A Companion to Lang's *Algebra*

(Last revised August 25, 2006; based on 3rd Ed., 4th printing of revised Springer version)

notes by George M. Bergman

In these notes, I attempt to bring together in an orderly arrangement materials I have given to my classes over the years when teaching Berkeley's basic graduate algebra course, Math 250, from Lang's *Algebra* – motivations, explanations, supplementary results and examples, advice on material to skip, etc.. I follow this expository material with some exercises not in the text that I particularly like, together with notes on a few of the exercises in the text.

This version of the handout is based on many passes through the first Semester of the course, Math 250A, but only one occasion when I taught 250B from Lang, in Spring 1997. Thus, the coverage of the 250A material is more complete and probably better written than that of the 250B material. Sections of Lang that I have not used in the course are in general not covered.

Order of topics. Math 250A generally treats group theory in some depth, gives some elementary category theory and basic results on rings and modules, and then an extensive development of Galois theory. (Linear algebra, though logically in the same class of fundamental algebraic subjects as the above, is generally postponed till the second Semester, 250B, so that Galois theory can be covered adequately in 250A.) As shown in the list below, in using Lang I follow approximately his ordering of this material, with a few deviations.

> Logical Prerequisites (pp.ix-x). Appendix 2.1. §§I.1-6. Appendices 2.2-2.4. §§I.7-12 except §I.10.
> §§II.1-2. §§III.1-4 except the latter half of §III.2 (from 3d line of p.125). §§II.3-5. §§III.5-6.
> §§IV.1-6 except §5. §§V.1-6. §§VI.1-9, but ending VI.2 at p.273 and VI.9 at p.299. §VIII.1.

In these notes, for the convenience of those not necessarily following the above order, I have put the commentary on the various sections in the order in which these occur in Lang. However, in cases where the order of the above list deviates from that in Lang, my comments on the sections occurring later in the above list may refer to some sections that occur earlier in that list but later in Lang; e.g., I assume Zorn's Lemma (Lang's Appendix 2.2) in discussing free abelian groups (Lang's §I.7), and I assume basic results on modules, and the concept of an algebra (§§III.1-4) when discussing the ring-theoretic ideas of §§II.3-5.

The material of 250B is less fixed than that of 250A. The additional sections that I covered the one recent time I taught 250B using Lang, and for which there are therefore notes in this Companion, are

> §§III.7-9, §X.4, §§XIII.1-7, §§XIV.1-3, §§XVI.1-8.

Pre-comments, post-comments, etc.. Some of my comments are written to be read *before* you read the indicated passage in Lang, to help prepare you for what is done there; others discuss what he has done, assuming you have already read the passage. I will mark these two kinds of comments [<] and [>] respectively. Comments which are to be read right when you reach a certain point in the text will be marked [=]. Finally, the mark [~] means that the order relative to the passage in Lang is not critical.

Long comments. Though most of the items in these notes are short comments on specific points in the text, there are also some longer passages. Some of these develop material which Lang leaves out but which I wished to include in the course (e.g., on structures of groups of small orders), others concern heuristics, or motivate and discuss an open problem. The beginnings and ends of such passages are marked with the symbols ⌉ and ⌋ in the right-hand margin.

Numbering. Results in Lang are referred to here by their chapter number, followed by the section-and-result number they have in that chapter of Lang. E.g., what Lang refers to as Theorem 5.1 of Chapter VI is referred to here as Theorem VI.5.1. Results proved in these notes will be numbered similarly, but with a ''c'' (for ''Companion'') before the final number; e.g., my notes to Lang's §I.6 contain a series of results, numbered Lemma I.6.c1, Lemma I.6.c2, etc.. Labeled displays in this Companion will likewise be numbered (c1), (c2), etc..

The collection of exercises near the end of this Companion is arranged according to the *section* of Lang to which the exercises are related, and numbered accordingly. E.g., the first exercise relating to Lang's §III.2 is numbered III.2:1. In addition, I give notes regarding some of the exercises in Lang. I refer to the

nth exercise at the end of Chapter M as ''Exercise M:Ln'' (e.g., VI:L8), since Lang does not classify his exercises by section.

Errata. I include, amid the notes to the text and to the exercises, indications of errata that I think are worth bringing to the reader's attention. I also made a practice of sending lists of errata to Lang, and many of these (especially those that are simple typographical errors) were corrected as of the next printing. In these notes I give after the Exercises a list showing the errata to earlier printings of the third edition that have subsequently been corrected (in case you have a copy from one of those printings), as well as errata to the current printing that are not important enough to note below (e.g., misspellings, and symbols in wrong fonts). If any reader discovers errata which are not listed either in the main part of these notes or in that supplementary list, I hope he or she will bring them to my attention.

How do you tell which printing of Lang you have? If you have the Springer version (the ''Revised Third Edition'', with the bright yellow-and-white cover), look at the back of the title page. Four lines from the bottom you will find a line beginning ''9 8 7 ...''. If it goes all the way down to ''1'', you have the *first* printing, if it only goes down to ''2'', the *second*, etc.. If you have the Addison-Wesley version (with a patterned pale blue cover), look at the last line of the back of the title page. If this line begins ''1 2 3 ...'' you have the first printing, if it begins ''2 3 4 ...'' you have the second printing, etc..

On Lang's ''Examples''. Lang frequently gives ''examples'' taken from advanced areas of mathematics which beginning graduate students cannot be expected to be familiar with. Students need not worry if they do not have the background assumed in such examples. I make clear below which they are.

Google Book Searching through Lang. If you want to look for some phrase in Lang and the index doesn't help, it may be useful to go to `https://books.google.com/`, type in

`inauthor:Lang intitle:algebra date:2002-2002`

and then, in the resulting display, click on ''About this book'', and type what you are looking for into the field ''Search in this book''. The results are of limited accuracy – I have run into important errors due to faulty character-recognition on poorly digitized pages; but sometimes the results have been helpful. I'd be interested in hearing whether you find this technique useful.

Additional handouts; web access. There are some handouts that I use in teaching 250A, either expository material or sets of exercises, which are not specific to this text. If I were to incorporate those handouts into the body of these notes, then every time I modified them, I would have to modify both the version in these notes, and the text-independent version, and I probably wouldn't do a good job coordinating the changes. Hence, they constitute separate sheets, which I have included at the end of this Companion. The points where these handouts are to be used will be noted in this Companion. The handouts are as follow. (Note that 2 pp. = one sheet.)

The Axiom of Choice, Zorn's Lemma, and all that, 4 pp.
A principal ideal domain that is not Euclidean, developed as a series of exercises, 1 p.
Lüroth's Theorem and some related results developed as a series of exercises, 2 pp.
Quadratic reciprocity developed from the theory of finite fields as a series of exercises, 2 pp.
Solution in radical of polynomial of degree ≤ 4*, developed as a series of exercises*, 2 pp.

Some other handouts, which I use in 250B, are not included with this Companion, but can (along with the above) be found through my web page, `/~gbergman` .

Other texts. Here are a few books other than Lang that you might like to look at for alternative developments of the same and related material:

Thomas Hungerford, *Algebra* (the other commonly used text for Math 250A)
B. L. van der Waerden, *Modern Algebra* (an old classic)
David S. Dummit and Richard M. Foote, *Abstract Algebra* (more elementary than Lang, with many detailed examples)
P. M. Cohn, *Basic Algebra*, and for some Math 250B material its sequel, *Further Algebra and Applications*.

The next few should also be good, but I haven't had a chance to look closely at them:

Nathan Jacobson, *Basic Algebra*

S. Mac Lane and G. Birkhoff, *Algebra* (not to be confused with G. Birkhoff and S. Mac Lane, *A Survey of Modern Algebra*, an undergraduate text)
I. M. Isaacs, *Algebra, a Graduate Course*
Joseph J. Rotman, *Advanced Modern Algebra.*
N. Bourbaki, *Éléments de Mathématique* (an encyclopedic development), various parts.

For a motivated development of the topics of category theory and universal algebra (a.k.a. ‘‘general algebra’’), introduced in §§I.11-12 of Lang, I recommend some course notes I have been developing for Math 245:

G. M. Bergman, *An Invitation to General Algebra and Universal Constructions. 398 pp.*. This can be bought at my office for \$40, or from the publisher Henry Helson for \$45 by mail, or viewed online via my web page. Below, I will refer to it as ‘‘my Math 245 notes’’.

NOTES ON THE TEXT

Re ‘‘Logical prerequisites’’. pp.ix-x.

P.*ix*, first two paragraphs [>]: There are three symbols, $\subseteq$, $\subset$, and $\subsetneqq$, one or another pair of which are used by different authors for the two concepts ‘‘is a subset of’’ and ‘‘is a proper subset of’’. Some authors use $\subseteq$ and $\subset$ respectively, paralleling the use of $\leq$ and $<$, while others (in particular, many Eastern Europeans, and Lang) use $\subset$ for ‘‘is a subset of’’ on the grounds that the more basic concept should have the simpler symbol, and then use $\subsetneqq$ (if anything) to explicitly write ‘‘is a proper subset of’’ (e.g., in Lang, p.112, second display). To avoid ambiguity, I will (unlike Lang) use $\subseteq$ for ‘‘is a subset of’’, while (like Lang) I will use $\subsetneqq$ for ‘‘is a proper subset of’’ on the rare occasions when a symbol for proper inclusion is needed.

When f is a map and A a subset of its domain, the notation $f(A)$ for the image of A under f is, strictly speaking, ambiguous, because some subset of the domain might also be an *element* of the domain. Like Lang, we shall use this as a convenient notation; but we will keep in mind that it should be avoided when it could lead to confusion.

Note (if you have taken a course using different notation) that functions will in general be written to the *left* of their arguments, and composed accordingly.

P.*x*, end of page [>]: A *commutative square* is so called because (as a member of the class once elegantly put it), it has the property that ‘‘going to the right commutes with going down’’. The term commutativity is extended to diagrams of other shapes by analogy with this case.

In Lang’s statement that to verify that a diagram composed of squares and triangles is commutative, ‘‘it suffices to verify’’ that each triangle and square in it is commutative, replace ‘‘suffices’’ by ‘‘often suffices (depending on the layout of the diagram)’’. Cf. Exercise 0:1 in this Companion for more details.

An indexed family of elements, which Lang writes $\{f(i)\}_{i\in I}$, I shall write $(f(i))_{i\in I}$. Lang’s notation is misleading, because it looks like set-notation, yet distinct families can have the same sets of elements. (Note, incidentally that the product set $A_1 \times ... \times A_n$ referred to by Lang near the bottom of this page is constructed as the set of all families $(a_i)_{i\in\{1,\,...\,,\,n\}}$ with $a_i \in A_i$ for all i.)

The concept of an *equivalence relation* on a set A may be thought of as an abstraction of the properties of the relation $\{(a,a') \mid f(a) = f(a')\}$ induced by a function $f\colon A \to B$. Its importance grows out of the fact that this relation determines the ‘‘form’’ of the image of A under f. Given an equivalence relation $\sim$ on a set A, the quotient set $A/\!\sim$ is a set B with a function $q\colon A \to B$ which induces (in the above sense) precisely the given equivalence relation $\sim$. Although the ‘‘set of equivalence classes’’ construction is a convenient way to prove the existence of such a B, I advise against thinking of the elements of $A/\!\sim$ always as subsets of A; conceptually, I prefer to think of $A/\!\sim$ as ‘‘a set obtained from A by gluing together certain of its elements in a way specified by $\sim$’’.

If $f: A \to B$ is a map, and $\sim$ the relation $\{(a_1, a_2) \mid f(a_1) = f(a_2)\}$, then there is a natural bijection between $A/\sim$ and $f(A) \subseteq B$. But there is a still more useful statement, (ii) below:

$$\begin{tikzcd} A \arrow[r, "q"] \arrow[dr, "\varphi"'] & A/\sim \arrow[d, "f"] \\ & B \end{tikzcd}$$

Lemma 0.c1. *Suppose* $\sim$ *is an equivalence relation on a set* A. *Then the canonical map* $q: A \to A/\sim$ *has the properties* (i) $a \sim a' \Leftrightarrow q(a) = q(a')$, *and* (ii) *if* $\varphi: A \to B$ *is any map such that* $a \sim a' \Rightarrow \varphi(a) = \varphi(a')$, *then there exists a unique map* $f: A/\sim \to B$ *such that* $\varphi = fq$, *i.e., making the diagram at right commute.* $\square$

Statement (ii) is a formalization of the ritual of verifying that the map $(A/\sim) \to B$ is ''well defined''. (Lang's discussion of this point does not distinguish clearly between the above ''f'' and ''φ''. It can be corrected by changing the words ''first give its value on an element'' on p.x, 9th and 10th lines from bottom, to ''express its value as a function of an element''.)

Lang mentions that Zorn's Lemma will be assumed. This is developed in his Appendix 2.2, and in my handout on ''The Axiom of Choice, Zorn's Lemma, and all that'', which, as indicated in the notes below, is to be read before §I.7.

Chapter I. Groups.

Re §I.1. Monoids.

P.3, bottom [>]: An element e satisfying $ex = x = xe$ for all x is variously called an ''identity element'', a ''unit'', or a ''neutral element''. The first term invites confusion with *identities* satisfied by particular groups or monoids (e.g., the commutative identity), the second is also the term used for any invertible element of a ring, and the last is awkward; so I can only recommend using whatever term has the least disadvantages in a given situation.

A set with an associative operation, but not necessarily having a neutral element, is called a *semigroup*. I actually prefer to call a monoid a ''semigroup with neutral element'', but the term ''monoid'' has become standard, so in these notes I will follow Lang and use it.

Lang will give some examples on pp.6-7, but let me give some quick examples of semigroups and monoids here. (a) *Fundamental example*: If X is a set, then the set of all maps $X \to X$ forms a monoid under composition, with the identity map as neutral element. (b) We can make any set X a semigroup by defining $xy = x$ for all $x, y \in X$. (This is called the ''left zero multiplication''. Setting $xy = y$, we likewise get the ''right zero multiplication''.) (c) If X is a set, we can make the set of all its subsets a monoid, by defining $st = s \cap t$.

How does one formalize the concept of a monoid as a set ''given with'' an operation? If one wants to be precise, one regards the monoid as an ordered pair, $G = (|G|, \mu)$, where $|G|$ is the underlying set of the monoid, and $\mu: |G| \times |G| \to |G|$ the operation. This is done in any course on universal algebra (and in my Math 245 notes). Note that where we now write $g \in G$, the above formalization would have us write $g \in |G|$. An alternative preferred by many workers in universal algebra, which allows one to keep writing $g \in G$, is to write $\mathbf{G} = (G, \mu)$. In this course, we simply note that the formalism of ordered pairs (or ordered 3-tuples, since there are good arguments for writing a monoid as $(|G|, \mu, e)$, even though the first two components determine the third) can be used when one needs to make things precise; but we will stick to standard notation here.

P.6, line after first display [>]:

On set-indexed products in commutative monoids.

Let me try to give a more precise and better organized derivation of some consequences of the commutative law. Worked out in detail, it becomes rather lengthy, but I think there is a certain elegance in the description that develops.

Lemma I.1.c1. *Let G be a commutative monoid. Then there is a unique way of associating to every finite set I, and I-tuple $(x_i)_{i\in I}$ of elements of G, an element of G which we will denote $\prod_I x_i$, in such a way that*

(i) (*Empty product*) *Taking $I = \varnothing$, we have*

$$\prod_{\varnothing} x_i \;=\; e.$$

(ii) (*Recursive step*) *If $(x_i)_{i\in I}$ is any I-tuple and j any element of I, we have*

$$\prod_I x_i \;=\; (\prod_{I-\{j\}} x_i)\, x_j.$$

Proof. We shall prove inductively that for each n, there is a unique way of defining products of I-tuples of elements of G for those I with $\mathrm{card}(I) \leq n$, such that (i) and (ii) hold.

The case $n = 0$ is clear from (i). So let $n > 0$, and assume by induction that our result is true for all index sets of cardinality $< n$.

Given a set I of cardinality n, the right-hand side of the display in (ii) is defined, by our inductive assumption, for every choice of $j \in I$; we claim that all these choices give the same value. Indeed, suppose j_1 and j_2 are distinct elements of I. Using j_1, we get the value $(\prod_{I-\{j_1\}} x_i)\, x_{j_1}$. Now applying (ii) to the set $I-\{j_1\}$, we see that this value equals $((\prod_{I-\{j_1,j_2\}} x_i)\, x_{j_2})\, x_{j_1}$. Similarly, when we use $j = j_2$, we get $((\prod_{I-\{j_1,j_2\}} x_i)\, x_{j_1})\, x_{j_2}$. We verify that these are equal using the associative and commutative laws:

$$\begin{aligned}((\prod_{I-\{j_1,j_2\}} x_i)\, x_{j_2})\, x_{j_1} \;&=\; (\prod_{I-\{j_1,j_2\}} x_i)(x_{j_2}\, x_{j_1})\\ &=\; (\prod_{I-\{j_1,j_2\}} x_i)(x_{j_1}\, x_{j_2}) \;=\; ((\prod_{I-\{j_1,j_2\}} x_i)\, x_{j_1})\, x_{j_2}. \quad \square\end{aligned}$$

We note two simple special cases, the descriptions of products over 1- and 2-element sets:

$$\prod_{\{i_0\}} x_i \;=\; x_{i_0}, \qquad\qquad \prod_{\{i_0,\, i_1\}} x_i \;=\; x_{i_0}\, x_{i_1}.$$

We also see that whenever I is a set $\{1, \dots, n\}$, our $\prod_I x_i$ agrees with the meaning given to this symbol for arbitrary monoids on p.4 of Lang. Another result quickly verified by induction is that for any finite set I,

$$\prod_I e \;=\; e.$$

The next result describes how, in products defined as above, one can collect terms:

Lemma I.1.c2. *Let $f\colon I \to J$ be a map of finite sets, and $(x_i)_{i\in I}$ an I-tuple of elements of a commutative monoid G. Then*

$$\prod_{i\in I} x_i \;=\; \prod_{j\in J}(\prod_{i\in f^{-1}(j)} x_i).$$

Proof. We shall again use induction on $\mathrm{card}(I)$.

When $\mathrm{card}(I) = 0$, point (i) of the preceding lemma shows that the left-hand side of the desired equation is e, and that the right-hand side is $\prod_J e$, which we have noted also equals e.

So now suppose $\mathrm{card}(I) > 0$, and assume inductively that the lemma holds for index sets of smaller cardinality. Take any $i_0 \in I$, and let $j_0 = f(i_0)$. By our recursive definition of $\prod_I x_i$, i.e., condition (ii) of the preceding lemma, we have

$$\prod_{i\in I} x_i \;=\; (\prod_{i\in I-\{i_0\}} x_i)\, x_{i_0}.$$

If we now let f' denote the restriction of $f: I \to J$ to $I-\{i_0\}$, our inductive hypothesis says we can write the right-hand side of the above equation as $(\prod_{j\in J}(\prod_{i\in f'^{-1}(j)} x_i))\, x_{i_0}$. Again applying condition (ii), this time to the index-set J and its element j_0, this becomes

$$((\prod_{j\in J-\{j_0\}}(\prod_{i\in f'^{-1}(j)} x_i))(\prod_{i\in f'^{-1}(j_0)} x_i))\ x_{i_0}.$$

We can now use associativity to bracket together the terms $(\prod_{i\in f'^{-1}(j_0)} x_i)\, x_{i_0}$, and apply condition (ii) to reduce their product to $\prod_{i\in f^{-1}(j_0)} x_i$. We also note that for $j \neq j_0$, the set $f'^{-1}(j)$ is the same as $f^{-1}(j)$; thus our product equals $(\prod_{j\in J-\{j_0\}}(\prod_{i\in f^{-1}(j)} x_i))(\prod_{i\in f^{-1}(j_0)} x_i)$, which by a final application of (ii) reduces to $\prod_{j\in J}(\prod_{i\in f^{-1}(j)} x_i)$, as claimed. □

Having gotten the above results in the context of finite products, where we could use induction, we may now generalize them to products that are possibly infinite but ''essentially finite'':

Lemma I.1.c3. *There is a unique way of associating to every* (*not necessarily finite*) *set* I, *and every* I*-tuple* $(x_i)_{i\in I}$ *of elements of* G *almost all of whose terms are* e, *an element* $\prod_I x_i \in G$, *in such a way that*

(iii) *For finite* I, *this value agrees with that defined by Lemma I.1.c1, and*

(iv) *If* $(x_i)_{i\in I}$ *is such an* I*-tuple, and* I_0 *is a subset of* I *containing all* i *such that* $x_i \neq e$, *then* $\prod_I x_i = \prod_{I_0} x_i$.

Moreover, this extended operation still has the properties listed in Lemmas I.1.c1 and I.1.c2.

Sketch of Proof. Given an I-tuple $(x_i)_{i\in I}$ as in the above statement, let us take any *finite* subset $I_0 \subseteq I$ containing all i such that $x_i \neq e$, and define $\prod_I x_i = \prod_{I_0} x_i$. To show that this is well-defined, we want to show that this product is the same whatever I_0 we choose. Let us first consider the case of two choices one of which is contained in the other, $I_0 \subseteq I_1$. Let $f: I_1 \to \{0,1\}$ be defined by $f(i) = 0$ if $i\in I_0$, $f(i) = 1$ otherwise, and let us apply the preceding lemma to this map f. This gives an expression of $\prod_{I_1} x_i$ as the product of two factors, one of which is $\prod_{I_0} x_i$, while the other is a product of e's, which we know equals e. Hence $\prod_{I_1} x_i = \prod_{I_0} x_i$. Now given two sets I_0, I_1 neither of which is assumed contained in the other, we can apply the above case twice to get $\prod_{I_1} x_i = \prod_{(I_1\cap I_0)} x_i = \prod_{I_0} x_i$, as required.

It is straightforward to verify that the properties we established earlier for finite products carry over to these ''essentially finite'' products, using (iii) and (iv) above. □

We see that the results at the tops of pages 5 and 6 of Lang are particular cases of Lemma I.1.c2 and its generalization to essentially finite products: we get the first by taking $I = J = \{1,\ldots,n\}$ and letting f be a permutation; the second by considering a product-set $I\times J$ and its projections onto the two factors I and J.

The concept of homomorphism allows us to state one more important property of set-indexed products, easily proved by induction on $\mathrm{card}(I)$ when I is finite, and using condition (iv) of the preceding lemma in the possibly infinite case:

Lemma I.1.c4. *If* $h: G \to H$ *is a homomorphism of commutative monoids, and* $(x_i)_{i\in I}$ *a family of elements of* G, *almost all of which equal* e, *then*

$$\prod_I h(x_i) \;=\; h(\prod_I x_i).\ \square$$

Re §I.2. Groups.

P.7, 7th line from bottom [=]: After ''*assume that every element has a left inverse*'', add ''*with respect to* e''.

P.8, after first 3 lines [=]: Lang has proved that one-sided inverses are two-sided inverses under appropriate assumptions. For an example where such conditions are not satisfied, consider the monoid G of maps of $\mathbf{Z}$ into itself, and the element $a\in G$ defined by $a(n) = 2n$; you should be able to find an inverse to a on one side, but show that a has no inverse on the other. Can you find a semigroup containing a one-sided unit that is not a two-sided unit?

P.8, first Example [<]: By $M(S,G)$, Lang means the set of all set-maps from S to G. This is more commonly written G^S.

P.8, second example [>]: Lang has probably included the word ''nonempty'' because some students are confused about maps involving the empty set $\varnothing$, and he doesn't consider it worth the bother to deal with. But if we ignore these problems now, they will take revenge later, so let us get things straight. Recall that a map $f\colon X \to Y$ is (or at least, is determined by) a set of ordered pairs, $\{(x,\ f(x)) \mid x\in X\}$. We now note that

if X *is a nonempty set, there is no map* $X \to \varnothing$, *but*

for any X, *there is a unique map* $\varnothing \to X$, *namely the map whose graph is the empty set of pairs.*

Hence, there is exactly one map $\varnothing \to \varnothing$, and this is a permutation. So we may delete the word ''nonempty'' from the definition of $\mathrm{Perm}(S)$, and note that $\mathrm{Perm}(\varnothing)$ is a one-element group.

P.8, ''**Example. The group of automorphisms**'' [<]: I *don't* advise following Lang's recommendation unless you are already familiar with category theory. If you are not, skip this ''example''.

P.9, line 2 [>]: The term ''cyclic'' is natural for this class of groups because of the form the finite ones have. By extension, the term is also frequently used to mean ''generated by a single element'' when referring to any sort of mathematical object, even if the forms of these objects are not ''cycle''-like.

It is interesting to examine the structures of *cyclic monoids*. Their description is not quite as simple as that of cyclic groups, but not too hard to work out. The details can be entertaining to develop in class, or you might like to work them out for yourself.

P.10, top paragraph [>]: Logically, the definition of a homomorphism of monoids belongs in the preceding section. Note that such a homomorphism is required not only to respect multiplication, but to take unit element to unit element. A map $G \to G'$ respecting multiplication will clearly send e_G to an element i of G', such that $i^2 = i$. Such an element is called an *idempotent*. In a *group*, it is easy to see that the only idempotent is the neutral element, so in the definition of a group homomorphism, many authors leave out the condition of respecting neutral elements. But a monoid can have many idempotent elements (e.g., in the monoid of subsets of a set, under the operation of intersection, all elements are idempotent), so the condition of respecting neutral elements is an essential part of the definition of a monoid homomorphism.

P.10, middle [>]: ''The existence of an isomorphism between two groups G and G' is sometimes denoted $G \approx G'$.'' Some other common notations for this relation are $G \cong G'$ and $G \simeq G'$.

P.10, italicized statement at the bottom of page [>]: Here are two ways of proving this. In each case, let us assume (modifying Lang's notation) that $p, q\colon G \to H$ are homomorphisms, both of which, when restricted to the generating set S, give the map $f\colon S \to H$. We wish to prove $p = q$.

Sketch of naive proof. Since G is generated by S, every element $g\in G$ may be written as a product of elements of S and their inverses. Use the fact that p and q respect composition and inverses to write $p(x)$ and $q(x)$ as expressions in the images of elements of S under f. We get the same expressions, hence $p(x)$ and $q(x)$ have the same value.

Good proof. It is immediate to verify that the set $A = \{x\in G \mid p(x) = q(x)\}$ is a subgroup of G, and contains S. Hence it contains the *least* subgroup containing S, i.e., the subgroup generated by S, which in this situation is assumed to be G. So A contains G, i.e., $p = q$.

The characterization used above of the subgroup of a group G generated by a set S, as the *least* subgroup containing S, is an example of what is called a *universal* characterization of a mathematical object, since it describes the subgroup in question in terms of its relation to *all* subgroups containing S. Universal characterizations sometimes yield much more direct proofs than do ''descriptions'' of the objects so characterized.

P.11, parenthetical remark containing first display [=]: This just repeats the proof of the first display on the preceding page.

P.11, italicized statement in middle of page [=]: For ''*a homomorphism*'' read ''*a group homomorphism*''.

P.11, Proposition I.2.1 [>]: This result may be looked at as the *converse* to an important observation: that if one forms the direct product G of two groups H and K, then H and K can be identified with subgroups of G (namely, $\{(h,e) \mid h\in H\}$ and $\{(e,k) \mid k\in K\}$), and that these two subgroups have trivial intersection, commute elementwise, and generate the whole group.

It is often convenient to state the conditions described in the conclusion of the above proposition, as saying that ''G decomposes as the direct product of the subgroups H and K''. But one should bear in mind that there is a subtle difference between the senses of ''being a direct product'' in this situation and as originally defined, since this G does not literally consist of ordered pairs of elements of H and of K (though the proof is based on associating to each $g\in G$ such a pair, namely the unique (h,k) such that $g = hk$); and inversely, in a direct product in the original sense, the given groups are not subgroups of G (though, as noted, they can be naturally identified with such subgroups). When one needs to distinguish these usages, one may say that a group G obtained by the original direct product construction is the ''external direct product'' of H and K, while in the situation of the proposition, G is the ''internal direct product'' of its subgroups H and K. When we come to category theory, we shall discover a unified concept of direct product which includes both of these.

P.12, middle, definition of ''coset representative'' [>]: An important concept (used in the next proposition) is that of ''a set of coset representatives''. This means a subset of G containing *exactly one* representative of each coset of H.

P.12, Proposition I.2.2 [=]: Lang asks us to think of the index-function $(G:H)$ as taking as values the finite cardinals and a nonspecific ''∞''. But as one can easily see from the second paragraph of this proposition, if one defines $(G:H)$ as the cardinality of the set of cosets of H in G, the displayed equation holds as a precise arithmetic statement about these cardinals, whether they are finite or infinite. (So, for instance, one can deduce that an uncountable group can't have a countable subgroup of countable index.)

P.13, end of §I.2 [>]:

Alias and alibi.

To avoid confusion in the study of groups of permutations, it is useful to be aware of what Hermann Weyl called ''the problem of alias and alibi''. As an example, suppose we are studying the permutation group on 5 elements, and we think of these permutations as corresponding to ways of putting marbles labeled $1,\ldots,5$ into holes labeled $1,\ldots,5$. Suppose we have a particular such permutation, say (23) (represented by the arrangement with marble 3 in hole 2, marble 2 in hole 3, and all other marbles in the holes with the same numbers), and we wish to describe the effect of following it by the permutation (12). Do we get this by interchanging the marbles that are currently in holes 1 and 2, or by interchanging the marbles with numbers 1 and 2? In the first case, we are letting S_5 act by *alibi* (Latin for ''elsewhere'', i.e., we are letting S_5 act by permuting the 5 *positions*); in the second, it acts by *alias* (Latin for ''otherwise''). For another example: consider the symmetry group of a square drawn on the blackboard. We might imagine a celluloid copy of this square made (say labeled $ABCD$, with A at the upper left), and we might then represent elements of this group as ways of placing this copy over the original square. But then, should the group contain operations such as ''flipping about the vertical axis of the blackboard square'' or ''flipping so as to interchange A with B, and C with D, in the celluloid copy''? If we start with the celluloid square in its original position, these have the same effect, but if we compose symmetries, so that one operation acts after another has moved the celluloid square, we in general get different results.

In these examples, note that what we have been calling ''permutations'' are really bijections between two sets, X (a set of ''places''), and Y (a set of objects), which are given us with some ''standard'' bijection between them. Our symmetry group can be regarded either as a group of permutations of X, or as a group of permutations of Y. Given a bijection $f\colon X \to Y$, the action by ''alias'' is that of composing f on the left with permutations of Y, and the action by ''alibi'' is that of composing f on the right with permutations of X. If we *identify* X with Y, using our ''standard'' bijection – so that the set of allowed bijections, and the symmetry group of X, and the symmetry group of Y all become the same set – then the action of an element g on the symmetry group by *alias* becomes the ''left translation'' map $x \mapsto gx$ on that group, and the action by *alibi* becomes the ''right translation'' map, $x \mapsto xg$.

Re §I.3. Normal subgroups.

P.14, top [~]: What general computational principle allows Lang to argue (in the first paragraph) that the condition $xH \subset Hx$ is equivalent to $xHx^{-1} \subset H$ and (in the next paragraph) to write $xHyH = xyHH = xyH$ if H is a normal subgroup? The key observation is that the *subsets* of a group or monoid G form a monoid under the multiplication of subsets defined on p.6:

$$SS' = \{xy \mid x \in S,\ y \in S'\}.$$

(Associativity of this operation is an immediate consequence of the associativity of G. The ''unit'' is the singleton $\{e\}$.) In this monoid, singletons clearly multiply like the corresponding elements: $\{x_0\}\{x_1\} = \{x_0 x_1\}$; also recall that we abbreviate the product $\{x\}S$ or $S\{x\}$ of a singleton $\{x\}$ and a set S as xS, respectively Sx.

Thus, given $xH \subset Hx$, we can multiply on the right by x^{-1} (i.e., $\{x^{-1}\}$) and get $xHx^{-1} \subset He = H$. You should be able to write down a similar argument showing that if H is a normal subgroup, then it ''commutes'' with all elements. Thus, we get $xHyH = xyHH = xyH$.

P.14, middle [<]: *Motivation for the construction of the factor group.* We noted in our comments on pp.*ix*-*x* that given a *set* G, to form a picture of the image of a map f with domain G it suffices to know the equivalence relation on G that describes which elements fall together under f. If G is a *group* and f a group homomorphism, we can then describe the *group structure* of the image, using the law $f(x)\,f(y) = f(xy)$; so the group structure of G and the above equivalence relation completely determine the structure of the group $f(G)$.

Also note that $f(x) = f(y)$ if and only if $f(xy^{-1}) = e$, hence to determine the equivalence relation, it suffices to know the set of elements which have image e. This set turns out to be a subgroup N, and to satisfy $y^{-1}Ny = N$ for all $y \in G$; a subgroup with these properties is called *normal*, and the normal subgroup of G associated in this way with a homomorphism f is called the *kernel* of f. Conversely, if N is a normal subgroup of G, then the quotient-set of G by the equivalence relation $\{(x,y) \mid xy^{-1} \in N\}$ can be made a group G/N in such a way that the quotient map $G \to G/N$ is a homomorphism having N as its kernel. These considerations show that homomorphic images of G are in one-to-one correspondence with normal subgroups of G. (To be more precise, suppose that given homomorphisms of a group G into two groups H_1 and H_2, we define an ''isomorphism of groups with homomorphisms of G into them'' to mean a group isomorphism $H_1 \to H_2$ making a commuting triangle with the given homomorphisms from G. Then normal subgroups of G correspond bijectively to homomorphic images of G, up to isomorphism as groups with homomorphisms of G to them.)

For monoids, on the other hand, the set of elements that a homomorphism sends to the neutral element does *not* determine the structure of the image; hence one must look at the whole equivalence relation to find that structure. Such an equivalence relation is called a *congruence* on a monoid, and one can write down necessary and sufficient conditions on an equivalence relation on a monoid for it to be a congruence.

Group-theorists generally write $H < G$ to mean ''H is a subgroup of G'', and $H \triangleleft G$ for ''H is a normal subgroup of G'' (and likewise $G > H$ or $G \triangleright H$ if they want to name the larger group first). We shall often use this notation.

Given any subset S of a group G, there is a *least* normal subgroup of G containing S. A few basic facts about this construction are noted in Exercise I.3:4. I recommend looking at that exercise and making sure you see why the assertions are true.

P.14, middle, the ''canonical map'' [>]: A ''canonical'' object means one determined in a special way by the data being considered.

Something you might find it interesting to think about: Suppose G is a group, H a subgroup, and S a set of left coset representatives of H in G. If H is normal, we know that left cosets are the same as right cosets, so these will also be a system of right coset representatives. If H is not normal, we may ask whether we can at least find *some* set of left coset representatives which is also a set of right coset representatives. You might consider both case where H has finite index in G and the general case.

P.15, three italicized sentences [=]: The first and last of these statements are fairly obvious (and, in fact, say the same thing); the middle one takes more thought, but is important. You should check that you can prove it.

P.15, final display [=]: Here ''0'' denotes the trivial group. That is a common symbol for it; ''1'' is often used when G is written multiplicatively. The explicit symbols $\{e\}$, $\{0\}$ and $\{1\}$ can also be used. (The notation ''1'' is implicit in the notation $(G:1)$ for the order of G.)

P.16, the commutative diagram [<]: You do not have to learn all the diagrammatic results proved over the next three pages; what you should concentrate on is understanding *how* they are proved. You should, however, learn the two boxed isomorphisms on p.17, which in most texts are stated as ''Isomorphism Theorems''. The ideas behind these are discussed below.

P.16, point (**ii**) [~]: Here we might just as well replace the subgroup H by an arbitrary subset S. The result obtained can be thought of as saying that the group G/N is *universal* among groups E given with homomorphisms $G \to E$ having S in the kernel.

P.17, (**iii**) and (**iv**) [>]:

Canonical isomorphisms.

The isomorphisms in the two boxes on p.17 relate subgroups and homomorphic images of a group G and of a factor-group thereof. An important fact, easily verified, but not mentioned by Lang, which will help us understand these results, is:

(c1) *Let $q: G \to G/N$ be the canonical map from a group to a factor-group. Then there is a bijective correspondence between subgroups of G/N and subgroups of G that contain N, the map from the former to the latter being given by $H \mapsto q^{-1}(H)$; the map the other way by $H \mapsto H/N = q(H)$. Moreover, H/N is normal in G/N if and only if H is normal in G.*

The first boxed isomorphism on p.17 can now be understood as follows: Suppose we have a group G, a homomorphic image of that group, and a homomorphic image of this homomorphic image. Thus, we can write the first of these images as G/K, where $K \triangleleft G$; and by the above observation, the second will have the form $(G/K)/(H/K)$ for some normal subgroup $H \triangleleft G$ containing K. Now we may also regard this group directly as a homomorphic image of G, i.e., its factor-group by the kernel of the composite map $G \to G/K \to (G/K)/(H/K)$. This kernel is clearly H, giving the asserted isomorphism $(G/K)/(H/K) \cong G/H$.

This result can be looked at in still another way: Given a group G and two normal subgroups K and H, with the former contained in the latter, the canonical map $G \to G/H$ has K within its kernel, hence induces a map $G/K \to G/H$, as in the triangular diagram in point (ii) on p.16. This expresses G/H as a homomorphic image of G/K, equivalently, as a factor-group thereof. The factor-group by what subgroup? H/K!

Note that the final statement of (c1) above concerns this same situation; but the boxed isomorphism tells us more than that statement, namely that corresponding normal subgroups of G and G/N have isomorphic factor-groups.

To motivate the other boxed isomorphism on p.17, again consider the canonical map from G to a factor-group G/K, and now let H be an arbitrary subgroup of G, and let us look for a description of the image of H in G/K under that map. On the one hand, this will be isomorphic to the factor-group of H by the subgroup of its elements that go to the identity element under the map; i.e., to $H/(H \cap K)$. On the other hand, it is a subgroup of G/K, hence has the form L/K, where L is the inverse image of this subgroup in G. This inverse image consists of those elements of G whose images in G/K are equal to images of elements of H. We easily verify that this says $L = HK$; so identifying our two descriptions of the image of H in G/K, we have $H/(H \cap K) \cong HK/K$.

(Lang actually assumes slightly less than we did in the above paragraph: where we assumed K normal, he just says that H lies in the normalizer of K. However, if we let G_0 denote that normalizer, then his assertion follows from our observations applied with the group G_0 in place of G.)

Important remark: Most often, a statement that something is ''isomorphic to'' something else is an incomplete statement. The full statement, which is frequently what one really uses when one applies the result, is that *a particular map* from one object to the other is an isomorphism; or that there is a *unique* isomorphism from one object to the other that *has a certain property*. This is indeed true in the above two cases. The first result should be formulated as saying that there is a unique isomorphism between $(G/K)/(H/K)$ and G/H that makes a commuting triangle with the canonical maps of G into these two groups – briefly, a unique ''isomorphism as images of G''; the second says, similarly, that there is a unique isomorphism between $H/(H \cap K)$ and HK/K as images of H.

P.18, definition of normal tower [>]: Though each G_i is assumed normal in the subgroup immediately above it in the tower, that does not mean it will be normal in *all* higher members of the tower: A normal subgroup of a normal subgroup need not be normal.

(However, if one defines a *characteristic subgroup* to be a subgroup carried into itself by *all* automorphisms of the group, then it is easy to see that a characteristic subgroup of a characteristic subgroup is characteristic. I leave it to you to think of what can be said of a characteristic subgroup of a normal subgroup, or a normal subgroup of a characteristic subgroup.)

P.18, proof of Proposition I.3.1 [<]: Before reading this, note the following

Observations on normal towers.

A key fact used in connection with normal towers is: If $G \triangleright N$ are a group and a normal subgroup, and we have a normal tower

$$G = H_1 \triangleright H_2 \triangleright \ldots \triangleright H_m = N,$$

then

$$(H_1/N) \triangleright (H_2/N) \triangleright \ldots \triangleright (H_m/N) = \{e\}$$

is a normal tower for G/N, and the successive factor-groups of this tower are isomorphic to those of the first tower via the first boxed isomorphism on p.17. Moreover, this construction gives a *bijective* correspondence between normal towers starting with G and ending with N, and normal towers starting with G/N and ending with $\{e\}$.

In particular, suppose we look at any pair of successive terms $G_i \triangleright G_{i+1}$ in a normal tower for some group G, and ask what sequences of terms may come between G_i and G_{i+1} in refinements of that tower. The above result shows that the possibilities correspond naturally to the normal towers that start with G_i/G_{i+1} and end with $\{e\}$.

This is used implicitly in the first sentence of the proof of Proposition I.3.1 to pass from a tower for the abelian group G_i/G_{i+1} to a tower between G_i and G_{i+1}, and again in the proof of Theorem I.3.5, where it is combined with the observation that a normal tower for a simple group merely has one factor isomorphic to the group and all other factors trivial.

P.18, fifth line of the proof of Proposition I.3.1 [=]: After ''tower in G''' add ''ending in $\{e\}$''.

P.19, proof of Theorem I.3.2 [<]: The proof that if G is solvable then G/H is solvable is essentially given on the preceding page, in the paragraph beginning ''Let $f: G \to G'$ be a homomorphism''.

P.20, above first example, definition of *simple group* [=]: This definition is a sensible one, but the usual usage of finite group-theorists is different: they call a group simple if it satisfies the property stated by Lang *and* is non-abelian. The reason is that the abelian simple groups, i.e., the cyclic groups of prime order, though straightforward in their structure, are quite different in properties from the other simple groups. In particular, note that for every group G, the subgroup G^c is normal, hence if G is simple (in Lang's sense), G^c is either G or $\{e\}$. The abelian simple groups are those for which G^c is $\{e\}$, while for all others it is G.

P.20, last two sentences of second paragraph of second example [>]: A group theorist tells me that the one known gap in the proof, the ''quasi-thin'' case, was filled in two papers totaling about 1,300 pages by Michael Aschbacher and Steve Smith, published in 2004.

Since the proof of the classification constitutes many thousands of pages of articles, there is still a possibility of hidden errors. People in the field have been working on simplifying parts of the proof, so the chance of undiscovered errors should decrease with time.

P.20, Lemma I.3.3 [=]:

Idea of the Butterfly Lemma

If G is a group, and U/u, V/v are homomorphic images of subgroups of G, one would like to describe the extent to which one can ''relate'' part of the structure of U/u and part of the structure of V/v, based on their common origin in G. If one takes the subgroup of U/u consisting of all elements which are also images of elements of V, one can write this $u(U \cap V)/u$. The analogous subgroup of V/v is $(U \cap V)v/v$. To get a common homomorphic image of these, we must divide each by the subgroup of those elements that are annihilated in the construction of the other. These subgroups are the denominators in the expressions at the bottom of p.20. The Butterfly Lemma says that after making these adjustments, we do get isomorphic groups, the ''common heritage'' of U/u and V/v.

Group-theorists call a *factor-group* of a *subgroup* of a group G a *subfactor* of G. Thus, given two subfactors U/u and V/v of a group G, the Butterfly Lemma characterizes their largest natural ''common subfactor''.

There are a couple of points in the proof of that lemma that Lang skips over. First, why are the four sets $u(U \cap V)$ etc. subgroups? Note that u is normal in U, so $U \cap V$ is contained in the normalizer of u. From this we see that $u(U \cap V)$ is a subgroup; the other cases are seen in the same way. Second, where Lang says of the diagram on p.21 that ''The intersection of two line segments going downward represents the intersection of the groups. Two lines going upwards meet in a point which represents the product of two subgroups'', this gives four descriptions of the one unlabeled point of the diagram; so these four groups must be shown equal! This is not hard once the task is recognized; cf. Exercise I.3:3.

Re §I.4. Cyclic groups.

P.23, start of §I.4 [<]: In §I.2, a ''generator'' of a group or monoid G meant a member of whatever generating set was under discussion – a relative rather than an absolute concept. But in this section, an element x is called a ''generator'' of G if the singleton $\{x\}$ is a generating set, i.e., if x is what is called a ''cyclic generator'' on p.8. Both usages are common in group-theoretic writing, so one needs to watch out which sense is meant when the term is used.

What Lang calls the *period* of an element of a group (p.23) is usually called its *order*. This concept is most useful when one wants to know ''all about'' some element of a particular group; the concept of an *exponent* of an element is more useful in general arguments, where one has limited information to work with.

P.25, end of proof of (**v**) and whole proof of (**vi**) [=]: Lang relies on results not yet proved. Here are self-contained arguments (though in the proof of (**vi**), I will leave two steps for you to justify):

In (**v**), having proved that the kernel of the map $\mathbf{Z} \to A\times B$ is $mn\mathbf{Z}$, we conclude that the image has order mn, which is the order of the whole group $A\times B$, so the map is surjective.

In proving (**vi**), let us use the group-theorists' notation ''$\langle \ldots \rangle$'' for ''subgroup generated by''; and let us also write our group operation additively, so that we can write things like nG for $\{nx \mid x\in G\}$. Let H be a subgroup of G *minimal* for the property of being noncyclic (i.e., such that $H \leq G$ is noncyclic, but all proper subgroups of H are cyclic. Such an H must exist because G is noncyclic and has only finitely many subgroups.) Now since multiplication by the integer $(H:1)$ is the trivial endomorphism of H, and this map is a composite of multiplications by primes, there must be some prime p such that multiplication by p is not an automorphism of H. Hence pH is properly smaller than H, hence by assumption it is cyclic. Let us write a generator of this subgroup as px, where $x\in H$. Since H is not cyclic, it properly contains $\langle x\rangle$, so we can choose a $y\in H$ not in $\langle x\rangle$. Since $py\in pH = \langle px\rangle$, we can write $py = mpx$ for some integer m. Let $z = y - mx$. Then $z\notin\langle x\rangle$ and $pz = 0$. There are now two cases:

Case 1: x has order divisible by p. Then $\langle x\rangle$ has an element w of order p (by (iv)), and you should be able to show that the subgroup of H generated by w and z is isomorphic to the direct product of two cyclic groups of order p, as desired.

Case 2: x has order not divisible by p. Then it is easy to verify that the element $x' = x + z$ has order divisible by p, and it clearly also has the property we assumed earlier for x, namely that $px' = px$ generates pH. Hence we can repeat the above argument with x' in place of x, and this time we end up in Case 1, and so get the desired conclusion.

Re §I.5. Operations of a group on a set.

P.26, end of top paragraph [=]:

General observation on the source of group and monoid theory.

The concept of group (respectively monoid) arose from consideration of the properties of the automorphisms (respectively endomorphisms) of a mathematical object. The concept of an *abstract group* was a great advance, allowing mathematicians to consider groups without reference to specific permutation representations. But group theory does not cut itself off from its roots – the concept of an action of a group by permutations, introduced in this section, allows us to work with abstract groups, and yet represent them, when convenient, by permutations of sets; and often provides good intuition in studying abstract group theory.

Of course, to know whether permutation representations of groups are really general enough to embrace all of abstract group theory, we may ask whether *every* abstract group G admits a representation as a group of permutations of some set X. Many of you have seen the answer in your undergraduate courses, but let us analyze the question pretending that we haven't. Such a representation of G would correspond to giving a G-set X, such that distinct elements of G have distinct actions on X. (A G-set X with this property is said to be *faithful*.) We observe that if X is any G-set, and x any element of X, then X must contain elements gx for all $g\in G$, and we can describe how G acts on the set of such elements – namely,

$$g'(gx) = (g'g)x.$$

Hence, let us try to construct a G-set by choosing a symbol x, letting X consist of all symbols gx as g ranges over G, and defining the action of G on X as above. It is easy to verify that this does make X a G-set, and that distinct elements of G act by distinct permutations. This gives the desired result, *Cayley's Theorem*, that every group is isomorphic to a group of permutations of a set. We may make a symbolic simplification: since the same symbol x appears at the end of all the elements of X, it provides no information, so we may instead take for our X the set of elements of G itself. This gives the standard proof of Cayley's Theorem, by the observation that the *translation representation* of G (which Lang will introduce at the bottom of this page) is faithful.

P.26, section on conjugation [>]: We can get perspective on the concept of *conjugation* in groups by beginning outside the group-theoretic context, and looking at a bijection

$$\alpha: X \to Y$$

from one set to another. This can clearly be used to ''carry'' any sort of structure on X over to Y. For instance, if we have an I-tuple of elements of X, $(x_i)_{i\in I}$, we get an I-tuple of elements of Y, $(\alpha(x_i))_{i\in I}$. If we have a real-valued function $h: X \to \mathbf{R}$, we find that the corresponding real-valued function on Y is $h\alpha^{-1}$. (How do we see this? Given any $y \in Y$, to find the value at y of the function on Y ''corresponding to h'', we look for the element of x that corresponds to y, namely $\alpha^{-1}(y)$, evaluate h there, and conclude that the result, $h(\alpha^{-1}(y))$, is the value we want.) Now given a map $f: X \to X$, we can likewise get a ''corresponding'' map $Y \to Y$; you should verify for yourself that this is $\alpha f \alpha^{-1}$. Note that this new map will be bijective if and only if f is. If we now turn to the case where X and Y are the same set X, and α and f are two permutations of X, we see that $\alpha f \alpha^{-1}$ still represents the operation on X which ''corresponds to f, under the transformation α of X into itself''.

Returning to the case of two sets X and Y for clarity, we can see intuitively that this operation of ''carrying a permutation of X to the corresponding permutation of Y'' must respect *composition of permutations*; hence in particular, this is true in the same-set case. Since every group can in fact be represented by permutations of a set, every group will have the property that conjugation by a fixed element respects composition, hence is an automorphism of the permutation group.

Finally, if in our ''different-sets'' situation, we have in addition to the bijection α between X and Y, a bijection β between Y and a set Z, we can see that carrying an endomap f of X to the endomap of Y to which it corresponds under α, and carrying this in turn to the endomap of Z to which it corresponds under β, must give the same result as carrying f directly to the endomap of Z to which it corresponds under $\beta\alpha$. Thus, a composite of conjugation maps gives conjugation by the composite map. In particular, for a group G, the construction sending each element α to the operation of conjugation by α is a *group homomorphism* $G \to \mathrm{Aut}(G)$.

All the results obtained above can be verified *computationally* in less space than the above discussion! The point of this exposition is to give us intuition on the subject. In particular, it often allows us to ''see'' without computation the *form* that the conjugate of one element by another in a group of permutations of a structure will take. For instance, in the symmetric group S_5 (which Lang will define on p.30, but most of you saw in your undergraduate courses), if we take the permutation acting by $1 \mapsto 2 \mapsto 3 \mapsto 1$ – in the notation Lang introduces on p.30, [123] – and conjugate it by the permutation [25], we must get [153].

If you haven't seen the symmetric group, you should read the short paragraph defining it on p.30 now, then return to p.26, since Lang will assume familiarity with it at the bottom of p.28.

P.26, bottom [=]: The maps T_x that Lang defines are more precisely called *left translations*. One also speaks of the *right translation* maps R_x given by $R_x(y) = yx$.

The map $x \mapsto R_x$ is not a homomorphism $G \to \mathrm{Perm}(G)$; rather, it satisfies $R_x R_y = R_{yx}$. Thus, these maps do not define an action of G on itself; nevertheless it is of some importance, so it is best to always speak of ''left translation'' and ''right translation'', rather than just ''translation''.

(In fact, $x \mapsto R_x^{-1}$ does define an action of G on itself. Moreover, every map T_x commutes with every map R_y: $T_x R_y = R_y T_x$. One can deduce from this and the fact that $x \mapsto R_x^{-1}$ gives an action of G on itself that the map $x \mapsto T_x R_x^{-1}$ also gives such an action. Namely?)

P.27, paragraph just below definition of **isotropy** group [<]: If you are reading this in my 250A or another course that has you read this section before §I.3, then simply understand that the ''normalizer'' mentioned in this paragraph is something you will learn about very soon. (Or better – work out for yourself what the isotropy subgroup of an element under the conjugation action is, then look up ''normalizer'' in the index and see whether it agrees with what you have gotten.)

P.28, first sentence of last paragraph [<]: S_n denotes the group which, in the notation of the second example of p.8, would be written $\mathrm{Perm}(\{1, \dots, n\})$. (Lang will introduce the notation S_n on p.30.)

The remaining sentences of this paragraph refer to subjects you should not expect to be familiar with.

P.28, bottom [>]:

Notes on group actions.

I have pointed out that the essential motivation for the concept of *group* is that it embodies the properties of the collection of all permutations of a set, or more generally, of the collection of all automorphisms of a mathematical object. Thus, to get intuition on some statement about groups, it is useful to think of what the statement means in the case of a group of permutations. What about the concept of a subgroup of a group G?

The most useful general source for that concept, I claim, is the construction of the *isotropy subgroup* G_x of an element x of a G-set X (also often called the *stabilizer* of x). Indeed, most of our examples of subgroups really arise in this way. Consider, for example, the symmetry group of the square, D_4, which is the subgroup of the permutation group S_4 of the vertices of a square comprising those permutations that preserve all distances between vertices. If we let X denote the set of all *metrics* on the 4-element set consisting of the vertices, $\{1, 2, 3, 4\}$, then S_4 acts on X: a permutation σ takes each metric $d(\ ,\ )$ to the metric $d(\sigma^{-1}(\ ), \sigma^{-1}(\ ))$. The definition of D_4 can now be seen to say that, if we let d denote the metric on $\{1, 2, 3, 4\}$ such that $d(i, j)$ is the distance between the ith and jth vertices of the unit square, then D_4 is the *isotropy subgroup* $(S_4)_d$ under this action of S_4 on metrics. Other descriptions of subgroups of various groups, as comprising those permutations that ''preserve'' certain structure on a certain object, translate similarly to characterizations of them as isotropy subgroups. (In considering particular cases, there is no need to always translate such definitions of subgroups into the language of isotropy subgroups. But having seen that they can be so translated gives us confidence in using the isotropy-subgroup concept as a general way to look at subgroups.)

Given a group G and a subgroup H, does there always exist a G-set X and an element $x \in X$ with $G_x = H$? Yes: as shown in exercise I.5:1, G/H is a G-set with such an element. That result could have been motivated in the same way as the proof of Cayley's Theorem above.

Some examples of the translation between group-theoretic concepts and G-set concepts: If A and B are isotropy subgroups of elements of two G-sets, $x \in X$ and $y \in Y$ respectively, then their *intersection*, $A \cap B$, is the isotropy subgroup of the element (x, y) of the G-set $X \times Y$. If H is a subgroup of a group G, say the isotropy subgroup of an element x of a G-set X, then the *conjugate* gHg^{-1} of H by an element g can be interpreted as the isotropy subgroup of the translate gx of x. (This fits in with the viewpoint that conjugation of g by α means taking ''the permutation corresponding to g under a shift of location given by α''.)

It follows that a subgroup H of a group of permutations G of a set X will be *normal* in G if H can be described in terms of some *G-invariant* structure on X.

We can use this observation to get some insight to the fact noted in our discussion of normal towers, that a normal subgroup of a normal subgroup need not be normal. E.g., in the group G of symmetries of the square, the subgroup N consisting of the four elements preserving both the horizontal and the vertical axes is normal, and within N the subgroup M generated by reflection through the vertical axis is normal, but M is not normal in G. These facts, easily verified computationally, can be interpreted conceptually by noting that N can be characterized in terms of structure on the square (namely, the set consisting of the horizontal and vertical axes) which is G-invariant, and M is normal in N because it can be characterized in terms of a structure (namely, the vertical axis) which is N-invariant; but, since the latter structure is not G-invariant, M may not be (and indeed, is not) normal in G.

If G is any group, then a group which, like the M of the above example, appears in a normal tower for G is called a *subnormal subgroup* of G. If, in this example, we replace the square by a regular 2^n-gon, one can show that the subgroup generated by reflection about the vertical axis occurs in a normal tower of height n, but not in any normal tower of smaller height.

P.30, last display [=]: Can you see that the first two equalities are correct? The difficulty one typically has in applying the definition Lang has given is a classic example of ''alias vs. alibi'' confusion (cf. comment in this Companion re p.13 of Lang).

Incidentally, on the next page Lang will take for f a polynomial function. If we regard polynomial functions as induced by members of the ring of polynomials over $\mathbf{Z}$ in n indeterminates, $X_1, \ldots, X_n$ (to be introduced formally in the next chapter), then the corresponding action of S_n on this polynomial ring is by permutation of the indeterminates, $\pi(\sigma)(X_i) = X_{\sigma(i)}$; and in this case we have $\pi(\sigma)\pi(\tau)\, f(X_1, \ldots, X_n) = \pi(\sigma)\, f(X_{\tau(1)}, \ldots, X_{\tau(n)})$, which nevertheless leads to $\pi(\sigma)\pi(\tau) = \pi(\sigma\tau)$. I leave it to you to ponder this paradox.

P.32, line after first display [=]: For $H_\nu / H_{\nu-1}$ read $H_{\nu-1} / H_\nu$.

P.32, Proof of Theorem I.5.5 [<]: I recommend not trying to follow the details, but working out for yourself that if N is any normal subgroup of S_n $(n \geq 5)$ and g a nonidentity element of N, then by playing with conjugates of g, you can come up with an element that moves few points, from that get a 3-cycle, and from this get the general element of A_n.

(Incidentally, in the first paragraph of Lang's proof, every printing through the current one has some confusion regarding the conditions on σ. In the current printing, there is merely the unnecessary phrase ''or the identity'' four lines from the bottom, though σ has been assumed a nonidentity element; the 3rd printing likewise had that phrase plus a following unnecessary sentence. The 4th and 5th printings, more seriously, left out the assumption on an earlier line that $\sigma \neq 1$, though without this, the maximality condition makes $\sigma = 1$, and one can deduce nothing about 3-cycles.)

P.33, end of §I.5 [>]: The material below, not in Lang, is required for my 250A class. It will be assumed in the analysis of groups of order pq a few pages from now.

Semidirect products.

When a turtle *retracts* into its shell, the shell stays fixed, and the turtle's legs, head and tail move into it. Similarly, a *retraction* of a group G onto a subgroup H means an endomorphism $r: G \to G$ such that r fixes all elements of H, and carries all elements of G into H. If $r: G \to G$ is such a retraction, and N its kernel, it is easy to verify that every element $g \in G$ can be written uniquely as a product of a member of H (namely, $r(g)$) and a member of N (namely, $r(g)^{-1} g$). To describe how such products are multiplied, we need to know the action of H on N by conjugation: If for fixed $h \in H$ we write $\psi(h)$ for the automorphism $x \mapsto h^{-1} x h$ of N, and write the action of such automorphisms exponentially, we see that

(c2) $$(h_1 x_1)(h_2 x_2) \;=\; (h_1 h_2)\, (x_1^{\psi(h_2)} x_2).$$

Conversely, given any groups H and N, and any action of H on N via a homomorphism $\psi: H \to \mathrm{Aut}(N)$, we can define a group whose elements are formal products hx $(h \in H,\ x \in N)$, with multiplication defined by (c2). To make the concept of ''formal product'' more precise, let us use the ordered pair (h, x) instead of the symbol hx. Thus, the law of multiplication becomes

(c3) $$(h_1, x_1)(h_2, x_2) \;=\; (h_1 h_2,\; x_1^{\psi(h_2)} x_2).$$

It is a straightforward exercise to verify that this is indeed a group operation. (Associativity is a quick verification, and the neutral element is, of course, (e_H, e_N). You should find it easy to solve the equation $(h, x)(h', x') = e$ for h' and x', and so get a formula for the inverse of (h, x).) Note that the resulting group has the same underlying set as the direct product group $H \times N$, but has a more general multiplication operation. (''More general'' because, if we take for ψ the trivial homomorphism $\psi\colon H \to \mathrm{Aut}(N)$ which takes all elements of H to the identity automorphism of N, we get precisely the direct product.) It is called ''the *semidirect* product of H and N determined by the map ψ'', and denoted $H \ltimes_\psi N$, or, when there is no danger of ambiguity, $H \ltimes N$.

I should backtrack here, and clarify a point of notation. I said in the very first item of this Companion that functions would in general be written to the *left* of their arguments, and composed accordingly. However, I have made an exception here, and written the action of $\psi(h)$ $(h \in H)$ on elements of N as an exponent, $(\;)^{\psi(h)}$. The reason for this goes back to our decision to write elements of G as products hx with $h \in H$ on the left, and $x \in N$ on the right. The result was that in expressing the product of two such elements as such an element, a member of H had to be ''pulled to the left'' past a member of N, using the formula $xh = h(h^{-1}xh)$; so to each $h \in H$ we associated the automorphism $\psi(h)\colon x \mapsto h^{-1}xh$. Now it is easy to check that if we apply to an element x this automorphism $\psi(h)$, and then apply to the result the automorphism $\psi(h')$ for another element $h' \in H$, this is equivalent to applying to the original element the automorphism $\psi(hh')$. Hence if we used left functional notation, we would have the unpleasant law $\psi(h')\psi(h)(x) = \psi(hh')(x)$; but when we write these automorphisms on the right (and use exponential notation to avoid confusion with multiplication), this takes the nicer form $(x^{\psi(h)})^{\psi(h')} = x^{\psi(hh')}$. So when I called ψ a homomorphism from H to $\mathrm{Aut}(N)$, I should have specified that by $\mathrm{Aut}(N)$ I meant the automorphism group under the composition operation one gets when one writes automorphisms to the *right* of their arguments; the reverse of the ordinary composition. I will note at the end of this discussion some variant forms of the semidirect product construction that one gets by making different notational choices.

We have motivated the general semidirect product construction by considering a group G with a retraction to a subgroup H. Conversely, if we form a semidirect product group $H \ltimes_\psi N$ from any two groups H and N and a homomorphism ψ as described above, we find that the map $(h, x) \mapsto (h, e)$ is a retraction of this group onto a subgroup isomorphic to H, and having kernel isomorphic to N. Hence an isomorphism of a group G with such a semidirect product is *equivalent* to a retraction of G onto a subgroup. As with direct products, we can, when the distinction needs to be made, call a group G constructed from two given groups H and N and a map ψ an ''external'' semidirect product, while when we are given a group G and discover that it has a retraction to a subgroup H, and hence is isomorphic to a group constructed in that way, we can call G an ''internal'' semidirect product.

What are some examples of groups expressible as semidirect products? As noted above, direct products are trivial examples of this construction. A nontrivial example is S_3: the endomorphism which sends every even permutation to the identity, and every odd permutation to (12), is easily seen to be a retraction to the subgroup $\langle(12)\rangle$. That subgroup is isomorphic to Z_2, the kernel of the retraction, A_3, is isomorphic to Z_3, and the action of $\langle(12)\rangle$ on A_3 by conjugation corresponds to the homomorphism $\psi\colon Z_2 \to \mathrm{Aut}(Z_3)$ that sends $[0]$ to the identity automorphism, and $[1]$ to the automorphism $[n] \mapsto [-n]$; hence $S_3 \cong Z_2 \ltimes_\psi Z_3$ for ψ so defined. Recall that S_3 is the symmetry group of an equilateral triangle. More generally, the symmetry group of the regular n-gon can be written as a semidirect product $Z_2 \ltimes_\psi Z_n$, where ψ is defined by the same formula as above. (This group is called the *dihedral group* of order $2n$. Likewise, $Z_2 \ltimes_\psi \mathbf{Z}$, with ψ defined in the analogous way, is called the *infinite dihedral group*.) The group Lang discusses in the top paragraph of p.15 can be seen to be a semidirect product of the additive group of real numbers and the multiplicative group of nonzero real numbers. Some examples of nonsimple groups that are *not* semidirect products of proper subgroups are the groups Z_{p^n} (p a prime, $n>1$), and the eight-element ''quaternion group'' that Lang introduces at the bottom of p.9.

Since semidirect product decompositions are extremely useful in describing groups, it is worth seeing various formulations of the condition for a group G to be a semidirect product. We have seen that if G can be *retracted* to a subgroup H, then G is a semidirect product of H and the kernel of the retraction. You should verify that an endomorphism r of a group G is a retraction from G to its image $r(G)$ if and only if $r^2 = r$. An entity satisfying the equation $r^2 = r$ is called *idempotent* (Latin for ''equal to its power(s)''); so semidirect product decompositions of G correspond to idempotent endomorphisms of G.

Next, suppose G and H are any two groups, and $f: G \to H$ a surjective homomorphism. It may or may not be possible to find within G a subgroup H_0 that is mapped isomorphically to H by the restriction of f. (For instance, if $G = S_3$, $H = Z_2$ and $f: G \to H$ is the homomorphism taking even permutations to $[0]$ and odd permutations to $[1]$, then $H_0 = \langle(12)\rangle$ is such a subgroup, while if $G = Z_{p^n}$ (p prime, $n>1$), $H = Z_p$, and f is the map sending each congruence class $[i]_{p^n}$ to its image $[i]_p$, there is no such subgroup.) If we have such an H_0, let us denote by g the homomorphism $g: H \to G$ that takes each $h \in H$ to its unique inverse image in H_0 under f. Then we see that

(c4) $$fg = \mathrm{id}_H.$$

What can we say about the other composite, gf? This turns out to be a *retraction* of G onto H_0! In other words, whenever a group homomorphism $f: G \to H$ has a right inverse g, the composite gf is a retraction of G to $g(H) \cong H$; so, in this situation, G is isomorphic to a semidirect product of H with the kernel of this retraction, which is also the kernel of f. (Terminological note: A homomorphism f having a right inverse g as in (c4) is called a *split surjective* homomorphism. Similarly, a homomorphism g having a left inverse f is called *split injective*.)

Finally, given a group G and two subgroups, H and N, we may ask when these will be the image and kernel of a retraction, and thus the ingredients in a semidirect product decomposition of G. You should not find it hard to check that the necessary and sufficient conditions are:

(c5) $$N \text{ is normal in } G, \qquad HN = G, \qquad \text{and} \qquad H \cap N = \{e\},$$

(cf. Lang, Exercise I:L12, (a) and (b)), and that H and N then uniquely determine the retraction.

I noted earlier that the form that our construction of the semidirect product took – in particular, the expression of the action of H on N as a homomorphism into the group of automorphisms of N composed in the way that goes with writing automorphisms on the *right* of their arguments – was a result of certain choices we had made. Different choices lead to different notations, so I should point out some other forms of this construction that you may encounter. To make the distinctions explicit, let us temporarily write $\mathrm{Aut}(N)_r$ for the group of automorphisms of N with multiplication defined in accordance with action of automorphisms on the right, and $\mathrm{Aut}(N)_l$ for the same set but with the multiplication that corresponds to writing automorphisms on the left, the opposite of the former multiplication. Thus, given a group G, a subgroup H, and a normal subgroup N, the map taking each $h \in H$ to the automorphism $x \mapsto h^{-1}xh$ of N is a homomorphism $H \to \mathrm{Aut}(N)_r$, while the map taking h to $x \mapsto hxh^{-1}$ is a homomorphism $H \to \mathrm{Aut}(N)_l$. Let us call these two homomorphisms ψ and θ respectively.

Thus, if we had insisted on writing automorphisms on the left of their arguments, we could have described the law of composition for products hx ($h \in H$, $x \in N$) in a group G with a retraction to a subgroup H as $(h_1x_1)(h_2x_2) = (h_1h_2)(\theta(h_2)^{-1}(x_1)\,x_2)$, and this would have led to an external semidirect product construction in which the law of multiplication (c3) took the form

(c6) $$(h_1, x_1)(h_2, x_2) = (h_1h_2,\ \theta(h_2)^{-1}(x_1)\,x_2)$$

for θ an arbitrary homomorphism $H \to \mathrm{Aut}(N)_l$. Another option would have been to write the typical element of our group G with a retraction r not as hx but as xh $(x\in N,\ h\in H)$. We would then have found that the multiplication of such products takes the form $(x_1h_1)(x_2h_2) = (x_1\theta(h_1)(x_2))(h_1h_2)$, leading to a construction of the general semidirect product as having underlying set $N\times H$, and multiplication

(c7) $$(x_1, h_1)(x_2, h_2) = (x_1\theta(h_1)(x_2),\ h_1h_2).$$

This version of the semidirect product is denoted $N\rtimes_\theta H$. (It is introduced by Lang in Exercise I:L12, but unfortunately, in all versions through the first Springer printing, the formula he gives for the multiplication is garbled.)

The above three constructions are all essentially equivalent. Indeed, if H and N are groups, ψ a homomorphism $H\to \mathrm{Aut}(N)_r$, and θ the homomorphism $H\to\mathrm{Aut}(N)_l$ given by $\theta(h) = \psi(h)^{-1}$, then (c3) and (c6) say exactly the same thing, and the group they define is *isomorphic* to the group $N\rtimes_\theta H$ defined by (c7). (To find the isomorphism between these groups, think about how to turn a product hx into a product $x'h'$ in the situation with which we began this discussion.) As a mnemonic, to remember which side which group goes on when one uses the symbols $\ltimes$ and $\rtimes$, you can think of these as containing normal subgroup symbols $\triangleright$ and $\triangleleft$, with the small end pointing to the subgroup that is to be normal; or, as suggested by a graduate student here, you can think of the open end of either symbol as an open jaw, and remember that ''the beastie eats the kernel''. I should add that these symbols, though common, are not universal. Lang introduces no symbol for the construction in Exercise I:L12; Hungerford uses $N\times_\psi H$; and another writer may just write HN, saying in words that the product is semidirect and has multiplication induced by ψ.

Throughout these notes, semidirect products will be written $H\ltimes_\psi N$ or $H\ltimes N$, and have multiplication defined by (c3).

Re §I.6. Sylow subgroups.

P.33, Proof of Lemma I.6.1 [<]: Until now, Lang has consistently written e for the unit element (also called the identity element) of a group that is not written additively. From this point on, he will vacillate between e and 1. Both notations are common in group theory.

P.33, third line from bottom, '' ... is divisible by p'' [>]: The point is that if all elements had orders (''periods'') not divisible by p, then the least common multiple of these orders would be an exponent n of G not divisible by p, hence the order of G would divide a power of n, contradicting the assumption that it is divisible by p.

P.34, middle (end of proof of Theorem I.6.2) [>]:

Alternative proof of the existence Sylow subgroups.

Note that if G is a group, then the set of *subsets* of G can be made a G-set by letting G act by left translation. It is easy to verify:

(c8) If G is a finite group and x any subset of G, then the isotropy subgroup G_x of x under left translation is the largest subgroup H of G such that x is a union of right cosets of H. In particular, the cardinality of x is a multiple of the order of G_x.

We can now give an elegant proof that every finite group G has a p-Sylow subgroup. Let $(G{:}1) = n = p^r s$, where p does not divide s. Let us write $G^{\{p^r\}}$ for the set all subsets of G of cardinality p^r. Then G acts by left translation on $G^{\{p^r\}}$. We shall soon prove that the cardinality of this set, $\binom{p^r s}{p^r}$, is $\equiv s \pmod p$, hence not divisible by p. Assuming this for the moment, we see that some orbit Gx $(x\in G^{\{p^r\}})$ must have cardinality not divisible by p, hence G_x has order a multiple of

p^r. But by (c8), the order of G_x is a divisor of $\#(x) = p^r$, hence it is exactly p^r; that is, G_x is a p-Sylow subgroup of G.

But how do we prove the number-theoretic result $\binom{p^r s}{p^r} \equiv s \pmod p$ used above? Surprisingly, we can get this by turning our original proof backwards. We start with some group G of cardinality $p^r s$ that we *know* has a p-Sylow subgroup P (for instance $Z_{p^r} \times Z_s$). If we let P act on $G^{\{p^r\}}$ by left translation, we see that the only orbits of cardinality 1 are the right cosets of P, of which there are exactly s. All other orbits have cardinalities divisible by p; so the cardinality of $G^{\{p^r\}}$, namely $\binom{p^r s}{p^r}$, must be congruent to s mod p.

The additional results on Sylow subgroups that Lang proves in Theorem I.6.4 can also be gotten by looking at the G-set $G^{\{p^r\}}$, but the resulting arguments are fairly close to those Lang gives, so I will not write them out here.

P.36, end of §I.6 [>]: The Sylow theorems are a powerful tool in the study of finite groups. Let us use them to determine the

Structures of groups of order pq.

We know that the groups of prime order are just the abelian groups Z_p (Lang, Proposition I.4.1, last sentence). Exercise I.6:1(a) = I:L24 similarly shows that groups of prime-squared order are just the abelian groups $Z_p \times Z_p$ and Z_{p^2}. We shall show below that groups whose factorizations have the next more complicated form, pq, where p and q are distinct primes, can be described as semidirect products of Z_p and Z_q. We will need the following observation (essentially, Lang's exercise I:L4):

(c9) If G is a finite group with subgroups H_1 and H_2, the set $H_1 H_2$ has cardinality $(H_1 : 1)(H_2 : 1)/(H_1 \cap H_2 : 1)$.

To see this, let us ask what elements fall together under the map $H_1 \times H_2 \to G$ given by $(h_1, h_2) \mapsto h_1 h_2$. If $h_1' h_2' = h_1 h_2$, then we have $h_1^{-1} h_1' = h_2 h_2'^{-1}$, and we see that this element (let us call it k) must belong to $H_1 \cap H_2$. One easily deduces that the elements of $H_1 \times H_2$ that fall together with (h_1, h_2) under the multiplication map are precisely those of the form $(h_1 k, k^{-1} h_2)$ with $k \in H_1 \cap H_2$; so the multiplication map is $(H_1 \cap H_2 : 1)$-to-one, proving (c9). We can now prove

Lemma I.6.c1. *Let G be a finite group, H a subgroup, and N a normal subgroup of G. Then*

(a) *G is a semidirect product $H \ltimes N$ if and only if $(G{:}1) = (H{:}1)(N{:}1)$ and $H \cap N = \{e\}$.*

(b) *In particular, the above conclusion holds if $(G{:}1) = (H{:}1)(N{:}1)$ with $(H{:}1)$ and $(N{:}1)$ relatively prime.*

Proof. Since N is assumed normal, the conditions given in (a) are the same as those given in (c5) for G to be a semidirect product, except that the equation $HN = G$ has been replaced by $(G{:}1) = (H{:}1)(N{:}1)$. But from (c9), we see that when $H \cap N = \{e\}$, these two conditions are equivalent.

Assertion (b) is the special case of (a) where the condition $H \cap N = \{e\}$ follows from the fact that the order of this intersection must be a common divisor of the orders of H and N. □

In what situations can we prove that a group G will have subgroups H and N satisfying the above conditions? If the order of G is a product of powers of two distinct primes, p and q, then the corresponding Sylow subgroups of G will have relatively prime orders whose product is the order of G, so we only need to know that one of these subgroups is normal. A convenient tool for showing this is Lemma I.6.7, p.36. Here is an alternative proof of that lemma (part (c) below), with some additional results given by the same method (parts (a) and (b)).

Lemma I.6.c2. *Let G be a group and K a subgroup of finite index in G. Then*

(a) *G has a* normal *subgroup N of finite index which is contained in K, and such that $(G:N)$ divides $(G:K)!$.*

Assuming G itself finite, we have the consequences

(b) *If every prime dividing the order of K is $\geq (G:K)$, then K is normal in G.*

(c) *Hence if $(G:K)$ equals the least prime dividing the order of G, then K is normal.*

Proof. Let $n = (G:K)$, and consider the action of G on the n-element G-set G/K. This is a homomorphism *into* the permutation group on that G-set, which has order $n!$, hence *onto* a subgroup S of that permutation group. The order of S will be a divisor of $n!$, and the kernel of that homomorphism will be a normal subgroup N whose index in G is that order, giving (a).

Note that $(K:N) = (G:N)/(G:K) = (G:N)/n$, a divisor of $(n-1)!$, hence a product of primes $\leq n-1$. But in the situation of statement (b), every prime divisor of $(K:N)$ must be $> n-1$, so there are no such divisors, and $K = N$. Statement (c) is a particular case of (b). □

We can now characterize all groups whose orders are products of two distinct prime factors.

Lemma I.6.c3. *Let $p < q$ be primes. Then every group of order pq is a semidirect product of a cyclic subgroup of order p and a normal cyclic subgroup of order q. If $q \not\equiv 1 \pmod{p}$, the only such group is the* direct *product of these subgroups, an abelian group isomorphic to $\mathbf{Z}/pq\mathbf{Z}$. If $q \equiv 1 \pmod{p}$, there also exists a nonabelian group of order pq.*

Proof. Let G be a group of order pq, and let H be a p-Sylow subgroup and N a q-Sylow subgroup of G. By Lemma I.6.c2(c) above, N is normal, hence by Lemma I.6.c1(b), G is a semidirect product of the indicated form.

Hence let us write $G \cong H \ltimes_{\psi} N$ where ψ is a homomorphism $H \to \mathrm{Aut}(N)$. This group will be nonabelian if and only if ψ is nontrivial, and such a nontrivial homomorphism can be found if and only if $\mathrm{Aut}(N)$ has order divisible by p (Lang, Theorem I.6.2 and Corollary I.6.6). By Proposition I.4.3(iii), the automorphism group of the cyclic group of order q has the same number of elements as the group has generators; since q is prime, all nonidentity elements are generators, so $\mathrm{Aut}(N)$ has order $q-1$. This will be divisible by p if and only if $q \equiv 1 \pmod{p}$, establishing the asserted condition for the existence of a nonabelian group.

On the other hand, if G is abelian, and hence is the direct product of H and N, we know from Proposition I.4.3(v) that G is cyclic. □

We now ask: If $q \equiv 1 \pmod{p}$, will there be just one, or more than one isomorphism class of nonabelian groups of order pq? To approach this question, suppose G is a nonabelian group of order pq, with a normal subgroup N cyclic of order q, and a p-Sylow subgroup H cyclic of order p. Let a be a generator of N. Any automorphism φ of N carries a to another generator of N, which will have the form a^r for some r relatively prime to q; given φ and a, this r is clearly uniquely determined mod q. Note that φ will carry any other generator $a' = a^i$ of N to $(a^r)^i = a^{ir} = a'^r$; hence the residue of r modulo q is in fact an invariant of the automorphism φ, independent of our choice of generator a of N. We see that composing two automorphisms corresponds to multiplying the corresponding residues modulo q. Now if we map each element hn of $G = HN$ to the automorphism of N that it induces, we can see that this will depend only on h (since N is abelian), and will determine by the preceding results a homomorphism of G into the group of nonzero residues modulo q. The image of this homomorphism will be a subgroup of this group of residues; since it equals the image of H under the same map, it will have order $(H:1) = p$.

This subgroup of the multiplicative group of integers modulo q will be an invariant of the structure of the group G; i.e., if G and G' are each a semidirect product of a cyclic subgroup of order p with a cyclic subgroup of order q, and they are isomorphic, they must determine the same order-p group of residues modulo q.

The converse is also true: if G and G' are semidirect products of a cyclic group of order p and a cyclic group of order q, and they determine the same group of residues, then they will be isomorphic. Indeed, let a be a generator of a normal subgroup $N \lhd G$ of order q, and b a generator of a subgroup $H < G$ of order p, and say conjugation by b carries a to a^r. We now want to find elements a', b' in our other group G' that will correspond to a and b under an isomorphism. Let a' be any generator of a normal subgroup $N' \lhd G'$ of order q, but within a subgroup $H' < G'$ of order p, let us be more careful: Because of our assumption about G and G' determining the same groups of residues, we know that there will be some element of H' conjugation by which carries a' to a'^r; let b' be such an element. Then it is not hard to verify that the map $G \to G'$ carrying each element $b^i a^j$ to $b'^i a'^j$ is an isomorphism.

The problem of whether there can be *more than one* isomorphism class of nonabelian groups of order pq therefore reduces to the question: Does the multiplicative group of nonzero residues modulo q contain more than one subgroup of order p?

Since this group of residues is abelian, it has a unique p-Sylow subgroup, which will contain all subgroups of order p. So the problem is what this p-Sylow subgroup looks like! If it has order p (i.e., if $q-1$, though divisible by p, is not divisible by p^2), then clearly it will be the unique subgroup of order p, and thus by the above reasoning there will be a unique isomorphism class of nonabelian groups of order pq. For instance, this happens if $pq = 6$, 14, 21 or 39. (E.g., it happens in the last case because $39 = 3\cdot 13$, and 3 divides $13-1$ just once.) What about cases where $q-1$ *is* divisible by p^2, e.g., $pq = 10$, 26 or 57? Checking by hand, we find that in each of the cases just listed, the p-Sylow subgroup of $\mathrm{Aut}(\mathbf{Z}/q\mathbf{Z})$ has only one subgroup of order p. In our notes on §IV.1, we shall see that simple ring-theoretic considerations prove that this is always so, and hence that under the conditions of the last sentence of Lemma I.6.c3, there is, up to isomorphism, a *unique* nonabelian group of order pq. In the mean time, we will live with the suspense of wondering how that will be proved!

The above results, together with Exercises I.6:1-2, give us the structures of groups of most ''small'' orders. The first case not covered, order 12, is not hard to analyze; I like to work it out in class. The next value, 16, is the fourth power of a prime, a difficult case, which is more work than we have time for, but the two after that, 18 and 20, can be analyzed with the help of some further easy results, which I give below, but do not make required reading for Math 250A.

Further notes on structures of finite groups. (Optional reading for my 250A.)

Lemma I.6.c4. *Let p and q be primes and i, j positive integers, and let G be a group of order $p^i q^j$. Then G has a normal q-Sylow subgroup if* either (a) $i = 1$ *and* $p < q$, or (b) $i = 2$ *and* $p < q-1$.

Thus, if either of these conditions holds, G is a semidirect product of this normal q-Sylow subgroup and any p-Sylow subgroup of G.

Proof. Case (a) can be done by the same method as Lemma I.6.c3 above, but let us indicate a different method, which parallels the one we shall use for (b). We know from Theorem I.6.4(iii) (Lang, p.35) that the number of conjugates of a q-Sylow subgroup Q of G is $\equiv 1 \pmod{q}$. But this number is the index in G of the normalizer of Q, which is a divisor of the index of Q itself, which is p. The only divisors of p are p and 1, and the former cannot be congruent to $1 \pmod{q}$ because it is strictly between 1 and q, hence the number of conjugates of Q must be 1; i.e., Q must be normal.

In case (b) we begin in the same way by noting that the number of conjugates of Q must be p^2, p or 1. As before, we cannot have $p \equiv 1 \pmod{q}$. Can p^2 be $\equiv 1 \pmod{q}$? This would mean that q divided $p^2 - 1 = (p-1)(p+1)$. Hence one of these factors must be divisible by q, hence must be *at least* q. But our hypothesis $p < q-1$ says that even the larger of these two factors, $p+1$, is $< q$. So as in case (a), Q must be normal.

The final sentence follows by Lemma I.6.c2. □

Corollary I.6.c5. *Any group of order $p^2 q$, where p and q are primes, has a normal p- or q-Sylow subgroup, and is thus a semidirect product of an abelian p-group and an abelian q-group.*

Sketch of Proof. If $p = q$, the first assertion is trivial, while the second follows from Exercise I.6:2. If $p > q$, then by *interchanging* the roles of p and q we are reduced to case (a) of the preceding lemma. If $p < q$, then we are OK by part (b) of that lemma unless $p = q-1$. But if two primes differ by 1, one of them must be even, so they must be 2 and 3 respectively. Thus, the order of G is $2^2{\cdot}3 = 12$, a case which, as I mentioned earlier, I like to analyze in class. (In that case, one finds examples both with 2-Sylow subgroup normal and with 3-Sylow subgroup normal.) □

I will not discuss here the question of *how many* nonisomorphic groups of order $p^2 q$ there are; in other words, the classification of actions of abelian groups of order p^2 on abelian groups of order q, and vice versa.

One might hope to stretch the method of Lemma I.6.c4 further, e.g., to prove that (perhaps with a small number of exceptions), for $p < q$, groups of order $p^3 q^j$ always have normal q-Sylow subgroups. But the argument does not generalize, essentially because in the factorization of $p^3 - 1$, namely $(p-1)(p^2 + p + 1)$, the second factor may be divisible by a prime significantly larger than p. In fact, whenever q is a prime dividing $p^2 + p + 1$, one can show that $\mathrm{Aut}((Z_p)^3)$ has order divisible by q, hence one gets a nontrivial map from Z_q to this automorphism group, which allows one to construct a group $Z_q \ltimes (Z_p)^3$ of order $p^3 q$ with non-normal q-Sylow subgroup.

Of course, in this situation one may try to prove that every group of order $p^3 q$ is *either* such a semidirect product $Z_q \ltimes (Z_p)^3$ *or* of the form $H \ltimes N$ where N is a normal q-Sylow subgroup. We shall note in the next paragraph one example for which this does not hold; I don't know whether it holds in all or most other cases. One can also give conditions which guarantee that $p^2 + p + 1$ is not divisible by q. One such condition is clearly $q > p^2 + p + 1$. Another is $q \equiv 2 \pmod 3$; you might try to discover a proof of this.

The smallest group that has *no* normal Sylow subgroup is S_4. Indeed, a 3-Sylow subgroup of S_4 is given by cyclic permutation of three of the four elements on which S_4 acts, and will not be normalized by a permutation that sends one of these three elements to the remaining element. Likewise a 2-Sylow subgroup will be the symmetry group of a square, where the elements $\{1, 2, 3, 4\}$ are arranged as the vertices of that square, and it is easy to see that a permutation that does not preserve the pairing of opposite vertices will not normalize it.

However, S_4 has a normal subgroup of index 2, which in turn has a normal 2-Sylow subgroup; thus S_4 is solvable. The smallest nonsolvable group is the 60-element group A_5.

Let us end by noting two further useful consequences of Lemma I.6.c2. The first is the simplest possible case of that lemma, and is a standard elementary exercise. (It is worked as an Example on p.28 of Lang.) The second yields a large class of examples that show that groups G of order n need not have subgroups of all orders dividing n.

Corollary I.6.c6 (to Lemma I.6.c2)**.** (a) *Any subgroup of index* 2 *in a group is normal.*

(b) *If a* simple *group* G *has a subgroup of finite index* r, *then* $(G:1) \mid r!$. □

In particular, we see from (b) that the simple group A_5 of order 60 can have no subgroups of order 15 or 20.

It is an amusing exercise to discover the structures of Sylow subgroups of symmetric groups. I like to do this interactively with the class, so I will not present it here; but I will give a terse summary of the result, which you can read after seeing the development.

Given any two finite groups G and H, let us form the direct product H^G of a family of copies of H indexed by the elements of G. Let G act on this product by translating the subscripts; i.e., by letting $g \in G$ send each G-tuple $(h_\gamma)_{\gamma \in G}$ to $(h_{g\gamma})_{\gamma \in G}$. Since this is an action of G on H^G by automorphisms, we can form the semidirect product induced by this action; this is called the *wreath product,* G wr H. (That term is reasonable in the light of the pictures we drew!)

Now let us write $P_{n,p}$ for the p-Sylow subgroup of S_n. One finds that for n a power of p, these groups are described inductively by

$$P_{p^m,p} \cong Z_p \text{ wr } P_{p^{m-1},p}.$$

To get a description for general n, we write n to base p: $n = a_m p^m + \ldots + a_0$, with $0 \leq a_i < p$. We then have

$$P_{n,p} \cong \Pi_i \, (P_{p^i,p})^{a_i}.$$

*Note on the much used word ''**structure**''.* In the last few pages, we have used that word many times. Let us observe that it has several senses in mathematics. On the one hand, looking at a group as a set with a multiplication operation satisfying appropriate conditions, one can call that multiplication operation a ''group structure'' on the set. (Similarly, a distance function gives a set a ''structure of metric space'', etc. etc..) On the other hand, when we spoke of studying the ''structures of groups of small orders'', this meant finding convenient descriptions of all such groups, up to isomorphism. (So in this usage, does ''structures'' just mean ''distinct isomorphism classes''? Not exclusively; for instance the above investigation of the Sylow subgroups of S_n helps us understand the ''structure'' of S_n, even though that involves no classification question.) In still another usage, one can speak of groups, fields, metric spaces, manifolds, etc. as various kinds of ''mathematical structures''. And these three senses merge into one another at times.

I consider it valuable to have words like this that do not have precise technical definitions, and so allow us to talk nontechnically about mathematics. In fact, I think it is unfortunate when someone takes one of these words and gives it a technical definition in some field of mathematics, depriving our communication of some of its flexibility.

Re §I.7. Direct sums and free abelian groups.

Before reading §I.7, I recommend reading Lang's set-theoretic Appendices 2.2-2.4, together with the comments on them in this Companion. (Those comments, in turn, indicate a point at which you should switch from Lang to my handout ''The Axiom of Choice, Zorn's Lemma, and all that''.) §I.7 can be read without that set-theoretic material, but then the important result Theorem I.7.3 is only proved in the finitely generated case. The discussion of that theorem below shows how to extend the proof to the general case.

P.38, line 4 [~]: Delete the condition ''if the family is not empty''! The empty set *is* a basis for the trivial group. (Lang will impose similar inappropriate nonemptiness conditions in several definitions in later sections as well. I will postpone discussing *why* I consider them inappropriate till later, though you can read that discussion, at the end of the material on §III.5, now if you wish.)

P.38, middle [~]: What Lang denotes $\mathbf{Z}\langle S\rangle$ I recommend writing as $\mathbf{Z}\,S$, to avoid conflict with ring-theorists' notation for free rings, and to agree with standard notation for group rings. (In any case, he uses the notation $\mathbf{Z}\langle S\rangle$ only briefly.)

P.38, last paragraph [>]: The comment that begins with the third sentence of this paragraph, ''If $g: S \to B$ is a mapping ... '', is very important! What is shown is that any set-map from a basis of a free abelian group F into any abelian group B extends uniquely to a *homomorphism* $F \to B$; intuitively, that we can choose homomorphisms $F \to B$ that send the elements of S to ''wherever we like''. It should be stated as a theorem; it will be used both by Lang and in this Companion. The result is known as the *universal property* of free abelian groups, and I will refer to it by that name below.

P.40, *Proof* of two italicized lines [~]: You have probably seen the construction of the field of fractions of a commutative integral domain (or at least the special case of the construction of $\mathbf{Q}$ from $\mathbf{Z}$) as a set of equivalence classes of ordered pairs. The construction Lang is about to describe is analogous to this; in fact, when applied to the abelian monoid of nonzero elements of such an integral domain under multiplication, it is exactly that construction (minus the consideration of 0 and addition). When applied to the *additive* monoid of nonnegative integers, it likewise gives the additive monoid of all integers; i.e., it ''brings in'' the negative integers. Lang assumes in the first sentence that students are more likely to have seen the integers-from-nonnegative integers construction than the field-of-fraction construction; though I think that under current US undergraduate curricula, the reverse is true.

In the second sentence, where Lang defines what is meant by ''(x,y) is equivalent to (x',y')'', you should check that this is indeed an equivalence relation. The verification uses the assumption that the cancellation law holds in M.

P.40, end of discussion of the Grothendieck construction [>]: Here are a few examples of that construction; you should be able to work out the verifications: If $M = \mathbf{N}$, i.e., the monoid of nonnegative integers, then $K(M) = \mathbf{Z}$. If M is a monoid generated by an element x such that the powers of x repeat cyclically with period n, starting with x^r (i.e., such that $x^r = x^{r+n}$, and thus $x^{r+i} = x^{r+i+n}$ for all $i \geq 0$, but no x^i with $j < r$ is equal to any other power of x), then $K(M)$ is the cyclic group of order n, i.e., the equation $x^j = x^{j+n}$ holds in $K(M)$ for all j. In particular, if $n = 1$, i.e., if the powers of x become constant from some point on, then $K(M)$ is the trivial group. If M is in fact a group, then $K(M) \cong M$; that is, the canonical map is an isomorphism.

P.41, proof of Lemma I.7.2 [~]: Recall from our discussion of semidirect products that a group homomorphism $f\colon G \to H$ is called a *split surjection* if f has a right inverse, $g\colon H \to G$, and that when this happens, gf is a *retraction* of G onto $g(H)$, and G is a semidirect product of $g(H)$ and the kernel N of f.

Now if G is abelian, this semidirect product will be the direct product (equivalently, the direct sum), since the action of H on G, and hence on N, by conjugation is trivial. Thus, a splitting of a surjective homomorphism of abelian groups, $f\colon A \to B$ corresponds to a decomposition of A as the direct sum of $C = \mathrm{Ker}(f)$ and a subgroup that is isomorphic to B (and, indeed, is mapped isomorphically via f to B).

Now suppose $f\colon A \to B$ is a surjective homomorphism of abelian groups, and B is *free*, say with a basis E. By surjectivity of f, we can find a set-map $E \to A$ taking each element of E to an inverse image in A, and by the universal property of free abelian groups, this extends to a homomorphism $g\colon B \to A$. Since $fg : B \to B$ agrees with the identity endomorphism of B on the basis E, it must equal the identity, so g is a right inverse to f. Thus, by our preceding observations, we get a decomposition of A as the direct sum of a subgroup isomorphic to B, and the subgroup $\mathrm{Ker}(f)$. This is the essentially Lang's proof of Lemma I.7.2, but he writes out as computations what we have embodied in general observations.

Remarks on the concepts of split surjections, split injections, and split exact sequences.

(Footnote to the above discussion.)

We have been using the terms ''split surjection'' for a homomorphism with a right inverse, and ''split injection'' for a homomorphism with a left inverse. It is clear that these are indeed surjections and injections; but why the term ''split''?

Recall that the relation between an abelian group A, its image B under a surjective homomorphism, and the kernel C of that homomorphism can be expressed by the *exact sequence*

$$0 \to C \to A \to B \to 0. \qquad \text{(c10)}$$

Exact sequences beginning and ending with ''0'' and with three groups in between are called ''short exact sequences''. Note that given any one-to-one group homomorphism f, i.e., any exact sequence $0 \to C \to$

A, if the image $f(C)$ is normal in A (as is automatic for abelian groups), then one can embed this in a short exact sequence (c10) in an essentially unique way, namely by taking $B = A/f(C)$. Likewise, given an onto homomorphism f, i.e., an exact sequence $A \to B \to 0$, one can embed this (without any additional assumptions) in a short exact sequence (c10) by taking $C = \mathrm{Ker}(f)$.

A short exact sequence (c10) is said to *split* if it is isomorphic to a direct sum of two ''trivial'' short exact sequences, $0 \to C \to C \to 0 \to 0$ and $0 \to 0 \to B \to B \to 0$ (with the nontrivial map the identity in each case), i.e., if it is isomorphic to a sequence

(c11) $$0 \to C \to C \oplus B \to B \to 0,$$

where the first map has the form $c \mapsto (c, 0)$ and the second the form $(c, b) \mapsto b$.

This use of the word ''split'' is, hopefully, quite natural. We can now justify the terms ''split surjective homomorphism'' and ''split injective homomorphism'' by showing that a surjective or injective homomorphism has an inverse on the appropriate side if and only if the corresponding short exact sequence splits. Indeed, it is straightforward to verify

Lemma I.7.c1. *Let*

(c12) $$0 \to C \xrightarrow{f} A \xrightarrow{f'} B \to 0$$

be a short exact sequence of abelian groups. Then the following conditions are equivalent.

(a) *The exact sequence* (c12) *splits; i.e., up to isomorphism, it has the form* (c11).

(b) *The image of* f, *equivalently, the kernel of* f', *is a direct summand in* A.

(c) *The surjective map* f' *splits, i.e., has a right inverse as a group homomorphism.*

(d) *The injective map* f *splits, i.e., has a left inverse as a group homomorphism.* □

P.41, proof of Theorem I.7.3 (continued on p.42) [=]: Let us do this in general, rather than limiting ourselves to the finitely generated case. Let X be a basis of A, and let us choose a *well-ordering* $\leq$ of X (cf. the handout on the Axiom of Choice, and/or Lang's Appendix 2.4). For each $s \in X$, let

$$A_{<s} = \text{subgroup of } A \text{ generated by } \{x \in X \mid x < s\},$$
$$A_{\leq s} = \text{subgroup of } A \text{ generated by } \{x \in X \mid x \leq s\}.$$

Note that $A_{<s} = \bigcup_{t<s} A_{\leq t}$. Given any $s \in X$, let $f_s : A \to \mathbf{Z}$ be the map taking each $a \in A$ to the *coefficient* of s in the unique expression of a as a linear combination of the elements of X. Thus, the restriction of f_s to $A_{\leq s}$ has kernel $A_{<s}$. Now let

$$S = \{s \in X \mid (B \cap A_{<s}) \neq (B \cap A_{\leq s})\}.$$

Thus, for $s \in S$, the image $f_s(B \cap A_{\leq s}) \subseteq \mathbf{Z}$ is a nonzero subgroup of an infinite cyclic group, hence is infinite cyclic; say $f_s(B \cap A_{\leq s}) = n_s \mathbf{Z}$. For each $s \in S$, let us choose $y_s \in B \cap A_{\leq s}$ such that $f_s(y_s) = n_s$, and let $Y = \{y_s \mid s \in S\}$. We claim that Y is a basis for B; this will prove B free, as required.

To show Y linearly independent, suppose

$$\Sigma_{s \in S}\, m_s y_s = 0 \qquad (m_i \in \mathbf{Z}).$$

If not all m_s are zero, let t be the *largest* element of S (under our ordering of X) such that $m_t \neq 0$. (A largest value exists because only finitely many m_s are nonzero.) Applying f_t to the above equation, we get $m_t n_t = 0$, a contradiction since both m_t and n_t were assumed nonzero.

To show that Y generates B, let us, for every nonzero $a \in A$, define $d(a)$ to be the largest element of X occurring with nonzero coefficient in the expression for a. Note that if $a \in B$, $d(a)$ is the least x such that $a \in A_{\leq x}$; thus $a \notin A_{<d(a)}$, so $d(a) \in S$. Now if Y does not generate B, let b be an element of B not in the subgroup generated by Y, and chosen, among all such elements, to *minimize* $d(b)$. (This is where we use the fact that $\leq$ is a *well*-ordering!) Thus $f_{d(b)}(b)$ has the form $r\, n_{d(b)}$. Hence

$b - ry_{d(b)}$ lies in $A_{<d(b)}$, i.e., $d(b - ry_{d(b)}) < d(b)$. Thus, by choice of b, the element $b - ry_{d(b)}$ *does* belong to the subgroup generated by Y, and adding $ry_{d(b)}$, we conclude that b lies in that subgroup; a contradiction. This completes the proof that B is free, with basis Y.

We now come to the proof of uniqueness of the rank (cardinality of a basis), which Lang gives in the top paragraph of p.42. I am torn between saying that Lang's proof, based on looking at the cardinality of B/pB is ''instructive'' and that it is ''a cheap trick''. If we regard the study of free abelian groups as the $R = \mathbf{Z}$ case of the study of free modules over a ring (to be defined in Chapter III), this counting trick cannot be extended to the general case; it just happens to work in this case because the ring $\mathbf{Z}$ has homomorphic images $\mathbf{Z}/p\mathbf{Z}$ which are finite. On the other hand, the idea of finding invariants of one kind of mathematical structure by applying to them a construction (what we will soon call a *functor*) taking them to structures of another sort that we have better information about, is a tool of great generality and power.

Well, let us be satisfied with this proof for now; we will see a different one that works for modules over general commutative rings later. Note, however, that this proof works only if one of the two bases is finite. This is not a fault of this particular proof; it is true of every proof of the finite case that I know, including those used in module-theoretic contexts. To get the infinite case, we shall use the following nice observation.

Lemma I.7.c2. *Let G be a group (not necessarily abelian) generated by an infinite set X. Then if Y is any other generating set for G, there is a subset $Y' \subseteq Y$ of cardinality $\leq \mathrm{card}(X)$ which also generates G.*

(*Likewise, if G is generated by a finite set X, then any generating set Y has a* finite *generating subset, though not necessarily of cardinality $\leq \mathrm{card}(X)$*).

Proof. For each $x \in X$, since x lies in the group generated by Y, it can be written as an expression in finitely many elements of Y; hence there is a finite subset $Y_x \subseteq Y$ such that x lies in the subgroup generated by Y_x. If we take the union of these sets Y_x, over all $x \in X$, we get a set $Y' \subseteq Y$ whose cardinality is $\leq \mathrm{card}(X)$ if X is infinite (cf. Theorem A2.3.3), or finite if X is finite, and such that the subgroup generated by Y' contains X, hence equals G. $\square$

Now if S and T are two infinite bases of a free abelian group B, then the above lemma shows that T has a subset T' of cardinality $\leq \mathrm{card}(S)$ which generates A. But it is easy to verify that a *proper* subset of a basis of A cannot generate A, hence T' must be T, showing that $\mathrm{card}(T) \leq \mathrm{card}(S)$. Similarly, $\mathrm{card}(S) \leq \mathrm{card}(T)$, proving the equality of these cardinalities. $\square$

This completes the proof of Theorem I.7.3.

It is interesting to note how the proof of the first part of that theorem, on existence of bases, which we carried out using Well-Ordering, would be proved using Zorn's Lemma. Given A, X and B as in that proof, let us, for any subset $S \subseteq X$, define A_S to be the subgroup of A generated by S. Let P denote the set of ordered pairs (S, Y) such that $S \subseteq X$, and Y is a basis of $B \cap A_S$, and let us write $(S, Y) \leq (S', Y')$ if $S \subseteq S'$ and $Y \subseteq Y'$. Since $(\varnothing, \varnothing) \in P$, P is nonempty. Clearly any chain $\{(S_i, Y_i) \mid i \in I\}$ in P has an upper bound, given by $\{(\cup S_i, \cup Y_i)\}$. Hence P has a maximal element (S, Y).

Having made this application of Zorn's Lemma, we are not at the end of the proof, but at the beginning of the group-theoretic part! We claim that from the maximality of (S, Y) in P, it follows that $S = X$. For suppose $x \in X - S$. There are two possibilities: If $B \cap A_{S \cup \{x\}} = B \cap A_S$, then we find that $(S \cup \{x\}, Y)$ is an element of P greater than (S, Y), contradicting the maximality of this pair. On the other hand, if $B \cap A_{S \cup \{x\}}$ is strictly larger than $B \cap A_S$, then we look at the nonzero map $f_x\colon B \cap A_{S \cup \{x\}} \to \mathbf{Z}$, choose an element $y \in B \cap A_{S \cup \{x\}}$ that is mapped to a generator of the cyclic image group, and show that $Y \cup \{y\}$ is a basis of $B \cap A_{S \cup \{x\}}$; so in this case, $(S \cup \{x\}, Y \cup \{b\})$ gives a contradiction to the maximality of (S, Y). (This step is a little easier than proving that Y is a basis for

B in the proof using Well-Ordering, because it simply involves adding one element to a set known to be a basis of a subgroup. You should work out the details.) Thus, the maximal pair (S, Y) satisfies $S = X$, so Y is a basis of $B \cap A_X = B$, as desired.

Re §I.8. Finitely generated abelian groups.

P.43, proof of Theorem I.8.2 (through p.45) [~]: To motivate the result, it is instructive to try to construct various sorts of finite abelian p-groups – e.g., by starting with cyclic p-groups, forming direct sums of these, and then forming factor-groups. We find that all the examples we can come up with can themselves be written (after a change of generators) as direct sums of cyclic p-groups. So we would like to see whether we can prove that every finite abelian p-group A has this form.

To get such a proof, we need a construction that, starting with an arbitrary finite abelian p-group, tells us how to find generators for a family of cyclic direct summands. Such a system of summands is not unique (although having read the statement of Theorem I.8.2, we know that it will be unique up to isomorphism). E.g., if $A = Z_{p^3} \oplus Z_p$, we find that along with the given cyclic summands, generated by $(1, 0)$ and $(0, 1)$ there are others, such as those generated by $(1, 1)$ and $(p^2, 1)$.

However, we cannot start with an arbitrary cyclic subgroup and expect it to form part of a direct sum decomposition. E.g., if $A = Z_{p^2}$, then the cyclic subgroup generated by p is not a direct summand. Looking at what goes wrong in that case, we might conjecture that any *maximal* cyclic subgroup of a finite abelian p-group A would be a direct summand. But even this is not so: in $A = Z_{p^3} \oplus Z_p$, the cyclic subgroup generated by $(p, 1)$ is maximal (since that element is not of the form pa, as we see by looking at its second coordinate), but it cannot be part of a direct-sum decomposition, since it has order p^2, and we know from the statement of Theorem I.8.2 that in any decomposition of A as a direct sum of cyclic subgroups, the summands must have orders p^3 and p.

In the proof of Theorem I.8.2 in Lang, the trick used is to choose a cyclic subgroup of *maximal order*, i.e., one whose order is an exponent of the whole group A. Using this same idea, we shall prove below a stronger result than that theorem, namely that any *not necessarily finite* abelian group of exponent p^n is a direct sum of cyclic subgroups. The key step, which shows that we can ''split away'' arbitrary direct sums of cyclic subgroups of maximal order, is the following:

Lemma I.8.c1. *Let p be a prime and n a positive integer.*

(a) *If B is a direct sum* (*or a direct product*) *of cyclic groups of order p^n, then it satisfies*

(c13) *Every element $x \in B$ of exponent p^i can be written $p^{n-i}y$ for some $y \in B$.*

(b) *If A is an abelian group of exponent p^n, and B a subgroup satisfying* (c13), *then B is a direct summand in A* (*i.e., there exists a subgroup $C \subseteq A$ such that $A = B \oplus C$*).

Proof. (a) is straightforward. To prove (b), we shall show how to ''build up'' a complementary summand C. Suppose we have a ''partial'' complementary summand; that is, a subgroup $D \subseteq A$ satisfying $B \cap D = \{0\}$, but such that $B + D$ is not all of A. Thus there exist elements of A not in $B + D$. If we multiply such an element by p enough times, we will get zero (since A has exponent p^n); in particular, after a certain (possibly smaller) number of multiplications by p, we get an element of $B + D$. Hence if we stop one step short of this, we get an element $x \notin B + D$ such that $px \in B + D$. Thus

$$px = b + d \quad \text{where} \quad b \in B,\ d \in D.$$

Since x is annihilated by p^n, if we multiply both sides of this equation by p^{n-1} we get $0 = p^{n-1}b + p^{n-1}d$. Hence $p^{n-1}b = -p^{n-1}d \in B \cap D = \{0\}$, so b is annihilated by p^{n-1}. Hence by assumption (c13), we can write $b = pb'$ $(b' \in B)$. Thus, if we let $y = x - b'$; the above displayed equation gives $py = d$. We claim that if we set $D' = D + \langle y \rangle$, then this will be a larger subgroup than

D, but will still satisfy $B \cap D' = \{0\}$.

To see that D' is strictly larger than D, it suffices to note that $B + D'$ contains $b' + y = x$, which by assumption is *not* in $B + D$. This observation can be translated to say that the image $\bar{y}$ of y in $A/(B + D)$ equals the image of x therein, and so is nonzero. Since by assumption $px \in B + D$, this element $\bar{y}$ has order p. We can now prove that $B \cap D' = \{0\}$. Indeed, suppose that an element of D', say $e + ry$ $(e \in D,\ r \in \mathbf{Z})$ lies in B. Then its image in $A/(B + D)$ is 0; but we can also write this image as $r\bar{y}$ (because $e \in D$ has image 0); so $r\bar{y} = 0$. Since $\bar{y}$ has order p, this says that r is divisible by p, hence ry is a multiple of $py \in D$. Hence $e + ry \in D$, hence, as $B \cap D = \{0\}$, we have $e + ry = 0$. Since $e + ry$ was an arbitrary element of $B \cap D'$, this shows that $B \cap D' = \{0\}$, as claimed.

This passage, from a partial complement D of B to a larger partial complement D', gives the general ''step'' of our desired construction. We now put such steps together using Zorn's Lemma: Let P be the set of all subgroups $D \subseteq A$ with the property $B \cap D = \{0\}$, partially ordered by inclusion. It is easy to verify that the union of any chain in P lies in P, and hence forms an upper bound for that chain. So by Zorn's Lemma, P has a maximal element C. We claim that

$$B + C \ = \ A.$$

For if this were not so, our preceding argument would give us a larger $C' \in P$, contradicting the maximality of C. Since $B \cap C = \{0\}$ by definition of P, we have $A = B \oplus C$, as desired. □

From this, we can now prove

Theorem I.8.c2. *If A is an abelian group of exponent p^n, then A is a direct sum of cyclic subgroups of orders dividing p^n.*

Proof. Let us begin by constructing the family of subgroups of order exactly p^n that will occur in our decomposition. To do this, let P be the set of all sets X such that X is a set of cyclic subgroups of A, each of order p^n, whose sum in A is direct. It is easy to verify that the union of any chain in P is a member of P, and is thus an upper bound for that chain. Hence P has a maximal element X; let $B \subseteq A$ be the sum of the subgroups comprising X. We note that by definition of P, this is a direct sum. By part (a) of the preceding lemma, B has property (c13), hence by part (b) of that lemma, we can write $A = B \oplus C$. Being a subgroup of A, C has exponent p^n; but we claim it has no element of order exactly p^n. For if c were such an element, then $X \cup \{\langle c \rangle\}$ would be a member of P larger than X, contradicting maximality. It follows that C has exponent p^{n-1}.

We may now assume by induction on n that C is a direct sum of various cyclic subgroups of orders p^i with $i \leq n-1$. Putting together this decomposition of C and our given direct sum decomposition of B, we get the desired decomposition of A. □

We noted earlier that a maximal cyclic subgroup of a group of exponent p^n need not be a direct summand; but our proof of the above theorem (via the preceding lemma) shows that a maximal cyclic subgroup *having largest possible order* is indeed a direct summand. Curiously, if among maximal cyclic subgroups we choose one having *smallest* order, it will also be a direct summand. This leads to another proof of the above theorem, sketched in Exercise I.8:7. We thus see that our example of a maximal cyclic subgroup which was *not* a direct summand had to avoid both these extremes: it in fact had order p^2 in a group whose largest- and smallest-order cyclic summands had orders p^3 and p respectively.

Still another way of proving the above theorem is developed in Exercise I.8:6. (The idea there is to choose a generating set of the exponent-p abelian group $A_p = \{a \in A \mid pa = 0\}$, in a way which takes account of the chain of subgroups consisting of elements that are ''multiples of p^{n-1}'', ''multiples of p^{n-2}'', etc.. One can also give a proof ''dual'' to this, in which one chooses a basis of the exponent-p abelian group A/pA in a way which takes account of the chain of subgroups given by those elements which are images of elements of A of exponent p, of exponent p^2, etc..)

For the proof of *uniqueness up to isomorphism* of these decompositions, when the group A is finitely generated, see Lang, p.44 (bottom) to 45. The idea is that, if we write A_p for the subgroup $\{a \in A \mid pa = 0\}$, then the number of cyclic summands of each order can be read off from the sizes of the subgroups $A_p \cap p^{i-1}A$. This idea can be adapted to our non-finitely-generated context with a little work; you will not be responsible for this generalization, but if you are interested, you might think about how this would be done.

Of course, from the above theorem we can get, using Theorem I.8.1 (p.42) the

Corollary I.8.c3. *If A is an abelian group of exponent m, then A is a direct sum of cyclic subgroups of orders dividing m.* □

Every time I teach this section, I end up getting fascinated by questions of ways that the main results of the section can and cannot be generalized; the number of exercises relating to this material that I have accumulated is testimony to this fact. There is, in fact, a well-developed theory of abelian groups, which I have never had time to study. If you would like to look further into it, a standard work in the field is L. Fuchs, *Abelian groups,* Pergamon Press, 1960. (A reference to a book that studies similar questions in a module-theoretic context is given at the end of my comments on §III.7 below.)

P.46, line 2 [=]: For ''a finite set'' read ''the finite set S''.

P.46, line preceding Theorem I.8.5 [~]: Lang writes ''mA is free, hence A is free''. The argument is that multiplication by m is a homomorphism $A \to mA$; it is clearly onto, and it is one-to-one because A is torsion-free. Hence $A \cong mA$, so if the latter is free, the former is, also.

Re §I.9. The dual group.

P.46, beginning of the section [~]: On pp.24-25, the ''standard'' cyclic group of order n was written $\mathbf{Z}/n\mathbf{Z}$; here Lang writes Z_n for an arbitrary cyclic group of order n. Such a group is isomorphic to $\mathbf{Z}/n\mathbf{Z}$, but not in a unique way. He will apply the results of this section in §VI.8, in a situation where the cyclic group in question is the group of nth roots of unity in a field (e.g., for $n = 4$, the group of complex numbers $\{1, i, -1, -i\}$), so that it would be inconvenient to have the results stated only for the ''standard'' group.

Note that if m is an exponent of the abelian group A, then any multiple rm of m is also an exponent of A. If we work with ''standard'' cyclic groups, we have a canonical embedding $\mathbf{Z}/m\mathbf{Z} \to \mathbf{Z}/rm\mathbf{Z}$, taking $[1]$ to $[r]$, whose image is precisely the group of elements of $\mathbf{Z}/rm\mathbf{Z}$ of exponent m. If we compute $A^\wedge$ with respect to each these cyclic groups, we find that all homomorphisms of A into $\mathbf{Z}/rm\mathbf{Z}$ have range in the image of this canonical embedding, so that there is a natural way of identifying the two groups called ''$A^\wedge$'' that we get by dualizing with respect to these two cyclic groups. Unfortunately, when we work with arbitrary cyclic groups, we lose this canonical identification; thus, Lang is being a bit sloppy in writing ''$A^\wedge$'' without specifying which cyclic group is being used. A good rule for interpreting his results precisely is that whenever finitely many groups are mentioned, m denotes some *common* exponent of these groups, and duals are then taken with respect to some fixed cyclic groups Z_m of order m.

P.47, Theorem I.9.1 [>]: This theorem makes two statements of isomorphism. I claim that the first of these is a ''natural'' isomorphism, and the second an ''unnatural'' one. Indeed, the theorem gives one a precise prescription for constructing an isomorphism $(B \times C)^\wedge \cong B^\wedge \times C^\wedge$; but the isomorphism $A^\wedge \cong A$ for A a finite abelian group depends on one's choice of a family of cyclic subgroups having A as their direct sum: pick a different decomposition, and you will get a different isomorphism. We shall see later that ''natural'' constructions have certain particularly useful properties that constructions requiring choices do not have. The isomorphisms described in the remaining two results in this section are again ''natural''.

P.48, last line [=]: ''is in fact'' should be ''is in effect'', or better, ''induces''. The same correction applies on the third line of the next page. There, the homomorphism of the second display is the map taking each coset $[x'] \in A'/B'$ (i.e., if additive notation is used, $x'+B'$, if multiplicative notation is used, $x'B'$) to the homomorphism $\psi_{x'}: A/B \to C$.

P.49, third line of Theorem I.9.2 [>]: Before the word ''*Assume*'' add the sentence ''*Then A/B and A'/B' are of exponent m.*'' This comes easily out of the proof given.

I will insert here an important result which Lang has omitted.

Proposition I.9.c1. *Let A be a finite abelian group, and let us define a map $\theta_A: A \to A$^^ by $\theta_A(x)(\alpha) = \alpha(x)$ ($x \in A$, $\alpha \in A$^). Then θ_A is an isomorphism. Moreover, if $f: A \to B$ is any homomorphism of abelian groups, then $\theta_B\, f = f$^^θ_A; i.e., the diagram*

$$\begin{array}{ccc} A & \xrightarrow{\quad f \quad} & B \\ \Big\downarrow{\scriptstyle \theta_A} & & \Big\downarrow{\scriptstyle \theta_B} \\ A\text{^^} & \xrightarrow{\quad f\text{^^} \quad} & B\text{^^} \end{array}$$

commutes.

Sketch of Proof. The fact that θ_A is a homomorphism, and the equation in the last statement (which intuitively says that the maps θ_A and θ_B identify the finite abelian groups A and B with their double duals in a way ''consistent with'' the construction $f \mapsto f$^^ on group homomorphisms), are straightforward verifications, which you should write out; the nontrivial result is that these maps are *isomorphisms*. The most direct way to get this is to write A as a direct sum of cyclic subgroups, and verify that θ_A carries each of these summands isomorphically to the corresponding summand in A^^.

An alternative way is to apply Theorem I.9.2, with A^ in the role of A, A in the role of A', Z_m in the role of C, and the evaluation function $(\alpha, x) \mapsto \alpha(x)$ in the role of the given bilinear map. Then you will find that the kernel on the left is zero (why?), hence, writing K for the kernel on the right, that Theorem says that Lang's canonical map ψ gives an isomorphism of A/K with (A^$/0$)^ = A^^. Since A and A^^ have the same order, the kernel K must also be zero, so ψ gives an isomorphism $A \cong A$^^. Moreover, comparing definitions, one sees that ψ is the map we have named θ_A. □

P.49, Proof of Corollary I.9.3 [=]: To get this as a ''special case'' of the theorem, consider the bilinear map of the last display of p.48, with $C = Z_m$, so that Hom(A, C) is just A^, then restrict the resulting bilinear map $A \times A$^ $\to Z_m$ to a map $B \times A$^ $\to Z_m$, and apply the theorem to this restriction. (The kernel of this map on the left will be trivial – why?)

P.49, end of §I.9 [>]: As we noted earlier, the dualities constructed as in this section using different cyclic groups Z_m and Z_n, where m and n are both exponents of some family of abelian groups, can be related by regarding Z_m as embedded in Z_n whenever m divides n. In fact, in some pleasant cases those embeddings are actually inclusions. This is true when Z_m is the multiplicative group of all mth roots of unity in an appropriate field, mentioned earlier. For a case accessible to us now, consider the factor group $\mathbf{Q}/\mathbf{Z}$. For each m, the subgroup of elements of exponent m is the cyclic group of order m generated by the image of $1/m$, i.e., $(1/m)\mathbf{Z}/\mathbf{Z}$, and clearly, whenever m divides n we have an inclusion of these finite groups. Thus, the various duals A^ of a finite abelian group A that one obtains using different cyclic subgroups of $\mathbf{Q}/\mathbf{Z}$ can all be written Hom(A, $\mathbf{Q}/\mathbf{Z}$), yielding a single coherent duality theory.

(If you some day study topological abelian groups, you will learn of ''Pontryagin duality'' among locally compact abelian groups, defined by A^ = Hom(A, $\mathbf{R}/\mathbf{Z}$), where ''Hom'' is now defined as the group of *continuous* homomorphisms, and is given an appropriate topology. It is easy to see that the torsion subgroup of $\mathbf{R}/\mathbf{Z}$ is $\mathbf{Q}/\mathbf{Z}$, from which it follows that for A finite, $\mathrm{Hom}(A, \mathbf{R}/\mathbf{Z}) = \mathrm{Hom}(A, \mathbf{Q}/\mathbf{Z})$. Hence Pontryagin duality, when restricted to finite groups, gives the duality treated in this section.)

We remark that Lang's symbol A^ is not standard. A much commoner symbol for dualities of various sorts is $\hat{A}$. (Perhaps Lang wishes to avoid this symbol because it has so many other uses, such as denoting various sorts of ''completion''.)

The interested student might like to think about which definitions and arguments in this section make sense for the constructions $\mathrm{Hom}(-, R)$ for an arbitrary abelian group R, and which are specific to cyclic groups.

Re §I.10. Inverse limit and completion.

This is an interesting subject, but one which we generally don't have time to cover in Math 250A. Some time I hope to go through this section and set down full commentary. I give below a few errata supplied by Bjorn Poonen who covered this section in Fall 2003, and some comments of my own resulting from reading the points to which those errata applied.

P.50, beginning of last line of first paragraph [=]: lim should be $\varprojlim$ (as on the preceding line); in earlier printings, the first two displays are also missing the arrow beneath the symbol.

I also suggest reading $A[p^n]$ in place of $A[p^{n+1}]$ in the first display. The limit group is the same whichever indexing is used, but the former indexing is more straightforward, and consistent with the third display.

P.51, third paragraph [>]: There are two sorts of ''limits'' of groups (or other mathematical structures), known as ''direct limits'' and ''inverse limits''. The former is constructed from what is called a ''directed system'' of groups, the latter from what is called an ''inversely directed system''. In this section, Lang only discusses the latter construction, and in this paragraph he writes ''(inversely) **directed system**'', meaning that he will often drop the word ''inversely'' for brevity. However, I would advise against doing so, since it can lead to confusion as soon as the other concept is also present; he will introduce that on p.160. (One might use an adjective to distinguish the two sorts of systems whenever there is ambiguity, and say ''directed system'' when the context indicates which sort is meant; the trouble is that I don't know of any adjective commonly used to specify the sort of system from which one constructs a direct limit of groups. Perhaps one could call it a ''forward directed system''.)

Roughly speaking, an inversely directed system is a (generally infinite) diagram of groups and homomorphisms among them, satisfying certain conditions on composition of these homomorphisms, and such that given any two groups in the diagram, there is at least one group that has arrows into both of them. One makes this precise by indexing the groups by an appropriate sort of partially ordered set I, and letting the system have a map from one group another whenever there is an inequality between the indices.

Of course, one has to decide whether the homomorphisms should go from the object with the smaller index to the one with the larger index or vice versa. Moreover, of these two possibilities, we might make the *same* choice for both kinds of limits, indexing our systems of groups by partially ordered sets with *opposite* properties, or make *opposite* choices for the two sorts of limits, indexing them by partially ordered sets with the *same* property! Different authors make different choices. Lang, at a later point where he discusses both constructions in a general setting (pp.160-161) uses in both constructions the same sort of ordered set, called a ''directed'' partially ordered set, and indexes the directed system in different ways in the two cases. The background of this choice is that the earliest cases that were considered, before the general concept was developed, involved ''sequences'' of groups, and people naturally indexed both the directed and inverse systems by the positive integers. However, be prepared to see different notations in different works. The properties of the diagrams defining ''direct'' and ''inverse'' limits will always be the same, but the formalism of their descriptions in terms of partially ordered sets will differ.

For an additional bit of confusion, the terms ''direct limit'' and ''inverse limit'' have synonyms, ''inductive limit'' and ''projective limit'' respectively. I find these unsuggestive, and often have trouble remembering which means which. Lang generally keeps to ''direct limit'' and ''inverse limit''; I will give corrections below in the few places where he slips and uses the other terms. (The term ''profinite group'', introduced toward the bottom of p.51, is short for ''projective limit of finite groups''. That is a standard term, and when I need to remind myself which of the above names refers to which kind of limit, I work backward from that.)

P.52, last two paragraphs [<]: Lang doesn't seem to have made up his mind whether he wants to tell us about the completion of a group G with respect to an arbitrary directed system $\mathfrak{F}$ of normal subgroups, or only about the special case where $\mathfrak{F}$ is the set of all normal subgroups of finite index, so what he says is a bit inconsistent. The latter construction, called the ''profinite completion'' of G, is important; but since he doesn't go into the theory of that completion, let us assume $\mathfrak{F}$ is a general directed system of normal subgroups. I will indicate the necessary adjustments in what he writes along with other corrections below.

P.52, first three sentences of next-to-last paragraph [=]: Lang defines a ''Cauchy family'' in G to be indexed by the directed partially ordered set of subgroups $\mathfrak{F}$; but this is not in general a natural choice. A more general definition makes a Cauchy family a system $(x_j)_{j\in J}$ indexed by any directed partially ordered set J, and satisfying the condition that for every $H\in\mathfrak{F}$ there exists a $j\in J$ such that for all k, $k' \geq j$ one has $x_k x_{k'}^{-1}\in H$. However, Lang has introduced the concept of Cauchy family only for perspective; he will not use the concept, so you do not have to learn one version or the other.

P.52, last two sentences of next-to-last paragraph [=]: Let us put off for the moment the question of why having $\mathfrak{F}$ denumerable means that ''in practice one can work with (Cauchy) sequences'', and merely note that Lang is here saying that the system of subgroups of finite index is often countable, and is often the system we are interested in.

P.52, final display [=]: The left-hand side of the display should read $\varprojlim_i G/(H_1\cap\ldots\cap H_i)$. Likewise, on the top of the next page, ''the sequence $\{H_i\}$'' should be ''the sequence $\{H_1\cap\ldots\cap H_i\}$''.

The key idea here is that from a countable cofinal *set* of subgroups $H_1, H_2, H_3, \ldots$ one can obtain a countable cofinal *chain*, $H_1 \supseteq H_1 \cap H_2 \supseteq H_1 \cap H_2 \cap H_3 \supseteq \ldots$, and that the inverse limit over *that* chain is isomorphic to the inverse limit of the original system. This last statement is not obvious, but, as Lang says, it makes a nice exercise.

(Actually, one can show from the cofinality assumption that there is a *subsequence* (H_{i_m}) of (H_i) which satisfies $H_{i_{m+1}} \subseteq H_{i_m}$ for all m, and is still cofinal in $\mathfrak{F}$. This gives the desired chain without taking intersections, and the statements Lang makes hold with H_{i_m} in place of H_i. This may be closer to what he had in mind, but more words would have to be inserted to express it.)

Re §I.11. Categories and functors.

P.53, definition of ''category'' [~]:

To motivate this concept it is best to begin by comparing it with some more familiar concepts.

The definition of a *monoid* is obtained by abstracting the properties of the set of all maps from a mathematical object X into itself: the set of such maps is closed under the operation of composition, which satisfies the associative law, and it contains an identity element, the composite of which with any other element f is f. A general system of elements with an operation named ''composition'', having these properties, *but not necessarily assumed to arise this way*, is called a *monoid*.

Similarly, the concept of a *group* is motivated by considering the properties of the set of *automorphisms* of a mathematical object. The set of such automorphisms has a structure of monoid, but is also closed under taking *inverses*; this gives another operation, which satisfies familiar identities.

If one considers the set of *subobjects* of a mathematical object, there are no operations one can automatically define on this set, but there is a relation of ''inclusion''. A generalization and abstraction of the properties of this relation motivates the definition of *partially ordered set*.

Next, an unfamiliar, but nevertheless similar situation: Suppose we consider two mathematical objects X_0 and X_1, and the set S of all maps among these. This set falls naturally into four pieces: S_{00}, consisting of all maps from X_0 to itself, S_{01}, consisting of maps from X_0 to X_1, etc.. One can compose various of these maps, but not every map is composable with every other, so one does not have a composition operation $S \times S \to S$, but rather, a family of 8 operations, $S_{jk} \times S_{ij} \to S_{ik}$ $(i, j, k \in \{0,1\})$. One easily sees that these 8 operations satisfy 16 associativity laws, corresponding to multiplication on the 16 sets $S_{jk} \times S_{ij} \times S_{hi}$, and that there are 2 identity elements, $\mathrm{id}_0 \in S_{00}$ and $\mathrm{id}_1 \in S_{11}$. One might abstract this situation, and call a family of four sets S_{00}, S_{01}, S_{10}, S_{11} given with operations satisfying the appropriate conditions a ''*bimonoid*''. Note that, as in the definitions of ''monoid'', ''group'' etc., this definition would *not* presuppose that the sets S_{ij} were actually sets of maps among two mathematical objects, nor that the composition operations were composition of such maps, but only that the system satisfied the algebraic laws motivated by that situation. (A virtue of this abstraction is that one could consider many ''representations'' of *the same* abstract ''bimonoid'' on different pairs of objects.)

One might go on to define ''trimonoid'', ''quadrimonoid'' etc., but this would clearly be a waste of paper: once the pattern is clear, one should make a general definition that covers all cases; namely, a model for the algebraic structure one gets when one considers a *family of mathematical objects and all the maps among these*. The resulting sort of structure is called a *category*. The family that replaces the index-set $\{0,1\}$ in the definition of ''bimonoid'', the index-set $\{0,1,2\}$ in the unstated definition of ''trimonoid'', etc., is called the class of *objects* of the category; the family that replaces the set S is called the class of *morphisms*.

One of my main reasons for giving the above discussion is to make clear that in the definition of a category, as in the preceding definitions, we do *not* assume that the structure we are considering actually consists of a family of mathematical objects of some sort and the homomorphisms among its members. The relation between an abstract category and a category arising by taking a class of mathematical objects and the homomorphisms among them is analogous to the relation between an abstract group and the group of automorphisms of a mathematical object. In each case, the abstract concept is *modeled on* the concrete concept; indeed, we might think of an abstract group as giving ''a potential structure for the automorphisms of an entity'', and of an abstract category as ''a potential structure for a class of entities and the homomorphisms among them''. But in each case, the abstract concept has less information than the motivating situation. Thus, though in talking about a group of permutations of a set one can ask a question such as, ''Does this group have any fixed points?'', in talking about an abstract group, one cannot. Similarly, in proving results about abstract categories, one cannot use arguments that begin ''Let x be an element of the object X.'' (The temptation to do so is strong for the beginner! In every class where I have introduced the concept of category, I have emphasized the above point, and yet in almost every such class, at least one student has used such an argument in his or her homework.)

P.54, end of second paragraph of **Examples** [>]: Lang says the axioms for a category are trivially satisfied by these familiar examples. The verification of the nontrivial axioms **CAT 2** and **CAT 3** is indeed trivial in these cases; but whether they satisfy ''trivial'' axiom **CAT 1** is an iffy question.

For instance, if A is the set of even integers and B the set of all integers, is the inclusion map of A in B the same as the identity map of A? By the usual definition of function that one learns in a course in set theory (as a certain sort of subset of the direct product of two sets), they are the same; hence if in $\mathbb{S}$ we define $\mathrm{Mor}(A, B)$ to be the set of all functions $A \to B$, then $\mathrm{Mor}(A, B)$ and $\mathrm{Mor}(A, A)$ are *not* disjoint, contradicting **CAT 1**.

There are various ways to fix this.

One is to slightly change our definition of function, letting a function $A \to B$ mean a 3-tuple (A, f, B), where f is what we previously called a function from A to B. Then functions from A to B in our new sense are in one-to-one correspondence with functions from A to B in the old sense, but the set of functions from A to B in the new sense is disjoint from the set of functions from A' to B', unless $A = A'$ and $B = B'$.

Another, only subtly different, is to keep our definition of function unchanged, but define a morphism from A to B in $\mathbb{S}$ to mean a 3-tuple (A, f, B) as above.

A third approach, taken in some developments of category theory, is to drop **CAT 1** from the axioms. This has some disadvantages; as one of many examples, if we are given a morphism f, the statement ''f is invertible'' is ambiguous; it only becomes precise if one says ''f is invertible as a map $X \to Y$''. Nevertheless, one may prefer to leave it out simply to minimize the gap between one's category-theoretic usage and conventional mathematical usage. (This is the decision I make in my Math 245 notes.)

We will follow the textbook and assume **CAT 1** here. I leave it to you which way you prefer to adjust your conventions to make this consistent with our non-categorical notation; it will not affect the material to follow.

(Incidentally, while in conventional usage, a set-map has a well-defined domain but not a well-defined codomain, if we leave out **CAT 1**, we have no guarantee that either the domain or the codomain of a morphism in a category will be well-defined.)

P.54, remaining Examples (after $\mathbb{S}$, **Grp**, **Mon**) [<]: I recommend skipping the material from here through p.57, except for the two paragraphs near the top of p.55 beginning ''Next we consider ... '' and ''More generally ... ''. (In the fourth line of the latter paragraph, for ''In practice'' read ''In many cases''.) Most of Lang's other examples involve material that few beginning students will have had, and are not clearly presented. (If you do continue on p.54, you will find that his comment three lines later, ''It is important to emphasize ...'', is quite off the wall, since none of the three examples he has given so far had the property that its morphism-sets formed groups.)

In place of (or in addition to) Lang's examples, read the examples and (optionally) the notes on category theory and set theory below, then go back to Lang where indicated at the end of those notes.

Examples. Here are a couple of interesting categories in which the morphisms are not set-maps. First, one can define a category **Rel** whose objects are the same as the objects of **Set** (what Lang calls $\mathbb{S}$; most authors use more mnemonic names), but where a morphism $X \to Y$ means any *relation* $R \subseteq X \times Y$, and where composition is defined to mean composition as relations; i.e., for $R \subseteq X \times Y$ and $S \subseteq Y \times Z$, one defines

$$SR \;=\; \{(x, z) \in X \times Z \mid (\exists y \in Y)\ \ (x, y) \in R,\, (y, z) \in S\}.$$

Secondly, for students who have seen the definition of the fundamental group $\pi_1(X)$ of a topological space X, observe that given such a space X, we may define a category $\boldsymbol{\pi}_1(X)$ in which the objects are the points of X, a morphism from the point x_1 to the point x_2 means a homotopy class of paths from x_1 to x_2, and homotopy classes of paths are ''composed'' as in the definition of the fundamental group. (A curious property this category is that every morphism is invertible. A category with this property is often called a *groupoid*.)

Another class of categories is described in exercise I.11:6(a).

Category theory and set theory (optional). If one wants to study concepts like ''the category of all groups'', one is faced with the problem that under conventional set theory, all groups form a *class*, but not a *set*, and there is very little one is allowed to do with classes. One often calls a category which forms a genuine set a ''small category'', and though one can work easily with these, they exclude most of the categories one is interested in.

Here is a creative way out of this dilemma, due to Alexander Grothendieck. Conventional set-theory

(ZFC, i.e., Zermelo-Fraenkel set theory with the axiom of Choice) has various axioms asserting that given certain sets, one can obtain certain other sets. Let us think of these as saying that the *class of all sets* considered in the set theory is closed under certain constructions. ZFC neither asserts nor denies that there may be *sets* $\mathbb{U}$ within the theory that are themselves closed under all these constructions (i.e., such that, for all $X \in \mathbb{U}$ every member of X belongs to $\mathbb{U}$, the power set of X belongs to $\mathbb{U}$, etc., etc.). Let us call such a set $\mathbb{U}$ a ''universe'' (or a ''Grothendieck universe''). Thus, a universe will look *internally* like a full ZFC set theory, though externally it will be a *member* of our set-theory. Let us now adjoin to the axioms of ZFC one further axiom, the ''Axiom of Universes'', saying that every set is a *member* of a universe. Given a choice of a particular universe $\mathbb{U}$, let us define a ''small set'' (relative to $\mathbb{U}$) to mean a member of $\mathbb{U}$, and then define a ''small group'' (or other mathematical object) to mean a group (etc.) which is small as a set. Within our larger set theory, there exists a set of *all small* sets (namely $\mathbb{U}$), hence also a set of all small groups, and similarly for other sorts of mathematical objects. Thus, in this set theory, the categories of all small sets, small groups, etc., are genuine sets (though not small sets!), and can be handled mathematically without kid gloves. (Even the collection of all categories of the above sort – categories having object-sets and morphism-sets which are subsets of $\mathbb{U}$ – will form a genuine set in our set-theory. It will not be a small set with respect to $\mathbb{U}$, but it will be small with respect to any larger universe $\mathbb{U}'$.)

Since ZFC was apparently sufficient for studying group theory before categories came along, one can regard the category of *small* groups (with respect to an arbitrary fixed universe $\mathbb{U}$) as embracing the subject of traditional group theory, and likewise for other sorts of mathematical objects; and these categories can be treated conveniently within set theory. If you like this approach, you should henceforth read references to ''the category of all groups'', etc., as referring to these categories.

We will not develop these ideas further in this course, but simply note that they give a way out of the set-theoretic dilemma with which we began this remark. I may, in fact, for consistency with common usage, sometimes speak of the objects of the category of groups as forming ''the class of all groups''. Lang, as you may have noticed, begs these questions by using the vague term ''collection'' in his definition of category on p.53. (In my 245 notes, I start with a similarly vague ''Provisional Definition'' of category when I motivate the concept in §6.1, but then introduce the Axiom of Universes in §6.4, revise the definition of a category to require that its objects and morphisms form sets, and thereafter use this modified set theory.)

Now return to Lang, at the paragraph on p.57 beginning ''Let $\mathcal{A}$ be a category ... ''.

P.57, line after square diagram, ''... Strictly speaking ...'' [>]: Recall the comment on **CAT 1** in my discussion of p.54. If $\mathcal{A}$ satisfies **CAT 1**, then defining morphisms of $\mathcal{C}$ as tuples (f, φ, ψ, f') would insure that $\mathcal{C}$ satisfied that condition. Lang's approach to the question seems to be that in theory, we make that definition, but when there is no danger of ambiguity, for the sake of brevity we just write (φ, ψ).

P.58, second paragraph of Examples [~]:

Some notational matters.

Lang wants to show us here that a ''free abelian group'' is an example of a universal repelling object. To understand this approach, recall that the universal property of the free abelian group $F_{\mathrm{ab}}(S)$ on a set S says that it is an abelian group with a map f of S into it, which is ''universal'' among all abelian groups with maps of S into them, in the sense that such that given any abelian group B and map $g\colon S \to B$, there is a unique group homomorphism $g_*\colon F_{\mathrm{ab}}(S) \to B$ respecting maps from S; i.e., such that $g(s) = g_*(f(s))$ for all $s \in S$. (Lang, p.38, last paragraph, ''If g: ...''.) This suggests that, given a set S, we should define a *category* whose objects are abelian groups given with maps of S into them, with morphisms defined as group homomorphisms which respect these maps (i.e., which form commuting triangles with them). The group $F_{\mathrm{ab}}(S)$, given with the map f, will then be a ''universal repelling object'' in this category.

The straightforward way to formalize the definition of this category takes for the objects the pairs (B, g) such that B is an abelian group and $g: S \to B$ a set-map. However, note that if such a map g is considered to intrinsically specify its codomain B, then g alone gives as much information as the ordered pair (B, g). Hence Lang, and many other authors, parsimoniously define the *objects* of their auxiliary category simply to be these maps g. Lang then forgets to mention that he is redefining the ''free group on S'' to mean the object of this category which we would call $(F_{\mathrm{ab}}(S), f)$, and which he simply writes f. Once this redefinition is noted, what he says is correct.

In fact, he will not be consistent – he will sometimes speak of the free abelian group as a map (as above), sometimes as a group given with such a map, and sometimes simply as a group. My own advice is to avoid the ''just a map'' definition, and speak of free abelian groups either (precisely) as certain ordered pairs (A, f), or (loosely) as certain groups A.

Part of the difficulty is more than just terminological. Though the basic concept of a ''free abelian group'' is that of an abelian group given with a map of a specified set into it and having an appropriate universal property (equivalently, an abelian group in which a specified family of elements forms a basis), we are sometimes interested instead in a group which merely has the property that there *exists* a map of some (unspecified) set into it with that property; i.e., which has *some* basis. For instance, when we say that ''a subgroup of a free abelian groups is free'', we are in the later context. So we not only have three closely related *formalizations* to choose between (object, object-with-morphism, and morphism alone), but also two closely related *concepts* (abelian groups free on specified bases, and abelian groups free on unspecified bases). However, if you keep the possible formalizations and concepts in mind, you should generally be able to see which of these an author is using, and hence what he or she means. Whenever the *universal property* of free abelian groups is being discussed, the concept in question is the one involving a specified basis.

In the immediately following passage in Lang, on products and coproducts, you will encounter a similar situation. The *product* P of two objects A and B (to take the case you have some familiarity with) will be defined to be an object given with morphisms f, g to A and B respectively, called ''projections'', such that P is universal among all objects C given with two morphisms $\varphi: C \to A$ and $\psi: C \to B$. Lang will characterize the product as the universal attracting object (P, f, g) in the category of all such 3-tuples (C, φ, ψ). (Following the same principle as for free abelian groups, he could take the objects of this category simply to be the pairs (φ, ψ). Mercifully, he does not.) But he subsequently vacillates, sometimes using the term ''product'' for the 3-tuple (P, f, g) and other times for the object P.

These difficulties are not a specialty of category theory. As we noted in our comments on §I.1, when we speak of ''a monoid G'' we should, more accurately, specify its operation, and write ''the monoid (G, μ)''. But since category-theory is quite new to most students at this level, these variations in preciseness are particularly noticeable.

As a much more petty distinction, note that in setting up a notation for the entity given by the product object and its projection maps, one has a choice between using 3-tuples (P, f, g) and ordered pairs $(P, (f, g))$. Why might one use the latter? To make it equally convenient to define the product P of an arbitrary family of objects $(A_i)_{i \in I}$, for which the natural notation is $(P, (f_i)_{i \in I})$. There is no absolute answer as to which notation should be used; but you should be aware that Lang uses both forms.

Incidentally, what Lang calls ''universal repelling'' and ''universal attracting'' objects in categories are usually called ''initial'' objects and ''terminal'' (or ''final'') objects respectively.

P.59, first **Example**, second line [=]: By the ''cartesian product'', Lang means the ordinary set-theoretic product, i.e., the set of all I-tuples (a_i) with $a_i \in A_i$. (It is named after René Des**cartes**, who introduced the idea that points of the plane could be described by ordered pairs of real numbers.)

P.59, next paragraph [~]: Lang says that the symbols $A \times B$ and $\prod A_i$ are used to denote products in a general category $\mathcal{C}$. This is correct if we understand ''product'' to mean the universal *object* of $\mathcal{C}$; if we mean the universal *tuple*, then these symbols denote its first component, and the whole tuple will have the form $(A\times B, f, g)$ or $(\prod A_i, (f_i))$. (I have here copied the symbols Lang is using for the projection maps; these vary from author to author; p_A, p_B, p_i $(i\in I)$ etc., or π_A etc. are frequently used.)

The same comments apply to the symbol $\coprod$ for ''coproduct'', introduced on the next page. (Convenient symbols for the coprojection maps associated with a coproduct are q_A, q_B, q_i etc..)

P.60, last **Example** [<]: Ignore this unless you have seen tensor products of rings. (These are developed in §XVI.6 of Lang; though they could easily be introduced in Chapter II.)

P.61, three lines following the triangular diagram [=]: Replace these lines by: ''The **product** in $\mathcal{C}_Z$ of two objects $f: X\to Z$ and $g: Y\to Z$ is called the **fibered product** of X and Y (or of f and g) over Z, and is denoted $X\times_Z Y$. More precisely, this term describes the 3-tuple $(X\times_Z Y, p_1, p_2)$, where p_1 and p_2 are the canonical maps to X and Y. This 3-tuple is universal for making a commuting diagram:''

Incidentally, why the term ''fibered''? Given a set X *over* (i.e., with a map into) a set Z, the inverse image in X of each element $z\in Z$ is called a ''fiber'' of the map $X\to Z$. (To see why, draw a picture in which Z is shown as a line, and X as a ''glob'' above it, projected ''down'' to it by the given map; and sketch the inverse images of a few points.) The ''fibered product'' P of two sets X and Y over Z is so called because for each $z\in Z$, the fiber of P at z is the direct product of the fibers of X and Y at z.

The term ''fibered'' is sometimes extended, as Lang will do on the next page, to the dual case of coproducts, though there is no intrinsic motivation for it there. (A name more often used in that case is ''the coproduct of X and Y with amalgamation of Z''.)

P.61, after the first diamond-shaped diagram [<]: The next 8 lines (starting with the italicized line ''*Fibered products and coproducts ...*'') should go on next page, just before the word **Functors**. One should also say what map to Z is used to make this subgroup an object of $\mathcal{C}_Z$. It is the map taking each pair (x, y) satisfying the indicated equation to the common value in Z of the two sides of that equation.

P.62, **Example** just before the heading **Functors** [=]: After ''their product'' add ''in $\mathbb{S}^Z$''.

P.62, definition of functor [>]: If we think of a category as a type of algebraic object, then *functor* is the name for a *homomorphism* among such objects. On the other hand, if we think of the typical category as ''a class of mathematical objects, and the morphisms among them'', then we find that functors generally correspond to *constructions* that build one sort of mathematical object from another. Indeed, in most such mathematical constructions, it turns out not only that for every object X of one sort one gets an object $F(X)$ of the other sort, but also that for every morphism $a: X\to Y$ of objects of the first sort one gets a morphism of constructed objects, $F(a): F(X)\to F(Y)$ (or sometimes $F(Y)\to F(X)$). Although these constructed *morphisms* may not be noticed at first, they eventually turn out to play as important a role in the theory as the constructed *objects*.

For instance, you have seen the construction of the free abelian group on a set. You should now verify that given any set-map $a: X\to Y$, there is a natural way to construct a group homomorphism $F_{\mathrm{ab}}(a): F_{\mathrm{ab}}(X)\to F_{\mathrm{ab}}(Y)$. Similarly, given a homomorphism $G\to H$ of abelian monoids, there is a natural way to get a map $K(G)\to K(H)$ of their Grothendieck groups. In each case, this can be verified from the construction, or as an easy application of the universal property. In each case, also, it is not hard to show that these constructions respect identity maps and composition of maps.

The class of functors which Lang calls ''stripping functors'' are called by essentially all other authors ''forgetful functors'', and we shall use the latter term here.

There is a trivial but important class of functors which Lang does not mention. Any category $\mathcal{C}$ has an *identity functor* $\mathrm{Id}_{\mathcal{C}}: \mathcal{C}\to\mathcal{C}$, which you should have no trouble defining precisely.

Incidentally, Lang defines a functor as a ''rule'' rather than a function for the same reason he defines a category as a ''collection'' rather than a set. Cf. note on *Category theory and set theory* above.

P.63, starting with second example [<]: The remaining examples on this page refer mostly to nonalgebraic material, and so should be skipped by students not familiar with it. (The second example refers to a construction defined in §I.10, and so should likewise be skipped if that section was not covered.) I suggest resuming the reading of Lang with the second example on p.64.

P.64, middle [>]: The standard symbols for the functors that Lang denotes M_A and M^B are h_A and h^B, and we shall use these here. The objects A and B respectively are called the *representing objects* for these functors.

P.65, middle [>]: I advise skipping the remainder of this page. (But if you read it, note that in addition to Grothendieck's work, a very important contribution to the study of what Lang calls ''representation functions'' was Freyd's paper *Algebra-valued functors in general and tensor products in particular*, Colloquium Mathematicum (Wrocław) **14** (1966) 89-106.)

Below are further remarks on the material of §I.11.

Further notes on functors.

I have suggested two ways of thinking about functors: as ''homomorphisms'' between categories, and as abstractions of what generally happens when we have a way of constructing one sort of object from another. Here is a third important viewpoint (or perhaps a refinement of the first): A functor from a category $\mathcal{A}$ to a category $\mathcal{B}$ can be thought of as a system of objects of $\mathcal{B}$, parametrized by the set of objects of $\mathcal{A}$, and tied together by a system of morphisms, likewise parametrized by the morphisms of $\mathcal{A}$, which satisfy composition relations corresponding to those satisfied in $\mathcal{A}$. A *morphism of functors* can then be thought of as a way of setting up morphisms from the *objects* of one of these ''systems'' to the corresponding objects of the other, which respects the connecting morphisms. Note that the existence of a concept of morphism of functors $\mathcal{A} \to \mathcal{B}$ makes the set of functors $\mathcal{A} \to \mathcal{B}$ itself into a category, often denoted $\mathcal{B}^{\mathcal{A}}$. An easy illustration is given in Exercise I.11:3. We will note more examples in our comments on the next section.

Lang has only indicated one class of examples of *contravariant* functors so far, the functors he calls M^B (p.64), and which we will write more standardly as h^B, defined by $h^B(A) = \mathrm{Mor}(A, B)$; but this is an extremely important class. You should draw diagrams to convince yourself that given a fixed object B and a morphism $X \to Y$, one does indeed get a morphism $h^B(Y) \to h^B(X)$, and that this construction respects composition and identity morphisms. Some examples: If $\mathcal{A}$ is the category of vector spaces over a field k, and B is the 1-dimensional space k, then h^B is the ''dual vector space'' construction. The behavior of this construction with respect to linear maps probably constitutes the first example of contravariance most students see, though, of course, they do not see it under that name. Here the hom-sets can be given structures not only of sets, but of vector spaces, giving us a functor from vector spaces to vector spaces. (The circumstances under which the sets arising as values of a representable functor, covariant or contravariant, acquire algebraic structure, are something that we won't be able to cover here. It is treated in §§9.1-9.3 of the 245 notes. However, we will note in passing some further cases in the next few examples below.) If $\mathcal{A}$ is the category of abelian groups of exponent n, then the contravariant functor h^{Z_n}, from this category to **Set**, can likewise be made a functor to abelian groups of exponent n; this is the duality developed by Lang in §I.9. If $\mathcal{A} = \mathbf{Set}$ and B is the object $2 = \{0, 1\}$, note that each object $h^B(X)$ can be identified with the set of all subsets of X, by the correspondence between subsets of X and their characteristic functions; this is the source of the notation 2^X for the set of all subsets of X. If $f\colon X \to Y$ is any map, then the induced map $2^Y \to 2^X$ is the ''inverse image under f'' construction. (The ''set of all subsets'' construction can also be made a *covariant* functor, by using images rather than inverse images. Both functors are important, but the contravariant functor has better properties, and is the more important.) We remark that this construction be made a functor to *Boolean rings*.

For an example of a *covariant* representable functor, let us fix any integer n, and consider the construction associating to every group G the set of its elements of exponent n, which we might write $E_n(G) = \{x \in G \mid x^n = e\}$. It is easy to see that a group homomorphism $G_1 \to G_2$ will carry elements of exponent n in G_1 to elements of exponent n in G_2, making E_n a functor **Group** $\to$ **Set**. I claim that this functor is in fact *isomorphic* to h_{Z_n}. Intuitively, the idea is that to determine a homomorphism from Z_n to a group G, one simply has to specify where the generator $[1] \in Z_n$ is to be sent, and the possible choices are precisely the elements of exponent n, so the set of homomorphisms $Z_n \to G$ is in natural bijective correspondence with the set of such elements. To formalize this, we define a *morphism of functors* $h_{Z_n} \to E_n$ taking each $f \in h_{Z_n}(G)$ to $f([1]) \in E_n(G)$, a morphism the other way taking each $x \in E_n(G)$ to the group homomorphism $[r] \mapsto x^r$, and observe that these are inverse to one another, hence constitute an *isomorphism* of functors. This says that the functor E_n: **Group** $\to$ **Set** is *representable* (Lang, p.65), with representing object Z_n.

As another example, if X is a set, and we define a functor S_X on the category **Ab** of abelian groups taking each group A to the set of all X-tuples of elements of A, then we find that S_X is representable, with representing object the free abelian group on X. Indeed, this is equivalent to the universal property of that group! Finally, suppose L is an abelian monoid, and we let N_L denote the construction associating to each abelian group A the set of monoid homomorphisms $L \to A$. It might seem that we should call this functor **Ab** $\to$ **Set** ''h_L'', but this is not quite correct, because L is not an object of **Ab**. However, there *is* an object of **Ab** which represents this functor – the Grothendieck group $K(L)$, again, by inspection from its universal property.

The above observations are obviously somehow related to the characterizations of Z_n, of free abelian groups, and of Grothendieck groups in terms of initial (''universal repelling'') objects: the group Z_n, with a cyclic generator specified, constitutes the initial object in the category of groups given with a distinguished element of exponent n, the free abelian group on X is initial in the category of abelian groups with an X-tuple of distinguished elements, and $K(L)$ is initial in the category of abelian groups given with monoid homomorphisms of L into them. In the Math 245 notes, the relationship between these viewpoints is developed in §§7.1-7.2. Here, let it suffice that you see that each of these examples illustrates both concepts.

Remark on the category of categories (optional). Though an ordinary category involves two sorts of entities, ''objects'' and ''morphisms'', when one studies the category of *categories*, one has three levels: categories, functors, and morphisms of functors. One can write down various laws relating composition of functors with composition of morphisms between functors, and abstract the resulting structure on the category of categories. Systems with such structure are studied under the name *2-categories*, but we will not look at them in this course. Cf. Mac Lane's *Categories for the Working Mathematician* (or for a brief discussion, §§6.10-6.11 of the 245 notes).

Notes on varying conventions.

An author writing about categories has to make certain choices of convention. Lang has chosen to write morphism-sets as $\mathrm{Mor}(X, Y)$; a more common notation is $\mathrm{Hom}(X, Y)$ (even though one says ''morphism'' rather than ''homomorphism''). When one wants to specify the category $\mathcal{C}$ unambiguously, one writes this morphism-set $\mathrm{Hom}_{\mathcal{C}}(X, Y)$; a simpler notation, also common, is $\mathcal{C}(X, Y)$. The roles of X and Y in these symbols are also occasionally reversed, i.e., some authors use one or another of these symbols to denote the set of morphisms from Y to X rather than the set of morphisms from X to Y! Independent of this choice is the choice of whether to write gf (as we do in this course) or fg for the composite of a morphism $f: X \to Y$ and a morphism $g: Y \to Z$.

We have also noted that developments differ in whether they assume the axiom **CAT 1** (morphism-sets are disjoint).

In the above matters I shall follow the choices Lang has made. On the other hand, as mentioned earlier, I will not follow Lang where he has introduced various pieces of very nonstandard terminology and notation; thus, what he calls ''universal repelling'' and ''universal attracting'' objects of a category we shall call ''initial'' and ''terminal'' objects, what he calls ''stripping'' functors we shall call ''forgetful functors'', and the covariant and contravariant functions that he writes M_A and M^A we will denote h_A and h^A. I have not seen his term ''representation functors'' used for these; they are generally called the covariant and contravariant *hom-functors* determined by the object A. (However, the term ''*representable* functor'' for a functor *isomorphic* to one of these is standard.)

Re §I.12. Free groups.

In this section, Lang introduces the construction of the free group, using the language of category theory. My own preference is to first present this construction as a solution to a problem posed in purely group-theoretic terms, then to develop a number of similar constructions, and finally to introduce the concept of category, showing how that allows us to unify what we have been doing. Since Lang is our text, we will follow his order. Students interested in understanding better the concept and construction of free group will find a many-sided motivated development in Chapter 2 of my Math 245 notes (for which the prerequisite is the short Chapter 1). Chapter 3 of those notes treats a series of other universal constructions, which can be discussed more briefly with the ideas from Chapter 2 as background. In particular, the first eleven sections of that Chapter concern constructions involving only groups and (occasionally) monoids.

P.67, third from last line [=]: Before ''We let F'' add ''Now given $g,\ F_0$ as in the preceding paragraph''.

P.68, middle [>]: Lang has constructed the free group, but given us no idea what it looks like! We can ''understand'' it, in the sense that we can see from the universal property that in the free group on n generators $x_1, \ldots, x_n$, the only equations satisfied by these n elements are the equations that are satisfied by *any* n elements in *any* group. (Note that the analogous statement is true of the free *abelian* group on n generators, with ''any abelian group'' in place of ''any group''.) We will, however, be able to deduce a very explicit description of the free group from results Lang will prove on coproducts of groups later in this section.

P.68, last paragraph [<]:

Presentations of groups by generators and relations.

Suppose G is a group, S a generating family for G, and T a set of *relations* satisfied by the elements of S in G – that is, equations $u = v$, where u and v are group-theoretic expressions in an S-tuple of variables, which are satisfied by the given family of generators of G. Then we say that G is the group *presented by the generators* S *and relations* T if the relations constituting T imply *all* relations that hold among the elements of the generating set S. This is equivalent to saying that G is the initial object (''universal repelling object'') in the category of groups given with S-tuples of elements satisfying the system of relations T. For example, our characterization of the group Z_n in our ''*Further notes on functors*'' above is equivalent to saying that it can be presented by one generator x and one relation, $x^n = e$.

In fact, for every set S, and family of equations T in an S-tuple of group-theoretic variables, there exists a group with the above property. We may see this as follows. Form the free group $F(S)$. For each relation ''$u = v$'' in T, evaluate the expression uv^{-1} in $F(S)$ using the universal S-tuple of elements of $F(S)$, and let $R \subseteq F(S)$ denote the set of elements (''relators'') obtained in this way from the relations comprising T. Let $N \lhd F(S)$ be the normal subgroup generated by R. Then $F(S)/N$, with the map of S into it induced by the canonical map $S \to F(S)$, will be the desired object. You should find its universal property straightforward to prove using the universal property of the free group, and the universal property of a factor group. This group, which as noted above is called the group ''presented by generators S and relations T'', is often written $\langle S \mid T\rangle$ or $\langle S \mid R\rangle$ by group-theorists. (E.g., the group presented by generators x, y and the one relation $x^2 = y^3$ is sometimes written $\langle x, y \mid x^2 = y^3\rangle$ and sometimes $\langle x, y \mid x^2 y^{-3}\rangle$.)

Note that to show that a given group G has presentation $\langle S \mid T\rangle$, one must verify not only that the generators S satisfy the relations T in G, but that the relations T imply *all* the relations satisfied by S in G. We do not have time in this course to examine how this is done in various situations (it can depend on the sense in which G is ''given''), nor the reverse problem of obtaining concrete descriptions of groups presented by generators and relations. Exercise I.12:2 gives a little practice; you might also turn back to p.9 in Lang, and see whether you can show that the two sets of equations displayed near the bottom of the page are indeed the relations in presentations of the two nonabelian groups of order 8. (In the second system of equations, note that k and m are abbreviations for ij and i^2, as indicated on the preceding line.) The student interested in learning more can turn to any of the various books whose titles are, approximately or exactly, *Combinatorial Group Theory,* the best known being the book of that title by Magnus, Karrass and Solitar. There is also a lively book, *Presentations of Groups,* by D. L. Johnson (London Mathematical Society Student Texts, vol.15, 1990) specifically on this topic, aimed at advanced undergraduates and beginning graduate students.

It is important to note the difference between relations and *identities*. E.g., the group with presentation $\langle a, b \mid a^3 = e,\ b^3 = e \rangle$ does not satisfy the identity $x^3 = e$: in order for a group generated by elements a and b to satisfy this identity, not only must the cubes of those generators equal e, but also the cubes of all other elements, e.g., $(ab)^3$, $(a^2b^2a^{-1})^3$, etc., and the two relations $a^3 = e$ and $b^3 = e$ do not imply these others, as may be seen by considering the elements $a = (123)$ and $b = (234)$ of A_4. (On the other hand, the nonabelian group of order 27 not containing a cyclic subgroup of order 9 found in exercise I.6:2 does satisfy this identity.)

Lang's terminology and notation differ in some small ways from what I have described: He writes ''determined by generators and relations'' in place of the standard ''presented by generators and relations''. He also starts off (on p.68) only considering relations of the form $r = e$, and calls the elements r the ''relations''. (The standard term is ''relators''.) Then, on the next page, he passes to relations written as equations, as discussed above.

P.69, **Example** in the middle of the page [<]: I recommend skipping everything from this example through the short paragraph **Further Examples** on p.70. Resume reading with Proposition I.12.3 on that page.

If you do read the examples mentioned, note that the ''exercise'' referred to on the line following the first display is not trivial; cf. Exercises I.12:7 and I.12:8 in this Companion. And on the next page, two lines above the first display, ''*the free group with ...*'' should be ''*the group presented by ...*''. (If we think of a free group on a set X as a group generated by an X-tuple of elements that satisfy no relations other than those implied by the group identities, and a group presented by generators and relations X and R as a group generated by an X-tuple of elements that satisfy no relations other than those implied by the relations in R and the group identities, we can understand why the latter was sometimes called ''the free group with those generators and relations''. But ''free'' now has the specific meaning defined at the bottom of p.66, and should only be used in that sense.)

P.71, Example, ending just before Proposition I.12.4 on the next page [<]: I suggest skipping this. (It is a continuation of one of the examples on p.69 that I suggested you skip.)

P.73, proof of Proposition I.12.5 (continued on next page) [~]: Lang gives the ''natural'' but messy proof of this result. There is a much neater one, due to van der Waerden, based on remembering that the group concept arises from looking at the properties of *permutations of sets*. Given two groups A and B, let us follow Lang in forming the set of sequences $(a_1, \ldots, a_n)$ (including the empty sequence), such that

(c14) All a_i lie in $(A - \{1\}) \cup (B - \{1\})$, and no two successive terms a_i, a_{i+1} lie in the *same* group A or B.

But rather than calling this set $A \circ B$ and putting a group structure on it, let us call it X, and construct a group of *permutations* of X.

This group will be generated by the homomorphic images of A and B corresponding to an action of each of these groups on X. The goal is that $(a_1, \ldots, a_n) \in X$ should be the sequence arising by starting with the empty sequence $() \in X$, applying a_n to this, then a_{n-1} to the result, etc.. This tells us how we ought to define certain cases of the actions of A and B on X, namely

(i) If $(a_1, \ldots, a_n)$ is an element of X, and $a \in A \cup B$ is a nonidentity element, and does not come from the same group A or B from which a_1 comes (which we take to be satisfied trivially if $n=0$), we define $a(a_1, \ldots, a_n)$ to be $(a, a_1, \ldots, a_n)$, which we see is indeed an element of X.

There are three other possible relations between elements $(a_1, \ldots, a_n) \in X$ and $a \in A \cup B$:

(ii) If $a = 1$, we of course define $a(a_1, \ldots, a_n) = (a_1, \ldots, a_n)$.

(iii) If $a \neq 1$ and this belongs to the same group as a_1, we define $a(a_1, \ldots, a_n)$ to be $(a\,a_1, \ldots, a_n)$ if this is permitted, i.e., if $a\,a_1 \neq 1$.

Finally

(iv) If $a \neq 1$ and belongs to the same group as a_1, and if $a\,a_1 = 1$, then we let $a(a_1, \ldots, a_n) = (a_2, \ldots, a_n)$ (understood to mean $()$ if $n=1$).

Restricting attention to the action of elements $a \in A$, we find that the map $A \times X \to X$ given by the above ''multiplication'' is indeed a group action. (There are several cases to consider in verifying that $a(a'x) = (aa')x$, but each is a quick computation.) Similarly, the map $B \times X \to X$ is an action of B. We define $A \circ B$ to be the group of permutations of X generated by the images of these two actions.

Now if we write $\bar{a}$ for the image of $a \in A \cup B$ in $\mathrm{Perm}(X)$, then, simply from the fact that the group $A \circ B$ is generated by the images of A and B under certain homomorphisms $a \mapsto \bar{a}$, it is not hard to verify that every element of $A \circ B$ can be reduced to an expression $\bar{a}_1 \ldots \bar{a}_n$ satisfying (c14). (Lang speaks of elements $\neq 1$ as having this form; I also consider ''1'' the case of this expression where $n = 0$.) Moreover, given such an element $\bar{a}_1 \ldots \bar{a}_n \in A \circ B$, it is easy to verify by induction that on applying this element to the empty sequence $() \in X$, we get the sequence $(a_1, \ldots, a_n) \in X$. (The assumption (c14) guarantees that at each step of the induction we will be in case (i) above.) Hence a product $\bar{a}_1 \ldots \bar{a}_n \in \mathrm{Perm}(X)$ satisfying (c14) uniquely determines its sequence of factors, so permutations with distinct expressions of this form must be distinct. We may finish the proof, as Lang does, by observing that ''identifying A and B with their images in $A \circ B$, we obtain a proof of the proposition'' (though we should discuss in class how to make that argument precise).

Although the above proof involves less messy computation than the one in Lang, there is still a step that is messier than it should be: verifying that the maps $A \times X \to X$ and $B \times X \to X$ described are really group actions, since we have to consider several cases, depending on whether our element of X begins with an element of A or an element of B, and whether the various products we get are 1 or not. But there is even a neat way around this: it involves setting up a little more machinery, but saves us work in the end. The trick is to form two bijective of copies of X, which we shall call X_A and X_B, one convenient for describing the action of A and the other for describing the action of B. Namely, given $x \in X$, define x_A to be the same element x if this begins with a member of A, while if $x = (a_1, \ldots, a_n)$ where either $a_1 \in B$ or $n=0$, we define $x_A = (1, a_1, \ldots, a_n)$. Thus every element of X_A begins with an element of A, and we can define the action of A on this set by in *all* cases letting $a(a_1, \ldots, a_n) = (a\,a_1, \ldots, a_n)$. The desired condition $a(a'x_A) = (aa')x_A$ is now immediate from the associativity of A! We define X_B and the action of B on this set in the exactly analogous manner. Because we have bijections $X_A \longleftrightarrow X \longleftrightarrow X_B$, these actions induce actions of A and B on the set X. The descriptions of these actions on X are easily seen to be those that we began with; the advantage of this approach is that the tedious verification that they *are* actions has been circumvented. This gives the most elegant proof of the proposition.

We have described the construction of a group $A \circ B$ from *two* groups, A and B, only for convenience. There is a minor way in which the two-group case is special: in any sequence of elements satisfying (c14), elements of $A - \{1\}$ and of $B - \{1\}$ occur alternately. If we had stated (c14) in terms of this alternation, we would have to modify this statement to get the general result; but since we did not, we can say that exactly the same argument shows that given any family $(A_i)_{i\in I}$ of groups whose pairwise intersections are $\{1\}$, there exists a group $\circ_I A_i$ containing all the A_i, such that all products $a_1 \ldots a_n$ satisfying (c14) are distinct.

P.74, Corollary I.12.6 [~]: Note that merely because the group $F(S)$ is generated by the elements of S, every element of this group can be written in the form $x_{i_1}^{\nu_1} \ldots x_{i_r}^{\nu_r}$ as described, for some $x_1, \ldots, x_n \in S$ and some family of exponents. This corollary tells us that every element of $F(S)$ has a *unique* expression in this form. This is the explicit description of the free group promised earlier; it allows us to compute easily in such groups.

P.74, Corollary I.12.8 [~]: Here Lang finally gets to the point of the construction $G_1 \circ \ldots \circ G_n$. In fact, that notation of his was purely temporary; the group in question is written $G_1 * \ldots * G_n$ by most group theorists and called their ''free product'', while workers with a more category-theoretic orientation call it the coproduct of groups.

The notation $\coprod X_i$ for the coproduct of an arbitrary family of objects of a category is standard, but there is no standard notation for the coproduct of two or more ''listed'' objects. I agree with Lang that $X_1 \amalg \ldots \amalg X_n$ is a good choice.

In stating this corollary, Lang refers to arbitrary groups $G_1, \ldots, G_n$ forgetting to mention the condition of intersecting only in 1 that he had in Proposition I.12.5. Actually, this condition fails to hold in many situations of interest, so let us note how one gets around it. Given *any* family of groups (which may even include repetitions of the same group), observe that one can achieve this disjointness condition in appropriate *isomorphic copies* G_i' of these groups. Then the same argument given in Lang shows that the map $G_1 \amalg \ldots \amalg G_n \to G_1' \circ \ldots \circ G_n'$ is an isomorphism; equivalently, that $G_1' \circ \ldots \circ G_n'$ with the appropriate maps of the G_i's into it is a coproduct of the G_i's. (And the same is true for coproducts $\coprod_I A_i$ of not necessarily finite families of groups.)

P.74, bottom [>]: Here are some further notes on topics in the above section.

General heuristics for constructing universal objects.

A category for which one wants to find an *initial* object will typically consist of mathematical objects of some standard sort, together with certain maps of fixed objects into them having particular properties. (Examples: Groups with maps of a fixed set X into them; abelian groups with homomorphisms of a fixed nonabelian group G into them; abelian groups with monoid homomorphisms of a fixed abelian monoid M into them; arbitrary groups with homomorphisms of two fixed groups A and B into them.) Similarly, a category for which one wants to find a terminal object typically consists of objects of a certain kind, together with maps of those objects *into* certain fixed objects, having specified properties. (Examples: Groups or sets with homomorphisms into two specified groups or sets A and B; groups with monoid homomorphisms into a fixed monoid M.) In each case, it is useful to begin the search for a universal object by thinking about a general (rather than a universal) object U with the indicated extra maps, and asking, ''What handle can we get on the structure of U in terms of these maps?'' In the first type of case, this ''handle'' typically consists of specifications of certain elements of U (namely, images of elements under the given maps, and other elements obtained from these by the operations), and relations among these. If one constructs abstractly an object consisting of just such a family of elements, subject only to the relations that must hold in every case, one gets, in favorable cases, a *universal* object with the indicated type of structure, corresponding to an initial object of the category in question. (For instance, the expressions one can form from an S-tuple of elements of a general group are precisely all expressions $x_{i_1}^{\nu_1} \ldots x_{i_r}^{\nu_r}$ in finite families $x_1, \ldots, x_n$ of such elements, and each such expression can be *reduced* to one in which successive elements x_i, x_{i+1} are distinct. The set of all such reduced expressions, with multiplication defined in the natural way, is the description of the *free group* $F(S)$ given in

Corollary I.12.6.) In the second case, the type of ''handle'' one can generally get is a *classification* of elements of U based on their behavior under the maps. Here it is the set of ''classes'' in this classification that turns out, in favorable circumstances, to form a universal object, i.e., a terminal object of the auxiliary category. (For instance, given an ''unknown'' group K with homomorphisms α, β into ''known'' groups G and H, we may ''classify'' elements $k \in K$ by the pairs of elements $\alpha(k) \in G$ and $\beta(k) \in H$. Given two such pairs $(\alpha(k), \beta(k))$ and $(\alpha(k'), \beta(k'))$, it is easy to write down the formula for the pair associated with kk', and this formula can be used to put a group operation on the set of all pairs consisting of an element of G and an element of H. The resulting group, $G \times H$, is easily shown to be the terminal object in the category of groups given with maps into G and H.) It is perhaps best to think about these abstract heuristics *after* one has digested a sufficient number of particular cases. How many are ''a sufficient number'' will depend on the student.

Notes on free groups and coproducts.

Is every subgroup of a free group free? Yes, but the proof is intricate (in contrast to the proof for free abelian groups), and we will not give it here. It can be found in the book of Magnus, Karrass and Solitar mentioned above. (Curiously, one useful way to visualize the result is topological. For the student with some familiarity with these matters: One shows that the fundamental group of any 1-complex is a free group, and that every free group occurs as such a fundamental group; one proves that subgroups of fundamental groups correspond to fundamental groups of covering spaces; and finally, one notes that a covering space of a 1-complex is a 1-complex.) Exercise I.12:1 shows that a free group on 2 generators can have a subgroup which is free on infinitely many generators, which is very different from the situation for free abelian groups (but again easily picturable in terms of covering spaces).

Why does Lang give an abstract proof of the existence of coproducts of groups in Proposition I.12.3 when he will later construct them explicitly? The abstract proof is extremely versatile; once you have digested it (or another such general proof), you can see instantly, on encountering any of a wide range of sorts of categories, that these have coproducts of arbitrary families of objects. The *explicit* description is particular to groups. When one encounters a new sort of algebraic object, and realizes by general arguments that these have coproducts, the next step is to ask whether there is, similarly, some elegant description of the *structures* of such coproducts.

Products and coproducts of groups presented by generators and relations.

Given two groups

$$G = \langle X \mid R \rangle, \qquad H = \langle Y \mid S \rangle,$$

with X and Y disjoint, it is straightforward to show by universal properties that the coproduct of G and H has the presentation $\langle X \cup Y \mid R \cup S \rangle$. The (direct) product, on the other hand, may be presented as $\langle X \cup Y \mid R \cup S \cup C(X, Y) \rangle$, where $C(X, Y)$ denotes the set of relations $xy = yx$ $(x \in X,\ y \in Y)$.

Some morphisms of functors.

As I mentioned in my comments to p.62 of Lang (and as Lang notes on p.63), one example of a functor $\mathbf{Set} \to \mathbf{Ab}$ is the construction associating to every set S the free abelian group $\mathbf{Z}S$. (Elsewhere, the group $\mathbf{Z}S$ has been characterized either as initial in a certain auxiliary category, or as the representing object for the functor $\mathbf{Ab} \to \mathbf{Set}$ taking an abelian group A to the set of all S-tuples of elements of A. The way in which this family of objects that were constructed separately as representing a family of functors, equivalently, as initial in a family of related auxiliary categories, can be ''tied together'' into one functor, is a case of a concept called *adjoint functors*, which I will sketch later in my comments on §II.3. However, without knowing anything about that general concept, you should not find it hard, using the universal properties of free abelian groups, to construct the morphisms among these groups needed to make the free abelian group construction a functor.) Similarly, the constructions of the *free group* on a set, and the *abelianization* G/G^{c} of a general group G, can each be made into a functor in a natural way. Let us give ad hoc names to these and one other functor:

F: **Set** $\to$ **Group** for the free group functor,

F_{ab}: **Set** $\to$ **Ab** for the free abelian group functor,

A: **Group** $\to$ **Ab** for the abelianization functor, and

I: **Ab** $\to$ **Group** for the inclusion functor (defined in the obvious way).

With this list of functors to work with, I can now give some more examples of morphisms and isomorphisms of functors. Namely, suppose we associate to each set X the unique homomorphism $h(X) : F(X) \to F_{\mathrm{ab}}(X)$ sending the canonical image in $F(X)$ of every $x \in X$ to the image of the same element in $F_{\mathrm{ab}}(X)$. It is easy to verify that these maps together comprise a morphism of functors $f : F \to I F_{\mathrm{ab}}$. You should likewise not find it hard to show that for every set X, there is a canonical isomorphism $i(X) : AF(X) \cong F_{\mathrm{ab}}(X)$, and that these constitute an isomorphism of functors $i : AF \cong F_{\mathrm{ab}}$. If you'd like two more examples to think about, you should be able to find an isomorphism between the composite AI and the identity functor of the category **Ab**, and a certain natural morphism from the identity functor of **Group** to the reverse composite, IA.