

Chapter 2. Free groups.

In this chapter, we introduce the idea of universal constructions through the particular case of free groups. We shall first motivate the free group concept, then develop three ways of constructing such groups.

2.1. Motivation. Suppose G is a group and we take (say) three elements $a, b, c \in |G|$, and consider what group-theoretic relations these satisfy. That is, letting T be the set of all group-theoretic terms in three symbols x, y and z , we look at pairs of elements $p(x, y, z), q(x, y, z) \in T$, and if $p_G(a, b, c) = q_G(a, b, c)$ in $|G|$, we say that (a, b, c) satisfies the relation $p = q$. We note:

Lemma 2.1.1. *Suppose F and G are groups, $a, b, c \in |F|$ are three elements generating F , and α, β, γ are any three elements of G . Then the following conditions are equivalent:*

- (i) *Every group-theoretic relation $p = q$ satisfied by (a, b, c) in F is also satisfied by (α, β, γ) in G .*
- (ii) *There exists a group homomorphism $h: F \rightarrow G$ under which $a \mapsto \alpha, b \mapsto \beta, c \mapsto \gamma$.*

Further, when these conditions hold, the homomorphism h of (ii) is unique.

If the assumption that a, b and c generate F is dropped, one still has (ii) \Rightarrow (i).

Proof. If h is a homomorphism as in (ii), then I claim that for all $p \in T$,

$$h(p_F(a, b, c)) = p_G(\alpha, \beta, \gamma).$$

Indeed, the set of $p \in T$ for which the above equation holds is easily seen to contain x, y and z and to be closed under the operations of T , hence it is all of T . Statement (i) follows. If, also, a, b and c generate F , then every element of $|F|$ can be written $p_F(a, b, c)$ for some p , so the above formula shows that h is uniquely determined by a, b, c, α, β and γ .

On the other hand, suppose a, b and c generate F and (i) holds. For each $g = p_F(a, b, c) \in |F|$, define $h(g) = p_G(\alpha, \beta, \gamma)$. To show that this gives a well-defined map from $|F|$ to $|G|$, note that if we have two ways of writing an element $g \in |F|$, $p_F(a, b, c) = g = q_F(a, b, c)$, then the relation $p = q$ is satisfied by (a, b, c) in F , hence by (i), it is satisfied by (α, β, γ) in G , hence the two values our definition prescribes for $h(g)$, namely $p_G(\alpha, \beta, \gamma)$ and $q_G(\alpha, \beta, \gamma)$, are the same.

That this set map is a homomorphism follows from the way evaluation of group-theoretic terms is defined. For instance, given $g \in |F|$, suppose we want to show that $h(g^{-1}) = h(g)^{-1}$. We write $g = p_F(a, b, c)$. Then $(\iota_T(p))_F(a, b, c) = g^{-1}$, so our definition of h gives $h(g^{-1}) = (\iota_T(p))_G(\alpha, \beta, \gamma) = p_G(\alpha, \beta, \gamma)^{-1} = h(g)^{-1}$. The same method works for products and for the neutral element. \square

Exercise 2.1:1. Show by example that if $\{a, b, c\}$ does not generate F , then condition (i) of the above lemma can hold and (ii) fail, and also that (ii) can hold but h not be unique. (You may replace (a, b, c) with a smaller family, (a, b) or (a) , if you like.)

Lemma 2.1.1 leads one to wonder: Among all groups F given with generating 3-tuples of elements (a, b, c) , is there one in which these three elements satisfy the *smallest* possible set of relations? We note what the above lemma would imply for such a group:

Corollary 2.1.2. *Let F be a group, and $a, b, c \in |F|$. Then the following conditions are equivalent:*

- (i) a, b, c generate F , and the only relations satisfied by a, b, c in F are those relations satisfied by every 3-tuple (α, β, γ) of elements in every group G .
- (ii) For every group G , and every 3-tuple of elements (α, β, γ) in G , there exists a unique homomorphism $h: F \rightarrow G$ such that $h(a) = \alpha$, $h(b) = \beta$, $h(c) = \gamma$. \square

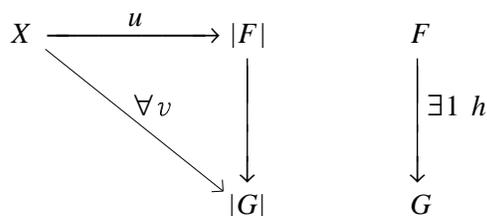
Only one point in the deduction of this corollary from Lemma 2.1.1 is not completely obvious; I will make it an exercise:

Exercise 2.1.2. In the situation of the above corollary, show that (ii) implies that a, b and c generate F . (Hint: Let G be the subgroup of F generated by those three elements.)

I've been speaking of 3-tuples of elements for concreteness; the same observations are valid for n -tuples for any n , and generally, for X -tuples for any set X . An X -tuple of elements of F means a set map $X \rightarrow |F|$, so in this general context, condition (ii) above takes the form given by the next definition. (But making this definition does not answer the question of whether such objects exist!)

Definition 2.1.3. *Let X be a set. By a free group F on the set X , we shall mean a pair (F, u) , where F is a group, and u a set map $X \rightarrow |F|$, having the following universal property:*

For any group G , and any set map $v: X \rightarrow |G|$, there exists a unique homomorphism $h: F \rightarrow G$ such that $v = hu$; i.e., making the diagram below commute.



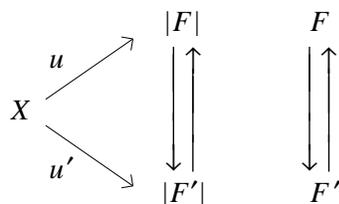
(In the above diagram, the first vertical arrow also represents the homomorphism h , regarded as a map on the underlying sets of the groups.)

Corollary 2.1.2 (as generalized to X -tuples) says that (F, u) is a free group on X if and only if the elements $u(x)$ ($x \in X$) generate F , and satisfy *no* relations *except* those that must hold in any group. In this situation, one says that these elements “freely” generate F , hence the term *free group*. Note that if such an F exists, then by definition, *every* X -tuple of members of any group G can be obtained, in a unique way, as the image under a group homomorphism $F \rightarrow G$ of the *particular* X -tuple given by u . Hence that X -tuple can be thought of as a “universal X -tuple of group elements”, and so the property characterizing it is called a *universal property*.

We note a few elementary facts and conventions about such objects. If (F, u) is a free group on X , then the map $u: X \rightarrow |F|$ is one-to-one. (This is easy to prove from the universal property, plus the well-known fact that there exist groups with more than one element. The student who has not seen free groups developed before should think this argument through.) Hence given a free group, it is easy to get from it one such that the map u is actually an inclusion $X \subseteq |F|$. Hence for notational convenience, one frequently assumes that this is so; or, what is approximately the same thing, one often uses the same symbol for an element of X and its image in $|F|$.

If (F, u) and (F', u') are both free groups on the same set X , there is a unique

isomorphism between them *as* free groups, i.e., respecting the maps u and u' . (Cf. diagram below.)



(If you haven't seen this result before, again see whether you can work out the details. For the technique you might refer to the proof of Proposition 3.3.3 below.) As any two free groups on X are thus “essentially” the same, one sometimes speaks of *the* free group on X .

One also often says that a group F “is free” to mean “there exists some set X and some map $u: X \rightarrow |F|$ such that (F, u) is a free group on X ”. When this holds, X can always be taken to be a subset of $|F|$, and u the inclusion map.

But it is time we proved that free groups exist. We will show three different ways of constructing them in the next three sections.

Exercise 2.1:3. Suppose one replaces the word “group” by “finite group” throughout Definition 2.1.3. Show that for any nonempty set X , *no* finite group exists having the stated universal property.

2.2. The logician’s approach: construction from group-theoretic terms. We know from Corollary 2.1.2 that if a free group F on three generators a, b, c exists, then each of its elements can be written $p_F(a, b, c)$ for some group-theoretic term p , and that *two* such elements, $p_F(a, b, c)$ and $q_F(a, b, c)$, are equal if and only if the equation “ $p = q$ ” is satisfied by every three elements of every group, i.e., follows from the group axioms. This suggests that we may be able to construct such a group by taking the set of all group-theoretic terms in three variables, constructing an equivalence relation “ $p \sim q$ ” on this set which means “the equality of p and q is a consequence of the group axioms”, taking for $|F|$ the quotient of our set of terms by this relation, and defining operations \cdot , $^{-1}$ and e on $|F|$ in some natural manner. This we shall now do!

Let X be any set, and $T = T_{X, \cdot, ^{-1}, e}$ the set of all group-theoretic terms in the elements of X . What conditions must a relation “ \sim ” satisfy for $p \sim q$ to be the condition “ $p_v = q_v$ ” for *some* map v of X into *some* group G ? Well, the group axioms tell us that it must satisfy

$$(2.2.1) \quad (\forall p, q, r \in T) (p \cdot q) \cdot r \sim p \cdot (q \cdot r),$$

$$(2.2.2) \quad (\forall p \in T) (p \cdot e \sim p) \wedge (e \cdot p \sim p),$$

$$(2.2.3) \quad (\forall p \in T) (p \cdot p^{-1} \sim e) \wedge (p^{-1} \cdot p \sim e).$$

Also, just the well-definedness of the operations of G tells us that

$$(2.2.4) \quad (\forall p, p', q \in T) (p \sim p') \Rightarrow ((p \cdot q \sim p' \cdot q) \wedge (q \cdot p \sim q \cdot p')),$$

$$(2.2.5) \quad (\forall p, p' \in T) (p \sim p') \Rightarrow (p^{-1} \sim p'^{-1}).$$

Finally, of course, \sim must be an equivalence relation:

$$(2.2.6) \quad (\forall p \in T) \quad p \sim p,$$

$$(2.2.7) \quad (\forall p, q \in T) \quad (p \sim q) \Rightarrow (q \sim p),$$

$$(2.2.8) \quad (\forall p, q, r \in T) \quad ((p \sim q) \wedge (q \sim r)) \Rightarrow (p \sim r).$$

So let us take for “ \sim ” the *least* binary relation on T satisfying conditions (2.2.1-8).

Let us note what this means, and why it exists: Recall that a binary relation on a set T is formally a subset $R \subseteq T \times T$; when we write $p \sim q$, this is understood to be an abbreviation for $(p, q) \in R$. “Least” means smallest with respect to set-theoretic inclusion. Our conditions (2.2.1-8) are in the nature of closure conditions, and, as with all sets defined by closure conditions, the existence of a least set satisfying them can be established in two ways:

We may capture this set “from above” by forming the *intersection* of all binary relations on T satisfying (2.2.1-8) – the set-theoretic intersection of these relations as subsets of $T \times T$. (Note, incidentally, that if we think of such relations as predicates rather than as sets, this intersection \cap becomes a (generally infinite) conjunction \wedge .) The key point to observe is that each of these conditions is such that an intersection of relations satisfying it again satisfies it. Hence the intersection of *all* relations satisfying (2.2.1)-(2.2.8) will be the least such relation.

Or we can “build it up from below”. Let R_0 denote the empty relation $\emptyset \subseteq T \times T$, and recursively construct the $i+1$ st relation R_{i+1} from the i th, by adding to R_i those elements that conditions (2.2.1-8) say must *also* be in R , given that the elements of R_i are there. Precisely, we let

$$\begin{aligned} R_{i+1} = R_i & && \text{(elements already constructed)} \\ \cup \{((p \cdot q) \cdot r, p \cdot (q \cdot r)) \mid p, q, r \in T\} & && \text{(elements arising by (2.2.1))} \\ \cup \dots & && \dots \\ \cup \{(p, r) \mid (\exists q) (p, q) \in R_i \wedge (q, r) \in R_i\}. & && \text{(elements arising by (2.2.8))} \end{aligned}$$

We now define $R = \cup_i R_i$. It is straightforward to show that R satisfies (2.2.1-8), and that any subset of $T \times T$ satisfying (2.2.1-8) must contain R ; so R , looked at as a binary relation \sim on T , is the desired least relation.

By (2.2.6-8), \sim is an equivalence relation; so define $|F| = T / \sim$, i.e., the set of equivalence classes of this relation. We shall denote the typical element of $|F|$, the equivalence class of an element $p \in T$, by $[p]$. We map X into $|F|$ by defining

$$u(x) = [x]$$

(or, if we do not identify $\text{symb}_T(x)$ with x in our construction of T , then $u(x) = [\text{symb}_T(x)]$).

We now define operations \cdot , $^{-1}$ and e on $|F|$ by

$$\begin{aligned} [p] \cdot [q] &= [p \cdot q], \\ [p]^{-1} &= [p^{-1}], \\ e &= [e]. \end{aligned}$$

That the first two of these are *well-defined* follows respectively from properties (2.2.4) and (2.2.5) of \sim ! (With the third there is no problem.) From properties (2.2.1-3) of \sim , it now follows that $(|F|, \cdot, ^{-1}, e)$ satisfies the group axioms. E.g., given $[p], [q], [r] \in |F|$, if we

evaluate $([p] \cdot [q]) \cdot [r]$ and $[p] \cdot ([q] \cdot [r])$ in $|F|$, we get $[(p \cdot q) \cdot r]$ and $[p \cdot (q \cdot r)]$ respectively, which are equal by (2.2.1). Finally, writing F for the group $(|F|, \cdot, {}^{-1}, e)$, it is clear from our construction of \sim that the only relations satisfied by the images in F of the elements of X are relations that follow logically from the group axioms; so by Corollary 2.1.2 (or more precisely, the generalization of that corollary with X -tuples in place of 3-tuples), F has the desired universal property.

To see this universal property more directly, suppose v is any map $X \rightarrow |G|$, where G is a group. Write $p \sim_v q$ to mean $p_v = q_v$ in G . Then clearly the relation \sim_v satisfies conditions (2.2.1-8), hence it contains the *least* such relation, our \sim . So a well-defined map $h: |F| \rightarrow |G|$ is given by $h([p]) = p_v \in |G|$, and it follows from the way the operations of F , and the evaluation of terms in G at the X -tuple v , are defined, that h is a homomorphism, and is the unique homomorphism such that $hu = v$. Thus

Proposition 2.2.9. *(F, u), constructed as above, is a free group on the given set X . \square*

So a free group on every set X does indeed exist!

Some further notes:

2.2.10. There is a viewpoint that goes along with this construction, which will be helpful in thinking about universal constructions in general. Suppose that we are given a set X , and that we know that G is a group, with a map $v: X \rightarrow |G|$. How much can we “say about” G from this fact alone? We can *name* certain elements of G , namely the $v(x)$ ($x \in X$), and all the elements that can be obtained from these by the group operations of G (e.g., $(v(x) \cdot v(y))^{-1} \cdot ((v(y)^{-1} \cdot e)^{-1} \cdot v(z))$ if $x, y, z \in X$). A particular G may contain more elements than those obtained in such ways, but we have no way of getting our hands on them from the given information. We can also derive from the identities for groups certain relations that these elements satisfy, (e.g., $(v(x) \cdot v(y))^{-1} = v(y)^{-1} \cdot v(x)^{-1}$). The elements $v(x)$ may, in particular cases, satisfy more relations than these, but again we have no way of deducing these additional relations. If we now gather together this limited “data” that we have about such a group G – the quotient of a set of labels for certain elements by a set of identifications among these – we find that this collection of “data” itself forms a group with a map of X into it; and is, in fact, a universal such group!

2.2.11. At the beginning of this section, I motivated our construction by saying that “ \sim ” should mean “equality that follows from the group axioms”. I then wrote down a series of eight rules, (2.2.1-8), all of which are clearly valid procedures for deducing equations which hold in all groups. What was not obvious was whether they would be sufficient to yield *all* such equations. But they were – the proof of the pudding being that $(T/\sim, \cdot, {}^{-1}, e)$ was shown to *be* a group.

This is an example of a very general type of situation in mathematics: Some class, in this case, a class of pairs of group-theoretic terms, is described “from above”, i.e., is defined as the class of all elements satisfying certain restrictions (in this case, those pairs $(p, q) \in T \times T$ such that the relation $p = q$ holds on all X -tuples of elements of all groups). We seek a way of describing it “from below”, i.e., of *constructing* or *generating* all members of the class. Some procedure which produces members of the set is found, and one seeks to show that this procedure yields the whole set – or, if it does not, one seeks to extend it to a procedure that does.

The inverse situation is equally important, where we are given a construction which “builds

up'' a set, and we seek a convenient way of characterizing the elements that result. Exercise 1.7:1 was of that form. You will see more examples of both situations throughout this course, and, in fact, in most every mathematics course you take.

Exercise 2.2:1. Prove directly from (2.2.1-8) that for $x, y \in X$, $(x \cdot y)^{-1} \sim y^{-1} \cdot x^{-1}$. (Your solution should show explicitly each application you make of each of those conditions.)

Exercise 2.2:2. Does the relation of the preceding exercise follow from (2.2.1-3) and (2.2.6-8) alone?

Note that in our recursive construction of the set R (that is, the relation \sim), repeated application of (2.2.1-3) was really unnecessary; these conditions give the same elements of R each time they are applied, so we might as well just have applied them the first time and only applied (2.2.4-8) from then on. Less obvious is the answer to:

Exercise 2.2:3. (A. Tourbaroff) Can the construction of R be done in three stages: First take the set P of elements given by (2.2.1-3), then form the closure Q of this set under applications of (2.2.4-5) (as before, by recursion or as an intersection), and finally, obtain R as the closure of Q under applications of (2.2.6-8) (another recursion or intersection)? This procedure will yield some subset of $T \times T$; the question is whether it is the R we want.

What if we do things in a different order – first (2.2.1-3), then (2.2.6-8), then (2.2.4-5)?

2.3. Free groups as subgroups of big enough direct products. Another way of getting a group in which some X -tuple of elements satisfies the smallest possible set of relations is suggested by the following observation. Let G_1 and G_2 be two groups, and suppose we are given elements

$$\alpha_1, \beta_1, \gamma_1 \in |G_1|, \quad \alpha_2, \beta_2, \gamma_2 \in |G_2|.$$

Then in the direct product group $G = G_1 \times G_2$ we have the elements

$$a = (\alpha_1, \alpha_2), \quad b = (\beta_1, \beta_2), \quad c = (\gamma_1, \gamma_2),$$

and we find that the set of relations satisfied by a, b, c in G is precisely the *intersection* of the sets of relations satisfied by $\alpha_1, \beta_1, \gamma_1$ in G_1 and by $\alpha_2, \beta_2, \gamma_2$ in G_2 . This may be seen from the fact that for any $s \in T$,

$$s_G(a, b, c) = (s_{G_1}(\alpha_1, \beta_1, \gamma_1), s_{G_2}(\alpha_2, \beta_2, \gamma_2)),$$

as is easily verified by induction.

More generally, if we take an arbitrary family of groups $(G_i)_{i \in I}$, and in each G_i three elements $\alpha_i, \beta_i, \gamma_i$, then in the product group $G = \prod G_i$ we can define the elements

$$a = (\alpha_i)_{i \in I}, \quad b = (\beta_i)_{i \in I}, \quad c = (\gamma_i)_{i \in I},$$

and the relations that these satisfy will be just those relations satisfied simultaneously by our 3-tuples in all of these groups.

This suggests that by using a *large enough* such family, we could arrive at a group with three elements a, b, c which satisfy a *smallest possible* set of relations.

How large a family $(G_i, \alpha_i, \beta_i, \gamma_i)$ should we use?

Well, we could be sure of getting the least set of relations if we could use the class of *all* groups and *all* 3-tuples of elements of these. But taking the direct product of such a family would give us set-theoretic indigestion.

We can cut down this surfeit of groups a bit by noting that for any group G_i and three

elements $\alpha_i, \beta_i, \gamma_i$, if we let H_i denote the subgroup of G_i generated by these three elements, it will suffice for our product to involve the group H_i , rather than the whole group G_i , since the relations satisfied by α_i, β_i and γ_i in the whole group G_i and in the subgroup H_i are the same. Now a finitely generated group is countable (meaning finite *or* countably infinite), so we see that it would be enough to let $(G_i, \alpha_i, \beta_i, \gamma_i)$ range over all *countable* groups, and all 3-tuples of elements thereof.

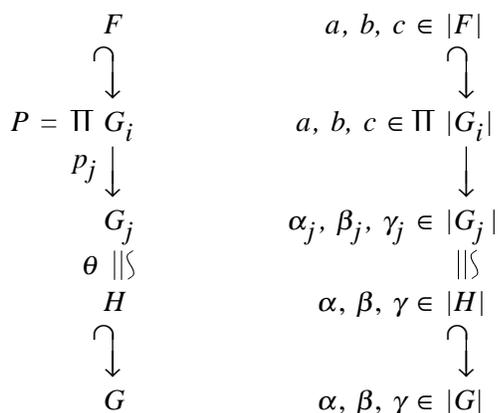
However, the class of all countable groups is still not a set. Indeed, even the class of one-element groups is not a set, because we get a different (in the strict set-theoretic sense) group for each choice of that one element. (For those not familiar with such considerations: In set theory, every element of a set is a set. If we had a set of *all* one-element groups, then we could form from this the set of all *members* of their underlying sets, which would be the set of *all* sets; and one knows that this does not exist.) But this is clearly just a quibble – obviously, if we choose any one-element set $\{x\}$, and take the unique group with this underlying set, it will serve as well as any other one-element group so far as honest group-theoretic purposes are concerned. In the same way, I claim we can find a genuine *set* of countable groups that *up to isomorphism* contains *all* the countable groups. Namely, let S be a fixed countably infinite set. Then we can form the set of all groups G whose underlying sets $|G|$ are subsets of S . Or, to hit more precisely what we want, let

$$(2.3.1) \quad \{(G_i, \alpha_i, \beta_i, \gamma_i) \mid i \in I\}$$

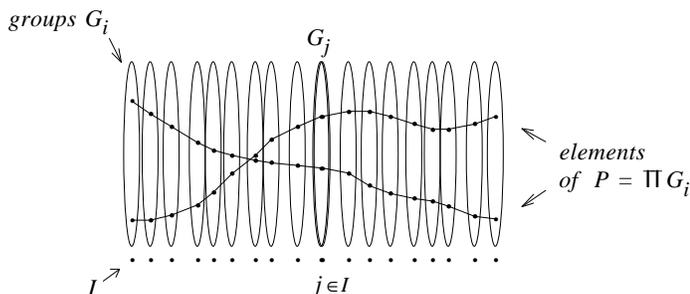
be the set of all 4-tuples such that G_i is a group with $|G_i| \subseteq S$, and α_i, β_i and γ_i are members of $|G_i|$. Now for any countable group H and three elements $\alpha, \beta, \gamma \in |H|$, we can clearly find an isomorphism θ between one of these groups, say G_j ($j \in I$), and H , such that $\theta(\alpha_j) = \alpha$, $\theta(\beta_j) = \beta$, $\theta(\gamma_j) = \gamma$; so (2.3.1) is “big enough” for our purpose.

So taking (2.3.1) as above, let P be the direct product group $\prod_I G_i$, let a, b, c be the I -tuples $(\alpha_i), (\beta_i), (\gamma_i) \in |P|$, and let F be the subgroup of P generated by a, b , and c . I claim that F is a free group on a, b , and c .

We could prove this by considering the set of relations satisfied by a, b, c in F as suggested above, but let us instead verify directly that F satisfies the universal property (2.1.3) characterizing free groups. Let G be any group, and α, β, γ three elements of G . We want to prove that there exists a unique homomorphism $h: F \rightarrow G$ carrying $a, b, c \in |F|$ to $\alpha, \beta, \gamma \in |G|$ respectively. Uniqueness will be no problem – by construction F is generated by a, b and c , so if such a homomorphism exists it is unique. To show the existence of h , note that the subgroup H of G generated by α, β, γ is countable, hence as we have noted, there exists for some $j \in I$ an isomorphism $\theta: G_j \cong H$ carrying $\alpha_j, \beta_j, \gamma_j \in |G_j|$ to $\alpha, \beta, \gamma \in |H|$. Now the projection map p_j of the product group $P = \prod G_i$ onto its j th coordinate takes a, b and c to $\alpha_j, \beta_j, \gamma_j$, hence composing this projection with θ , as shown in the diagram below, we get a homomorphism $h: F \rightarrow G$ having the desired effect on a, b, c .



For a useful way to picture this construction, think of P as the group of all functions on the base-space I , taking at each point i a value in G_i . F is the subgroup of functions generated by a , b and c .



Given α, β, γ in any group G , identify the subgroup of G that they generate with an appropriate G_j ($j \in I$). Then the homomorphism h that we constructed above may be thought of as taking each element of F to its value at the point j . We have chosen our space I and values for a, b and c sufficiently eclectically so that it is possible to choose points at which a, b and c take on (up to isomorphism) any 3-tuple of values in any group. Thus, the functions a, b and c are a “universal” 3-tuple of group-elements.

The same argument works if we replace “3-tuple” by “ X -tuple”, where X is any countable set. Here we use the observation that a group generated by a countable family of elements is countable. For X of arbitrary cardinality, one can easily show that a group H generated by an X -tuple of elements has cardinality $\leq \max(\text{card}(X), \aleph_0)$. Hence we get:

Proposition 2.3.2. *X be any set. Take a set S of cardinality $\max(\text{card}(X), \aleph_0)$, and let $\{(G_i, u_i) \mid i \in I\}$ be the set of all pairs such that G_i is a group with $|G_i| \subseteq S$, and u_i is a map $X \rightarrow |G_i|$ (i.e., an X -tuple of elements of G_i). Let $P = \prod_I G_i$, and map X into P by defining $u(x)$ ($x \in X$) to be the element with component $u_i(x)$ at each i . Let F be the subgroup of P generated by $\{u(x) \mid x \in X\}$.*

Then the pair (F, u) is a free group on the set X . \square

Digression: Let S_3 be the symmetric group on three letters. Suppose we had begun the above investigation with a less ambitious goal: merely to find a group J with three elements a, b, c such that

$$(2.3.3) \quad \begin{array}{l} \text{For every choice of three elements } \alpha, \beta, \gamma \in |S_3|, \\ \text{there exists a unique homomorphism } h: J \rightarrow S_3 \\ \text{taking } a, b, c \text{ to } \alpha, \beta, \gamma \text{ respectively.} \end{array} \quad \begin{array}{ccc} J & a, b, c \in |J| & \\ \downarrow h & & \downarrow \\ S_3 & \alpha, \beta, \gamma \in |S_3| & \end{array}$$

Then we could have performed the above construction just using 4-tuples $(S_3, \alpha, \beta, \gamma)$ ($\alpha, \beta, \gamma \in |S_3|$) as our $(G_i, \alpha_i, \beta_i, \gamma_i)$. There are $6^3 = 216$ such 4-tuples, so P would be the direct product of 216 copies of S_3 , and a, b, c would be elements of this product which, as one runs over the 216 coordinates, take on all possible combinations of values in S_3 . The subgroup J they generate would indeed satisfy (2.3.3). This leads to:

Exercise 2.3:1. Does condition (2.3.3) characterize (J, a, b, c) up to isomorphism? If not, is there some additional condition that (J, a, b, c) satisfies which together with (2.3.3) determines it up to isomorphism?

Exercise 2.3:2. Investigate the structure of the group J , and more generally, of the analogous groups constructed from S_3 using different numbers of generators. To make the problem concrete, try to determine, or estimate as well as possible, the *orders* of these groups, for 1, 2, 3 and generally, for n generators.

The two methods by which we have constructed free groups above go over essentially word-for-word with “group” replaced by “ring”, “lattice”, or a great many other types of mathematical objects. The determination of just what classes of algebraic structures allow this and related sorts of universal constructions is one of the themes of this course. The next exercise concerns a negative example.

Exercise 2.3:3. State what would be meant by a “free field on a set X ”, and show that no such object exists for any set X . If one attempts to apply the two methods of this and the preceding section to prove the existence of free fields, where does each of them fail?

Exercise 2.3:4. Let $\mathbb{Z}[x_1, \dots, x_n]$ be the polynomial ring in n indeterminates over the integers (= the free commutative ring on n generators – cf. §3.12 below). Its field of fractions $\mathbb{Q}(x_1, \dots, x_n)$, the field of “rational functions in n indeterminates over the rationals”, looks in some ways like a “free field on n generators”. E.g., one often speaks of evaluating a rational function at some set of values of the variables. Can some concept of “free field” be set up, perhaps based on a *modified* universal property, or on some concept of comparing relations in the *field* operations satisfied by n -tuples of elements in two fields, in terms of which $\mathbb{Q}(x_1, \dots, x_n)$ would indeed be the free field on n generators?

Exercise 2.3:5. A *division ring* (or *skew field* or *sfield*) is a ring (associative but not necessarily commutative) in which every nonzero element is invertible. If you find a satisfactory answer to the preceding exercise, you might consider the question of whether there exists in the same sense a *free division ring* on n generators. (This was a longstanding open question, which was finally answered in 1966, and then again, by a very different approach, about five years later. I can refer interested students to papers in this area.)

There are many hybrids and variants of the two constructions we have given for free groups. For instance, we might start with the set T of terms in X , and define $p \sim q$ (for $p, q \in T$) to mean that for every map v of X into a group G , one has $p_v = q_v$ in G . Now for each pair $(p, q) \in T \times T$ such that $p \sim q$ fails to hold, we can choose a map $u_{p,q}$ of X into a group $G_{p,q}$ such that $p_{u_{p,q}} \neq q_{u_{p,q}}$. We can then form the direct product group $P = \prod G_{p,q}$, take the induced map $u: X \rightarrow |P|$, and check that the subgroup F generated by the image of this map will satisfy condition (i) of Corollary 2.1.2. Interestingly, for X countable, this construction uses a product of *fewer* groups $G_{p,q}$ than we used in the version given above.

Finally, consider the following construction, which suffers from severe set-theoretic difficulties, but is still interesting. (I won't try to resolve these difficulties here, but will talk sloppily, as though they did not occur.)

Define a “generalized group-theoretic operation in three variables” as any function p which associates to every group G and three elements $\alpha, \beta, \gamma \in |G|$ an element $p(G, \alpha, \beta, \gamma) \in |G|$. We can “multiply” two such operations p and q by defining

$$(p \cdot q)(G, \alpha, \beta, \gamma) = p(G, \alpha, \beta, \gamma) \cdot q(G, \alpha, \beta, \gamma) \in |G|.$$

for all groups G and elements $\alpha, \beta, \gamma \in |G|$. We can similarly define the multiplicative inverse of such an operation p , and the constant operation e . We see that the class of generalized group-theoretic operations will satisfy the group axioms under the above three operations. Now consider the three generalized group-theoretic operations a, b and c defined by

$$a(G, \alpha, \beta, \gamma) = \alpha, \quad b(G, \alpha, \beta, \gamma) = \beta, \quad c(G, \alpha, \beta, \gamma) = \gamma.$$

Let us define a “derived generalized group-theoretic operation” as one obtainable from a, b and c by the operations of product, inverse, and neutral element defined above. Then the set of derived generalized group-theoretic operations will form a free group on the generators a, b and c . (This is really just a disguised form of our naive “direct product of *all* groups” idea.)

Exercise 2.3:6. Call a generalized group-theoretic operation p *functorial* if for every homomorphism of groups $f: G \rightarrow H$, one has $f(p(G, \alpha, \beta, \gamma)) = p(H, f(\alpha), f(\beta), f(\gamma))$ ($\alpha, \beta, \gamma \in |G|$). (We will see the reason for this term in Chapter 6.) Show that all derived group-theoretic operations are functorial. Is the converse true?

Exercise 2.3:7. Same question for functorial generalized operations on the class of all *finite groups*.

2.4. The classical construction: free groups as groups of words. The constructions discussed above have the disadvantage of not giving very explicit descriptions of free groups. We know that every element of a free group F on the set X arises from a term in the elements of X and the group operations, but we don't know how to tell whether two such terms – say $(b(a^{-1}b)^{-1})(a^{-1}b)$ and e – yield the same element; in other words, whether $(\beta(\alpha^{-1}\beta)^{-1})(\alpha^{-1}\beta) = e$ is true for all elements α, β of all groups. If it is, then by the results of §2.2 one can obtain this fact *somehow* by the procedures corresponding to conditions (2.2.1-8); if it is not, then the ideas of §2.3 suggest that we should try to prove this by looking for some particular elements for which it fails, in some particular group in which we know how to calculate. But these approaches are hit-and-miss.

In this section, we shall construct the free group on X in a much more explicit way. We will then be able to answer such questions by calculating in the free group itself.

We first recall an important consequence of the associative identity: that “products can be written without parentheses”. For example, given elements a, b, c of a group, the elements $a(c(ab)), a((ca)b), (ac)(ab), (a(ca))b$ and $((ac)a)b$ are all equal. It is conventional, and usually convenient, to say, “Let us therefore write their common value as $acab$.” However, we will soon want to relate these expressions to group-theoretic *terms*; so instead of dropping parentheses, let us agree to take $a(c(ab))$ as the common form to which we shall reduce the above five expressions, and generally, let us note that any product of elements can be reduced by the associative law to one with parentheses clustered to the right: $x_n (x_{n-1} (\dots (x_2 x_1) \dots))$.

In particular, given two elements written in this form, we can write down their product and

reduce it to this form by repeatedly applying the associative law:

$$(2.4.1) \quad \begin{aligned} & (x_n (\dots (x_2 x_1) \dots)) \cdot (y_m (\dots (y_2 y_1) \dots)) \\ &= x_n (\dots (x_2 (x_1 (y_m (\dots (y_2 y_1) \dots)))) \dots). \end{aligned}$$

If we want to find the inverse of an element written in this form, we may use the formula $(xy)^{-1} = y^{-1}x^{-1}$, another consequence of the group laws. By induction this gives $(x_n (\dots (x_2 x_1) \dots))^{-1} = (\dots (x_1^{-1} x_2^{-1}) \dots) x_n^{-1}$, which we may reduce, again by associativity, to $x_1^{-1} (\dots (x_{n-1}^{-1} x_n^{-1}) \dots)$.

More generally, if we started with an expression of the form

$$x_n^{\pm 1} (\dots (x_2^{\pm 1} x_1^{\pm 1}) \dots),$$

where each factor is either x_i or x_i^{-1} , and the exponents are independent, then the above method together with the fact $(x^{-1})^{-1} = x$ (another consequence of the group axioms) allows us to write its inverse as $x_1^{\mp 1} (\dots (x_{n-1}^{\mp 1} x_n^{\mp 1}) \dots)$, which is of the same form as the expression we started with; and, likewise, (2.4.1) shows that the *product* of two expressions of the above form reduces to an expression of the same form.

Note further that if two successive factors $x_i^{\pm 1}$ and $x_{i+1}^{\pm 1}$ are respectively x and x^{-1} for some element x , or are respectively x^{-1} and x for some x , then by the group axioms on inverses and the neutral element (and again, associativity), we can drop this pair of factors – unless they are the only factors in the product, in which case we can rewrite the product as e .

Finally, easy consequences of the group axioms tell us what the inverse of e is (namely e), and how to multiply anything by e . Putting these observations together, we conclude that *given any set X of elements of a group G* , the set of elements of G that can be written in one of the forms

$$(2.4.2) \quad \begin{aligned} & e, \text{ or } x_n^{\pm 1} (\dots (x_2^{\pm 1} x_1^{\pm 1}) \dots), \\ & \text{where } n \geq 1, \text{ each } x_i \in X, \text{ and no two successive factors are an} \\ & \text{element of } X \text{ and the inverse of the same element, in either order,} \end{aligned}$$

is closed under products and inverses. So this set must be the whole subgroup of G generated by X . In other words, any member of the subgroup generated by X can be *reduced* by the group operations to an expression (2.4.2).

In the preceding paragraph, X was a subset of a group. Now let X be an arbitrary set, and as in the preceding section, let T be a set of all group-theoretic terms in elements of X (Definition 1.5.1). For convenience, let us assume T chosen so as to contain X , with symb_T being the inclusion map. (If you prefer not to make this assumption, then in the argument to follow, you should insert “ symb_T ” at appropriate points.) Let $T_{\text{red}} \subseteq T$ (“red” standing for “reduced”) denote the set of terms of the form (2.4.2). If $s, t \in T_{\text{red}}$, we can form their product $s \cdot t$ in T , and then, as we have just seen, rearrange parentheses to get an element of T_{red} which is equivalent to $s \cdot t$ so far as evaluation at X -tuples of elements of groups is concerned. Let us call this element $s \odot t$. Thus, $s \odot t$ has the properties that it belongs to T_{red} , and that for any map $v: X \rightarrow |G|$ (G a group) one has $(s \cdot t)_v = (s \odot t)_v$. In the same way, given $s \in T_{\text{red}}$, we can obtain from $s^{-1} \in T$ an element we shall call $s^{(-)} \in T_{\text{red}}$, such that for any map $v: X \rightarrow |G|$, one has $(s^{-1})_v = (s^{(-)})_v$.

Are any further reductions possible? For a particular X -tuple of elements of a particular group there may be equalities among the values of different expressions of the form (2.4.2); but we are

only interested in reductions that can be done in all groups. No more are obvious; but can we be sure that some sneaky application of the group axioms wouldn't allow us to prove some two distinct terms (2.4.2) to have the same evaluations at all X -tuples of elements of all groups? (In such a case, we should not lose hope, but should introduce further reductions that would always replace one of these expressions by the other.) Let us formalize the consequences of the preceding observations, and indicate the significance of this question:

Lemma 2.4.3. *For each $s \in T$, there exists an $s' \in T_{\text{red}}$ (i.e., an element of T of one of the forms shown in (2.4.2)) such that*

$$(2.4.4) \quad \text{for every map } v \text{ of } X \text{ into any group } G, \quad s_v = s'_v \text{ in } |G|.$$

Moreover, if one of the following statements is true, all are:

- (i) For each $s \in T$, there exists a unique $s' \in T_{\text{red}}$ satisfying (2.4.4).
- (ii) If s, t are distinct elements of T_{red} , then “ $s = t$ ” is not an identity for groups; that is, for some G and some $v: X \rightarrow |G|$, $s_v \neq t_v$.
- (iii) The 4-tuple $F = (T_{\text{red}}, \odot, {}^{(-)}, e_T)$ is a group.
- (iv) The 4-tuple $F = (T_{\text{red}}, \odot, {}^{(-)}, e_T)$ is a free group on X .

Proof. We get the first sentence of the lemma by an induction, which I will sketch briefly. The assertion holds for elements $x \in X$: we simply take $x' = x$. Now suppose it true for two terms $s, t \in T$. To establish it for $s \cdot t \in T$, define $(s \cdot t)' = s' \odot t'$. One likewise gets it for s^{-1} using $s'^{(-)}$, and it is clear for e . It follows from condition (c) of the definition of “group-theoretic term” (Definition 1.5.1) that it is true for all elements of T .

The equivalence of (i) and (ii) is straightforward. Assuming these conditions, let us verify that the 4-tuple F defined in (iii) is a group. Take $p, q, r \in T_{\text{red}}$. Then $p \odot (q \odot r)$ and $(p \odot q) \odot r$ are two elements of T_{red} , call them s and t . For any $v: X \rightarrow |G|$, $s_v = t_v$ by the associative law for G . Hence by (ii), $s = t$, proving that \odot is associative. The other group laws for F are deduced in the same way.

Conversely, assuming (iii), we claim that for distinct elements $s, t \in T_{\text{red}}$, we can prove, as required for (ii), that the equation “ $s = t$ ” is not an identity by getting a counterexample to that equation in this very group F . Indeed, if we let v be the inclusion $X \rightarrow T_{\text{red}} = |F|$, we can check by induction on n in (2.4.2) that for all $s \in T_{\text{red}}$, $s_v = s$. Hence $s \neq t$ implies $s_v \neq t_v$, as desired.

Since (iv) certainly entails (iii), our proof will be complete if we can show, assuming (iii), that F has the universal property of a free group. Given any group G and map $v: X \rightarrow |G|$, we map $|F| = T_{\text{red}}$ to $|G|$ by $s \mapsto s_v$. From the properties of \odot and ${}^{(-)}$, we know that this is a homomorphism h such that $h|_X$ (the restriction of h to X) is v ; and since X generates F , h is the unique homomorphism with this property, as desired. \square

Well – are statements (i)-(iv) true, or not??

The usual way to answer this question is to test condition (iii) by writing down precisely how the operations \odot and ${}^{(-)}$ are performed, and checking the group axioms on them. Since a term of the form (2.4.2) is uniquely determined by the integer n (which we take to be 0 for the term e) and the n -tuple of elements of X and their inverses, $(x_n^{\pm 1}, \dots, x_1^{\pm 1})$, one describes \odot and ${}^{(-)}$ as operations on such n -tuples. E.g., one multiplies two tuples (w, \dots, x) and (y, \dots, z) (where each of w, \dots, z is an element of X or a symbolic inverse of such an element) by uniting them as

$(w, \dots, x, y, \dots, z)$, then dropping pairs of factors that may now cancel (e.g., x and y above if y is x^{-1}); and repeating this last step until no such cancelling pairs remain.

But checking the associative law for this recursively defined operation turns out to be very tedious, involving a number of different cases. (E.g., you might try checking associativity for $(v, w, x) \cdot (x^{-1}, w^{-1}, y^{-1}) \cdot (y, w, z)$, and for $(v, w, x) \cdot (x^{-1}, z^{-1}, y^{-1}) \cdot (y, w, z)$, where w, x, y and z are four distinct elements of X . Both cases work, but they are different computations.)

But there is an elegant trick, not as well known as it ought to be, which rescues us from the toils of this calculation. We construct a certain G which we *know* to be a group, using which we can verify condition (ii) – rather than condition (iii) – of the above lemma.

To see how to construct this G , let us go back to basics and recall where the group identities, which we need to verify, “come from”. They are identities which are satisfied by *permutations* of any set A , under the operations of composing permutations, inverting permutations, and the identity permutation. So let us try to describe a set A on which the group we want to construct should act by permutations in as “free” a way as possible, specifying the permutation of A that should represent the image of each $x \in X$.

To start our construction, let a be any symbol not in $\{x^{\pm 1} \mid x \in X\}$. Now define A to be the set of all strings of symbols of the form:

$$(2.4.5) \quad x_n^{\pm 1} x_{n-1}^{\pm 1} \dots x_1^{\pm 1} a$$

where $n \geq 0$, each $x_i \in X$, and no two successive factors $x_i^{\pm 1}$ and $x_{i+1}^{\pm 1}$ are an element of X and the inverse of that same element, in either order.

In particular, taking $n = 0$, note that $a \in A$.

Let G be the group of *all* permutations of A . Define for each $x \in X$ an element $v(x) \in |G|$ as follows. Given $b \in A$,

- if b does *not* begin with the symbol x^{-1} , let $v(x)$ take b to the symbol xb , formed by putting an x at the beginning of the symbol b ;
- if b does begin with x^{-1} , say $b = x^{-1}c$, let $v(x)$ take b to the symbol c , formed by removing x^{-1} from the beginning of b .

It is immediate from the definition of A that $v(x)(b)$ belongs to A in each case. To check that $v(x)$ is invertible, consider the map which sends a symbol xb to b , and a symbol c not beginning with x to the symbol $x^{-1}c$; we find that this is a 2-sided inverse to $v(x)$.

So we now have a map $v: X \rightarrow |G|$. As usual, this induces an evaluation map $s \mapsto s_v$ taking the set T of terms in X into $|G|$. Now consider any $s = x_n^{\pm 1} (\dots (x_2^{\pm 1} x_1^{\pm 1}) \dots) \in T_{\text{red}}$. It is easy to verify by induction on n that the permutation $s_v \in |G|$ takes our “base” symbol $a \in A$ to the symbol $x_n^{\pm 1} \dots x_1^{\pm 1} a$ (or if $s = e$, to a itself). It follows that if s and t are distinct elements of T_{red} , $s_v(a)$ and $t_v(a)$ are distinct elements of A , so $s_v \neq t_v$ in G , establishing (ii) of Lemma 2.4.3. By that lemma we now have

Proposition 2.4.6. $F = (T_{\text{red}}, \odot, (-), e)$ is a group; in fact, letting u denote the inclusion $X \rightarrow T_{\text{red}}$, the pair (F, u) is a free group on X .

Using parenthesis-free notation for products, and identifying each element of X with its image

in F , this says that every element of the free group on X can be written uniquely as

$$e, \quad \text{or} \quad x_n^{\pm 1} \cdots x_2^{\pm 1} x_1^{\pm 1},$$

where in the latter case, $n \geq 1$, each $x_i \in X$, and no two successive factors $x_i^{\pm 1}$ and $x_{i+1}^{\pm 1}$ are equal to an element of X and the inverse of that same element, in either order. \square

We have obtained what is called a *normal form* for elements of a free group on X – a set of expressions which contains a unique expression for each member of the group, such that we can algorithmically reduce any expression to one in this set. This indeed allows us to calculate explicitly in the free group; e.g., you should find it straightforward to do

Exercise 2.4:1. Determine whether each of the following equations holds for all elements x, y, z of all groups:

- (i) $(x^{-1}yx)^{-1}(x^{-1}zx)(x^{-1}yx) = (yx)^{-1}z(yx)$.
- (ii) $(x^{-1}y^{-1}xy)^2 = x^{-2}y^{-1}x^2y$.

In the next exercise, we will use the group theorists' abbreviations

$$x^y = y^{-1}xy \quad \text{and} \quad [x, y] = x^{-1}y^{-1}xy$$

for the *conjugate* of an element x by an element y in a group G , respectively the *commutator* of two elements x and y . If H_1, H_2 are subgroups of G , then $[H_1, H_2]$ denotes the *subgroup of G generated by all commutators $[h_1, h_2]$ with $h_1 \in H_1$ and $h_2 \in H_2$* .

Exercise 2.4:2. (i) Show that the commutator operation is not associative; i.e., that it is not true that for all elements a, b, c of every group G one has $[a, [b, c]] = [[a, b], c]$.

(ii) Prove a group identity of the form

$$[[x^{\pm 1}, y^{\pm 1}], z^{\pm 1}]y^{\pm 1} [[y^{\pm 1}, z^{\pm 1}], x^{\pm 1}]z^{\pm 1} [[z^{\pm 1}, x^{\pm 1}], y^{\pm 1}]x^{\pm 1} = e,$$

for some choice of the exponents ± 1 . (There is a certain amount of leeway in these exponents; you might try to adjust your choices so as to get maximum symmetry. The result is known as *Phillip Hall's identity*; however its form may vary with the text, depending on whether the above definition of $[x, y]$, preferred by most contemporary group theorists, is used, or the less common definition $xyx^{-1}y^{-1}$.)

(iii) Deduce that if A, B and C are subgroups of a group G such that two of $[[A, B], C]$, $[[B, C], A]$, $[[C, A], B]$ are trivial, then so is the third. (The “three subgroups theorem”.)

(iv) Deduce that if A and B are two subgroups of G , and $[A, [A, B]]$ is trivial, then so is $[[A, A], B]$. Is the converse true?

Incidentally, group theorists generally abbreviate $[[x, y], z]$ to $[x, y, z]$. If I worked with commutators every day I might do the same, but as an occasional visitor to the subject, I prefer to stick with more transparent notation.

The idea of finding normal forms, or other explicit descriptions, of objects defined by universal properties is a recurring one in algebra. The form we have found is specific to free groups. It might appear at first glance that corresponding forms could be obtained mechanically from any finite system of operations and identities; e.g., those defining rings, lattices, etc.; and thus that the results of this section should generalize painlessly (as those of the two preceding sections indeed do!) to very general classes of structures. But this is not so. An example we shall soon see (§3.5) is that of the Burnside problem, where a sweet and reasonable set of axioms obstinately refuses to yield a normal form. Other nontrivial cases are free Lie algebras [73] (cf. §8.7 below) and free

lattices [3, §VI.8], for which normal forms are known, but complicated; free modular lattices, for which it has been proved that the word problem is undecidable (no recursive normal form can exist); and groups defined by particular families of generators and relations (§3.3 below), for which the word problem has been proved undecidable in some cases, while nice normal forms exist in others. In general, normal form questions must be tackled case by case, but for certain large families of cases there *are* interesting general methods [40]. I hope eventually to add a chapter on that subject to these notes.

The trick that we used to show that the set of terms T_{red} constitutes a normal form for the elements of the free group is due to van der Waerden, who introduced it in [123] to handle the more difficult case of coproducts of groups (§3.6 below). Though the result we proved is, as we have said, specific to groups, the idea behind the proof is a versatile one: If you can reduce all expressions for elements of some universal structure F to members of a set T_{red} , and wish to show that this gives a normal form, then look for a “representation” of F (in whatever sense is appropriate to the structure in question – in the group-theoretic context this was “an action of the group F on a set A ”) which distinguishes the elements of T_{red} . A nice twist which often occurs, as in the above case, is that the object on which we “represent” F may be the set T_{red} itself, or some closely related object.

My development of Proposition 2.4.6 was full of motivations, remarks, etc.. You might find it instructive to write out for yourself a concise, direct, self-contained proof that the set of terms indicated in Proposition 2.4.6, under the operations described, forms a group, and that this has the universal property of the free group on X .

Exercise 2.4:3. If X is a set, and $s \neq t$ are two reduced group-theoretic terms in the elements of X (as in Lemma 2.4.3(ii)), will there in general exist a *finite* group G , and a map $v: X \rightarrow |G|$, such that $s_v \neq t_v$? (In other words, are the only identities satisfied by all finite groups those holding in all groups?)

If you succeed in answering the above question, you might try the more difficult ones in the next exercise.

Exercise 2.4:4. (i) If X is a set, F the free group on X , H a subgroup of F , and s an element of F such that $s \notin |H|$, will there in general exist a finite group G and a homomorphism $f: F \rightarrow G$ such that $f(s) \notin f(|H|)$?

(ii) Same question, under the assumption that the subgroup H is finitely generated.

Free groups can also be represented by matrices:

Exercise 2.4:5. Let $\text{SL}(2, \mathbb{Z})$ denote the group of all 2×2 matrices of integers with determinant 1, and let H be the subgroup thereof generated by the two elements $x = \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}$ and $y = \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix}$. Show that H is free on $\{x, y\}$. (Hint: Let c be the column vector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Examine the form of the column vector obtained by applying an arbitrary reduced group-theoretic word in x and y to c .)

If you do the above, you might like to think further about what pairs of (possibly distinct) integers, or for that matter, what pairs of real or complex numbers can replace the two “3”s in the above matrices. For integers the answer is known; for rational, real and complex numbers, there are many partial results (see [64]), but nothing close to a complete answer at present.