Notes of Zelmanov's Algebraic Lectures

taken by Li Yunsheng based on the contents of the lectures

Lecture Information

Lecturer: Efim Zelmanov
E-mail: efim.zelmanov@gmail.com
Website of the Course:
https://sustech-math.github.io/zelmanov.html
Office Hour: Tuesday 10:00 a.m.~12:00 a.m. at office at the Math Center
No midterm
No formal final exam, only oral presentation in person.
No textbook to follow

Contents

§1 Lecture 1

Review of Abstract Algebra

- Groups, Subgroups $H < G$, Homomorphisms
- Normal Subgroups $H \lhd G$, Quotient of Groups, the First Isomorphism Theorem
- Natural Homomorphism $G/H_1 \to G/H_2$, $H_1 \lhd H_2 \lhd G$.
- Commutation
- Groups Generated by a Subset: $X \subset G$,

$$\langle X \rangle := \{x_1^{\varepsilon_1} \cdots x_n^{\varepsilon_n} \mid x_1, \cdots, x_n \in X, \varepsilon_i = \pm 1\}$$

- Rings (with identity), Subrings, Homomorphisms
- Ideals $I \trianglelefteq R$, Quotient by Ideals (Factor Ring), the First Isomorphism Theorem
- Natural Homomorphism $R/I_1 \to R/I_2$, $I_1 \subset I_2 \trianglelefteq R$.
- Cartesian Products of Groups $\overline{\prod_{i \in I} G_i}$, Direct Products of Groups $\prod_{i \in I} G_i$
- Direct Sums of Rings $\bigoplus_{i \in I} R_i$
- Fields, (not necessarily commutative) Algebras over a Field A/F
- Algebras Generated by a Subset: $X \subset A$,

$$\langle X \rangle_F := \left\{ \sum \alpha x_1, \cdots, x_n \mid \alpha \in F; x_1, \cdots, x_n \in X \right\}$$

Note that the above formula holds for all kinds of (associative) algebras

- Vector Spaces, Linear Transformations $\mathrm{Lin}_F(V)$
- Modules over an Algebra: A/F, V a vector space over F. A bilinear map $A \times V \to V$ satisfying $a(bv) = ab \cdot v$ and $1_A v = v$ makes V into a (left) module over A. Note that a bilinear map is equivalent to a homomorphism of algebras $\varphi\colon A \to \mathrm{Lin}_F(V)$.
- Semigroup (with identity), Congruence Relation: $x \sim y \Rightarrow xz \sim yz$ and $zx \sim zy$.
- the "First Isomorphism Theorem" for semigroups:

  Let $\varphi : S_1 \to S_2$ be a homomorphism of semigroups. Define a congruence relation by that $x \sim y$ if and only if $\varphi(x) = \varphi(y)$, then $S_2 \cong S_1/\sim$
- Natural Homomorphism $S/\sim_1 \to S/\sim_2$, $\sim_1 \subset \sim_2$.

Plan of the Course

- Semigroups
- Free groups, Free Algebras
- Rings of Fractions
- Division Rings
- Ultraproduct

Full syllabus is available on the website.

§2 Lecture 2

Free Semigroup

Given a set X, the *free semigroup generated by* X is

$$X^* := \{\text{all words of elements in } X\}.$$

It enjoys the following universal property:

Proposition 2.1. Universal Property of Free Semigroup

Let S be a semigroup and $\varphi\colon X \to S$ be a mapping, then φ uniquely extends to a homomorphism $\varphi\colon X^* \to S$.

It is easy to observe the following: let S be a semigroup and $\mathscr{A} = \{a_i\}_{i\in I}$ be a set of generators of S, then the inclusion $\mathscr{A} \hookrightarrow S$ induces a surjective homomorphism $\varphi\colon \mathscr{A}^* \to S$.

Therefore $S \cong \mathscr{A}^*/\sim$, where $\sim$ is defined by $a \sim b$ if and only if $\varphi(a) = \varphi(b)$.

Generated Congruence

Let $\sim$ be a congruence on S, then it can be seen as a subset of $S \times S$. Let $R \subset S \times S$, we say that R generates $\sim$ if $\sim$ is the smallest congruence that contains R, or equivalently,

$$R = \bigcap \text{all congruences that contains } R.$$

The congruence generated by R always exists, noticing that $S \times S$ is itself a congruence.

Presentation

Let S be a semigroup and is isomorphic to the quotient of a free semigroup X^*, $S \cong X^*/\sim$. Let $R = \{a_k \times b_k\}_k$ be a subset of $X^* \times X^*$ that generates $\sim$, then we say that

$$S = \langle X \mid a_k = b_k \rangle,$$

and S is *presented* by generators X and relations R.

If $|X| < \infty$ and $|R| < \infty$, then we say that S is *finitely presented*. It turns out that the notion of finitely presented is independent of the choice of generators, see proposition 2.3.

Undecidable Word Problem

In general, given a presentation, it is algorithmatically undecidable to tell whether two words are equal under that presentation.

Extension Problem

Let $S = \langle s_i, i \in I \rangle$ and T be two semigroups. Let $\{t_i\}_{i\in I} \subset T$ and consider the map

$$\varphi\colon \{s_i\}_{i\in I} \to \{t_i\}_{i\in I} : s_i \mapsto t_i.$$

Question

Is φ extendable to a homomorphism?

The answer is quite simple. Write $S = \langle x_i, i \in I \mid a_k = b_k \rangle$ as is presented by $X = \{x_i\}_{i\in I}$ and $R = \{a_k \times b_k\}_k$, where $x_i \mapsto s_i$ gives the natural map $X^* \to S$, then

Proposition 2.2. Characterization of Extending

φ extends to a homomorphism if and only if it preserves all the defining relations, in other words, $a_k(\boldsymbol{t}) = b_k(\boldsymbol{t})$.

Proof. Just notice that we have naturally $S \cong X^*/\sim$ and $X^*/\sim_1 \hookrightarrow T$, and $\sim \subset \sim_1$ gives a homomorphism $X^*/\sim \to X^*/\sim_1$. $\square$

Finitely Presented

Let $S = \langle s_1, \cdots, s_m \rangle = \langle s'_1, \cdots, s'_k \rangle$ be generated by two different sets of generators.

Proposition 2.3. Finitely Presented is Well-defined

If S is finitely presented in $s_1, \cdots, s_m$, then it is also finitely presented in $s'_1, \cdots, s'_k$.

Proof. Write $S = \langle x_1, \cdots, x_m \mid a_1 = b_1, \cdots, a_n = b_n \rangle$ with $x_i \mapsto s_i$. Since s_i and s'_i generate S, we have

$$s_i = c_i(\boldsymbol{s}'), \quad 1 \le i \le m,$$

$$s'_j = d_j(\boldsymbol{s}), \quad 1 \le j \le k,$$

where $c_i(\boldsymbol{s})$ denotes some algebraic combination of s_j's, and similarly is $d_j(\boldsymbol{s})$ defined. These give rise to two kinds of relations that are satisfied in S:

$$a_l(\boldsymbol{c}(\boldsymbol{s}')) = b_l(\boldsymbol{c}(\boldsymbol{s}')), \quad 1 \le l \le n, \tag{I}$$

$$s'_j = d_j(\boldsymbol{c}(\boldsymbol{s}')), \quad 1 \le j \le k. \tag{II}$$

Let $Y^* = \langle y_i, \cdots, y_k \rangle$ be a free semigroup, the homomorphism $Y^* \to S \colon y_j \mapsto s'_j$ gives $S \cong Y^*/\sim_1$. Under the convention $y_j \leftrightsquigarrow s'_j$, the two kinds of relations above are all included by $\sim_1$, hence if we let them generate a relation $\sim_2$ on Y^*, then $\sim_2 \subset \sim_1$, giving rise to a homomorphism $Y^*/\sim_2 \to Y^*/\sim_1$.

On the other hand, the first kind of relations, with proposition 2.2, defines a homomorphism

$$S \cong X/\sim \to Y^*/\sim_2 \colon s_i \mapsto x_i \mapsto c_i(\boldsymbol{y}').$$

It remains only to show that these two homomorphisms are the inverse to each other.

Exercise 2.1

Complete the rest of the proof.

It is easy to see that $S \cong X/\sim \to Y^*/\sim_2 \to Y^*/\sim_1 \cong S$ is the identity on S. For the other direction, we have

$$\begin{array}{ccccc} Y^*/\sim_2 & \to & Y^*/\sim_1 \cong S \cong X/\sim & \to & Y^*/\sim_2 \\ y_j & \mapsto & y_j \leftrightsquigarrow s'_j \leftrightsquigarrow d_j(\boldsymbol{x}) & \mapsto & d_j(\boldsymbol{c}(\boldsymbol{y})) \overset{\text{by (II)}}{=\!=} y_j \end{array}$$

$\square$

Therefore, as we claimed before, the notion of finitely presented does not depend on the choice of the finite set of generators.

Length-lex (Lexicographical) Order

Consider a free semigroup $X^* = \langle x_i, i \in I \rangle$ whose index I is ordered. We define an order on X^* by the following: for any two elements

$$v = x_{i_1} \cdots x_{i_n}, \quad w = x_{j_1} \cdots x_{j_m},$$

1. if $n > m$ or $m < n$, then $v > w$ or $w < v$ respectively;
2. else $n = m$, then compare i_1 and j_1: if $i_1 > j_1$ then $v > w$; if $i_1 = j_1$, then compare i_2 and j_2 and so on.

This order on X^* is called *the length-lex order*, or *the lexicographical order*.

Definition 2.1. Minimality Condition

An ordered set satisfies *the minimal condition* if there does not exists an infinite descending chain $a_1 > a_2 > \cdots$ in it.

Theorem 2.4. Lexicographical Order Inherits Minimality Condition

If $X = (X, >)$ satisfies the minimal condition, then the length-lex order on X^* also satisfies the minimality condition.

Proof. Suppose that

$$v_1 > v_2 > \cdots$$

is an infinitely descending chain in X^*, then the length of v_i's forms a descending chain

$$\ell(v_1) \geq \ell(v_2) \geq \cdots$$

in $\mathbb{N}$. Thus there must exists n such that $\ell(v_n) = \ell(v_{n+1}) = \cdots$. Then $v_n > v_{n+1} > \cdots$ is an infinitely descending chain in X^* where all elements have the same length. The first letters of this sequences gives $x_{i_1} \geq x_{i_2} \geq \cdots$, which stabilizes since X has the minimality condition. Cut the sequence again and then consider the second letters, and so on. After $\ell(v_n)$ steps, we see that the original chain must stabilize, contradicting the assumption. $\square$

§3 Lecture 3

Free Semigroup Algebras

Let F be a field and S be a semigroup, we can consider the *semigroup algebra on S*,

$$FS := \{\alpha_1 s_1 + \cdots \alpha_n s_n\}.$$

Remark 3.1

Unlike building a ring on an abelian group, here the operation of the semigroup induces the multiplication in the algebra, instead of the summation.

Given any set X, we can consider the *free associative F-algebra on the set of free generators X*

$$F\langle X \rangle := FX^* := \{\alpha_1\omega_1 + \cdots + \alpha_n\omega_n \mid \alpha_i \in F, \omega_i \in X^*\}.$$

It enjoys the following universal property:

Proposition 3.1. Universal Property of Free Associative F-algebra

For any F-algebra A and an arbitrary mapping $\varphi\colon X \to A$, φ uniquely extends to a homomorphism of F-algebras $\varphi\colon F\langle X \rangle \to A$.

Consider the case where A is generated by $\{a_j\}_{j\in J}$, $X := \{x_j\}_{j\in J}$ and $\varphi\colon x_j \mapsto a_j$, then the induced homomorphism $\varphi\colon F\langle X \rangle \to A$ is surjective, giving $A \cong F\langle X \rangle / I$ where $I = \ker\varphi$.

Let $R \subset F\langle X \rangle$, we say that R generates I as an ideal if I is the smallest ideal containing R, or equivalently,

$$I = \left\{ \sum_j a_j \iota_j b_j \mid a_j, b_j \in F\langle X \rangle, \iota_j \in R \right\}.$$

Similar as the presentation of semigroups, when $A \cong F\langle X \rangle / I$ and I is generated by R, we may write

$$A = \langle X \mid R = 0 \rangle .$$

Also similarly, finitely presented is defined and can be proved to be well-defined.

Reduction to Irreducible Elements

The same undecidable word problem exists for presentations of algebras, but we do have some algorithm under certain conditions:

Keep the notations above. Now suppose X is (totally) ordered and satisfies the minimality condition. For an element $r \in R$, it can be written as

$$r = \alpha_1\omega_1 + \cdots + \alpha_n\omega_n,$$

with $\alpha_i \neq 0$ and $\omega_i \in X^*$ are distinct words. The maximal element among $\omega_1, \cdots, \omega_n$, $\bar{r} := \omega_i = \max(\omega_1, \cdots, \omega_n)$, is called *the leading monomial of* r, and its coefficient α_i is called *the leading coefficient of* r. In A, we have $r = 0$, which gives the following relation:

$$\alpha_i\omega_i = -\sum_{j\neq i}\alpha_j\omega_j \quad \Rightarrow \quad \omega_i = -\sum_{j\neq i}\frac{\alpha_j}{\alpha_i}\omega_j.$$

Definition 3.1. Reducible Word

A word $v \in X^*$ is called *reducible* if it contains a leading monomial $\bar{r}$ of some $r \in R$ as a *subword*, i.e. $v = v'\bar{r}v''$ for some $v', v'' \in X^*$.

Therefore if v is reducible, then in A, $v = \sum_k \alpha_k u_k$, $\alpha_k \in F$, $u_k < v$. This means that v can be *reduced* into a sum of "smaller" words.

Definition 3.2. Irreducible Word

A word is *irreducible* if it is not reducible.

Let us denote by $Ir \subset X^*$ the set of all irreducible words.

Proposition 3.2. Irreducible Words Span the Algebra

Ir spans A.

Proof. The result follows immediately by the reduction of reducible words and the minimality condition. □

Gröebuer-Shirshov Bases

We wish to use Ir as a basis of A, which would solve our undecidable word problem in this case completely since we can reduce any element in A to a linear combination of irreducible words within finitely many steps. Hence the following question arises:

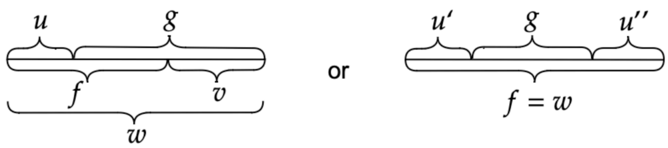
Question

When is Ir linearly independent in A?

Definition 3.3. Composition of Words

We say that two words v, w *admit a composition* if the end of one of these words equals the beginning of the other one, e.g. $v = x_3x_5x_1$ and $w = x_5x_1^2x_4$ admit a composition, or one of these words is a subword of the other one.

Suppose that $f, g \in R$ and the leading coefficients of f and g are 1. Suppose that $\bar{f}, \bar{g}$ admit a composition, i.e. they can be pieced together to get a word w as the following illustrates:



We define the composition of f and g, $(f,g)_w$, as

$$(f,g)_w = fv - ug \text{ or } f - u'gu'',$$

respectively to the two cases in the picture. Note that there may be several different choices of w associated to f and g, for example, $axyxb$ and $axyxyxb$ are both available w's of $f = axyx$ and $g = xyxb$.

Theorem 3.3. Gröebuer-Shirshov Bases

Ir is a basis of A if and only if for any two relations $f,g \in R$ that admit a composition, all these compositions $(f,g)_w$ reduce to 0 (in the free algebra) (instead of a nontrivial linear combination of irreducible elements).

By "reduce" we mean to substitute reducible words by the sum of smaller words given by the relations in R and obtain a new element in $F\langle X \rangle$ which is "closer" to a sum of irreducible words, and so on, until we get a linear combination of irreducible words (or 0).

The basis Ir found in this way is called the *Gröebuer-Shirshov basis.*

Remark 3.2. Reformulation of Reduction

The procedure of reduction can be reformulated as the following: An element $r \in R$ is of the form $r = \bar{r} - r'$. A word $v\bar{r}u$ is equal to $v(r + r')u = vru + vr'u$ in the free algebra, which lives in $vr'u + I(R)$. By passing through such equalities, any element is a linear combination of irreducible words whose leading monomial is smaller than the original one of the element, plus $I(R)$. The "old version" of reduction differs with this reformulation in the sense that it erases elements in $I(R)$. So far we can conclude that a necessary condition for R to be *closed under composition* (i.e. for any $f,g \in R$, $(f,g)_w$ reduces to 0) is that for any $f,g \in R$, we can write in the free algebra $(f,g)_w = \sum_j \alpha_j v_j r_j u_j$ with $\alpha_j \in F$, $v_j, u_j \in X^*$, $r_j \in R$ and $v_j \bar{r}_j u_j < w$ for all j.

Exercise 3.1

Show that the necessary condition above for R to be closed under composition is in fact sufficient.

Remark 3.3

A consequence of this theorem is that it does matter how we reduce words, since different non-trivial linear combination of irreducible words would be distinct.

One direction is easy: $(f,g)_w$ is always 0 in A since $f = g = 0$ in A and the procedure of reduction gives equality in A. Hence if $(f,g)_w$ does not reduce to 0, then we obtain that a nontrivial linear combination of irreducible words equals 0 in A, indicating that Ir is not linearly independent in A.

The other direction is straightforward, but is long and uninteresting. Before the complete proof, let us go through some examples of applications of the theorem:

Applications of Gröebuer-Shirshov Bases

Example 3.1

Consider $\langle x, y \mid yx - xy = 1 \rangle$ with order $x < y$. There is nothing to compose, and the irreducible words are words of the form $x^i y^j$, which form a basis by the theorem. This algebra is isomorphic to the *Weyl algebra*, the algebra generated by $y = \frac{d}{dt}$ and $x = t$ seen as linear operators on the space of differentiable functions.

Example 3.2

Consider $\langle x, y, z \mid [x,y] = z, [z,y] = 2y, [z,x] = -2x \rangle$ with order $x < y < z$, where $[a,b] := ab - ba$. The relations give the following reductions

$$\begin{cases} yx \to xy - z \\ zy \to yz + 2y \\ zx \to xz - 2x \end{cases}$$

Now that the first two elements admit a composition: $zy - yz - 2y$ and $yx - xy + z$. The composition is

$$(zy - yz - 2y)x - z(yx - xy + z) = -yzx - 2yx + zxy - z^2.$$

The reduction goes:

$$\begin{aligned}
&\to -y(xz - 2x) - 2(xy - z) + (xz - 2x)y - z^2 \\
&= -yxz + 2yz - 4xy + 2z + xzy - z^2 \\
&\to -(xy - z)z + 2(xy - z) - 4xy + 2z + x(yz + 2y) - z^2 = 0.
\end{aligned}$$

Since these are the only two elements in the relation that admit a composition, we see by theorem 3.3 that the irreducible words $x^i y^j z^k$ form a basis.

§4 Lecture 4

Example 4.1

This example shows that the order on X matters. Consider $\langle x, y \mid y^2x - xyx = 0 \rangle$. When $x < y$, the leading word is y^2x and there is no nontrivial composition with itself; when $y < x$, the leading word is xyx, which admits a nontrivial composition with itself, $xyxyx$. The first order, by the theorem, shows that the irreducible words (in that order) form a basis of A. However, for the second order we have nontrivial composition

$$(f, f)_{xyxyx} = (xyx - y^2x)yx - xy(xyx - y^2x) = -y^2xyx + xy^3x.$$

The relation gives reduction

$$xyx \to y^2x,$$

Hence the above composition reduces to

$$-y^4x + xy^3x,$$

which is a nontrivial linear combination of irreducible words, showing that the irreducible words (in this order) are not linearly independent.

The next example is about Lie algebras. Let us briefly recall some definitions.

Definition 4.1. Lie Algebra

A *Lie algebra* L is a vector space with a bilinear operation $[\cdot, \cdot]\colon L \times L \to L$ that satisfies

(1) (Antisymmetry) $[a, b] = -[b, a]$;

(2) (Jacobi identity) $[[a, b], c] + [b, c], a] + [[c, a], b] = 0$,

for any $a, b, c \in L$.

Clearly, a *homomorphism of Lie algebras* is defined as a linear map that commutes with the brackets.

Definition 4.2. Representation of Lie Algebra

A *representation of a Lie algebra* L is a homomorphism of Lie algebras $\varphi\colon L \to A^{(-)}$, where A is an associative algebra and $A^{(-)}$ is the Lie algebra with the bracket $[a, b] = ab - ba$. A *homomorphism of representations of* L is a homomorphism of Lie algebras $A^{(-)} \to B^{(-)}$ that makes the following triangle commutes.

$$\begin{tikzcd} A^{(-)} \arrow[r] & B^{(-)} \\ L \arrow[u] \arrow[ur] & \end{tikzcd}$$

A representation of L is called *universal* if it is initial in the category of representations of L.

Note that for a Lie algebra L there may not exists an algebra A such that $L = A^{(-)}$.

Lemma 4.1. Image of Lie Algebra Generates Universal Enveloping Algebra

Let $u\colon L \to U^{(-)}$ be a universal representation of L. Then U is generated by $u(L)$ as an associative algebra.

Proof. Let $\langle\, u(L)\, \rangle$ be the associative algebra generated by $u(L)$, then $u\colon L \to \langle\, u(L)\, \rangle^{(-)}$ is also a representation of L. The universality gives a homomorphism of representations of L, $U^{(-)} \to \langle\, u(L)\, \rangle^{(-)}$, which is identical to the identity when restricted on $\langle\, u(L)\, \rangle$; in particular it is surjective. Note that the inclusion $\langle\, u(L)\, \rangle^{(-)} \hookrightarrow U^{(-)}$ is also a homomorphism of representations of L. The composition $U^{(-)} \to \langle\, u(L)\, \rangle^{(-)} \hookrightarrow U^{(-)}$ gives a homomorphism of representations of L, which must be the identity on $U^{(-)}$, implying that $U^{(-)} \to \langle\, u(L)\, \rangle^{(-)}$ is injective. Therefore $U = \langle\, u(L)\, \rangle$. □

The initial property gives immediately that

Proposition 4.2. Uniqueness of Universal Enveloping Algebra

If a universal representation exists then it is unique up to isomorphism.

Also, in this case, the existence is always true:

Proposition 4.3. Existence of Universal Enveloping Algebra

The universal representation of a Lie algebra L always exists.

Proof. Let $\{e_i\}_{i \in I}$ be a basis of L (for infinite-dimensional L, use the Hamel basis), then we have

$$[e_i, e_j] = \sum_k \gamma_{ij}^k e_k,$$

for some $\gamma_{ij}^k \in F$ for any $i, j \in I$. Write $X = \{x_i\}_{i \in I}$ and consider

$$U \coloneqq \langle\, X \mid x_i x_j - x_j x_i - \sum_k \gamma_{ij}^k x_k = 0\, \rangle\,.$$

The homomorphism $\varphi\colon L \to U^{(-)}$ defined by $e_i \mapsto x_i$ gives the universal representation, as one can verify. □

Note that φ is never surjective, because $\varphi(L) = \operatorname{span}_F\{x_i \mid i \in I\} \subsetneq U$ since $U \ni x_i x_j \notin \operatorname{span}_F\{x_i \mid i \in I\}$. Let us call the unique U in the universal representation of L as *the universal enveloping algebra of the Lie algebra L*.

Example 4.2

Example. Consider the famous Lie algebra $sl_2(F) \coloneqq \{2 \times 2 \text{ matrices with zero trace}\}$. It has basis

$$e = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad f = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad h = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$[e, f] = h, \quad [f, h] = 2f, \quad [e, h] = -2e.$$

Its universal enveloping algebra is exactly $\langle\, x, y, z \mid [x, y] = z, [y, z] = 2y, [x, z] = -2x \,\rangle$, which has been discussed in example 3.2.

Poincare-Birkhoff-Witt Theorem

Let $R = \{x_i x_j - x_j x_i - \sum_k \gamma_{ij}^k x_k \mid i, j \in I\}$ and introduce any order on I with the minimality condition, which always exists provided the Axiom of Choice.

Theorem 4.4. Poincare-Birkhoff-Witt

R is closed with respect to compositions, i.e., for any $f, g \in R$, $(f, g)_w$ reduces to 0. And the irreducible words $\{x_{i_1} \cdots x_{i_k} \mid i_1 \le \cdots \le i_k\}$ form a basis of $U = \langle\, X \mid R = 0 \,\rangle$.

Proof. See the next lecture. □

Corollary 4.4.1. Lie Algebra Embeds Into Universal Enveloping Algebra

Let $\varphi\colon L \to U^{(-)}$ be the universal representation of L, then φ is an embedding.

Proof. For any $\sum_i \alpha_i e_i \neq 0$ in L, it is mapped by φ to $\sum_i \alpha_i x_i$ in U, which is a non-trivial linear combination of irreducible words, hence is nonzero by Poincare-Birkhoff-Witt Theorem. □

§5 Lecture 5

Proof of the Poincare-Birkhoff-Witt Theorem

We first show that R is closed with respect to compositions. Write $\{x_i, x_j\} \coloneqq \sum_k \gamma_{ij}^k x_k$, then $\operatorname{span}_F\{x_i \mid i \in I\}$ with this bracket is a Lie algebra that is isomorphic to L. In particular, the bracket $\{\cdot, \cdot\}$ satisfies the Jacobi identity. Any two elements in R that admits a nontrivial composition are of the form

$$\begin{aligned} &x_j x_i - x_i x_j - \{x_j, x_i\}, \\ &x_k x_j - x_j x_k - \{x_k, x_j\}, \end{aligned}$$

where $i < j < k$. Their composition is

$$\begin{aligned} &(x_k x_j - x_j x_k - \{x_k, x_j\}) x_i - x_k (x_j x_i - x_i x_j - \{x_j, x_i\}) \\ =& - x_j x_k x_i - \{x_k, x_j\} x_i + x_k x_i x_j + x_k \{x_j, x_i\} \\ \to& - x_j (x_i x_k + \{x_k, x_i\} - \{x_k, x_j\} x_i + (x_i x_k + \{x_k, x_i\}) x_j + x_k \{x_j, x_i\} \\ =& - x_j x_i x_k - x_j \{x_k, x_i\} - \{x_k, x_j\} x_i + x_i x_k x_j + \{x_k, x_i\} x_j + x_k \{x_j, x_i\} \\ \to& - \{x_j, x_i\} x_k - x_j \{x_k, x_i\} + x_i \{x_k, x_j\} + \{x_k, x_i\} x_j + x_k \{x_j, x_i\} \\ =& [\{x_i, x_j\}, x_k] + [\{x_k, x_i\}, x_j] + [\{x_j, x_k\}, x_i] \\ \to& \{\{x_i, x_j\}, x_k\} + \{\{x_k, x_i\}, x_j\} + \{\{x_j, x_k\}, x_i\} = 0, \end{aligned}$$

where $[\cdot, \cdot]$ denotes the commutator, i.e., $[a, b] = ab - ba$. This ends the first part. Keep in mind that the above procedure, though looks complicated, can be done automatically by a computer with only one single click of button.

The following displays one reason why the universal enveloping is important. Recall that if L is a Lie algebra and V is a vector space, then a homomorphism $L \to \operatorname{Lin}_F(V)^{(-)}\colon a \mapsto T_a$ defines an action of L on V, and we have $T_{[a,b]} = T_a T_b - T_b T_a$. Lifting this homomorphism to $U^{(-)} \to \operatorname{Lin}_F(V)^{(-)}$, we then obtain a homomorphism of associative algebras $U \to \operatorname{Lin}_F(V)$.

Joke

Associative algebras are in general easier to deal with than Lie algebras. However, the above procedure is still a trade off of difficulties: even if L is finitely dimensional, U is infinite dimensional. This shows the law of conservation of difficulty.

The rest of the Poincare-Birkhoff-Witt Theorem follows from the other direction of the theorem 3.3, which we are now going to prove.

Proof of the Groöebuer-Shirshov Bases Theorem

Before the proof, let us briefly recall the statement of the theorem:

Theorem. Groöebuer-Shirshov Bases

Ir is a basis in A if and only if for any two relations $f, g \in R$ that admit a composition, all these compositions $(f,g)_w$ reduce to 0.

Proof. Recall that the necessity has been proved right after theorem 3.3. For the other direction, we show that for any nonzero $f \in I(R)$ the leading monomial $\bar{f}$ is reducible, hence a nontrivial linear combination of irreducible words is never zero modulo $I(R)$, implying the linear independence of irreducible words.

For any $f \in I(R)$, we can write $f = \sum_i \alpha_i u_i r_i v_i$ for finitely many $\alpha_i \in F \setminus \{0\}$, $u_i, v_i \in X^*$ and $r_i \in R$; note that u_i and v_i are words, while r_i's are linear combinations of words.

Note that we have $\overline{u_i r_i v_i} = u_i \bar{r}_i v_i$. Write $w := \max_i\{\overline{u_i r_i v_i}\}$ and define for convention

$$S := \{i \in I \mid w = \overline{u_i r_i v_i}\}.$$

If $\#S = 1$, then $\bar{f} = w$ is reducible and we are done.

If $\#S > 1$, it may occur that $\sum_{i \in S} \alpha_i w = 0$ so that $\bar{f} \neq w$. To resolve this problem we use induction on $(w, \#S) \in \overline{I(R)} \times \mathbb{N}^*$, where $\overline{I(R)}$ denotes the set of leading monomials of elements in $I(R)$ and $\overline{I(R)} \times \mathbb{N}^*$ is equipped with the lexicographical order that compares w firstly and then $\#S$, which satisfies the minimality condition.

Since for $\#S = 1$ the statement is true no matter what w is, the initial condition is satisfied and we can proceed by induction, supposing that $\#S > 1$ and that the statement is true for all pairs less than $(w, \#S)$.

Now that $\#S > 1$, so there exists $i \neq j$ with $u_i \bar{r}_i v_i = u_j \bar{r}_j v_j = w$. We have

$$\alpha_i u_i r_i v_i + \alpha_j u_j r_j v_j = (\alpha_i + \alpha_j) u_i r_i v_i + \alpha_j (u_j r_j v_j - u_i r_i v_i).$$

The left hand side attributes 2 elements to S. For the right hand side, $(\alpha_i + \alpha_j) u_i r_i v_i$ attributes 1 or 0 denpending on whether $\alpha_i + \alpha_j$ is zero. If we can show that $u_j r_j v_j - u_i r_i v_i$ is of the form $u_j r_j v_j - u_i r_i v_i = \sum_k \beta_k u'_k r'_k v'_k$ with $\max_k\{\overline{u'_k r'_k v'_k}\} < w$, then by replacing the left hand side by the right hand side in the summation $f = \sum_i \alpha_i u_i r_i v_i$, we obtain a new summation that expresses f where either $\#S$ is smaller (while w is fixed) or w is smaller (while whatever $\#S$ becomes), consequently we are done by the induction.

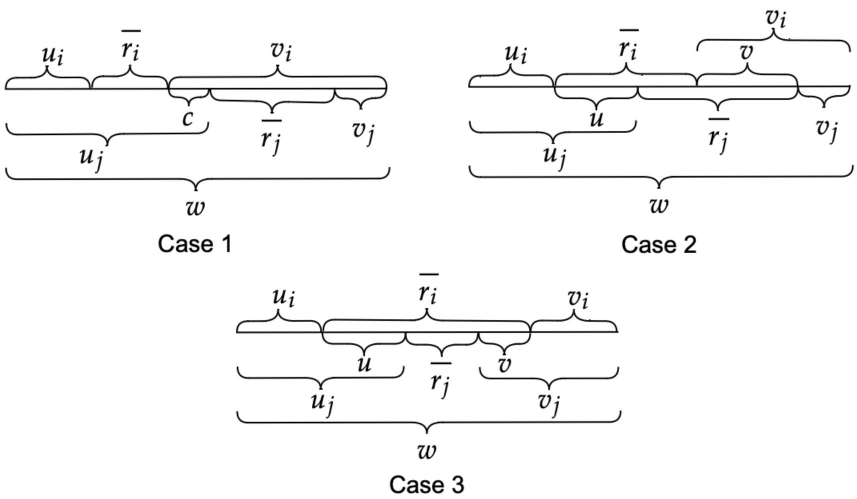
Therefore, for our purpose, there are now three cases to discuss:

1. $\bar{r}_i$ and $\bar{r}_j$ do not intersect in w, i.e. (up to a permutation) $w = u_i \bar{r}_i c \bar{r}_j v_j$ where $c \in X^*$, hence we have $v_i = c \bar{r}_j v_j$ and $u_j = u_i \bar{r}_i c$.

2. $\bar{r}_i$ and $\bar{r}_j$ intersect, but no one is contained by the other.

3. one of $\bar{r}_i$ and $\bar{r}_j$ is contained by the other.

Below gives the illustration of these three cases and the notations that we will use later in our proof.



Let us keep the notation that $r_i = \bar{r}_i - r_i'$ (and similarly for r_j) in the following.

For Case 1, we have

$$\begin{aligned}
u_j r_j v_j - u_i r_i v_i =& u_i \bar{r}_i c r_j v_j - u_i r_i c \bar{r}_j v_j \\
=& u_i((r_i + r_i')c r_j - r_i c(r_j + r_j'))v_j \\
=& u_i(r_i' c r_j - r_i c r_j')v_j = u_i r_i' c r_j v_j - u_i r_i c r_j' v_j.
\end{aligned}$$

Since $\max\{\overline{u_i r_i c r_j v_j}, \overline{u_i r_i c r_j' v_j}\} < \overline{u_i r_i c r_j v_j} = w$, we obtain the result as desired.

The other two cases will be discussed in the next lecture.

§6 Lecture 6

Now for Case 2, keeping the notation in the illustration above, we have

$$\begin{aligned}
u_j r_j v_j - u_i r_i v_i =& u_i u r_j v_j - u_i r_i v v_j \\
=& u_i(u r_j - r_i v)v_j \\
=& - u_i(r_i, r_j)_{u\bar{r}_j = \bar{r}_i v} v_j.
\end{aligned}$$

Recall that by remark 3.2, we have $(f, g)_w = \sum_k \alpha_k v_k r_k' u_k$ with $v_k \bar{r}_k u_k < w$ for all k for any $f, g \in R$, let $f = r_i$, $g = r_j$ and $w = u\bar{r}_j = \bar{r}_i v$ then we are done.

The argument for Case 3 goes similarly. □

Exercise 6.1

Finish the rest of the proof for Case 3.

Gröebuer-Shirshov Bases For Semigroups

Let us call a finite presentation $\langle X \mid R \rangle$ (either for semigroups or for algebras), where X is equipped with an order with the minimal condition, a *reduction system*.

A reduction system is *confluent*, if for every word in it, the result of reduction of the word is irrelevant to the choice of how the reduction is applied.

With this definition, the theorem which we just proved can be reformulated as that, for algebras, a reduction system is confluent if and only if R is closed under composition. The following lemma by Newman tells essentially the same thing for semigroups:

Lemma 6.1. Newman

For semigroups, a reduction system $S = \langle X \mid u_i = v_i \rangle$ with $u_i > v_i$ is confluent if and only if for any u_i and u_j such that $v' u_j = u_i v''$ for some $v', v'' \in X^*$, $v' v_j$ and $v_i v''$ has the same descendant*, i.e. they reduce to a same word after finitely many steps of reductions.*

Note that in a reduction system of a semigroup, every word can be represented by a irreducible word. The above gives a necessary and sufficient condition for all irreducible words to be different. If this condition is satisfied, then we call the irreducible words as *normal forms*, and we conclude that every word can be reduced to a unique normal form.

Further Applications of Gröebuer-Shirshov Bases

Let us give two general and important examples of algebras where the algorithm of reduction applies: graded algebras and commutative algebras.

Graded Algebras

Let $A = \oplus_{i=1}^{\infty} A_i$ be a graded algebra with $A_0 = F$ and $A = \langle A_1 \rangle$. Furthermore, we assume that $\dim_F A_i < \infty$ for all i, so that A is finitely generated as an algebra (by a basis of A_1).

Note that for homogeneous f and g, their composition (if exists) $(f,g)_w$ is also homogeneous and we always have $\deg(f,g)_w > \max\{\deg f, \deg g\}$, where $\deg 0$ is ∞ by convention.

Let $A = \langle X \mid R = 0 \rangle$ be a finite presentation where every relation in R is homogeneous. Define an order with the minimality condition on X so that we obtain a reduction system. Write $R_0 \coloneqq R$. Define inductively that

$$R_n \coloneqq R_{n-1} \cup \{(f,g)_w \mid f, g \in R_{n-1}\},$$

i.e., R_n is R_{n-1} union all possible compositions of elements in R_{n-1}. Since R_0 is finite, each R_n is also finite.

Write $R_\infty \coloneqq \bigcup_{i=0}^{\infty} R_i$, then $A = \langle X \mid R = 0 \rangle = \langle X \mid R_\infty = 0 \rangle$, since if $f, g \in R$, then $(f,g)_w = fv - ug \in I(R)$, where $I(R)$ is the ideal generated by R.

Now that R_∞ is closed under composition, so the Groöebuer-Shirshov bases theorem applies and we see that the set of irreducible words (with respect to R_∞) is a basis of A.

Although R_∞ might contain infinitely many elements, we still have an algorithm for reduction in this case: for any elemen a in A, there must exists $N \in \mathbb{N}$ such that every element in $R_\infty \setminus R_N$ is of degree strictly larger than the degree of any homogeneous component of a, because, by the construction, the minimal degree of elements in $R_\infty \setminus R_n$ strictly increases as n increases. Therefore to reduce a into a linear combination of irreducible words, we need only check the reduction relations in R_N, where there are only finitely many of them.

Commutative Algebras

Let A be a finitely generated commutative algebra, then we can find a surjective homomorphism

$$\varphi\colon F[x_1, \cdots, x_n] \twoheadrightarrow A,$$

so that $A \cong F[x_1, \cdots, x_n]/\ker\varphi$.

Recall the following lemma by Hilbert, which is a standard result in commutative algebra, c.f. Corollary 2.13 in [Kem11]:

Lemma 6.2. Hilbert

Every ideal of a polynomial ring $F[x_1, \cdots, x_n]$ is finitely generated.

Therefore every finitely generated commutative algebra admits a finite presentation, e.g. let $r_1, \cdots, r_m$ be a set of generators of $\ker\varphi$, then $A = \langle x_1, \cdots, x_n \mid r_i = 0 \rangle$.

Let us now define the composition of two elements in a polynomial ring $F[x_1, \cdots, x_n]$.

Without loss of generality, let the generators be ordered as $x_1 > \cdots > x_n$. To compare two monomials, we compare firstly their degrees, then the numbers of powers of x_1, and then the numbers of powers of x_2 and so on (i.e. from the largest generator to the smallest generator). Given a polynomial f, we define its *leading monomial* $\bar{f}$ as the largest monomial among all of its monomials.

Now for any two polynomials $f, g \in F[x_1, \cdots, x_n]$, we say that they are composable if their leading monomials have a non-constant common divisor, i.e. an element $d \in F[x_1, \cdots, x_n] \setminus F$ such that $\bar{f} = ad$ and $\bar{g} = bd$ for two elements $a, b \in F[x_1, \cdots, x_n]$. The composition of f and g with respect to this common divisor is thus defined as

$$(f,g)_{abd} \coloneqq bf - ag.$$

Similar to definition 3.1 and definition 3.2, reducible monomials and irreducible monomials are defined, and a similar argument shows that the Groöebuer-Shirshov bases theorem in this case is also true.

Now that given a commutative algebra A along with a finite presentation $A = \langle X \mid R = 0 \rangle$, we want to apply the Groöebuer-Shirshov bases theorem to it to obtain a basis of A along with an algorithm of reduction. Like what we did for graded algebras, we would like to consider $R_0 := R$ and then add compositions of elements in R_{n-1} to obtain R_n, while seeking for a way to obtain an algorithm. For this pupose, we consider the following proposition:

Proposition 6.3

Among every infinite set of monomials of finite many variables there exists two (distinct) monomials that one divides the other one.

Proof. Let us proceed by induction on the number of variables. The case where there is only one variable is trivial.

Suppose that the statement is true for $n-1$ variables. Let us identify monomials in n variables bijectively to elements in $\mathbb{N}^n$, hence we say that a monomial $(i_1, \cdots, i_n)$ divides another monomial $(j_1, \cdots, j_n)$ if and only if $i_k \leq j_k$ for all $k = 1, \cdots, n$. Suppose we have an infinite set of monomials S in which no monomial divides another one, then among all the corresponding tuples of the monomials we find a tuple whose first index is the smallest, say $(i'_1, \cdots, i'_n)$ with $i'_1 \leq i_1$ for all other tuples $(i_1, \cdots, i_n)$. For any other element $(i_1, \cdots, i_n) \in S$, there must be $i_k < i'_k$ for some $k = 2, \cdots, n$, hence if we define $S_k := \{(i_1, \cdots, i_n) \in S \mid i_k < i'_k\}$, then there must be

$$S = \{(i'_1, \cdots, i'_n)\} \cup \left(\bigcup_{k=2}^{n} S_k \right).$$

Therefore one of S_k's must be infinite (since S is). Up to a relabelling let us say that S_2 is infinite. Write $T_l := \{(i_1, \cdots, i_n) \in S \mid i_2 = l\}$, then we have

$$S_2 = \bigcup_{0 \leq l \leq i'_2 - 1} T_l.$$

Again, there must exist an $l \in \{0, \cdots, i'_2 - 1\}$ such that T_l is infinite. Now that the second index of elements in T_l is the constant l, hence there exists two n-variable monomials in T_l that one divides the other if and only if there exists two $(n-1)$-variable monomials in $T'_l := \{(i_1, i_3, \cdots, i_n) \mid (i_1, i_2, i_3, \cdots, i_n) \in T_l\}$ that one divides the other. By our induction assumption we see that there exists two monomials in $T_l \subset S$ that one divides another, contradicting the definition of S. $\square$

With this proposition, there must exist an $N \in \mathbb{N}$ such that the leading monomial of any element in $R_\infty \setminus R_N$ is divisible by the leading monomial of some element in R_N. Noticing that elements in $R_\infty \setminus R_N$ do not give any new reducible word other than those are given by R_N, we see that R_∞ and R_N define a same set of irreducible words. Therefore every element in A can be reduced to a linear combination of irreducible words using only the relations in R_N. By theorem 3.3, since R_∞ is closed under compositions, the set of irreducible words does give a basis of A. Since we need only the finitely many relations in R_N to operate the reduction, we obtain an algorithm. These fulfill our purpose completely.

Remark 6.1

This result for commutative algebra is called Buchberger's theorem, or Buchberger's algorithm.

§7 Lecture 7

Let us now look at the number of steps that we need to reduce a word in a reduction system of a semigroup.

Dehn function

In a reduction system (whether confluent or not), we say that two words are *equivalent* if they are *congruent*, i.e. they have the same descendent. This means that two congruent words can be transformed to each other by finitely many steps of substituting the relations, say.

$$u = w_1 \sim w_2 \sim \cdots \sim w_r = v.$$

Given two congruent words u and v, we denote by $\|u \times v\|$ the length of the smallest chain of substitution that we need to go through to transform u into v.

The *Dehn function* $D\colon \mathbb{N} \to \mathbb{R}_{>0}$ of a reduction system of a semigroup is now defined by

$$D(n) \coloneqq \max\{\|u \times v\| \mid u, v \text{ are equivalent with lengths no more than } n\}.$$

The maximum always exists since there are only finitely many u and v for a fixed n. Clearly, the Dehn function gives a measurement of the complexity of a system.

Given two functions $f, g\colon \mathbb{N} \to \mathbb{R}_{>0}$, we say that f is *asymptotically less or equal to* g, denoted as $f \preceq g$, if there exists $C \in \mathbb{N}$ such that

$$f(n) < Cg(Cn), \quad \forall n \in \mathbb{N}.$$

If $f \preceq g$ and $f \succeq g$, then we say that f and g are *asymptotically equivalent*, denoted by $f \sim g$.

Theorem 7.1. Dehn Functions Are Asymptotically Equivalent

Given two finite presentations of a semigroup, their corresponding Dehn functions are asymptotically equivalent.

Proof. Let $\langle X \mid R \rangle \cong \langle Y \mid R' \rangle$ be two finite presentations, then any generator $y_i \in Y$ can be written as a word $y_i(x)$ in X. Since there are only finitely many generators in Y, we can find an upper bound C of the lengths $\{\text{length}_X(y) \mid y \in Y\}$. For any two words $u = v$ with lengths less than n in Y, their lengths in X are thus less than Cn. Also, for each relation in R', it may be achieved by finitely many compositions of relations in R; since there are only finitely many relations, we may enlarge our C so that the number of needed compositions for each relation is always less than C. These give us

$$\|u \times v\|_Y \leq CD_X(Cn).$$

Take the maximum of the left hand side and then we obtain $D_Y(n) \leq CD_X(Cn)$. ☐

Therefore we may think Dehn function as an equivalence class of functions corresponding to a semigroup, regardless of the choice of presentations.

In a confluent reduction system, let us denote by $\gamma_{\min}(v)$ the minimum time of reduction that is needed to reduce v to its normal form and by $\gamma_{\max}(v)$ the maximum time of reduction. Note that by considering the time of reduction, we are requiring that each step gives a smaller word so that we cannot substitute a same relation back and forth, hence the maximum time is well-defined.

By linking two equivalent words with their normal form, which is a same irreducible word in a confluent reduction system, we see that

$$\|u \times v\| \leq \gamma_{\min}(u) + \gamma_{\min}(v).$$

Note that we have by definition that

$$\gamma_{\min}(u) = \|u \times \tilde{u}\|,$$

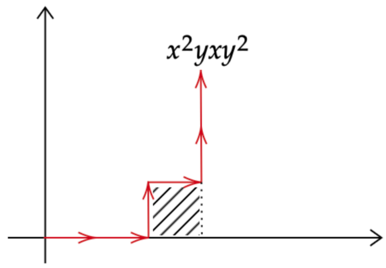
where $\tilde{u}$ denotes the normal form of u. Since the normal form of a word has length no longer than that word, we see that the Dehn function is asymptotically equivalent to the function $\gamma\colon \mathbb{N} \to \mathbb{R}_{>0}$ defined by

$$\gamma(n) \coloneqq \max\{\gamma_{\min}(u) \mid \text{length}(u) \leq n\}.$$

Let us take a look at some examples where we can compute the Dehn function.

Example 7.1

Consider $\langle x, y \mid xy = yx \rangle$ with $x < y$. Clearly the irredcubile words are words of the form $x^i y^j$. We can relate words in this system with 2-dimensional graphics: given any word, we start from the origin, and read each letter in the word from the left to the right. Each time we read x, we go right for a unit length, and each time we read y, we go up for a unit length. For example, the word x^2yxy^2 relates to the following graph:



Below the graph and the horizontal axis there is a region which is shadowed in the illustration. Let us call the area of the shadowed region the *area of the word*, say, the area of x^2yxy^2 is 1. We see that the irreducible words are exactly the words of 0 area, and each time we apply the reduction $yx \to xy$ to a word, its area goes down by 1. Therefore, in this case, $\gamma_{\max}(u) = \gamma_{\min}(u) = \mathrm{Area}(u)$. For a word of length n, its area is at maximum $\left(\frac{n}{2}\right)^2$, therefore we conclude that the Dehn function of this system is asymptotically equivalent to n^2.

Example 7.2

Consider $\langle x_1, \cdots, x_m \mid x_i x_j = x_j x_i, 1 \leq i, j \leq m \rangle$ with $x_1 < \cdots < x_m$. Write a word as $v = x_{i_1} \cdots x_{i_n}$, we can define its area as

$$\mathrm{Area}(v) := \#\{(k,l) \in \{1, \cdots, n\}^2 \mid x_{i_k} > x_{i_l},\ k < l\}.$$

For example, $\mathrm{Area}(x_3x_2x_2x_1) = 3+1+1 = 5$. Again, each reduction reduces the area by 1, hence the Dehn function is asymptotically equivalent to the maximum area. For a word of length n, its area is bounded above by n^2 by definition. Also, since $\mathrm{Area}(x_m^{n/2} x_1^{n/2}) = \left(\frac{n}{2}\right)^2$, the maximum area is bounded below by $\left(\frac{n}{2}\right)^2$. These conclude that the Dehn function of this system is asymptotically equivalent to n^2.

In general, if we can define the area of the words in a system in a way that the irreducible words are exactly those of area 0 and each step of reduction reduces an area that is bounded uniformly, i.e., irrelevant to whatever the word is, then we can estimate the Dehn function of the system. For example, if in a reduction system the reductions always reduce an area more than ε_1 and less than ε_2, i.e. for any word wuw' and relation $u = v$ with $u > v$ in that system, we have

$$\mathrm{Area}(wvw') + \varepsilon_1 < \mathrm{Area}(wuw') < \mathrm{Area}(wvw') + \varepsilon_2,$$

then the number K of required steps to reduce a word w is among $[\mathrm{Area}(w)/\varepsilon_2, \mathrm{Area}(w)/\varepsilon_1]$. Let $A(n) := \max_{\mathrm{length}(w) \leq n}\{\mathrm{Area}(w)\}$, then the Dehn function is estimated by

$$D(n) \in [A(n)/\varepsilon_2, A(n)/\varepsilon_1].$$

In particular, the Dehn function is asymptotically equivalent to the maximum-area function A.

Remark 7.1

Although one can define the Dehn function similarly for reduction systems of algebras in a naive way, the situation for algebras is more complicated and has been remained undecided. Say, let us define $D(n)$ as the maximum of the minimum number of steps of reduction for reducing an element in which each monomial has length no more than n. Since there are only finitely many generators, there are only finitely many monomials of length no more than n and finitely many ways to reduce even if we take the factorization into consideration, so the Dehn function is well-defined. However, when it comes to decide a way to reduce, the question becomes complicated: if we reduce the monomials term by term, then we may be too slow (at least m^n) compared to the real Dehn function; if we want to factorize firstly and then reduce, then how should we decide how to factorize? For example, the algebra $\langle x, y \mid x^2 = 0 \rangle$. It is very simple: once you see x^2 you kill that term. However, it may still cost about m^n steps to reduce if we look term by term, which means that the Dehn function defined above fails to measure the complexity of the system.

Maybe this can be solved when the quantum computer comes out, where we will be able to reduce all the monomials at the same time for one single step.

Now let us end this chapter and move on to talk about the free groups.

Free Groups

[Zelmanov introduced the definition, uniqueness and construction of free group which are not taken down here; for these things, see for example Chapter II, Section 5.1~5.3 of [Alu09].]

In short, let $X = \{x_i\}_{i\in I}$ be a set of generators and $Y = \{y_i\}_{i\in I}$ is another set of generators with the same index set I, then the following semigroup

$$\langle\, X, Y \mid x_i y_i = 1, y_i x_i = 1,\ i \in I\,\rangle,$$

is the free group generated by X, and it is a confluent system as one can verify using our previous theory. The normal forms in the above semigroup are called *reduced forms* of the free group generated by X. Explicitly, the reduced forms are

$$x_1^{\varepsilon_1} \cdots x_n^{\varepsilon_n},$$

where $\varepsilon_i = \pm 1$, $x_i \in X$ and we require that no subword is of the form xx^{-1} or $x^{-1}x$. We will keep this notation for reduced forms.

Let us denote the free group generated by X by $F(X)$. If $|X| = m < \infty$, then we may write $F(X) = F(m)$. It is obvious that the free groups are always isomorphic if their sets of generators have a same cardinality.

Question

For different cardinalities of the index, can the associated free groups be isomorphic? For example, is a free group generated by n elements isomorphic to a free group generated by m elements if $n \neq m$?

The answer is yes, for both finite and infinite cases. Here we only talk about the finite cases. We may answer this question firstly for the abelian case:

[Zelmanov introduced the definition, uniqueness and construction of free abelian group which are not taken down here; for these things, see for example Chapter II, Section 5.4 of [Alu09].]

Proposition 7.2

Two finitely generated free abelian groups are isomorphic if and only if they are freely generated by a same number of elements

Proof. Recall that two finitely generated free abelian groups must be of the form $\mathbb{Z}^n$ and $\mathbb{Z}^m$. If they are isomorphic, say $\varphi\colon \mathbb{Z}^n \to \mathbb{Z}^m$ is an isomorphism, then we have

$$\varphi^{-1}(2\mathbb{Z}^m) = 2\varphi^{-1}(\mathbb{Z}^m) = 2\mathbb{Z}^n.$$

Therefore $\mathbb{Z}^m/2\mathbb{Z}^m \cong \mathbb{Z}^n/\varphi^{-1}(2\mathbb{Z}^m) = \mathbb{Z}^n/2\mathbb{Z}^n$ by the first isomorphism theorem, see the diagram below.

$$\begin{tikzcd} \mathbb{Z}^n \arrow[r, "\varphi"] \arrow[d, two heads] & \mathbb{Z}^m \arrow[d, two heads] \\ \mathbb{Z}^n/2\mathbb{Z}^n = \mathbb{Z}^n/\varphi^{-1}(2\mathbb{Z}^m) \arrow[r, dashed, "\cong"'] & \mathbb{Z}^m/2\mathbb{Z}^m \end{tikzcd}$$

Since $\mathbb{Z}^n/2\mathbb{Z}^n \cong \bigoplus_{i=1}^n \mathbb{Z}/2\mathbb{Z}$, we thus obtain $2^m = 2^n$, concluding that there must be $m = n$. ☐

§8 Lecture 8

Let us recall some constructions in group theory.

- Given a group G, recall that the elements of the form $[a,b] = a^{-1}b^{-1}ab$ are called the commutators. Notice that $ab = ba[a,b]$, hence in some way the commutator measures how far the elements a and b are from being commutative. Note also that there is $[a,b]^{-1} = [b,a]$.
- Recall that $[G,G]$ the subgroup generated by all commutators in G is normal.
- Recall that $\mathrm{Aut}(G)$ is a group. For each element $x \in G$, we can define an automorphism $g \mapsto x^{-1}gx$ on G; automorphisms of such form are called *inner-automorphisms*. It is easy to verify that the set of inner-automorphisms forms a normal subgroup of $\mathrm{Aut}(G)$; let us denote it as $\mathrm{InAut}(G)$. Let us call the quotient group $\mathrm{Aut}(G)/\mathrm{InAut}(G)$ the group of outter-automorphisms, which we denote by $\mathrm{OutAut}(G)$.
- For an arbitrary normal group of G, the only thing we can say is that it is invariant under the inner-automorphisms, but not all automorphisms. However, the commutator subgroup $[G,G]$ is invariant under all automorphisms on G (just check the generators).
- Recall that for any normal subgroup H of G such that G/H is abelian, we have $[G,G] \subset H$. Hence we have that

$$[G,G] = \bigcap_{H \lhd G,\ G/H \text{ abelian}} H.$$

We are now ready to present the following lemma:

Lemma 8.1

$F(m)/[F(m),F(m)]$ is the free abelian group of rank m, i.e. the free abelian group generated by m elements.

Proof. Firstly we check that the quotient preserves that different x_i's in X to be distinct, i.e. we need to check that $x_i^{-1}x_j \notin [F(m),F(m)]$ for any $i \neq j$. Indeed, noticing that for any element in $[F(m),F(m)]$, the number of appearances of x_i must be equal to the number of appearances of x_i^{-1}. Hence we are done for this part.

The rest of the proof is a straightforward verification of the fact that $F(m)/[F(m),F(m)]$ satisfies the universal property of the free abelian groups. $\square$

The result that $F(m) \cong F(n)$ if and only if $m = n$ thus follows from the above lemma, the proof of proposition 7.2 and the proposition itself.

Schreier Theorem

We now proceed to another result, that all subgroups of free groups are free.

Recall that for any subgroup H, we have $G = \bigsqcup_i Hg_i$. Two elements $x, y \in G$ lies in a same coset Hg_i if and only if $xy^{-1} \in H$. In each coset Hg_i, let us select one representative with only one restriction that for H we choose the identity 1; for any element $g \in G$, we denote by $\bar{g}$ the representative that we chose in the coset Hg. Let S be the set of all representatives.

Remark 8.1

It is easy to verify that $\overline{\bar{a}b} = \overline{ab}$ for any $a, b \in G$.

Let X generate G and consider the elements of the form $sx^{\varepsilon}(\overline{sx^{\varepsilon}})^{-1}$ where $s \in S$ and $x \in X$. Note that $sx^{\varepsilon}(\overline{sx^{\varepsilon}})^{-1} \in H$.

Lemma 8.2

The set $\{sx(\overline{sx})^{-1} \mid s \in S, x \in X\}$ generates H.

Proof. Notice that we have that, since $s = \overline{\overline{sx^{-1}}x}$ (using remark 8.1),

$$\left(sx^{-1}\left(\overline{sx^{-1}}\right)^{-1}\right) \cdot \left(\overline{sx^{-1}}x\left(\overline{\overline{sx^{-1}}x}\right)^{-1}\right) = 1.$$

It follows that $sx^{-1}\left(\overline{sx^{-1}}\right)^{-1} \in \{sx(\overline{sx})^{-1} \mid s \in S, x \in X\}$.

Also, inspired by the above equality, we have the following algorithm: For any element $h \in H$, it has reduced form $h = x_{i_1}^{\varepsilon_1} \cdots x_{i_n}^{\varepsilon_n}$. Since $\overline{h} = 1$, we have

$$\begin{aligned} h =& x_{i_1}^{\varepsilon_1} \cdots x_{i_n}^{\varepsilon_n} \\ =& \left(1 \cdot x_1^{\varepsilon_1}\left(\overline{1 \cdot x_1^{\varepsilon_1}}\right)^{-1}\right)\left(\overline{x_{i_1}^{\varepsilon_1}} \cdot x_{i_2}^{\varepsilon_2}\left(\overline{x_{i_1}^{\varepsilon_1} \cdot x_{i_2}^{\varepsilon_2}}\right)^{-1}\right) \cdots \left(\overline{x_{i_1}^{\varepsilon_1} \cdots x_{i_{n-1}}^{\varepsilon_{n-1}}} \cdot x_{i_n}^{\varepsilon_n}\left(\overline{x_{i_1}^{\varepsilon_1} \cdots x_{i_n}^{\varepsilon_n}}\right)^{-1}\right), \end{aligned}$$

as desired. $\square$

Corollary 8.2.1

Let G be finitely generated group. Let $H < G$ be a subgroup of G with $|G : H| < \infty$. Then the group H is finitely generated.

Proof. Since $|G : H| < \infty$, S is finite. Since G is finitely generated, X is finite. Therefore the generating set of H given by lemma 8.2 is also finite. $\square$

More precisely, if $|X| = m$ and $|G : H| = n$, then H is generated by no more than mn elements.

We now move back to discussions about free groups. In the following, let H be a subgroup of $F(m)$.

Definition 8.1. Schreier System

We say that a set of representatives S of cosets of H in $F(m)$ is a Schreier system if for any $s \in S$, the reduced form $s = x_{i_1}^{\varepsilon_1} \cdots x_{i_k}^{\varepsilon_k}$ satisfies that $x_{i_1}^{\varepsilon_1} \cdots x_{i_j}^{\varepsilon_j} \in S$ for each $j = 1, \cdots, k$.

Lemma 8.3. Existence of Schreier System

For any subgroup $H < F(X)$, a Schreier system exists.

Proof. For any $g \in F(X)$ with reduced form $g = x_{i_1}^{\varepsilon_1} \cdots x_{i_k}^{\varepsilon_k}$, we say that k is the length of g. The length of the identity is defined to be 0. Given a coset $C = Hg$, the length of C is defined as $\min\{\text{length}(a) \mid a \in C\}$. We now proceed by induction on lengths.

Note that $\text{length}(H) = 0$, and H is the only coset with length 0. Suppose that for every coset C of length less than n, we have selected a representative $\bar{C}$ such that $\text{length}(\bar{C}) = \text{length}(C)$ and $\bar{C}$ satisfies the Schreier condition, i.e. if $\bar{C} = x_{i_1}^{\varepsilon_1} \cdots x_{i_k}^{\varepsilon_k}$ then $x_{i_1}^{\varepsilon_1} \cdots x_{i_j}^{\varepsilon_j} \in \overline{Hx_{i_1}^{\varepsilon_1} \cdots x_{i_j}^{\varepsilon_j}}$. Note that $\text{length}(Hx_{i_1}^{\varepsilon_1} \cdots x_{i_j}^{\varepsilon_j}) = j$, otherwise $\bar{C}$ can be replaced by a shorter word. For a coset C of length n, there exists some reduce form of length n, $x_{i_1}^{\varepsilon_1} \cdots x_{i_{n-1}}^{\varepsilon_{i_{n-1}}} x_{i_n}^{\varepsilon_n} \in C$, and we define

$$\bar{C} := \overline{x_{i_1}^{\varepsilon_1} \cdots x_{i_{n-1}}^{\varepsilon_{i_{n-1}}}} x_{i_n}^{\varepsilon_n},$$

whose reduced form is of length n and satisfies the Schreier condition as desired. $\square$

Theorem 8.4. Shreiez

Let $H < F(X)$ and S be a Schreier system corresponding to H. Then $\{sx(\overline{sx})^{-1} \neq 1 \mid s \in S,\ x \in X\}$ is a set of free generators of H.

Proof. We show firstly that if a reduction happens when we put sx^ε and $(\overline{sx^\varepsilon})^{-1}$ together for some $s \in S$, $x \in X$ and $\varepsilon = \pm 1$, then there must be $sx^\varepsilon(\overline{sx^\varepsilon})^{-1} = 1$. Let the reduced form of s be $s = x_{i_1}^{\varepsilon_1} \cdots x_{i_n}^{\varepsilon_n}$ and that of $\overline{sx^\varepsilon}$ be $\overline{sx^\varepsilon} = x_{j_1}^{\delta_1} \cdots x_{j_m}^{\delta_m}$, then

$$sx^\varepsilon(\overline{sx^\varepsilon})^{-1} = x_{i_1}^{\varepsilon_1} \cdots x_{i_n}^{\varepsilon_n} x^\varepsilon x_{j_m}^{-\delta_m} \cdots x_{j_1}^{-\delta_1},$$

hence a reduction happens if either $x_{i_n}^{\varepsilon_n} x^{\varepsilon} = 1$ or $x^{\varepsilon} x_{j_m}^{-\delta_m} = 1$. If $x_{i_n}^{\varepsilon_n} x^{\varepsilon} = 1$, then $sx^{\varepsilon} = x_{i_1}^{\varepsilon_1} \cdots x_{n-1}^{\varepsilon_{n-1}} \in S$ by the Schreier condition, hence $sx^{\varepsilon} = \overline{sx^{\varepsilon}}$, consequently $sx^{\varepsilon}(\overline{sx^{\varepsilon}})^{-1} = 1$. If $x^{\varepsilon} x_{j_m}^{-\delta_m} = 1$, then $\overline{sx^{\varepsilon}} = x_{j_1}^{\delta_1} \cdots x_{j_{m-1}}^{\delta_{m-1}} x^{\varepsilon}$. Write $s_1 := x_{j_1}^{\delta_1} \cdots x_{j_{m-1}}^{\delta_{m-1}}$, then $s_1 \in S$ by the Schreier condition, and we have $sx^{\varepsilon}(\overline{sx^{\varepsilon}})^{-1} = ss_1^{-1}$. Therefore it suffices to show that $ss_1^{-1} \in H$, so that s and s_1 live in a same coset and thus $s = s_1$. Indeed, we have

$$H \ni sx^{\varepsilon}(\overline{sx^{\varepsilon}})^{-1} = sx^{\varepsilon} x^{-\varepsilon} s_1^{-1} = ss_1^{-1},$$

done.

The rest of the proof will be given in the next lecture.

§9 Lecture 9

Continue of the Proof of Theorem 8.4. Consider a nontrivial reduced product in $\{sx(\overline{sx})^{-1} \neq 1 \mid s \in S,\ x \in X\}$,

$$s_1 x_1^{\varepsilon_1} \left(\overline{s_1 x_1^{\varepsilon_1}}\right)^{-1} \cdots s_k x_k^{\varepsilon_k} \left(\overline{s_k x_k^{\varepsilon_k}}\right)^{-1},$$

where by reduced we mean that no two adjacent $s_i x_i^{\varepsilon_i} \left(\overline{s_i x_i^{\varepsilon_i}}\right)^{-1}$ cancel. It suffices to show that cancellations won't touch $x_i^{\varepsilon_i}$'s, so that no nontrivial reduced product is equal to 1, which implies the freeness.

By what we have shown in the beginning of the proof, no cancellation happens in the middle of $s_i x_i$ nor $x_i \left(\overline{s_i x_i^{\varepsilon_i}}\right)^{-1}$, the only chance for anything to get cancelled lies in the middle of $\left(\overline{s_i x_i^{\varepsilon_i}}\right)^{-1} s_{i+1}$. We will check that such cancellation will not kill any of $x_i^{\varepsilon_i}$ and $x_{i+1}^{\varepsilon_{i+1}}$, which finishes the proof. It suffices to do this for $i = 1$ for simplicity of notation.

Let us check $x_1^{\varepsilon_1}$ firstly, assuming that $x_2^{\varepsilon_2}$ is not killed. the only possibility is that s_2 has reduced form $s_2 = \left(\overline{s_1 x_1^{\varepsilon_1}}\right) x_1^{-\varepsilon_1} \cdots$. By the Schreier condition, $\left(\overline{s_1 x_1^{\varepsilon_1}}\right) x_1^{-\varepsilon} \in S$. Since

$$\overline{\left(\overline{s_1 x_1^{\varepsilon_1}}\right) x_1^{-\varepsilon_1}} = \overline{s_1 x_1^{\varepsilon_1} x^{-\varepsilon_1}} = \overline{s_1} = s_1,$$

we obtain that $\left(\overline{s_1 x_1^{\varepsilon_1}}\right) x_1^{-\varepsilon} = s_1$. Therefore $s_1 x_1^{\varepsilon_1} \left(\overline{s_1 x_1^{\varepsilon_1}}\right)^{-1} = 1$, contradiction.

Then let us check x_2^{ε}, assuming that $x_1^{\varepsilon_1}$ is not killed. Again, the only chance is that $\left(\overline{s_1 x_1^{\varepsilon_1}}\right)^{-1} = \cdots x_2^{-\varepsilon_2} s_2^{-1}$, hence $\overline{s_1 x_1^{\varepsilon_1}} = s_2 x_2^{\varepsilon_2} \cdots$. By the Schreier condition, this means that $s_2 x_2^{\varepsilon_2} \in S$, therefore $s_2 x_2^{\varepsilon_2} \left(\overline{s_2 x_2^{\varepsilon_2}}\right)^{-1} = 1$, contradiction.

Finally, let us check that $x_1^{\varepsilon_1}$ and $x_2^{\varepsilon_2}$ cannot be killed simultaneously. If they are killed simultaneously, then $x_1^{\varepsilon_1} \left(\overline{s_1 x_1^{\varepsilon_1}}\right)^{-1} s_2 x_2^{\varepsilon_2} = 1$. Hence $s_2 x_2^{\varepsilon_2} = \overline{s_1 x_1^{\varepsilon_1}} x_1^{-\varepsilon_1}$. Therefore

$$s_1 \left(\overline{s_2 x_2^{\varepsilon_2}}\right)^{-1} = s_1 \left(\overline{\left(\overline{s_1 x_1^{\varepsilon_1}}\right) x_1^{-\varepsilon_1}}\right)^{-1} = s_1 s_1^{-1} = 1,$$

contradicting that the product is reduced. $\square$

We now obtain a set of free generators of H along with an algorithm, but the following question remains:

Question

What is the cardinality of $\{sx(\overline{sx})^{-1} \neq 1 \mid s \in S,\ x \in X\}$? That is to say, how many $sx\left(\overline{sx}\right)^{-1}$ is equal to 1?

Suppose that $|X| = m$. Assuming the Schreier condition, the answer to the latter question is $n - 1$, so that H is freely generated by $mn - n + 1$ elements. (Recall that n is the cardinality of S.)

In fact, if $sx\left(\overline{sx}\right)^{-1} = 1$, then a cancellation must happen in the middle of either sx or $x\left(\overline{sx}\right)^{-1}$. Conversely, we have shown that if a cancellation happens then $sx\left(\overline{sx}\right)^{-1} = 1$. For each nontrivial element s in S, its reduced form is $s = x_1^{\varepsilon_1} \cdots x_k^{\varepsilon_k}$. If $\varepsilon_k = 1$, then $\left(\overline{sx_k^{-1}}\right) x_k s^{-1} = 1$. If $\varepsilon_k = -1$, then

$sx_k\left(\overline{sx_k}\right)^{-1}=1$. By a simple argument using the Schreier condition, one sees that these two cases do not give repeated pairs $sx\left(\overline{sx}\right)^{-1} \leftrightsquigarrow (s,x)\in S\times X$. Therefore among the pairs $(s,x)\in S\times X$, there are exactly $n-1$ pairs that lead to $sx\left(\overline{sx}\right)^{-1}=1$.

A consequence of this result is that subgroups of a free group may have larger rank than the original free group. In fact, a subgroup of a finitely generated free group may have rank infinity, see the second set of exercises.

Note that this algorithm for computing generators of subgroups of free groups can be applied to any (finitely generated) group: by sending the free generators to generators of the group, we obtain an epimorphism $F(m)\twoheadrightarrow G$. For any subgroup $H<G$, the preimage of H gives a subgroup of $F(m)$. Apply the algorithm to that subgroup of $F(m)$ to compute its generators, and then bring the generators to G, and we obtain generators of H.

Cayley Graph of Group

Consider a group $G=\langle a_1,\cdots,a_m\rangle=\langle X\rangle$, the *cayley graph* of G with respect to generators a_i's , denoted as $\mathrm{Cay}(G,X)$, is constructed by the following:

> For each element in G we assign a vertice, thus vertices in the graph are identified with elements in G. For each pair of vertices g and a_ig, we connect them with an edge.

In $\mathrm{Cay}(G,X)$, every element $g=a_1^{\varepsilon_1}\cdots a_k^{\varepsilon_k}$ is connected to the identity via

$$g-(a_2^{\varepsilon_2}\cdots a_k^{\varepsilon_k})-\cdots-a_k^{\varepsilon_k}-1.$$

Therefore the cayley graph is always connected.

By claiming that each edge has length 1 and that the distance between any two vertices is the minimal length that one has to go through the edges, the cayley graph becomes a metric space. The space can be also assigned a norm of length, by $\mathrm{length}(g)=d(g,1)$.

A *cycle* in a graph is a loop without self-intersection. A graph without any cycle is called a *tree.* It is easy to see the following characterization of freeness:

Proposition 9.1

$Cay(G,X)$ is a tree if and only if G is a free group on free generators X.

Proof. It is easy to observe that a cycle exists if and only if a reduced word is trivial. □

For any element $a\in G$, we can consider its action on the cayley graph M of G, by sending a vertice g to ga, and edges $g-a_ig$ to $ga-a_iga$ accordingly. Let us denote this action by R_a, then clearly R_a preserves the distance, hence it is an isometry on M. Also, we have $R_aR_b=R_{ba}$. Therefore the map $a\mapsto R_a$ gives an embedding $G^{op}\hookrightarrow \mathrm{Isom}(M)$. Note that every isometry on a graph preserves vertices, because the vertices are exactly the points with integer lengths.

In general, we can define group actions on any metric space M: an action of G on M is a group homomorphism $G\to\mathrm{Isom}(M)$.

With such action, we have another perspective of freeness. Recall that a *fixed-point free action* is an action where the identity is the only element whose action has a fixed point.

Theorem 9.2. Serre

G is free if and only if there is a fixed-point free action of G on a tree.

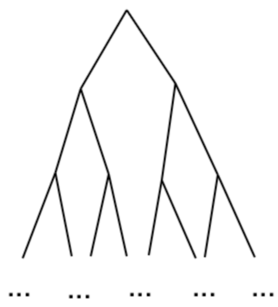
Note that we do not require the tree in the theorem to be a cayley graph of some group. It can be an arbitrary graph that is a tree.

If G acts freely on a tree, then any of its subgroup also acts freely on that tree. Therefore we obtain again that every subgroup of a free group is free, from this perspective.

Below gives an example of a group acting on a tree:

Example 9.1. Rooted Tree

Consider the *rooted tree*, which is illustrated below:



The dots indicate that the tree grows down indefinitely. The set of isometries on this tree is the set of reflections of its branches (let the length of the top point be zero and check the length of each vertice). The group of isometries naturally gives a group that acts on this tree. Every isometry must preserve the top point since it is the only point jointed with only two edges, hence no subgroup of this group of isometries is free.

We will talk about the rooted tree in details in a future lecture.

§10 Lecture 10

Let us talk about a little bit more about the word problem before proceeding.

Let us consider finitely generated groups. Say $G \cong F(m)/N$, $R \subset N$ generates N so that $G = \langle X \mid R = 1 \rangle$. The problem to determine whether two different words in $F(m)$ are equal when brought to G by quotienting N is the famous *word problem*. The research into the word problem significantly contributed to the development of computer science, in the sense that it made it clear what an algorithm is. It was proved by P. Novikov (1959) that there exists a finitely presented group for which no computer can ever exist that can decide whether an arbitrary word is equal to 1.

Still, we can have some discussion about this. Let us focus on the reduction of elements in N. Note that we have

$$N = \{(\tau_1^{g_1})^{\pm 1} \cdots (\tau_k^{g_k})^{\pm 1} \mid \tau_i \in R, g_i \in F(m)\},$$

where $\tau^g := g^{-1}\tau g$, the conjugation. Note that the conjugation indeed follows the rule of exponentiation: $(\tau^{g_1})^{g_2} = g_2^{-1}g_1^{-1}\tau g_1 g_2 = \tau^{g_1 g_2}$.

Recall the definition of Dehn function for semigroups in section 7. For groups, we can also define the Dehn function: for any $h \in N$, let $\|h\|$ denote the minimal possible k such that $h = (\tau_1^{g_1})^{\pm 1} \cdots (\tau_k^{g_k})^{\pm 1}$. Then

$$D(n) := \max\{\|h\| \mid h \in N, h \in B(n)\},$$

where $B(n)$ is the ball of radius n with center at 1, i.e. $h \in B(n)$ means that h is a word of length no more than n.

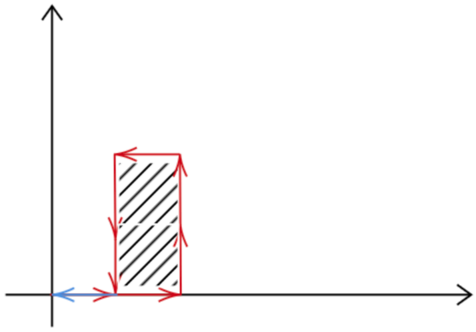
Exercise 10.1

Prove that the Dehn functions for different finite systems of a same object are asymptotically equivalent.

In fact, noticing that a step of reduction $\tau \to 1$ is the same as a step toward writing h as the form $h = (\tau_1^{g_1})^{\pm 1} \cdots (\tau_k^{g_k})^{\pm 1}$, one sees this immediately.

Example 10.1

Let us consider the group $\langle x, y \mid x^{-1}y^{-1}xy = 1 \rangle$. Like what we did in example 7.1, for each word, we start from the origin on the two-dimensional plane and look from left to the right. For each x, we go right by a unit length, for x^{-1} we go left, y we go up and for y^{-1} we go down. Thus any $x^{-1}y^{-1}xy$ would give a unit square. For example, the word $x^2y^2x^{-1}y^{-2}x^{-1}$ corresponds to the following graph:



So let us set $\mathrm{Area}(x^2y^2x^{-1}y^{-2}x^{-1}) := 2$. For this particular example, each replacement $yx^{-1}y^{-1} \to x^{-1}$ reduces the area by 1, hence it takes two steps to reduce the word $x^2y^2x^{-1}y^{-2}x^{-1}$ to its normal form. It follows that the Dehn function is asymptotically no less than $O(n^2)$, because it will take $n^2/16$ steps (cancellation of inverse elements is not counted) to reduce the square given by $x^{n/4}y^{n/4}x^{-n/4}y^{-n/4}$. Conversely, no element of length n can encircle a larger area than this square. Therefore the Dehn function is asymptotically equivalent to $O(n^2)$.
More precisely, if one wishes to stick to the definition, then the procedure is translated as the following:

$$\begin{aligned} x^2y^2x^{-1}y^{-2}x^{-1} =& x^2y^2(x^{-1}y^{-1}xy)y^{-1}x^{-1}y^{-1}x^{-1} \\ =& x^2y^2(x^{-1}y^{-1}xy)y^{-1}(x^{-1}y^{-1}xy)y^{-1}x^{-1}x^{-1} \\ =& (x^{-1}y^{-1}xy)^{y^{-2}x^{-2}}(x^{-1}y^{-1}xy)^{y^{-1}x^{-2}}. \end{aligned}$$

Free Products

Free Products of Algebras

Let F be a field. Let A and B be two F-algebras (with identity, and not necessarily commutative). The *free product* of A and B is the coproduct of A and B in the category of F-algebras, i.e. it is an F-algebra U along with two homomorphisms $\iota_A\colon A \to U$ and $\iota_B\colon B \to U$ that satisfies the universal property that for any other F-algebra C along with two homomorphisms $\varphi_A\colon A \to C$ and $\varphi_B\colon B \to C$, there exists a unique homomorphism $\varphi_U\colon U \to C$ such that $\varphi_U \circ \iota_A = \varphi_A$ and $\varphi_U \circ \iota_B = \varphi_B$. See the diagram below.

$$\begin{tikzcd} A \arrow[dr, "\iota_A"] \arrow[ddr, bend right, "\varphi_A"'] & & B \arrow[dl, "\iota_B"'] \arrow[ddl, bend left, "\varphi_B"] \\ & U \arrow[d, dashed, "\varphi_U"] & \\ & C & \end{tikzcd}$$

Assume for now that the free product always exists. Similar as we did for the universal enveloping of Lie algebras, let U denote the free product of A and B with $\iota_A\colon A \to U$ and $\iota_B\colon B \to U$, then U is generated by the images of A and B, i.e. $U = \langle\, \iota_A(A), \iota_B(B)\,\rangle$.

The uniqueness of free product is, again, permitted by the universal property. We now prove that it always exists:

Proposition 10.1

For any two F-algebras A and B, their free product exists.

Proof. Write $A = \langle\, X \mid R_A(X) = 0\,\rangle$ and $B = \langle\, Y \mid R_B(Y) = 0\,\rangle$. Define

$$U = \langle\, X \sqcup Y \mid R_A(X) = 0, R_B(Y) = 0\,\rangle,$$

then U is the free product of A and B with the obvious ι_A and ι_B, as one can check. $\square$

We can equip U with a more concrete system. Let $\{1, a_i \mid i \in I\}$ be a basis of A and $X = \{x_i\}_{i \in I}$. We have for any $i, j \in I$,

$$a_i a_j = \gamma_{ij}^0 1 + \sum_k \gamma_{ij}^k a_k,$$

for some $\gamma_{ij}^k \in F$. Then the set $R_A(X) = \{x_i x_j - \gamma_{ij}^0 1 - \sum_k \gamma_{ij}^k x_k \mid i, j \in I\}$ is closed under composition: we have

$$\begin{aligned} 0 &= \left(a_i a_j - \gamma_{ij}^0 - \sum_k \gamma_{ij}^k a_k \right) a_l - a_i \left(a_j a_l - \gamma_{jl}^0 - \sum_k \gamma_{jl}^k a_k \right) \\ &= \cdots \text{(progress of reduction)} = \text{(linear combination of } a_k\text{'s)} \end{aligned}$$

The basis condition then forces the coefficients to be all zeros. Replace the a's by x's and we see that $R_A(X)$ is closed under composition. Similarly $R_B(Y)$ is defined. Since no relation from $R_A(X)$ admit any composition with relations in $R_B(Y)$, $R_A(X)$ and $R_B(Y)$, put together, is still closed under composition.

Therefore the map $\iota_A \colon A \to U$, sending nontrivial linear compositions of $\{1, a_i\}$ to nontrivial linear compositions of $\{1, x_i\}$ which are irreducible in $\langle X \sqcup Y \mid R_A(X), R_B(Y) \rangle$, is injective. The same is true for ι_B. We thus obtain the following lemma.

Lemma 10.2

ι_A and ι_B are embeddings.

Notation 10.1

Let $A * B$ denote the free product of A and B.

With the above system, we have shown that the irreducible words of $A * B$ form a basis, which are exactly

$$\{1, c_1 \cdots c_k \mid (c_l \in \{a_i\} \text{ and } c_{l+1} \in \{b_j\}) \text{ or } (c_l \in \{b_j\} \text{ and } c_{l+1} \in \{a_i\}) \text{ for each } l = 1, \cdots, k-1\},$$

where $\{1, a_i\}$ and $\{1, b_j\}$ are the chosen basis of A and B respectively.

The free product for an arbitrary (set-valued) collection of F-algebras is defined similarly; it is just the coproduct in the categorical viewpoint. All constructions and arguments above pass immediately, giving exactly the same results. Also, we have the law of associativity, i.e. $(A * B) * C \cong A * (B * C)$; indeed, this law would follow immediately from a diagram chasing.

Free Products of Groups

Similarly, the free product of groups are defined as the coproduct in the category of groups. All arguments in the previous section pass to groups easily (including the presentation for the free product) except for the concrete system given after proposition 10.1. But we can still find a such system by bringing the question back to algebras:

Let F be a field. For any groups G_1, G_2 and G, group homomrphisms $\varphi \colon G_1 \to G$ and $\psi \colon G_2 \to G$, we have homomorphisms of F-algebras

$$F\varphi \colon FG_1 \to FG, \quad F\psi \colon FG_2 \to FG.$$

The free product of F-algebras gives us a unique homomorphism $F\varphi * F\psi \colon FG_1 * FG_2 \to FG$ such that the diagram commutes. Since G_1 and G_2 are basis of FG_1 and FG_2 respectively, our previous result tells that the set

$$\{1, c_1 \cdots c_k \mid (c_l \in G_1 \setminus \{1\},\ c_{l+1} \in G_2 \setminus \{1\}) \text{ or } (c_l \in G_2 \setminus \{1\},\ c_{l+1} \in G_1 \setminus \{1\}),\ l = 1, \cdots, k-1\},$$

forms a basis of $FG_1 * FG_2$. Noticing that the set along with the multiplication is a group, the claim is that this group (with the obvious embeddings of G_1 and G_2) is exactly the free product of G_1 and G_2; let us denote it by $G_1 * G_2$. Indeed, the uniqueness of the morphism that makes the coproduct diagram commutes is permitted by the above description of $G_1 * G_2$. With natural inclusions $G \subset FG$, the existence follows from the fact that $(Ff)|_G = f$ for any group homomorphism f whose domain is

G: restrict $F\varphi * F\psi$ to $G_1 * G_2$ and we see by the (apple-looking) commutative diagram below that the image of the restricted map should live in G.

$$\begin{tikzcd}
G_1 \arrow[r, dashed, bend left] \arrow[d, hook] \arrow[dddr, bend right=60, "\varphi"'] & G_1 * G_2 \arrow[d, hook] \arrow[ddd, dashed, bend right, "(F\varphi * F\psi)|_{G_1 * G_2}"'] & G_2 \arrow[l, dashed, bend right] \arrow[d, hook] \arrow[dddl, bend left=60, "\psi"] \\
FG_1 \arrow[r, hook] \arrow[dr, bend right, "F\varphi"'] & FG_1 * FG_2 \arrow[d, dashed, "F\varphi * F\psi"] & FG_2 \arrow[l, hook'] \arrow[dl, bend left, "F\psi"] \\
& FG & \\
& G \arrow[u, hook] &
\end{tikzcd}$$

Again, the free product for an arbitrary (set-valued) collection of groups is defined similarly, and all results pass over.

Example 10.2

We have immediately that $F(m) = \langle\, a_1 \,\rangle * \cdots * \langle\, a_m \,\rangle$, where each $\langle\, a_i \,\rangle$ is the infinite cyclic group generated by a_i.

§11 Lecture 11

Ping-Pong Lemma

From what we have got, we can easily see that if $G = \langle\, G_1, G_2 \,\rangle$ and every element of G can be written uniquely as an interchanging product of nonidentical elements[1] from G_1 and G_2, then $G \cong G_1 * G_2$ canonically. This observation leads to the following lemma:

Lemma 11.1. Ping-Pong

Consider a group G acting on a set X. Let G_1 and G_2 be two different subgroups with $|G_1| \geq 3$ and $|G_2| \geq 2$. Let X_1 and X_2 be two disjoint subsets of X such that

$$(G_1 \setminus \{1\})X_1 \subset X_2, \quad (G_2 \setminus \{1\})X_2 \subset X_1,$$

then $\langle\, G_1, G_2 \,\rangle \cong G_1 * G_2$.

Proof. It suffices to show that any interchanging product of nonidentical elements in G_1 and G_2 is not equal to 1. Let a denote elements in G_1 and b denotes elements in G_2, then the interchanging products can be divided into four cases:

Case 1. $a_1 b_1 a_2 b_2 \cdots a_{n-1} b_{n-1} a_n$. Since $a_1 b_1 a_2 b_2 \cdots a_{n-1} b_{n-1} a_n X_1 \subset X_2$, we are done.

Case 2. $b_1 a_1 b_2 a_2 \cdots b_{n-1} a_{n-1} b_n$. Since $b_1 a_1 b_2 a_2 \cdots b_{n-1} a_{n-1} b_n X_2 \subset X_1$, we are done.

Case 3. $a_1 b_1 a_2 b_2 \cdots a_n b_n$. Since $|G_1| \geq 3$, there exists $a \in G_1$ such that $a \neq 1$ and $a \neq a_1$. If $a_1 b_1 a_2 b_2 \cdots a_n b_n = 1$, then its conjugation by a is also equal to 1, i.e. $a^{-1} a_1 b_1 a_2 b_2 \cdots a_n b_n a = 1$. However, this cannot be true because the conjugation is in the form of Case 1.

Case 4. $b_1 a_1 \cdots b_n a_n$. Conjugate by an element $a \in G_1 \setminus \{1, a_n\}$ and we are back in Case 1. $\square$

As an important application of Ping-Pong Lemma, let us consider $SL(n, \mathbb{Z}) = \{A \in M_n(\mathbb{Z}) \mid \det(A) = 1\}$. It is in fact a group, because for any invertible matrix $A = (a_{ij})$, we have the formula $A^{-1} = \frac{1}{\det(A)}((-1)^{ij} \det(A_{ij}))^T$, which is in $SL(n, \mathbb{Z})$ provided that $a_{ij} \in \mathbb{Z}$ and $\det(A) = 1$.

Theorem 11.2

We have an embedding $F(2) \hookrightarrow SL(2, \mathbb{Z})$.

Such embedding is not canonical, though.

[1] i.e., elements that are not equal to 1.

Proof. Consider the action of $SL(2,\mathbb{Z})$ on $\mathbb{C}^2$ by matrix multiplication. The subgroups $G_1 := \left\langle \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \right\rangle = \left\{ \begin{pmatrix} 1 & 2n \\ 0 & 1 \end{pmatrix} \mid n \in \mathbb{Z} \right\}$ and $G_2 := \left\langle \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \right\rangle = \left\{ \begin{pmatrix} 1 & 0 \\ 2n & 1 \end{pmatrix} \mid n \in \mathbb{Z} \right\}$ are both cyclic. By example 10.2, we have $F(2) = G_1 * G_2$, hence it suffices to show that $G_1 * G_2 \cong \langle G_1, G_2 \rangle \subset SL(n,\mathbb{Z})$, which will be done using the Ping-Pong Lemma.

Indeed, consider the subsets $X_1 = \{(z_1,z_2)^T \in \mathbb{C} \mid |z_2| > |z_1|\}$ and $X_2 = \{(z_1,z_2)^T \in \mathbb{C} \mid |z_1| > |z_2|\}$. Then $\begin{pmatrix} 1 & 2n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} z_1 + 2nz_2 \\ z_2 \end{pmatrix}$. Since

$$|z_1 + 2nz_2| \geq 2|n||z_2| - |z_1| > |z_2|,$$

provided that $|z_2| > |z_1|$ and $n \neq 0$, we see that $(G_1 \setminus \{1\})X_1 \subset X_2$. Similarly $(G_2 \setminus \{1\})X_2 \subset X_1$ is seen. $\square$

A related theorem that is more general and much more difficult to prove is posted below without proof:

Theorem 11.3. J.Tits Alternative

Let H be a finitely generated subgroup of $GL(n,F)$. Then either $F(2)$ embeds into H or H contains a normal subgroup with $|H : N| < \infty$ and N is solvable.

Definition 11.1. A

roup G is *residually finite* if there exists a family of homomorphisms $\varphi_i \colon G \to G_i$ with $|G_i| < \infty$ and $\bigcap_i \ker \varphi_i = (1)$.

It is obvious that every subgroup of a residually finite group is again residually finite.

$F(2)$ is residually finite, as a result that $SL(n,\mathbb{Z})$ is residually finite: consider the homomorphisms $SL(n,\mathbb{Z}) \to SL(n,\mathbb{Z}/m\mathbb{Z})$ and we are done.

Recall that, let p be a prime number, G is a finite p-group if $|G| = p^s$. A group G is *residually-p* if there exists a family of homomorphisms $\varphi_i \colon G \to G_i$ where each G_i is a finite p-group and $\bigcap_i \ker \varphi_i = (1)$.

Again, every subgroup of a residually-p group is residually-p.

$F(2)$ is residually-p for any prime p: we have $F(2) = \left\langle \begin{pmatrix} 1 & p \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ p & 1 \end{pmatrix} \right\rangle \subset SL(2,\mathbb{Z},p)$, where $SL(n,\mathbb{Z},p)$ is defined by

$$SL(n,\mathbb{Z},m) := \{A \in SL(n,\mathbb{Z}) \mid A = I \bmod m\}.$$

The following exercise thus implies what we claimed:

Exercise 11.1

Show that $SL(2,\mathbb{Z},p)/SL(2,\mathbb{Z},p^s)$ is a finite p-group.

Solution. By Corollary 5.3 in [Hun80], it suffices to show that every element in $SL(2,\mathbb{Z},p)/SL(2,\mathbb{Z},p^s)$ has order a power of p. Indeed, every element in $SL(2,\mathbb{Z},p)$ is of the form $I + p \cdot A$ for some matrix $A \in SL(n,\mathbb{Z})$, and the binomial theorem implies that

$$(I + p \cdot A)^{p^s} = I + p^s \cdot (\text{something}).$$

Remark 11.1

With a similar argument (along with the Lagrange's theorem), one sees that any finite residually-p group is in fact a p-group.

Wreath Products

Recall that the Cartesian product $\prod_{i \in I} G_i$ of an arbitrary family of groups G_i index by I can be thought as a subset of functions from I to $\bigcup_{i \in I} G_i$.

Example 11.1

The universal property of the Cartesian product implies that residually-p groups are precisely groups that are embeddable into a Cartesian product of a family of finite p-groups.

The direct product $\overline{\prod}_{i\in I}G_i$ is the subgroup of $\prod_{i\in I} G_i$ where for each element $(g_i)_{i\in I}$ there are only finitely many components g_i that are not equal to 1. Every G_i embeds naturally into $\overline{\prod}_{i\in I}G_i$, and G_i and G_j commute for any $i \neq j$ seen as subgroups of $\overline{\prod}_{i\in I}G_i$.

Except the universal property as a final object that the direct product inherits as a subobject of Cartesian product, the finiteness makes the direct product a quotient of free product, which means that it enjoys a universal property as an initial object. To be explicit, the direct product is free product quotient the relation such that G_i and G_j are made commutative whenever $i \neq j$, hence the universal property is that, for any family $\{\varphi_i\}_{i\in I}$ of homomorphisms $\varphi_i\colon G_i \to G$ such that $\varphi_i(G_i)$ and $\varphi_j(G_j)$ commute in G whenever $i \neq j$, there exists a unique homomorphism $\overline{\varphi}\colon \overline{\prod}_{i\in I}G_i \to G$ such that $\varphi_i = \overline{\varphi}\circ\iota_i$ for each i where $\iota_i\colon G_i \hookrightarrow \overline{\prod}_{i\in I}G_i$ is the natural embedding. See the diagram below.

$$\begin{tikzcd}
G_i \arrow[r, hook, "\iota_i"] \arrow[dr, "\varphi_i"'] & \overline{\prod}_{i\in I}G_i \arrow[d, dashed, "\overline{\varphi}"] \\
& G
\end{tikzcd}$$

Using the presentation of free product, we thus obtain a presentation for the direct product: Suppose that each G_i is presented by $G_i = \langle\, X_i \mid R_i(X_i) = 1\,\rangle$, then

$$\overline{\prod_{i\in I}}G_i = \langle\, \sqcup_{i\in I}X_i \mid R_i(X_i) = 1,\ x_ix_j = x_jx_i,\ i\neq j\,\rangle\,.$$

Note that there is no similar thing that holds for the Cartesian product, since the Cartesian product is not generated by the groups G_i's.

Remark 11.2

Similarly one can define the Cartesian product and direct product for algebras. However, note that the direct product of infinitely many algebras does not contain a multiplicative identity.

§12 Lecture 12

Recall that so far we can conclude the normal forms in $\overline{\prod}_{i\in I}G_i$, which is explicitly, for any $x \in \overline{\prod}_{i\in I}G_i$, we have unique expression (up to permutation) that $x = g_{i_1}\cdots g_{i_k}$ for distinct i's with $g_{i_j}\in G_{i_j}$ for each $j = 1,\cdots,k$. Let A and B be two groups and B acts on A via a left action $B \to \mathrm{Aut}(A)$, which induces a right action of the opposite group B^{op} on A. The *semidirect product of B acting on A*, denoted as $A\rtimes B$ or $B\ltimes A$, is defined as the group whose underlying set is the set of pairs

$$A\rtimes B \coloneqq \{b\cdot a \mid b\in B,\ a\in A\} = B\times A,$$

and the multiplication is given by, for any b_1a_1 and b_2a_2, that

$$(b_1\cdot a_1)(b_2\cdot a_2) \sim b_1^{\mathrm{op}}b_2^{\mathrm{op}}\left((b_2^{\mathrm{op}})^{-1}a_1b_2^{\mathrm{op}}\right)a_2 \sim b_1b_2\cdot(a_1^{b_2}a_2),$$

where $a_1^{b_2}$ denotes the element in A obtained by the action of b_2 on a_1. The opposite is taken because the conjugation by b_2 in the form $b_2^{-1}a_1b_2$ is a right action, instead of left. With the opposite taken, the notation is compatible with conjugation in the sense that the following are all equal to $(b_1^{\mathrm{op}}b_2^{\mathrm{op}})^{-1}ab_1^{\mathrm{op}}b_2^{\mathrm{op}}$:

$$\left(a^{b_1}\right)^{b_2} \xlongequal{\text{left action}} a^{b_2b_1} = a^{(b_2b_1)^{\mathrm{op}}} = a^{b_1^{\mathrm{op}}b_2^{\mathrm{op}}} \xlongequal{\text{right action}} \left(a^{b_1^{\mathrm{op}}}\right)^{b_2^{\mathrm{op}}}.$$

In other words, we have an isomorphism

$$A\rtimes B \cong \langle\, A, B^{\mathrm{op}} \mid (b^{\mathrm{op}})^{-1}ab^{\mathrm{op}} = a^b,\ a\in A,\ b\in B\,\rangle,$$

given by $b\cdot a \mapsto b^{\mathrm{op}}a$. Note that A is normal seen as a subgroup of $A\rtimes B$.

Now let us consider the set of all functions from B to A, denoted by $\mathrm{Fun}(B, A)$. Endowed with the point-wise multiplication, it is a group isomorphic to the Cartesian product $A^B = \prod_B A$. Consider the (left) action of B on $\mathrm{Fun}(B, A)$ defined by, for any element $b \in B$ and $f \in \mathrm{Fun}(B, A)$, the action of b brings f to the function $f^b \colon B \to A$ given by $f^b(b') = f(bb')$ for any $b' \in B$.

With this action, the *wreath product of A by B*, denoted as $A \wr B$, is then defined by

$$A \wr B \coloneqq \mathrm{Fun}(B, A) \rtimes B.$$

Note that there is a natural embedding $A \hookrightarrow \mathrm{Fun}(B, A) \hookrightarrow A \wr B$ by sending elements $a \in A$ to the function $\dot{a} \colon B \to A$ defined by $\dot{a}(1_B) = a$ and $\dot{a}(b) = 1_A$ whenever $b \neq 1_B$. Let us denote the image of this embedding by $\dot{A}$.

For this lecture, we are more concerned about the *restricted wreath product*, which is defined by

$$A \bar{\wr} B \coloneqq \overline{\prod_B} A \rtimes B = \left\{ b \cdot f \mid b \in B,\ f \in \overline{\prod_B} A \subset \mathrm{Fun}(B, A) \right\}.$$

The natural embedding above also restricts to this case.

A basic but important observation is that elements in $\dot{A}^b$ commutes with $\dot{A}^{b'}$ whenever $b \neq b'$.

Remark 12.1

Note that since in $A \wr B$ the conjugation by B is not trivial as long as both A and B are not, the wreath product (either restricted or not) of two nontrivial groups is never abelian.

For the restricted wreath product, we have the following generating theorem:

Theorem 12.1

Suppose that $B = \langle b_1, \cdots, b_m \rangle$ and $A = \langle a_1, \cdots, a_n \rangle$, then $A \bar{\wr} B$ is generated by $b_1, \cdots, b_m, \dot{a}_1, \cdots, \dot{a}_n$.

Proof. Since every element in $A \bar{\wr} B$ is of the form bf, it suffices to show that $b_1, \cdots, b_m, \dot{a}_1, \cdots, \dot{a}_n$ generate $\overline{\prod_B} A$. Indeed, we have

$$f = \prod_{b \in B} \dot{f(b)}^{b^{-1}},$$

where only finitely many $\dot{f(b)}$ is nontrivial. Since $\dot{f(b)} \in \dot{A} = \langle \dot{a}_1, \cdots, \dot{a}_n \rangle$, we are done. ☐

Therefore, the restricted wreath product of any two finitely generated groups is finitely generated. However, it does not preserve the property of being finitely presented:

Theorem 12.2

The group of restricted wreath product $\mathbb{Z} \bar{\wr} \mathbb{Z} = \langle a \rangle \bar{\wr} \langle b \rangle$ is not finitely presented.

Before the proof of the theorem, let us firstly look at the following lemma:

Lemma 12.3

Let $G = \langle X \mid R = 1 \rangle$ be a presentation of group G where $|X| < \infty$. If G is finitely presented, then there exists a finite subset R_0 of R such that $G = \langle X \mid R_0 = 1 \rangle$.

Proof of Lemma 12.3. By proposition 2.3, we have shown that if G is finitely presented, then there exists a finite set of relations S such that

$$G = \langle X \mid S = 1 \rangle.$$

For each element $s \in S$, since R and S generate a same normal subgroup of the free group $F(X)$, there exists $r_1, \cdots, r_t \in R$ and words $g_1, \cdots, g_t$ such that

$$s = r_1^{g_1} \cdots r_t^{g_t}.$$

Since $|S| < \infty$, collect the $r_1, \cdots, r_t$ for each $s \in S$ and we obtain the desired R_0. ☐

Proof of Theorem 12.2. By theorem 12.1 we know that $\mathbb{Z}\bar{\wr}\mathbb{Z}$ is generated by b and $\dot{a}$. The fact that $\dot{A}^b$ and $\dot{A}^{b'}$ commutes whenever $b \neq b'$ thus gives an epimorphism

$$\langle\, x, y \mid [x^{y^i}, x^{y^j}] = 1,\ i, j \in \mathbb{Z} \,\rangle \to \mathbb{Z}\bar{\wr}\mathbb{Z},\ x \mapsto \dot{a},\ y \mapsto b,$$

where $x^{y^i} = y^{-i}xy^i$ and $[x^{y^i}, x^{y^j}] = (x^{y^j}x^{y^i})^{-1}x^{y^i}x^{y^j}$. This epimorphism is in fact an isomorphism: any word in the presentation can be written uniquely (up to permutation) as

$$y^k(x^{y^{i_1}})^{m_1}\cdots(x^{y^{i_j}})^{m_j},$$

where $i_1, \cdots, i_j$ are distinct, which is sent to $b^k(\dot{a}^{b^{i_1}})^{m_1}\cdots(\dot{a}^{b^{i_j}})^{m_j}$. Note that $b^k(\dot{a}^{b^{i_1}})^{m_1}\cdots(\dot{a}^{b^{i_j}})^{m_j} = 1$ only if $k = 0$, since $\mathbb{Z}\bar{\wr}\mathbb{Z} = \bigsqcup_{k\in\mathbb{Z}}\left(b^k \cdot \overline{\prod_B}\langle\, \dot{a} \,\rangle\right)$ and the only coset containing the identity is $\overline{\prod_B}\langle\, \dot{a} \,\rangle$. However, for $(\dot{a}^{b^{i_1}})^{m_1}\cdots(\dot{a}^{b^{i_j}})^{m_j} = 1$, the only chance is that $m_1 = \cdots = m_j = 0$, as one can see by evaluating it at elements in $\langle\, b \,\rangle$. Therefore we obtain a presentation

$$\mathbb{Z}\bar{\wr}\mathbb{Z} = \langle\, x, y \mid [x^{y^i}, x^{y^j}] = 1,\ i, j \in \mathbb{Z} \,\rangle\,.$$

Note that since we have

$$[x^{y^i}, x^{y^j}] = [x^{y^{i-j}}, x]^{y^j},$$

and

$$[a, b] = 1\ \Leftrightarrow [b, a] = 1,$$

there is in fact

$$\mathbb{Z}\bar{\wr}\mathbb{Z} = \langle\, x, y \mid [x^{y^i}, x] = 1,\ i \in \mathbb{N}^* \,\rangle\,.$$

By lemma 12.3, it now suffices to show that there does not exist a finite subset S of $\{[x^{y^i}, x] \mid i \in \mathbb{N}^*\}$ such that $\mathbb{Z}\bar{\wr}\mathbb{Z} = \langle\, x, y \mid S = 1 \,\rangle$. The existence of such S would give that

$$\mathbb{Z}\bar{\wr}\mathbb{Z} = \langle\, x, y \mid [x^{y^i}, x] = 1,\ i = 1, \cdots, n \,\rangle,$$

for some n that is large enough, hence it suffices to show that for any $n \in \mathbb{N}^*$, the relations $[x^y, x], \cdots, [x^{y^{n-1}}, x]$ cannot give $[x^{y^n}, x]$, so that no finite subset of $\{[x^{y^i}, x] \mid i \in \mathbb{N}^*\}$ would generate it. For this purpose we need only construct a specific group where we have $[x^{y^i}, x] = 1$ for $i = 1, \cdots, n-1$, while $[x^{y^n}, x] \neq 1$.

Let G be any nonabelian group, which means that there exists $u, v \in G$ such that $uv \neq vu$. Consider $G \wr \mathbb{Z} = G \wr \langle\, b \,\rangle$ and the function $f\colon \langle\, b \,\rangle \to G$ given by $f(1) = u$, $f(b^n) = v$, and $f(b^k) = 1_G$ for any $k \neq 0, n$. Now that in $G \wr \mathbb{Z}$, we have $[f^{b^i}, f] = 1$ for $i = 1, \cdots, n-1$ while $[f^{b^n}, f] \neq 1$ because $f^{b^n}(1) = v$. $\square$

Remark 12.2

With a similar argument, one sees that the restricted wreath product of any infinite group B acting on nontrivial A is not finitely presented. In fact, since $|A^B| \geq 2^{\mathbb{N}} = |\mathbb{R}|$, we would have that $A \wr B$ is uncountable.

Exercise 12.1

Show that the group of isometries on the finite rooted tree

is isomorphic to the iterated wreath product $(C_2 \wr C_2) \wr C_2$, where $C_2 = \mathbb{Z}/2\mathbb{Z}$.

§13 Lecture 13

Let us end this section with the following theorem:

Theorem 13.1. Krasner–Kaloujnine

Let $A \lhd G$ and $B = G/A$. Then there exists a natural embedding $G \hookrightarrow A \wr B$.

Proof. Note that B is a set of cosets. Let us denote the image of an element $g \in G$ in B by $\bar{g}$. Note that adding bar is a homomorphism since it is identical to the quotient map. Let $s\colon B \to G\colon b \mapsto b^s$ be any function of choosing representatives. Since elements in $A \wr B$ are of the form $b \cdot f$, it is natural to expect that the embedding $G \hookrightarrow A \wr B$ is of the form $g \mapsto \bar{g} f_g$ for some $f_g \in \mathrm{Fun}(B, A)$. Assuming this form, then for it to be a homomorphism we would need exactly that $f_{1_G} = 1_{\mathrm{Fun}(B,A)}$ and

$$\overline{g_1 g_2} f_{g_1 g_2} = (\overline{g_1} f_{g_1})(\overline{g_2} f_{g_2}) = \overline{g_1 g_2} f_{g_1}^{\overline{g_2}} f_{g_2}.$$

Therefore, it suffices to define f_g in the way that $f_{1_G} = 1_{\mathrm{Fun}(B,A)}$ and $f_{g_1 g_2} = f_{g_1}^{\overline{g_2}} f_{g_2}$. Let us define

$$f_g(b) \coloneqq ((\bar{g}b)^s)^{-1} g b^s.$$

It is easy to see that $((\bar{g}b)^s)^{-1} g b^s \in A$, since $\bar{g}b = \bar{g}\overline{b^s} = \overline{gb^s}$. Hence $f_g\colon B \to A$ is well-defined. Clearly $f_{1_G} = 1_{\mathrm{Fun}(B,A)}$. For $f_{g_1 g_2} = f_{g_1}^{\overline{g_2}} f_{g_2}$, we have

$$f_{g_1}^{\overline{g_2}}(b) f_{g_2}(b) = f_{g_1}(\overline{g_2} b) f_{g_2}(b) = ((\overline{g_1 g_2} b)^s)^{-1} g_1 (\overline{g_2} b)^s (g_1 (\overline{g_2} b)^s)^{-1} g_2 b^s = ((\overline{g_1 g_2} b)^s)^{-1} g_1 g_2 b^s = f_{g_1 g_2}(b).$$

Therefore the homomorphism is well-defined.

To show that it is an embedding, suppose that $g \in G$ is mapped to $1_{A \wr B}$. Then $\bar{g} f_g = 1_{A \wr B}$, which forces that $\bar{g} = 1_B$, hence $g \in A$. Also, we have $f_g = 1_{\mathrm{Fun}(B,A)}$, which means that

$$((\bar{g}b)^s)^{-1} g b^s = 1,$$

for any $b \in B$. Since $\bar{g}b = b$, we have that $(b^s)^{-1} g b^s = 1$, therefore $g = 1_G$. ☐

The Burnside Problem

Consider a finitely generated group G where every element $g \in G$ is of finite order. The condition that every $g \in G$ has a finite order is called that G is *torsion*.

The *General Burnside Problem* asks, if a group G is finitely generated and is torsion, then must G be finite?

The answer is: No. We will construct two counterexamples later, each of them are important by its own right.

The following states a less general version of the problem, which is known as The Burnside Problem:

Question. The Burnside Problem

If a group G is finitely generated and there exists $n \in \mathbb{N}^*$ such that $g^n = 1_G$ for all $g \in G$, then must G be finite?

For $n = 2$, this is trivial: since $g = g^{-1}$ for any $g \in G$, G is abelian. Since every element in G has order no larger than 2, every element is a word without any repetition of alphabets, hence $|G| \le 2^m$, where m is the number of generators of G.

However, the problem is highly nontrivial for any $n \ge 3$: Burnside himself proved the statement for the case $n = 3$, Sanov proved it for the case $n = 4$ and M.Hall proved it for the case $n = 6$. The problem for $n = 5$, however, remains open till now. In 1968, Novikov-Adian proved it for the cases of any odd $n \ge 4381$, which used a simultaneous induction on more than 100 indices and has length more than 300 pages; we will not go through that proof.

Noticing that any finite group can be embedded into the general linear group $\mathrm{GL}(n, F)$ for any field F (consider the inclusion $G \hookrightarrow FG$, where FG is the free vector space generated by G over F), Burnside proved the following restricted statement:

Theorem 13.2. Burnside

Every finitely generated torsion subgroup of $GL(n,\mathbb{C})$ is finite.

We will prove this theorem, admitting the following lemma which is again by Burnside. Let $V=\mathbb{C}^n$, we will consider $\mathrm{GL}(V)$ instead of $\mathrm{GL}(n,\mathbb{C})$, so that we do not specify any basis.

Lemma 13.3. Burnside

If a subset S of $GL(V)$ acts irreducibly on V, i.e. there is no non-trivial subspace of V that is invariant under all elements in S, then $\mathbb{C}S := span_{\mathbb{C}} S = End_{\mathbb{F}}(V)$.

The statement is easily verified for $S=\mathrm{GL}(V)$ since $\left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)=\frac{1}{2}\left(\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)+\left(\begin{smallmatrix}1&0\\0&-1\end{smallmatrix}\right)\right)$ and so on.

Observe that, if $W\subset V$ is invariant under all elements in S, then S acts naturally on V/W, and if we choose a basis of W and then extend it to a basis of V, then elements in S would have matrix representations of the form

$$\begin{pmatrix} * & * \\ 0 & * \end{pmatrix}.$$

With this observation, we state and prove the following two lemmas, which will be useful in our proof of Burnside's Theorem:

Lemma 13.4

For any subset $S\subset GL(V)$, there exists a finite chain of subspaces

$$\{0\}=V_0\subsetneq V_1\subsetneq V_2\subsetneq\cdots\subsetneq V_k=V,$$

such that all V_i's are S-invariant and S acts irreducibly on each fraction V_{i+1}/V_i.

Proof. If S acts irreducibly on V then there is nothing to prove. If not, let V_1 be a nontrivial S-invariant subspace of minimal dimension. If S acts irreducibly on V/V_1, then we are done; if not, let $V_2\subset V$ be the preimage of a nontrivial S-invariant subspace of minimal dimension in V/V_1, and repeat this procedure. We will be done in at most $\dim V=n$ steps. ☐

Lemma 13.5

In proving Burnside's theorem, it suffices to consider $G<GL(V)$ that acts irreducibly on V.

Proof. Suppose that the theorem has been proved for any G that acts irreducibly on V. Suppose now that we are given a finitely generated torsion subgroup $G\subset\mathrm{GL}(V)$ whose action on V is not necessarily irreducible. Let $\{0\}=V_0\subsetneq V_1\subsetneq V_2\subsetneq\cdots\subsetneq V_k=V$ be a chain as in lemma 13.4 with $S=G$. Choose a basis of V_1 and extend it to a basis of V_2, and so on until we obtain a basis of $V_k=V$. Then the matrix of an element $g\in G$ with respect to this basis is of the form

$$M(g)=\begin{pmatrix} M(g|_{V_1}) & * & 0 & 0 & 0 \\ 0 & M(g|_{V_2/V_1}) & * & 0 & \\ 0 & 0 & M(g|_{V_3/V_2}) & * & \vdots \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \cdots & M(g|_{V/V_{k-1}}) \end{pmatrix}.$$

Consider the map $G\to M_n(\mathbb{C})$ given by

$$g\mapsto\begin{pmatrix} M(g|_{V_1}) & 0 & 0 & 0 & 0 \\ 0 & M(g|_{V_2/V_1}) & 0 & 0 & \\ 0 & 0 & M(g|_{V_3/V_2}) & 0 & \vdots \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \cdots & M(g|_{V/V_{k-1}}) \end{pmatrix},$$

it is clear that it is a group homomorphism. Its image is a finite set, because by our assumption we know that the choice of each $M(g|_{V_{i+1}/V_i})$ is finite. The map is in fact an embedding, because if $g\mapsto 1$, then we have

$$M(g) = \begin{pmatrix} 1 & & * \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = I + U,$$

where $U \in M_n(\mathbb{C})$ is upper-triangular. Recall that the order of g should be finite, we have

$$I + dU + \binom{d}{2} U^2 \cdots + U^d = (I+U)^d = I,$$

for some positive integer d, hence $dU + \binom{d}{2} U^2 \cdots + U^d = (I+U)^d = 0$. If $U \neq 0$, then consider the upper diagonal in N which is the nearest to the diagonal among those whose entries are not identically zero. However, the particular upper diagonal in U^k for any $k \geq 2$ must be identically zero, giving that the upper diagonal itself must be zero, a contradiction. Therefore we must have $U = 0$, concluding the proof. □

Proof of Theorem 13.2. By lemma 13.5, we may assume that G acts irreducibly. Let us prove the theorem firstly with an additional assumption, and then remove the assumption. The additional assumption is that the orders of elements of G are bounded uniformly, i.e. there exists $d \in \mathbb{N}^*$ such that $g^d = 1$ for all $g \in G$. With this assumption, we then see that the diagonal of the Jordan form of an element $g \in G$ consists of d-th roots of the unity. Therefore, there are only finitely many available choices for the trace $\mathrm{Tr}(g)$ for elements $g \in G$, i.e. the image of the trace $\mathrm{Tr}\colon G \to \mathbb{C}\colon g \mapsto \mathrm{Tr}(g)$ is finite.[2] By lemma 13.3, since G acts irreducibly, G spans $\mathrm{End}_{\mathbb{C}}(V)$, so we can find elements $g_1, \cdots, g_{n^2} \in G$ that form a basis of $\mathrm{End}_{\mathbb{C}}(V)$. Consider the map

$$G \to \mathbb{C}^{n^2} : g \mapsto (\mathrm{Tr}(gg_1), \cdots, \mathrm{Tr}(gg_{n^2})),$$

then its image is again finite, since there are only finitely many choices for each entry. Now it suffices to show that this map is injective. Indeed, since Tr is linear, if $\mathrm{Tr}(gg_i) = \mathrm{Tr}(g'g_i)$ for all $i = 1, \cdots, n^2$, then $\mathrm{Tr}((g-g')g_i) = 0$ for all $i = 1, \cdots, n^2$. Since $g_1, \cdots, g_{n^2}$ spans $\mathrm{End}_{\mathbb{C}}(V)$, we thus see that $\mathrm{Tr}((g-g')M) = 0$ for any $M \in \mathrm{End}_{\mathbb{C}}(V)$. Since $\mathrm{Tr}((a_{ij})E_{kl}) = a_{lk}$, the only chance that this happens is $g - g' = 0$, concluding the injectivity.

The rest of the proof will be given in the next lecture. □

§14 Lecture 14

Continue of the Proof of Theorem 13.2. Let us now remove the assumption. We will use some field theory.

Lemma 14.1

For any fixed $n \geq 1$, there exists a sufficiently large N such that any torsion matrix from $GL(n, \mathbb{Q})$ has order no more than N. It follows that $A^{N!} = 1$ for all torsion matrix $A \in GL(n, \mathbb{Q})$.

Proof of Lemma 14.1. Let $A \in \mathrm{GL}(n, \mathbb{Q})$ be torsion. Brought to $\mathbb{C}$, our preceding results tell that any torsion matrix A must be diagonalizable over $\mathbb{C}$, and the entries λ_i's in the diagonal would all be roots of the unity. Suppose that λ_i has multiplicative order d, i.e. $\lambda_i^d = 1$, then basic field theory gives that $[\mathbb{Q}(\lambda_i) : \mathbb{Q}] \coloneqq \dim_{\mathbb{Q}} \mathbb{Q}(\lambda_i) = \varphi(d)$, where $\varphi\colon \mathbb{N}^* \to \mathbb{N}^*$ is the Euler function, i.e. $\varphi(d)$ is the number of integers less than d and are coprime to d. On the other hand, the characteristic polynomial of A tells that the λ_i's are roots of a polynomial of degree n, hence $[\mathbb{Q}(\lambda_i) : \mathbb{Q}] \leq n$. Therefore, for any λ_i on the diagonal, its multiplicative order is no more than the integer N such that $\varphi(N+m) > n$ for all $m \in \mathbb{N}^*$. Such N exists, because φ is asymptotically increasing in the sense that $\varphi(n)/n^{1-\delta} \to \infty$ as $n \to \infty$ for any $\delta > 0$. □

Let X be any set, we can consider the field $\mathbb{Q}(X)$ of all rational polynomials with variables in X, i.e. the field generated by elements of the form $\frac{f(x_1,\cdots,x_m)}{g(y_1,\cdots,y_k)}$, with $x_1, \cdots, x_m, y_1, \cdots, y_k \in X$, $f, g \in \mathbb{Q}[X]$ and $g \neq 0$. Note that since $\mathbb{Q}$ is an infinite field, by an induction on the number of variables one sees that for any polynomial $f \in \mathbb{Q}[X]$ that is not identically zero, there exists $\alpha_1, \cdots, \alpha_m \in \mathbb{Q}$ such

[2] In fact, by the proof of the preceding Lemma, the Jordan form of g must be diagonal; but this cannot tell the finiteness of G, because the Jordan basis for each element may be different. Note though that the trace is invariant under the choice of basis.

that $f(\alpha_1, \cdots, \alpha_m) \neq 0$. Moreover, given any finite set of nonzero polynomials $\{f_i\}_{i=1}^s$ over $\mathbb{Q}$, their multiplication $\prod_{i=1}^s f_i$ is nonzero, hence there exists scalar αs in $\mathbb{Q}$ such that $f_i(\alpha) \neq 0$ for all $i = 1, \cdots, s$.

With the same N as above, we have the following lemma

Lemma 14.2

For any fixed $n \geq 1$, any torsion matrix from $GL(n, \mathbb{Q}(X))$ has order dividing $N!$.

Proof of Lemma 14.2. Suppose $A^{N!} \neq I_n$, then we get a system of inequalities. Substitute scalar α's in $\mathbb{Q}$ into the indeterminants such that all the inequalities along with those that the dominator is nonzero hold and we obtain a contradiction to the preceding lemma. □

Finally, let G be a finitely generated subgroup of $GL(n, \mathbb{C})$, then there exists a finitely generated subfield $L = \mathbb{Q}(\alpha)$, where α is the set of entries of generators of G. By basic field theory, there exists a transcendental field $K = \mathbb{Q}(X)$ for some set X such that there is a chain of inclusions

$$\mathbb{Q} \subset K \subset L,$$

and $[L : K] < \infty$. Write $s := [L : K]$, then L naturally embedds into $M_s(K)$ by considering the right multiplication $L \hookrightarrow \mathrm{End}_K(L)\colon a \mapsto R_a$. With this embedding, we see that G can be seen as a subgroup of $\mathrm{GL}(ns, K)$ via the chain $G < \mathrm{GL}(n, L) < \mathrm{GL}(ns, K)$. If G is torsion, then the preceding lemma applies and we obtain the condition that the orders of elements of G are uniformly bounded. □

Remark 14.1

In general, it is true that every finitely generated torsion subgroup of $\mathrm{GL}(n, F)$ for any field F is finite. Our proof for $F = \mathbb{C}$ can be divided into two parts, one is that every subgroup where the orders of elements are uniformly bounded is finite, the other is that every finitely generated torsion subgroup satisfies the condition that the orders of elements are uniformly bounded. Our argument for the former only applies when the field has zero characteristic.

Let us move on to construct counterexamples of the General Burnside Problem. From now on we will not assume that every algebra has an identity element.

Let A be an algebra over a field F. An element in A is *nilpotent* if there exists $n \geq 1$ such that $a^n = 0$. The algebra A is *nilpotent* if there exists $N \geq 1$ such that $A^N = (0)$, which means that the product of any N elements in A is zero. A is called a nil algebra if every element in a is nilpotent.

In 1941, A. Kurosh formulated the following question, which turns out to be related to the General Burnside Problem:

Question. Kurosh Problem

Suppose A is finitely generated and nil, must there be that A is nilpotent and A is finite dimensional over F?

H. M. Wedderburn proved the following

Theorem 14.3. H. M. Wedderburn

If $\dim_F A < \infty$ and A is nil, then A is nilpotent.

Conversely, if A is finitely generated and nilpotent with $A^n = (0)$, then A is spanned by the products of its generators of length less than n, hence A is finite dimensional. Therefore, under the setup of Kruosh Problem, finite dimensionality and nilpotency of A are equivalent.

For an algebra without unit, we can consider the direct sum

$$\hat{A} := A \oplus F \cdot 1 = \{(a, \alpha \cdot 1) \mid a \in A, \alpha \in F\},$$

with the obvious structure of algebra. The algebra $\hat{A}$ can be seen as A added the unit, and is called the *unital hull* of A.

Let $a \in A$ be a nilpotent, then $1 + a \in \hat{A}$ is invertible, with inverse given by the finite sum $(1+a)^{-1} = 1 - a + a^2 - a^3 + a^4 - \cdots$. Suppose that $\mathrm{char} F = p > 0$, then $1 + a$ has a finite multiplicative order: suppose $a^n = 0$, choose k such that $p^k \geq n$ and then

$$(1+a)^{p^k} = 1 + \binom{p^k}{1} a + \cdots + \binom{p^k}{p^k - 1} a^{p^k - 1} + a^{p^k} = 1,$$

since each $\binom{p^k}{i}$ is divisible by p for $1 \leq i \leq p^k - 1$.

With these observations, we are now able to relate the Kurosh Problem with the General Burnside Problem:

Proposition 14.4

If there exists a counterexample to the KuroshProblem with char$F = p > 0$ *, then there exists a counterexample to the General Burnside Problem.*

Proof. Let A be a counterexample to the Kurosh Problem, say $A = \langle a_1, \cdots a_m \rangle$ is a nil algebra over F with char$F = p > 0$, and A is not nilpotent. Consider the multiplicative subgroup G of $\hat{A}$ generated by $1 + a_1, \cdots, 1 + a_m$. The observation above tells that G is torsion. If G is not a counterexample to the General Burnside Problem, then we have $|G| = d < \infty$. Then, any product $(1+a_{i_1}) \cdots (1+a_{i_d})$ of length d of generators $1 + a_1, \cdots, 1 + a_m$ would be equal to a shorter product of these generators, because the list in G,

$$1, 1 + a_{i_1}, (1 + a_{i_1})(1 + a_{i_2}), (1 + a_{i_1}) \cdots (1 + a_{i_d}),$$

has $d + 1$ elements, hence there must be

$$(1 + a_{i_1}) \cdots (1 + a_{i_t}) = (1 + a_{i_1}) \cdots (1 + a_{i_t})(1 + a_{i_{t+1}}) \cdots (1 + a_{i_{t+l}}),$$

for some t and l, which forces $(1+a_{i_{t+1}}) \cdots (1+a_{i_{t+l}}) = 1$. Since $(1+a_{i_{t+1}}) \cdots (1+a_{i_{t+l}})$ is a sub-product of $(1 + a_{i_1}) \cdots (1 + a_{i_d})$, replace it by 1 and we obtain a shorter product. Expand the bracket and move the terms, we then obtain

$$a_{i_1} \cdots a_{i_d} = \sum_{k<d} (\text{some coefficient in } \mathbb{Z}/p\mathbb{Z}) a_{j_1} \cdots a_{j_k}.$$

Therefore, since the choice of a_{i_k}'s are arbitrary, every product of generator a_i's of length no less than d can be expressed as a linear combination of products of generators of length less than d, hence A is spanned by products of generators of length less than d, which are finitely many. This concludes that $\dim_F A < \infty$, contradicting to Wedderburn's Theorem. $\square$

Therefore, to construct a counterexample to the General Burnside Problem, it suffices to construct a counterexample to the Kurosh Problem. We will construct it using graded algebras.

Bibliography

[Hun80] Thomas W. Hungerford. *Algebra*. Vol. 73. Graduate Texts in Mathematics. Reprint of the 1974 original. Springer-Verlag, New York-Berlin, 1980, pp. xxiii+502. ISBN: 0-387-90518-9.

[Alu09] Paolo Aluffi. *Algebra: chapter 0*. Vol. 104. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2009, pp. xx+713. ISBN: 978-0-8218-4781-7. DOI: `10.1090/gsm/104`. URL: `https://doi.org/10.1090/gsm/104`.

[Kem11] Gregor Kemper. *A Course in Commutative Algebra*. Graduate Texts in Mathematics. Springer, 2011. ISBN: 9783642035449.