MODERN ALGEBRA

$$R^* = \{x \in R \mid \exists y \in R, xy = 1$$

$$f(xy) = f(x)f(y$$

$$x^{m+n} = x^m x^n \text{ and } (xy)^n = x^n y$$

Jin-Gen Yang

$$\frac{x}{y} \pm wz = \frac{xz \pm yw}{yz}$$

SCIENCE PRESS

Alpha Science

Modern Algebra

Jin-Gen Yang

SCIENCE PRESS
Beijing

Alpha Science International Ltd.
Oxford, U.K.

Modern Algebra
174 pgs. | 8 figs. | 4 tbls.

Author
Jin-Gen Yang

Co-Published by:

Science Press
16 Donghuangchenggen North Street
Beijing 100717, China

and

Alpha Science International Ltd.
7200 The Quorum, Oxford Business Park North
Garsington Road, Oxford OX4 2JZ, U.K.

www.alphasci.com

ISBN 978-1-84265-760-7 (Alpha Science)

Printed in India

Foreword

For a sufficiently educated person, the word "algebra" often reminds him or her a great deal of high school mathematics, such like factorization of quadratic polynomials, solving an equation or a system of equations, exponential functions, logarithmic functions and so forth. These subjects are known as precalculus algebra or elementary algebra.

This course is quite beyond the precalculus algebra. It emphasizes the inner structures of groups, rings, fields and vector spaces as well as the maps between algebraic structures. For this reason this branch of mathematics is often referred to as "abstract algebra" or "modern algebra". The ideas of algebra evolved through many generations of mathematicians around the turn of 20th century. Among many prominent mathematicians we mention Emmy Noether and Emil Artin, who laid the foundation of modern algebra.

In the freshmen linear algebra, the students have encountered the algebraic structure of vector spaces over real number field or complex number field, as well as the structures of Euclidean spaces or unitary spaces after the introduction of inner product. If we look at the set of integers and the set of real polynomials in one variable closely, we may find their properties strikingly similar. From the algebraic point of view, their structures share many common properties. Roughly speaking, a so-called algebraic structure is a set with one or more operations satisfying certain conditions. This is one of the basic target of study in mathematics. Another important basic structure is topological structure, which is not in the scope of this course. The development of mathematics in the recent century has confirmed that modern algebra is indispensable.

Due to its importance, modern algebra (or abstract algebra) has become a basic standard course for math major students, usually offered for sophomores or juniors. An ideal duration of the course is one whole year. In recent years, only one semester of modern algebra is available for many major universities of China. It is not easy for instructors of this course to cover Galois theory in one semester. But Galois theory is recognized as a milestone of algebra. It will be regretful if a student does not know Galois theory after taking the course of algebra. Through many years of teaching the author has made careful selection of the materials for groups, rings and fields so that the students can reach the main theorem of Galois theory and the proof of the insolvability by radicals of equations of degrees greater than four.

Among all algebraic structures, the group is a basic structure. To shorten the first chapter on groups, I postpone some less fundamental but more technical subjects such like Sylow groups, finitely generated abelian groups and solvable groups to a later chapter. The first three chapters aim at basic knowledge on groups, rings and fields. For ring theory, emphasis is put on the residue ring $\mathbb{Z}/n\mathbb{Z}$ and polynomial rings, of one variable and of several variables. Through these concrete rings students can understand more abstract concepts such like principal ideal domain and unique factorization ring. The non-commutative rings are restricted to basic knowledge and a few standard examples such like matrix rings and quaternions.

Chapter 4 is a brief account of linear algebra over an arbitrary field. Since the students are assumed to have taken the first course of linear algebra already, the account is often sketchy.

The main purpose of this chapter is to emphasize the difference of vector spaces over fields of different characteristics and prepare for the theory of extensions of fields. Students should establish a point of view to regard the extension of a field as a vector space over the base field.

Chapter 5 covers Sylow groups, finitely generate abelian groups and solvable groups as I mentioned before. The last three chapters are devoted to the field theory exclusively. By author's experience, it is reasonable to finish all eight chapters in 15 or 16 weeks.

The appendices are for more enthusiastic students. A proof of quadratic reciprocity law displays elegant techniques of finite fields. The proof of the theorem concerning the structure of finite skew fields displays a clever use of group actions. The proofs of these two theorems are adapted from the ones of J.-P. Serre and A.Weil respectively. The author wishes that the interested readers can feel the beauty of a mathematical proof by reading these two appendices.

Due to the limit of time, in the last section only the insolvability of equations of degree greater than four is proved. Using Galois theory to deduce the formulas for the solutions of cubic and quartic equations are put in appendices.

The sections marked with an asterisk can be skipped on the first reading.

There are adequate amount of exercises throughout the book. The degree of difficulty varies. Some exercises are chosen from Ph.D qualifying exams. Hints or solutions are provided as an appendix. Most problems have many different proofs, which are impossible to be included. Students are strongly encouraged to find their own solutions and do not rely on the appendix too heavily. Only by laying one's own hand on hard problems, one can feel the charm of mathematical deduction.

In summary, the author has tried to give a concise and easy to understand account of basic knowledge and methods in algebra without loss of rigor. The main perspective readers are sophomores or juniors of math major.

Professor Li Kezheng offered many valuable comments and pointed out many mistakes. Many students of Fudan University who have taken this course in the past several years have provided useful comments too. I regret that I am not able to list their names. I thank all those people who have made contributions to this book. The assistance of Ms. Yao Lili of Science Press in the publication of this book is greatly appreciated.

J.-G. Yang,

Preliminaries and Notations

Since this course is intended for the undergraduate students in math major, we assume that the readers have already studied linear algebra and calculus (at least calculus in one variable). The readers are assumed to have the basic knowledge on sets and maps.

Some conventions and notations are given below.

A map f from one set S to another set T is an injection if $f(x) \neq f(y)$ wherever x and y are two distinct elements in S. The map f is a surjection if there exists $x \in S$ such that $f(x) = z$ for every $z \in T$. A map is a bijection if it is injective and surjective. Let U be a subset of S. Denote by $f|_U$ the restriction of f on U.

A map f from S to T carrying x to z is denoted by $f : S \to T, x \mapsto z$. For example $x \mapsto e^x$ is the exponential function. Sometimes the identity map from a set S into itself is denoted by id or 1.

Let f be a map from S to T and g be a map from T to U. The map $g \circ f : S \to U, x \mapsto g(f(x))$ is the composite of f and g. It can also be denoted by gf.

A subset of a set S is often defined by the notation $\{x \in S | P\}$, where P is the condition that x should satisfy. For example $\{x \in \mathbb{R} | 0 \leqslant x \leqslant 1\}$ is the closed interval $[0, 1]$. The union and the intersection of sets are denoted by $\cup$ and $\cap$ respectively. The difference $\{x \in A | x \notin B\}$ of the sets A and B is denoted by $A - B$ or $A \backslash B$, and we prefer to use the latter. A set with a few elements is usually denote by $\{\cdots\}$, in which $\cdots$ is the list of all elements. For example $\{a\}$ is a set consisting of a single element a, $\{0, 1\}$ is the set consisting of the elements 0 and 1. The empty set is denoted by $\emptyset$.

The knowledge of equivalence relations and equivalence classes is helpful, but is not indispensable. We will explain them when they are first used in the text.

Readers are required to know the number fields well. By definition, a number field is a subset of the complex number field closed under addition, subtraction, multiplication and division. The most frequently used number fields are complex number field, real number field and rational number field.

The equality $x = \pm a$ means that x is equal to either a or $-a$.

Some common notations are listed as follows:

$\mathbb{N}$ — set of natural numbers
$\mathbb{Z}$ — set of integers
$\mathbb{Q}$ — set of rational numbers
$\mathbb{R}$ — set of real numbers
$\mathbb{C}$ — set of complex numbers

Contents

Chapter 1
Elements of Groups

Many objects we encounter in mathematics are sets equipped with one or more operations. Such objects are usually called **algebraic structures**. In this course the algebraic structures we will study include groups, rings and vector spaces. In terms of the number of operations the group theory can be considered to be a natural starting point for this course, since a group involves only one basic operation. In this chapter the definition, basic properties of groups and homomorphisms between groups will be explained. More advanced topics on groups will be covered in later chapters.

1.1 Definitions and Examples

Let S be a set. Denote the Cartesian product of S with itself by $S \times S$. It consists of all pairs (a, b) with $a, b \in S$. Note that (a, b) and (b, a) are different elements in $S \times S$ if $a \neq b$. A map f from the set $S \times S$ to S is called a (binary) operation.

For example, let S be the set of all real numbers. Defined the map $f : S \times S \to S$ by $f(a, b) = a + b$. Then f is a binary operation. We may use all sorts of ways to define operations such as $f(a, b) = a^2 + b^3$, $f(a, b) = ab, \ldots$.

Since it is awkward to use the notation $f(a, b)$ for an operation, there are more convenient notations such as $a + b$, $[A, B]$, $\mathbf{u} \times \mathbf{v}, \ldots$, depending on occasions. The simplest notation is ab. We will use this notation when no confusion will be caused. In practice, it is better to use the notations that we are already familiar with. For instance, the addition of numbers is denoted by "+", the cross product of two vectors in $\mathbb{R}^3$ is denoted by "$\times$".

Let's look at some more examples.

(1) Let S be a 3-dimensional Euclidean space. The addition $\mathbf{u} + \mathbf{v}$ and cross product $\mathbf{u} \times \mathbf{v}$ are binary operations. The dot product $\mathbf{u} \cdot \mathbf{v}$ is not a binary operation since the result of the dot product is not a vector.

(2) Let S be the set of all $n \times n$ matrices. Then $A + B$, AB, $AB - BA$ are three different operations on S (the last one is a basic operation in Lie Algebra).

(3) Let S be the set of all everywhere well-defined real functions. Then the composite $f \circ g$ is a binary operation on S.

Unitary operations (involving one variable) appear quite often too, such as $-a$ (the negative of a number), a^2 (the square of a number), $\bar{a}$ (the conjugate of a complex number), A^{T} (the transpose of a square matrix), etc..

We say that a binary operation f on a set S satisfies the **law of associativity** if

$$f(a, f(b, c)) = f(f(a, b), c)$$

holds for any $a, b, c \in S$. By our convention this condition can be written simply as

$$a(bc) = (ab)c,$$

which is exactly the form we are familiar with.

In the previous examples many binary operations satisfy the law of associativity. The cross product of vectors and the operation $AB - BA$ for square matrices A, B do not satisfy the law of associativity. It is easy to verify that the operation $f(a, b) = a^2 + b^3$ on real numbers does not satisfy the law of associativity.

An element e in a set S equipped with a binary operation is called an **identity element** if $ea = ae = a$ for any $a \in S$.

Examples:

(1) 0 is an identity element in the set of real numbers under addition.

(2) 1 is an identity element in the set of real numbers under multiplication.

(3) The identity matrix I_n is an identity element in the set of $n \times n$ matrices under multiplication.

Proposition 1.1.1 *A set S with a binary operation has at most one identity element.*

Proof Let e and e' be two identity elements of S. Since e is an identity element, $ee' = e'$ holds. On the other hand, $ee' = e$ holds since e' is also an identity element. Therefore $e = ee' = e'$. □

Definition 1.1.1 A nonempty set S with a binary operation is called a **semigroup** if the law of associativity is satisfied. A semigroup containing an identity element is called a **monoid**.

According to Proposition 1.1.1 there is a unique identity element in a monoid.

Example 1.1.1 The set of natural numbers under addition is a semigroup but not a monoid.

Definition 1.1.2 Let a be an element in a monoid S and let e be the identity element of S. If there is an element $b \in S$ such that $ba = ab = e$, then a is called an **invertible element** and b is called the **inverse** of a.

Note that the article in front of the word "inverse" is "the" instead of "an". This will be justified by the following proposition.

Proposition 1.1.2 *There is a unique inverse for an invertible element in a monoid.*

Proof Let b, b' be inverses of a. Then $b = be = b(ab') = (ba)b' = eb' = b'$. □

It is obvious that the relation "inverse" is symmetric. That is to say, if a is invertible with b as its inverse, then b is also invertible and a is the inverse of b.

The inverse of an invertible element a is commonly denoted by a^{-1}.

Definition 1.1.3 A nonempty set G with a binary operation is defined to be a **group** if the following three conditions are satisfied:

(1) The law of associativity is satisfied;

(2) The identity element exists;

(3) Every element in G is invertible.

The first two conditions in the definition mean that every group is a monoid. Although the concepts of semigroup and monoid are more general than group, but we are primarily interested in groups.

The unitary operation $a \mapsto a^{-1}$ in a group is called the inverse operation. Since it is totally determined by the binary operation of the group, we do not consider the inverse as a basic operation of the group.

As we have mentioned before, the binary operation in a group uses different notations depending on occasions. The default notation is ab, or $a \cdot b$ occasionally. For this reason, the operation can also be referred to as "multiplication" and ab is called the product of a and b. In this case the inverse of a is denoted by a^{-1} and the identity element e can also be denoted by 1, or 1_G to avoid confusion. For a natural number n, the n-th power a^n is defined to be the product of a with itself n times. The power a^{-n} with negative integral exponent is defined to be $(a^{-1})^n$.

Let G be a group. If $ab = ba$ for any $a, b \in G$, then G is called a commutative group, or **abelian group**.

For many abelian groups, the binary operation is usually called "addition" and denoted by $a + b$. In this case the identity element is often called the zero (element) of the group and denoted by 0. The inverse of an element a is denoted by $-a$ and the n-th power of an element a becomes $na = a + \cdots + a$.

A groups consisting of a finite number of elements is called a **finite group**, otherwise it is called an **infinite group**. The number of elements in a group G is denoted by $|G|$, called the **order** of G. Hence $|G|$ is ∞ if G is an infinite group and $|G|$ is a natural number if G is a finite group.

Let g be an element of a group G. If there is no natural number n such that $g^n = 1$, then g is called an element of infinite order, otherwise, define the **order** of g to be the smallest natural number n such that $g^n = 1$, denoted by $o(g)$. Hence an element in G has order one if and only if it is the identity element. By convention, $o(g) = \infty$ if g is an element of infinite order.

Example 1.1.2 (1) Every number field under addition forms an abelian group, called the additive group of that number field. One thing to keep in mind is that the multiplication operation of that number field is neglected when it is regarded as a group, since by definition a group involves only one basic operation.

(2) The set $\mathbb{Z}$ of all integers under addition is an abelian group, called the additive group of integers.

(3) All $n \times n$ invertible matrices over a number field K under the (matrix) multiplication is a group called **general linear group** and denoted by $GL_n(K)$. It is not an abelian group when $n > 1$.

(4) A number field K under multiplication is a monoid, but not a group since the element 0 is not invertible. Denote $K^* = K \backslash \{0\}$. Then K^* is an abelian group under multiplication, which is essentially the same as $GL_1(K)$. Although this group is abelian, but it would be absurd to use "+" to denote the multiplication.

(5) Any vector space is an abelian group under addition.

(6) 3-dimensional Euclidean space $\mathbb{R}^3$ is not a semigroup under cross product, since the law of associativity is not valid.

(7) The smallest group has only one element. Such a group is called a trivial group.
(8) Empty set is not a group.

Proposition 1.1.3 (law of cancelation) *If the element a, b, c in a group satisfy $ab = ac$ or $ba = ca$, then $b = c$.*

Proof Assume that $ab = ac$. By left multiplying a^{-1} to both sides of $ab = ac$ the equality $a^{-1}(ab) = a^{-1}(ac)$ is obtained, so $(a^{-1}a)b = (a^{-1}a)c$ by the law of associativity. Hence $eb = ec$ by the definition of inverse. Finally $b = c$ follows from the definition of the identity element e.
The same argument shows that $ba = ca$ implies $b = c$. □

Corollary 1.1.1 *Let a, b be two elements of a group G.*
(1) If $ab = a$ or $ba = a$ then $b = 1_G$;
(2) If $ab = 1_G$ or $ba = 1_G$ then $b = a^{-1}$.

Proof (1) Apply the law of cancelation to $ab = a1_G$.
(2) Apply the law of cancelation to $ab = aa^{-1}$. □

Proposition 1.1.4 *Let a, b be two elements of a group G. Then $(ab)^{-1} = b^{-1}a^{-1}$.*

Proof The equality $(b^{-1}a^{-1})(ab) = b^{-1}(a^{-1}a)b = b^{-1}b = 1$ and Corollary 1.1.1 (2) imply $b^{-1}a^{-1} = (ab)^{-1}$. □

This property is the generalization of the well-known formula $(AB)^{-1} = B^{-1}A^{-1}$ in linear algebra, where A and B are two invertible $n \times n$ matrices.

As we have seen, any number field is a group under addition. For this reason we may consider addition as the first important operation of this number field. In fact, this is the first operation taught in elementary school. The subtraction is not regarded as a basic operation. Since $a - b = a + (-b)$, the subtraction can be considered to be the composite of the inverse and the addition.

Exercises 1.1

1. In the set $G = \{a \in \mathbb{R} | a > 0, a \neq 1\}$ define a binary operation $a * b = a^{\ln b}$. Is G a group under this operation? Here $\ln b$ is the natural logarithm of b.
2. Let A be the set of all strictly increasing continuous functions on $[0, 1]$ satisfying $f(0) = 0, f(1) = 1$. For any $f, g \in A$ define fg to be the composite of f and g, i.e., $(fg)(x) = f[g(x)]$ for any $x \in [0, 1]$. Prove that A is a group under this operation.
3. Let G be a semigroup satisfying the following two conditions:
(1) there is some $e \in G$ such that $ea = a$ for any $a \in G$;
(2) for every $a \in G$, there is some $b \in G$ such that $ba = e$.
Prove that G is a group.
4. Prove that every element of a finite group has finite order.
5. Let G be a group and $a, b \in G$. Prove that $o(ab) = o(ba)$.
6. Assume that every element a of a group G satisfies $a^{-1} = a$. Prove that G is an abelian group.

1.2 Subgroups

Definition 1.2.1 Let H be a nonempty subset of a group G satisfying the following two conditions:

(1) (closure under the multiplication) $ab \in H$ for any $a, b \in H$;
(2) (closure under the inverse operation) $a^{-1} \in H$ for any $a \in H$.
Then H is called a **subgroup** of G.

By the closure of multiplication, a subgroup H of G inherits the binary operation from G. Naturally this inherited operation preserves the law of associativity. Since H is nonempty, there is some element $a \in H$. Hence $a^{-1} \in H$ by the closure of inverse operation, which implies that $1 = a^{-1}a \in H$. So H is a group by itself. The readers may compare the concept of subgroup with that of subspace in linear algebra.

It is easy to see that any subgroup of an abelian group is abelian.

Example 1.2.1 (1) Every group G has two subgroups for free. They are $\{1\}$ and G. (They become the same if $|G| = 1$.) They are called trivial subgroups of G. A subgroup H of G is called a proper subgroup if $H \neq G$.

(2) Let n be a natural number. Let $n\mathbb{Z}$ denote the set of all integers divisible by n. Then $n\mathbb{Z}$ is a subgroup of $\mathbb{Z}$.

(3) Let $SL_n(K)$ be the set of all $n \times n$ matrices with determinant one over a number field K. It is a subgroup of $GL_n(K)$, called the **special linear group**.

(4) The set of matrices consisting of

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

is a subgroup of $SL_2(K)$ with $|H| = 2$.

(5) The real number field $\mathbb{R}$ is a group under addition and $\mathbb{R}^* = \mathbb{R}\backslash\{0\}$ is a group under multiplication. Although the latter is a subset of the former, but $\mathbb{R}^* = \mathbb{R} - \{0\}$ is not a subgroup of $\mathbb{R}$ because the binary operations of these two groups are different.

Proposition 1.2.1 (a criterion for subgroup) *Let H be a nonempty subset of a group G. If $ab^{-1} \in H$ for any $a, b \in H$, then H is a subgroup of G.*

Proof Let c be an arbitrary element in H. Then $1 = cc^{-1} \in H$ by the hypothesis of the proposition. For any $a \in H$, we have $a^{-1} = 1a^{-1} \in H$. Hence H is closed under the inverse operation.

Since $ab = a(b^{-1})^{-1} \in H$, H is also closed under multiplication. □

Proposition 1.2.2 *Let $\{H_i\}_{i \in I}$ be a set (not necessarily finite) of subgroups of a group G. Then $H = \bigcap_{i \in I} H_i$ is a subgroup of G.*

Proof Since $1 \in \bigcap_{i \in I} H_i$ the set H is not empty. Let $a, b \in H$. Then $ab^{-1} \in H_i$ for every $i \in I$, since every H_i is a subgroup of G. Hence $ab^{-1} \in H$. It follows from Proposition 1.2.1 that H is a subgroup of G. □

Definition 1.2.2 Let g be an element of a group G. The set $C(g) = \{a \in G | ag = ga\}$ is called the **centralizer** of g in G. For any nonempty subset S of G, the set $C(S) = \{a \in G | ag = ga \text{ for all } g \in S\}$ is called the centralizer of S in G. In particular, $C(G)$ is called the **center** of G.

It is easy to see that $C(g)$ and $C(S) = \bigcap_{g \in S} C(g)$ are subgroups of G and $C(G)$ is an abelian group. The group G is abelian if and only if $G = C(G)$.

Let S be a nonempty subset of a group G. Let $\langle S \rangle$ denote the intersection of all subgroups of G containing S. Then $\langle S \rangle$ is the smallest subgroup of G containing S in the sense that $\langle S \rangle \subseteq H$ for every subgroup H of G with $S \subseteq H$. The subgroup $\langle S \rangle$ is called the subgroup generated by S. If S is a finite set consisting of the element $a_1, \ldots, a_n$, then $\langle S \rangle$ can also be denoted by $\langle a_1, \ldots, a_n \rangle$. If $G = \langle S \rangle$, then we say that G is generated by S.

Example 1.2.2 (1) $\mathbb{Z} = \langle 1 \rangle = \langle -1 \rangle$.
(2) $GL_n(K)$ is generated by all $n \times n$ elementary matrices.

A group generated by a finite set is called a **finitely generated group**. In particular, a group generated by a single element is called a **cyclic group**. For instance, $\mathbb{Z}$ is a cyclic group while $GL_n(K)$ is not. It is obvious that every cyclic group is abelian.

Let a be an element of a group G. Then $\langle a \rangle$ is a subgroup of G, called a cyclic subgroup of G. It is easy to verify that $o(a) = |\langle a \rangle|$. In fact, if $o(a) = \infty$ then

$$\langle a \rangle = \{\ldots, a^{-2}, a^{-1}, 1, a, a^2, \ldots\}$$

and if $o(a) = n < \infty$ then

$$\langle a \rangle = \{1, a, a^2, \ldots, a^{n-1}\}.$$

Example 1.2.3 In $GL_n(K)$ the matrix

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

generates an infinite cyclic subgroup while the matrix

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

generates a cyclic subgroup of order 3.

Proposition 1.2.3 *Let S be a nonempty subset of a group G. Then*

$$\langle S \rangle = \{a_1^{e_1} a_2^{e_2} \cdots a_n^{e_n} | a_1, \ldots, a_n \in S, e_1, \ldots, e_n = \pm 1,$$

$$n \textit{ is an arbitrary nonnegative integer}\}.$$

Proof First note that the elements $a_1, a_2, \ldots, a_n$ involved in the expression are not necessarily distinct.

Denote the set of the right hand side by T. Then T is a subgroup containing S. Hence $\langle S \rangle \subseteq T$.

Let H be a subgroup of G such that $S \subseteq H$. By the definition of subgroup every element that can be expressed as $a_1^{e_1} a_2^{e_2} \cdots a_n^{e_n} (a_i \in S, e_j = \pm 1)$ is in H. Hence $T \subseteq H$. Therefore $T = \langle S \rangle$. □

At this point we want to determine all subgroups of $\mathbb{Z}$. First of all, as we have already seen before, $0, \mathbb{Z}, 2\mathbb{Z}, 3\mathbb{Z}, \ldots$ are all subgroups of $\mathbb{Z}$. We show that there are no other subgroups.

Let H be a nontrivial subgroup of $\mathbb{Z}$. Then there is some nonzero integer a in H. Since $-a$ is also in H by the definition of subgroup, there exists a natural number in H. Let n be the smallest natural number in H. Then $n\mathbb{Z} \subseteq H$. For any $m \in H$, there are integers q, r such that $m = qn + r$ in which $0 \leqslant r \leqslant n - 1$. Since $r = m - qn \in H$, r cannot be a natural number by the minimality of n. Hence $r = 0$. This implies that $m \in n\mathbb{Z}$. It follows that $H \subseteq n\mathbb{Z}$. Therefore every subgroup of $\mathbb{Z}$ is an infinite cyclic group.

Exercises 1.2

1. Let G be the set of all 3×3 real upper triangular matrices with all diagonal elements equal to 1. Show that G is a group under multiplication and determine the center of G.

2. Let X and Y be two subsets of a group G. Prove that
(1) if $X \subseteq Y$ then $C(X) \supseteq C(Y)$;
(2) $X \subseteq C(C(X))$;
(3) $C(X) = C(C(C(X)))$.
Here $C(X)$ denotes the centralizer of X.

3. Let H be a subgroup of a group G such that H is contained in every nontrivial subgroup of G. Prove that H is contained in the center of G.

4. An element a of a group G is called a perfect square if there exists $b \in G$ such that $a = b^2$. Assume that G is a cyclic group and $a, b \in G$ are not perfect squares. Show that ab is a perfect square. Give an example to show that this statement is not true for non-cyclic groups.

5. Let H be a nonempty subset of a finite group G. Show that H is a subgroup of G if $ab \in H$ for any $a, b \in H$.

6. Let A be an $n \times n$ real invertible matrix and let G be the set consisting of all $n \times n$ real matrices P such that $P^{\mathrm{T}}AP = A$. Show that G is a subgroup of $GL_n(\mathbb{R})$. Here P^{T} is the transpose of P.

1.3 Permutation Groups

Permutations and combinations have already been studied in high school mathematics. For instance, $2, 1, 4, 5, 3$ is a permutation (or rearrangement) of the sequence $1, 2, 3, 4, 5$. It is known that the number of permutations of $1, 2, 3, 4, 5$ is equal to $5! = 120$.

Let's look at the permutations from another point of view. The rearrangement $2, 1, 4, 5, 3$ is treated as a bijection σ of the set $\{1, 2, 3, 4, 5\}$ into itself, which carries 1 to 2, 2 to 1, 3 to 4, 4 to 5 and 5 to 3. We can use the following table to specify this bijection

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 5 & 3 \end{bmatrix}.$$

This interpretation suggests the following definition.

Definition 1.3.1 Let n be a natural number. A bijection from the set $\{1, 2, 3, \ldots, n\}$ to itself is a **permutation** of n objects.

As in the example, a permutation can be represented by a table

$$\begin{bmatrix} 1 & 2 & 3 & \cdots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \cdots & \sigma(n) \end{bmatrix}.$$

Here the numbers $1, 2, 3, \ldots, n$ do not have numerical meaning whatsoever. They are merely convenient labels for distinct objects. Denote the set of all permutations of n objects by S_n. Then S_n contains $n!$ (the factorial of n) elements.

Introduce a binary operation in S_n in the following way. For any $\sigma, \tau \in S_n$ define $\sigma\tau$ to be the composite of σ and τ, the map τ followed by σ. Thus $\sigma\tau(i) = \sigma[\tau(i)]$ for every $i \in \{1, 2, \ldots, n\}$. Evidently this is a well-defined binary operation of the set S_n.

Proposition 1.3.1 *The set S_n is a group under the binary operation of composite.*

Proof By the rule of composite of maps $\sigma(\tau\pi) = (\sigma\tau)\pi$ holds for any $\sigma, \tau, \pi \in S_n$. Hence the law of associativity is valid.

Let id be the identity map of $\{1, 2, 3, \ldots, n\}$, i.e., $\mathrm{id}(i) = i$ for every $i \in \{1, 2, 3, \ldots, n\}$. Then $\mathrm{id} \circ \sigma = \sigma \circ \mathrm{id} = \sigma$ for any $\sigma \in S_n$. Hence id is the identity element.

Since $\sigma \in S_n$ is a bijection of $\{1, 2, \ldots, n\}$, its inverse map τ exists. It means that $\sigma\tau = \tau\sigma = \mathrm{id}$. □

The product of two permutations can be read off from their tables. For example,

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 5 & 4 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 5 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 4 & 2 & 5 \end{bmatrix}.$$

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 5 & 4 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 5 & 1 & 4 & 3 \end{bmatrix}.$$

Definition 1.3.2 A subgroup of S_n is called a **permutation group**. S_n is called the **symmetric group** of n objects.

There are two reasons for introducing permutation groups at this point. First of all, they form a large class of very important finite groups. Secondly, many non-abelian groups can be found in permutation groups.

Let's look at some symmetric groups of low order. When $n \leqslant 2$, they are too simple to worth studying. The first nontrivial symmetric group is S_3. Its six elements can be enumerated as

$$\sigma_0 = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}, \quad \sigma_1 = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix},$$

$$\sigma_3 = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}, \quad \sigma_4 = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix}, \quad \sigma_5 = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}.$$

It is easy to verify that $\sigma_1\sigma_2 = \sigma_4$ and $\sigma_2\sigma_1 = \sigma_3$. Thus S_3 is not abelian.

It is not hard to verify that S_3 has following four nontrivial subgroups $\{\sigma_0, \sigma_1\}$, $\{\sigma_0, \sigma_3\}$, $\{\sigma_0, \sigma_4\}$, $\{\sigma_0, \sigma_2, \sigma_5\}$.

The relations between the subgroups of S_3 can be described by Figure 1.1.

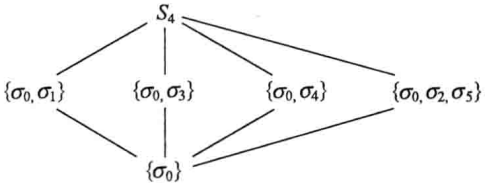

Figure 1.1

If two subgroups are directly connected by a straight line, the subgroup in the lower position is contained in the one in the upper position. The diagram makes it easy to see the relations of subgroups at a glance.

Serious readers may try to write all elements of S_4 and as many as possible subgroups and draw a diagram of subgroups.

You may probably notice that this two line notation for permutations is not economic. The first line is really not necessary. To be worse, in S_9 a permutation of "swapping 1 and 2" is denoted by

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 1 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{bmatrix},$$

which is hardly bearable. This suggests the following concepts.

Definition 1.3.3 Let $i_1, i_2, \ldots, i_d$ be d distinct objects in $\{1, 2, \ldots, n\}$. Let σ be an element in S_n such that

$$\sigma(i_1) = i_2, \quad \sigma(i_2) = i_3, \quad \ldots, \quad \sigma(i_d) = i_1$$

and $\sigma(i) = i$ for all $i \notin \{i_1, i_2, \ldots, i_d\}$. Then σ is called a d-**cycle**, denoted by $(i_1 i_2 \cdots i_d)$. The notation of a cycle is not unique. For instance, $(i_2 i_3 \cdots i_d i_1)$ and $(i_1 i_2 \cdots i_d)$ are the same cycle. Two cycles are called disjoint if every object in the first cycle does not appear in the second one. For instance, (142) and (36) are disjoint while (261) and (3245) are not.

A 2-cycle is called a **transposition.**

It is not hard to see that every permutation can be expressed as the product of mutually disjoint cycles. For example,

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 5 & 6 & 4 & 2 & 1 \end{bmatrix} = (136)(52).$$

Although the cycle notation is not as straightforward as the table notation, we will soon see that in some occasions it is indispensable. One thing we should mention is: $(135)(52)$ can denote an element in S_5 as well as an element in any S_n with $n > 5$. This ambiguity is tolerable, since the value of n is usually clear from the context.

The multiplication of two permutations under the cycle notation can be accomplished by changing them into the table notion first. This is not necessary. One may use his (or her) favorite method to write out the result directly from the cycle notation. For example, let $\sigma = (152)(34), \tau = (35)(41)$. Then $\sigma\tau = (152)(34)(35)(41) = (132)(45)$.

Proposition 1.3.2 *Every permutation can be written as the product of transpositions, not necessarily disjoint.*

Proof It suffices to show that every cycle can be written as the product of some transpositions.

It is easy to verify that $(i_1 i_2 \cdots i_d) = (i_1 i_2)(i_2 i_3) \cdots (i_{d-2} i_{d-1})(i_{d-1} i_d)$. □

Let $\sigma \in S_n$. Let r be the number of elements in the set

$$\{(i, j) | 1 \leqslant i < j \leqslant n, \sigma(i) > \sigma(j)\}.$$

In other words, r is the number of pairs out of order in the sequence $\sigma(1), \sigma(2), \cdots, \sigma(n)$.