Combinatorics over free groups

Alexei Myasnikov
City College of New York
New York, NY 10031

Vladimir Shpilrain

Table of contents

I. Free groups
1.1.Definitions.
1.2.Construction of a free group $F(X)$ with basis X.
1.3.The universal property of free groups.
1.4.Rank.
II. Elementary equations over free groups
2.1.General remarks.
2.1.1. Equations and some classical problems.
2.1.2. Cancellation schemes.
2.2. Equation $x^n = g$.
2.2.1. Cyclically reduced forms and the conjugacy problem.
2.2.2. Uniqueness of roots. Power problem.
2.3. Equation $[x, y] = g$.
2.3.1. Homogeneous equation $[x, y] = 1$.
2.3.2. Centralizers in $F(A)$. Commutative-transitive groups.
2.3.3. Wicks forms. Recognition of commutators.
2.3.4. Elementary transformations and minimal solutions.
III. Finitely generated subgroups of F
3.1. Graphs and automata.
3.1.1. Graphs.
3.1.2. Automata.

3.2. Stallings Folding Γ_H
3.2.1. Construction of Γ_H.
3.2.2. Membership problem.
3.2.3. Uniqueness of the core automaton.
3.2.4. Jitsukawa's Theorem.
3.3. Subgroup theorem.
3.4. Index of H.
3.5. Howson property.
3.6. H. Neumann conjecture.
3.7. Malnormal subgroups.
3.8. Free factors. M. Hall theorem.
3.9. Principal quotients.
3.10. Algebraic extensions.
3.11. Isolated subgroups.
3.12. Topological closures.
IV. Nielsen Method
1. N- reduced sets.
1.1. $N1$-reduced sets (Hoar result).
1.2. $N2$-reduced sets (construction with maximal tree).
1.3. N-reduced and geodesic trees.
2. Schreier method.
2.1. Characterization of Schreier transversals in terms of trees.
2.2. Minimal Schreier transversals.
2.3. Subgroup theorem (general case).
3. Shortlex (Shortwest) ordering on $F(X)$.
3.1. Shortwest and regular trees.
3.2. Subgroups theorem (elimination procedure).
4. Nielsen elementary transformations (Shortwest approach).
V. Automorphisms of F
1. Hopfian groups.
2. Nielsen automorphisms.
3. IA-automorphisms.
4. Automorphisms of F_2.
4.1. Chan's theorem.
4.2 Mal'cev's equation.
VI. Test elements
1. Test elements.
2. Retracts (Turner's theorem).

3. Diophantine sets and epimorphism problem.
4. 3-satisfiability and endomorphism problem.

VII. Whitehead Method
1. $AutF$-equivalence.
2. Whitehead theorem.
2.1. W-autos.
2.2. Pick reduction lemma (no proof).
2.3. Whitehead algorithm (graphs and complexity).

VIII. Fixed points and stabilizers
II.b. Equations in one variable.
1. Appel's theorem.
2. Number of irreducible components (relation to stab).
3. Isolated points (uniform bounds).

IX. Quadratic equations. Quadratic words
1. Quadratic words.
2. Basic solutions and regular stabilizers.
3. C and E result.

X. Calculus over free groups
1. Derivatives.
2. The chain rule.
3. Fox's theorem.
4. Inverse function theorem.
5. Implicit function theorem.

XI. Makanin-Rasborov Method
1. Generalized equations.
2. Elementary transformations.
3. Quadratic part.
4. Entire transformations.
5. Graphs and fundamental sequences.
6. T.Q.Q. systems.
7. Decidability of the Diophantine problem.
8. Complexity.

1 Free groups

1.1. Definitions.

1.2. Construction of a free group $F(X)$ with basis X.

1.3. The universal property of $F(X)$. Presentations of groups.

1.4. Rank.

1.5. Examples.

1.1 Definitions

Let G be a group. If H is a subgroup of G then we write $H \leq G$; if H is a normal subgroup of G we write $H \trianglelefteq G$.

Let X be a subset of G. By $\langle X \rangle$ we denote the subgroup of G generated by X, i.e., the least subgroup of G containing X (= the intersection of all subgroups of G containing X). It is easy to see that

$$\langle X \rangle = \{x_{i_1}^{\varepsilon_1}, \ldots, x_{i_n}^{\varepsilon_n} \mid \ x_{i_j} \in X, \ \varepsilon_j \in \{1, -1\}, \ n \in N\}$$

An expression of the type

$$w = x_{i_1}^{\varepsilon_1} \cdots x_{i_n}^{\varepsilon_n} \quad (x_{i_j} \in X, \varepsilon_j \in \{1, -1\}), \tag{1}$$

is called a *group word* in X. If $X \subseteq G$ then every group word w in X determines a unique element from G which is equal to the product $x_{i_1}^{\varepsilon_1} \cdots x_{i_n}^{\varepsilon_n}$ of the elements $x_{i_j}^{\varepsilon_j}$.

We assume that the empty word is a group word in X and it defines the identity of G. A word

$$w = x_{i_1}^{\varepsilon_1} \cdots x_{i_n}^{\varepsilon_n}$$

is called reduced if for each $j = 1, \ldots, n-1$ either $x_{i_j} \neq x_{i_{j+1}}$ or $\varepsilon_j + \varepsilon_{j+1} \neq 0$.

Definition 1 *A group G is called a free group if there exists a generating set X of G such that every non-empty reduced group word in X defines a non-trivial element of G.*

In this event X is called a *free basis* of G and G is called *free on* X or *freely generated by* X. It follows from the definition that every element of a free group on X can be defined by a reduced group word on X. Moreover, different reduced words on X define different elements. We will say more about this in the next section.

1.2 Construction of a free group with basis X

Let X be an arbitrary set. Consider

$$X^{-1} = \{x^{-1} | x \in X\},$$

where x^{-1} is just a formal expression obtained from x and -1. If $x \in X$ then the symbols x and x^{-1} are called *literals* in X. Denote by

$$X^{\pm 1} = X \cup X^{-1}$$

the set of all literals in X. For a literal $y \in X^{\pm 1}$ we define y^{-1} as follows

$$y^{-1} = \begin{cases} x^{-1}, & \text{if } y = x \in X; \\ x, & \text{if } y = x^{-1} \in X. \end{cases}$$

Now a group word

$$w = x_{i_1}^{\varepsilon_1} \cdots x_{i_n}^{\varepsilon_n}, \quad (x_{i_j} \in X, \varepsilon_i \in \{1, -1\}) \tag{2}$$

is just a word in the alphabet $X^{\pm 1}$. The number n is called the *length* of w, we denote it by $|w|$.

Observe, that the word

$$w = y_1 \dots y_n, \quad (y_i \in X^{\pm 1})$$

is reduced if and only if, $y_i \neq y_{i+1}^{-1}$ for any $i = i, \dots, n-1$, i.e., w does not contain a subword of the type yy^{-1}, for a literal $y \in X^{\pm 1}$.

An *elementary reduction* of a word w consists of deleting a subword of the type yy^{-1} ($y \in X^{\pm 1}$) from w.

For example, suppose $w = uyy^{-1}v$ for some words u, v in $X^{\pm 1}$. Then the elementary reduction of w with respect to the given subword yy^{-1} results in the word uv. In this event we write

$$uyy^{-1}v \to uv.$$

A *reduction* of w (or a reduction process starting at w) consists of consequent applications of elementary reductions starting at w and ending at a reduced word:

$$w \to w_1 \to \cdots \to w_n, \quad (w_n \textit{ is reduced}).$$

The word w_n is termed a *reduced form* of w.

In general, there may be different possible reductions of w. Nevertheless, it turns out that all possible reductions of w end up with the same reduced form.

Proposition 1 *Let w be a group word in X. Then any two reductions of w:*

$$w \to w_1' \to \cdots \to w_n'$$

$$w \to w_1'' \to \cdots \to w_m''$$

result in the same reduced form, i. e. , $w_n' = w_m''$.

Proof. Induction on $|w|$.

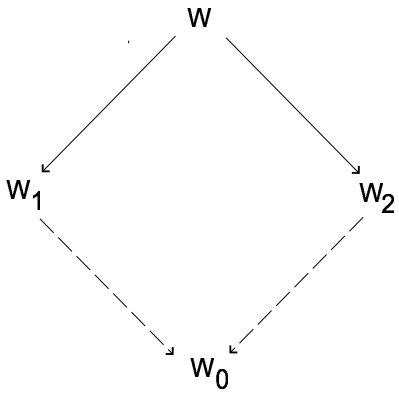
If $|w| = 1$ then w is reduced and there is nothing to prove.

Let

$$w \to w_1' \to \cdots \to w_n'$$
$$w \to w_1'' \to \cdots w_m''$$

be two reductions of w. We claim that to prove the proposition it suffices to prove the following result which is called the $\diamond$-*principal* or $\diamond$-*lemma*.

$\diamond$-Principal *For any two elementary reductions $w \to w_1$ and $w \to w_2$ there exist elementary reductions $w_1 \to w_0$ and $w_2 \to w_0$ to one and the same word w_0, i.e., the following diagram commutes:*



Proof. Indeed let

$$w \to w_1' \to \cdots \to w_n'$$
$$w \to w_1'' \to \cdots \to w_m''$$

be two arbitrary reductions of w. Then by the $\diamond$-principle we can find elementary reductions $w_1' \to w_0$ and $w_1'' \to w_0$ and then make a reduction of w_0 :

$$w_0 \to w_1 \to \cdots \to w_k.$$

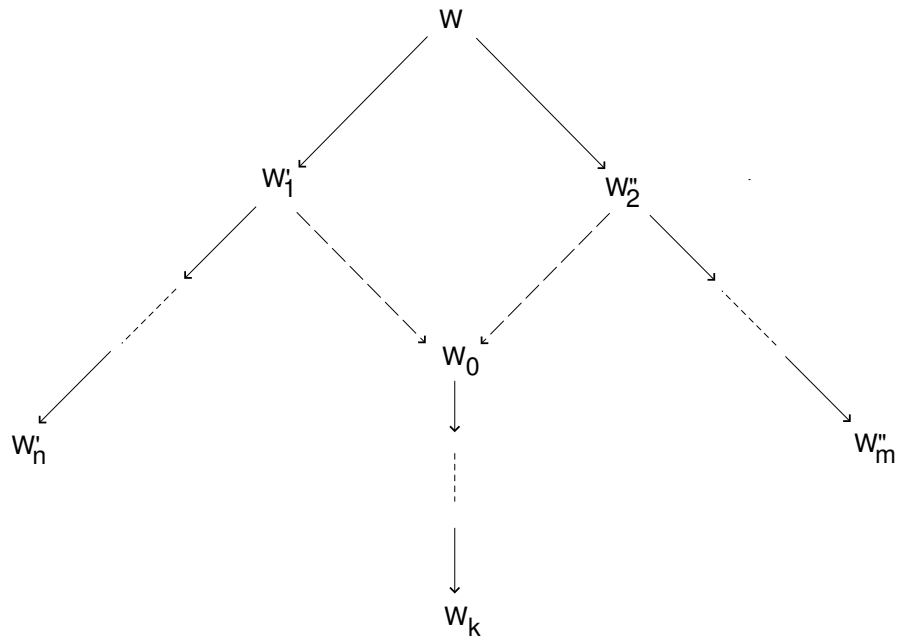
This corresponds to the following diagram:

By induction all reduced forms of the word w_1' are equal to each other, as well as all reduced forms of w_1''. Since w_k is a reduced form of both w_1' and w_1'', then $w_n' = w_k = w_m''$ as desired.

Now it is left to see that the $\diamond$-principal holds. Let $w \xrightarrow{\lambda_1} w_1$, and $w \xrightarrow{\lambda_2} w_2$ be elementary reductions of a word w. There are two possible ways to carry out the reductions λ_1 and λ_2.

Case a) (disjoint reductions). In this case

$$w = u_1 y_1 y_1^{-1} u_2 y_2 y_2^{-1} u_3, \quad (y_i \in X^{\pm 1})$$



and λ_i deletes the subword $y_i y_i^{-1}, i = 1, 2$. Then

$$w \overset{\lambda_1}{\to} u_1 u_2 y_2 y_2^{-1} u_3 \overset{\lambda_2}{\to} u_1 u_2 u_3$$

$$w \overset{\lambda_2}{\to} u_1 y_1 y_1^{-1} u_2 u_3 \overset{\lambda_1}{\to} u_1 u_2 u_3.$$

Hence the $\diamond$-principle holds.

Case b) (overlapping reductions). In this case $y_1 = y_2$ and w takes on the following form

$$w = u_1 y y^{-1} y u_2.$$

Then

$$w = u_1 y (y^{-1} y) u_2 \overset{\lambda_2}{\to} u_1 y u_2,$$

$$w = u_1 (y y^{-1}) y u_2 \overset{\lambda_1}{\to} u_1 y u_2;$$

and the principle holds. This proves the proposition.

For a word w by $\overline{w}$ we denote the unique reduced form of w.

Let $F(X)$ be the set of all reduced words in $X^{\pm 1}$. For $u, v \in F(X)$ we define multiplication $u \cdot v$ as follows:

$$u \cdot v = \overline{uv}.$$

Theorem 1 *The set $F(x)$ forms a group with respect to the multiplication $\cdot$. This group is free on X.*

Proof. The multiplication defined above is associative:

$$(u \cdot v) \cdot w = u \cdot (v \cdot w)$$

for any $u, v, w \in F(X)$. To see this it suffices to prove that

$$\overline{\overline{(uv)}w} = \overline{u\overline{(vw)}}$$

for given u, v, w. Observe, that each of the products above can be obtained form the word uvw by one of the possible reductions, which result in the same reduced form by Proposition 2. Hence

$$\overline{\overline{uv}w} = \overline{uvw} = \overline{u\overline{vw}}$$

Clearly, the empty word (we denote it by 1) is the identity in $F(X)$ with respect to the multiplication above, i.e.,

$$1 \cdot w = w \cdot 1,$$

for every $w \in F(X)$. Let $w = y_1 \cdots y_n$, $y_i \in X^{\pm 1}$. Then the word

$$w^{-1} = y_n^{-1} \cdots y_1^{-1}$$

is also reduced and

$$w \cdots w^{-1} = \overline{y_1 \cdots y_n y_n^{-1} \cdots y_1^{-1}} = 1.$$

Hence w^{-1} is the reverse of w. This shows that the set $F(X)$ satisfies all the axioms of a group.

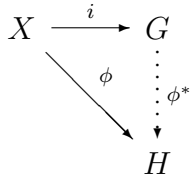
Notice that X is a generating set of $F(X)$. Moreover, every non-empty reduced word

$$w = x_{i_1}^{\varepsilon_1} \cdots x_{i_n}^{\varepsilon_n}$$

in $X^{\pm 1}$ defines a non-trivial element in $F(X)$ (the word w itself). Hence X is a free basis of $F(X)$, and $F(X)$ is free on X.

1.3 The universal property of free groups. Presentation of groups.

Theorem 2 *Let G be a group with a generating set $X \subseteq G$. Then G is free on X if and only if the following universal property holds. Every map $\varphi : X \to H$ from X into a group H can be extended to a unique homomorphism $\varphi^* : G \to H$, i. e., the following diagram commutes:*



(here $X \xrightarrow{i} G$ is the inclusion of X into G).

Proof. Let G be a free group freely generated by X. Suppose $\varphi : X \to H$ is an arbitrary map from X into a group H. Every element $g \in G$ is a product of elements from $X^{\pm 1}$, i.e., g corresponds to some group word w on X. Clearly the reduced form $\overline{w}$ of w defines the same element g in G. Now, we define the image of g under $\varphi*$ as follows. If

$$w = x_{i_1}^{\varepsilon_1} \cdots x_{i_n}^{\varepsilon_n}, \quad (x_{i_j} \in X, \varepsilon_i \in \{1, -1\}),$$

then

$$g^{\varphi*} = (x_{i_1}^{\varphi})^{\varepsilon_1} \dots (x_{i_n}^{\varphi})_{i_n}^{\varepsilon_n}. \tag{3}$$

Notice that equation (3) defines a map from G into H correctly. Indeed, since G is free on X then the reduced word $\overline{w}$ that defines g is unique.

We claim, that $\varphi*$ is a homomorphism. To see this choose arbitrary elements $g, h \in G$ and express them as reduced group words on X:

$$g = y_1 \cdots y_n z_1 \cdots z_m,$$

$$h = z_m^{-1} \cdots z_1^{-1} y_{n+1} \cdots y_k,$$

here $y_i, z_j \in X^{\pm 1}$ and $y_n \neq y_{n+1}^{-1}$ (we allow the subwords $y_1 \cdots y_n$, $z_1 \cdots z_m$, and $y_{n+1} \cdots y_k$ to be empty). Then

$$gh = y_1 \cdots y_n y_{n+1} \cdots y_k$$

is a reduced word on X presenting gh. Now

$$(gh)^{\varphi*} = y_1^{\varphi*} \dots y_n^{\varphi*} y_{n+1}^{\varphi*} \dots y_n^{\varphi*} =$$

$$= y_1^{\varphi*} \dots y_n^{\varphi*} z_1^{\varphi*} \dots z_m^{\varphi*} (z_m^{\varphi*})^{-1} \dots (z_1^{\varphi*})^{-1} \dots y_{n+1}^{\varphi*} \dots y_k^{\varphi*} = g^{\varphi*} h^{\varphi*}.$$

Hence $\varphi*$ is a homomorphism. Clearly, $\varphi*$ extends φ, so the universal property holds.

Suppose now that a group G with a generating set X satisfies the universal property. Take $H = F(X)$ and define a map $\varphi : X \to H$ as $x^{\varphi} = x$, $(x \in X)$. Then φ extends to a homomorphism $\varphi* : G \to F(X)$.

Let w be a non-empty reduced group word on X. Then w defines an element g in G for which $g^{\varphi} = w \in F(X)$. Hence $g^{\varphi^*} \neq 1$ and consequently $g \neq 1$ in G. It shows that G is a free group on X. This proves the theorem.

Observe, that the argument above implies the following result, which we state as a corollary.

Corollary 1 *Let G be a free group on X. Then the identical map $X \to X$ extends to an isomorphism $G \to F(X)$.*

This corollary allows us to identify a free group freely generated by X with the group $F(X)$. In what follows we usually refer to a free group on X as $F(X)$.

1.4 Rank of free groups

Theorem 3 *If G is free on X and also on Y, then $|X| = |Y|$.*

Proof. By the universal property of free groups any map $X \to \mathbf{Z}_2$ gives rise to a homomorphism of G into the cyclic group $\mathbf{Z}_2$ of order 2. Moreover, every homomorphism $G \to \mathbf{Z}_2$ can be obtained in this way (every homomorphism is completely defined by its values on a given generating set). Hence there are exactly $2^{|X|}$ different homomorphism form G into $\mathbf{Z}_2$. It implies that

$$2^{|X|} = 2^{|Y|}$$

and consequently $|X| = |Y|$. Here for infinite sets X and Y we are assuming Generalized Continuum Hypothesis from the set theory. But in the event when at least one of the sets X and Y is finite the implication

$$2^{|X|} = 2^{|Y|} \Longrightarrow |X| = |Y|$$

is obvious. This proves the theorem.

Corollary 2 *Let X and Y be sets. Then*

$$F(X) \simeq F(Y) \Leftrightarrow |X| = |Y|.$$

Theorem 3 shows that the cardinality of a basis of a free group G is an invariant of G, which is by the corollary above characterizes the free group G uniquely to an isomorphism.

Definition 2 *Let G be a free group on X. Then the cardinality of X is called the rank of G.*

For an ordinal n we denote by F_n a free group of rank n. By the remark above F_n is defined uniquely up to an isomorphism.

Notice that if $X \subseteq Y$ then the subgroup $< X >$ generated by X in $F(Y)$ is equal to the group $F(X)$, hence it is free on X. This implies that if m and n are cardinals and $n \leq m$, then F_n is embeddable into F_m.

We will show now that in some sense the reverse is also true. Namely we prove that any countable free group G is embeddable into a free group of rank 2. To this end it suffices to find a free subgroup of countable rank in a free group of rank 2.

Let F_2 be a free group with a basis $\{a, b\}$. Denote

$$x_n = b^{-n}ab^n \quad (n = 0, 1, 2, \ldots)$$

and put

$$X = \{x_0, x_1, \ldots\}.$$

We claim that X freely generates the subgroup $< X >$ in F_2. To see this it suffices to prove that every non-empty reduced group word on X defines a non-trivial element in F_2. To this end let

$$w = x_{i_1}^{\varepsilon_1} \ldots x_{i_n}^{\varepsilon_n}$$

be such a reduced word. Then w can also be viewed as a word in $\{a, b\}$:

$$w = b^{-i_1}a^{\varepsilon_1}b^{i_1}b^{-i_2}a^{\varepsilon_2}b^{i_2} \ldots b^{-i_n}a^{\varepsilon_n}b^{i_n}.$$

Since w is reduced on X, then for each $j = 1, \ldots, n-1$ either $i_j \neq i_{j+1}$ or $i_j = i_{j+1}$ and $\varepsilon_j + \varepsilon_{j+1} \neq 0$. In either case any reduction of w (as a word on $\{a, b\}$) does not affect a^{ε_j} and $a^{\varepsilon_{j+1}}$ in the subword

$$b^{-i_j}a^{\varepsilon_j}b^{i_j}b^{-i_{j+1}}a^{\varepsilon_{j+1}}b^{i_{j+1}};$$

i.e., the literals a^{ε_j} and $a^{\varepsilon_{j+1}}$ present in the reduced form of w as a word in $\{a, b\}^{\pm 1}$. Hence the reduced form of w is non-empty, so $w \neq 1$ in F_2. Clearly, $< X >$ is a free group of countable rank.

2 Elementary equations over free groups

2.1. General remarks.

2.l.1. Equations and some classical problems.

2.1.2. Cancellation schemes.

2.2. Equation $x^{-1}fx = g$.

2.2.1. Cyclically reduced words and conjugacy problem.

2.2.2. Solutions of the conjugacy equation $x^{-1}fx = g$.

2.3. Power equation $x^n = g$.

2.4. Equation $[x, y] = g$.

2.4.1. Homogeneous equation $[x, y] = 1$.

2.4.2. Centralizers in $F(X)$. Commutative-transitive groups.

2.4.3. Elementary transformations and minimal solutions.

2.4.4. Wick's theorem.

2.1 General remarks

2.1.1 Equations and some classical problems

Let G be a group and $X = \{x_1, \cdots, x_n\}$ be a set of variables.

An *equation* in variables $x_1, \cdots, x_n$ with coefficients in G is a formal expression of the form

$$g_1 {x_{i_1}}^{\varepsilon_1} g_2 {x_{i_2}}^{\varepsilon_2} \cdots {x_{i_k}}^{\varepsilon_k} g_n = 1 \tag{4}$$

where $g_j \in G,\ x_{i_j} \in X,\ \varepsilon_j \in \{1, 1\}$.

Notice, that we allow equations without coefficients (in which case $g_1 = \cdots = g_n = 1$). Such equations are also called *coefficient-free* equations. Usually, we write equations in functional notations:

$$f(x_1, \cdots, x_n,\ g_1, \cdots, g_n) = 1$$

or simply as

$$f(x_1, \cdots,\ x_n) = 1,$$

omitting coefficients.

Let

$$f(x_1, \cdots, x_n) = 1$$

be an equation in variables $x_1, \cdots, x_n$ with coefficients in G and let H be a group containing G as a subgroup. A tuple $(h_1, \ldots, h_n)$ of elements from H is called a *solution* of $f(x_1, \ldots, x_n) = 1$ in H if the element $f(h_1, \ldots, h_n)$ is trivial in H, i.e., the substitution $x_i \to h_i$ turns the formal expression 4 into equality in H.

By $V_H(f)$ we denote the solution set of $f(x_1, \ldots, x_n) = 1$ in H:

$$V_H(f) = \{(h_1, \ldots, h_n) \in H^n \mid f(h_1, \ldots, h_n) = 1\}.$$

There are two typical questions about a given equation $f(x_1, \ldots, x_u) = 1$: whether the equation has a solution in a given group H, and what are all possible solutions of this equation in H?

The following problem is one of the most intriguing which attracted a lot of attention over the years:

Diophantine Problem (DP) Let G be a group. Does there exist an algorithm which for any equation $f = 1$ with coefficients in G determines whether $f = 1$ has a solution in G or not.

If such an algorithm exists we call it a *decision algorithm* and say that the Diophatine problem is *decidable* over G.

Note. *A decision algorithm for DP takes equations* $f(x_1, \dots, x_n,\ g_1, \dots, g_n) = 1$ *as its inputs. Hence this equations have to be described in a constructive way. By definition these equations are group words in* X *and* G*. So it would suffices to describe effectively elements from* G*. Usually, we assume that the group* G *comes equipped with a given finite (sometimes countable) generating set* A*. In this event, every element from* G *can be represented by a word in* $A^{\pm 1}$*. From now on we will always (if not said otherwise) that elements of* G *are already given as words in* $A^{\pm 1}$*.*

G. Makanin proved that the DP is decidable over free groups [**?**]. In general DP is very difficult. In fact, decidability of the DP implies the decidability of the word problem and the conjugacy problem in G. To explain this let $S = 1$ be an arbitrary collection of equations over group G. We say that DP is *decidable for* S *over* G, if there exists an algorithm which for any equation $f = 1$ from $S = 1$ decides whether $f = 1$ has a solution in G or not.

The Word Problem in G (WP) Let $A = \{a_1, \dots, a_n\}$ be a finite generating set for G. The word problem is decidable in G if the DP is decidable for the set of equations

$$S = \{{a_{i_1}}^{\varepsilon_1} \cdots {a_{i_n}}^{\varepsilon_n} = 1 \mid a_{i_j} \in A, \varepsilon_j \in \{1, -1\}, n \in N\}.$$

The Conjugacy Problem in G (CP) The conjugacy problem is decidable in G if the DP is decidable for the collection of equations

$$S = \{x^{-1}gx = f \ \mid \ g, f \in G\}.$$

We mention here few more algorithmic problems in groups.

The Power Problem in G (PP) The power problem is decidable in G if the DP is decidable for the collection of equations

$$S = \{x^n = f \ \mid \ n \in \mathbf{N},\ f \in G\}.$$

The Commutator Problem in G (ComP) The commutator problem is decidable in G if the DP is decidable for the collection of equations

$$S = \{[x, y] = f \ \mid \ f \in G\}.$$

2.1.2 Cancellation schemes

Let $F(A)$ be a free group on $A = \{a_1, \dots, a_n\}$.

Suppose $u, v \in F(A)$ and the word uv is reduced as written, i.e., there is no cancellation in uv. In this event we denote the product uv by $u \circ v$. Plainly,

$$uv \equiv u \circ v \Longleftrightarrow |uv| = |u| + |v|$$

(here $x \equiv y$ means that the words x and y are equal words in the alphabet $A^{\pm 1}$).

Suppose now that

$$u = u_1 \circ p, v = p^{-1} \circ v_1.$$

Then

$$uv = u_1(pp^{-1})v_1 = u_1 v_1.$$

In this case we say that the subwords p and p^{-1} *cancel out* in forming the product uv. Graphically this can be expressed as

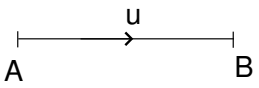
Notice, that we are not able to recover the cancelled subwords p, p^{-1} from the resulting word $u_1 v_1$. To keep the history of cancellation we will sometimes write

$$u_1(pp^{-1})v_1$$

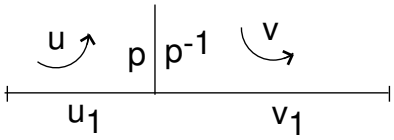
putting parentheses to mark the cancellation.

The graphical equivalent of this will be to "pinch" p and p^{-1} together and write the result in the following form:

To simplify the notations we will use the following agreement. The diagram



means that the edge from A to B has the word u as a label, and the inverse edge from B to A has a label u^{-1} (i.e. we read u going from A to B). Hence the cancellation above can be expressed as



Now, let $w_1, \ldots, w_n$ be reduced words in $A^{\pm 1}$. Then, in general, we have several different ways to reduce the word

$$w \equiv w_1 w_2 \cdots w_n$$

into its reduced form. Each such way of reduction can be completely described by consequently putting parentheses to mark each step of the reduction (cancellation). For example, let

$$w_1 \equiv y_1 y_2 y_3 y_1, \quad w_2 \equiv y_1^{-1} y_3^{-1} y_2^{-1} y_4, \quad w_3 \equiv y_4^{-1} y_2 y_1.$$

Then

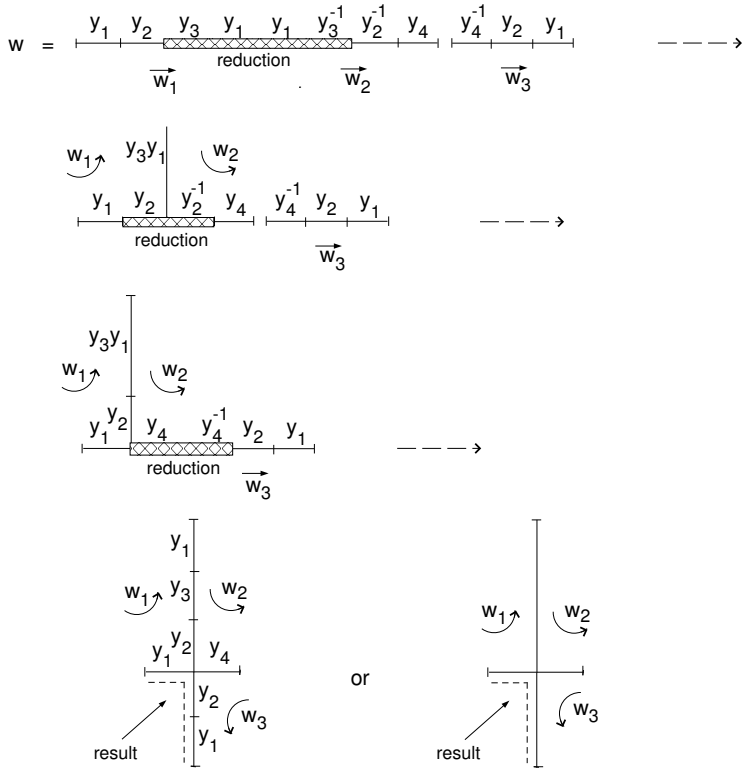
$$w = \underbrace{y_1 y_2 y_3 y_1}_{w_1} \underbrace{y_1^{-1} y_3^{-1} y_2^{-1} y_4}_{w_2} \underbrace{y_4^{-1} y_2 y_1}_{w_3}$$

and we have the following two different reduction processes of w:

$$w = \underbrace{y_1 y_2 (y_3 y_1}_{w_1} \underbrace{y_1^{-1} y_3^{-1})(y_2^{-1} y_4}_{w_2} \underbrace{y_4^{-1} y_2) y_1}_{w_3} = y_1 y_2 y_1;$$

$$w = \underbrace{y_1 (y_2 (y_3 y_1}_{w_1} \underbrace{y_1^{-1} y_3^{-1}) y_2^{-1})(y_4}_{w_2} \underbrace{y_4^{-1}) y_2 y_1}_{w_3} = y_1 y_2 y_1.$$

Graphically, we can express these reductions as the following labeled trees:



Such words with parentheses or such labeled trees are called *cancellation schemes* for the word $w \equiv w_1 \cdots w_n$.

Lemma 1 *For any positive integer n there are finitely many trees $T_1, \ldots, T_{\mu(u)}$ such that for any n-tuple of words $w_1, \ldots, w_n \in F(A)$ each cancellation scheme for $w \equiv w_1 \ldots w_n$ can be obtained by labelling one of the trees $T_1, \ldots, T_{\mu(u)}$ by elements from $F(A)$.*

Proof. It suffices to notice that each cancellation tree for $w \equiv w_1 \ldots w_n$ has no more than $2n$ vertices.

2.2 Equation $x^{-1}fx = g$

2.2.1 Cyclically reduced words and conjugacy problem

Let $F(A)$ be a free group with basis $A = \{a_1, \ldots, a_n\}$.

Definition 3 *A reduced word $w = y_1 \cdots y_m$, $(y_i \in A^{\pm 1})$ is called cyclically reduced if $y_1 \neq y_m^{-1}$.*

For example, the word ${a_1}^{-1}a_2a_2$ is cyclically reduced, but the word ${a_1}^{-1}a_2a_1$ is not.

Lemma 2 *Let w be a reduced word from $F(A)$. Then there exists a cyclically reduced word $\overline{w}$ and a reduced word w_1 such that*

$$w = {w_1}^{-1} \circ \overline{w} \circ w_1. \tag{5}$$

Moreover, such words $\overline{w}$ and w_1 are unique.

Proof. We will read of $\overline{w}$ and w_1 from w using the following marking process. Let $w = y_1 \cdots y_m,\ (y_i \in A^{\pm 1})$ be a reduced word. Compare the initial literal y_1 with the final literal y_m. If $y_1 = {y_m}^{-1}$ then mark them both, and repeat the process for the unmarked subword $y_2 \cdots y_{m-1}$. In the case $y_1 \neq {y_m}^{-1}$ we stop. Clearly, we will finish in $\leq |w|$ steps, and the unmarked subword is cyclically reduced, we denote it by $\overline{w}$. Meanwhile, the longest marked subword at the end of w is w_1. From the discussion above it follows that

$$w = {w_1}^{-1} \circ \overline{w} \circ w_1$$

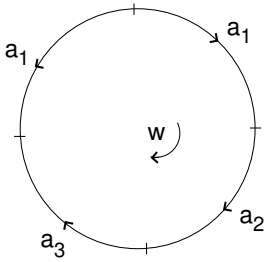
and the words $\overline{w}, w_1$ are uniquely determined by w. This proves the lemma.

Definition 4 *Let w be an element of $F(A)$. Then the above decomposition*

$$w = {w_1}^{-1} \circ \overline{w} \circ w_1$$

is called the cyclic decomposition of w, and the word $\overline{w}$ is called the cyclically reduced form of w.

The cyclically reduced form of w has a simple visual interpretation. Every word $w \in F(A)$ defines a *cyclic word* w, which is the word w written on a circle in the clockwise orientation. For example, $w = a_1a_2a_3{a_1}^{-1}$ defines the cyclic word



A *reduced cyclic word* is a cyclic word without neighbors of the type yy^{-1} ($y \in A^{\pm 1}$).

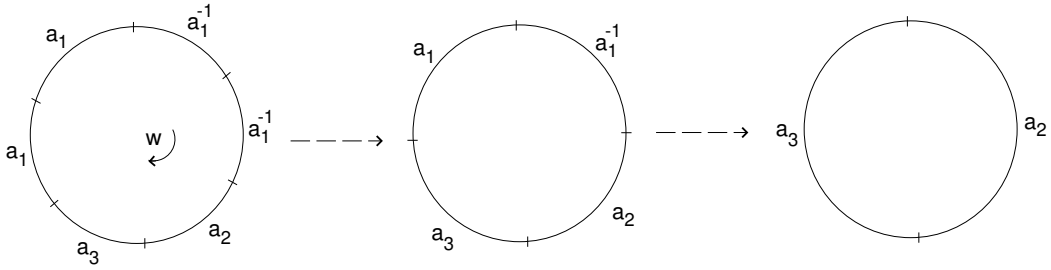
Every cyclic word can be transformed into a unique reduced cyclic word by means of elementary reductions exactly in he same way as for the ordinary words (see Section 1.2). Notice, that if w is a reduced word, then the reduction process for the cyclic word w is uniquely determined at each step, i.e., at each step there is only one possible elementary reduction to perform. For example, if

$$w = {a_1}^{-2} a_2 a_3 {a_1}^2$$

then the reduction process for w can be described as follows.

$$w = a_1^{-2} a_2 a_3 a_1^2 \ \rightarrow \ a_1^{-1} a_2 a_3 a_1 \ \rightarrow \ a_2 a_3$$

or graphically:



This reduction procedure reflects the marking process from Lemma 2 (only instead of marking literals we delete them from the cyclic word). Therefore, the reduction of a cyclic word w results in the cyclic word $\overline{w}$. Hence, the cyclically reduced form of a word w is equal (as a cyclic word) to the reduction of a cyclic word w.

Starting with a word w we can form the corresponding cyclic word defined by w and then cut it back into a (linear) word. The resulting word depends on the cutting point and might be different from the initial word w. To describe all words that can occur here we need the following definition.

Definition 5 *Let $w = y_1 \cdots y_m$ ($y_i \in A^{\pm 1}$) be a word in $A^{\pm 1}$. A cyclic permutation of w is a word of the form $y_{i+1} \cdots y_m y_1 \cdots y_i$ ($1 \leq i \leq n$).*

Clearly, two cyclic words defined by $u, v \in F(A)$ are equal if and only if u is a cyclic permutation of v (and hence, v is a cyclic permutation of u). Observe also, if $v = y_{i+1} \ldots y_m y_1 \ldots y_i$ is a cyclic permutation of $w = y_1 \cdots y_m$ then

$$(y_1 \cdots y_i)^{-1} \ w \ (y_1 \ldots y_i) = v.$$

So any cyclic permutation of w is a conjugate of w. Summarizing the discussion above we have the following

Proposition 2 *Let u, v be elements of $F(A)$. The following conditions are equivalent:*

1. *u and v are conjugate in $F(A)$;*
2. *the cyclic words defined by u and v are equal;*
3. *$\overline{u}$ is a cyclic permutation of $\overline{v}$.*

Corollary 3 *The conjugacy problem in $F(A)$ is decidable.*

Proof. Indeed, to check whether given elements $u, v \in F(A)$ are conjugate in F or not it suffices to find their cyclically reduced forms $\overline{u}$ and $\overline{v}$, and check whether the cyclic words defined by $\overline{u}$ and $\overline{v}$ are equal or not. This can be done effectively.

Notice, that the algorithm indicated above is a fast one, it has at most a quadratic time complexity. For it takes $O(|u| + |v|)$ steps to obtain the reduced cyclic words defined by u and v, and $O((|u| + |v|)^2)$ steps to verify whether these cyclic words are equal or not.

2.2.2 Solutions of the conjugacy equation $x^{-1}ux = v$

We have already seen in the previous section how one can effectively verify whether or not the equation

$$x^{-1}ux = v \tag{6}$$

has a solution in $F(A)$.

A similar method allows one also to find a particular solution (if it exists) of this equation. Indeed, suppose $x^{-1}ux = v$ for some $x \in F(A)$. Let

$$u = {u_1}^{-1} \circ \overline{u} \circ u_1, \quad v = {v_1}^{-1} \circ \overline{v} \circ v_1$$

be the cyclic decompositions of u and v. Then

$$x^{-1}{u_1}^{-1}\overline{u}u_1 x = v^{-1}\overline{v}v_1$$

and

$$v_1 x^{-1}{u_1}^{-1}\overline{u}u_1 x {v_1}^{-1} = \overline{v}.$$

Hence $\overline{v}$ must be a cyclic permutation of $\overline{u}$ (otherwise, (6) has no solutions in $F(A)$). It follows that some initial segment of $\overline{u}$ (viewed as a word in generators $A^{\pm 1}$), say u_2, conjugates $\overline{u}$ into $\overline{v}$. Hence

$$v = v_1^{-1} u_2^{-1} \overline{u} u_2 v_1$$

and the element $x = u_1^{-1} u_2 v_1$ is a solution of the equation (6) (straightforward verification).

Clearly, there exists an effective algorithm to find a solution of the type $x = u_1^{-1} u_2 v_1$ from above (check one by one all initial segments u_2 of the word $\overline{u}$).

To describe all solutions of the equation (6) we rewrite it in the following form

$$[u, x] = g, \quad (\textit{here } g = u^{-1} v). \tag{7}$$

This is a particular type of a *commutator equation* which we discuss in Section 3.4. Notice that if $x = b$ and $x = c$ are solutions of the equation 7 then $[u, b] = [u, c]$ which implies that $[u, bc^{-1}] = 1$. Hence, if b is a particular solution of (7), then all other solutions of (7) can be described as

$$x = y^{-1} b,$$

where y is a solution of the homogeneous equation

$$[u, x] = 1. \tag{8}$$

We will see in Section 4.1 how to describe all solutions of this homogeneous equation.

Note. *Observe, that the description above reminds remotely the description of solution sets of liner systems of equations over an arbitrary ring or over an abelian group. Unfortunately, for equations over non-abelian groups, as we will see later, this situation is rather an exceptional one.*

Question: can be similar methods applied for solving equations of the type $[x, u_1, \ldots, u_n] = v \quad (u_i, v \in F(A))$?

Here we define the commutator $[x, u_1, \ldots, u_n]$ inductively as follows:

$$[x, u_1, \ldots, u_n] = [[x, u_1, \ldots, u_{n-1}], u_n].$$

2.3 Equation $x^n = g$

For an element $g \in F(A)$ consider the power equation

$$x^n = g \quad (n \in \mathbf{N}) \tag{9}$$

Lemma 3 *The power equation $x^n = g$ has at most one solution in $F(A)$.*

Proof. Suppose $u^n = g$ for some $u \in F(A)$. Let

$$u = {u_1}^{-1} \circ (\overline{u}) \circ u$$

be the cyclic decomposition of u. Then

$$u^n = {u_1}^{-1} \circ (\overline{u})^n \circ u_1$$

is the cyclic decomposition of

$$g = u^n.$$

Hence

$$u_1 = g_1, (\overline{u})^n = (\overline{g}).$$

It follows that 9 has a solution in $F(A)$ if and only if, firstly, n divides $|(\overline{g})|$ and , secondly, if

$$(\overline{g}) = f_1 \circ f_2 \circ \cdots \circ f_n$$

is the partition of $(\overline{g})$ into n subwords of equal length, then

$$f_1 = f_2 \cdots = f_n$$

In this event

$$x = {g^{-1}}_1 f_1 g_1$$

is the only solution of 9 in $F(A)$.

Notice, that the argument above shows that for $n \geq |g|$ the equation $x^n = g$ has no solution in $F(A)$. Another remark is that if the power equation has a solution in $F(A)$, one can effectively find this solution in $O(|g|)$ steps.

Definition 6 *The Power Problem is decidable in a group G if there exists an algorithm which determines whether a power equation $x^n = g (g \in G, n \in N)$ has a solution in G or not.*

Corollary 4 *The Power Problem is decidable in a free group.*

Notice, that there are finitely presented groups in which the power problem is undecidable.

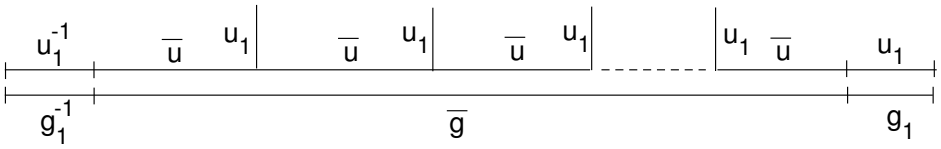
The uniqueness of solutions of power equations in $F(A)$ allows one to define n-th root of elements of $F(A)$.

Definition 7 *The solution of the power equation $x^n = g$ is called the n-th root of g in $F(A)$.*

The *maximal root* of g is the n-th root of g with n maximal possible.

Observe that the maximal root of a non-trivial element on $g \in F(A)$ exists and it is unique.

The following picture (a cancellation scheme) explains geometrically how to find a solution of a power equation $x^n = g$. If $x = u$ is a solution of this equation then we have the following cancellation scheme:



2.4 Equation $[x, y] = g$

2.4.1 Homogenous equation $[x, y] = 1$

We start with the classical approach using induction on the total length of solutions. Then we demonstrate how to solve the equation using cancellation schemes.

Lemma 4 *Let $u, v \in F(A)$. If $uv = vu$ then there exists $w \in F(A)$ and integers m, n such that*

$$u = w^m, v = w^n$$

$\triangleright$ Induction on $|u| + |v|$.

If $|u| + |v| = 1$, then $u = 1$ or $v = 1$ and in this case the conclusion of the Lemma is obvious.Suppose now that $|u|, |v| \geq 1$ and $|u| \leq |v|$.

Case 1. There is no cancellation in uv i.e., $|uv| = |u| + |v|$. Then $|vu| = |uv| = |u| + |v|$ and there is no cancellation in vu. Hence vu and uv are reduced. Now from $uv = vu$ and $|u| \leq |v|$ we deduce that $v = u \circ v_1$. Therefore $uv_1 = v_1u$ and $|v_1| < |v|$. By induction u and v_1 (hence u and v) are powers of some $w \in F(A)$, as desired.

Case 2. Suppose u cancels completely in v. Then $v = u^{-1}v_1$ and the equality $uv = vu$ takes the form $uu^{-1}v_1 = u^{-1}v_1u$. The latter can be written as

$$uv_1 = v_1u.$$

Since $|v_1| < |v|$ we can proceed by induction.

Case 3. There is a cancellation in uv, but u does not cancel completely in uv. In this event

$$u = u_1 \circ y, v = y^{-1} \circ v_1, y \in A^{\pm 1}.$$

Then $uv = vu$ can be rewritten as

$$u_1 v_1 = y^{-1} v_1 u_1 y \tag{10}$$

Since v does not cancel completely in uv, then vu begins with y^{-1}. Similarly, u_1 does not cancel completely in $u_1 v_1$ (otherwise u cancels out in uv), therefore u_1 begins with y^{-1}. Hence $u_1 = y^{-1} \circ u_2 \circ y, v_1 = y^{-1} \circ v_1$ and it follows from (10) that

$$u_2 v_1 = v_1 y^{-1} u_2 y \tag{11}$$

Notice (comparing lengths) that it must be a cancellation on the right in 11 and there is only one possible place to have the cancellation, namely, $v_1 = v_2 \circ y$. In this event, $u = y^{-1} \circ u_2 \circ y, v = y^{-1} \circ v_2 \circ y$ equality $uv = vu$ takes the form

$$u_2 v_2 = v_2 u_2$$

and the conclusion of the lemma follow by induction.

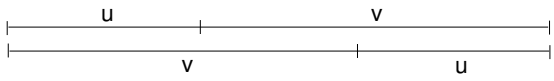
Corollary 5 *The set* $\{(w^n, w^m) | w \in F(A), n, m \in \mathbf{Z}\}$ *is the solution set of the equation* $[x, y] = 1$.

Now we illustrate the proof above by cancellation schemes.

Let $u, v \in F(A)$ and $[u, v] = 1$.

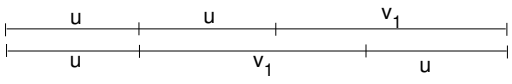
Consider all possible cancellation schemes in the product $uvu^{-1}v^{-1}$.

Case 1. No cancellation in uv . then the corresponding scheme is the following:

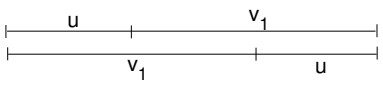


(we assume here that $|u| \leq |v|$).

Clearly, the point A divides v into two parts u and v_1 ($v = u \circ v_1$). Draw u on v and on the other occurrences of u in the scheme:



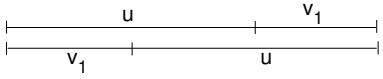
and cut out the common initial segment u from the scheme, the resulting scheme will be of the following type:



It is exactly the initial scheme provided $|u| \leq |v_1|$. We can repeat the process until this is possible and in finitely many steps (say, m_1 steps) we will have:

$$\begin{cases} v = u^{m_1} v_1, \ |u| > |v_1| \\ u = u \end{cases}$$

The resulting scheme will be like this:

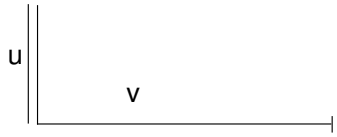


Now we see that $u = v_1 \circ u_1$ and we can again cut out the common initial segment v_1 from the scheme and proceed to do so until this is possible, i.e., until $|u_1|| \prec |v_1|$. In finitely many steps, say K_1 steps, we will have

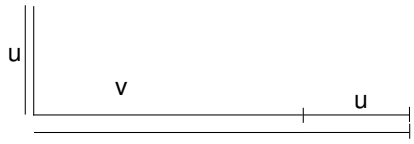
$$\begin{cases} v = u^{m_1} \circ v_1, \ |u_1| \leq |v_1| \\ u = v_1^{K_1} \circ u_1 \end{cases}$$

Now again we have the initial cancellation scheme and we can repeat the process again. Since, we cannot decrease the length of $|u|$ or $|v|$ forever, the process will stop eventually. At this point we will have $u_1 = 1$ or $v_1 = 1$. In any event, u and v will be powers of either u_1 or v_1.

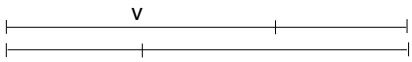
Case 2. u cancels out in uv.



Then there exists only one possible scheme:



If we draw it in the following way



then we can see that the process from Case 1 will work in this case also.

Case 3. There is a cancellation in uv. Then there is a cancellation in vu and the corresponding scheme is as follows:

Let a be the first literal in u, i.e., $u = a \circ u_1$. Draw a in the scheme

Then we see that $u = a \circ u_2 \circ a^{-1}, v = a \circ v_2 \circ a^{-1}$.

Clearly, we can cut out this segment a everywhere in the scheme and the resulting scheme will be the cancellation scheme for $[u_2, v_2] = 1$. It is easy to see that in finitely many steps we will cut out either the whole initial segment up to the point A, or we will cut out both pinches. In either case we wil get the scheme from case 1) or case 2).

2.4.2 Centralizers in $F(A)$. Commutative-transitive groups

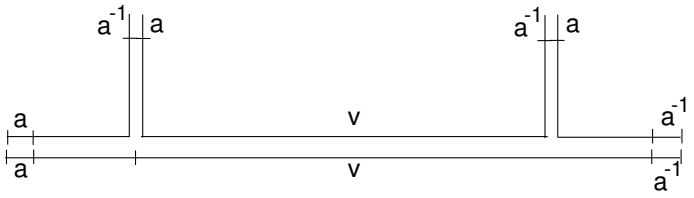
Definition 8 *Let G be a group and let M be a subset of G. Then the set*

$$C_G(M) = \{g \in G | [g, m] = 1 \ \forall m \in M\}$$

is called the centralizer of M in G.

It is easy to see that $C_G(M)$ is always a subgroup of G.

Proposition 3 *Let F be a free group and M be a subset of F. If $M \neq \{1\}$, then the centralizer $C_F(M)$ is cyclic.*



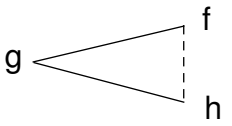
$\triangleright$. If $g \in M$, then $C_F(M) \subseteq C_F(g)$. Therefore to prove the proposition it suffices to prove that the centralizer $C_F(g)$ of any non-trivial element $g \in F$ is cyclic. Let q_0 be the maximal root of g in F, then $g = {q_0}^k$ for some positive k. We claim that for any $f \in F$ if $[{q_0}^n, f] = 1$, then $[q_0, f] = 1$. Indeed, if ${q_0}^n f = f{q_0}^n$, then

$$ {q_0}^k = f{q_0}^k f^{-1} = (fq_0f^{-1})^k $$

since k-roots of elements in F are unique (see section 2.3), then $q_0 = fq_0f^{-1}$ i.e., $[q_0, f] = 1$. It follows that $C_F(g) = C_F(q_0)$. Now if $f \in C_F(q_0)$, then by Lemma 4 q_0 and f are powers of some element $w \in F$. Since q_0 is not a proper power, then $q_0 = w^{\pm 1}$ and f is a power of q_0. Consequently, the maximal root q_0 is a generator of the centralizer $C_F(g)$.

Definition 9 *A group G is called commutative-transitive if the commutative is an equivalence relation on $G - \{1\}$, i.e., if $[g, f] = 1$ and $[g, h] = 1$ then $[f, h] = 1$ provided f, g, h are non-trivial elements of G.*

The commutative graph of G consists of elements from $G - \{1\}$ as vertices of this graph and two vertices $q_1 h$ are connected by an edge if and only if $[g, h] = 1$. So whenever we have



then there exists an edge between f and h. Now it is clear, that a group G is commutative-transitive if in its commutativity each connected component is a complete graph (i.e., any two vertices are connected by an edge).

Now the following lemma is obvious.

Lemma 5 *A group G is commutative-transitive if all proper centralizers of G are commutative.*

Corollary 6 *Every free group is commutative-transitive.*

Notice, that commutative transitive groups have very simple structure of centralizers: any two proper centralizers in such a group either are equal or intersect in identity.

Notice also, that many non-free groups are commutative-transitive. For example, proper centralizers in a torsion-free hyperbolic group are cyclic, hence each such group is commutative-transitive.

2.4.3 Wick's theorem

Let G be a group and g be an element of G. The equation

$$[x, y] = g$$

has a solution in G if and only if g is a commutator in G. It follows that the Diophantine problem is decidable in the class of equations

$$\{[x, y] = g | g \in G\}$$

if and only if there exists an effective procedure for recognizing commutators in G.

The following result shows that there exists an algorithm for recognizing commutators in free groups.

Theorem 4 *(Wick's). Let F be a free group, g an element in F, and $(\overline{g})$ the cyclically reduced form of g. Then g is a commutator in F if and only if some cyclic permutation of $(\overline{g})$ is of the form*

$$a \circ b \circ c \circ a^{-1} \circ b^{-1} \circ c^{-1} \tag{12}$$

for some $a, b, c \in F$.

$\triangleright$. Notice, that since g and $(\overline{g})$ are conjugate then if one of them is a commutator, then the other one also does.

Obviously, every commutator $u^{-1}v^{-1}uv$ can be presented as a product $abca^{-1}b^{-1}c^{-1}$ (say $a = u^{-1}, b = v^{-1}, c = 1$) in which it may be some cancellation. And vice versa, for any $a, b, c \in F$ we have

$$abca^{-1}b^{-1}c^{-1} = (ab)(ca^{-1})(ab)^{-1}(ca^{-1})^{-1}$$

hence $abca^{-1}b^{-1}c^{-1}$ is a commutator.

Now suppose g is a cyclically reduced commutator. As we have mentioned above any cyclic permutation of g can be presented in the form $abca^{-1}b^{-1}c^{-1}$. Among all such presentations chose one with the minimal total length $|a| +$

$|b| + |c|$. We claim, that in this event there is no cancellation in the product $abca^{-1}b^{-1}c^{-1}$. To prove this it suffices to consider all possible cases where cancellation can occur.

Suppose $b = b_1 \circ y, c = y^{-1} \circ c_1$.

Then

$$abca^{-1}b^{-1}c^{-1} = ab_1c_1a^{-1}y^{-1}{b_1}^{-1}{c_1}^{-1}y$$

Conjugating by a, we obtain the following cyclic permutation of g:

$$b_1c_1a^{-1}y^{-1}{b_1}^{-1}{c_1}^{-1}ya = b_1c_1(a^{-1}y^{-1})b_1^{-1}c_1^{-1}(a^{-1}y^{-1})^{-1}$$

with total length $|b_1| + |c_1| + |a| + 1 < |a| + |b| + |c|$-contradiction. Similar argument shows that cancellation does not occur in all places.

Corollary 7 *Commutators are effectively recognizable in free groups.*

Indeed, let g be an element of a free group F. We can effectively find the cyclically form $(\overline{g})$ of g. If $|(\overline{g})|$ is of odd length, then $(\overline{g})$ (as well as g) is not a commutator. If $|(\overline{g})|$ is of even length, then we can take the left half of $(\overline{g})$ and partition in into all possible products of the type $a \circ b \circ c$.

Now for every such partition abc we check whether the right half of $(\overline{g})$ is equal to $a^{-1}b^{-1}c^{-1}$ or not. If yes, then g is a commutator, if not, then we take one by one all cyclic permutations of $(\overline{g})$ and repeat the process for each of them.

What is the complexity of the algorithm described above?

For a given g we can find in linear time (on $(\overline{g})$) the cyclically reduced form $(\overline{g})$ of g. Now, if $(\overline{g})$ is of even length, then we have $\leq 1/2|(\overline{g})|$ choices for a, b and c. Hence, altogether we have to check $\leq c \cdot |g|^3$ combinations of a, b and c. The equality problem $(\overline{g}) = a \circ b \circ c \circ a^{-1} \circ b^{-1} \circ c^{-1}$ is of linear time on $|(\overline{g})|$. Thus the complexity of the algorithm above is at most $c|g|^3$ in time.

2.4.4 Elementary transformations and minimal solutions

The following lemma shows how one can obtain infinitely many solutions of the equation $[x, y] = g$ from a given particular solution $x = u, y = v$.

Lemma 6 *Let (u, v) be a solution of an equation*

$$[x, y] = g.$$

Then pairs

$$(v^m u, v) \text{ and } (u, u^m v), m \in \mathbf{Z}$$

are also solutions of this equation.

$\triangleright$. This follows immediately from the following identities:

$$[v^m u, v] = [v^m, v]^u [u, v] = [u, v] = g;$$

$$[u, u^m v] = [u, v][u, u^m]^v = [u, v] = g.$$

So starting from a solution (u, v) we can generate infinitely many other solutions:

$$(u, v) \to (v^{m_1} u, v) \to (v^{m_1} u, (v^{m_1} u)^{m_2} v) \to (((v^{m_1} u)^{m_2} v)^{m_3}, (v^{m_1} u)^{m_2} v) \to \dots$$

Now we introduce the following elementary transformations on pairs of elements from F:

$$U : (u, v) \to (vu, v)$$

$$V : (u, v) \to (u, uv).$$

Clearly, this transformation U, V has inverses

$$U^{-1} : (u, v) \to (v^{-1} u, v)$$

$$V^{-1} : (u, v) \to (u, u^{-1} v).$$

By U^m, V^m we denote the product of the m consequent transformation of the type U, V correspondingly.

We will say that two pairs (u_1, v_1) and (u_2, v_2) are equivalent ($(u_1, v_1) \sim (u_2, v_2)$) if one can transform (u_1, v_1) into (u_2, v_2) by a finite chain of elementary transformations. Clearly, $\sim$ is an equivalence relation on pairs

By Lemma 6 if (u, v) is a solution of $[x, y] = g$, then the equivalence class $[(u, v)]$ provides infinitely many other solutions of this equation. It follows that the solution set $V([x, y] = g)$ of the equation is a union of equivalence classes:

$$V([x, y] = g) = \cup_{i \in I} [(u_1, v_i)].$$

Our goal now is to prove that one can effectively find a finite collection of solutions $(u_1, v_1), \dots, (u_n, v_n)$ such that

$$V([x, y] = g) = [(u_1, v_1)] \cup \dots \cup [(u_n, v_n)].$$

To this end we need the following

Definition 10 *A solution (u, v) of the equation $[x, y] = g$ is called minimal if it has the minimal total length $|u| + |v|$ among all solutions in its equivalence class $[(u, v)]$.*

Lemma 7 *Let (u, v) be a minimal solution of the equation $[x, y] = g$. Then in each of the products*

$$uv, v^{-1}u, u^{-1}v^{-1}$$

the total cancellation can not be greater than

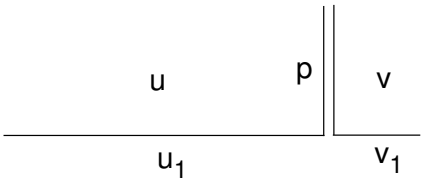
$$\min\{1/2|u|, 1/2|v|\}$$

Proof. Observe, that for arbitrary $u, v \in F(A)$ the following conditions are equivalent:

1) more than half of v cancels out in uv;

2) $|uv| < |u|$

This can be seen from the corresponding cancellation scheme:



Indeed,

$$|uv| = |u_1| + |v_1| < |u| \Leftrightarrow |v_1| < |p|.$$

Now, let (u, v) be a minimal pair.

Case 1. Consider the product uv.

Suppose, that more than half of v cancels out in uv. Then $|uv| < |u|$. In this case

$$(u, v) \sim (u, uv) \sim (v^{-1}, uv)$$

and $|v^{-1}| + |uv| = |v| + |uv| < |v| + |u|$ - contradiction with the minimality of (u, v).

Suppose, that more then half of u cancels out in uv. Then $|uv| < |v|$ and the equivalence

$$(u, v) \sim (u, uv)$$

shows that (u, v) is not minimal ($|u| + |uv| < |u| + |v|$).

Case 2. Suppose more than half of u cancels out in the product $v^{-1}u$. Then $|v^{-1}| < |v^{-1}| = |v|$. In this case

$$(u, v) \sim (u, uv)$$

and $|u| + |uv| < |u| + |v|$ - contradiction.

Suppose that more than half of v cancels out in $v^{-1}u$. Then

$$|v^{-1}u| = |u|.$$

Hence,

$$(u, v) \sim (v^{-1}u, v)$$

and $|v^{-1}u| + |v| < |u| + |v|$ - contradiction.

Case 3. $u^{-1}v^{-1}$.

If more than half of u^{-1} cancels out in $u^{-1}v^{-1}$, then $|u^{-1}v^{-1}| = |vu| < |v|$. Hence

$$(u, v) \sim (vu, v) \sim (vu, u^{-1})$$

and $|vu| + |u^{-1}| < |v| + |u|$ contradiction. If more than half of v^{-1} cancels out in $u^{-1}v^{-1}$ then $|u^{-1}v^{-1}| < |u^{-1}| = |u|$. Now

$$(u, v) \sim (vu, v)$$

and $|vu| + |v| = |u^{-1}v^{-1}| + |v|$ - contradiction.

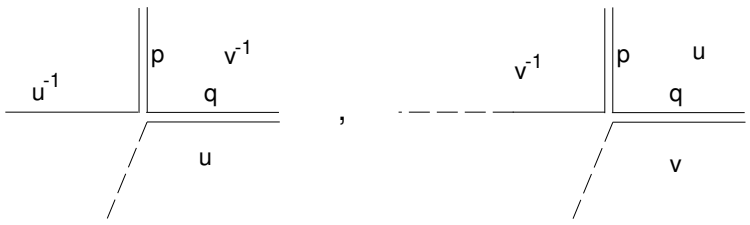
Actually, it suffices to consider just the first case, two others are similar.

Corollary 8 *Let (u, v) be a minimal solution of the equation $[x, y] = g, (g \neq 1)$. Then at least one literal from each of the factors u^{-1}, v^{-1}, u, v occurs in the reduced form of $u^{-1}v^{-1}uv$.*

$\triangleright$. By the lemma above at most half of each factor can cancel out in the products $u^{-1}v^{-1}, v^{-1}u, uv$. Hence it has odd length, then the literal in the middle of u does not cancel out in the product $v^{-1}uv$. Similarly, if v^{-1} is of odd length, then it does not cancel completely in $u^{-1}v^{-1}u$.

Suppose now that u is of even length, $u = pq, |p| = |q|$. If u cancels out completely in $v^{-1}uv$, then $p = q^{-1}$ and $u = 1$. This implies $g = 1$-contradiction. Hence u does not cancel completely in $v^{-1}uv$, as well as v^{-1} in $u^{-1}v^{-1}u$.

It follows that reducing v^{-1} in $u^{-1}v^{-1}uv$ we do not have cancellation schemes of the types:



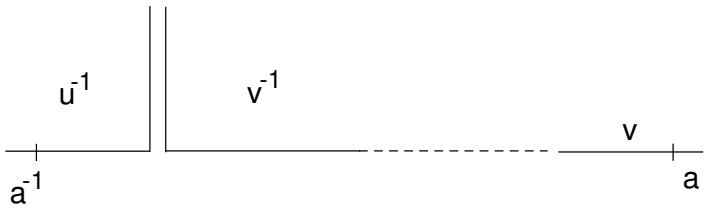
Hence, neither of the factors u^{-1}, v^{-1}, u, v cancels out completely in $u^{-1}v^{-1}uv$.

Lemma 8 *If (u, v) is a minimal solution of the equation $[x, y] = g(g \neq 1)$ then $|u| + |v| \leq |g|$.*

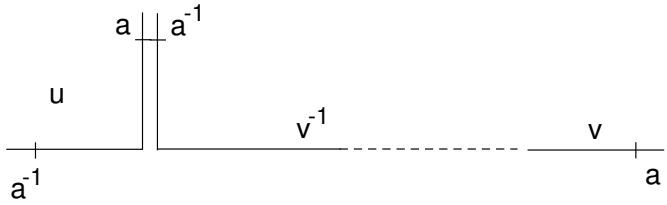
$\triangleright$. Suppose g is not cyclically reduced, say $g = a^{-1} \circ g_1 \circ a$. If we do not have pinches in a cancellation scheme for

$$u^{-1}v^{-1}uv,$$

then $|u^{-1}| + |v^{-1}| + |u| + |v| = |g|$ and the conclusion of the lemma holds. Suppose we have at least one pinch in the cancellation scheme for $u^{-1}v^{-1}uv$; say the scheme is of the type:



then we can see that we have:



hence $u = a^{-1} \circ u_1 \circ a, v = v_1 \circ a, g = {g_1}^a$.

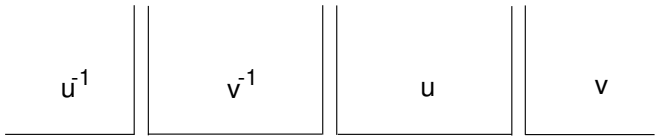
Now, $[u_1, av_1] = g_1$ and (u_1, av_1) is a minimal solution of the equation $[x, y] = g_1$. By induction on $|g|$ we have

$$|u_1| + |av_1| \leq |g_1|$$

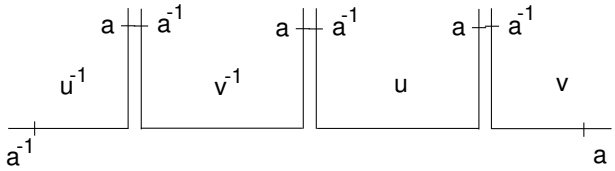
hence

$$|u| + |v| = |a^{-1} \circ u_1 \circ a| + |v_1 \circ a| \leq |a^{-1} \circ g_1 \circ a| = |g|.$$

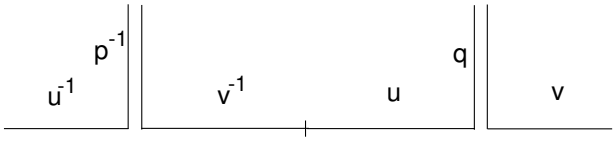
Suppose now, that g is cyclically reduced. Then there is no cancellation scheme with all three non-trivial pinches:



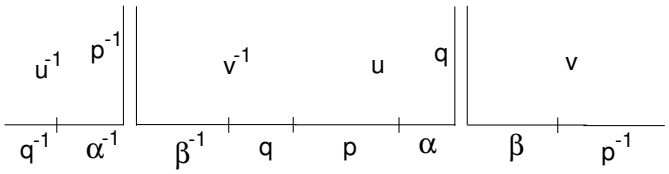
Indeed, in this event we have



So $u = a^{-1} \circ u_1 \circ a, v = a^{-1} \circ v_1 \circ a$, and hence g is not cyclically reduced-contradiction. It implies, that we have to consider the schemes of the type (with at least one trivial pinch):



Assume, $|p| \geq |q|$. Then we have the following scheme:



where $u = p\alpha q, v = q^{-1}\beta p^{-1}$. Hence

$$q^{-1}\alpha^{-1}\beta^{-1}qp\alpha\beta p^{-1} = g \tag{13}$$

Hence

$$|u| + |v| = |p| + |\alpha| + |q| + |q| + |\beta| + |p| \leq |g|.$$

Remark 1 *Notice, that from (13) we have*

$$q^{-1}\alpha^{-1}\beta^{-1}qp\alpha\beta p^{-1} = g.$$

Hence the cyclic permutation of g:

$$\alpha^{-1}\beta^{-1}(qp)\alpha\beta(p^{-1}q^{-1})$$

is the Wick's form of the type

$$a \circ b \circ c \circ a^{-1} \circ b^{-1} \circ c^{-1}.$$

It shows that Wick's for the cyclic word g come from the minimal solutions of $[x, y] = g$.

Corollary 9 *There are only finitely many minimal solutions of the equation*

$$[x, y] = g (q \neq 1).$$

If $(u_1, v_1), \ldots, (u_n, v_n)$ are minimal solutions of the equation above, then the solution set of this equation is a finite union of the equivalence classes:

$$[(u_1, v_1)] \cup \cdots \cup [(u_n, v_n)].$$

Moreover, we can effectively find the set of minimal solutions $(u_1, v_1), \cdots, (u_n, v_n)$.

Corollary 10 *Let $F = F(a, b, \ldots)$. Then the solution set of the equation $[x, y] = [a, b]$ is equal to the equivalence class $[(a, b)]$, and (a, b) is the only minimal solution of this equation.*

$\triangleright$. Indeed, if (u, v) is a minimal solution of $[x, y] = [a, b]$, then $|u|+|v| \leq 4$. If $|u| = 1$, then $|v| = 1$ (since in this case u does not cancel in $[u, v]$). In this event $u = 1, v = b$. If $|u| = 2 = |v|$, then

$$u^{-1} = a^{-1}p^{-1}, v^{-1} = pc$$

(since $[a, b]$ begins with a^{-1}) , and also $v = c^{-1}p^{-1}$ has to end on b. It follows $p = b^{-1}$, and substituting into $[u, v]$ we see that $[u, v] \neq [a, b]$. So the case $|u| = |v| = 2$ is not possible.

We described completely the solution set of the equation $[x, y] = g$.

It turns out that the methods we discussed here can be generalized into much more general situations. In fact, they are some of the most powerful methods in the theory of free groups. The transformations U, V are particular types of so-called Nielsen elementary transformations, the idea of minimal pairs gives rise to the notation of N-reduced sets, Wick's forms exists for an arbitrary quadratic equation over free groups and the idea of estimation of the length of a minimal solution of an equation in terms of length of its coefficients is the principal one in Makanin's solution of the Diophantine problem over free groups

3 Finitely generated subgroups of F

3.1. Graphs and automata.

3.1.1. Graphs.

3.1.2. Automata.

3.2. Stallings Folding Γ_H

3.2.1. Construction of Γ_H.

3.2.2. Membership problem.

3.2.3. Uniqueness of the core automaton.

3.2.4. Jitsukawa's Theorem.

3.3. Subgroup theorem.

3.4. Index of H.

3.5. Howson property.

3.6. H. Neumann conjecture.

3.7. Malnormal subgroups.

3.8. Free factors. M. Hall theorem.

3.9. Principal quotients.

3.10. Algebraic extensions.

3.11. Isolated subgroups.

3.12. Topological closures.

3.1 Graphs and automata

3.1.1 Graphs

A graph Γ consists of a set $V(\Gamma)$ of vertices of Γ, a set $E(\Gamma)$ of edges of Γ and two functions:

$$E \to V \times V \ , \ e \to (\sigma e, \tau e)$$

$$E \to E \ , \ e \to (\overline{e})$$

which satisfy the following properties

$$\overline{\overline{e}} = e, e \neq \overline{e}, \sigma e = \tau \overline{e}$$

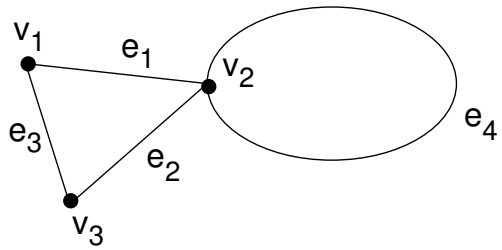
Elements from V are called *vertices* of Γ, while elements of E are *edges* of Γ, $\overline{e}$ is the *inverse* edge of e. We call σe the *initial* vertex of e or the *starting* vertex of e; τe is the *terminal* or the *final* vertex of e. We also call them the *end-points* of e. Two vertices are termed *adjacent* if they are end-points of one and some edge. An *orientation* of graph Γ is a subset $E_+ \subseteq E$ such that

$$E_+ \cap \bar{E}_+ = \emptyset, E = E_+ \cup \bar{E}_+$$

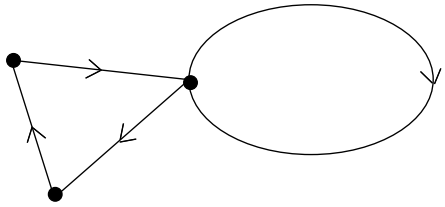
An *oriented* graph consists of a set of vertices V, a set of edges E and a function $E \to V \times V$,

$$e \to (\sigma e \ , \tau e) \ e \in E$$

Notice, that any orientation E_+ on a graph Γ determines an oriented graph Γ_+ and the function $E_+ \to V \times V$ which is the restriction of the corresponding function on E in Γ. And vice versa, any oriented graph Γ defines uniquely a graph Γ^* in which $E(\Gamma)$ is an orientation of Γ^*. Usually, we use pictures to present graphs:



in which we draw only one edge e from each pair $\{e \ \bar{e}\}$. If a graph is oriented we use arrows to show the orientation:



A *path* of *length* n in Γ is a finite sequence of edges $p = e_1...e_n$ such that $\sigma e_{i+1} = \tau e_i$, $i = 1, ..., n-1$. We say that the path p *starts* at σe_1, and *ends* at τe_n; in this event we also say that p *connects* σe_1 and τe_n. A path $p = e_1...e_n$ $n \geq 1$, is a *loop* if it starts and ends at the same point. a *simple loop* or a *circuit* is a loop $e_1...e_n$ such that all the vertices $\sigma(e_1), ..., \sigma(e_n)$ are different (no self-intersections) and which is not of the form $e\bar{e}$. A path $e_1...e_n$ is *reduced* if $e_i \neq \bar{e}_{i+1}$ for any $i = 1, ..., n-1$ (no back-tracking). By the definition above any simple loop is reduced.

A graph Γ is *connected* if any two vertices of Γ are connected by a path in Γ. It is easy to see, that any graph Γ is a disjoint union of its maximal connected subgraphs, which are called the *connected components* of Γ.

A *tree* is a connected graph without simple loops. Clearly, if Γ is a tree, then any two points in Γ are connected by a unique reduced path in Γ.

If u and v are two vertices in Γ, then a *geodesic* from u to v is a path in Γ which starts at u and ends at v and with the minimal length among all the paths connecting u and v. The length of a geodesic from u to v is called the distance between u and v. This defines a metric on Γ. In general, a geodesic between u and v is not unique, but if Γ is a tree then there exists only one geodesic between two points.

By the Zorn‘s lemma every graph Γ contains a maximal subtree. Let Γ be a connected graph, then a subtree T or Γ is maximal in Γ if and only if $V(\Gamma) = V(T)$. By this reason we refer sometimes to a maximal subtree of Γ as a spanning subtree of Γ.

Let Γ be an oriented graph and S be a set. We say that Γ is *labeled* by S or S-labeled graph, if Γ comes equipped with a function

$$\lambda : E(\Gamma) \to S.$$

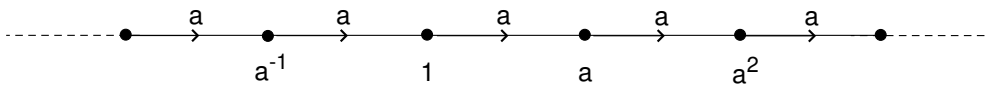
We call $\lambda(e)$ the *label* of an edge e.

Let Γ be a non-oriented graph. We say that Γ is labeled by S of there is a function $\lambda : E(\Gamma) \to S^{\pm 1}$ s.t.

$$\lambda(\bar{e}) = \lambda(\bar{e})^{-1}$$

Let G be a group and $S \subseteq G$. We define a graph $\Gamma(g, S)$ as a S-labeled graph with set G and oriented $E_+ = G \times S$, such that $\sigma(g\ , s) = g$
,
$\tau(g, s) = gs, (g \in G\ ,\ s \in S)$ and s is the label of $(g\ ,\ s) \in E_+$.

Example 1 *Let $G =< a >$ be a cyclic group generated by a, $S = \{a\} \subseteq G$. Then $\Gamma(G\ \ S)$ can be presented as follows:*

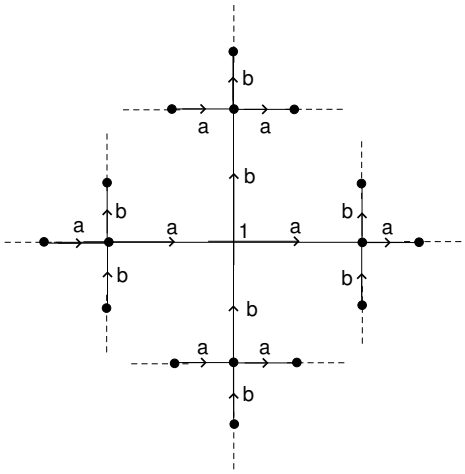


The graph $\Gamma(G\ \ S)$ is called the Cayley graph of G with respect to a subset $S \subseteq G$.

Lemma 9 *Let G be a group and $S \subseteq G$. Then:*
1) $\Gamma(G\ \ S)$ is connected $\Leftrightarrow$ S generates G;
2) $\Gamma(G\ \ S)$ is a tree $\Leftrightarrow$ G is free on S.

To prove the lemma observe that any path $p = e_1\ ...\ e_n$ from g to h in $\Gamma(G\ ,\ S)$ corresponds to a word $s_1\ ...\ s_n$ in the alphabet $S^{\pm 1}$, where s_i is the alphabet of $e_i\ ,\ i = 1\ ,...,\ n$. We call this word $s_1\ ...\ s_n$ the label of p. Now, for any $g \in G$ some paths connect 1 and g in $\Gamma(G\ ,\ S)$ if and only if $g \in < S >$. This proves 1). To prove 2) notice that $\Gamma(G\ \ S)$ is a tree if and only if different reduced paths in $\Gamma(G\ \ S)$ starting at 1 end at different vertices, and all vertices are connected to 1.It implies that, different reduced words in the alphabet $S^{\pm 1}$ define different elements of G, and each element of G is defined by such a word.. Hence, S is a basis of G and G is free on S.

Example 2 *Let $F(a\ ,\ b)$ be a free group on $\{a\ ,\ b\}$ and $S = \{a\ ,\ b\}$. Then we can present the Cayley graph $\Gamma(F(a\ ,\ b)S)$ by the following picture:*



3.1.2 Automata

Let S be a finite set and Γ be an S-labeled graph (oriented and non-trivial). We say that Γ has *One-way reading property*, if the following condition holds:

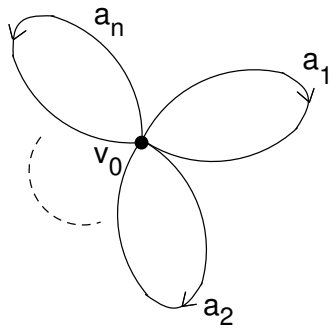
(OR) for each vertex v in Γ and each label $s \in S$ (or $s \in S^{\pm 1}$ in non-oriented case) there exists at most one edge in Γ starting at v with label s.

Suppose Γ has the property (OR). If we chose a vertex $V_0 \in V(\Gamma)$ and a set $Y \subseteq V(\Gamma)$ then the tuple $(\Gamma\ ,\ S\ ,\ V_0\ ,\ Y)$ defines an automaton A. This automaton A consists of the set of states $Q = V(\Gamma) \cup \oplus$ (here $\oplus$ is a new symbol which we call "fails" symbol), the alphabet S or $(S^{\pm 1})$, the starting state V_0, the set of accepting states Y and the transition function $\mu : Q \times X \to Q$ defined as follows:

$$\mu(v,s) = \begin{cases} w, \text{ if } v \in V(\Gamma) \text{ and there is an edge from } v \text{ to } w \text{ in } \Gamma \\ \quad \text{with the label } s; \\ \oplus, \text{ otherwise.} \end{cases}$$

We say that the automaton A *accepts* a word p in the alphabet S or $(S^{\pm 1})$ if there is a loop at V_0 in Γ with the label p. Notice, that in this event such a loop is unique since Γ has one-way reading property. Denote by $L(A)$ the set of words in S $(S^{\pm 1})$ accepted by A.

Example 3 *Let* $S = \{a_1\ , ...,\ a_n\}$. *Then a non-oriented graph*



accepts all words in the alphabet $S^{\pm 1}$.

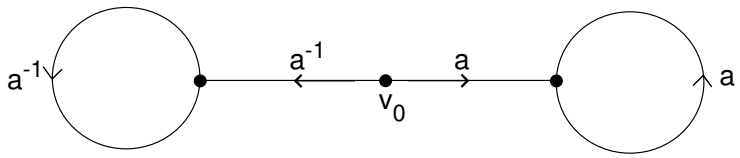
If we want to construct an automaton that will accept precisely all reduced words in $S^{\pm 1}$, then we have to use oriented graphs labeled by $S^{\pm 1}$.

A subset $L \subseteq S^*$ is called *regular* if $L = L(A)$ for some finite automata A (recall that by S^* we denote the set of all words in S).

It is worth to mention that if L is a regular subset of $(S^{\pm 1})^*$ then the set $\bar{L}$ which consists of all reduced forms of words from L, is also regular.

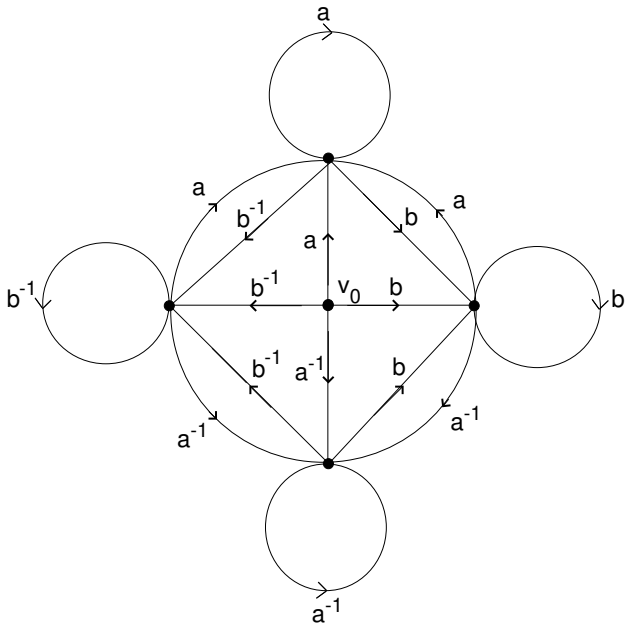
Here are some examples of oriented labeled graphs which accept only reduced words.

Example 4 *Let* Γ *be the following directed labeled graph with labels from* $S = \{a\}^{\pm 1}$*:*



If we put $Y = V(\Gamma)$*, then* Γ *gives rise to a finite automaton* A *which accepts all words of the type* $a^n, n \in \mathbf{Z}$ *and nothing else. Hence,* $L(A) = F(a)$*. In this case we say that* A *accepts the infinite cyclic group* $< a >$ *generated by* a*.*

Example 5 *Let* Γ *be the following directed labeled graph with labels from* $S = \{a\ ,\ b\}^{\pm 1}$*:*



If we put $Y = V(\Gamma)$, then Γ determines an automaton A which accepts precisely all reduced words in the alphabet $\{a\ ,\ b\}^{\pm 1}$, i.e., A accepts the free group $F(a, b)$.

Observe, that for any group G and any subset $S \subseteq G$ the Cayley graph $\Gamma(G\ ,\ S)$ satisfies the one-way reading property.

Now, if we again choose $V_0 = 1$ and $Y = \{1\}$ then the graph $\Gamma(G\ ,\ S)$ gives rise to an automaton, which is infinite if the group G is infinite. This automaton accepts a reduced word in the alphabet $S^{\pm 1}$ if and only if this word w is equal to 1 in G.

3.2 Stallings Foldings. Γ_H

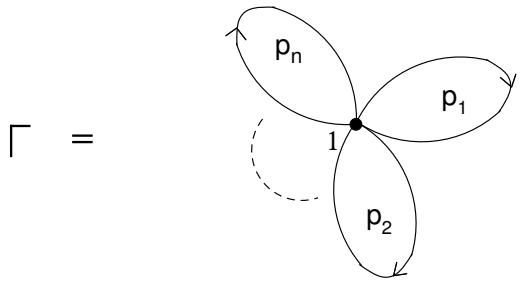
3.2.1 Construction of Γ_H

Let $F(A)$ be a free group with basis $A = \{a_1, \ldots, a_m\}$. Let H be a subgroup of F generated by elements $h_1\ \ldots,\ h_n \in F$.

We will show how to construct a finite automaton $\Gamma(H)$ which accepts precisely the reduced in $A^{\pm 1}$ words which belong to H (if we identify $F(A)$ with the set of all reduced words in $A^{\pm 1}$).

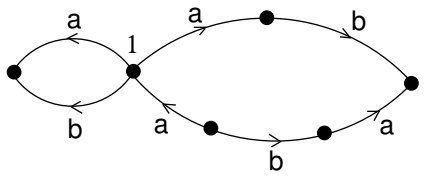
We start with a finite non-oriented labeled graph Γ with labels from $A^{\pm 1}$. The graph Γ has a distinguished vertex that we denote by 1 and n simple loops $p_1, \ldots, p_n$ with labels $h_1, \ldots, h_n$. Recall, that in any labeled graph if

x is a label of an edge e, then x^{-1} is the label of the edge $\bar{e}$. In the graph Γ the orientation E_+ is given by all edges with labels $a_1, \ldots, a_m$. As usual, we present the graph Γ by a picture in which we draw only edges from E_+. Thus



where p_i has label h_i , $i = 1, \ldots, n$.

Example 6 *Let $F = F(a\ ,\ b)$ and $H = \langle aba^{-1}b^{-1}a\ ,\ ab^{-1}\rangle$, then the corresponding graph Γ can be presented as follows:*



Observe, that for any subgroup $H = \langle h_1\ , \ldots,\ h_n\rangle$ the graph Γ satisfies the following property:

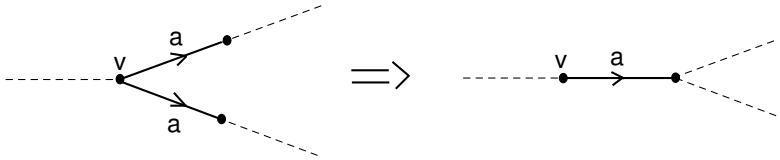
$(*)$ an element $f \in F$ belongs to H if and only if there exists a loop in Γ at the vertex 1 with label w, such that w defines $f \in F$.

Indeed, every loop p in Γ is a product of finitely many simple loops $p_1^{\pm 1}\ , \ldots,\ p_n^{\pm 1}$ (here if $p_i = e_1 \cdots e_k$, then $p_i^{-1} = \bar{e_k}\bar{e_{k-1}} \cdots \bar{e_1}$). Hence the label of p is a finite product of the elements $h_1^{\pm 1}\ , \ldots,\ h_n^{\pm 1}$, therefore it defines an element of H. The inverse is obvious.

Notice, that in general, Γ does not satisfy the one-way reading property (OR), hence Γ does not give rise to an automaton. For instance, in Example 6 there are two different edges from 1 with the same label a.

Now we will show how to produce a graph $\Gamma(H)$ which satisfies both properties (OR) and $(*)$.

$\Gamma(H)$ can be obtained from Γ by finitely many steps which are called *reductions*. Suppose in a labeled graph Δ there are two different edges starting at the same vertex v and labeled by the same label a:



then we transform the graph Δ into a graph Δ' identifying these two edges and their terminal vertices. This is called a reduction of Δ.

Lemma 10 *If Δ satisfies $(*)$, then Δ' also satisfies $(*)$.*

$\rhd$. To see this it suffices to notice that if $p = e_1 \cdots e_n$ is a loop in Δ, then $p = e_1 \ \cdots \ e_n$ is also a loop in Δ' and vice versa. $\lhd$

Now, starting with the graph Γ we transport it by finitely many reductions into a graph $\Gamma(H)$ which satisfies the property (OR) and also the property $(*)$:

$$\Gamma \longrightarrow \Gamma_1 \cdots \longrightarrow \Gamma_n = \Gamma(H).$$

Indeed, we do reductions until this is possible, the resulting graph $\Gamma(H)$ has no vertices with two different outgoing edges labeled by the same symbol. Hence, $\Gamma(H)$ satisfies (OR) and it also satisfies (*) by the lemma above.

Now, if we put $v_0 = 1$ and $Y = \{1\}$, then $\Gamma(H)$ gives rise to a finite automaton which accepts H, i.e., $\Gamma(H)$ accepts a reduced word in $A^{\pm 1}$ if and only if this word defines an element from H.

From the discussion above we see that the following theorem has been proved.

Theorem 5 *Let $H = \langle h_1 \ \ldots, \ h_n \rangle$ be a finitely generated subgroup of a free group $F = F(a_1 \ \ldots, \ a_m)$ of finite rank. Then one can effectively construct a finite automaton $\Gamma(H)$ which accepts H.*

Notice, that we need at most

$$c \ m \ (|h_1| + \ \cdots \ + |h_n|)^2$$

steps to construct $\Gamma(H)$, $c = const.$ so if the group F is fixed, then the procedure of constructing $\Gamma(H)$ is of quadratic-time complexity.

Corollary 11 *Let F be a free group of finite rank. The membership problem for F is decidable, i.e., for given words $w\ h_1\ \ldots,\ h_n \in F$ one can effectively decide whether $w \in \langle h_1\ \ldots,\ h_n\rangle$ or not and this takes at most $C\ (|w|+(|h_1|+\cdots+|h_n|)^2)$ steps, where C is a constant which depends linearly on the rank of F.*

There is a relation between graphs $\Gamma(F\ ,\ A)$ and $\Gamma(H)$.

Let G be a group, $A \subseteq G$, and $\Gamma = \Gamma(G\ ,\ A)$ be the Cayley graph of G with respect to A. The group G acts on $\Gamma(G\ ,\ A)$ by left multiplication:

$$v \xrightarrow{g} gv,\quad (v\ ,\ a) \xrightarrow{g} (gv\ ,\ a);$$

here $v \in V(\Gamma), (v\ ,\ a) \in E(\Gamma)\ ,\ g \in G$. This map is an isomorphism of Γ.

Let H be a subgroup of G. Then H also acts on Γ. Denote by Γ_H the quotient graph under this action:

$$\Gamma_H = \Gamma/H.$$

The vertices of Γ_H are the orbits of the action of H on Γ; hence they are just the right cosets of H in G:

$$Hf = \{g \in G | \exists\ h \in H,\ g = hf\}.$$

There is an edge with label $a \in A$ from Hf into Hg if and only if $Hfa = Hg$. This turns Γ_H into an A-labeled graph. Notice, that every loop in Γ_H at the vertex H defines an element from H. Moreover, Γ_H has one-way reading property. So, if we put $v_0 = H, Y = \{H\}$, then Γ_H gives rise to an automaton which accepts H. It is not hard to see, that $\Gamma(H)$ is embeddable into Γ_H, and any loop at H in Γ_H is a product of loops in $\Gamma(H)$ (we identify $\Gamma(H)$ with its image in Γ_H). We will prove this in section ???

3.2.2 Jitsukawa's Theorem

Let $K = \langle w_1, \ldots, w_m\rangle$ be a finitely generated subgroup of a finitely generated free group $F_n = \langle x_1, \ldots, x_n\rangle$, where each w_i is a reduced word in F_n. By the Stallings folding, we can obtain the core graph $\Gamma(K)$ which accepts exactly K. Every core graph has the following properties:

1) it is a finite connected gra[h with the base point,

2) the degree of each vertex is not more than $2n$, and is more than 1 except at the base point.

Let $\mathcal{J} = \{$graphs equipped with 1) and 2)$\}$.

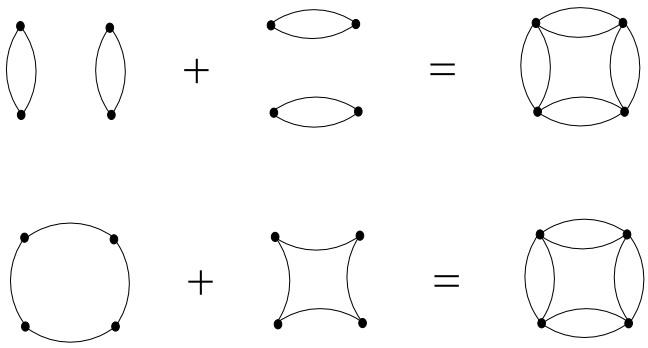
Theorem 6 *Let $G \in \mathcal{J}$. Then, there exists a finitely generated sungroup K of the finitely generated free group F_n such that $G = \Gamma(K)$ by assigning labels and directions appropriately.*

Definition 11 *A 2-factor of a graph is a spanning subgraph, that is a subgraph containing every vertex of the graph, so that every vertex has degree 2.*

Definition 12 *A graph is said to be 2-factorable if the graph can be decomposed into edge disjoint 2-factors.*

Theorem 7 *(Petersen) Every finite even degree regular graph is 2-factorable.*

Factorization of a graph is not unique. The following are two different 2-factorizations of a 4-regular graph.



Lemma 11 *Let G be a finite graph. Then, there exists a regular graph such that G is a spanning subgraph of the graph.*

$\triangleright$. Let m' be the maximum degree of the vertices of G and let

$$m = \begin{cases} m', \text{ if } m' \text{ is even,} \\ m' + 1, \text{ if } m' \text{ is odd.} \end{cases}$$

First, we add $\left\lfloor \frac{m - \deg(v)}{2} \right\rfloor$ loops to each vertex v of G. Then, the degree of each vertex is m or $m - 1$, and, there is even number of odd degree vertices since every graph has even number of odd degree vertices. Second, we pick two odd degree vertices if there are any, and connect them by an edge. Repeat this connection as long as possible. The resulting graph is an m-regular graph containing G as a spanning subgraph. $\triangleleft$

Proof of the Theorem. Let $\tilde{G}$ be a finite $2n$-regular graph, which contains G as a subgraph. By the Petersen's Theorem we can decompose $\tilde{G}$ into n 2-graphs. Then, each 2-factor consists of disjoint union of loops. Pick a 2-factor of $\tilde{G}$ and label all the edges by x_1, and direct each edge along each loop. Then, at each vertex of $\tilde{G}$ we have one outgoing x_1-edge and one incoming x_1-edge. Similarly, we repeat this assignment for all other 2-factors with other labels. Now, we look at subgraph G, which is folded. And, to obtain a set of generators of K, take a maximal tree of G, which involves the base point of G as a vertex. The Schreier's Method gives us a set of generators of K. $\lhd$

3.3 Subgroup Theorem

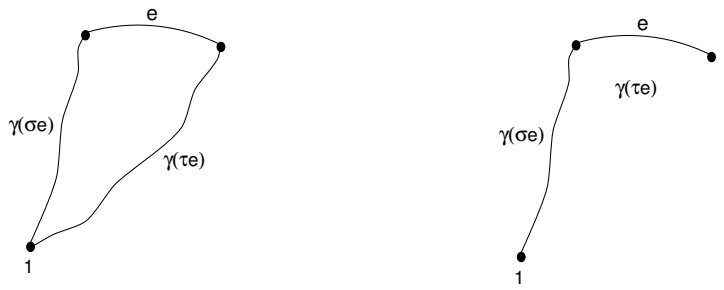
The subgroup theorem for free groups states that every subgroup H of a free group F is free. We will prove the subgroup theorem here for only finitely generated subgroups, leaving the infinite case for Section ???. Also, we will show how to find effectively a basis of H, hence we can compute effectively the rank of H.

Let F be a free group with basis $A = \{a_1, \ldots, a_m\}$. Let $H =< h_1, \ldots, h_n >$ be a subgroup of F. Denote by $\Gamma(H)$ the finite A-labeled graph from the Section 3.2.1 which gives rise to an automaton which accepts H. (We will say, that $\Gamma(H)$ accepts H). We saw already that a reduced word w belongs to H if there exists a loop at 1 in $\Gamma(H)$ with the label w. Now we will show how to construct a finite set of so called fundamental loops $p_1, \ldots, p_k$ in $\Gamma(H)$, such that their labels form a basis of H.

Let T be a maximal subtree of $\Gamma(H)$. For a vertex $v \in \Gamma(H)$ denote by $\gamma(v)$ the unique reduced path in T from 1 to v. Now, for an edge $e \in \Gamma(H)$ the path

$$p_e = \gamma(\sigma e) e \gamma(\tau e)^{-1}$$

is a loop at 1. Notice, that p_e is reduced if $e \in \Gamma(H) - T$.



$$e \in \Gamma(H) - T \qquad\qquad e \in T$$

Denote by s_e the label of p_e.

Claim 1. The set

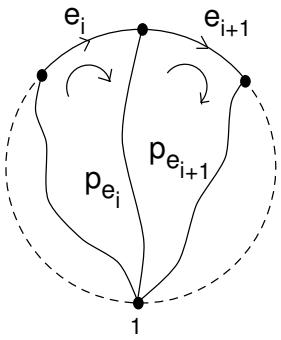
$$S_T = \{S_e \mid e \in \Gamma(H) - T\}$$

generates H.

Indeed, since p_e is a loop at 1 in $\Gamma(H)$, then $s_e \in H$. Now let w be an arbitrary element of H. We think of w as a reduced word in $A^{\pm 1}$. Since $\Gamma(H)$ accepts w, then there exists a loop

$$q = e_1 \ \cdots \ e_n$$

in Γ with the label w. Notice, that $q = p_{e_1} \ \cdots \ p_{e_n}$



Therefore, $w = s_{e_1} \ \cdots \ s_{e_n}$ in $F(A)$ hence $w \in \langle S_T \rangle$.

Claim 2. The set S_T is a basis of H.

Let

$$w = s_{e_1}^{\varepsilon_1} \ \cdots \ s_{e_k}^{\varepsilon_k}$$

be a reduced non-trivial word in $S_T^{\pm 1}$.

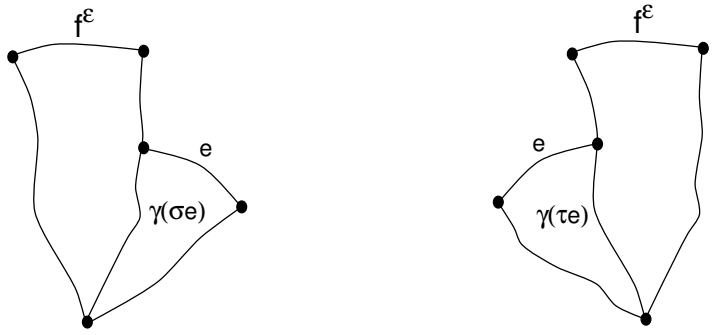
Then the edges $e_1^{\varepsilon_1} \ldots, \ e_k^{\varepsilon_k}$ can not be reduced in reducing the path $p_{e_1}^{\varepsilon_1} \ \cdots \ p_{e_k}^{\varepsilon_k} = q$, i.e., the labels of these edges, say, $\mu_1^{\varepsilon_1} \ , \ldots, \ \mu_k^{\varepsilon_k}$ can not cancel out in reducing w.

Indeed, for any element $s_e \in S_T$ and any other element $s_f \in S_T$, $(e \neq f)$ the longest initial segment of s_e that cancels out in the product

$$s_f^{\varepsilon} s_e \qquad (\varepsilon \in \{-1 \ , \ 1\})$$

is a subword of $\gamma(\sigma e)$, since $e, f \notin T$ and $e \neq f^{\pm 1}$. Similarly, the longest terminal subword of s_e that can be canceled out in a product of the type

$$s_e s_f^{\varepsilon} \quad , \quad (\varepsilon \in \{-1 \ , \ 1\})$$



is a subword of $\gamma(\tau e)^{-1}$.

Hence, the cancellation in w does not affect the labels $\mu_1^{\varepsilon_1}, \ldots, \mu_k^{\varepsilon_k}$ - the "midles" of $s_{e_1}^{\varepsilon_1}, \ldots, s_{e_k}^{\varepsilon_k}$. Thus $w \neq 1$.

It implies, that S_T is a basis of H.

Consequently H is a free group on S_T. Thus we proved the following

Theorem 8 *(Nielsen). Every finitely generated subgroup of F is free.*

Nielsen proved this theorem using a completely different argument , which we will discuss in the next chapter.

Observe, that we also prove the following results.

Corollary 12 *For a given finitely generated subgroup $H = \langle h_1, \ldots, h_n \rangle$ of F one can effectively find a basis of H.*

Corollary 13 *For a given f.g. subgroup $H = \langle h_1, \ldots, h_n \rangle$ of F one can effectively find the rank of H.*

Indeed, for a given H we can effectively construct the graph $\Gamma(H)$. For the finite graph $\Gamma(H)$ we can effectively find a maximal subtree T of $\Gamma(H)$ (this is a well known procedure in graph theory, we will discuss one particular method of finding a maximal subtree in Section ???). Then for a given maximal subtree T we form the basis S_T of H. Notice, that the rank of H is equal to

$$|S_T| = \text{the number of edges in } (\Gamma(H) - T).$$

If we denote by n the number of edges in $\Gamma(H)$ and by m the number of vertices in $\Gamma(H)$, then

$$\text{the number of edges in } T = \text{the number of vertices in } T - 1 = m - 1.$$

Hence,

$$\text{rank } H = n - m + 1$$

.

3.4 Finite index

Let H be a subgroup of a group G. A set $R \subseteq G$ is called a set of (right) *representatives* of H in G or a (right) *transversal* of G in H, if each right caset of H contains one and only one element from R. In this event

$$G = \bigcup_{r_i \in R} Hr_i.$$

The cardinality of a transversal R of H in G is called the index of H in G, we denote it by $[G : H]$.

Notice, that if G is a finitely generated group and H is a subgroup of finite index in G, then H is finitely generated. Indeed, let G be generated by a finite set

$$X = \{x_1 \ , \ldots , \ x_n\}$$

and $R = \{r_1 \ \ldots , \ r_n\}$ be a finite transversal of H in G, containing 1. Denote by $\bar{g}$ the representative of an element $g \in G$ in R. Put

$$t(r \ , \ x) = rx\overline{(rx)}^{-1}.$$

Then the set $S_{X,R} = \{t(r \ , \ x) \mid r \in R \ , \ x \in X\}$ generates H. Indeed, let

$$w = y_1 \ \cdots \ y_e \in H, \ (y_i \in X^{\pm 1}).$$

Then

$$w = (1 \cdot y_1\overline{1 \cdot y_1}^{-1})(\overline{y_1}y_2\overline{(\overline{y_1}y_2)}^{-1})(\overline{\overline{y_1}y_2}y_3\overline{\overline{\overline{y_1}y_2}y_3}^{-1})$$
$$\cdots(\overline{\overline{\overline{y_1}y_2 \ \ldots \ y_{n-1}} \cdot y_n}(\overline{\overline{y_1} \ \ldots \ y_n})^{-1})\overline{\overline{y_1} \ \cdots \ y_n}.$$

But since $\overline{\bar{y}_1 \ \cdots \ y_n} \in H$, then $\overline{\overline{y_1} \ \cdots \ y_n} = 1$, hence

$$w \in \langle S_{X,R} \rangle,$$

as desired.

Definition 13 *We say that the Index Problem is decidable in G if for any finitely generated subgroup H of G we can effectively find the index of H in G.*

It is understood, that the group G above if finitely presented.

Notice, that if the group G above is finitely presented and H is a finitely generated subgroup of G of finite index, then there exists an effective procedure, so-called Todd-Coxeter algorithm, to find a transversal R of H in G.

Moreover, in this case, for each $g \in G$ we can effectively find the representative of g in R. It implies, that if the Index Problem is decidable over G, then the membership problem is also decidable for G. Since there are groups with the undecidable membership problem, the there are groups with the undecidable Index Problem.

We will see shortly, that the index problem is decidable in an arbitrary free group. Let $F = F(A)$ be a free group with basis $A = \{a_1, \ldots, a_m\}$, and $H = \langle h_1, \ldots, h_n \rangle$ be a finitely generated subgroup of F. as before, we denote by $\Gamma(H)$ the finite graph corresponding to the subgroup H; $\Gamma(H)$ is an A-labeled graph such that a reduced word w in the alphabet $A^{\pm 1}$ belongs to H if there exists a loop at 1 in $\Gamma(H)$ with label w.

Definition 14 *We say that an A-labeled graph Γ is* regular *if for any vertex v in Γ and any label $a \in A$ there exists an edge in Γ starting at v and with label a.*

Theorem 9 *A finitely generated subgroup H of F has a finite index if and only if $\Gamma(H)$ is a regular $A^{\pm 1}$-labeled graph.*

Proof. Suppose $\Gamma(H)$ is not a regular A-labeled graph. Then there exists a vertex $v \in \Gamma(H)$ and a label $a \in A^{\pm 1}$, such that there is no edge in $\Gamma(H)$ starting at v and labeled by a. Let p be a reduced path from 1 into v in $\Gamma(H)$ and g be the label of the path p. Choose a letter $b \in A$ which is not equal to a or a^{-1} (we can assume that F is of rank more than 1), and define a word w as follows:

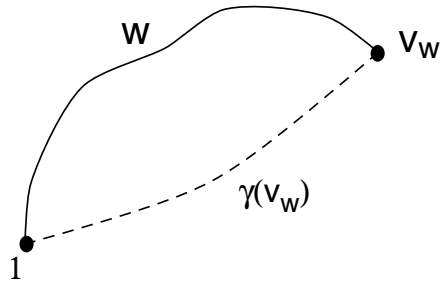
$$w = \begin{cases} ga, \text{ if } g \text{ does not begin with } a^{-1}, \\ gab, \text{ otherwise.} \end{cases}$$

Clearly, w is a cyclically reduced word. Hence any power w^k is also reduced as written. Notice, that $w^k \notin H$. Indeed, we cannot "read" ga inside $\Gamma(H)$, i.e., after reading of ga we go to the "fails" vertex $\otimes$ (see Section ???) Now, it follows, that

$$Hw^k = Hw^e \Leftrightarrow w^{k-e} \in H \Leftrightarrow k = e.$$

Hence, H is of infinite index in F.

Suppose now, that the graph $\Gamma(H)$ is a regular $A^{\pm 1}$-labeled graph. Let T be a maximal subtree of $\Gamma(H)$. For any vertex $v \in \Gamma(H)$ denote by $\gamma(v)$ the path from 1 into v inside T. If w is a reduced word in $A^{\pm 1}$, then we can read off w inside $\Gamma(H)$ starting at 1. Let v_w be the vertex where we finish reading w off inside $\Gamma(H)$.



Then $w\gamma(v_w)^{-1}$ is the label of a loop in $\Gamma(H)$ at 1, hence

$$w\gamma(v_w) \in H.$$

Therefore

$$Hw = H\gamma(v_w)$$

and $\gamma(v_w)$ is a representative of w modulo H. Put

$$R_T = \{\gamma(v) \mid v \in \Gamma(H)$$

then R_T is a transversal of H in F. Since $\Gamma(H)$ is a finite graph, then $|R_T| < \infty$ and H is of finite index in F.

Theorem is proved.

Corollary 14 *Let H be a f.g. subgroup of a free group F. If H contains a non-trivial normal subgroup of F, then H has a finite index in F.*

Proof. Let $N \lhd F$, $N \leq H$ and $1 \neq g \in N$. If F is of rank 1 then every non-trivial subgroup of F has finite index. Suppose now that F is of rank at lest 2. Then for any element $f \in F$ there exists an element $h \in H$ such that

$$w = fhgh^{-1}f^{-1}$$

is reduced as written (in fact, if a , b are part of a basis of F, then one can choose h to be either $a^{\pm 1}$ or $b^{\pm 1}$). Since, $w \in N$ the word w is readable in $\Gamma(H)$, hence arbitrary $f \in F$ is readable in $\Gamma(H)$. It implies, that $\Gamma(H)$ is a regular graph, so H has finite index in F.

Corollary 15 *Let H be of finite index in F, and T be a maximal subtree of $\Gamma(H)$. Then*

$$R_T = \{\gamma(v) \mid v \in \Gamma(H)\}$$

is a right transversal of H in F. In particular, $[F : H]$ equals to the number of vertices in $\Gamma(H)$.

Definition 15 *A transversal R of H in F is called a Schreier's transversal if for any $r \in R$ every initial segment of r also belongs to R, i.e., if $r \in R$ and $r = r_1 \circ r_2$, then $r_1 \in R$.*

Notice, that the transversal R_T is a Schreier's transversal of H in F.

Corollary 16 *Let H be a subgroup of finite index in F. Then one can effectively find a Schreier's transversal of H in F.*

Corollary 17 *Let H be a subgroup of finite index in F. Then*

$$[F : H] = \frac{\text{rank } H - 1}{\text{rank } F - 1}.$$

$\triangleright$. Let rank $F = n$, $m =$ the number of vertices in $\Gamma(H)$. Since $\Gamma(H)$ is regular, then exactly $2n$ different edges start at every vertex in $\Gamma(H)$. Then $(2n)m$ is twice the number of edges in $\Gamma(H)$ (each edge is counted once at each of its end-points). Hence

$$mn = \text{the number of edges in } \Gamma(H).$$

As we have mentioned in Section ???

rank $H-1 =$ the number of edges in $\Gamma(H)-$the number of vertices in $\Gamma(H) = nm-m$

. Notice, that

$$[F : H] = |R_T| = m.$$

Hence,

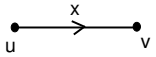
$$[F : H] = m = \frac{nm - m}{n - 1} = \frac{\text{rank } H - 1}{\text{rank } F - 1}.$$

Corollary is proved.

Now we will prove that if H is a subgroup of finite index of a finitely generated group G, then all Schreier's transversals of H can be obtained as R_T for some maximal tree T of some finite automaton Γ accepting H.

Proposition 4 *Let $G = \langle x_1, \ldots, x_n \rangle$ be a finitely generated group and H is a subgroup of finite index in G with a Schreier's transversal $R = \{r_1, \ldots, r_k\}$. Then there exists a finite $X^{\pm 1}$-labeled regular automaton Γ accepting H and a maximal subtree T of Γ such that*

$$R = R_T$$



$\triangleright$. Denote by V the set of all initial segment of all words in R (including elements from R themselves). Let a graph Γ be a graph with vertex set V in which for each label $x \in X^{\pm 1}$ and each vertex $v \in V$ there exists an edge starting at v with label x and the final vertex $v\bar{x}$, where $v\bar{x}$ is the representative of vx in R.

Clearly, Γ is a finite $X^{\pm 1}$-labeled regular graph. Let T consist of all edges of the type

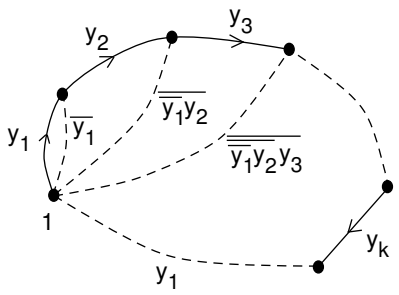
in which $v = u \circ x$ and v is a subword of some $r \in R$. Clearly T is a maximal subtree of Γ. Now, Γ accepts H and $R = R_T$.

To see that it is enough to repeat the argument from page 52. But the best way to present this argument, is by the following picture:

let

$$w = y_1 \ \cdots \ y_k \in H \ , y_i \in X^{\pm 1}$$

then



$$w = y_1 y_2 y_3 \ \cdots \ y_k = (y_1 \overline{y_1}^{-1})(\overline{y_1} y_2 \overline{\overline{y_1} y_2}^{-1})(\overline{\overline{y_1} y_2} y_3 \overline{\overline{\overline{y_1} y_2} y_3}^{-1})$$
$$\cdots (\overline{\overline{\overline{y_1} y_2} \cdots y_n}).$$

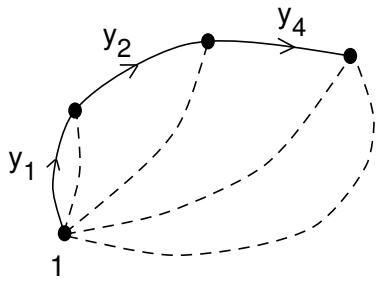
Hence, $\overline{\overline{\overline{y_1} y_2} \ \cdots \ y_n} \in H$ and therefore it is equal to 1. Now the picture above has the form

and $w = p_{y_1} p_{y_2} \ \cdots \ p_{y_k}$ (see Section ???), where

$$p_{y_i} = \gamma(\sigma y_1) y_i \gamma(\tau y_1)^{-1}$$

and $\gamma(v)$ is the unique path in T from 1 into v, $(v \in V)$.

It implies that $R = R_T$ as desired.



We will discuss Schreier's representatives for arbitrary (not only finitely generated) subgroups of F in Section ??? .

3.5 Howson property

Definition 16 *A group G has the* Howson property *if the intersection of any two finitely generated subgroups of G are again finitely generated.*

Example 7 *Finite groups and abelian groups have the Howson property. In fact, these groups satisfy an even stronger condition: every subgroup of finitely generated abelian (or finite) group is finitely generated.*

We say that a group G satisfies the *maximal condition on subgroups* (G has max) if every subgroup G is finitely generated. It is easy to see that G has max if and only if G satisfies ascending chain condition on subgroups, i.e., every strictly increasing chain of subgroups is finite.

It is known, that finitely generated nilpotent groups and polycyclic groups have max (see, []).

We saw already, that every non-abelian free group contains a subgroup of infinite (countable) rank, so non-abelian free groups do not have max. Nevertheless, we have the following theorem.

Theorem 10 *(Howson, 54). Free groups have the Howson property.*

$\triangleright$. Let H and K be finitely generated subgroups of a free group F. We can assume, that F is finitely generated (otherwise, we replace F by the subgroup $\langle H\ ,\ K\rangle$, which is finitely generated, hence it is free of finite rank). The easiest way to prove the theorem is the following. We have proved above that H and K are regular subsets of F. (i.e., there are some automata that accept them). Now it is a well-known result in the automata theory that the intersection of two regular sets is again regular, hence $H \cap K$ is accepted by

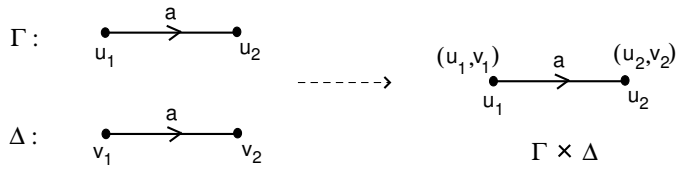
some finite automata Γ. Now as we saw in Section ??? the subgroup $H \cap K$ is finitely generated. That's it. But we would like to have a description of this automata Γ (we will use it in the discussion of H. Newmann conjecture). $\lhd$

To this end, let us consider the product operation on labeled graphs.

Let Γ and Δ be A-labeled graphs (directed or non-directed) with labels in a set A. By $\Gamma \times \Delta$ we denote the direct product of graphs Γ and Δ, i.e., $\Gamma \times \Delta$ is an A-labeled graph defined as follows

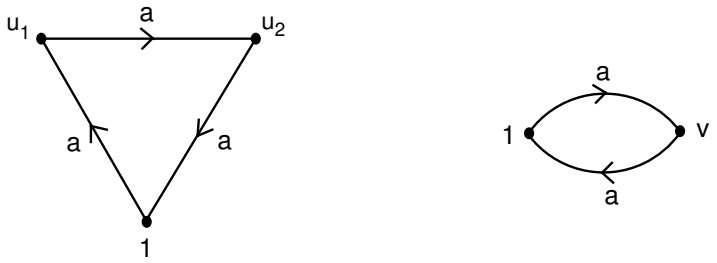
$$V(\Gamma \times \Delta) = V(\Gamma) \times V(\Delta)$$

and edges are defined coordinate-wise: for vertices $(u_1\ ,\ v_1), (u_2\ ,\ v_2) \in V(\Gamma \times \Delta)$ we define an edge in $\Gamma \times \Delta$ from $(u_1\ ,\ v_1)$ into $(u_2\ ,\ v_2)$ with a label $a \in A$ if and only if there is an edge in Γ from u_1 into u_2 with label a and there is an edge in Δ from v_1 into v_2 with the label a:



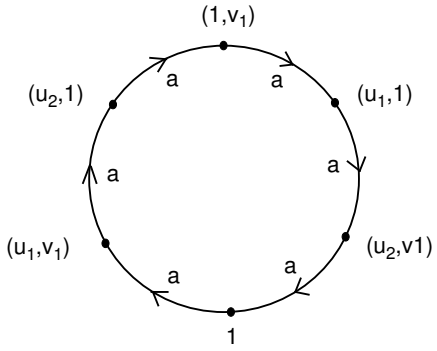
It is easy to see, that if $\Gamma(H)$ and $\Gamma(K)$ are finite state automata, which accept H and K, correspondingly, then $\Gamma = \Gamma(H) \times \Gamma(K)$ accepts $H \cap K$.

Example 8 *Let* $F = F(a\ ,\ b), H = \langle a^2 \rangle, K = \langle a^3 \rangle$.



Then according to the construction above Γ *has the following form:*

Thus, for given finitely generated subgroups $H = \langle h_1\ , \ldots,\ h_m \rangle, K = \langle k_1\ , \ldots,\ k_m \rangle$ of a free group F we can construct the automaton Γ, which



accepts $H \cap K$. Now, we can effectively construct a maximal subtree T of Γ, hence we can find the free basis S_T of $H \cap K$ corresponding to the tree T. We formulate this as:

Corollary 18 *There exists an algorithm to find a set of generators of the intersection of any two finitely generated subgroups of F given by some finite generating sets.*

3.6 Hanna Neumann Conjecture

3.6.1 HNC

In paper [**?**], improving the result of A. Howson, Hanna Neumann proved the following bound on the rank of intersection of two non-trivial finitely generated subgroups $H, K \neq F$:

$$r(H \cap K) - 1 \leq 2(r(H) - 1)(r(K) - 1).$$

In the same paper she asked if the factor 2 can be dropped. This became known as the H.Neumann Conjecture (HNC):

(HNC) Is is true that for every non-trivial f.g. subgroups $H, K \leq F$ the following inequality holds:

$$r(H \cap K) - 1 \leq (r(H) - 1)(r(K) - 1).$$

We show how Stallings' foldings can be applied to prove H.Neumann result. This exposition is based on papers by S. Gersten [**?**, **?**, G] Imrich, Nicholas, ???.

Obviously, HNC holds if at least one of the subgroups H, K has rank 1.

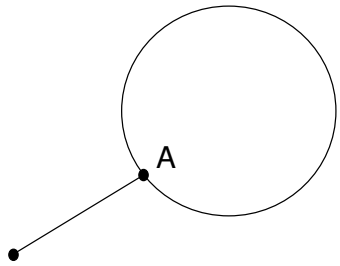
Tardos [**?**] proved that HNC holds if at least one of the subgroups H, K has rank 2. W.Dicks and E.Formanek showed that HNC holds if at least one of the subgroups has rank 3. Two reductions:

Reduction 1 *One can always assume that the free group F has rank 2.*

Indeed, let ϕ be a monomorphism $\phi : F \longrightarrow F_2$. Thus $r(H) = r(H^\phi), r(K) = r(K^\phi), r(H \cap K) = r(H^\phi \cap K^\phi)$. Hence the result.

Reduction 2 *One can always assume that the graphs Γ_H and Γ_K do not have vertices of degree 1.*

Proof. We have seen already that for every subgroup $H \leq F$ the graph H has at worst one vertex of degree 1 - the starting vertex 1_H. If 1_H has degree 1, then Γ_H has the following form:



Γ(H)

Let Γ_H^0 be the core of the graph Γ_H and p be the unique shortest path connecting 1_H to Γ_H^0. If $u \in F$ is the label of the path p then every element $h \in H$ is of the form uh_0u^{-1}, where h_0 is the label of some closed path in Γ_H^0 started at A - the endpoint of p. We turn Γ_H^0 into a deterministic finite automaton with starting and termination vertex A. In fact, Γ_H^0 is the Stallings folding of the subgroup $H_0 = u^{-1}Hu$.

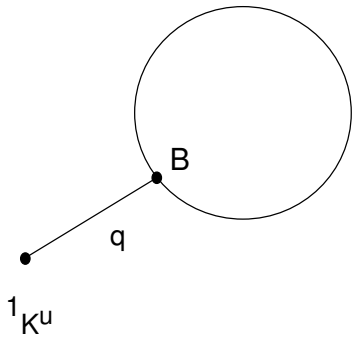
Now, replacing subgroups H, K by their conjugates H^u, K^u we can assume that the folding $\Gamma_{H^u}^0 = \Gamma_{H_0} = \Gamma_H^0$ is a core graph, i.e., it does not contain vertices of degree 1.

The argument above shows that Γ_{K^u} has the form:

Let q be the shortest path connecting i_K to the core(Γ_{K^u}) and B be the endpoint of q, and let v be the label of q.

If $H \cap K = 1$ then HNC obviously holds.

If $H \cap K \neq 1$ then there exists a non-trivial $w \in H \cap K$. Since Γ_{K^u} accepts w it follows that $w = uw_0u^{-1}$ as reduced word. Hence there is a path p_u in Γ_{H^u} starting at 1 with the label u. Let C be the endpoint of p_u. Consider subgroups H^{uv}, K^{uv}. The argument above shows that $\Gamma_{K^{uv}}$ has no vertices of degree 1. Observe that $\Gamma_{H^{uv}}$ is the graph Γ_{H^u} with distinguished point C. So it does not have vertices of degree 1 either.



Now HNC holds for (H, K) iff it golds for $(H^{uv}, K^{u,v})$. That finishes reduction 2.

Recall that

$$r(H) = |E(\Gamma_H)| - |V(\Gamma_H)| + 1$$

hence

$$r(H) - 1 = n - m.$$

Now, for a vertex v in a graph Γ, we denote by $d(v)$ the number of edges adjacent to v (i.e. $d(v)$ is the sum of all outgoing edges at v and all incoming edges).

Denote by $d_i(v)$ the number of vertices in Γ of degree i. Clearly:

$$m = \sum_i d_i(\Gamma), 2n = \sum_i i d_i(\Gamma)$$

(since each edge inputs one to incoming degree of its terminal vetrex, and inputs one to outcoming degree of its initial vertex, even though these vertices are equal).

Hence

$$r(H) - 1 = \frac{1}{2}\sum_i i d_i(\Gamma_H) - \sum_i d_i(\Gamma_H).$$

Now suppose $H \leq F_2$. Then

$$r(H) - 1 = \frac{1}{2}\sum_{i=1}^{4} i d_i(\Gamma_H) - \sum_{i=1}^{4} d_i(\Gamma_H) = d_4(\Gamma_H) + \frac{1}{2}d_3(\Gamma_H) - \frac{1}{2}d_1(\Gamma_H).$$

If $d_1(H) = 0$ then

$$r(H) - 1 = d_4(\Gamma_H) + \frac{1}{2}d_3(\Gamma_H).$$

This formula gives a powerful tool in investigating HNC. One can rewrite $(r(H) - 1)(r(K) - 1)$ as follows

$$(*) \quad (r(H)-1)(r(K)-1) = (d_4(\Gamma_H) + \frac{1}{2}d_3(\Gamma_H))(d_4(\Gamma_K) + \frac{1}{2}d_3(\Gamma_K))$$

$$= d_{4,H}d_{4,K} + \frac{1}{2}d_{4,H}d_{3,K} + \frac{1}{2}d_{3,H}d_{4,K} + \frac{1}{4}d_{3,H}d_{3,K}.$$

One can estimate $r(H\cap K)-1$ using the product automaton $\Gamma_H \times \Gamma_K$. Indeed, recall that $\Gamma_{H\cap K}$ is the core of the connected component of the graph $\Gamma_H \times \Gamma_K$ containing point $(1_H, 1_K)$. In any case $\Gamma_{H\cap K}$ is a subgraph of $\Gamma_H \times \Gamma_K$. Hence

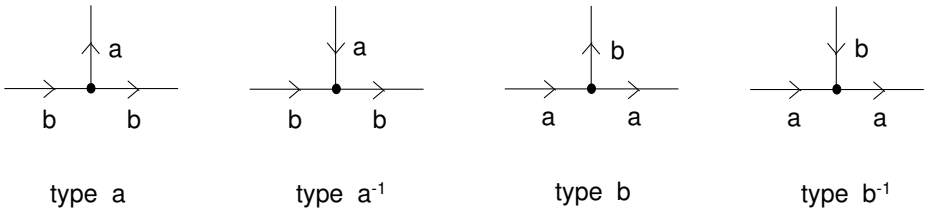
$$d_4(\Gamma_{H\cap K}) \leq d_4(\Gamma_H \times \Gamma_K),$$

$$d_3(\Gamma_{H\cap K}) \leq d_3(\Gamma_H \times \Gamma_K).$$

Clearly, a vertex $(u, v) \in \Gamma_H \times \Gamma_K$ has degree 4 iff vertices u, v have degree 4 in Γ_H, Γ_K. Hnece

$$d_4(\Gamma_H \times \Gamma_K) = d_4(\Gamma_H)d_4(\Gamma_K).$$

To estimate $d_3(\Gamma_H \times \Gamma_K)$ we need the following notation. There are four types of vertices of degree 3 in a graph Γ labelled by $\{a, b\}^{\pm 1}$:



Denote by $C_x(\Gamma)$ the number of vertices of type x in Γ. Clearly

$$d_3(\Gamma) = \sum_{x\in\{a,b\}^{pm1}} C_x(\Gamma).$$

Observe that if u, v are vertices of degree 3 in Γ_H, Γ_K then (u, v) has type x iff both u, v have type x. This shows that there are $C_x(\Gamma_H)C_x(\Gamma_K)$ vertices of type x in $\Gamma_H \times \Gamma_K$. Hence

$$d_3(\Gamma_H \times \Gamma_K) = d_4(\Gamma_H)d_3(\Gamma_K) + d_3(\Gamma_H)d_4(\Gamma_K) + \sum_x C_x(\Gamma_H)C_x(\Gamma_K).$$

To simplify notations we will write $d_{i,H}, C_{x,H}$ instead of $d_i(\Gamma_H), C_x(\Gamma_H)$. Now

$$(**) \quad r(H \cap K) - 1 = d_4(\Gamma_{H\cap K}) + \frac{1}{2}d_3(\Gamma_{H\cap K}) - \frac{1}{2}d_1(\Gamma_{H\cap K})$$

$$\leq d_4(\Gamma_{H\cap K}) + \frac{1}{2}d_3(\Gamma_{H\cap K})$$

$$\leq d_{4,H}d_{4,K} + \frac{1}{2}\left[d_{4,H}d_{3,K} + d_{3,H}d_{4,K} + \sum_x C_{x,H}C_{x,K}\right].$$

Clearly, if the righthand side of $(**)$ is less than the righthand side of $(*)$ then HNC holds for (H, K). This amounts to:

$$d_{4,H}d_{4,K} + \frac{1}{2}\left[d_{4,H}d_{3,K} + d_{3,H}d_{4,K} + \sum_x C_{x,H}C_{x,K}\right]$$

$$\leq d_{4,H}d_{4,K} + \frac{1}{2}d_{4,H}d_{3,K} + \frac{1}{2}d_{3,H}d_{4,K} + \frac{1}{4}d_{3,H}d_{3,K}.$$

Simplifying:

$$(***) \qquad \frac{1}{2}\sum_x C_{x,H}C_{x,K} \leq \frac{1}{4}d_{3,H}d_{3,K}.$$

We summarize the discussion above in the following lemma.

Lemma 12 *Let $H, K \leq F_2$. If Reduction 2 holds for H, K and*

$$\sum_x C_{x,H}C_{x,K} \leq \frac{1}{2}d_{3,H}d_{3,K}$$

then HNC holds for (H, K).

Now we deduce several corollaries from this result.

1. H.Neumann bound:

$$r(H \cap K) - 1 \leq 2(r(H) - 1)(r(K) - 1).$$

It sufficies to show that

$$\frac{1}{2}\sum_x C_{x,H}C_{x,K} \leq \frac{1}{4}d_{3,H}d_{3,K} + (r(H) - 1)(r(K) - 1).$$

From $(*)$ we see that

$$(r(H) - 1)(r(K) - 1) \geq \frac{1}{2}d_{3,H}d_{3,K}.$$

Hence

$$\frac{1}{2}d_{3,H}d_{3,K} + (r(H) - 1)(r(K) - 1) \geq d_{3,H}d_{3,K}.$$

But $\sum_x C_{x,H}C_{x,K} \leq \sum_x C_{x,H}d_{3,K} \leq d_{3,K}\sum_x C_{x,H} = d_{3,H}d_{3,K}$ as required.

2. R.Burns bound:

$$r(H\cap K)-1 \leq 2(r(H)-1)(r(K)-1) - \min\{r(H)-1, r(K)-1\}.$$

Again, it sufficies to show that

$$\frac{1}{2}\sum_x C_{x,H}C_{x,K} \leq \frac{1}{4}d_{3,H}d_{3,K} + (r(H)-1)(r(K)-1) - \min\{r(H)-1, r(K)-1\}.$$

By $(*)$:

$$\frac{1}{4}d_{3,H}d_{3,K} + (r(H)-1)(r(K)-1) =$$

$$= \frac{1}{2}d_{3,H}d_{3,K} + d_{4,H}d_{4,K} + \frac{1}{2}d_{4,H}d_{3,K} + \frac{1}{2}d_{3,H}d_{4,K}.$$

Since $\frac{1}{2}\sum_x C_{x,H}C_{x,K} \leq \frac{1}{2}d_{3,H}d_{3,K}$ to prove R.Burns' bound one needs to show that

$$(***) \qquad \min\{r(H)-1, r(K)-1\} \leq d_{4,H}d_{4,K} + \frac{1}{2}d_{4,H}d_{3,K} + \frac{1}{2}d_{3,H}d_{4,K}.$$

If $d_{4,H} \neq 0$ then $d_{4,H}d_{4,K} + \frac{1}{2}d_{4,H}d_{3,K} = d_{4,H}(d_{4,K} + \frac{1}{2}d_{3,K}) \geq r(K)-1$ and $(***)$ holds.

Similarly, for the case $d_{4,K} \neq 0$.

Now suppose $d_{4,H} = d_{4,K} = 0$. Then $r(H)-1 = \frac{1}{2}d_{3,H}, r(K)-1 = \frac{1}{2}d_{3,K}$ and it sufficies to show that

$$\frac{1}{2}\min\{d_{3,H}, d_{3,K}\} \leq \frac{1}{4}d_{3,H}d_{3,K}.$$

Suppose $d_{3,H} \leq d_{3,K}$. Then one needs to show

$$d_{3,H} \leq \frac{1}{2}d_{3,H}d_{3,K},$$

i.e.,

$$2 \leq d_{3,K}.$$

Observe that if $d_{3,K} = 0$ then Γ_K is a simple cycle and K is cyclic (hence HNC holds for (H,K)). Now, let $d_{3,K} \neq 0$.

Observe, that

$$2n = \sum_{i=1}^{4} id_i = 2d_2 + 3d_3$$

hence d_3 is even, i.e., $d_3 \geq 2$ as required.

One of the important devepopments toward proving HNC is due to W.Neumann. We start this discussion with the following sufficient condition for HNC from above:

$$\sum_x C_{x,H}C_{x,K} \leq \frac{1}{2}d_{3,H}d_{3,K}.$$

We rewrite it as follows:

$$d_{3,H}d_{3,K} - 2\sum_x C_{x,H}C_{x,K} \geq 0$$

$$d_{3,H}\sum_x C_{x,K} - 2\sum_x C_{x,H}C_{x,K} \geq 0$$

$$(\nabla) \qquad \sum_x (d_{3,H} - 2C_{x,H})C_{x,K} \geq 0.$$

Now if (H, K) is a counterexample to HNC then

$$(\nabla\nabla) \qquad \sum_x (d_{3,H} - 2C_{x,H})C_{x,K} < 0.$$

Hence there exists $x \in \{a, b\}^{\pm 1}$ such that

$$d_{3,H} - 2C_{x,H} < 0,$$

i.e.,

$$C_{x,H} \geq \frac{1}{2}d_{3,H},$$

i.e., there are more than half vertices of the type x in Γ_H.

Similarly, there exists $x \in \{a, b\}^{\pm 1}$ such that

$$C_{y,K} \geq \frac{1}{2}d_{3,K}.$$

Lemma 13 *(W.Neumann). Let $H, K \leq F_2$. If (H, K) is a counterexample to HNC then there exists $x \in \{a, b\}^{\pm 1}$ such that*

$$C_{x,H} \geq \frac{1}{2}d_{3,H}, C_{x,K} \geq \frac{1}{2}d_{3,K}.$$

Proof. Observe first that the argument in Reduction 2 shows that we can assume that $d_{1,H} = d_{1,K} = 0$. Indeed, the graph $\Gamma_{H^{uv}}$ and Γ_H, as well as $\Gamma_{K^{uv}}$ and Γ_K have the same vertices of degree 3 (since they have the same core graphs, just the strting point perhaps is different).

Now let for a given $x \in \{a, b\}^{\pm 1}$

$$C_{x,H} \geq \frac{1}{2} d_{3,H}.$$

For each $j = 1, 2, 3, 4$ consider a set

$$S_j = \{(a_1, \ldots, a_4) \in \mathbb{R}_+^4 \mid a_1 + \cdots + a_4 - 2a_j < 0, a_i \geq 0\}.$$

Put $S = S_1 \cup \cdots \cup S_4$. Fix a subgroup $H \leq F_2$ and enumerate

$$b_1 = C_{a,H}, b_2 = C_{a^{-1},H}, b_3 = C_{b,H}, b_4 = C_{b^{-1},H}.$$

Define a linear function

$$L_H : \mathbb{R}_+^4 \to \mathbb{R}_+, \quad \mathbb{R} = \{x \in \mathbb{R} \mid x \geq 0\}$$

$$L_H(a_1, \ldots, a_4) = \sum_i (d_{3,H} - 2b_i) a_i.$$

Put CE $= L_H^{-1}(-\infty, 0) = \{(a_1, \ldots, a_4) \mid L_H(a_1, \ldots, a_4) < 0\}$. Since $(-\infty, 0)$ is convex and L_H is linear then the set CE is convex. Observe that CE $\subseteq S$. Indeed, let $L_H(a_1, \ldots, a_4) < 0$. Then

$$L_H(a_1, \ldots, a_4) = \sum_i (d_{3,H} - 2b_i) a_i =$$

$$= \sum_i (\sum_i b_i - 2b_i) a_i =$$

$$= \sum_i (\sum_i a_i b_i - 2a_i b_i) =$$

$$= \sum_i (\sum_i a_i - 2a_i) b_i < 0.$$

Hence $\exists\ j\ \sum_i a_i - 2a_j < 0$, i.e., $(a_1, \ldots, a_4) \in S_j$.

Claim. $\exists\ j$ CE $\subseteq S_j$.

Proof. Indeed, suppose there are two vertices $u \in S_i, v \in S_j, i \neq j$ and $u, v \in$ CE. Since CE is convex, for any $t \in [0, 1]$

$$w_t = tu + (t - 1)v \in \text{CE},$$

if $t \to 1$, then $w_t \in S_i$, if $t \to 0$, then $w_t \in S_j$.

Fix u, v as above and put

$$L(t) = \pi_i w_t - \pi_j w_t,$$

where π_m is a projection onto m-th coordinate.

Clearly $t \to L(w_t)$ is a continuous function from $\mathbb{R}$ into $\mathbb{R}$.

Observe that

$$L(1) = \pi_i w_1 - \pi_j w_1 = \pi_i u - \pi_j u > 0,$$

(since $\pi_i u = u_i > \frac{1}{2}\sum u_i$)

$$L(1) = \pi_i v - \pi_j v < 0.$$

Hence there is a point $t_0 \in [0,1]$ such that $L(t_0) = 0$ then $\pi_i w_{t_0} = \pi_j w_{t_0}$. But then $w_{t_0} \notin S$ - contradiction.

Hence the claim holds.

Now observe that $u = (u_1, \ldots, u_4)$ with $u_i = 1, u_j = 0, j = 0$, belongs to CE:

$$\sum_j (d_{3,H} - 2b_i)a_j = d_{3,H} - 2b_i < 0.$$

Hence $j = i$.

3.6.2 Open problems related to HNC

1) Algorithmic HNC:

Is it true that there is an algorithm that for each positive integers m, n checks whether HNC holds or not for arbitrary (H, K) with $r(H) = m, r(K) = n$?

2) Fix rank m. How many topologically different foldings one can obtain by folding m arbitrary elements in F_2 ? Is this number finite ? If it is finite, then 1) holds.

Define by G_m the number of topologically different foldings of m-generated subgroups of F.

3) Spectrum problem.

Given $m, n \in \mathbb{N}$. Which numbers R can be realized as $R = r(H \cap K) - 1$ with $r(H) = m, r(K) = n$.

Denote by $\mathrm{Spec}(n, m) = \{e \mid e = r(H \cap K) - 1,\ r(H) = m,\ r(K) = n.$

4) $(n-1)(m-1) - 1 \in \mathrm{Spec}(n, m)$? i.e., H.Neumann bound minus 1.

5) Are there any *gaps* in $\mathrm{Spec}(n, m)$?

3.6.3 Strengthend HNC

Instead of asking about upper bounds of $r(H \cap K)$, one can ask about upper bounds of the rank of all intersections

$$x^{-1}Hx \cap y^{-1}Ky, \quad x, y \in F.$$

Clearly, any such intersection is conjugate to one of the form

$$H \cap z^{-1}Kz, \quad (z = yx^{-1}).$$

Moreover, if $y \in KxH$, then $H \cap x^{-1}Kx$ and $H \cap y^{-1}Ky$ are conjugate. Thus one needs only to estimate the ranks of intersections $H \cap K^x$., where$x$ runs through a set of representatives for double cosets KxH, $x \in F$. Furthermore, one needs only to consider such representatives x, for which $H \cap K^x \neq 1$. Denote such a set by T.

Strenthend Hanna Neumann Conjecture

$$\sum_{x \in T} [r(H \cap K^x) - 1] \leq (r(H) - 1)(r(K) - 1).$$

Exercise 1 *Show that all the results we discussed about HNC hold in the strenthend form.*

Hint: Observe that all the results we discussed above were about upper bounds for Euler characteristic of the product graph $\Gamma_H \times \Gamma_K$, i.e., about estimates for $d_4(\Gamma_H \times \Gamma_K), d_3(\Gamma_H \times \Gamma_K)$.

The following theorem gives a base for the exercise above.

Theorem 11 *Let H, K f.g. subgroups of F. Then for any $x \in F$ the subgroup $H \cap K^x$ is either trivial or it is conjugate to one of the subgroups from a finite set S:*

$$S = \left\{ \begin{array}{l} L \,|L = L(\Delta), \text{ where } \Delta \text{ is the core graph of one of the connected} \\ \quad |\text{components of the graph } \Gamma_H \times \Gamma_K \end{array} \right\}$$

Consider the following construction. Let $H, K \leq_{f.g.} F$. The graph $\Gamma_H \times \Gamma_K$ is finite, so if has only finitely many connected components, say $\Delta_1, \ldots, \Delta_m$. For each Δ_i denote by Δ_i^0 the core graph of Δ_i. Fix an arbitrary vertex of Δ_i. This turns Δ_i into a finite deterministic automaton, which accepts some subgroup, say L_i in F. Observe that L_i is finitely generated. It is not hard to see that l_i is of the form

$$L_i = H \cap K^{x_i},$$

for some $x_i \in F$.

3.7 Malnormal subgroups

Definition 17 *Let G be a group. A subgroup $H \leq G$ is called* malnormal *in G is for any $x \in G - H$, $H^x \cap H = 1$.*

Observe, that every malnormal subgroup of G is *centralizer-closed*, i.e., for any $1 \neq h \in H, C_G(h) \subseteq H$. In particular, every malnormal subgroup of G is *isolated*, i.e., for any $g \in G$ if $g^n \in H$, then $g \in H$.

Obviously, 1 and G are malnormal subgroups of G. If G is abelian, then 1 and G are the only malnormal subgroups of G.

Since malnormal subgroups are centralizer-closed then every abelian malnormal subgroup H of a group G must be maximal abelian, moreover for any $1 \neq h \in H$ $C_G(h) = H$. The reverse is also true if all proper centralizers in G are infinite cyclic.

Proposition 5 *[?]. Let G be a group in which every proper centralizer is infinite cyclic. Then every cantralizer of G is malnormal.*

Proof. Let G be a group in which every proper cantralizer is infinite cyclic. Then (see ????) G is commutative-transitive, i.e., commutation in G is transitive on the set $G - \{1\}$. Now let C be a proper centralizer in G. Then $C =< g >$ for some $g \in G$. Notice that g is not a proper power in G. Indeed, if $g = g_0^n$, then $[g \ \ g_0^n] = 1$ and $[g_0^n \ \ g_0] = 1$, hence $[g \ \ g_0] = 1$ (transitivity of commutation). It implies that $g_0 \in < g >$ and $n = \pm 1$.

Suppose now that

$$< g >^x \cap < g >= 1.$$

Then $(g^n)^x = g^m$ for some non-zero $m \ \ n \in \mathbb{Z}$. From transitivity of commutation we have $[g^x \ \ g] = 1$, therefore $g^x = g^k$ for some integer k. But then $g = (g^k)^{x^{-1}} = (g^{x^{-1}})^k$ and hence $n = \pm 1$ (g is not a proper power in G). If $n = 1$, then $[g \ \ x] = 1$ and $x \in < g >$. If $u = -1$, then $g^{x^2} = g$, so $[x^2 \ \ g] = 1$ and consequently $[x \ , \ g] = 1$, which implies $x \in < g >$. It follows that $< g >$ is malnormal in G.

Corollary 19 *Every centralizer in a free group is malnormal.*

Notice, that this result holds if one replaces "a free group" in the corollary above by a torsion-free hyperbolic group. Indeed, according to [**?**] every centralizer in a torsion-free hyperbolic group is cyclic.

Free factors provide another type of examples of malnormal subgroups in a free group F. To see this, let $F =< x_1 , \ldots , \ x_n >$ and $H =< x_1 , \ldots , \ x_k >$, $k < n$. Then for any element $h \in H$ and any element $f \in F$ which contains a

letter $x_j (j > k)$ in its reduced form, this letter x_j cannot be cancelled out in reducing the product $f^{-1}hf$. Hence for any $f \in F - H$ we have $H^f \cap H = 1$.

Now we are going to generalize the result above to arbitrary *retracts* of F. Recall, that a subgroup H is a retract of a group G if there exists a homomorphism $\varphi : G \to H$ s.t. $\varphi_H = id_H$.

Proposition 6 *Every retract of a free group F is malnormal in F.*

Proof. Let H be a refract of F and $\varphi : F \to H$ a homomorphism for which $\varphi_H = id_H$. Our proof rests just on the fact that all proper centralizers in F are infinite cyclic.

We prove first that every refract H is isolated in F. Indeed, if $g^n = h \in H$, then $(g^\varphi)^n = h^\varphi = h$, hence $(g^\varphi)^n = g^n$ in F. From transitivity of commutation in F we have $[g^\varphi\ ,\ g] = 1$ (unless $g = h = 1$), hence $g^\varphi = g$ (extraction of roots is unique in an infinite cyclic group).

It follows, that $g \in H$. Now suppose, $H^x \cap H \neq 1$, i.e., suppose $h^x = h_1$ for some $x \in F$ and $1 \neq h \in H$. Then, applying φ, we have $h^{\varphi(x)} = h_1$, therefore $h^{\varphi(x)} = h^x$ and $[h\ ,\ x\varphi(x)^{-1}] = 1$. since the centralizer $C_F(h)$ is cyclic and H is isolated in F, we have $x\varphi(x)^{-1} \in H$ and consequently $x \in H$. It shows that H is malnormal in F.

The argument above shows that the following result is also true.

Corollary 20 *Let G be a torsion-free hyperbolic group. Then every retract of G is malnormal.*

The following theorem shows that there exists a fast algorithm for detecting malnormality of finitely generated subgroups of a free group. To formulate the result we need the following definition.

Definition 18 *Let H be a subgroup of a group G. We call an element $f \in G - H$ a potential normalizer (pn) of H if $H^f \cap H \neq 1$.*

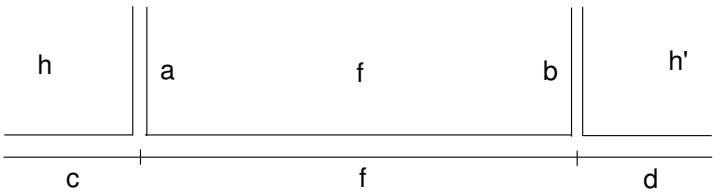
Denote by $pn(H)$ the set of all pn-elements of H. Clearly, H is malnormal in G iff $pn(H) = \emptyset$.

Observe, that of $f \in pn(H)$, then $HfH \subset pn(H)$.

Indeed, if $h^f \in H$, $h \in H$ and $f' = h_1 f h_2, (h_1\ ,\ h_2 \in H)$, then $(h_1 h h_1^{-1})^{f'} \in H$.

Theorem 12 *[?] Let F be a free group and H be a finitely generated subgroup of F with a Nielsen basis $Y = \{u_1\ , \ldots ,\ u_m\}$. Then H is not malnormal in F if there exists a pn-element $f \in pn(H)$ which is a product of two pieces of elements from $Y \cup Y^{-1}$ and each such piece has length at most $\frac{1}{2}\max\{|u_1|\ , \ldots ,\ |u_m|\}$.*

Proof. Let $f \in pn(H)$. We call f minimal if f has the minimal length in its double coset HfH. Notice, that if f is minimal, then for any $h \in H$ not more than half of h cancels out in fh and hf. Suppose H is not malnormal. Denote by f a minimal element in $pn(H)$. Then there are elements $1 \neq h\ ,\ h_1 \in H$ s.t. $f^{-1}hf = h_1$ or, equivalently, $hfh' = f$ $(h' = h_1^{-1})$. The following cancellation scheme shows that f has to cancel out completely in the product hfh'.

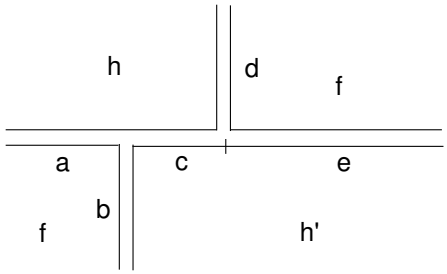


Indeed, since f is minimal we have

$$|a| \leq |c|,\ |b| \leq |d|.$$

From $|f| = |hfh'|$ we deduce that $|a| = |c|, |b| = |d|$, which implies $a = c\ ,\ b = d$. But then $h = aa^{-1} = 1$ - contradiction.

It follows that the cancellation scheme for $hfh'f^{-1} = 1$ is of the type:

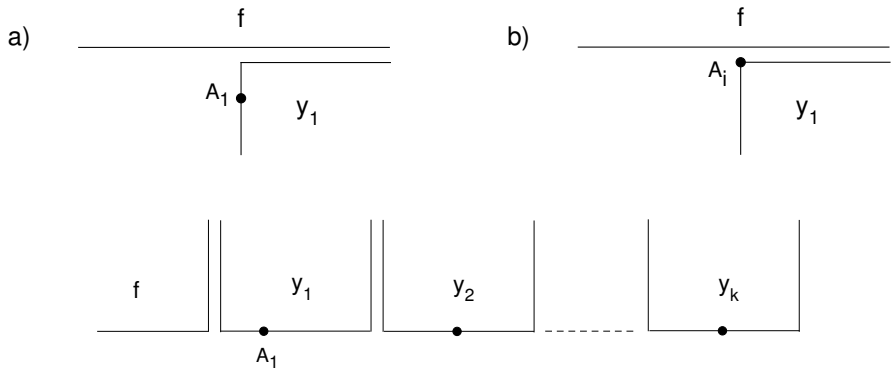


(it includes cases when one or more of the segements $a\ ,\ b\ ,\ c\ ,\ d\ ,\ e$ are of zero length).

To deal with this diagram consider first what kind of cancellation may occur in the product of the type

$$fy_1 \ \cdots \ y_k \quad , \quad y_i \in Y \cup Y^{-1}.$$

Denote by A_i the midpoint of y_i (if y_i is of odd length, then A_i is in the middle of the segment presenting the middle letter in y_i). Since f is minimal, we may have only two possible types of cancellation:



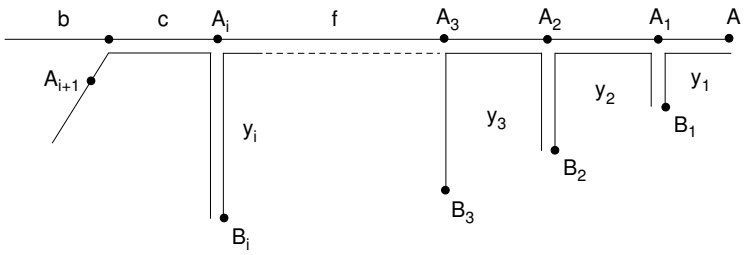
In the first case the cancellation scheme for $fy_1 \cdots y_k$ looks like

So here the part of f that cancels out completely is an initial piece of y_1 of length $\leq \frac{1}{2}|y_1|$.

In the second case exactly half of y_1 cancels out in fy_1. Then $|f| = |fy_1|$ and the element $f_1 = fy_1$ is also minimal in $pn(H)$. Notice that

$$fy_1 \cdots y_k = f_1 y_2 \cdots y_k$$

By induction on k we can assume that either the cancellation in $f_1 y_2$ is of the type a) above, and in this case the part of f_1 that cancels out is an initial piece of y_2 of length $\leq \frac{1}{2}|y_2|$; or the cancellation is of the type b). In this case, by induction, there exists an index i, $2 \leq i \leq k$, such that the cancellation in $f_1 y_2 \cdots y_k$, and hence $fy_1 \cdots y_k$, goes along the following scheme



(it includes cases where b or c (or both) are of zero length). Moreover, in this event we have

$$|A_1 B_1| = |A_1 A| \ , \ |A_2 B_2| = |A_2 A| \ , \ldots , \ |A_i B_i| = |A_i A|.$$

Now if we put $f_i = fy_1 \cdots y_i$, then f_i is a minimal element in $pn(H)$. Notice that the part of f_i that cancels out in $f_i y_{i+1} \cdots y_k$ is an initial piece of y_{i+1} of length $\leq \frac{1}{2}|y_{i+1}|$.

Consider now cancellation in the product $hfh', (h\ ,\ h' \in H)$.

We have $h = z_1\ \cdots\ z_m, h' = y_1\ \cdots\ y_k$ for some $y_i\ ,\ z_j \in Y \cup Y^{-1}$. Then we can rewrite the product

$$z_1\ \cdots\ z_m f y_1\ \cdots\ y_k$$

in the form

$$z_1\ \cdots\ z_m(f y_1\ \cdots\ y_i)y_{i+1}\ \cdots\ y_k = z_1\ \cdots\ z_m f_i y_{i+1}\ \cdots\ y_k.$$

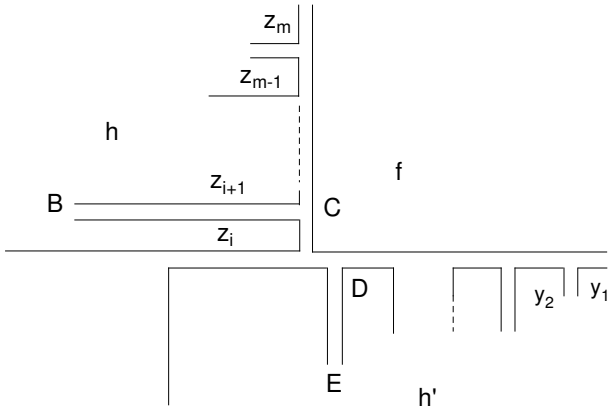
Cancellation in $z_1\ \cdots\ z_m f_i$ is similar to the cancellation in $f y_1\ \cdots\ y_k$. So from the argument above we can rewrite the product

$$z_1\ \cdots\ z_m f_i y_{i+1}\ \cdots\ y_k = z_1\ \cdots\ z_j(z_{j+1}\ \cdots\ z_m f_i)y_{i+1}\ \cdots\ y_k$$

such that $h = z_{j+1}\ \cdots\ z_m f y_1\ \cdots\ y_i$ is minimal in $pn(H)$ and $h = a \circ b$, where a cancels out in $z_j h$, and b cancels out in $h y_{i+1}$.

It follows, that a a is a piece of z_j, of length $\leq \frac{1}{2}|z_j|$ and b is a piece of y_{i+1} of length $\leq \frac{1}{2}|y_{i+1}|$.

Geometrically h can be seen in the following picture:



Corollary 21 *Let F be a free group. There exists an algorithm which determines whether a finitely generated subgroup $H \leq F$ (given by a finite set of generators) is malnormal or not.*

$\triangleright$. Let $H = \langle h_1\ , \ldots,\ h_n \rangle \leq F$.

Denote $M = \max\{|h_1|\ , \ldots,\ |h_n|\}$. In Section ???? we proved that one can effectively find a Nielsen basis $Y = \{u_1\ , \ldots,\ u_m\}$ for H in $O(nM)^2$ steps. Notice, that

$$\max\{|u_1|\ , \ldots,\ |u_m|\} \leq M.$$

Now, according to the theorem above, the subgroup H is malnormal if and only if there are no pn-elements among all products ab, where a and b are pieces of elements from $Y \cup Y^{-1}$ of length $\leq \frac{1}{2}M$. It takes $O(nM)$ steps to enumerate all pieces $a_1, \ldots, a_k$ of elements from $Y \cup Y^{-1}$ and $O(n^2M^2)$ steps to enumerate all products of the type $a_i a_j$, $1 \leq i, j \leq k$. Now, for each such product $a_1 a_j$ we can check effectively whether $H \cap H^{a_i a_j} = 1$ or not. For this we need to construct automata $\Gamma(H) \times \Gamma(H^{a_i a_j})$ and to check whether the connected component of this automata, containing the vertex $(1, 1)$, is a tree or not. As we mentioned in Section ???? it takes $O(|h_1| + \cdots + |h_n|)^2$ steps to construct the automaton $\Gamma(H)$. Similarly, to construct $\Gamma(H^{a_i a_j})$ we need $O(|h_1^{a_i a_j}| + \cdots + |h_n^{a_i a_j}|)^2 = O(nM^2)^2$ steps. Thus to construct $\Gamma(H)$ and $\Gamma(H^{a_i a_j})$ we need $O(nM^2)^2$ steps. To construct the product of two automata Γ and Δ we need $O(|E(\Gamma) \times E(\Delta)|)$ steps, the same amount of steps one needs to construct a maximal subtree of $\Gamma \times \Delta$ (?). In our case it comes down to $O(nM \cdot nM) = O(n^2M^2)$ steps. Thus for each pair $i, j, 1 \leq i, j \leq k$ we need $O(n^2M^4) \cdot O(n^2M^2) = O(n^4M^6)$ steps to check whether $a_i a_j$ is a pn-element or not. As we noticed above $k \leq n^2M^2$, hence altogether it takes

$$O(n^4M^6) \cdot n^2M^2 = O(n^6M^8)$$

steps to check whether H is malnormal or not.

Exercise 2 *Check whether the argument above is correct.*

Now we will give a description of finitely generated malnormal subgroups of a free group $F = F(X)$ in terms of graphs. This description is based on the following observation which is due to J. Stallings (or Imrich ?). Let H and K be free group subgroups of F, and f be an element of F. Denote

$$\Delta = \Gamma(H) \times \Gamma(K),$$

$$\Sigma = \Gamma(H) \times \Gamma(K^f).$$

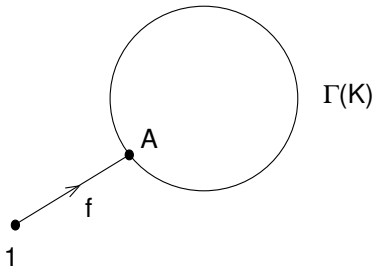
If $u \in H$ and $v \in K$ $(v \in K^f)$ then by $\Delta_{u,v}(\Sigma_{u,v})$ we denote the connected component of $\Delta(\Sigma)$ containing (u, v). We say that a word $W \in F$ is readable in $\Delta_{u,v}(\Sigma u, v)$ if there exists a path in $\Delta_{u,v}(\Sigma u, v)$ which starts at (u, v) and has the label w. Now observe the automaton $\Gamma(K^f)$ can be obtained from the graph

by elementary transformations, so $\Gamma(K^f)$ is one of the following two types:

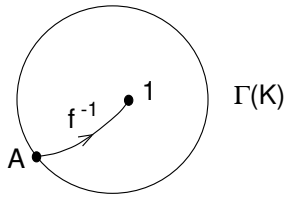
this corresponds to the case when f^{-1} is readable in $\Gamma(K)$;

in this case $f = f_1 f_2$, $(f_1 \neq 1)$ and f_2 is the longest terminal subword of f such that f_2^{-1} is readable in $\Gamma(K)$.

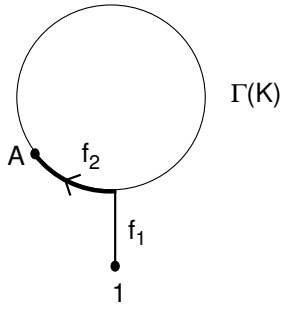
Then the following holds:



a)



b)



1) suppose $\Gamma(K^f)$ is of the type a). Then

$$\Sigma_{1,1} = \Delta_{1,v},$$

where v is the end-point of the unique path in $\Gamma(K)$ starting at 1 and with the label f^{-1}.

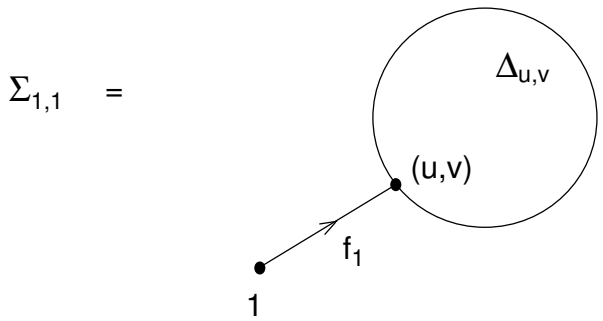
2) suppose $\Gamma(K^f)$ is of the type b). There are two subcases here. If f_1 is not readable in $\Gamma(H)$, then

$$\Sigma_{1,1} = \begin{array}{l}\text{the arc which corresponds to the longest initial segment} \\ \text{of } f_1 \text{ that is readable in } \Gamma(H)\end{array}$$

If f_1 is readable in $\Gamma(H)$ and u is the end-point of the path in $\Gamma(H)$ with the label f_1 starting at 1, then

$$\Sigma_{1,1} = \Delta_{u,v} \cup \left\{ \begin{array}{l}\text{an arc from a new vertex to the} \\ \text{vertex } (u,v) \text{ with the label } f_1\end{array} \right\}$$

here v is the end-point of the path in $\Gamma(K)$ starting at 1 and with the label f_2^{-1}.



It follows immediately from the observation above that the subgroup accepted by the automaton $\Sigma_{1,1}$ is conjugate to the subgroup accepted by the automaton $\Delta_{u,v}$.

Summarizing the discussion above we have the following theorem:

Theorem 13 *Let H and K be finitely generated subgroups of F. Then for any $f \in F$ the subgroup $H \cap K^f$ is conjugate to one of the finitely many subgroups of F, which correspond to the connected components of the graph $\Gamma(H) \times \Gamma(K)$.*

Corollary 22 *(Imrich). Let H and K be finitely generated subgroups of F. Then one can effectively find a finite set of elements $f_1, \ldots, f_k$ such that for any $f \in F$ the subgroup $H \cap K^f$ is conjugate to one of the subgroups $H \cap K^{f_i}, i = 1, \ldots, k$.*

Proof. In the view of the theorem above it suffices to notice that the component $\Delta_{u,v}$ of the graph $\Gamma(H) \times \Gamma(K)$ accepts exactly the subgroup $H \cap K^f$, where f is as above.

Our next goal is to clarify the structure of the fundamental groups of connected components of the graph $\Gamma(H) \times \Gamma(K) = \Delta$. As before, by $\Delta_{u,v}$ we denote the connected component of Δ containing (u, v). The graph $\Delta_{u,v}$ satisfies the one-way reading property, so if we chose the point (u, v) as the starting point in $\Delta_{u,v}$, then the graph $\Delta_{u,v}$ gives rise to a finite automaton. This automaton (we denote it by $\Delta_{u,v}$) accepts a subgroup of F, which is called the fundamental group of graph $\Delta_{u,v}$ (with the distinguished point (u, v)). We denote this subgroup by $\pi(\Delta_{u,v})$.

Lemma 14 *Let H, K be finitely generated subgroups of F, $\Delta = \Gamma(H) \times \Gamma(K)$, and $\Delta_{u,v}$ a connected component of Δ, containing a vertex (u, v). Then*

$$\pi(\Delta_{u,v}) = H^f \cap K^g,$$

where f is the label of some path in $\Gamma(H)$ from 1 to u, and g is the label of some path in $\Gamma(K)$ from 1 to v.

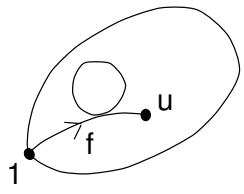
Proof. Suppose a word w is accepted by $\Delta_{u,v}$:

it translates into $\Gamma(H)$ and $\Gamma(K)$ as in the picture above. But then

$$fwf^{-1} \in H, gwg^{-1} \in K.$$

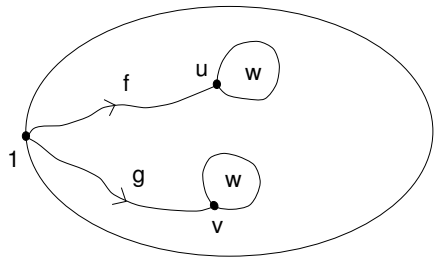
Hence $w \in H^f \cap K^g$, and it shows that $\pi(\Delta_{u,v} \subseteq H^f \cap K^g$. The reverse is also true. Indeed, let $w \in H^f \cap K^g$. Then $w^{f^{-1}} \in H$ and $w^{g^{-1}} \in K$. Notice, that $w^{f^{-1}}$ means that we have the following picture:

Maybe the easiest way to see this, is to notice that since f is readable in H then if H accepts the reduced word $w^{f^{-1}}$, then H accepts the word fwf^{-1}, which means $\Gamma(H)_u$ accepts w.



Theorem 14 *Let H be a finitely generated subgroup of F. Then H is malnormal in F if and only if all the connected components of $\Delta = \Gamma(H) \times \Gamma(H)$, other than $\Delta_{1,1}$, are trees.*

Proof. Suppose the component $\Delta_(u, v)$ for some (u, v) is not a tree, and suppose $\Delta_{u,v} \neq \Delta_{1,1}$, i.e., $u \neq v$. Then there exists a circuit, say p, in $\Gamma(H)$ starting at u with some label w, and also there exists another circuit in $\Gamma(H)$ at v with the same label w. Let f be the label of some path from 1 to u and g be the label of some path from 1 to v in $\Gamma(H)$. Then we have the following picture:



Then we have seen already

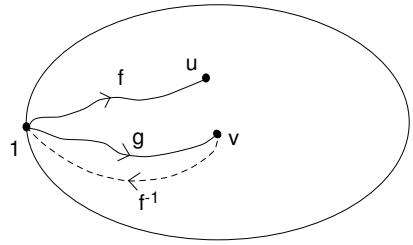
$$w \in H^f \cap H^g \Leftrightarrow w \in H \cap H^{gf^{-1}}.$$

We claim that $gf^{-1} \notin H$.

Suppose $gf^{-1} \in H$ then gf^{-1} is accepted by $\gamma(H)$, hence there is a path q from v to 1 with label f^{-1}.

But then q^{-1} has to coincide with the path p (they have the same label f). It would imply $u = v$ - contradiction. Hence H is not malnormal in F. Suppose now, that $H \cap H^f \neq 1$ for some $f \in F$. Then $H \cap H^f$ is isomorphic to the fundamental group of some connected component of $\Gamma(H) \times \Gamma(H)$, hence this connected component is not a tree.

Remark 2 *Notice that the argument in this first part of the proof above allows one to find a pn-element for H provided H is not malnormal. Moreover,*



it gives the same description of minimal pn-elements as products ab where a and b are pieces of some elements from $Y^{\pm 1}$*, where* Y *is an arbitrary Nielsen basis for* H*.*

Let H be a subgroup of a group G. Denote by $mal(H)$ the intersection of all malnormal subgroups of G containing H:

$$mal(H) = \cap\{K | H \leq K \leq G\ , K \leq_{mal} G\}.$$

It is easy to see that $mal(H)$ is malnormal in G, hence $mal(H)$ is the smallest malnormal subgroup of G containing H. We call $mal(H)$ the malnormal closure of H in G.

Theorem 15 *Let* H *be a finitely generated subgroup of a free group* F*. Then*

$$rank(mal(H)) \leq rank(H).$$

Moreover, there exists an algorithm to find a finite generating set for $mal(H)$*.*

Proof. The proof is based on the existence of the algorithm which determines whether a free group subgroup of F is malnormal or not, and if not it finds a *pn*-element for the subgroup.

Now, let H be a f.g. group subgroup of F. If H is malnormal, then $mal(H) = H$ and there is nothing to prove. Otherwise, we can effectively find a *pn*-element, say x_1, for H, i.e., $H^{x_1} \cap H \neq 1$. Put $H_1 = \langle H\ ,\ x_1 \rangle$. Notice that $x_1 \in mal(H)$. Indeed, $mal(H)^{x_1} \cap mal(H) \geq H^{x_1} \cap H \neq 1$, hence $x_1 \in mal(H)$. By induction, we either construct a strictly increasing chain of subgroups $H_i\ ,\ H_0 = H$, and elements $x_i \in F$ s.t. $H_{i+1} = \langle H_i\ ,\ x_{i+1} \rangle$

$$H = H_0 < H_1 < H_2 < \ \cdots,$$

where for each $i,\ x_i \in mal(H)$, i.e., $H_i \leq mal(H)$. Notice that for each i

$$rank(H_{i+1}) \leq rank(H_i).$$

Indeed, suppose H_i has a free basis $y_1, \ldots, y_n$. Then H_{i+1} is generated by $y_1, \ldots, y_n, x_{i+1}$, hence it has rank at most $n+1$.

If H_{i+1} has rank $n+1$, then $y_1, \ldots, y_n, x_{i+1}$ is a free basis for H_{i+1} hence H_i is a free factor in H_{i+1}, therefore H_i is malnormal in H_{i+1} - contradiction with the choice of x_{i+1}. It follows that

$$rank(H_{i+1}) \leq rank(H_i) \leq \cdots rank(H_0).$$

Since every ascending chain of subgroups of a free group of uniformly bamded rank stabilizes, then for some i, H_i is malnormal in F, hence $H_i = mal(H)$. In particular, $mal(H) = \langle H, x_1, \ldots, x_i \rangle$.

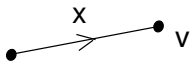
3.8 Marshall Hall's Theorem

Theorem 16 *(Marshall Hall) Let H be a f.g. subgroup of $F(X) = F(x_1, \ldots, x_n)$ then there exists a subgroup L of finite index in $F(X)$, such that H is free factor of L, $L = H * K$.*

Lemma 15 *(Back-step lemma). Suppose Γ is a folded finite graph , where edges are labeled by X. Let $x \in X \cup X^{-1}$ and $v \in V\Gamma$ be such that*

1) there exists an edge labeled x with end-point v;

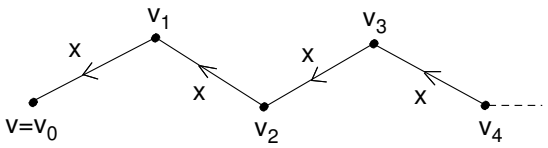
2) there exists no edge labeled x with origin v.

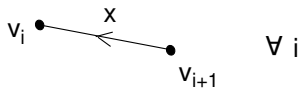


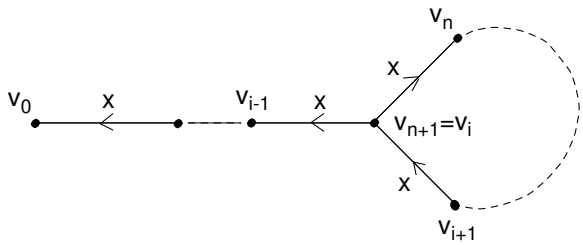
Then there exists another vertex $u \neq v$ of Γ such that there is no edge labeled x with end-point u in Γ.

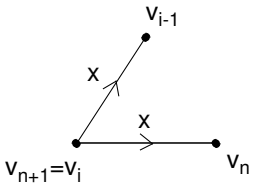
Proof. Suppose no such u exists, that is every vertex of Γ is an end-point of an edge labeled x.

Put $v_0 = v$, put v_1 to be the original of an x-labeled edge terminating at v, put v_2 to be the original of an x-labeled edge terminating at v_1, etc.









Put n to be the maximum number such that all vertices $v_0, v_1, \ldots, v_n$ are distinct (it exists since $\Gamma(H)$ is finite). Then $v_{n+1} = v_i$ for some $0 \leq i \leq n$. $i \neq 0$ since by assumption v_0 has no originating edges labeled x:

Then v_i has two edges labeled x with origin v_i:

$v_{i-1} \neq v_n$ by the choice of $n \Rightarrow \Gamma(H)$ is not folded-contradiction.

Proof of M.Hall's Theorem. By Corollary ?? and Theorem ?? it is enough to show that $\Gamma(H)$ can be extended to an X-complete connected folded graph (which will be $\Gamma(L)$). For each $v \in V\Gamma(H)$ define

$$def\ v = \left\{ \begin{array}{l} \text{the number of } x \in X \cup X^{-1} \text{ such that there's no edge} \\ \text{in } \Gamma(H) \text{ labeled } x \text{ with origin } v \end{array} \right\}$$

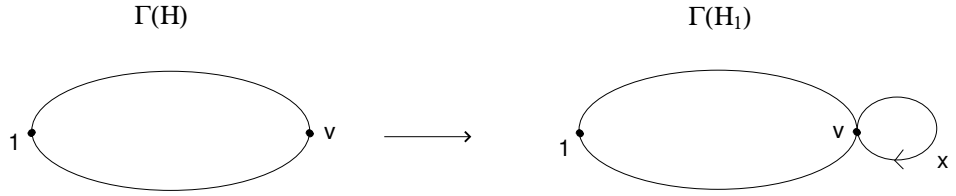
$$def\ V\Gamma(H) = \Sigma_{v \in V\Gamma(H)}\ def\ v\ -\ \text{deficit of } \Gamma(H)$$

$\Gamma(H)$ is X-complete if $def\ \Gamma(H) = 0$. If $def\ \Gamma(H) = 0$, then $|F(X) : H| < \infty$, $H = L$ and there's nothing to prove.

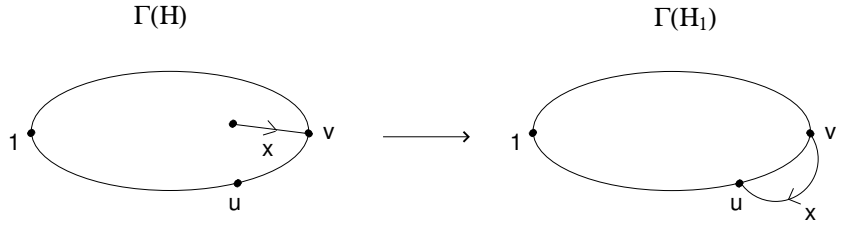
Suppose $def\ \Gamma(H) = d > 0$. We will prove the theorem by induction on $def\ \Gamma(H)$. If $def\ \Gamma(H) = d > 0$ and the theorem holds for all subgroups, whose graph has smaller deficit, we chose $v \in V\Gamma(H)$ such that $def\ v > 0$. Then $\exists x \in X \cup X^{-1}$ s.t. no x-labeled edge originates at v.

Case 1. Suppose that there's no x-labeled edge ending at v in $\Gamma(H)$. Then add a loop to $\Gamma(H)$ to get $\Gamma(H_1)$.

Clearly the new graph is folded and $def\ \Gamma(H_1) < def\ \Gamma(H) \Rightarrow$ by induction $\Gamma(H_1)$ can be extended to an X-complete connected folded graph $\Gamma(L) \Rightarrow$ theorem follows.



Case 2. Suppose there's an edge with endpoint v and label x in $\Gamma(H)$ (but no edge with origin v and label x). Then by Back-step Lemma there's a vertex $u \in V\Gamma(H)$ s.t there's no x-labeled edge in $\Gamma(H)$ which ends at $u \Downarrow$ add such an edge to get $\Gamma(H_1)$.

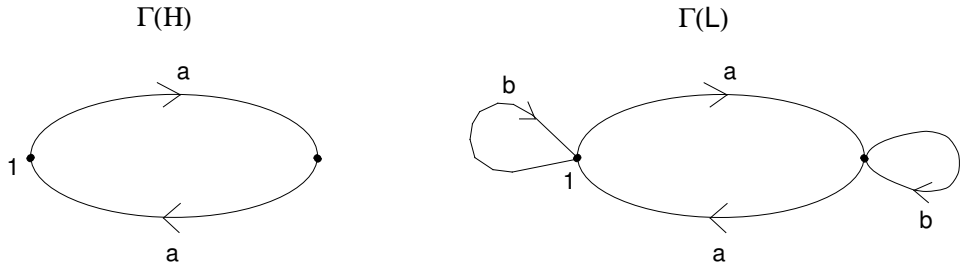


Then new graph $\Gamma(H_1)$ is connected, folded, contains $\Gamma(H)$ and has smaller deficit

$$def\ \Gamma(H_1) < def\ \Gamma(H)$$

$\Rightarrow$. Induction applies to $\Gamma(H_1)$ and Theorem follows.

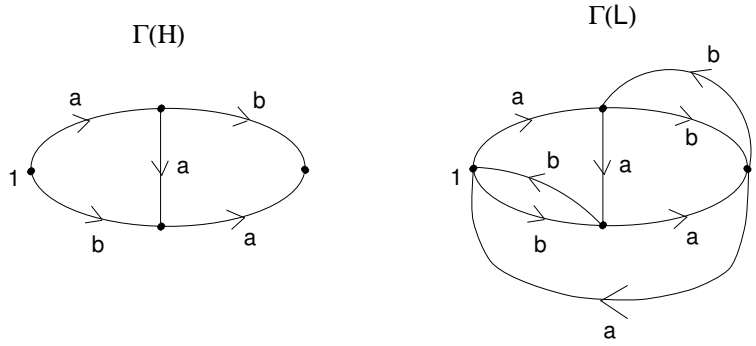
Example 9 $H = \langle a^2 \rangle \leq F(a\ ,\ b)$,



$L = \langle b\ ,\ aba^{-1}\ ,\ a^2 \rangle$ *has index* 3 *in* $F(a\ ,\ b)$.

Example 10 $H = \langle a^2b^{-1}\ ,\ aba^{-1} \rangle$

$L = \langle a^2b^{-1}\ ,\ aba^3\ ,\ a^2b\ ,\ ab^2a^{-1}\ ,\ aba \rangle$ *has index* 4 *in* $F(a\ ,\ b)$. $L = H * K\ ,\ K = \langle a^2b\ ,\ ab^2a^{-1}\ ,\ aba \rangle$.



Remark 3 *Note that index of L equals the number of vertices of $\Gamma(H)$ since $\Gamma(L)$ has the same set of vertices as $\Gamma(H)$, only edges are added to get $\Gamma(L)$ from $\Gamma(H)$. Also $\Gamma(L)$ has the same maximal tree as $\Gamma(H)$ does.*

3.9 Principal quotients and subgroups

In this section we want to investigate more carefully the relationship between $\Gamma(H)$ and $\Gamma(K)$ if $K \leq H$. To simplify the matters we will assume that both H and K are finitely generated, although it will be clear that one can consider a more general situation.

Definition 19 *Let $K \leq H \leq F(X)$ be two finitely generated subgroups of $F(X)$. Since $K = L(\Gamma(K), 1_K) \subseteq L(\Gamma(H), 1_H) = H$, for any path p in $\Gamma(K)$ from 1_K to 1_K there is a unique path $\alpha[p]$ in $\Gamma(H)$ from 1_H to 1_H such that the label of $\alpha[p]$ is the same as that of p.*

Define a subgraph $\Gamma_H(K)$ of $\Gamma(H)$ as

$$\Gamma_H(K) = \cup\{\alpha[p] \ , | \ , p\text{is a path in } \Gamma(K) \text{ from } 1_K \text{ to } 1_K\}.$$

In fact, constructing $\Gamma_H(K)$ can be considerably simplified, as the following lemma shows.

Lemma 16 *Let $K \leq H \leq F(X)$ be finitely generated subgroups of $F(X)$. Let $K = \langle k_1, \ldots, k_m \rangle$ where k_i are freely reduced words in $F(X)$. For each $i = 1, \ldots, m$ let p_i be the loop at 1_K in $\Gamma(K)$ with label k_i. Then*

$$\Gamma_H(K) = \cup_{i=1}^{n} \alpha[p_i]$$

The above lemma shows that in practical terms it is very easy to construct $\Gamma_H(K)$, given $\Gamma(K)$ and $\Gamma(H)$. For example, we can choose as a finite generating set for K the set T_K corresponding to some spanning tree in $\Gamma(K)$.

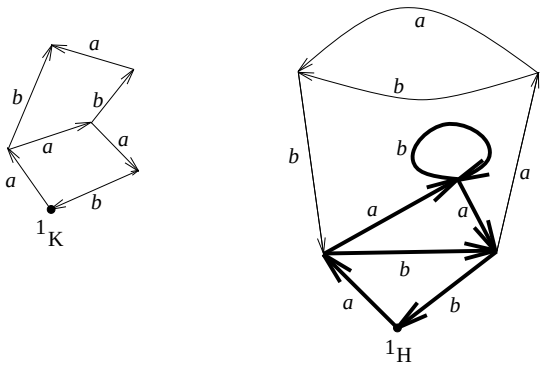


Figure 1: An example of $\Gamma_H(K)$. Here $K \leq H \leq F(a,b)$ and the graph $\Gamma_H(K)$ is highlighted. We suggest that the reader uses Lemma 16 to verify the validity of this picture

Lemma 17 *Let $K \leq H \leq F(X)$. Then there is a unique morphism of based X-digraphs $\alpha : (\Gamma(K), 1_K) \longrightarrow (\Gamma(H), 1_H)$; for this morphism we have $\alpha(\Gamma(K)) = \Gamma_H(K)$.*

Proof.

The existence and uniqueness of α follow from Proposition **??**.

It is obvious from the definition of $\Gamma_H(K)$ and the fact that any edge in $\Gamma(K)$ lies on a reduced path from 1_K to 1_K that $\alpha(\Gamma(K)) = \Gamma_H(K)$. $\triangleright$

Lemma 18 *Let $K \leq H \leq F(X)$. Then the graph $\Gamma_H(K)$ is a connected folded graph which is a core graph with respect to 1_H.*

Proof. It is obvious that $\Gamma_H(K)$ is folded, connected and that it contains 1_H. Let $\alpha : \Gamma(K) \longrightarrow \Gamma(H)$ be the unique morphism such that $\alpha(1_K) = 1_H$, so that $\alpha(\Gamma(K)) = \Gamma_H(K)$ (the existence of α follows from Lemma 17.

Assume that $\Gamma_H(K)$ is not a core-graph with respect to 1_H and that there is a vertex $u \neq 1_H$ of degree one on $\Gamma_H(K)$. Let e' be the unique edge of $\Gamma_H(K)$ with terminus u. Let e be an edge of $\Gamma(K)$ such that $\alpha(e) = e'$. Since $\Gamma(K)$ is a core graph with respect to 1_K, there exists a reduced loop p at 1_K passing through p. Denote the label of p by w. Since p is reduced and $\Gamma(K)$ is folded, w is freely reduced. The path $\alpha(p)$ is a loop at 1_H in $\Gamma_H(K)$ with label w and $\alpha(p)$ passes through e'. Since w is freely reduced and $\Gamma_H(K)$ is folded, the path $\alpha(p)$ is reduced. However, this contradicts our assumption that u has degree one in $\Gamma_H(K)$. $\triangleright$

Recall that by Lemma **??** (morphism factorization) an epimorphic image of an X digraph Γ can be obtained by first identifying some vertex subsets

of Γ into single vertices and then folding some edges with the same initial vertices and the same terminal vertices (so that the vertex set of the graph does not change at this step).

Definition 20 (Principal quotients) *Let K be a finitely generated subgroup of $F(X)$. Since $\Gamma(K)$ is finite, there exist only finitely many based X-digraphs $(\Gamma_1, v_1), \dots (\Gamma_s, v_s)$ such that:*
(a) each Γ_i is a finite connected folded X-digraph which is a core-graph with respect to v_i;
(b) for each $i = 1, \dots, s$ there is an epimorphism $f : \Gamma(K) \longrightarrow \Gamma_i$ with $f(1_K) = v_i$.

These graphs $(\Gamma_1, v_1), \dots (\Gamma_s, v_s)$ are called principal quotients *of $\Gamma(K)$.*

Remark 4 *Note that if (Γ, v) is a principal quotient of $\Gamma(K)$ then $(\Gamma, v) = (\Gamma(H), 1_H)$ for some finitely generated subgroup H of $F(X)$ with $K \leq H$. Moreover, in this case $\Gamma_H(K) = \Gamma$.*

Lemma 19 *Let F be a free group (of possibly infinite rank) and let K be a finitely generated free factor of F. Let $H \leq F$ be a finitely generated subgroup such that $K \leq H$. Then K is a free factor of H.*

Proof. Let $Y = Y_1 \sqcup Y_2$ be a free basis of F such that Y_1 is a finite free basis of K. Since H is finitely generated and involves only finitely many letters from Y_2, we may assume that Y_2 is also finite. It is clear that in this case the graph $\Gamma^Y(K)$ is just a wedge of circles, labeled by elements of Y_1, wedged at a vertex 1_K. This graph clearly has no epimorphic images other than itself, so $\Gamma^Y(K) = \Gamma^Y_H(K) \subseteq \Gamma^Y(H)$. Since $(\Gamma^Y(K), 1_K)$ is a subgraph of $(\Gamma^Y(H), 1_H)$, this implies that K is a free factor of H, as required. $\triangleright$

3.10 Algebraic extensions

Let F be a fixed free group with basis X. Let H, K be finitely generated (f.g.) subgroups of F. We say that K is an *extension* of H if $H \leq K$.

In this section we develop a theory of algebraic extensions of finitely generated subgroups of F which is analogues to the theory of algebraic extensions of fields.

Definition 21 *1) An extension $H \leq K$ is called free (or purely transcendental) if H is a free factor of K;*

2) An extension $H \leq K$ is called algebraic if H is not a subgroup of a proper free factor of K;

3) If $H \leq K$ is not algebraic then it is called transcendental, in this case $K = K' * C$ for some $K' \geq H$ and $C \neq 1$.

The following result gives a characterization of algebraic extensions in terms of graphs.

Let $H \leq K$. Recall (Section ???) that there exists a unique homomorphism of graphs

$$\alpha_{H,K} : \Gamma_H \to \Gamma_K.$$

The image $\alpha_{H,K}(\Gamma_H)$ is called the principal quotient of H inside Γ_K. The subgroup

$$L_{H,K} = L(\alpha_{H,K}(\Gamma_H))$$

is called the principal extension of H in K.

Lemma 20 *If $H \leq K$ is algebraic then*

$$\alpha_{H,K}(\Gamma_H) = \Gamma_K.$$

Proof. If $\alpha_{H,K}(\Gamma_H) \neq \Gamma_K$ then $L_{H,K}$ is a proper free factor of K and $H \leq L_{H,K}$ is trabscendental. This contradiction shows that $\alpha_{H,K}(\Gamma_H) = \Gamma_K$ as required.

Remark 5 *Observe that not every principal extension of H is algebraic.*

Lemma 21 *Let $H \leq K$ be f.g. subgroups of F. Then there exists an algebraic extension $H \leq L$ such that L is a free factor of K.*

Proof. Let L be a free factor of K of minimal possible rank containing H. Since H is f.g. such L exists. Thus

$$K = L * C.$$

If $H \leq L$ is transcendental then

$$L = L' * C'$$

for some $L' \geq H$ and $C' \neq 1$. Hence $rank(L') < rank(L)$ - contradiction.

Theorem 17 *Let $H \leq_{f.g.} K$. Then:*

1) there are only finitely many algebraic extensions of H;

2) one can find all algebraic extensions of H algorithmically;

3) one can check algorithmically whether a given extension $H \leq K$ is algebraic or not;

4) one can check algorithmically whether or not a given extension $H \leq K$ is free or not.

Proof. 1) By Lemma 20 every algebraic extension is a principal extension. We prove in Section ??? that there are only finitely many principal extensions. This proves 1).

2) In view of 1) one has to check which principal extensions are algebraic. Since one can find all principal extensions of H algorithmically then 2) follows from 3).

3) Let $H \leq K$. List all principal extensions $L_1, \ldots, L_k$ algorithmically and check which of them are subgroups of K. For the latter check whether there exists a principal extension of H which is a free factor of K. If there exists such then $H \leq K$ is obviously transcendental, if not then by Lemma 21 $H \leq K$ is algebraic. Now 3) follows from 4).

And 4) follows from Whitehead algorithm.

Lemma 22 *Let $H \leq K, F = K * C$. Then if $H \leq L$ is algebraic then $L \leq K$.*

Proof. Suppose $L \not\leq K$. Since $F = K * C$ then there exists a basis X of F such that $X = X_K \cup X_C$ where $X_K.X_C$ are bases in K, C correspondingly. Since $L \not\leq K$ then Γ_L contains an edge with some label from X_C. Similarly, Γ_H contains only edges with labels from X_K, as well as each principal quotient of Γ_H. This implies that

$$\alpha_{H,L}(\Gamma_H) \neq \Gamma_L.$$

Hence by Lemma 20 $H \leq L$ is not algebraic - contradiction, which proves the lemma.

Theorem 18 *Let $H \leq_{f.g.} F$. Then:*

1) there exists a unique maximal algebraic extension of H in F (we denote it by $acl(H)$);

2) $acl(H)$ is a free factor of F;

3) $acl(H)$ is the unique free factor of F of minimal rank containing H.

Proof. By Lemma 21 there exists an algebraic extension of H which is a free factor of F. By Lemma 21 this algebraic extension contains all other algebraic extensions of H, hence it is unique and maximal.

To prove 3) let $H \leq L$ and L be a free factor of F. Then by Lrmma 22 $alc(H) \leq L$. Now it is not hard to see that $alc(H)$ is a free factor of L. Hence $rank(alc(H)) \leq rank(L)$. And if $rank(alc(H)) = rank(L)$ then $alc(H) = L$.

Now we prove two results which are precise analogs of corresponding results from field theory.

Theorem 19 *Let $H \leq_{f.g.} F$. Then:*
1) If $H \leq K$ and $[K : H] < \infty$ then $H \leq K$ is algebraic.
2) If $H \leq K$ and $K \leq L$ are algebraic then $H \leq L$ is algebraic.
3) If $H \leq K_1, H \leq K_2$ are algebraic then $H \leq \langle K_1, K_2 \rangle$ is algebraic.

Proof. 1) is obvious.
2) Suppose $H \leq L$ is not algebraic. Then

$$L = L' * C, \quad H \leq L', \quad C \neq 1.$$

Since $H \leq K$ is algebraic then by Lemma 22 $K \leq L'$. Since $K \leq L$ is algebraic then again by Lemma 22 $L \leq L'$ - contradiction.
3) Let $K = \langle K_1, K_2 \rangle$. If $H \leq K$ is not algebraic then

$$K = K' * C, \quad H \leq K', \quad C \neq 1.$$

By Lemma 22 $K_1, K_2 \leq K'$, i.e., $K \leq K'$ - contradiction.

3.11 Closures of subgroups

In this section we study various closures of f.g. subgroups of F.

Let P be an abstract property of subgroups. Denote by S_P all subgroups of F which satisfy P.

We say that P is intersection-closed if S_P is closed under arbitrary (infinite) intersections. So P is IC if $H_i \in P, i \in I$, then

$$\cap_{i \in I} H_i \in S_P.$$

Proposition 7 *The following properties are IC:*
1) normal subgroup;
2) malnormal subgroup;
3) isolated subgroup.

Racall, that $H \leq G$ is called *malnormal* if

$$H^g \cap H = 1 \Rightarrow g \in H$$

for any $g \in G$.

A subgroup $H \leq G$ is called *isolated* if

$$g^n \in H \Rightarrow g \in H$$

for any $g \in G, n \neq 0$.

Lemma 23 *Let P be IC and G satisfy P. Then for every subgroup $H \leq G$ there exists a unique minimal subgroup of G satisfying P and containing H.*

Proof. Let $S_{P,H} = \{K \mid H \leq K, K \in S_P\}$. Since $G \in S_P$ then $G \in S_{P,H}$, so $S_{P,H} \neq \emptyset$. Put $cl_P(H) = \cap S_{P,H}$. Then $cl_P(H)$ - minimal in $S_{P,H}$.

We say that P is free-factor closed if S_P contains all free factors of F.

Proposition 8 *The following properties are intersection closed and free-factor closed:*

1) normal;

2) malnormal;

3) isolated.

We say that P is well-ordered if for any subgroup $H \leq F$ there exists a unique minimal subgroup K from S_P containing H, i.e., $S_{P,H}$ has a unique minimal element.

Observe that int.-closed P is well-ordered.

Example 11 *Property "being a free factor" is well-ordered.*

Theorem 20 *Let P be a property of subgroups such that*

a) every subgroup H of F has a unique minimal P-closure;

b) every free factor of F satisfies P.

Then for every subgroup $H \leq F$ its P-closure $cl_P(H)$ is an algebraic extension of H.

Proof. Suppose $H \leq cl_P(H)$ is not algebraic then $cl_P(H) = K' * C$ and $H \leq K', C \neq 1$. But then $K' \in S_{P,H}$ hence $cl_P(H) \leq K'$ - contradiction.

Now we describe a method to construct P-closures of H which are related to equations iver H.

Definition 22 *Let $H \leq F$ and $g \in F$. Then:*

1) g is equationally algebraic over H is g is solution of some non-trivial equation $e(x, h_1, \ldots, h_n) = 1$ in one variable x and constants $h_1, \ldots, h_n$.

2) g is equationally transcendental if it does not satisfy any non-trivial equation $e(x, H) = 1$.

Lemma 24 *Let $F = H * C$. Then every element $g \in F - H$ is transcendental over H.*

Proof. Let $e(x,H) = h_1 x^{\epsilon_1} h_2 x^{\epsilon_2} \cdots x^{\epsilon_n} h_{n+1}$, where $\epsilon_i \neq 0,\ h_2, \ldots, h_n \neq 1$.

Let $g \in F - H$ then $g = u^{-1} \circ \bar{g} \circ u$, where $\bar{g}$ is cyclically reduced.

Then $e(g,H) = h_1 u^{-1} \bar{g}^{\epsilon_1} u h_2 u^{-1} \bar{g}^{\epsilon_2} u \cdots h_{n+1} = \varepsilon$. Since $g \notin H$ then there exists a letter y from the basis X whcich is not in H.

????????

Let $\mathcal{E}(x,H) = \{e_i(x,H) | i \in I\}$ be a set of non-trivial equations with one variable x and constants from H.

Definition 23 *A subset $M \subset F$ is $\mathcal{E}(x,H)$-closed if for any equation $e(X,M) \in \mathcal{E}$ its every solution in F belongs to M.*

Lemma 25 *For every set $\mathcal{E}$ any free factor of F is $\mathcal{E}$-closed.*

Proof. Follows from the lemma above.

Let $H \leq F$.

Exercise 3 $\mathcal{E}_{mal,H} = \{x^{-1} h_1 x = h_2 \mid h_1, h_2 \in H\}$

Then H - malnormal $\Leftrightarrow$ H is $\mathcal{E}_{mal,H}$-closed.

Exercise 4 $\mathcal{E}_{iso,H} = \{x^n = h \mid n \neq 0,\ h \in H\}$

Then H - isolated $\Leftrightarrow$ H is $\mathcal{E}_{iso,H}$-closed.

Let $P_{\mathcal{E}}$ be the property "being $\mathcal{E}$-closed".

Lemma 26 *$P_{\mathcal{E}}$ is intersection-closed.*

Proof. Obvious.

Theorem 21 *Let $\mathcal{E}(x,\bar{y})$ be any set of one variable equations with parameters $y_1, \ldots, y_n$. Then for every subgroup $H \leq_{f.g.} F$ there exists $\mathcal{E}$-closure $cl_{\mathcal{E}}(H)$ of H which is an algebraic extension of H and $rank(H) \geq rank(cl_{\mathcal{E}}(H))$.*

Proof. Let $H_0 = H$. If H_0 is $\mathcal{E}$-closed we are done. If not then there exists an equation $e(x,h) \in \mathcal{E}$ and a solution g of $e(x,h) = 1$ such that $g \notin H_0$. Put $H_1 = \langle H_0, g \rangle$.

By induction we construct

$$H_0 < H_1 < H_2 < \cdots$$

Observe that:

1) $rank(H_{i+1}) \leq rank(H_i)$.

Indeed, obviously $rank(H_{i+1}) \leq rank(H_i)+1$. If $rank(H_{i+1}) = rank(H_i)+1 = m$ then every m-generating set of H_{i+1} is a basis. In particular, if Y is

a basis for H_i then $Y \cup \{g_i\}$ is a basis for H_{i+1}. But then g_i is equationally transcendental over H_i - contradiction. Hence the inequality.

2) $H_i \leq H_{i+1}$ is algebraic.

Indeed, if not, then

$$H_{i+1} = H'_{i+1} * C, \quad C \neq 1, \quad H'_{i+1} > H_i.$$

But g_{i+1} is equationally algebraic over H_i, hence over H'_{i+1}. It follows then $g_{i+1} \in H'_{i+1}$ and $\langle H_1, g_{i+1} \rangle \leq H'_{i+1}$ - contradiction.

Now $H_0 < H_1 < H_2 \cdots$ all algebraic over H_0 - hence the chain is finite.

Observe, that $H_i \leq cl_{\mathcal{E}}(H)$. Hence the last in the chain is $cl_{\mathcal{E}}(H)$.

IV. Nielsen's Method.
1. N-reduced sets
1.1. Reduced sets and bases of free.
1.2. Finite automaton and N-reduced sets.

IV. Nielsen's Method

1. Reduced sets.

1.1. Reduced sets and bases.

Let $F = F(A)$ be a free group on $A = \{a_1, ..., a_u\}$.

Definition. A set $S \subseteq F$ is called Nielsen-reduced or N-reduced if the following conditions hold:

N1) $1 \notin S$;

N2) $|uv| \geq |u|$ for all $u, v \in S^{\pm 1}$ with $uv \neq 1$;

N3) $|uvw| > |u| + |w| - |v|$ for all $u, v, w \in S^{\pm 1}$ with $uv \neq 1$ and $vw \neq 1$.

Notice that N2) implies also $|uv| \geq |v|$, since $|uv| = |v^{-1}u^{-1}| \geq |v^{-1}| = |v|$. Therefore N2) is equivalent to

$N2^1$) $|uv| >= max\{|u|, |v|\}$, for all $u, v \in S^{\pm 1}$ with $uv \neq 1$.

Geometrically, the conditions N2) and N3) amount to some restrictions on cancellation allowed among elements from $S^{\pm 1}$. Indeed we claim that:

1) a subset $S \subseteq F$ satisfies N2) if for any $u, v \in S^{\pm 1}, uv \neq 1$,not more then half of u (and v) cancels out in reducing uv;

2) a subset $S \subseteq F$ satisfies N3) is v does not cancel completely in every product uvw such that $u_1 v_1 w \in S^{\pm 1}, uv \neq 1, vw \neq 1$. these can be seen from the following pictures:

p1) $|uv| \geq |u| \Leftrightarrow |p| \leq |q|$

p2) $|uvw| > |u| - |v| + |w| \Leftrightarrow |v| > |p| + |q|$

The geometric characterization of N3) leads to the following definition. But before the definition we introduce a few notations.

Let S be a subset of F. For an element $v \in S^{\pm 1}$ denote:

$\lambda(v)$ = the longest initial segment of v which cancels out in some non-trivial product $uv, u \in S^{\pm 1}$;

$\rho(v)$ = the longest terminal segment of v which cancels out in some non-trivial product $vw, w \in S^{\pm 1}$.

Definition. We say that $S \subseteq F$ is $N3$)-reduced if for any $v \in S^{\pm 1}$ there exists a non-trivial word $\mu(v) \in F$ such that

$$v = \lambda(v) \circ \mu(v) \circ \rho(v)$$

Lemma 1. Let $S \subseteq F$ and $S \neq \{1\}$. Then S satisfies N3) if and only if S is N3)-reduced.

Proof follows from the geometric characterization of the condition N3).

Proposition 1. Let S be an N3)-reduced subset of F. Then S is a free basis of the subgroup $H =< S >$ generated by S.

Proof. By the definition of a free group from section 1.1 S is a basis of H if and only if S generates H (which is already given by the conditions of the

proposition) and every non-trivial reduced word in $S^{\pm 1}$ defines a non-trivial element in H. now if

$$w = u_1...u_n, u_i \in S^{\pm 1}$$

is a non-trivial reduced word in $S^{\pm 1}$, then

$$|w| \geq |\mu(u_1)| + . + |\mu(u_n)| > 0$$

(here — — is the length on F with respect to the initial set of generators of F), hence $w \neq 1$.

Theorem 1. Let $H =< h_1, ., h_n >$ be finitely generated subgroup of F. Then there exists an N3)-reduced set S of generators of H and a such set S can be found effectively.

Proof. Let $\Gamma(H)$ be a finite automaton accepting H, and let T be a maximal subtree of $\Gamma(H)$. We have proved in section 3.1 that the set of generators S_T of H is n3)-reduced, and this set S_T can be found effectively from the initial set of generators $h_1, ..., h_n$.

Now we generalize the condition N2) to obtain different type of basis of a free group.

Definition. A subset $S \subseteq F, 1 \not\in S \neq$ is called N2)-reduced (or weakly reduced) if for any reduced word $u_1...u_n$ in the alphabet $S^{\pm 1}$ the following holds: (1)

$$|u_1...u_n| \geq |u_2...u_n|$$

Notice, that N2)-reduced sets satisfy the condition N2), but the reverse is false.

Example 1. The set

$$s_1 = \{ab^{-1}, bc, c^{-1}a^{-1}\} \subseteq F(a, b, c)$$

satisfies the condition N2), but it is not N2)-reduced.

Notice also, that there are N3)-reduced sets which are not N2)-reduced, and N2)-reduced sets which are not N3)-reduced.

Example 2. The set

$$S_2 = \{a^2, b^2, ab\}$$

is N3)-reduced, but it does not satisfy N2).

Example 3. The set

$$S = \{ab^{-1}, bc^{-1}, cd^{-1}\} \subset F(a, b, c, d)$$

is N2)-reduced, but does not satisfy the condition N3).

We saw in Example 1 that it does not suffice to check the condition N2) to verify that a given set S is N2)-reduced. Thus in Example 1 the set S_1 satisfies the condition (1):

$$|u_1...u_n| \geq |u_2...u_n|, u_i \in S^{\pm 1}$$

for all products $u_1...u_n$ of length $n \leq 2$. but for $n = 3$, we have $(ab^{-1})(bc)(c^{-1}a^{-1})$, so (1) does not hold.

The following criterion of N2)-reducification shows that one has to check the condition (1) only for products of length $\leq 2^{|S|}$.

Proposition 2. A subset $S \subseteq F$ is N2)-reduced if and only if the inequality (1):

$$|u_1...u_n| \geq |u_2...u_n|$$

holds for every reduced product $u_1...u_n$ in the alphabet $S^{\pm 1}$ with $n \leq 2^{|S|}$.

Proof is in Combinatorial Group Theory: a topological approach, P. Cohen, L. Math. See student Texts, v.14].

For N2)-reduced sets we have a result similar to Proposition 1.

Proposition 3. Let S be an N2)-reduced subset of F. Then S is a free basis of the subgroup $H =< S >$ generated by S.

$\triangleright$. By the definition of the basis from section 1.1 it suffices to prove that every non-trivial reduced word in the alphabet $S^{\pm 1}$ defines a non-trivial element of F. If $u_1...u_n$ is a non-trivial reduced word in $S^{\pm 1}$, then

$$|u_1...u_n| \geq |u_2...u_n| \geq ... \geq |u_n| > 0$$

Hence, $u_1...u_n \neq 1$, as desired.

In [on length functions and Nielsen methods in free groups. J. Lou. Math Soc.,1976,14,p.188-192.] A. Hoare showed how to construct N2)-reduced sets in free groups.

Now we turn to the main objects of this section-N-reduced sets. There are two principal techniques for finding an N-reduced set of generators for a subgroup H of F which is given by a finite set of generators. The classical one rests on elementary Nielsen transformations, which allow one to transform a given set of generators of H into an N-reduced set. It is more difficult technically, but it has very important applications to the automorphisms of free groups. The second one is based on the finite automaton $\Gamma(H)$ accepting H and methods of graph theory, it provides fast and simple algorithms.

1.2 Geodesic maximal trees and N-reduced sets

Let H be a finitely generated subgroup of F. Denote by $\Gamma(H)$ a finite automaton which accepts H.

We have seen already that for each maximal subtree T of $\Gamma(H)$ one can construct the corresponding set of generators S_T of the subgroup H. Moreover, this set S_T is always N3)-reduced. Our goal now is to describe some particular maximal subtrees T of $\Gamma(H)$ for which the set of generators S_T is N-reduced. To this end we need the following definition.

Definition. Let Γ be a connected graph with a distinguished vertex v_0. A maximal subtree T of Γ is called geodesic at v_0 from Γ if the length of the path in T from v_0 to v is the shortest possible among all the paths from v_0 to v in Γ.

If for a vertex $v \in \Gamma$ we denote by $l(v)$ the length of the path in T from v_0 to v, then a geodesic maximal subtree T satisfies the following condition:

$$(*)|l(\sigma(e)) - l(\tau(e)| \leq 1$$

for every edge e in Γ (here $\sigma(e)$ is the initial vertex of e, and $\tau(e)$ is the terminal vertex of e)

A maximal subtree T satisfying $(*)$ is called balanced. Notice that not all balanced maximal subtrees are geodesic. For example: Clearly, the tree $T = \Gamma - \{e_1 e_2\}$ is balanced, but not geodesic.

Proposition 4. Let H be a finitely generated subgroup of F, and $\Gamma(H)$ is automaton which accepts H. If T is a balanced maximal subtree of $\Gamma(H)$, then the corresponding set of generators S_T is N-reduced.

Proof. Recall that the set S_T is defined. For a vertex $V \in \Gamma(H)$ we denote by $r(v)$ the path in T from 1 to v. We will the notation $r(v)$ also for the label of this path, which is an element from F. Then

$$S_T = \{r(\sigma(e))e\ r^{-1}(\tau(e))|e \in \Gamma - T\}$$

we denote by $S(e)$ the path $r(\sigma(e))e\ r^{-1}(\tau(e))$ as well as its label, which is an element from H. We will call e the mid-edge of $s(e)$. It was shown already that S_T satisfies the condition N3). notice, that the mid-edge e does not cancel in any product of the type $S(e)^{\pm 1}S(e^1)^{\pm 1}$ or $S(e^1)^{\pm 1}S(e)^{\pm 1}$, for $e^1 \in \Gamma - T\ \ e^1 \neq e$. Now since

$$||r(\sigma(e))| - |r(\tau(e))|| \leq 1$$

we see that S_T also satisfies N2).

Clearly, $S(e) \neq 1$ for any $e \in \Gamma - T$. It follows that S_T is N- reduced set of generators of H.

There is a well known procedure in graph theory (or computer science) which for any finite connected graph Γ finds a maximal geodesic subtree T of Γ. No we recall briefly this procedure.

Let Γ be a connected graph with a distinguished vertex v_0. Denote by $T(0)$ the graph with one vertex v_0 and no edges. We will expand the graph $T(0)$ into a geodesic at v_0 maximal subtree of $\Gamma(H)$ by finitely many elementary steps , which we describe below. Let T^1 be a subtree of Γ and v be a vertex in T^1. choose an edge $e \in \Gamma$ with $r(e) = v, \tau(e) \not\in T^1$, and replace T^1 by $T^1 v\{e\}$. Repeat this until possible. In finitely many steps ($\leq$ the degree of v) we will get a tree T'' which contains all the vertices a distance 1 from v. We call this process a 1-step expansion around v.

Now starting with $T(0)$ make 1-step expansion of $T(0)$ around v_0, and denote the resulting tree by $T(1)$. Clearly, $T(1)$ is a maximal subtree of the ball $(v_0, 1)$ which is geodesic at v_0. Enumerate al the vertices in $T(1) - T(0)$ as $v_{11}, ., v_{1k_1}$. Make 1-step expansion consequently at each vertex $v_{11}, ., v_{1k_1}$ and denote the resulting tree by $T(2)$. Again, $T(2)$ is a maximal subtree of the label $(v_0, 2)$ which is geodesic at v_0. Enumerate all the vertices of $T(2) - T(1)$ as $v_{21}, ., v_{2k_2}$, and make all the consequent 1-step expansions at these vertices. Repeating this process at finitely many steps we will construct a maximal subtree of Γ which is geodesic at v_0.

As a corollary of the foregoing discussion we have the following.

Theorem. Every non-trivial finitely generated subgroup of F has an N-reduced set of generators and such set can be effectively found.

Remark. Notice that for a given subgroup $H = \{h_1, ., h_m\} \leq F$ N-reduced set of generators can be found in $O(d \circ n)$ steps, where $d = |h_1| + ... + |h_m|\} \leq F$ an N-reduced set of generators can be found in $O(d \cdot n)$ steps, where $d = |h_1| + ... + |h_m|$ is the total length of the given set of generators of H and n is the rank of F.

Indeed, we need $O(d)$ steps to construct the graph $\Gamma(H)$. For a given vertex $v \in \Gamma(H)$ we need $\leq 2n$ steps to make 1-step expansion around v. since we have $\leq d + 1$ vertices in $\Gamma(H)$ the process described above will stop in $\leq (d + 1)2n$ steps. Altogether, we need $\leq O(dn)$ steps to construct the N-reduced set S_T of generators of H.

We did not define precisely what we meant by a "step" in the argument above. It is understood, that any formalization of this will amount to a constant number of elementary computations which have to be performed to make one "step".

V. Nielsen Transformations

In this section we will show how an arbitrary finite set of generators of a finitely generated subgroup H of a free group F can be transformed into an N-reduced set of generators of H by a finite sequence of so-called "elementary" transformations.

Let $H \leq F, H =< h_1, ..., h_n >$. As usual we draw a rose corresponding to the set of generators $U = \{h_1, ., h_n\}$: An elementary transformation of U consists of replacing some coordinate h_i by a new word $h\prime_i$ according to the following rule.

Choose any other coordinate $h_j, i \neq J$, and draw these two loops separately:

Then glue together the same edges starting from 1 on the upper parts of h_i and h_j until this is possible: and delete the common part: Replace h_i by $h\prime_i = q^{-1}r$.

Or glue together lower parts of h_i and h_j until this is possible: and h_i is obtained from the result of glueing by deleting the common part. Or else, twist h_i (or h_j or both) around 1 and then repeat the process above: Equivalently, we can choose one of the two orientations in each loop: and glue together the longest equal paths on this loop starting from 1 and going according to the chosen orientations.

Now suppose, that starting from U we will apply only those elementary transformations that strictly decrease the total length of the tuple: $U \to U_1 \to ... \to U_m$. Geometrically it means that if two loops h_i and h_j and we mark the midpoints on them: then we apply elementary transformations to h_i, h_j if and only if after gluing at least one of the midpoints occurred in the common path.

Plan:

1. $U \to U*$-the rose.
2. Elementary transform. gluing.
3. L-minimal $\Rightarrow N1, N2$.
4. Loops into wedges:
5. If not N3, then

$$u = ap^{-1}, v = pq^{-1}, w = qb^{-1}$$

in this event we perform one of the following transforms a), b)

Put in the tree Γ: we perform those which move edge to the left $p < q \to U \to U_1 \to ...$ in finitely many steps we cannot move to the left any more (no room on the left), so N-reduced. a)

b)

We can assume that our wedges sit inside the tree Γ. So the rule is if $p < q$, then replace $w = qb^{-1}$ by $pb^{-1} = w$ but if $q < p$ then replace $u = ap^{-1}$ by $aq^{-1} = uv$. The reason is that a such transformation leaves one branch of the wedge fixed and moves another branch to the left.

Clearly, if the total length $|U|$ is fixed there is only a finite number of such elementary transformations that move at least one branch to the left and fixes the other.

So in finitely many moves we will get to the most far left possible transform of U:

$$U \to ... \to U_k = V$$

Now V is N-reduced. If U does not satisfy N2 then it might happen only if there are three elements $u, v, w \in U^{\pm 1}$ such that $u = ap^{-1}, v = pq^{-1}, w = qb^{-1}, |p| = |q|$. In this event we can perform one of the following two elementary transformations:

a) replace $u = ap^{-1}$ by $uv = aq^{-1}$;

b) replace $w = qb^{-1}$ by $vw = pb^{-1}$.

We are going to specify the rule which one of these two perform. Since the gluing of two loops from U cannot go further any of their midpoints, then we will cut all loops at their midpoints and consider the resulting wedges:

Now our transformations look like:

If α and β are two geodesics in Γ going from root 1, then we say that α is going to the left from β (and write $\alpha < \beta$) if $\alpha = r \circ \alpha\prime, \beta = r \circ \beta\prime$ and the first edge in $\alpha\prime$ goes to the left from the first edge in $\beta\prime$:

Let

$$u$$

be a loop at 1. We can write $u = pq^{-1}$ where p and q are two geodesics starting at 1 with the property that $|p| = |q|$ (notice that the midpoint could be in the middle of some edge, so the length of $|p|$ is not integral in this case).

Put

$$L(u) = min\{p, q\}, R(u) = max\{p, q\}$$

So $L(w)$ is the most far left half of $u, R(u)$ is the most far right half of u. Similarly, if we have:

b) q is to the left of p, then

So the transformation $u \to uv$ moves this loop to the left.

Clearly, there are only finitely many elementary transformations which move at least one of the tuples to the left and keep the same total length of the tuple. So in finitely many moves

$$U_m \to U_{m+1} \to ... \to U_{m+k} = V$$

we will get a tuple V that cannot be moved to the left anymore. Then V is N-reduced.

The tuple U_m satisfying $N0$ and $N1$ may not satisfy N2 only in the following case: there are elements $u, v, w \in U_m^{\pm 1}$ such that $u = ap^{-1}, v = pq^{-1}, w = qb^{-1}$ and $|p| = |q|$. In this event we can apply an elementary transformation and replace

$$u = ap^{-1}$$

by

$$uv = aq^{-1}$$

or replace

$$w = qb^{-1}$$

by

$$vw = pb^{-1}$$

Let us cut our loops at the midpoint and put them into the graph into the graph Γ.

There are two possible cases: a) p is to the left of q in Γ, then so transformation $w \to vw$ moves our loop more to the left.

We say that a loop u is to the left of the loop v (and write $u < v$) if $L(u) < L(v)$ or $L(u) = L(v)$ and $R(u) < R(v)$.

Geometrically:

Clearly, for any loop u there are only finitely many loops v with equal length $|u| = |v|$ and to the left of u.

Now we can continue the chain of transformations

$$U \to ... \to U_m \to U_{m+1} \to .U_k = V$$

allowing transformations that keep the same total length but decreasing at least one of the loops with respect to the "left" ordering $<$ but replacing one of the loops by a loop to the left of it.

Claim: V is N-reduced. Suppose v does not satisfy N2. Since V satisfies N0 and N1 the only case when the condition N2 fails is the following: there are some loops $u, v, w \in V$ such that $u = qap^{-1}, v = pq^{-1}, w = qb$ is to the left of v if

$$L(u) < L(v)$$

or $L(u) = L(v)$ and $R(u) < R(v)$.

Shortleft order.

Let $A = \{a_1, ..., a_n\}$. List all elements in $A \cup A^{-1}$ in the following order:

$$a_1 < a_1^{-1} < a_2 < a_2^{-1} < ... < a_n < a_n^{-1}(*)$$

. Now we can draw the following plane tree with root 1:

i.e., from each vertex there are exactly $2n$ outgoing edges which are located from the left to the right with respect to the order $(*)$. $n = 2$

If α and β are two paths from 1 then we say that $\alpha < \beta \Leftrightarrow |\alpha| < |\beta|$ or $|\alpha| = |\beta|$ and α is to the left of β i.e., $\alpha = \gamma \circ \alpha\prime, \beta = \gamma\beta\prime$ and $\alpha\prime$ goes to the left from $\beta\prime$.

Let $a_1 < {a_1}^{-1} < a_2 < a_2^{-1}. < a_n < {a_n}^{-1}(*)$. We will draw the Cayley graph γ_F of F in the following way: i.e., from each vertex v all edges go from the left to the right with respect to the order $(*)$. Let u be a loop at 1 in our rose. Then we can write $u = pq^{-1}$ where $|p| = |q|$ notice that we allow for the midpoint of u to be in the middle of the edge, so $|p| = |q| = 2^{\frac{|U|}{2}}$. Now if we draw the geodesic p and q in the graph $\Gamma(F)$ then it is clear what means that one of them goes to the left from the other, we will call this one the left half of u and denote by $L(u)$, the other one is the right part of u and we denote it by $R(u)$. Now if u and v are two loops then we can cut both at midpoints and all the a halves in Γ_F. We say that u and $|p| = |q|$. Now there are two possibilities, if $L(v) = p$, then the loop $vw = pb$ is to the left of $w = qb$, contradiction with minimality of V. If $L(v) = q$, then uv is to the left of u- again a contradiction. Hence, this case cannot occur and V is N-reduced.

Clearly, for a given tuple U we can apply not more then $|U|$ elementary transformations, which decrease the total length. Suppose

$$U \to U_1 \to ... \to U_m$$

and there is no any elementary transformation that decreases the total length of U_m. Then the set of all non-trivial fundamental loops in the rose U_m satisfies the conditions N1 and N2.

Let us order the elements of $A \cup A^{-1}$ as follows:

$$a_1 < {a_1}^{-1} < a_2 < {a_2}^{-1} < . < a_n < {a_n}^{-1}$$

We can draw the Cayley graph of the group $F(a_1, ..., a_n)$ in the following way (left oriented) such that from each vertex the smallest edge goes to the left, then the next one and so on.

Automorphisms of F.

Let $F = F(x_1, ., x_n$ be a free group with a basis $x_1, ., x_n$.

Theorem. The following holds:

1) F cannot be generated by fewer than n elements;

2) if F is generated by a set $U = (U_1, ., U_n)$ of n elements, then U is a basis of F.

$\triangleright$. Suppose F is generated by $U = (u_1, .., u_m)$. Then there exists a chain of elementary Nielsen transformations $t_1, ..., t_k$ such that

$$V = Ut_1...t_k$$

is N-reduced, and therefore a basis for $F = \langle U \rangle$. Hence $|V| = n$. But $|U| \geq |V|$, and it follows that $m \geq n$.

Now suppose that $m = n$. Then in the notations above $V = Ut_1...t_k$ is a basis of F. Since $U = Vt_k{}^{-1}...t_1^{-1}$ and every regular elementary Nielsen transformation transforms a basis into another basis of F, then U is a basis of F (notice, that all the transformations $t_1, ..., t_k$ are regular).

Definition. A group G is called Hopfian if it is not isomorphic to any proper factor group of itself.

Observe, that G is Hopfian if every epimorphism of G is an automorphism.

Theorem. Every finitely generated free group is Hopfian.

$\triangleright$. Let $\varphi : F \to F$ be an epimorphism of F. Then $U = (x_1{}^{\varphi}, ..., x_n{}^{\varphi})$ generates F. Hence by the theorem above U is a basis of F. It implies that φ is an automorphism. Any regular elementary Nielsen transformation on $X = \{x_1, ..., x_u\}$ extends to an automorphism of F. Such automorphisms of F are called elementary Nielsen X-automorphisms of F.

There are $n + n(n-1) = n^2$ elementary Nielsen X-automorphisms.

Denote by $U = (u_1, ., u_n)$ a tuple in F^n. For an automorphism $\varphi \in AutF$ denote by $U^{\varphi} = (u_1{}^{\varphi}, ..., u_n{}^{\varphi})$. If t is a regular Nielsen transformation then for any $\varphi \in AutF$ we have

$$(Ut)^{\varphi} = U^{\varphi}t$$

This simple observation will be of use in the following result.

Theorem. Let $F = F(x_1, ..., x_u)$. Then the elementary Nielsen automorphisms generate the group $AutF$

$\triangleright$. Let $\varphi \in AutF$ and $U = X^{\varphi}$. Then U generates F and therefore there exists a chain of elementary regular Nielsen transformations $t_1, ..., t_n$ such that $Ut_1, ..., t_n$ is N-reduced, then every element of F, in particular all x_i's, can be expressed as a product of components of $Ut_1, ..., t_n$. But then every x_i is a component of $Ut_1, ..., t_n$. Hence the claim.

Now

$$U = Xt_k{}^{-1}.t_1{}^{-1}$$

Denote by φ_i the elementary Nielsen X-automorphism defined by t_i.

Then

$$U = Xt_n{}^{-1}...t_1^{-1} = (X^{\varphi_k^{-1}}t^{-1}{}_{k-1}...t_1^{-1} =$$
$$= (Xt^{-1}{}_{n-1}...t_1^{-1})^{\varphi_u^{-1}}$$

Hence

$$U^{\varphi_u} = Xt^{-1}{}_{k-1}...t_1$$

Continually this process we will get

$$U^{\varphi_k\cdots\varphi_1} = X$$

So $X^{\varphi\varphi_k\cdots\varphi_1} = X$ and $\varphi = \varphi_1{}^{-1}...\varphi_k{}^{-1}$.

Generators of $AutF_n$.

As we have mentioned already the automorphism group $AutF_n$ of a free group F_n of rank n is generated by elementary Nielsen automorphisms.

Let $F_n = F_n(x_1, ..., x_n)$. By definition elementary Nielsen X- automorphisms are the following ones:

1) $x_i \to x_i x_j, i \neq j$;

2) $x_i \to x_i{}^{-1}$ for all $1 \leq i \neq j \leq n$.

In fact, the group $AutF_n$ is generated by the following four automorphisms:

Indeed, to obtain all the automorphisms of the type 1) and 2) we need only automorphisms

$$\sigma : x_1 \to x_1{}^{-1}, U : x_1 \to x_1x_2$$

and all the permutations of the set X.

It is a well known result that the group of permutations $S(X)$ of the set X can be generated by only two autos:

$$P : x_1 \to x_2, x_2 \to x_1$$

$$Q : x_1 \to x_2, x_2 \to x_3, ..., x_n \to x_1$$

Corollary 1. Let $F = F(x_1, x_2)$ be a free group of rank 2. Then for any $\varphi \in AutF_n$ there exists $f \in F$ s.t.

$$[x_1{}^{\varphi}, x_2{}^{\varphi}] = f^{-1}[x_1, x_2]^{\pm 1}f$$

$\triangleright$. It is suffices to check only for elementary automorphisms. Indeed, let $\varphi = \varphi_1...\varphi_k$ be a product of elementary automorphisms. Then $y_1 = x_1{}^{\varphi_1\cdots\varphi_{k-1}}, y_2 = x_2{}^{\varphi_1\cdots\varphi_{k-1}}$ form a basis of F_2, hence

$$[y_1{}^{\varphi_k}, y_2^{\varphi_k}] = f_k{}^{-1}[y_1, y_2]f_k$$

and we apply induction to $y_1 = x_1{}^{\varphi_1 \cdots \varphi_{k-1}}, y_2 = x_2{}^{\varphi_1 \cdots \varphi_{k-1}}$.

Theorem (Nielsen). A pair (u, v) form a basis of $F_2(x_1, x_2)$ if and only if

$$[u, v] = f^{-1}[x_1, x_2]^{\pm 1} f$$

for some $f \in F_2$.

$\triangleright$. Indeed, if (u, v) is a basis of F_2, then $x_1 \to u, x_2 \to v$ is an automorphism of F_2 and the conclusion follows from the corollary above. Suppose, now

$$[u, v] = f^{-1}[x_1, x_2] f$$

then

$$[u^{f^{-1}}, v^{f^{-1}}] = [x_1, x_2]$$

and $u_1 = u^{f^{-1}}, v_1 = v^{f^{-1}}$ is a solution of the equation $[u_1, v_1] = [x_1, x_2]$. In this case (u_1, v_1) can be obtained from (x_1, x_2) by elementary transformations (see section 3.3), hence $x_1 \to u_1, x_2 \to v_2$ is an automorphism of F_2, as well as $x_1 \to u, x_2 \to v$.

Fixed points of endomorphisms.

Let F be a free group on $x_1, ..., x_n$, and $\varphi : F \to F$ be an endomorphism. Then the set of all fixed points of φ. $Fix\varphi = \{w \in F | w^\varphi = w\}$ is a subgroup of F.

Theorem. For any endomorphism $\phi : F \to F$ the subgroup $Fix\phi$ is finitely generated.

$\triangleright$. We construct an automaton Γ (not finite) which accepts $Fix\varphi$.

First we construct a non-oriented graph Γ that accepts $Fix\varphi$. Let elements from F be the vertices of Γ. Two vertices v and w are connected by and edge if there exists an element $x \in X \cup X^{-1}$ such that $x^{-1} v x^\varphi = w$. We label this edge by x. Obviously, $(x^{-1})^{-1} w (x^{-1})^\varphi = v$, we consider this edge as the inverse for the previous one.

The graph Γ accepts $Fix\varphi$. Indeed, if reading w along Γ we finished at 1, then

$$w^{-1} w^\varphi = 1, i.e., w^\varphi = w$$

Claim. The graph Γ has only finitely many fundamental loops, i.e., for any maximal subtree T of Γ the set $\Gamma - T$ is finite: (and hence $Fix\varphi$ is s finitely generated).

To prove this define the following orientation of edges in Γ. Suppose e is an arbitrary edge between v and w with label x. If v and w satisfy the condition:

(*) v does not cancel out completely in vx^φ w does not cancel out completely in $w(x^{-1})^\varphi$ then we orient the edge e as going from v into w if and only if v begins with x.

Notice that if v does not begin with x then w begins with x^{-1} and hence the edge e has orientation from w into v.

It shows, that in the event (*) the vertex v has only one outgoing edge.

Clearly, only initial segments of x^{φ} and $(x^{-1})^{\varphi}$ do not satisfy the condition (*). Hence outside the ball $B(1,r)$ with radius $r = max\{|x^{\varphi}||x \in XUX^{-1}\}$ each vertex has exactly one outgoing edge.

The graph $B(1,r)$ is finite, therefore the graph $\Gamma - B(1,r)$ has only finitely many connected components say $\Gamma_1, ..., \Gamma_m$.

Lemma. Let Δ be a connected oriented graph with exactly one outgoing edge at each vertex. Then either Δ is a tree or Δ has exactly one fundamental loop.

$\triangleright$. Proof. Let $p = e_1.e_n$ be a loop in Δ as a non-oriented graph. Then p is a loop in Δ with the given orientation. Now if we delete an edge from p then $\Delta - e$ will have a vertex with no loops at all. Clearly e lies at most in one loop in Δ. Hence Δ has at most one fundamental loop.

Now to finish the proof it suffices to notice that if we choose a maximal subtree T_i in Γ_i. Then the union UT_i can be enlarged to a maximal subtree T of Γ and $T - \Gamma$ is finite. To see this it suffices to notice, that every subforest of a connected graph can be extended to a maximal forest which contains all the vertices. Then we add some bridges between these connected components of the maximal forest, and turn it into a tree.

Schrier's method.

Let $F = F(a_1, ..., a_n)$ be a free group and H a subgroup of F. Recall, that a set of representatives $S = \{s_o = 1, s_1, ...\}$ of right cosets of H in F is called a Schrier set, if for every $s \in S$ each initial segment of s also belongs to S.

We have showed in the previous section that if H is a subgroup of finite index and T is a spanning subtree of $\Gamma_0(H)$ then the set S_T of all paths in T starting at 1 is a Schrier's transversal for H in F. Now we will prove that the reverse is also true, namely, for any subgroup H with Schrier's transversal S there exists a graph Γ accepting H and a spanning tree T in Γ such that $S = S_T$.

Indeed, consider the tree T with the given set S as vertices of T and for vertices $u, v \in S$ we draw an edge with label a_i from u into v if and only if $ua_i = v$. Clearly, T is a tree.

For an element $w \in F$ by $\bar{w}$ we denote the representative of w in S. now for each vertex $v \in T$ and each $a_i \in A$ we connect v with the vertex $v\bar{a}_i$ by the directed edge from v into $v\bar{a}_i$ and label this edge by a_i. The resulting regular graph is denoted by Γ_S. By the construction T is a spanning tree for Γ_S. For every $e \in \Gamma_S - T$ the fundamental path p_e defined by e is equal to $s_1 s a_i {s_2}^{-1}$. Observe, that $s_2 = s_1\bar{a}_1$, so p_e defines an element from H. Since

every closed path at 1 is a product of fundamental loops, it follows that Γ_S accepts H. Moreover, $S = S_T$ as desired.

Definition. Schrier's transversal S of H in F is called minimal if each $s \in S$ is of minimal length is its coset Hs.

Proposition. Let H be a subgroup of F. Suppose a graph Γ accepts H and T is a geodesic spanning subtree of Γ. Then the Schrier transversal S_T is minimal. The reverse is also true, if S is a minimal Schrier transversal then S is a geodesic spanning tree in Γ_S.

$\triangleright$. Obvious.

Corollary. If S is a minimal Schrier transversal for a subgroup H of F. Then the set R_S of all fundamental loop sin Γ_S with respect to the spanning subtree S is an N-reduced set generators of H.

That was the original Schrier's idea: find a minimal Schrier transversal S for H. Then the set R_S of all non-trivial elements of the form $sx(\bar{sx})^{-1}, s \in S, x \in A$, (all fundamental loops in Γ_S!) is a basis for H.

What is left to show is how to find a minimal Schrier transversal S for H. To this end, let $\leq$ be a Shortlex order in F. In each coset Hw of H choose the minimal element with respect to $\leq$, say $\bar{w}$. Then the set $S = \{\bar{w} | w \in F\}$ is a Schrier minimal transversal of H.

Nielsen-Schrier Theorem (general case).

Theorem. Ever subgroup of a free group F is free.

Before we start to prove the theorem, let us introduce a well-ordering $<<$ on F.

Let $<$ be an arbitrary ordering on $A \cup A^{-1}$ (recall that A is a basis of F), for example,

$$a_1 < a_2 < ... < a_n < {a_1}^{-1} < {a_2}^{-1} < ...$$

or if A is infinite:

$$a_1 < {a_1}^{-1}, a_2 < {a_2}^{-1} < ...$$

Thus $(A \cup A^{-1}, \leq)$ is a well ordered set.

Extend $\leq$ auto $F(A)$ as follows: if $u = a_{i_1}^{\varepsilon_1} ... {a_{i_n}}^{\varepsilon_n}, v = a_{j_1}^{\delta_1} ... a_{j_m}^{\delta_m}$ then $u < v$ is and only if

1) $|u| < |v|$, or

2) $|u| = |v|$, and there is some k such that

$$a_{i_e} = a_{j_e}, \varepsilon_e = \delta_e$$

for all $e < k$, and $a_{ie+1}^{\varepsilon_{e+1}} < a_{j_{e+1}}^{\delta_{e+1}}$.

Clearly, $<$ is a well-ordering on $F(A)$. Such orderings are called shortlex orderings.

Observe, that

$$u < v \Rightarrow w \circ u < w \circ v$$

$$u < v\& \ |w_1| \leq |w_2| \Rightarrow u \circ w_1 < v \circ w_2$$

The left $e(v)$ a word $v \in F$ is defined as the initial segment of v of length $\frac{[|v|+1}{2]}$ where [] is the integral part function, i.e., if $|v| = 2n$, then $e(v)$ is the initial segment of v of length n; if $|v| = 2n+1$, then $e(v)$ is the initial segment of length $n+1$.

Finally define an order $<$ on pairs $\{u, u^{-1}\}$ as follows $u^{\pm 1} < v^{\pm 1}$ if either

1) $|u| < |v|$; or

2) $|u| = |v|$ and either $lmin\{u, u^{-1}\} < min\{v, v^{-1}\}$ or this mins are equal and

$$max\{u, u^{-1}\} < max\{v, v^{-1}\}$$

Let H be a subgroup of F. For $\alpha \in H$ define

$$H_\alpha = \langle h \in H | h^{\pm 1} < \alpha^{\pm 1} \rangle$$

and put $A = \{a \in H | a \notin H_\alpha\}$. Let S be a subset of A which is obtained from A by throwing away exactly one of each inverse pair of elements.

Claim. S is N-reduced set of generators of H.

Proof. 1) S generates H. Suppose $\langle S \rangle \neq H$, and let $\alpha^{\pm 1}$ be the minimum pair in $H - \langle S \rangle$. Then $\alpha \notin \in A \Rightarrow \alpha \in H_\alpha$, but by minimality of α we have $H_\alpha \leq \langle S \rangle$-contradiction.

2) S is N-reduced. Indeed,

N1: $1 \notin S$ for $1 \in H_1 = \langle \phi \rangle = \{1\}$

N2: Let $u, v \in S \cup S^{-1}$, then $|uv| \geq |u|$, indeed, suppose $(uv)^{\pm 1} < v^{\pm 1}$. Then $v \in < uv, u > \Rightarrow v \in H_v$-contradiction with choice of A. Hence $(uv)^{\pm 1} > u^{\pm 1}, v^{\pm 1}$. But then

$$|uv| \geq |u|, |v|$$

N3: As in the Nielsen theorem.

Whitehead Method.

Let $F = F(x_1, ..., x_n)$ be a free group with basis $X = \{x_1, ..., x_n\}$. Let $a \in X^{\pm 1}$. We call an automorphism $\varphi \in AutF$ of the type (w) if φ maps each element $x_i \in X$ into one of the following:

$$x_i, x_i a, a^{-1}x_i, a^{-1}x_i a$$

Notice, that if φ is an automorphism of the type (w) then φ maps ${x_i}^{-1}$ also into one of the following:

$${x_i}^{-1}, {x_i}^{-1}a, a^{-1}x_i, a^{-1}x_i a$$

Also, is φ is of the type (w), then φ^{-1} is also of the type (w).

Definition. A Whitehead X-automorphism of the group F is an automorphism of one of the following types:

1) $x_i \Rightarrow x_j, x_j \Rightarrow x_i, x_k \to x_k, k \neq i, j$;

2) $x_i \Rightarrow {x_i}^{-1}, x_k \Rightarrow x_k, k \neq i$;

3) an automorphism of the type (w).

Notice, that Nielsen X-automorphisms are also Whitehead X- automorphisms. Denote by Ω the set of all Whitehead automorphisms of F. Then Ω generators $AutF$. We say that two tuples

$$U = (u_1, ..., u_m), V = (v_1, ..., v_m)$$

are $AutF$-equivalent (and write $U \sim_{Aut} V$) if there exists an automorphism $\varphi \in AutF$ s.t.

$$U^\varphi = V (i.e., u_i^\varphi = v_i, i = 1, ..., m)$$

Clearly, $\sim_{Aut}$ is an equivalence relation on F^m. Denote by $[U]$ the equivalence class containing U.

A tuple U is called $_{Aut}$-minimal, is U has a minimal total length

$$|U| = |u_1| + ... + |u_m|$$

in its class $[U]$

The following result shows how one can effectively find a minimal element in the class $[U]$.

Theorem (Whitehead). Let $U, V \in F^m$ and $U \sim_{Aut} V$. Then:

1) if $|U| > |V|$, then there exists an automorphisms $\varphi \in Aut$ of the type (w) s.t.

$$|U| > |U^\varphi|$$

2) if $|U| = |V|$, then there exists a finite sequence of Whitehead automorphisms $\varphi_1, ..., \varphi_k$ such that $U^{\varphi_1 \cdots \varphi_k} = V$ and

$$|U| = |U^{\varphi_1}| = |U^{\varphi_1 \varphi_2} = ... = |U^{\varphi_1 \ \cdots \ \varphi_k}| = |V|$$

Proof is difficult and we will omit it for a while. Now we can describe the Whitehead algorithm which finds a minimal element in $[U]$ and also allows one to decide effectively whether two tuples U and V are $AutF$-equivalent or not.

Whitehead Algorithm:

Apply automorphisms of the type (w) to the tuple U and see whether it is possible to reduce the total length. If this is not possible, then U is $\sim_{Aut}$-minimal. If there exists an automorphism φ_1 of the type (w) which decreases the total length of U, then consider $U_1 = U^{\varphi_1}$ and repeat the process. In finitely man steps the process will stop:

$$|U| > |U^{\varphi_1}| > ... > |U^{\varphi_1 \cdots \varphi_n}|$$

and $U^{\varphi_1 \cdots \varphi_n}$ is -minimal.

Now let U and V be tuples from F^m. Then one can effectively decide whether $U \sim V$ or not. To this end, let U^* and V^* is minimal in the classes $[U]$ and $[V]$. Then if $|U^*| \neq |V^*|$, then $U \not\sim V$. If $|U^*| = |V|$, then it suffices to do the following. Starting at U apply all Whitehead automorphisms which do not increase the length:In finitely many steps we will construct the graph Γ in which vertices are all the -minimal elements in $[U]$, and in which two vertices are connected by an edge if there exists a Whitehead automorphism transforming one vertex into the other. If V occurs among the vertices of this graph, then $V \sim U$, otherwise $V \not\sim U$.

Unfortunately, the Whitehead algorithm is not that effective. It takes about

$$c|U|^2$$

steps to find a minimal tuple in the class $[U]$, where

$$c \approx n8^n, \ n = rankF$$

Moreover, the number of steps required to construct the graph above can be estimated as

$$\leq C_1(2n)^{|U|}$$

where

$$C_1 = n!2^n$$

It is still unknown whether it is possible to verify that $U \sim_{Aut} V$ in a number of steps which is a polynomial on $|U|$. In this direction the following problems are of interest.

Problem 1. Is the number of minimal elements in $[U]$ is bounded from above by a polynomial on $|U|$? (i.e., is the number of vertices in the graph Γ above is bounded by a polynomial on $|U|$?)

Problem 2. Is the diameter of the graph constructed above is not greater then a polynomial on $|U|$?

Definition. An element $g \in F$ is called primitive if g can be included into some basis of F.

A system of elements $g_1, .., g_k \in F$ is called primitive if $g_1, ..., g_n$ is a part of a basis F.

Notice that a system $g_1, ..., g_n$ is primitive if and only if, there exists an automorphism $\varphi \in AutF$ such that

$$g_1^{\phi} = x_1, ..., g_k^{\phi} = x_k$$

Corollary. $U = (u_1, ..., u_k)$ is primitive in F iff there exists a finite sequence of Whitehead automorphisms $\varphi_1, ..., \varphi_s$ that

$$|U^{\varphi_1}| > |U^{\varphi_1\varphi_2}| > ... > |U^{\varphi_1 \ \cdots \ \varphi_s}|$$

and

$$U^{\varphi_1 \cdots \varphi_s} = (y_1, ..., y_k), y_i \in X^{\pm 1}$$

Hence there is an algorithm which decides whether a system $U = (u_1 ..., \ u_k)$ is primitive or not. In particular, the problem whether a subgroup $\langle u_1, ..., u_k \rangle$ is a free factor of F or not, is decidable.

Test elements.

Let $F = F(x_1, .., x_n)$ be a free group with basis $X = \{x_1, ..., x_n\}$. Any tuple of elements $(u_1, ..., u_m) \in F^n$ gives rise to an endomorphism of F:

$$\varphi x_i \to u_i, i = 1, ..., n$$

and every endomorphism of F can be obtained this way.

Denote by $End(F)$ the set of all endomorphisms of F.

Due to Whitehead theorem we can recognize whether a given endomorphism $\varphi \in EndF$ is an automorphism of F. But this is a very time consuming algorithm and we would like to have some other test for φ to be an automorphism.

There is a very simple and effective test for isomorphism in the case F has rank 2. According to the theorem of Nielsen an endomorphism $\varphi \in EndF_2$ is an automorphism if an if and only if

$$[x_1{}^{\varphi}, x_2{}^{\varphi}] = f^{-1}[x_1, x_2]^{\pm 1} f$$

for some $f \in F_2$, i.e., if and only if, either the cyclic words $[x_1{}^{\varphi}, x_2{}^{\varphi}]$ and $[x_1, x_2]$ are equal, or the cyclic words $[x_1{}^{\varphi}, x_2{}^{\varphi}]$ and $[x_2, x_1]$ are equal.

In particular, $\varphi \in End$ is an automorphism if

$$[x_1, x_2]^{\varphi} = [x_1, x_2]$$

i.e., $[x_1, x_2] \in Fix\varphi$.

Definition. A word $w \in F$ is a test word if for any $\varphi \in EndF$

$$w^{\varepsilon} = w$$

implies that φ is an automorphism of F.

For example, $[x_1, x_2]$ is a test word in $F_2(x_1, x_2)$ (as well as $[x_1, x_2]$).

Obviously, x_1 is not a test word, since fixes x_1 and is not an automorphism of F_2.

One can generalize the example above to the following.

Remark. It w belongs to a proper free factor of F, then w is not a test element. This leads to the following.

Definition. The rank of $w \in F$ is the smallest rank of a free factor of F containing w.

It is clear, that if w is a test word, then the rank of w is equal to the rank of F.

The reverse is not true.

Example. Let $\varphi : F_2 \to F_2$ be defined by then $\varphi(w) = w$ (check!) but the rank of w is 2 (we will check it later).

A free factor is a particular example of a more general type of subgroups of F, so called retracts of F.

Definition. An endomorphism $\varphi : G \to G$ of a group G is called a retraction (or retract) if $\varphi^2 = \varphi$. If φ is a retraction of G, then the image $\varphi(G)$ is called a retract of G.

It is easy to see that if $\varphi : G \to G$ is a retraction, then $(g^\varphi)^\varphi = g^\varphi$, hence φ is identical on $\varphi(G)$. So a subgroup $R \leq G$ is a retract if and only if the identical map $i : R \to R$ can be extended to an endomorphism $G \to R$.

Notice, that if R is a free factor of F, then any endomorphism of R (not only identical) can be extended to an endomorphism of F. Therefore, every free factor of F is a retract.

We will see shortly, that there are retracts of F which are not free factors.

Theorem (T. Turner). A word $w \in F$ is a test word if w does not lie in any proper retract of F.

Before we prove this theorem we will consider some properties of retracts of F.

Definition. let $\varphi : F \to F$ be an endomorphism of F. Then the subgroup

$$\varphi^\infty(F) = \cap\varphi^i(F)$$

is called the stable image of φ.

It is easy to see, that $\varphi^\infty(F)$ is invariant under action of φ, i.e.,

$$g \in \varphi^\infty(F) \Rightarrow g^\varphi \in \varphi^\infty(F)$$

Denote by φ^∞ the restriction of φ on $\varphi^\infty(F)$.

Exercise. Prove that φ^∞ is an automorphism of $\varphi^\infty(F)$.

Notice, that $Fix\varphi \leq \varphi^\infty(F)$, and $\varphi^\infty(F) = F \Leftrightarrow \varphi \in AutF$.

Proposition. Let $\varphi : F \to F$ be an endomorphism of F. Then $\varphi^\infty(F)$ is a retract of F. If φ is a monomorphism of F, then $\varphi^\infty(F)$ is a free factor of F.

Lemma. $\varphi^\infty(F)$ is a free factor of $\varphi^n(F)$ for almost all n.

We will show first, that the conclusion of the proposition follows from this lemma.

Let n be such n that

$$\varphi^n(F) = \varphi^\infty(F) \star H$$

If φ is a monomorphism, then

$$\varphi^n : F \to \varphi^n(F) = \varphi^\infty(F) \star H$$

is an isomorphism. Put

$$\bar{H} = (\varphi^n)^{-1}(H)$$

Then

$$F = (\varphi^n)^{-1}(\varphi^\infty(F)) \star (\varphi^n)^{-1}(H) = \varphi^\infty(F) \star \bar{H}$$

Hence $\varphi^\infty(F)$ is a free factor of F.

Now denote by $\pi : \varphi^n(F) \to \varphi^\infty(F)$ the projection of $\varphi^n(F)$ on the free factor $\varphi^\infty(F)$.π is defined as follows: if $\varphi^n(F) = F(a_1, ..., a_m, a_{m+1}, ..., a_{m+k})$ and $\varphi^\infty(F) = F(a_1, ..., a_m)$, $H = F(a_{m+1}, ..., a_{m+k})$, then We have the following chain of homomorphisms:

$$F \to^{\varphi^n} \varphi^n(F) \to^{\pi} \varphi^\infty(F) \to^{(\varphi^\infty)^{-1}} \varphi^\infty(F)$$

Denote by p the composition of this sequence of homomorphisms. Then

$$p : F \to \varphi^\infty(F)$$

and $p^2 = p$. Hence $\varphi^\infty(F)$ is a retract of F.

$$F \stackrel{\varphi^n}{\to} \varphi^\infty(F) * H$$

$$\rho\varphi^\infty(F) = id_F$$

Open Problems:

1. Let $\varphi \in EndF_n$ and φ takes every primitive element of F to a primitive element of F. Is φ an automorphism of F_n?

2. $1 \neq u \in F$ is called strong test element if $\forall \varphi \in EndF$ if $\varphi(u) \neq 1$ and $\varphi(u) \in ncl(u)$ then $\varphi \in AutF$. Are these strong test elements?

3. Let $S \leq F$, R be a retract of F. Is it true that $S \cap R$ is a retract of S?

Proof of the Lemma. We will prove it in a more general form.

Let

$$(1) H_1 > H_2 > .$$

be a descending chain of free groups and $r(H_1)$ is finite or countable. Then the subgroup

$$H = \cap H_i$$

has a free basis $a_1, a_2, .$ such that any finite set

$$a_1 ... a_k$$

freely generate a free factor in all but finitely many H_i.

$\triangleright$. H is countable. Let

$$H = \{1, w_1, w_2, w_3, ..\}$$

be an enumeration of H.

Let H_{ji} be the first term of the chain (1) in which w_1 has smallest possible length as a word in a set of free generators of H_{ji}. If

$$w_1 = {a_1}^{\varepsilon_1}.{a_n}^{\varepsilon_n}$$

then each a_{i_k} is in H. For if $a_{i_k} \in H$ for every $\forall k \leq q$, but $a_{i_{k+1}}$ is not in H_{j_t} for some $H_t, t.j_i$, we chose $a_{i_{k+1}}$ as its own representative in a Schrier system for H_{j_1} and H_t. Then w_1 has shorter length in H_t. (Rewrite w_1 in S-symbols). Now $\langle a_1, ..., a_{i_p} \rangle$ is a free factors of H_{j_1}.

Chains of Subgroups.

Let $G_1, G_2, G_3, ..., G_i ...(i \in \mathbf{N})$ be a sequence of subgroups. If $G_i \leq G_{i+1}$ for every $i \in \mathbf{N}$, then the sequence is called an ascending chain of groups. We write this as (1)

$$G_1 \leq G_2 \leq G_3 \leq .(i \in \mathbf{N})$$

The union $G = \cup G_i$ forms a group with respect to the multiplication induced by the multiplications on each G_i. We say that (1) is strictly ascending chain if $G_i < G_{i+1}$ for each i.

If there exists an index n such that $G_{i+1} \neq G_i$ for each $i \geq n$, then we say that (1) stabilizes.

Example 1. Let F be an infinite cyclic group, then every ascending chain of subgroups of F stabilizes. In fact, such a chain stabilizes at most after n steps, where n is the index of the first term of the chain in F.

Example 2. Let F be a free group of rank ≥ 2. Then F contains a subgroup H of infinite rank. If $\{a_1, a_2, ...\}$ is a basis of H, then (3)

$$< a_1 >< < a_1, a_2 >< < a_1, a_2, a_3 >< ...$$

is an infinite strictly ascending chain of subgroups of F. Notice, that the sequence of ranks of the subgroups in (3) is not bounded from above.

Let $\{F_i | i \in \mathbf{N}\}$ be a sequence of free groups. We say that the ranks of groups $F_i, i \in \mathbf{N}$, are uniformly bounded if there exists an integer r such that $r(F_i) \leq r$.

It turns out that every ascending chain of subgroups of a given true group stabilizes provided the ranks of these subgroups are uniformly bounded. To show this we need to prove first the following.

Theorem. Let (4)

$$F_1 \leq F_2 \leq ...$$

be an ascending chain of free groups with uniformly bounded ranks. Then the union

$$G = \cup F_i$$

is a Hopfian group.

Proof. Let γ_0 be the least positive integer for which there exists a chain of the type (4) with rank uniformly bounded by γ_0 and such that $G = \cup F_i$ is not Hopfian. Then there exists a normal non-trivial subgroup $N \triangle G$ such that

$$G \simeq G/N$$

Then

$$G \simeq G/N = \cup F_i N/N \simeq \cup F_i/N \cup F_i$$

Notice, that every free subgroup of G is contained in some F_i, hence it is free. In particular, the group $F_i N/N \simeq F_i/N \cap F_i$ is free. Since N is not trivial, then $N \cap F_i \neq 1$ for each $i \geq n$. Therefore, the rank of the free group $F_i N/N$ is strictly less then that of F_i, i.e., $r(F_i N/N) < \gamma_0$.

Now G is the union of the chain of free groups

$$F_n N/N \leq F_{n+1} N/N \leq .$$

with the rank uniformly bounded by $\gamma_0 - 1$-contradiction with the minimality of γ_0. It follows, that G is Hopfian.

Corollary. Let F be a free group. Then every ascending chain of subgroups of F, with uniformly bounded ranks stabilizes.

Proof. Indeed, let G be the union of an ascending chain of subgroups of F with uniformly bounded ranks. Then G is free as a subgroup of F and G is Hopfian by the theorem above. Since a free group of infinite rank is not Hopfian, and the chain stabilizes, as desired.

Let H be a subgroup of F and let γ be the rank of H. Then by the corollary above every chain of the type

$$H = F_1 \leq F_2 \leq .$$

where F_i are subgroups of F of the rank γ, stabilizes. Therefore, there exists a maximal subgroup $\bar{H}$ of rank γ in F, containing H, i.e., a subgroup of rank γ which contains H, but which is not a proper subgroup of any other subgroup of rank γ in F. We call such subgroups of F γ-maximal.

Proposition. For a given subgroup H of rank $\gamma = 1, 2$ in a free group F there are only finitely many γ-maximal subgroups in F containing H. Moreover, if H is given by a set of generators $h_1, ..., h_m$ then one can effectively enumerate all such γ-maximal subgroups.

Proof. Let $\bar{H}$ be a γ-maximal subgroup containing H. Then all the words $h_1, ..., h_m$ are readable in the automaton $\Gamma(\bar{H})$. Moreover, if we denote by Γ_H the subgraph of $\Gamma(\bar{H})$ consisting of all paths with the labels $h_1, ..., h_m$ in $\Gamma(\bar{H})$, then $\Gamma(\bar{H}) - \Gamma_H$ is a tree. Indeed, if T is a maximal subtree of Γ_H,

then we can extend T to a maximal subtree T of $\gamma(\bar{H})$. Now if there exists a loop in $\Gamma(\bar{H}) - \Gamma_H$ then there exists a fundamental loop in Γ_H with respect to the tree T such that the label of this loop does not belong to H. Hence the rank of $\bar{H}$ is greater then the rank of H-contradiction.

It shows, that the group $\bar{H}$ coincides with the subgroup accepted by the automaton Γ_H. Clearly, there are only finitely many ways to construct different graphs Γ_H starting with generators $h_1, ..., h_m$ (see section about test elements).

To finish the proof one needs to show that the rank of the subgroup accepted by Γ_H is not less then the rank of H. Indeed, if $\gamma(H) = 1$ or 2, then H is not contained in any subgroup of smaller rank. $\square$.

Notice, the proposition does not hold for $\gamma = 4$:

Example. Let H be a subgroup accepted by Now one can extend Γ_H by adding an arbitrary loop to a graph accepting a subgroup $\bar{H}$ containing H, so one can construct infinitely many such different (?) subgroups.

The example above shows, that at least in rank ≥ 4 the maximal extensions $\bar{H}$ can be very different. Now, if one would like to have an extensions $\bar{H}$ of the same rank as H and s.t. H does not contain in any proper free factor of $\bar{H}$, then we can describe such "algebraic" maximal extensions of H. Indeed, in this event the rank of Γ_H cannot collapse, i.e., $rank(\Gamma_H) = rank(H)$. And we have the description of "algebraic extensions" $\bar{H}$ similar to the proposition above.

Problem. Is any maximal extension $\bar{H}$ of H "algebraic" if rank $(H) = 3$.

Proposition. Let H be a finitely generated subgroup of F of finite index Then H is maximal, i.e., $H = \bar{H}$.

Proof. Let $H \leq \bar{H} \leq F$ and $[F : H] < \infty$. If $\gamma(H) = \gamma(\bar{H})$, then

$$[F : \bar{H}] = \overset{\gamma(\bar{H})-1}{\gamma(F)} - 1 = \overset{\gamma(H)-1}{\gamma(F)} - 1 = \lfloor F : H \rfloor$$

Hence $[\bar{H} : H\rfloor = 1$ and $H = \bar{H}$.

Descending Chain of Subgroups

If $G_i, i \in \mathbf{N}$ are subgroups of a group G and $G_i \geq G_{i+1}$ for every $i \in \mathbf{N}$, then the sequence (1) $G_1 \geq G_2 \geq ... \geq G_i \geq G_{i+1} \geq ...$ is called a descending chain of subgroups of G. (And strictly descending in the case when $G_i > G_{i+1}, i \in \mathbf{N}$).

If $F =< x >$ is an infinite cyclic group, then the following is an infinite strictly descending chain of subgroups of $< x >$:

$$< x >>< x^2 >>< x^3 >>$$

This example shows that if a group G has an element of infinite order, then there are infinite strictly descending chain of subgroups of G.

For a descending chain (1) denote by G_w the intersection of all terms in (1):

$$G_w = \cap G_i$$

Our primary concern in this section will be to investigate when $G_w = 1$.

Theorem (Takahashi). Let

$$F = F_1 > F_2 > F_3 > ...$$

be a strictly descending chain of subgroups of a free group F such that for each i, F_{i+1} does not contain any primitive element of F_i. Then $F_w = 1$.

Proof. By induction on i we prove that every non-trivial element $w \in F_i$ has length $|w| \geq i$ in F. $i = 1$ is obvious. $i = n+1$. Suppose every non-trivial element of F_i has length $\geq i$. Let X_i be a N-reduced basis of F_i. If $w \in F_{i+1}$, then

$$w = y_1 ... y_m, y_k \in X_i$$

If $m \leq 2$ then either w is primitive in F_i or $w = {y_1}^2$ and in this case $|w| = |{y_1}^2| > |y_1| \geq i$.

Corollary. Let

$$F = F_1 > F_2 > ... > F_n > ...$$

be a strictly descending chain of subgroups of F. Suppose that F_{i+1} is invariant under $AutF_i$, then $F_w = 1$.

Proof. Indeed, if F_{i+1} is invariant in F_i, then F_{i+1} does not contain any primitive element from F_i.

Regular Subsets of Free Groups.

1. Definition of an automaton over a monoid M. (Standard automaton, deterministic automaton).

1.1 Regular sets (as accepted by a finite automata closure of regular sets under $\cdot; *$.

2. Pumping lemma.

3. $\sum \to M \rightsquigarrow \sum^* \to M$

$$\sum = (XuX^{-1}), \sum^* \xrightarrow{\varphi} F(x) \subset \sum^*$$

Lemma. $L \subset \sum^*$ is regular $\Rightarrow \varphi(L)$ is regular $(\in \sum^{ast}$.

4. A subgroup is generated by a regular set $\Leftrightarrow$ it is finitely generated.

5. Regular sets: $\{x^2 | x \in F\}$ is regular. Then let $w \in F(x)$ and $W(F) = \{f | \exists y_1, ..., y_n \in F f = w(y_1, ..., y_n)\}$. Then $w(F)$ is regular $\Leftrightarrow (F) = F$ or $w(F) = 1$. (Nikita).

6. Then let $u(x, \bar{g}) = 1$ be an equation of one variable over F. Then the solution set in $FV_F(u) == a_1 b_1^x c_1 \cup a_2 b_2^z c_2 \cup \ ... \cup a_k b_k^z c_k$ for some $a_i, b_i, c_i \in F$. (prove without Appel).

7. Problem: Characterize quadratic equations over F with regular solution sets.

8. Characterize regular orbits of subgroups $A \leq AutF$ over F. For example, let $f \in F$, $\alpha\beta \in AutF$ when the orbit $< \alpha, \beta > f$ is regular?