§ 2. Free Groups and Graphs

John R. Stallings, U. C. Berkeley

2.1. Introduction

The idea of a "free group" was originally a group given by a set of generators among which there were no relations. This could be imagined to mean that the generators $\{x, y, \ldots\}$ could be composed with their inverses in such a way that if one had a finite sequence of such without any obvious cancellations, such as $x \cdot x^{-1}$, then the result is non-trivial. Another way to look at this is category-wise, as the left adjoint of the forgetful functor from groups to sets. These groups have been singled out since, at least, Dyck (*Math. Ann.* **20**, 1882), and results about them have been achieved by Dehn, Nielsen, Schreier, etc.

2.2. Free monoid

alphabet, n-tuple, words, length, empty word, concatenate, free monoid

Let A be a set (an **alphabet**). Define the set A^* to consist of the disjoint union of the sets A^n, for $n = 0, 1, \ldots$. By A^n is meant the set of n-**tuples** of elements of A, that is, the set of all functions from the set $\{1, 2, \ldots, n\}$ to A; call the elements of A^n the **words of length** n in A; if w has length n, write $|w| = n$. There is a unique word in A of length 0, called Λ, the **empty word**. (—In fact, the better analogy is not to alphabets and words, but to words and sentences.—)

Write a word of length n in A thus: $w = a_1 a_2 \cdots a_n$, for $a_i \in A$. Given a word w as above of length n and one of length k, say $u = b_1 b_2 \cdots b_k$, then **concatenate** the two to get a word of length $n + k$ and write $w \cdot u = c_1 c_2 \cdots c_{n+k}$, where $c_i = a_i$ if $i \leq n$ and $c_i = b_{i-n}$ if $i > n$. Thus, A^* has an associative binary operation (concatenation) defined on it, and Λ is a two-sided identity element; it is convenient to denote this triple of things by A^* alone, of course.

Call A^* the **free monoid on** A; this terminology is justified by the fact that, for every monoid M and every function f from A to the underlying set of M, there is a unique monoid-map $f^* : A^* \to M$ extending f. This A^* is an explicit way to describe a left adjoint to the forgetful functor from monoids to sets.

2.2.1. **Congruence relations.**

congruence, generated by a binary relation, normal subgroup

Suppose that M is a monoid, and that $\approx$ is an equivalence relation on the underlying set of M. Call $\approx$ a **congruence relation** if, for every $x, x', y, y' \in M$, if $x \approx x'$ and $y \approx y'$, then $xy \approx x'y'$.

The significance of the notion of congruence is this: If $\approx$ is a congruence relation on the monoid M, then there is a unique binary operation on the set of equivalence classes $M/\approx$, making it into a monoid, such that the quotient function $M \to M/\approx$ is a monoid-map. Furthermore, if $\phi : M \to N$ is a monoid-map, then the relation $\approx$ on M, defined by $x \approx y$ iff $\phi(x) = \phi(y)$, is a congruence relation and the monoid $M/\approx$ is monoid-equivalent to the image monoid $\phi(M)$.

If $\sim$ is a binary relation on M, define the congruence relation $\approx$ **generated** by $\sim$: Write $x \approx y$, when there is some n (in $\{0, 1, 2, \ldots\}$), and $z_0, z_1, \ldots, z_n \in M$, such that $x = z_0$, $y = z_n$, and such that for all $i = 1, \ldots, n$, there exist formulas $z_{i-1} = psq$, $z_i = ptq$, where either $s \sim t$ or $t \sim s$.

Since every group is a monoid, there are congruence relations in groups. Here the theory reduces to the notion of **normal subgroup**. If G is a group and $\approx$ is a congruence relation on G, define $N = \{x \in G \mid x \approx 1\}$. Then N is a subgroup of G satisfying the rule that for all $g \in G$ and all $x \in N$, the element gxg^{-1} belongs to N (that is the definition of "N is a normal subgroup of G", which is written $N \triangleleft G$). The relation $\approx$ can be defined in terms of N thus: $x \approx y$ iff $xy^{-1} \in N$ (which is equivalent to $x^{-1}y \in N$); and if $N \triangleleft G$ is any given normal subgroup, then the relation $\approx$ thus defined is indeed a congruence relation on G.

2.2.2. **Monoid Presentations.**

presentation, monoid presented by a presentation

A **presentation** of a monoid, $\mathcal{P} = \langle A \mid R \rangle$, consists of an alphabet A and a binary relation R on A^*. This binary relation R generates a congruence relation $\approx$ on A^*, and the quotient monoid $A^*/\approx$ is called the **monoid presented by** $\mathcal{P}$. It is convenient to imagine that A and R are finite, and to write a presentation by listing the elements of the alphabet A, and then listing the pairs of elements related by R in some form such as "u=v", when $(u, v) \in R$ might be more precise.

For instance, $\mathcal{P} = \langle a,\ b,\ c \mid ac = bc,\ ab = c,\ ba = \Lambda \rangle$. You can see that it might be hard to explicitly describe the monoid presented by such a presentation. For instance, in this case,

$$\begin{aligned} \underline{ac}ac \approx bcac = b\underline{\Lambda}cac \approx bb\underline{ac}ac \approx bbbcac \\ = bbbc\underline{\Lambda}ac \approx bbbcba\underline{ac} \approx bbbcb\underline{ab}c \approx bbbcbcc. \end{aligned}$$

Every monoid is monoid-equivalent to the monoid presented by some presentation. Various questions about presentations are hard to answer in particular cases. For example, a given monoid may or may not have a finite presentation; this is a subtle and hard to decide property of monoids.

2.3. Free group

bar operation, formal inverse, explicit left adjoint

Consider an alphabet A; construct a set disjoint from, and in 1-1 correspondence with, A, and call it $\overline{A}$. The element of $\overline{A}$ corresponding to $a \in A$ will be called $\overline{a}$. Extend the **bar operation** by saying that $\overline{\overline{a}} = a$. Thus, the bar operation is just the same as a free action of the group of order 2 on the set $A \cup \overline{A}$. Given a set X with an operation $x \mapsto \overline{x}$ on it, such that $x \neq \overline{x}$ and $\overline{\overline{x}} = x$, conversely, one can (using the Axiom of Choice) choose half of X to be the set A so that $X = A \cup \overline{A}$. Then look at the monoid presentation $\mathcal{P} = \langle A \cup \overline{A} \mid R \rangle$, where R consists of all the relations of the form $a\overline{a} = \Lambda$ and $\overline{a}a = \Lambda$, for all $a \in A$. The monoid of this presentation is in fact a group; the inverse of the element represented by a word $w = x_1 x_2 \cdots x_n$ (where $x_i \in A \cup \overline{A}$), is $\overline{x_n} \cdots \overline{x_2}\,\overline{x_1}$.

In fact, we could have started with any monoid presentation, added these **formal inverses** for the generators and added the relations making them into inverses in the monoid of the new presentation. The resulting monoid would be a group. This would provide a left-adjoint to the inclusion functor from groups into monoids. And so, pure thought would show that the group defined by the presentation described in the preceding paragraph is the free group on the set A.

The group of this presentation $\mathcal{P} = \langle A \cup \overline{A} \mid \{a\overline{a} = \Lambda : a \in A \cup \overline{A}\} \rangle$ is $F(A)$. The elements of A can be thought of as certain words of length 1, and then as equivalence classes of these words in the group of the presentation. Thus, there is a canonical set-map from A to $F(A)$. The universal property it has is that if $f : A \to G$ is any set-map from A to (the underlying set of) any group G, then there is a unique group-map $F(A) \to G$ consistent with the canonical $A \to F(A)$. In other words, this is an **explicit** way to describe the left adjoint of the forgetful functor from groups to sets.

2.3.1. **Reduced words.**

reduced, elementary reduction, unique reduced words

A word belonging to $(A \cup \overline{A})^*$ is said to be **reduced** if it contains no subword of the form $a\overline{a}$ or $\overline{a}a$, for $a \in A$. That is, $w = x_1 x_2 \cdots x_n$ is reduced (where $x_i \in A \cup \overline{A}$), is to mean that for all $i = 2, \ldots, n$, it is the case that $x_{i-1} \neq \overline{x_i}$. The empty word Λ is reduced, as is every word of length 1. The act of replacing a subword $a\overline{a}$ or $\overline{a}a$ by the empty word (within a word w) is called **elementary reduction**. Thus a reduced word is one to which no elementary reduction can be applied. Elementary reduction is the basic step in defining the equivalence relation in the monoid of the presentation of the free group $F(A)$; it decreases the length of a word by 2; and so cannot be done infinitely many times to any particular word. Thus, there is a sequence of elementary reductions from any word in $(A \cup \overline{A})^*$ to a reduced word. Thus, every element of $F(A)$ is represented by some reduced word. Let us call the set of reduced words in $(A \cup \overline{A})^*$ by the name $R(A)$.

2.3.2. Theorem [Uniqueness of Reduced Words]**.** *In every equivalence class of elements in $F(A)$, there is a unique reduced word.*

In other words, no matter how you get from one word to another by a sequence of insertions and deletions of $x\overline{x}$, if both words are reduced, then they happen to be equal as words. This very basic, very elementary result has several proofs, which shine light on the theory of groups in different directions.

In many of these proofs the following remark is useful:

2.3.3. Lemma. *Suppose that it is a fact that the only reduced word equivalent to the empty word is the empty word. Then the theorem follows.*

This has to do with the "groupness" of the situation: If $u \approx v$, then $uv^{-1} \approx \Lambda$, and so uv^{-1} is not reduced. However, if both u and v are reduced, the only thing that could happen to cause uv^{-1} not to be reduced is that the last letter of u and the last letter of v are equal. Remove this letter to get u' and v' two reduced words whose lengths are less, such that $u'(v')^{-1} \approx \Lambda$, which implies that $u' \approx v'$. And so, by induction on the minimum of the lengths of u and v, it would follow that $u' = v'$ and thus $u = v$.

2.3.4. **Mountain-leveling proof.**

This is an instance of something that happens often in PL topology and in certain kinds of group theory. M.H.A. Newman used this a bit.

Suppose that $u \approx v$ and that u and v are reduced words in $R(A)$. Thus, there is a sequence of intermediate words: $w_0, w_1, \ldots, w_n$ in $(A \cup \overline{A})^*$, such that $w_0 = u$, $w_n = v$, and such that w_{i-1} and w_i are related by one insertion or deletion of a word of the form $x\overline{x}$. If $n = 0$, then the desired result is true, that $u = v$.

Consider the lengths $|w_0|, \ldots, |w_n|$. Of these numbers, there is a maximum one; of the w_i with $|w_i|$ maximum, there is a largest subscript i. (If $|w_i| < |v|$ then v would not have been reduced.) It then follows that w_{i+1} is obtained by cancelling out of w_i a subword of the form $x\overline{x}$; and that w_{i-1} is similarly obtained by cancelling a subword of the form $y\overline{y}$ from w_i. If the instances of $x\overline{x}$ and $y\overline{y}$ are exactly the same or if they overlap, simply delete from the sequence $w_0, \ldots, w_n$ two terms, the ith and the $(i+1)$th, since $w_{i+1} = w_{i-1}$. If the instances are disjoint, then, first, from w_{i-1}, cancel out the $y\overline{y}$ instance, to get $(w_i)'$, and then in that word insert the $x\overline{x}$ instance, getting w_{i+1}. This process, of modifying the sequence of intermediate words, diminishes some complexity of the picture, which complexity takes values in some well-ordered set; and thus it must happen eventually that the case occurs where the picture of the lengths of the intermediate terms has no peak; since the endpoints are supposed to be reduced, this can only happen when $n = 0$, and the proof is finished.

As for the "complexity" which is reduced, various things work. The most elementary one is to define the complexity of the sequence $w_0, \ldots, w_n$ to be the ordered pair (k, ℓ), where k is the maximum of the lengths $|w_i|$, and where ℓ is the largest subscript such that $|w_\ell| = k$. The process above, in each case, either reduces ℓ while keeping k fixed, or else it reduces k. Thus, regarding complexities as belonging to the lexicographic product of the set of non-negative integers with itself, the complexity is reduced. This kind of complexity occurs in several other instances in group-theory. — However, in this special case, we could simply have taken the complexity to be the non-negative integer $|w_0| + \cdots + |w_n|$, and this would have been quite as good.

2.3.5. Artinoid proof.

Van der Waerden credits E. Artin with a proof (of a more general theorem) along the following lines:

For each $x \in A \cup \overline{A}$, define a function $\lambda_x : R(A) \to R(A)$ as follows: Let $u \in R(A)$. Recall that $R(A)$ is the set of reduced words. (a) If the leftmost letter of u is not $\overline{x}$, define $\lambda_x(u) = xu$. (b) If, on the other hand, $u = \overline{x}v$, then define $\lambda_x(u) = v$. Now, it is easy to calculate that for every $u \in R(A)$, and every $x \in A \cup \overline{A}$, it is the case that $\lambda_x(\lambda_{\overline{x}}(u)) = u$. Thus, $\lambda_{\overline{x}}$ is the inverse function to λ_x; and so λ_x belongs to the group of permutations of the set $R(A)$.

In other words, λ is a set-map from the set $A \cup \overline{A}$ into the group G of permutations of the reduced words $R(A)$. Therefore, by the universal property of free groups, it extends to a unique group-map $\lambda : F(A) \to G$. If the word $w \in (A \cup \overline{A})^*$ is written as $w = x_1 \cdots x_n$ for $x_i \in A \cup \overline{A}$, then λ_w is the composition of the functions $\lambda_{x_1}, \ldots, \lambda_{x_n}$. Now, suppose that an element α of $F(A)$ is represented by the reduced word w. Then $\lambda_\alpha(\Lambda) = \lambda_{x_1}(\cdots(\lambda_{x_n}(\Lambda))\cdots)$, and this latter expression is always computable using case (a) above, and comes out to be w itself. Thus $\alpha \in F(A)$ determines λ_α which determines $\lambda_\alpha(\Lambda) = w$, which is therefore the unique reduced word equivalent to w. This proof is finished.

2.3.6. Matrix proof.

Schottky, Macbeath, residual finiteness

This sort of proof is due to Schottky, Macbeath, and others.

For instance, look at $F(a, b)$, the free group on a two-element set.

It is first necessary to construct a set Y which can be written as the union of four sets Y_a, Y_b, $Y_{\overline{a}}$, $Y_{\overline{b}}$, such that no three of these are contained on any one of them. And then to construct permutations α and β of Y such that:

$$\alpha(Y_a \cup Y_b \cup Y_{\overline{b}}) \subset Y_a$$
$$\beta(Y_a \cup Y_{\overline{a}} \cup Y_b) \subset Y_b$$
$$\alpha^{-1}(Y_{\overline{a}} \cup Y_b \cup Y_{\overline{b}}) \subset Y_{\overline{a}}$$
$$\beta^{-1}(Y_a \cup Y_{\overline{a}} \cup Y_{\overline{b}}) \subset Y_{\overline{b}}$$

In such a situation, then any reduced word w of length ≥ 1 in $\{a, b, \overline{a}, \overline{b}\}$ will produce an expression (by going over to the Greek alphabet) as compositions of $\alpha^{\pm 1}, \beta^{\pm 1}$. The set of three terms corresponding to the rightmost letter of w will be mapped by this composed map into the single term corresponding to the leftmost letter of w; and so this reduced word w cannot represent the identity in the free group $F(a, b)$.

A specific example of this can be constructed by taking Y to be the plane as represented in Cartesian coordinates. Divide the plane into eight 45° regions bounded by the two coordinate axes and the coordinate axes rotated by 45°. The set Y_a is the union of two such sectors which are symmetric about the origin; and similarly for the other three sets. You will have to figure out how to label these, so that the following works:

The first choice of α and β will be given by the matrices

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$$

and another choice which works with the same division of the plane but with different labels is

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

[Residual finiteness]. A group G is "residually finite", when, for every $g \in G$, if $g \neq 1$, then there exists a normal subgroup of finite index in G, call it N, such that the image of g in the finite group G/N is unequal to the identity element.

Every subgroup of finite index in G contains a normal subgroup of finite index. Thus, another way to describe residual finiteness is to say that the intersection of all subgroups of finite index in G is $\{1\}$.

Every subgroup of a residually finite group is residually finite. In this section, it was shown that a free group (with two generators) is contained in a group of matrices with integer entries. By mapping the integers into the integers mod n for larger and larger n, it follows that this matrix group is residually finite, and therefore the free group also is residually finite.

2.3.7. **Infinite series.**

power series in non-commuting variables, residually a p-group

This idea is due to W. Magnus.

Let R denote one of the rings Z (the integers) or Z_p (the integers mod p). Let A be an alphabet. Construct the ring of formal power series in the non-commuting variables A with coefficients in R. Call this Γ.

The element $1 + a$ has formal inverse $1 - a + a^2 - a^3 + a^4 - \cdots$; thus, such elements are units in this ring and generate a group under multiplication. Thus, there is a function $\phi : F(A) \to \Gamma$, which is a homomorphism into the group of units of the ring, given by $\phi(a) = 1 + a$.

First look at $\phi(a^n)$, for n a positive integer. This starts with "1" and ends with "a^n". Thus, there is a lowest positive integer k such that the coefficient of a^k in $(1 + a)^n$ is non-zero; call this integer $k(n)$. (In the characteristic zero case, $k(n) = 1$ for all n; but in characteristic p, it depends on the largest power of p dividing n.) Since $(1 + a)^{-n}$ is the inverse of $(1 + a)^n$, it is easy to see that the lowest positive integer for which the coefficient of a^k in $(1 + a)^{-n}$ is non-zero is the same, $k(n) = k(-n)$.

Now, any reduced word in $A \cup \overline{A}$ can be written, by collecting the same letters next to each other together, and writing a^{-1} for $\overline{a}$, in the form $w = a_1^{n_1} a_2^{n_2} \cdots a_\ell^{n_\ell}$, where $a_i \in A$, n_i is a non-zero integer, and where, for $i = 2, \ldots, \ell$, it is the case that $a_{i-1} \neq a_i$.

Now look at $\phi(w)$ and compute the coefficient in this expression of the monomial $a_1^{k(n_1)} a_2^{k(n_2)} \cdots a_\ell^{k(n_\ell)}$. This turns out to be the product of the coefficients of $a_i^{k(n_i)}$ which occur in $(1 + a_i)^{n_i}$; these terms are non-zero and their product is non-zero since the coefficient ring has no zero-divisors.

Thus, if w is a non-empty reduced word, the element $\phi(w)$, which depends only on the element of the free group $F(A)$ represented by w, is not equal to 1. End of proof.

[**More residual finiteness**]. Consider the case where the prime p occurs, so that the above argument involves non-commuting infinite series with coefficient group Z_p. For any group G, consider the subgroup $G(n;p)$ which is recursively defined: $G(1;p) = G$, and $G(n+1;p)$ is generated by all commutators $[w,u] = wuw^{-1}u^{-1}$, with $w \in G$ and $u \in G(n;p)$, together with all pth powers u^p, for $u \in G(n;p)$. Then $G(n;p)$ is a normal subgroup of G; if G is the free group $F(A)$, with finite basis A, it is possible to find a finite bound on the size of $G/G(n;p)$ and to prove that this quotient group has order some power of p. Furthermore, under the homomorphism above of G into the group of units of the power-series ring Γ, the image of the group $G(n;p)$ consists of series starting with 1 and having no terms of degrees $< n$. Thus, when Γ is factored by I^n, where I is the two-sided ideal generated by A, the image of $G(n;p)$ is trivial. This implies that the intersection $\bigcap_{n=1}^{\infty} G(n;p) = \{1\}$. In other words, a free group with finite basis is not only residually finite, it is residually a finite p-group; this is true for every prime number p.

2.4. Basis

In the free monoid A^*, the elements A are algebraically determined. A is, for instance, the unique smallest subset of A^* which generates the entire monoid A^*. Thus, A is *the* free basis of A^*.

In the free group $F(A)$, we have the canonical set-map $A \to F(A)$. This is easily seen to be injective, and the image is identified with A. Call A a free basis of $F(A)$. Now, it happens in many circumstances that there is a group-equivalence $F(A) \to F(B)$. The image of A under this group-equivalence is another "free basis of $F(B)$". In other words, a free basis X of a group G can be defined to be a subset X of G, such that the homomorphism defined by the set map $X \to G$, giving $F(X) \to G$, is an isomorphism of groups.

There are various tricks to prove that the cardinal number of elements in a basis of a given free group F is algebraically determined by F. There are combinatorial proofs of various kinds. However this (not particularly combinatorial) one is easy to verify, although not particularly interesting: If one looks at the subgroup of F generated by all squares, that is, by all elements of the form ww for $w \in F$, then this forms a normal subgroup $N \triangleleft F$. The quotient group F/N is a vector space over the field of 2 elements and has dimension equal to the number of elements in a free basis of F. The "rank of F" is defined to be the number of elements in a free basis of F. Summarizing the above discussion:

$$\operatorname{rank}(F) = \log_2 ||F/F^2||.$$

2.5. The Category of Graphs

graph, vertex, edge, initial and terminal vertex, interval, loop, directed graph, map of graphs

The word "graph" has many meanings. One is simply a symmetric, irreflexive, binary relation on a set; the set consists of the "vertices" or "nodes", and two vertices are connected by an "edge" when they are related by the given binary relation. If one has, instead, an antisymmetric relation (so that whenever aRb, then it is not the case that bRa), this gives the notion of a "directed graph"; and the edge joining a and b is said to start at a and end at b, when aRb. It may be that one wants to allow several edges at once with the same pair of vertices, and that one wants to allow an edge to start and end at the same vertex; this is a more pictorial concept, for which graph theorists use words such as "multigraph". Serre's notion of graph is that of a multigraph in which an edge can be directed either way. Gersten came up with a type of graph which is close to that of Serre, but in which the concept of a map of graphs allows the collapsing of edges. What we shall do here is to define the Serre type of graph.

Definition. A **graph** (in Serre's sense) is a 4-tuple $\Gamma = (V(\Gamma), E(\Gamma), s_\Gamma, i_\Gamma)$, where $V(\Gamma)$ and $E(\Gamma)$ are sets, $s_\Gamma : E(\Gamma) \to V(\Gamma)$ is a function, and $i_\Gamma : E(\Gamma) \to E(\Gamma)$ is a function, satisfying the axiom: For all $e \in E(\Gamma)$, it is the case that $i_\Gamma(i_\Gamma(e)) = e$ and that $i_\Gamma(e) \neq e$.

The subscripts and the "(Γ)" may be omitted when we hope the context makes things clear.

The set V is the set of **vertices** of Γ. The set E is the set of **edges**. The vertex $s(e)$ is the **initial** or **starting** vertex of e. And $i(e)$ is the **reverse** or **inverse** of e. The function i essentially is a free action of the group of order two on the set E. Define the **terminal** vertex of e to be $t(e) = s(i(e))$. The inverse edge to e will often be denoted by a bar over the symbol for that edge, so that $\overline{e} = i(e)$.

An **interval** in a graph Γ consists of an edge e whose two vertices are distinct, together with those vertices. A **loop** is an edge e having its terminal and initial vertices equal.

A **directed graph**, as used earlier to define limit and colimit in a category, is basically the same thing as a Serre graph in which, out of each pair $\{e, i(e)\}$, one element is chosen.

A **map of graphs** $f : \Gamma \to \Delta$ consists of two functions $f_V : V(\Gamma) \to V(\Delta)$ and $f_E : E(\Gamma) \to E(\Delta)$, such that, for all $e \in E(\Gamma)$, it is the case that $f_V(s_\Gamma(e)) = s_\Delta(f_E(e))$ and $f_E(i_\Gamma(e)) = i_\Delta(f_E(e))$. A graph is a set and some functions among these sets satisfying certain axioms; this is the sort of thing which is considered in the subject of "universal algebra"; and then a map is a collection of functions between the sets making commutative diagrams of sets.

This then defines the **Category of (Serre) graphs**.

2.5.1. **Local properties of graph maps.**

link, locally injective, locally surjective, immersion, submersion, covering

When v is a vertex of the graph Γ, define the **link** of v in Γ thus: $\mathrm{Lk}_\Gamma(v) = \{\, e \in E : s(e) = v \,\}$.

A map of graphs $f : \Gamma \to \Delta$ takes $\mathrm{Lk}_\Gamma(v)$ to $\mathrm{Lk}_\Delta(f(v))$.

A graph map f is said to be **locally injective** (resp., **locally surjective**, **locally bijective**), when, for every vertex v, the function $\mathrm{Lk}_\Gamma(v) \to \mathrm{Lk}_\Delta(f(v))$ is injective (resp., surjective, bijective). There are other words for these concepts. A locally injective map is an **immersion**; a locally surjective map is a **submersion**; a locally bijective map is a **covering** or **covering projection**.

2.5.2. **Folds, pushouts.**

fold, nonsingular fold

The Category of graphs admits various category-theoretic constructions, limits, colimits, etc. One kind of pushout is known as a **fold**. Consider a graph A which has three vertices a, b, c and two pairs of edges $(x, \overline{x}), (y, \overline{y})$, so that $s(x) = s(y) = a$, and $t(x) = b$, $t(y) = c$. This maps to the graph B by identifying x to y, $\overline{x}$ to $\overline{y}$, and b to c. The graph B is just an interval with two vertices and one edge pair between them. We have $\phi : A \to B$. Now, given any graph Γ, there may be a map of graphs $f : A \to \Gamma$. This is the same as choosing in Γ a vertex $v = f(a)$ and two edges $e_1 = f(x)$ and $e_2 = f(y)$ having the same initial vertex v. The pushout of this diagram will exist if and only if $e_1 \neq \overline{e_2}$, and it consists of the graph Γ with e_1 identified to e_2 (and their inverses identified also), and with $t(e_1)$ identified to $t(e_2)$.

$$\begin{tikzcd} B \arrow[dr, "f'"'] & A \arrow[l, "\phi"'] \arrow[r, "f"] & \Gamma \arrow[dl, "\phi'"] \\ & \Gamma' & \end{tikzcd}$$

The map $\phi' : \Gamma \to \Gamma'$ thus defined is generally called a **fold**, said to be gotten by folding e_1 to e_2. The fold is called **nonsingular**, when $t(e_1) \neq t(e_2)$; this includes the case of two intervals with distinct terminal vertices folding together, and the case where an interval folds onto a loop. The other case, the singular situation, would involve the case where two intervals with the same pair of vertices fold together, and the case where two loops with the same vertex fold together.

2.5.3. **Pullbacks.**

Given a pair of graph-maps $\alpha : \Gamma_1 \to \Delta$ and $\beta : \Gamma_2 \to \Delta$, we can look at the pullbacks of the sets of vertices and edges. We get

$$V(P) = \{\, (e_1, e_2) \in V(\Gamma_1) \times V(\Gamma_2) \, : \, \alpha_V(e_1) = \beta_V(e_2) \,\}$$

and similarly for $E(P)$. There are obvious functions defining initial vertex and inverse edge on these sets, making P into a graph, and yielding maps of graphs

$\alpha' : P \to \Gamma_2$ and $\beta' : P \to \Gamma_1$.

$$\begin{tikzcd} & P \arrow[dl, "\beta'"'] \arrow[dr, "\alpha'"] & \\ \Gamma_1 \arrow[r, "\alpha"'] & \Delta & \Gamma_2 \arrow[l, "\beta"] \end{tikzcd}$$

This is a pullback diagram in the category of graphs.

2.6. Paths and the fundamental group

arc, circle, path in a graph, length of path, homotopy of paths, concatenation of paths, reduced path, reverse path, fundamental group of a graph

The **arc** of length n, where n is a non-negative integer, is the graph with vertices $\{0, 1, \ldots, n\}$, and n pairs of edges joining $i-1$ and i, for $i = 1, \ldots, n$. The **circle** of length n is the graph obtained from the arc by identifying the two vertices 0 and n.

Paths, homotopy. A **path** in a graph Γ of **length** n, is a graph-map of the arc of length n into Γ. For $n > 0$, this can be determined by an n-tuple of edges of Γ, say $(e_1, \ldots, e_n)$, such that for all $i = 2, \ldots, n$, it is the case that $t(e_{i-1}) = s(e_i)$. The case of length 0 is slightly peculiar, and a path of length 0 is the same, essentially, as a vertex. We say that the path p starts at $s(p) = p(0)$, and terminates at $t(p) = p(n)$ when the length of p is n. Given paths p and q of lengths m and n, if p terminates at the vertex at which q starts, then we can **concatenate** them, forming the path of length $m + n$ denoted by pq. If $p = ue\overline{e}v$, then uv is a path, said to be obtained from p by an elementary reduction. The equivalence relation on paths generated by elementary reduction is called **homotopy**. In the anomalous case that $p = e\overline{e}$, then the reduction produces a path of length zero that corresponds to the vertex $s(e)$. A path to which no elementary reduction can be applied is called a **reduced path**. There is a map of the arc of length n to itself which takes each vertex k to the vertex $n - k$. The composition of this with a path p of length n is called the **reverse** of p and is denoted by $\overline{p}$.

Fundamental group. A graph Γ together with a chosen vertex $v_0 \in V(\Gamma)$, will be called a graph with **basepoint** v_0. The **fundamental group of** Γ **based at** v_0 consists of all homotopy classes of paths in Γ which start and terminate at v_0; concatenation is defined between all such homotopy classes, thus producing a binary operation on this set; the path of length 0 is a 2-sided identity element; the homotopy class of $\overline{p}$ is the inverse of the homotopy class of p. Therefore, this is in fact a group. It is a functor from the Category of graphs with basepoint to the Category of groups. Because this is very similar to the topological concept of fundamental group defined by Poincaré, and later extended to higher dimensions by Hurewicz, this fundamental group is symbolically denoted $\pi_1(\Gamma, v_0)$. If Γ has only one vertex, or if the basepoint is otherwise clear from context, the symbol $\pi_1(\Gamma)$ may be used.

2.6.1. **Forests and trees.**

forest, tree

A **forest** is a graph Γ, such that, for every pair of vertices v, w, there is at most one reduced path that starts at v and terminates at w. A **tree** is a connected forest. Each component of a forest is a tree. If Γ is not a forest, then there is a subgraph of Γ which is graph-equivalent to a circle of some length $n > 0$. Given any totally ordered (by inclusion) set of subforests of Γ, its union is a forest. If T_1 and T_2 are two disjoint subtrees of Γ, and e is an edge of Γ with $s(e) \in T_1$ and $t(e) \in T_2$, construct the union of T_1 and T_2 together with the edges e and $\overline{e}$; this larger subgraph of Γ is a tree. If every component of a graph Γ is a tree, then Γ is a forest.

2.6.2. Theorem. *In every graph Γ there is contained a maximal subforest F; every maximal forest in Γ contains all the vertices of Γ; when Γ is connected, every maximal forest is a tree.*

Proof. By Zorn's Lemma, since every union of a totally ordered set of forests is a forest, there exists a maximal forest. If a subforest F does not contain a vertex v, then the union of F with the singleton vertex $\{v\}$ is a larger forest. If Γ is connected, and T is a component of a maximal forest F, and if v were a vertex in another component of F, then there would be a path p from a vertex w in T to the vertex v. The first vertex in p which does not belong to T will belong to another component T' of F, and will involve an edge joining T to T'; the union of these two subtrees with that edge will be a larger tree. Thus adding that edge to the forest F, all the components of the new graph will be trees, so that F was not maximal. $\square$

"Rewriting" and free fundamental group. Let Γ be a connected graph, with base vertex v_0. Let T be a maximal tree in Γ. For every vertex w of Γ, let $\tau(w)$ be the <u>unique</u> reduced path in T from v_0 to w. From each edge e, let $p(e) = \tau(s(e))e\overline{\tau(t(e))}$. Then $p(e)$ is a closed path based at v_0. Suppose that q is any closed path based at v_0, which goes across the edges in order: $q = e_1 e_2 \cdots e_n$. Then q is homotopic to the path $p(e_1)p(e_2)\cdots p(e_n)$.

Now, if e is an edge of the maximal tree T, then $p(e)$ is homotopic to the path of length 0 at v_0. If e is another edge, then $p(\overline{e}) = \overline{p(e)}$. From each edge pair $\{e, \overline{e}\}$ not in T, choose one. Let E be the set of these chosen edges. Let u be a reduced word in $E \cup \overline{E}$, so that u represents an element of the free group $F(E)$. Suppose that $u = e_1 e_2 \cdots e_n$ (where $e_i \in E \cup \overline{E}$), and let $\epsilon(u) = p(e_1)\cdots p(e_n)$. Then the length of the reduced path homotopic to $\epsilon(u)$ is at least as great as the length of u in $\{E \cup \overline{E}\}^*$. Conclude that the fundamental group of Γ based at v_0 is isomorphic to $F(E)$. Given a path between v_0 and v_1, define an isomorphism from $\pi_1(\Gamma, v_0)$ to $\pi_1(\Gamma, v_1)$, which depends on the given path. $\diamondsuit$

2.6.3. **Path lifting.**

Given a graph-map $f : \Delta \to \Gamma$, and base vertices v_0 in Δ, w_0 in Γ, with $f(v_0) = w_0$: Consider a path p in Γ which starts at w_0; it is said to **lift** to a path

q in Δ starting at v_0, when both paths have the same length, and thus are maps of the arc A_n into the respective graphs, and when it is the case that $fq = p$.

2.6.4. **Covering corresponding to a subgroup.** Let Γ be a 1-vertex graph, with fundamental group F, free with a basis consisting of chosen orientations of the edges. Let S be a subgroup of F. Construct a graph $\Delta = \Delta(S)$, as follows:

The vertices of Δ are the right cosets $\{Sw\}$ of S in F.

An edge of Δ is determined by a right coset Sw and an edge e of Γ. Denote this edge by (Sw, e).

The starting vertex of (Sw, e) is Sw; the terminal vertex is Swe.

The inverse edge to (Sw, e) is $(Swe, \overline{e})$.

Furthermore, there is a map of graphs $f : \Delta \to \Gamma$ which takes (Sw, e) to e. This map f is locally bijective; taking the base vertex of Δ to be the base coset $S1$, the image by f of $\pi_1(\Delta)$ is S. Call Δ the **covering of Γ corresponding to** S.

2.6.5. **The core of a graph.** Let Γ be a connected graph with base vertex v_0. It may well happen that some proper subgraph of Γ has the same fundamental group as that of Γ; more precisely, say that a subgraph Δ of Γ "has the same π_1", when: (a) v_0 is a vertex of Δ; (b) Δ is connected; and (c) the inclusion maps $\pi_1(\Delta, v_0)$ onto $\pi_1(\Gamma, v_0)$ (so that, since the inclusion is locally injective, the two fundamental groups are isomorphic).

The intersection of any family of such subgraphs with the same fundamental group is itself such a subgraph; therefore, there is a minimum subgraph of Γ which has the same fundamental group as Γ. This is called the **core** of Γ relative to the base vertex v_0. The core can be described intrinsically as the subgraph consisting of those vertices and edges which occur in some reduced closed path of Γ based at v_0. The graph Γ consists of its core, together with trees attached to the vertices.

2.7. Representation of subgroups by immersions

2.7.1. **Behavior of folding on fundamental group.** Consider a map of graphs $f : \Gamma \to \Delta$. If f is not locally injective, then there are two edges in Γ which have a common starting vertex and which map to the same edge in Δ. This image edge e is unequal to $\overline{e}$, and therefore the two edges being identified are not inverses of each other. Thus, the map f factors through a fold, as described in 2.5.2; call this fold $\phi : \Gamma \to \Gamma'$; thus, there is a map $f' : \Gamma' \to \Delta$, such that $f'\phi = f$.

2.7.2. Lemma. *A fold ϕ is surjective on fundamental groups. That is, if $\phi : \Gamma \to \Gamma'$ is a fold, then $\phi_* : \pi_1(\Gamma) \to \pi_1(\Gamma')$, with consistent base vertices, is surjective. Furthermore, if the fold is nonsingular (i.e., the fold identifies edges e_1 and e_2 with $s(e_1) = s(e_2)$, and $t(e_1) \neq t(e_2)$), then ϕ is an isomorphism on fundamental groups. And if the fold is singular (i.e., not nonsingular), ϕ has non-trivial kernel on the fundamental group level.*

2.7.3. Theorem. *If $f : \Gamma \to \Delta$ is a graph map, where Γ is a finite graph, then there is a finite sequence of folds $\phi_i : \Gamma_{i-1} \to \Gamma_i$, where $\Gamma = \Gamma_0$, ending with $\Gamma' = \Gamma_n$, and a graph map $f' : \Gamma' \to \Delta$, such that $f = f'\phi_n\phi_{n-1}\cdots\phi_1$ and such that f' is locally injective. The image of $\pi_1(\Gamma')$ in $\pi_1(\Delta)$ is equal to the image of $\pi_1(\Gamma)$.*

Proof. Each fold reduces the number of edges in Γ_i, and since this number is finite, the possibility of folding must terminate. If there is no possible fold, the end result map into Δ is locally injective. By the Lemma 2.7.2, each fold is surjective on π_1, and therefore the images of fundamental groups in the fundamental group of Δ are all the same. $\square$

Thus, one can say that every finitely generated subgroup S of the free group $F = \pi_1(\Delta)$ (suppose that Δ is a 1-vertex graph), can be **represented by an immersion** of a finite graph Γ'. The idea is that one starts with a set of generators $\{s_i\}$ of S; there is a 1-vertex graph B whose edges can be thought of as the given set of generators. Subdivide the loop of B corresponding to s_i into a circle of length equal to the length of s_i considered in the basis of F corresponding to the edges of Δ. Call the resulting subdivided B by the name Γ; choosing a maximal tree in Γ, one can identify the fundamental group of Γ with that of B; there is a graph map $f : \Gamma \to \Delta$, so that the image of the fundamental group is S. Then factor f through a finite sequence of folds and an immersion $f' : \Gamma' \to \Delta$. We say that f' is an immersion that represents S, since the image of the fundamental group is still S, by the theorem 2.7.3.

2.7.4. Theorem. *Suppose that Δ is a 1-vertex graph, which therefore has $\pi_1(\Delta) = F$ free on a particular basis. Suppose that S is any finitely generated subgroup of F. Then S is the π_1-image of an immersion $\alpha : \Gamma \to \Delta$, where Γ has a base vertex v_0. The following are equivalent:*

(a) $\alpha : \Gamma \to \Delta$ is an immersion, the image of $\pi_1(\Gamma, v_0$ is S, andΓ is a core graph with respect to v_0. (That is, Γ has no proper connected subgraph containing v_0, whose image in Δ is S.)

(b) Γ is a connected graph with $\alpha : \Gamma \to \Delta$, such that the image of $\pi_1(\Gamma, v_0)$ is S, and such that of all such graphs and maps α with this property, the number of edges in Γ is the minimum.

Furthermore, if $\alpha' : \Gamma' \to \Delta$ is another situation with these properties, then there is an equivalence of graphs $\phi : \Gamma \to \Gamma'$, taking base vertex v_0 of Γ to v_0' of Γ', such that $\alpha'\phi = \alpha$. (In other words, α and Γ are uniquely determined by S and the minimality condition (a) or (b).)

2.7.5. Howson's Theorem. *If A and B are finitely generated subgroups of the free group F, then $A \cap B$ is finitely generated.*

Proof. Represent this picture by graphs, so that we have $F = \pi_1(\Delta)$, A is the image of an immersion $f : \Gamma_A \to \Delta$, and B is the image of an immersion $g : \Gamma_B \to \Delta$. The graphs Γ_A and Γ_B are finite graphs.

There is a pullback graph (2.5.3), which will be called Γ_C. Implicit in the situation there are base vertices $v_A \in \Gamma_A$, and $v_B \in \Gamma_B$, whose images in Δ are equal. Thus, there is a vertex v_C in the pullback Γ_C which maps to both these vertices, and which will be used for the base vertex of Γ_C. Now, suppose that $c \in A \cap B$. This can be represented by reduced closed paths in each of Γ_A , Γ_B, and Δ; call these, respectively, p_A, p_B, and p. Since $f : \Gamma_A \to \Delta$ is an immersion, it follows that fp_A is a reduced closed path in Δ; and it is homotopic to the reduced closed path p; therefore it is exactly equal to p. Similarly, $gp_B = p$. By the pullback property, categorically, there is p_C, a closed based path in Γ_C, which maps to both p_A and p_B. This shows that c is in the image of $\pi_1(\Gamma_C)$ in $\pi_1(\Delta)$.

What this does is to verify that $A \cap B$ is represented by the given map of Γ_C to Δ. This map is in fact an immersion itself. The pullback graph Γ_C might not be connected, and so essentially what one wants is to consider only the component of Γ_C that contains the base vertex v_C. Now, Γ_C is a finite graph, being the pullback of two finite graphs is itself finite, and therefore its fundamental group is finitely generated. $\square$

2.8. Cancellation Bound

Let $f : \Gamma \to \Delta$ be a graph map. A number n is called a **cancellation bound** on f, when: For any two paths p, q in Γ, if p and q are reduced paths, and pq can be defined and is also reduced, and if fp^* and fq^* denote the reduced paths in Δ obtained from fp and fp by homotopies in Δ, then the amount of cancellation in the product path $(fp^*)(fq^*)$ in Δ is at most n. In other words, if $|fp|$ denotes the length of the reduced path in Δ homotopic to fp, then

$$|fp| + |fq| - |f(pq)| \leq 2n.$$

2.8.1. Lemma. *If $f : \Gamma_1 \to \Gamma_2$ has n as a cancellation bound, and $g : \Gamma_2 \to \Gamma_3$ has m as a cancellation bound, then $n+m$ is a cancellation bound for $gf : \Gamma_1 \to \Gamma_3$.*

Proof. Let p, q be reduced paths in Γ_1, such that pq is reduced. In other words, $|p| + |q| - |pq| = 0$. Because f has cancellation bound n, one can write the reduced words homotopic to fp etc., thus: $fp^* = p_1 r$ and $fq^* = \overline{r} q_1$, where p_1, q_1, and $p_1 q_1$ are all reduced paths in Γ_2, and where $|r| \leq n$. Then $gf(pq)$ is homotopic to $g(p_1 q_1)$, and the amount of cancellation between $g(p_1)^*$ and $g(q_1)^*$ is bounded by m. That plus the length of r is a bound on the amount of cancellation between $fg(p)$ and $fg(q)$. $\square$

2.8.2. Lemma. *Let $f : \Gamma \to \Gamma'$ be a nonsingular fold. Then a cancellation bound for f is* 1.

Proof. The fold will fold e_1 to e_2, where e_1 and e_2 have equal start vertices but different terminal vertices. If p does not terminate at the start vertex of e_1, and pq is defined, then to get the reduced version of fp^* amounts to cancelling $\overline{e_1} e_2$ and $\overline{e_2} e_1$ in the word for p in terms of Γ edges; the rightmost letter of fp remains unchanged; in this case, there is zero cancellation between fp^* and fq^*. Otherwise, to get the reduced version of fp^*, it may end on the right with one of the edges $\overline{e_1}$ or $\overline{e_2}$; and fq^* may start with e_1 or e_2. Because the fold is non-singular, one of the edges being folded, at least, is a segment and not a loop; if that is e_1, then it does not appear in any path twice in a row; since p and q cancel a zero amount, if there is any cancellation at all between fp^* and fq^*, then at most one edge can be cancelled. □

2.8.3. Theorem. *Suppose that $\phi : F \to G$ is a homomorphism of free groups with specified bases; suppose that F is finitely generated, and that ϕ is an injective homomorphism. Then ϕ has some cancellation bound.*

Proof. Represent G by a 1-vertex graph Δ and F by an map $f : \Gamma \to \Delta$. The graph Δ is made from edges in one-to-one correspondence with the given basis of G. The graph Γ is a subdivided version of the same thing with respect to the given basis of F. The concept of a "cancellation bound for ϕ" can be defined so that it is bounded by the cancellation bound for the map f. Now, by 2.7.3, the map f factors into a sequence of n folds and an immersion. Because of the fact that the map $\pi_1(\Gamma) \to \pi_1(\Delta)$ is injective, it follows (2.7.2) that each fold is nonsingular. An immersion, obviously, has a cancellation bound of zero. Each nonsingular fold has a cancellation bound of 1, by 2.8.2. Adding these up, by 2.8.1, the consequence is that f and hence ϕ has n as a cancellation bound. □

2.9. Finite generation of the fixed subgroup

In 1982, Gersten came up with a proof of the theorem below that the fixed subgroup of an automorphism of a finitely generated free group is itself finitely generated. Cooper found a proof independently. Both these proofs waited a long time to be published. Goldstein and Turner came up with their own proofs, one of which will be given in this section. The conjecture was that the rank of the fixed subgroup is bounded by the rank of the larger group; and this was proved by Bestvina and Handel, using quite different methodology.

Given a group G and an endomorphism $\alpha : G \to G$, define

$$\mathrm{Fix}(\alpha) = \{ g \in G : \alpha(g) = g \}.$$

This is the **fixed subgroup** of α.

2.9.1. Theorem. *Let F be a finitely generated free group. Let $\alpha : F \to F$ be an injective endomorphism of F. Then $S = Fix(\alpha)$ is finitely generated.*

Proof. Let F be represented as $\pi_1(\Gamma)$, where Γ is a finite 1-vertex graph. The subgroup S corresponds to a covering $h : \Delta \to \Gamma$. There is a certain base vertex of Δ, call it v_0. Let Δ' be the core of Δ, that is, the smallest subgraph carrying all the fundamental group of Δ, and containing v_0.

For each vertex w of Δ, and hence as well each vertex of Δ', define a label $\lambda(w)$ to be as follows: Choose a path p in Δ from v_0 to w; this represents a closed path, also called p, in Γ, and thus an element $p \in F$. Define $\lambda(w) = \alpha(p)^{-1}p$. If q is another choice of such a path from v_0 to w, then pq^{-1} is a closed path based at v_0 and thus belongs to $\mathrm{Fix}(\alpha)$. In other words,

$$\alpha(pq^{-1}) = pq^{-1}, \qquad \alpha(p)\alpha(q)^{-1} = pq^{-1}, \qquad \alpha(q)^{-1}q = \alpha(p)^{-1}p.$$

Thus the labeling λ is independent of the choice of paths p. Furthermore, if $\lambda(w) = \lambda(w')$, and p and p' are the choices of paths for w and w', then:

$$\alpha(p)^{-1}p = \alpha(p')^{-1}p', \qquad p'p^{-1} = \alpha(p'p^{-1}), \qquad p'p^{-1} \in \mathrm{Fix}(\alpha).$$

This shows that $w = w'$, since the vertices w of Δ are in one-to-one correspondence with the right cosets of S in F. Thus, the labeling λ is well-defined and each vertex of Δ has a unique label.

The claim now is that the number of vertices of Δ' of valence more than 2 is finite. And the valence of any such vertex is bounded by the valence of the vertex of Γ, which is itself finite. The rank of $\pi_1(\Delta') = \mathrm{Fix}(\alpha)$ can be determined and bounded by these numbers, and thus will be finite.

This claim is based on the fact (2.8.3) that α has a finite cancellation bound k. Let w be a vertex of Δ' of valence greater than 2; then there exist closed paths based at v_0, which both take the same path to w, and then diverge; call these paths r and s, and their common left segment from v_0 to w, call p. Thus, $r = pR$, and $s = pS$, with no cancellation of paths in these products, and furthermore, $r^{-1}s = R^{-1}S$, the latter product having zero cancellation as well. Then, $\alpha(r) = r$ and $\alpha(s) = s$, which are reduced words when considered in the free group F with its given basis (the reason being that $\Delta' \to \Gamma$ is locally injective), and $\alpha(rs^{-1}) = rs^{-1}$, but the reduced word for this is some reduction of RS^{-1}. Thus the amount of cancellation occurring in this product, under α, is at least the length of $\alpha(p)$. Since α is injective (and F is finitely generated), the set of those p for which the length of $\alpha(p)$ is at most k is finite. This is then a bound on the number of vertices of Δ' of valence more than 2. $\square$

2.10. Exercises and Examples

1. Look up how you show that addition in the positive integers is associative, given a recursive definition of addition in terms of the successor function. Then show that concatenation in A^* is associative.

2. If $\sim$ is a binary relation on M, in 2.2.1, there is the definition of the congruence relation $\approx$ generated by $\sim$: Define $x \approx y$, when there is some n (in $\{0, 1, 2, \ldots\}$), and $z_0, z_1, \ldots, z_n \in M$, such that $x = z_0$, $y = z_n$, and such that for all $i = 1, \ldots, n$, there exist formulas $z_{i-1} = psq$, $z_i = ptq$, where either $s \sim t$ or $t \sim s$. Show that this $\approx$ is in fact the smallest congruence relation containing the relation $\sim$.

3. (Look at 2.2.2.) There is a way to describe the monoid presented by $\mathcal{P}$ as the coequalizer in the Category of monoids of a pair of maps $B^* \to A^*$. Here, A is the alphabet of the presentation $\mathcal{P}$, and B is a set in 1-1 correspondence with the set of relations $\{u = v\}$; one of the maps takes the generator of B to its corresponding "u", and the other takes it to the corresponding "v".

4. Go through the section 2.3.0 and detail the proof that the given monoid presentation does indeed yield a free group, and more generally that it determines the left adjoint of the inclusion functor from groups to monoids.

5. The inclusion functor groups to monoids also has a right adjoint; discuss this.

6. Every subgroup of finite index in G contains a normal subgroup of finite index.

7. Go through the discussion in 2.3.7 in detail. In particular, verify the claim "under the homomorphism above of G into the group of units of the power-series ring Γ, the image of the group $G(n;p)$ consists of series starting with 1 and having no terms of degrees $< n$". As for the finiteness of the quotient groups $G/G(n;p)$, you may have some trouble with this at this point.

8. Draw some pictures to illustrate all the possibilities in 2.5.1.

9. The ease of making the definition in 2.5.3 gives a hint that the functors V and E from graphs to sets preserve limit, and therefore suggests that these functors have left adjoints. Given a set X, we would like to define a graph $D(X)$, so that graph-maps $D(X) \to \Gamma$ are in a natural 1-1 correspondence with set-maps $X \to V(\Gamma)$, and similarly for the edge-functor. Figure out what these left adjoints are.

10. Some details from 2.6.0. Show that when paths p and q are homotopic, then $s(p) = s(q)$ and $t(p) = t(q)$. In each homotopy class of paths, there is only one of shortest length; prove this by relating the situation to the free group on the edges of Γ. A reduced path is the same as an immersion of the arc of length n into the graph Γ. If p and q are homotopic, and r is a path such that pr is defined, then qr is defined, and pr is homotopic to qr. Show that $\overline{\overline{p}} = p$, and that $p\overline{p}$ is defined and is homotopic to the path of length 0 given by the vertex $s(p)$.

11. Say that two vertices $v, w \in V(\Gamma)$ are **in the same component**, when there is a path p in Γ that starts at v and terminates at w. This is an equivalence relation on the vertices of Γ. A graph Γ is **connected**, when all its vertices are in the same component. In general, a **component** of Γ is the subgraph consisting on one equivalence class of vertices and all edges whose initial and terminal vertices are in this equivalence class. Any graph is the disjoint union (coproduct in the Category of graphs) of its components.

12. If Γ has only one vertex, then a path in Γ is the same as a word in the set of edges. In this case, $\pi_1(\Gamma)$ is a free group.

13. If $f : \Delta \to \Gamma$ is a map of graphs, with $f(v_0) = w_0$, then:

(a) When f is locally surjective, then every path p in Γ starting at w_0 lifts to some path q in Δ starting at v_0.

(b) When f is locally injective, then there is at most one lift of p starting at v_0; furthermore, the homomorphism of fundamental groups $\pi_1(\Delta, v_0) \to \pi_1(\Gamma, w_o)$ determined by f is injective.

(c) When f is locally bijective (i.e., a covering projection), define $F = \pi_1(\Gamma, w_0)$, and define S to be the subgroup of F which is the image of $\pi_1(\Delta, v_0)$ by the map f. Then an element $\alpha \in F$ belongs to S, if and only if, for any closed path p of Γ based at w_0 which represents α, its unique lift q in Δ starting at v_0 also ends at v_0.

14. Go through the details of 2.6.4. In addition, consider the case that Γ is a more complicated, but connected, graph, and figure how to define the covering Δ corresponding to a subgroup of $\pi_1(\Gamma, v_0)$.

15. Consider the free group F on $\{a, b\}$, and its subgroup S generated by $\{a, bab^{-1}\}$. Invent a finite graph Δ and a map $f : \Delta \to \Gamma$, where Γ is the 1-vertex graph with fundamental group F, which represents the subgroup S. This can be done so that f is locally injective. Now, enlarge Δ to a covering space of Γ by adding on extra trees to the vertices. The old Δ will probably be the core of the new and improved Δ. Relate the new Δ to the right cosets of S in F, and state some results about this situation that can be read off from the picture you have of Δ. For instance: Suppose that u, v are two reduced words in $\{a, b, \overline{a}, \overline{b}\}$. Suppose that u starts (on the left) with bb and that v starts with $\overline{b}$; then u and v belong to distinct right cosets of S in F.

16. Using results about immersions and folds, and Exercise 2.13, show that there are simple algorithms, which can be programmed on your toy computer, with these properties:

(a) A program P1 that accepts as input the following: $x_1, \ldots, x_n$, a finite alphabet; $w_1, \ldots, w_k$, a finite set of elements of the free group $F = F(x_1, \ldots, x_n)$. It outputs a free basis for the subgroup of F generated by $\{w_1, \ldots, w_k\}$.

(b) A program P2 that accepts as input the following: $x_1, \ldots, x_n$, a finite alphabet; $w_1, \ldots, w_k$, a finite set of elements of the free group $F = F(x_1, \ldots, x_n)$; and $u \in F$. This program outputs: The answer as to whether or not $u \in S$, where S is the subgroup of F generated by $\{w_1, \ldots, w_k\}$. A specific instance of this is the solution to the question whether or not the subgroup S generated by a given finite set in the finitely generated free group F, is equal to F.

17. More on Howson's theorem, 2.7.5. By looking closely at core graphs, and putting the picture in the context where the target graph Δ has only vertices of valence 3 (the valence of a vertex being the cardinality of its link), one can count the ranks of these groups to a certain extent. The result of this kind of investigation is this:

If $C = A \cap B$ is non-trivial, then

$$\text{rank}\,(C) - 1 \leq 2 \cdot (\text{rank}\,(A) - 1) \cdot (\text{rank}\,(B) - 1).$$

If both or one of A or B has finite index in the group F, the factor "2" in the above inequality can be replaced by "1". It is a, to date unproved, conjecture by Hanna Neumann that the factor 2 can always be replaced by 1. $\diamondsuit$

18. Define the cancellation bound of ϕ in 2.8.3, intrinsically, without using graphs, but using free bases of the given groups. Examples can be made up out of automorphisms of free groups. For instance, in $F(a, b)$, consider the two elements $f(x) = ab^2ab$ and $f(y) = babab^2ab$. The map $f : F(x, y) \to F(a, b)$ is an isomorphism. By counting edges, we can figure how many folds there are between Γ, the graph made out of x, y, and Δ, the graph made out of a, b. There will be 11 folds. Let $p = xy$ and $q = x^{-1}y^{-1}$. Then

$$(fp) \cdot (fq) = ab^2abbabab^2ab \cdot b^{-1}a^{-1}b^{-2}a^{-1}b^{-1}a^{-1}b^{-2}a^{-1}b^{-1}a^{-1}b^{-1}.$$

The product $fp \cdot fq$ cancels down to $aba^{-1}b^{-1}$, and there are 11 cancellations.

19. Go through the proof of 2.9.1, and find all the points which are not totally obvious. Then prove them. In particular, in Δ', assuming that $\pi_1(\Delta')$ is non-trivial, if w is any vertex, and e is any edge having w as starting vertex, then there is a reduced path in Δ' from w to v_0, with first edge equal to e. Given any graph, in which all vertices have finite valence, compute the Euler characteristics of finite subgraphs, so as to bound the rank of the fundamental group by a formula involving the number of vertices of valence greater than 2, and their valences.

20. Note that the above proof does not yield a bound on the lengths of a basis for the fixed subgroup. Try to invent examples with long lengths.

References

[1] Dyck, W., Gruppentheoretische Studien, *Math. Ann.* 20 (1882), 1–45. [Early ideas about free groups and presentations.]

[2] Magnus, W., Beziehungen swischen Gruppen und Idealen in einem speziellen Ring, *Math. Ann.* 111 (1935), 259–280. [Infinite series and lower central series of free groups.]

[3] Hall, M., Coset representations in free groups, *Trans. Amer. Math. Soc.* 67 (1949), 421–432. [Finitely generated subgroup of a free group is the intersection of subgroups of finite index.]

[4] Stallings, J., Topology of finite graphs, *Inv. math.* 71 (1983), 551–565. [How graphs and free groups can be friends.]

[5] Goldstein, R. and Turner, E., Monomorphisms of finitely generated free groups have finitely generated equalizers, *Inv. math.* 82 (1985), 283–289.

[6] Bestvina, M. and Handel, M., Train tracks and automorphisms of free groups, *Ann. of Math.* 135 (1992), 1–51. [Proof of the conjecture of Peter Scott that the fixed subgroup of an automorphism of a free group of rank n has rank $\leq n$.]