A.I. Kostrikin I.R. Shafarevich (Eds.)

Algebra IV

Infinite Groups. Linear Groups

With 9 Figures

Contents

I. Infinite Groups

A.Yu. Ol'shanskij, A.L. Shmel'kin

Translated from the Russian
by J. Wiegold

Contents

Introduction

The achievements of Galois theory stimulated intensive study of permutation groups, and indeed in the early stages of its development, group theory was preoccupied almost exclusively with finite groups. However, under the influence of geometry, topology, and the theory of differential equations, there arose a pressing need to consider infinite groups of transformations. For example, Klein proposed that the classification of geometries should be linked with the description of the corresponding transformation groups. Parametric groups made their appearance in the works of Lie. Poincaré established the first contacts between combinatorial topology and group theory. Fedorov discovered a remarkable application of groups to the geometry of crystals.

There is an idea that has been developing over many years, and is by now completely obvious. It is this: whenever symmetry of a mathematical or physical object has a significant role to play – be the object of an algebraic nature, a differential equation, a crystallographic lattice or a geometrical invariant – it must have associated with it a group G of transformations preserving it (things like rigid motions, changes of variables, permutations of indices and so on). Groups thus emerged as measures of symmetry, and became indispensible tools for classifying the symmetries of a system.

The superposition of transformations in G is also in G (that is, it is a symmetry), and the term "group" came to be used for any set G with a law of composition defined on it: $(x, y) \mapsto xy \in G$, such that: 1) $(xy)z = x(yz)$, that is, the "product" is associative; 2) there is an element e in G such that $xe = ex = x$ for every x in G (e is called the identity element of G; for transformation groups, e is the identity mapping); 3) for every x in G, there exists an element x^{-1} in G (called the inverse of x) such that $xx^{-1} = x^{-1}x = e$. The study of group multiplication without restriction on the nature of the elements of G, and without the requirement that G be finite, was promoted in Shmidt's book entitled "Abstract Group Theory" and published in 1916. Groups are among the very first examples of abstract systems; at the beginning of this century, many other mathematical disciplines were re-examined on the group-theoretical model.

The breadth of the group axioms links groups with widely different branches of mathematics and natural science, and it causes new directions in the general theory to be closely ramified with the specifics inherent in them. The theory of topological groups, and other branches of mathematics where structures additional to the group structure are important, are outside the scope of our survey: their place is in specialist books on the relevant themes. Having once noted the various interconnections and motivations, we shall, in the main, be considering only properties of an abstract algebraical character, properties that do not depend on any analytical structure nor on the way transformation groups are represented.

Independently of the concrete nature of the elements, groups often appear together with natural generators and relations among them. The study of groups given in this way is the subject of combinatorial group theory. Although the combinatorial and structural theories cannot be divorced from each other (the object of investigations is the same, even if the methods differ), we have separated this survey into two chapters, for the sake of convenience.

With very few exceptions, no special preparation is required of the reader, if such common mathematical terms as "normal subgroup", "factor-group", "simple group", "Cartesian product of groups" are discounted. Some of the conventional notation we use is as follows: S_n is the group of all permutations on $\{1, 2, \ldots, n\}$, $GL_n(k)$ is the multiplicative group of all $n \times n$ non-singular matrices with coefficients in k (here $k = \mathbb{Z}$, the ring of integers, or a field), $\mathbb{Z}_n$ is the group (or ring) of residues modulo n: $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$. A list of terms is to be found in the subject index.

The authors have concentrated their attention on the principal examples of groups, and these have been selected because of their connection with adjacent disciplines such as topology, linear and homological algebra, and other sciences. However, it should be noted that the axiomatic reconstruction of the foundations has enabled group theory to disover its independence, and has called to life a large collection of self-standing research programmes. For example, the study of groups with various "finiteness" conditions has given rise to the formulation of problems that are typical for the investigation of infinite groups, as well as to the creation of new methods. The emergence of the strong internal springs of the theory has inevitably meant that some parts of it are no longer in intimate contact with their origins, namely the most important examples of groups and types of groups.

At the present time, it is impossible to descry sufficient grounds for enunciating an aim such as that of describing up to isomorphism all (or even of wide classes of) groups, and indeed it is possible that the coming years will see the relation between the abstract and the concrete in group theory changing once again to the advantage of the latter. Mathematical logic, topological manifolds, the study of automorphisms of metric spaces and infinite graphs, algebraic geometry, and perhaps contemporary physics, all lead to the formulation of new problems in group theory. The concept of group, just like that of transformation or symmetry, has established itself firmly as one of the most fundamental of all mathematical ideas.

The aim of the present survey is to acquaint mathematicians in contiguous disciplines with the most important achievements of group theory and the methods employed there. Results occur either in the most up-to-date version, or in a form that seems most natural to the authors. In a survey like this, it is not possible to trace the historical development of this or that idea, nor of group theory in general. Individual remarks simply indicate the *dramatis personae*, not their contribution in group theory.

Chapter 1
Combinatorial Group Theory

By this we understand the large section of group theory that studies groups given in terms of generators and relations. In fact, that is how fundamental groups of topological spaces are usually described, and it was this fact that provided such a powerful impetus for the investigation of infinite groups. The origin of the combinatorial approach appeared in the works of Poincaré, Dehn, and Nielsen. The fruitfulness of the method is explained firstly by its universal nature – every group can be defined by generators and relations. Secondly, the combinatorial analysis of words and relations among them, as well as the "free constructions" typical of this theory, are often associated with natural geometrical representations and constructions.

In recent decades, combinatorial group theory has been influenced very strongly by mathematical logic: algorithm theory and model theory. The first of these influences to appear was the theorem of Novikov asserting the existence of finitely presented groups with algorithmically insoluble word problem.

§1. Free Groups

The involutions $a = (1, 2)$ and $b = (1, 3)$ in the symmetric group S_3 satisfy the relations $a^2 = 1$, $b^2 = 1$, $(ab)^3 = 1$, where 1 is the identity permutation. These relations are, of course, interpreted as being different, but in order to distinguish formally between their left-hand sides, it is necessary to study not only the (finite) group S_3, but also the infinite group consisting of words in the alphabet a, b, a^{-1}, b^{-1}.

1.1. Definition of Free Group. Let X be some finite or infinite alphabet. By a (group) *word* in X we understand an expression w of the form

$$w \equiv x_{i_1}^{\varepsilon_1} x_{i_2}^{\varepsilon_2} \dots x_{i_n}^{\varepsilon_n}, \tag{1}$$

where the x_{i_k} are in X, $\varepsilon_k = \pm 1$ for $k = 1, 2, \dots, n$, and $\equiv$ means letter-for-letter equality. The number n is called the *length* of w: $|w| = n$. The *empty* word has zero length, and it is denoted by 1. A word is said to be *uncancellable* if it contains no subwords of the form $x_j x_j^{-1}$ or $x_j^{-1} x_j$. The deletion of such a pair, that is, the replacement of such words of length 2 by the empty subword, is termed a cancellation. It is not difficult to check that every word can be reduced after finitely many cancellations to *uncancellable* form independent of the process of cancellation.

The *free group* $F(X)$ consists of all uncancellable words in the alphabet X, the product of words v and w being the result of bringing vw to uncancellable form. It is easy to see that this multiplication does in fact satisfy the group axioms, that the identity of $F(X)$ is the empty word, and the inverse of (1) is $w^{-1} \equiv x_{i_n}^{-\varepsilon_n} \ldots x_{i_1}^{-\varepsilon_1}$. An equivalent definition of $F(X)$ gives its elements as classes of words having identical uncancellable form. With it, one does not need to worry about the constant reduction to uncancellable form of words appearing in any intermediate calculations in $F(X)$. The cardinality of $F(X)$ is called the *rank* of $F(X)$.

Free groups are torsion-free: every non-identity element w in $F(X)$ has infinite order, that is, $w^m \neq 1$ for every natural number m. Moreover, the equality $v^m = w^m$ in $F(X)$ implies that $v = w$ (that is, free groups are R-groups, groups with unique extraction of roots). The first property is obtained at once on representing w in the form $w = aba^{-1}$, where b is cyclically reduced, that is, its first and last letters are not mutually inverse. Then, w^m has uncancellable form $ab^m a^{-1}$, and this means that $w^m \neq 1$. The second property also follows from the definition and some simple combinatorial arguments. Equally easily, it can be shown that commuting elements v and w of $F(X)$ (that is, such that $vw = wv$ in $F(X)$) must be powers of one and the same element, in other words, must lie in one and the same cyclic subgroup.

Free groups have the following universal property. If G is any group and α any map $X \to G$, there exists a unique homomorphism $\bar{\alpha}$: $F(X) \to G$ extending α to $F(X)$ (if $w = x_{i_1}^{\varepsilon_1} \ldots x_{i_n}^{\varepsilon_n}$, one simply writes $\bar{\alpha}(w) = \alpha(x_{i_1})^{\varepsilon_1} \ldots \alpha(x_{i_n})^{\varepsilon_n}$). This property characterises free groups, in fact: if X is a subset of a group F which is universal in the sense indicated, then F is isomorphic with $F(X)$ in a natural way. The subset X is said to be a *basis* of F, and every group isomorphic to $F(X)$ is said to be free, that is, every group with a basis is free. The rank of a free group is an invariant: all its bases have the same cardinality.

1.2. Subgroups of Free Groups. As is clear from the definition given, a subset X of a group is a basis of a free subgroup F of G if and only if all products of the form (1) in which $x_{i_k}^{\varepsilon_k} x_{i_{k+1}}^{\varepsilon_{k+1}} \neq 1$ in G for all $k = 1, \ldots, n-1$ are not the identity. For example, it is easy to show that the subset

$$Y = \{y_i \equiv x_1^i x_2 x_1^{-i} | i = 0, \pm 1, \ldots\}$$

of the free group $F_2 = F(x_1, x_2)$ is a free basis for a subgroup H of F_2, namely the kernel of the homomorphism from F_2 onto the infinite cyclic group $F(x_1)$ such that $x_1 \mapsto x_1$, $x_2 \mapsto 1$. In particular, it follows from this that a free group of rank 2 can have a free subgroup of countably infinite rank.

This example is not an isolated one, since we have:

Theorem. *Every subgroup H of a free group F is a free group.*

The existence of bases for subgroups of groups of finite free rank was proved by Nielsen in 1921; Schreier did it in 1927 without restriction on the rank of F.

To describe Schreier's method, we recall the following definition.

Definition. A subset $M = \{a_i | i \in I\}$ of elements of a group G is said to be a set of *generators* if M is not contained in any proper subgroup of G. In other words, every element g in G is a product of the form $a_{i_1}^{\varepsilon_1} \ldots a_{i_n}^{\varepsilon_n}$, where $i_k \in I$, $\varepsilon_k = \pm 1$. We write $G = \langle a_i | i \in I \rangle$ to denote this.

If H is a subgroup of a group G generated by a set M, a *generating* set for H can be chosen in the following way. Write G as a disjoint union of coset modulo H: $G = \bigcup_\alpha Hs_\alpha$. Having fixed the representatives s_α of the cosets modulo H, we define the function $g \mapsto \bar{g}$, which associates with each element g the representative of the coset containing g; that is, $Hg = H\bar{g}$ and $\bar{g} = s_\alpha$ for some s_α. It is convenient to take $\bar{g} = 1$ when g is in H. In this situation, the products $s_\alpha a_i \overline{s_\alpha a_i}^{-1}$ lie in H and generate H. In fact, we have first that $Hs_\alpha a_i = H\overline{s_\alpha a_i}$, so that $s_\alpha a_i \overline{s_\alpha a_i}^{-1}$ is in H. Secondly, $\overline{\bar{g}_1 g_2} = \overline{g_1 g_2}$ in all cases, since $H\bar{g}_1 g_2 = Hg_1 g_2$. Thus, we have for $g = a_1^{\varepsilon_1} \ldots a_n^{\varepsilon_n} \in H$:

$$g = (1 \cdot a_1^{\varepsilon_1} \overline{1 \cdot a_1^{\varepsilon_1}}^{-1})(\overline{a_1^{\varepsilon_1}} a_2^{\varepsilon_2} \overline{a_1^{\varepsilon_1} a_2^{\varepsilon_2}}^{-1}) \ldots (\overline{a_1^{\varepsilon_1} \ldots a_{n-1}^{\varepsilon_{n-1}}} \cdot a_n \cdot \overline{a_1^{\varepsilon_1} \ldots a_n^{\varepsilon_n}}^{-1}),$$

since $\overline{a_1^{\varepsilon_1} \ldots a_n^{\varepsilon_n}} = \bar{g} = 1$. All we need is that

$$s_\alpha a_i^{-1} \overline{s_\alpha a_i^{-1}}^{-1} = (\overline{s_\alpha a_i^{-1}} \cdot a_i \cdot \overline{\overline{s_\alpha a_i^{-1}} \cdot a_i}^{-1})^{-1},$$

since $\overline{\overline{s_\alpha a_i^{-1}} \cdot a_i} = \overline{s_\alpha a_i^{-1} a_i} = \bar{s}_\alpha = s_\alpha$.

In particular, suppose that G is finitely generated (that is, some generating set M for it is finite), and that the number $|G:H|$ of cosets of G modulo H is finite (this is the *index* of H in G). Then H is also finitely generated.

For a free group F with basis X, it can be shown by induction that there is a so-called *Schreier system of representatives* modulo H. It has the property that every initial segment of the uncancellable representation of each s_α in terms of the letters in X is itself a coset representative. In this situation, the non-identity products $s_\alpha x_i \cdot \overline{s_\alpha x_i}^{-1}$ constitute a basis for H, so that H is free. Schreier established that a subgroup H of rank j in a free group of rank n has rank $1 + j(n-1)$, by counting the number of these products.

1.3. Nielsen's Method. Automorphisms of Free Groups. It was Nielsen who created the theory of *cancellation* in free groups, which is one of the main methods of studying them. For a finitely generated subgroup H of a free group $F(X)$, a so-called *N-reduced system* of generators can be extracted using three types of transformations:

N1. In a generating set $\{a_1, \ldots, a_m\}$ for H, replace a_i by a_i^{-1} for $i \in \{1, \ldots, m\}$.

N2. For $i, j \in \{1, \ldots, m\}$, $i \neq j$, replace a_i by $a_i a_j$.

N3. Reduce the number of generators by deleting any a_i such that $a_i = 1$ in $F(X)$.

A finite number of applications of these elementary *Nielsen transformations* reduces the system $\{a_i\}_{i=1}^n$ to an N-reduced system $\{b_j\}_{j=1}^l$, that is, a system such that:

1) $b_i \neq 1$ in $F(X)$ for $i = 1, \ldots, l$;

2) if $v \equiv b_i^\varepsilon b_j^\delta$, where $\varepsilon, \delta = \pm 1$, and $\varepsilon \neq \delta$ if $i = j$, then $|v'| \geqslant |b_i|, |b_j|$ (here v' denotes the uncancellable form of v);

3) if $b_i^\varepsilon b_j^\delta \neq 1$ and $b_j^\delta b_k^\xi \neq 1$ in $F(X)$, then

$$|(b_i^\varepsilon b_j^\delta b_k^\xi)'| > |b_i| - |b_j| + |b_k| \qquad (\varepsilon, \delta, \xi = \pm 1).$$

Theorem (Nielsen). *Every N-reduced system of uncancellable words is a basis of a subgroup H for which the sum of the lengths of the words contained in it with respect to the alphabet X is the smallest for all generating sets of H.*

When $X = \{x_1, \ldots, x_n\}$ and $H = F(X)$, it follows from this that every generating set for $F(X)$ can be carried into $\{x_1, \ldots, x_n\}$ by a finite number of elementary Nielsen transformations. Thus, the elementary automorphisms, that is, those given by the elementary Nielsen transformations, generate the automorphism group Aut $F(X)$. Aut $F(X)$ is finitely presented (see section 2.1). A set of defining relatons is to be found in Magnus, Karras and Solitar (1966).

Birman has established the following criterion in terms of free differential calculus (the necessary definitions are given in section 4.3).

Theorem. *A set of words $\{w_1, \ldots, w_n\}$ in the alphabet $X = \{x_1, \ldots, x_n\}$ is a basis for $F(X)$ if and only if the Fox matrix $\left(\frac{\partial w_i}{\partial x_j}\right)$ is invertible in the group ring $\mathbb{Z}[F(X)]$.*

Every Nielsen automorphism fixes almost all elements in the basis, so that Aut $F(X)$ is not generated by the Nielsen automorphisms when X is infinite. However, the subgroup $\mathrm{Aut}_N F(X)$ they generate is "dense" in Aut $F(X)$:

Theorem. *For every finite set of elements $w_1, \ldots, w_n$ in $F(X)$ and every α in* Aut $F(X)$, *there is an automorphism β in $\mathrm{Aut}_N F(X)$ such that $\alpha(w_i) = \beta(w_i)$ for each $i = 1, \ldots, n$.*

The free group $F_n = F(x_1, \ldots, x_n)$ has a natural homomorphism onto the free abelian group of rank n (see section 1 of Chapter 2). Thus, there is a canonical homomorphism ϕ of Aut F_n into $\mathrm{GL}_n(\mathbb{Z})$. Using the fact that $\mathrm{GL}_n(\mathbb{Z})$ is generated by elementary matrices, it is easy to see that ϕ is surjective, so that it is an epimorphism. Clearly, its kernel contains all the inner automorphisms (that is, automorphisms of the form α_a, $a \in F_n$, such that $\alpha_a(g) = aga^{-1}$ for all g in F_n). For $n = 2$, Nielsen proved the following result:

Theorem. *The kernel of the homomorphism ϕ*: Aut $F_2 \to \mathrm{GL}_2(\mathbb{Z})$ *consists exactly of the inner automorphisms of F_2.*

The analogous assertion for $n \geqslant 3$ is false.

There exists an algorithm allowing one to decide, for a pair u, v of words in $F(X)$, whether or not there is an automorphism carrying the one to the other. Rather than dwelling on particular results about the automorphism group of $F(X)$, we state next a powerful assertion, proved quite recently by Gersten. Unfortunately, an explanation of the proof would require the introduction of a large number of special terms.

Theorem. *The fixed-point subgroup of every automorphism of a free group of finite rank is of finite rank.*

Howson's theorem (see section 1.4) shows that this result can be extended to the fixed-point subgroup of any finitely generated subgroup G of Aut F_n.

1.4. Other Realisations of Free Groups. In addition to the representation of a free group as a group of words, there are some other useful descriptions. For example, the *Magnus representation* of $F(X)$ in the ring of formal power series in non-commuting variables has been a particularly fruitful one. The ring $A(\mathbb{Z}, X)$ consists of the formal sums of the form $\sum_{i=0}^{\infty} u_i$, where u_i is a homogeneous polynomial of degree i, that is, a finite integral linear combination of "non-commuting" monomials of the form

$$x_{i_1}^{l_1} \dots x_{i_n}^{l_n}, \quad \text{where} \quad i_k \neq i_{k+1} \quad \text{for} \quad k = 1, \dots, n-1, \quad l_k > 0, \quad \sum_{k=1}^{n} l_k = i.$$

The multiplication of monomials is simple juxtaposition, and this operation extends naturally to series. The set of invertible elements of $A(\mathbb{Z}, X)$ consists of the series with free term ± 1. For example, $(1 + x_j)^{-1} = 1 - x_j + x_j^2 - x_j^3 + \cdots$.

Lemma. *The elements* $1 + x_j$ *with* $x_j \in X$ *freely generate a subgroup of the multiplicative group isomorphic with* $F(X)$.

For the proof, we need to check that the series

$$(1 + x_{i_1})^{l_1} \dots (1 + x_{i_n})^{l_n},$$

where $n \geqslant 1$, $i_k \neq i_{k+1}$, $l_k \neq 0$, is different from 1. It can be rewritten in the form

$$(1 + l_1 x_{i_1} + \cdots) \dots (1 + l_n x_{i_n} + \cdots),$$

whence it is clear that the monomial $x_{i_1} \dots x_{i_n}$ occurs in the product with non-zero coefficient $l_1 \dots l_n$. This proves the lemma.

This description of $F(X)$ enabled Magnus to study the lower central series (see section 4.1 in Chapter 2) of a free group. It turned out that $\gamma_m(F(X))$ consists of the series $\sum_{i=0}^{\infty} u_i$ in $F(X)$ for which $u_1 = \cdots = u_{m-1} = 0$. The next theorem follows immediately from this.

Theorem. 1) $\bigcap_{m=1}^{\infty} \gamma_m(F(X)) = \{1\}$; 2) *the nilpotent factorgroups* $F(X)/\gamma_m(F(X))$ *are torsion-free*; 3) *the factors* $\gamma_m(F(X))/\gamma_{m+1}(F(X))$ *are free abelian.*

Witt discovered the *number of generators* of these abelian groups; if $F = F_n$ is free of rank n, the rank of $\gamma_m(F_n)/\gamma_{m+1}(F_n)$ as free abelian group (see section 1 of Chapter 2) is

$$\frac{1}{m} \sum_{d|m} \mu(d) n^{m/d},$$

where μ is the Möbius function.

Sanov found a beautiful linear representation of F_2, in 1947.

Theorem. *The correspondence* $x_1 \mapsto \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$, $x_2 \mapsto \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$ *defines an isomorphic embedding of* F_2 *into* $\mathrm{GL}_2(\mathbb{Z})$.

If we change the matrices to

$$A_\mu = \begin{pmatrix} 1 & \mu \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad B_\mu = \begin{pmatrix} 1 & 0 \\ \mu & 1 \end{pmatrix},$$

the corresponding representation is an isomorphism for all μ with $|\mu| \geqslant 2$. Clearly, all transcendentals μ are also "free". Several articles (see Lyndon and Schupp (1977)) are devoted to determining which complex numbers are free and which not, in the sense indicated. Interest in this problem has been stimulated by the *Tits alternative*: a finitely generated linear group either contains a free subgroup of rank 2, or else it has a soluble subgroup of finite index (see Part II, "Linear Groups", of this volume).

Residual finiteness of free groups is easily deduced from the Sanov representation. At this point, we give a more general definition.

Definition. Let $\mathscr{K}$ be a class of groups. A group G is said to be residually in $\mathscr{K}$ if, for every $g \in G\setminus\{1\}$, there exists an epimorphism ϕ of G onto a group K in $\mathscr{K}$ such that $\phi(g) \neq 1$.

A group G is residually in $\mathscr{K}$ if and only if it is isomorphic with a subgroup of a Cartesian product of groups from $\mathscr{K}$ such that the projection onto each component is an epimorphism.

If $\mathscr{K}$ is the class of all finite groups, the groups that are residually in $\mathscr{K}$ are called *residually finite*. A group G is residually finite if (and only if) the intersection of all normal subgroups of finite index is the identity subgroup.

To prove that the free group $F(X)$ is residually finite, it is enough to do it when $|X| < \infty$, since every word w in $F(X)$ can be written in terms of a finite subset of the alphabet, and maps to a nonempty word under an epimorphism to some F_r. The example in section 1.2 shows that F_r is embeddable in F_2, so that by Sanov's theorem it is embeddable in the group $SL_2(\mathbb{Z})$ of integer matrices of determinant 1. Finally, this last group is residually in the set of finite groups of the form $SL_2(\mathbb{Z}_n)$, the groups of unimodular matrices over the residue class rings $\mathbb{Z}_n$, $n = 1, 2, \dots$.

Definition. A group G is said to be *Hopfian* if every epimorphism $G \to G$ is an automorphism of G (that is, it has trivial kernel).

Every finitely generated residually finite group is Hopfian (see Magnus, Karrass and Solitar (1966)). Thus, the free group F_n is Hopfian, a fact that can also be deduced from the theorem of Magnus in 1.3.

The following assertion of Burns about free groups is significantly stronger than residually finiteness. (See section 2.2 for the definition of free product.)

Theorem. *Let A be a finite subset of a free group F, and H a finitely generated subgroup of F such that $H \cap A = \varnothing$. Then H is a free factor of a subgroup K of finite index in F such that $K \cap A = \varnothing$.*

It is easy to deduce from this that, if H is a finitely generated subgroup of F containing a non-trivial normal subgroup, then $|F:H| < \infty$. In particular, non-trivial finitely generated normal subgroups of free groups are of finite index. Howson's theorem can also be deduced from that of Burns:

Theorem. *The intersection of two finitely generated subgroups of a free group is finitely generated.*

§ 2. Defining Relations and Free Constructions

2.1. Presentations of Groups. Suppose that a set $M = \{a_i | i \in I\}$ of generators is chosen in a group G. As we saw in § 1, there is just one epimorphism of the free group $F(X)$ on the alphabet $X = \{x_i | i \in I\}$ onto G extending the map $x_i \mapsto a_i$. The epimorphism ϕ is called a *presentation* of G. It depends on the generating set for G. The elements of Ker ϕ are called relations in G (among the generators in M); however, if $x_{i_1}^{\varepsilon_1} \dots x_{i_n}^{\varepsilon_n} \in \operatorname{Ker} \phi$, it is usual to write the relation in the form $a_{i_1}^{\varepsilon_1} \dots a_{i_n}^{\varepsilon_n} = 1$ (this is, of course, a true equation in G). The expression $a_{i_1}^{\varepsilon_1} \dots a_{i_k}^{\varepsilon_k} = a_{i_n}^{-\varepsilon_n} \dots a_{i_{k+1}}^{-\varepsilon_{k+1}}$ is also permitted.

If the set R of relations is not contained in any normal subgroup N of $F(X)$ such that $N \subsetneqq \operatorname{Ker} \phi$, then R is called a *set of defining relations* for G (among the generators in M). Since $G \cong F(X)/\operatorname{Ker} \phi$, G is uniquely defined (up to isomorphism) by the pair X, R; it is suggested in Kargapolov and Merzlyakov (1982) that the pair be called a genetic code for G. The expression $G = \langle M; R\rangle$ is normally used to express the fact that the group is given by generators and relations. If all the relations in R hold for elements b_i $(i \in I)$ of some group H, the map $a_i \mapsto b_i$ extends to a homomorphism $G \to H$. This property is sometimes called von Dyck's theorem; defining relations first made their appearance in a paper published by von Dyck in 1882/83.

A typical situation in combinatorial group theory is when one has a presentation, that is, X and R are the initial data; often this group is not given in any other way (say, as a transformation group), and its properties are investigated by analyzing the consequences of the defining relations.

Definition. The group G is said to be *finitely presented* if M and R are finite in some presentation $G = \langle M; R\rangle$.

How can one prove that some set of relations in a known group are defining relations? To do this, it must be shown that every relation r lies in the normal closure of R in $F(X)$, that is, in the least normal subgroup of $F(X)$ containing R. This holds if and only if r can be written in the form

$$r = \prod u_k r_k u_k^{-1},$$

where $r_k \in R$, $u_k \in F(X)$. In other words, it must be shown that r is 1 in any group where all the words in R are 1.

We consider there the example provided at the beginning of the chapter. If $r = 1$ is a relation in S_3, then using the relations $a^2 = 1$ and $b^2 = 1$ in R, we can bring r to the form $(a)bab\dots aba(b)$ where the letters in brackets may be absent. The length of this expression can to be taken to be less than 6, as is easily seen using the relation $(ab)^3 = 1$. From this it follows that r is empty, since it is simple to check that the products , ab, aba, $abab$, $ababa$ are not 1 in S_3. It has thus been shown that $r = 1$ is a consequence of the three relations, that is,

$$S_3 = \langle a, b; a^2 = b^2 = (ab)^3 = 1\rangle.$$

It is even easier to show that the free abelian group of rank n (see § 1 of Chapter 2) can be given by the presentation

$$\langle a_1, \dots, a_n; a_i a_j a_i^{-1} a_j^{-1} = 1, 1 \leqslant i < j \leqslant n\rangle.$$

A useful method for determining sets of defining relations for finite groups, as well as proving that a group given by defining relations is actually finite, is the method of enumerating cosets (see Kargopolov and Merzlyakov (1982)). It is very convenient for implementation on a computer, and gives a description of the group in terms of permutations.

There is considerable arbitrariness in presenting groups in terms of generators and relations, both in the choice of generators and that of defining relations. However, any presentation can be taken into any other using *Tietze transformations*. For finitely presented groups, they are of the following type:

1) if $r = 1$ is a consequence of the relations in R, add r to R;

2) if $r \in R$, and the relation $r = 1$ can be deduced from the remaining relations in R, delete r from R;

3) if W is a word in the alphabet X, adjoin a letter w to the alphabet and add the relation $w = W$ to the set of defining relations;

4) if a defining relation has the form $w = W$, where the word W is an expression in the letters other than w, delete w from X, delete $w = W$ from the list of defining relations, and replace w by W in the remaining defining relations.

For example, in the generators $a = (1, 2)$, $c = (1, 2, 3)$, our group has presentation

$$\langle a, c; a^2 = 1, c^3 = 1, aca^{-1} = c^{-1}\rangle.$$

It is obtained from the example considered above using the following Tietze transformations:

$$\langle a, b; a^2 = 1, b^2 = 1, (ab)^3 = 1\rangle$$
$$\stackrel{(3)}{\to} \langle a, b, c; a^2 = 1, b^2 = 1, (ab)^3 = 1, c = ba\rangle$$
$$\stackrel{(1)}{\to} \langle a, b, c; a^2 = 1, b^2 = 1, (ab)^3 = 1, c = ba, b = ca^{-1}\rangle$$
$$\stackrel{(4)}{\to} \langle a, c; a^2 = 1, (ca^{-1})^2 = 1, (aca^{-1})^3 = 1, c = caa^{-1}\rangle$$
$$\stackrel{(2)}{\to} \langle a, c; a^2 = 1, (ca^{-1})^2 = 1, (aca^{-1})^3 = 1\rangle$$
$$\stackrel{(1)}{\to} \langle a, c; a^2 = 1, (ca^{-1})^2 = 1, (aca^{-1})^3 = 1, c^3 = 1\rangle$$

$$\overset{(2)}{\to} \langle a, c; a^2 = 1, c^3 = 1, (ca^{-1})^2 = 1\rangle$$
$$\overset{(1)}{\to} \langle a, c; a^2 = 1, c^3 = 1, (ca^{-1})^2 = 1, aca^{-1} = c^{-1}\rangle$$
$$\overset{(2)}{\to} \langle a, c; a^2 = 1, c^3 = 1, aca^{-1} = c^{-1}\rangle.$$

Tietze's theorem remains true for groups with infinitely many generators or relations. It is simply necessary to apply a single "elementary" transformation to add or delete infinitely many letters or relations all at once. Finally, we note that Tietze's theorem carries over without change to rings, semigroups and other algebraic systems.

A presentation can be found for a subgroup K of $G = \langle X; R\rangle$ using the *Reidemeister-Schreier rewriting process*. It comes to choosing a Schreier system of generators $s_\alpha x_i \overline{s_\alpha x_i}^{-1}$ (see section 1.2) for the subgroup H of $F(X)$ which is the complete inverse image of K under the homomorphism $F(X) \to G$; this corresponds to a new alphabet $Y = \{y_{\alpha,i}\}$. Further, words of the form $s_\alpha r s_\alpha^{-1}$ ($r \in R$) lie in H, and it is natural to compare them with the words in Y using the rewriting process, that is, with the expressions of the words $s_\alpha r s_\alpha^{-1}$ as products of elements $(s_\alpha x_i \overline{s_\alpha x_i}^{-1})^{\pm 1}$. The set U of all words obtained in this way constitutes a set of defining relations for H: $H = \langle Y; U\rangle$.

For example, consider

$$G = \langle a, b, c, d; aba^{-1}b^{-1}cdc^{-1}d^{-1} = 1\rangle,$$

the fundamental group of a two-sphere with two handles. The kernel of the homomorphism of G onto a cyclic group of order 2 such that $b \mapsto 1$, $c \mapsto 1$, $d \mapsto 1$ is of index 2 in G; $\{1, a\}$ is a Schreier system of representatives for the corresponding subgroup H, and

$$\{a^2, b, c, d, aba^{-1}, aca^{-1}, ada^{-1}\} \tag{2}$$

is a Schreier system of generators for H. In this case,

$$aba^{-1}b^{-1}cdc^{-1}d^{-1} = (aba^{-1})\cdot b^{-1}\cdot c\cdot d\cdot c^{-1}\cdot d^{-1}$$

and

$$a(aba^{-1}b^{-1}cdc^{-1}d^{-1})a^{-1}$$
$$= a^2\cdot b\cdot a^{-2}\cdot(aba^{-1})\cdot(aca^{-1})\cdot(ada^{-1})\cdot(aca^{-1})^{-1}\cdot(ada^{-1})^{-1}.$$

This means that in the alphabet $\{t, u, v, w, x, y, z\}$ corresponding to the set (2), we get

$$K = \langle t, u, v, w, x, y, z; xu^{-1}vwv^{-1}w^{-1} = 1, tut^{-1}xyzy^{-1}z^{-1} = 1\rangle.$$

Removing x by a Tietze transformation using the first relation, we get

$$K = \langle t, u, v, w, y, z; tut^{-1}wvw^{-1}v^{-1}uyzy^{-1}z^{-1} = 1\rangle.$$

It is easy to see, further, that Tietze transformations enable us to introduce letters $w_1 = uwu^{-1}$ and $v_1 = uvu^{-1}$ replacing w and u; having done that, we see that

$$K = \langle t, u, v_1, w_1, y, z; tut^{-1}u^{-1}w_1v_1w_1^{-1}v_1^{-1}yzy^{-1}z^{-1}\rangle$$

is in fact the fundamental group of a sphere with three handles.

2.2. Free Constructions. The free product, like the direct product, has a natural analogue in topology. The fundamental group of the Cartesian product of two path-connected spaces X_1 and X_2 is the direct product $G_1 \times G_2$. The fundamental group of the union of X_1 and X_2 joined at a single point is isomorphic to the free product $G_1 * G_2$.

Definition. The *free product of the groups* G_i, $i \in I$, is a group $G = *_{i\in I} G_i$, together with homomorphisms ϕ_i: $G_i \to G$ such that, for every group A and all homomorphisms α_i: $G_i \to A$, there exists just one homomorphism ϕ: $G \to A$ such $\phi\phi_i = \alpha_i$ for every i in I.

Such a group G exists, and is unique up to a natural isomorphism. If the G_i are give by defining relatons in disjoint generating sets, a presentation for G is obtained by uniting the generating sets and the sets of defining relations of the G_i. The homomorphisms ϕ_i are injective (that is, they are monomorphisms), and the G_i are usually identified with subgroups of G. Every non-identity element g in G has a canonical representation of the form $g = g_1 g_2 \ldots g_n$, where g_k and g_{k+1} are non-identity elements from different factors for $k = 1, \ldots, n-1$; n is the length of g.

Every free group $F(X)$ decomposes as the free product of the infinite cyclic subgroups $\langle x_i\rangle$, $x_i \in X$. The main structure theorem about free products is the Kurosh subgroup theorem:

Theorem. *Every subgroup H of a free product $G = *_{i\in I} G_i$ decomposes as a free product of subgroups, each of which is either conjugate*[1] *to a subgroup of one of the factors G_i, or is infinite cyclic.*

Suppose that $G = G_1 * \cdots * G_n$ and that G is generated by a finite set $M = \{g_1, \ldots, g_m\}$. Grushko's theorem says that Nielsen transformations can be applied to M to give a set

$$M' = \{g_{11}, \ldots, g_{1.m_1}; \ldots; g_{n1}, \ldots, g_{n.m_n}\}$$

such that $\langle g_{i1}, \ldots, g_{in_i}\rangle = G_i$. In particular, the minimum number of generators of G is the sum of the minimum numbers of generators of the factors G_i.

Definition. Let A_1 and A_2 be isomorphic subgroups of groups G_1 and G_2 respectively with presentations $G_1 = \langle X_1, R_1\rangle$, $G_2 = \langle X_2, R_2\rangle$. By the free product of G_1 and G_2 with subgroups A_1 and A_2 amalgamated according to the isomorphism ϕ: $A_1 \to A_2$ we mean the group $G = \langle G_1 * G_2, A_1 = A_2, \phi\rangle$ with

[1] We recall that an element of the form aga^{-1} (subgroup of the form aHa^{-1}) is said to be conjugate to the element g (to the subgroup H) in G.

generating set the disjoint union $X_1 \cup X_2$ and defining relations the union of R_1, R_2 and the set $\{a = \phi(a) | a \in A_1\}$.

The construction of the free product with amalgamation over any number of factors is fully analogous. (A single subgroup A_i is chosen in each G_i!) It does not depend on the choice of the presentations of the factors. The groups G_i are isomorphically embedded in G and are identified with their images there, and $G_i \cap G_j = A_i = A_j = A$.

Every element g of G has a canonical expression obtained after choosing coset representatives: $G_i = \bigcup_\alpha A_i s_{i\alpha}$. Namely,

$$g = a s_{i_1\alpha_1} s_{i_2\alpha_2} \cdots s_{i_n\alpha_n},$$

where the $s_{i_k\alpha_k}$ are not the identity, $i_k \neq i_{k+1}$ for $k = 1, \ldots, n$, and $a \in A$.

The simplest example of a free product is the group $\mathrm{PSL}_2(\mathbb{Z})$ of fractional linear transformations of the upper complex half-plane:

$$z \mapsto \frac{az + b}{cz + d}, \tag{3}$$

where a, b, c, d are integers and $ad - bc = 1$. It is the free product of the cyclic subgroups $\langle\phi\rangle$ of order 3 and $\langle\psi\rangle$ of order 2, where $\phi(z) = 1 - \frac{1}{z}$, $\psi(z) = -\frac{1}{z}$ (see section 3.2). This group is a natural homomorphic image of the group $SL_2(\mathbb{Z})$ of unimodular matrices: $A \mapsto \phi$, $B \mapsto \psi$, where $A = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ (generally, the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ corresponds to the transformation (3)). The kernel of this map consists exactly of the matrices $\pm E$. Thus

$$SL_2(Z) = \langle A, B; A^6 = 1, B^4 = 1, A^3 = B^2 \rangle$$

is the free product of cyclic groups of order six and four amalgamating subgroups of order 2.

The Seifert-van Kampen theorem points up the importance of studying free constructions in topology. If the space $Z = X \cup Y$ is the union of path-connected spaces X and Y with nonempty path-connected intersection V such that the natural homomorphisms $\pi_1(V) \to \pi_1(X)$ and $\pi_1(V) \to \pi_1(Y)$ of fundamental groups are monomorphisms, then $\pi_1(Z)$ is isomorphic with the free product of $\pi_1(X)$ and $\pi_1(Y)$ with $\pi_1(V)$ amalgamated.

There is another construction that is naturally studied alongside free products, namely the HNN-extensions (or Higman-Neumann-Neumann extensions).

Definition. Let G be a group, A and B subgroups of G, and $\phi: A \to B$ an isomorphism. The HNN-*extension* G^* of G with respect to ϕ is the group obtained by adding a single letter t to the list of generators of G and the relations $tat^{-1} = \phi(a)$ for all a in A to the set of defining relations.

Once again we need to say that G is embedded naturally in G^* (G is called the base subgroup of G^*), and that the definition of G^* is invariant with respect to the choice of presentation for G.

A sequence $g_0, t^{\varepsilon_1}, g_1, \ldots, t^{\varepsilon_n}, g_n$, where $g_i \in G$, $\varepsilon_i = \pm 1$, is said to be reduced if the elements t, g_i, t^{-1} with g_i in A do not occur consecutively in it, nor t^{-1}, g_i, t with g_i in B. The foundation for the combinatorial study of HNN-extensions is the Novikov-Britton lemma:

Lemma. *If $g_0, t^{\varepsilon_1}, g_1, \ldots, t^{\varepsilon_n}, g_n$ is a reduced sequence with $n \geqslant 1$, then $g_0 t^{\varepsilon_1} g_1 \ldots t^{\varepsilon_n} g_n$ is different from 1 in G^*.*

HNN-extensions also have a topological significance. Let X and Y be open path-connected subspaces of a path-connected space Z for which there exists a homeomorphism $X \to Y$ such that the fundamental groups $\pi_1(X)$ and $\pi_1(Y)$ are isomorphically embedded in $\pi_1(Z)$. We construct a space W by attaching the space $X \times [0, 1]$ (as handle) to Z, identifying $X \times \{0\}$ with X and $X \times \{1\}$ with Y. The fundamental group $\pi_1(W)$ of W is the HNN-extension of $\pi_1(X)$ relative to the isomorphism between its subgroups $\pi_1(X)$ and $\pi_1(Y)$.

HNN-extensions first found group-theoretical application in the proof of an embedding theorem in a paper by Higman, Neumann and Neumann in 1949.

Theorem. *Every countable group G is isomorphic with a subgroup of a two-generator group A.*

If G is finitely presented, so is A. It was also proved in that article that every countable group can be embedded in a countably infinite group in which all elements of given order are conjugate. As Goryushkin has proved, every countable group can even be embedded in a 2-generator simple group. (This is just one of many examples which show that 2-generator groups are just as complicated as arbitrary finitely generated groups. The same is true in relation to many special classes of groups, to be considered in the second chapter). The first examples of infinite finitely presented simple groups were found by R.J. Thompson while investigating the automorphism groups of free Cantor algebras (in this context, see also the survey "Identities" in the present series).

§ 3. Properties of Free Constructions

3.1. Subgroups of Free Products with Amalgamation and HNN-Extensions. There is no natural analogue of the Kurosh subgroup theorem for free products in the case of free products with amalgamation nor HNN-extensions, when these constructions are looked at on their own. It has turned out that the answers to the two subgroup problems come only when these basic constructions are studied together. Structure theorems were obtained by Karrass and Solitar in 1970–71.

In order to state subgroup theorems, we shall first generalise the concept of free product with amalgamation. Let T be a finite or infinite tree, that is, a connected graph without cycles. Suppose further that each vertex v in the vertex-set V is associated with a group G_v, and that to each (non-oriented) edge e in E (E is the edge set) there is a subgroup H_e embedded in G_v and G_w *via* monomorphisms $\phi_{e,v}$ and $\phi_{e,w}$ if e joins v and w.

Definition. Suppose that the groups $G_v = \langle X_v; R_v \rangle$ are given on disjoint generating sets X_v. Their *tree product* is the group G with generating set $\bigcup_{v \in V} X_v$ and defining relations consisting of $\bigcup_{v \in V} R_v$ together with all relations of the form

$$\varphi_{e,v}(h) = \varphi_{e,w}(h) \qquad \text{for all} \qquad h \in H_e, e \in E.$$

Every subgroup H of the free product $G = \langle G_1 * G_2, A_1 = A_2, \phi \rangle$ with amalgamation can be described as follows. Let $\{s_\alpha\}$ be a system of representatives of the double cosets of G modulo (H, G_1), so that $G = \bigcup_\alpha H s_\alpha G_1$. Similarly, let $\{t_\beta\}$ be a system of representatives modulo (H, G_2). We set

$$H_\alpha = H \cap s_\alpha G_1 s_\alpha^{-1}, \qquad H_\beta = H \cap t_\beta G_2 t_\beta^{-1}.$$

Theorem. *The subgroups H_α and H_β generate in H their tree product K with amalgamated subgroups of the form $H \cap s_\alpha A_1 s_\alpha^{-1}$ and $H \cap t_\beta A_2 t_\beta^{-1}$, and H is an* HNN-*extension with base subgroup K.*

However, the conjugacy is introduced not simply for a single pair of subgroups of K, but rather for infinite sets of pair of isomorphic subgroups.

A parallel theorem holds for subgroups of HNN-extensions. Every subgroup H of an HNN-extension G^* with base subgroup G can be given as an HNN-extension of a tree product of subgroups of the form $H \cap aGa^{-1}$, $a \in G^*$. (See Lyndon and Schupp (1977) for more details.)

3.2. Free Constructions as Transformation Groups. Free products (with amalgamation) and HNN-extensions can be characterised as transformation groups, as was observed in papers of Macbeath and Maskit. The statement of our first theorem is suggested by a simple example. Let $G = G_1 * G_2$ be a free product acting regularly on itself by left translations: $x \to gx$. Further, suppose that S_1 (S_2 respectively) is the set of non-identity elements of G whose normal forms start with factors in G_1 (in G_2 respectively). Clearly, $gS_2 \subseteq S_1$ for $g \in G_1 \backslash \{1\}$ and $gS_1 \subseteq S_2$ for $g \in G_2 \backslash \{1\}$. The converse is also true.

Theorem. *If G is a group of permutations of some set S generated by subgroups G_1 and G_2 in such a way that there are two disjoint subsets S_1 and S_2 of S for which $gS_2 \subseteq S_1$ when $g \in G_1 \backslash \{1\}$ and $gS_1 \subseteq S_2$ when $g \in G_2 \backslash \{1\}$, then $G = G_1 * G_2$, except possibly when $|G_1| = |G_2| = 2$ and $G = \langle a, b; a^2 = b^2 = (ab)^n = 1 \rangle$ is dihedral of order $2n$.*

Using this theorem, we shall show that the transformations $\phi(z) = 1 - \frac{1}{z}$ and $\psi(z) = -\frac{1}{z}$ of the upper complex half-plane mentioned in 2.2 really do generate free factors, that is, $PSL_2(\mathbb{Z}) = \langle\phi\rangle_3 * \langle\psi\rangle_2$. It is immediately obvious that ϕ^3 and ψ^2 are the identity. For S we take the upper half-plane, and for S_1 the first quadrant: $S_1 = \{z | \operatorname{Re} z > 0, \operatorname{Im} z > 0\}$. Similarly, S_2 is the second quadrant $\{z | \operatorname{Re} z < 0, \operatorname{Im} z > 0\}$. It is obvious that $\psi(S_1) \subseteq S_2$ and $\phi(S_2) \subseteq S_1$; since $\phi^2(z) = \frac{1}{1-z}$, we have $\phi^2(S_2) \subseteq S_1$. The requirements of the theorem are satisfied, so we just have to apply it!

Similarly, to show that the group $\langle A_\mu, B_\mu\rangle$ generated by the matrices $A_\mu = \begin{pmatrix} 1 & \mu \\ 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ \mu & 1 \end{pmatrix}$ is free when $|\mu| \geqslant 2$ (see 1.4), it is enough to observe that it acts as a group of fractional linear transformations of the Riemann sphere $\mathbb{C} \cup \{\infty\}$ given by $A_\mu(z) = z + \mu$ and $B_\mu(z) = \frac{z}{\mu z + 1}$. Now define $S_1 = \{z | |z| > 1\}$ and $S_2 = \{z | |z| < 1\}$ and note the following conclusions, which are obvious for $|\mu| \geqslant 2$:

$$A_\mu^k(S_2) \subseteq S_1 \quad \text{and} \quad B_\mu^k(S_1) \subseteq S_2 \quad \text{for} \quad k = \pm 1, \pm 2, \ldots.$$

The dihedral group really is an exception to the theorem. It is enough to consider its natural representation as the group of symmetries of a dihedron consisting of regular pyramids with a common base (Fig. 1); a and b are chosen to be the rotations of three-space through an angle π about axes that are axes of symmetry of the base of the dihedron with constitutent angle $\frac{\pi}{n}$. We set $S_1 = \{o_1\}$, $S_2 = \{o_2\}$, where o_1 and o_2 are opposite vertices of the pyramids.

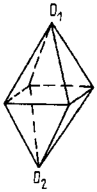

Fig. 1

We have the following characterisation for free products with amalgamation.

Theorem. *Let G be a group of permutations of a set S generated by subgroups G_1 and G_2 whose intersection H is a proper subgroup in each of them such that the indices $|G_1 : H|$ and $|G_2 : H|$ are not both 2. Suppose that there are disjoint subsets S_1 and S_2 of S such that $gS_2 \subseteq S_1$ for $g \in G_1 \backslash H$ and $gS_1 \subseteq S_2$ for $g \in G_2 \backslash H$, and also $hS_1 \subseteq S_1$, $hS_2 \subseteq S_2$ for $h \in H$. Then G is the free product of G_1 and G_2 with H amalgamated.*