{r_1}}{|\overline{r_1}|} > \frac{\overline{r_1}}{|\overline{r_1}|}$ is the resolution of $\overline{r_2}$ orthogonal $\overline{r_1}$. It turns out again that the area $\Omega = \sqrt{det(AA^t)}$, where

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}.$$

More generally, it follows by induction that the volume V of the parallelepiped in R^n whose co-terminus edges are given by the vectors $\{\overline{r_1}, \overline{r_2}, \ldots, \overline{r_m}\}$ is given by $V = \sqrt{det(AA^t)}$, where

$$A = \begin{bmatrix} \overline{r_1} \\ \overline{r_1} \\ . \\ . \\ . \\ \overline{r_m} \end{bmatrix}.$$

In particular, the volume V of the parallelepiped in $\mathbb{R}^n$ whose co-terminus edges are given by vectors $\{\overline{r_1}, \overline{r_2}, \ldots, \overline{r_n}\}$ is given by $V = det(A)$, where

$$A = \begin{bmatrix} \overline{r_1} \\ \overline{r_1} \\ . \\ . \\ . \\ \overline{r_n} \end{bmatrix}.$$

Example 5.3.20 The area Ω of the parallelogram in $\mathbb{R}^3$ with co-terminus edges given by vectors $[1, 0, 1]$ and $[2, 1, 1]$ is given by

$$\sqrt{Det(\begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 1 \end{bmatrix})} = \sqrt{Det(\begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix})} = \sqrt{6}.$$

Exercises

5.3.1 Let V be a vector space of dimension n and W a vector space of dimension m. Find a basis and also the dimension of $A_r(V, W)$. In particular, find a basis and show that the dimension of $A_r(V, F)$ is ${}^nC_r$.

5.3.2 Find the determinant of the matrix

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \end{bmatrix}.$$

Find the co-factor, adjoint, and also the inverse of the matrix.

5.3.3 Find the determinant of the matrix A_n given by

$$A_n = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ x_1^n & x_2^n & \cdots & x_n^n \end{bmatrix}.$$

5.3.4 Show that the determinant of an orthogonal matrix is ± 1. Let A be a 2×2 orthogonal matrix whose determinant is 1. Show that it is a rotation matrix in the sense that there exist a θ such that

$$\begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix}.$$

Show further that if $det(A) = -1$, it represents reflection in the plane about a line passing through origin.

5.3.5 Find the determinant of a $n \times n$ matrix $A = [a_{ij}]$, where $a_{ij} = max(i, j)$.

5.3.6 Show that the determinant of the $n \times n$ matrix A_n whose ith row is $[(i-1)n + 1, (i-1)n + 2, \ldots, in]$ is 0 for $n \geq 3$. What is the rank A_n?

5.3.7 Show that the determinant of a unitary matrix is a complex number whose modulus is 1. Conversely, show that any complex number with modulus 1 is determinant of a unitary matrix.

5.3.8 Let R be a commutative integral domain. Then it can be considered as a subring of a field F. Let A be a matrix with entries in R and so in F. Show that A has inverse with entries in R if and only if $det(A)$ is a unit in R in the sense that its inverse is in R. Deduce that a matrix A with entries in $\mathbb{Z}$ has inverse with entries in $\mathbb{Z}$ if and only if $det(A) = \pm 1$.

5.3.9 Let p be a permutation of degree n, and A_p is the matrix obtained by permuting the rows of the identity matrix through permutation p. Show that A_p is an orthogonal matrix whose determinant is $sign(p)$.

5.3.10 Suppose that A is invertible. Show that A^{adj} is also invertible. Is the converse true? Support.

5.3.11 Suppose that A is a 3×3 invertible matrix with determinant 3. Find the determinant of A^{adj}.

5.3.12 Suppose that A is a invertible 4×4 matrix such that $det(A^{adj}) = 8$. Find the determinant of A.

5.3.13 Can we find a invertible 4×4 rational matrix A such that $det(A^{adj}) = 2$? Support.

5.3.14 Let A be a real skew-symmetric $n \times n$ matrix, where n is odd. Show that $det(A) = 0$. Deduce that A is not invertible.

5.3.15 Find the solution of the following system of linear equation using the Cramer's rule:

$$x + y + z + t = 3$$

$$2x + 3y + 4z + t = 5$$

$$4x + 9y + 16z + t = 2$$

$$8x + 27y + 64z + t = 1.$$

5.3.16 Let $\{\overline{r_1}, \overline{r_2}, \ldots, \overline{r_{n-1}}\}$ be an ordered set of $n-1$ vectors in $\mathbb{R}^n$. Define a map f from $\mathbb{R}^n$ to $\mathbb{R}$ by

$$f(\overline{x}) = Det \left(\begin{bmatrix} \overline{x} \\ \overline{r_1} \\ \overline{r_2} \\ \cdot \\ \cdot \\ \cdot \\ \overline{r_{n-1}} \end{bmatrix} \right).$$

Show that f is a linear functional on $\mathbb{R}^n$. Deduce that there is a unique vector $\overline{u}$ such that $f(\overline{x}) = < \overline{x}, \ \overline{u} >$. Let us call this vector $\overline{u}$ the **vector product** of $\{\overline{r_1}, \overline{r_2}, \ldots, \overline{r_{n-1}}\}$. Observe that on $\mathbb{R}^3$, the concept agrees with that of usual vector product on $\mathbb{R}^3$. Show that the vector product on $\mathbb{R}^n$ defined above is a $n-1$ alternating form on $\mathbb{R}^n$.

5.3.17 Find the vector product in $\mathbb{R}^4$ of the set of three ordered vectors $\{(1, 1, 1, 1), (1, 2, 3, 4), (1, 4, 9, 16)\}$. Determine also the volume of the parallelepiped formed by the three given vectors as co-terminus edges.

5.3.18 Check if

$$| \, det \left(\begin{bmatrix} \overline{r_1} \\ \overline{r_2} \\ \cdot \\ \cdot \\ \cdot \\ \overline{r_n} \end{bmatrix} \right) | \leq | \, \overline{r_1} \, || \, \overline{r_2} \, | \ \cdots | \, \overline{r_n} \, | \, .$$

Determine the condition under which the equality holds.

5.4 Invariant Subspaces, Eigenvalues

Let T be a linear transformation on V. A subspace W of V is called an **invariant** subspace if $T(W) \subseteq W$. Clearly, the zero space $\{0\}$ and the whole space V are invariant subspaces. These invariant subspaces are called improper invariant subspaces. Other invariant subspaces are called proper invariant subspaces.

A linear transformation need not have any proper invariant subspaces. For example, the rotation in $\mathbb{R}^2$ through the angle $\frac{\pi}{4}$ radian has no proper invariant subspace. We shall be mainly interested in one-dimensional invariant subspaces. Let T be a linear transformation on a vector space V. An element $x \in V - \{0\}$ is called an **eigenvector** or a **characteristic** vector or a **proper** vector if there is a λ in F such that $T(x) = \lambda x$. Clearly, such a λ is unique, and it is called the **eigenvalue** or **characteristic value** or a **proper value** corresponding to the eigenvector x. If $x \neq 0$ is an eigenvector corresponding to the eigenvalue λ, then $T(\alpha x) = \alpha T(x) = \alpha\lambda x = \lambda\alpha x$. This means that the subspace $< x >$ generated by x is a one-dimensional invariant subspace of which any nonzero vector is an eigenvector corresponding to the same eigenvalue. Conversely, any nonzero element of a one-dimensional invariant subspace is an eigenvector, and all nonzero vector of this invariant subspace corresponds to same eigenvalue.

The eigenvalues and eigenvectors of a matrix A are defined to be the eigenvalues and eigenvectors of the linear transformation L_A. Thus, a nonzero column vector $\overline{X}^t$ in F^n is an eigenvector of A corresponding to the eigenvalue λ if $L_A(\overline{X}^t) = A \cdot \overline{X}^t = \lambda\overline{X}^t$.

Theorem 5.4.1 *Let V be a finite-dimensional vector space over a field F. Then $\lambda \in F$ is an eigenvalue of a linear transformation T on V if and only if $det(\lambda I - T) = 0$. In turn, $\lambda \in F$ is an eigenvalue of a $n \times n$ matrix A with entries in F if and only if $det(\lambda I_n - A) = 0$.*

Proof By the definition, λ is an eigenvalue of a linear transformation T on V if and only if there is a nonzero element $x \in V$ such that $0 = \lambda x - T(x) = (\lambda I - T)(x)$. This is equivalent to say that $det(\lambda I - T) = 0$. The rest of the statement follows if we apply the result for the linear transformation L_A determined by the matrix A. ♯

Let $A = [a_{ij}]$ be a $n \times n$ matrix with entries in F. Then $xI_n - A$ is a matrix with entries in the polynomial ring $F[x]$. The determinant of $xI_n - A = [x\delta_{ij} - a_{ij}]$ defined again by the formula

$$\Sigma_{p \in S_n} sign(p) \prod_{i=1}^{n} (x\delta_{p(i)i} - a_{p(i)i})$$

is a polynomial in $F[x]$. If we follow the rule of expansion of determinant, we see that it is a polynomial of degree n in $F[x]$. This polynomial is called the **characteristic polynomial** of A and is denoted by $\phi_A(x)$.

Example 5.4.2 Consider the matrix A given by

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

The characteristic polynomial $\phi_A(x)$ is given by

$$\phi_A(x) = Det(xI_3 - A) = Det\left(\begin{bmatrix} x-1 & 0 & -1 \\ 0 & x-1 & 0 \\ -1 & -1 & x-1 \end{bmatrix}\right) = x^3 - 3x^2 - 2x.$$

Definition 5.4.3 The determinant of a $r \times r$ submatrix of A all of whose diagonal entries are also the diagonal entries of A is called a **principal r − minor** of A.

Thus, the principal $1-$minors are precisely diagonal entries of A. There are 3 principal $2-$minors of a 3×3 matrix which are $det(A_{11})$, $det(A_{22})$ and $det(A_{33})$. How many principal $r \times r$ principal minors of a $n \times n$ matrix will be there? The following result follows immediately from the expansion rule of the determinant.

Proposition 5.4.4 *Let A be a $n \times n$ matrix. Then the characteristic polynomial $\phi_A(x)$ of A is given by*

$$\phi_A(x) = x^n - a_1x^{n-1} + a_2x^{n-2} + \cdots + (-1)^r a_r x^{n-r} + \cdots + (-1)^n a_n,$$

where a_r is sum of principal $r-$ minors of A. ♯

In particular, it follows that a_1 is the sum of diagonal entries of A. This is called the **trace** of A. Similarly, a_n is the determinant of A.

Corollary 5.4.5 *The eigenvalues of a matrix A with entries in a field F are precisely the roots of the characteristic polynomial $\phi_A(x)$ which are in F.*

Proof The roots of $\phi_A(x)$ are precisely those λ for which $det(\lambda I - A) = 0$. Equivalently, $\lambda I - A$ is singular. This is further equivalent to say that there is a nonzero vector $\overline{X}^t$ such that $A\overline{X}^t = \lambda\overline{X}^t$. ♯

Corollary 5.4.6 *Let A and B be similar matrices with entries in a field F. Then*
(i) $\phi_A(x) = \phi_B(x)$.
(ii) Sum of principal r-minors of A is same as the sum of the principal r-minors of B.
(iii) Trace of A is same as trace of B, and the determinant of A is also same as that of B.
(iv) Eigenvalues of A are same as those of B.

Proof (ii), (iii), and (iv) are consequence of (i). Thus, it is sufficient to show the (i). Suppose that $B = PAP^{-1}$. Then

$$\phi_B(x) = det(xI_n - B) = det(PxP^{-1} - PAP^{-1}) = detP(Det(xI_n - A))det(P^{-1}) = det(xI_n - A) = \phi_A(x). \quad \sharp$$

Remark 5.4.7 Matrices having same characteristic polynomial (and so same sum of principal r-minors, same trace, same determinant, and same eigenvalues) need not be similar. Consider, for example, a nonidentity uni-upper triangular $n \times n$ matrix A. Then the characteristic polynomial of A is clearly $(x - 1)^n$ which is same as that of the identity matrix. But identity matrix is similar only to identity matrix.

Example 5.4.8 The characteristic polynomial $\phi_A(x)$ of the matrix A in Example 5.4.2 is $\phi_A(x) = x^3 - 3x^2 - 2x$. Thus, the eigenvalues A (which are the roots of the characteristic polynomial) are 0, 1, and 2. We also find eigenvectors. Suppose that the column vector

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

is an eigenvector corresponding to 0. Then

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0 \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Solving, we get that $x_1 = -x_3$, and $x_2 = 0$. Thus, eigenvectors of A corresponding to the eigenvalue 0 are the set of nonzero vectors of the form

$$\begin{bmatrix} a \\ 0 \\ -a \end{bmatrix}$$

Similarly, eigenvectors of A corresponding to the eigenvalue 1 are the set of nonzero vectors of the form

$$\begin{bmatrix} a \\ -a \\ 0 \end{bmatrix},$$

and that corresponding to eigenvalue 2 are the set of nonzero vectors of the form

$$\begin{bmatrix} a \\ 0 \\ a \end{bmatrix}.$$

Remark 5.4.9 A square matrix A with entries in a field F need not have any eigenvalue (in F). For example, the characteristic polynomial of the matrix

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

is $x^2 + 1$ which has no real root and so the matrix has no eigenvalues.

Let T be a linear transformation on a finite-dimensional vector space V. Then the matrix representations of T with respect to different choices of bases are all similar. As such, we can define the characteristic polynomial of T to be the characteristic polynomial of any matrix representing T. Eigenvalues, trace, determinant of a linear transformation are related to the characteristic polynomial of T.

A linear transformation T on a finite-dimensional vector space V is said to be **semi − simple** or diagonalisable, if there is a basis of V with respect to which the matrix of T is a diagonal matrix. This is equivalent to say that there is a basis $\{x_1, x_2, \ldots, x_n\}$ of V such that $T(x_i) = \lambda_i x_i$ for some $\lambda_i \in F$. In other words T is diagonalisable if and only if there is a basis of V consisting of eigenvectors of T. We know that matrices corresponding to different bases are similar, and the similar matrices represent same linear transformation corresponding to different choices of bases. It is also clear that if T is diagonalisable, then any linear transformation similar to T is diagonalisable.

A $n \times n$ matrix A with entries in F is said to be **diagonalizable** or **semi − simple** is L_A is diagonalisable. This is equivalent to say that A is similar to a diagonal matrix. Thus, a $n \times n$ matrix A is digonalizable if and only if F^n has a basis consisting of eigenvectors of A.

Theorem 5.4.10 *Let T be linear transformation on a vector space V of finite dimension. Let $\lambda_1, \lambda_2, \ldots, \lambda_r$ be a set distinct eigenvalues of T. Let $x_1, x_2, \ldots, x_r$ be the corresponding eigenvectors. Then $\{x_1, x_2, \ldots, x_r\}$ is linearly independent.*

Proof Suppose contrary. Then $\{x_1, x_2, \ldots, x_r\}$ is linearly dependent. Since the eigenvectors are nonzero, there is a minimal linearly dependent subset of $\{x_1, x_2, \ldots, x_r\}$ which, of course, contains at least two elements. After rearranging, we may suppose that $\{x_1, x_2, \ldots, x_s\}$ is a minimal linearly dependent subset of $\{x_1, x_2, \ldots, x_r\}$. Then there exists $\alpha_1, \alpha_2, \ldots, \alpha_s$ not all zero such that

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_s x_s = 0 \cdots \cdots . \tag{5.1}$$

All α_i in Eq. 5.1 are nonzero, for otherwise it will contradict the assumption that $\{x_1, x_2, \ldots, x_s\}$ is a minimal linearly dependent subset. Applying the linear transformation T on Eq. 5.1, we get that

$$\alpha_1 \lambda_1 x_1 + \alpha_2 \lambda_2 x_2 + \cdots + \alpha_r \lambda_s x_s = 0 \cdots \cdots . \tag{5.2}$$

Multiplying Eq. 5.1 by λ_1, we get that

$$\lambda_1 \alpha_1 x_1 + \lambda_1 \alpha_2 x_2 + \cdots + \lambda_1 \alpha_s x_s = 0 \cdots \cdots . \tag{5.3}$$

Subtracting Eq. 5.3 from 2, we get that

$$\alpha_2(\lambda_2 - \lambda_1)x_1 + \alpha_3(\lambda_3 - \lambda_1)x_3 + \cdots + \alpha_s(\lambda_s - \lambda_s)x_s = 0.$$

Since each $\alpha_i \neq 0$ and $\lambda_i \neq \lambda_1$ for $i \geq 2$, it reduces to a contradiction to the supposition that $\{x_1, x_2, \ldots, x_s\}$ is minimal linearly dependent set. ♯

Corollary 5.4.11 *Let T be linear transformation on a vector space V of dimension n. Suppose that T has n distinct eigenvalues. Then T is diagonalisable.*

Proof Suppose that T has distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, and $\{x_1, x_2, \ldots, x_n\}$ the corresponding of eigenvectors of T. From the above theorem $\{x_1, x_2, \ldots, x_n\}$ is linearly independent. Since the dim V = n, it is a basis of V. Clearly, the matrix of T relative to this basis is the diagonal matrix $diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$. ♯

Corollary 5.4.12 *If a $n \times n$ matrix A with entries in F has n distinct eigenvalues in F, then A is similar to a diagonal matrix. Indeed, if $\lambda_1, \lambda_2, \ldots, \lambda_n$ are distinct eigenvalues of A, and $\overline{r_1}^t, \overline{r_2}^t, \ldots, \overline{r_n}^t$ are the corresponding column eigenvectors, then the matrix $P = [\overline{r_1}^t, \overline{r_2}^t, \ldots, \overline{r_n}^t]$ is a nonsingular matrix such that $P^{-1}AP = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$.*

Proof From Theorem 5.4.10 the set $\{\overline{r_1}^t, \overline{r_2}^t, \ldots, \overline{r_n}^t\}$ of column eigenvectors form a basis of F^n (here the elements of F^n are treated as column vectors). Thus, P is invertible. Suppose that P^{-1} is the matrix whose ith row is $\overline{s_i}$. Then $\overline{s_i}\ \overline{r_j}^t = 1$ if $i = j$ and 0 otherwise. Further, since the columns $\overline{r_j}^t$ are eigenvectors of A, $AP = [\lambda_1\overline{r_1}^t, \lambda_2\overline{r_2}^t, \ldots, \lambda_n\overline{r_n}^t]$. Hence the ith row jth column entry of $P^{-1}AP$ is $\lambda_j\overline{s_i}\ \overline{r_j}^t$. This is λ_j if $i = j$ and 0 otherwise. This confirms that $P^{-1}AP = Diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$.♯

Thus, any upper triangular matrix with all diagonal entries distinct is similar to a diagonal matrix because diagonal entries of triangular matrices are precisely the eigenvalues of the matrix. The above result need not hold in case all eigenvalues are not distinct. For example, a nonidentity uni-upper triangular matrix is not similar to any diagonal matrix. This is because all eigenvalues of unitriangular matrices are 1, the only diagonal matrix all of whose eigenvalues are 1 is the identity matrix, and the identity matrix is similar only to the identity matrix.

We illustrate the result by means of an example.

Example 5.4.13 Consider the matrix

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 3 \\ 0 & 0 & 3 \end{bmatrix}.$$

The eigenvalues of A are 1, 2, and 3 which are all distinct. Hence A is similar to a diagonal matrix. We find a nonsingular matrix P such that $P^{-1}AP = diag(1, 2, 3)$. We first find eigenvectors corresponding to these eigenvalues. Suppose that the vector

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

is an eigenvector corresponding to the eigenvalue 1. Then $A \cdot X = X$. Equating rows, we find that $x_1 + x_2 = x_1$, $2x_2 + 3x_3 = x_2$ and $3x_3 = x_3$. This implies that $x_2 = 0 = x_3$, and x_1 is arbitrary. Thus

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

is a typical eigenvector of A corresponding to the eigenvalue 1. Using same process we see that

$$e_1 + e_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

is a typical eigenvector corresponding to the eigenvalue 2, and

$$3e_1 + 6e_2 + 2e_3 = \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}$$

is an eigenvector corresponding to eigenvalue 3. Thus, the transformation matrix

$$P = \begin{bmatrix} 1 & 1 & 3 \\ 0 & 1 & 6 \\ 0 & 0 & 2 \end{bmatrix}$$

is a nonsingular matrix such that $P^{-1}AP = Diag(1, 2, 3)$ (confirm it).

Let T be a linear transformation on V, and λ an eigenvector of T. Consider $V_\lambda = \{v \in V \mid T(v) = \lambda v\}$. Then V_λ is a subspace of V (consisting of eigenvectors corresponding to λ together with 0). This subspace is called the **λ-eigenspace** of T.

Corollary 5.4.14 *A linear transformation T on a finite-dimensional vector space V is diagonalisable if and only if V is direct sum of the eigen subspaces of T.*

Proof Suppose that

$$V = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_r},$$

where V_{λ_i} is λ_i-eigenspace. Clearly, λ_i are distinct. Let S_i be a basis of V_i. Then $S = \bigcup_{i=1}^{r} S_i$ is a basis of V consisting of eigenvectors of T. ♯

Corollary 5.4.15 *A linear transformation T on V is diagonalisable if and only if there exists a set $\{\lambda_1, \lambda_2, \ldots, \lambda_r\}$ of distinct eigenvalues such that $dim V = \Sigma_{i=1}^{r} dim(V_{\lambda_i})$.*

Proof Since eigenvectors corresponding to distinct eigenvalues are linearly independent, under the assumption V becomes direct sum of its eigenspaces. ♯

Let $F[x]$ denote the set of all polynomials with coefficients in the field F. $F[x]$ is a commutative integral domain (with respect to the usual addition and multiplication of polynomials in $F[x]$) in the sense that it satisfies all the postulate of a field except the existence of the multiplicative inverse of a nonzero element in $F[x]$ (Indeed, there is no polynomial $f(x)$ such that $xf(x) = 1$). Let T be a fixed linear transformation on a vector space V over a field F, and

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

a polynomial in $F[x]$. Then $f(T)$ is a linear transformation on V defined by

$$f(T) = a_0I + a_1T + a_2T^2 + \cdots + a_nT^n.$$

We can extend the multiplication on the vector space V by the members of F to the multiplication by the members of $F[x]$ by defining $f(x) \cdot v = f(T)(v)$. Then V becomes a $F[x]$-module in the sense that it satisfies all the postulates of a vector space with the field F replaced by the polynomial ring $F[x]$.

Similarly, if A is a $n \times n$ matrix with entries in a field F, and $f(x)$ a polynomial in $F[x]$, then we have the matrix $f(A)$ defined by

$$f(A) = a_0I_n + a_1A + a_2A^2 + \cdots + a_nA^n.$$

It may be observed that if A is matrix of T with respect to certain basis, then $f(A)$ is the matrix of $f(T)$ with respect to same basis. It may also be observed that if λ is an eigenvalue of T with eigenvector v, then $f(T)(v) = f(\lambda)v$, and so $f(\lambda)$ is an eigenvalue of $f(T)$. If A is a $n \times n$ matrix, then F^n becomes a $F[x]$ module with respect to the external product defined by $f(x) \cdot \overline{X}^t = f(A) \cdot \overline{X}^t$, where $\overline{X}^t$ is a column vector in F^n. It is clear that the matrix product $(xI_n - A) \cdot \overline{X}^t = 0$ for all $\overline{X}^t \in F^n$. Matrix theory with entries in the polynomial ring $F[x]$ can be developed on the pattern it was developed for matrices with entries in F. For example, we can talk of adjoint of a matrix, determinant of a matrix, and the relation $A^{adj} \cdot A = det(A)I_n$ holds for the matrices with entries in a field $F[x]$ (the proof goes exactly on the same lines) also.

Following is one of the most fundamental results in linear algebra.

Theorem 5.4.16 (Cayley Hamilton Theorem) *Every square matrix satisfies its own characteristic polynomial. More precisely, if A is a square matrix, then* $\phi_A(A) = 0$.

Proof $\phi_A(x) = Det(xI - A)$. The matrix $xI - A$ is a matrix with entries in $F[x]$. From the discussion above, it follows that

$$(xI_n - A)^{adj} \cdot (xI_n - A) = Det(xI_n - A)I_n = \phi_A(x)I_n.$$

Hence,

$$\phi_A(A) \cdot X = \phi_A(x) \cdot X = \phi_A(x) I_n \cdot X = (xI_n - A)^{adj}(xI_n - A) \cdot X = 0$$

(see the discussion in the paragraph just above the theorem). This shows that the matrix $\phi_A(A) = 0$. ♯

Let A be a 3×3 unitriangular matrix. Then its characteristic polynomial $\phi_A(x) = (x-1)^3$. From the Cayley Hamilton theorem $(A - I_3)^3 = 0$. In other words, $A^3 - 3A^2 + 3A - I_3 = 0$. This shows that $A(A^2 - 3A + 3I_3) = I_3$ and so $A^2 - 3A + 3I_3$ is the inverse of A. Similarly, result holds for any $n \times n$ unitriangular matrices. This also says that if A is strictly triangular $n \times n$ matrix, then $A^n = 0$. If A is nonsingular, then the constant term $(-1)^n a_n$ in the characteristic polynomial $\phi_A(x)$, being the determinant of A, is nonzero. Since $\phi_A(A)$ is the zero matrix, $(-1)^n a_n I_n = -(A^n - a_1 A^{n-1} + a_2 A^{n-2} + \cdots + (-1)^r a_r A^{n-r} + \cdots + (-1)^{n-1} a_{n-1} A)$, where a_r is the sum of the principal r-minors of A. It follows that the inverse of a matrix A, if exists, is a polynomial in A. This also gives an algorithm to find the inverse of A.

A linear transformation T on V is said to be **triangulable** if there is a basis of V with respect to which the matrix is a triangular matrix. A matrix A with entries in F is said to be **triangulable** if L_A is triangulable. This is equivalent to say that A is similar to a triangular matrix. In general, a matrix in a field need not be similar to a triangular matrix. Consider the matrix

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

over the field $\mathbb{R}$ of real numbers. This is not similar to any triangular matrix over $\mathbb{R}$. For if it is similar to a triangular matrix over $\mathbb{R}$, then it will have its eigenvalues real (the diagonal terms of the triangular matrix to which it is similar). But this has no real eigenvalues.

Theorem 5.4.17 *A linear transformation on V is triangulable if and only if there exists an ascending chain*

$$\{0\} = V_0 \subset V_1 \subset V_2 \subset \cdots \subset V_n = V$$

of invariant subspaces, called a **flag** *of V, such that the dimension of V_i is i.*

Proof Suppose that such a chain of invariant subspaces exist. By induction, we show the existence of a basis $\{x_1, x_2, \ldots, x_n\}$ of V such that $\{x_{n-r+1}, x_{n-r+2}, \ldots, x_n\}$ is a basis of V_r for each r. Let $\{x_n\}$ be a basis of V_1. Since $\{x_n\}$ is linearly independent subset of V_2, it can be extended to a basis $\{x_{n-1}, x_n\}$ of V_2. Proceeding inductively, we find a basis $\{x_1, x_2, \ldots, x_n\}$ of V with the required property. Since each V_{n-r+1} which has basis $\{x_r, x_{r+1}, \ldots, x_n\}$ is invariant under T, it follows that

$$T(x_r) = a_{rr}x_r + a_{rr+1}x_{r+1} + \cdots + a_{rn}x_n$$

for each r. This means that the matrix of T with respect to this basis is triangular.

Conversely, suppose that T is triangulable. Then there is a basis $\{x_1, x_2, \ldots, x_n\}$ with respect to which it is a upper triangular matrix. But, then

$$T(x_r) = a_{rr}x_r + a_{rr+1}x_{r+1} + \cdots + a_{rn}x_n$$

for each r. Let V_r be the subspace of V generated by $\{x_{n-r+1}, x_{n-r+2}, \ldots, x_n\}$. Then it follows that V_r is invariant under T, dimension of V_r is r, and we have the chain

$$\{0\} \subset V_1 \subset V_2 \subset \cdots \subset V_n. \qquad \sharp$$

Corollary 5.4.18 *A matrix A is triangulable if and only if there is a chain*

$$\{0\} \subset V_1 \subset V_2 \subset \cdots \subset F^n$$

of subspaces of F^n such that dimension of V_r is r and $A \cdot \overline{X}^t = L_A(\overline{X}^t) \in V_r$ for all r, and for all $\overline{X}^t \in V_r$. $\sharp$

As observed earlier, a matrix need not be triangulable. The reason was that there need not be any eigenvalue of the matrix. A field F is called algebraically closed, if every polynomial in $F[x]$ has all its roots in F. It is a fact that every field can be enlarged to an algebraically closed field (see Chap. 9).

Theorem 5.4.19 *Every linear transformation on a vector space V over an algebraically closed field is triangulable.*

Proof Let T be a linear transformation on a vector space V over an algebraically closed field F. We have to show the existence of a chain

$$\{0\} \subset V_1 \subset V_2 \subset \cdots \subset V_n = V$$

of invariant subspaces. The proof is by the induction on the dimension of V. If $dim V = 1$, then there is nothing to prove. Assume that the result is true for all vector spaces of dimension less than n. Suppose that the dimension of V is $n \geq 2$. Since F is algebraically closed, the characteristic polynomial $\phi_T(x)$ has a root $\lambda \in F$. Then λ is an eigenvalue. Thus, there exists a nonzero vector v in V such that $T(v) = \lambda v$. Let V_1 be the subspace generated by v. Then V_1 is of dimension 1. Since $T(v) = \lambda v \in V_1$, V_1 is an invariant subspace. Consider the vector space $W = V/V_1$. Then the dimension of W is $n-1$, and since $T(V_1) \subseteq V_1$, T induces a linear transformation $\overline{T}$ on W defined by $\overline{T}(w + V_1) = T(w) + V_1$. By the induction hypothesis there is a chain

$$\{V_1\} = W_0 \subset W_1 = V_2/V_1 \subset W_2 = V_3/V_1 \subset \cdots \subset W_{n-1} = V_n/V_1 = W$$

of invariant subspaces of $\overline{T}$, such that dimension of W_r is r, and so dimension of V_{r+1} is $r+1$. We show that each V_r is invariant under T for each r. Already, V_1 is invariant under T. Let $x \in V_r,\ r \geq 2$. Then $\overline{T}(x + V_1) = T(x) + V_1$ belongs to $W_{r-1} = V_r/V_1$. This implies that $T(x) \in V_r$. ♯

Corollary 5.4.20 *Every square matrix A with entries in an algebraically closed field is similar to a triangular matrix.*

Proof To say that A is similar to a triangular matrix is to say that L_A is triangulable. The result follows from the above theorem. ♯

5.5 Spectral Theorem, and Orthogonal Reduction

Theorem 5.5.1 *Let V be a complex inner product space, and T a Hermitian linear transformation on V. Then all the eigenvalues of T are real.*

Proof Let λ be an eigenvalue of T. Then there exists a nonzero vector $x \in V$ such that $T(x) = \lambda x$. Since T is Hermitian, $< T(u),\ v > = < u,\ T(v) >$ for all $u, v \in V$, and hence

$$\lambda < x,\ x > = < \lambda x,\ x > = < T(x),\ x > = < x,\ T(x) > = < x,\ \lambda x > = \overline{\lambda} < x,\ x > .$$

Since $x \neq 0,\ < x,\ x > \neq 0$, and so $\lambda = \overline{\lambda}$. ♯

Corollary 5.5.2 *Let V be a finite-dimensional complex inner product space, and T a Hermitian linear transformation on V. Then all the roots of the characteristic polynomial of T are real.*

Proof Since the field $\mathbb{C}$ of complex numbers is algebraically closed, all the roots of the characteristic polynomial of A exist in $\mathbb{C}$, and they are the eigenvalues of A. The result follows from the above theorem. ♯

Corollary 5.5.3 *All eigenvalues of Hermitian matrices are real.*

Proof The matrix A is Hermitian if and only if L_A is Hermitian linear transformation on the standard complex inner product space $\mathbb{C}^n$. ♯

Corollary 5.5.4 *All roots of the characteristic polynomial of a real symmetric matrix are real. In particular, every real symmetric matrix has an eigenvalue.*

Proof A real symmetric matrix can also be taken to be a complex Hermitian matrix. The result follows from the above corollary. ♯

Corollary 5.5.5 *All roots of the characteristic polynomial of a symmetric linear transformation T on a real inner product space V are real. In particular, if T is a real symmetric linear transformation, then there is a real number λ, and $x \neq 0$ such that $T(x) = \lambda x$.* ♯

Corollary 5.5.6 *All nonzero eigenvalues of a skew-Hermitian matrix are purely imaginary.*

Proof We know that A is skew-Hermitian if and only if iA is Hermitian. Now, λ is an eigenvalue of A if and only if $i\lambda$ is an eigenvalue of iA. This shows that $i\lambda$ is real, and so λ is purely imaginary. ♯

Corollary 5.5.7 *Let A be a real skew-symmetric matrix. Then there is no nonzero eigenvalue of A. In other words, if $A \cdot \overline{X}^t = \lambda \overline{X}^t$, where λ is real, then $\lambda = 0$ or $\overline{X}^t = \overline{0}^t$.*

Proof A real skew-symmetric matrix is also a skew-Hermitian matrix. Hence, all the nonzero roots of its characteristic polynomial are purely imaginary. Thus, there is no $\overline{X}^t \neq \overline{0}^t$, and real number $\lambda \neq 0$ such that $A \cdot \overline{X}^t = \lambda \overline{X}^t$. ♯

Proposition 5.5.8 *Let A be an unitary linear transformation (matrix) on a complex inner product space, and λ an eigenvalue of A. Then $\mid \lambda \mid = 1$.*

Proof Let $x \neq 0$ be an eigenvector corresponding to eigenvalue λ. Then

$$\mid \lambda \mid^2 < x,\ x > = \lambda\overline{\lambda} < x,\ x > = < \lambda x,\ \lambda x > = < T(x),\ T(x) > = < T^\star T(x),\ x > = < x,\ x > .$$

Since $x \neq 0$, $< x,\ x > \neq 0$. Hence $\mid \lambda \mid^2 = 1$. ♯

Proposition 5.5.9 *Let A be an orthogonal linear transformation (matrix) on real inner product space, and λ a real eigenvalue of A. Then $\lambda = \pm 1$.*

Proof Let $x \neq 0$ be an eigenvector corresponding to a real eigenvalue λ. Then as in the previous proposition

$$\lambda^2 < x,\ x > = < x,\ x >,$$

and so $\lambda^2 = 1$. Hence $\lambda = \pm 1$. ♯

Proposition 5.5.10 *Let V be an inner product space. Let T be a linear transformation on V, and W is a subspace of V which is invariant under T. Then the orthogonal compliment $W^\perp$ of W is invariant under $T^\star$.*

Proof Since W is invariant under T, for each $y \in W$, $T(y) \in W$. Let $x \in W^\perp$. Then for each $y \in W$, $< y,\ T^\star(x) > = < T(y),\ x > = 0$. This shows that $T^\star(x) \in W^\perp$. ♯

Corollary 5.5.11 *Let T be a Hermitian linear transformation on an inner product space V, and W an invariant subspace of T. Then $W^\perp$ is also invariant under T.* ♯

Proposition 5.5.12 *Let T be a Hermitian linear transformation on a complex (real) inner product space V. Let x_1 and x_2 be eigenvectors corresponding to distinct eigenvalues λ_1 and λ_2 of T. Then $< x_1,\ x_2 > = 0$. In other words, $V_{\lambda_1} \perp V_{\lambda_2}$.*

Proof From previous results, λ_1 and λ_2 are real. Further,

$$\lambda_1 < x_1,\ x_2 > = < \lambda_1 x_1,\ x_2 > = < T(x_1),\ x_2 > = < x_1,\ T(x_2) > = < x_1,\ \lambda_2 x_2 > = \lambda_2 < x_1,\ x_2 > .$$

Since $\lambda_1 \neq \lambda_2$, we see that $< x_1,\ x_2 > = 0$. ♯

If we apply the above result for L_A on the standard inner product space, then we have the following corollary:

Corollary 5.5.13 *Let A be a Hermitian (real symmetric) matrix with eigenvectors X_1 and X_2 corresponding to distinct eigenvalues. Then $X_1^\star X_2 = 0$ ($X_1^t X_2 = 0$).* ♯

Theorem 5.5.14 (Spectral Theorem) *Let T be a Hermitian linear transformation on a finite-dimensional complex (real) inner product space V. Then there is an orthonormal basis consisting of eigenvectors of T.*

Proof The proof is by the induction on $dim V$. If $dim V = 1$, then take any nonzero vector of V and divide it by its length to get a unit vector v. Clearly, $\{v\}$ is an orthonormal basis of V, and since $dim V = 1$, v is an eigenvector of T. Assume that the result holds for all Hermitian linear transformations on vector spaces of dimensions less than n. Let T be a Hermitian linear transformation on a complex (real) inner product space V of dimension n. Now, T, being a Hermitian linear transformation on a complex (real) inner product space V, has an eigenvector x_1. Dividing x_1 by its length, we may assume that x_1 is a unit vector. Let W be the subspace of V generated by x_1. Then $V = W \oplus W^\perp$. Clearly, W is invariant under T. Since T is Hermitian (real symmetric), $W^\perp$ is also invariant under T. It is clear that the restriction $T/W^\perp$ of T to $W^\perp$ is also Hermitian (real symmetric), and the dimension of $W^\perp$ is $n-1$. By the induction hypothesis, $W^\perp$ has an orthonormal basis $\{x_2, x_3, \ldots, x_n\}$ consisting of eigenvectors of $T/W^\perp$ (and so of T). Clearly, then $\{x_1, x_2, \ldots, x_n\}$ is an orthonormal basis of V consisting of eigenvectors of T. ♯

Corollary 5.5.15 *Let T be a Hermitian linear transformation on a complex(real) inner product space. Let $\{\lambda_1, \lambda_2, \ldots, \lambda_r\}$ be the set of all distinct eigenvalues of T. Then*

$$V = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_r}.$$

Proof Since eigenspaces corresponding to distinct eigenvalues are orthogonal, the result follows from the above theorem. ♯

Corollary 5.5.16 *The matrix representation of a Hermitian linear transformation with respect to a suitable orthonormal basis is a diagonal matrix.* ♯

Corollary 5.5.17 *Let A be a Hermitian (real symmetric) matrix. Then A is similar to a diagonal matrix. In fact there exists a unitary matrix U(an orthogonal matrix O) such that $U^\star AU$ ($O^t AO$) is a diagonal matrix.*

Proof A is Hermitian (real symmetric) matrix if and only if L_A is Hermitian (real symmetric) linear transformation on the standard complex (real) inner product space. Thus, there exists an orthonormal basis $\{\overline{X_1}^t, \overline{X_2}^t, \ldots, \overline{X_n}^t\}$ of the standard complex (real) inner product space consisting of eigenvectors of L_A (and so of A also). Now, the standard inner product is given by $< \overline{X_i},\ \overline{X_j} > = \overline{X_i} \cdot \overline{X_j}^\star$ in complex case, and by $< \overline{X_i},\ \overline{X_j} > = \overline{X_i} \cdot \overline{X_j}^t$ in real case, where $\cdot$ in R.H.S. is the matrix multiplication. Thus, $\overline{X_i} \cdot \overline{X_i}^\star = 1$ $(\overline{X_i} \cdot \overline{X_i}^t = 1)$, and for $i \neq j$, $\overline{X_i} \cdot \overline{X_j}^\star = 0$ $(\overline{X_i} \cdot \overline{X_j}^t = 0)$. Let U (respectively O) denote the matrix whose ith row is $\overline{X_i}$. Then the above observation says that $U(O)$ is unitary (orthogonal) such that

$$UAU^\star = \begin{bmatrix} \overline{X_1} \\ \overline{X_2} \\ \cdot \\ \cdot \\ \cdot \\ \overline{X_n} \end{bmatrix} \cdot A \cdot [\overline{X_1}^\star, \overline{X_2}^\star, \ldots, \overline{X_n}^\star] =$$

$$\begin{bmatrix} \overline{X_1} \\ \overline{X_2} \\ \cdot \\ \cdot \\ \cdot \\ \overline{X_n} \end{bmatrix} \cdot [\lambda_1 \overline{X_1}^\star, \lambda_2 \overline{X_2}^\star, \ldots, \lambda_n \overline{X_n}^\star] = diag(\lambda_1, \lambda_2, \ldots, \lambda_n),$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are eigenvalues of A. Similarly, if A is a real symmetric matrix, then $OAO^t = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$. ♯

Remark 5.5.18 We have an algorithm to find an orthonormal basis of V consisting of eigenvectors of T provided that we have an algorithm to solve the characteristic polynomial of T (indeed, we have algorithms to solve a nth degree equation for $n \leq 4$ (see the Chap. 9 on Fields and Galois theory)). After getting the distinct eigenvalues, we can find the corresponding eigenspaces, and then use Gram–Schmidt process to find orthonormal basis of each eigenspaces. In turn, it gives an orthonormal basis consisting of eigenvectors. This also gives us a method to diagonalize a Hermitian, and also a real symmetric matrix. We illustrate it by means of an example.

Example 5.5.19 Consider the matrix

$$A = \begin{bmatrix} \frac{3}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{3}{2} \end{bmatrix}.$$

This is a real symmetric matrix. We find orthogonal matrix O such that O^tAO is a diagonal matrix. The characteristic polynomial $\phi_A(x)$ is given by

$$\phi_A(x) = Det(xI - A) = (x - \tfrac{3}{2})(x - 1)(x - \tfrac{3}{2}) - \tfrac{(x-1)}{4}.$$

The roots of the characteristic polynomial are 1,1, and 2. We find the eigenspace $\mathbb{R}^3_1$. Suppose that

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix}$$

belongs to $\mathbb{R}^3_1$. Then

$$A \cdot \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} u \\ v \\ w \end{bmatrix}.$$

Equating rows, we get that $u = w$. Thus

$$\mathbb{R}^3_1 = \{ \begin{bmatrix} u \\ v \\ w \end{bmatrix} \textit{ such that } u = w \}.$$

This subspace is clearly of dimension 2. Putting $u = 1 = w = v$, we get a nonzero member of R^3_1. Another nonzero member of $\mathbb{R}^3_1$ which is not a multiple of the previous element is obtained by taking $u = 0 = w$ and $v = 1$. Hence

$$\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \}$$

is a basis of $\mathbb{R}^3_1$. Using Gram–Schmidt process we find the orthonormal basis

$$\{ \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ \sqrt{\frac{2}{3}} \\ -\frac{1}{\sqrt{6}} \end{bmatrix} \}$$

of $\mathbb{R}^3_1$. Similarly, $\mathbb{R}^3_2$ is subspace of dimension 1 which has a singleton

$$\{ \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \}$$

as an orthonormal basis. This gives us an orthonormal basis

$$\{ \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ \sqrt{\frac{2}{3}} \\ -\frac{1}{\sqrt{6}} \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \}$$

In turn, we get an orthogonal matrix

$$O = \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \sqrt{\frac{2}{3}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

such that $O^tAO = Diag(1, 1, 2)$.

Example 5.5.20 Let A be a $n \times n$ real symmetric matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ counted with their multiplicities. Let $f(x)$ be a polynomial with real coefficients. Let $\mu_1, \mu_2, \ldots, \mu_n$ be real numbers with $f(\mu_i) = \lambda_i$ for all i. Then we can find a real symmetric matrix B such that $f(B) = A$ as follows: From Corollary 5.5.17, there exists an orthogonal matrix O such that $O^tAO = Diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Let $B = O\, Diag(\mu_1, \mu_2, \ldots, \mu_n)O^t$. Then $f(B) = ODiag(f(\mu_1), f(\mu_2), \ldots, f(\mu_n))\, O^t = O\, Diag(\lambda_1, \lambda_2, \ldots, \lambda_n)\, O^t = A$. Thus, then, B is a solution of $f(X) = A$. In particular, if we take $B = ODiag(1, 1, \sqrt{2})O^t$, where O is as in the above example, then $B^2 = A$, where A is as in the above example. Can we count the number of solutions of $X^2 = A$, where X is an unknown in the set of real symmetric matrices?

Let B be any nonsingular complex (real) matrix. Then the matrix $A = BB^\star$ (BB^t) is a Hermitian (real symmetric) matrix. Further, let λ is an eigenvalue of A. Then since $< \overline{X}BB^\star, \overline{X} > = < \overline{X}B, \overline{X}B >$ is non-negative for all row vector $\overline{X}$, it follows that all the eigenvalues of A are positive. Conversely, if all the eigenvalues of a Hermitian (real symmetric) matrix A are positive (non-negative), then, as described above, we can find a positive definite (positive) Hermitian (real symmetric) matrix B such that $A = B^2 = BB^\star(BB^t)$:

Polar Decomposition

Every nonzero complex number $z = a + ib$ is nonsingular 1×1 matrix which is uniquely expressible in polar form as $z = |z|\, u = re^{i\theta}$, where r is positive definite 1×1 Hermitian matrix, and $u = e^{i\theta}$ is 1×1 unitary matrix. More generally, every nonsingular complex square matrix (indeed, every complex matrix) A can be uniquely expressed as $A = B + iC$, where B and C are Hermitian matrices. Following is the multiplicative analog of this identity called the **polar decomposition**.

Proposition 5.5.21 *Every nonsingular square complex matrix A can be uniquely expressed as $A = PU$, where P is a positive definite Hermitian matrix, and U is a unitary matrix.*

Proof Consider the matrix $AA^\star$. Since A is nonsingular, $AA^\star$ is a Hermitian matrix all of whose eigenvalues are positive. As such, $AA^\star$ is positive definite Hermitian matrix. Let P be a positive definite Hermitian matrix which is square root of $AA^\star$. Take $U = P^{-1}A$. Then $UU^\star = P^{-1}AA^\star(P^{-1})^\star$. Again, since P is Hermitian P^{-1} is also Hermitian (indeed, $(P^{-1})^\star = (P^\star)^{-1} = P^{-1}$). Thus, $UU^\star = P^{-1}AA^\star(P^{-1})^\star = P^{-1}AA^\star P^{-1} = I$. This shows that U is unitary, and $A = PU$. ♯

Corollary 5.5.22 *Every nonsingular matrix A with real entries can be uniquely expressed as $A = PO$, where P is a positive definite real symmetric matrix, and O is an orthogonal matrix.* ♯

Example 5.5.23 Consider the matrix

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{3}} \\ 0 & -\frac{\sqrt{2}}{\sqrt{3}} & \frac{2}{\sqrt{3}} \end{bmatrix}.$$

This matrix is nonsingular. We find its polar decomposition. The matrix $B = AA^t$ is given by

$$B = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

The eigenvalues of B are 1, 1, 4. Using the algorithm as described in Remark 5.5.18 (see Example 5.5.19), we find an orthogonal matrix O given by

$$O = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & -\frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix},$$

such that $B = O^t\, Diag[1,\ 1,\ 4]\, O$. The positive square root P of B is given by $P = O^t\, Diag[1,\ 1,\ 2]\, O$. As described in Proposition 5.5.21, $U = P^{-1}A$ is unitary, and $A = PU$ is the polar decomposition of A.

Singular Value Decomposition

Let A be a $n \times m$ matrix with entries in a field F, where F is $\mathbb{C}$ or F is $\mathbb{R}$. The matrix $AA^\star$ is a positive matrix in the sense that all its eigenvalues are non-negative. This is because if λ is an eigenvalue of $AA^\star$, then there is a nonzero row vector $\overline{x}$ such that $AA^\star\overline{x}^\star = \lambda\overline{x}^\star$. Hence $(\overline{x}A)(\overline{x}A)^\star = \lambda\overline{x}\overline{x}^\star$. Since $\overline{x}$ is nonzero vector, it follows that λ is non-negative.

Definition 5.5.24 The non-negative square root of eigenvalues of $AA^\star$ is called a **singular value** of A.

If A is a Hermitian, then λ is an eigenvalue of A if and only if λ^2 is an eigenvalue of $A^2 = AA^\star$. As such, the singular values of Hermitian matrices are precisely the absolute values of their eigenvalues.

Example 5.5.25 The singular values of the matrix

$$\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

are $\sqrt{2}$, $\sqrt{2}$, for the eigenvalues of

$$AA^t = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

are 2, 2.

Proposition 5.5.26 *Let A be a $n \times m$ matrix and $\overline{x}^\star$ is a unit column eigenvector of $AA^\star$ with λ as corresponding eigenvalue. Then $(|| \overline{x}A ||)^2 = \lambda$. In turn, $|| \overline{x}A ||$ is the corresponding singular value. Further, if $\overline{x}$ is a row eigenvector of $AA^\star$, and $\overline{y}$ is a row vector orthogonal to $\overline{x}$, then $\overline{x}A$ and $\overline{y}A$ are orthogonal to each other.*

Proof Under the hypothesis, $(|| \overline{x}A ||)^2 = \overline{x}A(\overline{x}A)^\star = \overline{x}AA^\star\overline{x}^\star = \lambda\overline{x}\,\overline{x}^\star = \lambda$. Next, suppose that $\overline{x}$ is a row eigenvector of $AA^\star$ with associated eigenvalue λ. Then $\overline{x}AA^\star = \lambda\overline{x}$. In turn,

$$\overline{y}A(\overline{x}A)^\star = \overline{y}AA^\star\overline{x}^\star = \overline{y}(\overline{x}AA^\star)^\star = \lambda\overline{y}\overline{x}^\star.$$

The result is evident. ♯

Corollary 5.5.27 *Let A be a $n \times m$ matrix. Then rank of A is ρ if and only if there are exactly ρ strictly positive eigenvalues of $AA^\star$. Equivalently, there are exactly ρ nonzero singular values of A.*

Proof Suppose that $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_\rho$ are nonzero eigenvalues of $AA^\star$ and the rest of the $n - \rho$ eigenvalues are 0. Let $\{\overline{r_1}^\star, \overline{r_2}^\star, \ldots, \overline{r_\rho}^\star, \ldots, \overline{r_n}^\star\}$ be an orthonormal basis of F^n (considering F^n as space of columns) consisting of column eigenvectors of $AA^\star$ with $\overline{r_1}^\star, \overline{r_2}^\star, \ldots, \overline{r_\rho}^\star$ corresponding to eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\rho$, respectively. It follows from the above proposition that $|| \overline{r_i}A ||$ is nonzero if and only if $i \leq \rho$. Further, $\overline{r_i}A(\overline{r_j}A)^\star = \overline{r_i}AA^\star\overline{r_j}^\star = \lambda_j\overline{r_i}\,\overline{r_j}^\star = \lambda_i$ for $i = j$ and 0, otherwise. This shows that $\{\overline{r_1}A, \overline{r_2}A, \ldots, \overline{r_\rho}A\}$, being an orthogonal set, is linearly independent. Since the rest of $\overline{r_i}A$, $i > \rho$ are zero, it follows that ρ is the rank of A. ♯

Theorem 5.5.28 (Singular value decomposition) *Let A be a $n \times m$ matrix. Then there exists a unitary/orthogonal $n \times n$ matrix U, and a $m \times m$ unitary/orthogonal matrix V such that $UAV = \Sigma$, where Σ is a $n \times m$ matrix whose first ρ diagonal entries are nonzero singular values $\sigma_1, \sigma_2, \ldots, \sigma_\rho$ of A in non-ascending order, and the rest of the entries are 0. In turn, $A = U^\star\Sigma V^\star$, where $U^\star$ and $V^\star$ are again unitary/orthogonal matrices, and Σ as described.*

Proof Let $\{\overline{r_1}^\star, \overline{r_2}^\star, \ldots, \overline{r_\rho}^\star, \ldots, \overline{r_n}^\star\}$ be an orthonormal basis of F^n (considering F^n as space of columns) consisting of column eigenvectors of $AA^\star$ with $\overline{r_1}^\star, \overline{r_2}^\star, \ldots, \overline{r_\rho}^\star$ corresponding to nonzero eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\rho$. From Proposition 5.5.26 and Corollary 5.5.27, it follows that $\{A^\star\overline{r_1}^\star, A^\star\overline{r_2}^\star, \ldots, A^\star\overline{r_\rho}^\star\}$ is an orthogonal set of column vectors in F^m. Let $\overline{s_j}^\star$ denote the column vector $\frac{1}{\sigma_j}A^\star\overline{r_j}^\star$. Then $\{\overline{s_1}^\star, \overline{s_2}^\star, \ldots, \overline{s_\rho}^\star\}$ is an orthonormal set of column vectors in F^m. Embed it in to

an orthonormal basis $\{\overline{s_1}^\star, \overline{s_2}^\star, \ldots, \overline{s_m}^\star\}$ of F^m. Take U to be the $n \times n$ matrix whose *ith* row is $\overline{r_i}$, and V the $m \times m$ matrix whose *jth* column is $\overline{s_j}^\star$. Then U and V are unitary/ orthogonal matrices. The *ith* row *jth* column entry c_{ij} of UAV is given by $\overline{r_i}A\,\overline{s_j}^\star$. Further, $\overline{r_i}A = \overline{0}$ for all $i > \rho$, $\overline{r_i}A(\overline{r_j}A)^\star = 0$ for $i \neq j$ and $\overline{r_i}A(\overline{r_i}A)^\star = \lambda_i = \sigma_i^2$. Now, it is evident that $c_{ii} = \sigma_i$ for all $i \leq \rho$ and 0, otherwise. ♯

The proof of the theorem is algorithmic. We illustrate it by means of the following example.

Example 5.5.29 Consider the matrix A given by

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

Then

$$AA^t = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

The eigenvalues of AA^t are 3, 1. Thus, the singular values of A are $\sqrt{3},\ 1$. A unit eigenvector $\overline{r_1}$ corresponding to 3 is $[\frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{2}}]$, and that $\overline{r_2}$ corresponding to 1 is $[\frac{1}{\sqrt{2}},\ -\frac{1}{\sqrt{2}}]$. $\overline{r_1}A = [\frac{1}{\sqrt{2}},\ \sqrt{2},\ \frac{1}{\sqrt{2}}]$, and $\overline{r_2}A = [-\frac{1}{\sqrt{2}},\ 0,\ \frac{1}{\sqrt{2}}]$. Thus, $\overline{s_1} = [\frac{1}{\sqrt{6}},\ \sqrt{\frac{2}{3}},\ \frac{1}{\sqrt{6}}]$, and $\overline{s_2} = [-\frac{1}{\sqrt{2}},\ 0,\ \frac{1}{\sqrt{2}}]$. We extend $\{\overline{s_1}, \overline{s_2}\}$ to an orthonormal basis by adjoining $\overline{s_3} = [\frac{1}{\sqrt{3}},\ -\frac{1}{\sqrt{3}},\ \frac{1}{\sqrt{3}}]$. The matrix U is given by

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix},$$

the matrix V is the transpose of

$$\begin{bmatrix} \frac{1}{\sqrt{6}} & \sqrt{\frac{2}{3}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix},$$

and the matrix Σ is given by

$$\Sigma = \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Further, $A = U^t \Sigma V^t$.

Geometry of Orthogonal Transformation

Recall that a subspace H of dimension $n-1$ of the Euclidean space $\mathbb{R}^n$ is called a **hyperplane**, and a translate $\overline{x} + H$ is called an **affine hyperplane**.

Proposition 5.5.30 *Let H be a hyperplane in the Euclidean space $\mathbb{R}^n$. Then there is a unit vector $\overline{x} \in H^{\perp}$. Further, if $\overline{y}$ is any unit vector in $H^{\perp}$, then $\overline{y} = \pm\overline{x}$.*

Proof Let $\{\overline{s_1}, \overline{s_2}, \ldots, \overline{s_{n-1}}\}$ be an orthonormal basis of H. Using Gram–Schmidt process, enlarge it to an orthonormal basis $\{\overline{s_1}, \overline{s_2}, \ldots, \overline{s_{n-1}}, \overline{x}\}$ of $\mathbb{R}^n$. Then $\overline{x}$ is a unit vector in $H^{\perp}$. Let

$$\overline{y} = a_1\overline{s_1} + a_2\overline{s_2} + \cdots + a_{n-1}\overline{s_{n-1}} + a\overline{x}$$

be a unit vector which is a member of $H^{\perp}$. Then

$$1 = <\overline{y}, \overline{y}> = a <\overline{y}, \overline{x}> = a^2 \cdot <\overline{x}, \overline{x}> = a^2.$$

Hence $a = \pm 1$. Further, $0 = <\overline{y}, \overline{s_i}> = a_i$ for all i. This shows that $\overline{y} = \pm\overline{x}$. ♯

If $\overline{x}$ is a unit vector, the hyperplane H consisting of vectors orthogonal to $\overline{x}$ is denoted by $H_{\overline{x}}$. It follows that $H_{\overline{x}} = H_{\overline{y}}$ if and only if $\overline{x} = \pm\overline{y}$.

Proposition 5.5.31 *Let $\overline{x}$ be a unit vector in $\mathbb{R}^n$, and $H_{\overline{x}}$ the corresponding hyperplane. Then the map $\sigma_{\overline{x}}$ from $\mathbb{R}^n$ to itself defined by*

$$\sigma_{\overline{x}}(\overline{y}) = \overline{y} - 2 <\overline{y}, \overline{x}> \overline{x}$$

is the unique linear transformation which fixes the members of $H_{\overline{x}}$, and takes $\overline{x}$ to its negative. Further, $\sigma_{\overline{x}}$ is an orthogonal transformation with determinant -1 (refer to the Exercises 4.4.35–4.4.40).

Proof Clearly, $\sigma_{\overline{x}}$ is a linear transformation. If $<\overline{y}, \overline{x}> = 0$, then by the definition, $\sigma_{\overline{x}}(\overline{y}) = \overline{y}$. Also $\sigma_{\overline{x}}(\overline{x}) = -\overline{x}$. Further,
$< \sigma_{\overline{x}}(\overline{y}), \sigma_{\overline{x}}(\overline{z}) >$
$= <\overline{y} - 2 <\overline{y}, \overline{x}> \overline{x}, \overline{z} - 2 <\overline{z}, \overline{x}> \overline{x} >$
$= <\overline{y}, \overline{z}>$.
This shows that $\sigma_{\overline{x}}$ is an orthogonal transformation. Also, the matrix representation of $\sigma_{\overline{x}}$ with respect to the orthonormal basis $\{\overline{s_1}, \overline{s_2}, \ldots, \overline{s_{n-1}}, \overline{x}\}$ of $\mathbb{R}^n$, where $\{\overline{s_1}, \overline{s_2}, \ldots, \overline{s_{n-1}}\}$ is an orthonormal basis of $H_{\overline{x}}$ is the diagonal matrix $diag(1, 1, \ldots, 1, -1)$. Hence, its determinant is -1. Finally, if τ is any linear transformation with the required property, then again the matrix representation of τ with respect to the basis $\{\overline{s_1}, \overline{s_2}, \ldots, \overline{s_{n-1}}, \overline{x}\}$ is $diag(1, 1, \ldots, 1, -1)$. Hence, $\tau = \sigma_{\overline{x}}$. ♯

Definition 5.5.32 The transformation $\sigma_{\overline{x}}$ is called a **hyperplane reflection** (indeed, it is reflection about the hyperplane $H_{\overline{x}}$).

Note that $\sigma_{\overline{x}} = \sigma_{\overline{y}}$ if and only if $\overline{x} = \pm\overline{y}$
Recall that an inner product space V is said to be the **orthogonal sum** of its subspaces $V_1, V_2, \ldots, V_r$ if

$$V = V_1 + V_2 + \cdots + V_r,$$

and for $i \neq j$, the elements of V_i are orthogonal to V_j. Symbolically, we write it as

$$V = V_1 \perp V_2 \perp \cdots \perp V_r.$$

In particular, if W is a subspace of V, then $V = W \perp W^{\perp}$.

Theorem 5.5.33 *Let T be an Euclidean orthogonal transformation on $\mathbb{R}^n$. Then there exist subspaces V, W, and two-dimensional subspaces $P_1, P_2, \ldots, P_l$ together with angles $0 < \theta_1 \leq \theta_2 \leq, \ldots, \leq \theta_l < \pi$ such that the following hold:*

1. $\mathbb{R}^n = V \perp W \perp P_1 \perp P_2 \perp \cdots \perp P_l$.

2. $T(\overline{x}) = \overline{x}$ for all $\overline{x} \in V$.

3. $T(\overline{x}) = -\overline{x}$ for all $\overline{x} \in W$.

4. The restriction of T to P_k is a rotation through an the angle θ_k in the plane P_k. In other words, if $\{\overline{u_k}, \overline{v_k}\}$ is an orthonormal basis of P_k, then

(i) $T(\overline{u_k}) = cos\theta_k \overline{u_k} + sin\theta_k \overline{v_k}$, and

(ii) $T(\overline{v_k}) = -sin\theta_k \overline{u_k} + cos\theta_k \overline{v_k}$ for all $k \leq l$.

Proof The proof is by the induction on n. For $n = 1$, it follows trivially. Assume that the result is true for all $m \leq n$. Let T be an orthogonal transformation on $\mathbb{R}^{n+1}$. Then T induces a unique linear transformation $\tilde{T}$ from the standard complex vector space $\mathbb{C}^{n+1}$ to itself by putting $\tilde{T}(\overline{x}) = \overline{x}M(T)$, where $\overline{x} = [x_1\, x_2\, \cdots\, x_{n+1}]$ is a row vector in $\mathbb{C}^{n+1}$, and $M(T)$ is the matrix representation of T with respect to the standard basis of $\mathbb{R}^{n+1}$. Note that the matrix representation of $\tilde{T}$ with respect to the standard basis of $\mathbb{C}^{n+1}$ is the same as $M(T)$. Let λ be an eigenvalue of $M(T)$ which may be a complex number. Then there is a nonzero complex unit vector $\overline{x} = [x_1\, x_2\, \cdots\, x_{n+1}]$ such that $\overline{x}M(T) = \lambda\overline{x}$. Since $M(T)$ is a real matrix,

$$\overline{\overline{x}}M(T) = \overline{\overline{x}M(T)} = \overline{\lambda}\overline{x},$$

where $\overline{\overline{x}} = [\overline{x_1}\, \overline{x_2}\, \cdots\, \overline{x_{n+1}}]$ denote the complex conjugate of the complex vector $\overline{x}$, and $\overline{\lambda}$ the complex conjugate of λ. This shows that $\overline{\lambda}$ is also an eigenvalue of $M(T)$, and $\overline{\overline{x}}$ is a corresponding eigenvector. Since $M(T)$ is orthogonal,

$$|\lambda|^2 = \lambda\overline{x}\overline{\lambda}\overline{\overline{x}}^t = \overline{x}M(T)(\overline{\overline{x}}M(T))^t = \overline{x}(\overline{\overline{x}})^t = ||\overline{x}||^2.$$

This shows that $|\lambda| = 1$, and so $\lambda = e^{i\theta}$ for some θ.

Now, suppose that $M(T)$ has a real eigenvalue λ. Then $\lambda = \pm 1$. Suppose again that $\lambda = 1$ is an eigenvalue of $M(T)$, and $\overline{x}$ is a unit eigenvector associated to 1. Clearly, then $\overline{x}$ is a real vector. Consider the corresponding hyperplane $H_{\overline{x}}$. The dimension of $H_{\overline{x}}$ is n. Further, if $\overline{y} \in H_{\overline{x}}$, then

$$< \overline{x},\ T(\overline{y}) > = < T(\overline{x}),\ T(\overline{y}) > = < \overline{x},\ \overline{y} >$$

(for T is orthogonal). This shows that T restricted to $H_{\overline{x}}$ is an orthogonal transformation on $H_{\overline{x}}$. Also $\mathbb{R}^{n+1} = < \overline{x} > \perp H_{\overline{x}}$. By the induction hypothesis, there exist

subspaces V', W, and two-dimensional subspaces $P_1, P_2, \ldots, P_l$ of $H_{\overline{x}}$ together with angles $0 < \theta_1 \leq \theta_2 \leq, \ldots, \leq \theta_l < \pi$ such that the following hold:

1. $H_{\overline{x}} = V' \perp W \perp P_1 \perp P_2 \perp \cdots \perp P_l$.
2. $T(\overline{x}) = \overline{x}$ *for all* $\overline{x} \in V'$.
3. $T(\overline{x}) = -\overline{x}$ *for all* $\overline{x} \in W$.
4. The restriction of T to P_k is a rotation through an the angle θ_k in the plane P_k for each $k \leq l$.

Taking $V = < \overline{x} > \perp V'$, the result holds for $\mathbb{R}^{n+1}$. If 1 is not an eigenvalue but -1 is an eigenvalue, then a similar argument proves the result with $V = \{0\}$.

Assume that $M(T)$ has no real eigenvalues. Note that in this case $n+1$ will be even $2m$ (say). Let θ_1 be the smallest positive real number such that $\lambda = e^{i\theta_1}$ is an eigenvalue, and $\overline{x}$ a corresponding complex eigenvector. As observed earlier, $e^{-i\theta_1}$ is also an eigenvalue with $\overline{\overline{x}}$ a corresponding eigenvector. Then $\overline{x} + \overline{\overline{x}}$ and $i(\overline{x} - \overline{\overline{x}})$ are nonzero real vectors. Take $\overline{u_1} = \frac{\overline{x}+\overline{\overline{x}}}{||\overline{x}+\overline{\overline{x}}||}$, and $\overline{v_1} = \frac{i(\overline{x}-\overline{\overline{x}})}{||\overline{x}-\overline{\overline{x}}||}$. Then $\{\overline{u_1}, \overline{v_1}\}$ is an orthonormal subset of $\mathbb{R}^{n+1}$ which generates a subspace P_1. Then $T(\overline{u_1}) = \overline{u_1}M(T) = \frac{1}{||\overline{x}+\overline{\overline{x}}||}(e^{i\theta_1}\overline{x} + e^{-i\theta_1}\overline{\overline{x}}) = cos\theta_1\overline{u_1} + sin\theta_1\overline{v_1}$, and similarly, $T(\overline{v_1}) = -sin\theta_1\overline{u_1} + cos\theta_1\overline{v_1}$. Let $U = P_1^{\perp}$. Then U is of dimension $2(m-1)$, $\mathbb{R}^{n+1} = P_1 \perp U$, and T restricted to U is an orthogonal transformation on U. By the induction hypothesis, there exist two-dimensional subspaces $P_2, P_3, \ldots, P_m$ together with angles $0 < \theta_2 \leq \theta_3 \leq, \ldots, \leq \theta_m < \pi$ with $\theta_1 \leq \theta_2$ such that $U = P_2 \perp P_3 \perp \cdots \perp P_m$, and T restricted to each P_k is a rotation through the angle θ_k. Clearly, $\mathbb{R}^{n+1} = P_1 \perp P_2 \perp P_3 \perp \cdots \perp P_m$. ♯

Corollary 5.5.34 *Suppose that the dimension of W in the above theorem is m, and $\{\overline{w_1}, \overline{w_2}, \ldots, \overline{w_m}\}$ an orthonormal basis of W. Then*

$$T = \sigma_{\overline{w_1}} o \sigma_{\overline{w_2}} o \cdots o \sigma_{\overline{w_m}} o \rho_{P_1} o \rho_{P_1} o \cdots o \rho_{P_l},$$

where $\sigma_{\overline{w_i}}$ is the reflection about the hyperplane $H_{\overline{w_i}}$, and ρ_{P_j} denote the rotation through an angle θ_j in the plane P_j, and it is given by

(i) $\rho_{P_j}(\overline{x}) = \overline{x}$ *for all* $\overline{x}$ *in* V, W *and* P_k, $k \neq j$,

(ii) $\rho_{P_j}(\overline{u_j}) = cos\theta_j\overline{u_j} + sin\theta_j\overline{v_j}$,

(iii) $\rho_{P_j}(\overline{v_j}) = -sin\theta_j\overline{u_j} + cos\theta_j\overline{v_j}$, *where $\{u_j, v_j\}$ is an orthonormal basis of $P_j, j \leq l$.* ♯

Corollary 5.5.35 *The transformation T in Theorem 5.5.33 is a composition of $m + 2l$ hyperplane reflections, where $m = dim W$.*

Proof From the above corollary, it is sufficient to show that the rotation ρ_{P_j} is composition of two hyperplane reflections. Since,

$$\begin{bmatrix} cos\theta_j & sin\theta_j \\ -sin\theta_j & cos\theta_j \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} cos\theta_j & sin\theta_j \\ sin\theta_j & -cos\theta_j \end{bmatrix},$$

it follows that $\rho_{P_j} = \sigma_{\overline{e_j}} o \sigma_{\overline{v_j}}$, where $\overline{v_j} = cos\frac{\theta_j}{2}\overline{u_j} + sin\frac{\theta_j}{2}\overline{v_j}$. ♯

Corollary 5.5.36 *The matrix representation of the orthogonal transformation T on $\mathbb{R}^n$ described in Theorem 5.5.33 with respect to a suitable orthonormal basis is $A = [a_{ij}]$, where*

(i) $a_{ii} = 1$ for all $i \leq r = dimV$ (r may be 0 also),
(ii) $a_{ii} = -1$ for all $i = r + j \leq r + m$, where $m = dimW$ (m may also be 0),
(iii) $a_{ii} = cos\theta_j$ for $i = r + m + 2j - 1$ and $i = r + m + 2j, j \leq l$,
(iv) $a_{ii+1} = sin\theta_j$ and $a_{i+1i} = -sin\theta_j$ for $i = r + m + 2j - 1, j \leq l$,
(v) the rest of the entries are 0.

Proof Taking orthonormal bases of V, W, and of the two-dimensional subspaces $P_1, P_2, \ldots, P_l$ together, we get an orthonormal basis of $\mathbb{R}^n$ with respect to which the matrix representation is the required one. ♯

Corollary 5.5.37 *Every orthogonal matrix A is orthogonally similar to a matrix of the form described in the above corollary. More explicitly, there is an orthogonal matrix O such that OAO^t is the matrix of the form described in the above corollary.* ♯

Corollary 5.5.38 *An orthogonal transformation with determinant 1 is composition of even number of hyperplane reflections, and with determinant -1 is composition of odd number of hyperplane reflections.* ♯

Corollary 5.5.39 *Two orthogonal matrices A and B are similar if and only if they have same set of eigenvalues counted with their multiplicities.* ♯

Exercises

5.5.1 Find invariant subspaces of the differential operator D on the space $\wp_n$ of polynomials of degree at most n over reals. Find its characteristic polynomial, and also the eigenvalues.

5.5.2 Consider the matrix

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

Show that the cube roots of 1 are precisely the eigenvalues of A. Show that the matrix is diagonalisable over the field $\mathbb{C}$ of complex numbers. Find a complex matrix P such that $P^{-1}AP$ is a diagonal matrix. Is it diagonalisable over the field $\mathbb{R}$ of reals?

5.5.3 Show that the matrix

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 4 & 7 \\ 0 & 0 & 2 \end{bmatrix}$$

is diagonalisable over the field $\mathbb{R}$ of real numbers, and find P such that $P^{-1}AP$ is a diagonal matrix.

5.5.4 Show that the matrix

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 0 & 2 & 5 \\ 0 & 0 & 2 \end{bmatrix}$$

is not diagonalisable even over the field $\mathbb{C}$ of complex numbers.

5.5.5 Show that the matrix

$$A = \begin{bmatrix} 1 - cos\alpha sin\alpha & cos^2\alpha \\ -sin^2\alpha & 1 + sin\alpha cos\alpha \end{bmatrix}$$

is similar to an upper triangular matrix over $\mathbb{R}$. Find P such that PAP^{-1} is upper triangular. What is PAP^{-1}? Show that it is not similar to a diagonal matrix even over the field $\mathbb{C}$ of complex numbers.

5.5.6 Show that a 2×2 matrix over reals all of whose off diagonal entries are positive have all its eigenvalues real. Determine a necessary and sufficient condition on the entries of a 2×2 matrix for its diagonalisability.

5.5.7 Suppose that $T^2 - 5T + 6 = 0$. Determine all possible eigenvalues of T.

5.5.8 Suppose that A is nonsingular. Show that AB and BA have same eigenvalues.

5.5.9 Let A be a $n \times n$ matrix. Suppose that $A^m = 0$ for some $m \geq 1$. Show that $A^n = 0$.

5.5.10 Let A be a nilpotent $n \times n$ matrix. Show that $I_n + A$ is invertible, and $det(I_n + A) = 1$.

5.5.11 Let A and B be complex $n \times n$ matrices such that $AB - BA$ commutes with A. Show that $(AB - BA)^n = 0$.

5.5.12 Let A be a $n \times n$ matrix. Define a map M_A from $M_n(F)$ to itself by $M_A(B) = A \cdot B$. Show that M_A is a linear transformation. Relate the trace of A and that of M_A.

5.5.13 Suppose that $A^t = A^2$. What are possible eigenvalues of A?

5.5.14 Show that the determinant of a Hermitian matrix is always real.

5.5.15 Show that the determinant of a skew-symmetric real matrix of odd order is 0.

5.5.16 Let A be $n \times n$ skew-Hermitian matrix. Show that the determinant of A is either 0, or it is purely imaginary if n is odd. Show that it is purely real if n is even.

5.5.17 Show that if A is $n \times n$ Hermitian, then $iI_n + A$ is invertible.

5.5.18 Show that if A is skew real symmetric (skew-Hermitian) $n \times n$ matrix, then $I_n + A$ is nonsingular.

5.5.19 Show that if $X^\star AX$ is real for all complex vector X, then A is Hermitian.

5.5.20 Find an orthogonal matrix O such that O^tAO is diagonal, where

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 3 \end{bmatrix}.$$

Find a real symmetric matrix B, if possible, such that $B^2 = A$.

5.5.21 Show that all eigenvalues of $A^\star A$ are real, and $A^\star A$ is unitarily similar to a diagonal matrix.

5.5.22 Let A be a real matrix. Show that A^tA is similar to a diagonal matrix.

5.5.23 Show that every real symmetric matrix A can be expressed as $A = PDP^t$, where D is a diagonal matrix.

5.5.24 Show that every nonsingular real symmetric matrix A can be expressed as $A = LDL^t$, where L is a lower triangular matrix, and D a diagonal matrix.

5.5.25 Let A be a $n \times n$ matrix with entries in $\mathbb{R}$. Show that the map $<,>$ from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$ defined by $< \bar{x}, \bar{y} > = \bar{x}A\,\bar{y}^t$ is an inner product if and only if A is symmetric, and all the eigenvalues of A are positive.

5.5.26 Which of the following matrices are positive or positive definite?
(i)

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \\ 1 & 1 & 3 \end{bmatrix}$$

(ii)

$$A = \begin{bmatrix} 2 & 3 & 1 \\ 3 & 1 & 1 \\ 1 & 1 & 3 \end{bmatrix}$$

(iii)

$$A = \begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & 1 \\ 5 & 1 & 3 \end{bmatrix}$$

(iv)

$$A = \begin{bmatrix} 3 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 3 \end{bmatrix}$$

(v)

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

For each of the above matrices, find orthogonal matrices O such that O^tAO is a diagonal matrix. Express the positive definite matrices as BB^t.

5.5.27 Find the cube roots of the matrices in Exercise 5.5.26.

5.5.28 Show that $I + iA^\star A$ is nonsingular for all complex matrices A.

5.5.29 Show that the matrix

$$A = \begin{bmatrix} cos\alpha & sin\alpha \\ -sin\alpha & cos\alpha \end{bmatrix}$$

is diagonalisable over $\mathbb{R}$ if and only if $\alpha = n\pi$ for some n. Diagonalize it over the field $\mathbb{C}$ of complex numbers.

5.5.30 Show that every orthogonal 2×2 matrix with determinant 1 is a matrix of the form given in the above exercise.

5.5.31 Show that the group $O(2)$ is isomorphic to $SO(2) \times \mathbb{Z}_2$. Show that the group $SO(2)$ is isomorphic to the circle group S^1.

5.5.32 Show that $O(3)$ is isomorphic to $SO(3) \times \mathbb{Z}_2$.

5.5.33 Let $A \in SO(3)$. Show that $A - I_3$ is singular. Deduce that 1 is always an eigenvalue of A. What are other possible eigenvalues of A.
Hint. $Det(A - I_3) = Det(A^t(A - I_3)) = Det(I_3 - A)$.

5.5.34 Use the above exercise to show that every matrix A in $SO(3)$ is similar to a matrix of the form

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\alpha & sin\alpha \\ 0 & -sin\alpha & cos\alpha \end{bmatrix}.$$

Deduce that every matrix A in $SO(3)$ represents a rotation in R^3 about an axis through an angle α, where trace of A is $2cos\alpha + 1$. In particular, deduce that $-1 \leq Tr(A) \leq 3$. This justifies the name **rotation** group for $SO(3)$.

5.5.35 Show that two orthogonal 3×3 matrices are similar if and only they have the same trace, and also the same determinant.

5.5.36 Show that $SO(3)$ is generated by reflections.

5.5.37 Describe the closed subgroups of $SO(3)$.

5.5.38 Show that the group $O(n)$ acts transitively on the set $V(n, r)$ of r-dimensional subspaces of $\mathbb{R}^n$. Describe the isotropy subgroup of the subspace $W \in V(n, r)$ consisting of vectors with last $n - r$ coordinates 0.

5.5.39 Show that the group $O(n)$ acts transitively on the $n-1$ sphere $S^{n-1} = \{\overline{x} \in \mathbb{R}^n \mid || \overline{x} || = 1\}$ in $\mathbb{R}^n$. Describe the isotropy group $O(n)_{\overline{e_1}}$.

5.5.40 Consider the sphere $S^2 = \{\overline{x} \in \mathbb{R}^3 \mid || \overline{x} || = 1\}$ in $\mathbb{R}^3$. Define a map d_{S^2} from $S^2 \times S^2$ to $\mathbb{R}^+ \bigcup \{0\}$ by $cosd_{S^2}(\overline{x}, \overline{y}) = < \overline{x}, \overline{y} >$. Show that d_{S^2} is a metric called the **spherical metric**. Use Exercise 5.5.38 to show that the map $d_{S^{n-1}}$ from $S^{n-1} \times S^{n-1}$ to $\mathbb{R}^+ \bigcup \{0\}$ defined by $cosd_{S^{n-1}}(\overline{x}, \overline{y}) = < \overline{x}, \overline{y} >$ is metric. The metric space (S^n, d_{S^n}) is called the **spherical n-space**. Describe the Geodesics (path of shortest distance) in S^n.

5.5.41 Consider the upper part of the hyperboloid $H^2 = \{\overline{x} = (x_1, x_2, x_3) \mid x_1^2 - x_2^2 - x_3^2 = 1, x_1 > 0\}$. Show that $\overline{x}, \overline{y} \in H^2$ implies that $x_1y_1 - x_2y_2 - x_3y_3 \geq 2$. Show that the map d_{H^2} from $H^2 \times H^2$ to $\mathbb{R}^+ \bigcup \{0\}$ defined by $cosh(d_{H^2}(\overline{x}, \overline{y}) = x_1y_1 - x_2y_2 - x_3y_3$ is a metric (called the **hyperbolic metric**). (How to generalize it for arbitrary n?).

5.5.42 Show that every matrix in $SO(3)$ is similar to a diagonal matrix over the field $\mathbb{C}$ of complex numbers.

5.5.43 Show that a square matrix A with entries in the field $\mathbb{C}$ of complex numbers is similar to a diagonal matrix if and only if it is a normal matrix in the sense that $AA^\star = A^\star A$.

5.5.44 Find the polar decomposition of the following complex matrix.

$$\begin{bmatrix} 1 & i \\ 1 & 1 \end{bmatrix}.$$

5.5.45 Find the polar decomposition of the following real matrix.

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}.$$

5.5.46 Find a singular value decomposition of the following matrices:

$$\begin{bmatrix} 1 & i \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

5.5.47 The map $\times$ from $\mathbb{R}^3 \times \mathbb{R}^3$ to $\mathbb{R}^3$ defined by $\overline{a} \times \overline{b} = (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1)$ is called the vector product on $\mathbb{R}^3$. Show that the vector product is uniquely characterized by the requirement that it is bilinear, and it satisfies
(i) $\overline{a} \times \overline{a} = \overline{0}$,

(ii) $< \overline{a} \times \overline{b},\ \overline{a} > = 0 = < \overline{a} \times \overline{b},\ \overline{b} >$ for all $\overline{a}, \overline{b} \in \mathbb{R}^3$.
(iii) $< \overline{e_1} \times \overline{e_2},\ \overline{e_3} > = 1$.
Further, show that a nonzero alternating map from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}^n$ exists if and only if $n = 3$. In particular, the concept of vector product exists only on $\mathbb{R}^3$.

5.6 Bilinear and Quadratic Forms

In this section, we discuss bilinear and quadratic forms, and their canonical reduction.

Definition 5.6.1 Let V be a finite-dimensional vector space over a field F. A map f from $V \times V$ to F is called a **bilinear** form if f is linear in each coordinate in the sense that
(i) $f(ax + by, z) = af(x, y) + bf(y, z)$, and
(ii) $f(x, ay + bz) = af(x, z) + bf(x, z)$
for all $x, y, z \in V$, and $a, b \in F$.

Thus, a map f from $V \times V$ to F is a bilinear form if and only if the maps f_x and f^y from V to F defined by $f_x(y) = f(x, y)$ and $f^y(x) = f(x, y)$ are linear functionals. Further, then the maps $x \rightsquigarrow f_x$, and $y \rightsquigarrow f^y$ denoted by L_f and R_f, respectively, are linear transformations from V to $V^\star$ (verify).

The zero map from $V \times V$ to F is a bilinear form on V. Any inner product on a real vector space is a bilinear form. The determinant map from $F^2 \times F^2$ to F is an other bilinear form on F^2, where F is a field.

Example 5.6.2 Consider the vector space F^n of column vectors over F. Let A be a $n \times n$ matrix over F. The map f from $F^n \times F^n$ to F defined by

$$f(X, Y) = X^t AY$$

is a bilinear form on F^n (verify). We shall see below that these are all bilinear forms on F^n. In fact, given any bilinear form f on a vector space V of dimension n over F, there exists an isomorphism T from V to F^n (corresponding to each choice of basis), and a matrix A such that $f(x, y) = T(x)^t AT(y)$. Thus, essentially these are all bilinear forms on a vector space of dimension n.

Example 5.6.3 Let ϕ and ψ be linear functional on a vector space V of dimension n. Then the map f from $V \times V$ to F defined by $f(x, y) = \phi(x) \cdot \psi(y)$ is a bilinear form on V. Is it true that every bilinear form on V is of this form? Support.

Let f and g be bilinear forms on V and $a, b \in F$. Then it is easily seen that $af + bg$ defined by $(af + bg)(x, y) = af(x, y) + bg(x, y)$ is a bilinear form on V. Further, the zero map which takes every thing to 0 is already a bilinear form. Thus, the set $BL(V, F)$ of bilinear forms on V is a vector space over F with respect to the

operations defined above. Let us fix an ordered basis $\{u_1, u_2, \ldots, u_n\}$ of V. Define a map $M_{u_1,u_2,\ldots,u_n}$ from $BL(V, F)$ to $M_n(F)$ by

$$M_{u_1,u_2,\ldots,u_n}(f) = [a_{ij}],$$

where $a_{ij} = f(u_i, u_j)$. This map is a linear transformation (verify), and it is called the **matrix representation** map relative to the ordered basis $\{u_1, u_2, \ldots, u_n\}$. The matrix $M_{u_1,u_2,\ldots,u_n}(f)$ is called the **matrix representation** of the bilinear form f.

Theorem 5.6.4 *The matrix representation map $M_{u_1,u_2,\ldots,u_n}$ is a vector space isomorphism from $BL(V, F)$ to $M_n(F)$.*

Proof It is already seen to be a linear transformation. Thus, it is sufficient to show that $M_{u_1,u_2,\ldots,u_n}$ is bijective. Suppose that $M_{u_1,u_2,\ldots,u_n}(f) = M_{u_1,u_2,\ldots,u_n}(g)$. Then $f(u_i, u_j) = g(u_i, u_j)$ for all i, j. Let $x = \Sigma_{i=1}^n x_i u_i$ and $y = \Sigma_{i=1}^n y_i u_i$ be any two members of V. Then using the bilinearity of f and g, we get

$$f(x, y) = f(\Sigma_{i=1}^n x_i u_i, \Sigma_{i=}^n y_i u_i) = \Sigma_{i=1}^n x_i \Sigma_{j=1}^n f(u_i, u_j) y_j = \Sigma_{i,j} x_i f(u_i, u_j) y_j = \Sigma_{i,j} x_i g(u_i, u_j) y_j = g(x, y).$$

This shows that $f = g$. Observe that $f(x, y) = X^t A Y$, where $A = [a_{ij}]$ is the matrix representation of f with respect to the ordered basis $\{u_1, u_2, \ldots, u_n\}$, X and Y are column vectors whose ith row entries are x_i and y_i, respectively. Conversely, let $A = [a_{ij}]$ be a matrix in $M_n(F)$. Define a map f from $V \times V$ to F by

$$f(x, y) = X^t A Y,$$

where $x = \Sigma_{i=1}^n x_i u_i$, $y = \Sigma_{i=1}^n y_i u_i$, X a column vector with ith row entry x_i, and Y a column vector with ith row entry is y_i. It is easy to observe that f defined above is a bilinear form. Since $f(u_i, u_j) = a_{ij}$, the matrix of f is A. This shows that the matrix representation map is an isomorphism. ♯

Corollary 5.6.5 *Let V be a vector space and $\{u_1, u_2, \ldots, u_n\}$ an ordered basis of V. Let $\{u_1^\star, u_2^\star, \ldots, u_n^\star\}$ be the dual basis. Then $\{f_{ij} = u_i^\star u_j^\star \mid 1 \leq i \leq n, 1 \leq j \leq n\}$ forms a basis of $BL(V, F)$.*

Proof The matrix representation map $M_{u_1,u_2,\ldots,u_n}$ takes f_{ij} to e_{ij}, and since $\{e_{ij} \mid 1 \leq i, j \leq n\}$ forms a basis of $M_n(F)$, the result follows from the above theorem. ♯

Effect of Change of Basis on Matrix Representation

Theorem 5.6.6 *The matrix representations of a bilinear form on a vector space V with respect to different choices of bases are all congruent*

Proof Let f be a bilinear form on V. Let $\{u_1, u_2, \ldots, u_n\}$ and $\{v_1, v_2, \ldots, v_n\}$ be ordered bases of V. Let $P = [p_{ij}]$ be the matrix of transformation from the ordered

basis $\{u_1, u_2, \ldots, u_n\}$ to the basis $\{v_1, v_2, \ldots, v_n\}$. This means that $v_j = \Sigma_{i=1}^{n} p_{ij} u_i$. Clearly, P is nonsingular. Further, suppose that

$$M_{u_1, u_2, \ldots, u_n}(f) = [a_{ij}]$$

and

$$M_{v_1, v_2, \ldots, v_n}(f) = [b_{ij}].$$

Then $a_{ij} = f(u_i, u_j)$ and $b_{ij} = f(v_i, v_j)$. Thus, using bilinearity of f we get

$$b_{ij} = f(v_i, v_j) = f(\Sigma_{k=1}^{n} p_{ki} u_k, \Sigma_{l=1}^{n} p_{lj} u_l) = \Sigma_{k=1}^{n} p_{ki} (\Sigma_{l=1}^{n} f(u_k, u_l) p_{lj}).$$

This shows that b_{ij} is the ith row jth column entry of $P^t M_{u_1, u_2, \ldots, u_n}(f) P$. Hence

$$M_{v_1, v_2, \ldots, v_n}(f) = P^t M_{u_1, u_2, \ldots, u_n}(f) P,$$

where P is the matrix of transformation from the ordered basis $\{u_1, u_2, \ldots, u_n\}$ to $\{v_1, v_2, \ldots, v_n\}$. ♯

Definition 5.6.7 Let f and g be bilinear forms on V. We say that f is **congruent** to g if there is an isomorphism T from V to V such that $g(x, y) = f(T(x), T(y))$ for all $x, y \in V$.

The main problem in the theory of bilinear form is the classification of bilinear forms up to congruence over different fields. We shall give a solution to this problem for symmetric ($f(x, y) = f(y, x)$) bilinear forms. Following theorem reduces this problem to the problem of classifying matrices up to congruence. More precisely, we need to determine a unique member from each congruence class of matrices and determine an algorithm to reduce a matrix to the unique representative of the congruence class determined by that matrix.

Theorem 5.6.8 *A bilinear form f is congruent to a bilinear form g on V if and only if their matrix representations corresponding to any choice of ordered bases are congruent.*

Proof Let f and g be congruent bilinear forms on V. Then there is an isomorphism T from V to V such that $g(x, y) = f(T(x), T(y))$ for all $x, y \in V$. Let $\{u_1, u_2, \ldots, u_n\}$ be an ordered basis of V. Then $\{T(u_1), T(u_2), \ldots, T(u_n)\}$ is also an ordered basis of V. Also $g(u_i, u_j) = f(T(u_i), T(u_j))$ for all i, j. This shows that

$$M_{u_1, u_2, \ldots, u_n}(g) = M_{T(u_1), T(u_2), \ldots, T(u_n)}(f).$$

Since matrix of a bilinear form associated to any two ordered basis of V are congruent, it follows that the matrices of f and g corresponding to any choice of ordered bases are congruent. Conversely, suppose that $M_{u_1, u_2, \ldots, u_n}(g) = P^t M_{v_1, v_2, \ldots, v_n} P$. Then

$$g(u_i, u_j) = \Sigma_{k=1}^n p_{ki}(\Sigma_{l=1}^n f(v_k, v_l)p_{lj}) = f(\Sigma_{k=1}^n p_{ki}v_k, \Sigma_{l=1}^n p_{lj}v_l) = f(w_i, w_j),$$

where $w_i = \Sigma_{k=1}^n p_{ki}v_k$, and $w_j = \Sigma_{l=1}^n p_{lj}v_l$. Since P is nonsingular $\{w_1, w_2, \ldots, w_n\}$ form an ordered basis of V. Thus, if T denotes the isomorphism from V to V which takes u_i to w_i, then $g(u_i, u_j) = f(T(u_i), T(u_j))$ for all i, j. Since $\{u_1, u_2, \ldots, u_n\}$ is an ordered basis of V, and f and g are bilinear forms, it follows that $g(x, y) = f(T(x), T(y))$ for all $x, y \in V$. ♯

Example 5.6.9 We have seen above that any bilinear form on F^n is given by $f(X, Y) = X^tAY = \Sigma_{i,j}x_ia_{ij}y_j$, where $A = [a_{ij}]$, X is the column vector whose ith row is x_i, and Y is the column vector whose ith row is y_i. Note that the matrix of this bilinear form with respect to the standard ordered basis is A. Consider the bilinear form on $\mathbb{R}^3$ given by

$$f(X, Y) = x_1y_2 + 2x_1y_3 + x_2y_1 + x_2y_3 + 2x_3y_1 + x_3y_2.$$

The matrix A of this bilinear form with respect to the standard ordered basis is

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}.$$

From Example 2.7.5, it follows that this matrix is congruent to the diagonal matrix $Diag(2, -\frac{1}{2}, -4)$, and the matrix P of transformation is

$$P = \begin{bmatrix} 1 & -\frac{1}{2} & -1 \\ 0 & \frac{1}{2} & -2 \\ 0 & 0 & 1 \end{bmatrix}.$$

Thus, the bilinear form g on $\mathbb{R}^3$ given by

$$g(X, Y) = 2x_1y_1 - \tfrac{1}{2}x_2y_2 - 4x_3y_3$$

is congruent to f. The isomorphism T from $\mathbb{R}^3$ to itself which takes e_i^t to the ith column of P is an isomorphism such that $g(X, Y) = f(T(X), T(Y))$ for all $X, Y \in \mathbb{R}^3$. In fact, the substitution $x_1 \rightsquigarrow x_1 - \frac{1}{2}x_2 - x_3$, $x_2 \rightsquigarrow \frac{1}{2}x_2 - 2x_3$, $x_3 \rightsquigarrow x_3$, and $y_1 \rightsquigarrow y_1 - \frac{1}{2}y_2 - y_3$, $y_2 \rightsquigarrow \frac{1}{2}y_2 - 2y_3$, $y_3 \rightsquigarrow y_3$ transforms f to g.

Theorem 5.6.10 *Let f be a bilinear form on a vector space V of finite dimension n over a field F. Then*

$$\rho(L_f) = \rho(A) = \rho(R_f),$$

where A is a matrix of f corresponding to any choice of basis, and ρ denotes the rank.

Proof Let us first observe that matrices of f corresponding to different ordered bases are congruent, and so they all have the same rank. Because of the rank-nullity theorem, it is sufficient to show that $\nu(L_f) = \nu(A) = \nu(R_f)$, where ν denotes the nullity. Now,

$$\nu(L_f) = dim(\{x \in V \mid L_f(x) = f_x = 0\}) = dim(\{x \in V \mid f(x, y) = 0 \text{ for all } y \in V\}).$$

If we fix an ordered basis $\{u_1, u_2, \ldots, u_n\}$ of V, then the map T from V to the vector space F^n, which associates to $x = \Sigma_{i=1}^n x_i u_i$ the column vector X whose ith row is x_i, is an isomorphism, and then $f(x, y) = T(x)^t AT(y) = X^t AY$. This isomorphism takes the subspace $\{x \mid f(x, y) = 0 \text{ for all } y \in V\}$ of V isomorphically to the subspace $\{X \in F^n \mid X^t AY = 0 \text{ for all } Y \in F^n\}$. Thus $\nu(L_f) = dim(\{X \in F^n \mid X^t AY = 0 \text{ for all } Y\})$. Next, over any field if C is a column vector such that $C^t Y = 0$ for all $Y \in F^n$, then $C = 0$ (verify). Hence $\{X \in F^n \mid X^t AY = 0 \text{ for all } Y \in F^n\} = \{X \in F^n \mid X^t A = 0\} = \{X \in F^n \mid A^t X = 0\}$. This shows that $\nu(L_f) = \nu(A^t)$. Since A is a square matrix, $\nu(A^t) = \nu(A)$. This shows that $\rho(L_f) = \rho(A)$. Similarly, we can show that $\rho(R_f) = \rho(A)$. ♯

Definition 5.6.11 Let f be a bilinear form on V. Then the common number $\rho(L_f) = \rho(A) = \rho(R_f)$ is called the **rank** of f.

Corollary 5.6.12 *Let f be a bilinear form on a vector space V of dimension n. Then the following conditions are equivalent.*

1. *Rank of f is n.*
2. *$f_x(y) = 0$ for all y implies that $x = 0$.*
3. *$f^y(x) = 0$ for all x implies that $y = o$.* ♯

Definition 5.6.13 A bilinear form f on V is called **non-degenerate**, or **nonsingular** bilinear form if it satisfies any one (and hence all) of the above three conditions in the corollary.

Symmetric Bilinear Forms

Now, we try to describe bilinear forms which has a nice diagonal representation

$$f(x, y) = \Sigma_{i=1}^n a_i x_i y_i$$

with expect to an ordered basis $\{u_1, u_2, \ldots, u_n\}$, $x = \Sigma_{i=1}^n x_i u_i$, and $y = \Sigma_{i=1}^n y_i u_i$. This is equivalent to say that the matrix representation of f with respect to some ordered basis is diagonal. Since matrix representation of a bilinear form with respect to any two bases is congruent, we need to characterize those bilinear forms whose matrix representations are congruent to diagonal matrices. If A is a matrix which is congruent to a diagonal matrix D, then there is a nonsingular matrix P such that $P^t AP = D$ or equivalently, $A = QDQ^t$, where $Q = (P^{-1})^t$. Thus, $A^t = QD^t Q^t = QDQ^t = A$. This shows that A is symmetric matrix. In other words, all matrices associated to the bilinear form f are symmetric. This is so if and only if

$f(x, y) = f(y, x)$ for all $x, y \in V$ (verify). Such a bilinear form is called a **symmetric** bilinear form.

Corollary 5.6.14 *A necessary condition for a bilinear form f to have a diagonal representation is that it is a symmetric bilinear form.* ♯

Let f be a symmetric bilinear form on a finite-dimensional vector space V. A pair of vectors $x, y \in V$ is said to be orthogonal to each other if $f(x, y) = 0$. Let W be a subspace of V. Then $W^{\perp} = \{x \in V \mid f(x, y) = 0 \text{ for all } y \in W\}$ is a subspace of V (verify), and it is called the orthogonal compliment of W with respect to the bilinear form f. Observe that unlike the case of inner product space, $W \bigcap W^{\perp}$ may be different from $\{0\}$. Clearly, $kerL_f = kerR_f = \{x \in V \mid f_x = 0\} = V^{\perp}$. If V is a vector space over the field $\mathbb{R}$ of real numbers, then f is called **positive** (**negative**) if $f(x, x) \geq 0 (\leq 0)$ for all $x \in V$. It is said to be **positive** (**negative**) **definite** if it is positive (negative), and $f(x, x) = 0$ if and only if $x = 0$. To say that f has a diagonal representation with respect to the basis $S = \{x_1, x_2, \ldots, x_n\}$ is to say that S is orthogonal basis in the sense that the members of S are pairwise orthogonal.

The following result is the converse of the above corollary in case the field is of characteristic different from 2.

Theorem 5.6.15 *Let f be a symmetric bilinear form on a finite-dimensional vector space V over a field F of characteristic different from 2. Then there is an orthogonal basis of V, or equivalently, there is an ordered basis with respect to which f has diagonal representation.*

Proof The proof is by the induction on the dimension of V. If dimension of V is 0 or 1, then there is nothing to do. Assume that the result is true for symmetric bilinear forms on vector spaces of dimension n. Let f be a symmetric bilinear form on a vector space V of dimension $n + 1$. If f is zero bilinear form, then there is nothing to do. Suppose that $f \neq 0$. We claim that there is a $x \in V - \{0\}$ such that $f(x, x) \neq 0$. Suppose not. Then $f(x, x) = 0$ *for all* $x \in V$. Now, since f is symmetric bilinear form

$$f(x + y, x + y) - f(x - y, x - y) = 2f(x, y) + 2f(y, x) = 4f(x, y)$$

for all $x, y \in V$. Hence $4f(x, y) = 0$ *for all* $x, y \in V$. Since the field is of characteristic different from 2, $f(x, y) = 0$ *for all* $x, y \in V$. This is a contradiction to the supposition that $f \neq 0$. Let $u_1 \in V - \{0\}$ such that $f(u_1, u_1) \neq 0$. Let $W = \{au_1 \mid a \in F\}$ be the subspace generated by u_1. Then dimension of W is 1. Consider $W^{\perp} = \{v \in V \mid f(u_1, v) = 0\}$. Then $W^{\perp}$ is a subspace of V(verify). Suppose that $au_1 \in W^{\perp}$. Then $0 = f(u_1, au_1) = af(u_1, u_1)$. Since $f(u_1, u_1) \neq 0$, it follows that $a = 0$. Thus, $W \bigcap W^{\perp} = \{0\}$. Further, let $v \in V$. Then

$$f(u_1, v - \tfrac{f(u_1, v)}{f(u_1, u_1)} u_1) = f(u_1, v) - f(u_1, v) = 0.$$

Put $w = v - \frac{f(u_1, v)}{f(u_1, u_1)} u_1$. Then $w \in W^{\perp}$, and

$$v = \frac{f(u_1,v)}{f(u_1,u_1)}u_1 + w$$

belongs to $W + W^\perp$. This shows that $V = W \oplus W^\perp$. Since restriction of f to $W^\perp$ is also symmetric bilinear form on $W^\perp$ and dimension of $W^\perp$ is n, it follows that there is an ordered basis $\{u_2, u_3, \ldots, u_{n+1}\}$ with respect to which f restricted to $W^\perp$ has diagonal representation. In other words $f(u_i, u_j) = 0$ *for all* $i \neq j, i \geq 2$ and $j \geq 2$. Already $f(u_1, u_j) = 0$ *for all* $j \geq 2$. Thus, f has the diagonal representation with respect to the ordered basis $\{u_1, u_2, \ldots, u_{n+1}\}$. ♯

Corollary 5.6.16 *Let A be a symmetric matrix with entries in a field F of characteristic different from* 2. *Then there is a nonsingular matrix P such that P^tAP is a diagonal matrix.* ♯

Remark 5.6.17 The proof of the above theorem and the corollary is algorithmic. It gives an algorithm to reduce a symmetric bilinear form (symmetric matrix) over a field of characteristic different from 2 to a diagonal bilinear form (matrix). An algorithm to reduce a symmetric matrix over a field of characteristic different from 2 congruently to a diagonal matrix is given in the proof of Theorem 2.7.2 which is further illustrated in Example 2.7.5. This also gives an algorithm (see Example 5.6.9) to reduce a symmetric bilinear form to diagonal form.

Corollary 5.6.18 *Let f be a symmetric bilinear form on a finite-dimensional vector space V over the field $\mathbb{C}$ of complex numbers (or over a field F of characteristic different from* 2, *and which contains square root of each of its elements). Then there is a basis $\{u_1, u_2, \ldots, u_n\}$ of V such that*

(i) $f(u_i, u_j) = 0$ *for* $i \neq j$,
(ii) $f(u_i, u_i) = 1$ *for* $i \leq r$, *where r is rank of f, and*
(iii) $f(u_i, u_i) = 0$ *for* $i \geq r + 1$.

More precisely, the matrix of f with respect to some ordered basis is in normal form.

Proof From Theorem 5.6.15, we can find an ordered basis $\{v_1, v_2, \ldots, v_n\}$ such that the matrix of f with respect to this ordered basis is diagonal. Suppose that the rank of f is r. We may assume that the first r diagonal entries are different from 0, and the rest of the diagonal entries are 0. This means that $f(v_i, v_j) = 0$ for $i \neq j, f(v_i, v_i) \neq 0$ for $i \leq r$, and $f(v_i, v_i) = 0$ for $i \geq r + 1$. Since the field contains square root of each of its elements, we have $\sqrt{f(v_i, v_i)} \in \mathbb{C}$. Take $u_i = \frac{1}{\sqrt{f(v_i,v_i)}}v_i$ for $i \leq r$, and $u_j = v_j$ for $j \geq r + 1$. Clearly, the ordered basis $\{u_1, u_2, \ldots, u_n\}$ has the required properties. ♯

Corollary 5.6.19 *Any symmetric matrix over $\mathbb{C}$ is congruent to a matrix in normal form.* ♯

Since any two congruent bilinear forms (matrices) have same rank, we have the following corollary.

Corollary 5.6.20 *Any two symmetric bilinear forms (matrices) over the field $\mathbb{C}$ of complex numbers are congruent if and only if they have same rank.* ♯

The above corollary is not true over the field $\mathbb{R}$ of real numbers. I_n and $-I_n$ have same rank where as they are not congruent over the field $\mathbb{R}$ of real numbers (verify). However, over the field $\mathbb{R}$ of real numbers we have the following results.

Proposition 5.6.21 *Let f be a symmetric bilinear form of rank r on a finite-dimensional real vector space. Let $\{u_1, u_2, \ldots, u_n\}$ be an orthogonal basis of V with $f(u_i, u_i) \neq 0$ for all $i \leq r$, and $f(u_i, u_i) = 0$ for all $i \geq r+1$. Then f is positive (negative) if and only if $f(u_i, u_i) \geq 0$ (≤ 0). It is positive (negative) definite if and only if $f(u_i, u_i) > 0$ (< 0) for all i.*

Proof If $x = \Sigma_{i=1}^n a_i u_i$, then $f(x, x) = \Sigma_{i=1}^n a_i^2 f(u_i, u_i)$. Since a_i^2 is always non-negative, the result follows. ♯

Theorem 5.6.22 (Sylvester) *Let f be a symmetric bilinear form on a real vector space (more generally over a sub field of the field of real numbers in which all positive members have square root). Then there is an ordered basis $\{u_1, u_2, \ldots, u_n\}$ of V, and non-negative integers r, p and q such that*

(i) $f(u_i, u_i) = 1$ for all $i \leq p$,

(ii) $f(u_i, u_i) = -1$ for all $i, p+1 \leq i \leq r$,

(iii) $f(u_i, u_i) = 0$ for all $i \geq r+1$, and

(iv) $f(u_i, u_j) = 0$ for all $i \neq j$. Further, integers r, p, and q are independent of choice of such bases.

Proof From Theorem 5.6.15, it follows that there is an ordered basis $\{v_1, v_2, \ldots, v_n\}$ such that $f(v_i, v_j) = 0$ for $i \neq j$, and $f(v_i, v_i) = 0$ for all $i \geq r+1$, where r is the rank of f. Changing the order of $\{v_1, v_2, \ldots, v_n\}$, we may assume that $f(v_i, v_i) > 0$ for $i \leq p$, and $f(v_i, v_i) < 0$ for $p+1 \leq i \leq r$. Take $u_i = \frac{1}{\sqrt{|f(u_i, u_i)|}} v_i$ for $i \leq r$, and $u_i = v_i$ for all $i \geq r+1$. Then it is clear that $\{u_1, u_2, \ldots u_n\}$ has the required property, where $q = r - p$. It remains to show that r, p and q are independent of the choice of basis. Since r is the rank of f, and which is invariant (congruent matrices have same rank), it is independent of the choice of basis. It is sufficient to show that p is independent of the choice of basis. Let $\{v_1, v_2, \ldots, v_n\}$ be an other ordered orthogonal basis such that $f(v_i, v_i) = 1$ for $i \leq p'$, and $f(v_i, v_i) = -1$ for $p' + 1 \leq i \leq r$. It is clear from the above proposition that $V^\perp$ has $\{u_{r+1}, u_{r+2}, \ldots, u_n\}$ and $\{v_{r+1}, v_{r+2}, \ldots, v_n\}$ as bases. We show that

$$X = \{v_1, v_2, \ldots, v_{p'}, u_{p+1}, u_{p+2}, \ldots, u_r, u_{r+1}, \ldots, u_n\}$$

is linearly independent. Suppose that

$$a_1 v_1 + a_2 v_2 + \cdots + a_{p'} v_{p'} + a_{p+1} u_{p+1} + a_{p+2} u_{p+2} + \cdots + a_r u_r + a_{r+1} u_{r+1} + \cdots + a_n u_n = 0.$$

Then $u + v + w = 0$, where $u = a_1 v_1 + a_2 v_2 + \cdots + a_{p'} v_{p'}$, $v = a_{p+1} u_{p+1} + a_{p+2} u_{p+2} + \cdots + a_r u_r$, and $w = a_{r+1} u_{r+1} + a_{r+2} u_{r+2} + \cdots + a_n u_n$. Clearly, $f(u, u) \geq 0$, $f(v, v) \leq 0$, and $f(x, w) = 0$ for all $x \in V$ (this is because $w \in V^\perp$). Thus,

$$0 = f(u, u+v+w) = f(u,u)+f(u,v)+f(u,w) = f(u,u)+f(u,v),$$

and similarly, since f is symmetric,

$$0 = f(v, u+v+w) = f(v,v)+f(u,v).$$

From the above two equations, $f(u,u) = f(v,v)$. Since f restricted to the subspace generated by $\{v_1, v_2, \ldots, v_{p'}\}$ is positive definite, and f restricted to the subspace generated by $\{u_{p+1}, u_{p+2}, \ldots, u_r\}$ is negative definite, it follows that $f(u,u) = 0 = f(v,v)$. This, in turn, implies that $u = 0 = v$, and so also $w = 0$. Since $\{v_1, v_2, \ldots, v_{p'}\}$, $\{u_{p+1}, u_{p+2}, \ldots, u_r\}$, and $\{u_{r+1}, u_{r+2}, \ldots, u_n\}$ are linearly independent, it follows that X is linearly independent. Hence $n - p + p' \leq n$, and so $p' \leq p$. Interchanging the role of the bases, we see that $p \leq p'$. The result follows. ♯

Remark 5.6.23 p is the largest among the dimensions of subspaces of V over which f is positive definite. Similarly, q is the largest among the dimensions of the subspaces of V over which f is negative definite.

Corollary 5.6.24 *Every real symmetric matrix is congruent to a unique diagonal matrix of the form*

$$Diag(\underbrace{1, 1, \ldots, 1}_{p}, \underbrace{-1, -1, \ldots, -1}_{q}, \underbrace{0, 0, \ldots, 0}_{n-p-q}).$$ ♯

Since for any $r \leq n$, there are $r+1$ pairs of non-negative integers p, q such that $p + q = r$, we have the following corollary.

Corollary 5.6.25 *There are* $\frac{(n+1)(n+2)}{2}$ *distinct congruence classes of* $n \times n$ *real symmetric matrices.* ♯

Definition 5.6.26 If f (A) is a real symmetric bilinear form (matrix), the uniquely determined integer $p - q$ is called the **signature** of f (A).

Real Skew-Symmetric Forms (Matrices)

Recall that a bilinear form f is skew-symmetric if $f(x,y) = -f(y,x)$. It follows that f is skew-symmetric if and only if its matrix with respect to any basis is skew-symmetric. If the field F is of characteristic 0, then $f(x,x) = 0$ for all $x \in V$.

Proposition 5.6.27 *Let* f *be a skew-symmetric bilinear form on a finite-dimensional vector space* V *over a field of characteristic 0. Suppose that* $f(x,y) \neq 0$. *Then* $\{x, y\}$ *is linearly independent. Further, if* $z \in W$, *where* W *is the subspace generated by* $\{x, y\}$, *then* $z = \frac{f(z,y)}{f(x,y)}x - \frac{f(z,x)}{f(x,y)}y$.

Proof Suppose that $ax + by = 0$. Then $0 = f(ax+by, x) = af(x,x) + bf(y,x) = -bf(x,y)$. Since $f(x,y) \neq 0$, $b = 0$. Similarly, $0 = f(ax+by, y) = af(x,y) + bf(y,y) = af(x,y)$. Again, since $f(x,y) \neq 0$, $a = 0$. This proves that

$\{x, y\}$ is linearly independent. Next, if $z = ax + by$, then $f(z, x) = af(x, x) + bf(y, x) = -bf(x, y)$. Thus, $b = -\frac{f(z,x)}{f(x,y)}$. Similarly, $a = \frac{f(z,y)}{f(x,y)}$. Hence

$$z = \tfrac{f(z,y)}{f(x,y)}x - \tfrac{f(z,x)}{f(x,y)}y. \qquad \sharp$$

Proposition 5.6.28 *Under the hypothesis of the above proposition* $V = W \oplus W^{\perp}$.

Proof Let $v \in V$ and $w = \frac{f(v,y)}{f(x,y)}x - \frac{f(v,x)}{f(x,y)}y$. Then

$$f(v - w, x) = f(v, x) + \tfrac{f(v,x)}{f(x,y)}f(y, x) = f(v, x) - f(v, x) = 0.$$

Similarly, $f(v - w, y) = 0$. This shows that $v - w \in W^{\perp}$. Thus $V = W + W^{\perp}$. Suppose that $v = ax + by \in W^{\perp}$. Then $0 = f(v, x) = -bf(x, y)$, and $0 = f(v, y) = af(x, y)$. Since $f(x, y) \neq 0$, it follows that $a = 0 = b$. Hence $v = 0$. $\sharp$

Theorem 5.6.29 *Let f be a skew-symmetric bilinear form on a finite-dimensional vector space V over a field F of characteristic 0. Then there is a non-negative integer r, and an ordered basis*

$$\{u_1, v_1, u_2, v_2, \ldots, u_r, v_r, u_{r+1}, u_{r+2}, \ldots, u_n\}$$

of V such that

(i) $f(u_i, v_i) = 1 = -f(v_i, u_i)$ *for all* $i \leq r$,
(ii) $f(u_i, u_j) = 0 = f(v_i, v_j)$ *for all* i, j *and*
(iii) $f(u_i, v_j) = 0$ *for all* $i \neq j$.

Proof The proof is by the induction on the dimension of V. If dimension of V is 0, there is nothing to do. If dimension of V is 1, then also any skew-symmetric bilinear form on V is 0, and there is nothing to do. Assume that the result is true over all vector spaces of dimension less than n. Let V be a vector space of dimension n, and f a skew-symmetric bilinear form on V. If f is 0, then there is nothing to do. Suppose that $f \neq 0$. Then there exists $u_1, v_1 \in V$ such that $f(u_1, v_1) \neq 0$. Multiplying by a suitable scalar to u_1, we may suppose that $f(u_1, v_1) = 1$. From above results, it follows that $\{u_1, v_1\}$ is linearly independent, and if W is the subspace generated by $\{u_1, v_1\}$, then $V = W \oplus W^{\perp}$. Clearly, the dimension of $W^{\perp}$ is $n - 2$, and the restriction of f to $W^{\perp}$ is skew-symmetric. By the induction hypothesis, there exists an ordered basis $\{u_2, v_2, u_3, v_3, \ldots, u_r, v_r, u_{r+1}, u_{r+2}, \ldots, u_n\}$ of $W^{\perp}$ such that

(i) $f(u_i, v_i) = 1 = -f(v_i, u_i)$ for all i, $2 \leq i \leq r$,
(ii) $f(u_i, u_j) = 0 = f(v_i, v_j)$ for all $i, j \geq 2$ and
(iii) $f(u_i, v_j) = 0$ for all $i \neq j$, $i, j \geq 2$.

Clearly, $\{u_1, v_1, u_2, v_2, \ldots, u_r, v_r, u_{r+1}, u_{r+2}, \ldots, u_n\}$ has the required properties. $\sharp$

Corollary 5.6.30 *The rank of a skew-symmetric bilinear form on a vector space over a field of characteristic 0 is always even.* $\sharp$

Following is the matrix form of the theorem.

Corollary 5.6.31 *Every $n \times n$ skew-symmetric matrix A with entries in a field F of characteristic 0 is congruent to a matrix of the form*

$$\begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix},$$

where $2r$ is the rank of A. ♯

It is clear that there is no nondegenerate skew-symmetric bilinear form on a vector space of odd dimension.

Suppose that f is a nondegenerate skew-symmetric bilinear form on a vector space V of even dimension $2r$ over a field of characteristic 0. Arranging the basis vectors $u_1, v_1, u_2, v_2, \ldots, u_n, v_n$ of the theorem as $u_1, u_2, \ldots, u_n, v_1, v_2, \ldots, v_n$ and looking at the matrix representation, we get the following corollary.

Corollary 5.6.32 *Every nonsingular $2n \times 2n$ skew-symmetric matrix with entries in a field F of characteristic 0 is congruent to a matrix of the form*

$$\begin{bmatrix} 0_n & J \\ -J & 0_n \end{bmatrix},$$

where 0_n is the $n \times n$ zero matrix and

$$J = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
1 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

♯

Quadratic Forms, Orthogonal Reduction

Let V be a finite-dimensional vector space over a field F. A map q from V to F is called a **quadratic form**, if there is a bilinear form f on V such that $q(v) = f(v, v)$

for all $v \in V$. If f is skew-symmetric, then $q(v) = f(v, v) = 0$. We assume that the field is of characteristic different from 2. Then every bilinear form $f = f_s + f_{ss}$, where f_s is symmetric and f_{ss} is skew-symmetric. But, then $q(v) = f_s(v)$. Thus, for every quadratic form q, there is a symmetric bilinear form f such that $q(v) = f(v, v)$ for all $v \in V$. The following proposition says that the symmetric bilinear form is uniquely determined by the quadratic form.

Proposition 5.6.33 *Let q be a quadratic form corresponding to a symmetric bilinear form f. Then*

$$f(v, w) = \tfrac{1}{4}(q(v + w) - q(v - w)).$$

Proof $f(v + w, v + w) - f(v - w, v - w) = 4f(v, w)$. ♯

A quadratic form on a vector space V with respect to an ordered basis $\{u_1, u_2, \ldots, u_n\}$ is given by $q(v) = \Sigma_{i,j} x_i a_{ij} x_j$, where $a_{ij} = f(u_i, u_j)$, f being the symmetric bilinear form representing q, and $v = \Sigma_{i=1}^n x_i u_i$. The following two results follow from the corresponding results on symmetric bilinear forms over the field $\mathbb{C}$ of complex numbers, and over the field $\mathbb{R}$ of real numbers.

Corollary 5.6.34 *Let q be a quadratic form on a vector space V over the field $\mathbb{C}$ of complex numbers. Then there is an ordered basis $\{u_1, u_2, \ldots, u_n\}$ of V such that the representation of q with respect to this basis is*

$$q(v) = x_1^2 + x_2^2 + \cdots + x_r^2,$$

where r is the rank of q (rank of q is defined to be the rank of corresponding f), and $v = \Sigma_{i=1}^n x_i v_i$. ♯

Corollary 5.6.35 *Let q be a quadratic form on a real vector space V. Then there is an ordered basis $\{u_1, u_2, \ldots, u_n\}$ of V such that*

$$q(v) = x_1^2 + x_2^2 + \cdots + x_p^2 - x_{p+1}^2 - x_{p+2}^2 - \cdots - x_r^2,$$

where r is the rank of q, and $2p - r$ is the signature of q(signature of q is defined to be the signature of the corresponding symmetric bilinear form). ♯

Example 5.6.36 Consider the bilinear form f on $\mathbb{R}^3$ given in Example 5.6.9. Its reduced diagonal form and the matrix of transformation P is given in that example. Clearly, the form is further congruent to $x_1y_1 - x_2y_2 - x_3y_3$, and the corresponding matrix is congruent to $diag(1, -1, -1)$. The rank is 3, and the signature is -1. The matrix of transformation is given by $diag(\frac{1}{\sqrt{2}}, \sqrt{2}, 2)P$ (check it). Let q be the quadratic form on $\mathbb{R}^3$ given by

$$q(X) = 2x_1x_2 + 4x_1x_3 + 2x_2x_3.$$

It can be seen easily that the symmetric bilinear form of q is the bilinear form given in Example 5.6.9. The congruent reduction of q to the normal form is

$$q(X) = x_1^2 - x_2^2 - x_3^2.$$

The matrix of transformation is as above. One can obtain the ordered basis of $\mathbb{R}^3$ using the matrix of transformation with respect to which the quadratic form is in reduced form as given above.

Using Corollary 5.5.17 (orthogonal reduction), we get the following proposition.

Proposition 5.6.37 *Let q be a quadratic form on a real inner product space V. Then there is an orthonormal ordered basis $\{u_1, u_2, \ldots, u_n\}$ of V such that*

$$q(v) = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2,$$

where $v = \Sigma_{i=1}^n x_i u_i$, and $\lambda_1, \lambda_2, \ldots, \lambda_n$ are eigenvalues of the matrix of q corresponding to a basis of V. ♯

Remark 5.6.38 Corollary 5.5.17 gives an algorithm to find an orthogonal transformation which reduces the given quadratic form to the diagonal form. This reduction is called an orthogonal reduction.

Surfaces in $\mathbb{R}^3$ Represented by the Equations of Second Degree

A general equation of second degree representing a surface in $\mathbb{R}^3$ is given by

$$f(x, y, z) = ax^2 + by^2 + cz^2 + 2hxy + 2fyz + 2gxz + 2ux + 2vy + 2wz + d = 0.$$

Let us denote the column vector

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

by X. Consider the quadratic form q on $\mathbb{R}^3$ given by

$$q(X) = ax^2 + by^2 + cz^2 + 2hxy + 2fyz + 2gxz.$$

Then $f(x, y, z) = q(X) + 2ux + 2vy + 2wz + d$. The matrix A of the quadratic form q is given by

$$A = \begin{bmatrix} a & h & g \\ h & b & f \\ g & f & c \end{bmatrix},$$

and $q(X) = X^t AX$. Using Corollary 5.5.17, we can find orthogonal matrix O such that $O^t AO = diag(\lambda_1, \lambda_2, \lambda_3)$. Put $X' = O^t X$. Then $X = OX'$. Substituting $X = OX'$, the quadratic form reduces to the form

$$q(X') = \lambda_1 x'^2 + \lambda_2 y'^2 + \lambda_3 z'^2.$$

Suppose that

$$O = \begin{bmatrix} l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \\ l_3 & m_3 & n_3 \end{bmatrix}.$$

The fact that O is orthogonal means that the rows represent direction cosines of three perpendicular axes, and the columns also represent the direction cosines of three perpendicular axes. Further, then $f(x, y, z)$ reduces to

$$f(x', y', z') = \lambda_1 x'^2 + \lambda_2 y'^2 + \lambda_3 z'^2 + 2u'x' + 2v'y' + 2w'z' + d,$$

where $u' = ul_1 + vl_2 + wl_3$, $v' = um_1 + vm_2 + wm_3$, $w' = un_1 + vn_2 + wn_3$. If the quadratic form q is nondegenerate, then all λ_i are nonzero, and then making perfect squares $f(x', y', z')$ it reduces to

$$\lambda_1(x' + \tfrac{u'}{\lambda_1})^2 + \lambda_2(y' + \tfrac{v'}{\lambda_2})^2 + \lambda_3(z' + \tfrac{w'}{\lambda_3})^2 + d',$$

where $d' = d - (\frac{u'}{\lambda_1})^2 - (\frac{v'}{\lambda_2})^2 - (\frac{w'}{\lambda_3})^2$. Substituting $x'' = x' + \frac{u'}{\lambda_1}$, $y'' = y' + \frac{v'}{\lambda_2}$, $z'' = z' + \frac{w'}{\lambda_3}$, the equation reduces to

$$\lambda_1 x''^2 + \lambda_2 y''^2 + \lambda_3 z''^2 = -d'.$$

If $d' = 0$, then it represents a cone. If all λ_i together with d' are positive, then such a surface does not exists. Suppose that all λ_i are positive, and d' is negative. Then it represents an ellipsoid with center given by $x'' = y'' = z'' = 0$, and the principal axes given by the lines $x'' = 0$, $y'' = 0$, $z'' = 0$. Expressing x'', y'', and z'' in terms of x', y', z', and then, in turn, expressing x', y', z' in terms of x, y, z with help of the orthogonal transformation O, we get center, principal axes, and principal planes in terms of original coordinate systems.

If two of λ_i are positive, and the other is negative, and also d' is negative, then it represents a one-sheeted hyperboloid whose invariants can be obtained as above. If one of them is positive, and other two are negative, then it represents two-sheeted hyperboloid.

Next, suppose that $\lambda_3 = 0$. Then the above equation will reduce to $\lambda_1 x''^2 + \lambda_2 y''^2 = 2az''$, or to $\lambda_1 x''^2 + \lambda_2 y''^2 = a$. In case 1 it represents elliptic paraboloid if both λ_1, λ_2 are positive, and it represents hyperbolic paraboloid otherwise. Further, in case 2 it represents elliptic cylinder, or hyperbolic cylinder. If two eigenvalues are 0, then it reduces to the form $x''^2 = 4y''$, or to the form $x''^2 = a$. In case 1 it represents parabolic cylinder, and in case 2 it represents pair of parallel planes.

We illustrate the above discussion by means of an example.

Example 5.6.39 Consider the second-degree equation

$$\tfrac{3}{2}x^2 + y^2 + \tfrac{3}{2}z^2 - xz + x - 1 = 0.$$

The matrix of the quadratic form associated to this equation is the matrix A of Example 5.5.19. Its eigenvalues are 1, 1, 2. If we transform the equation using the orthogonal transformation O (see Example 5.5.19), the equation is transformed to

$$x'^2 + y'^2 + 2z'^2 + \tfrac{1}{\sqrt{2}}x' - \tfrac{1}{\sqrt{2}}y' = 1.$$

Completing the square it reduces to

$$(x' + \tfrac{1}{2\sqrt{2}})^2 + (y' + \tfrac{1}{2\sqrt{2}})^2 + 2z'^2 = \tfrac{5}{4}.$$

This represents ellipsoid with axes $\frac{\sqrt{5}}{2}, \frac{\sqrt{5}}{2}, \sqrt{5}$. The center is given by $x' = -\frac{1}{2\sqrt{2}}$, $y' = -\frac{1}{2\sqrt{2}}$, and $z' = 0$. Since $X = OX'$, substituting the values we get that $x = -\frac{1}{4}$, $y = -\frac{1}{2\sqrt{2}}$, $z = -\frac{1}{4}$. The principal planes are given by $x' = -\frac{1}{2\sqrt{2}} = y'$, and $z' = 0$. Using the transformation $X' = O^tX$, we see that the principal planes are $\frac{x+z}{\sqrt{2}} = \frac{1}{2\sqrt{2}}$, $y = \frac{1}{2\sqrt{2}}$, and $\frac{-x+z}{\sqrt{2}} = 0$.

Exercises

5.6.1 Show that $f(X, Y) = x_1y_1 + 2x_2y_1 + 3x_1y_2 + x_2y_3 + 4x_3y_3$ defines a bilinear form on $\mathbb{R}^3$. Find its matrix representation with respect to the standard basis, and also with respect to the ordered basis $\{e_1 + e_2, e_2 + e_3, e_1 + e_3\}$. Conclude that these two matrices are congruent to each other. Is this bilinear form symmetric? Find its rank. Is this nondegenerate?

5.6.2 Let $V = M_n(F)$ denote the vector space of all $n \times n$ matrices with entries in F. Define a map f from $V \times V$ to F by $f(A, B) = Tr(A^tCB)$, where C is a fixed matrix. Is this symmetric? Find its rank in terms of the matrix C.

5.6.3 Determine which of the following define a bilinear form on $\mathbb{R}^3$.
(i) $f(X, Y) = x_1^2 + y_1^2 + x_1y_2 + x_3y_3$.
(ii) $f(X, Y) = 2$ for all $X, Y \in \mathbb{R}^3$.
(iii) $f(X, Y) = x_1x_2 + y_1y_2 + x_1y_3$.
(iv) $f(X, Y) = x_1y_3 - x_2y_2 + x_3y_2$.

5.6.4 Let V be the vector space of 3×3 matrices over $\mathbb{R}$ and

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 4 \\ 3 & 4 & 5 \end{bmatrix}.$$

Define a map f from $V \times V$ to $\mathbb{R}$ by $f(X, Y) = Tr(X^tAY)$. Show that f is a bilinear form. Is it a symmetric bilinear form (observe that A is symmetric). Find the matrix of f relative to the ordered basis $\{e_{11}, e_{12}, e_{13}, e_{21}, e_{22}, e_{23}, e_{31}, e_{32}, e_{33}\}$, where e_{ij} is the matrix whose ith row jth column entry is 1 and the rest of the entries are 0. Find its rank.

5.6.5 Let V be as above. Define a map f from $V \times V$ to $\mathbb{R}$ by $f(A, B) = Tr(AB) + Tr(A)Tr(B)$. Show that it is a symmetric bilinear form. Find its rank and signature.

5.6.6 Show that a bilinear form on V is product of linear functionals if and only if it is of rank 1.

5.6.7 Let f be a nondegenerate form on a finite-dimensional vector space, and g a bilinear form. Show that there exists a unique linear transformation T_1 on V given by $g(v, w) = f(T_1(v), w)$, and also there exists a unique linear transformation T_2 on V such that $g(v, w) = f(v, T_2(w))$.

5.6.8 Reduce the following symmetric bilinear forms on $\mathbb{C}^3$ congruently to the normal form, and find the matrices of transformations.

(i) $f(X, Y) = x_1y_2 + ix_1y_3 + x_2y_1 + ix_3y_1$.
(ii) $g(X, Y) = (1 + i)x_1y_1 + x_1y_3 + x_3y_1 + ix_2y_3 + ix_3y_2$.
(iii) $h(X, Y) = x_1y_3 + x_2y_3 + x_3y_1 + x_3y_2$.

5.6.9 Check if the following pairs of bilinear forms on $\mathbb{C}^3$ are congruent.

(i) (f, g).
(ii) (g, h).
(iii) (f, h),

where f, g, h are bilinear forms defined in Exercise 5.6.8.

5.6.10 Reduce the following complex symmetric matrix congruently to a matrix in normal form.

$$\begin{bmatrix} 1 & 2 & i \\ 2 & 4 & 7 \\ i & 7 & -i \end{bmatrix}.$$

5.6.11 Reduce the following symmetric bilinear forms over $\mathbb{R}^3$ congruently to normal form. Find the matrix of transformation in each case, and also rank and signatures:

(i) $f(X, Y) = 2x_1y_1 + 3x_1y_2 + x_1y_3 + 3x_2y_1 + x_2y_2 - x_2y_3 + x_3y_1 - x_3y_2$.
(ii) $g(X, Y) = x_1y_2 + x_2y_3 + x_2y_1 + x_3y_2$.
(iii) $h(X, Y) = x_1y_3 - x_2y_2 + x_3y_1$.

5.6.12 Check if the following pairs of real symmetric bilinear forms are congruent.

(i) (f, g).
(ii) (g, h).
(ii) (f, h),

where f, g, h are as in the above exercise.

5.6.13 Find the ranks and signatures of the following matrices by congruently reducing them in to the standard canonical forms.

(i)

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 3 & 4 \\ 2 & 4 & 6 \end{bmatrix}$$

(ii)

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$$

(iii)

$$C = \begin{bmatrix} 3 & 4 & 5 \\ 4 & 5 & 6 \\ 5 & 6 & 7 \end{bmatrix}$$

5.6.14 Determine which of the following pair of matrices are congruent.
(i) (A, B).
(ii) (A, C).
(iii) (B, C),
where $A,\ B,\ C$ are as in the above

5.6.15 Can we have a nondegenerate skew-symmetric bilinear form on a complex vector space of dimension 3? Support.

5.6.16 Find the number of congruence classes of skew-symmetric bilinear forms on a real vector space of dimension 3.

5.6.17 Show that any two nondegenerate skew-symmetric $2n \times 2n$ real matrices are congruent.

5.6.18 Reduce, orthogonally, the following quadratic forms in to standard canonical form.
(i) $y^2 + z^2 + yz + zx - xy$.
(ii) $4x^2 + 3y^2 + 2z^2 + 4yz - 4xy$.
(iii) $xy + yz + zx$.

5.6.19 Find the eigenvalues of the real symmetric matrix A, and also an orthogonal matrix O such that O^tAO is diagonal, where

$$A = \begin{bmatrix} 2 & -2 & 5 \\ -2 & -1 & 10 \\ 5 & 10 & -22 \end{bmatrix}.$$

5.6.20 Reduce the following surfaces in to standard form, find their nature, and also their invariants such as center, principal axis, and principal planes.

(i) $y^2 + z^2 + yz + zx - xy - 2x + 2y - 2z + 1 = 0$.
(ii) $4x^2 + 3y^2 + 2z^2 + 4yz - 4xy - 4x - 6y - 8z - 6 = 0$.
(iii) $3x^2 + 6y^2 + 11z^2 + 8yz + 10zx + x - y + z - 4 = 0$.

5.6.21 Let f be a nondegenerate bilinear form on a vector space V. Let $O(f)$ denote the set of all linear transformations T which preserve f in the sense that $f(T(x), T(y)) = f(x, y)$. Show that $O(f)$ is a group.

5.6.22 The bilinear form $< >_L$ on $\mathbb{R}^{n+1}$ defined by

$$< \overline{x}, \overline{y} >_L = x_1y_1 + x_2y_2 + \cdots + x_ny_n - x_{n+1}y_{n+1}$$

is called the **Lorentz inner product**. Show that $< >_L$ is a symmetric nondegenerate bilinear form of signature $n - 1$. The transformations preserving Lorentz inner product are called the **Lorentz transformations**. Show that the set $O(n, 1)$ of all Lorentz transformations form a group under composition of maps. This group is called the **Lorentz Group**.

5.6.23 A set $\{\overline{u_1}, \overline{u_2}, \ldots, \overline{u_r}\}$ is called Lorentz orthonormal set if $< \overline{u_i}, \overline{u_j} >_L = 0$, and $< \overline{u_i}, \overline{u_i} >_L = \pm 1$. Use the Selvester's law to show that at most one i will be such that $< \overline{u_i}, \overline{u_i} >_L = -1$.

5.6.24 Show that every Lorentz orthonormal set is linearly independent. Show also that any Lorentz orthonormal set can be enlarged to a Lorentz orthonormal basis.

5.6.25 A $(n + 1) \times (n + 1)$ matrix A is called **Lorentz matrix** if $AJA^t = J$, where $J = Diag(1, 1, \ldots, 1, -1)$. Show that a linear transformation is a Lorentz transformation if and only if its matrix representation with respect to any Lorentz orthonormal basis is a Lorentz matrix.

5.6.26 Show that the determinant of a Lorentz matrix is ± 1.

5.6.27 Call a Lorentz matrix a **positive Lorentz** matrix if $< A\overline{e_{n+1}}, \overline{e_{n+1}} > > 0$. Show that the set $PSO(n, 1)$ of positive Lorentz matrices of determinant 1 is a group under the product of matrices. This group is called the special positive Lorentz group.

5.6.28 Let $A \in PSO(1, 1)$. Show that there is a unique $x \geq 0$ such that

$$A = \begin{bmatrix} coshx & sinhx \\ sinhx & coshx \end{bmatrix}.$$

Describe the group $PSO(1, 1)$.

5.6.28*. Try to describe the geometry of Lorentz transformations. More explicitly, show that any matrix $A \in PSO(n, 1)$ is similar to a matrix of the form

$$\begin{bmatrix} B & 0_{(n-1)\times 2} \\ 0_{2\times(n-1)} & C \end{bmatrix},$$

where $B \in PSO(n - 1)$, and $C \in SO(1, 1)$. Compare with the geometry of orthogonal transformation.

Chapter 6
Canonical Forms, Jordan and Rational Forms

In the previous chapter, we studied congruence classes of matrices over some special type of fields. This chapter is devoted to describe similarity classes of matrices with entries in some special type of fields. For the purpose, we first introduce the concept of a module over a ring, and obtain the structure theory of modules over a principal ideal domain. The reader is referred to Algebra 1 for the definition and some basic properties of rings.

6.1 Concept of a Module over a Ring

A module over a ring R is a structure obtained by replacing a field F in the definition of a vector space over F by a ring R. Thus,

Definition 6.1.1 Let R be a ring with identity 1. A **left R-module** is an abelian group $(M, +)$ together with a map $\cdot$ from $R \times M$ to M (the image of (a, x) under $\cdot$ is denoted by $a \cdot x$) such that

(i) $(a + b) \cdot x = a \cdot x + b \cdot x$
(ii) $a \cdot (x + y) = a \cdot x + a \cdot y$
(iii) $(ab) \cdot x = a \cdot (b \cdot x)$
(iv) $1 \cdot x = x$

for all $a, b \in R$ and $x, y \in M$.

In the similar manner we can define **right modules**.
We also say that M is a left(right)R-module or M is a left(right) module over R.

Remark 6.1.2 If a left R-module structure on M is such that $(ab) \cdot x = (ba) \cdot x$ for all $a, b \in R$, and $x \in M$, then this left R-module M can also be viewed as a

R. Lal, *Algebra 2*, Infosys Science Foundation Series in Mathematical Sciences,
DOI 10.1007/978-981-10-4256-0_6

right R-module by defining $x \cdot a = a \cdot x$. In particular, if R is a commutative ring, then every left R-module can also be considered as a right R-module. In this case we simply say that M is a R-module.

A module over a field F is simply a **vector Space** over F.

We shall develop the theory of left modules. The theory of right modules can be developed on the same lines.

Let M be a left R-module. Let $a \in R$. Define a map f_a from M to M by $f_a(x) = a \cdot x$. Since $a \cdot (x + y) = a \cdot x + a \cdot y \, for \, all \, x, \, y \in M$, $f_a \in End(M, +)$. Thus, we have a map f from R to $End(M, +)$ given by $f(a) = f_a$. Since

$$f_{a+b}(x) = (a + b) \cdot x = a \cdot x + b \cdot x = f_a(x) + f_b(x) = (f_a + f_b)(x),$$

and

$$f_{ab}(x) = (ab) \cdot x = a \cdot (b \cdot x) = f_a(f_b(x)) = (f_a of_b)(x)$$

for all $a, b \in R$, and $x \in M$, we see that

$$f(a + b) = f(a) + f(b), \text{ and } f(ab) = f(a) of(b)$$

for all $a, b \in R$. Also

$$f(1)(x) = f_1(x) = 1 \cdot x = x = I_M(x)$$

for all $x \in M$. Hence $f(1) = I_M$, the identity of the ring $End(M, +)$. It turns out that every left R-module M gives rise to a ring homomorphism f from R to $End(M, +)$ defined by $f(a)(x) = a \cdot x$.

Conversely, given any ring homomorphism f from R to $End(M, +)$ (a ring homomorphism is assumed to preserve identity), $(M, +)$ becomes an R-module with respect to the external product $\cdot$ given by $a \cdot x = f(a)(x)$, and this in turn, induces the same homomorphism f. Thus, a left R-module can be viewed as a representation of R in a ring of endomorphism of an abelian group.

Similarly, a right R-module determines and is determined uniquely by an anti-homomorphism f from R to $End(M, +)$ in the sense that $f(ab) = f(b) of(a)$ for all $a, b \in r$.

Let M be a left R-module and $a \in R$. Then the map f_a is group homomorphism from $(M, +)$ to itself. Thus,

$$a \cdot 0 = f_a(0) = 0, \text{ and } a \cdot (-x) = f_a(-x) = -(f_a(x)x) = -a \cdot x$$

for all $a \in R$ and $x \in M$.

Also, since the map $f : R \longrightarrow End(M, +)$ defined by $f(a) = f_a$ is a homomorphism, $f(0) = f_0$ is the zero map. Thus, we have

$$0 \cdot x = f_0(x) = 0,$$

and

$$(-a) \cdot x = f_{-a}(x) = -f_a(x) = (-a \cdot x)$$

for all $a \in R$ and $x \in M$.

Remark 6.1.3 Unlike in vector spaces, $\alpha \cdot x = 0$ need not imply that $[\alpha = 0 \, or \, x = 0]$. If in a module this implication holds, then we say that the module is **torsion free** module. We say that it is **torsion module** if for each $x \neq 0$, there is an element $a \neq 0$ in R such that $ax = 0$

Example 6.1.4 Every abelian group is naturally a $\mathbb{Z}$ module. It is a torsion-free module if and only if every nonzero element is of infinite order, and it is a torsion module if and only if all elements are of finite order.

Example 6.1.5 Let R be a ring with identity. Then $(R, +)$ is a R-module. Here $\cdot$ is the ring multiplication.

Example 6.1.6 Let R be a ring with identity. Let R^n denote the n times Cartesian product of R with itself. Thus, $R^n = \{(a_1, a_2, \ldots, a_n) \mid a_i \in R\}$. Clearly, R^n is an abelian group with respect to the coordinate-wise addition. Define the external operation $\cdot$ from $R \times R^n$ to R^n by

$$a \cdot (a_1, a_2, \ldots, a_n) = (a \cdot a_1, a \cdot a_2, \ldots, a \cdot a_n).$$

Then $(R^n, +)$ is a left R-module. It can also be made a right R-module.

Definition 6.1.7 Let M be a left R-module. A subset N of M is called a **submodule** of M, *if*
(i) N is a subgroup of $(M, +)$
(ii) The map $\cdot$ from $R \times M$ to M induces a map from $R \times N$ to N. In other words $a \cdot x \in N$ *for all* $a \in R$ *and* $x \in N$.

Thus, a submodule is a module at its own right.

Remark 6.1.8 If we consider a ring R with identity as left(right) module over itself, then the submodules, by definition, are the left(right) ideals of R.

The proofs of the following propositions are imitations of the corresponding results in vector spaces.

Proposition 6.1.9 *Let M be a left R-module. Then a nonempty subset N of M is a submodule if and only if $ax + by \in N$ for all $a, b \in R$, and $x, y \in N$.* ♯

Proposition 6.1.10 *Intersection of a family of submodules is a submodule.* ♯

Proposition 6.1.11 *Let N_1 and N_1 be submodules of a R-module M. Then $N_1 \bigcup N_2$ is a submodule if and only if $N_1 \subseteq N_2$ or $N_2 \subseteq N_1$.* ♯

Proposition 6.1.12 *Let N_1 and N_2 be submodules of a left R-module. Then $N_1 + N_2 = \{x + y \mid x \in N_1, y \in N_2\}$ is also a submodule (called the* **sum** *of N_1 and N_2) which is generated by $N_1 \bigcup N_2$.* ♯

Proposition 6.1.13 *Union of a chain of submodules is a submodule.* ♯

Let M be a left R-module, and S a subset of M. Then the smallest submodule of M containing S exists, and it is in fact the intersection of all submodules of M containing S. This submodule is called the **submodule generated by S** or the **submodule spanned by S**, and it is denoted by $< S >$. If $< S > = M$, then we say that **S generates** M, or S is a **set of generators** of M. A module M is said to be **finitely generated** if it has a finite set of generators.

Thus, $< \emptyset > = \{0\}$.

Definition 6.1.14 Let S be a nonempty subset of a left R-module M. An element $x \in M$ is called a **linear combination** of members of S if

$$x = a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

for some $a_1, a_2, \ldots, a_n \in R$ and $x_1, x_2, \ldots, x_n \in S$.

Proposition 6.1.15 *Let S be a nonempty subset of a left R-module M. Then $< S >$ is the set of all linear combination of members of S.*

Proof Imitate the proof of the corresponding result in vector space case. ♯

A subset S of a left R-module M is called **independent** *if*
(i) $o \notin S$,
and
(ii) given a finite subset $\{x_1, x_2 \ldots, x_n\}$ of S with $x_i \neq x_j$, for $i \neq j$,

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0 \textit{ implies that } a_ix_i = 0 \textit{ for all } i.$$

A subset which is not independent is called a **dependent** set.

Definition 6.1.16 A subset S of a module M is called **linearly independent** if given a finite subset $\{x_1, x_2, \cdots x_n\}$ of S, $x_i \neq x_j$ for $i \neq j$,

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0 \textit{ implies that } a_i = 0 \textit{ for all } i.$$

A subset S which is not linearly independent is called a **linearly dependent** subset.

Clearly, a subset of a linearly independent(independent) set is always linearly independent(independent).

Proposition 6.1.17 *Every linearly independent subset is independent.*

Proof Let S be a linearly independent subset. Then $0 \notin S$, for otherwise $1 \cdot 0 = 0$, whereas $1 \neq 0$. Next, since $a_i = 0$ *for all* i implies that $a_i x_i = 0$ *for all* i, the result follows. ♯

Remark 6.1.18 An independent subset need not be linearly independent subset. For example, consider the $\mathbb{Z}$-module $\mathbb{Z}_6$. Then $\{\overline{2}, \overline{3}\}$ is independent but not linearly independent(verify).

Since in a vector space $a_i x_i = 0$ and $x_i \neq 0$ implies that $a_i = 0$, we have the following proposition.

Proposition 6.1.19 *A subset S of a vector space V is linearly independent if and only if it is independent.* ♯

Proposition 6.1.20 *Union of a chain of linearly independent(independent) subsets is linearly independent(independent).*

Proof Let $\{S_\alpha \mid \alpha \in \Lambda\}$ be a family of linearly independent(independent) subsets. Then $0 \notin S_\alpha$ *for all* α, and hence $0 \notin \bigcup_{\alpha\in\Lambda} S_\alpha$. Let $x_1, x_2, \ldots, x_n$ be distinct elements $\bigcup_{\alpha\in\Lambda} S_\alpha$. Suppose that $x_i \in S_{\alpha_i}$. Since $\{S_\alpha \mid \alpha \in \Lambda\}$ is a chain, *there exists* α_r such that $S_{\alpha_i} \subseteq S_{\alpha_r}$ *for all* i. Thus, $x_1, x_2, \ldots, x_n$ all belong to S_{α_r}. Since S_{α_r} is linearly independent(independent),

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0 \text{ implies that } a_i = 0 (a_i x_i = 0) \text{ for all } i.$$

This shows that the union is linearly independent(independent). ♯

Proposition 6.1.21 *Every linearly independent(independent) subset can be embedded in to a maximal linearly independent(independent) subset.*

Proof Let S be a linearly independent(independent) subset of a module M. Let X be the set of all linearly independent(independent) subsets which contain S. Then $X \neq \emptyset$, for $S \in X$. Thus, $(X, \subseteq)$ is a nonempty partially ordered set. From the previous proposition, it follows that every chain in X has an upper bound. By the Zorn's Lemma X has a maximal element T (say). Clearly T is also maximal linearly independent(independent). ♯

Remark 6.1.22 A maximal linearly independent subset of a module may be far from a set of generators. For example the additive group $(\mathbb{Q}, +)$ is a $\mathbb{Z}$-module. Every singleton $\{\alpha\}$, $\alpha \neq 0$ is a maximal linearly independent: given $\alpha = \frac{m}{n}$, and $\beta = \frac{p}{q}$, $pn\alpha - qm\beta = 0$. We also know that it has no finite set of generators. However, we have the following proposition.

Proposition 6.1.23 *Let S be a maximal independent subset of a left R-module M. Let $x \in M$. Then there exists $\alpha \in R - \{0\}$ such that $\alpha x \in < S >$*

Proof If $x \in S$, then $1x = x \in S$ and $S \subseteq < S >$. If $x = 0$, then $0 = 1 \cdot 0 \in < S >$. Suppose $x \neq 0$, and $x \notin S$. Since S is supposed to be a maximal independent subset, $S \bigcup \{x\}$ is dependent. Thus, there exist $\alpha, \alpha_1, \cdots \alpha_n$ in R, and $x_1, x_2, \cdots x_n$ in S, $x_i \neq x_j$ *for* $i \neq j$ such that

$$\alpha x + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0,$$

where not all of $\alpha x,\ \alpha_1 x_1,\ \alpha_2 x_2,\ \ldots,\ \alpha_n x_n$ are zero. If $\alpha x = 0$, then

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0,$$

where not all of $\alpha_1 x_1, \alpha_2 x_2, \ldots, \alpha_n x_n$ are 0. This contradicts the supposition that S is independent. Thus, $\alpha x \neq 0$, and

$$\alpha x = -\alpha_1 x_1 - \alpha_2 x_2 - \cdots - \alpha_n x_n$$

belongs to $< S >$. Since $\alpha x \neq 0,\ \alpha \neq 0$. ♯

Proposition 6.1.24 *Let M be a left R-module and S a set of generators for M. If T is a subset of M such that S is properly contained in T, then T is dependent.*

Proof If $0 \in T$, then there is nothing to do. Let $x \in T - S,\ x \neq 0$. Since $< S > = M$, there exists $x_1, x_2, \ldots, x_n \in S,\ x_i \neq x_j$ for $i \neq j$ and $\alpha_1, \alpha_2, \ldots, \alpha_n \in R$ such that $x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$. But, then

$$1x + (-\alpha_1)x_1 + (-\alpha_2)x_2 + \cdots + (-\alpha_n)x_n = 0,$$

where $1x \neq 0$. Hence T is dependent. ♯

Definition 6.1.25 Let M be a left R-module. A subset S of M is called a **minimal or irreducible** set of generators of M if $< S > = M$, and no proper subset of S generates M.

Remark 6.1.26 A set of generators need not contain any minimal set of generators. In fact, a module need not have any minimal set of generators. For example, the $\mathbb{Z}$-module $(\mathbb{Q}, +)$ does not have any minimal set of generators. However, we have already seen that this is true in vector spaces.

Direct Sum of Modules

Let $M_1, M_2, \ldots, M_r$ be left R-modules. Then $M = M_1 \times M_2 \times \cdots \times M_r$ is an abelian group with respect to the coordinate-wise addition. If we define the external multiplication $\cdot$ by $\alpha \cdot (x_1, x_2, \ldots, x_r) = (\alpha x_1, \alpha x_2, \ldots, \alpha x_r)$, then M becomes a left R-module. This is called the **external direct sum** of $M_1, M_2, \ldots, M_r$.

A module M is said to be **direct sum**(**internal direct sum**) of its submodules $M_1, M_2, \ldots, M_r$, if every element x of M has a unique representation as

$$x = x_1 + x_2 + \cdots + x_r,$$

where $x_i \in M_i$ *for all* i. The notation

$$M = M_1 \oplus M_2 \oplus \cdots \oplus M_r$$

stands to say that M is direct sum of its submodules $M_1, M_2, \ldots, M_r$.

Proposition 6.1.27 *Let $M_1, M_2, \ldots, M_r$ be submodules of a module M. Then the following conditions are equivalent.*

(1) $M = M_1 \oplus M_2 \oplus \cdots \oplus M_r$.

(2) (i). $M = M_1 + M_2 + \cdots + M_r$, and
(ii). $M_i \bigcap M^i = \{0\}$ for all i, where M^i denotes the submodule $(M_1 + M_2 + \cdots + M_{i-1} + M_{i+1} + M_{i+2} + \cdots + M_r)$.

Proof $1 \Longrightarrow 2$. Assume 1. Since every element x of M has a unique representation as $x = x_1 + x_2 + \cdots + x_r$, 2(i) is evident. Let $x \in M_i \bigcap M^i$. Then $x = x_1 + x_2 + \cdots + x_{i-1} + x_{i+1} + x_{i+2} + \cdots x_r$, where $x_j, j \neq i$ belong to M_j. Thus,

$$0 = x_1 + x_2 + \cdots + x_{i-1} - x + x_{i+1} + x_{i+2} + \cdots + x_r = 0 + 0 + \cdots + 0.$$

From the uniqueness of the representation of 0, it follows that each x_j is zero, and so x is also 0.

$2 \Longrightarrow 1$. Assume 2. Since every element of M is sum of elements of $M_1, M_2, \ldots, M_r$, it follows that every element x of M has a representation $x = x_1 + x_2 + \cdots + x_r$, where $x_i \in M_i$ for all i. Now, we prove the uniqueness of the representation. Suppose that

$$x = x_1 + x_2 + \cdots + x_r = y_1 + y_2 + \cdots + y_r,$$

where $x_i, y_i \in M_i$ for all i. Then $x_i - y_i \in M_i \bigcap M^i = \{0\}$. This shows that $x_i = y_i$ *for all* i. ♯

Remark 6.1.28 If M is direct sum of $M_1, M_2, \ldots, M_r$, then M^i as defined in the above proposition is direct sum of $M_1, M_2, \ldots, M_{i-1}, M_{i+1}, M_{i+2}, \ldots, M_r$.

Quotient Modules

Let M be a left R-module, and N a submodule of M. Then N is a subgroup of $(M, +)$. Consider the quotient group

$$M/N = \{x + N \mid x \in M\}.$$

Here the coset $x + N = \{x + n \mid n \in N\}$, and the addition is defined by

$$(x + N) + (y + N) = (x + y) + N.$$

Define the external multiplication $\cdot : R \times M/N \longrightarrow M/N$ by

$$\alpha \cdot (x + N) \;=\; (\alpha \cdot x) + N.$$

Then M/N is a left module, and it is called the **quotient module** of M modulo N.

Remark 6.1.29 In general, submodule of a finitely generated module need not be finitely generated. For example, the polynomial ring $\mathbb{Z}[X_1, X_2, \ldots]$ over $\mathbb{Z}$ in infinitely many variables is a module over itself which is generated by the identity of the ring, whereas the submodule generated by the set of all indeterminate is not finitely generated(verify).

Module Homomorphisms, Isomorphisms

Let M_1 and M_2 be left R-modules over a ring R. A map f from M_1 to M_2 is called a R-module homomorphism if $f(ax + by) \;=\; af(x) + bf(y)$ for all $a, b \in R$ and $x, y \in M_1$. A bijective homomorphism is called an isomorphism. The proofs of the following results are imitation of the proofs of the corresponding results in vector space theory, and are left as simple exercises.

Proposition 6.1.30 *Let f be a homomorphism from a left R-module M_1 to a left R-module M_2. Then,*

(i) $f(0) \;=\; 0$,
(ii) $f(-x) \;=\; -f(x)$ *for all* $x \in M_1$,
(iii) the image of a submodule of M_1 under the map f is a submodule of M_2, and
(iv) the inverse image of a submodule of M_2 under the map f is a submodule of M_1.

♯

In particular, $f^{-1}(\{0\})$ is a submodule of M_1, called the kernel of the homomorphism, and it is denoted by *kerf*.

Proposition 6.1.31 *A homomorphism f is injective if and only if* $kerf \;=\; \{0\}$. ♯

Theorem 6.1.32 (Fundamental theorem of homomorphism). *Let f be a homomorphism from a left R-module M_1 to a left R-module M_2. Let N_1 be a submodule of M_1. Then there exists a homomorphism $\overline{f}$ from M_1/N_1 to M_2 such that $\overline{f} o\nu \;=\; f$ if and only if $N_1 \subseteq kerF$. Also if such a homomorphism exists, then it is unique. Further, then $\overline{f}$ is injective if and only if $N_1 \;=\; kerf$. Finally, $\overline{f}$ is an isomorphism if and only if f is surjective, and $N_1 \;=\; kerf$.* ♯

Theorem 6.1.33 (Noether isomorphism theorem). *Let N_1 and N_2 be submodules of a left R-module M. Then $(N_1 \;+\; N_2)/N_2$ is isomorphic to $N_1/N_1 \bigcap N_2$.* ♯

Proposition 6.1.34 *Let M_1 and M_2 be left R-modules and f a homomorphism from M_1 to M_2. Then,*

(i) f is surjective if and only if it takes a set of generators to a set of generators.
(ii) f is injective if and only if it takes an independent set to an independent set. ♯

6.2 Modules over P.I.D

In this section, we shall be mainly interested in finitely generated modules. Let M be a finitely generated R-module. We say that a finite subset $S = \{x_1, x_2, \ldots, x_n\}$ is a basis of M if every element $x \in M$ can be written uniquely as

$$x = a_1x_1 + a_2x_2 + \cdots + a_nx_n.$$

This amounts to say that S generates M, and S is linearly independent.

A R-module M is said to be a **free** module over R if it has a basis.

Thus, every vector space V (a module over a field) over a field F is a free F-module. This is not true for modules over an arbitrary ring. For example, $\mathbb{Z}_2$ is a $\mathbb{Z}_4$-module but not free $\mathbb{Z}_4$-module.

Proposition 6.2.1 *Let M be a free R-module with S as a basis. The every map f from S to a R-module N can be extended uniquely to a homomorphism from M to N.*

Proof Let $S = \{x_1, x_2, \ldots, x_n\}$ be a basis of M, and f a map from S to N. Since every element of M can be written uniquely as $a_1x_1 + a_2x_2 + \cdots + a_nx_n$, f can be extended to a map $\overline{f}$ from M to N defined by

$$\overline{f}(a_1x_1 + a_2x_2 + \cdots + a_nx_n) = a_1f(x_1) + a_2f(x_2) + \cdots + a_nf(x_n).$$

Clearly, $\overline{f}$ is a homomorphism which extends f. It is also clear that the definition of $\overline{f}$ is forced, and so it is unique. ♯

Corollary 6.2.2 *A R-module M is free with a basis containing n elements if and only if it is isomorphic to R^n.*

Proof The R-module R^n has $\{e_1, e_2, \ldots e_n\}$ as a basis, where e_i is a row with n columns in which i^{th} column is 1 and the rest of the columns are 0 (This basis is called the standard basis). Thus, R^n is a free R-module. Since an isomorphism takes a basis to a basis, any isomorphic image of R^n is a free R-module with a basis containing n elements. Conversely, if M is a free R-module with a basis $\{x_1, x_2, \ldots, x_n\}$, then the map which takes x_i to e_i extends to an isomorphism from M to R^n. ♯

Remark 6.2.3 Unlike in the case of vector spaces, for arbitrary ring R, R^n isomorphic to R^m does not imply that $n = m$. For example, if we take R to be the ring of endomorphism on an infinite dimensional vector space V, then R^2 is isomorphic to R (verify). However, if R is a P.I.D., then R^n isomorphic to R^m implies that $n = m$. The proof of this fact will follow soon.

Proposition 6.2.4 *Let M be a R-module, and N a submodule such that M/N is free. Then M is isomorphic to $N \oplus M/N$.*

Proof Let $T = \{x_1 + N, x_2 + N, \ldots, x_r + N\}$ be a basis of M/N. Then since $a_1x_1+a_2x_2+\cdots+a_rx_r = 0$ implies that $a_1(x_1+N)+a_2(x_2+N)+\cdots+a_r(x_r+N) = N$ (the zero of M/N), and since T is a basis of M/N, it follows that $a_i = 0$ for all i. Thus, $S = \{x_1, x_2, \ldots, x_r\}$ is a linearly independent subset of M. Let L be the submodule of M generated by S. Then L is free with S as a basis, and the map $x_i \rightsquigarrow x_i + N$ extends an isomorphism from L to M/N. Let $x \in M$. Then $x + N = a_1(x_1 + N) + a_2(x_2 + N) + \cdots + a_r(x_r + N)$ for some $a_1, a_2, \ldots, a_r \in R$. But, then $x - (a_1x_1 + a_2x_2 + \cdots + a_rx_r)$ belongs to N. Hence $M = N + L$. Next, suppose that $y + z = 0$, where $y \in N$ and $z \in L$. Then $z + N = z + y + N = N$, and since $z \rightsquigarrow z + N$ is an isomorphism from L to M/N, it follows that $z = 0$. In turn, $y = 0$. This shows that $M = N \oplus L \approx N \oplus M/N$.

Proposition 6.2.5 *If L and N are free submodules of M such that $M = L \oplus N$, then M is a free module.*

Proof Suppose that S_1 is a basis of L, and S_2 is a basis of N. Then $S_1 \bigcup S_2$ is linearly independent, and also generates M (verify). ♯

Recall that a commutative ring is a principal ideal domain (P.I.D.) if it is integral domain, and every ideal of R is of the form Ra for some $a \in R$. Note that $Ra = Rb$ if and only if a and b differ by a unit in R. For further details see the Algebra 1.

Theorem 6.2.6 *Let R be a principal ideal domain, and M a free R-module with a finite basis containing n elements. Then every nonzero submodule of M is a free module with a basis containing at most n elements.*

Proof The proof is by the induction on n. If $n = 1$, then $M = < x_1 > = Rx_1$, where $x_1 \neq 0$, and $ax_1 = 0$ implies that $a = 0$. The map $a \rightsquigarrow ax_1$ is clearly a R-isomorphism from R to M. Therefore, it is sufficient to show that every nonzero submodule of R considered as a module over R is free. A nonzero submodule of R is an ideal of R, and it is of the form Ra, where $a \neq 0$. Clearly, Ra is free with $\{a\}$ as a basis.

Assume that the result is true for $n < m$. Let M be a free module with $S = \{x_1, x_2, \ldots, x_m\}$ as a basis of M. Then the submodule $< x_1 >$ generated by x_1 is free with $\{x_1\}$ as a basis. Consider the quotient module $L = M/ < x_1 >$, and the quotient map ν. Since S generates M, $\nu(S)$ generates L. Since $\nu(x_1)$ is the zero of L, it follows that $S' = \{\nu(x_2), \nu(x_3) \ldots, \nu(x_m)\}$ also generates L. We show that S' is linearly independent. Suppose that

$$a_2\nu(x_2) + a_3\nu(x_3) + \cdots + a_m\nu(x_m) = < x_1 > .$$

Then

$$a_2x_2 + a_3x_3 + \cdots + a_mx_m + < x_1 > = < x_1 >,$$

or equivalently,

$$a_2x_2 + a_3x_3 + \cdots + a_mx_m = -a_1x_1$$

for some $a_1 \in R$. But, then

$$a_1x_1 + a_2x_2 + \cdots + a_mx_m = 0.$$

Since S is linearly independent, each $a_i = 0$. This shows that L is free, and S' is a basis of L. By the induction hypothesis, every submodule of $L = M/ < x_1 >$ is free with a basis containing at most $m - 1$ elements. Now, let N be a nonzero submodule of M. If $\nu(N)$ is zero submodule of L, then N is a nonzero submodule of $< x_1 >$, and so from the previous case, it is free with a singleton basis. Suppose that $\nu(N)$ is nonzero, and so it is free with a basis containing at most $m - 1$ elements. The restriction ν/N of ν to N is a surjective homomorphism from N to $\nu(N)$ whose kernel is $N \bigcap < x_1 >$. Since $N/N \bigcap < x_1 >$ is a submodule of $M/ < x_1 >$, it is free with a basis containing at most $m - 1$ elements. The result follows from the Propositions 6.2.4 and 6.2.5. ♯

Remark 6.2.7 In fact every submodule of a free R-module is free. The proof uses transfinite induction.

$\mathbb{Z}$-modules are abelian groups, and free $\mathbb{Z}$-modules are called free abelian groups. Since $\mathbb{Z}$ is a P.I.D., we have the following corollary.

Corollary 6.2.8 *Every subgroup of a finitely generated free abelian group is free.* ♯

Proposition 6.2.9 *Every finitely generated module over R is isomorphic to quotient of a free module.*

Proof If M is generated by $\{x_1, x_2, \ldots, x_n\}$, then the map f from R^n to M defined by $f(a_1, a_2, \ldots, a_n) = a_1x_1 + a_2x_2 + \cdots + a_nx_n$ is a surjective homomorphism. Since R^n is free, the result follows from the fundamental theorem of homomorphism. ♯

Corollary 6.2.10 *Every submodule of a finitely generated module over a P.I.D. is finitely generated.*

Proof Let $M = < \{x_1, x_2, \ldots, x_n\} >$ be a finitely generated module. Then the map f from R^n to M defined by $f(a_1, a_2, \ldots, a_n) = a_1x_1 + a_2x_2 + \cdots + a_nx_n$ is a surjective homomorphism from the free module R^n to M. Let N be a submodule of M. Then $f^{-1}(N)$ is a submodule of R^n. Since R^n is a free module with a basis consisting of n elements, it follows that $f^{-1}(N)$ is a free submodule of R^n with a basis containing at most n elements. Hence $N = f(f^{-1}(N))$ is also generated by at the most n elements. ♯

A module M is called a **cyclic** module if M is generated by a single element. Thus, M is cyclic if there exists an element $x \in M$ such that $M = < x > = Rx$.

The map f defined by $f(a) = ax$ is a surjective homomorphism from R to M. Since R is cyclic over R, and homomorphic images of cyclic modules are cyclic modules, it follows that a module M is cyclic if and only if it is homomorphic image of R. Suppose that R is P.I.D., and $M = Rx$ a cyclic module. Let N be a submodule of M. Then $f^{-1}(N)$ is a submodule of R, and so it is an ideal Ra for some $a \in R$. Thus, $f^{-1}(N)$ is cyclic. Hence $N = f(f^{-1}(N))$ is cyclic. This shows that submodule of a cyclic module over a P.I.D. is cyclic. Is this assertion true if R is not a P.I.D.?

Let M be a R-module, where R is a P.I.D. Let $x \in M$. Consider the map f from R to M given by $f(a) = a \cdot x$. Then f is clearly a R-homomorphism. The kernel of f is an ideal of R. Since R is a principal ideal domain, $kerf = Ra$ for some $a \in R$. If $a = 0$, then the kernel of f is $\{0\}$, and in this case the submodule $< a >$ generated by a is isomorphic to R. In this case we say that x is of **period 0**. Thus, x is of period 0 if and only if $ax = 0$ implies that $a = 0$. Such an element is also called a **torsion free** element of M. If $kerf = Ra$ is nonzero, then $a \neq 0$ and $ax = 0$. In this case x is called a **torsion** element, and a, where $kerf = Ra$, is called a **period** of x. If a and b are periods of a torsion element x, then $Ra = Rb$, and so $a \sim b$. Thus, period of a torsion element is unique up to associates. It is clear that a is a period of x in M if and only if $bx = 0$ *if and only if* a/b. A period of an element x is denoted by $o(x)$.

Suppose that $M = < x >$ is a cyclic module generated by x. If x is of period 0, then M is isomorphic to R, and if a period of x is $a \neq 0$, then M is isomorphic to R/Ra. Thus, a cyclic R-module is isomorphic R, or it is isomorphic to R/Ra for some $a \neq 0$. In case of abelian groups ($\mathbb{Z}$-modules), period corresponds to order of the element.

Definition 6.2.11 A module M over a ring R is called a **torsion** module if every element of M is a torsion element. It is said to be torsion free, if every nonzero element of M is torsion free. A module which is neither torsion nor torsion free is called a **mixed** module.

Every finite abelian group is torsion $\mathbb{Z}$-module. A torsion $\mathbb{Z}$-module is also called a periodic group. The additive group $\mathbb{Z}$ of integers is torsion-free $\mathbb{Z}$-module.

Proposition 6.2.12 *Let M be a R-module, and let $T(M)$ denote the set of all torsion element of M. Then $T(M)$ is a torsion submodule of M, and $M/T(M)$ is torsion-free module.*

Proof Suppose that $x, y \in T(M)$, and $a, b \in R$. Since x, y are torsion elements, there exist nonzero elements c and d such that $cx = 0 = dy$. Clearly, then $cd \neq 0$, and $cd(ax + by) = 0$. This shows that $ax + by \in T(M)$. Thus, $T(M)$ is a submodule of M. Next, let $x + T(M)$ be a nonzero element of $M/T(M)$. Then $x + T(M) \neq T(M)$. This means that x is not a torsion element of M. Suppose that $a(x+T(M)) = T(M)$. Then $ax \in T(M)$. Hence there exists $b \neq 0$ such that $bax = 0$. Since x is torsion free, $ba = 0$. Again, since $b \neq 0$, $a = 0$. This shows that $M/T(M)$ is torsion free. ♯

Definition 6.2.13 $T(M)$ is called the **torsion** part of M, and $M/T(M)$ is called the **torsion free** part of M.

If M is finitely generated over a P.I.D., then so are $T(M)$ and $M/T(M)$.

Theorem 6.2.14 *Every finitely generated torsion-free module over a P.I.D. is free.*

Proof Let $S = \{x_1, x_2, \ldots, x_n\}$ be a set of generators of M. We may assume without any loss that each $x_i \neq 0$. Since M is torsion free, $\{x_i\}$ is linearly independent. Let T be a maximal linearly independent subset of S. Without any loss, we may suppose that $T = \{x_1, x_2, \ldots, x_r\}$, $r \leq n$. Let N be a submodule of M generated by T. Then N is free. Since T is maximal linearly independent, $\{x_1, x_2, \ldots, x_r, x_{r+i}\}$ is linearly dependent for all i, $1 \leq i \leq n - r$. Hence, there are $a_1, a_2, \ldots, a_r, a_{r+i}$ in R not all 0 such that

$$a_1x_1 + a_2x_2 + \cdots + a_rx_r + a_{r+i}x_{r+i} = 0.$$

Since T is linearly independent, $a_{r+i} \neq 0$, for otherwise each $a_i = 0$. Also $a_{r+i}x_{r+i} = -a_1x_1 - a_2x_2 - \cdots - a_rx_r$ belongs to N. Let $a = a_{r+1}a_{r+2}\cdots a_n$. Then $a \neq 0$, and $ax_i \in N$ for all i. Since S generates M, $ax \in N$ for all $x \in M$. Define a map f from M to N by $f(x) = ax$. Clearly, this is a module homomorphism. Since M is torsion free, and $a \neq 0$, $ax = 0$ implies that $x = 0$. This means that f is injective. Thus, M is isomorphic to a submodule of N. Since N is free with a finite basis, and submodule of a free module with finite basis is free (Theorem 6.2.6), it follows that M is free. ♯

Corollary 6.2.15 *If M is a finitely generated module over a P.I.D., then $M = T(M) \oplus M/T(M)$.*

Proof Since M is finitely generated, $M/T(M)$ is finitely generated and torsion free. From the above theorem $M/T(M)$ is free. From Proposition 6.2.4, it follows that $M = T(M) \oplus M/T(M)$. ♯

Since every finitely generated free module over R is isomorphic to R^n for some n, we have the following corollary.

Corollary 6.2.16 *Every finitely generated module over a P.I.D. is isomorphic to the direct sum of a finitely generated torsion module and R^n for some n.* ♯

Since every finitely generated torsion abelian group is finite, we have the following:

Corollary 6.2.17 *Every finitely generated abelian group is isomorphic to the direct sum of a finite abelian group with $\mathbb{Z}^n$ for some n.* ♯

Thus, to study the structure of finitely generated modules over principal ideal domains, it is sufficient to study the structure of finitely generated torsion modules over principal ideal domains.

Proposition 6.2.18 *Let M be a finitely generated torsion module. Then there exists $a \neq 0$ such that $ax = 0$ for all $x \in M$.*

Proof Suppose that $M = < \{x_1, x_2, \ldots, x_n\} >$. Let a_i be a period of x_i. Then $a_i x_i = 0$. Let $a = a_1 a_2 \cdots a_n$. Then $a \neq 0$, and $ax = 0$ for all $x \in M$. ♯

Let $A = \{a \in R \mid ax = 0 \text{ for all } x \in M\}$. Then A is an ideal of R, and since R is P.I.D., $A = Rm$ for some $m \in R$. Such a m is called an **exponent** of M. It is clear that exponent of M is unique up to associates.

Let M be a torsion module over R, and p a prime element of R. We say that M is a p-module if given any element $x \in M$, there exists $n \in \mathbb{N}$ such that $p^n \cdot x = 0$.

Let M be a torsion module and p a prime of R. Let $M_p = \{x \in M \mid p^n x = 0 \text{ for some } n \in \mathbb{N}\}$. Then M_p is a submodule of M, and it is called the p-part of M.

Theorem 6.2.19 *Let M be a torsion module, and a an exponent of M. Let $\{p_1, p_2, \ldots, p_n\}$ be a set of primes dividing a such that p_i and p_j are not associate for $i \neq j$, and also each prime divisor of a is an associate of p_i for some i. Then*

$$M = M_{p_1} \oplus M_{p_2} \oplus \cdots \oplus M_{p_n}.$$

Proof Let $x \in M$. Since a is exponent of M, period of x divides a. We may suppose that

$$o(x) = p_1^{t_1} p_2^{t_2} \cdots p_n^{t_n}.$$

Let $q_i = \frac{o(x)}{p_i^{t_i}}$. Then $(q_1, q_2, \ldots, q_n) \sim 1$. Since R is a P.I.D., there exist $u_1, u_2, \ldots, u_n$ in R such that

$$u_1 q_1 + u_2 q_2 + \cdots + u_n q_n = 1.$$

Hence

$$x = u_1 q_1 x + u_2 q_2 x + \cdots + u_n q_n x.$$

Now, $p^{t_i} u_i q_i x = u_i p^{t_i} q_i x = u_i o(x) x = 0$. This shows that $u_i q_i x \in M_{p_i}$. Hence

$$M = M_{p_1} + M_{p_2} + \cdots + M_{p_n}.$$

Further, suppose that

$$x_1 + x_2 + \cdots + x_n = 0,$$

where $x_i \in M_{p_i}$. Suppose that $o(x_i) \sim p_i^{t_i}$. Let $q_i = p_1^{t_1} p_2^{t_2} \cdots p_{i-1}^{t_{i-1}} p_{i+1}^{t_{i+1}} \cdots p^{t_n}$. Then $0 = q_i(x_1 + x_2 + \cdots + x_n) = q_i x_i$. Since $(p_i^{t_i}, q_i) \sim 1$, there exists u_i, v_i such that $u_i p_i^{t_i} + v_i q_i = 1$. Hence $x_i = u_i p_i^{t_i} x_i + v_i q_i x_i = 0$. This shows that the

representation of an element x as sum of elements of M_{p_i} is unique. Hence M is the direct sum $M_{p_1} \oplus M_{p_2} \oplus \cdots \oplus M_{p_n}$. ♯

Now, we describe the structure of finitely generated p-modules, where p is a prime. First observe that if M is a torsion module generated by $\{x_1, x_2, \ldots, x_n\}$, then the exponent of M is l.c.m. of $o(x_1), o(x_2), \ldots o(x_n)$. Thus, if M is a p-module generated by $\{x_1, x_2, \ldots, x_n\}$, where $o(x_i) \sim p^{n_i}$, and m is the maximum of n_i, then p^m will be an exponent of M.

Theorem 6.2.20 *Let M be a finitely generated p-module over a P.I.D.(p a prime). Then M is direct sum*

$$< x_1 > \oplus < x_2 > \oplus \cdots \oplus < x_m >$$

of cyclic modules, where $o(x_i) \sim p^{n_i}$, $n_1 \geq n_2 \geq \cdots \geq n_m$.

Proof (The proof is the imitation of the proof of the Theorem 7.3.1 of the Algebra 1) Let M be a p-module generated by $\{x_1, x_2, \ldots, x_m\}$, where $x_i \neq 0$ for all i. The proof is by the induction on m. If $m = 1$, then $M = < x_1 >$, and then there is nothing to do. Assume that the result is true for $m = r$. We prove it for $r + 1$. Let $S = \{x_1, x_2, \ldots, x_{r+1}\}$ be a set of generators for M, where $x_i \neq 0$ for all i. Suppose that $o(x_i) \sim p^{n_i}$. We may assume that $n_1 \geq n_2 \geq \cdots \geq n_{r+1}$. Thus, p^{n_1} is an exponent of M. Consider the quotient module $M/ < x_1 >$, and the quotient map ν from M to $M/ < x_1 >$. Then $M/ < x_1 >$ is generated by $\{\nu(x_2), \nu(x_3), \ldots, \nu(x_{r+1})\}$. Clearly, $M/ < x_1 >$ is a p-module of exponent p^t, where $t \leq n_1$. By the induction hypothesis,

$$M/ < x_1 > = < \nu(y_2) > \oplus < \nu(y_3) > \oplus \cdots \oplus < \nu(y_s) >$$

for some $y_2, y_3, \ldots, y_s$ in M such that $o(\nu(y_i)) = p^{n_i}$, $n_1 \geq n_2 \geq n_3 \geq \cdots \geq n_s$. We show that *there exists* $z_i \in M$ for all $i \geq 2$ such that $\nu(z_i) = \nu(y_i)$, and $o(z_i) = o(\nu(y_i)) = o(\nu(z_i))$. Since $p^t y_i = 0$ implies that $p^t(\nu(y_i)) = < x_1 >$ (the zero of $M/ < x_1 >$), it follows that $o(\nu(y_i))$ divides $o(y_i)$. Since $o(\nu(y_i)) = p^{n_i}$, $p^{n_i}\nu(y_i) = < x_1 >$. This means that $p^{n_i} y_i \in < x_1 >$. Suppose that $p^{n_i} y_i = p^{t_i} a_i x_1$, where $(p, a_i) \sim 1$, and $t_i \leq n_1$. If $t_i = n_1$, then $p^{n_i} y_i = 0$, and $o(y_i)$ divides p^{n_i}. Hence $o(y_i) \sim p^{n_i}$, and then there is nothing to do. Suppose that $t_i < n_1$. Since $(a_i, p^{n_1}) \sim 1$, *there exist* $u, v \in R$ such that $ua_i + vp^{n_1} = 1$. But, then $x_1 = ua_i x_1$. Hence $o(a_i x_i) = o(x_1) = p^{n_1}$. Thus, $o(p^{n_i} y_i) = o(p^{t_i} x_1) = p^{n_1 - t_i}$. This shows that $o(y_i) = p^{n_1 - t_i + n_i}$. Since p^{n_1} is exponent of M, it follows that $n_1 - t_i + n_i \leq n_1$. Hence $n_i \leq t_i$. Take $z_i = y_i - p^{t_i - n_i} a_i x_1$. Then $\nu(z_i) = \nu(y_i)$, and $o(z_i) = p^{n_i} = o(\nu(z_i))$. Now, we show that

$$M = < x_1 > \oplus < z_2 > \oplus < z_3 > \oplus \cdots \oplus < z_s > .$$

Let $x \in M$. Since $\{\nu(z_2), \nu(z_3), \ldots, \nu(z_s)\}$ generates $M/ < x_1 >$, it follows that $\nu(x) = a_2\nu(z_2) + a_3\nu(z_3) + \cdots + a_s\nu(z_s)$ for some $a_2, a_3, \ldots, a_s$ in R. Hence

$$x - a_2z_2 - a_3z_3 - \cdots - a_sz_s = a_1x_1$$

for some $a_1 \in R$. Thus,

$$x = a_1x_1 + a_2z_2 + a_3z_3 + \cdots + a_sz_s.$$

Next, suppose that

$$a_1x_1 + a_2z_2 + a_3z_3 + \cdots + a_sz_s = 0.$$

Then

$$a_2\nu(z_2) + a_3\nu(z_3) + \cdots + a_s\nu(z_s) = < x_1 > .$$

Since

$$M/ < x_1 > = < \nu(z_2) > \oplus < \nu(z_3) > \oplus \cdots \oplus < \nu(z_s) >,$$

$a_i\nu(z_i) = < x_1 >$ for all $i \geq 2$. But, then $o(z_i) = o(\nu(z_i))$ divides a_i for all $i \geq 2$. Hence $a_iz_i = 0$ *for all* $i \geq 2$. In turn, a_1x_1 is also 0. Thus, every element x can be written uniquely as

$$x = w_1 + w_2 + \cdots + w_s,$$

where $w_1 \in < x_1 >$, $w_i \in < z_i >$ for $i \geq 2$. ♯

Combining the above results, we obtain the following:

Corollary 6.2.21 *Every finitely generated module M over a P.I.D. is isomorphic to direct sum of finitely many cyclic modules, some of them isomorphic to R, and some of them isomorphic to R/Rp^n for different primes p and for different $n \in \mathbb{N}$.* ♯

Proposition 6.2.22 *Let R be an integral domain. Then R^m is isomorphic to R^n as R-modules if and only if $n = m$.*

Proof Let $\{e_1, e_2, \ldots, e_m\}$ be the standard basis of R^m. Let f be an isomorphism from R^m to R^n. Let F be the field of quotients of R. Then f can be extended to a vector space homomorphism $\overline{f}$ from F^m to F^n by

$$\overline{f}(a_1e_1 + a_2e_2 + \cdots + a_me_m) = a_1f(e_1) + a_2f(e_2) + \cdots + a_mf(e_m).$$

It is clear that $\overline{f}$ is injective. Hence $m \leq n$. Similarly, considering f^{-1} we see that $n \leq m$. ♯.

The proof of the following proposition is straightforward verification.

Proposition 6.2.23 *Let R be a principal ideal domain. Then two R-modules M and M' are isomorphic if and only if $T(M)$ is isomorphic to $T(M')$, and $M/T(M)$ is isomorphic to $M'/T(M')$.* ♯

It follows from the Proposition 6.2.22, that there is a unique $n \in \mathbb{N}$ such that $M/T(M)$ is isomorphic to R^n. This n is the rank of M. The following proposition is also easy to observe.

Proposition 6.2.24 *A finitely generated torsion module M over a P.I.D. is isomorphic to M' if and only if M_p is isomorphic to M'_p for all prime p.* ♯

The proof of the following proposition is also an imitation of the proof of Theorem 7.3.3 of Algebra 1.

Proposition 6.2.25 *Let M and M' be finitely generated p-modules. Suppose that*

$$M = < x_1 > \oplus < x_2 > \oplus \cdots \oplus < x_m >,$$

where $o(x_i) \sim p^{r_i}$, $r_1 \geq r_2 \geq \cdots r_m$, and

$$M' = < y_1 > \oplus < y_2 > \oplus \cdots \oplus < y_n >,$$

where $o(y_j) \sim p^{s_j}$, $s_1 \geq s_2 \geq \cdots \geq s_n$. Then M is isomorphic to M' if and only if $m = n$ and $r_i = s_i$ for all i. ♯

Proof Suppose that $m = n$ and $r_i = s_i$ for all i. Then $< x_i > \approx R/Rp^{n_i} \approx < y_i >$ for all i. Further, if $P \approx P'$ and $Q \approx Q'$, then $P \oplus Q \approx P' \oplus Q'$. Thus, $< x_1 > \approx < y_1 >$, $< x_1 > \oplus < x_2 > \approx < y_1 > \oplus < y_2 >$. Proceeding inductively, we find that M is isomorphic to M'. The proof of the converse is by the induction on $max(m, n)$. If $max(m, n) = 1$, then $M = < x_1 >$ is cyclic of exponent p^{r_1}, and $M' = < y_1 >$ is cyclic of exponent p^{s_1}. Since isomorphic modules have same exponent, it follows that $r_1 = s_1$. Assume that the result is true for $max(m, n) = m, n \leq m$. Let

$$M = < x_1 > \oplus < x_2 > \oplus \cdots \oplus < x_{m+1} >,$$

where $o(x_i) \sim p^{r_i}$, $r_1 \geq r_2 \geq \cdots r_{m+1}$, and

$$M' = < y_1 > \oplus < y_2 > \oplus \cdots \oplus < y_k >,$$

where $k \leq m + 1$, $o(y_j) \sim p^{s_j}$, $s_1 \geq s_2 \geq \cdots \geq s_k$ be isomorphic modules. Let σ be an isomorphism from M to M'. Clearly, exponent of M is p^{r_1}, and the exponent of M' is p^{s_1}. Since isomorphic modules have same exponents, it follows that $r_1 = s_1$. Since σ is an isomorphism $p^{r_1} = o(x_1) = o(\sigma(x_1)) = o(y_1)$. Suppose that

$$\sigma(x_1) = \beta_1 y_1 + \beta_2 y_2 + \cdots + \beta_k y_k \cdots . \tag{6.1}$$

Since $o(\sigma(x_1)) = p^{r_1}$, $o(\beta_j y_j) = p^{r_1}$ for some j. After rearranging, we may assume that $o(\beta_1 y_1) = p^{r_1}$. We show that

$$M' = < \sigma(x_1) > \oplus < y_2 > \oplus \cdots \oplus < y_k > .$$

Since p^{r_1} is an exponent of M', and $o(\beta_1 y_1)$ divides $o(y_1)$, it follows that $o(y_1) \sim o(\beta_1 y_1) \sim p^{r_1}$. Hence the $(\beta_1, p^{r_1}) \sim 1$. Since R is a principal ideal domain, there exist $u, v \in R$ such that

$$u\beta_1 + vp^{r_1} = 1.$$

Hence

$$y_1 = u\beta_1 y_1 = u(\sigma(x_1) - \beta_2 y_2 - \beta_3 y_3 - \cdots - \beta_k y_k).$$

Since $\{y_1, y_2, \ldots, y_k\}$ generates M', $\{\sigma(x_1), y_2, \ldots, y_k\}$ also generates M'. Next, suppose that

$$\delta_1 \sigma(x_1) + \delta_2 y_2 + \cdots + \delta_k y_k = 0.$$

Substituting the value of $\sigma(x_1)$ from (6.1), we find that

$$\delta_1 \beta_1 y_1 + (\delta_1 \beta_2 + \delta_2) y_2 + \cdots + (\delta_1 \beta_k + \delta_k) y_k = 0.$$

Since

$$M' = < y_1 > \oplus < y_2 > \oplus \cdots \oplus < y_k >,$$

$$\delta_1 \beta_1 y_1 = (\delta_1 \beta_2 + \delta_2) y_2 = \cdots = (\delta_1 \beta_k + \delta_k) y_k = 0.$$

Since $o(\beta_1 y_1) = p^{r_1}$, p^{r_1} divides δ_1. Hence $\delta_1 y_j = 0$ for all $j \geq 2$. In turn, $\delta_j y_j = 0$ for all $j \geq 2$. Consequently, $\delta_1 \sigma_1$ is also 0. This shows that

$$M' = < \sigma(x_1) > \oplus < y_2 > \oplus \cdots \oplus < y_k > .$$

Since σ is an isomorphism from M to M' such that $\sigma(< x_1 >) = < \sigma(x_1) >$, it induces an isomorphism from $M/ < x_1 >$ to $M'/ < \sigma(x_1) >$. Clearly,

$$M/ < x_1 > \approx < x_2 > \oplus < x_3 > \oplus \cdots \oplus < x_{m+1} >,$$

and

$$M'/ < \sigma(x_1) > \approx < y_2 > \oplus < y_3 > \oplus \cdots \oplus < y_k > .$$

By the induction assumption, $m+1 = k$, and $r_i = s_i$ for all i. ♯

If $\{m_1, m_2, \ldots, m_r\}$ is a set of pairwise co-prime members of R, then it follows, from the Chinese remainder theorem, that R/Rm is isomorphic to $R/Rm_1 \oplus R/Rm_2 \oplus \cdots \oplus R/Rm_r$. Using this fact, and the above results, we obtain the following theorems.

Theorem 6.2.26 *Let M be a finitely generated torsion module over R, where R is a P.I.D. Then there exists an ordered set $\{a_1, a_2, \ldots, a_t\}$ of elements of R such that a_i divides a_{i+1} for all i, and M is isomorphic to*

$$R/Ra_1 \oplus R/Ra_2 \oplus \cdots \oplus R/Ra_t.$$

Further, such an ordered set $\{a_1, a_2, \ldots, a_t\}$ is unique in the sense that if M is also isomorphic to

$$R/Rb_1 \oplus R/Rb_2 \oplus \cdots \oplus R/Rb_s,$$

where b_j divides b_{j+1} for all j, then $t = s$ and a_i is an associate of b_i for all i. ♯

Theorem 6.2.27 *Let M be a finitely generated module over a principal ideal domain R. Then there exists a nonnegative integer n together with an ordered set $\{a_1, a_2, \ldots, a_t\}$ of elements of R such that a_i divides a_{i+1} for all i, and M is isomorphic to*

$$R^n \oplus R/Ra_1 \oplus R/Ra_2 \oplus \cdots \oplus R/Ra_t.$$

Further, n and the ordered set $\{a_1, a_2, \ldots, a_t\}$ is unique in the sense that if M is also isomorphic to

$$R^m \oplus R/Rb_1 \oplus R/Rb_2 \oplus \cdots \oplus R/Rb_s,$$

where b_j divides b_{j+1} for all j, then $m = n$, $t = s$, and a_i is an associate of b_i for all i. ♯

Exercises

6.2.1 Describe all torsion abelian groups of exponent 24 which are generated by at the most three elements.

6.2.2 Describe all torsion $\mathbb{R}[X]$- modules which are of exponents $(X-1)^2(X^2+1)^2$, and which are generated by at the most two elements.

6.2.3 Describe the torsion modules of exponent $(X^2+1)^3$ which cannot be generated by less than three elements.

6.3 Rational and Jordan Forms

Let V be a finite-dimensional vector space over a field F, and T be a linear transformation on V. Then V becomes a $F[X]$-module with respect to the external operation $\cdot$ defined by $f(X) \cdot v = f(T)(v)$ (verify that it is indeed a $F[X]$ module). This module will be referred as $F[X]$ module associated to the linear transformation T. By the Cayley Hamilton theorem $\Phi_T(T) = 0$, where $\Phi_T(X)$ is the characteristic polynomial of T. Hence $\Phi_T(X) \cdot v = \Phi_T(T)(v) = 0$. Thus, V (being finite dimensional)is a finitely generated $F[X]$ module which is torsion module. Since $F[X]$ is a P.I.D., using the structure theory of finitely generated torsion module over P.I.D. developed in the previous section, we study the linear transformation T by looking at its matrix representation with respect to suitable bases. Let $m_T(X)$ be an exponent of the module V. Then $m_T(T) = 0$, and whenever $f(T) = 0$, $m_T(X)$ divides $f(X)$. In particular, $m_T(X)$ divides $\Phi_T(X)$. If $m_T(X)$ is assumed to be monic (leading coefficient 1), then $m_T(X)$ is unique, and it is called the **minimum polynomial** of T.

Proposition 6.3.1 *Let T_1 and T_2 be linear transformations on V. Then T_1 is similar to T_2 if and only if the $F[X]$ module V associated to T_1 is isomorphic to the $F[X]$ module associated to T_2.*

Proof Suppose that $T_2 = PT_1P^{-1}$, where P is a nonsingular linear transformation on V. Then given any polynomial $f(X) \in F[X]$, we have $f(T_2) = Pf(T_1)P^{-1}$. It follows that P is in fact a module isomorphism from the module V associated to T_2 to the module associated to T_1. Conversely, suppose that P is an isomorphism from the $F[X]$ module V associated to T_1 to the $F[X]$ module associated to T_2. Then P is clearly a nonsingular linear transformation on V, and $P(T_1(v)) = P(x \cdot v) = x \cdot P(v) = T_2(P(v))$ for all $v \in V$. Hence $T_2 = PT_1P^{-1}$. ♯

Let T be a linear transformation on a finite-dimensional vector space V over a field F such that the associated $F[X]$ module V is cyclic module generated by $v \in V$. Then it is clear that the period $o(v)$ of v is precisely the minimum polynomial $m_T(X)$ of T. Thus, we have an $F[X]$ module isomorphism η from $F[X]/F[X]m_T(X)$ to the $F[X]$ module V given by $\eta(f(X) + F[X]m_T(X)) = f(X) \cdot v = f(T)(v)$. Suppose that

$$m_T(X) = a_0 + a_1X + a_2X^2 + \cdots + a_{n-1}X^{n-1} + X^n.$$

Every element of $F[X]/F[X]m_T(X)$ is uniquely expressible as $r(X) + F[X]m_T(X)$, where $r(X)$ is a polynomial of degree at most $n-1$. Thus, $\{1 + F[X]m_T(X), X + F[X]m_T(X), \ldots, X^{n-1} + F[X]m_T(X)\}$ is a basis of the vector space $F[X]/F[X]$ $m_T(X)$ over F. Since η (being a $F[X]$ module isomorphism) is a F-isomorphism, and $\eta(X^i + F[X]m_T(X)) = T^i(v)$, it follows that $\{v, T(v), T^2(v), \ldots, T^{n-1}(v)\}$ is a basis of V. Since $m_T(T) = 0$, we have

$$T^n(v) = -a_0v - a_1T(v) - a_2T^2(v) - \cdots - a_{n-1}T^{n-1}(v).$$

Hence the matrix representation of T with respect to this ordered basis is clearly

$$\begin{bmatrix} 0\,0\,0\cdots 0 & -a_0 \\ 1\,0\,0\cdots 0 & -a_1 \\ 0\,1\,0\cdots 0 & -a_2 \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot & \cdot \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot & \cdot \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot & \cdot \\ 0\,0\,0\cdots 0 & -a_{n-2} \\ 0\,0\,0\cdots 1 & -a_{n-1} \end{bmatrix}$$

The above matrix is termed as the **companian** matrix of the polynomial

$$a_0 + a_1X + a_2X^2 + \cdots + a_{n-1}X^{n-1} + X^n.$$

Definition 6.3.2 A matrix A with entries in a field F is said to be in **rational** canonical form if there exists an ordered set $\{f_1(X), f_2(X), \ldots, f_r(X)\}$ of polynomials such that $f_i(X)$ divides $f_{i+1}(X)$ for all i and

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & A_r \end{bmatrix},$$

where A_i is the companion matrix of $f_i(X)$, and each 0 is zero matrix of appropriate order.

Using the Theorem 6.2.26, and above discussion, we obtain the following theorem.

Theorem 6.3.3 *Let T be a linear transformation on V. Then there is a basis of V such that the matrix of T with respect to this basis is in rational canonical form.* ♯

Corollary 6.3.4 *Every square matrix with entries in a field F is similar to a unique matrix in rational canonical form.* ♯

The following example illustrates as to how to find/reduce a matrix having all its eigenvalues in the field to a matrix in rational canonical form. We chose a simple triangular matrix for convenience of the computation.

Example 6.3.5 Consider the matrix A given by

$$A = \begin{bmatrix} 1\,2\,2 \\ 0\,1\,1 \\ 0\,0\,2 \end{bmatrix}.$$

The eigenvalues of A are 1, 1, 2, and the characteristic polynomial $\phi_A(x)$ of A is given by $\phi_A(x) = (x-1)^2(x-2)$. The minimum polynomial $m_A(x)$ is a divisor of the characteristic polynomial having all eigenvalues as roots. Thus, the possibilities for $m_A(x)$ are $(x-1)^2(x-2)$ and $(x-1)(x-2)$. Since $(A-I)(A-2I) \neq 0$, it follows that $m_A(x) = \phi_A(x) = (x-1)^2(x-2)$. Now, $\mathbb{R}^3$ is a $\mathbb{R}[x]$ module associated to the matrix A. The exponent of this module is the minimum polynomial $m_A(x)$. The primes dividing the exponent are $x-1$ with multiplicity 2, and $x-2$ with multiplicity 1. The $x-1$ part $\mathbb{R}^3_{(x-1)}$ of the module is given by $\mathbb{R}^3_{(x-1)} = \{\overline{v} \in \mathbb{R}^3 \mid (x-1)^2 \cdot \overline{v} = \overline{0}\} = \{\overline{v} \mid (A-I)^2\overline{v}^t = \overline{o}\} = \{(v_1, v_2, v_3) \mid v_3 = 0\} = \mathbb{R}^2 \times \{0\}$. It can be easily checked that this is a cyclic submodule generated by $(1, 1, 0)$. This submodule is isomorphic to $\mathbb{R}[x]/\mathbb{R}[x](x-1)^2$. Further, the $x-2$ part $\mathbb{R}^3_{(x-2)}$ is again the null space $\{(4\alpha, \alpha, \alpha) \mid \alpha \in \mathbb{R}\}$ of $A - 2I$, and it is a cyclic submodule of $\mathbb{R}^3$ isomorphic to $\mathbb{R}[x]/\mathbb{R}[x](x-2)$. Indeed, since $(x-1)^2$, and $(x-2)$ are co-prime, $\mathbb{R}^3$ itself is a cyclic module (for example generated by $(5, 2, 1)$), and it is isomorphic to the module $\mathbb{R}[x]/\mathbb{R}[x]m_A(x)$. Thus, the rational form of A is the companion matrix of the minimum polynomial $m_A(X)$ of A. As such the rational form of the matrix is given by

$$\begin{bmatrix} 0 & 0 & 2 \\ 1 & 0 & -5 \\ 0 & 1 & 4 \end{bmatrix}$$

To get the matrix P such that PAP^{-1} is the rational form of A, we need to find the matrix P of transformation from the standard basis to the basis $\{\overline{v}^t, A\overline{v}^t, A^2\overline{v}^t\}$, where $\overline{v} = (5, 2, 1)$ is the generator of the module. Indeed, $\{\overline{v}^t, A\overline{v}^t, A^2\overline{v}^t\}$ are the columns of the matrix P.

Let T be a linear transformation on a finite-dimensional vector space V over a field F. Suppose that $m_T(X) = (X-\lambda)^n$, where $\lambda \in F$, and $F[X]$ module V associated to T is a cyclic module (note that $(X-\lambda)$ is a prime element of $F[X]$) generated by v. Then $V = F[X]v = \{f(X) \cdot v \mid f(X) \in F[X]\}$. Since period of v is $(X-\lambda)^n$, it follows that $f(X) \cdot v$ is uniquely expressible as $r(X) \cdot v$, where $r(X)$ is the remainder obtained when $f(X)$ is divided by $(X-\lambda)^n$. Further, every polynomial $r(X)$ is uniquely expressible as polynomial $s(X-\lambda)$ in $X-\lambda$ which is of same degree as of $r(X)$ (write $r(X) = r(X-\lambda+\lambda)$, and use the binomial theorem). This shows that every element of V is unique F-linear combination of $\{v, (X-\lambda)\cdot v, (X-\lambda)^2 \cdot v, \ldots, (X-\lambda)^{n-1} \cdot v\}$. Also by the definition of $F[X]$-module $(X-\lambda)^i \cdot v = (T-\lambda I)^i(v)$. Hence $\{v, (T-\lambda I)(v), (T-\lambda I)^2(v), \ldots, (T-\lambda I)^{n-1}(v)\}$ is a basis of V. Further $T((T-\lambda I)^i(v)) = (T-\lambda I)^{i+1}(v) + \lambda(T-\lambda)^i(v)$. This shows that the matrix of T with respect to the above ordered basis is the $n \times n$ matrix A given by

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

Such a matrix is called a **Jordan block** of order n. We have established the following proposition.

Proposition 6.3.6 *Let T be a linear transformation on V with minimum polynomial $(X-\lambda)^n$, and it is such that the corresponding $F[X]$-module V is cyclic. Then there is a basis of V with respect to which the matrix of T is a Jordan block as given above.* ♯

Example 6.3.7 Consider the nonidentity uni-upper triangular matrix A given by

$$A = \begin{bmatrix} 1 & \alpha & \beta \\ 0 & 1 & \gamma \\ 0 & 0 & 1 \end{bmatrix}.$$

Clearly, the characteristic polynomial $\phi_A(X)$ of A is $(X-1)^3$. Further, $A - I \neq 0$, and

$$(A-I)^2 = \begin{bmatrix} 0 & 0 & \alpha\gamma \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Suppose that $\alpha \neq 0 \neq \gamma$. Then $(A-I)^2 \neq 0$. This means that $m_A(X) = \phi_A(X) = (X-1)^3$. Let $\overline{v} = [v_1, v_2, v_3]$ be a nonzero vector in $\mathbb{R}^3$. Then $(A-I)\cdot\overline{v}^t = [\alpha v_2 + \beta v_3, \gamma v_3, 0]^t$, and $(A-I)^2\cdot\overline{v}^t = [\alpha\gamma v_3, 0, 0]^t$. This shows that the period $o(\overline{e_3}^t)$ of the column vector $\overline{e_3}^t$ is $(X-1)^3$. Let us consider the $\mathbb{R}[X]$-submodule of $\mathbb{R}^3$ generated by $\overline{e_3}^t$. Since the set

$$\{\overline{e_3}^t, A\cdot\overline{e_3}^t = [\beta, \gamma, 1]^t, A^2\cdot\overline{e_3}^t = [2\beta + 2\gamma, 2\gamma, 1]^t\}$$

is a basis of $\mathbb{R}^3$, it follows that $\mathbb{R}^3$ is a cyclic $\mathbb{R}[X]$-module generated by $\overline{e_3}^t$. Hence A is similar to the Jordan block of order 3 all of whose diagonal entries are 1. In particular, all such 3×3 matrices are similar.

From the structure Theorem 6.2.20 of finitely generated p-module over a P.I.D., and the Proposition 6.3.6, we have the following more general result.

Corollary 6.3.8 *Let T be a linear transformation on a finite-dimensional vector space V of dimension n over a field F such that the minimum polynomial $m_T(X) = (X-\lambda)^{m_1}$. Then there exist integers $m_2 \geq m_3 \geq \cdots \geq m_r$ with $m_1 \geq m_2$, and a basis*

$$\{v_1, v_2, \ldots, v_{m_1}, v_{m_1+1}, v_{m_1+2}, \cdots, v_{m_1+m_2}, v_{m_1+m_2+1}, \ldots, \cdots, v_{m_1+m_2+\cdots+m_r}\}$$

of V with respect to which the matrix of T is

$$\begin{bmatrix} A_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & A_2 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & A_r \end{bmatrix},$$

where A_i is a Jordan block of order m_i all of whose diagonal entries are λ. ♯

Example 6.3.9 Consider the nonidentity uni-upper triangular matrix A given by

$$A = \begin{bmatrix} 1 & 0 & \beta \\ 0 & 1 & \gamma \\ 0 & 0 & 1 \end{bmatrix}.$$

Clearly, the characteristic polynomial $\phi_A(X)$ of A is $(X-1)^3$. Further, $A - I \neq 0$, and $(A-I)^2 = 0$. This means that $m_A(X) = (X-1)^2$. Let $\overline{v} = [v_1, v_2, v_3]$ be a nonzero vector in $\mathbb{R}^3$. Then $(A-I)\cdot \overline{v}^t = [\beta v_3, \gamma v_3, 0]^t$. Since $(\beta, \gamma) \neq (0,0)$, it follows that the period $o(\overline{e_3}^t)$ of $\overline{e_3}^t$ in the corresponding $\mathbb{R}[X]$-module $\mathbb{R}^3$ is $(X-1)^2$. Thus, the $\mathbb{R}[X]$-submodule of $\mathbb{R}^3$ generated by $\overline{e_3}^t$ is the subspace W of $\mathbb{R}^3$ generated by the set

$$\{\overline{e_3}^t, A\cdot\overline{e_3}^t = [\beta, \gamma, 0]^t\}.$$

Clearly, the dimension of W is 2. Consider the vector $[u, v, 0]^t$ such that $\beta v - u\gamma \neq 0$. Then $\{\overline{e_3}^t, [\beta, \gamma, 0]^t, [u, v, 0]^t\}$ is a basis of $\mathbb{R}^3$. Also the period of $[u, v, 0]^t$ is $(X-1)$. The subspace U generated by $[u, v, 0]^t$ is a $\mathbb{R}[X]$ - submodule such that the module $\mathbb{R}^3$ is the direct sum $W \oplus U$. Evidently, the matrix A is similar to the matrix

$$\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix},$$

where A_1 is a Jordan block of order 2 with diagonal entries 1, A_2 is the Jordan block of order 1 with diagonal entry 1, and zeros are the zero matrices of appropriate orders. If P is a matrix with first, second, and the third columns as $\overline{e_3}^t$, $[\beta, \gamma, 0]^t$, and $[u, v, 0]^t$ respectively, then $P^{-1}AP$ is the matrix

$$\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}.$$

Consequently, all such matrices are similar.

Definition 6.3.10 A matrix A is said to be in Jordan canonical form if there exist $\lambda_1, \lambda_2, \ldots, \lambda_r$ all distinct elements in F, and integers

$$m_1, m_2, \ldots, m_{t_1}, m_{t_1+1}, m_{t_1+2}, \ldots, m_{t_2}, \ldots, m_{t_r}$$

such that $m_k \geq m_l$ for all $t_i + 1 \leq k \leq l \leq t_{i+1}$ for all i, and Jordan Blocks $A_1, A_2, \ldots, A_{t_r}$, where A_j is $m_j \times m_j$ Jordan block with diagonal entries λ_i for all $J, t_i + 1 \leq j \leq t_{i+1}$ such that

$$A = \begin{bmatrix} A_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & A_2 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & A_{t_r} \end{bmatrix}$$

The following result is immediate consequence of the structure theorems (Theorems 6.2.19 and 6.2.20) of finitely generated torsion modules over a P.I.D.

Theorem 6.3.11 *Let T be a linear transformation on a vector space V over a field F such that all the characteristic roots of T are in F. Then there is a basis of V with respect to which the matrix of T is in Jordan canonical form.*

Proof Since all the characteristic roots of T are in F, the minimum polynomial $m_T(X)$ of T is given by

$$m_T(X) = (X - \lambda_1)^{m_1}(X - \lambda_2)^{m_2} \cdots (X - \lambda_r)^{m_r},$$

where $\lambda_1, \lambda_2, \ldots, \lambda_r$ are distinct characteristic roots of T. Thus, the $F[X]$ module V associated to T is a finitely generated torsion module of exponent $m_T(X)$. Using the structure theorem of finitely generated torsion module over a P.I.D., together with the above results, we find that there is a basis of V with respect to which the matrix of T is in Jordan canonical form. ♯

Corollary 6.3.12 *Let T be a linear transformation on a finite-dimensional vector space V over an algebraically closed field F. Then there is a basis of V with respect to which the matrix of T is in Jordan canonical form.* ♯

Since the matrices with respect to different bases are similar, we have the following corollary.

Corollary 6.3.13 *Every square matrix with entries in an algebraically closed field is similar to a matrix in Jordan canonical form.* ♯

Following corollary follows from the uniqueness theorem for the decomposition of finitely generated torsion modules over a P.I.D. as direct sum of cyclic p-modules for different primes.

Corollary 6.3.14 *Let A and B be two $n \times n$ matrices with entries in an algebraically closed field F. Then A and B are similar if and only if they are similar to matrices in Jordan canonical forms with same set of Jordan blocks.* ♯

Corollary 6.3.15 *A square matrix A is similar to a diagonal matrix if and only if its minimum polynomial has all its roots distinct.* ♯

We illustrate the reduction of a matrix in to its Jordan canonical form by means of an example.

Example 6.3.16 Consider the matrix

$$A = \begin{bmatrix} 1 & 2 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}$$

of the Example 6.3.5. As already observed, in Example 6.3.5, that the $\mathbb{R}[x]$-module $\mathbb{R}^3$ associated to the matrix A is the direct sum of the cyclic $(x-1)$-submodule $\mathbb{R}^2 \times \{0\}$ with a generator $(1, 1, 0)$ (isomorphic to the direct sum of $\mathbb{R}[x]/\mathbb{R}[x](x-1)^2$), and the cyclic $(x-2)$ submodule $\{(4\alpha, \alpha, \alpha) \mid \alpha \in \mathbb{R}\}$ (isomorphic to $\mathbb{R}[x]/\mathbb{R}[x](x-2)$). As such the representation of the matrix relative to the basis $\{(1, 1, 0)^t, (A - I)(1, 1, 0)^t, (4, 1, 1)^t\}$ is the Jordan canonical form

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

of the matrix A. The matrix P of transformation is the matrix with columns $\{(1, 1, 0)^t, (A - I)(1, 1, 0)^t, (4, 1, 1)^t\}$.

Recall that a linear transformation is said to be a semi-simple linear transformation, or it is said to be a diagonalizable linear transformation if its matrix representation with respect to certain basis is diagonal. T is said to be nilpotent if $T^n = 0$ for some n. A square matrix A is said to be a semi-simple matrix, or it is said to be a diagonalizable matrix if it is similar to a diagonal matrix. It is said to be nilpotent if $A^n = 0$ for some n.

Theorem 6.3.17 (Jordan–Chevalley) *Let T be a linear transformation on a finite-dimensional vector space V over an algebraically closed field F (or at least all characteristic roots of T are in F). Then T can be expressed uniquely as $T = T_s + T_n$, where T_s is semi-simple, T_n is nilpotent, and T_s and T_n commute. Further, there are polynomials $g(X)$ and $h(X)$ without constant terms such that $g(T) = T_s$ and $h(T) = T_n$.*

Proof Suppose that $m_T(X) = \prod_{i=1}^{r}(X-\lambda_i)^{m_i}$, where $\lambda_1, \lambda_2, \ldots, \lambda_r$ are the distinct eigenvalues of T. Since every finitely generated torsion module over a P.I.D. is direct sum of p - submodules for different primes p dividing the exponent of the module, we have $V = V_1 \oplus V_2 \oplus \cdots \oplus V_r$, where V is $F[X]$-module associated to the linear transformation T, and V_i is the $(X-\lambda_i)$-submodule of V. Clearly, $V_i = Ker(T-\lambda_i I)^{m_i}$. Let T_s be the linear transformation defined on V by the requirement that $T_s(x) = \lambda_i x$ for all $x \in V_i$, $1 \leq i \leq r$. Then T_s is clearly a semi-simple linear transformation. Take $T_n = T - T_s$. Then T_n is nilpotent, for the matrix of T_n relative to the basis of V obtained by taking the union of bases of V_i is strictly lower triangular. Thus, $T = T_s + T_n$. We show that T_s and T_n have the required property. Since $\lambda_1, \lambda_2, \ldots, \lambda_r$ are all distinct, the set $\{(X-\lambda_1)^{m_1}, (X-\lambda_2)^{m_2}, \ldots, (X-\lambda_r)^{m_r}\}$ is a set of pairwise co-prime elements of $F[X]$. By the Chinese remainder theorem, there exists a polynomial $g(X)$ such that $g(X) \equiv \lambda_i (mod (X-\lambda_i)^{m_i})$ for all i, and also $g(X) \equiv 0 (mod X)$. Then it is clear that $T_s = g(T)$, and if we take $h(X) = X - g(X)$, then $T_n = T - T_s = h(T)$. Since any two polynomial in T will commute with each other, it follows that T_s and T_n commute with each other. Next, suppose that $T = T_1 + T_2$ is another such decomposition. Then $T_s - T_1 = T_2 - T_n$. Since T_s, T_1, and also T_2, T_n commute, it follows that $T_s - T_1$ is semi-simple as well as nilpotent. But, then $T_s - T_1 = 0$ (note that a diagonal matrix is nilpotent if and only if it is 0). Hence $T_s = T_1$, and so also $T_n = T_2$. ♯

Definition 6.3.18 The linear transformation T_s in the above theorem is called the **semi - simple** part of T, and T_n is called the **nilpotent** part of T.

Corollary 6.3.19 (Jordan–Chevalley) *Let A be a square matrix with entries in an algebraically closed field (or at least all the characteristic roots of A are in F).z Then A can be expressed uniquely as $A = A_s + A_n$, where A_s is diagonalizable, A_n is nilpotent, and A_s and A_n commute. Further, there exist polynomials $g(X)$ and $h(X)$ without constant terms such that $A_s = g(A)$, and $A_n = h(A)$.* ♯

Corollary 6.3.20 *Let T be a linear transformation on a finite-dimensional vector space V over an algebraically closed field F. Then a linear transformation S on V commutes with T if and only if it commutes with its semi-simple and nilpotent parts.*

Proof Clearly if S commutes with T_s and T_n, then it commutes with $T = T_s + T_n$. Conversely, if S commutes with T, then it commutes with $f(T)$ for all polynomials $f(X)$, and since T_s and T_n are polynomials in T, it commutes with T_s as well as with T_n. ♯

Recall that a linear transformation T is unipotent if all of its all characteristic roots are 1.

Corollary 6.3.21 (Multiplicative Jordan–Chevalley theorem) *Let T be a nonsingular linear transformation on a finite-dimensional vector space V over an algebraically closed field F (or at least all characteristic roots of T are in F). Then T is uniquely expressible as $T = T_s T_u$, where T_s is semi-simple, T_u is unipotent, and $T_s T_u = T_u T_s$. Further, T_s and T_u are polynomials in T.*

Proof We know that T is uniquely expressible as $T = T_s + T_n$, where T_s is semi-simple, T_n is nilpotent, $T_sT_n = T_nT_s$, and also T_s, T_n are polynomials in T. Since T is nonsingular, T_s is also nonsingular. Set $T_u = I + T_s^{-1}T_n$. Since T_s and T_n commute, and T_n is nilpotent, it follows that $T_s^{-1}T_n$ is nilpotent. Hence T_u is unipotent. Clearly $T = T_sT_u$. The rest follows from the properties of T_s and T_n.

Application to Differential Equations

Consider the first-order linear differential equation

$$\frac{dx}{dt} = ax.$$

The general solution to the above differential equation is $x(t) = ce^{at}$, where c is an arbitrary constant.

In complete analogy to the above differential equation, we discuss the solution to a system of n-homogeneous first-order linear differential equations with constant coefficients. In the matrix form, this system of equations can be expressed as

$$\frac{d\overline{X}(t)}{dt} = A \cdot \overline{X}(t),$$

where $\overline{X}(t)$ is a smooth column vector point function (smooth function from $\mathbb{R}$ to $\mathbb{R}^n$) and A a $n \times n$ real matrix.

First, let us introduce e^A. Identify $M_n(\mathbb{R})$ with the Euclidean space $\mathbb{R}^{n^2}$ with Euclidean metric. Consider the sequence $\{T_m\}$ of functions from the metric space $M_n(\mathbb{R})$ to itself defined by

$$T_m(A) = I + A + \frac{A^2}{2!} + \cdots + \frac{A^m}{m!}.$$

It can be seen easily that the above sequence is uniformly convergent on any compact subset of $M_n(\mathbb{R})$. Let us denote e^A by

$$Lim_{m\to\infty} T_m(A).$$

This defines a map *exp* from $M_n(\mathbb{R})$ to $M_n(\mathbb{R})$ by $exp(A) = e^A$. Using elementary analysis, we observe that *exp* is continuous, in fact, differentiable, and its Jacobian at 0 is the identity matrix of order n^2. By the inverse function theorem, it follows that *exp* is local diffeomorphism. Again using the Abel's result, we can show that $e^{A+B} = e^A \cdot e^B$ provided that $AB = BA$. In particular, it follows that

$$e^{-A} \cdot e^A = e^{-A+A} = e^0 = I = e^A \cdot e^{-A}.$$

Hence e^A is always nonsingular. Thus, *exp* is a local diffeomorphism from $M_n(\mathbb{R})$ to $GL(n, \mathbb{R})$. The map $t \rightsquigarrow e^{tA}$ is a group homomorphism from $(\mathbb{R}, +)$ to $GL(n, \mathbb{R})$ for all $A \in M_n(\mathbb{R})$. These are called one-parameter family of subgroups of $GL(n, \mathbb{R})$.

Theorem 6.3.22 *The columns of the matrix e^{tA} form a basis of the space of solutions of the system of homogeneous linear differential equations expressed in matrix form by*

$$\frac{d\overline{X}(t)}{dt} = A \cdot \overline{X}(t),$$

where $\overline{X}(t)$ is a smooth column vector point function.

Proof Using the theorem on term by term differentiation of a uniformly convergent series, it follows that

$$\frac{de^{tA}}{dt} = A \cdot e^{tA}.$$

Since $\frac{d[a_{ij}(t)]}{dt} = [b_{ij}(t)]$, where $b_{ij}(t) = \frac{da_{ij}(t)}{dt}$, each column $\overline{Y}_i(t)$ of e^{tA} satisfies

$$\frac{d\overline{Y}_i(t)}{dt} = A \cdot \overline{Y}_i(t).$$

Thus, each column of e^{tA} is a solution of the given system of differential equations. Conversely, suppose that $\overline{X}(t)$ is a solution. Then

$$\frac{de^{-tA}\overline{X}(t)}{dt} = -Ae^{-tA}\overline{X}(t) + e^{-tA}\frac{d\overline{X}(t)}{dt} = -Ae^{-tA}\overline{X}(t) + e^{-tA}A\overline{X}(t).$$

Since A and e^{-tA} commute, it follows that

$$\frac{de^{-tA}\overline{X}(t)}{dt} = 0.$$

Hence $e^{-tA}\overline{X}(t) = \overline{C}$, where $\overline{C}$ is a constant column vector. It follows that $\overline{X}(t) = e^{tA} \cdot \overline{C}$, and so every solution of the system of equations is a linear combination of the columns of e^{tA}. Since e^{tA} is nonsingular the columns are linearly independent. ♯

Thus, the problem is to compute e^{tA}. Let us observe the following:
(i) If $A = Diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$, then $e^A = Diag(e^{\lambda_1}, e^{\lambda_2}, \ldots, e^{\lambda_n})$.
(ii) If

$$A = \begin{bmatrix} 0 & t & 0 & \cdots & 0 \\ 0 & 0 & t & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 0 & t \\ 0 & 0 & \cdot & \cdots & \cdot & 0 \end{bmatrix},$$

then

$$e^A = \begin{bmatrix} 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{n-1}}{(n-1)!} \\ 0 & 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{n-2}}{(n-2)!} \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & 1 & t \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

(iii) If

$$A = \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix},$$

then

$$e^A = \begin{bmatrix} e^{B_1} & 0 \\ 0 & e^{B_2} \end{bmatrix}.$$

(iv) If $A = CBC^{-1}$, then $e^A = Ce^BC^{-1}$.

(v) If A and B commute, then $e^{A+B} = e^A \cdot e^B$.

(vi) If A is a real $n \times n$ matrix, and $\overline{X}(t)$ a complex-valued solution of the system of differential equations

$$\frac{d\overline{X}(t)}{dt} = A \cdot \overline{X}(t),$$

then the real and imaginary parts of $\overline{X}(t)$ are also solutions of the above system of equations.

Using the Jordan–Chevalley decomposition $A = A_s + A_n$, where A_s is similar to a diagonal matrix, and A_n is similar to direct sum of the matrices of the form

$$\begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 0 & 1 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 \end{bmatrix}$$

Further, A_s and A_n commute, and so $e^A = e^{A_s} \cdot e^{A_n}$. Using the above observations, we can compute e^{tA}, and thereby get the general solution of the given homogeneous system of linear differential equations. We illustrate the above discussion by means of an example.

Example 6.3.23 Consider the system of differential equations given in the matrix form by

$$\frac{d\overline{X}(t)}{dt} = A\overline{X}(t),$$

where

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Then $A_s = I$, and

$$A_n = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Since $tA = tA_s + tA_n$, it follows that $(tA)_s = tA_s$, and $(tA)_n = tA_n$. Again, since tA_s and tA_n commute, $e^{tA} = e^{tA_s} \cdot e^{tA_n}$. Clearly, $e^{tA_s} = e^t I$, and as discussed above

$$e^{tA_n} = \begin{bmatrix} 1 & t & \frac{t^2}{2!} \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix}$$

Thus,

$$e^{tA} = \begin{bmatrix} e^t & te^t & \frac{t^2}{2!}e^t \\ 0 & e^t & te^t \\ 0 & 0 & e^t \end{bmatrix}.$$

The columns of the above matrix form a basis of the space of solutions.

Exercises

6.3.1 Consider the following linear transformations on $\mathbb{R}^3$ whose matrix representations with respect to the standard ordered basis are given by

(i)

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix},$$

(ii)

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix},$$

(iii)

$$\begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

(iv)

$$\begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix}, \text{ and}$$

(v)

$$\begin{bmatrix} 1 & 3 & 3 \\ 3 & 1 & 3 \\ -3 & -3 & -5 \end{bmatrix}$$

In each case, find the minimum polynomial, the decomposition of the corresponding $\mathbb{R}[X]$-module $\mathbb{R}^3$ as direct sum of cyclic modules, also a basis of $\mathbb{R}^3$ with respect to which the matrix representation is in rational canonical forms.

6.3.2 Reduce the matrices in the Exercise 6.3.1 to rational canonical form.

6.3.3 Determine the pairs of matrices in Exercise 6.3.1 which are similar.

6.3.4 Reduce the following two matrices over $\mathbb{Z}_5$ into rational canonical forms, and determine if they are similar to each other.

(i)

$$\begin{bmatrix} \overline{1} & \overline{3} & \overline{7} \\ \overline{2} & \overline{0} & \overline{4} \\ \overline{0} & \overline{4} & \overline{1} \end{bmatrix}$$

(ii)

$$\begin{bmatrix} \overline{1} & \overline{2} & \overline{0} \\ \overline{0} & \overline{1} & \overline{3} \\ \overline{1} & \overline{0} & \overline{1} \end{bmatrix}$$

6.3.5 Reduce the matrices in Exercise 6.3.1 in Jordan canonical form considering them as matrices over the field $\mathbb{C}$ of complex numbers. Determine which pairs are similar over $\mathbb{C}$.

6.3.6 Reduce the following complex matrices into Jordan canonical form, and also, in each case, find a nonsingular matrix P such that PAP^{-1} is in Jordan canonical form. Determine which pair of matrices are similar to each other.

(i)

$$\begin{bmatrix} i & 1 & 0 \\ 0 & 2i & -1 \\ 1 & 0 & 1+i \end{bmatrix}$$

(ii)

$$\begin{bmatrix} i & 1+i & 0 \\ 1 & i & 0 \\ 0 & 1 & i \end{bmatrix}$$

(iii)

$$\begin{bmatrix} 0 & 1 & i \\ i & i & 1 \\ 0 & 0 & i \end{bmatrix}$$

6.3.7 Show that a linear transformation S commutes with T if and only it commutes with T_s as well as with T_n.

6.3.8 Let T_1 and T_2 be linear transformations on a vector space V of dimension 3 over a field. Show that the module V over $F[X]$ associated to T_1 is isomorphic to the $F[X]$-module V associated to T_2 if and only if they have same characteristic polynomials and minimum polynomials. Deduce that any two 3×3 matrix over F are similar if and only if they have same characteristic and minimum polynomials. Is this result true for 4×4 matrices? Support.

6.3.9 Let A be a complex matrix all of whose characteristic roots are real. Show that A is similar to a real matrix in Jordan form.

6.3.10 Let A be a $n \times n$ real matrix such that $A^2 + I = 0$. Show that $n = 2r$ is even. Show also that A is similar to

$$\begin{bmatrix} 0_n & -I_n \\ I_n & 0_n \end{bmatrix}$$

6.3.11 Let T be a nilpotent transformation on a complex finite dimensional vector space V. Let $f(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$. Find the semi-simple part of $f(T)$.

6.3.12 Find e^A for

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & 2 \end{bmatrix},$$

and also the solution of the system of linear equations given in matrix form by

$$\frac{d\overline{X}(t)}{dt} = A\overline{X}(t).$$

6.3.13 Reducing the matrix

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

in to Jordan canonical form, find e^A, and then solve the corresponding system of differential equations.

6.3.14 If λ is an eigenvalue of A, then show that e^λ is an eigenvalue of e^A. Deduce from this fact that e^A is nonsingular.

6.3.15 Show that $Det(e^A) = e^{trA}$.

6.3.16 Show that the map *exp* induces a map from the space $sl(n, \mathbb{R})$ of $n \times n$ matrices with trace 0 to the group $SL(n, \mathbb{R})$ of matrices of determinant 1. Show further that it is a local diffeomorphism. Determine the dimension of the group $SL(n, R)$.

6.3.17 Show that the map *exp* induces local diffeomorphism from the space $SS_n(\mathbb{R})$ of skew symmetric matrices to the group $O(n)$ of orthogonal matrices. Determine the dimension of $O(n)$.

6.3.18 Is *exp* surjective from $M_n(\mathbb{R})$ to $GL(n, \mathbb{R})$? Support.

6.3.19 Give an example to show that e^{A+B} need not be $e^A \cdot e^B$.

Chapter 7
General Linear Algebra

The present chapter is devoted to the study of Noetherian rings, Projective modules, Injective Modules, Tensor product of modules, Grothendieck, and Whitehead groups of rings.

7.1 Noetherian Rings and Modules

Over an arbitrary ring, we note that left and right modules are in general distinct. Recall, further, that all subspaces of a finitely generated vector space over a field with a given choice of basis determines and is uniquely determined by a matrix with entries in F. This is a consequence of the fact that every subspace of F^n is finitely generated. However, this fact is not true in general rings. Rings over which all submodules of left module R^n can be described by matrices are essentially left noetherian rings, which we describe in this section. The theory of right noetherian modules and right noetherian rings can be developed exactly on the same lines. A module will always mean a left module, unless stated otherwise.

A module M over R is said to satisfy **ascending chain condition** $(A.C.C)$, if given any chain

$$N_1 \subseteq N_2 \subseteq \cdots \subseteq N_r \subseteq N_{r+1} \subseteq \cdots$$

of submodules of M, $there\ exists\ n_0 \in \mathbb{N}$ such that $N_r = N_{n_0}$ *for all* $r \geq n_0$.

A module M is said to satisfy **maximal condition**, if given any nonempty family $\{M_\alpha \mid \alpha \in \Lambda\}$ of submodules of M, it has a maximal member.

Theorem 7.1.1 *Let M be a module over R. Then the following conditions are equivalent.*

R. Lal, *Algebra 2*, Infosys Science Foundation Series in Mathematical Sciences,
DOI 10.1007/978-981-10-4256-0_7

1. *M satisfies A.C.C.*
2. *Every submodule of M is finitely generated.*
3. *M satisfies maximal condition.*

Proof $1 \Longrightarrow 3$. Let $X = \{M_\alpha \mid \alpha \in \Lambda\}$ be a nonempty family of submodules of M. Suppose that X has no maximal member. Let $M_{\alpha_1} \in X$. Since M_{α_1} is not a maximal member of the family, there is a member $M_{\alpha_2} \in X$ such that $M_{\alpha_1} \subset M_{\alpha_2}$. Again, since M_{α_2} is not a maximal member, there is a member $M_{\alpha_3} \in X$ such that $M_{\alpha_2} \subset M_{\alpha_3}$. Using induction, we arrive at a properly ascending chain of submodules of M. This is a contradiction to 1 (note that we have used axiom of choice in some form).

$3 \Longrightarrow 2$. Assume 3. Let L be a submodule of M. Let X be the family of all finitely generated submodules of L. Clearly, $\{0\} \in X$, and so X is a nonempty family. From 3, it has a maximal member L_0 (say). We claim that $L_0 = L$. Suppose not. Then there is $x \in L - L_0$. But, then $L_0 + < x >$ is also finitely generated, and so it belongs to X. This is a contradiction to the choice of L_0. Thus, L is finitely generated.

$2 \Longrightarrow 1$. Assume 2. Let

$$M_1 \subseteq M_2 \subseteq \cdots \subseteq M_r \subseteq M_{r+1} \subseteq \cdots$$

be an ascending chain of submodules of M. Then $L = \bigcup_{r \geq 1} M_r$ is a submodule. By 2, L is finitely generated. Suppose that $L = < \{x_1, x_2, \ldots, x_n\} >$. Then $x_i \in M_{r_i}$ for some r_i. Let r_0 be the maximum of all r_i. Then $x_i \in M_{r_0}$ for each i. It follows that $L = M_{r_0}$, and so $M_r = M_{r_0}$ *for all* $r \geq r_0$. ♯

A module M over R is said to be **noetherian** module if it satisfies any one, and hence all of the conditions in the above theorem. A ring R is said to be a **noetherian ring** if it is a noetherian module over itself.

If we consider a ring R as a left module over itself, then submodules are precisely the left ideals. Thus, a ring R is a left noetherian ring if and only if all its left ideals are finitely generated.

Since submodule of a submodule is again a submodule of the module, we have

Proposition 7.1.2 *Every submodule of a noetherian module is a noetherian module.* ♯

Proposition 7.1.3 *Any homomorphic image of a noetherian module is noetherian.*

Proof Let $f : M_1 \rightsquigarrow M_2$ be a surjective homomorphism, where M_1 a noetherian module. Let L be a submodule of M_2. Then $f^{-1}(L)$ is a submodule of M_1. Since M_1 is noetherian, $f^{-1}(L)$ is finitely generated. Since image of a finitely generated module under a homomorphism is finitely generated, $L = f(f^{-1}(L))$ (f being surjective $f(f^{-1}(L)) = L$) is finitely generated. Thus, every submodule of M_2 is finitely generated. Hence M_2 is a noetherian module. ♯

The argument used in the proof of the above proposition is valid for rings(inverse image of an ideal under a homomorphism of rings is an ideal), and so we have the following proposition.

Proposition 7.1.4 *Any homomorphic image of a noetherian ring is noetherian.* ♯

Corollary 7.1.5 *Quotient of a noetherian module (ring) is a noetherian module(ring).* ♯

Proposition 7.1.6 *Let M be a module over a ring R. Let L be a submodule of M such that L and M/L are noetherian. Then M is noetherian.*

Proof Let U be a submodule of M. Then, $U + L/L$, being a submodule of M/L, is finitely generated. By the second isomorphism theorem $U/U \bigcap L$ is isomorphic to $U + L/L$. Hence $U/U \bigcap L$ is finitely generated. Further, $U \bigcap L$, being submodule of a northerian module L, is noetherian. Hence $U \bigcap L$ is finitely generated. We know that if N and M/N are finitely generated, then M is also finitely generated. Thus, U is finitely generated. This shows that M is noetherian. ♯

Proposition 7.1.7 *Let $M_1, M_2, \ldots, M_r$ be modules over R. Then $M = M_1 \times M_2 \times \cdots \times M_r$ is noetherian if and only if each M_i is noetherian.*

Proof For each i, the projection p_i is a surjective homomorphism from M to M_i. Sine homomorphic image of a noetherian module is noetherian, if M is noetherian, then each M_i is noetherian. Conversely, suppose that each M_i is noetherian. We have to show that M is noetherian. By the induction, it is sufficient to prove the result for $r = 2$. Suppose that M_1 and M_2 are noetherian. The projection p_2 of M on to M_2 is a surjective homomorphism whose kernel is $M_1 \times \{0\}$. By the fundamental theorem of homomorphism $M/M_1 \times \{0\}$ is isomorphic to M_2. Thus, $M/M_1 \times \{0\}$ is noetherian. Also $M_1 \times \{0\}$ is isomorphic to M_1 (the map $x \rightsquigarrow (x, 0)$ is an isomorphism), and so it is noetherian. From the previous proposition M is noetherian. ♯

Theorem 7.1.8 *Let R be a noetherian ring. Then every finitely generated module over R is a noetherian module.*

Proof Let R be a noetherian ring, and M a finitely generated module over R which is generated by $S = \{x_1, x_2, \ldots, x_r\}$. Define a map η from R^r to M by

$$\eta(\alpha_1, \alpha_2, \ldots, \alpha_r) = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_r x_r.$$

Clearly, η is a surjective homomorphism. Since R is noetherian, from the previous result, R^n is also noetherian. Since homomorphic image of a noetherian module is noetherian, M is noetherian. ♯

Remark 7.1.9 It follows that R is noetherian if and only if every finitely generated module over R is noetherian.

Example 7.1.10 Every P.I.D. is a noetherian ring because every ideal is generated by a singleton. Thus, every finitely generated module over a P.I.D is noetherian. In particular, every submodule of a finitely generated module over a P.I.D is finitely generated. Note that this is not true over an arbitrary ring (give an example). In turn, every subgroup of a finitely generated abelian group ($\mathbb{Z}$-module) is finitely generated.

Example 7.1.11 The polynomial ring $\mathbb{Z}[X_1, X_2, \ldots, X_n \ldots]$ over $\mathbb{Z}$ in countably infinite set of indeterminates $\{X_1, X_2, \ldots, X_n, \ldots\}$ is not a noetherian ring, for the ideal generated by $\{X_1, X_2, \ldots, X_n, \ldots\}$ is not finitely generated. This also shows that, in general, submodule of a finitely generated module need not be finitely generated. Check if it is a U.F.D.

Example 7.1.12 Subring of noetherian ring need not be a noetherian ring: The ring $\mathbb{Z}[X_1, X_2, \ldots, X_n \ldots]$ is an integral domain which is not noetherian. However, its field of fractions is noetherian.

Example 7.1.13 Let R be noetherian integral domain. Then every nonzero nonunit element of R can be written as finite product of irreducible elements of R. To prove this, it is sufficient to show that there is no infinite chain $a_1, a_2, \ldots, a_n, \ldots$ such that a_{n+1} is proper divisor of a_n for all n, or equivalently, there is no infinite properly ascending chain of principal ideals. This is true, for R is a noetherian ring.

Theorem 7.1.14 (Hilbert Basis Theorem). *Let R be a commutative ring with identity. Then the polynomial ring $R[X]$ is noetherian if and only if R is noetherian.*

Proof Suppose that $R[X]$ is noetherian. Define a map η from $R[X]$ to R by $\eta(f(X)) = f(0)$ (the constant term of $f(X)$). Then η is a surjective homomorphism. Since homomorphic image of a noetherian ring is a noetherian ring, R is noetherian.

Conversely, suppose that R is noetherian. Then, we have to show that $R[X]$ is noetherian. Let A be an ideal of $R[X]$. We show that A is finitely generated. Let $n \in \mathbb{N} \bigcup \{0\}$. Define A_n by

$$A_n = \{a \in R \mid \textit{there exists } f(X) = a_0 + a_1X + \cdots + a_{n-1}X^{n-1} + aX^n \in A\}.$$

Clearly, $0 \in A_n$, for $0 = 0 + 0X + \cdots 0X^n \in A$. Let $a, b \in A_n$. Then *there exist* $f(X) = a_0 + a_1X + \cdots + a_{n-1}X^{n-1} + aX^n$, *and* $g(X) = b_0 + b_1X + \cdots + b_{n-1}X^{n-1} + bX^n \in A$. Since A is an ideal, $f(X) - g(X) \in A$, and also $\alpha f(X) \in A$. Hence $a - b \in A_n$, and also $\alpha a \in A_n$. This shows that A_n is an ideal of R. Further, let $a \in A_n$, and $f(X) \in A$ be such that aX^n is the leading term of $f(X)$. Since $Xf(X) \in A$, $a \in A_{n+1}$. Thus, $A_n \subseteq A_{n+1}$ for all n, and we have an ascending chain

$$A_0 \subseteq A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq \cdots$$

of ideals of R. Since R is noetherian, *there exists* $m \in \mathbb{N}$ such that $A_n = A_m$ *for all* $n \geq m$. Again, since R is noetherian, A_n is finitely generated ideal of R for all n. Let $\{a_{i1}, a_{i2}, \ldots, a_{in_i}\}$ be a set of generators of the ideal A_i. Let $f_{ij}(X), 0 \leq i \leq m, 1 \leq j \leq n_i$ be a polynomial in A whose leading term is $a_{ij}X^i$. We show that $S = \{f_{ij}(X) \mid 0 \leq i \leq m, 1 \leq j \leq n_i\}$ is a set of generators of the ideal A of $R[X]$. Let $f(X) \in A$. We show that $f(X)$ is linear combination of members of S with coefficient in $R[X]$. The proof is by the induction on degree of $f(X)$ (clearly 0 is linear combination of members of S). If degree of $f(X)$ is 0, then

$f(X)$ is constant, and so it belongs to A_0. But, then it is a linear combination of $\{a_{01}, a_{02}, \ldots, a_{0n_0}\}$ with coefficients in R. Clearly, $a_{0j} = f_{oj}$, and so in this case $f(X)$ is a linear combination of members of S with coefficients in $R \subset R[X]$. Thus, the result is true if the degree of $f(X)$ is 0. Assume that the result is true for all those polynomials in A whose degree is less than r. Let $f(X) \in A$, and degree $f(X)$ is r. There are two cases:

(i) $r \geq m$.
(ii) $r \leq m - 1$.

Consider the case (i). Let $f(X) = a_0 + a_1X + \cdots + a_rX^r$ be a member of A. Then $a_r \in A_r = A_m$. Since A_m is generated by $\{a_{m1}, a_{m2}, \ldots, a_{nm}\}$, $a_r = \alpha_{m1}a_{m1} + \alpha_{m2}a_{m2} + \cdots + \alpha_{mn_m}a_{mn_m}$ for some $\alpha_{m1}, \alpha_{m2}, \ldots, \alpha_{mn_m}$ in R. Then,

$$f_1(X) = f(X) - X^{r-m}(\alpha_{m1}f_{m1}(X) + \alpha_{m2}f_{m2}(X) + \cdots + \alpha_{mn_m}f_{mn_m}(X))$$

is a member of A, and it is 0, or it is of degree less than r. By the induction hypothesis $f_1(X)$ is a linear combination of members of S with coefficients in $R[X]$. In turn, $f(X)$ is also a linear combination of members of S with coefficients in $R[X]$.

Consider the case (ii). In this case $r \leq m - 1$, and so $f_{rj}(X)$, $1 \leq j \leq n_r$ are in S. Now,

$$f_1(X) = f(X) - \alpha_{r1}f_{r1}(X) - \alpha_{r2}f_{r2}(X) - \cdots - \alpha_{rn_r}f_{rn_r}(X)$$

belongs to A, and it is 0, or it is of degree less than r. Again, by the induction hypothesis, $f_1(X)$ is a linear combination of members of S. Hence, $f(X)$ is also a linear combination of members of S with coefficients in $R[X]$. ♯

Using the induction on n, we get the following corollary.

Corollary 7.1.15 *If R is noetherian ring, then $R[X_1, X_1, \ldots, X_n]$ is also noetherian.* ♯

Remark 7.1.16 Although in a noetherian domain every nonzero nonunit element is product of irreducible elements, it need not be a U.F.D. For example, consider $Z[\sqrt{-5}]$. This is not a U.F.D. Now, $Z[X]$ is a noetherian ring (by the Hilbert basis theorem), and the map $f(X) \rightsquigarrow f(\sqrt{-5})$ is a surjective homomorphism of rings(verify). Since homomorphic image of a noetherian ring is noetherian $Z[\sqrt{-5}]$ is a noetherian ring. Also observe that a U.F.D. need not be a noetherian ring. For example $Z[X_1, X_2, \ldots, X_n, \cdots]$ is a U.F.D. but it is not noetherian.

Exercises

7.1.1 Show that a noetherian domain is a U.F.D. if and only if g.c.d. exists in R.

7.1.2 Show that every proper ideal of a noetherian ring can be embedded in a maximal ideal.

7.1.3 Give an example of an integral domain in which every nonzero nonunit element is expressible as product of irreducible elements but still it is not a noetherian ring.

7.1.4 Is a direct product of noetherian rings a noetherian ring? Support.

7.1.5 Show that a vector space is noetherian if and only if it is finite dimensional.

7.1.6 Show that $\mathbb{Z}[\sqrt{n}]$ is a noetherian ring for all integers n.

7.1.7 Show that an abelian group is a noetherian $\mathbb{Z}$-module if and only if it is finitely generated.

7.1.8 Give an example of a noetherian module which does not satisfy D.C.C. for submodules.

7.1.9 Let G be a finite commutative semigroup with identity. Show that the semigroup ring $R(G)$ is noetherian if and only if R is noetherian.

7.1.10 Suppose that R is noetherian, and G is group which is also noetherian in the sense that every subgroup of G is finitely generated. Is $R(G)$ also noetherian? Support.

7.1.11 Show that if R is noetherian, then the ring $R[[X]]$ of formal power series is also noetherian.
Hint. Imitate the proof of the Hilbert basis theorem by taking order function on the power series instead of degree of a polynomial.

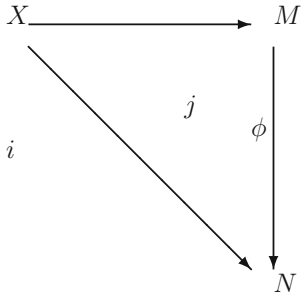
7.2 Free, Projective, and Injective Modules

In the last section, we described rings over which all modules possesses one of the most important and crucial property of vector spaces (modules over fields), viz., all submodules of finitely generated modules are finitely generated. Following are other two important properties of vector spaces: (i). Given a vector space W over a field F, and a surjective homomorphism f from a vector space V over F to W, there is a vector space homomorphism t from W to V such that $tof = I_W$. (ii). Given an injective homomorphism i from W to V, there is a homomorphism s from V to W such that $soi = I_W$. In this section, we discuss modules over arbitrary rings with these important crucial properties. Later in Chap. 9 on representation theory of finite groups, we shall describe rings (semi-simple rings) over which all modules have both of these crucial properties.

Let R be a ring(not necessarily commutative) with identity, and X be a set. We have the following universal problem:

"Does there exists a pair (M, i), where M is a left R-module, i a map from X to M, with the property that given any such pair (N, j), there is a unique R-homomorphism ϕ from M to N such that $\phi oi = j$?"

As in case of free groups, the solution to this problem is unique up to isomorphism. More precisely, if (M, i) and (N, j) are solutions to the above problem, then there exists an isomorphism ϕ from M to N such that the following diagram is commutative. (imitate the proof in the case of free groups).



We show the existence of solution to the above problem. If X is finite set $\{x_1, x_2, \ldots, x_n\}$ containing n elements. Then the pair (R^n, i), where $i(x_j) = \overline{e_j}$, is a solution to the problem. We do the construction of solution to the above problem for an arbitrary set X. Let $F(X)$ denote the set of all maps from X to R which vanish at all but finitely many points of X. More precisely, $F(X) = \{f : X \longrightarrow R \mid \textit{there is a finite subset } J \textit{ such that } f(x) = 0 \textit{ for all } x \in X - J\}$. Let $f, g \in F(X)$. Define a map $f + g$ from X to R by $(f + g)(x) = f(x) + g(x)$. Observe that $f + g \in F(X)$. This defines a binary operation $+$ on $F(X)$ such that $(F(X), +)$ is an abelian group. Define a multiplication $\cdot$ on $F(X)$ by elements of R by $(a \cdot f)(x) = a \cdot f(x)$, where $\cdot$ in the R.H.S. is the product in the ring. It is easy to see that $F(X)$ is a left R-module. Define a map i from X to $F(X)$ by $i(x)(y) = 1$ if $x = y$, and 0 otherwise. It is clear that i is an injective map. We show that $(F(X), i)$ is a solution to the above problem. Let $f \in F(X) - \{0\}$. Let $\{x_1, x_2, \ldots, x_n\}$ the finite subset of X such that $f(x_i) = a_i \neq 0$, and $f(x) = 0$ if $x \neq x_i$. Then it is clear that $f = a_1 i(x_1) + a_2 i(x_2) + \cdots + a_n i(x_n)$, and such a representation is unique. More precisely, $i(X)$ is a basis for $F(X)$. Let N be a left R-module, and j be a map from X to N. Define a map ϕ from $F(X)$ to N by

$$\phi(a_1 i(x_1) + a_2 i(x_2) + \cdots + a_n i(x_n)) = a_1 j(x_1) + a_2 j(x_2) + \cdots + a_n j(x_n).$$

Then ϕ is a homomorphism such that $\phi o i = j$. Since $i(X)$ is a basis, such a map is unique. This completes the proof of the existence of the solution to the above problem.

Definition 7.2.1 The solution to the above universal problem is called the **free** left R-module on X. Thus, $(F(X), i)$ is the free left R-module on X.

Proposition 7.2.2 *Every left R-module is quotient of a free left R-module.*

Proof Let M be a left R-module, and $(F(M), i)$ the free left R-module on M. The identity map I_M is a map from the set M to the left R-module M. From the universal

property of free left R-module, there is a unique homomorphism ϕ from $F(M)$ to M such that $\phi oi = I_M$. This shows that ϕ is surjective homomorphism. By the fundamental theorem of homomorphism M is isomorphic to $F(M)/ker\phi$. ♯

In Chap. 6 Sect. 6.1, we defined direct sum of finitely many R-modules. Now, we define direct sum of an arbitrary family of submodules. Let $\{M_\alpha \mid \alpha \in \Lambda\}$ be a family of R-modules. Then the Cartesian product

$$\prod_{\alpha\in\Lambda} M_\alpha = \{x : \Lambda \longrightarrow \bigcup_{\alpha\in\Lambda} M_\alpha \mid x(\alpha) \in M_\alpha \text{ for all } \alpha\}$$

is a left R-module with respect to the operations defined by $(x+y)(\alpha) = x(\alpha)+y(\alpha)$ and $(a \cdot x)(\alpha) = a \cdot x(\alpha)$. Define a map i_α from M_α to $\prod_{\alpha\in\Lambda} M_\alpha$ by $i_\alpha(x)(\beta) = x$ if $\beta = \alpha$, and 0 otherwise. It is clear that i_α is an injective homomorphism. Further, the α^{th} projection p_α from $\prod_{\alpha\in\Lambda} M_\alpha$ to M_α defined by $p_\alpha(x) = x(\alpha)$ is a surjective homomorphism such that $i_\alpha op_\alpha = I_{M_\alpha}$. The submodule of $\prod_{\alpha\in\Lambda} M_\alpha$ generated by $\bigcup_{\alpha\in\Lambda} i_\alpha(M_\alpha)$ is clearly

$$\{x \in \prod_{\alpha\in\Lambda} M_\alpha \mid x(\alpha) = 0 \text{ except for finitely many } \alpha\}.$$

This submodule is denoted by $\oplus\Sigma_{\alpha\in\Lambda} M_\alpha$, and it is called the **external direct** sum of the family $\{M_\alpha \mid \alpha \in \Lambda\}$. If M^α denotes the submodule of $\oplus\Sigma_{\alpha\in\Lambda} M_\alpha$ generated by $\bigcup_{\beta\neq\alpha} i_\beta(M_\beta)$, then $i_\alpha(M_\alpha) \bigcap M^\alpha = \{0\}$.

Proposition 7.2.3 *Let M be a module over a ring R. Let $\{M_\alpha \mid \alpha \in \Lambda\}$ be a family of its submodules. Then the following conditions are equivalent.*

1. *(i) M is generated by $\bigcup_{\alpha\in\Lambda} M_\alpha$.*
   (ii) $M_\alpha \bigcap M^\alpha = \{0\}$, where M^α is the submodule of M generated by $\bigcup_{\beta\neq\alpha} M_\beta$.
2. *For every nonzero element $x \in M$, there is a unique finite subset $\{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ of distinct elements of Λ together with unique nonzero elements $x_{\alpha_i} \in M_{\alpha_i}$ for each i, $1 \leq i \leq r$ such that*

$$x = x_{\alpha_1} + x_{\alpha_2} + \cdots + x_{\alpha_r}.$$

Proof $(1 \Rightarrow 2)$ Assume 1. Let x be a nonzero element of M. From 1(i), there exist a finite subset $\{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ of Λ together with nonzero elements $x_{\alpha_i} \in M_{\alpha_i}$ for each i, $1 \leq i \leq r$ such that

$$x = x_{\alpha_1} + x_{\alpha_2} + \cdots + x_{\alpha_r}.$$

Next, we prove the uniqueness. Suppose that

$$x = x_{\alpha_1} + x_{\alpha_2} + \cdots + x_{\alpha_r} = y_{\beta_1} + y_{\beta_2} + \cdots + y_{\beta_s},$$

where $\{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ and $\{\beta_1, \beta_2, \ldots, \beta_s\}$ are sets of distinct elements of Λ, $x_{\alpha_i} \in M_{\alpha_i} - \{0\}$ *for all* i, $1 \leq i \leq r$, and $y_{\beta_j} \in M_{\beta_j} - \{0\}$ *for all* j, $1 \leq j \leq s$. We need

to show the following: (i) $r = s$, (ii) after some rearrangement $\alpha_i = \beta_i$ for all i, and (iii) $x_{\alpha_i} = y_{\alpha_i}$ for all i. We prove it by the induction on $max(r, s)$. Suppose that $max(r, s) = 1$. Clearly, $r = 1 = s$. If $\alpha_1 \neq \beta_1$, then $x \in M_{\alpha_1} \bigcap M^{\alpha_1} = \{0\}$. This means that $x = 0$, a contradiction. Hence $\alpha_1 = \beta_1$, and then $x = x_{\alpha_1} = y_{\beta_1}$. Thus the result is true for $max(r, s) = 1$. Assume that result is true for $max(r, s) = n$. Let x be a nonzero element having representations

$$x = x_{\alpha_1} + x_{\alpha_2} + \cdots + x_{\alpha_{n+1}} = y_{\beta_1} + y_{\beta_2} + \cdots + y_{\beta_m},$$

where $\{\alpha_1, \alpha_2, \ldots, \alpha_{n+1}\}$ and $\{\beta_1, \beta_2, \ldots, \beta_m\}$ are sets of distinct elements of Λ, $m \leq n+1, x_{\alpha_i} \in M_{\alpha_i} - \{0\} for all\ i, 1 \leq i \leq n+1$, and $y_{\beta_j} \in M_{\beta_j} - \{0\} for all\ j, 1 \leq j \leq m$. We show that $\alpha_1 = \beta_j$ for some j. Suppose not. Then

$$x_{\alpha_1} = -x_{\alpha_2} - \cdots - x_{\alpha_{n+1}} + y_{\beta_1} + y_{\beta_2} + \cdots + y_{\beta_m}$$

belongs to $M_{\alpha_1} \bigcap M^{\alpha_1} = \{0\}$. Hence $x_{\alpha_1} = 0$. This is a contradiction to the supposition that $x - \alpha_1 \neq 0$. Thus, $\alpha_1 = \beta_j$ for some j. After rearranging, we may assume that $\alpha_1 = \beta_1$. Further,

$$x_{\alpha_1} - y_{\alpha_1} = -x_{\alpha_2} - \cdots - x_{\alpha_{n+1}} + y_{\beta_2} + \cdots + y_{\beta_m}.$$

Hence $x_{\alpha_1} - y_{\alpha_1}$ belongs to $M_{\alpha_1} \bigcap M^{\alpha_1} = \{0\}$. This shows that $x_{\alpha_1} = y_{\alpha_1}$. In turn,

$$x_{\alpha_2} + x_{\alpha_3} + \cdots + x_{\alpha_{n+1}} = y_{\beta_2} + y_{\beta_3} + \cdots + y_{\beta_m}.$$

By the induction hypothesis, $n + 1 = m, \alpha_i = \beta_i$, and $x_{\alpha_i} = y_{\beta_i}$ for all i.

$(2 \Rightarrow 1)$ Assume 2. Evidently, 1(i) follows. From the uniqueness of the representation of a nonzero element in M, it follows that $M_\alpha \bigcap M^\alpha$ cannot contain any nonzero element of M. Thus, $M_\alpha \bigcap M^\alpha = \{0\}$. ♯

Definition 7.2.4 We say that a module M over a ring R is an **internal direct sum** of the family $\{M_\alpha \mid \alpha \in \Lambda\}$ of its submodules if it satisfies any one, and hence both of the conditions in the above proposition is satisfied.

Proposition 7.2.5 *Let M be an internal direct sum of the family $\{M_\alpha \mid \alpha \in \Lambda\}$ of its submodules. Then M is isomorphic to the external direct sum $\oplus\Sigma_{\alpha\in\Lambda} M_\alpha$.*

Proof The map η from the external direct sum $\oplus\Sigma_{\alpha\in\Lambda} M_\alpha$ to M defined by $\eta(x) = \Sigma_{\alpha\in\Lambda} x(\alpha)$ is easily seen to be an isomorphism. ♯

From now onward, we shall not distinguish the internal and the external direct sums.

It follows from the construction of a free R- module $F(X)$ on a set X that $F(X)$ is isomorphic to the direct sum $\oplus\Sigma_{\alpha\in X} M_\alpha$, where $M_\alpha = R$ for all α. Thus, $F(X)$ is precisely the direct sum of X copies of R.

Consider a chain

$$\longrightarrow M_{n+1} \xrightarrow{\alpha_{n+1}} M_n \xrightarrow{\alpha_n} M_{n-1} \longrightarrow$$

where M_n is an R-module for all n, and α_n is a homomorphism for all n. This chain is called an **exact** sequence at M_n if $ker\alpha_n = image\alpha_{n+1}$. It is said to be **exact** sequence, if it is exact at all M_n.

An exact sequence

$$0 \longrightarrow M_1 \xrightarrow{\alpha} M_2 \xrightarrow{\beta} M_3 \longrightarrow 0$$

where 0 is the trivial module, is called a **short** exact sequence. Clearly, the above sequence is a short exact sequence if and only if (i) α is injective, (ii) β is surjective, and (iii) $ker\beta = image\alpha$.

If N is a submodule of a module M, then

$$0 \longrightarrow N \xrightarrow{i} M \xrightarrow{\nu} M/N \longrightarrow 0$$

is a short exact sequence, where i is the inclusion map, and ν is the quotient map.

The sequence $0 \longrightarrow M_1 \longrightarrow M_2$ is exact if and only if $M_1 \longrightarrow M_2$ is injective. The sequence $M_2 \longrightarrow M_3 \longrightarrow 0$ is exact if and only if $M_2 \longrightarrow M_3$ is surjective, and the sequence $0 \longrightarrow M_1 \longrightarrow M_2 \longrightarrow 0$ is exact if and only if $M_1 \longrightarrow M_2$ is an isomorphism.

Theorem 7.2.6 (Five Lemma) *Consider the following commutative diagram where rows are exact, and vertical maps are homomorphisms.*

$$\begin{array}{ccccccccc}
M_1 & \xrightarrow{\alpha_1} & M_2 & \xrightarrow{\alpha_2} & M_3 & \xrightarrow{\alpha_3} & M_4 & \xrightarrow{\alpha_4} & M_5 \\
\downarrow{\scriptstyle f_1} & & \downarrow{\scriptstyle f_2} & & \downarrow{\scriptstyle f_3} & & \downarrow{\scriptstyle f_4} & & \downarrow{\scriptstyle f_5} \\
N_1 & \xrightarrow{\beta_1} & N_2 & \xrightarrow{\beta_2} & N_3 & \xrightarrow{\beta_3} & N_4 & \xrightarrow{\beta_4} & N_5
\end{array}$$

(i) If f_1 is surjective, f_2 and f_4 are injective, then f_3 is injective.
(ii) If f_5 is injective, f_4 and f_2 are surjective, then f_3 is surjective.
(iii) If f_1, f_2, f_4, f_5 are isomorphisms, then f_3 is also an isomorphism.

Proof (i). Suppose that f_1 is surjective, f_2 and f_4 are injective. We have to show that f_3 is injective. Suppose that $f_3(m) = 0$. Then $f_4(\alpha_3(m)) = \beta_3(f_3(m))$ (commutativity of the diagram) $= \beta_3(0) = 0$. Since f_4 is injective, $\alpha_3(m) = 0$. Thus, $m \in ker\alpha_3 = image\alpha_2$ (exactness), and hence

there is an element $m_2 \in M_2$ such that $\alpha_2(m_2) = m$. Further, $0 = f_3(m) = f_3(\alpha_2(m_2)) = \beta_2(f_2(m_2))$ (commutativity of the diagram). Thus, $f_2(m_2) \in ker\beta_2 = image\beta_1$ (exactness). Hence, there exists $n_1 \in N_1$ such that $\beta_1(n_1) = f_2(m_2)$. Since f_1 is surjective, there is an element $m_1 \in M_1$ such that $f_1(m_1) = n_1$. Now, $f_2(\alpha_1(m_1)) = \beta_1(f_1(m_1)) = \beta_1(n_1) = f_2(m_2)$. Since f_2 is injective, $\alpha_1(m_1) = m_2$. But, already $\alpha_2(m_2) = m$. Hence $m = \alpha_2(\alpha_1(m_1)) = 0$, for $image\alpha_1 = ker\alpha_2$. This shows that f_3 is injective.

(ii). Suppose that f_5 is injective, f_2 and f_4 are surjective. We have to show that f_3 is surjective. Let $n \in N_3$. We have to show the existence of an element $m \in M_3$ such that $f_3(m) = n$. Now, $\beta_3(n) \in N_4$. Since f_4 is surjective, there is an element $m_4 \in M_4$ such that $f_4(m_4) = \beta_3(n)$. Now $f_5(\alpha_4(m_4)) = \beta_4(f_4(m_4))$ (commutativity of the diagram)$= \beta_4(\beta_3(n)) = 0$ (exactness). Since f_5 is injective, $\alpha_4(m_4) = 0$. Thus, $m_4 \in ker\alpha_4 = image\alpha_3$. Hence, there is an element $m_3 \in M_3$ such that $\alpha_3(m_3) = m_4$. Since $\beta_3(f_3(m_3)) = f_4(\alpha_3(m_3)) = f_4(m_4) = \beta_3(n)$, $\beta_3(n - f_3(m_3)) = 0$. Thus, $n - f_3(m_3) \in ker\beta_3 = image\beta_2$. Hence there exists $n_2 \in N_2$ such that $\beta_2(n_2) = n - f_3(m_3)$. Since f_2 is surjective, there is an element $m_2 \in M_2$ such that $f_2(m_2) = n_2$. Now $n - f_3(m_3) = \beta_2(n_2) = \beta_2(f_2(m_2)) = f_3(\alpha_2(m_2))$. This shows that $f_3(m_3 + \alpha_2(m_2)) = n$, and so f_3 is surjective.

(iii). Follows from (i) and (ii). ♯

Remark 7.2.7 The technique used in the proof of the above theorem is known as diagram chasing.

Corollary 7.2.8 *Consider the following commutative diagram*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & M_1 & \longrightarrow & M_2 & \longrightarrow & M_3 & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & N_1 & \longrightarrow & N_2 & \longrightarrow & N_3 & \longrightarrow & 0 \end{array}$$

where rows are exact, vertical arrows are homomorphisms, and the extreme vertical arrows are isomorphisms. Then the middle vertical arrow is also an isomorphism. ♯

A short exact sequence

$$0 \longrightarrow M_1 \xrightarrow{\alpha} M_2 \xrightarrow{\beta} M_3 \longrightarrow 0$$

is said to be a **split** exact sequence, if there exists a homomorphism t from M_3 to M_2 such that $\beta ot = I_{M_3}$. The homomorphism t is called a splitting of the exact sequence.

Proposition 7.2.9 *A short exact sequence*

$$0 \longrightarrow M_1 \xrightarrow{\alpha} M_2 \xrightarrow{\beta} M_3 \longrightarrow 0$$

is split exact if and only if there exists a homomorphism s from M_2 to M_1 such that $so\alpha = I_{M_1}$. Further, there is a bijective correspondence between the set of splittings of the short exact sequence and the set of all homomorphisms s from M_2 to M_1 satisfying $so\alpha = I_{M_1}$.

Proof Let t be a splitting. Then $\beta ot = I_{M_3}$. Let $x \in M_2$. Then $\beta(x - t(\beta(x))) = \beta(x) - \beta(t(\beta(x))) = \beta(x) - \beta(x) = 0$. Hence $x - t(\beta(x)) \in ker\beta = image\alpha$. Since α is injective, there is a unique $s(x) \in M_1$ such that $\alpha(s(x)) = x - t(\beta(x))$. Using the defining property of s and the injectivity of α, it can be seen that s is a homomorphism from M_2 to M_1. Also $\alpha(s(\alpha(y))) = \alpha(y) - t(\beta(\alpha(y))) = \alpha(y)$ (for $\beta o\alpha = 0$). Since α is injective $s(\alpha(y)) = y$ for all $y \in M_1$. Hence $so\alpha = I_{M_1}$. We show that the correspondence which associates an splitting t with s defined above is bijective. Suppose that t_1 and t_2 are splittings which associates to same s. Then $\alpha(s(x)) = x - t_1(\beta(x)) = x - t_2(\beta(x))$. Since β is surjective, $t_1 = t_2$. Let s be a homomorphism from M_2 to M_1 such that $so\alpha = I_{M_1}$. Let $y \in M_3$. Since β is surjective, there is an element $x \in M_2$ such that $\beta(x) = y$. Define a binary relation t from M_3 to M_2 by $t(\beta(x)) = x - \alpha(s(x))$. Suppose that $\beta(x_1) = \beta(x_2)$. Then $x_1 - x_2 \in ker\beta = image\alpha$. Hence there exists $z \in M_1$ such that $\alpha(z) = x_1 - x_2$. Now, $s(x_1 - x_2) = s(\alpha(z)) = z$. Hence $\alpha(s(x_1 - x_2)) = \alpha(z) = x_1 - x_2$. Thus, $x_1 - \alpha(s(x_1)) = x_2 - \alpha(s(x_2))$. This shows that t is a map from M_3 to M_1. It can easily be seen that t is a homomorphism. Also $\beta(t(\beta(x))) = \beta(x - \alpha(s(x))) = \beta(x) - \beta(\alpha(s(x))) = \beta(x)$, for $\beta o\alpha = 0$. Thus, t is splitting, and since $x - t(\beta(x)) = \alpha(s(x))$, the homomorphism from M_2 to M_1 associated to the splitting t is s. ♯

A homomorphism s such that $so\alpha = I_{M_1}$ is also called a splitting.
Let M_1 and M_3 be R-modules. Then

$$0 \longrightarrow M_1 \xrightarrow{i_1} M_1 \oplus M_3 \xrightarrow{p_2} M_3 \longrightarrow 0$$

is a split exact sequence, where $i_1(x) = (x, 0)$, *and* $p_2(x, y) = y$. The map i_2 from M_3 to $M_1 \oplus M_3$ given by $i_2(y) = (0, y)$ is a splitting, and the associated splitting from $M_1 \oplus M_3$ to M_1 is the first projection p_1.

Proposition 7.2.10 *Let*

$$0 \longrightarrow M_1 \xrightarrow{\alpha} M_2 \xrightarrow{\beta} M_3 \longrightarrow 0$$

be a split exact sequence. Then there exists an isomorphism from M_2 to $M_1 \oplus M_3$ such that the diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & M_1 & \xrightarrow{\alpha} & M_2 & \xrightarrow{\beta} & M_3 & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle I_{M_1}} & & \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle I_{M_3}} & & \\ 0 & \longrightarrow & M_1 & \xrightarrow{i_1} & M_1 \oplus M_3 & \xrightarrow{p_2} & M_3 & \longrightarrow & 0 \end{array}$$

is commutative. Further, if t and s are associated splittings, then

$$0 \longrightarrow M_3 \xrightarrow{t} M_2 \xrightarrow{s} M_1 \longrightarrow 0$$

is also a spilt exact sequence.

Proof Let t be a homomorphism from M_3 to M_2 which is a splitting, and s be the associated splitting. Define a map f from M_2 to $M_1 \oplus M_3$ by $f(x) = (s(x), \beta(x))$. Then f is a homomorphism which makes the diagram commutative(verify). By the five lemma f is an isomorphism. Finally, since f is an isomorphism $t = f^{-1} o i_2$, and $s = p_1 o f$. The result follows if we observe that

$$0 \longrightarrow M_3 \xrightarrow{i_2} M_1 \oplus M_3 \xrightarrow{p_1} M_1 \longrightarrow 0$$

is split exact. ♯

Let M and N be left R-modules. Let $Hom_R(M, N)$ denote the set of all R-homomorphisms. Let $f, g \in Hom_R(M, N)$. Define a map $f + g$ from M to N by $(f + g)(x) = f(x) + g(x)$. It is easy to observe that $f + g$ is also a member of $Hom_R(M, N)$. This defines an addition in $Hom_R(M, N)$ with respect to which it is an abelian group. We may be tempted to define a module structure on $Hom_R(M, N)$ by defining $(a \cdot f)(x) = a \cdot f(x)$. But $a \cdot f$ need not be a member of $Hom_R(M, N)$, and so it will not work in general. However, if R is a commutative ring, then it is indeed a member of $Hom_R(M, N)$, and then $Hom_R(M, N)$ becomes a R-module. Note that every R-module is also a $\mathbb{Z}$-module, and $Hom_R(M, N)$ is a subgroup of $Hom_{\mathbb{Z}}(M, N)$. Let $f \in Hom_{\mathbb{Z}}(M, N)$, and $a \in R$. Define a map $f \cdot a$ from M to N by $(f \cdot a)(x) = f(a \cdot x)$. Clearly, $f \cdot a \in Hom_{\mathbb{Z}}(M, N)$. It is easy to observe that $Hom_{\mathbb{Z}}(M, N)$ is a right R-module with respect to the above right multiplication. Also, if R is a commutative ring, then $Hom_R(M, N)$ is a right R-submodule of $Hom_{\mathbb{Z}}(M, N)$.

Let M_1 and M_2 be left R-modules, and α a R-homomorphism from M_1 to M_2. Let N be a left R-module. Then we have a map $\alpha^\star$ from $Hom_R(M_2, N)$ to $Hom_R(M_1, N)$

defined by $\alpha^\star(f) = fo\alpha$. It can be easily seen that $\alpha^\star$ is a group homomorphism. Similarly, we have a group homomorphism $\alpha_\star$ from $Hom_R(N, M_1)$ to $Hom_R(N.M_2)$ given by $\alpha_\star(f) = \alpha of$. Let β be a homomorphism from M_2 to M_3. We leave it to the reader to verify that (i) $(\beta o\alpha)^\star = \alpha^\star o\beta^\star$, and (ii) $(\beta o\alpha)_\star = \beta_\star o\alpha_\star$. It is also clear that $0^\star$ and $0_\star$ are the corresponding zero homomorphisms. Further, it is straight forward to see that $(I_M)^\star$ and $(I_M)_\star$ are the corresponding identity maps.

Theorem 7.2.11 *Hom is a left exact functor in the following sense.*
(i) If

$$M_1 \xrightarrow{\alpha} M_2 \xrightarrow{\beta} M_3 \longrightarrow 0$$

is an exact sequence of left R-modules, and N a left R-module, then the sequence

$$0 \longrightarrow Hom_R(M_3, N) \xrightarrow{\beta^\star} Hom_R(M_2, N) \xrightarrow{\alpha^\star} Hom_R(M_1, N)$$

is exact.
(ii) If

$$0 \longrightarrow M_1 \xrightarrow{\alpha} M_2 \xrightarrow{\beta} M_3$$

is an exact sequence, then

$$0 \longrightarrow Hom_R(N, M_1) \xrightarrow{\alpha_\star} Hom_R(N, M_2) \xrightarrow{\beta_\star} Hom_R(N, M_3)$$

is also exact.

Proof (i). Let $f \in Hom_R(M_3, N)$ such that $\beta^\star(f) = 0$. Then $fo\beta = 0$. Since β is surjective (exactness), it follows that $f = 0$. This shows that $\beta^\star$ is injective. Next, since $\beta o\alpha = 0$ (exactness of the given sequence), $\alpha^\star o\beta^\star = (\beta o\alpha)^\star = 0^\star = 0$. Hence image $\beta^\star \subseteq ker\alpha^\star$. Let $f \in ker\alpha^\star$. Then $\alpha^\star(f) = fo\alpha = 0$, and so $kerf \supseteq image\alpha = ker\beta$. By the fundamental theorem of homomorphism, there is a unique homomorphism $\overline{f}$ from $M_2/ker\beta$ to N such that $\overline{f}ov = f$. Also, since β is surjective, we have an isomorphism $\overline{\beta}$ from $M_2/ker\beta$ to M_3 such that $\overline{\beta}ov = \beta$. Then $\beta^\star(\overline{f}o\overline{\beta}^{-1}) = \overline{f}o\overline{\beta}^{-1}o\beta = \overline{f}ov = f$. Thus $f \in image\beta^\star$. This shows that $ker\alpha^\star = image\beta^\star$.

The proof of (ii) is similar, and it is left as an exercise. ♯

Remark 7.2.12 Hom is not a right exact functor, for even if α is injective $\alpha^\star$ need not be surjective, and even if β is surjective, $\beta_\star$ need not be surjective. Consider, for example, the multiplication f_m from $\mathbb{Z}$ to $\mathbb{Z}$ by m, where $m > 1$. Then $f_m^\star$ from $Hom_\mathbb{Z}(\mathbb{Z}, \mathbb{Z})$ to itself is not surjective (verify). The quotient map ν frim $\mathbb{Z}$ to

$\mathbb{Z}_m$ is a surjective homomorphism but $\nu_\star$ from $Hom_{\mathbb{Z}}(\mathbb{Z}_m, \mathbb{Z})$ to $Hom_{\mathbb{Z}}(\mathbb{Z}_m, \mathbb{Z}_m)$ is not surjective for the simple reason that $Hom_{\mathbb{Z}}(\mathbb{Z}_m, \mathbb{Z}) = \{0\}$, whereas $Hom_{\mathbb{Z}}(\mathbb{Z}_m, \mathbb{Z}_m) \approx \mathbb{Z}_m \neq \{0\}$.

Proposition 7.2.13 *If*

$$0 \longrightarrow M_1 \xrightarrow{\alpha} M_2 \xrightarrow{\beta} M_3 \longrightarrow 0$$

is a split exact sequence, and N a module, then

$$0 \longrightarrow Hom_R(M_3, N) \xrightarrow{\beta^\star} Hom_R(M_2, N) \xrightarrow{\alpha^\star} Hom_R(M_1, N) \longrightarrow 0$$

and

$$0 \longrightarrow Hom_R(N, M_1) \xrightarrow{\alpha_\star} Hom_R(N, M_2) \xrightarrow{\beta_\star} Hom_R(N, M_3) \longrightarrow 0$$

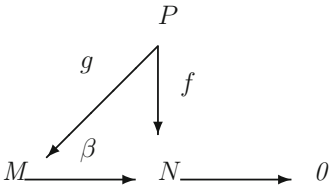
are also split exact sequence.

Proof Let t and s be associated splittings. Then $(s^\star o \alpha^\star) = (\alpha o s)^\star = I_{Hom_R(M_1, N)}$, and similarly, $t_\star o \beta_\star$ is also the identity map. This shows that $\alpha^\star$ and $\beta_\star$ are surjective maps, and the above sequence splits. ♯

Definition 7.2.14 A left R-module P is called a **projective** left R-module if given an exact sequence

$$M \xrightarrow{\beta} N \longrightarrow 0$$

(or equivalently, β is a surjective homomorphism from M to N), and a homomorphism f from P to N, there is a homomorphism g from P to M such that the diagram



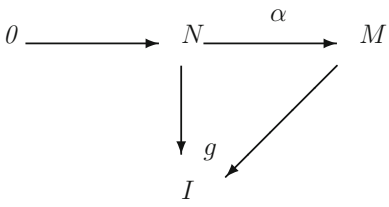
is commutative.

Dually, we have

Definition 7.2.15 A left R-module I is called an **injective** left R-module if given any exact sequence

$$0 \longrightarrow N \xrightarrow{\alpha} M$$

(or equivalently, α an injective homomorphism), and a homomorphism f from N to I, there is a homomorphism g from M to I such that the diagram



is commutative.

Proposition 7.2.16 *If P is a projective R-module, then every short exact sequence*

$$0 \longrightarrow M \xrightarrow{\alpha} N \xrightarrow{\beta} P \longrightarrow 0$$

splits. Dually, if M is injective, then also it splits.

Proof Suppose that P is projective. Since

$$N \xrightarrow{\beta} P \longrightarrow 0$$

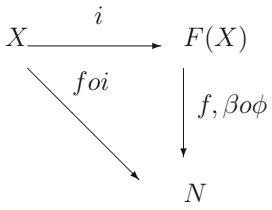
is exact, and I_P is a homomorphism from P to P, there is a homomorphism t from P to N such that $\beta o t = I_P$. Thus, t is a splitting. The rest also follows similarly. ♯

The following result follows from Propositions 7.2.10 and 7.2.13.

Corollary 7.2.17 *If the last but one term in a short exact sequence is a projective module, or the second term in a short exact sequence is injective module, then Hom takes the short exact sequence to a split exact sequence.* ♯

Proposition 7.2.18 *Every free R-module is projective.*

Proof Let $(F(X), i)$ be a free R-module on X, and β a surjective homomorphism from M to N. Then $\beta^{-1}\{(foi)(x)\} \neq \emptyset$ *for all* $x \in X$. From the axiom of choice, there is a map c from X to M such that $c(x) \in \beta^{-1}\{foi(x)\}$ *for all* $x \in X$. This means $\beta oc = foi$. Since $(F(X), i)$ is a free R-module on X, there is a unique homomorphism ϕ from $F(X)$ to M such that $\phi oi = c$. Hence $\beta o\phi$ and f both make the triangle



commutative. Since $(F(X), i)$ is a free R-module, $\beta o\phi = f$. This shows that $F(X)$ is a projective module. ♯

A submodule N of a module M over a ring R is called a **direct summand** of M if there is a submodule L of M such that $M = N \oplus L$.

Proposition 7.2.19 *Direct summand of a projective (injective) module is projective(injective).*

Proof Suppose that $P \oplus Q$ is projective. We show that P is projective. Let β be a surjective homomorphism from M to N, and f a homomorphism from P to f. Then fop_1 is a homomorphism from $P \oplus Q$ to N. Since $P \oplus Q$ is projective, there is a homomorphism ϕ from $P \oplus Q$ to M such that $\beta o\phi = fop_1$. We have the homomorphism ϕoi_1 from P to M such that $\beta o\phi oi_1 = fop_1 oi_1 = foI_P = f$. This shows that P is projective. The rest can be proved similarly. ♯

Theorem 7.2.20 *A left R-module P is projective if and only if it is direct summand of a free R-module.*

Proof Since a free R-module is projective, and direct summand of a projective module is projective, direct summand of a free module is projective. Next suppose that P is projective. From Proposition 7.2.2 we have a surjective homomorphism β from from $F(P)$ to P. This gives us an exact sequence

$$0 \longrightarrow Ker\beta \longrightarrow F(P) \longrightarrow P \longrightarrow o.$$

Since P is projective, the sequence splits. Hence from Proposition 7.2.10. P is direct summand of $F(P)$. ♯

We have the following corollary.

Corollary 7.2.21 *A left R-module P is projective if and only if every short exact sequence*

$$0 \longrightarrow M \longrightarrow N \longrightarrow P \longrightarrow 0$$

splits.

Proof Suppose that every short exact sequence

$$0 \longrightarrow M \longrightarrow N \longrightarrow P \longrightarrow 0$$

splits. Then, in particular,

$$0 \longrightarrow Ker\beta \longrightarrow F(P) \longrightarrow P \longrightarrow o$$

splits. From the Proposition 7.2.10, P is a direct summand of $F(P)$. From Theorem 7.2.20, it follows that P is projective. The converse follows from the Proposition 7.2.16. ♯

Corollary 7.2.22 *Let $\{P_\alpha \mid \alpha \in \Lambda\}$ be a family of left R-modules. Then $P = \oplus\Sigma_{\alpha\in\Lambda} P_\alpha$ is projective if and only if each P_α is projective.*

Proof If P is projective, then, since each P_α is direct summand of P, each P_α is projective. Conversely, suppose that each P_α is projective, then P_α is direct summand of a free module F_α. But, then P is a direct summand of $\oplus\Sigma_{\alpha\in\Lambda} F_\alpha$. Since direct sum of free modules are free, the result follows. ♯

Example 7.2.23 Every vector space (being free) is projective. It is also injective.

Example 7.2.24 Direct sum of infinite cyclic groups are $\mathbb{Z}$-projective, for they are free.

Example 7.2.25 Since submodules of free modules over a P.I.D. are free, projective, and free modules over P.I.D. are same. In particular, all projective modules over $F[X]$, where F is field, is free. It is a fact that all finitely generated projective module over $F[X_1, X_2, \ldots, X_n]$ is free. This fact was conjectured by J.P. Serre, and it was proved by D. Quillen and Suslin simultaneously, and independently in 1976.

Example 7.2.26 $\mathbb{Z}_m$ is not $\mathbb{Z}$-projective for it is not free. $\mathbb{Z}$ is $\mathbb{Z}$-projective but it is not injective.

Example 7.2.27 Submodule of a free module need not be free. For example, $\mathbb{Z}_6$ is free over $\mathbb{Z}_6$, but $\mathbb{Z}_3$ being an ideal of $\mathbb{Z}_6$ is a submodule, and it is not free. Also $\mathbb{Z}_6 = \mathbb{Z}_3 \oplus \mathbb{Z}_2$, and so $\mathbb{Z}_3$ is projective but not free.

Example 7.2.28 Submodule of a projective module need not be a projective module. $\mathbb{Z}_4$ is free module over $\mathbb{Z}_4$, and so it is projective. However w $\mathbb{Z}_2$ is a submodule of $\mathbb{Z}_4$ which is not direct summand of a free module, and so it cannot be a projective module. Quotient of a projective module need not be projective, for otherwise every module will become projective.

Theorem 7.2.29 *Let I be a left R-module. Then I is injective if and only if given any left ideal A of R and a R-homomorphism f from A (considered as R-module) to I, there exists a R-homomorphism $\overline{f}$ from R to I such that $\overline{f}/A = f$.*

Proof If I is injective, then by the definition, every homomorphism from A to I can be extended to a homomorphism from R to I.

Conversely, suppose that every R-homomorphism from every ideal A to I can be extended to a homomorphism from R to I. Let ξ be an injective homomorphism

from a left R-module to M to a left R-module N, and ϕ a R-homomorphism from M to I. We have to show that ϕ can be extended to a homomorphism from N to I. Let X be the set of all pairs (L, ψ) such that L is a submodule of N containing $\xi(M)$, and ψ a homomorphism from L to I such that $\psi o\xi = \phi$. Further, $X \neq \emptyset$, for $(\xi(M), \psi)$ belongs to X, where $\psi(\xi(x)) = \phi(x)$ *for all* $x \in M$. Define a relation $\leq$ on X by $(L, \psi) \leq (L', \chi)$, if $L \subseteq L'$ and $\chi/L = \psi$. Clearly, $(X, \leq)$ is a nonempty partially ordered set. Let $\{(L_\alpha, \psi_\alpha) \mid \alpha \in \Lambda\}$ be a chain in X. Since union of a chain of submodules is a submodule, $L = \bigcup_{\alpha\in\Lambda} L_\alpha$ is a submodule of N containing $\xi(M)$. We have a unique homomorphism ψ from L to N defined by the property that $\psi/L_\alpha = \psi_\alpha$ for all α. Clearly, (L, ψ) is an upper bound of the chain. By the Zorn's Lemma X has a maximal element (L_0, ψ_0) (say). We show that $L_0 = N$. Suppose that $L_0 \neq N$, and $x_0 \in N - L_0$. Let $A = \{\lambda \in R \mid \lambda x_0 \in L_0\}$. Then A, being the inverse image of L_0 under the homomorphism $\lambda \rightsquigarrow \lambda x_0$ from R to N, is an ideal of R, and the map f from A to I defined by $f(\lambda) = \psi_0(\lambda x_0)$ is a homomorphism. From our supposition, we have a R-homomorphism $\overline{f}$ from R to I such that $\overline{f}/A = f$. Let $L_1 = L_0 + < x_0 >$ be the submodule of N generated by $L_0 \bigcup \{x_0\}$. Then L_0 is a proper submodule of L_1. Any element of L_1 is of the form $u + \lambda x_0$, where $u \in L_0$ and $\lambda \in R$. Suppose that $u_1 + \lambda_1 x_0 = u_2 + \lambda_2 x_0$, where $u_1, u_2 \in L_0$ and $\lambda_1, \lambda_2 \in R$. Then $(\lambda_2 - \lambda_1)x_0 = u_1 - u_2 \in L_0$. Hence $\lambda_2 - \lambda_1 \in A$, and $\overline{f}(\lambda_2 - \lambda_1) = f(\lambda_2 - \lambda_1) = \psi_0((\lambda_2 - \lambda_1)x_0) = \psi_0(u_1 - u_2) = \psi_0(u_1) - \psi_0(u_2)$. Hence $\overline{f}(\lambda_2) - \overline{f}(\lambda_1) = \psi_0(u_1) - \psi_0(u_2)$, and so $\psi_0(u_1) + \overline{f}(\lambda_1) = \psi_0(u_2) + \overline{f}(\lambda_2)$. Thus, we have a map ψ_1 from L_1 to I given by $\psi_1(u + \lambda x_0) = \psi_0(u) + \overline{f}(\lambda)$. clearly, ψ_1 is a homomorphism, and $\psi_1/L_0 = \psi_0$. Hence $(L_1, \psi_1) \in X$, and $(L_1, \psi_1) > (L_0, \psi_0)$. This is a contradiction to the maximality of (L_0, ψ_0). Thus, $L_0 = N$, and ψ_0 is a homomorphism from N to I such that $\psi_0 o\xi = \phi$. This proves that I is injective. ♯

Now, we describe $\mathbb{Z}$-injective modules. Recall the following:

Definition 7.2.30 An abelian group A is called a **divisible** if for all $a \in A$ and $n \in \mathbb{Z} - \{0\}$, there is a $b \in A$ such that $nb = a$.

Corollary 7.2.31 *An abelian group A is $\mathbb{Z}$-injective if and only if it is divisible.*

Proof An ideal of $\mathbb{Z}$ is of the form mZ for some nonnegative integer m. From the above theorem, an abelian group A is $\mathbb{Z}$-injective if and only if for each m, every homomorphism from $m\mathbb{Z}$ to A can be extended to a homomorphism from $\mathbb{Z}$ to A.

Suppose that A is $\mathbb{Z}$-injective. Let $a \in A$ and $n \in \mathbb{Z} - \{0\}$. The map f from $n\mathbb{Z}$ to A defined by $f(nm) = ma$ is a homomorphism. Since A is injective, there is a homomorphism $\overline{f}$ from $\mathbb{Z}$ to A such that $\overline{f}/nZ = f$. Now $\overline{f}(n) = f(n) = 1 \cdot a = a$. Suppose that $\overline{f}(1) = b$. Then $a = \overline{f}(n) = n \cdot \overline{f}(1) = n \cdot b$. This shows that A is divisible.

Conversely, suppose that A is divisible. Let f from $n\mathbb{Z}$ to A be a homomorphism. Suppose that $f(n) = a$. Since A is divisible, there is $b \in A$ such that $nb = a$. We have a homomorphism $\overline{f}$ from $\mathbb{Z}$ to A defined by $\overline{f}(m) = mb$. Also if $m = nr$, then $\overline{f}(m) = nrb = rnb = ra = f(nr) = f(m)$. This means that $\overline{f}/n\mathbb{Z} = f$. It follows from the above theorem that A is injective. ♯

Example 7.2.32 The groups $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$, $(\mathbb{C}, +)$, and $(S^1, \cdot)$ are all divisible (verify), and so they are all $\mathbb{Z}$-injective.

Example 7.2.33 Homomorphic images, and also the quotients of divisible groups are divisible groups (verify). In particular, $\mathbb{Q}/\mathbb{Z} \approx P$ is $\mathbb{Z}$-injective. Submodule of an injective module need not be injective, for $(\mathbb{Z}, +)$ is not injective whereas $(\mathbb{Q}, +)$ is injective.

Example 7.2.34 No nontrivial finite group can be divisible, for if $| A | = n$ and $a \neq 0$, then we cannot find a $b \in A$ such that $nb = a$.

We have seen that every module is quotient of a projective module. Dually, we show that every module is submodule of an injective module.

Proposition 7.2.35 *A left R-module I is injective if and only if for every left ideal A of R and every R homomorphism f from A to I, there exists a $x \in I$ such that $f(a) = ax$ for all $a \in A$.*

Proof Suppose that I is injective, A a left ideal of R, and f a R-homomorphism from A to I. Then there exists a R-homomorphism ϕ from R to I such that $\phi/A = f$. Suppose that $\phi(1) = x$. Then $\phi(a) = \phi(a \cdot 1) = a \cdot \phi(1) = a \cdot x$ for all $a \in R$. In particular, $f(a) = a \cdot x$ *for all* $a \in A$. Conversely, suppose that such a $x \in I$ exists. Then the map ϕ from R to I defined by $\phi(a) = a \cdot x$ is a homomorphism which is an extension of f. The result follows from the Theorem 7.2.29. ♯

Let R be a ring with identity, and A be an abelian group. Then the set $Hom_{\mathbb{Z}}(R, A)$ of all additive group homomorphisms from $(R, +)$ to A is an abelian group with respect to the pointwise addition. $Hom_{\mathbb{Z}}(R, A)$ becomes a left R-module with respect to the external multiplication $\cdot$ defined by $(a \cdot f)(b) = f(ba)$. If A is also a left R-module, then $Hom_R(R, A)$ is a subgroup of $Hom_{\mathbb{Z}}(R, A)$. If $f \in Hom_R(R, A)$, then $a \cdot f \in Hom_R(R, A)$, for $(a \cdot f)(bc) = f(bca) = bf(ca)$ (for f is a $R-$ homomorphism)$= b(a \cdot f)(c)$. This shows that $Hom_R(R, A)$ is a left submodule of $Hom_{\mathbb{Z}}(R, A)$ for all left R-module A.

Proposition 7.2.36 *Let M be a left R-module. Then the map ϕ from $Hom_R(R, M)$ to M defined by $\phi(f) = f(1)$ is an isomorphism of R-modules.*

Proof Clearly, $\phi(f + g) = (f + g)(1) = f(1) + g(1) = \phi(f) + \phi(g)$, and $\phi(a \cdot f) = (a \cdot f)(1) = f(1 \cdot a) = f(a) = a \cdot f(1)$ (for f is a R-homomorphism)$= a \cdot \phi(f)$. This shows that ϕ is a homomorphism. Next, suppose that $\phi(f) = \phi(g)$. Then $f(1) = g(1)$. Since f and g are R-homomorphisms, $f(a) = a \cdot f(1) = a \cdot g(1) = g(a)$ for all $a \in R$. This shows that $f = g$, and so ϕ is injective. Lastly, let $x \in M$. Define a map f from R to M by $f(a) = a \cdot x$. Then $f \in Hom_R(R, M)$ and $\phi(f) = x$. This shows that ϕ is also surjective. ♯

Proposition 7.2.37 *Let A be a divisible group and R a ring. Then the left R-module $Hom_{\mathbb{Z}}(R, A)$ is a left injective module over R.*

Proof Let B be a left ideal of R, and f be a R-homomorphism from B to $Hom_{\mathbb{Z}}(R, A)$. Then the map χ from B to A defined by $\chi(b) = f(b)(1)$ is clearly a group homomorphism from $(B, +)$ to A. Since A, being divisible, is $\mathbb{Z}$-injective, we can extend χ to a group homomorphism $\overline{\chi}$ from the group $(R, +)$ to A. Now, for $b \in B$, $f(b)(1) = \chi(b) = \overline{\chi}(b) = b\overline{\chi}(1)$. This show that $f(b) = b \cdot \overline{\chi}$ for all $b \in B$. From the Proposition 7.2.35, the result follows. ♯

The proof of the following proposition is an easy verification.

Proposition 7.2.38 *Direct sum of divisible groups are divisible.* ♯

Proposition 7.2.39 *Every abelian group can be embedded in to a divisible group.*

Proof Let A be an abelian group. Then A is quotient of the free abelian group $F(A)$ on A. Suppose that $A \approx F(A)/L$. Now, $F(A)$ is direct sum of A copies of $\mathbb{Z}$, and so it is a subgroup of direct sum of A copies of $(Q, +)$. Thus, A is isomorphic to a subgroup of quotient group of the direct sum of A copies of Q. Since direct sum of divisible groups are divisible, and also the quotient group of divisible groups are divisible, the result follows. ♯

Theorem 7.2.40 *Every left R-module can be embedded in a left injective R-module.*

Proof Let M be a left R-module. From the above proposition, $(M, +)$ is subgroup of a divisible group D. Since Hom is a left exact functor, $Hom_{\mathbb{Z}}(R, M)$ is isomorphic to a submodule of $Hom_{\mathbb{Z}}(R, D)$. Since D is divisible, $Hom_{\mathbb{Z}}(R, D)$ is injective over R. Also $M \approx Hom_R(R, M)$ is a submodule of $Hom_{\mathbb{Z}}(R, M)$. The result follows. ♯

Corollary 7.2.41 *A left R-module I is injective if and only if every short exact sequence of the type*

$$0 \longrightarrow I \longrightarrow M \longrightarrow N \longrightarrow 0$$

splits.

Proof If I is injective, then it is already seen that the sequence will split. Conversely, suppose that every such exact sequence splits. Since every module can be embedded in an injective module, there is an injective module M such that I is a submodule of M. This gives us an exact sequence

$$0 \longrightarrow I \longrightarrow M \longrightarrow M/N \longrightarrow o.$$

By our hypothesis, the above exact sequence splits. Hence I is direct summand of an injective module M. Since direct summand of an injective module is an injective module, I is an injective module. ♯

Exercises

7.2.1 State and prove the Five lemma for groups.

7.2.2 Develop the concept and the theory of projective and injective groups. Try to characterize them.

7.2.3 A commutative integral domain R is said to be a **Dedekind domain** if given any pair of ideals A and B of R such that $B \subset A$, there is an ideal C such that $B = AC$. For example, every PID is Dedekind domain. $\mathbb{Z}[\sqrt{-5}]$ is a Dedekind domain (prove it). Indeed, if we have a subfield F of $\mathbb{C}$ which is a finite-dimensional vector space over its subfield $\mathbb{Q}$ (such a field is called a **number field**) , and R the set of elements of F which are roots of monic polynomials with rational coefficients (called the ring of algebraic integers F), then R is a Dedekind domain. Let R be a Dedekind domain. Let A be an ideal of R. Show that A considered as a module over R is a finitely generated projective module.
Hint. Let $a \in A, a \neq 0$. Then $Ra \subset A$. Let B be an ideal such that $Ra = BA$. Suppose that $a = b_1a_1 + b_2a_2 + \cdots + b_na_n$. Check that $(u_1, u_2, \ldots, u_n) \mapsto u_1a_1 + u_2a_2 + \cdots + u_na_n$ is a module homomorphism with the inverse map given by $x \mapsto (v_1, v_2, \ldots, v_n)$, where $v_ia = xb_i$.

7.2.4 Let R be a Dedekind domain. Using induction on n and the fact that projection maps from R^n to R are module homomorphisms, show that every finitely generated projective module is direct sum of finitely many ideals of R.

7.2.5 A ring R with identity is called a **local ring** if the set $M = R - R^\star$ of non-units form a left ideal of R. Show that M is a two-sided ideal which is maximal ideal. Deduce that R/M is a division ring. Let $[a_{ij}]$ be a $n \times n$ matrix such that the matrix $[a_{ij} + M]$ is invertible in R/M. Show that A is invertible.
Hint. If $[a_{ij} + M]$ is the identity matrix in R/M, then using elementary operations $[a_{ij}]$ can be reduced to identity matrix.

7.2.6 Use the Exercise 7.2.5 to show that every finitely generated projective module over a local ring is free.

7.2.7 R be a ring with identity, and A be a $n \times n$ idempotent matrix with entries in R. Show that R^nA is a finitely generated projective module, where elements of R^n are treated as row matrices. Conversely, show that any finitely generated projective module is isomorphic to such a module.

7.2.8 Let A and B be $m \times m$ idempotent matrices with entries in a ring R. Suppose that there is a invertible $m \times m$ matrix P such that $PAP^{-1} = B$. Show that the projective modules R^mA and R^mB are isomorphic as a module over R.

7.3 Tensor Product and Exterior Power

Let R be a ring with identity. Let M be a right R-module, N a left R-module, and L an abelian group. A map f from $M \times N$ to L is called a **balanced** map if it satisfies the following two conditions:

(i) The map f is additive in both the coordinates in the sense that $f(x+y,u) = f(x,u)+f(y,u)$, $f(x,u+v) = f(x,u)+f(x,v)$ for all $x, y \in M$, and for all $u, v \in N$.
(ii) $f(xa,u) = f(x,au)$ for all $x \in M$, $a \in R$, and $u \in N$.

If further, M, N, and L are both sided R-modules, and in addition to (i) and (ii), we have $f(xa,u) = af(x,u)$, then we say that f is a **bilinear** map.

If f is a balanced map, then it follows from the additivity that $f(0,u) = 0 = f(x,0)$.

Let M be a right R-module, and N be a left R-module. We have the following universal problem:

"Does there exists a pair (L, f), where L is an abelian group, f a balanced map from $M \times N$ to L with the property that if (L', f') is another such pair, then there is a unique homomorphism ϕ from L to L' such that $\phi of = f'$?"

As in earlier cases solution to above problem, if exists, is unique upto isomorphism. For the existence, consider the free abelian group $(F(M \times N), i)$ on $M \times N$. Let A be the subgroup of $F(M \times N)$ generated by the elements of the types

(i) $i(x+y,u) - i(x,u) - i(y,u)$,
(ii) $i(x,u+v) - i(x,u) - i(x,v)$,

and

(iii) $i(xa,u) - i(x,au)$.

Let $L = F(M \times N)/A$, and $f = voi$. We show that (L, f) is a solution to the above problem. Let L' be an abelian group, and g a balanced map from $M \times N$ to L'. From the universal property of a free abelian group, there is a unique homomorphism ϕ from $F(M \times N)$ to L' such that $\phi oi = g$. Since g is a balanced map $\phi(i(x+y,u)-i(x,u)-i(y,u)) = \phi(i(x+y,u))-\phi(i(x,u))-\phi(i(y,u)) = g(x+y,u)-g(x,u)-g(y,u) = 0$. Thus, the elements of the type (i) are contained in the kernel of ϕ. Similarly, elements of the types (ii) and (iii) are also contained in the kernel of ϕ. This shows that A is contained in the kernel of ϕ. From the fundamental theorem of homomorphism, there is a unique homomorphism η from $L = F(M \times N)/A$ to L' such that $\eta ov = \phi$. But, then $\eta of = \eta ovoi = \phi oi = g$. This completes the proof of the fact that (L, f) is the solution to the above universal problem. The abelian group L is denoted by $M \otimes_R N$, and it is called the **tensor** product of M and N. The image $f(m,n) = i(m,n) + A$ is denoted by $m \otimes n$. Thus, $(m,n) \rightsquigarrow m \otimes n$ is a balanced map, and hence

(i) $(x+y) \otimes u = x \otimes u + y \otimes u$,
(ii) $x \otimes (u+v) = x \otimes u + y \otimes v$,

and

(iii) $xa \otimes u = x \otimes au$

for all $x, y \in M$, $a \in R$, and $u, v \in N$.

Also $0 \otimes u = 0 = x \otimes 0$ for all $x \in M$ and $u \in N$.

Further, if L' is an abelian group, and g a balanced map from $M \times N$ to L', then we have a unique homomorphism ϕ from $M \otimes_R N$ to L' defined by the property $\phi(m \otimes n) = g(m, n)$.

Definition 7.3.1 Let R and S be rings with identities. An abelian group M which is a left R-module, and also a right S-module is called a $\mathbf{Bi} - (\mathbf{R}, \mathbf{S})$ module if $(a \cdot x) \cdot b = a \cdot (x \cdot b)$ for all $x \in M$, $a \in R$, and $b \in S$.

Observe that if R is a commutative ring with identity, then a left R-module M is also a right R-module (define $x \cdot a = a \cdot x$). In fact, it is a bi-(R, R) module.

Proposition 7.3.2 *Let M be a right R-module and N a bi-(R, S) module. Then $M \otimes_R N$ has unique right S-module structure defined by $(x \otimes u) \cdot b = (x \otimes (u \cdot b))$. If M is bi-(S, R) module, and N a left R-module, then $M \otimes_R N$ is a left S-module.*

Proof Let M be a right R-module, and N be a bi-(R, S) module. Let $b \in S$. Define a map f_b from $M \times N$ to $M \otimes N$ by $f_b(x, u) = x \otimes ub$. It is easy to observe (using the fact that N is a bi-(R, S) module) that f_b is a balanced map. From the universal property of the tensor product, we have a unique homomorphism ϕ_b from $M \otimes_R N$ to itself defined by the property $\phi_b(x \otimes u) = x \times ub$. Define an external multiplication on $M \otimes_R N$ by elements of S from right by $z \cdot b = f_b(z)$ for all $z \in M \otimes_R N$, and $b \in S$. Since f_b is a homomorphism for all $b \in S$, and $f_{b_1 b_2} = f_{b_2} o f_{b_1}$ for all $b_1, b_2 \in S$, it follows that $M \otimes_R N$ is a right S-module with respect to the external multiplication defined above. The rest can be proved similarly. ♯

In particular, we have the following corollary.

Corollary 7.3.3 *If R is a commutative ring, then $M \otimes_R N$ is a both sided R-module.* ♯

Proposition 7.3.4 *Let M and N be bi-(R, R) modules. Then we have a unique isomorphism f from $M \otimes_R N$ to $N \otimes_R M$ such that $f(x \otimes y) = y \otimes x$.*

Proof The map ϕ from $M \times N$ to $N \otimes_R M$ defined by $\phi(x, y) = y \otimes x$ is a balanced(in fact bilinear) map. From the universal property of the tensor product, we have a unique homomorphism f subject to the condition $f(x \otimes y) = y \otimes x$. Similarly, we have a unique homomorphism g from $N \otimes_R M$ to $M \otimes N$ subject to the condition $g(y \otimes x) = x \otimes y$. Clearly, $gof(x \otimes y) = x \otimes y$ for all $x \in M$ and $y \in N$. Since $\{x \otimes y \mid x \in M, y \in N\}$ is a set of generators of $M \otimes_R N$, it follows that $gof = I_{M \otimes_R N}$. Similarly, fog is also the identity map. This shows that f is an isomorphism. ♯

In particular, we have the following corollary.

Corollary 7.3.5 *Let R be a commutative ring, and M and N be are R-modules. Then $M \otimes_R N$ is isomorphic to $N \otimes_R M$.* ♯

Proposition 7.3.6 *Let M be a right R-module, N a bi-(R, S) module, and L a left S-module. Then there is a unique isomorphism ϕ from $(M \otimes_R N) \otimes_S L$ to $M \otimes_R (N \otimes_S L)$ subject to the condition $\phi((x \otimes y) \otimes z) = x \otimes (y \otimes z)$ for all $x \in M$, $y \in N$, and $z \in L$.*

Proof Let $x \in M$. The map $(y, z) \rightsquigarrow (x \otimes y) \otimes z$ defines a balanced map from $N \times L$ to $(M \otimes_R N) \otimes_S L$. Hence, there is a unique homomorphism ϕ_x from $N \otimes_S L$ to $(M \otimes_R N) \otimes_S L$ subject to the condition $\phi_x(y \otimes z) = (x \otimes y) \otimes z$. The map $(x, u) \rightsquigarrow \phi_x(u)$, where $u \in N \otimes_S L$, is also a balanced map from $M \times (N \otimes_S L)$ to $(M \otimes_R N) \otimes_S L$. Thus, there is a unique homomorphism ϕ from $M \otimes_R (N \otimes_S L)$ to $(M \otimes_R N) \otimes_S L$ subject to the condition $\phi(x \otimes (y \otimes z)) = (x \otimes y) \otimes z$. Similarly, we have a unique homomorphism ψ from $(M \otimes_R N) \otimes_S L$ to $M \otimes_R (N \otimes_S L)$ subject to the condition $\psi((x \otimes y) \otimes z) = x \otimes (y \otimes z)$. It is clear that ϕ and ψ are inverses of each other. ♯

Remark 7.3.7 The above result, in particular, says that if R is a commutative ring with identity, $M_1, M_2, \ldots, M_n$ are R-modules, then the tensor product of $M_1, M_2, \ldots, M_n$ taken in same order with respect to any two bracket arrangements are naturally isomorphic. Thus, we can define the tensor product $M_1 \otimes_R M_2 \otimes_R \cdots \otimes_R M_n$ unambiguously. It is universal with respect to n-linear maps in the sense that if ϕ is a n-linear map from $M_1 \times M_2 \times \cdots \times M_n$ to an R-module L, then there is a unique homomorphism ψ from $M_1 \otimes_R M_2 \otimes_R \cdots \otimes_R M_n$ to L subject to $\psi(x_1 \otimes x_2 \otimes \cdots \otimes x_n) = \phi(x_1, x_2, \ldots, x_n)$.

Proposition 7.3.8 *Let R be a ring with identity. Then there is a unique R-isomorphism f from $R \otimes_R M$ to M defined by $f(a \otimes x) = ax$.*

Proof The map $(a, x) \rightsquigarrow ax$ is clearly a balance map from $R \times M$ to M. Hence there is a unique homomorphism ϕ from $R \otimes_R M$ to M such that $\phi(a \otimes x) = ax$. Also the map ψ from M to $R \otimes_R M$ defined by $\psi(x) = 1 \otimes x$ is a homomorphism. Now, $(\psi o \phi)(a \otimes x) = \psi(ax) = 1 \otimes ax = 1a \otimes x = a \otimes x$. Thus, $\psi o \phi = I_{R \otimes_R M}$. Similarly, $\phi o \psi = I_M$. This shows that ψ is a group isomorphism. Further, $\psi(ax) = 1 \otimes ax = a \otimes x = a \cdot (1 \otimes x) = a\psi(x)$. This shows that ψ is a R-isomorphism. ♯

Proposition 7.3.9 *Let $\{M_\alpha \mid \alpha \in \Lambda\}$ be a family of right R-modules and N a left R-module. Then there is a unique isomorphism $\overline{\phi}$ from $(\oplus\Sigma_{\alpha \in \Lambda} M_\alpha) \otimes_R N$ to $\oplus\Sigma_{\alpha \in \Lambda}(M_\alpha \otimes_R N)$ such that $\overline{\phi}(f \otimes n)(\alpha) = f(\alpha) \otimes n$. Similar result holds if N is a right R-module and M_α is left R-module for each α.*

Proof The map ϕ from $(\oplus\Sigma_{\alpha \in \Lambda} M_\alpha) \times N$ to $\oplus\Sigma_{\alpha \in \Lambda}(M_\alpha \otimes_R N)$ defined by $\phi((f, n))(\alpha) = f(\alpha) \otimes n$ is easily seen to be a balanced map. Hence there is a unique homomorphism $\overline{\phi}$ such that $\overline{\phi}(f \otimes n)(\alpha) = f(\alpha) \otimes n$. The inverse map is an obvious map. The proof of the second part is similar. ♯

Remark A free left R-module is isomorphic to direct sum of several copies of R, and which are also bi-(R, R) modules. Thus, a free left (right) R-module is also a free bi-(R, R) module.

Corollary 7.3.10 *Tensor product of free left R-modules is a free left R-module. In turn, the tensor product $P \otimes Q$ of a projective bi-(R, R) module P with a projective left R-module Q is a projective left R-module.*

Proof Since a free left R-module is direct sum of so many copies of R, and since $R \otimes R$ is isomorphic to R, the first part of the result follows from the above proposition. Further, let P be a projective bi-(R, R) module, and Q a projective left R-module. Then there exists a right R-module L, and a left R-module M such that such that $P \oplus L$ is a free R-module, and $Q \oplus M$ is also a free R-module. Since tensor product of free R-modules are free R-modules, $(P \oplus L) \otimes (Q \oplus M)$ is a free R-module. From the previous proposition, $(P \otimes Q) \oplus U$ is free, where $U = (P \otimes M) \oplus (L \otimes Q) \oplus (L \otimes M)$. Hence $P \otimes Q$ is a projective module. ♯

Corollary 7.3.11 *Let V and W be finite-dimensional vector spaces over a field F. Then $dim(V \otimes_F W) = dim(V) \cdot dim(W)$.*

Proof Suppose that $dim(V) = n$, and $dim(W) = m$. Then V is isomorphic to direct sum of n copies of F, and W is isomorphic to m copies of F. From the above proposition it follows that $V \otimes W$ is isomorphic to the direct sum of nm copies of F. The result follows. ♯

Remark 7.3.12 Let $\{e_1, e_2, \ldots, e_n\}$ be a basis of V, and $\{f_1, f_2, \ldots, f_m\}$ be a basis of W. Then $\{e_i \otimes f_j \mid 1 \leq i \leq n, 1 \leq j \leq m\}$ is a set of generators of $V \otimes W$, and so it is a basis of $V \otimes W$.

Example 7.3.13 We show that $\mathbb{Z}_m \otimes_{\mathbb{Z}} \mathbb{Z}_n$ is isomorphic to $\mathbb{Z}_d$, where d is the g.c.d of m and n: Suppose that $\overline{a} = \overline{a'}$ in $\mathbb{Z}_m$, and $\overline{b} = \overline{b'}$ in $\mathbb{Z}_n$. Then m divides $a - a'$, and n divides $b - b'$. This means d divides $a - a'$, and it also divides $b - b'$. In turn, d divides $ab - a'b'$. Hence $\overline{ab} = \overline{a'b'}$ in $\mathbb{Z}_d$. Thus, we have a map f from $\mathbb{Z}_m \times \mathbb{Z}_n$ to $\mathbb{Z}_d$ defined by $f(\overline{a}, \overline{b}) = \overline{ab}$. Evidently, f is a balanced map, and so it induces a unique homomorphism $\overline{f}$ from $\mathbb{Z}_m \otimes_{\mathbb{Z}} \mathbb{Z}_n$ to $\mathbb{Z}_d$ such that $\overline{f}(\overline{a} \otimes \overline{b}) = \overline{ab}$. Clearly, $\overline{f}$ is surjective. Suppose that $\overline{ab} = \overline{0}$ in $\mathbb{Z}_d$. Then d divides ab. By the Euclidean algorithm, there are integers u and v such that $d = um + vn$. Since d divides ab, there are integers r and s such that $rm + sn = ab$. But, then

$$\overline{a} \otimes \overline{b} = \overline{a} \otimes b\overline{1} = \overline{ab} \otimes \overline{1} = \overline{ab} \otimes \overline{1} = \overline{rm + sn} \otimes \overline{1} = \overline{sn} \otimes 1 = \overline{s} \otimes n\overline{1} = \overline{s} \otimes \overline{0} = 0$$

This shows that kernel of $\overline{f}$ is $\{0\}$, and so $\overline{f}$ is also injective.

Let f_1 be a R-homomorphism from a right R-module M_1 to a R-module M_2, and g_1 a R-homomorphism from a left R-module N_1 to a left R-module N_2. The map $f_1 \times g_1$ from $M_1 \times N_1$ to $M_2 \otimes_R N_2$ defined by $(f_1 \times g_1)(x, y) = f_1(x) \otimes g_1(y)$ is a balanced map(verify). Hence it induces a unique homomorphism $f_1 \otimes g_1$, called the tensor product of f_1 and g_1, from $M_1 \otimes_R N_1$ to $M_2 \otimes_R N_2$ such that $(f_1 \otimes g_1)(x \otimes y) = f_1(x) \otimes g_1(y)$. Since $\{u \otimes v \mid u \in M_2, v \in N_2\}$ is a set of generators of $M_2 \otimes_R N_2$,

it follows that tensor product of any two surjective homomorphisms is a surjective homomorphism. Suppose further that f_2 is a homomorphism from M_2 to M_3, and g_2 that from N_2 to N_3. Then

$$(f_2 o f_1) \otimes (g_2 o g_1) \;=\; (f_2 \otimes g_2) o (f_1 \otimes g_1).$$

If N is a left R-module, then the homomorphism $f \otimes I_N$ from $M_1 \otimes_R N$ to $M_2 \otimes_R N$ is denoted by $f_\star$. It is clear that $(gof)_\star = g_\star o f_\star$, and $0_\star = 0$. Also $I_M \otimes I_N = I_{M \otimes_R N}$.

Tensoring is a right exact functor in the sense of the following theorem.

Theorem 7.3.14 *Let*

$$M_1 \xrightarrow{\alpha} M_2 \xrightarrow{\beta} M_3 \longrightarrow 0$$

be an exact sequence of right R-modules, and N be a left R-module. Then the sequence

$$M_1 \otimes_R N \xrightarrow{\alpha_\star} M_2 \otimes_R N \xrightarrow{\beta_\star} M_3 \otimes_R N \longrightarrow 0$$

is also exact.

Proof Since β is surjective, and the tensor product of surjective homomorphisms are surjective, $\beta_\star$ is surjective. Thus, we need to show that $ker\beta_\star = image\alpha_\star$. Again, since $\beta o \alpha = 0$, we have $0 = (\beta o \alpha)_\star = \beta_\star o \alpha_\star$. Hence $image\alpha_\star \subseteq ker\beta_\star$. Put $image\alpha_\star = L$. By the fundamental theorem of homomorphism, we have a unique homomorphism ϕ from $(M_2 \otimes_R N)/L$ to $M_3 \otimes_R N$ defined by $\phi((m_2 \otimes n) + L) = \beta(m_2) \otimes n$. It is sufficient to show that ϕ is an isomorphism. We construct its inverse. If $\beta(m_2) = \beta(m_2')$, then $m_2 - m_2'$ belongs to $ker\beta = image\alpha$. Hence $(m_2 \otimes n) - (m_2' \otimes n) = (m_2 - m_2') \otimes n$ is in L. This ensures that we have a map $(m_3, n) \rightsquigarrow m_2 \otimes n + L$ from $M_3 \times N$ to $(M_2 \otimes_R N)/L$ where $\beta(m_2) = m_3$. This is a balanced map(verify). Hence we have a unique homomorphism ψ from $M_3 \otimes_R N$ to $(M_2 \otimes_R N)/L$ such that $\psi(m_3 \otimes n) = m_2 \otimes n + L$, where $\beta(m_2) = m_3$. Now $(\phi o \psi)(m_3 \otimes n) = \phi(m_2 \otimes n + L) = \beta(m_2) \otimes n = m_3 \otimes n$. This shows that $\phi o \psi$ is the identity map. Similarly, $\psi o \phi$ is also the identity map. This proves that ϕ is an isomorphism, and so $L = ker\beta_\star$. ♯

Remark 7.3.15 Consider the homomorphism f from $\mathbb{Z}$ to $\mathbb{Z}$ defined by $f(a) = 5a$. Then f is injective but $f_\star$ from $\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}_5$ to itself is the zero map, for $f_\star(m \otimes \overline{a}) = f(m) \otimes \overline{a} = 5m \otimes \overline{a} = m \otimes 5\overline{a} = 0$. Since $\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}_5 \approx \mathbb{Z}_5$ is nontrivial, $f_\star$ is not injective. This shows that tensoring is not left exact.

Example 7.3.16 Let A be an abelian group. Then $\mathbb{Z}_m \otimes_{\mathbb{Z}} A$ is isomorphic to A/mA: Consider the exact sequence

$$0 \longrightarrow \mathbb{Z} \xrightarrow{\alpha} \mathbb{Z} \xrightarrow{\nu} \mathbb{Z}_m \longrightarrow 0$$

where α is the multiplication by m. Taking tensor product with A, and observing the fact that tensoring is right exact, we see that $\mathbb{Z}_m \otimes_{\mathbb{Z}} A$ is isomorphic to $(\mathbb{Z} \otimes_{\mathbb{Z}} A)/ker\nu_\star$. Again $ker\nu_\star = image\alpha_\star$. The isomorphism f from $\mathbb{Z} \otimes_{\mathbb{Z}} A$ to A given by $f(n \otimes a) = na$ takes $image\alpha_\star$ to mA. The assertion follows from the fundamental theorem of homomorphism.

Let V be a vector space over a field F. Let $\overset{s}{\otimes} V$ denote s times tensor product of V with itself, and $\overset{r}{\otimes} V^\star$ denote r times tensor product of the dual space $V^\star$ with itself. Let V_s^r denote the tensor product $(\overset{s}{\otimes} V) \otimes (\overset{r}{\otimes} V^\star)$. The members of V_s^r are called **tensors** of the type (r, s). Tensor product induces a multiplication in $T(V) = \oplus \Sigma_{r,s} V_s^r$ with respect to which it is an associative algebra called the **tensor algebra** of V. The Riemann's metric tensor is an example of a tensor of order $(2, 0)$.

Let V be a vector space over F. Let W denote the subspace of $\overset{r}{\otimes} V$ generated by $\{x_1 \otimes x_2 \otimes \cdots \otimes x_r \mid x_i = x_j \, for \, some \, i \neq j\}$, and let $\bigwedge^r V$ denote the quotient space $(\overset{r}{\otimes} V)/W$. Let us denote the coset $x_1 \otimes x_2 \otimes \cdots \otimes x_r + W$ by $x_1 \bigwedge x_2 \bigwedge \cdots \bigwedge x_r$. The map f from V^r to $\bigwedge^r V$ defined by

$$f(x_1, x_2, \ldots, x_r) = x_1 \bigwedge x_2 \bigwedge \cdots \bigwedge x_r$$

is r-alternating, and the pair $(\bigwedge^r V, f)$ is universal in the sense that if g is any r-alternating map from V^r to a space U, then there is a unique linear transformation η from $\bigwedge^r V$ to U such that

$$\eta(x_1 \bigwedge x_2 \bigwedge \cdots \bigwedge x_r) = g(x_1, x_2, \ldots, x_r).$$

The pair $(\bigwedge^r V, f)$ is called the rth **exterior power** of V.

If T is a linear transformation from V to W. Then the map $\times^r T$ from V^r to $\bigwedge^r W$ defined by $(\times^r T)(x_1, x_2, \ldots, x_r) = T(x_1) \bigwedge T(x_2) \bigwedge \cdots \bigwedge T(x_r)$ is an r - alternating map, and so it induces a unique homomorphism $\bigwedge^r T$ from $\bigwedge^r V$ to $\bigwedge^r W$ which takes $x_1 \bigwedge x_2 \bigwedge \cdots \bigwedge x_r$ to $T(x_1) \bigwedge T(x_2) \bigwedge \cdots \bigwedge T(x_r)$. It is easy to verify that $\bigwedge^r (T'0T) = (\bigwedge^r T')o(\bigwedge^r T)$, and $\bigwedge^r I_V = I_{\bigwedge^r V}$. In particular, if T is an isomorphism, then $\bigwedge^r T$ is an isomorphism for all r.

If V is vector space of dimension n, then any m-alternating map on V for $m > n$ is zero map (for an r alternating map takes any linearly dependent r tuple to 0). Thus, we have the following proposition.

Proposition 7.3.17 *Let V be a vector space of dimension n. Then $\bigwedge^m V = \{0\}$ for all $m > n$.* ♯

Theorem 7.3.18 *Let V be a vector space of dimension n. Then $dim \bigwedge^n V = 1$.*

Proof Let $\{e_1, e_2, \ldots, e_n\}$ be an ordered basis of V. If f is an n-alternating map, then for any ordered n-tuple $\{x_1, x_2, \ldots, x_n\}$, $f(x_1, x_2, \ldots, x_n) = detA \cdot f(e_1, e_2, \ldots, e_n)$, where $x_j = \Sigma_{i=1}^n a_{ij} e_i$, and $A = [a_{ij}]$. Thus, f is determined uniquely by its value $f(e_1, e_2, \ldots, e_n)$. This shows that $dim \bigwedge^n V$ is at most 1. Also the map f defined by $f(x_1, x_2, \ldots, x_n) = detA$ defines a nonzero n-alternating map(indeed, $f(e_1, e_2, \ldots, e_n) = 1$). This shows that the dimension of $\bigwedge^n V$ is 1. ♯

Let V be a vector space of dimension n, and T a linear transformation on V. Then $\bigwedge^n T$ is a linear transformation on $\bigwedge^n V$. Since $\bigwedge^n V$ is of dimension 1, the linear transformation $\bigwedge^n T$ is multiplication by a scalar. This scalar is called the determinant of T. It is easy to observe that this definition of determinant agrees with the definition of determinant in the Chap. 5.

Theorem 7.3.19 *Let V be a vector space of dimension n, and $r \leq n$. Then $dim \bigwedge^r V = {}^nC_r$.*

Proof For $r = n$, the result is the content of the above theorem. Let $\{e_1, e_2, \ldots, e_n\}$ be a basis of V. Then as observed in the above theorem $e_1 \bigwedge e_2 \bigwedge \cdots \bigwedge e_n$ is nonzero. Consider the subset $S = \{e_{i_1} \bigwedge e_{i_2} \bigwedge \cdots \bigwedge e_{i_r} \mid i_1 < i_2 < \cdots < e_{i_r}\}$ of $\bigwedge^r V$. Clearly, every member of $\bigwedge^r V$ is a linear combination of members of S, and so it is a set of generators. We show that S is linearly independent. Suppose that

$$\Sigma_{i_1 < i_2 < \cdots < i_r} a_{i_1 i_2 \cdots i_r} (e_{i_1} \bigwedge e_{i_2} \bigwedge \cdots \bigwedge e_{i_r}) = 0.$$

Fix $j_1 < j_2 < \cdots < j_r$. Suppose that

$$\{1, 2, \ldots, n\} - \{j_1, j_2, \ldots, j_r\} = \{j_{r+1}, j_{r+2}, \ldots, j_n\}.$$

Taking the exterior product with $j_{r+1} \bigwedge j_{r+2} \bigwedge \cdots \bigwedge j_n$ we obtain that

$$a_{j_1 j_2 \cdots j_r} e_{j_1} \bigwedge e_{j_2} \bigwedge \cdots \bigwedge e_{j_r} \bigwedge e_{j_{r+1}} \bigwedge \cdots \bigwedge e_{j_n} = 0.$$

Since $j_k \neq j_l$ for all $k \neq l$, it follows that

$$e_{j_1} \bigwedge e_{j_2} \bigwedge \cdots \bigwedge e_{j_n} = \pm(e_1 \bigwedge e_2 \bigwedge \cdots \bigwedge e_n) \neq 0.$$

Hence $a_{j_1 j_2 \cdots j_n} = 0$. This shows that S is linearly independent, and so it is a basis. Clearly, the number of elements in S is ${}^nC_r$. ♯

The exterior product gives us an external multiplication from $\bigwedge^r V \times \bigwedge^s V$ to $\bigwedge^{r+s} V$, and this can be extended linearly to a multiplication on $E(V) =$

$\oplus\Sigma_{r=0}^{\infty}(\bigwedge^r V)$ with respect to which it is an associative algebra, and it is called the **exterior algebra** of V. Also,

$$dim(E(V)) = \Sigma_{r=0}^{n} Dim \bigwedge^{r} V = {}^nC_0 + {}^nC_1 + \cdots + {}^nC_n = 2^n.$$

Exercises

7.3.1 Show that $\mathbb{Z}_5 \otimes_{\mathbb{Z}} \mathbb{Z}_3$ is the trivial group.

7.3.2 Show that $A \otimes_{\mathbb{Z}} \mathbb{Z}_m$ is trivial whenever A is divisible. Deduce that $\mathbb{Q} \otimes_{\mathbb{Z}} \mathbb{Z}_m$ is trivial.

7.3.3 Let A be an abelian group of exponent m. Show that $A \otimes_{\mathbb{Z}} \mathbb{Z}_m$ is isomorphic to A.

7.3.4 Show that tensoring takes a split exact sequence to a split exact sequence.

7.3.5 Find $\mathbb{Q} \otimes_{\mathbb{Z}} \mathbb{Q}$.

7.3.6 Let R be a commutative ring. Show that $Hom_R(A \otimes_R B, C)$ is isomorphic to $Hom_R(A, Hom_R(B, C))$.

7.3.7 Show that the definition of determinant in this section agrees with the definition of determinant given in the previous chapter, and establish all properties of determinant using this definition.

7.3.8 Let V be a vector space of dimension at least 2. Show that a linear transformation T on V is an isomorphism if and only if $\bigwedge^r T$ is an isomorphism for all $r \geq 2$.

7.4 Lower K-theory

In this section, we shall introduce and discuss the functors K_0 and K_1 from the category of rings to the category of abelian groups. Let R be a ring with identity. Let $\wp(R)$ denote the set of isomorphism classes of finitely generated projective left R-modules (Note that $\wp(R)$ is, indeed, a set). The isomorphism class of projective module determined by P will be denoted by $[P]$. Let $K_0(R)$ denote the abelian group generated by $\wp(R)$ subject to the relation

$$[P] + [Q] = [P \oplus Q].$$

More precisely, $K_0(R) = F/N$, where F is the free abelian group with basis $\wp(R)$, and N is the subgroup of F generated by the set of elements of the type $[P] + [Q] - [P \oplus Q]$. The group $K_0(R)$ is also called the **Grothendieck**

group of the ring R. It is also called the Grothendieck group of the category P_R of finitely generated left projective modules over R. The coset $[P] + N$ is denoted by $< P >$. Thus, the elements of the type $< P >$ generate $K_0(R)$. Clearly, any element of $K_0(R)$ is expressible as $< P > - < Q >$.

Definition 7.4.1 Two finitely generated projective R-modules P and Q are said to be **stably isomorphic**, if $P \oplus R^n$ is isomorphic to $Q \oplus R^n$ for some n. A projective R-module P is said to be stably free if it is stably isomorphic to a free R-module.

Remark 7.4.2 Clearly, isomorphic projective modules are stably isomorphic. However, two stably isomorphic projective modules need not be isomorphic. For example, let V be an infinite-dimensional vector space over a field F, and R the ring of endomorphisms of V. Then the module R over R is isomorphic to the module $R \oplus R$ (check it). As such, the trivial module is stably isomorphic to the module R, but R is not isomorphic to the trivial module.

Proposition 7.4.3 *$< P > = < Q >$ if and only if P and Q are stably isomorphic. In turn, $< P > - < Q > = < P' > - < Q' >$ if and only if $P \oplus Q'$ is stably isomorphic to $P' \oplus Q$.*

Proof Suppose that P is stably isomorphic to Q. Then $[P \oplus R^n] = [Q \oplus R^n]$ for some n. This means that $< P > + < R^n > = < Q > + < R^n >$. This shows that $< P > = < Q >$. Conversely, suppose that $< P > = < Q >$. Then $[P] - [Q] \in N$. Since N is the subgroup of F generated by the elements of the type $[P] + [Q] - [P \oplus Q]$, there exist elements $[P_i], [Q_i], i = 1, 2, \ldots, n$, and $[L_j], [M_j], j = 1, 2, \ldots, m$ in F such that

$$[P] - [Q] = \Sigma_{i=1}^{n}([P_i]+[Q_i] - [P_i \oplus Q_i]) - \Sigma_{j=1}^{m}([L_j]+[M_j] - [L_j \oplus M_j])$$

in F. Equivalently,

$$[P] + \Sigma_{i=1}^{n}[P_i \oplus Q_i] + \Sigma_{j=1}^{m}([L_j] + [M_j]) = [Q] + \Sigma_{j=1}^{m}[L_j \oplus M_j] + \Sigma_{i=1}^{n}([P_i] + [Q_i]).$$

Since F is free abelian on $\wp(R)$, there is a bijective correspondence between the set of terms in the sum of the LHS to the set of terms in the sum of RHS so that the corresponding terms represent same elements in $\wp(R)$. This ensures the existence of a finitely generated projective module U such that $P \oplus U$ is isomorphic to $Q \oplus U$. Since U is projective, there is a module V such that $U \oplus V$ is isomorphic to R^n for some n. But, then $P \oplus R^n$ is isomorphic to $Q \oplus R^n$. This shows that P is stably isomorphic to Q. The rest follows immediately. ♯

Let f be a homomorphism from a ring R_1 to a ring R_2. A ring homomorphism is always assumed to preserve the identities of the rings. R_2 can be treated as a right R_1-module by defining $r \cdot a = rf(a), r \in R_2, a \in R_1$. In fact, then R_2 is a $bi-(R_2, R_1)$ module. If M is a left R_1-module, then $R_2 \otimes_{R_1} M$ is a left R_2-module.

We denote this R_2-module by $f_\sharp(M)$. Since tensor product distributes over direct sum, the following assertions can be easily verified.

(i) $f_\sharp(P \oplus Q) \approx f_\sharp(P) \oplus f_\sharp(Q)$,
(ii) $f_\sharp$ takes finitely generated modules to finitely generated modules,
(iii) $f_\sharp$ takes free modules to free modules,
(iv) $f_\sharp$ takes projective modules to projective modules,
(v) $f_\sharp$ defines a map from $\wp(R_1)$ to $\wp(R_2)$, and it respects the relation $[P] + [Q] = [P \oplus Q]$.

In turn, f induces a homomorphism $K_0(f)$ from $K_0(R_1)$ to $K_0(R_2)$ given by $K_0(f)(< P > - < Q >) = < f_\sharp(P) > - < f_\sharp(Q) >$. Further, (i) $K_0(I_R) = I_{K_0(R)}$, and (ii) $K_0(gof) = K_0(g)oK_0(f)$. In the language of category theory, it says that K_0 is a functor from the category of rings to the category of abelian groups.

If R is a commutative ring, P and Q are finitely generated projective modules, then $P \otimes Q$ is again a finitely generated projective module. Since tensor product distributes over direct sum, we have a product $\cdot$ on $K_0(R)$ given by $(< P > - < Q >) \cdot (< P' > - < Q' >) = < P \otimes Q > + < Q \otimes Q' > - < P \otimes Q' > - < Q \otimes P' >$. It follows that $K_0(R)$ is a commutative ring. Thus, K_0 defines a functor from the category of commutative rings to the category of commutative rings also.

Proposition 7.4.4 *Let R be a ring such that the following hold:*

(i) R^n is isomorphic R^m if and only if $n = m$.
(ii) Every finitely generated projective module is free.

Then $K_0(R)$ is isomorphic to the group of integers.

Proof Under the hypothesis of the proposition $\wp(R) = [R^n]$, and R^n is stably isomorphic to R^m if and only if $n = m$. Thus, $[R^n] = < R^n >$. The map η from the set $\{< R^n > \mid n \in \mathbb{N}\}$ to $\mathbb{N}$ defined by $\eta(< R^n >) = n$ is a bijective map such that $\eta < R^n \oplus R^m > = \eta(< R^n >) + \eta(< R^m >)$. Suppose that $< R^n > - < R^m > = < R^r > - < R^s >$. Then $< R^{n+s} > = < R^{m+r} >$. This means that $n + s = m + r$. Thus, η can be extended to a map $\overline{\eta}$ from $K_0(R)$ to $\mathbb{Z}$ given by $\overline{\eta}(< R^n > - < R^m >) = n - m$. Clearly, $\overline{\eta}$ is an isomorphism. ♯

Corollary 7.4.5 *(i) If D is a division ring, then $K_0(D)$ is the group of integers.*
(ii) If R is a principal ideal domain, then $K_0(R)$ is the group of integers.
(iii) If R is a local ring, then $K_0(R)$ is the group of integers.

Proof In each of the cases, the hypothesis of the above proposition is satisfied (for local ring see Exercise 5, 6, and 7 of the Sect. 7.2. ♯

Now, we introduce the functor K_1 from the category of rings to the category of abelian groups. Let R be a ring with identity. Let us denote by $GL(n, R)$ the group of invertible $n \times n$ matrices with entries in R. There is a natural embedding of $GL(n, R)$ in to $GL(n + 1, R)$given by

$$A \rightsquigarrow \begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix}.$$

Under this embedding, $GL(n, R)$ can be treated as a subgroup of $GL(n+1, R)$. We get a chain of groups

$$GL(1, R) \subseteq GL(2, R) \subseteq \cdots \subseteq GL(n, R) \subseteq GL(n+1, R) \subseteq \cdots\cdots$$

The union of this chain is a group denoted by $GL(R)$, and it is called the **general linear group** over the ring R. The elementary $n \times n$ matrices E_{ij}^{λ} (also called transvections) are members of $GL(n, R)$. The Steinberg relations still hold among these matrices. Let $E(n, R)$ denote the subgroup of $GL(n, R)$ generated by these elementary matrices. We have a chain

$$E(1, R) \subseteq E(2, R) \subseteq \cdots \subseteq E(n, R) \subseteq E(n+1, R) \subseteq \cdots\cdots$$

of subgroups of $GL(R)$. The union $E(R)$ of this chain is a subgroup of $GL(R)$. Recall that for a field F, every matrix of determinant 1 can be reduced to the identity matrix using the elementary operations corresponding to transvections. In other words, the special linear group $SL(F) \subseteq E(F)$. Already, the elements of $E(F)$ are of determinant 1. Thus, $SL(F) = E(F)$.

Proposition 7.4.6 *$E(R)$ is a perfect group in the sense that $[E(R), E(R)] = E(R)$.*

Proof One of the Steinberg relation is $[E_{ij}^{\lambda}, E_{jl}^{\mu}] = E_{il}^{\lambda\mu}$. Taking $\lambda = 1$, we observe that every transvection is a member of $[E(R), E(R)]$. Hence, $[E(R), E(R)] = E(R)$. ♯

Proposition 7.4.7 *Matrices of the type*

$$\begin{bmatrix} I & A \\ 0 & I \end{bmatrix}, \begin{bmatrix} I & 0 \\ A & I \end{bmatrix}, \text{ and } \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix}.$$

are members of $E(R)$.

Proof The result follows, if we observe that the matrices described in the proposition can be reduced to the identity matrix by applying the elementary operations associated to the matrices E_{ij}^{λ}. ♯

Corollary 7.4.8 *Let $A \in GL(R)$. Then the matrix*

$$\begin{bmatrix} A & 0 \\ 0 & A^{-1} \end{bmatrix}.$$

is a member of $E(R)$.

Proof Follows from the above proposition and the identity

$$\begin{bmatrix} A & 0 \\ 0 & A^{-1} \end{bmatrix} = \begin{bmatrix} I & A \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ -A^{-1} & I \end{bmatrix} \begin{bmatrix} I & A \\ 0 & I \end{bmatrix} \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix}.$$

♯

Lemma 7.4.9 ***Whitehead Lemma***. $[GL(R), GL(R)] = E(R)$.

Proof Already $E(R) = [E(R), E(R)] \subseteq [GL(R), GL(R)]$. Thus, it is sufficient to observe that any commutator $ABA^{-1}B^{-1}$ in $GL(n, R)$ treated as an element of $GL(2n, R)$ can be expressed as

$$ABA^{-1}B^{-1} = \begin{bmatrix} A & 0 \\ 0 & A^{-1} \end{bmatrix} \begin{bmatrix} B & 0 \\ 0 & B^{-1} \end{bmatrix} \begin{bmatrix} (BA)^{-1} & 0 \\ 0 & BA \end{bmatrix}.$$

♯

Definition 7.4.10 The abelian group $GL(R)/E(R)$ is called the **Whitehead group** of the ring R, and it is denoted by $K_1(R)$.

Thus, the Whitehead group $K_1(R)$ can be viewed as the group of equivalent matrices in $GL(R)$, where two matrices A and B in $GL(R)$ is said to be equivalent one can be obtained from the other by using elementary operations associated to transvections. For example, over fields, two matrices are equivalent if and only if they have same determinant. Note that a nonsingular matrix A with entries in a field F can be reduced to the matrix $diag(1, 1, \ldots, 1, det A)$ by using the elementary operations associated to transvections.

If f is a homomorphism from a ring R_1 to a ring R_2, then it induces a map from $M_n(R_1)$ to $M_n(R_2)$ which takes $A = [a_{ij}]$ to $f(A) = [b_{ij}]$, where $b_{ij} = f(a_{ij})$. In fact, it maps $GL(R_1)$ to $GL(R_2)$, and $E(R_1)$ to $E(R_2)$. In turn, it induces a homomorphism $K_1(f)$ from $K_1(R_1)$ to $K_1(R_2)$. It can be easily observed that (i) $K_1(gof) = K_1(g)oK_1(f)$, and (ii) $K_1(I_R) = I_{K_1(R)}$. In the language of category theory, K_1 defines another functor from the category of rings to the category of abelian groups.

If R is a commutative ring, then determinant of a square matrix with entries in R makes sense, and then every element of $E(R)$ is of determinant 1. Thus, $E(R) \subseteq SL(R)$. In general $E(R) \neq SL(R)$. We denote the group $SL(R)/E(R)$ by $SK_1(R)$. It follows that

$$K_1(R) = SK_1(R) \oplus U(R),$$

where $U(R)$ is the group of units of R. For most of the commutative rings R, $SL(R) = E(R)$, and in such cases $K_1(R) = U(R)$. For example, if R is a Field, or a Local ring, or an Euclidean domain, or the ring of integers in a number field, the matrices of determinant 1 can be reduced to the identity matrix by using elementary operations associated to the transvections E_{ij}^{λ}. Thus, in these cases

$SL(R) = E(R)$, and so $K_1(R) = U(R)$. However, there are commutative rings in which a matrix of determinant 1 may not be reducible to the identity matrix by using elementary operations associated to transvections. For example, consider the ring $R = \mathbb{R}[x, y]/\Gamma$, where Γ is the ideal generated by $x^2 + y^2 - 1$. Then the matrix

$$A = \begin{bmatrix} \tilde{x} & -\tilde{y} \\ \tilde{y} & \tilde{x} \end{bmatrix},$$

where $\tilde{x} = x + \Gamma$ and $\tilde{y} = y + \Gamma$ is a matrix of determinant 1. Using topological arguments (see "Algebraic K-Theory" by Milnor, p. 58), it can be shown that no nontrivial power of A can be in $E(R)$. Thus, $SK_1(R)$ contains an element of infinite order.

Exercises

7.4.1 Compute $K_0(\mathbb{Z}_6)$, and also $K_1(\mathbb{Z}_6)$.

7.4.2 Find $K_0(R)$, where R is the ring of endomorphisms of an infinite-dimensional vector space V.

7.4.3 Determine $K_0(M_2(\mathbb{C}))$.

7.4.4 Show that $K_0(R_1 \times R_2) \approx K_0(R_1) \times K_0(R_2)$.

7.4.5 Determine $K_0(\mathbb{Z}[i])$ and $K_1(\mathbb{Z}[i])$.

Chapter 8
Field Theory, Galois Theory

This chapter is devoted to the theory of fields, Galois theory, geometric constructions by ruler and compass, and the theorem of Abel–Ruffini about the polynomials equations of degree n, $n \geq 5$. We also discuss cubic and biquadratic equations.

8.1 Field Extensions

Let K be a subfield of a field L. Then we say that L is a **field extension** of K. The notation L/K is used to say that L is a field extension of K. If L is a field extension of K, then $(L, +)$ is a vector space over K (the multiplication by scalars being the field multiplication). If the dimension of $(L, +)$ over K is infinite, then we say that L is infinite extension of K. If the dimension of $(L, +)$ over K is finite, then the dimension of $(L, +)$ over K is called the **degree of the extension**, and it is denoted by $[L : K]$.

Proposition 8.1.1 *Let K be a finite field. Then the number of elements in K is p^n for some prime p, and for some $n \in \mathbb{N}$.*

Proof Since K is a finite field, its characteristic is a prime p. The map $\bar{i} \rightsquigarrow i \cdot 1$ is an injective homomorphism from the field $\mathbb{Z}_p$ to the field K. Thus, $\mathbb{Z}_p$ can be considered as a subfield of K, and since K is finite, it is a finite-dimensional vector space over $\mathbb{Z}_p$. Suppose that the dimension of K over $\mathbb{Z}_p$ is n. Then K, as a vector space over $\mathbb{Z}_p$, is isomorphic to $\mathbb{Z}_p^n$. This shows that $| K | = p^n$. ♯

Proposition 8.1.2 *Let L/K and K/F be finite field extensions. Then L/F is also finite field extension, and $[L : F] = [L : K][K : F]$.*

Proof Suppose that $[L : K] = n$ and $[K : F] = m$. Let $\{x_1, x_2, \ldots, x_n\}$ be a basis of the vector space L over K, and $\{y_1, y_2, \ldots, y_m\}$ be a basis of K over F. We show that $S = \{x_i y_j \mid 1 \leq i \leq n, 1 \leq j \leq m\}$ is a basis of L over F. Let $x \in L$. Since $\{x_1, x_2, \ldots, x_n\}$ is a basis of L over K, $x = a_1x_1 + a_2x_2 + \cdots + a_nx_n$ for some $a_1, a_2, \ldots, a_n$ in K. Further, since $\{y_1, y_2, \ldots, y_m\}$ is a basis of K over

R. Lal, *Algebra 2*, Infosys Science Foundation Series in Mathematical Sciences,
DOI 10.1007/978-981-10-4256-0_8

$F, a_i = \Sigma_{j=1}^{m} \alpha_{ji} y_j$ for some $\alpha_{ji} \in F$. Thus, $x = \Sigma_{j,i} \alpha_{ji} x_i y_j$. This shows that S generates the vector space L over F. Now, we show that S is linearly independent over F. Suppose that $\Sigma_{j,i} \alpha_{ji} x_i y_j = 0$. Then $\Sigma_{i=1}^{n} (\Sigma_{j=1}^{m} \alpha_{ji} y_j) x_i = 0$. Since $\{x_1, x_2, \ldots, x_n\}$ is linearly independent over K, $\Sigma_{j=1}^{m} \alpha_{ji} y_j = 0$ *for all* i. Again, since, $\{y_1, y_2, \ldots, y_m\}$ is linearly independent over F, $\alpha_{ji} = 0$ *for all* j, i. ♯

Corollary 8.1.3 *Let L be a finite extension of K, and a F a subfield of L containing K. Then $[L : F]/[L : K]$, and also $[F : K]/[L : K]$.* ♯

Corollary 8.1.4 *Let L be a finite field containing p^n elements. Let K be a subfield of L. Then K contains p^m elements, where m divides n.*

Proof Since K is a subfield of L, $charK = charL = p$. Thus K contains p^m elements. Suppose that $[L : K] = r$. Then L, as a vector space over K, is isomorphic to K^r, and so $p^n = \mid L \mid = (p^m)^r = p^{mr}$. Hence $n = mr$. ♯

Let L be a field extension of K. Let S be a subset of L. The subring of L generated by $K \bigcup S$ will be denoted by $K[S]$. Clearly, $K[S]$ is the intersection of all subrings of L containing $K \bigcup S$. The subfield of L generated by $K \bigcup S$ is the intersection of all subfields containing $K \bigcup S$, and it is denoted by $K(S)$. Clearly, $K(S)$ is the field of fractions of $K[S]$. If S is finite, and $K(S) = L$, then we say that L is a **finitely generated field extension** of K. The ring $K[S]$ is called the **finitely generated ring extension** of K.

Let $S = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be a finite subset of L. Let $f(X_1, X_2, \ldots, X_n)$ be a polynomial in $K[X_1, X_2, \ldots, X_n]$. Then $f(\alpha_1, \alpha_2, \ldots, \alpha_n)$ denotes the element of L which is obtained by substituting α_i at the place of X_i in the polynomial $f(X_1, X_2, \ldots, X_n)$ for each i. It is clear that $f(\alpha_1, \alpha_2, \ldots, \alpha_n)$ belongs to each subring which contains $K \bigcup S$. Consider the map η from $K[X_1, X_2, \ldots, X_n]$ to L defined by $\eta(f(X_1, X_2, \ldots, X_n)) = f(\alpha_1, \alpha_2, \ldots, \alpha_n)$. Clearly, the map η is a ring homomorphism whose image is the subring of L generated by $K \bigcup S$. This subring is denoted by $K[\alpha_1, \alpha_2, \ldots, \alpha_n]$. By the fundamental theorem of homomorphism, $K[X_1, X_2, \ldots, X_n]/ker\eta$ is isomorphic to $K[\alpha_1, \alpha_2, \ldots, \alpha_n]$. If $ker\eta = \{0\}$, then we say that the set $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is **algebraically independent**. More explicitly, $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is said to be algebraically independent if there is no nonzero polynomial $f(X_1, X_2, \ldots, X_n)$ such that $f(\alpha_1, \alpha_2, \ldots, \alpha_n) = 0$. It is said to be **algebraically dependent** otherwise. It follows that $K[\alpha_1, \alpha_2, \ldots, \alpha_n]$ is the subfield $K(\alpha_1, \alpha_2, \ldots, \alpha_n)$ of L generated by $K \bigcup \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ if and only if the $ker\eta$ is a maximal ideal. We shall see that this is, indeed, the case if and only if each α_i is a root of a nonzero polynomial in $K[X]$. Such elements are called **algebraic elements** over K. We first consider the case when $n = 1$.

Let L be a field extension of K and $\alpha \in L$. Consider the map η from $K[X]$ to $K[\alpha]$ defined by $\eta(f(X)) = f(\alpha)$. Then η is a surjective ring homomorphism. Thus, by the fundamental theorem of homomorphism $K[X]/ker\eta$ is isomorphic to $K[\alpha]$. There are two cases:

(i) $ker\eta = \{0\}$.
(ii) $ker\eta \neq \{0\}$.

In case (i), η is an isomorphism from $K[X]$ to $K[\alpha]$, and we say that α is a **transcendental** element over K. More explicitly, α is a transcendental element over K if α is not a root of any nonzero polynomial $f(X) \in K[X]$. For example, π and the exponential e are transcendental over $\mathbb{Q}$.

In case (ii), there is a nonzero polynomial $f(X) \in K[X]$ such that $f(\alpha) = 0$, and in this case we say that α is an **algebraic** element over K. For example, a primitive cube root $\omega = e^{\frac{2\pi i}{3}}$ of unity, being a root of the nonzero polynomial $X^2 + X + 1 \in \mathbb{Q}[X]$, is algebraic over $\mathbb{Q}$.

Suppose that α is an algebraic element over K. Then $ker\eta \neq \{0\}$. Since $K[X]$ is a P.I.D, $ker\eta$ is a nontrivial principal ideal. Suppose that $ker\eta = < p(X) > = K[X]p(X)$. Let $p'(X)$ be another polynomial in $K[X]$ such that $ker\eta = < p'(X) > = K[X]p'(X)$. Then $p(X)$ divides $p'(X)$, and also $p'(X)$ divides $p(X)$. In turn, $p(X) = ap'(X)$ for some nonzero element $a \in K$. Thus, there is a unique monic polynomial $p(X) \in K[X]$ such that $ker\eta = < p(X) >$. The unique monic polynomial $p(X)$, thus obtained, is called the **minimum polynomial** of α over K, and it is denoted by $min_K(\alpha)(X)$. Clearly, the minimum polynomial $min_K(\alpha)(X)$ of α over K is the monic polynomial of smallest degree having α as a root. Conversely, suppose that $f(X)$ is a monic polynomial of smallest degree having α as a root. Then $min_K(\alpha)(X)$ divides $f(X)$. By the division algorithm, there exist polynomials $q(X)$ and $r(X)$ such that $min_K(\alpha)(X) = q(X)f(X) + r(X)$, where $r(X) = 0$ or else $deg(r(X)) < degf(X)$. Hence $r(\alpha) = 0$. This implies that $r(X) = 0$, for other wise we shall arrive at a contradiction to the supposition that $f(X)$ is the smallest degree polynomial having α as a root. Hence $f(X)$ divides $min_K(\alpha)(X)$. It follows that $f(X) = min_K(\alpha)(X)$.

Proposition 8.1.5 *Let L be a field extension of K, and $\alpha \in L$ be an algebraic element over K. Then the minimum polynomial $min_K(\alpha)(X)$ of α over K is an irreducible polynomial in $K[X]$, and the ideal $< min_K(\alpha)(X) > = K[X]min_K(\alpha)(X)$ is a maximal ideal of $K[X]$.*

Proof Suppose that $min_K(\alpha)(X) = f(X)g(X)$, where $f(X)$ and $g(X)$ are nonconstant polynomials in $K[X]$. Then $degf(X) < degmin_K(\alpha)(X)$, and also $degg(X) < degmin_K(\alpha)(X)$. Further, $0 = min_K(\alpha)(\alpha) = f(\alpha)g(\alpha)$. Since L is a field, $f(\alpha) = 0$ or $g(\alpha) = 0$. But, then $min_K(\alpha)(X)$ divides $f(X)$, or it divides $g(X)$. This is a contradiction. Hence $min_K(\alpha)(X)$ is an irreducible polynomial. Since the ideal generated by an irreducible element in a principal ideal domain is a maximal ideal, $< min_K(\alpha)(X) >$ is a maximal ideal. ♯

Proposition 8.1.6 *Let L be a field extension of K, and $\alpha \in L$. Then α is algebraic if and only if $K[\alpha] = K(\alpha)$. Further, then $[K(\alpha) : K] = degmin_K(\alpha)(X)$.* ♯

Proof Suppose that α is algebraic. The map η from $K[X]$ to $K[\alpha]$ given by $\eta(f(X)) = f(\alpha)$ is a surjective homomorphism. From the fundamental theorem of homomorphism, $K[X]/ker\eta$ is isomorphic to $K[\alpha]$. By the above proposition, $ker\eta = < min_K(\alpha)(X) >$ is a maximal ideal. Hence $K[\alpha]$ is a field, and so

$K[\alpha] = K(\alpha)$. Conversely, suppose that $K[\alpha] = K(\alpha)$. Then $K[\alpha]$ is a subfield of L. Hence $\alpha^{-1} \in K[\alpha]$. Suppose that

$$\alpha^{-1} = a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_n\alpha^n,$$

where at least one a_i is nonzero. Clearly,

$$f(X) = -1 + a_0X + a_1X^2 + \cdots + a_{n+1}X^{n+1}$$

is a nonzero polynomial in $K[X]$ such that $f(\alpha) = 0$. It follows that α is algebraic. Finally, suppose that α is algebraic, and

$$min_K(\alpha)(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n.$$

It is sufficient to show that the set $S = \{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$ is a basis of $K(\alpha)$ over K. Since $K(\alpha) = K[\alpha]$, every element of $K(\alpha)$ is of the form $f(\alpha)$, where $f(X)$ is a polynomial in $K[X]$. By the division algorithm, there exist polynomials $q(X)$ and $r(X)$ such that $f(X) = q(X)min_K(\alpha) + r(X)$, where $r(X) = 0$, or else $degr(X) \leq n-1$. But, then $f(\alpha) = r(\alpha)$. Thus, $f(\alpha) = r(\alpha)$ is linear combination of members of S. This shows that S generates the vector space $K(\alpha)$ over K. Next, suppose that

$$a_0 + a_1\alpha + a_2\alpha^2 + \cdots a_{n-1}\alpha^{n-1} = 0,$$

where $a_i \in K$ for each i. Then each a_i is 0, for otherwise we get a nonzero polynomial

$$g(X) = a_0 + a_1X + a_2X^2 + \cdots a_{n-1}X^{n-1}$$

of degree less than the degree of $min_K(\alpha)(X)$ such that $g(\alpha) = 0$. This shows that S is also linearly independent. ♯

Proposition 8.1.7 *Let L be a field extension of K. An element $\alpha \in L$ is algebraic over K if and only if $K(\alpha)$ is a finite extension of K.*

Proof Suppose that α is algebraic over K. From the above proposition, it follows that $K(\alpha)$ is a finite extension of K. Conversely, suppose that $K(\alpha)$ is a finite extension of degree n over K. Then the dimension of $K(\alpha)$ over K is n. Hence the set $\{1, \alpha, \alpha^2, \ldots, \alpha^n\}$ is linearly dependent. Thus, there exist $a_0, a_1, \ldots, a_n$ not all 0 such that

$$a_0 + a_1\alpha + a_2\alpha^2 + \cdots a_{n-1}\alpha^n = 0.$$

This gives us a nonzero polynomial

$$g(X) = a_0 + a_1x + a_2X^2 + \cdots a_nX^n$$

such that $g(\alpha) = 0$. It follows that α is algebraic over K. ♯

Proposition 8.1.8 *Let L be a field extension of K, and $\alpha_1, \alpha_2, \ldots, \alpha_n$ be elements of L which are algebraic over K. Then $K(\alpha_1, \alpha_2, \ldots, \alpha_n) = K[\alpha_1, \alpha_2, \ldots, \alpha_n]$. In particular, $ker\eta$ is a maximal ideal of $K[X_1, X_2, \ldots, X_n]$.*

Proof The proof is by the induction on n. For $n = 1$, the result is the above proposition. Assume that the result is true for n. Let $\alpha_1, \alpha_2, \ldots, \alpha_n, \alpha_{n+1}$ be elements of L which are algebraic over K. By the induction hypothesis,

$$K[\alpha_1, \alpha_2, \ldots, \alpha_{n+1}] = K[\alpha_1, \alpha_2, \ldots, \alpha_n][\alpha_{n+1}] = K(\alpha_1, \alpha_2, \ldots, \alpha_n)[\alpha_{n+1}].$$

Since α_{n+1} is algebraic over K, it is also algebraic over $K(\alpha_1, \alpha_2, \ldots, \alpha_n)$. From the above proposition,

$$K(\alpha_1, \alpha_2, \ldots, \alpha_n)[\alpha_{n+1}] = K(\alpha_1, \alpha_2, \ldots, \alpha_{n+1}).$$ ♯

Proposition 8.1.9 *Let L be a field extension of K, and F be a subfield of L containing K. Let $\alpha \in L$ be algebraic over K. Then α is algebraic over F, and $min_F(\alpha)(X)$ divides $min_K(\alpha)(X)$ in $F[X]$.*

Proof Since $min_K(\alpha)(X)$ belongs to $K[X]$, it also belongs to $F[X]$. Again, since α is a root of $min_K(\alpha)(X)$, the result follows. ♯

Definition 8.1.10 A field extension L of K is said to be an **algebraic extension** of K if every element of L is algebraic over K.

Evidently, we have the following proposition:

Proposition 8.1.11 *Let L be an algebraic extension of K and F be a subfield of L containing K. Then L is an algebraic extension of F.* ♯

Proposition 8.1.12 *Let L be a finite extension of K. Then L is an algebraic extension of K.*

Proof Let L be a finite extension of K and $\alpha \in L$. The $[K(\alpha) : K] \leq [L : K] < \infty$. Thus from the above proposition α is algebraic over K. ♯

Proposition 8.1.13 *Let L be a field extension of K. Let L_0 be the set of all algebraic elements of L over K. Then L_0 is a subfield of L containing K which is the largest algebraic extension of K contained in L.*

Proof Every element $\alpha \in K$ is algebraic over K, for it is a root of $X - \alpha$. Thus $K \subseteq L_0$. Let $\alpha, \beta \in L_0$. Since α is algebraic over K, $K(\alpha)$ is a finite extension of K. Further since β is algebraic over K, it is also algebraic over $K(\alpha)$. Thus $K(\alpha, \beta) = K(\alpha)(\beta)$ is a finite extension of $K(\alpha)$. It follows that $K(\alpha, \beta)$ is also a finite extension of K. Hence $K(\alpha, \beta)$ is an algebraic extension of K. Hence $\alpha \pm \beta$ and $\alpha \cdot \beta^{-1}$ are also algebraic over K, and so both of them belong to L_0. This shows that L_0 is a subfield. Clearly, this is the largest field contained in L which is algebraic over K. ♯

Definition 8.1.14 The subfield L_0 of L in the above proposition is called the **algebraic closure** of K in L.

Corollary 8.1.15 *Let L be an algebraic extension of F, and F be an algebraic extension of K. Then L is an algebraic extension of K.*

Proof Let $\alpha \in L$. Since L is algebraic over F, α is algebraic over F. Let

$$min_F(\alpha)(X) = X^n + a_1X^{n-1} + \cdots + a_n.$$

Then α is algebraic over $F' = K(a_1, a_2, \ldots a_n)$. Clearly, $F'(\alpha)$ is a finite extension of F', and $[F'(\alpha) : F'] = n$. Further, since a_1 is algebraic over K, $K(a_1)$ is a finite extension of K. Again, since a_2 is algebraic over K, and so also over $K(a_1)$, it follows that $K(a_1)(a_2) = K(a_1, a_2)$ is a finite extension of $K(a_1)$. In turn, it follows that $K(a_1, a_2)$ is a finite extension of K. Proceeding inductively, we find that F' is a finite extension of K. But, then $F'(\alpha)$ is also a finite extension of K. Hence every element of $F'(\alpha)$ is algebraic over K. In particular α is algebraic over K. ♯

Definition 8.1.16 An extension L of K is called a **simple** extension if there is an element $\alpha \in L$ such that $L = K(\alpha)$. Such an element α is called a **primitive** element of the extension.

Theorem 8.1.17 *Let L be a finite extension of K. Then L is simple over K if and only if there are only finitely many intermediary field between L and K.*

Proof Suppose that $L = K(\alpha)$ is a finite simple extension of K. Let F be a subfield of L containing K. Then α is algebraic over K and also over F. Clearly, $min_F(\alpha)(X)$ is a divisor of $min_K(\alpha)(X)$. We show that F is uniquely determined by the factor $min_F(\alpha)(X)$ of $min_K(\alpha)(X)$. Suppose that

$$min_F(\alpha)(X) = a_0 + a_1X + a_2X^2 + \cdots + a_{n-1}X^{n-1} + X^n,$$

where $a_i \in F$. Consider the subfield $K' = K(a_0, a_1, \ldots a_{n-1})$ of F. Then $min_{K'}(\alpha)(X) = min_F(\alpha)(X)$, and so $[L : F] = [L : K']$. Hence $F = K'$. This shows that F is uniquely determined by $min_F(\alpha)(X)$. Since the number of monic polynomial divisors of $min_K(\alpha)(X)$ are finitely many, we have only finitely many intermediary fields between L and K.

Conversely, suppose that there only finitely many intermediary fields. Then we have to show that L is a simple extension of K. Suppose first that K is a finite field,

and L is a finite extension of K. Then L is also a finite field. Further, we know that the multiplicative group $L^\star$ of nonzero elements of L is cyclic. Suppose that it is generated by α. Then it is clear that $L = K(\alpha)$.

Assume, now, that K is infinite. Since $[L : K]$ is finite, L is finitely generated extension of K. Suppose that $L = K(\alpha_1, \alpha_2, \ldots, \alpha_n)$. The proof is by the induction on n. If $n = 1$, then there is nothing to do. Suppose that the result is true for n. Suppose that $L = K(\alpha_1, \alpha_2, \ldots, \alpha_n, \alpha_{n+1})$. Then, by the induction hypothesis $K(\alpha_1, \alpha_2, \ldots, \alpha_n) = K(\alpha)$ for some $\alpha \in L$. Clearly, $L = K(\alpha, \alpha_{n+1})$. Consider the set $\{K(a\alpha + \alpha_{n+1}) \mid a \in K\}$. Since the there are only finitely many intermediary fields between K and L, and K is infinite, there are distinct elements $a, b \in K$ such that $K(a\alpha + \alpha_{n+1}) = K(b\alpha + \alpha_{n+1}) = F$ (say). But, then $\alpha = ((a\alpha + \alpha_{n+1}) - (b\alpha + \alpha_{n+1}))(a - b)^{-1}$ belongs to $K(a\alpha + \alpha_{n+1})$. This also shows that $\alpha_{n+1} \in K(a\alpha + \alpha_{n+1})$. Hence $L = K(\alpha, \alpha_{n+1}) = K(a\alpha + \alpha_{n+1})$. The proof is complete. ♯

Now, we have some examples.

Example 8.1.18 The field $\mathbb{C}$ of complex numbers is an extension field of $\mathbb{R}$. Since $\{1, i\}$ is a basis of the vector space $\mathbb{C}$ over $\mathbb{R}$, we have $[\mathbb{C} : \mathbb{R}] = 2$. The field $\mathbb{R}$ of real numbers is an extension of the field $\mathbb{Q}$ of rational numbers. The dimension of $\mathbb{R}$ considered as a vector space over $\mathbb{Q}$ is infinite(since $\mathbb{Q}^n$ is countable, any finite-dimensional vector space over $\mathbb{Q}$ is countable, but $\mathbb{R}$ is uncountable). $\sqrt{2}$ is an element of $\mathbb{R}$ which is algebraic over $\mathbb{Q}$, for it is a root of the polynomial $X^2 - 2$. Also, since $X^2 - 2$ is the smallest degree monic polynomial over $\mathbb{Q}$ of which $\sqrt{2}$ is a root, $min_Q(\sqrt{2})(X) = X^2 - 2$. Further $\mathbb{Q}(\sqrt{2}) = \mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$. Note that $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$.

Example 8.1.19 Let ω be a primitive cube root of unity. Then ω is algebraic over $\mathbb{Q}$, for it is a root of the equation $X^3 - 1$. Further, since $\omega \notin Q$, and it is also a root of $X^2 + X + 1$, it follows that $X^2 + X + 2$ is the minimum polynomial of ω over $\mathbb{Q}$. Clearly, $\mathbb{Q}(\omega) = \mathbb{Q}[\omega] = \{a + b\omega \mid a, b \in \mathbb{Q}\}$, and $[\mathbb{Q}(\omega) : \mathbb{Q}] = 2$.

Example 8.1.20 $\sqrt[3]{2}$ and ω are algebraic over $\mathbb{Q}$. Hence $\mathbb{Q}[\sqrt[3]{2}, \omega] = \mathbb{Q}(\sqrt[3]{2}, \omega)$ is a finite extension of $\mathbb{Q}$. We show that it is a simple extension of $\mathbb{Q}$(in fact we shall show later that every finite extension of a field of characteristic 0 is simple). We find an element $\alpha \in \mathbb{C}$ such that $\mathbb{Q}(\alpha) = \mathbb{Q}(\sqrt[3]{2}, \omega)$. Observe that $\sqrt[3]{2} + \omega$ belongs to $\mathbb{Q}(\sqrt[3]{2}, \omega)$. Put $\alpha = \sqrt[3]{2} + \omega$. We show that $\mathbb{Q}(\alpha) = \mathbb{Q}(\sqrt[3]{2}, \omega)$. We have $(\alpha - \omega)^3 = 2$. Using the binomial expansion and the fact that $\omega^2 = -1 - \omega$, we find that

$$\omega = (\alpha^3 - 3\alpha - 3)(3\alpha^2 + 3\alpha)^{-1}.$$

Note that $\alpha^2 + \alpha \neq 0$. Thus, $\omega \in \mathbb{Q}(\alpha)$, and so $\sqrt[3]{2}$ also belongs to $\mathbb{Q}(\alpha)$. This shows that $\mathbb{Q}(\alpha) = \mathbb{Q}(\sqrt[3]{2}, \omega)$. Next, we find the minimum polynomial of α, and also the degree $[\mathbb{Q}(\alpha) : Q]$. Note that $\mathbb{Q}(\alpha)$ is an extension of $\mathbb{Q}(\sqrt[3]{2})$, and it is also an extension of $\mathbb{Q}(\omega)$. Since $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3(X^3 - 2$ is the minimum polynomial of $\sqrt[3]{2})$ and $[\mathbb{Q}(\omega) : \mathbb{Q}] = 2$, it follows that 2 and 3 both divide $[\mathbb{Q}(\alpha) : Q]$. Thus 6 divides $[\mathbb{Q}(\alpha) : \mathbb{Q}]$. Next,

$$\alpha^3 - 3\alpha - 3 = \omega(3\alpha^2 + 3\alpha).$$

Putting the value of ω from the above equation in the equation $\omega^2 + \omega + 1 = 0$, we get that

$$(\alpha^3 - 3\alpha - 3)^2 + (3\alpha^2 + 3\alpha)^2 + (3\alpha^2 + 3\alpha)(\alpha^3 - 3\alpha - 3) = 0.$$

This gives us a six degree monic polynomial

$$(X^3 - 3X - 3)^2 + (3X^2 + 3X)^2 + (3X^2 + 3X)(X^3 - 3X - 3) = 0.$$

of which α is a root. Since $[\mathbb{Q}(\alpha) : \mathbb{Q}]$ is at least 6, it follows that this is the minimum polynomial of α, and $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 6$.

Example 8.1.21 Let p and q be distinct prime numbers. Then $\sqrt{p}$ and $\sqrt{q}$ are algebraic over $\mathbb{Q}$ with minimum polynomials $X^2 - p$ and $X^2 - q$ respectively. Clearly, $\mathbb{Q}(\sqrt{p}) = \{a + b\sqrt{p} \mid a, b \in \mathbb{Q}\}$, and $[\mathbb{Q}(\sqrt{p}) : \mathbb{Q}] = 2$. Also $\sqrt{q} \notin \mathbb{Q}(\sqrt{p})$, and so $\sqrt{q}$ is algebraic over $\mathbb{Q}(\sqrt{p})$ with minimum polynomial $X^2 - q$. Thus, $[\mathbb{Q}(\sqrt{p}, \sqrt{q}) : \mathbb{Q}] = [\mathbb{Q}(\sqrt{p}, \sqrt{q}) : \mathbb{Q}(\sqrt{p})][\mathbb{Q}(\sqrt{p}) : \mathbb{Q}] = 4$. We show that $\mathbb{Q}(\sqrt{p}, \sqrt{q}) = \mathbb{Q}(\sqrt{p} + \sqrt{q})$. Note that $\mathbb{Q}(\sqrt{p} + \sqrt{q}) \subseteq \mathbb{Q}(\sqrt{p}, \sqrt{q})$. Put $\alpha = \sqrt{p} + \sqrt{q}$. Then

$$(\alpha - \sqrt{p})^2 = \alpha^2 + p - 2\alpha\sqrt{p} = q.$$

This shows that $\alpha^2 + p - 2\alpha\sqrt{p}$ belongs to $\mathbb{Q}(\alpha)$. Hence $\sqrt{p} \in Q(\alpha)$. Similarly, $\sqrt{q} \in \mathbb{Q}(\alpha)$. This shows that $\mathbb{Q}(\alpha) = \mathbb{Q}(\sqrt{p}, \sqrt{q})$. Next,

$$(\alpha^2 + p - q)^2 - 4p\alpha^2 = 0.$$

This shows that α satisfies a monic polynomial of degree 4 which is the minimum polynomial of α.

Example 8.1.22 Let p be a positive prime integer. By the Eisenstein irreducibility criteria, $X^n - p$ is irreducible in $\mathbb{Q}[X]$. Hence $\sqrt[n]{p}$ is an algebraic element over Q of which $X^n - p$ is the minimum polynomial. It follows that $[\mathbb{Q}(\sqrt[n]{p}) : \mathbb{Q}] = n$.

Example 8.1.23 π and the exponential e are transcendental (not algebraic). It was Hermite who proved the transcendence of e in 1873. Earlier, the irrationality of e was established by Liouville. The transcendence of π was established by Lindemann. The proof of this fact can be found in Algebra by S. Lang. You may also refer to the corollary following the theorem 8.7.5. It is not known if π is transcendental over $\mathbb{Q}(e)$, or equivalently, it is not known if e is transcendental over $\mathbb{Q}(\pi)$.

Proposition 8.1.24 *Let $K(X)$ be the field of fractions of the polynomial ring $K[X]$ (This is also called the* **function field** *over K in one variable). Let $T \in K(X) - K$. Then $K(X)$ is an algebraic extension of $K(T)$, and T is transcendental over K.*

Further, if $T = \frac{f(X)}{g(X)}$, *where* $f(X)$ *and* $g(X)$ *are co-prime in* $K[X]$, *then* $h(Y) = Tg(Y) - f(Y) \in K(X)[Y]$ *is a minimum degree polynomial over* $K(T)$ *of which* X *is a root. In turn,* $[K(X) : K(T)] = max(degf(X), g(X))$.

Proof Put $L = K(T)$, where $T = \frac{f(X)}{g(X)}$. We find the minimum degree polynomial in $K(T)[Y]$ of which X is a root. Consider the polynomial $h(Y) = Tg(Y) - f(Y)$ in $K(T)[Y]$. We first observe that $h(Y) \neq 0$ and $degh(Y) = max(degf(X), degg(X))$. If not, then the leading term of $Tg(Y)$ and $f(Y)$ should be same. Now, the leading coefficient of $Tg(Y)$ is Ta for some $a \in K$, and the leading coefficient of $f(Y)$ is b for some $b \in K$. This would mean that $T = ba^{-1} \in K$, a contradiction to the supposition that $T \notin K$. Clearly, X is a root of $h(Y)$, and so X is algebraic over $K(T)$. This also shows that $K(X)$ is algebraic extension of $K(T)$. It is sufficient, therefore, to show that $h(Y)$ is irreducible in $K(T)[Y]$. We first observe that T is transcendental over K (in other words every element of $K(X) - K$ is transcendental over K). For if not, then T would be algebraic over K. This says that $K(T)$ is algebraic extension of K. Since $K(X)$ is already algebraic extension of $K(T)$, this would mean that $K(X)$ is algebraic extension of K, a contradiction to the fact that X is transcendental over K. Thus, $K[T]$ is isomorphic to the polynomial ring over K. Since $f(X)$ and $g(X)$ are co-prime in $K[X]$, $h(Y) = Tg(Y) - f(Y)$ is a primitive polynomial of degree 1 in $K[Y][T]$. Hence it is irreducible in $K[Y][T] = K[T][Y]$, and so also it is irreducible in $K(T)[Y]$. The result follows. ♯

Remark 8.1.25 Luroth proved in 1876 a fundamental result in the theory of algebraic function fields that any subfield F of $L = K(X)$ containing K is again of the form $K(T)$, where, of course, T is transcendental over K. When K is algebraically closed field in the sense that there is no proper algebraic extension of K, then a result of Castelnuovo says that every subfield of $K(X, Y)$ containing K is of the form $K(T)$, or it is of the form $K(T, S)$. The result of this kind is not true for algebraic function field in 3 variables.

Example 8.1.26 Let $L = K(X)$ be the field of rational functions over K in one variable. We determine the Galois group of $K(X)$ over K (the group of automorphisms of $K(X)$ fixing the members of K). Let $T \in K(X)$ be such that $K(X) = K(T)$. Then $T \notin K$. Suppose that $T = \frac{f(X)}{g(X)}$, where $f(X)$ and $g(X)$ are co-prime in $K[X]$. Then $[K(X) : K(T)] = 1$, and so from the previous proposition, $max(degf(X), g(X)) = 1$. Thus, $T = \frac{aX+b}{cX+d}$ for some $a, b, c, d \in K$. Interchanging the role of X and T we observe that $X = \frac{uT+v}{pT+q}$ for some $u, v, p, q \in K$. Thus

$$X(p(\frac{aX+b}{cX+d}) + q) = u(\frac{aX+b}{cX+d}) + v,$$

or equivalently,

$$p(aX^2 + bX) + q(cX^2 + dX) = uaX + ub + vcX + vd.$$

Thus, comparing the coefficient of same powers of X, we obtain that $pa + qc = 0 = ub + vd$, and $pb + qd = ua + vc \neq 0$. This shows that the matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is a nonsingular 2×2 matrix, and so it belongs to the group $GL(2, K)$. Conversely, given such a matrix in $GL(2, K)$, we can solve X in terms of $T = \frac{aX+b}{cX+d}$, and so $K(X) = K(T)$. It follows that any element

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

of $GL(2, K)$ determines an element of $G(K(X)/K)$ which takes X to $\frac{aX+b}{cX+d}$. This defines a surjective homomorphism η (say) from $GL(2, K)$ to $G(K(X)/K)$. Since the $T = X$ if and only if $a = d$ and $b = 0 = c$, the kernel of η is the normal subgroup of $GL(2, K)$ consisting of the scalar matrices. The subgroup of the scalar matrices is precisely the center of $GL(2, K)$. The quotient group of $GL(2, K)$ modulo its center is called the projective general linear group, and it is denoted by $PGL(2, K)$. Thus, the group $G(K(X)/K)$ of K—automorphisms of $K(X)$ is isomorphic to $PGL(2, K)$.

Exercise

8.1.1 Let K be a field, and $K((X)) = \{\Sigma_{n=m}^{\infty} a_n X^n \mid a_n \in K, m \in \mathbb{Z}\}$ be the set of formal power series with coefficients in K. Define $+$ by

$$\sum_{n}^{\infty} a_n X^n + \sum_{n}^{\infty} b_n X^n = \sum_{n}^{\infty} (a_n + b_n) X^n.$$

and the multiplication $\cdot$ by

$$(\sum_{n=m}^{\infty} a_n X^n) \cdot (\sum_{n=r}^{\infty} b_n X^n) = \sum_{n=m+r}^{\infty} (\sum_{l=m}^{n-r} a_n b_{n-l}) X^n.$$

Show that $K((X))$ is a field with respect to the above operation, and it is a field extension of $K(X)$.

8.1.2 Give an example of an algebraic extension which is not a finite extension.

8.1.3 Show that $\mathbb{Q}(\sqrt[4]{5}, \sqrt{7})$ is a finite extension of $\mathbb{Q}$. Find its degree. Find also a primitive element of the extension together with the minimum polynomial of that primitive element.

8.1.4 Find the minimum polynomial of a primitive 11th root of unity over $\mathbb{Q}$.

8.1.5 Let L be a field extension of K. Let $\alpha_1, \alpha_2, \ldots, \alpha_r$ be elements of L which are algebraic over K. Show that

$$[K(\alpha_1, \alpha_2, \ldots, \alpha_r) : K] \leq \prod_{i=1}^{r} [K(\alpha_i) : K].$$

Show, by means of an example, that the strict inequality may hold. Find a sufficient condition for the equality.

8.1.6 Show that the fields $\mathbb{Q}(\sqrt{p})$ and $\mathbb{Q}(\sqrt{q})$, where p and q are distinct primes, are not isomorphic whereas they are isomorphic as vector spaces over Q.

8.1.7 Let L be an algebraic extension of K. Show that any subring of L which contains K is a field.

8.1.8 Let L be a finite field extension of K. Let F_1 and F_2 be intermediary fields. The smallest subfield of L containing F_1 and F_2 is called the **composite** of F_1 and F_2, and it is denoted by F_1F_2. Show that

$$[F_1F_2 : K] \leq [F_1 : K][F_2; K].$$

Further, show that the equality holds provided that $[F_1 : K]$ and $[F_2 : K]$ are co-prime. Give an example to show that the strict inequality may hold.

8.1.9 Find the degree of $\mathbb{Q}(\sqrt[5]{2}, \sqrt{7})$ over $\mathbb{Q}$, and also find a primitive element for the extension.

8.1.10 Let $\overline{\mathbb{Q}}$ be the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$. Show that $[\overline{\mathbb{Q}} : \mathbb{Q}]$ is infinite.

8.1.11 Let m be co-prime to $[K(\alpha) : K]$. Show that $K(\alpha) = K(\alpha^m)$.

8.1.12 Show that the composite of two field extensions F_1 and F_2 is algebraic if and only if both are algebraic.

8.1.13 Show, by means of an example, that $[L : K] = 3$ need not mean that $L = K(\sqrt[3]{\alpha})$ for some α.

8.1.14 Show that $Sinm^0$ is algebraic for all rational m.

8.2 Galois Extensions

To each polynomial $f(X) \in K[X]$, Galois attached a group, called the **Galois group** of the polynomial $f(X)$. It is essentially a group of permutations of roots of $f(X)$ in certain extension field L of K over which $f(X)$ splits into linear factors. He showed that the polynomial equation $f(X) = 0$ can be solved using field and radical operations if and only if the Galois group of $f(X)$ is a solvable group. Here, we follow a slightly different but equivalent approach due to Artin in which we proceed with the Galois group of an extension.

Let L and L' be two extensions of K. A ring homomorphism f from L to L' such that $f(a) = a$ *for all* $a \in K$ is called a **K-homomorphism**. Clearly, a K-homomorphism f from L to L' is also a vector space homomorphism from L to L' considered as vector spaces over K, for $f(\alpha a) = f(\alpha)f(a) = \alpha f(a)$ *for all* $\alpha \in K$, *and* $a \in L$. Since L is a field $ker f = \{0\}$, or $ker f = L$. Since f takes identity to identity, $ker f = \{0\}$. This shows that f is injective. If f is also a bijection, then f is called a **K-isomorphism**. Thus, if $[L : K] = [L' : K] < \infty$, then f is an isomorphism (an injective linear transformation between vector spaces of same dimensions is an isomorphism). In particular, if L is a finite field extension of K, then any K-homomorphism from L to L is an isomorphism. This is called a **K-automorphism** of L.

Definition 8.2.1 Let L be a field extension of K. Then the set of all K - automorphisms of L form a group under composition of maps. This group is called the **Galois group** of the extension L of K, and it is denoted by $G(L/K)$.

Definition 8.2.2 Let L be a field, and X be a subset of the group $Aut(L)$. Let $F(X)$ denote the set of all elements $a \in L$ such that $\sigma(a) = a$ *for all* $\sigma \in X$. Then, $F(X)$ is a subfield of L, and it is called the **fixed field** of X. It is clear that $F(X) = F(< X >)$.

Observe that if $X \subseteq G(L/K)$, then $F(X)$ is an intermediary field of the field extension L of K.

Let $S(G(L/K))$ denote the set of all subgroups of the Galois group $G(L/K)$ of the field extension L of K, and $SF(L/K)$ denote the set of all intermediary fields. Then, we have a map Φ from $S(G(L/K))$ to $SF(L/K)$ defined by $\Phi(H) = F(H)$ (the fixed field of H), and a map Ψ from $SF(L/K)$ to $S(G(L/K))$ defined by $\Psi(F) = G(L/F)$.

The aim of this and the following two sections will be to show that in case L is a finite extension of K, these two maps are inverses to each other if and only if $K = F(G(L/K))$.

Definition 8.2.3 Let L be an algebraic extension of K. We say that L is a **Galois extension** of K if $K = F(G(L/K))$.

Proposition 8.2.4 *Let L be a field extension of K. Then,*

1. $K_1 \subseteq K_2, K_1, K_2 \in SF(L/K) \Longrightarrow G(L/K_2) \subseteq G(L/K_1)$.
2. $H_1 \subseteq H_2, H_1, H_2 \in S(G(L/K)) \Longrightarrow F(H_2) \subseteq F(H_1)$.
3. $S \subseteq G(L/K) \Longrightarrow F(S) = F(< S >)$.
4. $K' \in SF(L/K) \Longrightarrow K' \subseteq F(G(L/K'))$.
5. $S \subseteq G(L/K) \Longrightarrow < S > \subseteq G(L/F(< S >))$.
6. $F(H) = F(G(L/F(H)))$ *for all* $H \in S(G(L/K))$.
7. $G(L/K') = G(L/F(G(L/K')))$ *for all* $K' \in SF(L/K)$.

Proof 1 and 2 follow from the definitions. Since $S \subseteq < S >$, if a is fixed by every element of $< S >$, then it is also fixed by every element of S. Hence $F(< S >) \subseteq F(S)$. Further, suppose that a is fixed by every element of S. Then it is fixed by every power of elements of S, and so also by the products of powers of elements of S. Thus, $F(S) \subseteq F(< S >)$. This proves 3. 4 and 5 also follow from the definitions.

Now, we prove 6. Let $a \in F(H)$. If $\sigma \in G(L/F(H))$, then by the definition $\sigma(a) = a$. This shows that $a \in F(G(L/F(H)))$. Thus, $F(H) \subseteq F(G(L/F(H)))$. Next, by 5, $H \subseteq G(L/F(H))$. By 2, $F(G(L/F(H))) \subseteq F(H)$. This completes the proof of 6.

Finally, we prove 7. From 4, it follows that $K' \subset F(G(L/K'))$, and so by 1, $G(L/F(G(L/K'))) \subseteq G(L/K')$. Let $\sigma \in G(L/K')$. Then by the definition σ will fix every member of $F(G(L/K')$, and hence it belongs to $G(L/F(G(L/K')))$. This completes the proof of 7. ♯

Example 8.2.5 $G(\mathbb{C}/\mathbb{R}) = \{I_{\mathbb{C}}, \sigma\}$, where σ denotes the complex conjugation of $\mathbb{C}$: Let σ be a nonidentity $\mathbb{R}$-automorphism of $\mathbb{C}$. Let $a + ib \in \mathbb{C}$. Then $\sigma(a + ib) = \sigma(a) + \sigma(i)\sigma(b) = a + \sigma(i)b$. Since σ is an automorphism $-1 = \sigma(-1) = \sigma(i^2) = \sigma(i)\sigma(i) = (\sigma(i))^2$. This shows that $\sigma(i) = \pm i$. Since σ is nonidentity, $\sigma(i) = -i$. Thus, $\sigma(a + ib) = a - ib = \overline{a + ib}$. It follows that $\mathbb{R} = F(G(\mathbb{C}/\mathbb{R}))$, and so the extension $\mathbb{C}$ of $\mathbb{R}$ is a Galois extension.

Example 8.2.6 As in the above example it can be seen that $G(\mathbb{Q}(\sqrt{2})/\mathbb{Q}) = \{I_{Q(\sqrt{2})}, \sigma\}$, where $\sigma(a + b\sqrt{2}) = a - b\sqrt{2}$. It follows that the extension $Q(\sqrt{2})$ is also a Galois extension of $\mathbb{Q}$.

Example 8.2.7 Consider the extension $\mathbb{Q}(\sqrt[3]{5})$ of $\mathbb{Q}$, where $\sqrt[3]{5}$ is the real cube root of 5. It is not a Galois extension: We first show that $G(\mathbb{Q}(\sqrt[3]{5})/\mathbb{Q}) = \{I_{\mathbb{Q}(\sqrt[3]{5})}\}$. Clearly, $\mathbb{Q}(\sqrt[3]{5}) = \{a + b5^{\frac{1}{3}} + c5^{\frac{2}{3}} \mid a, b, c \in \mathbb{Q}\}$. The other cube roots of 5 are not in $\mathbb{Q}(\sqrt[3]{5})$, for they are not real numbers. If σ is an automorphism of $\mathbb{Q}(\sqrt[3]{5})$, then $5 = \sigma(5) = (\sigma(5^{\frac{1}{3}})^3)$. Thus, $\sigma(5^{\frac{1}{3}}) = 5^{\frac{1}{3}}$. Hence σ is the identity map. Thus, the Galois group of the extension is trivial, and $F(G(\mathbb{Q}\sqrt[3]{5})) \neq Q$. Hence it is not a Galois extension.

Since any field automorphism of the field $\mathbb{R}$ of real numbers fixes $\mathbb{Q}$, and since it also preserves order, it is the trivial identity automorphism. Thus, the Galois group of $\mathbb{R}$ over $\mathbb{Q}$ is trivial. In turn, it follows that $\mathbb{R}$ is not a Galois extension of $\mathbb{Q}$.

Proposition 8.2.8 *Let L_1 and L_2 be a field extensions of K. Let σ be a K-homomorphism from L_1 to L_2. Let α be an element of L_1 which is algebraic over K. Then $\sigma(\alpha)$ is also algebraic over K, and $min_K(\alpha)(X) = min_K(\sigma(\alpha)(X)$. Further, σ permutes the roots of $min_K(\alpha)(X)$ if $L_1 = L_2$.*

Proof Since α is algebraic over K, there exists a nonzero polynomial $f(X) \in K[X]$ such that $f(\alpha) = 0$. Suppose that

$$f(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n,$$

where each $a_i \in K$. Then, since σ is a K-homomorphism,

$$0 = \sigma(0) = \sigma(f(\alpha)) = \Sigma_{i=0}^{n}\sigma(a_i)\sigma(\alpha)^i = \Sigma_{i-0}^{n}a_i\sigma(\alpha)^i = f(\sigma(\alpha)).$$

Thus, $\sigma(\alpha)$ is also algebraic. Further, $min_K(\alpha)(\sigma(\alpha)) = 0$, and so $min_K(\sigma(\alpha))(X)$ divides $min_K(\alpha)(X)$, and since the later is irreducible, $min_K(\alpha)(X) = min_K(\sigma(\alpha))(X)$. ♯

Definition 8.2.9 Let L be a field extension of K. We say that an element $\alpha \in L$ is K-conjugate to an element $\beta \in L$ if there exists an element $\sigma \in G(L/K)$ such that $\sigma(\alpha) = \beta$.

Corollary 8.2.10 *If α is an algebraic element of L over K, then there are at most $deg(min_K(\alpha)(X))$ conjugates of α in L over K.*

Proof It is clear from the above result that if β is conjugate to α, then β is a root of $min_K(\alpha)(X)$, and $min_K(\alpha)(X) = min_K(\beta)(X)$. ♯

Example 8.2.11 The number of conjugates to an algebraic element may be strictly less than $deg(min_K(\alpha)(X))$: Consider the field $\mathbb{Z}_p(X)$ of rational functions in one variable over the prime field $\mathbb{Z}_p$. Let

$$L = \mathbb{Z}_p(X)(X^{\frac{1}{p}}) = \mathbb{Z}_p(X)[X^{\frac{1}{p}}] = \{a_0(X) + a_1(X)X^{\frac{1}{p}} + \cdots + a_r(X)X^{\frac{r}{p}} \mid a_i(X) \in \mathbb{Z}_p(X),\ 0 \leq r \leq p-1\}.$$

Then it can be easily seen that L is a field with respect to the usual addition and multiplication of polynomials, and it is a field extension of $\mathbb{Z}_p(X)$. The element $X^{\frac{1}{p}}$ is algebraic over $\mathbb{Z}_p(X)$, for it is a root of $Y^p - X$ in $\mathbb{Z}_p(X)[Y]$. By the Eisenstein irreducibility criteria $Y^p - X$ is irreducible in $\mathbb{Z}_p(X)[Y]$, and so it is the minimum polynomial of $X^{\frac{1}{p}}$. Since $Y^p - X = (Y - X^{\frac{1}{p}})^p$, it follows that $X^{\frac{1}{p}}$ is self conjugate, and no other element is conjugate to it.

Proposition 8.2.12 *Let L be a field extension of K. Let α and β be elements of L which are algebraic over K. Suppose that they are K-conjugate. Then there is a K-isomorphism from $K(\alpha)$ to $K(\beta)$ which takes α to β.*

Proof It follows from the results above that if α and β are conjugates, then $min_K(\alpha)(X) = \min_K(\beta)(X)$. The map $f(X) \rightsquigarrow f(\alpha)$ defines a surjective homomorphism from $K[X]$ to $K[\alpha] = K(\alpha)$ whose kernel is the ideal $< min_K(\alpha)(X) >$. Thus, by the fundamental theorem of homomorphism, we have an isomorphism σ from $K[X]/ < min_K(\alpha)(X) >$ to $K(\alpha)$ such that $\sigma(f(X) + < \min_K(\alpha)(X) >) = f(\alpha)$. Clearly, $\sigma(a + < min_K(\alpha)(X)) = a$ *for all* $a \in K$ and $\sigma(X + < min_K(\alpha)(X)) = \alpha$. Similarly, we have an isomorphism τ from $K[X]/ < min_K(\beta)(X) > = K[X]/ < min_K(\alpha)(X) >$ to $K(\beta)$ such that $\tau(f(X) + < min_K(\alpha)(X)) = f(\beta)$. Clearly $\tau o \sigma^{-1}$ is a K-isomorphism from $K(\alpha)$ to $K(\beta)$ which takes α to β. ♯

Corollary 8.2.13 *Let α be an algebraic element of L over K. Then there is a bijective map η from $G(K(\alpha)/K)$ to the set of roots of $min_K(\alpha)(X)$ defined by $\eta(\sigma) = \sigma(\alpha)$. In particualr, $\mid G(K(\alpha)/K) \mid$ is the number of distinct roots of $min_K(\alpha)(X)$ which are in $K(\alpha)$.*

Proof If σ is a K-automorphism of $K(\alpha)$, then it is uniquely determined by $\sigma(\alpha)$ which has the same minimum polynomial as α. Conversely if β is a root of the minimum polynomial of α which belongs to $K(\alpha)$, then it follows from the above proposition that there is a K-isomorphism from $K(\alpha)$ to $K(\beta)$. But, since $\beta \in K(\alpha)$, and $[K(\alpha) : K] = deg(min_K(\alpha)(X)) = deg(min_K(\beta)(X)) = [K(\beta) : K]$ it follows that $K(\beta) = K(\alpha)$. The result follows. ♯

Example 8.2.14 In this example, we calculate the Galois group of $\mathbb{Q}(2^{\frac{1}{3}}, \omega) = \mathbb{Q}(2^{\frac{1}{3}} + \omega)$ over $\mathbb{Q}$: Looking at the minimum polynomial of $2^{\frac{1}{3}} + \omega$ over $\mathbb{Q}$, which we have already found in Example 8.1.20, we see that the other 5 roots of the minimum polynomial of $2^{\frac{1}{3}} + \omega$ are $2^{\frac{1}{3}} + \omega^2$, $2^{\frac{1}{3}}\omega + \omega$, $2^{\frac{1}{3}}\omega^2 + \omega^2$, $\omega 2^{\frac{1}{3}} + \omega^2$, *and* $\omega^2 2^{\frac{1}{3}} + \omega$. The Galois group of this extension is, therefore, of order 6. Denote these elements by $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6$. Let σ_i be the automorphism which takes α_1 to α_i. Then, σ_1 is the identity automorphism. It can be checked that $\sigma_2^2, \sigma_3^3, \sigma_4^2, \sigma_5^2, \sigma_6^3$ are all identity maps. Thus, the Galois group of this extension is the symmetric group S_3 of degree 3.

Proposition 8.2.15 *Let L be a finite field extension of K. Then the Galois group $G(L/K)$ is finite.*

Proof Let $\{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ be a basis of L over K. Clearly, $L = K(\alpha_1, \alpha_2, \cdots, \alpha_r)$, and any K-automorphism of L is uniquely determined by its effect on $\alpha_1, \alpha_2, \ldots, \alpha_r$. Since any K-automorphism of L will take an element of L to one of its conjugates, and since there are only finitely many conjugates to an algebraic element α (at most $deg(min_K(\alpha)(X))$ conjugates of α are there), it follows that a K-automorphism of L can take finitely many values on each α_i. Hence $G(L/K)$ is finite. ♯

Definition 8.2.16 Let K be a field and G a group. A group homomorphism from G to $K^\star$ (the multiplicative group of nonzero members of K) is called a **character** of G in K. Thus, a character of G over K is just one dimensional representation of G over K.

Consider the set $F(G, K)$ of all maps from G to K. This is a vector space over K with respect to the point wise addition and multiplication by scalars. The dimension of this space is the cardinality $\mid G \mid$ of G (verify). Let $Ch(G, K)$ denote the set of all characters of G in K. Then $Ch(G, K) \subseteq F(G, K)$. We have the following result due to Dedekind.

Theorem 8.2.17 (Dedekind) *$Ch(G, K)$ is linearly independent subset of the vector space $F(G, K)$.*

Proof Suppose the contrary. Then there is a finite subset of $Ch(G, K)$ which is linearly dependent. Let $S = \{\sigma_1, \sigma_2, \ldots, \sigma_n \mid \sigma_i \neq \sigma_j \, for \, i \neq j\}$ be a minimal finite linearly dependent subset of $Ch(G, K)$. Then there exist $\alpha_1, \alpha_2, \ldots, \alpha_n$ in K not all zero such that

$$\alpha_1\sigma_1 + \alpha_2\sigma_2 + \cdots + \alpha_n\sigma_n = 0 \cdots\cdots, \tag{8.2.1}$$

where 0 in the RHS is the zero of $F(G, K)$. This means that

$$\alpha_1\sigma_1(g) + \alpha_2\sigma_2(g) + \cdots + \alpha_n\sigma_n(g) = 0 \cdots\cdots \quad (8.2.2)$$

for all $g \in G$. Indeed, all α_i are nonzero, for otherwise we shall get a proper subset of S which is linearly dependent, a contradiction to the minimality of S. Since $\sigma_1 \neq \sigma_2$ *there exists* $h \in G$ such that $\sigma_1(h) \neq \sigma_2(h)$. Multiplying the Eq. (8.2.2) by $\sigma_1(h)$ we get that

$$\Sigma_{i=1}^{n}\sigma_1(h)\alpha_i\sigma_i(g) = 0 \cdots\cdots \quad (8.2.3)$$

for all $g \in G$. Further substituting hg at the place of g in the Eq. (8.2.2), we get

$$\Sigma_{i=1}^{n}\alpha_i\sigma_i(hg) = 0$$

for all $g \in G$. Since each σ_i is a homomorphism, we have

$$\Sigma_{i=1}^{n}\alpha_i\sigma_i(h)\sigma_i(g) = 0 \cdots\cdots \quad (8.2.4)$$

for all $g \in G$. Subtracting the Eq. 8.2.4 from the Eq. 8.2.3, we get that

$$\Sigma_{i=2}^{n}(\sigma_1(h) - \sigma_i(h))\alpha_i\sigma_i(g) = 0$$

for all $g \in G$. Put $b_i = (\sigma_1(h) - \sigma_i(h))\alpha_i,\ i \geq 2$. Then, since $\sigma_1(h) \neq \sigma_2(h)$ and each $\alpha_i \neq 0$, it follows that $b_2 \neq 0$. Also

$$b_2\sigma_2(g) + b_3\sigma_3(g) + \cdots + b_n\sigma_n(g) = 0$$

for all $g \in G$, where $b_2 \neq 0$. This means that

$$b_2\sigma_2 + b_3\sigma_3 + \cdots + b_n\sigma_n = 0,$$

where 0 in the RHS is the zero of $F(G, K)$. This is a contradiction to the minimality of S. ♯

Corollary 8.2.18 *Let L be a finite field extension of K. Then* $| G(L/K) | \leq [L : K]$.

Proof We have already seen that under the hypothesis of the corollary $G(L/K)$ is finite. Suppose that $G(L/K) = \{\sigma_1, \sigma_2, \ldots \sigma_n\}$ where $\sigma_i \neq \sigma_j$ for $i \neq j$. Suppose the contrary. Then $[L : K] < n$. Suppose that $[L : K] = m$ and $\{x_1, x_2, \ldots, x_m\}$ is a basis of L over K. Consider the elements $C_1, C_2, \ldots, C_n$ of the set L^m consisting of row vectors given by

$$C_i = (\sigma_i(x_1), \sigma_i(x_2), \ldots, \sigma_i(x_m)).$$

Since L^m is a m-dimensional vector space over L and $n > m$, the set $\{C_1, C_2, \ldots, C_n\}$ is linearly dependent. Thus, there exist $\alpha_1, \alpha_2, \ldots, \alpha_n$ in L not all zero such that

$$\alpha_1 C_1 + \alpha_2 C_2 + \cdots + \alpha_n C_n = 0,$$

where 0 in the RHS is the zero of L^m. This means that

$$\alpha_1 \sigma_1(x_j) + \alpha_2 \sigma_2(x_j) + \cdots + \alpha_n \sigma_n(x_j) = 0$$

for all j. Let $x \in L$. Then, since $\{x_1, x_2, \ldots, x_m\}$ is a basis of L over K, $x = \Sigma_{j=1}^m \beta_j x_j$. But then since each σ_i is a K automorphism, we have

$$\Sigma_{i=1}^n \alpha_i \sigma_i(x) = \Sigma_{j=1}^m \beta_j \Sigma_{i=1}^n \alpha_i \sigma_i(x_j) = 0.$$

Since not all α_i are zero, $\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ is linearly dependent. Further each σ_i is a character of $L^\star$ in L, and hence by the Dedekind theorem $\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ should be linearly independent in $F(L^\star, L)$. This is a contradiction. Hence the assumption that $m < n$ is false. ♯

Theorem 8.2.19 *Let K be the fixed field of a finite group G of automorphisms of a field L. Then $\mid G \mid = [L : K]$.*

Proof By the definition $G \subseteq G(L/K)$, and so by the previous theorem $\mid G \mid \leq \mid G(L/K) \mid \leq [L : K]$. Suppose that $\mid G \mid < [L : K]$. Let $G = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$, where $\sigma_i \neq \sigma_j$ for $i \neq j$. Then since n is supposed to be strictly less than $[L : K]$, there exists a subset $\{x_1, x_2, \ldots, x_{n+1}\}$ of L which contains $n + 1$ elements, and which is linearly independent over K. Consider the subset $\{C_1, C_2, \ldots, C_{n+1}\}$ of L^n consisting of rows with n columns and entries in L, where

$$C_i = (\sigma_1(x_i), \sigma_2(x_i), \ldots, \sigma_n(x_i))$$

$i = 1, 2, \ldots, n+1$. Since the dimension of L^n over L is n, this set is linearly dependent. Rearranging if necessary, we may assume that $\{C_1, C_2, \ldots, C_m\}$ is a minimal subset which is linearly dependent. Then there exist $\alpha_1, \alpha_2, \ldots, \alpha_m$ in L not all zero such that

$$\alpha_1 C_1 + \alpha_2 C_2 + \cdots + \alpha_m C_m = 0,$$

where 0 in the RHS is the zero of L^n. The minimality assumption implies that all α_i are nonzero. Dividing by α_1 we may assume that $\alpha_1 = 1$. Thus, we have

$$\Sigma_{i=1}^m \alpha_i \sigma_j(x_i) = 0 \cdots\cdots \tag{8.2.5}$$

for all j, $1 \leq j \leq n$, where $\alpha_1 = 1$, and no α_i is zero. Let σ be an arbitrary element of G. Then applying σ on the above equation we get that

$$\Sigma_{i=1}^m \sigma(\alpha_i) \sigma\sigma_j(x_i) = 0$$

for all j. Since multiplication by an element of the group to the elements of the group permutes the elements of the group, we have

$$\Sigma_{i=1}^{m}\sigma(\alpha_i)\sigma_j(x_i) \;=\; 0 \cdots\cdots \tag{8.2.6}$$

for all j. Subtracting the Eq. 8.2.2 from the Eq. 8.2.1, and observing that $\alpha_1 \;=\; 1$, we see that

$$\Sigma_{i=2}^{m}(\alpha_i \;-\; \sigma(\alpha_i))\sigma_j(x_i) \;=\; 0$$

for all j. It follows from the minimality assumption of m that $(\alpha_i \;-\; \sigma(\alpha_i)) \;=\; 0$ *for all* i. Thus, $\sigma(\alpha_i) \;=\; \alpha_i$ *for all* i. Since σ is an arbitrary element of G, each α_i belongs to K the fixed field of G. But, then, since each $\sigma_i \in G$, we see that

$$\sigma_j(\Sigma_{i=1}^{m}\alpha_i x_i) \;=\; 0$$

for all j. Since σ_j is an automorphism, we have

$$\Sigma_{i=1}^{m}\alpha_i x_i \;=\; 0.$$

This is a contradiction to the fact that each α_i is nonzero, and $\{x_1, x_2, \ldots, x_{n+1}\}$ is linearly independent over K. ♯

Corollary 8.2.20 *Under the hypothesis of the above theorem,* $G \;=\; G(L/K)$. *In other words* $G(L/F(G)) \;=\; G$.

Proof Already $G \;\subseteq\; G(L/K)$. Also from the Dedekind theorem, and the above theorem, we have $\mid G(L/K) \mid \leq [L : K] \;= \mid G \mid$. The result follows. ♯

Corollary 8.2.21 *Let* L *be a finite field extension of* K. *The* L *is a Galois extension of* K *if and only if* $\mid G(L/K) \mid \;=\; [L : K]$.

Proof Suppose that L is a Galois extension of K. Then by the definition $K \;=\; F(G(L/K))$. Hence from the above theorem $\mid G(L/K) \mid \;=\; [L : K]$. Conversely, suppose that $\mid G(L/K) \mid \;=\; [L : K]$. Let $L_1 \;=\; F(G(L/K))$. Then $K \subseteq L_1$. From the hypothesis and from the previous theorem $[L : K] \;= \mid G(L/K) \mid \;=\; [L : L_1] \leq [L : K]$. Thus, $[L : L_1] \;=\; [L : K]$, and so $K \;=\; L_1$. ♯

Corollary 8.2.22 *Let* L *be a field extension of* K, *and* α *be an element of* L *which is algebraic over* K. *Then* $K(\alpha)$ *is Galois extension of* K *if and only if there are* $deg min_K(\alpha)(X)$ *distinct roots of* $min_K(\alpha)(X)$ *all belonging to* $K(\alpha)$.

Proof We have already seen that $\mid G(K(\alpha)/K) \mid$ is equal to the number of distinct roots of $min_K(\alpha)(X)$ which belong to $K(\alpha)$. Further, we know that $[K(\alpha) : K] \;=\; deg(\min_K(\alpha)(X))$. From the previous theorem, it follows that $K(\alpha)$ is a Galois extension of K if and only if $\mid G(K(\alpha)/K) \mid \;=\; [K(\alpha) : K]$. The result follows. ♯

Example 8.2.23 $\mathbb{Q}(2^{\frac{1}{3}}, \omega) \;=\; \mathbb{Q}(2^{\frac{1}{3}} + \omega)$ is a Galois extension of $\mathbb{Q}$, for $\mid G(\mathbb{Q}(2^{\frac{1}{3}} + \omega)/\mathbb{Q}) \mid \;= \mid S_3 \mid \;=\; 6 \;=\; [\mathbb{Q}(2^{\frac{1}{3}} + \omega) : \mathbb{Q}]$.

Example 8.2.24 Consider the field $L = K(X_1, X_2, \ldots, X_n)$ of rational functions in n variable over the field K. It is the field of fractions of the polynomial ring $K[X_1, X_2, \ldots, X_n]$. If $p \in S_n$ is a permutation of degree n, then we have an unique isomorphism from $K[X_1, X_2, \ldots, X_n]$ to itself given by $f(X_1, X_2, \ldots, X_n) \rightsquigarrow f(X_{p(1)}, X_{p(2)}, \ldots, X_{p(n)})$ which extends uniquely to an automorphism of L. Thus, S_n is isomorphic to a subgroup of $Aut(L)$. Let G be a subgroup of S_n. Then from what we have done above, it follows that L is a Galois extension of $F(G)$, and the Galois group of this extension is G. This, in particular, says that every finite group will appear as a Galois group of some Galois extension.

Exercises

8.2.1 Determine the Galois group of the following extensions. Which of them are Galois extensions:

(i) $\mathbb{Q}(e^{\frac{2\pi i}{5}})$ over $\mathbb{Q}$.
(ii) $\mathbb{Q}(5^{\frac{1}{3}})$ over $\mathbb{Q}$.
(iii) $\mathbb{Q}(\sqrt{p}, \sqrt{q})$ over $\mathbb{Q}$, where p and q are distinct primes.
(iv) $\mathbb{R}$ over $\mathbb{Q}$.
(v) $\mathbb{C}$ over $\mathbb{Q}$.

8.2.2 We have seen that $\mathbb{Q}(2^{\frac{1}{3}}, \omega)$ is Galois extension of $\mathbb{Q}$ with Galois group S_3. Find the fixed field K of A_3. Is K a Galois extension of Q? Support. Find also a fixed field of a subgroup of order 2 in S_3. Is that a Galois extension of $\mathbb{Q}$? support.

8.2.3 Suppose that L is a Galois extension of K, and F a subfield of L containing K. Show that L is also a Galois extension of F. Show by means of an example that F need not be a Galois extension of K.

8.2.4 Is $\mathbb{Q}(\sqrt{3}i)$ isomorphic to $\mathbb{Q}(\sqrt{3})$? support.

8.2.5 Let K and L be two extensions of $\mathbb{Q}$ of degrees 2. Are they necessarily isomorphic? Support.

8.2.6 Find the Group of automorphisms of $\mathbb{Z}_p(X)$. Show that it is finite. Find its order.

8.2.7 Let ρ be the primitive 4th root of unity. Find the Galois group of $\mathbb{Q}(\rho)/\mathbb{Q}$. Is this a Galois extension? Support.

8.2.8 Show that any field extension L of degree 2 of a field K of characteristic different from 2 is a Galois extension.

8.2.9 Let L be a finite field extension of K, and L' any extension of K. Let $\Sigma_K(L, L')$ be the set of all field homomorphisms from L to L' which fix the members of K (in particular $\Sigma_K(L, L) = G(L/K)$). Use the Dedekind theorem and imitate the Proof of Corollary 8.2.18 to show that $| \Sigma_K(L, L') | \leq [L : K]$.

8.2.10 Show that a finite field extension L of K is a Galois extension if and only if the following two conditions hold.

(i) There is a field extension L' of L such that $| \Sigma_K(L, L') | = [L : K]$.
(ii) Given any field homomorphism σ from L to a field extension L' of L which fixes every element of K, $\sigma(L) = L$.

8.2.11 Prove the following generalization of the result in Exercise 8.2.9: Let K and K' be fields. Let L be a finite extension of K and L' an extension of K'. Show that the number of extensions of σ to homomorphisms from L to L' is at most $[L : K]$.

8.3 Splitting Field, Normal Extensions

A finite field extension L of K may cease (see the Exercise 8.2.10 of the previous section) to be a Galois extension because of any of the following two reasons:

(i) There is field extension L' of L and a field homomorphism σ from L to L' fixing each element of K such that $\sigma(L) \neq L$.
(ii) Given any field extension L' of L, $| \Sigma_K(L, L') | < [L : K]$.

In this section, we study those finite extensions L of K for which (i) is not the reason.

Definition 8.3.1 A finite extension L of K is called a **normal** extension if given any field extension L' of L, and a field homomorphism σ from L to L' which fixes every element of K, $\sigma(L) = L$.

Definition 8.3.2 A finite extension L of K is called a **separable** extension if there exists an extension L' of L such that $| \Sigma_K(L, L') | = [L : K]$.

Thus, a finite extension is a Galois extension if and only if it is separable as well as normal.

Separable extensions will be the subject matter of study of the next section.

Theorem 8.3.3 *Let K be a field, and $f(X)$ be a nonconstant polynomial of degree n over K. Then there exists a field extension L of K such that $[L : K] \leq n$, and $f(X)$ has a root in L.*

Proof Every nonconstant polynomial in $K[X]$ is product of irreducible polynomials of positive degrees in $K[X]$, and if an element α in a field extension L of K is a root of an irreducible factor of $f(X)$, then it is also a root of $f(X)$. Thus, it is sufficient to prove the result for irreducible polynomials over $K[X]$ of positive degrees. Let $p(X)$ be an irreducible polynomial in $K[X]$ of degree n. Since $K[X]$ is a principal ideal domain, the ideal $< p(X) >$ generated by $p(X)$ is a maximal ideal. Thus, the quotient ring $K[X]/< p(X) >$ is a field. Let us denote it by F. Define a map η from K to F by $\eta(a) = a + < p(X) >$. It is easily seen that η is a ring homomorphism.

Suppose that $\eta(a) = \eta(b)$. Then $a + < p(X) > = b + < p(X) >$. This means that $a - b$ belongs to $< p(X) >$. Since $p(X)$ is irreducible polynomial of positive degree, this is possible if and only if $a - b = 0$. Thus $a = b$. This shows that η is injective homomorphism and so it is an embedding. Any element of F is of the form $f(X) + < p(X) > = r(X) + < p(X) >$, where $r(X)$ is the remainder obtained when we divide $f(X)$ by $p(X)$. It is also clear that if $r_1(X) + < p(X) > = r_2(X) + < p(X) >$, and $deg(r_i(X)) < deg(p(X)),\ i = 1, 2$, then $r_1(X) = r_2(X)$. Let $\wp_n(K)$ denote the set of all polynomials in $K[X]$ whose degrees are less that n. Then from what we have proved above it follows that the map ρ from $\wp_n(K)$ to $K[X]/ < p(X) >$ defined by $\rho(r(X)) = r(X) + < p(X) >$ is a bijective mapping whose restriction to K is η. Pullback the operations of $K[X]/ < p(X) >$ to those of $\wp_n(K)$ with the help of the map ρ. The operations $\oplus$ and $\star$ on $\wp_n(K)$, thus obtained, are given by

$$r(X) \oplus s(X) = r(X) + s(X),$$

and

$$r(X) \star s(X) = t(X),$$

where $t(X)$ is the remainder obtained when $r(X) \cdot s(X)$ is divided by $p(X)$. Clearly, $\wp_n(K)$ becomes a field such that ρ is an isomorphism, and K is a subfield of $\wp_n(K)$. Further, $X \in \wp_n(K)$, and X is a root of $p(X)$, for $p(X)$ represents 0 in $\wp_n(K)$. ♯

Let $f(X)$ be a polynomial in $K[X]$. We say that $f(X)$ has all its roots in a field extension L of K if $f(X)$ splits into product of linear factors in $L[X]$ in the sense that $f(X) = a(X - a_1)(X - a_2) \ldots (X - a_n)$ for some $a \in K$ and $a_1, a_2, \ldots, a_n$ in L. We also say that $f(X)$ splits over L.

Corollary 8.3.4 *Let $f(X) \in K[X]$ be a polynomial of degree n. Then there is a field extension L of degree at most $n!$ such that $f(X)$ has all roots in L.*

Proof The proof is by the induction on the degree of $f(X)$. If $f(X)$ is of degree 1, then it has only one root which is in K. Assume that the result is true for all those polynomials whose degrees are less than n. Let $f(X)$ be a polynomial of degree n. From the above theorem it follows that there is an extension L of K which has a root α (say) of $f(X)$, and $[L : K]$ is at most n. Then $f(X) = (X - \alpha)g(X)$, where $g(X)$ is a polynomial in $L[X]$ of degree $n - 1$. By the induction hypothesis there is an extension F of L of degree at most $(n - 1)!$ in which $g(X)$ has all its roots. Clearly, $f(X)$ has all its roots in F, and $[F : K] = [F : L][L : K]$ is at most $n!$. ♯

Definition 8.3.5 Let K be a field, and $f(X)$ be a polynomial in $K[X]$. A minimal field extension L of K such that $f(X)$ splits over L is called a **splitting field** of $f(X)$ over K.

We have seen that given any polynomial $f(X) \in K[X]$, there is a field extension L of K such that $f(X)$ splits over K. Let $a_1, a_2, \ldots, a_n$ be all roots of $f(X)$ in L. Then $K(a_1, a_2, \ldots, a_n)$ is a minimal field extension of K in which $f(X)$ splits. Thus, we have

Corollary 8.3.6 *Let K be a field, and $f(X)$ be a polynomial over K. Then $f(X)$ has a splitting field over K.* ♯

The above result says that a splitting field of a polynomial exists. Our next aim is to show that it is unique up to K-isomorphism.

Proposition 8.3.7 *Let K and K' be fields, and σ be an isomorphism from K to K'. Let L be a field extension of K, and L' be an extension of K'. Let a be an element of L which is algebraic over K with minimal polynomial $p(X)$, and b an element of L' which is a root of $p^\sigma(X)$. Then there is an extension Σ of σ to an isomorphism from $K(a)$ to $K'(b)$ such that $\Sigma(a) = b$.*

Proof Since σ is an isomorphism, and $p(X)$ is irreducible, $p^\sigma(X)$ is also irreducible in $K'[X]$. In particular, $p^\sigma(X)$ is the minimum polynomial of b. The map η from $K[X]/ < p(X) >$ to $K'[X]/ < p^\sigma(X) >$ defined by $\eta(f(X)+ < p(X) >) = f^\sigma(X)+ < p^\sigma(X) >$ is an isomorphism (verify). Further, we have an isomorphism ϕ from $K(a)$ to $K[X]/ < p(X) >$ which fixes the members of K and takes a to $X+ < p(X) >$. Similarly, we have an isomorphism ψ from $K'(b)$ to $K'[X]/ < p^\sigma(X) >$ which fixes the members of K' and takes b to $X+ < p^\sigma(X) >$. It is clear that the map $\psi^{-1}o\eta o\phi$ has the desired property. ♯

Corollary 8.3.8 *Let K and K' be fields, and σ be an isomorphism from K to K'. Let $f(X) \in K[X]$. Let L be a splitting field of $f(X)$ over K, and L' be a splitting field of $f^\sigma(X)$ over K'. Then σ can be extended to an isomorphism from L to L'.*

Proof The proof is by the induction on the degree of $f(X)$. If degree of $f(X)$ is 1, then there is nothing to do. Assume that the result is true for all those polynomials whose degrees are less than n. Let $f(X)$ a polynomial of degree n. Suppose that $f(X)$ has a root $a \in K$. Then $f(X) = (X-a)g(X)$, where $g(X)$ is a polynomial in $K[X]$ of degree $n-1$. Clearly, L is a splitting field of $g(X)$, and L' is a splitting field of $g^\sigma(X)$ over K'. By the induction hypothesis, σ can be extended to an isomorphism from L to L'. Suppose that $f(X)$ has no root in K. Let $p(X)$ be an irreducible factor of $f(X)$. Then degree of $p(X)$ is greater than 1, and $p^\sigma(X)$ is an irreducible factor of $f^\sigma(X)$. Let a be a root of $p(X)$ in L, and b be a root of $p^\sigma(X)$ in L'. From the previous result, σ can be extended to an isomorphism σ' from $K(a)$ to $K'(b)$ such that $\sigma'(a) = b$. Now $(X-a)$ divides $f(X)$ over $K(a)$, and $X-b$ divides $f^\sigma(X)$ over $K'(b)$. Suppose that $f(X) = (X-a)g(X)$, and $f^\sigma(X) = (X-b)h(X)$, where $g(X) \in K(a)[X]$, and $h(X) = g^{\sigma'}(X)$ in $K'(b)[X]$. Clearly, L is a splitting field of $g(X)$ over $K(a)$, and L' is the splitting field of $h(X) = g^{\sigma'}(X)$ over $K'(b)[X]$. By the induction hypothesis, σ' can be extended to an isomorphism from L to L'. Thus, σ can be extended to an isomorphism from L to L'. ♯

Corollary 8.3.9 *Splitting field of a polynomial $f(X)$ in $K[X]$ is unique upto K-isomorphism.*

Proof Take $\sigma = I_K$ the identity map on K, and apply the above corollary. ♯

Corollary 8.3.10 *If L is a splitting field of a polynomial $f(X) \in K[X]$ of degree n, then $[L : K] \leq n!$.* ♯

Example 8.3.11 The equality may also hold in the above corollary. Consider the field $L = K(X_1, X_2, \ldots, X_n)$ of rational functions in n indeterminates over the field K. Define a map η from the symmetric group S_n of degree n to the group $Aut(L)$ of automorphisms of L by

$$\eta(p)\left(\frac{f(X_1, X_2, \ldots, X_n)}{g(X_1, X_2, \ldots, X_n)}\right) = \frac{f(X_{p(1)}, X_{p(2)}, \ldots, X_{p(n)})}{g(X_{p(1)}, X_{p(2)}, \ldots, X_{p(n)})}$$

where $p \in S_n$. It is clear that η is an injective homomorphism. Let $F = F(\eta(S_n))$ the fixed field of $\eta(S_n)$. Let us denote $\eta(S_n)$ by G. Let $s_1, s_2, \cdots, s_n$ be n elementary symmetric polynomials given by

$$s_k = \Sigma_{i_1 < i_2 < \cdots < i_k} x_{i_1} x_{i_2} \ldots x_{i_k},$$

where summation is taken over all k-tuples of distinct elements $i_1 < i_2 < \cdots < i_k$. In particular, $s_1 = x_1 + x_2 + \cdots + x_n$, and $s_n = x_1 x_2 \ldots x_n$. We show that $F = K(s_1, s_2, \ldots, s_n)$, and L is the splitting field of the polynomial

$$f(X) = X^n - s_1 X^{n-1} - s_2 X^{n-2} - \cdots - (-1)^n s_n$$

over $K(s_1, s_2, \ldots, s_n)$. Clearly, each $s_k \in F(G) = F$, and so $K(s_1, s_2, \ldots, s_n) \subseteq F$. Further, since F is the fixed field of G, we have $[L : F] = \mid G \mid = \mid S_n \mid = n!$. Next, we note that

$$f(X) = (X - x_1)(X - x_2) \ldots (X - x_n)$$

(proof follows from easy expansion). Thus, L is the splitting field of $f(X)$ over $K(s_1, s_2, \ldots, s_n)$. In turn, we have

$$n! = [L : F] \leq [L : K(s_1, s_2, \ldots, s_n)] \leq n!.$$

The first inequality is true because $K(s_1, s_2, \ldots, s_n) \subseteq F$, and the second inequality is true because of the previous corollary. This shows that $F = K(s_1, s_2, \ldots, s_n)$. ♯

Example 8.3.12 The field $\mathbb{C}$ of complex numbers is the splitting field the polynomial $X^2 + 1$ over the field $\mathbb{R}$ of real numbers. It is also the splitting field of $X^2 + 3$ over $\mathbb{R}$.

Example 8.3.13 $K = \mathbb{Q}(2^{\frac{1}{3}}, \omega)$ is the splitting field of $X^3 - 2$ over $\mathbb{Q}$. This is because K is the smallest field containing all the roots $2^{\frac{1}{3}}, 2^{\frac{1}{3}}\omega,$ *and* $2^{\frac{1}{3}}\omega^2$ of $X^3 - 2$.

Algebraically Closed Field and Algebraic Closure

Definition 8.3.14 A field K is said to be an **algebraically closed field** if every polynomial $f(X)$ in $K[X]$ has all its roots in K, or equivalently, every irreducible element of $K[X]$ is a linear polynomial of the type $aX + b, a, b \in K, a \neq 0$.

Proposition 8.3.15 *The following conditions on a field K are equivalent.*

1. *K is algebraically closed field.*
2. *If $f(X)$ is any polynomial in $K[X]$, then K is the splitting field of $f(X)$.*
3. *Every nonconstant polynomial $f(X) \in K[X]$ has a root in K.*
4. *There is no proper algebraic extension of K.*
5. *There is no proper finite extension of K.*

Proof $1 \Rightarrow 2$. Assume 1. Let $f(X)$ be a polynomial in $K[X]$. Since K is algebraically closed, $f(X)$ has all its roots in K, and so K is the splitting field of $f(X)$.

$2 \Rightarrow 3$. Assume 2. Let $f(X)$ be a nonconstant polynomial in $K[X]$. By 2, K is the splitting field of $f(X)$, and so K contains all roots of $f(X)$. Since $f(X)$ is nonconstant, it has at least one root.

$3 \Longrightarrow 4$. Assume 3. Let L be an algebraic extension of K. Let $a \in L$. Suppose that $a \notin K$. Then the minimum polynomial $min_K(a)(X)$ is an irreducible polynomial of degree greater than 1, and so it will have no root in K. This a contradiction to the assumption.

$4 \Longrightarrow 5$. Assume 4. Since every finite extension is an algebraic extension, there is no finite extension of K.

$5 \Longrightarrow 1$. Assume 5. Then there is no finite extension of K. Let $f(X) \in K[X]$ be a polynomial of positive degree. Let L be the splitting field of $f(X)$. Then L is a finite extension of degree at most $n!$, where n is the degree of $f(X)$. By 5, $L = K$, and so K has all roots of $f(X)$. By the definition, K is algebraically closed. ♯

Example 8.3.16 We shall see later that the field $\mathbb{C}$ of complex numbers is an algebraically closed field. This is known as the fundamental theorem of algebra which was first proved by Gauss.

Proposition 8.3.17 *No finite field can be algebraically closed.*

Proof Let $K = \{a_1, a_2, \ldots, a_n\}$ be a finite field. Then $f(X) = (X - a_1)(X - a_2) \ldots (X - a_n) + 1$ has no root in K. ♯

Now, we shall show that every field can be enlarged to an algebraically closed field.

Definition 8.3.18 An algebraic extension L of K is called an **algebraic closure** of K if L is an algebraically closed field. Thus, a maximal algebraic extension of K, if exists, is called an algebraic closure of K.

Now, we shall show that every field K has an algebraic closure and it is unique upto K-isomorphism.

The following proposition is essential to escape some set theoretic logical problems in proving the existence of algebraic closure.

Proposition 8.3.19 *Let L be an algebraic extension of K. If K is finite, then the cardinality $|\ L\ |$ of L is at most that of the set $\mathbb{N}$ of natural numbers(i.e., it is finite or countably infinite). If K is infinite, then the cardinality $|\ L\ |$ of L is the same as the cardinality $|\ K\ |$ of K.*

Proof Let X_K denote the set of irreducible monic polynomials over the field K. To every member $p(X) \in X_K$, we associate the subset $Y_{p(X)}$ consisting of roots of $p(X)$ in L. $Y_{p(X)}$ may be empty set also. Clearly, $Y_{p(X)}$ is a finite subset of L containing at most n elements, where n is the degree of $p(X)$. It is clear that $L = \bigcup_{p(X)\in X_K} Y_{p(X)}$. Now, each irreducible monic polynomial of degree n is determined uniquely by a choice of an ordered n-tuple in K. Thus, the cardinality of X_K is same as that of the set $\bigcup_{n\in\mathbb{N}} K^n$. If K is finite, then since a countable union of disjoint finite sets is again countable, it follows, in this case, that X_K has the same cardinality as that of $\mathbb{N}$. Next, if K is infinite, then K^n has same cardinality as K. Since K is infinite, a countable union of sets having the same cardinality as that of K again has the same cardinality as that of K. Since $Y_{p(X)}$ is finite, the same argument shows that if K is finite, then the cardinality of L is at most that of $\mathbb{N}$, and if K is infinite, then its cardinality is same as that of K. ♯

Theorem 8.3.20 *Every field has an algebraic closure.*

Proof Let K be a field. Observe that there is no set containing all algebraic extensions of K. Indeed, we need to consider a set of algebraic extensions of K so that every algebraic extension of K is K-isomorphic to a member of the set. For this purpose, let Σ be a set containing K, and whose cardinality is strictly larger than that of $K \bigcup \mathbb{N}$. This is possible because power set of a set always has larger cardinality than that of the set. Let Ω be the set of all fields which are algebraic extensions of K, and whose set part is contained in Σ.

Clearly, Ω is nonempty set for $K \in \Omega$. Define a partial order $\leq$ in Ω by $L_1 \leq L_2$ if L_1 is a subfield of L_2. Thus, $(\Omega, \leq)$ is a nonempty partially ordered set. Let $\{L_\alpha \mid \alpha \in \Lambda\}$ be a chain in Ω. Let $L_0 = \bigcup_{\alpha\in\Lambda} L_\alpha$. Then L_0 is also a field contained in Σ of which all L_α are subfields, and which is an algebraic extension of K. Thus, $L_0 \in \Omega$ is an upper bound of the chain. By the Zorn's lemma Ω has a maximal member $\overline{L}$ (say). Then $\overline{L}$ is an algebraic extension of K. We show that $\overline{L}$ is an algebraic closure of K by showing that it is an algebraically closed field. Suppose not. Then there is a proper algebraic extension F of $\overline{L}$. From the above proposition, it follows that the cardinality of F is strictly less than that of Σ. Hence there is a subset L' of Σ containing $\overline{L}$ properly, and a bijective map η from L' to F which is identity on $\overline{L}$. We can pull back the operations of F to that of L' so that L' becomes a proper algebraic extension of $\overline{L}$. Clearly, L' is also an algebraic extension of K, and $L' \in \Sigma$. This contradicts the supposition that $\overline{L}$ is a maximal member of Ω. This completes the proof of the fact that $\overline{L}$ is an algebraic closure of K. ♯

Theorem 8.3.21 *Let σ be an isomorphism from a field K to a field K'. Let $\overline{K}$ be an algebraic closure of K, and $\overline{K'}$ be an algebraic closure of K'. Then σ can be extended to an isomorphism from $\overline{K}$ to $\overline{K'}$.*

Proof Let Ω be the set of triples (L, σ', L'), where L is a subfield of $\overline{K}$ containing K, L' is a subfield of $\overline{K'}$ containing K', and σ' is an extension of σ to an isomorphism from L to L'. The set Ω is a nonempty set, for (K, σ, K') is a member of Ω. Define a relation $\leq$ on Ω by

$$(L_1, \sigma_1, L_1') \leq (L_2, \sigma_2, L_2') \Longleftrightarrow L_1 \subseteq L_2, L_1' \subseteq L_2' \text{ and } \sigma_2 \text{ an extension of } \sigma_1.$$

Thus, $(\Omega, \leq)$ is a nonempty partially ordered set. Let $\{(L_\alpha, \sigma_\alpha, L_\alpha') \mid \alpha \in \Lambda\}$ be a chain in Ω. Let $L_0 = \bigcup_{\alpha\in\Lambda} L_\alpha$, $L_0' = \bigcup_{\alpha\in\Lambda} L_\alpha'$, and σ_0 is a map from L_0 to L_0' defined by the property that σ_0 restricted to L_α is σ_α. Then it is easy to observe that the triple (L_0, σ_0, L_0') is a member of Ω which is an upper bound of the chain. By the Zorn's lemma Ω has a maximal member $(\overline{L}, \overline{\sigma}, \overline{L'})$ (say). We show that $\overline{L} = \overline{K}$ and $\overline{L'} = \overline{K'}$ which will complete the proof the theorem. Suppose that $a \in \overline{K} - \overline{L}$. Then a is algebraic over $\overline{L}$. Let $p(X)$ be the minimum polynomial of a over $\overline{L}$. Since $a \notin \overline{L}$, $p(X)$ is an irreducible polynomial of degree greater than 1. Since $\overline{\sigma}$ is an isomorphism, it follows that $p^{\overline{\sigma}}(X)$ is an irreducible polynomial over $\overline{L'}$ of degree greater than 1. Since $\overline{K'}$ is algebraically closed there is a root b in $\overline{K'}$ of $p^{\overline{\sigma}}(X)$. By the Proposition 8.3.7, we get an extension τ of $\overline{\sigma}$ to an isomorphism from $\overline{L}(a)$ to $\overline{L'}(b)$. Thus, the triple $(\overline{L}(a), \tau, \overline{L'}(b))$ belongs to Ω, and this is a contradiction to the maximality of the triple $(\overline{L}, \overline{\sigma}, \overline{L'})$. Hence $\overline{L} = \overline{K}$. But, then $\overline{\sigma}(\overline{K})$ is also an algebraically closed field of which $\overline{K'}$ is an algebraic extension. This means that $\overline{\sigma}(\overline{K}) = \overline{K'}$. This completes the proof. ♯

Taking $K = K'$, and $\sigma = I_K$ in the above theorem, we get the following corollary:

Corollary 8.3.22 *Algebraic closure of K is unique upto K-isomorphism.* ♯

The algebraic closure of K will usually be denoted by $\overline{K}$.

Corollary 8.3.23 *Let L be an algebraic extension of K. Then the algebraic closure $\overline{L}$ of L is K-isomorphic to the algebraic closure $\overline{K}$ of K.*

Proof Follows from the fact that the algebraic closure $\overline{L}$ of L is also an algebraic extension of K which is algebraically closed. ♯

Corollary 8.3.24 *Let K be a field and S a set of polynomials over K. Then there is a field extension L of K in which all the polynomials in S has a root. Further, given any two field extensions K_1 and K_2 of K such that all the polynomials in S split over K_1 as well as over K_2, let L_1 be the subfield of K_1 generated by K and the roots of the members of S belonging to K_1, and L_2 be the subfield of K_2 generated by K and the roots of the members of S belonging to K_2. Then L_1 and L_2 are K-isomorphic.*

Proof If $\overline{K}$ is the algebraic closure of K, then all roots of S split over $\overline{K}$. Let L_1 and L_2 be as in the hypothesis of the corollary. Then L_1 and L_2 are both algebraic extension of K. Let $\overline{L_1}$ be algebraic closure of L_1, and $\overline{L_2}$ be the algebraic closure of L_2. Then both of them are algebraic closure of K also. Hence there exists a K-isomorphism σ from $\overline{L_1}$ to $\overline{L_2}$. Clearly σ will take the roots of a polynomial in S to a root of the same polynomial in S. Thus, σ restricted to L_1 will be an isomorphism from L_1 to L_2. ♯

Definition 8.3.25 The unique (upto K-isomorphism) field described in the above corollary is called the **splitting field of the set S** of polynomials over K. It is in fact the smallest field upto injective embeddings over which all polynomials in S split.

Remark 8.3.26 The algebraic closure of K is the splitting field of the set of all polynomials over K.

Theorem 8.3.27 *Let L be an algebraic extension of K. Then the following conditions are equivalent.*

1. *L is splitting field of a family S of polynomials over K.*
2. *If σ is a K-homomorphism from L to its algebraic closure $\overline{L}$, then $\sigma(L) = L$.*
3. *Every member $\sigma \in G(\overline{L}/K)$ restricted to L is a member of $G(L/K)$.*
4. *If $f(X)$ is an irreducible polynomial in $K[X]$ which has a root in L, then it has all its roots in L.*

Proof $1 \implies 2$. Assume 1. Then by the definition of the splitting field of a family of polynomials, it follows that L is the subfield of $\overline{L}$ generated by K and the roots of the members of S. Since any K-homomorphism from L to $\overline{L}$ takes root of a polynomial in S to a root of the same polynomial, it follows that the σ takes the roots of members of S to the roots of members of S, and it also takes K to K. Hence $\sigma(L) = L$.

$2 \implies 3$. Assume 2. If $\sigma \in G(\overline{L}/K)$, then σ restricted to L is a K-homomorphism from L to $\overline{L}$. By 2, $\sigma(L) = L$, and so σ restricted to L is a member of $G(L/K)$.

$3 \implies 4$. Assume 3. Let $f(X)$ be an irreducible polynomial in $K[X]$ which has a root $a \in L$. Let b be another root of $f(X)$ in $\overline{L}$. Then by Proposition 8.3.7, there is K-isomorphism σ from $K(a)$ to $K(b)$ such that $\sigma(a) = b$. Clearly, $\overline{L}$ is an algebraic closure of $K(a)$ as well as of $K(b)$. By the proposition 8.3.21 there is an isomorphism τ which extends σ. From 3, $\tau(L) = L$. Hence $b \in L$.

$4 \implies 1$. Assume 4. Let S be the set of all polynomials in $K[X]$ having a root in L. Then from 4, all the roots of members of S are in L. Further, since L is algebraic over K, every element of L is a root of some polynomial in $K[X]$. It follows that L is the splitting field of S over K. ♯

Definition 8.3.28 An algebraic extension L of K is called a **normal extension** if it satisfies any one (and hence all) of the above conditions.

Corollary 8.3.29 *A finite extension L of K is a normal extension if and only if L is splitting field of a polynomial over K.*

Proof If L is splitting field of a polynomial over K, then by the definition, L is a normal extension of K. Conversely, suppose that L is a finite normal extension. Then $L = K(a_1, a_2, \ldots, a_n)$ for some $a_1, a_2, \ldots, a_n$ in L. Let $f_i(X)$ be the minimum polynomial of a_i. Let $f(X)$ be the product of $f_1(X), f_2(X), \ldots, f_n(X)$. Then clearly L is splitting field of $f(X)$ (note that by the definition L has all roots of $f(X)$). ♯

Corollary 8.3.30 *Every Galois extension is a normal extension.*

Proof Follows from the Exercise 8.2.10. ♯

Remark 8.3.31 A normal extension need not be a Galois extension: Let $K = Z_p(X)$ the field of rational functions over the field Z_p in one variable. Consider the polynomial $Y^p - X$ in $K[Y]$. Let L be the splitting polynomial of $Y^p - X$ over K. Then from the definition, L is a normal extension of K. $Y^p - X$ is irreducible $K[Y]$ by the Eisenstein irreducibility criteria (for X is a prime element in $Z_p[X]$ of which K is the field of fractions). Let a be a root of this polynomial in L. Then $a^p = X$. Thus, $Y^p - X = Y^p - a^p = (Y - a)^p$. Hence a is the only root of $Y^p - X$ which is denoted by $X^{\frac{1}{p}}$. This means that $L = K(X^{\frac{1}{p}})$. If σ is any automorphism in $G(L/K)$, then it permutes the roots of $Y^p - X$. Hence $\sigma(X^{\frac{1}{p}}) = X^{\frac{1}{p}}$. This shows that σ is the identity map. Thus, $G(L/K)$ is the trivial group, but the degree $[L : K] = deg(Y^p - X) = p$. Hence this extension is not a Galois extension.

Corollary 8.3.32 *Let L be a normal extension of K. Let F be any intermediary field. Then L is also a normal extension of F whereas F need not be a normal extension of K.*

Proof Let L be a normal extension of K and F an intermediary field. Then L is the splitting field of a family S of polynomials over K. Since every polynomial over K is also a polynomial over F, L is also the splitting field of the same set S of polynomials over F. This shows that L is a normal extension of F. Next, we show that F need not be a normal extension of K. Consider the extension $\mathbb{Q}(2^{\frac{1}{3}}, \omega)$ of $\mathbb{Q}$. This is the splitting field over $\mathbb{Q}$ of the polynomial $X^3 - 2$, and so it is a normal (in fact it is already seen to be a Galois extension). Further, $Q(2^{\frac{1}{3}})$ is an intermediary field which is not normal. Indeed, $X^3 - 2$ has a root in $Q(2^{\frac{1}{3}})$ but not all its roots are in $Q(2^{\frac{1}{3}})$. ♯

Remark 8.3.33 If L is a normal extension of F, and F is a normal extension of K, then L need not be a normal extension of K. For example, $\mathbb{Q}(\sqrt{2})$ is a normal extension (splitting field of X^2-2) of $\mathbb{Q}$, and $\mathbb{Q}(\sqrt{\sqrt{2}})$ is a normal extension of $\mathbb{Q}(\sqrt{2})$ whereas $\mathbb{Q}(\sqrt{\sqrt{2}})$ is not a normal extension of $\mathbb{Q}$. Clearly, $X^4 - 2$ is the minimum polynomial of $\sqrt{\sqrt{2}}$ over $\mathbb{Q}$ where as not all roots of $X^4 - 2$ are in $Q(\sqrt{\sqrt{2}})$. For example, $\sqrt{\sqrt{2}}i$ is also a root of the polynomial $X^4 - 2$, and it is not in this field (Find the splitting field of this polynomial over $\mathbb{Q}$).

Corollary 8.3.34 *Let F be a finite extension of K. Then there is a smallest subfield L of $\overline{F}$ which contains F, and which is a normal extension of K.*

Proof Suppose that $F = K(a_1, a_2, \ldots, a_n)$. Let $f_i(X)$ be minimum polynomial of a_i over K. Let $f(X)$ be the product of these polynomials, and $L \subseteq \overline{F}$ be the splitting field of $f(X)$ over F. L is also the splitting field of $f(X)$ over K, and it is the smallest normal extension of K containing F. ♯

Definition 8.3.35 The field L described in the above corollary is called the **normal closure** of F over K.

Proposition 8.3.36 *Let F be a finite extension of K of degree m, and $f(X)$ be an irreducible polynomial in $K[X]$ of degree n. Suppose that m and n are co-prime. Then $f(X)$ is also irreducible in $F[X]$.*

Proof Let L be the splitting field of the polynomial $f(X)$ over F. We may assume that $n > 1$. Now, $f(X)$ can not have any of its roots in F, for if a is a root of $f(X)$ which is in F, then $K(a) \subseteq F$. But, then $n = degf(X) = [K(a) : K]$ will divide $[F : K] = m$. This is a contradiction to the supposition. Let a be a root of $f(X)$ in L, which as shown above, is not in F. Consider $F(a)$. Then, since $K(a) \subseteq F(a)$, $[K(a) : K]$, and $[F : K]$ both divide $[F(a) : K] = [F(a) : F][F : K]$. Thus, $m \cdot n$ divides $[F(a) : F] \cdot m$. This shows that n divides $[F(a) : F]$. Since a is a root of $f(X)$, it follows that $[F(a) : F] \leq n$. Hence $[F(a) : F] = n = degf(X)$, and so $f(X)$ is irreducible over F. ♯

As an application of the above proposition we have the following example:

Example 8.3.37 The polynomial $f(X) = X^7 - 10X^4 + 5X^2 + 15X + 10$ is irreducible over $Q(2^{\frac{1}{3}}, \omega)$: By the Eisenstein irreducibility criteria, $f(X)$ is irreducible over Q. Next, we have already seen that $Q(2^{\frac{1}{3}}, \omega)$ is a Galois extension of degree 6 which is co-prime to 7. The assertion follows from the above proposition.

Exercises

8.3.1 Show that $\mathbb{Q}(\omega)$ is the splitting field of $X^4 + X^2 + 1$ over $\mathbb{Q}$, where ω is a primitive cube root of unity.

8.3.2 Show that the splitting field of $X^n - 1$ is $\mathbb{Q}(\rho_n)$, where $\rho_n = e^{\frac{2\pi i}{n}}$ is the primitive nth root of unity. Show that, in case $n = p$ a prime, its degree is $p - 1$. More generally, we shall show that its degree is $\phi(n)$.

8.3.3 Determine the degrees of the splitting fields over $\mathbb{Q}$ of the following polynomials:

(i) $X^4 + 1$.
(ii) $X^3 + X + 1$.
(iii) $X^9 + X^3 + 1$.
(iv) $X^4 - 5$.

8.3.4 Determine the degrees of the splitting fields of the following polynomials:

(i) $X^3 + X + \overline{1}$ over $\mathbb{Z}_5$.
(ii) $X^7 - \overline{5}$ over $\mathbb{Z}_{11}$.

8.3.5 Give examples of rational numbers r and s such that the splitting field of $X^3 + rX + s$ has degree 3 over $\mathbb{Q}$. Can we characterise such r and s?

8.3.6 Use the fact that if σ is a K-automorphism of a field extension L of K, and $f(X)$ is a polynomial in $K[X]$, then σ takes the roots of $f(X)$ to that of $f(X)$, to show the following:

(i) If z is a complex number which is a root of a polynomial $f(X)$ with real coefficients, then the conjugate $\bar{z}$ is also a root of $f(X)$.
(ii) If r is a rational number which is not a square of a rational number, and $a+b\sqrt{r}$ is a root of a polynomial $f(X)$ in $\mathbb{Q}[X]$, then $a-b\sqrt{r}$ is also a root of $f(X)$.

8.3.7 Show that $K = \mathbb{Q}(\sqrt{2}+\sqrt{3})$ is a normal extension of $\mathbb{Q}$ of degree 4. Show that every irreducible polynomial of odd degree over $\mathbb{Q}$ is also irreducible over K.

8.3.8 Let K be a field of characteristic p. Show that

(i) $X^p - a$, where $a \in K$ is either irreducible, or it has all its roots in K.
(ii) $X^p - X - a$ is irreducible, or it has all its roots in K. Deduce that if $a \neq 0$, then it is irreducible over $\mathbb{Z}_p$.

8.4 Separable Extensions

Now, we describe the finite extensions L of K for which the number of K embeddings of L into $\overline{L} = \overline{K}$ is $[L : K]$.

Consider the case when $L = K(a)$. Let $p(X)$ be the minimum polynomial of a. We know that $[K(a) : K] = degp(X)$. Since any K-embedding of L will take a to a root of $p(X)$, there are at the most as many K-embeddings of $K(a)$ into $\overline{K}$ as many distinct roots of $p(X)$. Further, given any root b of $p(X)$, there is a unique Kisomorphism σ from $K(a)$ to $K(b) \subseteq \overline{K}$ which takes a to b. This shows that there are exactly as many K-embeddings of $K(a)$ into $\overline{K}$ as many distinct roots of the minimum polynomial $p(X)$ of a. Thus, we have proved the following:

Proposition 8.4.1 *The number of K-embeddings of $K(a)$ into $\overline{K}$ is $[K(a) : K]$ if and only if all the roots of $p(X)$ are distinct.* ♯

The above proposition motivates to have the following definition.

Definition 8.4.2 An irreducible polynomial $p(X)$ in $K[X]$ is called a **separable polynomial** if all its roots are distinct in its splitting field. A polynomial $f(X)$ (not necessarily irreducible) is said to be **separable** if all its irreducible factors are separable. An algebraic element a of an extension field L of K is said to be **separable** if the minimum polynomial of a is separable. An algebraic extension L of K is said to be a separable extension if every element of L is separable over K. An algebraic extension L of K which is not separable is said to be an **inseparable** extension.

Proposition 8.4.3 *Let L be a separable extension of K and F be an intermediary field. Then L is a separable extension of F, and F is also a separable extension of K.*

Proof Suppose that L is a separable extension of K. Then every element of L is separable over K. In particular every element of F is separable over K. This shows that F is separable over K. Further, if a is an element of L, and $p(X)$ is the minimum polynomial of a over K, then the minimum polynomial of a over F is a factor of $p(X)$. Since $p(X)$ has all its roots distinct, any factor of $p(X)$ will also have its roots distinct. Thus, a is separable over F also. ♯

Definition 8.4.4 Let L be a finite extension of K. Let $[L : K]_s$ denote the number of distinct K-embeddings of L into $\overline{L}$. The number $[L : K]_s$ is called the **degree of separability** of the extension L of K (the justification for this terminology will be clear a little later).

Thus, the above proposition says that $a \in L$ is separable if and only if $[K(a) : K]_s = [K(a) : K]$. We shall show that L is a separable extension of K if and only if $[L : K]_s = [L : K]$.

Proposition 8.4.5 *Let L be a finite extension of K. Let σ be an isomorphism from K to K'. Then the number of extensions of σ to an embedding of L to $\overline{K'}$ is $[L : K]_s$.*

Proof Let $\overline{K}$ be an algebraic closure of K containing L. The it follows by 8.3.21 that σ can be extended to an isomorphism $\overline{\sigma}$ from $\overline{K}$ to $\overline{K'}$. Let $\Sigma_\sigma(L, \overline{K'})$ be the set of extensions of σ to homomorphisms from L to $\overline{K'}$. We have to show that $| \Sigma_\sigma(L, \overline{K'}) | = [L : K]_s$. Define a map η from $\Sigma_\sigma(L, \overline{K'})$ to $\Sigma_K(L, \overline{K})$ by $\eta(g) = \overline{\sigma}^{-1} o g$. Now, $\eta(g_1) = \eta(g_2)$ implies that $\overline{\sigma}^{-1} o g_1 = \overline{\sigma}^{-1} o g_2$. Since $\overline{\sigma}$ is bijective, $g_1 = g_2$. Thus, η is injective. Let $h \in \Sigma_K(L, \overline{K})$. Then $g = \overline{\sigma} o h \in \Sigma_\sigma(L, \overline{K'})$ and $\eta(g) = h$. This shows that η is surjective. Thus $| \Sigma_\sigma(L, \overline{K'}) | = | \Sigma_K(L, \overline{K}) |$. By the definition $| \Sigma_K(L, \overline{K}) | = [L : K]_s$. The result follows. ♯

Corollary 8.4.6 *Let L be a finite field extension of K, and let F be an intermediary field. Then $[L : K]_s = [L : F]_s \cdot [F : K]_s$.*

Proof Let $\overline{K}$ be an algebraic closure of K containing L. Then from the above proposition it follows that every K-embedding of F into $\overline{K}$ has exactly $[L : F]_s$ extensions to embeddings of L into $\overline{K}$. Further, since there are $[F : K]_s$, K-embeddings of F into $\overline{K}$, it follows that there are $[L : F]_s \cdot [F : K]_s$ K-embeddings of L into $\overline{K}$. The result follows from the definition. ♯

Corollary 8.4.7 *Let L be a finite extension of K. Then L is a separable extension of K if and only if $[L : K]_s = [L : K]$.*

Proof The proof is by the induction on $[L : K]$. If $[L : K] = 1$, then $L = K$, and then there is nothing to do. Suppose that the result is true for all extensions whose degrees are less than n. Let L be a separable extension of K of degree n. Then every element of L is separable over K. Let $a \in L - K$. Then by the definition a is separable over K. Let $\overline{K}$ be an algebraic closure of K containing L. We have already seen that the number of K-embeddings of $K(a)$ in $\overline{K}$ is the number of distinct roots of the minimum polynomial $p(X)$ of a, and it is the same as $deg(p(X)) =$

$[K(a) : K]$. Thus, $[K(a) : K]_s = [K(a) : K]$. We have also observed that L is separable over any intermediary field, and so L is separable over $K(a)$. By the induction assumption, $[L : K(a)]_s = [L : K(a)]$. From the previous corollary, we have $[L : K]_s = [L : K(a)]_s \cdot [K(a) : K]_s = [L : K(a)] \cdot [K(a) : K] = [L : K]$. Conversely, suppose that $[L : K]_s = [L : K]$. Let $a \in L$. We have to show that a is separable over K. It is sufficient to show that $[K(a) : K]_s = [K(a) : K]$. Suppose that $[K(a) : K]_s < [K(a) : K]$. Then since $[L : K(a)]_s \leq [L : K(a)]$, we see that $[L : K]_s = [L : K(a)]_s \cdot [K(a) : K]_s$ is strictly less than $[L : K]$. This is a contradiction to the hypothesis. ♯

Corollary 8.4.8 *Let L be finite separable extension of F and F a finite separable extension of K. Then L is a separable extension of K.*

Proof Under the hypothesis of the corollary, $[L : K]_s = [L : F]_s \cdot [F : K]_s = [L : F] \cdot [F : K] = [L : K]$. ♯

Corollary 8.4.9 *Let L be a finite extension of K. Let K_s^L denote the set of all elements of L which are separable over K. Then K_s^L is a subfield of L.*

Proof Let $a, b \in K_s^L$. Then we have already seen that $K(a)$ and $K(b)$ are separable extensions of K. Since b is separable over K, it follows from the previous result that $K(a)(b) = K(a, b)$ is a separable extension of K. Thus $K(a, b) \subseteq K_s^L$. Hence $a - b$, $a \cdot b$ and a^{-1}, $a \neq 0$ are in K_s^L. This shows that K_s^L is a subfield of L. ♯

Definition 8.4.10 The subfield K_s^L is called the **separable closure** of K in L. The separable closure of K in its algebraic closure is called the **separable closure** of K

Corollary 8.4.11 *A finite extension L of K is a Galois extension if and only if it is separable as well as normal.*

Proof We know that a finite extension L of K is a Galois extension if and only if $\mid G(L/K) \mid = [L : K]$. Suppose that L is a finite Galois extension. Since each member of $G(L/K)$ can be viewed as K-embedding of L into $\overline{K}$, there are at least $[L : K]$ K-embeddings of L into $\overline{K}$. There can not be more. Hence $[L : K]_s = [L : K]$. This shows that L is a separable extension of K. We have already seen that a Galois extension of K is also a normal extension. Conversely, suppose that L is a separable normal extension. Since it is separable, there are $[L : K]$ K-embeddings of L into $\overline{K}$. Since L is also a normal extension of K any K-embedding of L into $\overline{K}$ takes L to itself. This shows that $\mid G(L/K) \mid = [L : K]$, and so L is a Galois extension of K. ♯

Corollary 8.4.12 *A finite field extension L of K is Galois extension if and only if it is splitting field of a separable polynomial over K.*

Proof A finite extension is a normal extension if and only if it is splitting field of a polynomial. Thus, it is sufficient to show that the splitting field L of a polynomial $f(X)$ over K is a separable extension if and only if $f(X)$ is a separable polynomial. Suppose that L is splitting field of a separable polynomial $f(X) \in K[X]$. Let

$L = K(a_1, a_2, \ldots, a_n)$, where $a_1, a_2, \ldots, a_n$ are all roots of $f(X)$. Then minimum polynomials of each a_i is a divisor of $f(X)$. Since factor of a separable polynomial is separable, it follows that each a_i is separable. This shows that the separable closure of K in L is $K(a_1, a_2, \ldots, a_n) = L$. In other words L is a separable extension of K. Conversely, suppose that $L = K(a_1, a_2, \ldots, a_n)$ is a splitting field of $f(X)$ where $a_1, a_2, \ldots, a_n$ are roots of $f(X)$, and which is a separable extension of K. Then nonconstant irreducible factors of $f(X)$ are precisely the minimum polynomials of $a_1, a_2, \ldots, a_n$ which are all separable elements. Hence $f(X)$ is separable. ♯

Corollary 8.4.13 *A finite extension F of K is a separable extension if and only if there is a Galois extension L of K such that F is contained in L.*

Proof Since F is a finite extension, $F = K(a_1, a_2, \ldots, a_n)$, and since it is also separable, the minimum polynomial $f_i(X)$ of a_i is separable for each i. Let $f(X) = f_1(X)f_2(X)\ldots f_n(X)$. Then $f(X)$ is separable. Let L be the splitting field of $f(X)$ which contains F. Then from the above corollary, it follows that L is a Galois extension of K. ♯

Our next aim is to have a test for the separability of a polynomial. We first define the concept of formal derivative of a polynomial.

Definition 8.4.14 The **formal derivative** $f'(X)$ of a polynomial

$$f(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$$

is defined to be the polynomial

$$f'(X) = a_1 + 2a_2X + 3a_3X^2 + \cdots + na_nX^{n-1}.$$

The proof of the following proposition is straightforward and it is left as an exercise.

Proposition 8.4.15 *Let $f(X), g(X)$ be polynomials in $K[X]$ and $a, b \in K$. Then*

(i) $(af + bg)'(X) = af'(X) + bg'(X)$.
(ii) $(f \cdot g)'(X) = f'(X)g(X) + f(X)g'(X)$.
(iii) $(fog)'(X) = f'(g(X))g'(X)$.
(iv) $f(X + a) = f(a) + Xf'(a) + \frac{X^2}{2!}f''(a) + \cdots + \frac{X^n}{n!}f^n(a)$. ♯

Proposition 8.4.16 *Let $f(X)$ be a nonconstant polynomial in $K[X]$. Then a is a multiple root of $f(X)$ in a splitting field L of $f(X)$ over K if and only if it is a common root of $f(X)$ and $f'(X)$.*

Proof If $f(X) = (X - a)g(X)$, then $f'(X) = (X - a)g'(X) + g(X)$. Thus, a is a root of $g(X)$ if and only if a is a common root of $f(X)$ and $f'(X)$. ♯

Proposition 8.4.17 *Let $f(X)$ and $g(X)$ be polynomials in $K[X]$. Let L be a field extension of K. Then $f(X)$ and $g(X)$ are co-prime in $K[X]$ if and only if they are co-prime in $L[X]$.*

Proof Suppose that $f(X)$ and $g(X)$ are co-prime in $K[X]$. Then by the Euclidean algorithm there are polynomials $u(X)$ and $v(X)$ in $K[X]$ such that $u(X)f(X) + v(X)g(X) = 1$. Since a polynomial in $K[X]$ are also polynomials in $L[X]$, this is an identity in $L[X]$ also. This shows that they are co-prime in $L[X]$. Conversely, suppose that they are co-prime in $L[X]$. Then again by the Euclidean algorithm there are polynomials $h(X)$ and $k(X)$ in $L[X]$ such that $h(X)f(X) + k(X)g(X) = 1$. If $d(X)$ in $K[X]$ divides $f(X)$ as well as $g(X)$ in $K[X]$, then they divide $f(X)$ and $g(X)$ in $L[X]$ also. Hence $d(X)$ divides 1 in $L[X]$. Hence $d(X)$ is a unit. This means that $f(X)$ and $g(X)$ are co-prime in $K[X]$ also. ♯

Proposition 8.4.18 *Let $f(X) \in K[X]$ be a nonconstant polynomial. Let L be an splitting field of $f(X)$ over K. Then all the roots of $f(X)$ in L are distinct if and only if $f(X)$ and $f'(X)$ are co prime in $K[X]$.*

Proof Suppose that $f(X)$ and $f'(X)$ are co-prime in $K[X]$. Suppose that a is a multiple root of $f(X)$. Then $f(X) = (X-a)^2 g(X)$ for some $g(X) \in L[X]$. Now $f'(X) = 2(X-a)g(X) + (X-a)^2 g'(X)$. This shows that $X-a$ divides $f(X)$ as well as $f'(X)$. Hence $f(X)$ and $f'(X)$ are not co-prime in $L[X]$. From the above proposition, $f(X)$ and $f'(X)$ are not co-prime in $K[X]$. Conversely, suppose that all roots of $f(X)$ are distinct in L. Let L' be the splitting field of $f'(X)$ over L. Suppose that $f(X)$ and $f'(X)$ have a common root a in L', and so also in L (note that all roots of $f(X)$ are supposed to be in L). Suppose that $f(X) = (X-a)g(X)$. Then $f'(X) = g(X) + (X-a)g'(X)$. Since a is also a root of $f'(X)$, it follows that $(X-a)$ divides $g(X)$. But, then $(X-a)^2$ divides $f(X)$. This is a contradiction to the supposition that $f(X)$ has no multiple root. Hence $f(X)$ and $f'(X)$ have no common root in L'. This also shows that $f(X)$ and $f'(X)$ are co-prime in $L'[X]$, and so (by the above proposition) they are also co-prime in $K[X]$. ♯

Corollary 8.4.19 *Let $f(X)$ be an irreducible polynomial in $K[X]$. Then $f(X)$ is separable (i.e., all its roots distinct) if and only if $f(X)$ does not divide $f'(X)$.*

Proof Since $f(X)$ is irreducible greatest common divisor $(f(X), f'(X))$ is a unit or it is an associate of $f(X)$. By the definition $f(X)$ is separable if and only if it has no multiple roots. From the above proposition, this is equivalent to say that $(f(X), f'(X))$ is a unit. This in turn is equivalent to say that that $f(X)$ does not divide $f'(X)$. ♯

Corollary 8.4.20 *An irreducible polynomial $f(X)$ in $K[X]$ is separable if and only if $f'(X) \neq 0$.*

Proof If $f'(X) \neq 0$, then it is a polynomial of lower degree than $f(X)$, and so $f(X)$ can not divide $f'(X)$. The result follows from the above proposition. ♯

Since in a field of characteristic 0, $f'(X) = 0$ if and only if $f(X)$ is a constant polynomial (verify), the following corollary is immediate.

Corollary 8.4.21 *Every polynomial over a field K of characteristic 0 is separable.* ♯

Corollary 8.4.22 *Let K be a field of characteristic 0. Then every algebraic extension L of K is separable.*

Proof By the definition L is separable over K if and only if all elements of L are separable over K. This, in turn, means that the minimum polynomial of each element of L over K is separable. The result follows. ♯

Corollary 8.4.23 *Let K be a field of characteristic $p \neq 0$. Let $f(X)$ be an irreducible polynomial in $K[X]$. Then $f(X)$ is separable if and only if there is no polynomial $g(X) \in K[X]$ such that $f(X) = g(X^p)$.*

Proof We have seen that an irreducible polynomial $f(X)$ is separable if and only if $f'(X) \neq 0$. Let K be a field of characteristic $p \neq 0$. Let

$$f(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$$

be an irreducible polynomial in $K[X]$. Then

$$f'(X) = a_1 + 2a_2X + 3a_3X^2 + \cdots + na_nX^{n-1} = 0$$

if and only if $ia_i = 0$ for all i. This means that if p does not divide i, $a_i = 0$. This shows that $f'(X) = 0$ if and only if $f(X)$ is of the form

$$f(X) = a_0 + a_pX^p + a_{2p}X^{2p} + \cdots + a_{rp}X^{rp}$$

for some r. This is equivalent to say that $f(X) = g(X^p)$, where

$$g(X) = a_0 + a_pX + a_{2p}X^2 + \cdots + a_{rp}X^r.$$

The result follows. ♯

Proposition 8.4.24 *Let K be a field of characteristic $p \neq 0$. Let $a \in K$. Then the polynomial $X^p - a$ is irreducible over K if and only if it has no root in K.*

Proof If $X^p - a$ is irreducible, then obviously it has no roots. Conversely, suppose that it has no root in K. Let L be a splitting field of this polynomial, and let $b \in L$ be a root of $X^p - a$. Then $b \notin K$ and $b^p = a$. Further,

$$(X^p - a) = (X^p - b^p) = (X - b)^p.$$

Now, any nonunit factor of $X^p - a$ in $K[X]$ will be of the form $(X - b)^r$ for some r, $1 \leq r \leq p$. Suppose that $(X - b)^t \in K[X]$. Then $1 < t$ for $b \notin K$.

Suppose that $t < p$. Then since $(X - b)^t \in K[X]$, it follows that $tb \in K$. Since t is co-prime to p, $b \in K$. This is a contradiction to the supposition. Hence, in this case, it is irreducible. ♯

Corollary 8.4.25 *Let K be a field of characteristic $p \neq 0$. Then every algebraic extension of K is separable over K if and only if $K^p = \{a^p \mid a \in K\} = K$.*

Proof To say that every algebraic extension of K is separable and is equivalent to say that every polynomial over K is separable. This, in turn, is equivalent to say that every irreducible polynomial over K is separable. Suppose that every irreducible polynomial over K is separable. Let a be a nonzero element of K. Consider the polynomial $X^p - a$. If it is irreducible, then since its derivative is 0, it is not separable. Hence $X^p - a$ is not irreducible. From the previous proposition, it follows that $X^p - a$ has a root $b \in K$. This shows that $a = b^p \in K^p$. Conversely, suppose that $K^p = K$. We show that every irreducible polynomial over K is separable. Suppose contrary. Let $f(X)$ be an irreducible polynomial in $K[X]$ which is not separable. Then we have $f(X) = g(X^p)$ for some polynomial $g(X) \in K[X]$. Suppose that

$$g(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n.$$

Since $K^p = K$, we have $b_i \in K$ such that $b_i^p = a_i$ *for all* i. Hence

$$f(X) = g(X^p) = (b_0 + b_1X + b_2X^2 + \cdots + b_nX^n)^p$$

This contradicts the supposition that $f(X)$ is irreducible. ♯

Definition 8.4.26 A field K is said to **perfect** if every algebraic extension of K is separable.

The above results can be restated in the light of the above definition:

Corollary 8.4.27 *Every field of characteristic* 0 *is a perfect field. A field K of characteristic $p \neq 0$ is perfect if and only if $K^p = K$.* ♯

Corollary 8.4.28 *Every finite field is perfect.*

Proof Let K be a finite field of characteristic p. Consider the map σ from K to K defined by $\sigma(a) = a^p$. Then $\sigma(a) = \sigma(b)$ implies that $a^p = b^p$. Now, $(a-b)^p = a^p - b^p = 0$. Since K is a field, $a - b = 0$. Thus, σ is injective. Since K is finite, it is surjective. This means that $K^p = K$. From the previous result, K is perfect. ♯

We know that the order of every finite field is p^n for some prime p and $n \geq 1$.

Theorem 8.4.29 *Given any prime p and $n \geq 1$, there is one and only one (up to isomorphism) field of order p^n.*

Proof Consider the field $\mathbb{Z}_p$ of residue classes modulo p. Consider the polynomial $X^{p^n} - X$ in $\mathbb{Z}_p[X]$. Since its derivative is $-\overline{1} \neq \overline{0}$, all its roots in the splitting field L of this polynomial are distinct. We show that L is precisely the set of roots of $X^{p^n} - X$. It is sufficient to show that the set of roots of $X^{p^n} - X$ form a field. Let a and b be roots of this polynomial. Then $a^{p^n} = a$ and $b^{p^n} = b$. Now,

$$(a+b)^{p^n} = a^{p^n} + b^{p^n} = a + b,$$

$$(ab)^{p^n} = a^{p^n} b^{p^n} = ab,$$

and if $a \neq 0$, then

$$(a^{-1})^{p^n} = (a^{p^n})^{-1} = a^{-1}.$$

This shows that $a+b$, ab and a^{-1} are all roots of the above polynomial. This completes the existence of a field of order p^n. For uniqueness, let L_1 and L_2 be fields of order p^n. Then both are splitting fields of $X^{p^n} - X$ over their respective prime subfields. Since their prime subfields are isomorphic (to $\mathbb{Z}_p$), it follows from Corollary 8.3.9 that L_1 and L_2 are isomorphic. ♯

Corollary 8.4.30 *Every finite extension of a finite field is a Galois extension. If K is a field containing q elements, and L is a field extension of degree n, then the Galois group $G(L/K)$ is a cyclic group of order n generated by σ, where σ is defined by $\sigma(a) = a^q$.*

Proof Clearly, L is the splitting field of $X^{q^n} - X$ over K. Hence L is normal as well as separable extension of K. By the Corollary 8.4.11, it is Galois extension. Further, since $a^q = a$ *for all* $a \in K$, it is clear that the map σ defined by $\sigma(a) = a^q$ is a member of $G(L/K)$. Also $\sigma^n(a) = a^{q^n} = a$ *for all* $a \in L$, and so $\sigma^n = I_L$. If $m < n$, then σ^m can not be the identity map, for other wise every element of L would turn out to be a root of $X^{q^m} - X$. This shows that σ generates a cyclic subgroup of G of order n. Since L is Galois over K, $| G(L/K) | = [L : K] = n$. Thus, σ generates $G(L/K)$. ♯

Remark 8.4.31 Consider the algebraic closure L of $\mathbb{Z}_p$. Let K be a subfield of L order p^n, and F a subfield of order p^m. Then $K \subseteq F$ if and only if n divides m. This is clear, for if $K \subseteq F$, then F is a vector space over K of dimension r (say), and then F should contain exactly $(p^n)^r = p^{nr}$ elements. This means that $m = nr$. Conversely, if $m = nr$, then $X^{p^n} - X$ divides $X^{p^m} - X$. In fact,

$$(X^{p^{nr}} - X) = (X^{p^n} - X)(1 + X^{p^n} + X^{p^{2n}} + \cdots + X^{p^{(r-1)n}}).$$

This shows that the splitting field K of $X^{p^n} - X$ is contained in the splitting field F of $X^{p^m} - X$.

Proposition 8.4.32 *Let K be a finite field containing $q = p^m$ elements. Let $f(X)$ be a monic irreducible polynomial of degree n over K. Let a be a root of $f(X)$ in a*

field extension L of K. Then $K(a)$ is the splitting field of $f(X)$, and all roots of $f(X)$ are of the form a^{q^r}, $r \geq 1$.

Proof Clearly, $[K(a) : K] = degf(X) = n$. Thus, $K(a)$ is a field extension of K of degree n. From the above results, $K(a)$ is a Galois extension, and it is splitting field of $X^{q^n} - X$ over K. Thus, all the roots of $f(X)$ lie in $K(a)$. Further, $G(K(a)/K)$ is the cyclic group of order n generated by σ where σ is defined by $\sigma(b) = b^q$. This shows that $\{\sigma^r(a) \mid 1 \leq r \leq n\} = \{a^{q^r} \mid 1 \leq r \leq n\}$ is the set of all distinct roots of $f(X)$. ♯

These results say that to find a field K of order p^n, it is sufficient to find irreducible polynomials of degree n over $\mathbb{Z}_p$, and then K is simply isomorphic to the field $Z_p[X]/ < f(X) >$, where $f(X)$ is an irreducible polynomial over $\mathbb{Z}_p$ of degree n. There is an effective procedure using the division algorithm in $\mathbb{Z}_p[X]$ to enumerate the elements of K, and also to determine the addition and multiplications in K. How to determine irreducible polynomials of degree n in $\mathbb{Z}_p[X]$? As observed if $f(X)$ is a monic irreducible polynomial over $\mathbb{Z}_p$ of degree n, then the splitting field of $f(X)$ is same as the splitting field of $X^{p^n} - X$. If a is a root of $f(X)$, then it is also a root of $X^{p^n} - X$. Since $f(X)$ is the minimum polynomial of a, it divides $X^{p^n} - X$. This shows that all irreducible monic polynomials of degree n are the irreducible factors of $X^{p^n} - X$. Further, let $g(X)$ be any monic irreducible polynomial over $\mathbb{Z}_p[X]$ of degree m, where m divides n. Then the splitting field of $g(X)$ is also the splitting field of $X^{p^m} - X$, and it is contained in the splitting field of $X^{p^n} - X$. Hence any root of $g(X)$ is also a root of $X^{p^n} - X$. Conversely, if $g(X)$ is an irreducible monic polynomial of degree m which is a factor of $X^{p^n} - X$, then the splitting field of $g(X)$ is of order p^m, and it is contained in the splitting field of $X^{p^n} - X$. Thus, m divides n. This shows that all irreducible polynomials of degrees m, where m divides n are factors of $X^{p^n} - X$, and they are the only(upto associate) factors of $X^{p^n} - X$. Since the roots of this polynomial are all distinct, it has no repeated irreducible factors. The arguments above combine to give the following:

Theorem 8.4.33 *The polynomial $X^{p^n} - X$ in $\mathbb{Z}_p[X]$ is the product of distinct irreducible monic polynomials in $\mathbb{Z}_p[X]$ whose degrees are divisors of n.* ♯

One can develop a computer program to factorize $X^{p^n} - X$ in $\mathbb{Z}_p[X]$ for small primes p and for small n.

Example 8.4.34 Consider the case $p = 2$ and $n = 3$. We wish to find all monic irreducible polynomial of degree 3 in $\mathbb{Z}_2[X]$, and also factorize $X^{2^3} - X$ as product of irreducible factors. Since 1 and 3 are the only divisors of 3, irreducible factors of $X^8 - X$ are irreducible polynomials of degree 1 and irreducible polynomials of degree 3. The irreducible polynomials of degree 1 are X and $X + 1$ only. Consider the irreducible polynomials of degree 3. Let us enumerate the polynomials of degree 3. They are X^3, $X^3 + X^2$, $X^3 + X$, $X^3 + 1$, $X^3 + X^2 + X$, $X^3 + X^2 + 1$, $X^3 + X + 1$ and $X^3 + X^2 + X + 1$. Polynomials of degree 3 which have no roots in $\mathbb{Z}_2$ are $X^3 + X^2 + 1$ and $X^3 + X + 1$. They are all irreducible also. Thus,

$$X^8 - X = X(X+1)(X^3 + X^2 + 1)(X^3 + X + 1).$$

If a is a root of $X^3 + X^2 + 1$, then $a^3 = a^2 + 1$. Note that a^2 and a^4 are also roots of $X^3 + X^2 + 1$. If b is a root of $X^3 + X + 1$, then $b^3 = b + 1$. $\mathbb{Z}_2[X]/ < X^3 + X^2 + 1 >$ and $Z_2[X]/ < X^3 + X + 1 >$ are both fields of order 8. Determine an explicit isomorphism between them. Compare this with the example of a field of order 8 given in Chap. 7, of algebra 1.

Example 8.4.35 Consider the case when $p = 2$ and $n = 5$. Since 1 and 5 are the only divisors of 5, it follows that only irreducible polynomials of degrees 1 and 5 are factors of $X^{32} - X$. Irreducible polynomials of degree 1 (as above) are X and $X + 1$. Irreducible polynomials in $\mathbb{Z}_2[X]$ of degree 5 are precisely

$$X^5 + X^3 + 1, X^5 + X^2 + 1, X^5 + X^4 + X^3 + X + 1,$$
$$X^5 + X^4 + X^2 + X + 1, X^5 + X^4 + X^3 + X^2 + 1,$$

and $X^5 + X^3 + X^2 + X + 1$. It is easily seen that $X^{32} - X$ is product of these irreducible polynomials.

Exercises

8.4.1 Give an example of a inseparable polynomial over some field.
Hint. Consider the field $K = \mathbb{Z}_p(X)$ and take the polynomial $Y^p - X$ in $K[Y]$.

8.4.2 Let K be a field of characteristic $p \neq 0$. Show that the field $K(X)$ is not a perfect field.

8.4.3 Let K be a field of characteristic $p \neq 0$. Let $f(X)$ be an irreducible polynomial in $K[X]$. Show that there exists $n \geq 1$ and a separable polynomial $g(X)$ in $K[X]$ such that $f(X) = g(X^{p^n})$.

8.4.4 Let L be a field extension of K and the characteristic of K is $p \neq 0$. Let a be an element of L which is algebraic over K. Show that there is a positive integer n such that a^{p^n} is separable over K.

8.4.5 Call an element a of the extension L of a field K of characteristic $p \neq 0$ a **purely inseparable** element if it is algebraic over K, and its minimum polynomial has only one root namely a. Show that a is purely inseparable over K if and only if $a^{p^n} \in K$ for some n.

8.4.6 Show that an element a of L is separable as well as purely inseparable if and only if $a \in K$.

8.4.7 Show that if $a \in L$ is purely inseparable over K, then $K(a)$ is splitting field of the minimum polynomial of a, and $G(K(a)/K)$ is trivial.

8.4.8 Let L be an algebraic extension of K. Show that every element of L is purely inseparable over the separable closure K_s of K in L.

8.4.9 Call an algebraic extension L of K to be a **purely inseparable** extension if every element of L is purely inseparable over K. Show that L is purely inseparable of K_s.

8.4.10 Show that if L is a finite purely inseparable extension of K, then $[L : K] = p^n$ for some prime p and n.

8.4.11 Let L be a finite extension of K, and F be an intermediary field. Show that L is purely inseparable extension of K if and only if L is purely inseparable over F, and F is purely inseparable over K.

8.4.12 Let L be a field extension of K. Let K_i denote the set of all purely inseparable elements of L. Show that K_i form a subfield of L called **purely inseparable closure** of K in L. Observe that L need not be separable over K_i.

8.4.13 Define $[L : K]_i = [L : K_i]$ and call it the degree of inseparability. Show that $[L : K]_i = [L : F]_i[F : K]_i$, where F is an intermediary field.

8.4.14 Let L be a finite normal extension of K. Show that K_s is a Galois extension of K, and $G(L/K)$ is isomorphic to $G(K_s/K)$. Deduce that $\mid G(L/K) \mid = [L : K]_s$.

8.4.15 Let K be a field of characteristic $p \neq 0$ and $a \in K$ such that $a \notin K^p$. Show that $X^p - a$ is irreducible over K.

8.4.16 Find all irreducible polynomials of degree 4 over $\mathbb{Z}_2$. Factorize $X^{16} - X$ as product of irreducible factors over $\mathbb{Z}_2$. Determine the structure of a field of order 16.

8.4.17 Determine the cubic irreducible polynomials over $\mathbb{Z}_3$, and factorize $X^{27} - X$ over $\mathbb{Z}_3$. Determine the structure of a field of order 27.

8.4.18 Express $X^4 + 1$ as product of irreducible elements in $\mathbb{Z}_3[X]$. Determine its splitting field.

8.4.19 Show that $X^4 - 7$ is irreducible over $\mathbb{Z}_5$.
Hint. Observe that 7 is not quadratic residue mod 5.

8.4.20 Determine irreducible polynomials of degree 5 over $\mathbb{Z}_3$, and factorize $X^{243} - X$ as product of irreducible elements in $\mathbb{Z}_3[X]$.

8.4.21 Let $\psi(q, d)$ denote the number of irreducible polynomials of degree d over a field K_q containing q elements. Then show that

$$q^n = \Sigma_{d/n} d\psi(q, d).$$

Use the inversion formula to show that

$$n\psi(q, n) = \Sigma_{d/n} \mu(d) q^{\frac{n}{d}}.$$

8.4.22 Find the number of irreducible polynomials of degree 9 over a field K of order 125.

8.5 Fundamental Theorem of Galois Theory

In this section, we relate the intermediary fields of Galois extensions with the subgroups of the Galois groups. We translate problems in the field theory to the problems in group theory. As a simple application, we prove the fundamental theorem of *algebra*. Other applications of fundamental theorem of Galois theory will follow in the following sections.

Theorem 8.5.1 (Fundamental theorem of Galois theory). *Let L be a finite Galois extension of K. Let $S(G(L/K))$ denote the set of all subgroups of the Galois group $G(L/K)$. Let $S(L/K)$ denote the set of all intermediary subfields of the field extension L of K. Then we have a bijective map ϕ from $S(G(L/K))$ to $S(L/K)$ given by $\phi(H) = F(H)$ (the fixed field of H), and a bijective map ψ from $S(L/K)$ to $S(G(L/K))$ given by $\psi(F) = G(L/F)$ such that ϕ and ψ are inverses of each other. Further, the following conditions hold.*

(i) $H_1 \subseteq H_2 \Longrightarrow F(H_2) \subseteq F(H_1)$.
(ii) $F_1 \subseteq F_2 \Longrightarrow G(L/F_2) \subseteq G(L/F_1)$.
(iii) $| H | = [L : F(H)]$, *and* $[F(H) : K] = [G(L/K) : H]$.
(iv) H *is normal subgroup of* $G(L/K)$ *if and only if* $F(H)$ *is a Galois extension of* K, *and then* $G(F(H)/K)$ *is isomorphic to the quotient group* $G(L/K)/H$.

Proof Clearly, ϕ and ψ are maps. By the Corollary 8.2.20, $H = G(L/F(H))$. Also since L is a Galois extension of K, for any intermediary field F, L is also a Galois extension of F, and so it follows from the definition of Galois extension (Definition 8.2.3) that $F(G(L/F)) = F$. This shows that ϕ and ψ are inverses of each other. In particular, they are bijective maps also.

(i) and (ii) are restatements of Proposition 8.2.4. The part (iii) follows from the Theorem 8.2.19.

(iv) Suppose that H is a normal subgroup of $G(L/K)$. We have to show that $F(H)$ is a Galois extension of K. Since L is a Galois, and so separable extension of K, every element of L is separable over K. In particular, every element of $F(H)$ is separable over K. Thus, in any case $F(H)$ is a separable extension of K. We show that under the assumption that H is normal subgroup of $G(L/K)$, $F(H)$ is also a normal extension of K. It suffices to show that $\sigma(F(H)) \subseteq F(H)$ *for all* $\sigma \in G(L/K)$. Let $a \in F(H)$, and $\sigma \in G(L/K)$. Let $\tau \in H$. Then $\tau(\sigma(a)) = \sigma((\sigma^{-1} o \tau o \sigma)(a)) = \sigma(a)$, for $\sigma^{-1} o \tau o \sigma$ belongs to H and $a \in F(H)$. This shows that $\sigma(a) \in F(H)$ *for all* $\sigma \in G(L/K)$. Thus, $F(H)$ is a Galois extension of K. Conversely, suppose that $F(H)$ is a Galois extension of K. Then any $\sigma \in G(L/K)$ restricted to $F(H)$ is an automorphism of $F(H)$. This enables us to define a map ϕ from $G(L/K)$ to $G(F(H)/K)$ by $\phi(\sigma) = \sigma/F(H)$ (the restriction of σ to $F(H)$). Clearly, this is a homomorphism. Since L is a Galois extension of K, any element of $G(F(H)/K)$ can be extended to an element of $G(L/K)$. Thus, ϕ is a surjective homomorphism. Further, $Ker\phi = G(L/F(H)) = H$. This shows that H is a normal subgroup of $G(L/K)$, and by the fundamental theorem of homomorphism, $G(L/K)/H$ is isomorphic to $G(F(H)/K)$. ♯

Corollary 8.5.2 *Let L be a finite Galois extension of K. Then there are only finitely many intermediary fields. In particular, it is simple extension.*

Proof Since $G(L/K)$ is finite of order $[L : K]$, it has only finitely many subgroups. By the fundamental theorem of Galois theory, there is a bijective correspondence between the set of subgroups of $G(L/K)$ to the set of intermediary fields. The result follows. Finally by the Theorem 8.1.17, it follows that L is a simple extension of K.

Corollary 8.5.3 *Every finite separable extension is simple.*

Proof Let L be a finite separable extension of K. Suppose that $L = K(a_1, a_2, \ldots, a_n)$. Let $f_i(X)$ be the minimum polynomial of a_i. Then each $f_i(X)$ is separable. Let L' be the splitting field of $f(X) = f_1(X)f_2(X)\ldots f_n(X)$ containing L. Then L' is a finite Galois extension of K. Hence there are only finitely many intermediary field between L' and K. In particular, there are only finitely many intermediary field in between L and K. Again by the Theorem 8.1.17, L is a simple extension of K. ♯

Corollary 8.5.4 *Every finite field extension of a field K of characteristic* 0 *is simple.*

Proof Since every finite extension of a field of characteristic 0 is separable, the result follows. ♯

Our next aim will be to use the fundamental theorem of Galois theory to enumerate intermediary fields in a Galois extension L of K by calculating the Galois group $G(L/K)$, enumerating all subgroups of $G(L/K)$, and finding fixed fields of these subgroups. In this section we give some simple examples to illustrate it.

Example 8.5.5 Recall Examples 8.2.14, 8.2.23, and 8.3.13. The extension $L = \mathbb{Q}(2^{\frac{1}{3}}, \omega) = \mathbb{Q}(2^{\frac{1}{3}} + \omega)$ is a Galois extension of $\mathbb{Q}$ with Galois group isomorphic to S_3. Since S_3 has 6 subgroups, there are 6 intermediary fields. We enumerate them. Clearly, the fixed field $F(S_3)$ of S_3 is $\mathbb{Q}$. The fixed field of the trivial subgroup is the field $\mathbb{Q}(2^{\frac{1}{3}} + \omega)$. Consider the subgroup $< \sigma >$ of $G(L/\mathbb{Q})$ generated by the element σ, where σ takes $2^{\frac{1}{3}}$ to $\omega 2^{\frac{1}{3}}$, and takes ω to itself. Clearly, $< \sigma >$ is the unique normal subgroup of $G(L/\mathbb{Q})$ of order 3, and it is isomorphic to A_3. Thus, $F(< \sigma >)$ is a Galois extension of $\mathbb{Q}$ of degree $[G(L/\mathbb{Q}) :< \sigma >] = 2$. Also $\omega \in F(< \sigma >)$. Thus, $\mathbb{Q}(\omega) \subseteq F(< \sigma >)$. Since $[\mathbb{Q}(\omega) : \mathbb{Q}] = 2$. Hence $F(< \sigma >) = \mathbb{Q}(\omega) = \mathbb{Q}(\omega^2)$. Let τ_1, τ_2, τ_3 be elements of $G(L/K)$ given by $\tau_1(2^{\frac{1}{3}}) = 2^{\frac{1}{3}}$, $\tau_1(\omega) = \omega$; $\tau_2(2^{\frac{1}{3}}) = \omega 2^{\frac{1}{3}}$, $\tau_2(\omega) = \omega^2$, and $\tau_3(2^{\frac{1}{3}}) = \omega^2 2^{\frac{1}{3}}$, $\tau_3(\omega) = \omega^2$. It is clear that $< \tau_1 >$, $< \tau_2 >$, and $< \tau_3 >$ are all 3 subgroups of order 2 of $G(L/K)$. Now, τ_1 fixes the field $\mathbb{Q}(2^{\frac{1}{3}})$, and $[\mathbb{Q}(2^{\frac{1}{3}}) : \mathbb{Q}] = 3 = [G(L/K) :< \tau_1 >]$. Thus, $F(< \tau_1 >) = \mathbb{Q}(2^{\frac{1}{3}})$. Similarly, $F(< \tau_2 >) = \mathbb{Q}(\omega^2 2^{\frac{1}{3}})$, and $F(< \tau_3 >) = \mathbb{Q}(\omega 2^{\frac{1}{3}})$. This determines all intermediary fields.

Example 8.5.6 Consider the polynomial $X^4 - 5$ in $\mathbb{Q}[X]$. By the Eisenstein irreducibility criteria, it is irreducible over $\mathbb{Q}$. Let L be the splitting field of this polynomial over $\mathbb{Q}$. Then L is a Galois extension of $\mathbb{Q}$. We find its Galois group, and

also all intermediary fields. Let α be a real fourth root of 5. Then $\pm\alpha, \pm i\alpha$ are all four roots of $X^4 - 5$. Thus, $\pm i \in L$. Hence $L = \mathbb{Q}(\alpha, i)$. It is easy to verify that $L = \mathbb{Q}(\alpha + i)$. $[\mathbb{Q}(i) : \mathbb{Q}] = 2$. Consider the polynomial $X^4 - 5$ over $\mathbb{Q}(i)$. Note that $\mathbb{Q}(i)$ is the field of fractions of $\mathbb{Z}[i]$. It is easy to observe that $1 + 2i$ is an irreducible element of $\mathbb{Z}[i]$ which divides 5 ($5 = (1 + 2i)(1 - 2i)$) but $(1 + 2i)^2$ does not divide 5. By the Eisenstein irreducibility criteria, $X^4 - 5$ is irreducible over $\mathbb{Q}(i)$. Since α is a root of $X^4 - 5$, $L = \mathbb{Q}(i)(\alpha)$ is of degree 4 over $\mathbb{Q}(i)$. Thus, $[L : \mathbb{Q}] = [L : \mathbb{Q}(i)][\mathbb{Q}(i) : \mathbb{Q}] = 8$. Now, we find $G(L/\mathbb{Q})$. We first find $G(L/\mathbb{Q}(i))$. Consider the automorphism σ defined by $\sigma(\alpha) = \alpha i$ and $\sigma(i) = i$. Clearly, $\sigma \in G(L/\mathbb{Q}(i))$, and it generates a cyclic group of order 4. Thus, $< \sigma > = G(L/\mathbb{Q}(i))$. It has a unique proper subgroup $< \sigma^2 >$ of order 2. Since $\sigma^2(\alpha^2) = \alpha^2$, it follows that the fixed field of $< \sigma^2 >$ is $\mathbb{Q}(\alpha^2, i)$ (find β such that $\mathbb{Q}(\alpha^2, i) = \mathbb{Q}(\beta)$). Thus, we have a maximal tower $\mathbb{Q} \subseteq \mathbb{Q}(i) \subseteq \mathbb{Q}(\alpha^2, i) \subseteq L$ of intermediary subfields. Next, consider $\mathbb{Q}(\alpha)$ which is the fixed field of the subgroup $< \tau >$ of order 2, where τ is defined by $\tau(i) = -i$, and $\tau(\alpha) = \alpha$ (in fact τ is complex conjugation), and which is degree 4 extension of $\mathbb{Q}$. Clearly, this is not a Galois extension of $\mathbb{Q}$, as such τ does not lie in the center. Therefore, the Galois group is neither abelian nor the quaternion group. It is, therefore, the dihedral group with presentation $< \sigma, \tau; \sigma^4 = I_L = \tau^2, \tau\sigma\tau = \sigma^3 >$. The fixed field of the subgroup $< \sigma^2, \tau >$ which is a normal subgroup of $G(L/\mathbb{Q})$ of order 4 is clearly, $\mathbb{Q}(\alpha^2)$. This gives us another maximal tower $\mathbb{Q} \subseteq Q(\alpha^2) \subseteq \mathbb{Q}(\alpha) \subseteq L$ of intermediary fields. Consider the element $\sigma^2\tau$. This element takes α to $-\alpha$ and i to $-i$. This is an element of order 2. The fixed field of $< \sigma^2\tau >$ is $\mathbb{Q}(\alpha i)$ which is a field extension of $\mathbb{Q}$ of degree 4, and it is isomorphic to $\mathbb{Q}(\alpha)$. Clearly, it contains $\mathbb{Q}(\alpha^2)$. This gives another maximal tower $\mathbb{Q} \subseteq Q(\alpha^2) \subseteq \mathbb{Q}(\alpha i) \subseteq L$. Similarly, looking at other towers of subgroups of the group $G(L/K)$, we can find all towers of intermediary subfields. This is left as an exercise.

Example 8.5.7 Let $K = \mathbb{C}(X)$ the field of fractions of $\mathbb{C}[X]$. Then, since X is a prime element in $\mathbb{C}[X]$, by the Eisenstein irreducibility criteria, $Y^n - X$ is irreducible in $K[Y]$. Let L be the splitting field of this polynomial over K. We determine the Galois group, and also all the intermediary fields. Let α be a root of $Y^n - X$. Let $\rho = e^{\frac{2\pi i}{n}}$ be a primitive nth root of unity. Then $\rho^i\alpha$, $1 \leq i \leq n$ are all roots of the polynomial $Y^n - X$. Since $\rho \in \mathbb{C} \subseteq K$, $L = K(\alpha)$ is the splitting field of $Y^n - X$. This shows that L is a Galois extension of K, and the Galois group is of order $[L : K] = n$. We show that this is a cyclic group of order n. Thus, there will be $\tau(n)$ (the number of divisors of n) subgroups of the Galois group, and so also the intermediary subfields. We determine them. We have an automorphism σ in $G(L/K)$ given by $\sigma(\alpha) = \rho\alpha$. Then $\sigma^i(\alpha) = \rho^i\alpha$. This shows that σ is an element of order n. Thus, $G(L/K) = < \sigma >$. Corresponding to any positive divisor m of n, there is a unique subgroup $< \sigma^r >$ of order m, where $n = mr$. Consider $\sigma^r(\alpha^k) = \rho^{rk}\alpha^k$. This shows that $K(\alpha^m)$ is contained in the fixed field $F(< \sigma^r >)$. Since α^m is a root of $Y^r - X$, and $Y^r - X$ is irreducible, it follows that $[K(\alpha^m) : K] = r = |< \sigma^r >|$. This shows that $K(\alpha^m)$ is the fixed field of $< \sigma^r >$. This determines all $\tau(n)$ intermediary fields.

Now, we prove the fundamental theorem of algebra as an application of the fundamental theorem of Galois Theory. The fundamental theorem of algebra states that the field $\mathbb{C}$ of complex numbers is algebraically closed. This result was first proved by Gauss. The routine proof is usually given in a standard complex analysis course using the fact that there is no bounded function which is analytic throughout the complex plane. We first prove some basic results in the form of Lemmas.

Lemma 8.5.8 *There is no proper odd degree extension of the field $\mathbb{R}$ of real numbers.*

Proof Let L be a field extension of the field $\mathbb{R}$ of reals such that $[L : \mathbb{R}]$ is odd. Since $\mathbb{R}$ is of characteristic 0, it is a separable and since every finite separable extension is simple, there is an element $a \in L$ such that $L = \mathbb{R}(a)$. Let $p(X)$ be the minimum polynomial of a. Then $p(X)$ is irreducible polynomial of odd degree greater than 1. It is sufficient therefore to show that no polynomial of odd degree greater than 1 is irreducible over $\mathbb{R}$. Let $f(X)$ be a polynomial of degree $2n + 1$ over $\mathbb{R}$. Then $Lim_{n\longrightarrow -\infty} \frac{f(X)}{X^{2n}} = -\infty$ and $Lim_{n\longrightarrow \infty} \frac{f(X)}{X^{2n}} = \infty$. Thus, there exists a such that $f(a) > 0$ and $f(-a) < 0$. By the intermediate value theorem, there is a $c \in \mathbb{R}$ such that $f(c) = o$, and so $f(X)$ can not be irreducible. ♯

Lemma 8.5.9 *There is no Galois extension L of the field $\mathbb{C}$ of complex numbers such that $[L : C] = 2^n, n \geq 1$.*

Proof Suppose the contrary. Let L be a Galois extension of $\mathbb{C}$ such that $| G(L/K) | = 2^n$, where $n \geq 2$. Then $G(L/K)$ has a maximal normal subgroup H of order 2^{n-1}. Since H is a normal subgroup, $F(H)$ is an extension of $\mathbb{C}$ of order 2. Suppose that $F(H) = \mathbb{C}(a)$. Then $a \notin \mathbb{C}$, and $a^2 \in \mathbb{C}$. This is impossible, for if $a^2 = re^{i\theta}$, then $a = \pm r^{\frac{1}{2}} e^{\frac{i\theta}{2}}$ belongs to $\mathbb{C}$. ♯

Theorem 8.5.10 (Fundamental Theorem of Algebra). *The field $\mathbb{C}$ of complex numbers is algebraically closed.*

Proof Let K be a finite extension of $\mathbb{C}$. We have to show that $K = \mathbb{C}$. Suppose not. Then $K = \mathbb{C}(a_1, a_2, \ldots, a_n) = \mathbb{R}(i, a_1, a_2, \ldots, a_n)$, and $[K : \mathbb{C}] \geq 2$. Let $f_i(X)$ be minimum polynomial of a_i over $\mathbb{R}$. Let $f(X) = (X^2 + 1)f_1(X)f_2(X) \ldots f_n(X)$. Let L be the splitting field of $f(X)$ over $\mathbb{R}$. Then L is a Galois extension of $\mathbb{R}$ containing K. Since $[L : \mathbb{R}] = [L : \mathbb{C}][\mathbb{C} : \mathbb{R}] = 2[L : \mathbb{C}]$ is even, $G(L/\mathbb{R})$ is of even order, and so it has a Sylow 2—subgroup H of order 2^m (say). Consider the fixed field $F(H)$ of H. Then $[F(H) : \mathbb{R}] = [G(L/R) : H]$ is odd. Since $\mathbb{R}$ has no proper extension of odd degree, we have $F(H) = \mathbb{R}$. This means that $G(L/\mathbb{R}) = H$ is a $2-$ group. Hence $G(L/\mathbb{C}) \subseteq G(L/\mathbb{R})$ is also of order 2^r for some $r \geq 1$. This is a contradiction to the above lemma. ♯

Exercises

8.5.1 Let K be a field, and $f(X) \in K[X]$. Find the splitting field L, the Galois group $G(L/K)$, and all intermediary subfields in each of the following cases.

(i) $K = \mathbb{Q}$, $f(X) = X^4 - 11$.
(ii) $K = \mathbb{Q}$, $f(X) = X^8 - 10$.
(iii) $K = \mathbb{Z}_5$, $f(X) = X^4 - 2$.
(iv) $K = \mathbb{Z}_2$, $f(X) = X^3 + X + 1$.
(v) $K = \mathbb{Q}$, $f(X) = X^5 - 11$.

8.5.2 Find the Galois group of $L = \mathbb{Q}(\sqrt{2}, \sqrt{3})$ over $\mathbb{Q}$. Find all intermediary fields.

8.5.3 Let L be a Galois extension of K with Galois group $\mathbb{Z}_{100}$. Find the number of intermediary subfields and also find them.

8.5.4 Let L be a finite Galois extension of K and $L = K(a)$. Show that $\{\sigma(a) \mid \sigma \in G(L/K)\}$ is a basis of L over K (This result is known as normal basis theorem).

8.5.5 Let L be a finite Galois extension of K. Let $L_1 \subseteq L_2$ be intermediary fields which are Galois extensions of K. Show that $G(L_1/K)$ is isomorphic to $G(L_2/K)/G(L_2/L_1)$.

8.5.6 Let L_1 and L_2 be intermediary fields of a Galois extension L of K. Suppose that L_2 is Galois extension of K. Show that L_1L_2 (the smallest subfield of L containing L_1 and also L_2) is a Galois extension of L_1, and $G(L_1L_2/L_1)$ is isomorphic to $G(L_2/L_1 \bigcap L_2)$.

8.5.7 Let L be a Galois extension of K. Let

$$K = K_1 \subseteq K_2 \subseteq \cdots \subseteq K_n = L \cdots \tag{8.5.1}$$

and

$$K = L_1 \subseteq L_2 \subseteq \cdots \subseteq L_m = L \cdots\cdots \tag{8.5.2}$$

be two towers of intermediary extensions of the Galois extension L of K, where K_{i+1} is a Galois extension of K_i, and L_{j+1} is a Galois extension of L_j for all i and j. Show that there are refinements

$$K = F_1 \subseteq F_2 \subseteq \cdots \subseteq F_t = L$$

and

$$K = F_1' \subseteq F_2' \subseteq \cdots \subseteq F_t = L$$

of 1 and 2 respectively such that after some rearrangement $G(F_{i+1}/F_i)$ is isomorphic to $G(F'_{i+1}/F'_i)$ for all i.

8.5.8 State and prove the analogue of the Jordan Holder theorem for towers of intermediary subfields of a Galois extensions.

8.5.9 Let L be a finite Galois extension of K with no proper intermediary fields. Show that $G(L/K)$ is a cyclic group of prime order.

8.5.10 Let L be a finite Galois extension with $G(L/K)$ simple. Let F be an intermediary field such that $K \neq F \neq L$. Show that there is a K-automorphism σ of L such that $\sigma(F) \neq F$.

8.5.11 Let L be a Galois extension of degree 15 over a field K. Find the number of intermediary fields.

8.5.12 Let L be a finite Galois extension of K, and H a subgroup of $G(L/K)$. Show that $F(N(H))$ is the smallest subfield contained in $F(H)$ such that $F(H)$ is a Galois extension of $F(N(H))$.

8.5.13 Let L be a finite Galois extension of K of degree 35. Let F be any intermediary field. Show that F is Galois extension of K.

8.5.14 Let L be a finite extension of K such that p^n is the largest power of prime p dividing $[L : K]$. Show that any two intermediary fields F_1 and F_2 such that $[F_1 : K] = [F_2 : K] = p^n$ are K-isomorphic. If m is the number of intermediary fields which are extensions of degrees p^n, then show that m divides $[L : K]$, and it is of the form $1 + kp$.

8.5.15 Let L be a finite Galois extension of K, and H be a subgroup of $G(L/K)$. Let F be the normal closure of $F(H)$.
Show that $G(L/F) = \bigcap_{\sigma \in G(L/K)} \sigma F(H) \sigma^{-1}$.

8.5.16 Let L be a finite Galois extension of K, and H a subgroup of $G(L/K)$. Let $F(H) = K(a)$, and S a left transversal of $G(L/K)$ modulo H. Show that

$$\prod\nolimits_{\sigma \in S} (X - \sigma(a))$$

is the minimum polynomial of a. Deduce that

$$\prod\nolimits_{\sigma \in G(L/K)} (X - \sigma(a)) = p(X)^{\frac{n}{r}},$$

where n is the degree $[L : K]$, r is the index of H in $G(L/K)$, and $p(X)$ is the minimum polynomial of a over K. Further, deduce the normal basis theorem (Exercise 8.5.4).

8.5.17 Let F be a field of characteristic $p \neq 0$. Let $L = F(X, Y)$ be the field of fractions of $F[X, Y]$ and $K = F(X^p, Y^p)$. Show that $[L : K] = p^2$ is a finite extension which is not simple. Show that there are infinitely many intermediary fields.

8.6 Cyclotomic Extensions

Definition 8.6.1 A finite extension L of K is called a **cyclotomic extension** of K if there is an element $\xi \in L$ such that $L = K(\xi)$, and $\xi^n = 1$ for some n. A Galois extension L of K is called an **abelian** extension if the Galois group $G(L/K)$ is abelian. It is called **cyclic** if the Galois group is cyclic.

A root of $X^n - 1$ in a field K is called a **nth root** of unity. An element $\alpha \in K$ is called **primitive nth root** of unity if its order in the multiplicative group $K^\star$ is n. It follows that any root of unity is a primitive root of unity for some n.

Let K be a field. Consider the polynomial $X^n - 1$ in $K[X]$. Suppose first that characteristic of K is $p \neq 0$. Suppose that p divides n and $n = p^r m$, where p and m are co-prime. Then $X^n - 1 = (X^m - 1)^{p^r}$. Thus, the roots of $X^n - 1$ and that of $X^m - 1$ in any extension field are same. In other words, the splitting field of $X^n - 1$, and the splitting field of $X^m - 1$ are same. Let L be the splitting field of $X^n - 1$, and so also of $X^m - 1$. Since p does not divide m, all the roots of $X^m - 1$ are distinct. Let G be the group of roots $X^m - 1$ in L. Since it is a finite subgroup of $L^\star$ of order m, it is a cyclic group of order m. A generator ξ_m of G is a primitive mth root of unity. Thus, $G = \{\xi_m^r \mid 0 \leq r \leq m-1\}$.

Note that if ξ_m is a primitive mth root of unity, then any mth primitive root of unity is of the form ξ_m^r, where $1 \leq r \leq m-1$ and $(r, m) = 1$. Clearly, $L = K(\xi_m)$. Any K-automorphism η of L is uniquely determined by its effect on ξ_m, and whose restriction to G is an automorphism of G. This defines an injective group homomorphism from the Galois group $G(L/K)$ to $Aut(G)$. From the theory of cyclic groups, it follows that $Aut(G)$ is isomorphic to the group U_m of prime residue classes modulo m.

Next, if the characteristic of K is 0, then all roots of $X^n - 1$ are distinct, and $e^{\frac{2\pi i}{n}}$ is a primitive nth root of unity. The above argument is valid when m is replaced by n. We can be summarize the above discussion in the following theorem.

Theorem 8.6.2 *Let K be a field, and ξ be a root of unity. Then $K(\xi)$ is a Galois extension of K. Further, if ξ_m is a primitive mth root of unity, then $K(\xi_m)$ is an Abelian extension of K, and $G(K(\xi_m)/K)$ is isomorphic to a subgroup of U_m. In particular, $[K(\xi_m) : K]$ divides $\phi(m)$.* ♯

In general, the Galois group $G(K(\xi_m)/K)$ need not be exactly U_m. It, in general, depends on the field K. For example, $G(\mathbb{Q}(i)(\xi_8)/\mathbb{Q}(i))$ is a subgroup of order 2 of the group U_8, and U_8 is of order 4. Note that $G(\mathbb{Q}(\xi_8)/\mathbb{Q})$ is isomorphic to U_8.

Remark 8.6.3 In general, minimum polynomials of distinct primitive nth roots of unities are distinct over fields of characteristic $p \neq 0$. For example, we have the factorization

$$X^7 - 1 = (X - 1)(X^3 + X + 1)(X^3 + X^2 + 1)$$

over $\mathbb{Z}_2$. Clearly, $X^3 + X + 1$ is the minimum polynomial of three of its roots in the splitting field which are all primitive 7th roots of unity, and similarly, $X^3 + X^2 + 1$

is minimum polynomial of the rest of the primitive 7th roots of unity. We shall see below that all the primitive nth roots of unity over $\mathbb{Q}$ have same minimum polynomial, and we shall describe it.

Definition 8.6.4 Let $\{\rho_1, \rho_2, \ldots, \rho_{\phi(n)}\}$ denote the set of all primitive nth roots of unity in the field $\mathbb{C}$ of complex numbers. The polynomial

$$\phi_n(X) = \prod_{i=1}^{\phi(n)} (X - \rho_i)$$

is called the **nth cyclotomic** polynomial.

Thus, the degree of $\phi_n(X) = \phi(n)$. Further, since every nth root of unity is a primitive dth root of unity for a unique positive divisor of n, we have

Proposition 8.6.5 $X^n - 1 = \prod_{d/n} \phi_d(X)$. ♯

Proposition 8.6.6 $\phi_n(X) \in \mathbb{Z}[X]$.

Proof Consider $L = \mathbb{Q}(\rho_1)$, where ρ_1 is a primitive nth root of unity over $\mathbb{Q}$. Then L is the splitting field of $X^n - 1$ over $\mathbb{Q}$. Since any $\mathbb{Q}$-automorphism of L takes a primitive root to a primitive root, it follows that $\phi_n(X)$ is fixed by all members of the Galois group $G(L/\mathbb{Q})$. Since L is a Galois extension of $\mathbb{Q}$, it follows that $\phi_n(X)$ is a polynomial in $\mathbb{Q}[X]$ whose leading coefficient is 1. First, observe that if $f(X), g(X) \in \mathbb{Z}[X]$, and $h(X) \in \mathbb{Q}[X]$ are polynomials with leading coefficient 1 such that $f(X) = g(X)h(X)$, then $h(X) \in \mathbb{Z}[X]$. Now, we prove, by the induction on n, that $\phi_n(X) \in \mathbb{Z}[X]$ for each n. Clearly, $\phi_1(X) = X + 1 \in \mathbb{Z}[X]$. Assume that the result holds for all m less than n. From the previous proposition, we have

$$X^n - 1 = (\prod_{d/n, d<n} \phi_d(X))\phi_n(X).$$

The left hand side is a monic polynomial in $\mathbb{Z}[X]$, and by the induction hypothesis the first factor in the RHS (being product of monic polynomials in $\mathbb{Z}[X]$) is a monic polynomial in $\mathbb{Z}[X]$. From the earlier observation, $\phi_n(X) \in \mathbb{Z}[X]$. ♯

Example 8.6.7 The above proposition gives inductive procedure to find nth cyclotomic polynomial. We illustrate it in this example. If p is prime, then all pth roots of 1 except 1 are primitive roots. Thus,

$$\phi_p(X) = \frac{X^p - 1}{X - 1} = 1 + X + X^2 + \cdots + X^{p-1}.$$

Clearly, $\phi_1(X) = X - 1$, $\phi_2(X) = X + 1$, $\phi_3(X) = X^2 + X + 1$. Hence

$$X^6 - 1 = (X - 1)(X + 1)(X^2 + X + 1)\phi_6(X).$$

Thus,

$$\phi_6(X) = \frac{X^6 - 1}{(X^2 - 1)(X^2 + X + 1)}.$$

Theorem 8.6.8 *$\phi_n(X)$ is irreducible over $\mathbb{Q}$ for each positive integer n.*

Proof Suppose the contrary. Then $\phi_n(X)$ is reducible in $\mathbb{Q}[X]$. Since $\phi_n(X)$ is a monic polynomial in $\mathbb{Z}[X]$, it follows that $\phi_n(X)$ is reducible in $\mathbb{Z}[X]$. Let $f(X)$ be an monic irreducible factor of $\phi_n(X)$ in $\mathbb{Z}[X]$. Suppose that $\phi_n(X) = f(X)g(X)$, where $g(X)$ is also a monic polynomial in $\mathbb{Z}[X]$ of positive degree. Let ρ be a root of $f(X)$ in the splitting field. Then ρ is a primitive nth root of 1. We show that ρ^r are roots of $f(X)$ for all r co-prime to n. The proof is by the induction on r. If $r = 1$, then there is nothing to prove. Assume that the statement is true for all t less than $r \geq 2$ and co-prime to n. Let p be a prime dividing r and $r = ps$. Then since r is co-prime to n, it follows that p does not divide n. Thus, ρ^p is also a primitive nth root of 1, and so it is a root of $\phi_n(X)$. We claim that it is a root of $f(X)$. Suppose not. Then ρ^p is a root of $g(X)$. But then ρ is a root of $g(X^p)$. Since $f(X)$ is the minimum polynomial of ρ, it follows that $f(X)$ divides $g(X^p)$ in $\mathbb{Q}[X]$. Since $f(X)$ is monic, it follows that $f(X)$ divides $g(X^p)$ in $\mathbb{Z}[X]$. Reducing it modulo p, we see that $\overline{f(X)}$ divides $\overline{g(X^p)} = (\overline{g(X)})^p$ in $\mathbb{Z}_p[X]$. Let $\overline{h(X)}$ be an irreducible factor of $\overline{f(X)}$ (note that $\overline{f(X)}$ in $\mathbb{Z}_p[X]$ is neither zero nor a unit). Then $\overline{h(X)}$ divides $\overline{g(X)}$ in $\mathbb{Z}_p[X]$. This means that $\overline{h(X)}^2$ divides $\overline{\phi_n(X)} = \overline{f(X)g(X)}$. But, then $\overline{\phi_n(X)}$ will have repeated roots in splitting field of $X^n - \overline{1}$ over $\mathbb{Z}_p[X]$. Since $\overline{\phi_n(X)}$ divides $X^n - \overline{1}$ in $\mathbb{Z}_p[X]$, it follows that $X^n - \overline{1}$ has repeated roots in its splitting field. This is impossible for p does not divide n. Thus, we see that ρ^p is a root of $f(X)$. By the induction hypothesis $\rho^r = \rho^{ps} = (\rho^p)^s$ is a root of $f(X)$. Thus, all roots of $\phi_n(X)$ are also roots of $f(X)$, and since all the roots are distinct, $f(X) = \phi_n(X)$. This is a self contradiction. ♯

Corollary 8.6.9 *All the primitive nth roots of* 1 *have same minimum polynomials over $\mathbb{Q}$. If ρ is a primitive nth root of* 1*, then $\mathbb{Q}(\rho)$ is a Galois extension of $\mathbb{Q}$ with Galois group U_n. In particular, $\mathbb{Q}(\rho)$ is an abelian extension, and $[\mathbb{Q}(\rho) : \mathbb{Q}] = \phi(n)$.*

Proof We have already seen that $\mathbb{Q}(\rho)$ is a Galois extension with Galois group isomorphic to a subgroup of U_n. From the above result it follows that $[\mathbb{Q}(\rho) : \mathbb{Q}] = \phi(n) = | U_n |$. The result follows. ♯

Now, we shall describe the cyclic extensions. Recall that a Galois extension L of K is called a cyclic extension if the Galois group is cyclic. Since the structure of a cyclic group is easy to describe, one can describe the intermediary fields easily. We shall first discuss cyclic extensions L of K of order n, where K contains a primitive nth root of 1.

Theorem 8.6.10 *Let K be a field which contains a primitive nth root ρ of* 1*. Let L be a Galois extension of K such that $G(L/K)$ is a cyclic group of order n generated by σ. Then there is a nonzero element $\alpha \in L$ such that $\rho = \sigma(\alpha)\alpha^{-1}$. Further, $L = K(\alpha)$, where $\alpha^n = \beta \in K$, i.e., $L = K(\sqrt[n]{\beta})$ for some $\beta \in K$.*

Proof We first show that $\rho \in K$ is an eigenvalue of σ, or equivalently, ρ is a root of the characteristic polynomial of σ. Since σ is of order n, it satisfies the polynomial $X^n - 1$. Further, by the Dedekind theorem $G(L/K) = \{I, \sigma, \sigma^2, \ldots, \sigma^{n-1}\}$ is linearly independent. Hence σ can not be a root of lower degree polynomial. This shows that $X^n - 1$ is the minimum polynomial of σ. Since the dimension of L over K is n, the characteristic polynomial is also $X^n - 1$. Since ρ is a primitive nth root of 1, it is a root of the characteristic polynomial of σ. Hence, $\sigma(\alpha) = \rho\alpha$ for some α. It follows that $\sigma^r(\alpha) = \rho^r\alpha = \alpha$ if and only if n divides r. This means that $G(L/K(\alpha)) = \{I\}$. This shows that $L = K(\alpha)$. Also $\sigma(\alpha^n) = (\sigma(\alpha))^n = \rho^n\alpha^n = \alpha^n$. This shows that $\alpha^n \in F(G(L/K)) = K$. Take $\beta = \alpha^n$. ♯

The following theorem is converse of the above theorem in some sense.

Theorem 8.6.11 *Let K be a field which contains a primitive nth root of* 1. *Let L be a finite extension of K. Let $\alpha \in L$ such that $\alpha^n \in K$. Then $K(\alpha)$ is a cyclic extension of K. Further, if r is the order of the Galois group of $K(\alpha)$ over K, then r divides n and $\alpha^r \in K$, i.e., $K(\alpha) = K(\sqrt[r]{\beta})$ for some β in K.*

Proof Suppose that $\alpha^n = a \in K$. Then α is a root of $X^n - a$. Since K has a primitive nth root of 1, it follows that characteristic of K does not divide n, and so all the roots of $X^n - a$ in its splitting field are distinct (the derivative of $X^n - a$ is nonzero). Let ρ be a primitive nth root of 1. Then each $\rho^i\alpha$ is a root of $X^n - a$. It follows that $K(\alpha)$ is the splitting field of $X^n - a$, and so it is a Galois extension. Let G denote the group of roots of 1 in K. Then G is a cyclic group of order n. Define a map η from $G(K(\alpha)/K)$ to G by $\eta(\sigma) = \sigma(\alpha)\alpha^{-1}$ (note that $\sigma(\alpha) = \rho^r\alpha$ for some r). It is easy to see that η is a homomorphism which is injective (a member of $G(K(\alpha)/K)$ is uniquely determined by its effect on α). Since subgroup of a cyclic group is cyclic, it follows that the Galois group of $K(\alpha)$ over K is cyclic, and it is isomorphic to a subgroup of G. Suppose that the order of the Galois group $G(K(\alpha)/K)$ is r. Then r divides n. Let σ be a generator of $G(K(\alpha)/K)$. Then $\sigma^m = I$ if and only if r divides m. Suppose that $\sigma(\alpha) = \rho^s\alpha$. Then ρ^s is of order r. This means that $s = \frac{n}{r}$. The result follows if we take $\beta = \rho^s\alpha$. Clearly ρ^s is primitive rth root of 1.

Theorem 8.6.12 (Artin-Schreier) *Let K be a field of characteristic $p \neq 0$. Let L be a cyclic Galois extension of K of degree p. Then there is $a \in K$ and $\alpha \in L$ such that $L = K(\alpha)$, where $\alpha^p - \alpha - a = 0$. Conversely, if there is $a \in K$ such that there is no $\alpha \in K$ such that $\alpha^p - \alpha = a$, then $X^p - X - a$ is irreducible in $K[X]$, and its splitting field is cyclic Galois extension of K of degree p.*

Proof Suppose that L is a cyclic Galois extension of K of degree p. Let σ be a generator of $G(L/K)$. Then $\sigma^p - I = 0$. Thus, σ satisfies the polynomial $X^p - 1$. Since $G(L/K) = \{I, \sigma, \sigma^2, \ldots, \sigma^{p-1}\}$ is linearly independent (Dedekind theorem), σ can not satisfy a polynomial of lower degree. In other words, $X^p - 1$ is the minimum polynomial of σ. Let T denote the linear transformation $\sigma - I$. Then $T^p = (\sigma - I)^p = \sigma^p - I = 0$. Thus, image$T^{p-1} \subseteq KerT$. $T^{p-1} \neq 0$, for otherwise σ will satisfy a polynomial of degree $p - 1$. Further, $KerT = \{a \in L \mid$

$0 = T(a) = \sigma(a) - a\} = \{a \in L \mid \sigma(a) = a\} = F(G(L/K)) = K$. Since image$T^{p-1}$ is a nonzero subspace of $KerT = K$, it follows that $imageT^{p-1} = K$. Let $\beta \in K$ such that $T^{p-1}(\beta) = 1$. Take $\alpha = T^{p-2}(\beta)$. Then $T(\alpha) = 1$, and so $\sigma(\alpha) = \alpha + 1$. Since α is not fixed by σ, it follows that $\alpha \notin K$. Since $[K(\alpha) : K]$ divide $[L : K] = p$, it follows that $L = K(\alpha)$. Further, $\sigma(\alpha^p - \alpha) = \sigma(\alpha)^p - \sigma(\alpha) = (\alpha + 1)^p - (\alpha + 1) = \alpha^p + 1 - \alpha - 1 = \alpha^p - \alpha$. This shows that $\alpha^p - \alpha$ belongs to K. Put $a = \alpha^p - \alpha$. Then $X^p - X - a$ is the minimum polynomial of α. Conversely, Let $a \in K$ be such that there is no $\beta \in K$ such that $\beta^p - \beta = a$. Let L be the splitting field of $f(X) = X^p - X - a$. Let $\alpha \in L$ be a root of $f(X)$. By the hypothesis $\alpha \notin K$. Since $i^p = i$ for all i in the prime field of K, we have p distinct roots $\alpha, \alpha + 1, \alpha + 2, \ldots, \alpha + p - 1$. Thus all the roots of $f(X)$ are in $K(\alpha)$. This shows that $L = K(\alpha)$. Now, we show that $f(X)$ is irreducible. Suppose not. Let $f(X) = p_1(X)p_2(X) \ldots p_m(X)$, $m > 1$ be the factorization of $f(X)$ as product of irreducible factors. Let α_i be a root of $p_i(X)$. Then it is also a root of $f(X)$. From the earlier argument, it follows that $L = K(\alpha_i)$ for each i. Thus, $degp_i(X) = [L : K]$. This shows that $p = degf(X) = [L : K]m$.This is a contradiction, for p is prime, $[L : K] > 1$ and $m > 1$. This shows that $f(X)$ is irreducible, and so $[L : K] = p$. Hence the Galois group of L over K is prime cyclic. ♯

At last, in this section, we prove a very important and useful result known as Hilbert theorem 90. First we have some definitions.

Let L be a finite field extension of K. Let $a \in L$. The map L_a from L to L defined by $L_a(x) = ax$ is a K—linear endomorphism of the vector space L over K.

Definition 8.6.13 Let L be a finite field extension of K. We define two functions $N_{L/K}$ and $T_{L/K}$ from L to K, the **norm** and the **trace** functions, by $N_{L/K}(a) = DetL_a$ and $T_{L/K}(a) = TraceL_a$.

It follows easily that $L_{ab} = L_aL_b$, $L_{a+b} = L_a + L_b$ and $L_{\alpha a} = \alpha L_a$ for all $a, b \in L$ and $\alpha \in K$. Since the determinant of product of two linear transformations is the product of their determinants, and the trace is a linear functional, it follows that $N_{L/K}(ab) = N_{L/K}(a)N_{L/K}(b)$, and $T_{L/K}$ is a linear functional on L. Also $L_{a^{-1}} = (L_a)^{-1}$, and so $N_{L/K}$ is a group homomorphism from $L^\star$ to $K^\star$.

Proposition 8.6.14 *Let L be a finite extension of degree n, and $a \in L$. Let $p(X)$ be the minimum polynomial of a over K. Then $p(X)$ is also the minimum polynomial of the linear transformation L_a. Further, if $degp(X) = m$, then the characteristic polynomial of L_a is $p(X)^{\frac{n}{m}}$.*

Proof The map η from L to $End_K(L)$ defined by $\eta(a) = L_a$ is easily seen to be an injective algebra homomorphism (observe that both sides are algebras over K). This shows that the minimum polynomial of a over K is same as the minimum polynomial of L_a. If $\chi(X)$ is the characteristic polynomial of L_a, then it is a fact of elementary linear algebra (Cayley Hamilton Theorem) that the minimum polynomial $p(X)$ of L_a divides the characteristic polynomial, and they have same irreducible factors. This shows that $\chi(X) = p(X)^r$ for some r. Comparing the degrees, and observing that

the degree of the characteristic polynomial is same as the dimension n of the vector space L over K, we obtain that $\chi(X) = p(X)^{\frac{n}{m}}$. ♯

Theorem 8.6.15 *Let L be a finite Galois extension of K of degree n. Let $a \in L$. Then the characteristic polynomial $\chi(X)$ of L_a is given by*

$$\chi(X) = \prod_{\sigma \in G(L/K)} (X - \sigma(a)).$$

Proof Let us denote the polynomial $\prod_{\sigma \in G(L/K)}(X - \sigma(a))$ by $\psi(X)$. Since a is a root of $\psi(X)$, it follows that the minimum polynomial $p(X)$ of a divides $\psi(X)$. Also given any member $\sigma \in G(L/K)$, $p^{\sigma}(X) = p(X)$, and so $\sigma(a)$ is a root of $p(X)$ for all $\sigma \in G(L/K)$. In other words, each $\sigma(a)$ has same minimum polynomial $p(X)$. This also says that the only irreducible factor of $\psi(X)$ is $p(X)$. Comparing the degrees, we obtain that $\psi(X) = p(X)^{\frac{n}{m}}$. From the previous proposition, it follows that $\chi(X) = \psi(X)$. ♯

Corollary 8.6.16 *Let L be a finite Galois extension of K, and $a \in L$. Then*

$$N_{L/K}(a) = \prod_{\sigma \in G(L/K)} \sigma(a),$$

and

$$T_{L/K}(a) = \sum_{\sigma \in G(L/K)} \sigma(a).$$

Proof Since the determinant of a linear transformation is product of the characteristic roots, and the trace is the sum of the characteristic roots, the result follows from the above theorem. ♯

Definition 8.6.17 Let L be a Galois extension of K. A map f from $G(L/K)$ to $L^{\star}$ is called a 1-**cocycle** of $G(L/K)$ in $L^{\star}$ if f satisfies the following **Emmy Noether** equation.

$$f(\sigma\tau) = f(\sigma)\sigma(f(\tau)) \text{ for all } \sigma, \tau \in G(L/K).$$

A 1-cocycle is also called a **crossed** homomorphism:

Proposition 8.6.18 *Let L be a finite Galois extension of K, and f be a 1-cocycle of $G(L/K)$ in $L^{\star}$. Then there is an element $a \in L$ such that $f(\sigma) = \sigma(a)a^{-1}$ for all $\sigma \in G(L/K)$.*

Proof Since $f(\sigma) \neq 0$ *for all* $\sigma \in G(L/K)$, and by the Dedekind theorem, $G(L/K)$ is linearly independent, it follows that $\Sigma_{\sigma \in G(L/K)} f(\sigma)\sigma \neq 0$. Hence there is an element $b \in L$ such that $\Sigma_{\sigma \in G(L/K)} f(\sigma)\sigma(b) \neq 0$. Let $a = (\Sigma_{\sigma \in G(L/K)} f(\sigma)\sigma(b))^{-1}$. Since f is a 1-cocycle, for any $\tau \in G(L/K)$, we have

$$f(\tau)(\tau(a))^{-1} = f(\tau)\tau(a^{-1}) = f(\tau)\tau(\sum_{\sigma\in G(L/K)} f(\sigma)\sigma(b)) =$$
$$\sum_{\sigma\in G(L/K)} f(\tau)\tau(f(\sigma))(\tau\sigma)(b) = \sum_{\sigma\in G(L/K)} f(\tau\sigma)(\tau\sigma)(b) = a^{-1}.$$

This shows that $f(\tau) = \tau(a)a^{-1}$ for all $\tau \in G(L/K)$. ♯

Theorem 8.6.19 (Hilbert Theorem 90) *Let L be a cyclic Galois extension of K of degree n, and σ be a generator of $G(L/K)$. Then the kernel of the homomorphism $N_{L/K}$ from $L^\star$ to $K^\star$ is $\{\sigma(a)a^{-1} \mid a \in L^\star\}$.*

Proof It follows from Corollary 8.6.16 that $N_{L/K}(a) = N_{L/K}(\sigma(a))$ for all $a \in L$ and $\sigma \in G(L/K)$. Hence $\sigma(a)a^{-1}$ is in the kernel of $N_{L/K}$ for all $a \in L^\star$ and $\sigma \in G(L/K)$.

Conversely, suppose that u belongs to the kernel of $N_{L/K}$. Then $N_{L/K}(u) = 1$. Define a map f from $G(L/K)$ to $L^\star$ by $f(I_L) = 1$, and for $1 \leq i \leq n-1$, we define $f(\sigma^i) = u\sigma(u)\sigma^2(u)\ldots\sigma^{i-1}(u)$. We show that f is a 1-cocycle of $G(L/K)$ in $L^\star$. Suppose that $0 \leq i, j \leq n-1$. There are two cases: (i) $i+j \leq n-1$ and (ii) $i+j \geq n$. In case (i)

$$f(\sigma^i\sigma^j) = f(\sigma^{i+j}) = u\sigma(u)\sigma^2(u)\ldots\sigma^{i+j-1} = f(\sigma^i)\sigma^i(f(\sigma^j)).$$

Now, consider the case (ii). In this case

$$f(\sigma^i\sigma^j) = f(\sigma^{i+j-n}) = u\sigma(u)\sigma^2(u)\ldots\sigma^{i+j-n-1}(u).$$

Also

$$f(\sigma^i)\sigma^i(f(\sigma^j)) = u\sigma(u)\sigma^2(u)\ldots\sigma^{i-1}(u)\sigma^i(u\sigma(u)\ldots\sigma^{j-1}(u)) =$$
$$u\sigma(u)\ldots\sigma^{i+j-n-1}(u)\sigma^{i+j-n}(u\sigma(u)\ldots\sigma^{n-1}(u)) =$$
$$f(\sigma^i\sigma^j)\sigma^{i+j-n}(N_{L/K}(u)) = f(\sigma^i\sigma^j).$$

This shows that f is a 1-cocycle. From the earlier proposition, there is a $a \in L$ such that $f(\sigma^i) = \sigma^i(a)a^{-1}$. In particular, $u = f(\sigma) = \sigma(a)a^{-1}$. ♯

Exercises

8.6.1 Find out a primitive 16th root of 1, and also the minimum polynomial of a primitive root of 1 over $\mathbb{Q}$. What is the degree of the corresponding cyclotomic extension over $\mathbb{Q}$. Find also the intermediary subfields.

8.6.2 Let $\mathbb{Q}_n$ denote the splitting field of $X^n - 1$ over $\mathbb{Q}$, and $\mathbb{Q}_m$ the spitting field of $X^m - 1$. Show that $\mathbb{Q}_n \bigcap \mathbb{Q}_m = \mathbb{Q}_d$, where d is the g.c.d of n and m.

8.6.3 Show that the nth cyclotomic polynomial $\phi_n(X)$ over $\mathbb{Q}$ is also given by

$$\phi_n(X) = \prod_{d/n}(X^{\frac{n}{d}} - 1)^{\mu(d)},$$

where μ is the Mobius function.
Hint. Use the Mobius inversion formula and Proposition 8.6.5.

8.6.4 Find all subfields of $\mathbb{Q}_{20}$.

8.6.5 Show that $cos\frac{2\pi}{n}$ and $sin\frac{2\pi}{n}$ are both algebraic numbers.

8.6.6 Is $\mathbb{Q}(cos\frac{2\pi}{n})$ Galois over $\mathbb{Q}$. If yes, what is the Galois group?

8.6.7 Let L be a Galois extension of K, and F be an intermediary field such that F is a Galois extension of K. Show that $N_{L/K} = N_{F/K} o N_{L/F}$ and $T_{L/K} = T_{F/K} o T_{L/F}$.

8.6.8 Use the theorem Hilbert 90 to prove Theorem 8.6.10.

8.6.9 Let L be a cyclic Galois extension of K. A map f from $G(L/K)$ to L is called a 1-cocycle of $G(L/K)$ in L if $f(\sigma\tau) = f(\sigma) + \sigma(f(\tau))$ for all $\sigma, \tau \in G(L/K)$. Use the methods of the Proposition 8.6.18, and Theorem 8.6.12 to prove that the null space of $T_{L/K}$ is $\{\sigma(a) - a \mid a \in L\}$. This is known as the additive Hilbert Theorem 90.

8.7 Geometric Constructions

We shall be basically interested in problem of constructions using straight edge and compass.

First, we try to understand the meaning of geometric constructions. We start with two points O and P in a plane and take the length of the segment OP as unit. We draw the line OP indefinitely and take it as X axis with O as origin and P as marked point. We construct points, lines, and circles inductively. At each point of construction, we draw a line through an already constructed point, or draw a circle with center as one of the constructed point and radius as segment through that point and another constructed point, and then take the intersection of newly constructed line or circle with already existing lines and circles to construct new points. We can construct a line through O and perpendicular to the line OP to draw Y axis. By drawing a circle with center as O and radius OP, we get a point Q on Y axis whose coordinate is (0,1). Thus, by drawing perpendicular to OQ at Q and to OP at P, and then taking their intersection we determine a point whose coordinate is (1, 1). This is how, we proceed, and do the constructions. A real number a is said to be **constructible** if we can construct two points which are at a distance $\mid a \mid$ apart.

We recall some standard school level geometric constructions by ruler and compass.

(i). We can draw a line perpendicular to a line from any point on that line.
(ii). We can draw perpendicular to a line from any point outside the line.
(iii). We can draw a perpendicular bisector to any segment of a line.

(iv). Given any segment of a line we can draw an equilateral triangle with the given segment as a base. In particular, we can construct a 60° angle.

(v). Given a quadrilateral, we can construct a triangle which has the same area as the given quadrilateral.

(vi). Given segments of lengths a and b units, we can construct segments of lengths $a+b,\ a-b, a\ >\ 0,\ ab$, and a/b. In particular, given a unit segment, we can construct a segment of length $r\ >\ o$, where r is a rational number.

(vii). Given any segment of length a, we can construct a segment of length $\sqrt{a}$. In particular, given segments of lengths $|\ a\ |$ and $|\ b\ |$, we can construct segments of lengths $|\ \alpha\ |$ and $|\ \beta\ |$ where α and β are solutions of the equation $X^2\ -\ (a+b)X\ +\ ab$.

If a point (a, b) is constructible, then drawing perpendicular from that point to X-axis and also to Y-axis, we see that $(a, 0)$ and $(0, b)$ are also constructible points. Conversely, if $(a, 0)$ and $(0, b)$ are constructible points, then drawing perpendicular on X-axis through $(a, 0)$ and on Y-axis through $(0, b)$, and then taking the intersection, we construct the point (a, b). Next, suppose that the points U and V are constructible points, and the length of the segment UV is l. Then we can construct the point $(l, 0)$ (and also $(0, l)$) as follows: We can draw a parallelogram with sides OU and UV. Let W be the other vertex of the parallelogram. Then the length of the segment OW is l. Draw a circle with center O and radius on the segment OW. The point of the intersection of this circle with X axis is the point $(l, 0)$. The above discussion concludes to the following proposition.

Proposition 8.7.1 *Let L denote the set of all constructible numbers. Then $L \times L$ is the set of all constructible points. Further, L is subfield of $\mathbb{R}$, and every positive member of L has a square root in L. In other words, there is no proper real quadratic extension of L.* ♯

Proposition 8.7.2 *Let a be a positive real number such that there is a tower $Q\ =\ K_0 \subseteq K_1 \subseteq K_2 \subseteq\ \cdots \subseteq K_n$ of extensions such that $K_n \subseteq \mathbb{R}$, $[K_{i+1} : K_i] \leq 2$, and $a \in K_n$. Then a is constructible.*

Proof The proof is by the induction on n. If $n\ =\ 0$, then $a \in Q$, and since the set of all constructible numbers is a subfield, a is constructible. Assume the result for n. Let $a \in K_{n+1}$, where $Q\ =\ K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots\ \subseteq K_n \subseteq K_{n+1}$ is a tower of extensions such that $[K_{i+1} : K_i] \leq 2$. By the induction hypothesis each member of K_n is constructible. Since $[K_{n+1} : K_n] \leq 2$, $K_{n+1}\ =\ K_n$, or else $[K_{n+1} : K_n]\ =\ 2$. If $a \in K_n$, then there is nothing to do. Suppose that $a \notin K_n$, then $[K_{n+1} : K_n]\ =\ 2$, and so $a\ =\ \sqrt{b}$ for some $b \in K_n$. By the induction hypothesis b is constructible. Hence from our basic constructions a is also constructible. ♯

Now, we aim to prove the converse of the above proposition. First, we have some definitions.

Definition 8.7.3 Let K be a subfield of $\mathbb{R}$. An element of $K \times K$ will be called a **point** in a plane of K. A line passing through two points of the plane of K is called

a **line** in the plane of K. A circle with center in the plane of K, and radius a positive member of K, is called a **circle** in the plane of K

Proposition 8.7.4 *A line $aX + bY = c$ in $\mathbb{R}$ is a line in the plane of K if and only if there is a $\lambda \in \mathbb{R}$ such that $\lambda a, \lambda b, \lambda c \in K$.*

Proof Suppose that $aX + bY = c$ is a line in the plane of K. Then there are distinct points (α_1, β_1) and (α_2, β_2) in the plane of K which lie on the line. Thus,

$$a\alpha_1 + b\beta_1 = c = a\alpha_2 + b\beta_2.$$

Then $a(\alpha_1 - \alpha_2) = b(\beta_2 - \beta_1)$. Since the points are distinct, $\alpha_1 \neq \alpha_2$ or $\beta_1 \neq \beta_2$. Suppose that $\alpha_1 \neq \alpha_2$. Then $a = b\frac{\beta_2-\beta_1}{\alpha_1-\alpha_2}$ and $c = b\frac{\beta_2-\beta_1}{\alpha_1-\alpha_2} + b\beta_2$. $b \neq 0$, for otherwise $a = 0 = c$ which is not possible. Now, it is clear that $\lambda = b^{-1}$ will serve the purpose. For converse, we may assume that $a, b, c \in K$. Suppose that $a \neq 0 \neq b$. Then, the line passes through the points $(ca^{-1}, 0)$ and $(0, cb^{-1})$ in the plane of K. Suppose that $a = 0 \neq b$. Then the line passes through $(0, cb^{-1})$ and $(1, cb^{-1})$ which are in the plane of K. Similarly, if $a \neq 0 = b$, then the line passes through the points $(ca^{-1}, 0)$, and $(ca^{-1}, 1)$ in the plane of K. ♯

Proposition 8.7.5 *Suppose that $X^2 + Y^2 + 2uX + 2vY + w = 0$ is a circle in K. Then $u, v, w \in K$. Conversely, suppose that $u, v, w \in K$. Then it represents a circle with center in the plane of K, and the radius in $K(\sqrt{u^2 + v^2 - w})$.*

Proof If the given circle is in the plane of K, then the center $(-u, -v)$ is a point in the plane of K, and also the radius $\sqrt{u^2 + v^2 - w}$ belongs to K. This means that u, v and $u^2 + v^2 - w$ belongs to K, and so $u, v, w \in K$. Conversely, suppose that $u, v, w \in K$. Then the center $(-u, -v)$ is in the plane of K, and the radius is $\sqrt{u+v^2 - w}$. ♯

Proposition 8.7.6 *If two lines in the plane of K intersect, they intersect in the plane of K.* ♯

Proof Let $aX + bY = c$ and $a'X + b'Y = c'$ be two lines in the plane of K. We may assume that all the coefficients are in K. Since these lines are not parallel, the simultaneous equation has a unique solution in terms of the coefficients which are in K. Thus, these two lines intersect in the plane of K. ♯

Proposition 8.7.7 *The intersection of a line in the plane of K and a circle in the plane of K is either empty set, or a point in the plane of K, or it consists of two points in the plane of $K\sqrt{d}$ for some positive d in K.*

Proof Let $aX + bY = c$ be a line in the plane of K, and

$$X^2 + Y^2 + uX + vY + w = 0$$

a circle in the plane of K. We may assume that all the coefficients are in K and $b \neq 0$. Substituting $y = \frac{(-aX + c)}{b}$ in the equation of the circle, and then solving, we get that either the solutions are imaginary, and in this case they do not intersect, or they intersect at one point in the plane of K (this is when the discriminant of the quadratic equation obtained is 0), or it intersects at two points in the plane of $K(\sqrt{d})$, where d is the discriminant of the quadratic equation obtained after the substitution of the value $y = \frac{-aX + c}{b}$ of y in the equation of the circle. ♯

Proposition 8.7.8 *Given two circles in the plane of K one and only one of the following holds:*

(i). They do not intersect.
(ii). They touch each other at a point in the plane of K.
(iii). They intersect at two points in the plane of $K(\sqrt{d})$ for some positive $d \in K$.

Proof Let

$$X^2 + Y^2 + 2uX + 2vY + w = 0$$

and

$$X^2 + Y^2 + 2u'X + 2v'Y + w' = 0$$

be two circles in the plane of K. Then u, v, w, u', v', w' belong to K. The intersection of these circles is the same as the intersection of the plane

$$(u - u')X + (v - v')Y + w - w' = 0$$

with any of the two given circles. Now, the result follows from the previous proposition. ♯

Theorem 8.7.9 *A real number a is constructible if and only if there is a tower*

$$\mathbb{Q} = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_n$$

of field extensions such that $a \in K_n$ and $[K_{i+1} : K_i] \leq 2$.

Proof From the Proposition 8.7.2, it follows that if a lies in such a tower, then it is constructible. We prove the converse. Suppose that a is constructible. Then the point $(a, 0)$ can be constructed from a finite number of steps starting from the points in the plane of $\mathbb{Q}$. This is obtained by taking intersections of constructible lines and circles starting from lines and circles in the plane of $\mathbb{Q}$. In the first step the point will lie in the plane of $\mathbb{Q}$ or in the plane of $K_1 = \mathbb{Q}(\sqrt{a_1})$ for some positive $a_1 \in \mathbb{Q}$. In the second step the point will lie in the plane of K_1 or in the plane of $K_2 = K_1(a_2)$. Proceeding inductively we arrive at the result. ♯

Corollary 8.7.10 *A real number a is constructible only if $[\mathbb{Q}(a) : \mathbb{Q}]$ is a power of 2.*

Proof Suppose that a is constructible. From the above theorem, $a \in K_n$, where $[K_n : \mathbb{Q}] = 2^m$ for some m. Since $[\mathbb{Q}(a) : \mathbb{Q}]$ divides $[K_n : \mathbb{Q}]$, the result follows. ♯

We are now ready to answer the classical problems on geometric constructions.

Proposition 8.7.11 *An angle θ is constructible if and only if $cos\theta$, or equivalently, $sin\theta$ is constructible.*

Proof Suppose that $cos\theta$ is constructible. Then we can construct the point $P(cos\theta, 0)$. Draw a perpendicular line to the X-axis from this point, and also a unit circle with center origin. Let R be the point of intersection of the line with this unit circle. Then the angle $< POR$ is the required angle. Conversely, suppose that the angle θ is constructible. Let OQ be a line making angle θ with X-axis. Draw a unit circle with O as center, and let R be the intersection of this circle with line OQ. Draw the perpendicular from R to X-axis which meats it at a point P (say). Then OP is a segment of length $cos\theta$. ♯

Theorem 8.7.12 *It is impossible to construct* 20° *angle using ruler and compass only.*

Proof From the previous result, if we can construct 20° angle, then $a = cos20^\circ$ is constructible. Now

$$\frac{1}{2} = cos60^\circ = 4cos^3 20^\circ - 3cos20^\circ = 4a^3 - 3a.$$

Thus, a is a root of the polynomial $8X^3 - 6X - 1$. Since this polynomial has no rational root (prove it), it is irreducible. Hence $[Q(a) : Q] = 3$. From the Proposition 8.7.2, a is not constructible. ♯

Corollary 8.7.13 *Trisection of a* 60° *angle is impossible by ruler and compass construction.*

Proof Since 60° angle can be constructed by ruler and compass, if it is possible to trisect 60° angle, then 20° angle is constructible. This is impossible because of the above theorem. ♯

Corollary 8.7.14 *It is impossible to duplicate a unit cube by the ruler compass constructions.*

Proof If we can duplicate a unit cube by ruler and compass, then we can construct a segment of length $2^{\frac{1}{3}}$. Since $[\mathbb{Q}(2^{\frac{1}{3}}) : \mathbb{Q}] = 3$ is not a power of 2, this is impossible. ♯

Recall that a complex number a is called algebraic number if it is algebraic over $\mathbb{Q}$. We state a result without proof.

Theorem 8.7.15 (Lindemann–Weierstrass). *Let $a_1, a_2, \ldots, a_n$ be n distinct algebraic numbers. Then $\{e^{a_1}, e^{a_2}, \ldots, e^{a_n}\}$ is linearly independent set over $\mathbb{Q}$.* ♯

Corollary 8.7.16 *π and e are not algebraic over $\mathbb{Q}$.*

Proof If e is a root of a nonzero polynomial over $\mathbb{Q}$, then there are rational numbers $\alpha_0, \alpha_1, \ldots, \alpha_n$ not all zero such that

$$\alpha_0 + \alpha_1 e^1 + \cdots + \alpha_n e^n = o.$$

This means that $\{e^0, e^1, \ldots, e^n\}$ is linearly dependent. Since $0, 1, 2, \ldots, n$ are algebraic over $\mathbb{Q}$, we arrive at a contradiction to the theorem of Lindemann and Weierstrass. Thus, e is not algebraic over $\mathbb{Q}$.

Since e^0 and $e^{\pi i}$ are rational numbers, they are linearly dependent over $\mathbb{Q}$. By the theorem of Lindemann and Weierstrass it follows that 0 and πi both are not algebraic. Since 0 is algebraic, it follows that πi is not algebraic. Again, since i is algebraic, it follows that π is not algebraic. ♯

Remark 8.7.17 It is not known if π is algebraic over $\mathbb{Q}(e)$.

Theorem 8.7.18 *It is impossible to construct a square by ruler and compass whose area is the area bounded by the unit circle.*

Proof Suppose that it is possible to construct such a square. Then $\sqrt{\pi}$ will be a constructible number. Since π is not algebraic, $\sqrt{\pi}$ is also not algebraic. But, then $\mathbb{Q}(\sqrt{\pi})$ is of infinite degree over $\mathbb{Q}$. This is a contradiction. ♯

Theorem 8.7.19 *A regular polygon of n side is constructible (by ruler and compass) if and only if $\phi(n)$ is a power of 2.*

Proof A regular polygon of n side is constructible if and only if the angles $\frac{2\pi}{n}$ at the center are constructible. This is possible if and only if $cos\frac{2\pi}{n}$ is constructible. Let $\rho = e^{\frac{2\pi i}{n}}$ denote the primitive nth root of 1. Since $cos\frac{2\pi}{n} = \frac{\rho+\rho^{-1}}{2}$, it follows that $cos\frac{2\pi}{n} \in \mathbb{Q}(\rho)$. Since $\rho \notin \mathbb{R}$ and $cos\frac{2\pi}{n} \in \mathbb{R}$, it follows that $\mathbb{Q}(cos\frac{2\pi}{n})$ is a proper subfield of $Q(\rho)$. Further, ρ is a root of the polynomial $X^2 - 2cos\frac{2\pi}{n}X + 1$, it follows that $[\mathbb{Q}(\rho) : \mathbb{Q}(cos\frac{2\pi}{n})] = 2$. Hence if $cos\frac{2\pi}{n}$ is constructible, then $\phi(n) = [\mathbb{Q}(\rho) : \mathbb{Q}]$ is a power of 2.

Conversely, suppose that $\phi(n)$ is a power of 2. Since $\mathbb{Q}(\rho)$ is a Galois extension, and which is abelian (isomorphic to U_n) of degree $\phi(n) = 2^m$, all the subgroups of the Galois group $G(\mathbb{Q}(\rho)/\mathbb{Q})$ are normal. In particular, $G(\mathbb{Q}(\rho)/\mathbb{Q}(cos\frac{2\pi}{n}))$ is normal, and $H = G(\mathbb{Q}(cos\frac{2\pi}{n})/\mathbb{Q})$ is an abelian group of order 2^{m-1}. From the theory of Abelian groups, we have a normal series

$$H = H_1 \trianglerighteq H_2 \trianglerighteq \cdots \trianglerighteq H_m = \{I\},$$

where $[H_i : H_{i+1}] = 2$. Taking the fixed fields, we obtain a chain

$$Q \subset F(H_{m-1}) \subset F(H_{m-2}) \cdots \subset F(H_1) = Q(cos\frac{2\pi}{n})$$

such that $[F(H_i) : F(H_{i-1})] = 2$. The result follows from Proposition 8.7.2. ♯

In particular, it is impossible to construct a regular polygon with 9 sides by ruler and compass. It is of course possible to construct a regular polygon of 17 ($\phi(17) = 2^4$) sides. An explicit algorithm to construct a regular polygon of 17 sides was given by Gauss in 1801.

Exercises

8.7.1 Can we divide a right angle in 10 equal parts by ruler and compass? Support.

8.7.2 Let a and b be positive rationals. Can we construct a square whose area is same as that of the ellipse with major axis $2a$ and minor axis $2b$? Support.

8.7.3 Can we construct an angle $\frac{2\pi}{9}$? Support.

8.7.4 Can we construct a circular arc of length $\frac{2\pi}{17}$? Support.

8.7.5 Show that a regular polygon of n side is constructible by ruler and compass if and only if all odd prime divisors of n are of the form $2^m + 1$.

8.7.6 Think of a machine which can construct cube root of a rational number.

8.8 Galois Theory of Equation

In this section, we determine a necessary and sufficient condition (due to Galois) for the solvability of polynomial equations by the field and the radical operations.

Definition 8.8.1 Let L be a field extension of K. We say that L is a **radical** extension of K if there exists elements $a_1, a_2, \ldots, a_r$ in L, and positive integers $n_1, n_2, \ldots, n_r$ such that $L = K(a_1, a_2, \ldots, a_r)$, where $a_i^{n_i} \in K(a_1, a_2, \ldots a_{i-1})$ *for all* $i \geq 1$. If we can take $n = n_i$ *for all* i, then we say that it is n - **radical** extension of K.

Thus, every radical extension is an algebraic extension. To say that L is a radical extension of K is to say that there is tower of finite extensions $K = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_r = L$ of finite simple extensions. Further, any radical extension is n radical extension for $n = n_1 n_2 \ldots n_r$. Thus, every cyclotomic extension is a radical extension. Also observe that if L is a n-radical extension of F, and F is also a n-radical extension of K, then L is n-radical extension of K.

Proposition 8.8.2 *Let L be a n-radical extension of K, and L′ the normal closure of the extension L of K. The L′ is also n-radical extension of K.*

Proof Let $L = K(a_1, a_2, \ldots, a_r)$, where $a_i^n \in K(a_1, a_2, \ldots, a_{i-1})$. The proof is by the induction on r. Suppose that $L = K(a_1)$ such that $a_1^n \in K$. Let $p_1(X)$ be the minimum polynomial of a_1 over K. Since a_1 is a root of $X^n - a$, where $a = a_1^n \in K$, it follows that $p_1(X)$ divides $X^n - a$. Thus, every root β of $p_1(X)$ satisfies the relation $\beta^n = a \in K$. The normal closure $L' = K(\beta_1, \beta_2, \ldots \beta_s)$, where β_j are all roots

of $p_1(X)$ over K is therefore n-radical extension of K. Assume that the result is true for r. Let $L = K(a_1, a_2, \ldots, a_r, a_{r+1})$ be n-radical extension of K. Let L_0 be the normal closure of $K(a_1, a_2, \ldots, a_r)$ over K. Then L_0 is the splitting field of the set $\{p_i(X) \mid 1 \leq i \leq r\}$, where $p_i(X)$ is irreducible polynomial of a_i over K, and by the induction hypothesis L_0 is n-radical extension of K. Then the normal closure L' of L over K is the splitting field of $p_{r+1}(X)$ over L_0. Let $\beta_1, \beta_2, \ldots, \beta_t$ be the roots of $p_{r+1}(X)$. Then from the assumption, it follows that $\beta_j^n \in K(a_1, a_2, \ldots, a_r) \subseteq L_0$. Thus L' is n-radical extension of L_0, and L_0 is n-radical extension of K. This shows that L' is n-radical extension of K. ♯

Definition 8.8.3 Let K be a field and $f(X) \in K[X]$. Then $f(X)$ is said to be **solvable by radical** operations if the splitting field of $f(X)$ is contained in a radical extension of K.

Proposition 8.8.4 *Let K be a field containing nth root of* 1*, and L be an abelian Galois extension such that the exponent of $G(L/K)$ divides n. Then L is a n-radical extension of K.*

Proof From the structure theorem of finite abelian groups, we have

$$G(L/K) = C_1 \times C_2 \times \cdots \times C_r,$$

where C_i are prime power order cyclic groups such that order of each cyclic group divides n. Let L_i denote the fixed field of

$$H_i = C_1 \times C_2 \times \cdots \times C_{i-1} \times C_{i+1} \times \cdots \times C_r.$$

Then $G(L_i/K)$ is isomorphic to C_i for each i. Thus, L_i over K is a cyclic Galois extension of degree m_i, and since m_i divides n, K also contains primitive m_ith root of 1. From the Theorem 8.6.10, we see that $L_i = K(a_i)$, where $a_i^{m_i} \in K$, and so $a_i^n \in K$ for all i. Clearly, $K(a_1, a_2, \ldots, a_r) = L_1 L_2 \ldots L_r$ is the fixed field of $\bigcap_i H_i = \{I\}$. Hence $K(a_1, a_2, \ldots, a_r) = L$. This shows that L is n-radical extension of K. ♯

Theorem 8.8.5 (Galois) *Let K be a field of characteristic* 0 *and $f(X) \in K[X]$. Let L be the splitting field of $f(X)$. Then $f(X)$ is solvable by radical if and only if the Galois group $G(L/K)$ is a solvable group.*

Proof Suppose that $f(X)$ is solvable by radical operations. Let F be a n-radical extension of K containing the splitting field L of $f(X)$. Since char K=0, we may assume, by Proposition 8.8.2, that F is also a Galois extension. Let L' be the splitting field of $X^n - 1$ over F. Since characteristic K is $0, L' = F(\rho)$, where ρ is a primitive nth root of 1. From the previous proposition, L' is also n-radical extension of $K(\rho)$. Thus, we have tower

$$K = K_0 \subseteq K_1 = K(\rho) \subseteq K_2 \subseteq K_3 \subseteq \cdots \subseteq K_r = L',$$

where $K_{i+1} = K_i(a_i), i \geq 1$ for some a_i such that $a_i^n \in K_i$. Since each $K_i, i \geq 1$ contains a primitive nth root of 1, it follows by Theorem 8.6.11 that K_{i+1} is a cyclic Galois extension of K_i for all $i \geq 1$. Since K_1 is a cyclotomic extension it follows that it is abelian. It is also clear that L' is a Galois extension of each K_i, and we have a normal series

$$G(L'/K) \trianglerighteq G(L'/K_1) \trianglerighteq G(L'/K_2) \trianglerighteq \cdots \trianglerighteq G(L'/K_r) = \{I\}$$

whose factors $G(L'/K_i)/G(L'/K_{i+1}) \approx G(K_{i+1}/K_i)$ are all abelian (as observed above). Thus, $G(L'/K)$ is a solvable group. By the fundamental theorem of Galois theory, it follows that $G(L/K)$ is isomorphic to the quotient $G(L'/K)/G(L'/L)$. This shows that $G(L/K)$ is solvable.

Conversely, suppose that $G(L/K)$ is solvable, and

$$G(L/K) = G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_r = \{I\}$$

a normal series of $G(L/K)$ with abelian factors. Let $F_i = F(G_i)$. Then from the fundamental theorem of Galois theory, it follows that F_{i+1} is Galois over F_i with Galois group isomorphic to abelian group G_i/G_{i+1}. Let ρ be a primitive nth root of 1, where n is the exponent of $G(L/K)$. Note that it exists because the characteristic of K is 0. Let $L_i = F_i(\rho)$. Then we get a tower of extensions

$$K \subseteq L_0 \subseteq L_1 \subseteq \cdots \subseteq L_r = L(\rho).$$

It follows, by induction, that $L_{i+1} = L_i F_{i+1}$, and by the Exercise 8.6.6, $G(L_{i+1}/L_i)$ is isomorphic to a subgroup of the abelian group $G(F_{i+1}/F_i)$. Clearly exponent of $G(L_{i+1}/L_i)$ divides n. From the above proposition, it follows that L_{i+1} is n radical extension of L_i for each i. Since $L_0 = K(\rho)$ is already a n-radical extension of K, it follows that $L(\rho)$ is n-radical extension of K, and so L is contained in a radical extension of K. This means that $f(X)$ is solvable by the radical operations. ♯

Let L denote the field $K(X_1, X_2, \ldots X_n)$ of fractions of the polynomial ring $K[X_1, X_2, \ldots X_n]$, where K is a field of characteristic 0. Then every permutation $p \in S_n$ defines uniquely an automorphism of L. Indeed, S_n is isomorphic to a subgroup of $Aut(L)$ (see Example 8.3.11) which we denote by S_n again. Let F denote the fixed field of S_n. Then we have seen (Example 8.3.11) that $F = K(s_1, s_2, \ldots, s_n)$. The polynomial

$$f(X) = X^n - s_1 X^{n-1} - \cdots - (-1)^n s_n = (X - X_1)(X - X_n) \ldots (X - X_n)$$

in $F[X]$ is called a **general nth** degree polynomial over K. The question is: can we determine a formula for $X_1, X_2, \ldots, X_n$ in terms of the symmetric polynomials using field and radical operations? In other words, can we solve a general nth degree equation over K by the field and the radical operations. Since the Galois group of

$f(X)$ over F is S_n, and S_n is not a solvable group for $n \geq 5$, the following result of Abel and Ruffini follows from the above theorem of Galois.

Theorem 8.8.6 (Abel–Ruffini) *A general nth degree equation, $n \geq 5$, over a field K of characteristic 0 is not solvable by radicals.* ♯

Example 8.8.7 Let $f(X)$ be an irreducible polynomial over Q of degree p, where $p \geq 5$ is a prime such that it has exactly 2 imaginary roots (which are of course conjugate to each other). Then the Galois group of $f(X)$ over $\mathbb{Q}$ is S_p: We prove it. Let L be the splitting field of $f(X)$. Then since $f(X)$ is irreducible, p divides $\mid G(L/\mathbb{Q}) \mid$. Also $G(L/Q)$ is a subgroup of S_p (it acts faithfully on the roots of $f(X)$ which are all distinct). Since p divides $\mid G(L/\mathbb{Q}) \mid$, by the Cauchy theorem, $G(L/\mathbb{Q})$ has an element of order p. This means that it is a cycle of length p. Further, the complex conjugation is a member of $G(L/\mathbb{Q})$ which interchanges two roots, and therefore represents a transposition. Since a transposition together with a cycle of length p generates S_p, it follows that the Galois group $G(L/Q)$ is S_p. Since S_p is not solvable, by the above theorem of Galois, $f(X)$ is not solvable by radicals. In particular, using elementary calculus, and the Eisenstein irreducibility criteria, we see that $X^5 - 4X + 2$ satisfies the conditions mentioned above, and so it is not solvable by radical.

Since $S_n, n \leq 4$ is solvable, every general equation of degree $n, n \leq 4$ is solvable by radical. In what follows we determine formula for these lower degree general polynomial equations.

1. **Quadratic Equations**. For general quadratic polynomial $X^2 - s_1X + s_2$, the formula for X_1 and X_2 is given by the Sridharacharya formula. It is given by

$$X_i = \frac{s_1 \pm \sqrt{s_1^2 - 4s_2}}{2}.$$

2. **Cubic Equations**. We give the **Cardano′s method** to solve a cubic equation by radical operations. Consider the general 3° equation

$$X^3 - s_1X^2 + s_2X - s_3 = 0 \cdots\cdots \tag{8.8.1}$$

Substituting $X = Y + \frac{s_1}{3}$ in the above equation, we get

$$Y^3 + pY + q = 0 \cdots\cdots \tag{8.8.2}$$

where

$$p = s_2 - \frac{s_1^2}{3} \cdots\cdots \tag{8.8.3}$$

and

$$q = -2\frac{s_1^3}{9} + \frac{s_1 s_2}{3} - s_3 \cdots\cdots \tag{8.8.4}$$

If $p = 0$, then the roots are $\alpha, \alpha\omega$ and $\alpha\omega^2$, where $\alpha = (-q)^{\frac{1}{3}}$, and ω a primitive cube root of 1. Suppose that $p \neq 0$. Substitute $Y = U + V$ in 2, and then the equation reduces to

$$U^3 + V^3 + q + (3UV + p)(U + V) = 0 \cdots\cdots \tag{8.8.5}$$

We set two equations

$$U^3 + V^3 + q = 0 \cdots\cdots \tag{8.8.6}$$

and

$$3UV + p = 0 \cdots\cdots \tag{8.8.7}$$

in U and V. Since $p \neq 0$, U and V both are nonzero. Substituting the value of V obtained from the Eq. 8.8.7 in the Eq. 8.8.6, we get that

$$4U^6 + qU^3 - \frac{p^3}{27} = 0 \cdots\cdots \tag{8.8.8}$$

This is a quadratic equation in U^3. Solving we get $U^3 = -\frac{q}{2} \pm \sqrt{-\frac{D}{108}}$, where $D = -4p^3 - 27q^3$ (called the discriminant of $Y^3 + pY + q$). If we take $U^3 = -\frac{q}{2} + \sqrt{-\frac{D}{108}}$, and denote it by P, then by symmetry $V^3 = -\frac{q}{2} - \sqrt{-\frac{D}{108}}$, and this we denote by Q. Pairing the three cube roots of P and Q each such that their product is $-\frac{p}{3}$, we get the roots $\sqrt[3]{P} + \sqrt[3]{Q}$, $\omega\sqrt[3]{P} + \omega^2\sqrt[3]{Q}$, $\omega^2\sqrt[3]{P} + \omega\sqrt[3]{Q}$ of $Y^3 + pY + q = 0$. This gives all roots of the given cubic equation.

3. **Bi − quadratic equations**. We describe **Ferrari′s method** of solving the general 4° equation

$$X^4 - s_1X^3 + s_2X^2 - s_3X + s_4 = 0 \cdots\cdots \tag{8.8.9}$$

The above equation can also be written as

$$(X^2 + \frac{-s_1X + Y}{2})^2 = AX^2 + BX + C,$$

where $A = \frac{s_1^2}{4} - s_2 + Y$, $B = \frac{-s_1Y}{2} + s_3$, and $C = \frac{Y^2}{4} - s_4$. We chose Y so that the RHS becomes a perfect square. This is so if $B^2 - 4AC = 0$. This gives us a cubic equation in Y which can be solved by the Cardano's method. Suppose that the RHS becomes $(PX + Q)^2$. Then

$$X^2 + \frac{-s_1X + Y}{2} = \pm(PX + Q).$$

At last, this quadratic equation can be solved, and we get the solutions of the given bi-quadratic equation.

Exercises

8.8.1 Solve the cubic equation $X^3 - 3X^2 + 2X + 1 = 0$.

8.8.2 Solve the $X^4 + 4X - 1 = 0$.

8.8.3 Find a polynomial of degree 7 over Q whose Galois group is S_7.

Chapter 9
Representation Theory of Finite Groups

In this chapter, we develop the elementary theory of linear representations of finite groups over a field F. We shall assume that the characteristic of F does not divide the order $| G |$ of G. The representations over fields F where characteristic of F divides the order of G are called **Modular** representations or **Brauer** representations, and the theory was developed by $Brauer$. We shall have occasions, of course rare, to make some comments about modular representation theory.

9.1 Semi-simple Rings and Modules

One of the crucial properties of a field (skew field) F is that given any module M over F and a submodule N of M there is a submodule L of M such that M is direct sum of N and L. In this section, we study rings with this property. All the rings in this section are assumed to be rings with identities.

Definition 9.1.1 A left R-module M is called a **simple** left module if it has no proper submodules.

Nontrivial simple left modules over a division ring R are precisely one-dimensional spaces. Simple $\mathbb{Z}$-modules are precisely prime cyclic groups.

Example 9.1.2 Let D be a division ring, and $M_n(D)$ be the ring of $n \times n$ matrices with entries in D. Let D^n denote the additive group of column vectors. Then D^n is a left $M_n(D)$-module with respect to the matrix multiplication. Let X be a nonzero column vector in D^n, and $Y \neq 0$ be any other nonzero column vector. It is an elementary fact of linear algebra that there is a matrix A in $M_n(D)$ such that $A \cdot X = Y$. Thus, D^n has no proper M_n-submodules, and so it is simple. Let D_i denote the subset of $M_n(D)$ consisting of those matrices all of whose columns except possibly ith column is zero. Then an easy calculation shows that D_i is a minimal nonzero left

R. Lal, *Algebra 2*, Infosys Science Foundation Series in Mathematical Sciences,
DOI 10.1007/978-981-10-4256-0_9

ideal. As a module this is isomorphic to the module D^n. Also observe that the subset D_i' of $M_n(D)$ consisting of matrices whose ith column is zero is also a left ideal, and it is maximal because $M_n(D)/D_i'$ is isomorphic to the simple left module D^n. An elementary calculation shows that $M_n(D)$ has no proper two-sided ideal.

Proposition 9.1.3 *A left R-module M is simple if and only if there is a maximal left ideal A of R such that M as a module is isomorphic to R/A.*

Proof Since a submodule of R/A is of the form B/A, where B is a left ideal containing A, it follows that R/A is simple if and only if A is a maximal left ideal. Let M be a simple left R-module. Let $x \neq 0$ be an element of M. Since M is simple $Rx = M$. Thus the map $a \rightsquigarrow ax$ is a surjective R-homomorphism. By the fundamental theorem R/A is isomorphic to M where A is the kernel of the map. Since M is simple, A is maximal left ideal. ♯

Proposition 9.1.4 (Schur's Lemma) *Let M and N be simple left R modules. Then a nonzero homomorphism from M to N is an isomorphism. In particular, $End_R(M)$ is a division ring with respect to the addition of endomorphisms, and the product as composition of maps.*

Proof Let f be a nonzero homomorphism from M to N. Then $f(M)$ is a nonzero submodule of N. Since N is simple, $f(M) = N$ and so f is surjective. Next $Kerf$ is also a submodule of M different from M, and since M is also simple, $Kerf$ is $\{0\}$. This means that f is injective. In particular, every nonzero element of $End_R(M)$ is invertible, and so $End_R(M)$ is a division ring. ♯

Theorem 9.1.5 *Let M be a left R-module. Then the following conditions are equivalent.*

1. *M is sum of its simple submodules (equivalently M is generated by its simple submodules).*
2. *M is direct sum of simple submodules.*
3. *Every submodule of M is direct summand.*
4. *Every short exact sequence of the type*

$$0 \longrightarrow N \xrightarrow{f} M \xrightarrow{g} L \xrightarrow{0}$$

splits.

Proof $(1 \Longrightarrow 2)$. Suppose that $M = \Sigma_{\alpha\in\Lambda} M_\alpha$, where $\{M_\alpha \mid \alpha \in \Lambda\}$ is the family of all simple submodules of M. Let

$$X = \{J \subseteq \Lambda \mid \Sigma_{\alpha\in J} M_\alpha = \oplus\Sigma_{\alpha\in J} M_\alpha\}.$$

Then each $\{\alpha\} \in X$. Hence $X \neq \emptyset$. X with inclusion is a nonempty partially ordered set. Let $\{J_\mu \mid \mu \in \Omega\}$ be a chain in X. Then $J = \bigcup_{\mu\in\Omega} J_\mu$ is a member of X, for

$\Sigma_{\alpha\in J}M_\alpha = \oplus\Sigma_{\alpha\in J}M_\alpha$ (verify). Thus, J is an upper bound of the chain. By the Zorn's Lemma X has a maximal element J_0 (say). We show that $M = \oplus\Sigma_{\alpha\in J_0}M_\alpha$. Put $N = \oplus\Sigma_{\alpha\in J_0}M_\alpha$. Suppose that $M_\beta \nsubseteq N$ for some $\beta \in \Lambda$. Then $\beta \notin J_0$. Since M_β is simple, $M_\beta \bigcap N$ is $\{0\}$, or it is M_β. Since it is assumed that $M_\beta \nsubseteq N$, it follows that $M_\beta \bigcap N = \{0\}$. Thus, $M_\beta + N = M_\beta \oplus N = \oplus\Sigma_{\alpha\in J_0\bigcup\{\beta\}}M_\alpha$. This shows that $J_0 \bigcup\{\beta\} \in X$. This is a contradiction to the supposition that J_0 is a maximal element of X. Hence $M_\alpha \subseteq N$ for each $\alpha \in \Lambda$. This shows that $N = M$.

$2 \Longrightarrow 1$ is obvious.

$2 \Longrightarrow 3$. Assume 2. Let $M = \oplus\Sigma_{\alpha\in J}M_\alpha$, where each M_α is a simple submodule of M. Let N be a proper submodule of M. As in the proof of $1 \Longrightarrow 2$, consider

$$X = \{F \subseteq J \mid N + \Sigma_{\alpha\in F}M_\alpha = N \oplus \Sigma_{\alpha\in F}M_\alpha\}.$$

Since $N \neq M$, there is M_α, $\alpha \in J$ such that M_α is not contained in N. Since M_α is simple, $N \bigcap M_\alpha = \{0\}$, and so $N + M_\alpha = N \oplus M_\alpha$. Hence such a $\{\alpha\}$ is in X, and so X is nonempty set. Order it through inclusion. As in the proof of $1 \Longrightarrow 2$, X has a maximal element F_0(say), and then $N \oplus \Sigma_{\alpha\in F_0}M_\alpha = M$. This shows that N is a direct summand.

$3 \Longrightarrow 4$. Assume 3. Then $f(N)$ is a submodule of M, and $M = f(N) \oplus K$ for some K. Every element of M is uniquely expressible as $f(n) + k$, where $n \in N$ and $k \in K$. It is clear that the map s from M to N defined by $s(f(n) + k) = n$ defines a splitting.

$4 \Longrightarrow 3$. Assume 4. Let N be a submodule of M. Then we have a short exact sequence

$$0 \longrightarrow N \longrightarrow M \longrightarrow M/N \longrightarrow 0,$$

where the map from N to M is inclusion, and the map from M to M/N is quotient map. From 4, it is a split exact sequence, and so $M = N \oplus K$ for some submodule K (isomorphic to M/N) of M.

$3 \Longrightarrow 1$. Let $M \neq \{0\}$ be a left R-module such that every submodule of M is direct summand of M. We first show that every submodule N of M also has this property. Let K be a submodule of N. Then K is also a submodule of M, and hence there is a submodule L of M such that $M = K \oplus L$. We show that $N = K \oplus (L \bigcap N)$. Clearly, $K \bigcap (L \bigcap N) = \{0\}$. Let $x \in N$. Then $x = y + z$, where $y \in K$ and $z \in L$. Since $x, y \in N$, $z \in N$. Hence $z \in L \bigcap N$, and so $N = K + (L \bigcap N)$. Thus, every submodule of N is direct summand of N. Next, we show that every nonzero submodule N of M contains a nonzero simple submodule. Let $x \in N$, $x \neq 0$. Then $Rx \neq \{0\}$ is a submodule of N. Consider the surjective homomorphism f from R to Rx defined by $f(a) = ax$. Since f is surjective, and $Rx \neq \{0\}$, it follows that $Kerf \neq R$. Suppose that $A = Kerf$. Then A is a proper left ideal of R. By the Krull's theorem, A can be embedded in a maximal left ideal B (say). By the first isomorphism theorem $R/B \approx Rx/f(B) = Rx/Bx$. Since B is maximal, R/B is simple, and so Rx/Bx is also simple. Since Rx is a submodule of M, and Bx is a submodule of Rx, $Rx = Bx \oplus T$ for some nonzero submodule T of Rx. T being

isomorphic to Rx/Bx is simple. Thus, N contains a nonzero simple submodule of M. Let M_0 be the sum of all simple submodules of M. Suppose that $M_0 \neq M$. Then from 3, there exists a nonzero submodule N_0 of M such that $M = M_0 \oplus N_0$. From what we have proved above N_0 contains a nonzero simple submodule L_0 of M. But, then L_0 is a simple submodule of M not contained in M_0. This is a contradiction to the choice of M_0. Hence $M_0 = M$. ♯

Definition 9.1.6 A left R-module M is said to be **semi-simple** if it satisfies any one (and hence all) of the four conditions in the above theorem.

Every vector space is semi-simple, for all subspaces are direct summands. Since $\mathbb{Z}$-simple modules are prime cyclic groups, $\mathbb{Z}$ semi-simple modules are precisely direct sum of prime cyclic groups. More generally, simple R-modules over a P.I.D. are isomorphic to R/Rp, where p is irreducible element of R. In particular, a P.I.D. R is a left semi-simple module over itself if and only if it is a field.

Proposition 9.1.7 *Every submodule of a semi-simple left R-module is semi-simple. Every homomorphic image (and so quotient) module of a semi-simple left module is semi-simple.*

Proof Let M be a semi-simple left module, and N be a submodule of M. Then as observed in the proof of $3 \Longrightarrow 1$ (Theorem 9.1.5), it follows that every submodule of N is a direct summand of N. Hence N is semi-simple. Let β be a surjective homomorphism from M to N. Since M is semi-simple, the exact sequence

$$0 \longrightarrow Ker\beta \longrightarrow M \longrightarrow N \longrightarrow 0$$

splits. Hence, there exists a homomorphism t from N to M such that $\beta ot = I_N$. Clearly, t is injective homomorphism, and as a result N is isomorphic to the submodule $t(N)$ of M. Since submodule of a semi-simple left module is semi-simple, $t(N)$, and so N is semi-simple. ♯

Since a direct sum of direct sums of simple left modules is a direct sum of simple modules, we have the following proposition.

Proposition 9.1.8 *Direct sum of a family of semi-simple left modules is a semi-simple left module.* ♯

Definition 9.1.9 A ring R is said to be a Left **semi-simple** ring if it is a semi-simple left module over itself.

A field is a semi-simple ring, for it itself is a simple module over itself. $\mathbb{Z}$ is not semi-simple, for it cannot be direct sum of simple left ideals. Subring of a left semi-simple ring need not be left semi-simple. For example, $\mathbb{Q}$ is semi-simple ring but $\mathbb{Z}$ is not.

Theorem 9.1.10 *A ring R is left semi-simple ring if and only if every left module over R is semi-simple.*

Proof If every left module over R is semi-simple, then in particular, R is semi-simple left module over itself. By the definition, R is left semi-simple. Conversely, suppose that R is left semi-simple ring. Then every free left R-module, being direct sum of copies of R, is semi-simple left module. Since every left module is quotient of a free left module and quotient of a semi-simple left module is semi-simple, it follows that every left module over R is semi-simple. ♯

The following corollary is immediate from the previous results.

Corollary 9.1.11 *Let R be a ring. Then the following conditions are equivalent.*

1. *R is left semi-simple.*
2. *Every short exact sequence of left R-modules splits.*
3. *Every left R-module is semi-simple.*
4. *Every left R-module is projective.*
5. *Every left R-module is injective.* ♯

It follows from the definition of left semi-simple ring that a ring R is left semi-simple if it is direct sum of minimal nonzero left ideals. Since $M_n(D)$, where D is a division ring, is the direct sum of its minimal nonzero left ideals D_i, it follows that $M_n(D)$ is left semi-simple.

Definition 9.1.12 A left semi-simple ring is said to be **simple** if it has no nonzero proper two-sided ideals.

Example 9.1.13 Since $M_n(D)$ has no nonzero proper two-sided ideals, it follows from the above discussion that $M_n(D)$ is a left simple ring. We shall see that every simple ring is isomorphic to $M_n(D)$ for some n, and for some division ring D.

Let R_1 and R_2 be rings. Then left ideals of $R_1 \times \{0\}$, and those of $\{0\} \times R_2$ are also left ideals of $R_1 \times R_2$. Thus, if R_1 and R_2 are direct sum of minimal left ideals, then $R_1 \times R_2$ is also direct sum of minimal nonzero minimal left ideals. This proves the following proposition.

Proposition 9.1.14 *Direct product of left semi-simple rings is a left semi-simple ring.* ♯

In particular, we have the following corollary.

Corollary 9.1.15 *Let $D_1, D_2, \ldots, D_r$ be division rings, and $n_1, n_2, \ldots, n_r$ be positive integers. Then*

$$M_{n_1}(D_1) \times M_{n_2}(D_2) \times \cdots \times M_{n_r}(D_r)$$

is a left semi-simple ring. ♯

One of the main results of this section is to show that any left semi-simple ring is isomorphic to such a ring. In particular, left semi-simple rings and right semi-simple rings are same.

Let F be a field and G a group. Recall (see the Sect. 7.6 on polynomial rings, Algebra 1) the definition of the group ring $F(G)$. $F(G)$ is at first a vector space over F with members of G as basis. Thus, the members of $F(G)$ are formal sums $\Sigma_{g\in G}\alpha_g g$, where all but finitely many α_g are 0. The multiplication $\cdot$ in $F(G)$ defined by

$$(\Sigma_{g\in G}\alpha_g g)\cdot(\Sigma_{g\in G}\beta_g g) \;=\; (\Sigma_{g\in G}\gamma_g g),$$

where $\gamma_g \;=\; \Sigma_{hk\,=g}\alpha_h\beta_k$, makes $F(G)$ a ring. Thus, $F(G)$ is an algebra over F, and it is called a **group algebra**. The following result is the first basic result in the representation theory of groups.

Theorem 9.1.16 (Maschke) *Let F be a field, and G be a finite group such that the characteristic of the field F does not divide the order of the group G. Then $F(G)$ is a semi-simple ring.*

Proof Assume that the characteristic of F does not divide the order of the group G. We shall show that every left $F(G)$-module is semi-simple. Let M be a left $F(G)$-module and N a submodule of M. Since F is a subfield of $F(G)$, M is a vector space over F, and N is a subspace of M. Hence, there is a F-subspace L of M such that M is a vector space direct sum $N \oplus_F L$ of N and L. Let p_1 be the first projection from M to N. Then p_1 is a F-homomorphism from M to N such that $p_1(x) \;=\; x$ for all $x \in N$. We average p_1 to make it a $F(G)$-homomorphism. Define a map $\overline{p_1}$ from M to N by

$$\overline{p_1}(m) \;=\; (n\cdot 1)^{-1}\Sigma_{g\in G} g\cdot p_1(g^{-1}\cdot m),$$

where n is the order of G(note that characteristic of F does not divide n, and so $n\cdot 1 \neq 0$ in F). Further,

$$\overline{p_1}(m_1+m_2) \;=\; (n\cdot 1)^{-1}\Sigma_{g\in G} g\cdot p_1(g^{-1}(m_1+m_2)) \;=\; (n\cdot 1)^{-1}\Sigma_{g\in G} g\cdot p_1(g^{-1}m_1+ g^{-1}m_2) \;=\; (n\cdot 1)^{-1}\Sigma_{g\in G}(g\cdot p_1 g^{-1}(m_1)+g\cdot p_1 g^{-1}(m_2)) \;=\; \overline{p_1}(m_1)+\overline{p_1}(m_2),$$

and

$$\overline{p_1}(\alpha h\cdot m) \;=\; (n\cdot 1)^{-1}\Sigma_{g\in G} g\cdot p_1(g^{-1}\alpha hm).$$

Putting $g^{-1}h \;=\; x$ in the above equation, we get that

$$\overline{p_1}(\alpha hm) \;=\; (n\cdot 1)^{-1}\alpha h\Sigma_{x\in G} x p_1(x^{-1}m) \;=\; \alpha h\overline{p_1}(m)$$

for all $\alpha \in F, h \in G$, and $m \in M$. This shows that $\overline{p_1}$ is a $F(G)$-homomorphism. Also since N is a $F(G)$-submodule, for each $x \in N$, $g^{-1}x \in N$, and so $p_1(g^{-1}x) = g^{-1}x$. Hence for each $x \in N$, we have

$$\overline{p_1}(x) = (n \cdot 1)^{-1} \Sigma_{g \in G} g \cdot p_1(g^{-1}x) = (n \cdot 1)^{-1} \Sigma_{g \in G} g g^{-1} x = (n \cdot 1)^{-1} n x = x.$$

Thus, $\overline{p_1} o i = I_N$, where i is the inclusion map from N to M. Hence, M is $F(G)$-direct sum of N and $Ker\,\overline{p_1}$. This completes the proof of the fact that M is $F(G)$-semi-simple. ♯

Our next aim is to determine the structure of a semi-simple ring.

Let

$$M = M_1 \oplus M_2 \oplus \cdots \oplus M_n$$

and

$$N = N_1 \oplus N_2 \oplus \cdots \oplus N_m$$

be some direct sum decompositions of left R-modules M and N. Let M_{MN} denote the set of matrices $[\phi_{sr}]$, where $\phi_{rs} \in Hom_R(M_r, N_s)$. Let f_{sr} denote the homomorphism $p_s o f o i_r$ from M_r to N_s, where i_r is the natural inclusion of M_r in M, and p_s the sth projection of N to N_s. It is easy to verify that the map η_{MN} from $Hom_R(M, N)$ to M_{MN} defined by

$$\eta_{MN}(f) = [f_{sr}]$$

is a bijective map. In fact every $x \in M$ is uniquely expressed as $x_1 + x_2 + \cdots + x_n$, where $x_i \in M_i$, and then $f(x) = \Sigma_{r=1}^{n} f_{sr} x_r$. In matrix form it is expressed by

$$f(x) = \begin{bmatrix} f_{11} & f_{12} & \cdot\cdot & f_{1n} \\ f_{21} & f_{22} & \cdot\cdot & f_{2n} \\ \cdot & \cdot & \cdot\cdot & \cdot \\ f_{m1} & f_{m2} & \cdot\cdot & f_{mn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix}.$$

Further, let M_{NL} denote the set of matrices with respect to the above given direct decomposition of N, and with respect to a direct decomposition of L. Then the matrix multiplication induces an external multiplication from $M_{MN} \times M_{NL}$ to M_{ML} (multiplication of entries are composition of maps). It is also easy to observe that $\eta_{ML}(gof) = \eta_{NL}(g) \cdot \eta_{MN}(f)$. In particular, if L is a left R-module with a given direct sum decomposition, and M_{LL} denote the corresponding set of matrices, then M_{LL} is a ring with respect to matrix addition and multiplication. Clearly, this ring is isomorphic to the ring $End_R(L)$.

Let us denote by M^n the n times direct sum of M.

Proposition 9.1.17 *Let $M = M_1^{n_1} \oplus M_2^{n_2} \oplus \cdots \oplus M_r^{n_r}$, where $\{M_1, M_2, \ldots, M_r\}$ is a set of pairwise nonisomorphic simple left R-modules. Then the ring $End_R(M)$ is isomorphic to the direct product*

$$M_{n_1}(D_1) \times M_{n_2}(D_2) \times \cdots \times M_{n_r}(D_r),$$

where $D_i = End_R(M_i)$ is a division ring, and $M_{n_i}(D_i)$ is the ring of $n_i \times n_i$ matrices with entries in D_i.

Proof Since each M_i is simple left R-module, by the Schur's Lemma $End_R(M_i)$ is a division ring, and for $i \neq j$, $Hom_R(M_i, M_j) = \{0\}$. It follows from the discussion prior to the proposition that $End_R(M)$ is isomorphic to the ring of matrices of the type

$$\begin{bmatrix} A_1 & 0 & \cdots & \cdot & 0 \\ 0 & A_2 & 0 \cdots & \cdot & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & A_{r-1} & 0 \\ 0 & 0 & \cdots & 0 & A_r \end{bmatrix}$$

, where $A_i \in M_{n_i}(D_i)$. The map which associates to each matrix of the above type to $(A_1, A_2, \ldots, A_r)$ defines an isomorphism from $End_R(M)$ to $M_{n_1}(D_1) \times M_{n_2}(D_2) \times \cdots M_{n_r}(D_r)$. ♯

Proposition 9.1.18 *Let R be a ring with identity. Let $a \in R$. Then the map f_a from R to R defined by $f_a(x) = x \cdot a$ is a member of $End_R(R, +)$, where $(R, +)$ is treated as a left R-module. Further, the map f from R to $End_R(R, +)$ defined by $f(a) = f_a$ is an anti-isomorphism.*

Proof $f_a(x + y) = (x + y) \cdot a = x \cdot a + y \cdot a = f_a(x) + f_a(y)$. Also $f_a(b \cdot x) = b \cdot x \cdot a = b \cdot f_a(x)$. Thus, $f_a \in End_R(R, +)$. Next, $f(a + b)(x) = f_{a+b}(x) = (a + b) \cdot x = a \cdot x + b \cdot x = f_a(x) + f_b(x) = f(a)(x) + f(b)(x) = (f(a) + f(b))(x)$ for all $a, b, x \in R$. Thus, $f(a+b) = f(a)+f(b)$. Next, $f_{ab}(x) = x \cdot ab = f_b(f_a(x))$ for all $a, b, x \in R$. This shows that $f(ab) = f(b)f(a)$ for all $a, b \in R$, and so f is a anti-homomorphism. Suppose that $f(a) = f(b)$. Then $a = a \cdot 1 = f_a(1) = f(a)(1) = f(b)(1) = f_b(1) = b \cdot 1 = b$. Thus, f is an injective anti-homomorphism. Further, given any $f \in End_R(R, +)$, $f = f(f(1))$. This shows that f is also surjective. ♯

Theorem 9.1.19 *Let R be a left semi-simple ring. Then there are only finitely non-isomorphic simple left modules each isomorphic to a simple left ideal of R. Let $\{M_1, M_2, \ldots, M_r\}$ be a set of pairwise nonisomorphic simple left R-modules such that each simple left R-module is isomorphic to M_i for some i. Let D_i denote the*

division ring $End_R(M_i)$. *Then there exists positive integers* $n_1, n_2, \ldots, n_r$ *such that the ring* R *is isomorphic to the ring*

$$M_{n_1}(D_1) \times M_{n_2}(D_2) \times \cdots \times M_{n_r}(D_r).$$

Proof Let R be a left semi-simple ring. Let M be a nontrivial simple left R-module. Then there is a maximal left ideal A of R such that R/A is isomorphic to M. Since R is left semi-simple, there is a left ideal B of R such that $R = A \oplus B$. But, then B, being isomorphic to R/A, is a simple (minimal nontrivial)left ideal, and it is isomorphic to M. Since R is a left semi-simple ring, it is direct sum of simple left ideals. Suppose that

$$R = \oplus\Sigma_{\alpha\in\Lambda} A_\alpha,$$

where A_α is a simple left ideal for each $\alpha \in \Lambda$. Thus, $1 \in R$ can be uniquely expressed as

$$1 = e_{\alpha_1} + e_{\alpha_2} + \cdots + e_{\alpha_t},$$

where $\alpha_i \in \Lambda$, and each $e_{\alpha_i} \neq 0$ is a member of A_{α_i}. Since each A_{α_i} is a simple left ideal, $Re_{\alpha_i} = A_{\alpha_i}$. Thus,

$$R = R \cdot 1 = Re_{\alpha_1} + Re_{\alpha_2} + \cdots + Re_{\alpha_r} = A_{\alpha_1} \oplus A_{\alpha_2} \oplus \cdots \oplus A_{\alpha_r}.$$

In turn, R is direct sum of finitely many simple left ideals. Suppose that

$$R = A_1 \oplus A_2 \oplus \cdots \oplus A_t,$$

where A_i is simple left ideal of R. We show that any nontrivial simple left module is isomorphic to A_i for some i. Let M be a nonzero simple left R-module. Then $\{0\} \neq M = RM$. Hence $A_i M \neq \{0\}$ for some i. Since $A_i M$ is a nonzero submodule of M and M is simple, it follows that $A_i M = M$. In turn, it follows that $A_i x \neq \{0\}$ for some $x \in M$. Since $A_i x$ is also a submodule of M, we have $A_i x = M$. Define a map f from A_i to M by $f(a) = a \cdot x$. Then f is clearly a surjective R-homomorphism. Since A_i is simple and $Kerf \neq A_i$, it follows that $Kerf = \{0\}$. Thus, f is an isomorphism from A_i to M. This shows that every simple left R-module is isomorphic to A_i for some i, and so there are only finitely many nonisomorphic simple left R-modules. Let $\{M_1, M_2, \ldots, M_r\}$ be a set of pairwise nonisomorphic simple left R-modules such that every simple left R-module is isomorphic to M_i for some i. Suppose that n_i of $A_1, A_2, \ldots, A_t$ are isomorphic to M_i. Then as left R-module, R is isomorphic to

$$M_1^{n_1} \oplus M_2^{n_2} \oplus \cdots \oplus M_r^{n_r}.$$

Thus, $End_R(R, +)$ is isomorphic to

$$M_{n_1}(D_1) \times M_{n_2}(D_2) \times \cdots \times M_{n_r}(D_r).$$

The map $A \rightsquigarrow A^t$ defines an anti-isomorphism from $M_{n_i}(D_i)$ to itself. Thus, the above ring is anti-isomorphic to itself. Further, we have seen that R is anti-isomorphic to $End_R(R, +)$. Since composition of two anti-isomorphisms is isomorphisms, the result follows. ♯

Corollary 9.1.20 *Every left semi-simple ring is anti-isomorphic to itself.* ♯

It is clear that R is left semi-simple if and only if the opposite ring of R is right semi-simple. Thus, we have the following corollary.

Corollary 9.1.21 *Left semi-simple rings and right semi-simple rings are same.* ♯

From now onward, we shall simply write semi-simple ring instead of left semi-simple or right semi-simple.

Corollary 9.1.22 *A semi-simple ring is commutative if and only if it is direct product of fields.*

Proof The result follows if we observe that $M_n(D)$ is commutative if and only if D is a field, and $n = 1$. ♯

Let M be a left R-module, and $S = End_R(M)$. Then M is also a S-left module with respect to $\cdot$ defined by $f \cdot x = f(x)$.

Theorem 9.1.23 (Jacobson Density Theorem) *Let M be a left simple R-module and $S = End_R(M)$. Let $f \in End_S(M)$, and $x_1, x_2, \ldots, x_r \in M$. Then there exists $a \in R$ such that $f(x_i) = ax_i$ for all i.*

Proof We first prove the result for $r = 1$, and for a semi-simple left R-module M. Assume that M is semi-simple. Then there is a submodule N of M such that $M = Rx_1 \oplus N$. The first projection p_1 is a member of $End_R(M)$. Hence $f(x_1) = f(p_1(x_1)) = p_1 \cdot f(x_1) = p_1(f(x_1))$. Thus, $f(x_1) \in Rx_1$, and so $f(x_1) = ax_1$ for some $a \in R$. Next, assuming that M is simple, we prove the result for arbitrary r. Consider the map f^r from M^r to M^r defined by

$$f^r(u_1, u_2, \ldots, u_r) = (f(u_1), f(u_2), \ldots, f(u_r)).$$

Clearly, $S' = End_R(M^r) = M_r(S)$ is the ring of $r \times r$ matrices with entries in S. Now, f^r preserves addition, and

$$f^r(h \cdot (u_1, u_2, \ldots, u_r)) = f^r(\Sigma_{i=1}^r h_{i1} \cdot u_1, \Sigma_{i=1}^r h_{i2} \cdot u_2, \ldots, \Sigma_{i=1}^r h_{ir} \cdot u_r),$$

where $h = [h_{ij}] \in S'$. Applying the definition of f^r, and observing that $f \in End_S(M)$, it follows that the above is same as $h \cdot f^r(u_1, u_2, \ldots, u_r)$. This shows

that $f^r \in End_{S^r}(M^r)$. Since M^r is semi-simple, for $(x_1, x_2, \ldots, x_r) \in M^r$, there is a $a \in R$ such that $f^r(x_1, x_2, \ldots, x_r) = a \cdot (x_1, x_2, \ldots, x_r)$. This shows that $f(x_i) = ax_i$ for all i. ♯

Remark 9.1.24 The term density theorem for the above result is justified in the following sense. Give M the discrete topology. The set M^M of all maps from M to M can be considered as product of M copies of M. Give the product topology to M^M. $End_S(M)$ is a subset of M^M. Give the subspace topology to $End_S(M)$. Let $a \in R$. The map f_a from M to M defined by $f_a(x) = a \cdot x$ is easily seen to be a member of $End_S(M)$, and the map f from R to $End_S(M)$ defined by $f(a) = f_a$ is a ring homomorphism. The Jacobson density theorem can be restated by saying that $f(R)$ is dense in $End_S(M)$(justify).

Corollary 9.1.25 (Burnside) *Let V be a finite-dimensional vector space over an algebraically closed field F. Then V is simple $End_F(V)$-module, and it cannot be simple over any proper sub-algebra of $End_F(V)$.*

Proof Let v and w be nonzero members of V. Then there is a member T of $End_F(V)$ such that $T(v) = w$. This says that V is a simple left $End_F(V)$-module. Now, let R be a sub-algebra of $End_F(V)$ such that V is left simple over R. We show that $R = End_F(V)$. Since V is a simple left R-module, $End_R(V)$ is a division ring. Let $a \in F$. The map f_a from V to V defined by $f_a(v) = a \cdot v$ is a member of $End_R(V)$, for $f_a(g \cdot v) = a \cdot g(v) = g(a \cdot v) = g \cdot f_a(v)$ for all $g \in End_F(V)$. The map f from F to $End_R(V)$ defined by $f(a) = f_a$ is an embedding of F into $End_R(V)$, for $f_a = f_b$ implies that $a \cdot v = b \cdot v$ for all $v \in V$, and this, in turn, implies that $a = b$. Also if $h \in End_R(V)$, then $(f_a oh)(v) = a \cdot h(v) = h(a \cdot v) = (hof_a)(v)$ for all $a \in F$. This shows that F is embedded as subfield of $End_R(V)$ contained in the center. We identify the embedded subfield by F. More precisely, we identify a and f_a. Let $h \in End_R(V)$. Let $F(h)$ denote the subfield of $End_R(V)$ generated by F and h(note that h commutes with each element of F). Further, observe that $End_R(V)$ is a F-subspace of $End_F(V)$ which is finite dimensional. Hence, $F(h)$ is also a finite-dimensional subspace over F. Thus, $F(h)$ is a finite field extension of F. Since F is algebraically closed, it follows that $h \in F$. This shows that $F = End_R(V)$. Let $\{v_1, v_2, \ldots, v_r\}$ be a basis of V over F. Let $T \in End_F(V) = End_{End_R(V)}(V)$. By the Jacobson density theorem, there is a $h \in R$ such that $T(v_i) = h(v_i)$ for all i, and so $T = h$. This shows that $R = End_F(V)$. ♯

Let F be a field, and G be a finite group. Let V be a finite-dimensional vector space over F which is also a $F(G)$-module. Then the map f from $F(G)$ to $End_F(V)$ defined by $f(\Sigma_{g \in G} \alpha_g g)(v) = (\Sigma_{g \in G} \alpha_g g) \cdot v$ is an injective algebra homomorphism. Thus, $F(G)$ can be thought of as a sub-algebra of $End_F(V)$. The following corollary is restatement of the above corollary in this situation.

Corollary 9.1.26 *Let F be an algebraically closed field, and G be a finite group. Let V be a simple $F(G)$-left module which as F-space is of dimension n. Then the set $\{f_g \mid g \in G\}$ generates $End_F(V)$ as a F-space, where f_g is the linear transformation*

given by $f_g(v) = g \cdot v$. *In particular,* G *contains at least* n^2 *elements, and we have a subset* S *of* G *containing* n^2 *elements such that the set* $\{f_g \mid g \in S\}$ *is a basis of* $End_F(V)$. ♯

Let F be a field, and G be a subgroup of the general linear group $GL(n, F)$. Then every element $\Sigma_{A\in G}\alpha_A A$ of $F(G)$ can also be viewed as a member of $M_n(F) = End_F(F^n)$. In other words, we have a algebra homomorphism from $F(G)$ to $M_n(F)$. This makes F^n a left $F(G)$-module. We say that G is an **irreducible** subgroup if the $F(G)$ - module described above is simple. This amounts to say that given $v \neq 0$ in F^n, and any $w \in F^n$, there is an element $\Sigma_{A\in G}\alpha_A A$ in $F(G)$ such that $\Sigma_{A\in G}\alpha_A A \cdot v = w$, or equivalently, the subspace generated by $\{A \cdot v \mid A \in G\}$ is V. An other way to express this is to say that F^n has no nontrivial proper G-invariant subspace.

Corollary 9.1.27 *Let* F *be an algebraically closed field, and* G *be an irreducible subgroup of* $GL(n, F)$. *Suppose further that the set* $\{Tr(A) \mid A \in G\}$ *is finite, and it contains* m *elements. Then* G *is finite and contains at most* m^{n^2} *elements.*

Proof It follows from the above corollary that the F-subspace of $M_n(F)$ generated by G is $M_n(F)$. We can therefore choose a basis of $M_n(F)$ out of elements of G. Let $\{A^1, A^2, \ldots, A^{n^2}\}$ be a basis of $M_n(F)$, where $A^p = [a_{ij}^p] \in G$ for all $p \leq n^2$. Let $A = [a_{ij}]$ be an arbitrary element of G. Let us denote the trace of $A^p A$ by α_A^p. Thus,

$$\Sigma_{i,j=1}^n a_{ij}^p a_{ji} = \alpha_A^p.$$

This shows that the entries a_{ji} of A are a solution to the system of n^2 linear equations

$$\Sigma_{i,j=1}^n a_{ij}^p x_{ji} = \alpha_A^p$$

in unknowns x_{ji}. Since $\{A^1, A^2, \ldots, A^{n^2}\}$ is a basis of $M_n(F)$, the above system of linear equations has a unique solution. Hence, A is uniquely determined by the trace α_A^p of $A^p A$. Since the number of traces of elements of G is at most m, we have at most m choices for α_A^p for each p. The choices for A, therefore, are at most m^{n^2}. This shows that G contains at most m^{n^2} elements. ♯

Corollary 9.1.28 *Let* G *be a subgroup of* $GL(n, F)$ *with finitely many conjugacy classes, where* F *is an arbitrary field. Then* G *is finite.*

Proof Since $GL(n, F)$ is a subgroup of $GL(n, \overline{F})$, where $\overline{F}$ is algebraic closure of F, we may assume that F is algebraically closed. We may assume that G is irreducible. Since G has only finitely many conjugacy classes, and conjugate elements have same trace, it follows that there are only finitely many traces of elements of G. The result follows from the above corollary. ♯

We give few applications of the above result to the linear groups.

Theorem 9.1.29 (Burnside) *Let F be a field of characteristic 0. Then all finite exponents subgroups of $GL(n, F)$ are finite. Indeed, if G is of exponent m, then it contains at most m^{n^3} elements.*

Proof Clearly, $GL(n, F)$ is a subgroup of $GL(n, \overline{F})$, where $\overline{F}$ is algebraic closure of F. Thus, there is no loss in assuming that F is algebraically closed. The proof is by the induction on n. If $n = 1$, then G is subgroup of $F^\star$ of finite exponent m. Since the number of solutions of the equation $X^m = 1$ in F is at most m, order of a subgroups of $F^\star$ of exponent m is at most $m = m^{1^3}$. Assume that the result is true for subgroups of $GL(r, F)$, where $r < n$. Then we prove the result for subgroups of $GL(n, F)$. Let G be a subgroup of $GL(n, F)$ of finite exponent m, where F is an algebraically closed. Suppose that G is irreducible. Since $A^m = I$ for all $A \in G$, all eigenvalues of elements of G are the mth roots of 1. Since trace of a matrix is sum of its eigenvalues, there are at most m^n traces of the members of G. It follows from the above corollary that G is of order at most $(m^n)^{n^2} = m^{n^3}$. Now, suppose that G is reducible. Then there is a nontrivial proper subspace W of F^n which is invariant under the multiplication by elements of G. Clearly, $DimW = s < n$, and $DimF^n/W = t < n$. We have a homomorphism ρ from G to $GL(W)$ defined by $\rho(A) = A/W$, where A/W is the restriction of the matrix multiplication on F^n to W. Then $\rho(G)$ is a subgroup of $GL(W)$ of exponent at most m. By the induction hypothesis, $\rho(G)$ is finite of order at most m^{s^3}. Let H_1 be the kernel of ρ. Then by the fundamental theorem of homomorphism, H_1 is a normal subgroup of G of index at most m^{s^3}. Clearly, $H_1 = \{A \in G \mid A \cdot w = w \text{ for all } w \in W\}$. Next, since W is invariant under multiplication by the members of G, we have a homomorphism η from G to $GL(F^n/W)$ defined by $\eta(A)(v + W) = A \cdot v + W$. By the induction hypothesis, $\eta(G)$ being of exponent at most m is finite of order at most m^{t^3}. Let H_2 be the kernel of η. Then H_2 is also a normal subgroup of G of index at most m^{t^3}. Let $H = H_1 \bigcap H_2$. Then H is also normal of index at most $m^{s^3} \cdot m^{t^3} = m^{s^3+t^3}$. Since $s + t = n$ and $s^3 + t^3 \leq (s + t)^3$, it follows that G/H is of order at most m^{n^3}. Also H acts trivially on W as well as on F^n/W. Hence, we can find a basis of F^n with respect to which representation of all elements are upper triangular, and all of whose diagonal entries are 1. Since F is of characteristic 0, all nonidentity unitriangular matrices are of infinite order, and so H is trivial. This shows that G is of order at most m^{n^3}. ♯

The group $GL(n, F)$ can be thought of as a subgroup of $GL(n + 1, F)$ by identifying a $n \times n$ matrix A by

$$\begin{bmatrix} A & 0_{n,1} \\ 0_{1,n} & 1 \end{bmatrix}.$$

The union of the chain

$$GL(1, F) \subset GL(2, F) \subset \cdots \subset GL(n, F) \subset GL(n + 1, F) \subset \cdots$$

is a group denoted by $GL(F)$. A subgroup of $GL(F)$ is called a linear group. It is clear that every finitely generated linear group is subgroup of $GL(n, F)$ for sufficiently large n. Thus, we have the following corollary.

Corollary 9.1.30 *Every finitely generated linear group over a field of characteristic 0 is finite if and only if it is of finite exponent.* ♯

Burnside conjectured that all finitely generated groups of finite exponent is finite. This conjecture turns out to be false. In fact, we have uncountably many 2-generator infinite simple groups all of whose nontrivial proper subgroups are cyclic groups of same prime order p (p sufficiently large). As such, another modified conjecture known as restricted Burnside conjecture was framed. The restricted Burnside conjecture asserts that for all n and r, there is a finite group $RB(n, r)$ of exponent r which is generated by n elements such that every n-generator finite group of exponent r is quotient of $RB(n, r)$. This conjecture was finally settled by Zelmanov in 1994.

Theorem 9.1.31 (Schur) *Every torsion subgroup of $GL(n, Q)$ is finite. In fact, there exists a function f from $\mathbb{N}$ to $\mathbb{N}$ such that order of every torsion subgroup of $GL(n, Q)$ is less than or equal to $f(n)$.*

Proof It is sufficient to show the existence of a function g such that order of each finite-order element of $GL(n, Q)$ is at most $g(n)$, for then, using the above result of Burnside, order of each torsion subgroup of $GL(n, \mathbb{Q})$ is at the most $f(n)$, where $f(n) = g(n)^{n^3}$. We show that if m is order of an element of $GL(n, \mathbb{Q})$, then $\phi(m) \leq n$, where ϕ is the Euler's phi function. This, in turn, implies that there is a function g on $\mathbb{N}$ such that m is bounded by a function $g(n)$. The proof of the assertion is by the induction on n. If $n = 1$, then $GL(1, Q) = \mathbb{Q}^\star$. The only elements of $\mathbb{Q}^\star$ of finite order are 1 and -1. Since $\phi(1) = \phi(2) = 1$, the result follows for $n = 1$. Assume the result for all $GL(r, \mathbb{Q})$, where $r < n$. Consider the subgroup $G = < x >$ of $GL(n, \mathbb{Q})$, where x is an element of order m. Suppose that G is irreducible. Then $\mathbb{Q}^n$ is a simple $\mathbb{Q}(G)$ module. Thus, $End_{\mathbb{Q}(G)}(\mathbb{Q}^n) = D$ is a division algebra over $\mathbb{Q}$. Clearly, the center F of D is a subfield containing $\mathbb{Q}$ and x. Since x is of order m, it is root of the cyclotomic polynomial $\Phi_m(X)$ over $\mathbb{Q}$, and $\Phi_m(X)$ is irreducible of degree $\phi(m)$ over $\mathbb{Q}$. Since $x \notin \mathbb{Q}$, it is irreducible polynomial of x. Thus, there exists $v \in Q^n$, $v \neq 0$ such that $\{v, xv, x^2v, \ldots, x^{\phi(m)-1}v\}$ is linearly independent; otherwise, x will be a root of a polynomial of lower degree. This means that $\phi(m) \leq n$. This completes the proof of the theorem. ♯

A matrix A in $M_n(F)$ is called unipotent if all its characteristic roots are 1. Clearly, unipotent matrices are nonsingular.

Proposition 9.1.32 *Let F be an algebraically closed field. Every subgroup G of $GL(n, F)$ consisting of unipotent matrices is conjugate to a subgroup of the subgroup $U(n, F)$ of uni-upper triangular matrices.*

Proof To say that G is conjugate to a subgroup of $U(n, F)$ is to say that there is a basis of F^n such that the matrix representation of linear transformations from F^n to

F^n determined by the multiplications by elements of G are uni-upper triangular. The proof is by the induction on n. If $n = 1$, then there is nothing to do. Assume that the result is true for all subgroups of unipotent transformations in $GL(r, F)$ for $r < n$. We prove it for a subgroup G of $GL(n, F)$. Suppose that G is irreducible. Then, since each member of G is unipotent, the trace of each member of G is n. Hence, G contains at most $1^{n^2} = 1$ element. But, then G is the trivial group. Assume that G is not irreducible. Then there is a subspace W of F^n such that W is invariant under multiplication by elements of G. This defines a homomorphism ρ from G to $GL(W)$ defined by $\rho(A) = A/W$, where A/W is the restriction of the multiplication by A to W. Clearly, $\rho(G)$ consists of unipotent transformations in $GL(W)$. By the induction assumption, we can find a basis $\{w_1, w_2, \ldots, w_s\}$ of elements of W such that the matrix representation of elements of $\rho(G)$ with respect to this basis is uni-upper triangular. Also elements of G induce unipotent transformations on F^n/W. This gives us a homomorphism η from G to $GL(F^n/W)$ such that $\eta(G)$ is a group of unipotent transformations. By the induction hypothesis, we can find a basis $\{v_1 + W, v_2 + W, \ldots, v_t + W\}$ of F^n/W so that the matrix representation of each member of $\eta(G)$ with respect to this basis is uni-upper triangular. Clearly, $\{w_1, w_2, \ldots, w_s, v_1, v_2, \ldots, v_t\}$ is a basis of F^n with respect to which all members of G are uni-upper triangular. ♯

Exercises

9.1.1 Describe all semi-simple modules over a P.I.D.

9.1.2 Describe all simple modules over $\mathbb{R}[X]$.

9.1.3 What are integral domains which are semi-simple rings?

9.1.4 Let G be a cyclic group of order p. Show that $\mathbb{Z}_p(G)$ is not semi-simple.
Hint. $\mathbb{Z}_p^2$ is a $\mathbb{Z}_p$ vector space. It is G-module with respect to the multiplication defined by $a^i \cdot (\overline{u}, \overline{v}) = (\overline{u + iv}, \overline{v})$. Show that $\mathbb{Z}_p \times \{0\}$ is a $\mathbb{Z}_p(G)$ submodule but it is not a direct summand.

9.1.5 Let G be a finite p-group. Show if a $\mathbb{Z}_p(G)$ module M is simple, then it is of dimension 1 over $\mathbb{Z}_p$.

9.1.6 Show that the theorem of Schur is not true in $GL(n, \mathbb{C})$. Is it true in $GL(n, \mathbb{R})$?

9.1.7 Generalize the last result of the section for arbitrary fields.

9.1.8 Let V be a vector space of dimension n over a field F with basis $\{e_1, e_2, \ldots, e_n\}$. Define a $F(S_n)$ module structure on V by $p \cdot \Sigma_{i=1}^n a_i e_i = \Sigma_{i=1}^n a_i e_{p(i)}$. Show that it is not simple. Determine a simple submodule of V and its direct compliment.

9.2 Representations and Group Algebras

Let G be a group and V a vector space over a field F. A homomorphism ρ from G to $GL(V)$ is called a **linear representation** of G over F. We shall be interested in case when V is finite dimensional. Such representations are called finite-dimensional representations. The dimension of V is called the **degree** of the representation. If we fix a basis of V, then we get an isomorphism from $GL(V)$ to $GL(n, F)$, and so a homomorphism from G to $GL(n, F)$. This is also called a representation, or matrix representation of G of degree n.

Let ρ be a representation of a group G on a vector space V over a field. Then V becomes a left $F(G)$-module with respect to the external multiplication $\cdot$ defined by $(\Sigma_{g\in G}\alpha_g g)\cdot v = \Sigma_{g\in G}\alpha_g\rho(g)(v)$. This module will be termed as module associated to the representation ρ. Conversely, suppose that V is a left $F(G)$-module. Then already V is a vector space over F, and for each $g \in G$, the map $\rho(g)$ from V to V defined by $\rho(g)(v) = g \cdot v$ is a linear transformation such that $\rho(g_1g_2) = \rho(g_1)o\rho(g_2)$ for all $g_1, g_2 \in G$, and $\rho(e) = I_V$. It follows that $\rho(g)$ is bijective and $(\rho(g))^{-1} = \rho(g^{-1})$ for all $g \in G$. This says that ρ is a representation of G on V. This representation will be termed as the representation associated to the $F(G)$-module. This correspondence between representations over F and $F(G)$-modules is faithful in the sense that each can be recovered from the other. We have already developed the language of modules. The language of module theory and the representations correspond in the following manner.

1. $F(G)$-module $\longleftrightarrow$ representation over F.
2. $F(G)$ submodule $\longleftrightarrow$ sub-representation.
3. Simple $F(G)$-module $\longleftrightarrow$ irreducible representation.
4. Direct sum of $F(G)$ modules $\longleftrightarrow$ direct sum of representation.
5. If V and W are left $F(G)$ modules, then both of them are vector spaces over F. We can make $V \otimes W$ a $F(G)$ module by defining $g \cdot (v \otimes w) = (g \cdot v) \otimes (g \cdot w)$. The representation thus obtained is called the tensor product of the representation corresponding to $F(G)$ module V and the $F(G)$ module W.
6. Let ρ be a representation of G associated to the $F(G)$ module V. Then the rth exterior power $\bigwedge^r V$ is a vector space of dimension ${}^nC_r$, where n is the dimension of V. This can be made a $F(G)$-module by defining

$$g \cdot (v_1 \bigwedge v_2 \bigwedge \cdots \bigwedge v_r) = g \cdot v_1 \bigwedge g \cdot v_2 \bigwedge \cdots \bigwedge g \cdot v_r.$$

The representation thus obtained is called the rth exterior power of ρ, and it is denoted by $\bigwedge^r \rho$.

7. Let ρ be the representation associated to the $F(G)$-module V. Let $S^r(V)$ denote the rth symmetric power of V. More precisely, $S^r(V) = (\otimes^r V)/A_r)$, where A_r is the subspace of $\otimes^r V$ generated by the elements of the type

$$v_1 \otimes v_2 \otimes \cdots \otimes v_r - v_{p(1)} \otimes v_{p(2)} \otimes \cdots \otimes v_{p(r)},$$

where p is a permutation in S_r. We already have a $F(G)$-module structure on $\otimes^r V$ defined above which affords the rth tensor power of the representation ρ associated to the $F(G)$-module V. It is easily noticed that A_r is a $F(G)$-submodule of $\otimes^r V$. Hence, $S^r V$ is also a $F(G)$ module. The associated representation is called the rth symmetric power of ρ, and it is denoted by $S^r\rho$.

Consider the quotient map ν from $\otimes^2 V$ to $S^2 V$. The kernel of this map is $\bigwedge^2 V$ (verify). Thus, $\otimes^2\rho = S^2\rho \oplus \bigwedge^2 \rho$.

8. Representations associated to isomorphic $F(G)$-modules are called **equivalent**. Suppose that ρ is the representation of G associated to the $F(G)$-module V, and η a representation associated to the $F(G)$-module W. Then ρ is equivalent to η if and only if there is a $F(G)$-module isomorphism T from V to W. Clearly, T is a vector space isomorphism from V to W such that $T(g \cdot v) = g \cdot T(v)$ for all $g \in G$ and $v \in V$, or equivalently, $T(\rho(g)(v)) = \eta(g)(T(v))$ for all $g \in G$ and $v \in V$. This means that $\eta(g) = T\rho(g)T^{-1}$ for all $g \in G$. Thus, ρ is equivalent to η if there is a nonsingular linear transformation T from V to W such that $\eta(g) = T\rho(g)T^{-1}$ for all $g \in G$.

Given a representation ρ from G to $GL(V)$, a representation η from G to $GL(W)$ is called a subrepresentaion if $\rho(g)(W) \subseteq W$ for all $g \in G$, and then $\eta(g)$ is the restriction $\rho(g)/W$ for all $g \in G$. A representation ρ from G to $GL(V)$ is irreducible if there is no nontrivial proper subspace W of V such that $\rho(g)(W) \subseteq W$ for all $g \in G$. The Mashcke's theorem can be restated as follows:

Theorem 9.2.1 *Let G be a group and F a field. Suppose that characteristic of F does not divide $|G|$. Then every representation of G over F is direct sum of irreducible representations over F.* ♯

Thus, to determine all representations of a finite group G over a field whose characteristic does not divide $|G|$, it is sufficient to determine nonequivalent irreducible representations of G.

Example 9.2.2 Let G be a group, and V be a vector space over F. The trivial homomorphism which takes every element of G to the identity map I_V on V is a representation called the trivial representation on V. It is irreducible if and only if $dim V = 1$.

A one-dimensional representation of a group G over a field F is exactly homomorphism from G to $F^\star$. These are also irreducible representation of G. Distinct one-dimensional representations over F are all nonequivalent (why?).

Proposition 9.2.3 *Every irreducible representation of an abelian group over an algebraically closed field is one dimensional.*

Proof Let G be an abelian group, F an algebraically closed field, and ρ an irreducible representation of G on V. Consider $\rho(g)$. Since F is algebraically closed field, $\rho(g)$ has an eigenvalue $\lambda_g \in F$(say). The corresponding eigen subspace $V_{\lambda_g} \neq \{0\}$. Let $h \in G$ and $v \in V_{\lambda_g}$. Then $\rho(g)(\rho(h)(v)) = \rho(gh)(v) = \rho(hg)(v) = \rho(h)(\rho(g)(v)) = \lambda_g\rho(h)(v)$. Thus, V_{λ_g} is invariant under G. Since ρ is irreducible $V_{\lambda_g} = V$, and so $\rho(g)$ is multiplication by a scalar for each $g \in G$. This shows that

each subspace of V is invariant under G. Since ρ is irreducible representation, there should not be any proper subspace of V, and so V is one dimensional. ♯

Thus, to find irreducible representations of abelian groups over an algebraically closed field, it is sufficient to find all homomorphisms from G to $F^\star$.

By the Maschke's theorem, $F(G)$ is semi-simple if the characteristic of F does not divide $\mid G \mid$. By Theorem 9.1.19, it follows that every simple $F(G)$-module is isomorphic to a left ideal of $F(G)$ which is also direct summand of $F(G)$ considered as a left module. It is, therefore, necessary to find the structure of the group algebra $F(G)$. First, we find the division rings $End_{F(G)}(V)$, where V is a simple $F(G)$-module. In case F is an algebraically closed field, we have the following proposition.

Proposition 9.2.4 *Let F be an algebraically closed field, G a finite group, and V a simple $F(G)$ module. Then for any $T \in End_{F(G)}(V)$, there exists a unique $\lambda_T \in F$ such that T is multiplication by λ_T. Further, the map λ from $End_{F(G)}(V)$ to F defined by $\lambda(T) = \lambda_T$ is an isomorphism.*

Proof Let $T \in End_{F(G)}(V)$. Then T is a linear transformation on V. Since F is algebraically closed, T has an eigenvalue λ_T (say). Consider the eigenspace $V_{\lambda_T} \neq \{0\}$. Given any $v \in V_{\lambda_T}$ and $h \in G$, $T(h \cdot v) = h \cdot T(v) = h \cdot \lambda_T v = \lambda_T h \cdot v$. This shows that V_{λ_T} is a $F(G)$-submodule of V. Since V is simple, $V = V_{\lambda_T}$, and so T is multiplication by λ_T. The map λ which takes T to λ_T is clearly an isomorphism. ♯

Theorem 9.2.5 *Let F be an algebraically closed field, and G be a finite group such that characteristic of F does not divide $\mid G \mid$. Then there are only finitely many nonequivalent irreducible representations of degrees $n_1, n_2, \ldots, n_r$ such that the following holds.*

(i) The group algebra $F(G)$ is isomorphic as F algebra to

$$M_{n_1}(F) \times M_{n_2}(F) \times \cdots \times M_{n_r}(F).$$

(ii) $n_1^2 + n_2^2 + \cdots + n_r^2 = \mid G \mid$.

(iii) $n_1 = 1$ *corresponds to the degree of the trivial representation.*

(iv) The number r of nonequivalent irreducible representations is the number of conjugacy classes of G (called the class number of G).

Proof From Theorem 9.1.19 and the above proposition, it follows that there are positive integers $n_1, n_2, \ldots, n_r$ such that $F(G)$ as F algebra is isomorphic to

$$M_{n_1}(F) \times M_{n_2}(F) \times \cdots \times M_{n_r}(F).$$

Comparing the F-dimension of $F(G)$ and that of

$$M_{n_1}(F) \times M_{n_2}(F) \times \cdots \times M_{n_r}(F),$$

we obtain (ii). Clearly, the simple left ideals of the above algebra are isomorphic to the simple left ideals of $M_{n_i}(F)$ for $i = 1, 2, \ldots r$. All simple left ideals of $M_{n_i}(F)$ are isomorphic, and are of dimension n_i. Thus, $n_1, n_2, \ldots, n_r$ represent the dimensions over F of simple left $F(G)$-modules and so they represent the degrees of the irreducible representations of G over F. Since the trivial representation is of degree 1, we may assume that $n_1 = 1$.

Finally, we prove (iv) by comparing the dimension of the center of $F(G)$ and that of algebra

$$M_{n_1}(F) \times M_{n_2}(F) \times \cdots \times M_{n_r}(F).$$

The center of $M_n(F)$ is the ring of all scalar matrices, which is a vectorspace of dimension 1 over F. Hence, the dimension of the center of

$$M_{n_1}(F) \times M_{n_2}(F) \times \cdots \times M_{n_r}(F)$$

is r. Let $\{C_1, C_2, \ldots, C_t\}$ be the set of all distinct conjugacy classes of G. Let $u_i = \Sigma_{x \in C_i} x$. Since $g u_i g^{-1} = \Sigma_{x \in C_i} g x g^{-1} = \Sigma_{y \in C_i} y = u_i$, it follows that u_i is in the center of $F(G)$ for each i. We show that $\{u_1, u_2, \ldots, u_t\}$ is a basis of the center of $F(G)$. Since distinct conjugacy classes are disjoint, and the set G is linearly independent in $F(G)$, it follows that $\{u_1, u_2, \ldots, u_t\}$ is linearly independent. Let $\Sigma_{g \in G} \alpha_g g$ be a member of the center of $F(G)$. Then $h \Sigma_{g \in G} \alpha_g g h^{-1} = \Sigma_{g \in G} \alpha_g g$ for each $h \in G$. Since G is linearly independent, comparing the coefficients, we get that $\alpha_g = \alpha_{hgh^{-1}}$ for all $h \in G$. Thus, $\alpha_g = \alpha_h$ whenever g and h are in same conjugacy class. Let $\alpha_i = \alpha_g$ for each $g \in C_i$. Then $\Sigma_{g \in G} \alpha_g g = \Sigma_{i=1}^{t} \alpha_i u_i$. This shows that $\{u_1, u_2, \ldots, u_t\}$ form a basis of the center of $F(G)$. Thus, t is the dimension of the center of G. Hence, $r = t$ is the number of conjugacy classes of G. ♯

Remark 9.2.6 We shall show later that the degrees of irreducible representations divide the order of the group.

Example 9.2.7 Let F be an algebraically closed field of characteristic different from 2. We find the irreducible representations of the Klein's four group V_4. From the above results, it follows that there are 4 irreducible representations of V_4, and they are all of degree 1. Thus, we have to find all 4 distinct group homomorphisms from V_4 to $F^\star$. We list them. Let ρ_1 denote the trivial homomorphism which maps each element of V_4 to 1. Let ρ_2 denote the map which takes e and a to 1 and b and c to -1. Check that it is indeed a homomorphism. Let ρ_3 denote the map which takes e and b to 1 and the rest to -1. ρ_4 is the map which takes e and c to 1 and the rest to -1. These are the only irreducible representations of V_4. Note that all these irreducible representations of V_4 are realized on any field of characteristic different from 2.

Example 9.2.8 We find all irreducible representations of the Quaternion group Q_8 over an algebraically closed field F of characteristic different from 2. There will be as many irreducible representations of Q_8 as many conjugacy classes of Q_8. There

are 5 conjugacy classes of Q_8. They are $\{1\}$, $\{-1\}$, $\{i, -i\}$, $\{j, -j\}$ and $\{k, -k\}$. Thus, there are 5 irreducible representations of degrees $1, n_2, n_3, n_4, n_5$ such that $1+n_2^2+n_3^2+n_4^2+n_5^2 = 8$. The only possible solution is $n_2 = n_3 = n_4 = 1$, and $n_5 = 2$. In other words, there are 4 irreducible representations including the trivial representation of degrees 1, and there is a unique two-dimensional irreducible representation. We list them. All one-dimensional representations are just homomorphisms from Q_8 to $F^\star$. Note that the kernel of any homomorphism from Q_8 to $F^\star$ contains the commutator subgroup $\{1, -1\}$ of Q_8 (for $F^\star$ is abelian). Since $Q_8/\{1, -1\}$ is isomorphic to the Klein's four group, we get the four homomorphisms from Q_8 to $F^\star$ as in the above example. Thus, we have 4 one-dimensional representations, viz., ρ_1 the trivial representation, ρ_2 the homomorphism which takes 1 and -1 to $1, i, -i$ to 1, and the rest of them to -1. Similarly, we have two other homomorphisms from Q_8 to $F^\star$. Finally, we determine the two-dimensional irreducible representation. Since F is algebraically closed field of characteristic different from 2, $X^4 - 1 = 0$ has 4 distinct roots, which form a cyclic group of order 4. Let ξ denote the primitive 4 roots of unity. Then the map

$$i \rightsquigarrow \begin{bmatrix} \xi & 0 \\ 0 & -\xi \end{bmatrix}, j \rightsquigarrow \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, k \rightsquigarrow \begin{bmatrix} 0 & \xi \\ \xi & 0 \end{bmatrix}$$

defines a representation which is irreducible.

Example 9.2.9 Let F be an algebraically closed field of characteristic different from 2 and 3. We find all the irreducible representations of the symmetric group S_3 over F. Since there are 3 conjugacy classes of S_3, there are 3 irreducible representation of S_3 over F. Suppose that there degrees are $1, n_2, n_3$. Then $1+n_2^2+n_3^2 = 6$. The only possible solution is $n_2 = 1$ and $n_3 = 2$. Thus, there are 2 one-dimensional irreducible representations ρ_1, ρ_2, and 1 two-dimensional irreducible representation ρ_3. The one-dimensional representations ρ_1 and ρ_2 are just homomorphisms from S_3 to $F^\star$. We have a trivial homomorphism ρ_1 from S_3 to $F^\star$ which maps every member of S_3 to 1, and a nontrivial homomorphism ρ_2 from S_3 to $F^\star$ given by $\rho_2(p) = \chi(p)$, where χ is the alternating map. Now, we describe two-dimensional irreducible representation ρ_3. Let V be a vector space over F of dimension 3 with a basis $\{e_1, e_2, e_3\}$. Consider the representation ρ from S_3 to $GL(V)$ defined by $\rho(p)(x_1e_1 + x_2e_2 + x_3e_3) = x_1e_{p(1)} + x_2e_{p(2)} + x_3e_{p(3)}$. Consider the subspace $U = \{\alpha(e_1 + e_2 + e_3) \mid \alpha \in F\}$ of V. Clearly, U is such that $\rho(g)(U) \subseteq U$. In the language of modules U is a $F(S_3)$-submodule of V. The sub-representation thus obtained is the trivial representation ρ_1. Consider the subspace $W = \{x_1e_1 + x_2e_2 + x_3e_3 \mid x_1 + x_2 + x_3 = 0\}$ of V. Clearly, W is of dimension 2, and it is also a $F(S_3)$-submodule. The corresponding representation ρ_3 is 2 dimensional. We show that it is irreducible by showing that this is simple $F(S_3)$ module. Let $w = x_1e_1 + x_2e_2 + x_3e_3$ be a nonzero element of W. Then at least two of x_1, x_2, x_3 are nonzero, and $x_1 + x_2 + x_3 = 0$. Suppose that $x_1 \neq 0 \neq x_2$. We show that there is a permutation $p \in S_3$ such that w and $\rho_3(p)(w) = p \cdot w$ are linearly independent. Suppose not. Then w and $p \cdot w$ are linearly dependent for all $p \in S_3$. Thus, for each $p \in S_3$, there is a scalar α_p such

that $w = \alpha_p p \cdot w$. Taking $p = (2, 3)$ and comparing the coefficients of e_1, e_2, e_3, we find that $\alpha_p = 1$, and $x_2 = x_3$. Similarly, $x_1 = x_2$. Since $x_1 + x_2 + x_3 = 0$, we see that $x_i = 0$ for all i. This is a contradiction. Hence, there is a $p \in S_3$ such that w and $p \cdot w$ are linearly independent. This shows that W has no nontrivial $F(S_3)$-submodule, and so ρ_3 is the two-dimensional irreducible representation. The exterior power $\bigwedge^2 \rho_3$ is a one-dimensional representation which maps even permutations to 1, and the odd permutations to -1 (this representation is called the sign representation).

Exercises

9.2.1 Find all irreducible representations of a group of order 15 over the field $\mathbb{C}$ of complex numbers as well as over the field $\mathbb{Q}$ of rational numbers.

9.2.2 Find all irreducible representations of the dihedral group D_8 and also of A_4 over $\mathbb{C}$ and also over $\mathbb{Q}$.

9.2.3 Find the number of irreducible representations of S_4 over $\mathbb{C}$. Find also the degrees. Determine the structure of $\mathbb{C}(S_4)$. Find some of the irreducible representations of S_4 using the method of the last example of this section.

9.2.4 Show that over any field the number of irreducible representations of a group G can be at the most the class number of G.

9.2.5 Find the number of nonequivalent complex irreducible representation of each of the extra special p-groups of order p^3. Find also their degrees.

9.3 Characters, Orthogonality Relations

Let ρ be a representation of a group G on a finite-dimensional vector space V over a field F. The map χ_ρ from G to F defined by $\chi_\rho(g) = trace\rho(g)$ is called the **character** of G afforded by the representation ρ. Characters afforded by irreducible representations are called **irreducible** characters.

Proposition 9.3.1 *Characters are class functions in the sense that they are constants on conjugacy classes of G.*

Proof Let χ_ρ be the representation afforded by the representation ρ. Then $\rho(ghg^{-1}) = \rho(g)\rho(h)\rho(g)^{-1}$ for all $g, h \in G$. Since similar transformations have same trace, it follows that $\chi_\rho(h) = \chi_\rho(ghg^{-1})$ for all $g, h \in G$. ♯

Proposition 9.3.2 *Equivalent representations afford same characters.*

Proof Let ρ and η be equivalent representations on vector spaces V and W, respectively. Then there is an isomorphism T from V to W such that $\eta(g) = T\rho(g)T^{-1}$ for all $g \in G$. Hence, $\chi_\eta(g) = trace(\eta(g)) = trace(\rho(g)) = \chi_\rho(g)$ for all $g \in G$. ♯

Proposition 9.3.3 *Let ρ_1 and ρ_2 be representations. Then $\chi_{\rho_1 \oplus \rho_2} = \chi_{\rho_1} + \chi_{\rho 2}$ and $\chi_{\rho_1 \otimes \rho_2} = \chi_{\rho_1} \cdot \chi_{\rho_2}$.*

Proof The result follows from the fact that $trace(T_1 + T_2) = traceT_1 + traceT_2$ and $trace(T_1 \otimes T_2) = trace(T_1) \cdot trace(T_2)$. ♯

From the above result, it follows that sums and products of characters are characters, and the set of characters form a semi-ring. We complete this semi-ring to ring by putting negatives of characters called the virtual characters. The ring thus obtained is called the **character** ring, and it is denoted by $Ch(G)$.

Let ρ be the representation afforded by the $F(G)$-module M, μ the representation afforded by the $F(G)$-submodule N of M, and ν the representation afforded by the quotient module M/N. It follows from elementary linear algebra that $trace\rho(g) = trace\mu(g) + trace\nu(g)$ for all $g \in G$. Next, since M is finite-dimensional vector space, there is a composition series of $F(G)$-module M whose factors are simple. This proves the following proposition.

Proposition 9.3.4 *Every character (even if the characteristic F divides the order of the group) is sum of irreducible characters.* ♯

The members of $F(G)$ can be viewed as function from G to F. Indeed, we identify the member $\Sigma_{g \in G} \alpha_g g$ by the function α from G to F defined by $\alpha(g) = \alpha_g$. A character χ of G is, therefore, a member of $F(G)$. Since characters are class functions, they belong to the center of $F(G)$.

Let ρ be a representation of G on a finite-dimensional vector space V over a field F. Let $\{x_1, x_2, \ldots, x_n\}$ be a basis of V. Then we get n^2 functions ρ_{ij} from G to F defined by

$$\rho(g)(x_j) = \Sigma_{i=1}^{n} \rho_{ij}(g) x_i.$$

The character χ_ρ of ρ is given by $\chi_\rho(g) = \Sigma_{i=1}^{n} \rho_{ii}(g)$.

Suppose that the characteristic of F does not divide $\mid G \mid$. Define a map $< , >$ from $F(G) \times F(G)$ to F by

$$< \alpha, \beta > = (\mid G \mid \cdot 1)^{-1} \Sigma_{g \in G} \alpha(g) \beta(g^{-1}),$$

where 1 denotes the identity of the field F. It is easy to observe that $< , >$ is a symmetric bilinear form on $F(G)$. Suppose that $< \alpha, \beta > = 0$ for all β in $F(G)$. Then for each $h \in G$, we have $\alpha(h) = < \alpha, i_{h^{-1}} > = o$, where $i_{h^{-1}}$ is the map from G to F which takes h^{-1} to 1, and the rest of the elements to 0. This shows that $\alpha = 0$, and so $< , >$ is a nondegenerate symmetric bilinear form on $F(G)$. Such a bilinear form is also called an inner product on $F(G)$. Now, we shall show that the set of irreducible characters of G over F form an orthonormal basis of the center of $F(G)$. The results which follow are due to Frobenius, and are called the **orthogonality relations**.

Theorem 9.3.5 *Let G be a finite group, and F be an algebraically closed field whose characteristic does no divide the order of G. Let ρ and η be nonequivalent irreducible representations. Then*

$$(\mid G \mid \cdot 1)^{-1} \Sigma_{g \in G} \rho_{ik}(g^{-1}) \eta_{pj}(g) \; = \; 0$$

for all i, j, k and p.

Proof Let V and W be $F(G)$ modules corresponding to representations ρ and η, respectively. Let $\{x_1, x_2, \ldots, x_n\}$ be a basis of V, and $\{y_1, y_2, \ldots, y_m\}$ be a basis of W. Let T_{ji} be the linear transformation from V to W which takes x_i to y_j and x_k to 0 for $k \neq i$. T_{ji} need not be a $F(G)$-module homomorphism. We average it to make a $F(G)$-module homomorphism. Define $\overline{T_{ji}}$ by

$$\overline{T_{ji}}(v) \; = \; (\mid G \mid \cdot 1)^{-1} \Sigma_{g \in G} \eta(g) T_{ji}(\rho(g^{-1})(v)).$$

Then $\overline{T_{ji}}$ is a $F(G)$-module homomorphism from V to W (see the proof of the Maschke's theorem). Since V and W are simple and nonisomorphic $F(G)$-modules, any $F(G)$-homomorphism from V to W is the 0 map. Since $\overline{T_{ji}}$ is a $F(G)$-homomorphism from V to W, it follows that $\overline{T_{ji}} \; = \; 0$. Hence,

$$0 \; = \; \overline{T_{ji}}(x_k) \; = \; (\mid G \mid \cdot 1)^{-1} \Sigma_{g \in G} \eta(g) (T_{ji}(\Sigma_{l=1}^{n} \rho_{lk}(g^{-1}) x_l).$$

This in turn gives

$$(\mid G \mid \cdot 1)^{-1} \Sigma_{g \in G} \rho_{ik}(g^{-1}) \Sigma_{p=1}^{m} \eta_{pj}(g) y_p \; = \; 0.$$

Since $\{y_1, y_2, \ldots, y_m\}$ is a basis, we see that

$$(\mid G \mid \cdot 1)^{-1} \Sigma_{g \in G} \rho_{ik}(g^{-1}) \eta_{pj}(g) \; = \; 0$$

for all i, j, k and p. ♯

Corollary 9.3.6 *Let χ_ρ and χ_η be distinct irreducible characters of G over an algebraically closed field whose characteristic does not divide $\mid G \mid$. Then $< \chi_\rho, \chi_\eta > \; = \; 0$.*

Proof Since $\chi_\rho \neq \chi_\eta$, ρ and η are nonequivalent. In the above theorem putting $k \; = \; i$ and $p \; = \; j$, and then summing over all i and j, we see that $< \; \chi_\rho, \chi_\eta \; > = \; 0$. ♯

Theorem 9.3.7 *Let ρ be an irreducible representation of a finite group G on a vector space V of dimension n over an algebraically closed field whose characteristic does not divide $\mid G \mid$. Let $\{x_1, x_2, \ldots, x_n\}$ be a basis of V, and $[\rho_{ij}(g)]$ be the matrix of $\rho(g)$ with respect to this basis. Then*

(i) $\Sigma_{g\in G}\rho_{ij}(g^{-1})\rho_{kl}(g) = 0$ *if* $j \neq k$ *or* $i \neq l$ *and*
(ii) $n \cdot \Sigma_{g\in G}\rho_{ij}(g^{-1})\rho_{ji}(g) = \mid G \mid \cdot 1$.

Proof V is a simple $F(G)$-module. By Proposition 9.2.4 every $F(G)$-endomorphism of V is multiplied by a scalar. Fix i and j, and then consider the average $\overline{T_{ji}}$ of the linear transformation T_{ji} which maps x_i to x_j. Suppose that this endomorphism $\overline{T_{ji}}$ of V is multiplication by α_{ji}. Then $\overline{T_{ji}}(x_k) = \alpha_{ji}x_k$ for all i, j and k. Now

$$\overline{T_{ji}}(x_k) = (\mid G \mid \cdot 1)^{-1}\Sigma_{g\in G}\rho(g)(T_{ji}(\rho(g^{-1})(x_k))).$$

In turn, we get that

$$\alpha_{ji}x_k = (\mid G \mid \cdot 1)^{-1}\Sigma_{g\in G}\Sigma_{l=1}^{n}\rho_{lj}(g)\rho_{ik}(g^{-1})x_l.$$

Since $\{x_1, x_2, \ldots, x_n\}$ is a basis, we get that

$$(\mid G \mid \cdot 1)^{-1}\Sigma_{g\in G}\rho_{lj}(g)\rho_{ik}(g^{-1}) = 0$$

if $k \neq l$ and

$$\alpha_{ji} = (\mid G \mid \cdot 1)^{-1}\Sigma_{g\in G}\rho_{kj}(g)\rho_{ik}(g^{-1}).$$

Applying the above argument again, we see that $\alpha_{ji} = 0$ if $i \neq j$. Also

$$\alpha_{ii} = (\mid G \mid \cdot 1)^{-1}\Sigma_{g\in G}\rho_{ki}(g)\rho_{ik}(g^{-1}) = \alpha_{kk}$$

for all i, k. Now,

$$\begin{aligned} n\alpha_{ii} &= (\mid G \mid \cdot 1)^{-1}(\Sigma_{k=1}^{n}(\Sigma_{g\in G}\rho_{ki}(g)\rho_{ik}(g^{-1}))) \\ &= (\mid G \mid \cdot 1)^{-1}(\Sigma_{g\in G}(\Sigma_{k=1}^{n}\rho_{ki}(g)\rho_{ik}(g^{-1}))). \end{aligned}$$

Next, $1 = \rho_{ii}(g^{-1}g) = \Sigma_{k=1}^{n}\rho_{ki}(g)\rho_{ik}(g^{-1})$. Hence

$$n\alpha_{ii} = (\mid G \mid \cdot 1)^{-1}(\mid G \mid \cdot 1) = 1.$$

This shows that

$$n \cdot (\mid G \mid \cdot 1)^{-1}\Sigma_{g\in G}\rho_{ki}(g)\rho_{ik}(g^{-1}) = 1.$$

Multiplying $\mid G \mid \cdot 1$ we get the result. ♯

Corollary 9.3.8 (Orthogonality relation) *Let G be a finite group and F an algebraically closed field whose characteristic does not divide the order of G. Then the set of irreducible characters of G over F form an orthonormal basis of the center of $F(G)$ (which can be interpreted as vector space of class functions on G).*

Proof By Corollary 9.3.6, it follows that distinct irreducible characters are orthogonal. Next, if χ_ρ is an irreducible character afforded by the irreducible representation ρ, then from the above theorem it follows that

$$\begin{aligned} &< \chi_\rho, \chi_\rho > = (\mid G \mid \cdot 1)^{-1} \Sigma_{g \in G} \chi_\rho(g) \chi_\rho(g^{-1}) \\ &= (\mid G \mid \cdot 1)^{-1} (\Sigma_{g \in G} (\Sigma_{i=1}^{n} \rho_{ii}(g) \Sigma_{k=1}^{n} \rho_{kk}(g^{-1}))) \\ &= n \cdot (\mid G \mid \cdot 1)^{-1} \Sigma_{g \in G} \rho_{ii}(g) \rho_{ii}(g^{-1}) = 1. \end{aligned}$$

This shows that the set of irreducible characters form an orthonormal set. Since the number of irreducible characters is the class number of G, and the class number is the dimension of the center (the space of class functions on G) of $F(G)$, it follows that the set of irreducible characters form an orthonormal basis of the center of $F(G)$. ♯

Corollary 9.3.9 *Let G be a finite group, and F be an algebraically closed field of characteristic 0. Then a representation ρ over F is equivalent to a representation η over F if and only if $\chi_\rho = \chi_\eta$.*

Proof Clearly, equivalent representations have same characters. Conversely, suppose that ρ and η are representation such that $\chi_\rho = \chi_\eta$. Let $\{\rho_1, \rho_2, \ldots, \rho_r\}$ be the set of pairwise nonequivalent irreducible representations such that each irreducible representation is equivalent to one of them (r is the class number of G). Then there exists nonnegative integers $n_1, n_2, \ldots, n_r$, and $m_1, m_2, \ldots, m_r$ such that ρ is equivalent to $n_1\rho_1 \oplus n_2\rho_2 \oplus \cdots \oplus n_r\rho_r$, and η is equivalent to $m_1\rho_1 \oplus m_2\rho_2 \oplus \cdots \oplus m_r\rho_r$. Then $\chi_\rho = n_1\chi_{\rho_1} + n_2\chi_{\rho_2} + \cdots + n_r\chi_{\rho_r}$, and $\chi_\eta = m_1\chi_{\rho_1} + m_2\chi_{\rho_2} + \cdots + m_r\chi_{\rho_r}$. Since $\chi_\rho = \chi_\eta$, by the orthogonality relation, we have $n_i \cdot 1 =< \chi_\rho, \chi_{\rho_i} > =< \chi_\eta, \chi_{\rho_i} > = m_i \cdot 1$. Since F is of characteristic 0, we get that $n_i = m_i$ for all i. This shows that ρ is equivalent to η. ♯

Corollary 9.3.10 *Let ρ be a representation of G over an algebraically closed field of characteristic 0. Then ρ is irreducible if and only if $< \chi_\rho, \chi_\rho > = 1$.*

Proof Let $\{\rho_1, \rho_2, \ldots, \rho_r\}$ be a complete set of pairwise nonequivalent irreducible representations. Then $\rho = m_1\rho_1 \oplus m_2\rho_2 \oplus \cdots \oplus m_r\rho_r$, where each m_i is a nonnegative integer. Then $\chi_\rho = m_1\chi_{\rho_1} + m_2\chi_{\rho_2} + \cdots + m_r\chi_{\rho_r}$. From the orthogonality relation, we find that $< \chi_\rho, \chi_\rho > = m_1^2 + m_2^2 + \cdots + m_r^2$. Thus, $< \chi_\rho, \chi_\rho > = 1$ if and only if $m_i = 1$ for a unique i, and the rest of $m_j = 0$. Thus, $< \chi_\rho, \chi_\rho > = 1$ if and only if ρ is equivalent to ρ_i for some i. ♯

Let ρ be a irreducible representation of a finite group G over a field F whose characteristic does not divide $\mid G \mid$. Then V is a simple $F(G)$-module, and we have a homomorphism $\overline{\rho}$ from the ring $F(G)$ to $End_F(V)$ defined by $\overline{\rho}(\Sigma_{g \in G} \alpha_g g) = \Sigma_{g \in G} \alpha_g \rho(g)$. If $\Sigma_{g \in G} \alpha_g g$ is in the center of $F(G)$, then $\overline{\rho}(\Sigma_{g \in G} \alpha_g g)$ commutes with $\rho(g)$ for all $g \in G$, and so it belongs to $End_{F(G)}(V)$. Since V is a simple $F(G)$-module, members of $End_{F(G)}(V)$ are multiplications by scalars. Let $\{C_1, C_2, \ldots, C_r\}$ be the set of distinct conjugacy classes of G. Let $u_i = \Sigma_{g \in C_i} g$. Then as observed, $\{u_1, u_2, \ldots, u_r\}$ form a basis for the center of $F(G)$. From the previous observation $\overline{\rho}(u_i)$ is multiplication by a scalar α_i (say).

Proposition 9.3.11 *Let G be a finite group, F an algebraically closed field of characteristic 0 and ρ, an irreducible representation of G over F. Then the scalars α_i described in the above paragraph are algebraic integers.*

Proof Let $u_i = \Sigma_{g \in C_i} g$, where $\{C_1, C_2, \ldots, C_r\}$ is the set of all distinct conjugacy classes of G. Let $v \in C_k$ and a_{ij}^k denote the cardinality of the set $X_{ij}^v = \{(g, h) \in C_i \times C_j \mid gh = v\}$. If $w \in C_k$, then there is $x \in G$ such that $w = xvx^{-1}$. The map $(g, h) \rightsquigarrow (xgx^{-1}, xhx^{-1})$ is clearly a bijective map from X_{ij}^v to X_{ij}^w. Thus, the integer a_{ij}^k depends only on i, j, k, and not on the choice of $v \in C_k$. This also shows that

$$u_i u_j = \Sigma_{k=1}^r a_{ij}^k u_k.$$

Thus,

$$\overline{\rho}(u_i)\overline{\rho}(u_j) = \overline{\rho}(u_i u_j) = \Sigma_{k=1}^r a_{ij}^k \overline{\rho}(u_k).$$

The left-hand side is multiplication by $\alpha_i \alpha_j$, and the R.H.S. is multiplication by $\Sigma_{k=1}^r a_{ij}^k \alpha_k$. This shows that

$$\alpha_i \alpha_j = \Sigma_{k=1}^r a_{ij}^k \alpha_k$$

for all i, j and k. We can take C_1 to be the conjugacy class $\{e\}$, and so $u_1 = e$. Since $\overline{\rho}(u_1) = \rho(e) = I_V$, $\alpha_1 = 1 \neq 0$. The above equation shows that the column vector

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ . \\ . \\ . \\ \alpha_r \end{bmatrix}$$

is an eigenvector of the matrix $[b_{jk}]$, where $b_{jk} = a_{ij}^k$, and the corresponding eigenvalue is α_i. Thus, α_i is a root of the monic polynomial $det(xI_r - [b_{jk}])$ whose coefficients are all integers (note that $b_{jk} = a_{ij}^k$ are all nonnegative integers). This shows that each α_i is an algebraic integer. ♯

Corollary 9.3.12 *Let G be a finite group, and F be an algebraically closed field of characteristic 0. Let ρ be an irreducible representation over F of degree n. Let $g \in G$. Let $m = [G : C_G(g)]$ be the number of conjugates to g. Then $\frac{m\chi_\rho(g)}{n}$ is an algebraic integer(observe that every algebraically closed field of characteristic 0 contains the field of algebraic numbers).*

Proof Let C_i be the conjugacy class determined by g. Then the trace of $\rho(g) =$ the trace of $\rho(x)$ for each $x \in C_i$. Thus, $trace\overline{\rho}(u_i) = m \cdot trace\rho(g) = m \cdot \chi_\rho(g)$. Since

$\overline{\rho}$ is multiplication by α_i, and the degree of ρ is n, it follows that $trace\overline{\rho}(u_i) = n \cdot \alpha_i$. Thus, $\alpha_i = \frac{m\chi_\rho(g)}{n}$. The result follows from the above theorem. ♯

Corollary 9.3.13 *The degree of every irreducible representation of a finite group G over an algebraically closed field F of characteristic 0 divides the order of the group.*

Proof Let ρ be an irreducible representation of a finite group G of degree n over an algebraically closed field F. Let $\{C_1, C_2, \ldots, C_r\}$ be the set of all distinct conjugacy classes of G. Then χ_ρ is constant on each C_i. Let β_i be the value of χ_ρ on C_i. Then from the above corollary, it follows that $\frac{m_i\beta_i}{n}$ is an algebraic integer, where m_i is the number of elements in C_i. Let p be the permutation of $\{1, 2, \ldots, r\}$ defined by $C_i^{-1} = C_{p(i)}$, where C_i^{-1} is the set of inverses of the members of C_i (note that C_i^{-1} is again a conjugacy class). By the orthogonality theorem, we have

$$1 = < \chi_\rho, \chi_\rho > = \frac{1}{\mid G \mid} \Sigma_{g \in G} \chi_\rho(g)\chi_\rho(g^{-1}).$$

Thus, $\mid G \mid = \Sigma_{i=1}^{r} m_i\beta_i\beta_{p(i)}$. In turn,

$$\frac{\mid G \mid}{n} = \Sigma_{i=1}^{r} \frac{m_i\beta_i}{n}\beta_{p(i)}.$$

From the previous result $\frac{m_i\beta_i}{n}$ is an algebraic integer. Also each β_j is trace of $\rho(g)$ for $g \in C_j$. Since G is finite, $g^t = e$ for some t, and so $\rho(g)^t = I_V$. This shows that the eigenvalues of $\rho(g)$, being roots of unity, are algebraic integers. Since sum of algebraic integers are algebraic integers, it follows that $\beta_{p(i)} = trace\rho(g) =$ the sum of the eigenvalues of $\rho(g)$, $g \in C_{p(i)}$, is an algebraic integer. Again, since sums and products of algebraic integers are algebraic integers, it follows from the above identity that $\frac{|G|}{n}$ is an algebraic integer. We also know that a rational number is an algebraic integer if and only if it is an integer. Thus, $\frac{|G|}{n}$ is an integer. This means that n divides $\mid G \mid$. ♯

Following is a simple application of representation theory.

Proposition 9.3.14 *Let G be a finite simple group of order n, and p be a prime such that the number of conjugacy classes of G is greater than $\frac{n}{p^2}$. Then Sylow p-subgroups of G are abelian.*

Proof We may assume that p^2 divides n. Since G is simple, every nontrivial complex representation of G is injective. Now, $n = 1+n_2^2+n_3^2+\cdots+n_r^2$, where $n_2, n_3, \ldots, n_r$ are the degrees of nontrivial irreducible representations, and r the class number of G. Since the class number r is greater than $\frac{n}{p^2}$, there is $i \geq 2$ such that $n_i < p$. Consider the corresponding irreducible representation ρ_i. Let P be a Sylow p-subgroup of G. Consider the restriction ρ_i/P to P. Then ρ_i/P is a faithful representation of P of degree less than p. The degrees of the irreducible components of ρ_i/P must divide

the $| P | = p^t$, where $t \geq 2$. Since the degree of $\rho_i/P < p$, it follows that all irreducible components of ρ_i/P are of degree 1. Since ρ_i/P is faithful, P is abelian. ♯

Proposition 9.3.15 *Let χ_ρ be an irreducible character afforded by the irreducible representation ρ of degree n of a finite group G over an algebraically closed field F of characteristic 0. Let g be an element of G with m conjugates such that m and n are co-prime. Then $\frac{\chi_\rho(g)}{n}$ is an algebraic integer.*

Proof By the Euclidean algorithm, there exists integers u and v such that $um + vn = 1$. Hence $\frac{\chi_\rho(g)}{n} = u \cdot \frac{m \cdot \chi_\rho(g)}{n} + v\chi_\rho(g)$. By the previous result, $\frac{m \cdot \chi_\rho(g)}{n}$ is an algebraic integer. Also since eigenvalues of $\rho(g)$ are roots of unity (note that $(\rho(g)^{|G|} = I_V)$), it follows that $\chi_\rho(g)$ is an algebraic integer. Since sums and products of algebraic integers are algebraic, it follows that $\frac{\chi_\rho(g)}{n}$ is an algebraic integer. ♯

Proposition 9.3.16 *Under the hypothesis of the above proposition, $\chi_\rho(g) = 0$, or else $\rho(g)$ is multiplication by a scalar.*

Proof We first show that $\rho(g)$ is multiplication by scalar if and only if all the eigenvalues of $\rho(g)$ are same. One way is evident. Suppose that all the eigenvalues of $\rho(g)$ are same. Consider the restriction $\rho/ < g >$ of ρ to the cyclic subgroup generated by g. By the Maschke's theorem $\rho/ < g >$ is direct sum of irreducible representations of $< g >$. Since irreducible representations of $< g >$ are one dimensional, it follows that $\rho(g)$ is diagonalizable. Since all the eigenvalues of $\rho(g)$ are same, it is multiplication by a scalar.

Now, suppose that $\rho(g)$ is not multiplication by a scalar. Then all eigenvalues of $\rho(g)$ are not same. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be eigenvalues of $\rho(g)$. Then each λ_i is a root of unity, and so $| \lambda_i | = 1$. Since all of them are not same $| \chi_\rho(g) | = | \Sigma_{i=1}^n \lambda_i | < n$. By the previous proposition, $\frac{\chi_\rho(g)}{n}$ is an algebraic integer such that $| \frac{\chi_\rho(g)}{n} | < 1$. Let σ be an automorphism of a finite Galois extension K of $\mathbb{Q}$ containing each λ_i. Then $\sigma(\frac{\chi_\rho(g)}{n})$ is also an algebraic integer whose modulus is less than 1. Let $z = \prod_{\sigma \in Aut(K)} \sigma(\frac{\chi_\rho(g)}{n})$. Then z is an algebraic integer and $| z | < 1$. It is clear that $\sigma(z) = z$ for all $\sigma \in Aut(K)$. Since K is a Galois extension of $\mathbb{Q}$, it follows that $z \in \mathbb{Q}$. The only rational algebraic integers are integers, and therefore $z \in \mathbb{Z}$. Since $| z | < 1$, it follows that $z = 0$. This shows that $\frac{\chi_\rho(g)}{n} = 0$, and so $\chi_\rho(g) = 0$. ♯

Following is a criteria, due to Burnside, for non-simplicity of a finite group.

Theorem 9.3.17 *Let G be a finite group which has a conjugacy class containing p^m elements, where p is a prime and $m \geq 1$. Then G can not be simple.*

Proof If G is abelian, then there is nothing to do. Assume that G is non-abelian. Suppose that G is simple, and $g \in G$ be such that there are exactly p^m conjugates to g. Let ρ be a nontrivial irreducible complex representation of degree n such that p does not divide n. Since G is simple and ρ is nontrivial, it follows that ρ is injective, and so G is isomorphic to $\rho(G)$. Suppose that $\chi_\rho(g) \neq 0$. Then from the previous

result $\rho(g)$ is multiplication by a scalar. This means that $\rho(g)$ is in the center of $\rho(G)$. Since $\rho(G)$ is simple, $\rho(g) = I_V$. Again, since ρ is injective, $g = e$ the identity of G. This is a contradiction to the supposition that g has exactly $p^m > 1$ conjugates. Hence, $\chi_\rho(g) = 0$ whenever the degree of ρ is not divisible by p. Let χ_{reg} denote the character of the regular representation ρ_{reg}. Then

$$\chi_{reg} = \chi_1 + n_2\chi_2 + \cdots + n_r\chi_r,$$

where χ_1 is the trivial character, and χ_i is the irreducible character of degree n_i. From Proposition 9.3.16 and the previous observation, it follows that $\chi_{reg}(g) \equiv 1(mod p)$. One also observes that the matrix of $\rho_{reg}(g)$ with respect to the basis G of $F(G)$ has no nonzero entry in the diagonal. Hence, $\chi_{reg}(g) = 0$. This is a contradiction. ♯

Corollary 9.3.18 (Burnside) *Every group of order $p^r q^s$ is solvable, where p and q are primes.*

Proof Assume contrary. Let G be a counter example of the smallest order. Then if H is a nontrivial proper normal subgroup of G, then H and G/H are both solvable. But this will mean that G is solvable. Hence G is simple. Suppose that $| G | = p^r q^s$. Clearly, $r, s \geq 1$. Let Q be a Sylow q-subgroup of G. Let $g \neq e$ be a member of the center of Q. Then $C_G(g)$ contains Q, and it is not G (for then g will be in the center, a contradiction to the assumption of simplicity of G). This shows that $[G : C_G(g)] = p^m$ for some $m \geq 1$. Again from the previous theorem G cannot be simple, a self-contradiction. ♯

Remark 9.3.19 The representation theoretic proof of the above result came quite early in the twentieth century. A nonrepresentation theoretic proof of the result was given by Thomson, Bender, and Goldschmidt quite late around 1976. Now, we have a more general result due to Kegel and Wielandt which says that product of any two nilpotent groups is solvable.

Exercises

9.3.1 Show by means of an example that nonequivalent representations over a field of positive characteristic may have same character.

9.3.2 Show by means of an example that the degree of an irreducible representation over a field F need not divide the order of the group.

9.3.3 Find all irreducible representations of Q_8 and D_8 over $\mathbb{C}$.

9.3.4 Can we realize all complex irreducible representations of a finite group G over $\mathbb{Q}$.

9.3.5 Determine the number of irreducible complex representations of S_4, and also their degrees. Find them explicitly.

9.3.6 Determine the number of irreducible complex representations of A_4, and also find their degrees.

9.3.7 Let F be an algebraically closed field of characteristic 0, and G be a finite group. Let ρ and η be representations associated to $F(G)$ modules V and W, respectively. Show that $< \chi_\rho, \chi_\eta > = Dim_F(Hom_{F(G)}(V, W))$.

9.3.8 Let $G = < a >$ be a cyclic group of prime order p. Consider the group algebra $\mathbb{Q}(G)$. Let $\sigma = \frac{1}{p}\Sigma_{i=0}^{p-1} a^i$ and $\tau = a - \sigma$. Show that the following holds.

(i) $\sigma^2 = \sigma$.
(ii) $\mathbb{Q} \cdot \sigma$ is a subfield isomorphic to $\mathbb{Q}$.
(iii) $(1 - \sigma)^2 = 1 - \sigma)$.
(iv) $\mathbb{Q} \cdot (1 - \sigma)$ is also a subfield isomorphic to $\mathbb{Q}$.
(v) τ is a root of the irreducible polynomial $X^{p-1} + X^{p-2} + \cdots + X + 1$ over $Q \cdot (1 - \sigma)$.
(vi) The subring F of $\mathbb{Q}(G)$ generated by $\mathbb{Q} \cdot (1 - \sigma)$ and τ is isomorphic to $\mathbb{Q}(e^{\frac{2\pi i}{p}})$.
(vii) Every element of $\mathbb{Q}(G)$ can be written uniquely as sum of an element of $\mathbb{Q} \cdot \sigma$ and F.
(viii) Product of any element of $\mathbb{Q} \cdot \sigma$ with an element of F is 0.
(ix) $\mathbb{Q}(G)$ is isomorphic to $\mathbb{Q} \times Q(e^{\frac{2\pi i}{p}})$.

9.3.9 Determine irreducible representations of a cyclic group of order 3 over $\mathbb{Q}$.

9.3.10 Let $G = H \times K$ be the direct product of finite groups H and K. Let F be an algebraically closed field of characteristic 0. Let ρ and η be irreducible representations of H and K on vector spaces V and W, respectively. Let $\rho \otimes \eta$ be the representation of G on $V \otimes W$ defined by $(\rho \otimes \eta)(h, k)(v \otimes w) = \rho(h)(v) \otimes \eta(k)(w)$. Suppose that μ and ν are also irreducible representations of H and K, respectively. Show that

$$< \chi_{\rho\otimes\eta}, \chi_{\mu\otimes\nu} > = 1,$$

if $\rho = \mu$ and $\eta = \nu$, and

$$< \chi_{\rho\otimes\eta}, \chi_{\mu\otimes\nu} > = 0,$$

otherwise. Deduce that these are irreducible representations, and every irreducible representation of G is obtained in this manner.

9.3.11 Show that the Grothendieck group of the group algebra $\mathbb{C}(G)$ is the character ring $Ch(G)$ of G. Find the Grothendieck groups of
$\mathbb{C}(\mathbb{Z}_m)$, $\mathbb{C}(V_4)$, $\mathbb{C}(S_3)$, $\mathbb{C}(Q_8)$ and $\mathbb{C}(D_8)$.

9.4 Induced Representations

Let H be a subgroup of a group G, and ρ be a representation of a group G. Then the restriction of ρ to H denoted by ρ_H is a representation of H. One may observe, by means of an example, that the restriction of an irreducible representation need not be an irreducible representation(the two-dimensional irreducible representation of S_3 when restricted to A_3 is not irreducible).

Now, we describe the adjoint to the restriction. Let H be a subgroup of G, and ρ be a representation of H on W. Then W is a left $F(H)$-module. Since $F(H)$ is a sub-algebra of $F(G)$, we see that $F(G)$ is a bi-$(F(G), F(H))$ module. Hence $V = F(G) \otimes_{F(H)} W$ is a left $F(G)$-module. This gives us a representation of G which we denote by ρ^G, and call it the **induced** representation of G induced by the representation ρ of the subgroup H of G. Let S be a left transversal to H in G. Then $F(G)$ as right $F(H)$-module can be written as $\oplus\Sigma_{x\in S} xF(H)$. Thus, V can be written as

$$V = \oplus\Sigma_{x\in S} x \otimes W.$$

Consider an element $x \otimes w$, $w \in W$, $x \in S$ in one of the direct summands of V. Suppose that $gx = yh, h \in H$ and $y \in S$. Then $\rho^G(g)(x \otimes w) = g(x \otimes w) = gx \otimes w = yh \otimes w = y \otimes hw = y \otimes w'$, where $w' = hw = \rho(h)(w)$. Clearly, $DimV = DimW \cdot [G : H]$. Thus, $deg\rho^G = deg\rho \cdot [G : H]$. If $\{w_1, w_2, \ldots, w_r\}$ is a basis of W and $S = \{x_1, x_2, \ldots, x_s\}$. Then $\{x_i \otimes w_j \mid 1 \leq i \leq s, 1 \leq j \leq r\}$ is a basis of V. The character χ_{ρ^G} of G is denoted by χ_ρ^G, and it is called the **induced character**.

Proposition 9.4.1 *Let H be a subgroup of a finite group G and F a field whose characteristic does not divide $\mid G \mid$. Let ρ be a representation of H. Then*

$$\chi_\rho^G(g) = \frac{1}{\mid H \mid}\Sigma_{x\in G}\chi'_\rho(xgx^{-1}),$$

where $\chi'_\rho = \chi_\rho$ on H and 0 on $G - H$.

Proof Let ρ be a representation of H on W, and $\{w_1, w_2, \ldots, w_r\}$ be a basis of W. Let $S = \{x_1, x_2, \ldots x_s\}$ be a left transversal to H in G. Let $V = F(G) \otimes_{F(H)} W$. Then as observed, $\{x_i \otimes w_j\}$ form a basis of V. Now, the basis element $x_i \otimes w_j$ will contribute in the diagonal entry of $\rho^G(g)$ only if $gx_i = x_i h$ for some $h \in H$, and then $\rho^G(g)(x_i \otimes w_j) = x_i \otimes \rho(h)(w_j)$. Thus, for such a x_i, the sum of the contributions in the diagonal entries of $\rho^G(g)$ corresponding to the set $\{x_i \otimes w_j, \mid, 1 \leq j \leq r\}$ is $\chi_\rho(x_i^{-1}gx_i)$ to the diagonal entry. This shows that

$$\chi_\rho^G(g) = \Sigma_{i=1}^s \chi'_\rho(x_i^{-1}gx_i) = \frac{1}{\mid H \mid}\Sigma_{x\in G}\chi'_\rho(x^{-1}gx).$$

The last equality holds because χ_ρ is a class function. ♯

Theorem 9.4.2 (Frobenius reciprocity Law) *Let H be a subgroup of a finite group G. Let ρ be a representation of H, and η be a representation of G over a field whose characteristic does not divide $\mid G \mid$. Then*

$$< \chi_\rho^G, \chi_\eta >_G = < \chi_\rho, \chi_{\eta_H} >_H,$$

where η_H denotes the restriction of η to H, $<, >_G$ denote the inner product in $F(G)$, and $<, >_H$ the inner product in $F(H)$.

Proof We have

$$< \chi_\rho^G, \chi_\eta >_G = \frac{1}{|G|}\Sigma_{g\in G}\chi_\rho^G(g)\chi_\eta(g^{-1}) = \frac{1}{|G|}\Sigma_{g\in G}(\frac{1}{|H|}\Sigma_{x\in G}\chi'_\rho(x^{-1}gx)\chi_\eta(x^{-1}g^{-1}x)) = \frac{1}{|H|}\Sigma_{y\in G}\chi'_\rho(y)\chi_\eta(y^{-1}) = \frac{1}{|H|}\Sigma_{h\in H}\chi_\rho(h)\chi_{\eta_H}(h^{-1}) = < \chi_\rho, \chi_{\eta_H} >_H .$$

♯

In practice, to determine irreducible representations of a group G, we look at the representations of some special type of subgroups, induce it to G, and then decompose it into irreducible representations.

Remark 9.4.3 Observe that the Frobenious reciprocity holds even if we replace characters by the class functions.

Example 9.4.4 Let H be a subgroup of a finite group G. Let S be a left transversal to H in G. Let 1_H denote the trivial representation of H over a field F whose characteristic does not divide $\mid G \mid$, and $V = \oplus\Sigma_{x\in S}x \otimes F$ the right vector space over F with $\tilde{S} = \{x \otimes 1 \mid x \in S\}$ as a basis. Then the induced representation 1_H^G is the representation of G on V given by $1_H^G(g)(x \otimes 1) = gx \otimes 1 = yk \otimes 1 = y \otimes 1_H(k)(1) = y \otimes 1$, where $gx = yk, y \in S, k \in H$. The character $\chi_{1_H^G}$ is given by $\chi_{1_H^G}(g) = trace 1_H^G(g) = \mid \{x \in S \mid gx = xk \text{ for some } k \in H\} \mid = \mid \{x \in S \mid gxH = xH\} \mid$ for all $g \in G$. Using the Frobenious reciprocity law,

$$< \chi_{1_H^G}, \chi_{1_G} >_G = < \chi_{1_H}, \chi_{1_G}/H >_H = < \chi_{1_H}, \chi_{1_H} >_H = 1.$$

It follows that the trivial representation 1_G of G appears once and only once in the representation of 1_H^G as the direct sum of irreducible representations. More explicitly, $1_H^G = 1_G \oplus s_H(G)$, where $s_H(G)$ is the representation of G with no summands as 1_G. We shall call $s_H(G)$ as the standard representation of G induced by the subgroup H of G. What is $s_{\{e\}}(G)$? Describe the representation $s_H(G)$. Further,

$$\frac{1}{\mid G \mid}\Sigma_{g\in G}\chi_{1_H^G}(g)\chi_{1_G}(g^{-1}) = 1.$$

In turn,

$$\frac{1}{\mid G \mid}\Sigma_{g\in G} \mid \{x \in S \mid \ gxH \ = \ xH\} \mid = \ 1.$$

Now, let θ be a left transitive action of G on X. Then θ induces a representation ρ of G on the vector space FX over F with X as a basis. If H is the isotropy subgroup of the action at a point $x_1 \in X$, then X can be realized as a left transversal to H in G, and the representation ρ is equivalent to 1_H^G. Thus, in this case,

$$\frac{1}{\mid G \mid}\Sigma_{g\in G} \mid \{x \in S \mid \ g\theta x \ = \ x\} \mid = \ 1.$$

More generally, let G be a finite group which acts on a finite set X through a left action θ. Let F be a field whose characteristic does not divide $\mid G \mid$, and V a vector space over F with X as a basis. The action θ of G on X determines a representation ρ of G on V. Let $\{X_1, X_2, \ldots, X_r\}$ be the set of distinct orbits of the action. The action of G on X induces transitive actions of G on each X_i. Further, $V \ = \ FX \ = \ FX_1 \oplus FX_2 \oplus \cdots \oplus FX_r$, and ρ induces representations ρ_i of G on FX_i for each i with $\rho \ = \ \rho_1 \oplus \rho_2 \oplus \cdots \oplus \rho_r$. Let H_i denote the isotropy subgroup of the action at a point $x_i \in X_i$. Then as observed above, $\rho_i \ = \ 1_{H_i}^G$ for each i, and the character $\chi_\rho \ = \ \Sigma_i \chi_{\rho_i}$. In turn, using the Frobenius reciprocity,

$$< \chi_\rho, \ \chi_{1_G} >_G = \ \Sigma_i \ < \chi_{1_{H_i}}, \ \chi_{1_G} >_G = \ \Sigma_i \ < \chi_{1_{H_i}}, \ \chi_{1_{H_i}} >_{H_i} = \ r.$$

We get

$$\frac{1}{\mid G \mid}\Sigma_{g\in G} \mid \{x \in S \mid \ g\theta x \ = \ x\} \mid = \ r,$$

where r is the number of orbits of the action (see Exercise 9.1.11 of Algebra 1). Also note that $1_{\{e\}}^G$ is the regular action ρ_{reg} of G.

Example 9.4.5 Let G be a finite group which acts transitively on a finite set X through a left action θ. Let H denote the isotropy subgroup of the action at a point $x \in X$. Then H also acts on X. Let ρ denote the representation of G associated to the action θ. Then $\rho \ = \ 1_H^G$, and the representation of H associated to the induced action of H on X is the restriction $1_H^G/H$. It follows from the discussion in the above example that the number r of H-orbits of the action is given by

$$r \ = < \chi_{1_H^G/H}, \ \chi_{1_H} >_H = < \chi_{1_H^G}, \ \chi_{1_H^G} >_G = \ \frac{1}{\mid G \mid}\Sigma_{g\in G}\chi_\rho(g)\chi_\rho(g^{-1}).$$

Further, $\chi_\rho(g) \ = \ trace\rho(g)$ is the number of fixed points of the action of the element g on X, and which is the same as the number of fixed points of g^{-1}. This shows that $\chi_\rho(g) \ = \ \chi_\rho(g^{-1})$. Thus,

$$r = \frac{1}{|G|}\Sigma_{g\in G}(\chi_\rho(g))^2 = \frac{1}{|G|}\Sigma_{g\in G}(|\{x \in X \mid g\theta x = x\}|)^2.$$

Let us further assume that G acts doubly transitively on X. Then the isotropy subgroup H of the action at $x_0 \in X$ acts transitively on $X - \{x_0\}$, and so the number of orbits of the action of H on X is 2. From the above discussion, it follows that

$$\frac{1}{|G|}\Sigma_{g\in G}(|\{x \in X \mid g\theta x = x\}|)^2 = 2$$

(see Exercise 9.1.23, Algebra 1). Also $1_H^G = 1_G \oplus s_H(G)$, where 1_G does not appear as a summand in $s_H(G)$. Hence

$$< \chi_{s_H(G)}, \chi_{s_H(G)} >_G = 1.$$

This shows that the standard representation $s_H(G)$ of G is irreducible provided that the action of G on X is transitive as well as doubly transitive (For example, S_n or $A_n, n \geq 4$ acts transitively as well as doubly transitively on a set containing n elements).

Let H and K be subgroups of a group G. A subset $KgH = \{kgh \mid k \in K \text{ and } h \in H\}$ is called a **(K, H) double coset**. The set of all (K, H) double cosets will be denoted by $[K, G, H]$. What are $[\{e\}, G, H]$, $[H, G, \{e\}]$ and $[G, G, H]$? It is easily observed that $[K, G, H]$ is a partition of G. The set of representatives obtained by choosing one and only one member from each (K, H)-double coset is called a **double coset representative system**. For convenience, we choose e to represent double coset KH. Let S be a left transversal to H in G. Then $[K, G, H] = \{KsH \mid x \in S\}$. Further, G and so also K acts on S in a natural manner. It follows that the number of K-orbits of this action is precisely the number $|[K, G, H]|$ of (K, H)-double cosets. Using the arguments in Examples 9.4.4 and 9.4.5, we see that

$$|[K, G, H]| = \frac{1}{|K|}\Sigma_{k\in K}|\{x \in S \mid k\theta x = x\}|.$$

Since $k\theta x = x$ if and only if $k \in x^{-1}Hx \bigcap K$, it follows that

$$|[K, G, H]| = \frac{1}{|K|}\Sigma_{x\in S}|H^{x^{-1}} \bigcap K|.$$

We state few results due to Brauer and Artin without proof.

Theorem 9.4.6 (Artin) *Every character of G over $\mathbb{C}$ is a rational linear combination of characters induced from characters of cyclic subgroups.* ♯

Theorem 9.4.7 (Brauer) *Every character of G is integral linear combination of characters induced by one-dimensional characters of subgroups of G.* ♯

Exercises

9.4.1 Show that the regular ρ_{reg} representation of G is the same as the induced representation of the trivial representation of the trivial subgroup.

9.4.2 Let H be a subgroup of G of finite index. Let W be the F-vector space with $(G/H)^l$ as a basis. Then the action of G on $(G/H)^l$ gives rise to a representation of G. Show that this representation is the representation induced by the trivial representation of H. When can this representation be irreducible?

9.4.3 Describe irreducible components of all representations of S_3 induced by the representations of proper subgroups.

9.4.4 Call a group G to be a Frobenius group, if it has a proper subgroup H such that $H \bigcap xHx^{-1} = \{e\}$ for all $x \in G - H$. It is a fact, which can be proved using induced representation theory, that $N = G - \bigcup_{g\in G} g(H - \{e\})g^{-1}$ is a normal subgroup, called the Frobenius kernel, such that $G = HN$ and $H \bigcap N = \{e\}$. Show that a finite group G is a Frobenius group if and only if it is a transitive nonregular permutation group in which no nonidentity element has more than one fixed point.

9.4.5 Show that D_{4n+2} is a Frobenius group.

9.4.6 Let G be a Frobenius group with Frobenius kernel N, and the Frobenius compliment H. Show that $\mid H \mid$ divides $\mid N \mid -1$.

9.4.7 Let H be a cyclic group of order n_H. Define a map μ_H from H to $\mathbb{Z}$ by $\mu_H(h) = n_H$ if h is a generator of H, and 0 otherwise. Show that μ_H is a class function. Let $\nu_H = \phi(n_H)\chi_{reg} - mu_H$, where χ_{reg} is the regular character of H (note that ν_H is zero map on trivial cyclic group). Show that ν_H is also a class function on H.

9.4.8 Let G be a finite group of order m. Let $\chi = \chi_{reg} - \chi_{I_G}$. Using the Frobenious reciprocity, show that for any class function η on G,

$$< m\chi,\ \eta >_G = \Sigma_{H\in\Omega} < \nu_H^G,\ \eta >_G,$$

where Ω is the set of subgroups of G. Deduce that $m\chi = \Sigma_{H\in\Omega}\nu_H^G$.

9.4.9* Let η be a degree 1 character of a cyclic group H. Show that

$$< \nu_H,\ \eta >_H = \Sigma_{h\in X}(1 - \eta(h)),$$

where X is the set of generators of H. Using the fact that $\eta(h)$ is an algebraic integer, deduce that $< \nu_H,\ \eta >_H$ is positive integer.

9.4.10 Using the above exercises, show that χ (defined in Exercise 9.4.8) is positive linear combination of characters induced by degree 1 characters of cyclic subgroups of G.

9.4.11 Let H be a subgroup of a finite group G, and ρ be a representation of H on a finite-dimensional vector space W. Let ρ^G be the induced representation of G. Let K be a subgroup of G and S a set of (K, H)-double coset representative system. Let H_x denote the subgroup $x^{-1}Hx \bigcap K$. Let ρ_x denote the representation of H_x defined by $\rho_x(a) = \rho(xax^{-1})$ and ρ_x^K the induced representation of K induced by ρ_x. Let $res_K(\rho^G)$ denote the restriction of ρ^G to K. Show that

$$res_K(\rho^G) = \oplus \Sigma_{x \in S} \rho_x^K.$$

9.4.12 Refer to Exercise 9.4.11 with $K = H$. Show that ρ^G is irreducible if and only if ρ is irreducible, and $< \chi_{\rho_x}, \chi_{res_{H_x}(\rho)} > = 0$. This result is termed as **Mackey irreducibility criteria**.

Chapter 10
Group Extensions and Schur Multiplier

The Chap. 8 was devoted to the field extensions and Galois Theory. This chapter centers around the study of group extension and Schur multiplier. The guiding problem in Group Theory is to classify groups up to isomorphisms. The solution, in general, is beyond the dream to mathematicians. However, mathematicians always roam around this problem. Let us restrict our self to the problem of classifying finite groups upto isomorphisms. Every finite group has a composition series, and the composition length is an invariant of the group. If

$$G = G_1 \rhd G_2 \rhd \cdots \rhd G_n \rhd G_{n+1} = \{e\}.$$

is a composition series of G, then G_i/G_{i+1} is a finite simple group for each i. As such, the problem of classifying finite groups reduces to the following two problems:

1. Classify all finite simple groups.

2. Given a finite group H and a finite simple group K, to classify all groups G (up-to isomorphism) having H as a normal subgroup such that G/H is isomorphic to K.

Finite simple groups have been classified. They are of the following four types:

(i) Prime Cyclic groups.
(ii) The alternating groups A_n, $n \geq 5$.
(iii) Finite simple groups of Lie types such as $PSL(n, q)$.
(iv) 26 Sporadic simple groups.

The reader is referred to the book "Finite simple groups: An introduction to their Classification, by D. Gorenstein" for their detailed description.

The solution to the problem 2 is still beyond the dream to mathematicians, and it is addressed in the theory of extensions of groups and co-homology theory of groups. In this chapter, for convenience, we may frequently use the language of category theory. The reader may refer to the appendix of the Algebra 1 for the purpose.

R. Lal, *Algebra 2*, Infosys Science Foundation Series in Mathematical Sciences,
DOI 10.1007/978-981-10-4256-0_10

10.1 Schreier Group Extensions

In this section, we shall describe Schreier theory of group extensions.

A sequence

$$\cdots \overset{\alpha_{n-2}}{\to} G_{n-1} \overset{\alpha_{n-1}}{\to} G_n \overset{\alpha_n}{\to} G_{n+1} \overset{\alpha_{n+1}}{\to} \cdots$$

of groups G_n together with homomorphisms α_n is said to be an **exact** sequence at G_n if *image* $\alpha_{n-1} = ker\alpha_n$. The sequence is said to be **exact** if it is exact at each G_n.

A finite term exact sequence of the type

$$1 \longrightarrow H \overset{\alpha}{\to} G \overset{\beta}{\to} K \longrightarrow 1$$

with 1 representing the trivial group is called a **short exact** sequence. Thus, to say that the above sequence is exact is to say that α is injective, β is surjective, and *image* $\alpha = ker\ \beta$. In particular, $\alpha(H)$ is a normal subgroup of G such that β induces an isomorphism from $G/\alpha(H)$ to K. The above short exact sequence is also termed as an **extension** of H by K. By the abuse of language, we also say that G is an extension of H by K.

Example 10.1.1 For any positive integer m, we have the short exact sequence

$$\{0\} \longrightarrow m\mathbb{Z} \overset{i}{\to} \mathbb{Z} \overset{\nu}{\to} \mathbb{Z}_m \longrightarrow \{0\},$$

where i is the inclusion map, and ν is the quotient map. Thus, this is an extension of mZ by $\mathbb{Z}_m$. We have another extension of $m\mathbb{Z}$ by $\mathbb{Z}_m$ given by the short exact sequence

$$\{0\} \longrightarrow m\mathbb{Z} \overset{i_1}{\to} m\mathbb{Z} \oplus \mathbb{Z}_m \overset{p_2}{\to} \mathbb{Z}_m \longrightarrow \{0\},$$

where i_1 is the inclusion in the first component, and p_2 is the second projection. Note that $\mathbb{Z}$ is not isomorphic to $m\mathbb{Z} \oplus \mathbb{Z}_m$.

Example 10.1.2 We have the exact sequence

$$\{0\} \longrightarrow A_3 \overset{i}{\to} S_3 \overset{\chi}{\to} \{1, -1\} \longrightarrow \{0\},$$

where i is the inclusion map, and χ is the alternating map. Note that A_3 is a cyclic group of order 3, and $\{1, -1\}$ is cyclic group of order 2. We have another extension of a cyclic group of order 3 by a cyclic group of order 2 given by the exact sequence

$$\{0\} \longrightarrow \mathbb{Z}_3 \overset{i}{\to} \mathbb{Z}_6 \overset{\nu}{\to} \mathbb{Z}_2 \longrightarrow \{0\},$$

where $\mathbb{Z}_3$ is included as a cyclic group of order 3 in $\mathbb{Z}_6$, and ν is the corresponding quotient map. Note that S_3 is not isomorphic to $\mathbb{Z}_6$.

Example 10.1.3 Let G be a group, η the natural map from G to $Aut(G)$ given by $\eta(g) = f_g$ the inner automorphism determined by g ($f_g(x) = gxg^{-1}$), and ν the natural quotient map from $Aut(G)$ to $Out(G)$. Then the sequence

$$\{e\} \longrightarrow Z(G) \overset{i}{\to} G \overset{\eta}{\to} Aut(G) \overset{\nu}{\to} Out(G) \longrightarrow 1$$

is an exact sequence.

Example 10.1.4 The sequences

$$\{0\} \longrightarrow \mathbb{Z} \overset{(i_1,0)}{\to} \mathbb{Z} \oplus \mathbb{Z} \overset{p_2}{\to} \mathbb{Z} \longrightarrow \{0\}$$

and

$$\{0\} \longrightarrow \mathbb{Z} \overset{(i_1,-i_2)}{\to} \mathbb{Z} \oplus \mathbb{Z} \overset{p_1+p_2}{\to} \mathbb{Z} \longrightarrow \{0\},$$

where $(i_1, 0)$ is the first inclusion given by $n \mapsto (n, 0)$, p_2 is the second projection, $(i_1, -i_2)$ is the map given by $n \mapsto (n, -n)$, and $p_1 + p_2$ is the map given by $(n, m) \mapsto n + m$, are short exact sequences. As such, both are extensions of $\mathbb{Z}$ by $\mathbb{Z}$. Note that the middle term is also same.

Let ***EXT*** denote the category (see appendix of algebra 1 for the notions in category theory) whose objects are short exact sequences

$$1 \longrightarrow H \overset{\alpha}{\to} G \overset{\beta}{\to} K \longrightarrow 1$$

of groups, and a morphism between two extensions E_1 and E_2 given by the short exact sequences

$$1 \longrightarrow H_1 \overset{\alpha_1}{\to} G_1 \overset{\beta_1}{\to} K_1 \longrightarrow 1$$

and

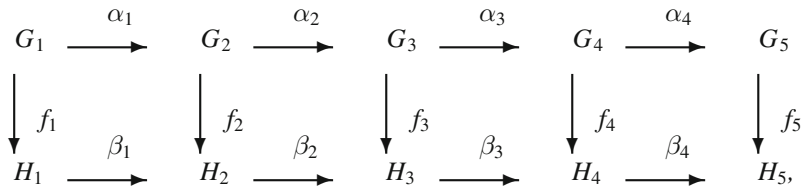
$$1 \longrightarrow H_2 \overset{\alpha_2}{\to} G_2 \overset{\beta_2}{\to} K_2 \longrightarrow 1$$

is a triple (λ, μ, ν), where λ is a homomorphism from H_1 to H_2, μ is a homomorphism from G_1 to G_2, and ν is a homomorphism from K_1 to K_2 such that the following diagram is commutative:

$$\begin{array}{ccccccccc}
1 & \longrightarrow & H_1 & \overset{\alpha_1}{\longrightarrow} & G_1 & \overset{\beta_1}{\longrightarrow} & K_1 & \longrightarrow & 1 \\
 & & \downarrow{\scriptstyle\lambda} & & \downarrow{\scriptstyle\mu} & & \downarrow{\scriptstyle\nu} & & \\
1 & \longrightarrow & H_2 & \overset{\alpha_2}{\longrightarrow} & G_2 & \overset{\beta_2}{\longrightarrow} & K_2 & \longrightarrow & 1
\end{array}$$

The category ***EXT*** is called the category of **Schreier extensions** of groups. The isomorphisms in this category are called the **equivalences of extensions**.

Theorem 10.1.5 (Five Lemma) *Consider the following commutative diagram*



where rows are exact sequences of groups, and the vertical maps are homomorphisms.

(i) If f_1 is surjective and f_2, f_4 are injective, then f_3 is injective.

(ii) If f_5 is injective and f_4, f_2 are surjective, then f_3 is surjective.

(iii) If f_1, f_2, f_4, f_5 are isomorphisms, then f_3 is also an isomorphism.

Proof (i). The proof is the imitation of the proof of the five lemma (Theorem 7.2.3) for modules. However, we repeat those arguments again. Suppose that f_1 is surjective, f_2 and f_4 are injective. We have to show that f_3 is injective. Let $g_3 \in G_3$ such that $f_3(g_3) = e$ (e will denote the identity of all the groups under consideration). Then $e = \beta_3(e) = \beta_3(f_3(g_3)) = f_4(\alpha_3(g_3))$ (from the commutativity of the diagram). Since f_4 is injective, $\alpha_3(g_3) = e$. Thus, $g_3 \in ker\alpha_3 = image\alpha_2$ (exactness), and hence there is an element$g_2 \in G_2$ such that $\alpha_2(g_2) = g_3$. Further, $e = f_3(g_3) = f_3(\alpha_2(g_2)) = \beta_2(f_2(g_2))$ (commutativity of the diagram). Thus, $f_2(g_2) \in ker\beta_2 = image\beta_1$ (exactness). Hence, there exists $h_1 \in H_1$ such that $\beta_1(h_1) = f_2(g_2)$. Since f_1 is surjective, there is an element$g_1 \in G_1$ such that $f_1(g_1) = h_1$. Now, $f_2(\alpha_1(g_1)) = \beta_1(f_1(g_1)) = \beta_1(h_1) = f_2(g_2)$. Since f_2 is injective, $\alpha_1(g_1) = g_2$. But, already $\alpha_2(g_2) = g_3$. Hence $g_3 = \alpha_2(\alpha_1(g_1)) = e$ (for $image\alpha_1 = Ker\alpha_2$). This shows that f_3 is injective.

(ii). Suppose that f_5 is injective, f_2 and f_4 are surjective. Let $h_3 \in H_3$. We have to show the existence of a $g_3 \in G_3$ such that $f_3(g_3) = h_3$. Now, $\beta_3(h_3) \in H_4$. Since f_4 is surjective, there is an element $g_4 \in G_4$ such that $f_4(g_4) = \beta_3(h_3)$. Now, $f_5(\alpha_4(g_4)) = \beta_4(f_4(g_4))$ (commutativity of the diagram)$= \beta_4(\beta_3(h_3)) = e$ (exactness). Since f_5 is injective, $\alpha_4(g_4) = e$. Thus, $g_4 \in ker\alpha_4 = image\alpha_3$. Hence there is an element $g_3 \in G_3$ such that $\alpha_3(g_3) = g_4$. Since $\beta_3(f_3(g_3)) = f_4(\alpha_3(g_3)) = f_4(g_4) = \beta_3(h_3)$, $\beta_3(h_3(f_3(g_3))^{-1}) = e$. Thus, $h_3(f_3(g_3))^{-1} \in ker\beta_3 = image\beta_2$. Hence there exists $h_2 \in H_2$ such that $\beta_2(h_2) = h_3(f_3(g_3))^{-1}$. Since f_2 is surjective, there is an element $g_2 \in G_2$ such that $f_2(g_2) = h_2$. Now $h_3(f_3(g_3))^{-1} = \beta_2(h_2) = \beta_2(f_2(g_2)) = f_3(\alpha_2(g_2))$. This shows that $f_3(\alpha_2(g_2)g_3) = h_3$, and so f_3 is surjective.

(iii). Follows from (i) and (ii). ♯

Corollary 10.1.6 *Let the triple (λ, μ, ν) be a morphism between two extensions E_1 and E_2. Suppose that λ and ν are isomorphisms. Then μ is also an isomorphism.* ♯

Corollary 10.1.7 *The triple* (λ, μ, ν) *is an isomorphism in the category* ***EXT*** *if and only if* λ *and* ν *are isomorphisms between corresponding groups.* ♯

Now, we shall give another description of an equivalence class in this category. Let

$$1 \longrightarrow H \overset{\alpha}{\rightarrow} G \overset{\beta}{\rightarrow} K \longrightarrow 1 \cdots\cdots \tag{10.1}$$

be an extension of H by K. Since β is surjective, there is a map t (not necessarily a homomorphism) from K to G with $t(e) = e$ (called a **section** or a **transversal**) such that $\beta ot = I_K$ (note that we are using the axiom of choice). $\alpha(H) = ker\beta$ is a normal subgroup of G. Thus, for each $x \in K$ and $h \in H$, $t(x)\alpha(h)t(x)^{-1}$ belongs to $\alpha(H)$. Since α is injective, there is a unique element $\sigma_x^t(h)$ in H depending on x and h such that

$$t(x)\alpha(h)t(x)^{-1} = \alpha(\sigma_x^t(h)) \cdots\cdots . \tag{10.2}$$

This gives us a map σ_x^t from H to H given by (10.2). Suppose that $\sigma_x^t(h_1) = \sigma_x^t(h_2)$. Then $t(x)\alpha(h_1)t(x)^{-1} = \alpha(\sigma_x^t(h_1)) = \alpha(\sigma_x^t(h_2)) = t(x)\alpha(h_2)t(x)^{-1}$. Hence $\alpha(h_1) = \alpha(h_2)$. Since α is injective, $h_1 = h_2$. This shows that σ_x^t is an injective map from H to H. Next, let $h \in H$. Then there is an element $a \in H$ such that $t(x)^{-1}(\alpha(h))t(x) = \alpha(a)$. Now, $t(x)\alpha(a)t(x)^{-1} = \alpha(h)$. By the definition $\sigma_x^t(a) = h$. This shows that σ_x^t is also a surjective map from H to H. Again,

$\alpha(\sigma_x^t(h_1h_2))$
$= t(x)\alpha(h_1h_2)t(x)^{-1}$
$= t(x)\alpha(h_1)\alpha(h_2)t(x)^{-1}$
$= t(x)\alpha(h_1)t(x)^{-1}t(x)\alpha(h_2)t(x)^{-1}$
$= \alpha(\sigma_x^t(h_1))\alpha(\sigma_x^t(h_2))$
$= \alpha(\sigma_x^t(h_1)\sigma_x^t(h_2))$.

Since α is injective, $\sigma_x^t(h_1h_2) = \sigma_x^t(h_1)\sigma_x^t(h_2)$. This shows that σ_x^t is an automorphism of H.

Thus, given an extension

$$1 \longrightarrow H \overset{\alpha}{\rightarrow} G \overset{\beta}{\rightarrow} K \longrightarrow 1$$

of H by K, every section t from K to H determines a map σ^t from K to $Aut(H)$ given by the Eq. 10.2 (note that σ^t depends on the chosen section t). Since $t(e) = e$,

$$\sigma_e^t = I_H \cdots\cdots \tag{10.3}$$

Further, $\beta(t(x)t(y)) = \beta(t(x))\beta(t(y)) = xy = \beta(t(xy))$. Hence $(t(x)t(y))(t(xy))^{-1}$ belongs to $ker\beta = image\alpha$. Thus, there is a unique element $f^t(x, y) \in H$ depending on t, x and y such that

$$t(x)t(y) = \alpha(f^t(x, y))t(xy) \cdots\cdots . \tag{10.4}$$

Again, as $t(e) = e$,

$$f^t(e, y) = e = f^t(x, e) \cdots\cdots \tag{10.5}$$

for all $x, y \in K$.

For $x, y, z \in K$,

$$(t(x)t(y))t(z) = \alpha(f^t(x, y))t(xy)t(z) = \alpha(f^t(x, y))\alpha(f^t(xy, z))t((xy)z) = \alpha(f^t(x, y)f^t(xy, z))t((xy)z)$$

On the other hand,

$$t(x)(t(y)t(z)) = t(x)\alpha(f^t(y, z))t(yz) = t(x)\alpha(f^t(y, z))t(x)^{-1}t(x)t(yz) = \alpha(\sigma_x^t(f^t(y, z))\alpha(f^t(x, yz))t(x(yz)) = \alpha(\sigma_x^t(f^t(y, z))f^t(x, yz))t(x(yz))$$

Equating both the expression for $t(x)t(y)t(z)$, we find that

$$\alpha(f^t(x, y)f^t(xy, z)) = \alpha(\sigma_x^t(f^t(y, z))f^t(x, yz)).$$

Since α is injective,

$$f^t(x, y)f^t(xy, z) = \sigma_x^t(f^t(y, z))f^t(x, yz) \cdots\cdots . \tag{10.6}$$

Next, for $x, y \in K$ and $h \in H$

$$(t(x)t(y))\alpha(h) = \alpha(f^t(x, y))t(xy)\alpha(h) = \alpha(f^t(x, y))t(xy)\alpha(h)t(xy)^{-1}t(xy) = \alpha(f^t(x, y))\alpha(\sigma_{xy}^t(h))t(xy) = \alpha(f^t(x, y)\sigma_{xy}^t(h))t(xy).$$

On the other hand,

$$t(x)(t(y)\alpha(h)) = t(x)(t(y)\alpha(h)t(y)^{-1}t(y)) = t(x)(\alpha(\sigma_y^t(h))t(y)) = t(x)(\alpha(\sigma_y^t(h)))t(x)^{-1}t(x)t(y) = \alpha(\sigma_x^t(\sigma_y^t(h)))\alpha(f^t(x, y))t(xy) = \alpha(\sigma_x^t(\sigma_y^t(h))f^t(x, y))t(xy).$$

Equating both the expression for $t(x)t(y)\alpha(h)$, and using the injectivity of α, we find that

$$f^t(x, y)\sigma_{xy}^t(h) = \sigma_x^t(\sigma_y^t(h))f^t(x, y) \cdots \tag{10.7}$$

We are prompted to have the following definition.

Definition 10.1.8 A Quadruple (K, H, σ, f), where K and H are groups, σ a map from K to $aut(H)$, and f a map from $K \times K$ to H, is called a **factor System** if the following conditions hold:
(i) $\sigma_e = I_H$ (For convenience we denote the image of x under the map σ by σ_x).
(ii) $f(x, e) = 1 = f(e, y)$ for all $x, y \in K$, where 1 denotes the identity of H, and e denotes the identity of K.
(iii) $f(x, y)f(xy, z) = \sigma_x(f(y, z))f(x, yz)$ for all $x, y, z \in K$.
(iv) $f(x, y)\sigma_{xy}(h) = \sigma_x(\sigma_y(h))f(x, y)$ for all $x, y \in K$ and $h \in H$.

Remark 10.1.9 The condition (iii) can be viewed as the non-abelian version of a 2 co-cycle.

The proof of the following proposition follows from the discussions which motivated the Definition 10.1.8.

Proposition 10.1.10 *Every extension E of H by K with a choice of a section t determines a factor system $Fac(E, t) \equiv (K, H, \sigma^t, f^t)$, where σ^t and f^t are described by the Eqs. (10.2) and (10.4) above. This factor system is termed as a factor system associated to the extension E.* ♯

Conversely, we have the following proposition.

Proposition 10.1.11 *Let (K, H, σ, f) be a factor system. Then there exists an extension E of H by K, and a section t of E such that $Fac(E, t) = (K, H, \sigma, f)$.*

Proof Let $G = H \times K$. Define a product $\cdot$ in G by

$$(a, x) \cdot (b, y) = (a\sigma_x(b)f(x, y), xy).$$

Using (i) and (ii) in the definition, $(1, e) \cdot (b, y) = (1\sigma_e(b)f(e, y), ey) = (b, y)$, and also $(a, x) \cdot (1, e) = (a\sigma_x(1)f(x, e), xe) = (a, x)$. This shows that $(1, e)$ is the identity of G. Let $(a, x) \in G$. To find the inverse of (a, x), we have to find a (b, y) so that $(1, e) = (a, x) \cdot (b, y) = (a\sigma_x(b)f(x, y), xy)$. Obviously, then, $y = x^{-1}$, and b should be such that $a\sigma_x(b)f(x, x^{-1}) = 1$. Since σ_x is a bijective map on H, $b = \sigma_x^{-1}(a^{-1}f(x, x^{-1})^{-1})$. More precisely, $(\sigma_x^{-1}(a^{-1}f(x, x^{-1})^{-1}), x^{-1})$ is the inverse of (a, x). Finally, to ensure that G is a group, we have to establish the associativity of $\cdot$. Now,

$$((a, x) \cdot (b, y)) \cdot (c, z) = (a\sigma_x(b)f(x, y), xy) \cdot (c, z) = (a\sigma_x(b)f(x, y)\sigma_{xy}(c)f(xy, z), (xy)z).$$

On the other hand,

$$(a, x) \cdot ((b, y) \cdot (c, z)) = (a, x) \cdot (b\sigma_y(c)f(y, z), yz) = (a\sigma_x(b\sigma_y(c)f(y, z))f(x, yz), x(yz)).$$

Since the multiplication in K is already associative, to ensure the associativity in G, we need to show that

$$a\sigma_x(b)f(x,y)\sigma_{xy}(c)f(xy,z) \;=\; a\sigma_x(b\sigma_y(c)f(y,z))f(x,yz)$$

for all a, x, b, y, c, z. Since σ_x is an automorphism, we need to verify that

$$f(x,y)\sigma_{xy}(c)f(xy,z) \;=\; \sigma_x(\sigma_y(c))\sigma_x(f(y,z))f(x,yz)$$

Using the (iii) part of the Definition 10.1.8, the RHS is transformed to

$$\sigma_x(\sigma_y(c))f(x,y)f(xy,z).$$

Hence we need to verify that

$$f(x,y)\sigma_{xy}(c) \;=\; \sigma_x(\sigma_y(c))f(x,y)$$

This is true because of the part (iv) of the Definition 10.1.8. Thus, G is a group with respect to the product $\cdot$ defined above. The map α from H to G defined by $\alpha(h) = (h, e)$, and the map β from G to K defined by $\beta(a, x) = x$ are easily seen to be homomorphisms. In turn, we get the extension

$$E \equiv 1 \longrightarrow H \overset{\alpha}{\to} G \overset{\beta}{\to} K \longrightarrow 1$$

of H by K. Consider the section t of E given by $t(x) = (1, x)$. Then $t(x) \cdot t(y) = (1, x) \cdot (1, y) = (f(x, y), xy) = (f(x, y), e) \cdot (1, xy) = \alpha(f(x, y)) \cdot t(xy)$. This shows that $f^t(x, y) = f(x, y)$ for all $x, y \in K$. Thus, $f^t = f$. Again, $t(x) \cdot \alpha(h) \cdot t(x)^{-1} = (1, x) \cdot (h, e) \cdot (1, x)^{-1} = (\sigma_x(h)f(x, e), x) \cdot ((\sigma_x)^{-1}((f(x, x^{-1}))^{-1}), x^{-1}) = (\sigma_x(h)(f(x, x^{-1}))^{-1}f(x, x^{-1}), e) = (\sigma_x(h), e) = \alpha(\sigma_x(h))$ (note that $(1, x)^{-1} = ((\sigma_x)^{-1}((f(x, x^{-1}))^{-1}), x^{-1})$). This shows that $\sigma^t_x(h) = \sigma_x(h)$ for all $x \in K$ and $h \in H$. It follows that $\sigma^t = \sigma$ and $f^t = f$. Thus, $Fac(E, t)$ is the given factor system (K, H, σ, f). ♯

Now, we describe the category ***EXT*** of extensions as the category of Factor systems. Let (λ, μ, ν) be a morphism between the extensions E_1 and E_2 given by the following commutative diagram:

$$\begin{array}{ccccccccc}
1 & \longrightarrow & H_1 & \overset{\alpha_1}{\longrightarrow} & G_1 & \overset{\beta_1}{\longrightarrow} & K_1 & \longrightarrow & 1 \\
 & & \downarrow{\scriptstyle\lambda} & & \downarrow{\scriptstyle\mu} & & \downarrow{\scriptstyle\nu} & & \\
1 & \longrightarrow & H_2 & \overset{\alpha_2}{\longrightarrow} & G_2 & \overset{\beta_2}{\longrightarrow} & K_2 & \longrightarrow & 1
\end{array}$$

Let t_1 be a section of E_1, and t_2 be a section of E_2. Let $(K_1, H_1, \sigma^{t_1}, f^{t_1})$ and $(K_2, H_2, \sigma^{t_2}, f^{t_2})$ be the corresponding factor systems. Let $x \in K_1$. Then $\mu(t_1(x)) \in G_2$ and $\beta_2(\mu(t_1(x))) = \nu(\beta_1(t_1(x))) = \nu(x) = \beta_2(t_2(\nu(x)))$. Thus, $\mu(t_1(x))(t_2$

$(\nu(x)))^{-1} \in ker\beta_2 = image\alpha_2$. In turn, we have a unique $g(x) \in H_2$ such that $\alpha_2(g(x)) = \mu(t_1(x))(t_2(\nu(x)))^{-1}$. Equivalently,

$$\mu(t_1(x)) = \alpha_2(g(x))t_2(\nu(x)) \cdots\cdots . \tag{10.8}$$

Since $t_1(e) = e = t_2(e)$, it follows that

$$g(e) = 1 \cdots\cdots \tag{10.9}$$

Now, using the commutativity of the diagram and the Eq. 10.8, we have

$$\mu(t_1(x)t_1(y)) = \mu(\alpha_1(f^{t_1}(x,y))t_1(xy)) = \alpha_2(\lambda(f^{t_1}(x,y))\mu(t_1(xy)) = \alpha_2(\lambda(f^{t_1}(x,y))\alpha_2(g(xy))t_2(\nu(xy)) = \alpha_2(\lambda(f^{t_1}(x,y))\alpha_2(g(xy))t_2(\nu(x)\nu(y))$$

On the other hand, since μ is a homomorphism, using again the Eq. 10.8,

$$\mu(t_1(x)t_1(y)) = \mu(t_1(x))\mu(t_1(y)) = \alpha_2(g(x))t_2(\nu(x))\alpha_2(g(y))t_2(\nu(y)) = \alpha_2(g(x))t_2(\nu(x))\alpha_2(g(y))(t_2(\nu(x)))^{-1}t_2(\nu(x))t_2(\nu(y)) = \alpha_2(g(x))\alpha_2(\sigma^{t_2}_{\nu(x)}(g(y)))t_2(\nu(x))t_2(\nu(y)) = \alpha_2(g(x))\alpha_2(\sigma^{t_2}_{\nu(x)}(g(y)))\alpha_2(f^{t_2}(\nu(x),\nu(y)))t_2(\nu(x)\nu(y)).$$

Equating the two expressions for $\mu(t_1(x)t_1(y))$, and observing that α_2 is an injective homomorphism, we obtain the following identity:

$$\lambda(f^{t_1}(x,y))g(xy) = g(x)\sigma^{t_2}_{\nu(x)}(g(y))f^{t_2}(\nu(x),\nu(y)) \cdots \tag{10.10}$$

Further, by Eq. 10.2,

$$t_1(x)\alpha_1(h)(t_1(x))^{-1} = \alpha_1(\sigma^{t_1}_x(h)).$$

Applying the homomorphism μ on the above equation, and using the commutativity of the diagram, we get

$$\mu(t_1(x))\alpha_2(\lambda(h))(\mu(t_1(x)))^{-1} = \alpha_2(\lambda(\sigma^{t_1}_x(h)).$$

Using the Eq. 10.8,

$$\alpha_2(g(x))t_2(\nu(x))\alpha_2(\lambda(h))(t_2(\nu(x)))^{-1}\alpha_2(g(x)^{-1}) = \alpha_2(\lambda(\sigma^{t_1}_x(h)).$$

Using the Eq. 10.2 for the extension E_2, we get

$$\alpha_2(g(x))\alpha_2(\sigma^{t_2}_{\nu(x)}(\lambda(h)))\alpha_2(g(x)^{-1}) = \alpha_2(\lambda(\sigma^{t_1}_x(h)).$$

Since α_2 is an injective homomorphism,

$$g(x)\sigma^{t_2}_{\nu(x)}(\lambda(h))g(x)^{-1} \;=\; \lambda(\sigma^{t_1}_x(h)) \;\cdots \tag{10.11}$$

Thus, a morphism (λ, μ, ν) between extensions E_1 and E_2 together with choices of sections t_1 and t_2 of the corresponding extensions, induces a map g from K_1 to H_2 such that the triple (ν, g, λ) satisfies (10.9), (10.10) and (10.11), and it may be viewed as a morphism from the factor system $(K_1, H_1, \sigma^{t_1}, f^{t_1})$ to $(K_2, H_2, \sigma^{t_2}, f^{t_2})$.

Let $(\lambda_1, \mu_1, \nu_1)$ be a morphism from an extension

$$E_1 \;\equiv\; 1 \longrightarrow H_1 \overset{\alpha_1}{\to} G_1 \overset{\beta_1}{\to} K_1 \longrightarrow 1$$

to an extension

$$E_2 \;\equiv\; 1 \longrightarrow H_2 \overset{\alpha_2}{\to} G_2 \overset{\beta_2}{\to} K_2 \longrightarrow 1,$$

and $(\lambda_2, \mu_2, \nu_2)$ be that from the extension E_2 to

$$E_3 \;\equiv\; 1 \longrightarrow H_3 \overset{\alpha_3}{\to} G_3 \overset{\beta_3}{\to} K_3 \longrightarrow 1.$$

Let t_1, t_2 and t_3 be corresponding choice of the sections. Then as in (10.8)

$$\mu_1(t_1(x)) \;=\; \alpha_2(g_1(x))t_2(\nu_1(x)).$$

and

$$\mu_2(t_2(u)) \;=\; \alpha_3(g_2(u))t_3(\nu_2(u)),$$

where g_1 is the uniquely determined map from K_1 to H_2, and g_2 is that from K_2 to H_3. In turn,

$$\begin{gathered}\mu_2(\mu_1(t_1(x))) \;=\; \mu_2(\alpha_2(g_1(x)))\mu_2(t_2(\nu_1(x))) \;=\\ \mu_2(\alpha_2(g_1(x)))\alpha_3(g_2(\nu_1(x)))t_3(\nu_2(\nu_1(x))) \;=\\ \alpha_3(\lambda_2(g_1(x)))\alpha_3(g_2(\nu_1(x)))t_3(\nu_2(\nu_1(x))) \;=\; \alpha_3(g_3(x))t_3(\nu_2(\nu_1(x))).\end{gathered}$$

It follows that the composition $(\lambda_2 \circ \lambda_1, \mu_2 \circ \mu_1, \nu_2 \circ \nu_1)$ induces the triple $(\nu_2 \circ \nu_1, g_3, \lambda_2 \circ \lambda_1)$, where $g_3(x) \;=\; \lambda_2(g_1(x))g_2(\nu_1(x))$ for each $x \in K_1$.

Prompted by the above discussion, we introduce the category **FACS** whose objects are factor systems, and a morphism from $(K_1, H_1, \sigma^1, f^1)$ to $(K_2, H_2, \sigma^2, f^2)$ is a triple (ν, g, λ), where ν is a homomorphism from K_1 to K_2, λ a homomorphism from H_1 to H_2, and g a map from K_1 to H_2 such that

(i) $g(e) \;=\; 1$

(ii) $\lambda(f^1(x, y))g(xy) \;=\; g(x)\sigma^2_{\nu(x)}(g(y))f^2(\nu(x), \nu(y))$

(iii) $g(x)\sigma^2_{\nu(x)}(\lambda(h))g(x)^{-1} \;=\; \lambda(\sigma^1_x(h))$

The composition of morphisms (ν_1, g_1, λ_1) from $(K_1, H_1, \sigma^1, f^1)$ to $(K_2, H_2, \sigma^2, f^2)$ and the morphism (ν_2, g_2, λ_2) from $(K_2, H_2, \sigma^2, f^2)$ to $(K_3, H_3, \sigma^3, f^3)$ is the triple

$(\nu_2 \circ \nu_1, g_3, \lambda_2 \circ \lambda_1)$, where g_3 is given by $g_3(x) = g_2(\nu_1(x))\lambda_2(g_1(x))$ for each $x \in K_1$.

The following theorem is the consequence of the above discussion.

Theorem 10.1.12 *Let t_E be a choice of a section of the extension E of a group by another group (such a choice function t exists because of axiom of choice). Then the association Fac which associates to each extension E the factor system $Fac(E, t_E)$ is an equivalence between the category **EXT** of extensions to the category* **FACS** *of factor systems.* ♯

Fix a pair H and K of groups. We try to describe the equivalence classes of extensions of H by K. Let G be an extension of H by K given by the exact sequence

$$E \equiv 1 \longrightarrow H \overset{\alpha}{\to} G \overset{\beta}{\to} K \longrightarrow 1.$$

Let (λ, μ, ν) be an equivalence from this extension to an other extension G' of H' by K' given by the exact sequence

$$E' \equiv 1 \longrightarrow H' \overset{\alpha'}{\to} G' \overset{\beta'}{\to} K' \longrightarrow 1$$

Then it follows that G' is also an extension of H by K given by the exact sequence

$$E'' \equiv 1 \longrightarrow H \overset{\alpha' \circ \lambda}{\to} G' \overset{\nu^{-1} \circ \beta'}{\to} K \longrightarrow 1$$

such that (I_H, μ, I_K) is an equivalence from E to E''. Also, $(\lambda, I_{G'}, \nu)$ is an equivalence from E'' to E'.

As such, there is no loss of generality in restricting the concept of equivalence on the class $E(H, K)$ of all extensions of H by K by saying that two extensions

$$E_1 \equiv 1 \longrightarrow H \overset{\alpha_1}{\to} G_1 \overset{\beta_1}{\to} K \longrightarrow 1.$$

and

$$E_2 \equiv 1 \longrightarrow H \overset{\alpha_2}{\to} G_2 \overset{\beta_2}{\to} K \longrightarrow 1.$$

in $E(H, K)$ are equivalent if there is an isomorphism ϕ from G_1 to G_2 such that the diagram

$$\begin{array}{ccccccccc}
1 & \longrightarrow & H & \xrightarrow{\alpha_1} & G_1 & \xrightarrow{\beta_1} & K & \longrightarrow & 1 \\
 & & \downarrow{\scriptstyle I_H} & & \downarrow{\scriptstyle \phi} & & \downarrow{\scriptstyle I_K} & & \\
1 & \longrightarrow & H & \xrightarrow{\alpha_2} & G_2 & \xrightarrow{\beta_2} & K & \longrightarrow & 1
\end{array}$$

is commutative. Indeed, for any extension E in $\boldsymbol{EXT}$ which is equivalent to a member E' of $E(H, K)$, there is a member E'' of $E(H, K)$ such that E is equivalent to E'' in the category $\boldsymbol{EXT}$ and E'' in $E(H, K)$ is equivalent E' in the sense described above.

Let

$$E \equiv 1 \longrightarrow H \xrightarrow{\alpha} G \xrightarrow{\beta} K \longrightarrow 1.$$

be an extension of H by K. Let t be a section of the extension. Then t induces a map σ^t from K to $Aut(H)$ as described by the Eq. 10.2. In turn, it induces a map Ψ_E^t from K to $Out(H)$ given by $\Psi_E^t(x) = \sigma_x^t Inn(H)$. Let

$$E' \equiv 1 \longrightarrow H \xrightarrow{\alpha'} G' \xrightarrow{\beta'} K \longrightarrow 1.$$

be another equivalent extension in $E(H, K)$. Let (I_H, Φ, I_K) be an equivalence from E to E', and t' be a section of E'. It induces an equivalence (I_K, g, I_H) from the factor system (K, H, σ^t, f^t) to the factor system $(K, H, \sigma^{t'}, f^{t'})$, where g is a map from K to H. From (10.10) and (10.11), we have

$$f^{t_1}(x, y)g(xy) = g(x)\sigma_x^{t_2}(g(y))f^{t_2}(x, y) \cdots\cdots \tag{10.12}$$

and

$$g(x)\sigma_x^{t'}((h))g(x)^{-1} = \sigma_x^t(h)) \cdots\cdots \tag{10.13}$$

for all $x, y \in K$ and $h \in H$. Thus, σ_x^t and $\sigma_x^{t'}$ differ by an inner automorphism of H. Hence $\Psi_E^t(x) = \sigma_x^t Inn(H) = \sigma_x^{t'} Inn(H) = \Psi_{E'}^{t'}(x)$ for each $x \in K$. This shows that the map Ψ_E^t from K to $Out(H)$ is independent of a representative E of the equivalence class, and also independent of the choice of a section t. Thus, without any ambiguity we can denote Ψ_E^t by $\Psi_{[E]}$, where $[E]$ denote the equivalence class determined by the extension E. Further, from (10.7), it follows that σ_{xy}^t and $\sigma_x^t \circ \sigma_y^t$ differ from an inner automorphism of H. Hence $\Psi_{[E]}(xy) = \Psi_{[E]}(x)\Psi_{[E]}(y)$ for all $x, y \in K$.

Definition 10.1.13 A homomorphism from K to $Out(H) = Aut(H)/Inn(H)$ is called a **coupling** or an **abstract kernel** of K to H.

We have established the following theorem:

Theorem 10.1.14 *Let* Ext (H, K) *denote the set of all equivalence classes of extensions in* $E(H, K)$. *Then there is a natural map* Ψ *from* Ext (H, K) *to the set* $Hom(K, Out(H))$ *of all abstract kernels (couplings) of* K *to* H *given by* $\Psi([E]) = \Psi_{[E]}$ *as defined above.* ♯

The map Ψ described in the above theorem is called the **abstract kernel** map or the **coupling** map.

Example 10.1.15 The abstract kernel map Ψ need not be injective. In other words, two non-equivalent extensions H by K may induce same abstract kernels of K to H. For example, consider the extensions

$$E_1 \equiv \{0\} \longrightarrow \mathbb{Z} \overset{(i_1,0)}{\to} \mathbb{Z} \oplus \mathbb{Z}_3 \overset{p_2}{\to} \mathbb{Z}_3 \longrightarrow \{0\}$$

and

$$E_1 \equiv \{0\} \longrightarrow \mathbb{Z} \overset{m_3}{\to} \mathbb{Z} \overset{\nu}{\to} \mathbb{Z}_3 \longrightarrow \{0\}$$

of the group $\mathbb{Z}$ by $\mathbb{Z}_3$, where m_3 is the multiplication by 3, and ν is the natural quotient map. These two extensions E_1 and E_2 are not equivalent as $\mathbb{Z} \oplus \mathbb{Z}_3$ and $\mathbb{Z}$ are not isomorphic. Since $Out(\mathbb{Z})$ is a group of order 2 and $\mathbb{Z}_3$ is a group of order 3, the only abstract kernel of $\mathbb{Z}_3$ to $\mathbb{Z}$ is the trivial map. As such, $[E_1] \neq [E_2]$, where as $\Psi([E_1]) = \Psi([E_2])$.

We shall see that the map Ψ may not be surjective also. Indeed, we have two basic problems in the theory of extensions of groups.

1. To determine the abstract kernels $\eta \in Hom(K, Out(H))$ which are realizable from an extension E of H by K in the sense that $\Psi([E]) = \eta$.

2. Given an abstract kernel $\eta \in Hom(K, Out(H))$ which is realizable from an extension, to determine and classify all extensions E up to equivalence such that $\Psi(E) = \eta$. Such abstract kernels are call couplings.

Theorem 10.1.16 *Let H be a group with $Z(H) = \{e\}$. Then the map Ψ from* Ext (H, K) *to the set $Hom(K, Out(H))$ is bijective. More explicitly, every abstract kernel η of K to H determines and is determined uniquely by an equivalence class of extensions in* Ext (H, K).

Proof Let $\eta \in Hom(K, Out(H))$ be an abstract kernel of K to H. Consider the Pull Back Diagram

$$\begin{array}{ccc} G & \xrightarrow{p_2} & K \\ \downarrow{\scriptstyle p_1} & & \downarrow{\scriptstyle \eta} \\ Aut(H) & \xrightarrow{\nu} & Out(H) \end{array}$$

More explicitly, G is the subgroup of the direct product $Aut(H) \times K$ given by $G = \{(\sigma, x) \mid \sigma \in Aut(H) \text{ and } \sigma InnH = \eta(x)\}$, p_1 the first projection and p_2 the second projection. Clearly, p_2 is a surjective homomorphism from G to K. The $kerp_2 = \{(\sigma, e) \mid \sigma InnH = \eta(e) = Inn(H)\} = Inn(H) \times \{e\}$. Since the center $Z(H)$ of H is trivial, the map α from H to G defined by $\alpha(h) = (i_h, e)$ (i_h denotes the inner automorphism determined by h) is an injective homomorphism with $image\alpha = kerp_2$. This gives an extension E of H by K given by the exact sequence

$$E \equiv 1 \longrightarrow H \overset{\alpha}{\to} G \overset{p_2}{\to} K \longrightarrow 1.$$

Using the axiom of choice, there is a map ξ from K to $Aut(H)$ such that $\xi(x)Inn(H) = \eta(x)$. This determines a section t of the extension E given by $t(x) = (\xi(x), x)$.

Recall that the abstract kernel $\Psi(E)$ associated to the extension E is given by $\Psi(E)(x) = \sigma_x^t Inn(H)$, where σ_x^t is given by (see Eq. 10.2)

$$t(x)\alpha(h)t(x)^{-1} = \alpha(\sigma_x^t(h)).$$

Now,

$$\alpha(\sigma_x^t(h)) = t(x)\alpha(h)t(x)^{-1} = (\xi(x), x)(i_h, e)(\xi(x), x)^{-1} = (\xi(x)i_h(\xi(x))^{-1}, e) = (i_{\xi(x)(h)}, e) = \alpha(\xi(x)(h)).$$

Thus, $\sigma_x^t(h) = \xi(x)(h)$ for all $h \in H$. In turn, $\sigma_x^t = \xi(x)$. By the definition, $\Psi(E)(x) = \sigma_x^t Inn(H) = \xi(x)Inn(H) = \eta(x)$ for all $x \in K$. This shows that $\Psi(E) = \eta$, and so Ψ is surjective.

To prove the injectivity, suppose that $\Psi(E_1) = \Psi(E_2)$, where E_1 and E_2 are extensions of H by K given by

$$E_1 \equiv 1 \longrightarrow H \overset{\alpha_1}{\to} G_1 \overset{\beta_1}{\to} K \longrightarrow 1.$$

and

$$E_2 \equiv 1 \longrightarrow H \overset{\alpha_2}{\to} G_2 \overset{\beta_2}{\to} K \longrightarrow 1.$$

Let t_1 be a section of E_1 with corresponding factor system $(K, H, \sigma^{t_1}, f^{t_1})$, and t_2 be a section of E_2 with the corresponding factor system $(K, H, \sigma^{t_2}, f^{t_2})$. Under our assumption

$$\sigma_x^{t_1} Inn(H) = \Psi(E_1)(x) = \Psi(E_2)(x) = \sigma_x^{t_2} Inn(H)$$

for all $x \in K$, where $\sigma_x^{t_1}$ and $\sigma_x^{t_2}$ are given by the equations

$$t_1(x)\alpha_1(h)t_1(x)^{-1} = \alpha_1(\sigma_x^{t_1}(h)).$$

and

$$t_2(x)\alpha_2(h)t_2(x)^{-1} = \alpha_2(\sigma_x^{t_2}(h)).$$

Since $\sigma_x^{t_1} Inn(H) = \sigma_x^{t_2} Inn(H)$ for all $x \in K$, and since H is center less, there is a unique map g from K to H such that

$$\sigma_x^{t_1} = i_{g(x)} \circ \sigma_x^{t_2} \cdots\cdots \tag{10.14}$$

for all $x \in K$. Again by (10.7), we have,

$$f^{t_1}(x, y)\sigma_{xy}^{t_1}(h) = \sigma_x^{t_1}(\sigma_y^{t_1}(h))f^{t_1}(x, y) \cdots\cdots \tag{10.15}$$

and

$$f^{t_2}(x, y)\sigma_{xy}^{t_2}(h) = \sigma_x^{t_2}(\sigma_y^{t_2}(h))f^{t_2}(x, y) \cdots\cdots \tag{10.16}$$

Using (10.14), (10.15), and (10.16), we get

$$\begin{aligned} & f^{t_1}(x,y)g(xy)\sigma_{xy}^{t_2}(h)g(xy)^{-1} \\ &= f^{t_1}(x,y)\sigma_{xy}^{t_1}(h) \\ &= \sigma_x^{t_1}(\sigma_y^{t_1}(h))f^{t_1}(x,y) \\ &= g(x)\sigma_x^{t_2}(\sigma_y^{t_1}(h))g(x)^{-1}f^{t_1}(x,y) \\ &= g(x)\sigma_x^{t_2}(g(y)\sigma_y^{t_2}(h)g(y)^{-1})g(x)^{-1}f^{t_1}(x,y) \\ &= g(x)\sigma_x^{t_2}(g(y))\sigma_x^{t_2}(\sigma_y^{t_2}(h))\sigma_x^{t_2}(g(y)^{-1})g(x)^{-1}f^{t_1}(x,y) \ \cdots\cdots \end{aligned} \tag{10.17}$$

In turn,
$\sigma_x^{t_2}(g(y)^{-1})g(x)^{-1}f^{t_1}(x,y)g(xy)\sigma_{xy}^{t_2}(h)(\sigma_x^{t_2}(g(y)^{-1})g(x)^{-1}f^{t_1}(x,y)g(xy))^{-1}$
$= \sigma_x^{t_2}(\sigma_y^{t_2}(h))$
$= f^{t_2}(x,y)\sigma_{xy}^{t_2}(h)(f^{t_2}(x,y))^{-1}.$
for all $h \in H$. Since $\sigma_{xy}^{t_2}$ is a bijective map on H, and H is center less,

$$f^{t_2}(x,y) = \sigma_x^{t_2}(g(y)^{-1})g(x)^{-1}f^{t_1}(x,y)g(xy),$$

or equivalently,

$$f^{t_1}(x,y)g(xy) = g(x)\sigma_x^{t_2}(g(y))f^{t_2}(x,y) \ \cdots\cdots . \tag{10.18}$$

The Eqs. (10.14) and (10.18) tells that (I_H, g, I_K) is an isomorphism from the factor system $(K, H, \sigma^{t_1}, f^{t_1})$ to $(K, H, \sigma^{t_2}, f^{t_2})$ in the category *FACS*. From Theorem 10.1.12, E_1 is equivalent to E_2. ♯

Indeed, the proof of the above theorem establishes the following more general result.

Proposition 10.1.17 *Let E_1 and E_2 be extensions of H by K with $\Psi(E_1) = \Psi(E_2)$. Then the following induced extensions E_1' and E_2' of $H/Z(H)$ by K given below are equivalent:*

$$E_1' \equiv 1 \longrightarrow H/Z(H) \xrightarrow{\overline{\alpha_1}} G_1/\alpha_1(Z(H)) \xrightarrow{\overline{\beta_1}} K \longrightarrow 1,$$

$$E_2' \equiv 1 \longrightarrow H/Z(H) \xrightarrow{\overline{\alpha_2}} G_2/\alpha_2(Z(H)) \xrightarrow{\overline{\beta_2}} K \longrightarrow 1. \qquad ♯$$

Split Extensions, Semi-direct Products

Definition 10.1.18 An extension

$$E \equiv 1 \longrightarrow H \xrightarrow{\alpha} G \xrightarrow{\beta} K \longrightarrow 1.$$

of H by K is called n **split extension**, if there is a section t which is a homomorphism. Such a section t is called a **splitting** of the extension. The corresponding factor system (K, H, σ^t, f^t) is such that f^t is trivial in the sense that $f^t(x, y) = 1$ for all $x, y \in K$, and then σ^t is a homomorphism from K to $aut(H)$.

The Example 10.1.2 (both the extensions), and the Example 10.1.4 (both the extensions) are split extensions, whereas the extension

$$\{0\} \longrightarrow m\mathbb{Z} \overset{i}{\rightarrow} \mathbb{Z} \overset{\nu}{\rightarrow} \mathbb{Z}_m \longrightarrow \{0\}$$

is not a split extension as the only homomorphism from $\mathbb{Z}_m$ to $\mathbb{Z}$ is the zero homomorphism.

Recall that a group G is said to be the **semi-direct product** of a normal subgroup H of G with a subgroup K of G if
(i) $G = HK$, and
(ii) $H \bigcap K = \{e\}$.
Symbolically, we write it as $H \succ K$.

Proposition 10.1.19 *Let*

$$E \equiv 1 \longrightarrow H \overset{\alpha}{\rightarrow} G \overset{\beta}{\rightarrow} K \longrightarrow 1.$$

be a split extension of H by K with a splitting t. Then $G = \alpha(H) \succ t(K)$ is the semi-direct product of $\alpha(H)$ with $t(K)$. Conversely, if $G = H \succ K$, then there is a natural projection p from G to K such that

$$1 \longrightarrow H \overset{i}{\rightarrow} G \overset{p}{\rightarrow} K \longrightarrow 1.$$

is a split extension of H by K.

Proof Suppose that E is a split extension with splitting t. Clearly, $\alpha(H) = ker\beta$ is a normal subgroup of G. Let $g \in G$. Then $\beta(gt(\beta(g^{-1}))) = \beta(g)\beta(t(\beta(g^{-1}))) = \beta(g)\beta(g^{-1}) = e$. This shows that $gt(\beta(g^{-1})) \in ker\beta = image\alpha$. Hence there is a unique $h \in H$ such that $gt(\beta(g^{-1})) = \alpha(h)$. In turn, $g = \alpha(h)t(\beta(g))$. Thus, $G = \alpha(H)t(K)$. Since t is an injective homomorphism, $t(K)$ is a subgroup of G isomorphic to K. Let $\alpha(h) \in \alpha(H) \bigcap t(K)$, $h \in H$. There is a $k \in K$ such that $\alpha(h) = t(k)$. But, then $e = \beta(\alpha(h)) = \beta(t(k)) = k$. Since t is a homomorphism, $e = t(e) = t(k) = \alpha(h)$. This shows that $\alpha(H) \bigcap t(K) = \{e\}$.

Conversely, suppose that $G = H \succ K$. Every element $g \in G$ is expressible as $g = hk$, where $h \in H$ and $k \in K$. Suppose that $h_1k_1 = h_2k_2$. Then $h_2^{-1}h_1 = k_2k_1^{-1} \in H \bigcap K = \{e\}$. This implies that $h_1 = h_2$ and $k_1 = k_2$. Thus, every element $g \in G$ is uniquely expressible as $g = hk$, where $h \in H$ and $k \in K$. This gives us a surjective map p from G to K given by $p(g) = k$, where $g = hk$. Also $(h_1k_1)(h_2k_2) = h_1k_1h_2k_1^{-1}k_1k_2$, where $h_1k_1h_2k_1^{-1} \in H$ and $k_1k_2 \in K$. It follows that p is a surjective homomorphism from G to K with kernel H. We get the extension

$$1 \longrightarrow H \xrightarrow{i} G \xrightarrow{p} K \longrightarrow 1.$$

of H by K with the inclusion map from K to G as splitting. ♯

Following are some applications of the above results.

Definition 10.1.20 A group H is called a **complete group** if the homomorphism $h \mapsto i_h$ (i_h being the inner automorphism determined by h) is an isomorphism from H to $aut(H)$. More explicitly, the center $Z(H)$ of H is trivial, and all automorphisms of H are inner.

Example 10.1.21 There are many complete groups. The symmetric group S_n, $n \neq 6$, the group $aut(G)$ of automorphisms of a non-abelian simple group G (or more generally, automorphism groups of direct products of non-abelian simple groups) are all complete groups. Let H be a cyclic group of odd order. Consider the symmetric group $Sym(H)$ on the set H. Let $\rho(H)$ denote the image of the Cayley representation of H in $Sym(H)$. Then the subgroup G of $Sym(H)$ generated by $\rho(H) \bigcup aut(H)$ is also a complete group. We shall give proof of some of them.

Proposition 10.1.22 *Let H be a complete group. Then any extension of H by K is equivalent to direct product extension*

$$1 \longrightarrow H \xrightarrow{i_1} H \times K \xrightarrow{p_2} K \longrightarrow 1,$$

where i_1 is the inclusion $h \mapsto (h, 0)$, and p_2 is the second projection. More explicitly, if G is any group containing H as a normal subgroup, then there is a subgroup K of G such that G is direct product of H and K.

Proof Since H is complete, the center $Z(H)$ of H is trivial, and also $Out(H)$ is trivial. Thus, for any group K there is only one abstract kernel of K to H which is the trivial homomorphism from K to $Out(H)$. It follows from the Theorem 10.1.16 that there is only one extension of H by K (up to equivalence) which, of course, is the one given in the proposition. ♯

Conversely, we have the following result due to Baer.

Proposition 10.1.23 (Baer) *Let H be a center less group. Suppose that for any group K, there is only one extension (up to equivalence) of H by K, then H is a complete group.*

Proof Let H be a center-less group such that all extensions of H are direct product extensions. Let $\alpha \in Aut(H)$. We wish to show that α is an inner automorphism of H. Consider the symmetric group $Sym(H)$ of permutations on the set H. For $h \in H$, let L_h denote the left multiplication by h on H. The map χ from H to $Sym(H)$ defined by $\chi(h) = L_h$ is an injective homomorphism from H to $Sym(H)$. Let G denote the subgroup of $Sym(H)$ generated by $\chi(H) \bigcup \{\alpha\}$. Let $L_h \in \chi(H)$. Then $\alpha \circ L_h \circ \alpha^{-1} =$

$L_{\alpha(h)} \in \chi(H)$. This shows that the subgroup $\chi(H)$ is a normal subgroup of G, and we have an extension

$$1 \longrightarrow H \overset{\chi}{\rightarrow} G \overset{\nu}{\rightarrow} G/\chi(H) \longrightarrow 1.$$

By our assumption, this is a direct product extension. Hence there exists a subgroup K of G isomorphic to $G/\chi(H)$ such that $G = \chi(H) \oplus K$. As such, elements of $\chi(H)$ will commute with elements of K, and every element of G is uniquely expressible as product of an element of $\chi(H)$ and an element of K. Suppose that $\alpha = L_x k$, $x \in H$ and $k \in K$. Now, $L_{\alpha(h)} = \alpha L_h \alpha^{-1} = L_x k L_h k^{-1} (L_x)^{-1} = L_x L_h L_{x^{-1}} = L_{xhx^{-1}}$. Since χ is injective, $\alpha(h) = xhx^{-1}$ for all $h \in H$. This shows that α is the inner automorphism determined by x. ♯

Following is an other characterization of a complete group.

Proposition 10.1.24 *A group H is a complete group if and only if it has a characteristic subgroup K with trivial centralizer in H such that all the automorphisms of K are induced by inner automorphisms of H.*

Proof Suppose that H is complete. Then we can take $K = H$, which is a characteristic subgroup of H, and since H is complete, every automorphism of H is an inner automorphism of H. Also $\{e\} = Z(H) = C_H(H)$. Conversely, let H be a group with a characteristic subgroup K whose centralizer $C_H(K)$ in H is trivial, and all of whose automorphisms are those induced by inner automorphisms of H. Since $C_H(K)$ is trivial, H is center less. It is sufficient (Proposition 10.1.23), therefore, to show that for any group G containing H as a normal subgroup, there is a subgroup L of G such that G is direct product of H and L. Let G be a group containing H as a normal subgroup. Then K, being a characteristic subgroup of H, is normal in G. Let g be any element of G. Then the inner automorphism i_g restricted to K is an automorphism of K. By our hypothesis, there is an element $h \in H$ such that i_h and i_g agree on K. This means that $h^{-1}g \in C_G(K)$. Thus, $G = HC_G(K)$. Since K is a normal subgroup of G, $C_G(K)$ is a normal subgroup of G. Also $H \bigcap C_G(K) = C_H(K) = \{e\}$. This shows that G is direct product of H and $C_G(K)$. ♯

Corollary 10.1.25 *Let G be a non-abelian simple group. Then $Aut(G)$ is a complete group.*

Proof Since G is non-abelian simple, $Inn(G)$ is isomorphic to G, and so $Inn(G)$ is simple. In the light of the above proposition, it is sufficient to show that $Inn(G)$ is a characteristic subgroup of $Aut(G)$ whose centralizer in $Aut(G)$ is trivial, and all automorphisms of $Inn(G)$ are induced by the inner automorphisms of $Aut(G)$. $Inn(G)$ is already seen to be a normal subgroup of $Aut(G)$. Let $\alpha \in C_{Aut(G)}(Inn(G))$. Then $\alpha \circ i_g = i_g \circ \alpha$ for all $g \in G$. Hence $\alpha(g)\alpha(x)\alpha(g)^{-1} = g\alpha(x)g^{-1}$ for all $x, g \in G$. This shows that $g^{-1}\alpha(g) \in Z(G)$ for all $g \in G$. Since G is center less, $\alpha(g) = g$ for all $g \in G$. This means that $\alpha = I_G$. Thus, $C_{Aut(G)}(Inn(G)) = \{I_G\}$. Next, we show that $Inn(G)$ is a characteristic subgroup

of $Aut(G)$. Let $\alpha \in Aut(Aut(G))$. Since $Inn(G)$ is a normal subgroup of $Aut(G)$, $\alpha(Inn(G))$ is also a normal subgroup of $Aut(G)$, and so $\alpha(Inn(G)) \bigcap Inn(G)$ is a normal subgroup of $Inn(G)$, and also of $\alpha(Inn(G))$. Since $Inn(G)$ and $\alpha(Inn(G))$ are simple, $\alpha(Inn(G)) \bigcap Inn(G) = \{I_G\}$ or else $\alpha(Inn(G)) \bigcap Inn(G) = Inn(G) = \alpha(Inn(G))$. Suppose that $\alpha(Inn(G)) \bigcap Inn(G) = \{I_G\}$. Then the elements of $\alpha(Inn(G))$ commute with elements of $Inn(G)$. This is a contradiction to the fact that $C_{Aut(G)}(Inn(G)) = \{I_G\}$. Hence $\alpha(Inn(G)) \bigcap Inn(G) = Inn(G) = \alpha(Inn(G))$. This shows that $Inn(G)$ is a characteristic subgroup of $Aut(G)$.

Let $\chi \in Aut(Inn(G))$. Then there is a bijective map μ from G to G such that $\chi(i_g) = i_{\mu(g)}$ for all $g \in G$. Further, $i_{\mu(g_1g_2)} = \chi(i_{g_1g_2}) = \chi(i_{g_1}i_{g_2}) = \chi(i_{g_1})\chi(i_{g_2}) = i_{\mu(g_1)}i_{\mu(g_2)}$. This shows that $\mu(g_1g_2) = \mu(g_1)\mu(g_2)$ for all $g_1, g_2 \in G$. Thus, $\mu \in Aut(G)$. Now, $(\mu i_g \mu^{-1})(x) = i_{\mu(g)}(x) = \chi(i_g)(x)$ for all $g, x \in G$. This shows that $\chi(i_g) = \mu i_g \mu^{-1}$ for all $g \in G$. Thus, χ is the automorphism of $Inn(G)$ induced by an inner automorphism of $Aut(G)$. ♯

We described the extensions of groups with trivial centers. Let us consider the other extreme case when center of the group is the group itself. More explicitly, we describe the extensions of abelian groups. Let H be an abelian group. We shall adopt the additive notation $+$ for the binary operation of H. The group $Out(H)$ is naturally identified with $Aut(H)$. An abstract kernel of K to H is a homomorphism σ from K to $Aut(H)$. We discuss the following problem:

Problem Let H be an abelian group. Classify all extensions of H by K (up to equivalence) with the given abstract kenel σ.

Let us denote by $EXT_\sigma(H, K)$ the set of equivalence classes of extensions of an abelian group H by a group K with the given abstract kernel σ. We have at least one such extension, viz., the semi-direct product extension of H by K associated to the homomorphism σ. Clearly, the factor system associated to the split extension is (K, H, σ, f_0), where f_0 is trivial in the sense that $f_0(x, y) = 0$ for all $x, y \in K$. Let $Z^2_\sigma(K, H)$ denote the set of factor systems (K, H, σ, f) associated to the abstract kernel σ. Indeed, a factor system in $Z^2_\sigma(K, H)$ determines, and it is uniquely determined by the corresponding map f which satisfies the condition

$$f(x, y) + f(xy, z) = \sigma_x(f(y, z)) + f(x, yz)$$

for all $x, y, z \in K$. By the abuse of language, we shall call such a f as a factor system in $Z^2_\sigma(K, H)$. f is also called a 2-co-cycle associated to (K, H, σ). The justification for the notation $Z^2_\sigma(K, H)$, and the 2- co-cycle terminology will follow later. Suppose that f and f' are two members of $Z^2_\sigma(K, H)$. Then

$$\begin{aligned}
&(f + f')(x, y) + (f + f')(xy, z) \\
&= f(x, y) + f(xy, z) + f'(x, y) + f'(xy, z) \\
&= \sigma_x(f(y, z)) + f(x, yz) + \sigma_x(f'(y, z)) + f'(x, yz) \\
&= \sigma_x((f + f')(y, z)) + (f + f')(x, yz).
\end{aligned}$$

This shows that $f + f'$ is also a factor system associated to the abstract kernel σ. Also $-f \in Z^2_\sigma(K, H)$ for all $f \in Z^2_\sigma(H, K)$. Thus, $Z^2_\sigma(K, H)$ is an abelian group with

respect to the operation defined above. f_0 is the identity of the group. Let $B^2_\sigma(K,H)$ denote the set of factor systems which are equivalent to the trivial factor system f_0. More precisely, from (10.12), $f \in B^2_\sigma(K,H)$ if and only if there is a map g from K to H with $g(e) = 0$ such that $f(x,y) = \sigma_x(g(y)) - g(xy) + g(x)$ (see Eq. 10.12 written additively). Note that for any map g with $g(e) = 0, f$ defined by $f(x,y) = \sigma_x(g(y)) - g(xy) + g(x)$ is a factor system. The members of $B^2_\sigma(K,H)$ are called the 2- co-boundaries associated to (K,H,σ). The quotient group $Z^2_\sigma(K,H)/B^2_\sigma(K,H)$ is called the **second co - homology group** associated to (K,H,σ), and it is denoted by $H^2_\sigma(K,H)$.

Theorem 10.1.26 *Let H be an abelian group, and K be a group. Let σ an abstract kernel of K to H. Then, there is a natural bijective correspondence Γ between the set $EXT_\sigma(H,K)$ of equivalence classes of extensions of H by K with the given abstract kernel σ and the second co-homology group $H^2_\sigma(K,H)$.*

Proof Let E be an extension of H by K with the abstract kernel σ. Let t be a section of the extension, and (K,H,σ,f^t) be the corresponding factor system. Then $f^t \in Z^2_\sigma(K,H)$. Let E' be another equivalent extension of H by K, and t' be a section of the extension E'. Let $(K,H,\sigma,f^{t'})$ be the corresponding factor system. Then (see the Eq. 10.18) there is a map g from K to H with $g(e) = 0$ such that $f^t(x,y) + g(xy) = \sigma_x(g(y)) + g(x) + f^{t'}(x,y)$. This shows that $f^t + B^2_\sigma(K,H) = f^{t'} + B^2_\sigma(K,H)$. Thus, the association $(E,t) \mapsto f^t$ induces a map Γ from $EXT_\sigma(H,K)$ to $H^2_\sigma(K,H)$ given by $\Gamma([E]) = f^t + B^2_\sigma(K,H)$, where t is a section of E. Let $f \in Z^2_\sigma(K,H)$. Then by the theorem 10.1.11, there is an extension E of H by K, and a section t such that $f^t = f$. This shows that Γ is surjective. Let E_1 and E_2 be extensions of H by K with sections t_1 and t_2 and abstract kernel σ such that $\Gamma([E_1]) = \Gamma([E_2])$. Then $f^{t_1} + B^2_\sigma(K,H) = f^{t_2} + B^2_\sigma(K,H)$. Hence there exists a map g from K to H with $g(e) = 0$ such that $f^{t_1}(x,y) + g(xy) = \sigma_x(g(y)) + g(x) + f^{t_2}(x,y)$. It follows that the factor system f^{t_1} is equivalent f^{t_2}. Hence the corresponding extensions E_1 and E_2 are equivalent. ♯

Let H be a group (not necessarily abelian), and K be a group. Though, H may not be abelian, we use the additive notation $+$ for the operation in H, and also for the operation in any extension G of H by K. The operation of K is denoted by juxtaposition. Thus, the identity of H is denoted by 0, and that of K by e. Let $\psi : K \longmapsto Out(H) = Aut(H)/Inn(H)$ be an abstract kernel. Since the center $Z(H)$ of H is a characteristic subgroup of H, we have a homomorphism $\chi : Aut(H) \longmapsto Aut(Z(H))$ given by $\chi(\alpha) = \alpha/(Z(H)$. Let $\sigma : K \longmapsto Aut(H)$ be a lifting of ψ with $\sigma(e) = I_H$ in the sense that $\nu o\sigma = \psi$, where ν is the quotient map from $Aut(H)$ to $Out(H)$. Since ψ is a homomorphism, $\sigma(xy)Inn(H) = (\sigma(x)\sigma(y))Inn(H)$. Hence there is a map f from $K \times K$ to H such that $\sigma(x)\sigma(y) = i_{f(x,y)}\sigma(xy)$ (recall that i_h denote the inner automorphism determined by h). It follows that $(\chi o\sigma)(xy) = (\chi o\sigma)(x)(\chi o\sigma)(y)$ *for all* $x, y \in K$. This means that $\chi o\sigma$ is a homomorphism from K to $Aut(Z(H))$. Let τ be another lifting of ψ. Then $\sigma(x)Inn(H) = \tau(x)Inn(H)$ *for all* $x \in K$. Hence there is a map g from K to H with $g(e) = 0$ such that $\sigma(x) = i_{g(x)}\tau(x)$ for all $x \in K$. But, then $(\chi o\sigma(x)) = (\chi o\tau(x))$ for all $x \in K$. Thus, $\chi o\sigma$ depends only on ψ and not on any

particular lifting σ. In turn, χ induces a map $\overline{\chi}$ from the set $Hom(K,\ Out(H))$ of abstract kernels from K to H to the set $Hom(K,\ Z(H))$ of abstract kernel from K to $Z(H)$ given by $\overline{\chi}(\psi) = \chi o\sigma$, where σ is a lifting of ψ.

Proposition 10.1.27 *Let*

$$E \equiv 1 \longrightarrow H \stackrel{\alpha}{\rightarrow} G \stackrel{\beta}{\rightarrow} K \longrightarrow 1.$$

and

$$E' \equiv 1 \longrightarrow H \stackrel{\alpha'}{\rightarrow} G' \stackrel{\beta'}{\rightarrow} K \longrightarrow 1.$$

be extensions of H by K such that $\psi_E = \psi_{E'} = \psi$. Then there is a section t of E and a section t' of E' such that $\sigma^t = \sigma^{t'} = \overline{\chi}(\psi)$, and $-f^t(x,y) + f^{t'}(x,y) \in Z(H)$ for all $x, y \in K$. Further, then the map h from $K \times K$ to $Z(H)$ defined by $h(x,y) = -f^t(x,y) + f^{t'}(x,y)$ is a 2 co-cycle in $Z^2_{\overline{\chi}(\psi)}(K, Z(H))$.

Proof Let t be a section of E, and s be a section of E'. Since $\psi_E = \psi_{E'}$, $\sigma^t(x)Inn(H) = \sigma^s(x)Inn(H)$ for all $x \in K$. This means that there is a function g from K to H with $g(e) = 0$ such that $\sigma^t(x) = i_{g(x)}\sigma^s(x)$ *for all* $x \in K$. The map t' from K to G' given by $t'(x) = g(x) + s(x)$ is also a section of E'. Further, then $\sigma^{t'}(x) = i_{g(x)}\sigma^s(x) = \sigma^t(x)$ for all x in K. This shows that $\sigma^t = \sigma^{t'}$. Now $f^t(x,y) = t(x) + t(y) - t(xy)$ and $f^{t'}(x,y) = t'(x) + t'(y) - t'(xy)$. Hence $i_{f^t(x,y)} = \sigma^t(x)\sigma^t(y)(\sigma^t(xy))^{-1} = \sigma^{t'}(x)\sigma^{t'}(y)(\sigma^{t'}(xy))^{-1} = i_{f^{t'}(x,y)}$ for all $x, y \in K$. Thus, $i_{-f^t(x,y)\,+\,f^{t'}(x,y)} = I_H$. This shows that $-f^t(x,y) + f^{t'}(x,y) \in Z(H)$ for all $x, y \in K$. Put $h(x,y) = -f^t(x,y) + f^{t'}(x,y)$. Then

$$\begin{aligned}
&h(x,y) + h(xy,z)\\
&= -f^t(x,y) + f^{t'}(x,y) - f^t(xy,z) + f^{t'}(xy,z)\\
&= -f^t(xy,y) - f^t(x,y) + f^{t'}(x,y) + f^{t'}(xy,z)\\
&= -(f^t(x,y) + f^t(xy,z)) + f^{t'}(x,y) + f^{t'}(xy,z)\\
&= -(\sigma^t_x(f^t(y,z)) + f^t(x,yz)) + f^{t'}(x,y) + f^{t'}(xy,z)\\
&= -f^t(x,yz) - \sigma^t_x(f^t(y,z)) + \sigma^{t'}_x(f^{t'}(y,z)) + f^{t'}(x,yz)\\
&= -f^t(x,yz) + \sigma^t_x(-f^t(y,z) + f^{t'}(y,z)) + f^{t'}(x,yz), \text{ for } \sigma^t = \sigma^{t'}\\
&= -f^t(x,yz) + f^{t'}(x,yz) + \sigma^t_x(-f^t(y,z) + f^{t'}(y,z))\\
&= \sigma^t_x(-f^t(y,z) + f^{t'}(y,z)) - f^t(x,yz) + f^{t'}(x,yz)\\
&= \sigma^t_x(h(y,z)) + h(x,yz)
\end{aligned}$$

for all $x, y, z \in K$. This shows that $h \in Z^2_{\overline{\chi}(\psi)}(K, Z(H))$. ♯

Theorem 10.1.28 *Let $\psi : K \longmapsto Out(H)$ be an abstract kernel from K to H which is realizable by an extension of H by K. Then the second co-homology group $H^2_{\overline{\chi}(\psi)}(K, Z(H))$ acts sharply transitively on the set $EXT_\psi(H,K)$ of equivalence classes of extensions of H by K associated to the abstract kernel ψ.*

Proof Let E be an extension of H by K which realizes the abstract kernel ψ, and t be a section of E. Let (K, H, σ^t, f^t) be the corresponding factor system. Then

$\psi(x) = \sigma_x^t Inn(H)$ *for all* $x \in K$. Let $h \in Z^2_{\overline{\chi}(\psi)}(K, Z(H))$. It is easily seen that $(K, H, \sigma^t, f^t + h)$ is again a factor system. Let $E \star h$ denote the corresponding extension. Clearly $E \star h$ also realizes ψ. Let h' be another 2 co-cycle in $Z^2_{\overline{\chi}(\psi)}(K, Z(H))$ such that the co-homology class $[h] = h + B^2_{\overline{\chi}(\psi)}(K, Z(H)) = [h'] = h' + B^2_{\overline{\chi}(\psi)}(K, Z(H))$ in $H^2_{\overline{\chi}(\psi)}(K, Z(H))$. Then there is a map $g : K \longmapsto Z(H) \subseteq H$ with $g(e) = 0$ such that $h'(x, y) = \partial g(x, y) + h(x, y)$ for all $x, y \in K$. Clearly $f^t + h' = f^t + h + \partial g$. Hence $(K, H, \sigma^t, f^t + h)$ is equivalent to $(K, H, \sigma^t, f^t + h')$. This shows that $[E \star h] = [E \star h']$. Let E and E' be equivalent extensions of H by K which realize ψ and $h \in Z^2_{\overline{\chi}(\psi)}(K, Z(H))$. By the Theorem 10.1.12, we have sections t and t' of E and E' respectively such that (K, H, σ^t, f^t) is equivalent to $(K, H, \sigma^{t'}, f^{t'})$. Hence, there is a map g from K to H with $g(e) = 0$ such that $f^t(x, y) + g(xy) = g(x) + \sigma^t(x)(g(y)) + f^{t'}(x, y)$ for all $x, y \in K$. Clearly, $(K, H, \sigma^t, f^t + h)$ is equivalent to $(K, H, \sigma^{t'}, f^{t'} + h)$, and so $[E \star h] = [E' \star h]$. Thus, we get an action $\star$ of $H^2_{\overline{\chi}(\psi)}(K, Z(H)$ on $EXT_\psi(H, K)$ given by $[E] \star [h] = [E \star h]$. We show that this action is sharply transitive. Let E and E' be extensions realizing the abstract kernel ψ. By the Proposition 10.1.27, there is a section t of E, and there is a section t' of E' such that $\sigma^t = \sigma^{t'} = \overline{\chi}(\psi)$, and the map h from $K \times K$ to $Z(H)$ defined by $h(x, y) = -f^t(x, y) + f^{t'}(x, y)$ is a 2 co-cycle in $Z^2_{\overline{\chi}(\psi)}(K, Z(H))$. Clearly, $[E] \star [h] = [E']$. This shows that the action $\star$ is transitive. Next, suppose that $[E] \star [h] = [E]$. Then there is a section t of E such that the factor system (K, H, σ^t, f^t) is equivalent to $(K, H, \sigma^t, f^t + h)$. Hence there is a map g from K to H with $g(e) = 0$ such that $f^t(x, y) + h(x, y) + g(xy) = g(x) + \sigma_x^t(g(y)) + f^t(x, y)$ for all $x, y \in K$ and also $g(x) + \sigma_x^t(h) - g(x) = \sigma_x^t(h)$ for all $x \in K$ and $h \in H$. Since σ_x^t is an automorphism of H, it follows that $g(x) \in Z(H)$ for all $x \in K$. Thus, $h(x, y) = \sigma_x^t(g(y)) - g(xy) + g(x)$ for all $x, y \in K$. This shows that $h = \partial g$, where g is a map from K to $Z(H)$ with $g(e) = 0$. It follows that $[h] = 0$. This completes the proof of the fact that the action $\star$ is sharply transitive. ♯

Corollary 10.1.29 *There is a bijective correspondence between* $EXT_\psi(H, K)$ *to* $H^2_{\overline{\chi}(\psi)}(K, Z(H))$ *provided there is an extension of* H *by* K *which realizes* ψ. ♯

Let H be an abelian group and K a group. As $H^2_\sigma(K, H)$ is an abelian group, the bijective map Γ (see Theorem 10.1.26) induces a group structure on the set $EXT_\sigma(H, K)$ of equivalence classes of extensions of H by K with the given abstract kernel σ. We shall try to describe the induced addition called the **Baer sum** on the class of extensions. Let

$$E_1 \equiv 1 \longrightarrow H \xrightarrow{\alpha_1} G_1 \xrightarrow{\beta_1} K \longrightarrow 1.$$

and

$$E_2 \equiv 1 \longrightarrow H \xrightarrow{\alpha_2} G_2 \xrightarrow{\beta_2} K \longrightarrow 1.$$

be two extensions of H by K with abstract kernel σ. We have the extension $E_1 \oplus E_2$ of $H \oplus H$ by $K \oplus K$ given by

$$E_1 \oplus E_2 \equiv 1 \longrightarrow H \oplus H \overset{\alpha_1 \oplus \alpha_2}{\rightarrow} G_1 \oplus G_2 \overset{\beta_1 \oplus \beta_2}{\rightarrow} K \oplus K \longrightarrow 1.$$

Using this, we construct an other extension of H by K. Let Δ denote the diagonal map from K to $K \oplus K$ defined by $\Delta(x) = (x, x)$. We have the pull back diagram

$$\begin{array}{ccc} L & \overset{\chi}{\longrightarrow} & K \\ {\scriptstyle i}\downarrow & & \downarrow{\scriptstyle \Delta} \\ G_1 \oplus G_2 & \overset{\beta_1 \oplus \beta_2}{\longrightarrow} & K \oplus K, \end{array}$$

where $L = \{(g_1, g_2) \mid \beta_1(g_1) = \beta_2(g_2)\}$, i the inclusion map, and χ is given by $\chi(g_1, g_2) = \beta_1(g_1)$ (ensure that it is a pull back diagram). Clearly, χ is a surjective homomorphism, $ker\chi = \{(g_1, g_2) \in L \mid \chi(g_1, g_2) = \beta_1(g_1) = e\} = \{(g_1, g_2) \mid e = \beta_1(g_1) = \beta_2(g_2)\} = H \oplus H$. Thus, we have an extension $\Delta^\star(E \oplus E)$ of $H \oplus H$ by K given by

$$\Delta^\star(E_1 \oplus E_2) \equiv 1 \longrightarrow H \oplus H \overset{\alpha_1 \oplus \alpha_2}{\rightarrow} L \overset{\chi}{\rightarrow} K \longrightarrow 1.$$

Now, let ∇ denote the co-diagonal map from $H \oplus H$ to H given by $\nabla(h_1, h_2) = h_1 + h_2$. Since H is an abelian group, ∇ is a homomorphism. Let $D = \{(h, h^{-1}) \mid h \in H\}$. Then D is a normal subgroup of L, and we have the push out diagram (verify)

$$\begin{array}{ccc} H \oplus H & \overset{\alpha_1 \oplus \alpha_2}{\longrightarrow} & L \\ {\scriptstyle \nabla}\downarrow & & \downarrow{\scriptstyle \nu} \\ H & \overset{\eta}{\longrightarrow} & \overline{G}, \end{array}$$

where $\overline{G} = L/D$, ν the quotient map, and η is given by $\eta(h) = (h, 0) + D$. Clearly, η is a homomorphism. Suppose that $\eta(h) = D$. Then $(h, 0) \in D$. This implies that $h = 0$. Hence η is injective. Again the map χ from L to K takes $(h, -h)$ to $\beta_1(h) = e$. This shows that χ induces a surjective homomorphism $\overline{\chi}$ from $\overline{G}$ to K given by $\overline{\chi}((g_1, g_2) + D) = \beta_1(g_1)$. Also $ker\overline{\chi} = \{(g_1, g_2) + D \mid e = \beta_1(g_1) = \beta_2(g_2)\} = \{(h_1, h_2) + D \mid h_1, h_2 \in H\} = \{(h_1 + h_2) + D \mid h_1, h_2 \in H\} = image\eta$. Thus, we get an other extension

$$1 \longrightarrow H \overset{\eta}{\rightarrow} \overline{G} \overset{\overline{\chi}}{\rightarrow} K \longrightarrow 1.$$

of H by K, called the **Baer sum** of E_1 and E_2, and it is denoted by $E_1 \biguplus E_2$. Further, let t_1 be a section of E_1, and t_2 be a section of E_2 with corresponding factor system f^{t_1} and f^{t_2}. Then we have the section $t_1 + t_2$ of $E_1 \biguplus E_2$ given by $(t_1 + t_2)(x) = (t_1(x), t_2(x)) + D$. It can be easily seen that $f^{t_1 + t_2} = f^{t_1} + f^{t_2}$. It follows that the bijective map Γ from the set $EXT_\psi(H, K)$ of equivalence classes of extensions of

H by K to the second co-homology group $H^2_\psi(K,H)$ respects the addition. In turn, $EXT_\psi(H,K)$ is an abelian group with respect to the Baer sum, and it is isomorphic to $H^2_\psi(K,H)$. ♯

Proposition 10.1.30 *Let H be an abelian group, and K be a group of order m. Then $mH^2_\sigma(K,H) = \{0\}$ for any abstract kernel $\sigma : K \longmapsto Aut(H)$.*

Proof Let $f \in Z^2_\sigma(K,H)$. Consider the map g from K to H given by $g(x) = \Sigma_{z\in K} f(x,z)$. Clearly $g(e) = 0$. Now,

$$\begin{aligned}
&\partial g(x,y) = \sigma_x(g(y)) - g(xy) + g(x)\\
&= \sigma_x(\Sigma_{z\in K} f(y,z)) - \Sigma_{z\in K} f(xy,z) + \Sigma_{z\in K} f(x,z)\\
&= \Sigma_{z\in K}(\sigma_x(f(y,z))) - f(xy,z) + f(x,z)\\
&= mf(x,y) + \Sigma_{z\in K} - f(x,y) - f(xy,z) + \sigma_x(f(y,z)) + f(x,z)\\
&= mf(x,y) + \Sigma_{z\in K}\sigma_x(f(y,z)) - f(xy,z) + f(x,yz) - f(x,y) \text{ (for } H \text{ is abelian)}\\
&= mf(x,y) + \Sigma_{z\in K}\partial f(x,y,z)\\
&= mf(x,y) \text{ (for } f \in Z^2_\sigma(K,H)).
\end{aligned}$$

Hence $mf \in B^2_\sigma(K,H)$. This shows that $m[f] = 0$. ♯

Corollary 10.1.31 *Let H be an abelian group of order n, and K a group of order m such that $(m,n) = 1$. Then $H^2_\sigma(K,H) = \{0\}$ for all abstract kernel $\sigma : K \longmapsto Aut(H)$.*

Proof From the above proposition, $mH^2_\sigma(K,H) = \{0\}$. Since $nf = 0$ for all maps f from $K \times K$ to H, it follows that $nH^2_\sigma(K,H) = \{0\}$. Since m and n are co-prime, $H^2_\sigma(K,H) = \{0\}$. ♯

Corollary 10.1.32 *Let H be a finite abelian group of order n, and K a group of order m, where $(m,n) = 1$. Then every extension of H by K splits.*

Proof It follows from Theorem 10.1.26 that $EXT_\sigma(H,K)$ is in bijective correspondence with $H^2_\sigma(K,H)$. From the above corollary, it is evident that there is only one equivalence class of extension, and indeed, it is the split extension. ♯

Let H and K be finite groups of co-prime orders. $Z(H)$ and K are also of co-prime orders. It follows from the Theorem 10.1.28 and the above corollary that EXT_ψ contains at most one element. More explicitly, an abstract kernel $\psi : K \longmapsto Out(H)$ is either not realizable from an extension of H by K or there is only one equivalence class of extensions of H by K associated to the abstract kernel ψ. However, it is not clear if the unique extension is split extension. The following theorem asserts that it is, indeed, a split extension even if H is non-abelian.

Theorem 10.1.33 (Schur–Zassenhauss) *Let G be a finite group having a normal subgroup H such that H and G/H are of co-prime orders. Then G is a split extension of H by G/H (equivalently, G is semi-direct product of H with G/H).*

Proof If H is abelian subgroup of G, then the result follows from the above corollary. We prove it for general case. The proof is by induction on the order $\mid G \mid$ of G. If $\mid G \mid = 1$, then there is nothing to do. Assume that the result is true for all those groups L for which $\mid L \mid < \mid G \mid$. We prove it for G. Let H be a normal subgroup of G such that $\mid H \mid$ is co-prime to $\mid G/H \mid$. Suppose that there is a proper subgroup K of G such that $G = HK$. Then $K \bigcap H$ is a normal subgroup of K with $(\mid (K \bigcap H) \mid , \mid K/K \bigcap H \mid) = 1$ (note that $K/K \bigcap H$ is isomorphic to $KH/H = G/H$). Since $\mid K \mid < \mid G \mid$, by the induction hypothesis, there is a compliment L of $K \bigcap H$ in K. But, then $K = (K \bigcap H)L$ and $K \bigcap H \bigcap L = \{e\}$. Hence $G = H(K \bigcap H)L = HL$ and $H \bigcap L = \{e\}$ (for $L \subseteq K$). This proves the result for G in case G has a proper normal subgroup K such that $G = HK$. Next, assume that there is no proper subgroup K of G such that $G = HK$. Suppose that there is a nontrivial normal subgroup M of G which is properly contained in H. Then H/M is a normal subgroup of G/M such that H/M and $(G/M)/(H/M) \approx G/H$ are of co-prime orders. By the induction hypothesis, there is a subgroup L/M of G/M such that $G/M = (H/M)(L/M)$ and $(H/M) \bigcap (L/M) = \{M\}$. In other words $G = HL$ and $H \bigcap L = M$. But, then $L = G$ and so $M = H \bigcap L = H \bigcap G = H$, a contradiction to the supposition that M is properly contained in H. Thus, H is a minimal normal subgroup of G. Let p be a prime dividing the order $\mid H \mid$ of H. Let P be a sylow p - subgroup of H. Since $(p, \mid G/H \mid) = 1$, P is a sylow p - subgroup of G also. Further, since all the sylow p - subgroups of G are conjugate, and they are all contained in H, $G = HN_G(P)$. Hence $N_G(P) = G$. In other words P is a normal subgroup of G which is contained in H. Since H is a minimal normal subgroup of G, $H = P$ is a p - subgroup. Thus, the center $Z(H) \neq \{e\}$. Since $Z(H)$ is a characteristic subgroup of H, and H is normal in G, it follows that $Z(H)$ is normal in G. Again the minimality of H ensures that $H = Z(H)$ is abelian. From the previous corollary, G is split extension of H by G/H. ♯

10.2 Obstructions and Extensions

Let us discuss the conditions under which an abstract kernel ψ of K to H can be realized from an extension of H by K. Here, H is not assumed to be an abelian group. Let σ be a map from K to $Aut(H)$ with $\sigma_e = I_H$ such that $\psi(x) = \sigma_x Inn(H)$ for each $x \in K$ (axiom of choice ensures that such a map exists). Such a map σ will be termed as lifting of ψ. Since ψ is a homomorphism, $\sigma_x \sigma_y Inn(H) = \sigma_{xy} Inn(H)$. Let f be a map from $K \times K$ to H with $f(e, x) = 1 = f(x, e)$ for all $x \in K$ such that

$$\sigma_x \sigma_y = i_{f(x,y)} \sigma_{xy} \cdots\cdots \tag{10.19}$$

for all $x, y \in K$ (existence of such a f is ensured by the axiom of choice). Now,

$$\sigma_x \sigma_y \sigma_z = i_{f(x,y)} \sigma_{xy} \sigma_z = i_{f(x,y)} i_{f(xy,z)} \sigma_{xyz}.$$

On the other hand,

$$\sigma_x\sigma_y\sigma_z \;=\; \sigma_x i_{f(y,z)}\sigma_{yz} \;=\; \sigma_x i_{f(y,z)}(\sigma_x)^{-1}\sigma_x\sigma_{yz} \;=\; i_{\sigma_x(f(y,z))}i_{f(x,yz)}\sigma_{xyz}.$$

Equating both the expressions,

$$i_{f(x,y)f(xy,z)} \;=\; i_{\sigma_x(f(y,z))f(x,yz)}.$$

for all $x, y, z \in K$. Hence there is a map ϕ from $K \times K \times K$ to $Z(H)$ with $1 = \phi(e, y, z) = \phi(x, e, z) = \phi(x, y, e)$ such that

$$f(x, y)f(xy, z) \;=\; \phi(x, y, z)\sigma_x(f(y, z))f(x, yz) \;\cdots\cdots\; . \tag{10.20}$$

Clearly, f is a factor system in $Z^2_\sigma(K, H)$ if and only if ϕ is identically trivial. It is natural, therefore, to term ϕ as the **obstruction** to f for being a factor system. This is also call an **obstruction** associated to the abstract kernel ψ.

We have the following proposition.

Proposition 10.2.1 *An abstract kernel ψ can be realized from an extension if and only if one of its obstruction is trivial.* ♯

We further analyze the obstruction ϕ, and its dependence on the choice of σ and the function f. First $\phi(x, y, z) \in Z(H)$, and the center is a characteristic subgroup. Thus, $\sigma_x(\phi(y, z, t)) \in Z(H)$ for all $x, y, z, t \in K$. For convenience, we adopt the additive notation for the operation of H. Note that H need not be abelian. However, $Z(H)$ is abelian. Thus, the Eq. (10.20) reads as

$$f(x, y) \;+\; f(xy, z) \;=\; \phi(x, y, z) \;+\; \sigma_x(f(y, z)) \;+\; f(x, yz) \;\cdots\cdots\; . \tag{10.21}$$

Proposition 10.2.2 *An obstruction ϕ associated to an abstract kernel ψ is a 3-***cocycle** *in the sense that*

$$\sigma_x(\phi(y, z, t)) \;-\; \phi(xy, z, t) \;+\; \phi(x, yz, t) \;-\; \phi(x, y, zt) \;+\; \phi(x, y, z) \;=\; 0$$

Proof Using (10.21), we express $f(x, y) + f(xy, z) + f(xyz, t)$ in two different ways.

$$\begin{aligned}
&f(x, y) + f(xy, z) + f(xyz, t)\\
&= \phi(x, y, z) + \sigma_x(f(y, z)) + f(x, yz) + f(xyz, t)\\
&= \phi(x, y, z) + \sigma_x(f(y, z)) + \phi(x, yz, t) + \sigma_x(f(yz, t)) + f(x, yzt)\\
&= \phi(x, y, z) + \phi(x, yz, t) + \sigma_x(f(y, z) + f(yz, t)) + f(x, yzt)\\
&= \phi(x, y, z) + \phi(x, yz, t) + \sigma_x(\phi(y, z, t) + \sigma_y(f(z, t) + f(y, zt)) + f(x, yzt)\\
&= \phi(x, y, z) + \phi(x, yz, t) + \sigma_x(\phi(y, z, t)) + \sigma_x(\sigma_y(f(z, t)) + \sigma_x(f(y, zt)) + f(x, yzt)
\end{aligned}$$

$= \phi(x, y, z) + \phi(x, yz, t) + \sigma_x(\phi(y, z, t)) + \sigma_x(\sigma_y(f(z, t)) - \phi(x, y, zt) + f(x, y) + f(xy, zt)$
$= \phi(x, y, z) + \phi(x, yz, t) + \sigma_x(\phi(y, z, t)) - \phi(x, y, zt) + f(x, y) + \sigma_{xy}(f(z, t)) + f(xy, zt).$

On the other hand,

$f(x, y) + f(xy, z) + f(xyz, t)$
$= f(x, y) + \phi(xy, z, t) + \sigma_{xy}(f(z, t)) + f(xy, zt)$
$= \phi(xy, z, t) + f(x, y) + \sigma_{xy}(f(z, t)) + f(xy, zt)$

Equating the two expressions for $f(x, y) + f(xy, z) + f(xyz, t)$, we get the desired result. ♯

Since $Z(H)$ is a characteristic subgroup of H, and σ_x is an automorphism of H, $\sigma_x/Z(H) \in Aut(Z(H))$. Again, since $\sigma_x\sigma_y = i_{f(x,y)}\sigma_{xy}$, $\sigma_x/Z(H)\sigma_y/Z_H = \sigma_{xy}/Z(H)$. Thus, σ induces a homomorphism from K to $Aut(Z(H)$ which associates x to $\sigma_x/Z(H)$. This induced homomorphism is again denoted by σ. Further, if τ is another lifting of ψ, then $\sigma_x Inn(H) = \tau_x Inn(H)$ for all $x \in K$. Hence there is a map g from K to H with $g(e) = 1$ such that $\sigma_x = i_{g(x)}\tau_x$ for all x. This means that the induced homomorphisms σ and τ from K to $Aut(Z(H))$ are same. It follows that the induced homomorphism σ from K to $Aut(Z(H))$ is independent of the lifting σ, and it depends only on the abstract kernel ψ. Let $Z^3_\psi(K, Z(H)) = Z^3_\sigma(K, Z(H))$ denote the set of 3 co-cycles associated to the abstract kernel ψ. Then this is an abelian group with respect to the obvious addition, and it is called the group of 3 co-cycles. Thus, for each choice of f satisfying the Eq. 10.19, we obtain an obstruction ϕ in $Z^3_\sigma(K, Z(H))$ described by the Eq. 10.21.

Now, we examine as to how the obstruction changes with different choices of the function f satisfying (10.19). Let f' be another map from $K \times K$ to H with $f'(e, y) = f'(x, e) = 0$ such that $\sigma_x\sigma_y = i_{f'(x,y)}\sigma_{xy}$. Let ϕ' be the obstruction to f'. Then there is another map p from $K \times K$ to $Z(H)$ such that

$$f'(x, y) = f(x, y) + p(x, y)$$

for all $x, y \in K$. Also

$$f'(x, y) + f'(xy, z) = \phi'(x, y, z) + \sigma_x(f'(y, z)) + f'(x, yz).$$

Putting the values of $f'(x, y)$, we get

$$f(x, y) + f(xy, z) + p(x, y) + p(xy, z) = \phi'(x, y, z) + \sigma_x(f(y, z) + p(y, z)) + f(x, yz) + p(x, yz).$$

In turn,

$$\phi(x, y, z) + \sigma_x(f(y, z)) + f(x, yz) + p(x, y) + p(xy, z) = \phi'(x, y, z) + \sigma_x(f(y, z)) + \sigma_x(p(y, z)) + f(x, yz) + p(x, yz).$$

Thus, there is a map p from $K \times K$ to $Z(H)$ with $p(e, y) = 0 = p(x, e)$ for all $x, y \in K$ such that

$$\phi(x, y, z) - \phi'(x, y, z) = \sigma_x(p(y, z)) - p(xy, z) + p(x, yz) - p(x, y).$$

Let us call a map δ from $K \times K \times K$ to $Z(H)$ a **3 co-boundary** if there is a map p from $K \times K$ to $Z(H)$ with $p(e, y) = 0 = p(x, e)$ for all $x, y \in K$ such that

$$\delta(x, y, z) = \partial p(x, y, z) = \sigma_x(p(y, z)) - p(xy, z) + p(x, yz) - p(x, y).$$

for all $x, y, z \in K$. It can be easily verified that a 3 co-boundary is also a 3 co-cycle, and the set $B^3_\sigma(K, Z(H))$ of co-boundaries is a subgroup of the group $Z^3_\sigma(K, Z(H))$. The quotient group $Z^3_\sigma(K, Z(H))/B^3_\sigma(K, Z(H))$ is called the third co-homology group denoted by $H^3_\sigma(K, Z(H))$. It follows that $\phi + B^3_\sigma(K, Z(H)) = \phi' + B^3_\sigma(K, Z(H))$. Thus, $\phi + B^3_\sigma(K, Z(H))$ is independent of the choice of f, and we have a map Obs (called the **obstruction** map) from the set $Hom(K, Out(H))$ of abstract kernels to the third co-homology group $H^3_\sigma(K, Z(H))$ defined by $Obs(\psi) = \phi + B^3_\sigma(K, Z(H))$, where ϕ is an obstruction to the choice f.

Proposition 10.2.3 *An abstract kernel ψ can be realized by an extension if and only if $Obs(\psi) = 0$.*

Proof Suppose that ψ can be realized by an extension. Then there is an extension E of H by K. Let t be a section of E. Then it gives rise to a factor system (K, H, σ^t, f^t) with $\psi(x) = \sigma^t_x Inn(H)$ (note that σ^t restricted to the center $Z(H)$ is independent of t). In turn,

$$\sigma^t_x \sigma^t_y = i_{f^t(x,y)} \sigma^t_{xy}$$

and

$$f^t(x, y) + f^t(xy, z) = \sigma^t_x(f^t(y, z)) + f^t(x, yz).$$

This shows that the obstruction ϕ to the choice f^t is 0. Hence $Obs(\psi) = 0$. Conversely, suppose that $Obs(\psi) = 0$. Then there is a map σ from K to $Aut(H)$, and a map f from $K \times K$ to H such that $\sigma_e = I_H$, $f(e, y) = 0 = f(x, e)$, $\psi(x) = \sigma_x Inn(H)$ and $\sigma_x \sigma_y = i_{f(x,y)} \sigma_{xy}$ for all $x, y \in K$. By our assumption, the obstruction ϕ to f is a co-boundary. Hence there is a map p from $K \times K$ to $Z(H)$ with $p(e, y) = 0 = p(x, e)$ for all $x, y \in K$ such that

$$\phi(x, y, z) = \sigma_x(p(y, z)) - p(xy, z) + p(x, yz) - p(x, y).$$

In other words, by (10.21),

$$f(x, y) + f(xy, z) = \sigma_x(p(y, z)) - p(xy, z) + p(x, yz) - p(x, y) + \sigma_x(f(y, z)) + f(x, yz).$$

for all $x, y, z \in K$. Take $f' = f - p$. Then $i_{f(x,y)} = i_{f'(x,y)}$ for all $x, y \in K$ and

$$f'(x, y) + f'(xy, z) = \sigma_x(f'(y, z)) + f'(x, yz).$$

This shows that (K, H, σ, f') is a factor system. Let E be the corresponding extension of H by K. Clearly, the associated abstract kernel is the given abstract kernel ψ. ♯

Proposition 10.2.4 *Let H and K be groups, and ϕ be a 3 co-cycle representing an element in $H^3_\sigma(K, Z(H))$, where σ is a homomorphism from K to $Aut(Z(H))$. Assume that K contains more than two elements. Then there is a group (free in some sense) G with $Z(G) = Z(H)$, and an abstract kernel $\psi \in Hom(K, Out(G))$ inducing the homomorphism σ from K to $Aut(G)$ such that $Obs(\psi) = [\phi]$.*

Proof Let ϕ be a 3 co-cycle in $Z^3_\sigma(K, Z(H))$. Let $X = K \times K - K \times \{e\} - \{e\} \times K$ and $F(X)$ the free group on X. For convenience, we use the additive notation for $F(X)$ also, although it is non-commutative. Let $G = Z(H) \times F(X)$ be the direct product of $Z(H)$ and $F(X)$. Then $Z(H)$ and $F(X)$ are naturally identified as subgroups of G. Indeed, for convenience, an element $(a, u) \in G$, $a \in Z(H)$, $u \in F(X)$ will be written as $a + u$. Since X contains more than one element, it follows that $Z(G) = Z(H)$. For each $x \in K$, we extend the map σ_x to an endomorphism $\overline{\sigma}_x$ of G by defining it on the free generating set X of $F(X)$ as follows:

$$\overline{\sigma_x}(y, z) = \phi(x, y, z) + (x, y) + (xy, z) - (x, yz) \cdots\cdots \tag{10.22}$$

for all $x, y, z \in K - \{e\}$. Since $\phi(x, y, z) = 0$ whenever any of the x, y, z are e, the identity (10.22) will make sense for all $x, y, z \in K$ if we identify (x, e) and (e, y) with the identity of $F(X)$ for all $x, y \in K$. Clearly, $\overline{\sigma}_e = I_G$. We show that

$$\overline{\sigma_x}\,\overline{\sigma_y} = i_{(x,y)}\overline{\sigma_{xy}} \cdots\cdots . \tag{10.23}$$

Since $\overline{\sigma_x}$ is an extension of σ_x for all $x \in K$, and σ is a homomorphism from K to $Aut(Z(H))$, for $a \in Z(H)$,

$$\overline{\sigma_x}(\overline{\sigma_y}(a)) = \sigma_x(\sigma_y(a)) = \sigma_{xy}(a) = (x, y) + \sigma_{xy}(a) - (x, y).$$

Thus, both sides of (10.23) are equal when restricted to $Z(H) = Z(G)$. It is sufficient, therefore, to show that both sides of (10.23) evaluated on (z, t) give the same result for all $z, t \in K$. Using (10.22) and the fact that ϕ is a 3-co-cycle, we get

$$\begin{aligned}
&\overline{\sigma_x}(\overline{\sigma_y}((z, t))) \\
&= \overline{\sigma_x}(\phi(y, z, t) + (y, z) + (yz, t) - (y, zt)) \\
&= \sigma_x(\phi(y, z, t)) + \overline{\sigma_x}((y, z)) + \overline{\sigma_x}((yz, t)) - \overline{\sigma_x}((y, zt)) \\
&= \sigma_x(\phi(y, z, t)) + \phi(x, y, z) + (x, y) + (xy, z) - (x, yz) + \phi(x, yz, t) + \\
&(x, yz) + (xyz, t) - (x, yzt) - (\phi(x, y, zt) + (x, y) + (xy, zt) - (x, yzt)) \\
&= \sigma_x(\phi(y, z, t)) + \phi(x, y, z) + (x, y) + (xy, z) - (x, yz) + \phi(x, yz, t) + \\
&(x, yz) + (xyz, t) - (x, yzt) + (x, yzt) - (xy, zt) - (x, y) - \phi(x, y, zt) \\
&= \sigma_x(\phi(y, z, t)) + \phi(x, y, z) + \phi(x, yz, t) - \phi(x, y, zt) + (x, y) + (xy, z) - \\
&(x, yz) + (x, yz) + (xyz, t) - (x, yzt) + (x, yzt) - (xy, zt) - (x, y)
\end{aligned}$$

$= (x, y) + \phi(xy, z, t) + (xy, z) + (xyz, t) - (xy, zt) - (x, y)$
$= i_{(x,y)}(\overline{\sigma_{xy}}(y, z))$

Thus,

$$\overline{\sigma_x}\,\overline{\sigma_y} = i_{(x,y)}\overline{\sigma_{xy}}$$

for all $x, y \in K$. In particular,

$$\overline{\sigma_x}\,\overline{\sigma_{x^{-1}}} = i_{(x,x^{-1})}\overline{\sigma_{xx^{-1}}} = i_{(x,x^{-1})}$$

This shows that $\overline{\sigma_x}$ is surjective, and $\overline{\sigma_{x^{-1}}}$ is injective for all $x \in K$. Hence $\overline{\sigma_x}$ is an automorphism of G. It follows that $\overline{\sigma}$ induces a homomorphism ψ from K to $Out(G)$ whose obstruction is the given co-cycle ϕ. ♯

The abstract kernel ψ, thus obtained, is universal in some sense. Let χ be a homomorphism from K to $Out(H)$ such that $Z(H) = Z(K)$, and the obstruction of χ is the obstruction of ψ. Then there is a map τ from K to $Aut(H)$ with $\tau_e = I_H$ and a map f from $K \times K$ to H with $f(e, y) = 0 = f(x, e)$ such that

(i) $\tau_x\tau_y = i_{f(x,y)}\tau_{xy}$,
(ii) $\tau_x = \sigma_x$ when restricted to $Z(H) = Z(G)$ and
(iii) $f(x, y) + f(xy, z) - f(x, yz) - \tau_x(f(y, z)) = \phi(x, y, z) = (x, y) + (xy, z) - (x, yz) - \overline{\sigma_x}(y, z)$

for all $x, y, z \in K$. Using the universal property of the free group $F(X)$, we get a unique homomorphism ρ from G to H subject to $\rho((x, y)) = f(x, y)$ and $\rho(h) = h$ for all $h \in Z(H) = Z(G)$. In turn, $\rho \circ \overline{\sigma_x} = \tau_x \circ \rho$ for all $x \in K$.

Example 10.2.5 In this example, we discuss the extensions of a group by a free group F (say). Let ψ is an abstract kernel of F to H. In other words ψ is a homomorphism from F to $Out(H)$. Since F is free, we have a homomorphism σ from F to $Aut(H)$ such that $\nu \circ \sigma = \psi$. The semi-direct product extension induced by σ is an extension of H by F with abstract kernel ψ. Further, since F is free, any extension E given by

$$E \equiv 1 \longrightarrow H \xrightarrow{\alpha} G \xrightarrow{\beta} F \longrightarrow 1.$$

splits. Thus, corresponding to any abstract kernel of F to H, there is one and only one equivalence class of extension of H by F which is a split extension. In particular, $H^2_\sigma(F, H) = \{0\}$

Example 10.2.6 In this example we discuss the extensions of a group by a cyclic group. The case of infinite cyclic group is already included in the above example. We discuss the extension of a group H by the cyclic group $\mathbb{Z}_m$, $m \geq 2$. Let ψ be an abstract kernel of $\mathbb{Z}_m$ to H. Let $\sigma_{\bar{1}}$ be an automorphism of H such that $\sigma_{\bar{1}} Inn(H) = \psi(\bar{1})$. Then the map σ from $\mathbb{Z}_m$ to $Aut(H)$ given by $\sigma_{\bar{i}} = (\sigma_{\bar{1}})^i$ is a lifting of the abstract kernel ψ. Clearly, $(\sigma_{\bar{1}})^m \in Inn(H)$. Note that, for convenience, the image of $\bar{i}$ under the map σ is being denoted by $\sigma_{\bar{i}}$. Let $G = H \times \mathbb{Z}_m$. Let $h_0 \in H$ such that $\sigma_{\bar{1}}(h_0) = h_0$.

Note that at least one such h_0 exists, for if worst comes $h_0 = 1$ will do. Define a product in G by

$$(a, \bar{i})(b, \bar{j}) = (a\sigma_{\bar{i}}(b), \overline{i+j})$$

in case $i + j \leq m - 1$, and

$$(a, \bar{i})(b, \bar{j}) = (a\sigma_{\bar{i}}(b)h_0, \overline{i+j}),$$

otherwise. It can be easily seen that G is a group. The identity is $(1, \bar{0})$, the inverse of $(h, \bar{i})$ is $(k, \overline{m-i})$, where $\sigma_i(k) = h^{-1}(h_0)^{-1}$. We have an extension E of H by $\mathbb{Z}_m$ given by

$$E \equiv 1 \longrightarrow H \overset{\alpha}{\rightarrow} G \overset{\beta}{\rightarrow} \mathbb{Z}_m \longrightarrow 1,$$

where $\alpha(a) = (a, \bar{0})$ and β is the second projection. Consider the section t of E given by $t(\bar{i}) = (1, \bar{i})$. Then $(\sigma_{\bar{i}}^t(h), 0) = t(\bar{i})(h, 0)(t(\bar{i}))^{-1} = (1, \bar{i})(h, 0)(1, \bar{i})^{-1} = (\sigma_{\bar{i}}(h), \bar{i})(h_0^{-1}, \overline{m-i}) = (\sigma_{\bar{i}}(h)h_0^{-1}h_0, \bar{0}) = (\sigma_{\bar{i}}(h), 0)$. This shows that $\sigma^t = \sigma$. It follows that the abstract kernel associated to this extension is ψ. Thus, every abstract kernel of $\mathbb{Z}_m$ to H can be realized from an extension of H by $\mathbb{Z}_m$. Note that $f^t(\bar{i}, \bar{j}) = 1$ for $i + j \leq m - 1$ and h_0 otherwise. Observe that $(t(\bar{1}))^m = ((\bar{1}, \bar{1}))^m = (h_0, \bar{0})$ and $(\sigma_{\bar{1}})^m = i_{h_0}$.

It also follows from the above discussion that any extension of H by $\mathbb{Z}_m$ with given abstract kernel ψ is determined by a choice of an element h_0 of H which is fixed by σ. There is a corresponding section t such that $t(\bar{1})^m = h_0$. Let t' be an other section of the extension with $\sigma^{t'} = \sigma^t = \sigma$. Then there is an element $a \in H$ such that $t'(\bar{1}) = (a, \bar{0})t(\bar{1}) = (a, \bar{0})(1, \bar{1}) = (a, \bar{1})$ and $(\sigma_{\bar{1}}(h), \bar{0}) = i_{(\bar{1},\bar{1})}((h, \bar{0})) = i_{(a,\bar{1})}(h, \bar{0}) = (a, \bar{1})(h, \bar{0})((a, \bar{1}))^{-1} = (a\sigma_{\bar{1}}(h)a^{-1}, \bar{0})$. This shows that $a\sigma_{\bar{1}}(h)a^{-1} = \sigma_{\bar{1}}(h)$ for all $h \in H$. This means that $a \in Z(H)$. Also, $t'(\bar{1})^m = (a, \bar{1})^m = (N_\sigma(a)h_0, 0)$, where N_σ is a map from H to H given by $N_\sigma(h) = h\sigma(h)\sigma^2(h)\cdots\sigma^{m-1}(h)$. The map N_σ is called **norm** of σ. Clearly, N_σ maps $Z(H)$ to itself, and when restricted to $Z(H)$, it is a homomorphism. Also $N_\sigma(a)$ and $N_\sigma(a)h_0$ are invariant under σ. Clearly, the choice $h_1 = N_\sigma(a)h_0$ determines another equivalent extension of H by K with prescribed abstract kernel ψ in the manner described above, and also any equivalent extension determines such an element *a in* $Z(H)$. To summarize, we have the following:

"Let H be group, and m be a natural number. Let ψ be an abstract kernel of $\mathbb{Z}_m$ to H, i.e., ψ is a homomorphism from $\mathbb{Z}_m$ to $Out(H)$. There is a lifting map σ (not necessarily a homomorphism) from $\mathbb{Z}_m$ to $Aut(H)$with $\sigma(\bar{0}) = I_H$such that $\sigma^i(h)Inn(H) = \psi(h)$ for all $h \in H$. Let $Fix(\sigma)$ denote the set $\{h \in H \mid \sigma(h) = h\}$ of elements of H fixed by σ. Clearly, $Fix(\sigma)$ is a subgroup of H. Also, $N_\sigma(Z(H))$ is a normal subgroup of $Fix(\sigma)$. The construction described above which constructs, for each $h \in Fix(\sigma)$ an extension of H by $\mathbb{Z}_m$ induces a natural bijection from the group $Fix(\sigma)/N_\sigma(Z(H))$ to the set $EXT_\sigma(\mathbb{Z}_m, H)$ of equivalence classes of extensions of H by $\mathbb{Z}_m$. In turn, if H is an abelian group, then it induces an isomorphism from $Fix(\sigma)/N_\sigma(H)$ to the second co-homology group $H^2_\sigma(\mathbb{Z}_m, H)$. In particular, if σ is trivial, then $Fix(\sigma) = H$ and $N_\sigma(H) = \{h^m \mid h \in H\}$. Thus, in this case $H^2(Z_m, H)$

is the quotient group H/H^m. In particular, $H^2(\mathbb{Z}_m, \mathbb{Z})$ is isomorphic to $\mathbb{Z}_m$. Also if D is a group such that every element of H is a m^{th} power, then $H^2(\mathbb{Z}_m, H) = \{0\}$"

Exercises

10.2.1 Find all extensions of (i) $\mathbb{Z}$ by $\mathbb{Z}$,
(ii) $\mathbb{Q}$ by $\mathbb{Z}$,
(iii) $\mathbb{Z}$ by $\mathbb{Z} \times \mathbb{Z}$,
(iv) a finite group by $\mathbb{Z}$,
(v) Q_8 by $\mathbb{Z}_2$,
(vi) D_8 by $\mathbb{Z}_m$,
(vii) S_5 by $\mathbb{Z}_2$,
(viii) A_4 by $\mathbb{Z}_2$,
(ix) $\mathbb{Z}_2$ by $\mathbb{Z}_2 \times \mathbb{Z}_2$,
(x) $\mathbb{Z}_p \times \mathbb{Z}_p$ by $\mathbb{Z}_p$,
(xi) Q_8 by V_4
up to equivalence.

10.2.2 Characterize groups all of whose extensions are split extensions.

10.2.3 There are several splittings of a split extension. How are they related?

10.2.4 Give a proof of the Proposition 10.1.17.

10.2.5 Show that the kernel of the natural homomorphism η from $Aut(G)$ to $Aut(G/Z(G))$ is $C_{Aut(G)}(Inn(G))$.

10.2.6 Show that a group G is free if and only if every extension by G splits.

10.2.7 Show, by means of an example, that the number of non-isomorphic groups G having a normal subgroup H such that G/H is isomorphic to a fixed group K may be strictly less than the number of equivalence classes of extensions of H by K. Hint. Look at the Exercise 10.2.1 (ix).

10.2.8 Describe the extensions of H by $\mathbb{Z}_m \times \mathbb{Z}_n$.

10.2.9 Describe all extensions of a group of order 5 by a group of order 4. Hence describe all groups of order 20.

10.2.10 Let τ be the automorphism of the Kleins four group V_4 given by $\tau(a) = b$, $\tau(b) = c$. Find $H^2_\sigma(\mathbb{Z}_3, V_4)$, where σ is the homomorphism from $\mathbb{Z}_3$ to $Aut(V_4)$ given by $\sigma_{\overline{1}} = \tau$.

10.3 Central Extensions, Schur Multiplier

In this section we study those extensions of abelian groups for which the associated abstract kernels are trivial homomorphisms. In other words, we are interested in extensions G of an abelian group A by a group G for which A is a subgroup of the center of G. More precisely, we have the following definition.

Definition 10.3.1 An extension E of H by K given by

$$E \equiv 1 \longrightarrow H \stackrel{\alpha}{\rightarrow} G \stackrel{\beta}{\rightarrow} K \longrightarrow 1.$$

is called a **central extension** if $\alpha(H) \subseteq Z(G)$.

Example 10.3.2 Any group G is naturally a central extension of its center $Z(G)$ by the group $Inn(G)$ of its inner automorphisms. Thus, the quaternion group Q_8 is a central extension of $\mathbb{Z}_2$ by the Kliens four group V_4.

Let F be a field, and $GL(n, F)$ be the general linear group of invertible $n \times n$ matrices with entries in the field F. The center $Z(GL(n, F))$ of $GL(n, F)$ consists of all scalar matrices aI_n, $a \in F^\star = F - \{0\}$. The quotient group $GL(n, F)/Z(GL(n, F))$ is called the **projective general linear group**, and it is denoted by $PGL(n, F)$. Thus, $GL(n, F)$ is a central extension of its center by the projective general linear group $PGL(n, F)$. We have the short exact sequence

$$1 \longrightarrow F^\star \stackrel{\alpha}{\rightarrow} GL(n, F) \stackrel{\nu}{\rightarrow} PGL(n, F) \longrightarrow 1 \cdots, \tag{10.24}$$

where $F^\star$ is the multiplicative group of the field, α is the map given by $\alpha(a) = aI_n$, and ν is the quotient map. The above exact sequence represents a central extension of $F^\star$ by $PGL(n, F)$. Mostly, groups can be represented as a subgroup of a linear group. Recall that a homomorphism ρ from a group G to the group $GL(n, F)$ is called a linear representation of G of degree n over the field F (see Chap. 9). A homomorphism from a group K to $PGL(n, F)$ is called a **projective representation** of K over F. It may not be possible to lift a projective representation ρ from a group K to $PGL(n, F)$ to a linear representation from K to $GL(n, F)$. For example, the identity projective representation from $PGL(n, F)$ to itself cannot be lifted to a linear representation as the exact sequence (10.24) does not split. A natural question is "When can we lift a projective representation to a linear representation". This question was first tackled by Schur in the beginning of the twentieth century.

Let A be an abelian group, and G be a group. Then $Hom(G, A)$ is again an abelian group with respect to the obvious operation. Let α be a homomorphism from G to a group K. Then α induces a homomorphism $\alpha^\star$ from $Hom(K, A)$ to $Hom(G, A)$ defined by $\alpha^\star(\eta) = \eta \circ \alpha$.

Proposition 10.3.3 *Let*

$$1 \longrightarrow H \stackrel{\alpha}{\rightarrow} G \stackrel{\beta}{\rightarrow} K \longrightarrow 1.$$

be an exact sequence of groups, and A be an abelian group. Then the sequence

$$0 \longrightarrow Hom(K, A) \stackrel{\beta^\star}{\rightarrow} Hom(G, A) \stackrel{\alpha^\star}{\rightarrow} Hom(H, A)$$

is exact.

Proof The proof of the proposition is again an imitation of the proof of the Theorem 7.2.11, and it is left as an exercise. ♯

Remark 10.3.4 (i) As observed in the Remark 7.2.12, the sequence, in general, is not exact if we adjoin 0 in the right side of the sequence.
(ii) In the language of category theory, the above proposition is expressed by saying that for all abelian groups A, $Hom(-, A)$ is a left exact contra-variant functor from the category of groups to the category of abelian groups.

Let

$$1 \longrightarrow H \xrightarrow{\alpha} G \xrightarrow{\beta} K \longrightarrow 1 \cdots . \tag{10.25}$$

be a central extension of H by K, and A be an abelian group. Let us denote by $H^2(K, A)$ the second co-homology group $H^2_\sigma(K, A)$ in case the homomorphism σ from K to $Aut(A)$ is trivial. We define a homomorphism δ(called a connecting homomorphism) from $Hom(H, A)$ to $H^2(K, A)$ as follows:

Let t be a section of (10.25). Since the extension is central, σ^t is trivial in the sense that $\sigma^t_x(h) = h$ for all $x \in K$ and $h \in H$. The function f^t from $K \times K$ to H is given by $t(x)t(y) = \alpha(f^t(x, y))t(xy)$. Then f^t belongs to group $Z^2(K, H)$ of 2 co-cycles. Though σ^t is independent of the choice of the section t, f^t depends on the choice of t. However, if t' is another section of the extension, then f^t and $f^{t'}$ differ by a 2 co-boundary in $B^2(K, H)$. Let $\eta \in Hom(H, A)$. Then $\eta \circ f^t$ is a map from $K \times K$ to A. Since η is a homomorphism and f^t is a 2 co-cycle in $Z^2(K, H)$, $\eta \circ f^t$ is a 2 co-cycle in $Z^2(K, A)$. Also $\eta \circ f^t$ and $\eta \circ f^{t'}$ differ by a 2 co-boundary. This defines, unambiguously, an element $\eta \circ f^t + B^2(K, A)$ in $H^2(K, A)$. We define $\delta(\eta) = \eta \circ f^t + B^2(K, A)$. Clearly, δ defines a homomorphism.

We have the following fundamental exact sequence associated to the central extension (10.25).

Proposition 10.3.5 *For any abelian group A, we have the following natural* **fundamental exact sequence**

$$0 \longrightarrow Hom(K, A) \xrightarrow{\beta^\star} Hom(G, A) \xrightarrow{\alpha^\star} Hom(H, A) \xrightarrow{\delta} H^2(K, A) \cdots \tag{10.26}$$

associated to the central extension given by (10.25)

Proof In the light of the Proposition 10.3.3, it is sufficient to prove the exactness at $Hom(H, A)$. Let $\chi \in Hom(G, A)$. Then by the definition, $\delta(\alpha^\star(\chi)) = (\chi \circ \alpha) \circ f^t + B^2(K, A)$. Now, $t(x)t(y) = \alpha(f^t(x, y))t(xy)$. Hence $\chi(t(x)) + \chi(t(y)) = \chi(\alpha(f^t(x, y))) + \chi(t(xy))$. Thus, we have the map g from K to A given by $g(x) = \chi(t(x))$ with $g(e) = 0$ such that $\chi(\alpha(f^t(x, y))) = g(y) - g(xy) + g(x)$. This shows that $(\chi \circ \alpha) \circ f^t \in B^2(K, A)$, and so $\delta \circ \alpha^\star = 0$. It follows that $image\alpha^\star \subseteq ker\delta$. Let $\eta \in ker\delta$. Then $\eta \circ f^t \in B^2(K, A)$. Let g be a map from K to A with $g(e) = 0$ such that

$$\eta(f^t(x, y)) = g(y) - g(xy) + g(x) \cdots\cdots . \tag{10.27}$$

Every element of G is uniquely expressed as $\alpha(a)t(x)$ for unique $a \in H$ and $x \in K$. Define a map χ from G to A by $\chi(\alpha(a)t(x)) = \eta(a) + g(x)$. Then

$$\begin{aligned}
&\chi(\alpha(a)t(x)\alpha(b)t(y)) \\
&= \chi(\alpha(a)\alpha(\sigma_x^t(b))\alpha(f^t(x,y))t(xy)) \\
&= \chi(abf^t(x,y)t(xy)) \\
&= \eta(abf^t(x,y)) + g(xy) \\
&= \eta(a) + \eta(b) + \eta(f^t(x,y)) + g(xy) \\
&= \eta(a) + \eta(b) + g(x) + g(y) \text{ (by (10.27))} \\
&= \chi(\alpha(a)t(x)) + \chi(\alpha(b)t(y)).
\end{aligned}$$

This shows that χ is a homomorphism. Also $\chi(\alpha(a)) = \eta(a)$ for all $a \in H$. Thus, $\eta = \chi \circ \alpha = \alpha^\star(\chi)$. It also follows that $Ker\delta \subseteq image\alpha^\star$. ♯

In particular, we have the following exact sequence.

$$0 \longrightarrow Hom(K,H) \xrightarrow{\beta^\star} Hom(G,H) \xrightarrow{\alpha^\star} Hom(H,H) \xrightarrow{\delta} H^2(K,H) \cdots . \quad (10.28)$$

The main problem is to determine all central extensions by a group K up to equivalence. Some author term it as an extension of K instead of by K. But, we shall stick to our earlier terminology by calling it an extension by K.

Definition 10.3.6 A central extension

$$E \equiv 1 \longrightarrow H \xrightarrow{\alpha} G \xrightarrow{\beta} K \longrightarrow 1.$$

by K is said to be a **free central** extension if given any central extension

$$E' \equiv 1 \longrightarrow H' \xrightarrow{\alpha'} G' \xrightarrow{\beta'} K' \longrightarrow 1,$$

and a homomorphism η from K to K', there exists a morphism (ρ, τ, η) (not necessarily unique) from the extension E to the extension E'.

Let

$$1 \longrightarrow R \xrightarrow{i} F \xrightarrow{\nu} K \longrightarrow 1 \cdots\cdots . \quad (10.29)$$

be a presentation of K, i.e., F is a free group, R a normal subgroup of F, i the inclusion, and ν a surjective homomorphism with kernel R (note that R is also free by the Neilson Schreier theorem). The subgroup $[R,F] = <\{[r,x] = rxr^{-1}x^{-1} \mid r \in R, x \in F\}>$ is a normal subgroup of F contained in R. As such, we get an extension

$$1 \longrightarrow R/[R,F] \xrightarrow{\bar{i}} F/[R,F] \xrightarrow{\bar{\nu}} K \longrightarrow 1 \cdots . \quad (10.30)$$

by K induced by the presentation (10.29) of K. Clearly, this is a central extension by K.

Proposition 10.3.7 *The central extension given by (10.30) is a free central extension by K.*

Proof Let

$$1 \longrightarrow H' \xrightarrow{\alpha} G' \xrightarrow{\beta} K' \longrightarrow 1 \cdots\cdots . \tag{10.31}$$

be a central extension by K', and η be a homomorphism from K to K'. Since F is a free group, from the projective property of a free group, there is a homomorphism γ from F to G' such that $\beta \circ \gamma = \eta \circ \nu$. Since $\beta(\gamma(R)) = \eta(\nu(R)) = \{e\}, \gamma(R) \subseteq ker\beta$. By the exactness of (10.31), $\gamma(R) \subseteq \alpha(H')$. Again, since $\alpha(H') \subseteq Z(G')$, it follows that $\gamma([R, F]) = [\gamma(R), \gamma(F)] \subseteq [\alpha(H'), G'] = \{e\}$. Thus, γ induces a homomorphism τ from $F/[R, F]$ to G' such that $\beta \circ \tau = \eta \circ \overline{\nu}$. Clearly, $\beta(\tau(R/[R, F])) = \{e\}$. By the exactness, $\tau(R/[R, F]) \subseteq \alpha(H')$. Since α is injective, it induces a homomorphism ρ from $R/[R, F]$ to H' such that $\alpha \circ \rho$ is τ restricted to $R/[R, F]$. Thus, (ρ, τ, η) is a morphism from the extension (10.30) to (10.31). ♯

Proposition 10.3.8 *Let*

$$E \equiv 1 \longrightarrow H \xrightarrow{\alpha} G \xrightarrow{\beta} K \longrightarrow 1.$$

be a free central extension. Then the map δ in the associated fundamental sequence described in the Proposition 10.3.5 is surjective. More explicitly, for any abelian group A, the sequence

$$0 \longrightarrow Hom(K, A) \xrightarrow{\beta^\star} Hom(G, A) \xrightarrow{\alpha^\star} Hom(H, A) \xrightarrow{\delta} H^2(K, A) \longrightarrow 0$$

is exact.

Proof Let μ be a 2 co-cycle in $Z^2(K, A)$. Then (K, A, σ, μ) with σ the trivial map from K to $Aut(A)$ is a factor system. The corresponding associated extension

$$E' \equiv 1 \longrightarrow A \xrightarrow{\alpha'} G' \xrightarrow{\beta'} K \longrightarrow 1.$$

is a central extension with a section t' such that $t'(x)t'(y) = \alpha'(\mu(x, y))t'(xy)$. Since E is a free central extension, we have a homomorphism ρ from H to A and a homomorphism τ from G to G' such that the diagram

$$\begin{array}{ccccccccc}
1 & \longrightarrow & H & \xrightarrow{\alpha} & G & \xrightarrow{\beta} & K & \longrightarrow & 1 \\
 & & \downarrow{\scriptstyle \rho} & & \downarrow{\scriptstyle \tau} & & \downarrow{\scriptstyle I_K} & & \\
1 & \longrightarrow & A & \xrightarrow{\alpha'} & G' & \xrightarrow{\beta'} & K & \longrightarrow & 1
\end{array}$$

is commutative. For each $x \in K$, chose a $t(x) \in G$ such that $\tau(t(x)) = t'(x)$. Then $\beta(t(x)) = \beta'(t'(x)) = x$. This shows that t is a section of E. Now, $t(x)t(y) = \alpha(f^t(x, y))t(xy)$, where f^t is a 2 co-cycle in $Z^2(K, H)$. Further, $\alpha'(\mu(x, y))t'(xy) = t'(x)t'(y) = \tau(t(x))\tau(t(y)) = \tau(t(x)t(y)) = \tau(\alpha(f^t(x, y))t(xy)) = \tau(\alpha(f^t(x, y)))\tau(t(xy)) = \tau(\alpha(f^t(x, y)))t'(xy)$. This shows that $\alpha'(\rho(f^t(x, y))) = \alpha'(\mu(x, y))$. Since α' is injective, $\rho(f^t(x, y)) = \mu(x, y)$. By the definition, $\delta(\rho) = \mu + B^2(K, A)$. It follows that δ is surjective. ♯

Next, we try to describe the image of the connecting homomorphism δ in the fundamental exact sequence associated to a central extension under the assumption that A is a divisible group. Recall that an abelian group A is called a **divisible** group if for any $a \in A$ and an integer n, there is an element $b \in A$ such that $nb = a$. For example, $(\mathbb{Q}, +)$, $(\mathbb{Q}/\mathbb{Z}, +)$, $(\mathbb{R}, +)$, $(\mathbb{C}, +)$, $(\mathbb{C}^\star, \cdot)$, and the circle group $(S^1, \cdot)$ are all divisible groups. From the Corollary 7.2.31, a group D is a divisible group if and only if given any subgroup H of an abelian group G, all homomorphisms f from H to D are restrictions of homomorphisms from G to D. Equivalently, the functor $Hom(-, D)$ from the category of abelian groups to itself takes a short exact sequence to a short exact sequence.

Proposition 10.3.9 *Let*

$$E \equiv 1 \longrightarrow H \xrightarrow{\alpha} G \xrightarrow{\beta} K \longrightarrow 1.$$

be a central extension, and D be a divisible group. Then the image of δ in the fundamental exact sequence

$$0 \longrightarrow Hom(K, D) \xrightarrow{\beta^\star} Hom(G, D) \xrightarrow{\alpha^\star} Hom(H, D) \xrightarrow{\delta} H^2(K, D)$$

is isomorphic to $Hom([G, G] \bigcap \alpha(H), D)$. In particular, if the central extension is free central extension, then $H^2(K, D)$ is isomorphic to $Hom([G, G] \bigcap \alpha(H), D)$.

Proof By the fundamental theorem of homomorphism, $image\ \delta \approx Hom(H, D)/Ker\delta \approx Hom(H, D)/image\alpha^\star$. The map α induces injective homomorphism $\overline{\alpha}$ from $H/H \bigcap \alpha^{-1}([G, G])$ to $G/[G, G]$. Since D is divisible, $\overline{\alpha}^\star$ is a surjective homomorphism from $Hom(G/[G, G], D)$ to $Hom(H/H \bigcap \alpha^{-1}([G, G]), D)$. Also, since D is abelian, $\nu^\star$ from $Hom(G/[G, G], D)$ to $Hom(G, D)$ is an isomorphism, where ν is the quotient map. Further the diagram

$$\begin{array}{ccc} Hom(G/[G, G], D) & \xrightarrow{\overline{\alpha}^\star} & Hom(H/H \bigcap \alpha^{-1}([G, G]), D) \\ \Big\downarrow{\scriptstyle \nu^\star} & & \Big\downarrow{\scriptstyle \nu^\star} \\ Hom(G, D) & \xrightarrow{\alpha^\star} & Hom(H, D) \end{array}$$

is commutative. It follows that the image of $\alpha^\star$ is the image of $\nu^\star$. Again, since D is divisible, the following sequence is exact:

$$0 \longrightarrow Hom(H/H \bigcap \alpha^{-1}([G,G]), D) \stackrel{\nu^\star}{\rightarrow} Hom(H, D)$$
$$\stackrel{i^\star}{\rightarrow} Hom(H \bigcap \alpha^{-1}([G,G]), D) \longrightarrow 1.$$

Thus, $Hom(H, D)/image\alpha^\star$ is isomorphic to $Hom(H \bigcap \alpha^{-1}([G,G]), D)$. Clearly, $Hom(H \bigcap \alpha^{-1}([G,G]), D)$ is isomorphic to $Hom([G,G] \bigcap \alpha(H),\ D)$. This shows that *image* δ is isomorphic to $Hom([G,G] \bigcap \alpha(H),\ D)$. The last assertion follow from the Proposition 10.3.8. ♯

Corollary 10.3.10 *Given a central extension*

$$E \equiv 1 \longrightarrow H \stackrel{\alpha}{\rightarrow} G \stackrel{\beta}{\rightarrow} K \longrightarrow 1.$$

by K, $Hom([G,G] \bigcap \alpha(H),\ \mathbb{C}^\star)$ is a subgroup of $H^2(K, \mathbb{C}^\star)$. ♯

Corollary 10.3.11 *Given a presentation*

$$1 \longrightarrow R \stackrel{i}{\rightarrow} F \stackrel{\nu}{\rightarrow} K \longrightarrow 1.$$

of K, $H^2(K, \mathbb{C}^\star)$ is isomorphic to $Hom(([F,F] \bigcap R)/[R,F],\ \mathbb{C}^\star)$.

Proof The given presentation induces the free central extension

$$1 \longrightarrow R/[R,F] \stackrel{\bar{i}}{\rightarrow} F/[R,F] \stackrel{\bar{\nu}}{\rightarrow} K \longrightarrow 1.$$

by K. The result follows from Proposition 10.3.9. ♯

Proposition 10.3.12 *Let K be a finite group of order n. Then $H^2(K, \mathbb{C}^\star)$ is also finite abelian group in which order of each element divide n.*

Proof Let $f \in Z^2(K, \mathbb{C}^\star)$. Then

$$f(x,y)f(xy,z) = f(y,z)f(x,yz)$$

for all $x, y, z \in K$. Taking the product of the equation over all $z \in K$, we get

$$f(x,y)^n \prod_{z\in K} f(xy,z) = \prod_{z\in K} f(y,z) \prod_{z\in K} f(x,z)$$

Define a map g from K to $\mathbb{C}^\star$ by $g(x) = \prod_{z\in K} f(x,z)$. Then $g(e) = 1$ and the above equation reads as

$$f(x,y)^n = g(y)g(xy)^{-1}g(x).$$

This means that f^n is a co-boundary. Hence order of each element of $H^2(K, \mathbb{C}^\star)$ divides n. Selecting a n^{th} root $u(x)$ of $g(x)$ for each $x \in K$, with $u(e) = 1$, we get a map u from K to $\mathbb{C}^\star$. Define a map f' from $K \times K$ to $\mathbb{C}^\star$ by

$$f'(x, y) = f(x, y)u(y)^{-1}u(xy)u(x)^{-1}.$$

It follows that $f' + B^2(K, \mathbb{C}^\star) = f + B^2(K, \mathbb{C}^\star)$ and $f'(x, y)^n = 1$. Thus, for each $x, y \in K$, there are only finitely many possibilities for $f'(x, y)$. Since K is finite, $H^2(K, \mathbb{C}^\star)$ is finite. ♯

Corollary 10.3.13 (Schur) *Let G be a group such that $G/Z(G)$ is finite. Then the commutator subgroup $[G, G]$ of G is finite.*

Proof Suppose that $n = \mid G/Z(G) \mid$. Then by the Proposition 10.3.12, $H^2(G/Z(G), \mathbb{C}^\star)$ is finite, and order of each of its element divide n. By the Proposition 10.3.10, $Hom([G, G] \bigcap Z(G), \mathbb{C}^\star)$ is isomorphic to a subgroup of $H^2(G/Z(G), \mathbb{C}^\star)$. Thus, $Hom([G, G] \bigcap Z(G), \mathbb{C}^\star)$ is finite, and order of each element of $Hom([G, G] \bigcap Z(G), \mathbb{C}^\star)$ divides n. If $[G, G] \bigcap Z(G)$ contains an element a of infinite order, then $Hom(< a >, \mathbb{C}^\star) \approx \mathbb{C}^\star$ is infinite. Since $\mathbb{C}^\star$ is a divisible group, $i^\star$ from $Hom(< a >, \mathbb{C}^\star)$ to $Hom([G, G] \bigcap Z(G), \mathbb{C}^\star)$ is injective. This contradicts the fact that $Hom([G, G] \bigcap Z(G), \mathbb{C}^\star)$ is finite. This shows that $[G, G] \bigcap Z(G)$ is a torsion group. Suppose that G is finitely generated. Then $Z(G)$, being a subgroup of finite index, is also finitely generated. Hence $[G, G] \bigcap Z(G)$ is finitely generated torsion abelian group, and so it is finite. Suppose, now, that G is not finitely generated. Let S be a transversal to $Z(G)$. Then S is finite. Let $L = < S >$ be the subgroup generated by S. Then L is finitely generated subgroup of G. Let $x, y \in G$. Then there are elements $a, b \in Z(G)$ and $u, v \in S$ such that $x = au$ and $y = bv$. But, then $[x, y] = [u, v]$. This shows that $[L, L] = [G, G]$. Since $G = Z(G)L$, the center $Z(L)$ of L is contained in $Z(G)$. Thus, $Z(G) \bigcap L = Z(L)$. Further, $G/Z(G) = LZ(G)/Z(G) \approx L/Z(G) \bigcap L = L/Z(L)$. This means that $L/Z(L)$ is finite. It follows from the earlier proved fact that $[L, L]$ is finite. Hence $[G, G]$ is finite. ♯

Corollary 10.3.14 (Schur–Hopf Formula) *Let*

$$1 \longrightarrow R \xrightarrow{i} F \xrightarrow{\nu} K \longrightarrow 1.$$

be a free presentation of a finite group K. Then $H^2(K, \mathbb{C}^\star)$ is isomorphic to $([F, F] \bigcap R)/[R, F]$.

Proof By the Corollary 10.3.11, $H^2(K, \mathbb{C}^\star)$ is isomorphic to $Hom(([F, F] \bigcap R)/[R, F], \mathbb{C}^\star)$. Further, $R/[R, F] \subseteq Z(F/[R, F])$. Since F/R, being isomorphic to K, is finite, it follows that $(F/[R, F])/Z(F/[R, F])$ is finite. From Corollary 10.3.13, it follows that the commutator $[F, F]/[R, F]$ of $F/[R, F]$ is finite. In turn, $([F, F] \bigcap R)/[R, F]$ is finite abelian. Clearly, $Hom(\mathbb{Z}_m, \mathbb{C}^\star) \approx \mathbb{Z}_m$. Also Hom respects direct sums in the sense that $Hom(A \oplus B, C) \equiv Hom(A, C) \oplus Hom(B, C)$. Since every finite

abelian group is direct sum of finite cyclic groups, it follows that for any finite abelian group A, $Hom(A, \mathbb{C}^\star) \approx A$. This shows that for finite groups K, $H^2(K, \mathbb{C}^\star)$ is isomorphic to $([F, F] \bigcap R)/[R, F]$. ♯

It follows from the above result that for a finite group K, the group $([F, F] \bigcap R)/[R, F]$ is independent of the choice of a free presentation of K. Indeed, we show that for any group (not necessarily finite), it is independent of the choice of a free presentation of the group.

Proposition 10.3.15 *Let*

$$E_F \equiv 1 \longrightarrow R \xrightarrow{i} F \xrightarrow{\nu} K \longrightarrow 1.$$

be an extension by K representing a free presentation of K, and

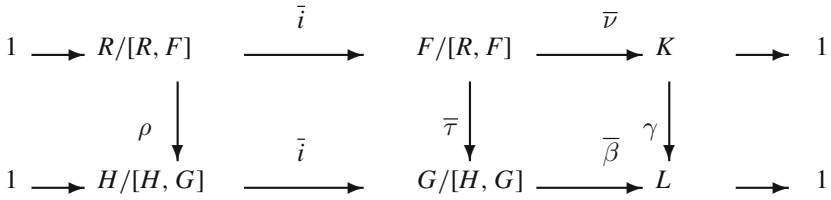
$$E \equiv 1 \longrightarrow H \xrightarrow{i} G \xrightarrow{\beta} L \longrightarrow 1.$$

an extension by L, where i denotes the inclusion map. Let γ be a homomorphism from K to L. Then there is a homomorphism τ from F to G (not necessarily unique) such that $(\tau/R, \tau, \gamma)$ is a morphism from the extension E_F to E. Further, the morphism $(\tau/R, \tau, \gamma)$ induces a homomorphism $\overline{\tau}$ from $[F, F]/[R, F]$ to $[G, G]/[H, G]$ such that the diagram

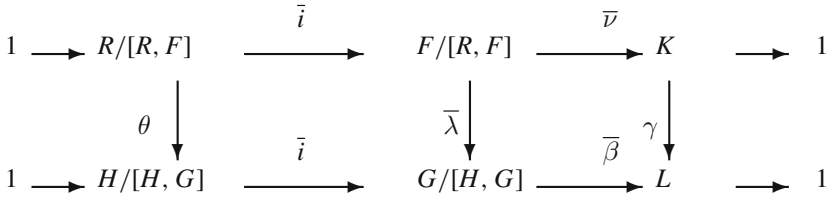
$$\begin{array}{ccccccccc}
1 & \longrightarrow & ([F,F] \bigcap R)/[R,F] & \xrightarrow{\bar{i}} & [F,F]/[R,F] & \xrightarrow{\overline{\nu}} & [K,K] & \longrightarrow & 1 \\
 & & \downarrow{\scriptstyle \rho} & & \downarrow{\scriptstyle \overline{\tau}} & & \downarrow{\scriptstyle \overline{\gamma}} & & \\
1 & \longrightarrow & ([G,G] \bigcap H)/[H,G] & \xrightarrow{\bar{i}} & [G,G]/[H,G] & \xrightarrow{\overline{\beta}} & [L,L] & \longrightarrow & 1
\end{array}$$

is commutative, where the maps $\bar{i}$ and $\overline{\nu}$ in the rows are the obvious induced maps, while ρ and $\overline{\gamma}$ are the restrictions of $\overline{\tau}$ and γ respectively. Further, if λ is an other homomorphism from F to G such that $(\lambda/R, \lambda, \gamma)$ is a morphism from the extension E_F to E. Then the induced homomorphism $\overline{\lambda}$ is same as $\overline{\tau}$.

Proof Since F is free, we have a homomorphism (not necessarily unique) τ from F to G such that $\beta \circ \tau = \gamma \circ \nu$. Clearly, $(\tau/R, \tau, \gamma)$ is a morphism from the extension E_F to E. Also τ maps $[R, F]$ to $[H, G]$. Thus, τ induces a homomorphism $\overline{\tau}$ from $F/[R, F]$ to $G/[H, G]$ such that the diagram



is commutative, where ρ is the restriction of $\overline{\tau}$. Again, since ν maps $[F, F]$ to $[K, K]$, β maps $[G, G]$ to $[L, L]$, and $\overline{\tau}$ maps $[F, F]/[R, F]$ to $[G, G]/[H, G]$, the diagram in the statement of the proposition is commutative. Suppose that there is an other homomorphism λ from F to G such that $(\lambda/R, \lambda, \gamma)$ is a morphism from the extension E_F to E. Then as for τ, λ induces a homomorphism $\overline{\lambda}$ from $F/[R, F]$ to $G/[H, G]$ such that the diagram



is commutative, where θ is the restriction of $\overline{\lambda}$. Let $\overline{x} \in F/[R, F]$. Then, $\overline{\beta}(\overline{\lambda}(\overline{x})) = \gamma(\overline{\nu}(\overline{x})) = \overline{\beta}(\overline{\tau}(\overline{x}))$. Hence $\overline{\lambda}(\overline{x}) = u(\overline{x})\overline{\tau}(\overline{x})$ for some $u(\overline{x}) \in H/[H, G]$. Since $H/[H, G]$ is contained in the center of $G/[H, G]$, $\overline{\lambda}([\overline{x}, \overline{y}]) = [\overline{\lambda}(\overline{x}), \overline{\lambda}(\overline{y})] = [u(\overline{x})\overline{\tau}(\overline{x}), u(\overline{y})\overline{\tau}(\overline{y})] = [\overline{\tau}(\overline{x}), \overline{\tau}(\overline{y})] = \overline{\tau}([\overline{x}, \overline{y}])$. This shows that the induced homomorphisms $\overline{\lambda} = \overline{\tau}$ when restricted to $[F, F]/[R, F]$, and so also to $([F, F] \bigcap R)/[R, F]$. ♯

Corollary 10.3.16 *Given two free presentations*

$$E_F \equiv 1 \longrightarrow R \xrightarrow{i} F \xrightarrow{\nu} K \longrightarrow 1.$$

and

$$E_{F'} \equiv 1 \longrightarrow R' \xrightarrow{i} F' \xrightarrow{\nu'} K \longrightarrow 1.$$

of K, the groups $([F, F] \bigcap R)/[R, F]$ and $([F', F'] \bigcap R')/[R', F']$ are naturally isomorphic.

Proof From the Proposition 10.3.15, for the identity map I_K from K to K, we have a unique homomorphism ρ from $([F, F] \bigcap R)/[R, F]$ to $([F', F'] \bigcap R')/[R', F']$ which is induced by a morphism $(\tau/R, \tau, I_K)$ from E_F to $E_{F'}$ and also a unique homomorphism ρ' from $([F', F'] \bigcap R')/[R', F']$ to $([F, F] \bigcap R)/[R, F]$ which is induced

by a morphism $(\tau'/R', \tau', I_K)$ from $E_{F'}$ to E_F. Thus, $((\tau' \circ \tau)/R, \tau' \circ \tau, I_K)$, and $(I_F/R, I_F, I_K)$ are both morphisms from E_F to itself and so they induce same homomorphisms from $([F, F] \bigcap R)/[R, F]$ to itself. This means that $\rho' \circ \rho$ is a homomorphism from $([F, F] \bigcap R)/[R, F]$ to itself which is induced by $(I_F/R, I_F, I_K)$. Hence $\rho' \circ \rho$ is the identity map on $([F, F] \bigcap R)/[R, F]$. Similarly $\rho \circ \rho'$ is the identity map on $([F', F'] \bigcap R')/[R', F']$. This shows that ρ and ρ' are isomorphisms. ♯

Since the group $([F, F] \bigcap R)/[R, F]$ is independent of a particular choice of the presentation of K, we have right to have the following definition.

Definition 10.3.17 Let

$$1 \longrightarrow R \xrightarrow{i} F \xrightarrow{\nu} K \longrightarrow 1.$$

be a free presentation of a group K. Then $([F, F] \bigcap R)/[R, F]$ is called the **Schur Multiplier** of K, and it is denoted by $M(K)$.

Corollary 10.3.18 *For finite groups K, $M(K)$ is finite, and it is isomorphic to $H^2(K, \mathbb{C}^\star)$. If order of K is n, then the order of each element of $M(K)$ divides n.*

Proof Follows from Corollary 10.3.14 and Proposition 10.3.12. ♯

Corollary 10.3.19 *The Schur multiplier M defines a co-variant functor from the category of groups to the category of abelian groups.*

Proof Let E_K denote the standard free multiplication presentation of K. More precisely,

$$E_K \equiv 1 \longrightarrow R_K \xrightarrow{i} F_K \xrightarrow{\mu} K \longrightarrow 1,$$

where F_K is the free group on K, μ the unique homomorphism from F_K to K induced by I_K and R_K the kernel of μ. Then the Schur multiplier$M(K) = (R_K \bigcap [F_K, F_K])/[R_K, F_K]$. Let λ be a homomorphism from K to L. Then from the Proposition 10.3.15, λ induces a unique homomorphism $M(\lambda)$ from $M(K) = (R_K \bigcap [F_K, F_K])/[R_K, F_K]$ to $M(L) = (R_L \bigcap [F_L, F_L])/[R_L, F_L]$. Further, if η is a homomorphism from the group L to a group U, $M(\eta) \circ M(\lambda)$ is the unique homomorphism which is induced by $\eta \circ \lambda$. Hence $M(\eta \circ \lambda) = M(\eta) \circ M(\lambda)$. Clearly, $M(I_K) = I_{M(K)}$. ♯

Proposition 10.3.20 *Let*

$$E \equiv 1 \longrightarrow R \xrightarrow{i} F \xrightarrow{\mu} K \longrightarrow 1$$

be a free presentation of a finite group K, where F is free group on a set $\{x_1, x_2, \ldots, x_n\}$ consisting of n elements. Then, $M(K) = (R \bigcap [F, F])/[R, F]$ is the finite torsion subgroup of $R/[R, F]$, and the torsion-free part $(R/[R, F])/((R \bigcap [F, F])/[R, F]) \approx R/(R \bigcap [F, F])$ is the free abelian of rank n. In turn, $R/[R, F]$ is isomorphic to the direct sum of $M(K)$ and $R/(R \bigcap [F, F])$.

Proof Since $R/[R,F]$ is contained in the center of $F/[R,F]$, it is abelian. Also $F/[F,F]$ is free abelian of rank n. Further $R/(R\bigcap[F,F])$ is isomorphic to $R[F,F]/[F,F]$ which is a subgroup of $F/[F,F]$. Since subgroup of a free abelian group is a free abelian group, $R/(R\bigcap[F,F])$ is a free abelian group of rank at the most n. Also $(F/[F,F])/(R[F,F]/[F,F]) \approx F/R[F,F]$. Hence $(F/[F,F])/(R[F,F]/[F,F])$ is finite. This means that $R[F,F]/[F,F]$ and so also $R/(R\bigcap[F,F])$ is free abelian of rank n. Next, by the Corollary 10.3.18, $M(K) = (R\bigcap[F,F])/[R,F]$ is a finite subgroup of $R/[R,F]$ such that $(R/[R,F])/((R\bigcap[F,F])/[R,F]) \approx R/(R\bigcap[F,F])$ is free abelian. This shows that $M(K)$ is a torsion part of $R/[R,F]$ and $R/(R\bigcap[F,F])$ is torsion-free part of $R/[R,F]$. It also follows that $R/[R,F]$ is direct sum of $M(K)$ and $R/(R\bigcap[F,F])$. ♯

Corollary 10.3.21 *Let*

$$E \equiv 1 \longrightarrow R \xrightarrow{i} F \xrightarrow{\mu} K \longrightarrow 1$$

be a free presentation of a finite group K, where F is free group on a set $\{x_1, x_2, \ldots, x_n\}$ consisting of n elements, and R the normal subgroup of F generated as normal subgroup by a set $\{w_1, w_2, \ldots, w_r\}$ consisting of r relators. Suppose that m is the minimum number of generators for $M(K)$. Then $r \geq n+m$. Equivalently, any set of generators of $M(K)$ contains at least $r-n$ elements.

Proof Since R is generated as a subgroup by the set $\{w_1, w_2, \ldots, w_r\}$ and its conjugates and $w_i[R,F] = ww_iw^{-1}[R,F]$ for all $w \in F$, it follows that $R/[R,F]$ is generated by the set $\{w_1[R,F], w_2[R,F], \ldots, w_r[R,F]\}$. From Proposition 10.3.20, it follows that $R/[R,F]$ is generated by at least $n+m$ elements. Thus, $r \geq n+m$. ♯

Corollary 10.3.22 *Let K is a finite group having a presentation with generating set $\{x_1, x_2, \ldots, x_n\}$, and the set $\{w_1, w_2, \ldots, w_r\}$ as irreducible set of defining relations. Then $r \geq n$. If $r = n$, then the Schur multiplier $M(K)$ is trivial. Further if $r = n+1$, then $M(K)$ is cyclic. If $r = n+2$, then $M(K)$ is either a finite cyclic group, or it is a $p-$ group which is direct product of two cyclic groups.*

Proof From the above corollary, $r \geq n$. If $r = n$, then the minimum number m for generating set of $M(K)$ is 0. Hence $M(K)$ is trivial. Suppose that $r = n+1$, then the minimum number m for generators of $M(K)$ is at most 1 and so $M(K)$ is finite cyclic. Suppose that $r = n+2$. Then the minimum number m for generators for $M(K)$ is at most 2. Since $M(K)$ is a finite abelian group, it is a direct product finite cyclic groups of prime power orders. Since direct product of cyclic groups of co-prime orders is a cyclic group, it follows that $M(K)$ is either cyclic or else it is a $p-$ *group* which is direct product of two cyclic $p-$ groups. ♯

Example 10.3.23 If F is a free group, then

$$1 \longrightarrow \{e\} \xrightarrow{i} F \xrightarrow{I_F} F \longrightarrow 1.$$

is a free presentation of F. Hence by the definition $M(F) = \{0\}$. In particular, Schur multiplier of an infinite cyclic group is trivial. Further,

$$\{0\} \longrightarrow m\mathbb{Z} \stackrel{i}{\to} \mathbb{Z} \stackrel{\nu}{\to} \mathbb{Z}_m \longrightarrow 1\{0\}.$$

is a free presentation of $\mathbb{Z}_m$. As such, the Schur multiplier $M(\mathbb{Z}_m) = \{0\}$. Alternatively, using, Corollary 10.3.18, $M(\mathbb{Z}_m) \approx H^2(\mathbb{Z}_m, \mathbb{C}^\star)$ and then, using, Example 10.2.6, we see that $H^2(Z_m, \mathbb{C}^\star) = \{0\}$. Using the fundamental theorem of finite abelian groups, we can find the Schur multipliers of all finite abelian groups provided we have a formula which relates $M(A \times B)$, $M(A)$ and $M(B)$ for all finite abelian groups. This will follow in sequel.

Example 10.3.24 Consider the quaternion group Q_8. It has a presentation $< i, j; i^4, i^2j^{-2}, iji^{-1} = j^{-1} >$. Indeed, i^4 is derivable from the other two relators as follows. $i^2 = ii^2i^{-1} = ij^2i^{-1} = j^{-2} = i^{-2}$. Hence $i^4 = 1$. Thus, Q_8 has a presentation $< i, j; i^2j^{-2}, iji^{-1} = j^{-1} >$ with two generators and two defining relations. As such, by Corollary 10.3.22, $M(Q_8)$ is trivial. More generally, a generalized quaternion group Q_{4n} of order $4n$ has a presentation $< x, y; x^{2n}, x^ny^{-2}, yxy^{-1}x >$. It is easily seen that this is a group of order $4n$. Here also x^{2n} is derivable from the other 2 relations as follows. We have $x^n = y^2 = yy^2y^{-1} = yx^ny^{-1} = x^{-n}$. Hence $x^{2n} = e$. Thus, Q_{4n} is a finite group generated by two elements with two defining relations. As such, by Corollary 10.3.22, $M(Q_{4n})$ is trivial.

Proposition 10.3.25 *Let*

$$1 \longrightarrow R \stackrel{i}{\to} F \stackrel{\mu}{\to} K \longrightarrow 1$$

be a free presentation of a finite group K, where F is a free group of rank n. Suppose that L is finite subgroup of $R/[R, F]$ such that the quotient of $R/[R, F]$ modulo L is generated by n elements. Then $M(K) \approx L$.

Proof Let A be the torsion part of $R/[R, F]$ and B the torsion-free part of $R/[R, F]$. Then B is free abelian of rank n. Also $R/[R, F] \approx A \oplus B$. Since L is finite, $L \subseteq A$. Thus, $(R/[R, F])/L \approx (A/L) \oplus B$. If $L \neq A$, then $(R/[R, F])/L$ cannot be generated by n elements. This shows that L is the Torsion subgroup of $R/[R, F]$. The result follows from Proposition 10.3.20. ♯

Example 10.3.26 The Dihedral group D_{2n} of order $2n$ is given by a presentation $< x, y; x^n, y^2, yxy^{-1} = x^{-1} >$. This is generated by two elements with three defining relations. Thus, $M(D_{2n})$ has to be cyclic. Every element of D_{2n} has unique representation as x^iy^j, $0 \leq i \leq n-1$, $0 \leq j \leq 1$, and so D_{2n} is a group of order $2n$. D_{2n} has a free central extension given by

$$1 \longrightarrow R/[R, F] \stackrel{i}{\to} F/[R, F] \stackrel{\mu}{\to} D_{2n} \longrightarrow 1,$$

where F is the free group on $\{x, y\}$, and R the subgroup of F generated by $\{x^n, y^2, yxy^{-1}x\}$ and its conjugates. As already described in Proposition 10.3.20, $R/[R, F]$ is in the center of $F/[R, F]$. The torsion-free part of $R/[R, F]$ is a free abelian group of rank 2, and its torsion part is $M(D_{2n})$. Since $uwu^{-1}[R, F] = w[R, F]$ for all $u \in F$ and $w \in R$, it follows that $R/[R, F]$ is an abelian subgroup of $F/[R, F]$ which is generated by $\{x^n[R, F], y^2[R, F], yxy^{-1}x[R, F]\}$. Denote $x^n[R, F]$, $y^2[R, F]$ and $yxy^{-1}x[R, F]$ by a, b and c respectively. Then $R/[R, F]$ is generated by $\{a, b, c\}$. Now, $a = x^n[R, F] = yx^n y^{-1}[R, F] = (yxy^{-1})^n[R, F] = (yxy^{-1}xx^{-1})^n[R, F] = (yxy^{-1}xx^{-1}[R, F])^n = (yxy^{-1}x[R, F](x^{-1}[R, F])^n = (yxy^{-1}x[R, F])^n x^{-n}[R, F] = c^n a^{-1}$. Thus $a^2 = c^n$. Suppose that $n = 2m$ is even. Put $d = a^{-1}c^m$. Then $d^2 = e$. Suppose that $d = e$. Then $a = c^m$ and so $x^n(yxy^{-1}x)^{-m} \in [R, F]$. Since $[R, F] \subseteq R$, it follows that x^n is derivable from the relation $yxy^{-1}x$. But the group given by a presentation $< x, y; y^2, yxy^{-1} = x^{-1} >$ is infinite Dihedral group. Hence d is an element of order 2 in $R/[R, F]$. Also the quotient group of $R/[R, F]$ modulo the subgroup $< d >$ generated by d is generated by $\{b < d >, c < d >\}$. It follows from the above proposition that $< d >$ is the torsion part of $R/[R, F]$. This shows that $M(D_{4m})$ is the cyclic group of order 2.

Next, suppose that $n = 2m + 1$ is odd. Now, put $d = ac^{-m}$. Then $d^2 = c$. Hence $a, c \in < d >$. Thus, in this case, $R/[R, F]$ is generated by $\{b, d\}$. Already, $R/[R, F]$ is direct sum of $M(D_{4m+2})$ with a free abelian group of rank 2. If $M(D_{4m+2})$ is nontrivial, $R/[R, F]$ can not be generated by two elements. Hence $M(D_{4m+2})$ is trivial.

Example 10.3.27 Consider the group G having presentation

$$< x, y; x^5, y^3, (xy)^2 > .$$

If we take $x = (12345)$ and $y = (152)$ in A_5, or $x = \overline{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}}$ and $y = \overline{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}$ in $PSL(2, 5)$, then $x^5 = y^3 = (xy)^2$ represent the identities in the respective groups. Also they generate the respective groups. As such there is a surjective homomorphism from G to A_5, and also a surjective homomorphism from G to $PSL(2, 5)$. Using the coset enumeration method of Coxeter and Todd, one finds that the order of G is 60 which is same as that of A_5, and also that of $PSL(2, 5)$. It follows that $< x, y; x^5, y^3, (xy)^2 >$ is presentation of A_5, and also a presentation of $PSL(2, 5)$. It also turns out that A_5 is isomorphic to $PSL(2, 5)$. Let F denote the free group on $\{x, y\}$ and R the normal subgroup of F generated by $\{x^5, y^3, (xy)^2\}$. To find the Schur multiplier, we need to find the Torsion part of $R/[R, F]$. Put $a = x^5[R, F]$, $b = y^3[R, F]$ and $c = (xy)^2[R, F]$. Then $R/[R, F]$ is a central subgroup of $F/[R, F]$ generated by $\{a, b, c\}$. We find relations between a, b, and c in $R/[R, F]$. Indeed, we show that $c^{30} = a^{12}b^{20}$. Since c is in the center of $F/[R, F]$, $c^2 = (xy)^2[R, F](xy)^2[R, F] = x(xy)^2yxy[R, F] = x^2yxy^2xy[R, F]$. Again inserting $(xy)^2$ between x^2 and y in the above expression, we find that $c^3 = x^3y(xy^2)^2xy[R, F]$. Repeating this process by putting $(xy)^2$ in between x^3 and y and again in turn putting $(xy)^2$ in between x^4 and y in the resulting expression, we find that $c^5 = x^5y(xy^2)^4xy[R, F] = ay(xy^2)^4xy[R, F]$. Since c and a commute with

all the elements of $F/[R,F]$, $a^{-1}c^5y^{-1}[R,F] = a^{-1}y^{-1}[R,F]c^5 = (xy^2)^4xy[R,F]$. Hence

$$c^5 = a(xy^2)^5[R,F]$$

Putting again $(xy)^2$ in between x and y^2 in the above expression, we get $c^{10} = a(x^2yxy^3[R,F])^5 = a(y^3[R,F])^5(x^2yx[R,F])^5 = ab^5(x^2yx[R,F])^5 = ab^5(x^2y(x^3y)^4x[R,F]$. Further, since c, a, b commute with all elements of $F/[R,F]$, we have $(x^2y(x^3y)^4[R,F] = y^{-1}x^{-3}(ab^5)^{-1}c^{10}$. In turn

$$c^{10} = ab^5(x^3y[R,F])^5$$

Again, iterating the same procedure, i.e., putting c in between x^3 and y in the above expression, we find

$$c^{15} = ab^5(x^4yx^2[R,F])^5$$

Iterating finally, we get

$$c^{30} = a^{12}b^{20}$$

Thus, if we put $p = a^6b^{10}c^{-15}$, then p^2 is the identity of $R/[R,F]$. Also if we put $q = a^2b^3c^{-5}$ and $r = ab^2c^{-3}$, then $R/[R,F] = < a, b, c > = < p, q, r >$. Hence the quotient group $(R/[R,F])/ < p >$ is generated by two elements. Since torsion-free part of $R/[R,F]$ is a free abelian group of rank 2, $< p >$ is the torsion part of $R/[R,F]$. This shows that the Schur multiplier of $A_5(PSL(2,5))$ is a group of order at most 2. Further, we have a central extension

$$1 \longrightarrow A \xrightarrow{i} SL(2,5) \xrightarrow{\mu} PSL(2,5) \longrightarrow 1,$$

where A is the center of $SL(2,5)$. Thus A is the group $\{\alpha I_2 \mid \alpha \in Z_5^\star, \alpha^2 = 1\} = \{I_2, -I_2\}$. This means that A is a group of order 2. Since $SL(2,5)$ is a perfect group, by the Proposition 10.3.9, $Hom(A, \mathbb{C}^\star) \approx A$ is a subgroup of the Schur multiplier $H^2(PSL(2,5), \mathbb{C}^\star)$ of $PSL(2,5)$. Thus, the Schur multiplier $M(PSL(2,5)) \approx M(A_5)$ is a group of order 2.

Five-Term Exact Sequence

For convenience, in all the group extensions of the type

$$1 \longrightarrow H \xrightarrow{\alpha} G \xrightarrow{\beta} K \longrightarrow 1,$$

α will be treated as inclusion map i. Thus, H is treated as a normal subgroup of G. Needless to say that there is no loss of generality.

Theorem 10.3.28 *To every group extension E given by*

$$E \equiv 1 \longrightarrow H \stackrel{i}{\rightarrow} G \stackrel{\beta}{\rightarrow} K \longrightarrow 1,$$

there is an associated connecting homomorphism $\delta(E)$ from $M(K)$ to $H/[H,G]$, and in turn the five-term exact sequence

$$M(G) \stackrel{M(\beta)}{\rightarrow} M(K) \stackrel{\delta(E)}{\rightarrow} H/[H,G] \stackrel{\overline{i}}{\rightarrow} G_{ab} \stackrel{\overline{\beta}}{\rightarrow} K_{ab} \longrightarrow 1,$$

which is natural in the sense that given any extension E' of H' by K', and a morphism $(\mu/H, \mu, \nu)$ from E to E', the diagram

$$\begin{array}{ccccccccc}
M(G) & \xrightarrow{M(\beta)} & M(K) & \xrightarrow{\delta(E)} & H/[H,G] & \xrightarrow{\overline{i}} & G_{ab} & \xrightarrow{\overline{\beta}} & K_{ab} \longrightarrow 1 \\
\downarrow{\scriptstyle M(\mu)} & & \downarrow{\scriptstyle M(\nu)} & & \downarrow{\scriptstyle \overline{\mu}} & & \downarrow{\scriptstyle \overline{\mu}} & & \downarrow{\scriptstyle \overline{\nu}} \\
M(G') & \xrightarrow{M(\beta')} & M(K') & \xrightarrow{\delta(E')} & H'/[H',G'] & \xrightarrow{\overline{i'}} & G'_{ab} & \xrightarrow{\overline{\beta'}} & K'_{ab} \longrightarrow 1
\end{array}$$

is commutative, where $G_{ab} = G/[G,G]$ and $K/[K,K]$ are the abelianizers of G and K respectively.

Proof By the Corollaries 10.3.16 and 10.3.17, M is a functor from the category of groups to the category of abelian groups. Thus, it is sufficient to establish the five-term exact sequence with a choice of a free presentation of G and that of K. Let

$$1 \longrightarrow R \stackrel{i}{\rightarrow} F \stackrel{\mu}{\rightarrow} G \longrightarrow 1,$$

be a free presentation of G. Let $S = \mu^{-1}(H)$. Then $R \subseteq S$ and we have a free presentation

$$1 \longrightarrow S \stackrel{i}{\rightarrow} F \stackrel{\beta o \mu}{\rightarrow} K \longrightarrow 1,$$

of K. Clearly, μ takes S to H, and indeed, $[S,F]$ to $[H,G]$. In turn, we have a natural map $\delta(E)$ from $M(K) \approx (S \bigcap [F,F])/[S,F]$ to $H/[H,G]$ given by $\delta(E)(s[S,F]) = \mu(s)[H,G]$. This gives us a five-term sequence

$$M(G) \stackrel{M(\beta)}{\rightarrow} M(K) \stackrel{\delta(E)}{\rightarrow} H/[H,G] \stackrel{\overline{i}}{\rightarrow} G_{ab} \stackrel{\overline{\beta}}{\rightarrow} K_{ab} \longrightarrow 1.$$

We prove the exactness of the sequence. Since β is surjective, $\overline{\beta}$ is surjective. Again, since $\beta o i$ is the trivial map, *image* $\overline{i} \subseteq ker \overline{\beta}$. Suppose that $\overline{\beta}(g[G,G]) = [K,K]$. Then $\beta(g) \in [K,K]$. Hence there is a $u \in [G,G]$ such that $\beta(g) = \beta(u)$. In turn, there is a $h \in H$ such that $g = hu$. Clearly, $\overline{i}(h[H,G]) = g[G,G]$. This proves the exactness at G_{ab}.

Next, since $\mu(s) \in H \bigcap [G, G]$ for all $s \in S \bigcap [F, F]$, it follows that

$$\bar{i}(\delta(E)(s[S, F])) = \bar{i}(\mu(s)[H, G]) = \mu(s)[G, G] = [G, G].$$

This shows that $image\delta(E) \subseteq ker\bar{i}$. Let $h[H, G] \in ker\bar{i}$. Then $h \in H \bigcap [G, G]$. This means that $h \in \mu(S \bigcap [F, F])$. It follows that $h[H, G] \in image\delta(E)$. Finally, we prove the exactness at $M(K)$. Let $r[R, F] \in M(G)$, $r \in R \bigcap [F, F]$. Then by definition, $\delta(E)(M(\beta)(r[R, F])) = \delta(E)(r[S, F]) = \mu(r)[H, G] = [H, G]$, for $\mu(r) = 0$. This shows that $imageM(\beta) \subseteq ker\delta(E)$. Further, suppose that $\delta(E)(s[S, F])\ \mu(s)[H, G] = [H, G]$, where $s \in S \bigcap [F, F]$. Then $\mu(s) \in [H, G] = \mu([S, F])$. Hence there is a $t \in [S, F]$ such that $\mu(s) = \mu(t)$. In turn, $\mu(st^{-1}) = e$. Thus, $s = rt$ for some $r \in R$. But, then $s[S, F] = rt[S, F] = r[S, F] = M(\beta)(r[R, F])$. It follows that $imageM(\beta) = ker(\delta(E))$. ♯

We give another interpretation of the group $M(G)$ as the group of commutator relations. For any group G, the commutator operation $[x, y] = xyx^{-1}y^{-1}$ can be easily seen to satisfy the following relations called the trivial commutator relations.
(i) $[x, x] = e$.
(ii) $[x, y][y, x] = e$.
(iii) $[xyx^{-1}, xzx^{-1}][x, z][z, xy] = e$.
(iv) $[xyx^{-1}, xzx^{-1}][z, y][yzy^{-1}z^{-1}, x] = e$.

The group $M(G)$ can be viewed as the group of nontrivial commutator relations in G. More precisely, consider the free group $F(X)$ on X, where $X = G \times G - (G \times \{e\} \bigcup \{e\} \times G)$. We identify (x, e) and (e, x) with identity of $F(X)$. From the universal property of a free group, we have a unique homomorphism η from $F(X)$ to G given by $\eta((x, y)) = [x, y] = xyx^{-1}y^{-1}$. Let $\Omega(G)$ denote the normal subgroup of $F(X)$ generated by the set of elements of the types (x, x), $(x, y)(y, x)$, $(xyx^{-1}, xzx^{-1})(x, z)(z, xy)$ and $(xyx^{-1}, xzx^{-1})(z, y)(yzy^{-1}z^{-1}, x)$. Clearly, $\Omega(G)$ is contained in the kernel of η. As such, it induces a unique homomorphism denoted, again, by η from $F(X)/\Omega(G)$ on to the commutator subgroup $[G, G]$ of G. The proof of the following theorem involves some computations which we leave and refer to the book "Schur Multiplier" by Karpilovski for the details of the computations.

Theorem 10.3.29 *The kernel of the above-described map η is the Schur multiplier $M(G)$, and we have the natural short exact sequence*

$$1 \longrightarrow M(G) \xrightarrow{i} F(X)/\Omega(G) \xrightarrow{\eta} [G, G] \longrightarrow 1.$$ ♯

Thus, $M(G)$ can be viewed as the group of commutator relations in G modulo the trivial commutator relations.

Definition 10.3.30 The group $F(X)/\Omega(G)$ introduced above is called the **nonabelian exterior power** of G, and it is denoted by $G \wedge G$.

Corollary 10.3.31 *If G is an abelian group, then $M(G) \approx G \wedge G$.* ♯

Tensor Product and Exterior Power of Groups

Let K and L be groups. Let G be another group. A map η from $K \times L$ to G is called a **bi-multiplicative map**, if

(i) $\eta(kk', l) = \eta(k, l)\eta(k', l)$ and
(ii) $\eta(k, ll') = \eta(k, l)\eta(k, l')$

for all $k, k' \in K$ and $l, l' \in L$. Note that if η is a bi-multiplicative map, then $\eta(e, l) = e = \eta(k, e)$ for all $k \in K$ and $l \in L$.

Proposition 10.3.32 *For any pair of groups K and L, there is a pair $(K \otimes L, \eta)$, where $K \otimes L$ is a group with η a bi-multiplicative map from $K \times L$ to $K \otimes L$ which is universal in the sense that for any pair (G', η') with η' a bi-multiplicative map from $K \times L$ to G', there is a **unique** homomorphism μ from $K \otimes L$ to G' such that $\mu \circ \eta = \eta'$.*

Proof Take $K \otimes L$ to be the group with presentation $< X;\ R >$, where the generating set $X = \{(k, l) \in K \times L \mid k \neq e \neq l\}$ and the set R of relators is given by $R = \{(kk', l)((k', l))^{-1}((k, l))^{-1},\ (k, ll')((k, l'))^{-1}((k, l'))^{-1} \mid k, k' \in K \text{ and } l, l' \in L\}$. Thus, $K \otimes L = F(X)/H$, where H is the normal subgroup generated by R. The map η is given by $\eta(k, l) = (k, l)H$. We denote $\eta(k, l)$ by $k \otimes l$. Clearly, η is a bi-multiplicative map, that is, $kk' \otimes l = (k \otimes l)(k' \otimes l)$ and also $k \otimes ll' = (k \otimes l)(k \oplus l')$. Let (G', η') with η' a bi-multiplicative map from $K \times L$ to G' be an other pair. From the universal property of free group, we have a unique homomorphism χ from $F(X)$ to G' such that $\chi(k, l) = \eta'(k, l)$. The supposition that η' a bi-multiplicative map from $K \times L$ to G' ensures that χ takes the relators R to e. This means that H is contained in the kernel of $Ker\chi$. By the fundamental theorem of homomorphism, χ induces a unique homomorphism μ from $K \otimes L$ to G' such that $\mu(\eta(k, l)) = \mu((k, l)H) = \chi(k, l) = \eta'(k, l)$. ♯

Proposition 10.3.33 *The pair $(K \otimes L, \eta)$ introduced above is unique in the sense that if (G', η') is another such pair, then there is a unique isomorphism μ from $K \otimes L$ to G' such that $\mu \circ \eta = \eta'$.*

Proof From the universal property of the pair $(K \otimes L, \eta)$ established in the above proposition, there is a unique homomorphism μ from $K \otimes L$ to G' such that $\mu \circ \eta = \eta'$. Since the pair (G', η') is also assumed to have the same universal property, there is a unique homomorphism ν from G' to $K \otimes L$ such that $\nu \circ \eta' = \eta$. Thus, $\nu \circ \mu$ and $I_{K \otimes L}$ are both homomorphisms from $K \otimes L$ to itself such that $(\nu \circ \mu) \circ \eta = \eta = I_{K \otimes L} \circ \eta$. From the universal property of the pair $(K \otimes L, \eta)$, $\nu \circ \mu = I_{K \otimes L}$. Reversing the role of $(K \otimes L, \eta)$ and (G', η'), we get that $\mu \circ \nu = I_{G'}$. Thus, μ is an isomorphism with the required property. ♯

Definition 10.3.34 The pair $(K \otimes L, \eta)$ is called the **tensor product** of K and L. By the abuse of language we also say that $K \otimes L$ is a tensor product of K and L. The image $\eta((k, l))$ is denoted by $k \otimes l$.

Thus, (i) $(kk' \otimes l) = (k \otimes l)(k' \otimes l)$, (ii) $(k \otimes ll') = (k \otimes l)(k \otimes l')$. In turn, $e \otimes l = e = k \otimes e$ and $k^{-1} \otimes l = (k \otimes l)^{-1} = k \otimes l^{-1}$.

Proposition 10.3.35 *Let η be a bi-multiplicative map from $K \times L$ to G. Then the image $\eta(K \times L)$ of η generates an abelian subgroup of G.*

Proof For $k, k' \in K$ and $l, l' \in L$,

$$\eta(kk', ll') = \eta(kk', l)\eta(kk', l') = \eta(k, l)\eta(k', l)\eta(k, l')\eta(k', l')$$

On the other hand,

$$\eta(kk', ll') = \eta(k, ll')\eta(k', ll') = \eta(k, l)\eta(k, l')\eta(k', l)\eta(k', l').$$

Comparing, $\eta(k', l)\eta(k, l') = \eta(k, l')\eta(k', l)$ for all $k, k' \in K$ and $l, l' \in L$. This means that the elements of $\eta(K \times L)$ commute pairwise. ♯

Corollary 10.3.36 *The tensor product $K \otimes L$ of any two group K and L is abelian group, and it is isomorphic to $K_{ab} \otimes L_{ab}$, where K_{ab} denote the abelianizer $K/[K, K]$ of K.*

Proof Since $K \otimes L$ is generated by $\{k \otimes l \mid k \in K, l \in L\}$, it follows from the above proposition that $K \otimes L$ is abelian. Let us denote the coset $k[K, K]$ by $\overline{k}$. Define a map η from $K \times L$ to $K_{ab} \otimes L_{ab}$ by $\eta(k, l) = \overline{k} \otimes \overline{l}$. Evidently, η is a bi-multiplicative map. As such, it induces a unique homomorphism $\overline{\eta}$ from $K \otimes L$ to $K_{ab} \otimes L_{ab}$ subject to $\overline{\eta}(k \otimes l) = \overline{k} \otimes \overline{l}$ which is clearly surjective. We show that it is bijective by constructing its inverse. Now, $[k, k'] \otimes l = kk'k^{-1}k'^{-1} \otimes l = (k \otimes l)(k' \otimes l)(k^{-1} \otimes l)(k'^{-1} \otimes l) = e$ for all $k, k' \in K$ and $l \in L$. Since every element of $[K, K]$ is product of commutators, and taking tensor product is bi-multiplicative, it follows that $u \otimes l = e$ for all $u \in [K, K]$. Similarly, $k \otimes v = e$ for all $k \in K$ and $l \in [L, L]$. It follows that $\overline{k} = \overline{k'}$ implies that $k \otimes l = k' \otimes l$ for all $l \in L$ and also $\overline{l} = \overline{l'}$ implies that $k \otimes l = k \otimes l'$ for all $k \in K$. Thus, $\overline{k} = \overline{k'}$ and $\overline{l} = \overline{l'}$ implies that $k \otimes l = k' \otimes l'$. This ensures that we have a map χ from $K_{ab} \times L_{ab}$ to $K \otimes L$ defined by $\chi(\overline{k}, \overline{l}) = k \otimes l$. Clearly, χ is a bi-multiplicative map, and as such it induces a homomorphism $\overline{\chi}$ from $K_{ab} \otimes L_{ab}$ to $K \otimes L$ given by $\overline{\chi}(\overline{k} \otimes \overline{l}) = k \otimes l$. Clearly, $\overline{\chi}$ is inverse of $\overline{\eta}$. ♯

It is evident from the above corollary that the theory of tensor product of groups reduces to the theory tensor product of abelian groups through their abelianizers. As such, we state few results which follow from the corresponding results on the tensor products of abelian groups (modules over $\mathbb{Z}$) (refer to the Chap. 7 of the book).

Corollary 10.3.37 *$K \otimes L$ is isomorphic to $L \otimes K$.* ♯

Proposition 10.3.38 *Let*

$$1 \longrightarrow H \xrightarrow{\alpha} G \xrightarrow{\beta} K \longrightarrow 1$$

be an exact sequence, and L be a group. Then the sequence

$$H \otimes L \overset{\alpha\otimes I_L}{\to} G \otimes L \overset{\beta\otimes I_L}{\to} K \otimes L \longrightarrow 1$$

is exact. ♯

Proposition 10.3.39 *Let H, K, and L be groups. Then $(H \oplus K) \otimes L$ is naturally isomorphic to $(H \otimes L) \oplus (K \otimes L)$.* ♯

Proposition 10.3.40 *Let H, K, and L be groups. There is a tautological isomorphism from $(H \otimes K) \otimes L$ to $H \otimes (K \otimes L)$ which maps $(h \otimes k) \otimes l$ to $h \otimes (k \otimes l)$.* ♯

Proposition 10.3.41 *Let H, K, and L be groups with L being abelian. Then there is a natural isomorphism from $Hom(H, Hom(K, L))$ to $Hom(H \otimes K, L)$.* ♯

We, further state few results without proof which can be used for some computations. For the proof we refer to the book "Schur Multiplier" by Karpilovski.

Proposition 10.3.42 *Let H and K be groups. Then $M(H \oplus K) \approx M(H) \oplus M(K) \oplus (H \otimes K)$.* ♯

Thus, for finitely generated abelian group, $M(A \oplus B) \approx A \otimes B$, and so the Schur multiplier of a finitely generated abelian group is easily determined. For free products, we have the following.

Proposition 10.3.43 $M(H \star K) \approx M(H) \oplus M(K)$. ♯

Exercises

10.3.1 Let K be a subgroup of a group G, and A be an abelian group which is a trivial G-module. Show that we have a homomorphism $res_{(G,K)}$ from $H^2(G, A)$ to $H^2(K, A)$ given by $res_{(G,K)}(f + B^2(G, A)) = f/K \times K + B^2(K, A)$. The homomorphism $res_{(G,K)}$ is called the restriction homomorphism from G to K. Observe that if L is a subgroup of K, then $res_{(K,L)} o\, res_{(G,K)} = res_{(G,L)}$.

10.3.2 Let K be a subgroup of a group G of finite index n. Let $= \{e = x_1, x_2, \ldots, x_n\}$ be a right transversal to K in G. Given any element $g \in G$, for each $x_i \in S$, there is a unique element $\sigma_{x_i}(g) \in K$, and a unique element $x_i \star g \in S$ such that $x_i g = \sigma_{x_i}(g) x_i \star g$. Let f be a 2 co-cycle in $Z^2(K, A)$, where A is a trivial K-module. Show that the map $\bar{f}$ from $G \times G$ to A defined by

$$\prod_{i=1}^{n} f(\sigma_{x_i}(x), \sigma_{x_i}(xy))$$

is a 2 co-cycle in $Z^2(G, A)$. Show also that $f \in B^2(K, A)$ implies that $\bar{f} \in B^2(G, A)$. Deduce that we have a co-restriction homomorphism $cores_{(K,G)}$ from $H^2(K, A)$ to $H^2(G, A)$ given by $cores_{(K,G)}(f + B^2(K, A)) = \bar{f} + B^2(G, A)$. Show also that $cores_{(K,G)} o\, cores_{(L,K)} = cores_{(L,G)}$.

10.3.3 Let $a = f + B^2(K, A)$ be an element of $H^2(G, A)$. Show that $(cores_{(K,G)} \, o \, res_{(G,K)})(a) = a^n$, where K is a subgroup of index n.

10.3.4 Let K be a normal subgroup of G of index n, and $a \in H^2(K, A)$. Show that $(res_{(G,K)} o \, cores_{(K,G)})(a) = a^n$.

10.3.5 Use the fact that the group $\mathbb{R}^\star$ has only one nonidentity element -1 of finite order to show that $H^2(G, \mathbb{R}^\star) \approx H^2(G, \mathbb{Z}_2)$ for any finite group G.

10.3.6 Let G be a finite group, and D be a divisible group. Show that $H^2(G, D) = H^2(G, T(D))$, where $T(D)$ is the torsion part of D. Deduce that $M(G) = H^2(G, \mathbb{Q}/\mathbb{Z})$ for all finite groups G.

10.3.7 Compute $Q_8 \wedge Q_8$, and hence also $M(Q_8)$.

10.3.8 Use the five-term exact sequence associated to the extension

$$1 \longrightarrow \{1, -1\} \xrightarrow{i} Q_8 \xrightarrow{\nu} V_4 \longrightarrow 1. \qquad \sharp$$

to show that $M(Q_8) = \{0\}$.

10.3.9 Compute $P \wedge P$, where P is a non-abelian group of order p^3, p a prime. Hence compute $M(P)$.

10.3.10 Compute the Schur multiplier of a non-abelian group of order pq, p, and q are primes.

10.3.11 Find the Schur multipliers of A_4, and also of S_4.

10.3.12 Let G be a finite nilpotent group. Show that the Schur multiplier of G is the direct products of Schur multipliers of its Sylow subgroups.

10.3.13 Let m, n, r be positive integers such that $r^n \equiv 1 (mod\ m)$, $(m, n) = 1 = (n, r - 1)$. Let G be a group having a presentation $< \{x, y\}; \mid x^m = 1 = y^n = y^{-1}xy^{-r} >$. Using the Tietze transformation (see Algebra 1), reduce the number of relators to 2, and then show that $M(G)$ is trivial.

10.3.14 Show that there is a surjective homomorphism from $M(GL(R))$ to $M(K_1(R))$.

10.4 Lower K-Theory Revisited

In Chap. 7, Sect. 7.4, we introduced the Grothendieck group $K_0(R)$ and the Whitehead group $K_1(R)$ of a ring R. Recall that $K_1(R) = GL(R)/E(R)$, where $E(R)$ is the group generated by the elementary matrices. By the Whitehead lemma, $E(R)$ is the commutator subgroup $[GL(R), GL(R)]$ of $GL(R)$. In this section, we introduce

Milnor group $K_2(R)$ of a ring R which can be viewed in two ways: (i) The Schur multiplier $M(E(R))$ of the group of commutator relations among the elements of $E(R)$ modulo the trivial commutator relations, and (ii) the group of relations among the transvections E_{ij}^{λ} modulo the group of trivial relations, viz., the group of Steinberg relations. We describe it in detail.

Note: Usually, in the literature, an extension

$$E \equiv 1 \longrightarrow H \stackrel{i}{\rightarrow} G \stackrel{\beta}{\rightarrow} K \longrightarrow 1$$

is termed as extension of K, but we shall adhere to our terminology by calling it extension by K.

Proposition 10.4.1 *Let K be a perfect group in the sense that $[K, K] = K$, and*

$$E \equiv 1 \longrightarrow H \stackrel{i}{\rightarrow} G \stackrel{\beta}{\rightarrow} K \longrightarrow 1$$

be a central extension by K. Then the commutator subgroup $[G, G]$ is perfect, and

$$1 \longrightarrow H \bigcap [G, G] \stackrel{i}{\rightarrow} [G, G] \stackrel{\beta}{\rightarrow} K \longrightarrow 1$$

is also a central extension by K.

Proof Since K is perfect, $\beta([G, G]) = [K, K] = K$. Thus, given any element $a \in G$, there is an element $u \in [G, G]$ such that $\beta(a) = \beta(u)$. This means that every element in G is of the type hu for some $h \in H$ and $u \in [G, G]$. Let $a = hu$, $b = h'u'$, where $h, h' \in H \subseteq Z(G), u, u' \in [G, G]$ be arbitrary elements of G. Then $[a, b] = [u, u'] \in [[G, G], [G, G]]$. This shows that $[G, G] \subseteq [[G, G], [G, G]]$. Hence $[G, G]$ is perfect. The rest is evident. ♯

Proposition 10.4.2 *Let*

$$E \equiv 1 \longrightarrow H \stackrel{i}{\rightarrow} G \stackrel{\beta}{\rightarrow} K \longrightarrow 1$$

be a central extension by K, where G is perfect. Let λ and μ be a homomorphisms from G to G' inducing the morphisms $(\lambda/H, \lambda, I_K)$ and $(\mu/H, \mu, I_K)$ from E to a central extension E' given by

$$E' \equiv 1 \longrightarrow H' \stackrel{i}{\rightarrow} G' \stackrel{\beta}{\rightarrow} K \longrightarrow 1.$$

Then $\lambda = \mu$.

Proof Since G is perfect, the commutators generate G. Thus, it is sufficient to show that $\lambda([a, b]) = \mu([a, b])$ for all $a, b \in G$. For each $x \in G$, $\beta(\lambda(x)) = \beta(\mu(x))$, and so there is an element $u(x) \in H$ such that $\lambda(x) = u(x)\mu(x)$. Now, $\lambda([a, b]) = [\lambda(a), \lambda(b)] = [u(a)\mu(a), u(b)\mu(b)] = [\mu(a), \mu(b)] = \mu([a, b])$. ♯

Definition 10.4.3 A central extension

$$\Omega_K \equiv 1 \longrightarrow H \xrightarrow{\alpha} U \xrightarrow{\beta} K \longrightarrow 1$$

is called a **universal central extension** by K if given any central extension

$$E \equiv 1 \longrightarrow L \xrightarrow{\alpha} G \xrightarrow{\beta} K \longrightarrow 1$$

by K, there is a **unique** homomorphism ϕ from U to G inducing a morphism (ξ, ϕ, I_K) from Ω_K to E.

Proposition 10.4.4 *Universal central extension by K is unique up to equivalence.*

Proof Let

$$\Omega_K^{'} \equiv 1 \longrightarrow H' \xrightarrow{\alpha'} U' \xrightarrow{\beta'} K \longrightarrow 1$$

be another universal central extension by K. Then there is a unique homomorphism ϕ from U to U' inducing a morphism (ξ, ϕ, I_K) from Ω_K to $\Omega_K^{'}$, and there is a unique homomorphism ϕ' from U' to U inducing a morphism (ξ', ϕ', I_K) from $\Omega_K^{'}$ to Ω_K. But, then we have homomorphisms $\phi' o \phi$ and I_U inducing morphisms $(\xi' o \xi, \phi' o \phi, I_K)$ and (I_H, I_U, I_K) respectively. From the universal property of Ω_K, $\phi' o \phi = I_U$. Similarly, using the universal property of $\Omega_K^{'}$, $\phi o \phi' = I_{U'}$. This shows that Ω_K is equivalent to $\Omega_K^{'}$. ♯

Proposition 10.4.5 *If*

$$\Omega_K \equiv 1 \longrightarrow H \xrightarrow{\alpha} U \xrightarrow{\beta} K \longrightarrow 1$$

is a universal central extension, then U is perfect ($U = [U, U]$).

Proof Suppose that U is not perfect. Then $U/[U, U]$ is a nontrivial abelian group. Consider the direct product extension

$$1 \longrightarrow U/[U, U] \xrightarrow{i_1} U/[U, U] \times K \xrightarrow{p_2} K \longrightarrow 1$$

by K, where i_1 is the first inclusion, and p_2 is the second projection. Clearly, this is a central extension by K. Further, the map (ν, β) defined by $(\nu, \beta)(u) = (u[U, U], \beta(u))$, and the map$(0, \beta)$ defined by $(0, \beta)(u) = ([U, U], \beta(u))$ are two distinct homomorphisms from U to $U/[U, U] \times K$ which induce morphism from Ω_K to the given direct product extension. Hence Ω_K can not be a universal central extension. ♯

Since homomorphic image of a perfect group is a perfect group, we have

Corollary 10.4.6 *If K admits a universal central extension by K, then K is a perfect group.* ♯

Conversely,

Proposition 10.4.7 *Every perfect group K admits (of course, a unique) universal central extension by K.*

Proof Suppose that K is perfect. Let

$$F_K \equiv 1 \longrightarrow R \xrightarrow{i} F \xrightarrow{\eta} K \longrightarrow 1$$

be a free presentation of K. We have a central extension

$$E_K \equiv 1 \longrightarrow R/[R,F] \xrightarrow{\bar{i}} F/[R,F] \xrightarrow{\bar{\eta}} K \longrightarrow 1$$

by K. Since K is perfect, $[F/[R,F], F/[R,F]] = [F,F]/[R,F]$ is perfect (by the Proposition 10.4.1), and we have a central extension

$$\Omega_K \equiv 1 \longrightarrow R \bigcap [F,F]/[R,F] \xrightarrow{\bar{i}} [F,F]/[R,F] \xrightarrow{\bar{\eta}} K \longrightarrow 1$$

by K. We prove that this is a universal central extension by K. Let

$$E \equiv 1 \longrightarrow H \xrightarrow{i} G \xrightarrow{\beta} K \longrightarrow 1$$

be a central extension by K. Since F is free, there is a homomorphism ϕ from F to G inducing a morphism $(\phi/R, \phi, I_K)$ from F_K to E. Since E is a central extension by K, it induces a morphism $(\bar{\phi}/R, \bar{\phi}, I_K)$ from E_K to E, which in turn, induces a morphism from Ω_K to E. Since $[F,F]/[R,F]$ is perfect, by the Proposition 10.4.2 such a morphism is unique. ♯

Corollary 10.4.8 *If K is perfect, then $K \wedge K$ is also a perfect group, and it is the universal central extension of $M(K)$ by K. More precisely,*

$$1 \longrightarrow M(K) \xrightarrow{i} K \wedge K \xrightarrow{c} K \longrightarrow 1$$

is a universal central extension, where c is the commutator map given by $c(x \wedge y) = [x,y]$. ♯

Proposition 10.4.9 *A central extension*

$$\Omega \equiv 1 \longrightarrow H \xrightarrow{i} U \xrightarrow{\beta} K \longrightarrow 1$$

by K is a universal central extension by K if and only if U is perfect and every central extension by U splits.

Proof Suppose that given extension Ω is a universal central extension. Then by Proposition 10.4.5, U is perfect. Let

$$E \equiv 1 \longrightarrow L \xrightarrow{i} G \xrightarrow{\delta} U \longrightarrow 1$$

be a central extension by U. Then consider the extension

$$E' \equiv 1 \longrightarrow ker(\beta o\delta) \xrightarrow{i} G \xrightarrow{\beta o\delta} K \longrightarrow 1$$

by K. We first show that this is a central extension by K. Let $g \in ker(\beta o\delta)$. We need to show that $g \in Z(G)$. Since $(\beta(\delta(g))) = e$, $\delta(g) \in ker\beta = H \subseteq Z(U)$. Let $x \in G$. Since $\delta(g) \in Z(U)$, $\delta(xgx^{-1}) = \delta(x)\delta(g)\delta(x)^{-1} = e$. This shows that $xgx^{-1} \in L \subseteq Z(G)$. Again, since $Z(G)$ is a characteristic subgroup of G, it follows that $g \in Z(G)$. In turn, it follows that E' is a central extension by K. Since Ω is universal central extension, there is a unique homomorphism ϕ from U to G such that $(\beta o\delta)o\phi = \beta$. This shows that $(\delta o\phi)$ and I_U are homomorphisms from U to U which induce morphisms from Ω to itself. Since Ω is universal central extension, it follows that $(\delta o\phi) = I_U$. This shows that E is split exact sequence.

Conversely, suppose that U is perfect and every central extension by U splits. Let

$$E'' \equiv 1 \longrightarrow P \xrightarrow{i} G \xrightarrow{\delta} K \longrightarrow 1$$

be a central extension by K. In the light of the Proposition 10.4.2, it is sufficient to show the existence of a homomorphism η from U to G inducing a morphism from Ω to E''. Consider the subgroup $U \times_K G = \{(u, g) \mid \beta(u) = \delta(g)\}$ of $U \times G$. We have the extension

$$E''' \equiv 1 \longrightarrow H \xrightarrow{i_1} U \times_K G \xrightarrow{p_1} U \longrightarrow 1$$

which is clearly a central extension by U. From our hypothesis, the sequence splits. Let t be an splitting. Then there is a homomorphism ϕ from U to G such that $t(u) = (u, \phi(u)) \in U \times_K G$. But, then $\beta(\phi(u)) = u$. Thus, ϕ induces a morphism from Ω to E''. ♯

Let R be a commutative ring. Recall that the group $E(R)$ is perfect. As such, we have the universal central extension

$$1 \longrightarrow M(E(R)) \xrightarrow{i} E(R) \wedge E(R) \xrightarrow{c} E(R) \longrightarrow 1$$

by $E(R)$. The group $E(R) \wedge E(R)$ represent the group of commutator relations in the group $E(R)$ modulo the trivial commutator relations. We shall have another interpretation of this group.

We have the following definition.

Definition 10.4.10 The group $M(E(R))$ is called the **Milnor group** of the ring R, and it is denoted by $K_2(R)$.

We shall have another way to see the group $K_2(R)$. If f is a homomorphism from a ring R to a ring R', then f induces a natural homomorphism $E(f)$ from $E(R)$ to $E(R')$ given by $E(f)[a_{ij}] = [b_{ij}]$, where $b_{ij} = f(a_{ij})$. Clearly, $E(gof) = E(g)oE(f)$ and $E(I_R) = I_{E(R)}$. Further, since M defines a functor from the category of groups to the category of abelian groups, it follows that K_2 is a functor from the category of rings to the category of abelian groups in the sense that if f is a homomorphism from a ring R to R', it induces a homomorphism $K_2(f)$ from $K_2(R)$ to $K_2(R')$ such that (i) $K_2(gof) = K_2(g)oK_2(f)$ and (ii) $K_2(I_R) = I_{K_2(R)}$.

The following natural exact sequence relates K_1 and K_2 functors.

$$1 \longrightarrow K_2(R) \xrightarrow{i} E(R) \wedge E(R) \xrightarrow{c} GL(R) \xrightarrow{\nu} K_1(R) \longrightarrow 1,$$

where c represents the commutator map given by $c(x \wedge y) = [x, y]$.

Recall that the $n \times n$, $n \geq 3$ elementary matrices E_{ij}^{λ} with entries in a ring R with identity satisfy the following relations termed as Steinberg relations.

(i) $E_{ij}^{\lambda} E_{ij}^{\mu} = E_{ij}^{\lambda+\mu}$.
(ii) $[E_{ij}^{\lambda}, E_{kl}^{\mu}] = I_n$ for $i \neq l$ and $j \neq k$.
(iii) $[E_{ij}^{\lambda}, E_{jl}^{\mu}] = E_{il}^{\lambda\mu}$, $i \neq l$.
(iv) $[E_{ij}^{\lambda}, E_{ki}^{\mu}] = E_{jk}^{-\mu\lambda}$.

For each $n \geq 3$, let $St(n, R)$ denote the group generated by the set

$$\{x_{ij}^{\lambda} \mid 1 \leq i \leq n, 1 \leq j \leq n, \lambda \in R\}$$

subject to the relations

(i) $x_{ij}\lambda x_{ij}^{\mu} = x_{ij}^{\lambda+\mu}$.
(ii) $[x_{ij}^{\lambda}, x_{kl}^{\mu}] = e$ for $i \neq l$ and $j \neq k$.
(iii) $[x_{ij}^{\lambda}, x_{jl}^{\mu}] = x_{il}^{\lambda\mu}$, $i \neq l$.
(iv) $[x_{ij}^{\lambda}, x_{ki}^{\mu}] = x_{jk}^{-\mu\lambda}$.

For each n, we have the natural surjective homomorphism ϕ_n from $St(n, R)$ to $E(n, R)$. Clearly, $St(n, R)$ is a subgroup of $St(n + 1, R)$ in a natural way, and we have a chain

$$St(3, R) \subseteq St(4, R) \subseteq \cdots \subseteq St(n, R) \subseteq St(n + 1, R) \subseteq \cdots\cdots$$

of groups. The union $St(R)$ of the chain is a group called the **Steinberg group** of the ring R. Note that the maps ϕ_n respect the inclusion maps in the sense that $i_n o \phi_n = \phi_{n+1} o i_n$, where i_n are the respective inclusion maps. In turn, in limit, ϕ_n induces a surjective homomorphism ϕ from $St(R)$ to $E(R)$.

Theorem 10.4.11 *The short exact sequence*

$$1 \longrightarrow Ker\phi \xrightarrow{i} St(R) \xrightarrow{\phi} E(R) \longrightarrow 1$$

is a universal central extension by $E(R)$.

Before proceeding to prove the above theorem, let us have a corollary.

Corollary 10.4.12 *We have natural isomorphisms* $ker\phi \approx K_2(R) \approx M(R)$ *and* $St(R) \approx E(R) \wedge E(R)$. ♯

Thus, $K_2(R)$ can be viewed as the group of nontrivial relations satisfied by the elementary matrices E_{ij}^{λ}. Further, we have the exact sequence

$$1 \longrightarrow K_2(R) \xrightarrow{i} St(R) \xrightarrow{\phi} GL(R) \xrightarrow{\nu} K_1(R) \longrightarrow 1.$$

Lemma 10.4.13 $Ker\phi$ *is the center* $Z(St(R))$ *of* $St(R)$.

Proof Let $a \in Z(St(R))$. Then $\phi(a) \in Z(E(R))$. Since $E(R)$ is center less (a matrix commutes with all elementary matrices if and only if it is a scalar matrix), $\phi(a)$ is identity. This means that $a \in Ker\phi$. Thus, $Z(St(R)) \subseteq Ker\phi$. Suppose that $\phi(a) = e$. We need to shows that $a \in Z(St(R))$. Let C_n denote the subgroup of $St(R)$ generated by the elements of the type x_{ij}^{λ}, where $i \neq n \neq j$. Clearly, there is a $n \in \mathbb{N}$ such that $a \in C_n$. Fix a $n \in \mathbb{N}$ such that $a \in C_n$. Let X_n denote the subgroup of $St(R)$ generated by the elements of the types x_{in}^{λ}, where $i \neq n$. Since any two such elements commute (see the Steinberg relations), X_n is abelian. Further, any nonidentity element x of X_n is expressible as

$$x = x_{i_1 n}^{\lambda_1} x_{i_2 n}^{\lambda_2} \cdots x_{i_r n}^{\lambda_r},\ i_1 < i_2 < \cdots < i_r,$$

where $i_k \neq n$. Now, $\phi(x) = E_{i_1 n}^{\lambda_1} E_{i_2 n}^{\lambda_2} \cdots E_{i_r n}^{\lambda_r}$ is the matrix all of whose diagonal entries are 1, i_k^{th} row and n^{th} column entry is λ_k, $k \leq r$ and the rest of the entries are 0. This shows that the representation of an element x as given above is unique, and ϕ is an injective homomorphism when restricted to X_n. Similarly, let Y_n denote the subgroup of $St(R)$ generated by the elements of the type x_{nj}^{λ}, where $j \neq n$. Then Y_n is also abelian, and ϕ restricted to Y_n is injective. Consider x_{ij}^{λ}, where $i \neq n \neq j$. Then for any $x_{kn}^{\mu} \in X_n$, $x_{ij}^{\lambda} x_{kn}^{\mu} x_{ij}^{-\lambda}$ is $x_{in}^{\lambda\mu} x_{kn}^{\mu}$, if $j = k$ and x_{kn}^{μ}, otherwise. This means that $x_{ij}^{\lambda} X_n x_{ij}^{-\lambda} \subseteq X_n$. Similarly, $x_{ij}^{\lambda} Y_n x_{ij}^{-\lambda} \subseteq Y_n$. It follows C_n is contained in the normalizer of X_n as well as in the normalizer of Y_n. Thus, $aX_n a^{-1} \subseteq X_n$, and also $aY_n a^{-1} \subseteq Y_n$. Let $u \in X_n$. Then $\phi(aua^{-1}) = \phi(a)\phi(u)\phi(a)^{-1} = \phi(u)$. Since *phi* is injective when restricted to X_n, it follows that $aua^{-1} = u$ for all $u \in X_n$. Thus, a commutes with each element of X_n. Similarly, a commutes with all elements of Y_n. It follows that a commutes with x_{in}^{λ} and also with x_{nj}^{μ} for all $i \neq n \neq j$. Consider x_{kl}^{λ}, where $k \neq n$ and $l \neq n$. Then $x_{kl}^{\lambda} = [x_{kn}^{\lambda}, x_{n1}^{\mu}]$, and so a commutes with x_{kl}^{λ} also. This means that a commutes with all the generators of $St(R)$. Hence $a \in Z(St(R))$. ♯

In the light of the Proposition 10.4.9, to complete the proof of the Theorem 10.4.11, it is sufficient to establish the following Lemma.

Lemma 10.4.14 *Every central extension by $St(R)$ splits.*

Proof Let

$$1 \longrightarrow H \xrightarrow{i} G \xrightarrow{\xi} St(R) \longrightarrow 1$$

be a central extension by $St(R)$. To show the existence of a splitting, it is sufficient to show the existence of a set $\{s_{ij}^{\lambda} \in G \mid \lambda \in R\}$ of elements of G which satisfy the Steinberg relations and $\xi(s_{ij}^{\lambda}) = x_{ij}^{\lambda}$. Let t be a section of the extension. Then $\xi([t(x_{ik}^{\lambda}), t(x_{kj}^{1})]) = [\xi(t(x_{ik}^{\lambda})), \xi(t(x_{kj}^{1})] = [x_{ik}^{\lambda}, x_{kj}^{1}] = x_{ij}^{\lambda}$, where i, j, k are distinct. Further, if t' is another section of the extension, then $t(x_{ik}^{\lambda}) = u_{ik}^{\lambda} t'(x_{ik}^{\lambda})$ for some $u_{ik}^{\lambda} \in H$. Since $H \subseteq Z(G)$, it follows that $[t(x_{ik}^{\lambda}), t(x_{kj}^{1})] = [t'(x_{ik}^{\lambda}), t(x_{kj}^{1})]$. Thus, $[t(x_{ik}^{\lambda}), t(x_{kj}^{1})]$ is independent of the choice of a section. In fact, using trivial commutator relations, and observing the fact that $[t(x_{ij}^{\lambda}, t(x_{kl}^{\mu})]([t(x_{ip}^{\lambda}, t(x_{ql}^{\mu})])^{-1} \in Z(G)$ for all $i, j \neq k, l, p \neq q$ and also $[t(x_{ij}^{\lambda}, t(x_{jl}^{\mu})]([t(x_{ip}^{\lambda}, t(x_{pl}^{\mu})])^{-1} \in Z(G)$ for all $i, j, p \neq q$, it can be shown that $[t(x_{ik}^{\lambda}), t(x_{kj}^{1})]$ is also independent of $k, i \neq k, j \neq k$. Take $\{s_{ij}^{\lambda} = [t(x_{ik}^{\lambda}), t(x_{kj}^{1})]$. Using the basic commutator relations, it may be further verified that $\{s_{ij}^{\lambda}\}$ respects the Steinberg relations. ♯

Bibliography

1. Artin, M.: Algebra. Pearson Education (2008)
2. Artin, E.: Galois Theory, New edn. Dover Publication (1998)
3. Birkoff, G., MacLane, S.: A Survey of Modern Algebra, 3rd edn. Macmillan, New York (1965)
4. Curtis, R.: Representation Theory of Finite Groups and Associative Algebras. New edn, AMS (2006)
5. Curtis, M.L.: Matrix Groups. Springer (1984)
6. Fulton, H.: Representation Theory. GTM, Springer, Berlin (1999)
7. Halmos, P.R.: Linear Algebra. UTM, Springer, Berlin (1958)
8. Herstein, I.N.: Topics in Algebra, 2nd edn. Wiley, New York (1975)
9. Hoffman, Kunze: Linear Algebra, 2nd edn. Prentice-Hall (1998)
10. Hungerford, T.W.: Algebra, 8th edn. GTM, Springer, Berlin (2003)
11. Jacobson, N.: Basic Algebra I. II. Freeman, San Francisco (1980)
12. Lang, S.: Algebra, 2nd edn. Addison-Wesley, Boston, MA (1965)
13. Morandi, P.: Field and Galois Theory. GTM, Springer, Berlin (1996)
14. Robinson, D.J.S.: A Course in the Theory of Groups, 2nd edn. Springer (1995)
15. Rotman, J.J.: An Introduction to the Theory of Groups, 4th edn. GTM, Springer, Berlin (1999)
16. Saikia, P.: Linear Algebra. Pearson (2009)
17. Serre, J.P.: Linear Representations of Finite Groups. GTM, Springer, Berlin (1996)
18. Suzuki, M.: Group Theory I and II. Springer (1980)

R. Lal, *Algebra 2*, Infosys Science Foundation Series in Mathematical Sciences,
DOI 10.1007/978-981-10-4256-0

Index

R. Lal, *Algebra 2*, Infosys Science Foundation Series in Mathematical Sciences,
DOI 10.1007/978-981-10-4256-0