Chapter II. Rings.

Re §II.1. Rings and homomorphisms.

P.83, definition of a ring [>]:

Just as the fundamental examples of groups are groups of permutations of sets, so there is a fundamental class of examples of rings: the rings of endomorphisms of abelian groups.

Indeed, recall that one can add and subtract members of $\mathrm{Hom}(A, B)$ for any abelian groups A and B, making this hom-set an abelian group. Taking $A = B$, this gives us an abelian group structure on $\mathrm{End}(A) = \mathrm{Hom}(A, A)$. Moreover, one can compose such endomorphisms, getting a ''multiplication'' on this set, for which the identity endomorphism will be a neutral element; and it is straightforward to verify that these additive and multiplicative structures together make $\mathrm{End}(A)$ a ring.

Just as every group G can be represented faithfully by permutations of a set, namely, via its action by left translations on its own underlying set, so a ring R can be represented faithfully by endomorphisms of an abelian group, namely, via its action by left multiplication on its own underlying additive group.

Note that the distributive law states that multiplication is a *bilinear* map with respect to the additive group structure.

P.85, **the convolution product** [~]: This is an important class of examples, but I will postpone discussion until Lang returns to it at greater length, two sections from now (on p.104). (The background of the term ''convolution product'' will get a discussion in itself.)

Note that in the sentence containing the last display, Lang modifies his viewpoint in two ways: First, by regarding the formal sums of the preceding display as functions on G, with a corresponding change of notation; and second, by not insisting that all but finitely many $f(x)$ be zero; though as he notes in the next sentences, *some* condition is needed to make the summations finite, and the most obvious such condition is the finiteness one. (But there are others. E.g., in the exercise he refers to, if we think of the monoid of positive integers under multiplication as a submonoid of the group of positive rationals, then the condition that only members of that submonoid have nonzero coefficient makes the multiplication work.)

Here is an interesting example of a ring of endomorphisms of an abelian group. Let A be the additive group of all polynomials in one indeterminate x with real coefficients. (Though we have not yet read the section on polynomials, in this and some subsequent examples I will assume familiarity with elementary properties of polynomials over the reals.) Let $X: A \to A$ denote the endomorphism of this abelian group given by multiplication by x, let $Y: A \to A$ denote the endomorphism given by *differentiation* with respect to x, and for each real number r, let the same symbol r denote the endomorphism of scalar multiplication by r. The ring R that these operators generate is called the *Weyl algebra* over the real numbers. From the product law for differentiation, one gets the formula

(c15) $$YX - XY = 1.$$

Using this formula, one can reduce any member of R to the form $\Sigma\, r_{ij} X^i Y^j$ with real coefficients r_{ij}. These expressions look like ordinary polynomials in two variables, but the multiplication, based on (c15), is clearly different. (This algebra is important in quantum mechanics, where multiplying the wave function ψ of a particle by a coordinate function x corresponds to measuring the x-coordinate of the *position* of the particle, differentiating ψ with respect to x corresponds to measuring the x-coordinate of the particle's *momentum*, and the noncommutativity of these operators somehow corresponds to the impossibility of measuring these two quantities simultaneously!)

P.86, middle [=]: Change the sentence containing the word **principal** to ''A left ideal is called **principal** if it can be generated by one element, i.e., has the form Aa for some $a \in A$.''

Later in the same paragraph, Lang introduces the notation $(a_1, \dots, a_n)$, which he says is used for the *left* ideal generated by $a_1, \dots, a_n$. Actually, this notation is rarely used except in commutative ring theory, where left, right, and two-sided ideals are the same. Moreover, I generally prefer the more explicit notation $a_1 A + \dots + a_n A$ (except in cases where the symbol for the ring is complicated!) With that explicit notation, one can distinguish in the noncommutative case between the right ideal $a_1 A + \dots + a_n A$, the left ideal $Aa_1 + \dots + Aa_n$, and the two-sided ideal $Aa_1 A + \dots + Aa_n A$.

Same page, 5 lines from bottom [~]: Where Lang writes ''a **principal** ring'', he is translating the French term; the standard term in English, which we will use, is ''principal ideal ring''.

After the above line [>]: In the remainder of this section, Lang will a couple of times use the term ''prime ideal'', which he doesn't define till the beginning of the next section, and whose definition depends on a term defined at the end of this section. So turn now to the paragraph beginning on the bottom of p.91, and read through the first paragraph of §II.2 on p.92, where ''prime ideal'' is defined, then return to the next paragraph of this Companion... .

Having told you to read Lang's definition of ''entire ring'' on p.91, I must now say that this term is simply not used. As Lang notes, the standard term in English is ''integral domain'', and we will use it. Where Lang feels the need for an adjective, as in ''R is entire'', one says ''R is an integral domain''. ''Integral domain'' is often shortened to ''domain'', especially in longer phrases, so that what Lang calls a ''principal entire ring'', i.e., a principal ideal ring that is an integral domain, is commonly called a ''principal ideal domain''. This is an important concept, often abbreviated ''PID''.

Now return to p.86 of Lang, bottom paragraph.

P.87, third and fourth paragraphs [<]: Skip these unless you are familiar with the topics (rings of algebraic integers in number fields, and the ring theory of analytic functions on the complex plane).

P.88, top two paragraphs [<]: I recommend skipping these.

P.88, **Example** occupying two long paragraphs [<]: This is another interesting topic which would have to be treated at greater length to get anything out of it. So read it, skim it, or skip it as you like.

P.90, first line of paragraph containing first display [=]: For ''commuting with'' read ''centralizing''. (A set X is said to *centralize* a set Y if every element of X commutes with every element of Y. A statement that a set X ''commutes with'' Y is vague; it might be taken to mean $XY = YX$ as sets of products.) Note further that although the concept of adjoining to a ring A a set S of elements centralizing A is of interest in ring theory, one should definitely not, as Lang does, limit the definition of ''ring generators for B over A'' to the situation where S centralizes A! Finally, note that noncommutative ring theorists generally write the ring generated over a subring A by a set of elements S

as $A\langle S\rangle$.

P.90, line after first display [=]: Change ''n-tuples $(i_1, \ldots, i_n)$ of integers ≥ 0'' to ''strings $(i_1, \ldots, i_n)$ where n and $i_1, \ldots, i_n$ are nonnegative integers''. (In other words, n is not assumed the same in all terms of the sum.)

Since the string $s_1, \ldots, s_n$ can be different in different summands, the notation $a_{i_1 \ldots i_n}$, which only shows the exponents and not the s's, is poor. In fact, since we must allow repetitions among the s_i, there is no need for exponents. So a better form for the display would be $\Sigma\, a_{s_1, \ldots, s_n}\, s_1 \ldots s_n$.

P.90, last sentence before Example [=]: By ''may not'' he means ''need not''.

P.91, three lines from bottom, definition of ''entire ring'' [=]: See my comment above.

P.92, end of §II.1 [>]. One good feature of Lang relative to Hungerford is that he does *not* introduce rings without unit. These are of some use in ring theory, but their place is secondary. Rather than try to develop the fundamentals of the theory of rings with unit in the context of the ''more general'' theory of rings without unit, it is most convenient, when basic facts on rings without unit are needed, to use a certain trick which reduces the study of rings without unit to that of rings with unit. We will not take the time to discuss this trick here; if you need it some day, ask me!

Re §II.2. Commutative rings.

P.92, beginning of the section [=]: Observe the proviso in italics! The results of this section are *not* in general true for noncommutative rings.

P.92, italicized result, *Every maximal ideal is prime* [~]: This does not need a separate proof: it will follow immediately from the result on the next page, that if $\mathbf{m}$ is maximal, $A/\mathbf{m}$ is a field, and the observation that a field has no zero-divisors.

(By the way, the text-formatting system that I am using doesn't have fraktur type, so I will use boldface lower-case letters $\mathbf{m}$, $\mathbf{p}$, $\mathbf{a}$, $\mathbf{b}$ etc. where Lang uses lower case fraktur letters for ideals.)

P.93, proof of first italicized statement [>]: In Exercise A2.2:1 it is shown that Zorn's Lemma can be used to obtain maximal proper subgroups of *finitely generated* groups, but not in general of infinitely generated groups. Why, then, can it be used here to get maximal ideals of not necessarily finitely generated rings? Because any ring A, even if not finitely generated as a *ring*, is finitely generated as an *ideal*, namely by $\{1\}$.

P.94, first display [=]: Lang says that the ring $\mathbf{Z}/n\mathbf{Z}$ will be abbreviated $\mathbf{Z}(n)$. This is very nonstandard; it is generally either written Z_n (like the group $\mathbf{Z}/n\mathbf{Z}$), or written out in full, as $\mathbf{Z}/n\mathbf{Z}$. In fact, Lang will rarely if ever use this notation he has introduced.

P.94, proof of the Chinese Remainder Theorem [~]: The proof Lang gives is messy. The easy-to-remember proof is as follows. Let A' denote the image of A in $\prod_{i=1,\ldots,n} A/\mathbf{a}_i$. For any pair of indices $i \neq j$, the hypothesis $\mathbf{a}_i + \mathbf{a}_j = A$ means that we can write $1 = u + v$, where u is an element of A whose image in A' has component 0 in the ith position, and v an element whose image has component 0 in the jth position. Thus, the image of v in A' has 1 in the ith position and 0 in the jth. Now fixing i, and multiplying together the images in A' of the ''v''s we get in this way for each j, we get an element having 1 in the ith position, and 0 everywhere else. Multiplying by an arbitrary element of A', we get an element having an arbitrary member of $A/\mathbf{a}_i$ in the ith position and 0 everywhere else. By summing a family of such elements, one chosen for each i, we can get an element of A' with arbitrary component in each place, proving that A' is all of $\prod A/\mathbf{a}_i$, which is the content of the theorem.

P.95, top paragraph [~]: The proof of the asserted result can be called ''trivial'', but it is not obvious if you haven't seen it before. Multiply the given equation $\mathbf{a}_1 + \ldots + \mathbf{a}_n = A$ by itself $\nu_1 + \ldots + \nu_n$ times. (Actually, $(\nu_1 - 1) + \ldots + (\nu_n - 1) + 1$ times is enough.) On the left-hand side of the resulting equation, note that every term contains *either* ν_1 factors $\mathbf{a}_1$, *or* ν_2 factors $\mathbf{a}_2$, *or* ν_3 factors $\mathbf{a}_3$, etc.. Hence that left-hand side is contained in $\mathbf{a}_1^{\nu_1} + \ldots + \mathbf{a}_n^{\nu_n}$. Hence, since that left-hand side equals the right-hand side,

A, so does $\mathbf{a}_1^{\nu_1} + ... + \mathbf{a}_n^{\nu_n}$.

P.96, first two lines [~]: One can prove this result more easily by counting integers relatively prime to p^r.

P.96, statement of Theorem II.2.3 [~]: Here Lang is being sloppy (as he would say, he is ''abusing notation''), since he sometimes treats k as an element of $\mathbf{Z}/n\mathbf{Z}$ and other times as a member of $\mathbf{Z}$. I point this out because I don't want you to take such sloppiness as a model! The same applies to p.97, lines 3 and 4, and to the last line of the section.

In connection with the last line of the section, it intuitively makes sense to exponentiate an element of exponent n by a member of $\mathbf{Z}/n\mathbf{Z}$; but if you want to use this observation in a formal argument, you should explicitly state a convention, ''If x is an element of exponent n in a group G, and k is an element of $\mathbf{Z}/n\mathbf{Z}$, then we shall understand x^k to mean x^c, where $c \in \mathbf{Z}$ is any representative of the residue class k,'' and observe that this is well-defined, and that the ''laws of exponents'' hold for these generalized exponents.

It's not clear why Lang calls Theorem II.2.3 an ''application''. Perhaps he meant to apply the Chinese Remainder Theorem to express the rings $\mathbf{Z}/n\mathbf{Z}$ as direct products of rings $\mathbf{Z}/p^r\mathbf{Z}$, but forgot to go on to this point.

P.97, end of §II.2 [>].

More examples of commutative rings.

An important class of examples of commutative rings is given by factor rings of polynomial rings. Let us regard the ring $\mathbf{R}[X, Y]$ of polynomials in two variables over the real numbers as consisting of certain functions on the plane. Then if we restrict these functions to the circle $X^2 + Y^2 = 1$, we get the ring A of those real-valued functions on the circle that are given by polynomials in the two coordinate-functions X and Y. The restriction map $\mathbf{R}[X, Y] \to A$ is a surjective ring homomorphism, hence A can be identified with the factor-ring $\mathbf{R}[X, Y]/I$ for some ideal I. This ideal turns out to be the principal ideal generated by $X^2 + Y^2 - 1$.

In general, for an ideal I of $\mathbf{R}[X, Y]$, let us think of the factor-ring $\mathbf{R}[X, Y]/I$ as consisting of polynomial functions which are ''missing some information''. We shall see below that it may or may not be possible to express these, as in the above case, as restrictions of polynomial functions to subsets of the plane.

One ring-theoretic phenomenon with which we have had little experience is that of zero-divisors, and the various forms they take, so let us look at some examples of rings $\mathbf{R}[X, Y]/I$ with zero-divisors.

The ring $\mathbf{R}[X, Y]/(XY)$ can be identified with the ring of polynomial functions on the union of the X-axis and the Y-axis. (To show this, one must show that the kernel of the restriction map $\mathbf{R}[X, Y] \to$ **{**functions on the union of the axes**}** is precisely (XY). ''$\supseteq$'' is immediate, and ''$\subseteq$'' is an easy argument.) It has zero-divisors, because X and Y are nonzero elements with product 0. (Where I am writing X and Y, I mean, strictly speaking, the residue classes of these elements of $\mathbf{R}[X, Y]$ in this factor-ring. I am using the same symbols for elements of the polynomial ring and of the factor-ring for notational simplicity.) One can get another sort of description of this ring: Note that a function on the union of the axes is determined by its restrictions to each of the axes, which are a polynomial in X and a polynomial in Y respectively. It is not hard to show that a pair of polynomials $f(X)$ and $g(Y)$ can arise in this way if and only if $f(0) = g(0)$. Hence our ring can be described, up to isomorphism, as the subring of $\mathbf{R}[X] \times \mathbf{R}[Y]$ consisting of all pairs (f, g) such that $f(0) = g(0)$. (The zero-divisors noted above are $(X, 0) \cdot (0, Y) = (0, 0)$ in this description.) The use of two variable-symbols X and Y is just a holdover from the original picture; we can equally well identify our ring with the subring of $\mathbf{R}[X] \times \mathbf{R}[X]$ consisting of pairs satisfying $f(0) = g(0)$. This is, incidentally, an example of what is called a *subdirect product* (not to be confused with a *semidirect product*), meaning a subobject of a direct product, whose projection maps onto the two factors are both surjective.

We note a few more examples similar to the above. $\mathbf{R}[X, Y]/(Y(Y-1))$ can be identified with the ring of polynomial functions restricted to the union of the two parallel lines $Y = 0$ and $Y = 1$. The restrictions to those two lines can be two arbitrary polynomials in X (verify this!), so this ring is isomorphic to the full direct product $\mathbf{R}[X]\times\mathbf{R}[X]$. $\mathbf{R}[X,Y]/(Y(Y-X^2))$ can be identified with polynomial functions restricted to the union of the line $Y = 0$ and the parabola $Y = X^2$. Restricting to these curves separately, we note that any polynomial function on either can be expressed as a polynomial in the X-coordinate alone, so our ring is again a subdirect product of two copies of $\mathbf{R}[X]$; in this case, it can be shown to consist of all pairs (f, g) whose values *and* first derivatives agree at 0. As a last example of this sort, $\mathbf{R}[X,Y]/((Y-X)Y(Y+X))$, the ring of polynomial functions restricted to the union of the three lines $Y = X$, $Y = 0$ and $Y = -X$, can be identified with the ring of 3-tuples of polynomials in X, (f, g, h), such that $f(0) = g(0) = h(0)$, and $f'(0) - 2g'(0) + h'(0) = 0$. (Proving this takes some thought, but still turns out to be a straightforward computation.)

We run into a new phenomenon when we look at $\mathbf{R}[X,Y]/(X^2)$. The subset of the plane defined by the equation $X^2 = 0$ is simply the line $X = 0$. But when we map $\mathbf{R}[X,Y]$ by restriction to functions on this line, the kernel consists of all multiples of X, which is larger than the ideal (X^2). Hence the ring of functions on the line induced by polynomials on the plane is not isomorphic to the ring we want to look at.

The trouble is that when we restrict to the line, we ''forget'' too much. The image in $\mathbf{R}[X,Y]/(X^2)$ of a polynomial $f(X,Y)$ ''remembers'' not only the terms without X, but also the terms *linear* in X. These terms specify the *partial derivative* of f in the normal direction to the line $X = 0$, all along that line. Algebraic geometers picture $\mathbf{R}[X,Y]/(X^2)$ as the ring of polynomial functions on an infinitesimally ''thickened'' version of the line $X = 0$, where there is just enough ''thickening'' so that the restriction of a function to it determines the normal derivative of that function, but not higher derivatives.

That $\mathbf{R}[X,Y]/(X^2)$ has zero-divisors is shown by the relation $X^2 = 0$. In terms of our ''values and derivatives'' interpretation, this is an instance of the observation that two differentiable functions whose values are zero all along the line $X = 0$ have product whose value *and* normal derivative are zero along that line.

A variant of this example is $\mathbf{R}[X,Y]/(X^2, XY)$. An element of this ring specifies the values of a polynomial function along the line $X = 0$, and its normal derivative to that line at the point $(0,0)$ only, so it is a line ''thickened'' only at that point.

I have given examples in terms of polynomials over the real numbers because these are familiar and easy to visualize; but the relation between ideals of this ring and subsets of $\mathbf{R}^2$ has weaknesses; e.g., ideals such as (X^2+1) or (X^2+Y^2+1) cannot be characterized by their behavior on $\mathbf{R}^2$. One gets much better behavior by looking at ideals in polynomial rings over the complex numbers $\mathbf{C}$; or for that matter by looking at ideals in polynomial rings over the real numbers, rational numbers, etc., in terms of the induced functions on space $\mathbf{C}^n$.

The study of ideals in the ring of polynomials in n indeterminates (equivalently, of homomorphic images of this ring), and the subsets of $\mathbf{C}^n$ that these determine, is the starting point of algebraic geometry. Lang proves in §IX.2 (which we do not read in 250A) some results on zero-sets of polynomials, and in §IX.5 he introduces the abstract object, the ''prime spectrum'' of a ring, which modern algebraic geometers study in place of these polynomially defined sets. But he does not make explicit the relationship between these spectra and the classical polynomially defined subsets, nor does any standard introduction to the subject that I have seen. I discussed that relationship in an expository article, *The Zariski topology and generalizations*, which appeared in a somewhat obscure volume (*Topology and its Applications*, the proceedings of a symposium in Budva, Jugoslavia, August 1972, published in Beograd in 1973). I can give anyone interested a reprint. (The article assumes a knowledge of elementary point-set topology.)

Re §II.3. Polynomials and group rings.

Before reading this section of Lang, I recommend reading sections III.1-III.4, and the notes on those sections in this Companion. I will often assume familiarity with that material below.

P.99, first line [=]: Lang says the elements of the free generating set of a polynomial ring will be called ''variables'' (and makes a similar remark on p.104). This is an old-fashioned and misleading term. Although the concept of polynomial ring is motivated by that of a ring of polynomial *functions* of a real or complex variable, in the abstract algebraic definition, nothing is ''varying''. The preferred term is ''indeterminate''.

P.99, middle, definition of **reduction map** [>]: The term is motivated by the case where φ takes $\mathbf{Z}[X]$ to $\mathbf{Z}_n[X]$ for some n (usually prime) by ''reducing coefficients modulo n''. Though the term is not as intuitive in other cases, e.g., when φ is the embedding of the integers in the rational numbers, Lang makes the general definition by analogy with the motivating case.

P.102, second paragraph [>]: I claim that what Lang has shown here is essentially that the polynomial ring $A[X_1,\dots,X_n]$ is the *free commutative A-algebra on n generators*. Indeed, given a commutative A-algebra B, and elements $x_1,\dots,x_n\in B$, let us show that there is a unique A-algebra homomorphism $A[X_1,\dots,X_n]\to B$ taking each X_i to x_i. To fit our presentation to Lang's, let us first consider the case where A is a subring of B, and the homomorphism $A\to B$ defining the A-algebra structure of B is the inclusion. Then what Lang calls the *evaluation map* determined by $x_1,\dots,x_n\in B$ will be the unique A-algebra homomorphism with the desired property.

In the general case, if we let A' denote the image of A under the homomorphism that makes B an A-algebra, then by combining the ''reduction map'' $A[X_1,\dots,X_n]\to A'[X_1,\dots,X_n]$ with the ''evaluation map'' $A'[X_1,\dots,X_n]\to B$ determined by $x_1,\dots,x_n$, we get the desired unique homomorphism. But there is really no need to divide the construction of this homomorphism into two stages as Lang has done; if $a\mapsto\bar a$ is the homomorphism defining the A-algebra structure of B, one can define the homomorphism required by the universal property (for which ''substitution map'' is still an appropriate term) in one step, as $\Sigma\, a_{(\nu)}\, X_1^{\nu_1}\dots X_n^{\nu_n}\ \mapsto\ \Sigma\, \bar a_{(\nu)}\, x_1^{\nu_1}\dots x_n^{\nu_n}$.

P.102, last paragraph, definition of **occur** [>]: Lang here gives the word ''occur'' a technical meaning, to allow him to conveniently refer to the set of monomials which have nonzero coefficient in f. A more standard way of doing this is to call that set, $\{M_{(\nu)}(X)\mid a_{(\nu)}\neq 0\}$, the *support* of the polynomial f.

P.104, beginning of material on group and monoid rings, through end of §II.3 on p.107 [~]:

The idea of a group (or monoid) ring.

Suppose A is a commutative ring, and we are interested in an action of a group or monoid G by module endomorphisms on an A-module M. In this situation, note that the underlying additive group of M has two systems of endomorphisms given: the actions of the elements of A, and the actions of the elements of G. Now on the one hand, these systems of abelian group endomorphisms are subject to the usual conditions defining an A-module structure, respectively a G-set structure; on the other hand, the statement that G acts by *A-module endomorphisms* means that the action of each element of G *respects* the operations of the elements of A; in other words, the two sorts of operations commute with one another: $gax = agx$ ($g\in G$, $a\in A$, $x\in M$). We may bring these systems of operations together by looking at the ring of additive-group endomorphisms of M that they generate. We find that we can write elements of this ring as linear combinations of operations induced by members of G, with coefficients taken from A, and that composition of such linear combinations is described by the law given on p.104 (and also p.85).

This suggests that we construct a ring whose elements are formal linear combinations of elements of G with coefficients in A, and whose multiplication is given by that rule. We find that this construction indeed gives a ring AG, and we verify that a left AG-module is *equivalent to* an A-module with an action of G on it by A-module endomorphisms!

Let us be more precise about how AG is defined: to pin down the concept of ''formal linear combinations of elements of G with coefficients in A'', we observe that such an element should be determined by the coefficients of the elements of G, and that all but finitely many of these coefficients should be zero. That is, such an element will be specified by a function associating to each element of G an element of A, such that the element associated to almost all elements of G is 0. Hence, the set of such functions is taken as the formal definition of the underlying set of AG (Lang, p.104). We then introduce the convenient abbreviation of using the symbol for an element of G to denote the image of that element in AG; formally the function which has value 1 on that element and 0 elsewhere, so that all elements can be written $\Sigma_{x\in G}\, a_x x$. (We are assuming G written multiplicatively. In the contrary case, we have to do some adjusting. E.g., to describe $A\mathbf{Z}$, we take a group written multiplicatively that is isomorphic to $\mathbf{Z}$, say $\langle x\rangle = \{x^i \mid i\in\mathbf{Z}\}$, then let our ring consist of formal linear combinations $\Sigma\, a_i x^i$ $(a_i\in A,\ i\in\mathbf{Z})$.)

Actually, this rigamarole of ''almost everywhere zero functions from G to A'' is just a way of describing the *free A-module* on the set of elements of G, which Lang cannot refer to directly at this point because he hasn't introduced modules! So the easier description of the underlying set of the monoid ring is as the free A-module on the underlying set of G.

Once the underlying set of this ring has been described as a free A-module, the natural way to describe its multiplication is to say that it is the unique A-bilinear map $\beta\colon AG\times AG\to AG$ which, on the basis G, agrees with the multiplication of G. The existence of such a bilinear map follows from Lemma III.4.c1. To see that it is associative, note that the ternary operations whose values on (u,v,w) are $\beta(u,\beta(v,w))$ and $\beta(\beta(u,v),w)$, respectively, are trilinear, and agree on elements of $X\times X\times X$; hence by the uniqueness part of that lemma, they are equal.

Incidentally, where Lang writes $A[G]$, I have substituted the symbol AG, preferred by group theorists. The general principle is that if X is a set, then a symbol of the form $A[X]$, since it resembles the symbol for a polynomial ring, is used to denote rings of linear combinations of *products* of elements of X with coefficients in A, while AX is used for rings or modules of linear combinations of elements of X alone with coefficients in A. The monoid algebra construction is of the latter sort. (Actually, there is a further distinction made. Suppose, for concreteness, that X consists of two elements, x, y. Then $A[X]$ is generally used for rings whose elements can be written as linear combinations of terms $x^i y^j$, while if one does not have enough commutativity or similar conditions to reduce to such a form, but needs to include separately terms such as x^2y, xyx and yx^2, one generally writes the algebra as $A\langle X\rangle$ or $A\langle x,y\rangle$.)

A terminological point: the ring AG should logically be called a group or monoid *algebra*. Both the terms ''group (or monoid) ring'' and ''group (or monoid) algebra'' are used in the literature. Lang, of course, cannot say ''algebra'' here, because he won't define that concept till Chapter III.

Some easily verified but important properties of monoid algebras are given in the following

Lemma II.3.c1. *Let A be a commutative ring. Then*

(i) *For any monoid G, the monoid algebra AG has the universal property that given an A-algebra R and a homomorphism h of G into the multiplicative structure of R, there exists a unique A-algebra homomorphism $AG\to R$ extending h.* (*I.e., AG is initial in the category of A-algebras given with homomorphisms of G into their multiplicative structures.*)

(ii) *For any set X, the monoid algebra on the free abelian monoid on X is the free commutative A-algebra on X; i.e., the polynomial algebra $A[X]$.*

(iii) *For any set X, the monoid algebra on the free monoid on X is the free A-algebra on X* (*generally written $A\langle X\rangle$*).

(iv) *An A-module M together with an action of a monoid G on M by A-module endomorphisms* (*with $1\in M$ understood to act by the identity*) *is equivalent to an AG-module, under the correspondence sketched earlier in this remark.*

Sketch of Proof. (i): Because AG is free as a module on the underlying set of G, the universal property of free modules allows us to extend the set-map h to an A-module-homomorphism $AG \to R$. If one now calls on the fact that $h: G \to R$ is assumed to respect multiplication, one can verify that A-linear combinations of elements $h(g)$ $(g \in G)$ multiply in the ring R by the same formula that defines the multiplication in the monoid algebra, and hence that the extension of h respects ring multiplication.

(ii): It will suffice to show that this monoid algebra has the universal property of the polynomial algebra. Given any map $f: X \to R$, where R is a commutative A-algebra, the universal property of the free abelian monoid G on X allows us to extend this in a unique manner to a homomorphism from that monoid to the multiplicative monoid of R. The universal property proved in (i) above for the monoid algebra now allows us to extend this map in a unique manner to an algebra homomorphism $AG \to R$, establishing the desired universal property.

(iii) is proved in the same way. Incidentally, the free monoid on X (respectively, the free abelian monoid on X) can be shown to be isomorphic to the sub*monoid* generated by X in the free group (respectively the free abelian group) on X.

(iv) is the result in terms of which we motivated the construction of AG! To formalize the argument, note that an action of G on an A-module M means a monoid homomorphism $G \to \mathrm{End}_A(M)$. By (i), this is equivalent to an A-algebra homomorphism $AG \to \mathrm{End}_A(M)$, which in particular gives a ring homomorphism from AG to the ring of abelian group endomorphisms of M, i.e., an AG-module structure on M. Conversely, every AG-module structure gives an A-module structure and an action by G respecting that structure. It is straightforward to verify that these ways of going between one and the other structure on M are inverse to one another. □

Note that statement (iv) of the above lemma means that linear actions of G on A-modules (in particular, on A-vector-spaces if A is a field) may be studied by looking at the structure of the algebra AG, and its modules. This is an important principle in the theory of groups and their representations. Chapter XVIII of Lang (not covered in 250A) gives an introduction to this technique in the case where G is a finite group.

There is a pair of longstanding open questions concerning group algebras:

Open Questions II.3.c2. *Let K be a field and G a torsion-free group, i.e., a group with no nonidentity elements of finite order. Must the following statements be true?*

(i) (*Zero-divisors conjecture.*) *KG is a ring without zero-divisors.*

(ii) (*Trivial units conjecture.*) *The only invertible elements of KG are the products cg $(c \in K-\{0\},$ $g \in G)$.*

When Lang first sketchily introduced group algebras as an Example on p.85, he called their multiplication operation ''the convolution product''. For those who are curious, let me describe the

Background of the term ''convolution'' (optional).

This term actually goes back to a very different line of thought.

Suppose f is a ''non-smooth'' function on the real line, such as the function which is 1 on $[0, 100]$ and 0 elsewhere, and we seek to ''smooth'' it out by ''blurring'' it. Intuitively, a nice way to do this would be if we could replace the ''value'' at each point by a bell-shaped curve, and integrate these together. In fact, if we write $g(x) = e^{-x^2}/\sqrt{\pi}$, we find that the above idea can be made precise, by taking for the ''blurred'' variant of f the function $f * g$ defined by

$$(f * g)(x) = \int_{-\infty}^{+\infty} f(x-y)\, g(y)\, dy. \tag{c16}$$

The idea is that the above definition makes the value of $f * g$ at x depend not only on the value of f at

x, but also on the value at nearby points $f(x-y)$, with the contributions of these points ''weighted'' by the smooth function $g(y)$.

The above is just one way of motivating the operation (c16). I hope you can see that in general, (c16) describes an interesting way of combining two functions f and g (satisfying appropriate conditions to make the integral converge) to get a third. The name given to this construction is ''convolution'', meaning ''twisting together''. I will mention without proof one other property of this construction: the Fourier transform of the convolution of two functions is the *product* of their Fourier transforms; and conversely, the Fourier transform of their product is the convolution of their Fourier transforms.

If we look at functions on the integers instead of on the real line, we may define an analog of (c16) by using a sum instead of an integral:

(c17) $$(f*g)(n) \ = \ \Sigma_{m=-\infty}^{+\infty} \ f(n-m)g(m).$$

As before, we need a condition to insure that these infinite sums will make sense. The easiest such condition is that the functions f and g be zero almost everywhere; this allows us to make sense of the above definition for functions with values in an arbitrary ring A, where we do not necessarily have a concept of convergence. But in that case f and g can be thought of as elements of the group algebra $A\mathbf{Z}$; and our ''convolution multiplication'' turns out to be the multiplication of the group algebra. (To see this, note that the arguments of f and g in (c17), $n-m$ and m, range over all pairs of integers summing to n.) Hence, by extension, the term ''convolution multiplication'' is often, as in Lang, used for the multiplication operation of any group algebra.

If in (c17) we restrict f and g to be functions on the *nonnegative* integers, modifying the range of summation appropriately, we of course get the description of multiplication of polynomials, looked at as ''sequences of coefficients''.

P.105, first line of text [=]: ''triples $(u\ v,\ y)$'' should be ''triples (u, v, y)''.

P.106, Proposition II.3.1 [~]: Observe that the condition that $h(a) = a$ for all $a \in A$ is equivalent to saying that we are dealing with a homomorphism of A-algebras (though Lang does not introduce that concept till the next Chapter).

P.107, end of §II.3 [>].

The concepts of this section have given us enough examples of rings so that we can give some further classes of

Examples of modules.

Let A be a commutative ring, and as in Lemma II.3.c1 let $A\langle x,y\rangle$ denote the free associative A-algebra on two generators x, y, equivalently, the monoid algebra over A on the free monoid $\langle x,y\rangle$ on two generators. What does a module over $A\langle x,y\rangle$ look like? Let us think of such a module as given by an A-algebra homomorphism $A\langle x,y\rangle \to \mathrm{End}_A(M)$ for some A-module M (cf. Lemma III.1.c3(b)). The universal property of the free algebra $A\langle x,y\rangle$ says that such a homomorphism is determined uniquely if we specify, in any way, elements of $\mathrm{End}_A(M)$ to which the free generators x and y should be mapped. So a module over $A\langle x,y\rangle$ is determined by choosing any A-module M, and any two A-module endomorphisms ξ and η of M, and letting multiplication by x act as ξ, and multiplication by y act as η. You can make up examples ad lib, taking, say, $A = \mathbf{Z}$, letting M be the free $\mathbf{Z}$-module (free abelian group) on n generators, using any two $n\times n$ matrices over $\mathbf{Z}$ to specify two endomorphisms, and letting $\mathbf{Z}\langle x,y\rangle$ act as just described.

Modules over the polynomial algebra $A[x,y]$ have an identical description, except that ξ and η, rather than being any two endomorphisms, must be two endomorphisms that commute with one another.

I chose the case of two indeterminates simply for concreteness; the analogous characterizations hold for any number of indeterminates. A particularly simple and important case is that of a single indeterminate. In that case we see that $A\langle x\rangle$ and $A[x]$ are the same; let us use the latter symbol. Thus, a module over

$A[x]$ is equivalent to an A-module together with a single endomorphism ξ. In particular, if A is a field, this is a vector-space given with a single linear operator. In §IV.1 we shall see that if A is a field, $A[x]$ is a principal ideal domain. Now the theory of finitely generated modules over a PID (Lang, §III.7, not covered in 250A) is quite similar to the theory of finitely generated $\mathbf{Z}$-modules (a.k.a. finitely generated abelian groups) that we saw developed in §I.8. This theory can be used to give an elegant analysis of the structure of a finite-dimensional vector spaces given with a single linear operator; equivalently, a theory of similarity invariants for matrices over a field; Lang will do this in Chapter XIV.

The next topic could have been introduced back when we were reading §I.11 or I.12, or could be put off to a later section than this; but let me introduce it here, taking the monoid algebra concept as a springboard.

Adjoint functors – a sketch.

The universal property of the monoid algebra AG involves considering homomorphisms from a *monoid* G to a *k-algebra* R (point (i) of the lemma on monoid algebras in this Companion). But if we look at what this means, we see that we are really considering homomorphisms from G to the *monoid* of elements of R under multiplication. Thus, if we write U for the ''forgetful functor'' from A-algebras to monoids that forgets the A-module structure of an algebra and remembers only the multiplication, we can say that we are dealing with homomorphisms $G \to U(R)$. This is more sensible, since both objects now lie in the category of monoids. The universal property of the monoid algebra then says that AG is a A-algebra given with a homomorphism $\eta_G: G \to U(AG)$, such that for any A-algebra R, and homomorphism $f: G \to U(R)$, there is a unique A-algebra homomorphism $h: AG \to R$ such that $f = U(h)\eta_G$.

You should now check for yourself that the universal properties of *free group*, *free abelian group*, *free commutative A-algebra*, and *Grothendieck group* all have analogous forms, gotten by substituting for ''the forgetful functor from A-algebras to monoids'' the forgetful functor from groups to sets, from abelian groups to sets, etc.. (The symbol ''U'' and its alphabetical neighbors are often used for these forgetful functors because many of them can be described as ''underlying set'' functors, or the like. Similarly, ''F'' and its alphabetical neighbors are often used for what we shall soon be calling the ''left adjoints'' of these functors, because of the prototypical case of *free* object constructions.)

Given a monoid homomorphism $\varphi: G \to H$, Lang has noted that we get a ring homomorphism $AG \to AH$. The existence of this homomorphism also follows from our universal property: the canonical homomorphism $\eta_H: H \to U(AH)$, composed with the given homomorphism $\varphi: G \to H$ yields a homomorphism $G \to U(AH)$; the universal property of AG now implies that this arises from a homomorphism $\bar\varphi: AG \to AH$; precisely, that there is a unique $\bar\varphi$ such that $\eta_H\,\varphi = U(\bar\varphi)\eta_G$. It is not hard to show that the construction of $\bar\varphi$ from φ makes the monoid algebra construction a *functor*, which is again analogous to what we have seen is true for free groups, free rings, etc..

Let us state the phenomenon that we are observing in a general category-theoretic context:

Let $\mathbf{C}$ and $\mathbf{D}$ be any categories, and $U: \mathbf{C} \to \mathbf{D}$ a functor. Suppose that for every object D of $\mathbf{D}$, there exists an object $F(D)$ in $\mathbf{C}$, and a morphism $\eta_D: D \to U(F(D))$, with the universal property that for every object C of $\mathbf{C}$ and morphism $f: D \to U(C)$, there exists a unique morphism $g: F(D) \to C$ such that $f = U(g)\eta_D$. Then we can make F into a functor $\mathbf{D} \to \mathbf{C}$ in a natural way, and this is called the *left adjoint* to the given functor U. Likewise, U is called the *right adjoint* to F; the relation between U and F is called an *adjunction*. Given such an adjunction, one finds that for every pair of objects $C \in \mathrm{Ob}(\mathbf{C})$, $D \in \mathrm{Ob}(\mathbf{D})$, the relation between morphisms $f: D \to U(C)$ and $g: F(D) \to C$ described above is a *bijection*

$$\mathrm{Hom}_{\mathbf{D}}(D, U(C)) \longleftrightarrow \mathrm{Hom}_{\mathbf{C}}(F(D), C).$$

(The names ''right'' and ''left'' adjunction refer to the fact that U occurs in the *right* slot in one hom-set, and F in the *left* slot in the other.)

A result which you should not find hard to prove, once you have digested the above concept (but for which you are not responsible) is

Lemma II.3.c3. *If* $\mathbf{C} \xrightarrow{U} \mathbf{D} \xrightarrow{V} \mathbf{E}$ *are functors having left adjoints* $\mathbf{C} \xleftarrow{F} \mathbf{D} \xleftarrow{G} \mathbf{E}$, *then the composite functor* VU *has left adjoint* FG. $\square$

For example, if we denote by $\mathbf{Comm}_A$ the category of commutative A-algebras, and by **AbMonoid** the category of abelian monoids, then we know that the forgetful functors $\mathbf{Comm}_A \to \mathbf{AbMonoid} \to \mathbf{Set}$ (where the first functor forgets the additive structure and remembers the multiplication, while the second forgets even that) have for left adjoints the *monoid algebra* and *free monoid* functors respectively, and that their composite, the forgetful functor from commutative A-algebras to sets, has for left adjoint the *polynomial algebra* functor. Thus, the above lemma shows why point (ii) of the lemma on monoid algebras proved in this Companion, saying that polynomial algebras are monoid algebras on free commutative monoids, had to be true. Looking at not-necessarily-commutative A-algebras and monoids, one likewise gets point (iii) of that lemma. You can discover some further results of this sort by applying the above lemma to the forgetful functors $\mathbf{Ab} \to \mathbf{Group} \to \mathbf{Set}$, and likewise $\mathbf{Ab} \to \mathbf{AbMonoid} \to \mathbf{Set}$.

Though the above examples of functors having left adjoints were all *forgetful* functors, there are many that do not fit this description. For instance, you might find it interesting to prove that for any positive integer n, the functor taking every commutative ring A to the group $\mathrm{SL}_n(A)$ of $n\times n$ matrices over A with determinant 1 has a left adjoint.

We will not be able to study the subject of adjunctions further in this course, aside from noting a few more examples. For further development, see any text on category theory, in particular, my 245 notes or Mac Lane's *Categories for the Working Mathematician*. One of the most striking results developed in the 245 notes is a theorem of Peter Freyd (in Chapter 9) characterizing precisely the functors among varieties of algebras that have left adjoints. (The term ''variety of algebras'', defined there, includes all of the categories of mathematical objects that have arisen so far in this course, in particular **Set**, **Monoid**, **Group**, **Ring**, the category $\mathbf{Comm}_A$ for A a fixed commutative ring, the category R-**Module** for R any ring, and the category G-**Set** for G a fixed group.)

Note on automorphisms of polynomial algebras (*optional;* > *Exercise II.3:1*).

For A a commutative ring, Exercise II.3:1(a)-(c) gives a simple description of the automorphism group of the polynomial algebra $A[X]$. However, the study of automorphisms of polynomial algebras in several indeterminates is more complicated. For instance, for every polynomial $f(X) \in A[X]$, the algebra $A[X,Y]$ has an automorphism acting by $X \mapsto X$, $Y \mapsto Y + f(X)$. (What is its inverse?) Likewise, for every $g \in A[Y]$ we get an automorphism $X \mapsto X + g(Y)$, $Y \mapsto Y$. Though it is easy to see how automorphisms of the first type compose with each other, and how those of the second type compose with each other, maps gotten by alternately applying automorphisms of the two different types can get arbitrarily complicated. Nevertheless, it has been proved that if A is a field, the group of A-algebra automorphisms of the polynomial algebra $A[X,Y]$ is generated by the automorphisms of these two sorts, together with those of the more trivial sort $X \mapsto aX$, $Y \mapsto Y$ $(a \in A - \{0\})$. But for polynomials in more than two indeterminates, it is not known even whether the analog of this statement is true. (The subgroup of the automorphism group of $A[X_1, \ldots, X_n]$ generated by the automorphisms of these types is called the group of *tame automorphisms*, so the standard formulation of this question is, ''Is every automorphism of a polynomial algebra over a field tame?'' Incidentally, in the 2-variable case, though the automorphisms are, in the above sense, ''known'', there is another longstanding open question concerning these, the *Jacobian conjecture*, which I will mention at the end of the comments concerning §IV.1.)

We remark that when A is *not* a field, part (d) of Exercise II.3:1 can fail. Indeed, let k be any integral domain (say a field), and let $A = k[Y]$. Then the k-algebra automorphism of $k[X,Y] = A[X]$

that acts by $X \mapsto X$, $Y \mapsto Y+X$ is a ring automorphism of $A[X]$ which does not take A into itself.

Re §II.4. Localization.

P.108, construction of the ring $S^{-1}A$ [~]:

What is being constructed is a universal example of a ring A' with a homomorphism $u : A \to B$ which sends all elements of S to invertible elements of B.

The construction of this sort that most of you have probably seen is that of the field of fractions of an integral domain A (where $S = A - \{0\}$). To understand the details of the general construction, however, we must also think about the case where A has zero-divisors.

Suppose that in a commutative ring A one has a relation $sa = 0$, and that f is any homomorphism of A into a commutative ring A' such that $f(s)$ is invertible. The equation $sa = 0$ gives us $f(s)f(a) = 0$, and multiplying by $f(s)^{-1}$, we get $f(a) = 0$. This shows that if we wish to map a ring A to another ring in which every member of a multiplicative submonoid $S \subseteq A$ becomes invertible, then every element of A which is annihilated by a member of S must go to zero. Now suppose we want to test whether two elements, $f(a_1)f(s_1)^{-1}$ and $f(a_2)f(s_2)^{-1}$ are equal. By bringing them to the common denominator $f(s_1 s_2)$, we may write their difference $f(s_2 a_1 - s_1 a_2)f(s_1 s_2)^{-1}$. This will be zero if and only if $f(s_2 a_1 - s_1 a_2) = 0$, and by our preceding discussion, this will necessarily happen if $s_2 a_1 - s_1 a_2$ is annihilated by some element $s' \in S$.

Now a general principle in attempting to construct a mathematical object with certain properties is to collect all the conditions we can come up with on how it should look, and then try to build a model based on these conditions. If we have collected enough conditions, the model will ''work''; if not, we see how it fails to work, and try to extract from this further conditions. Having found that when a commutative ring A is mapped to another ring A', in such a way that elements of a given multiplicative submonoid S all become invertible, two elements of the form $f(a_1)f(s_1)^{-1}$ and $f(a_2)f(s_2)^{-1}$ will be equal in the latter ring if

(c18) $s_2 a_1 - s_1 a_2$ is annihilated by an element of S,

we can try constructing our ring to consist of formal expressions intended to represent elements $f(a)f(s)^{-1}$ (namely, ordered pairs (a, s)), and identify two expressions (a_1, s_1) and (a_2, s_2) precisely when (c18) holds. As noted in the reading in Lang, this ''works''. Since the arithmetic conditions we have imposed on ordered pairs (a, s) are conditions that must hold among the elements $f(a)f(s)^{-1}$ for *any* homomorphism f of A into a ring which maps elements of S to invertible elements, the map of A into the ring we have constructed is universal.

There are certain improvements one can make in Lang's verification that the above construction does indeed ''work''. For instance, because S is a multiplicative submonoid, it is easy to see that for any two equivalence classes u and v, there exist elements $(a, s) \in u$ and $(a', s) \in v$ with the same second component s; intuitively, that ''any two fractions can be written to a common denominator''. Now given $u, v \in S^{-1}A$, suppose we choose representatives, (a, s) and (a', s), with the same second component, and define $u+v$ to be the equivalence class of $(a+a', s)$. One must verify that this is well-defined (independent of our choice of expressions for u and v having common denominator). Once this is done, the verification that addition of fractions is *associative* is much more straightforward than when it is defined by a messy formula. One can also put the condition for two elements (a, s) and (a', s') to be equivalent in a form that is slightly more elegant and convenient to calculate with; namely that there exist $t, t' \in S$ such that $(ta, ts) = (t'a', t's')$.

The above construction is called ''localization'', but it will be easiest for me to explain why after Lang has introduced the concept of ''local ring'', later in this section.

Another way of approaching localization (sketch, optional).

How does a child who knows about the integers learn the meaning of ''$1/2$''? One of the things he or she learns is that certain integers can be ''halved''. In other words, ''$1/2$'' corresponds to a certain *partial map* from integers to integers: $0 \mapsto 0$, $2 \mapsto 1$, $4 \mapsto 2$, etc.. Similarly, $1/4$ corresponds to the partial map $0 \mapsto 0$, $4 \mapsto 1$, $8 \mapsto 2$, etc.. This suggests that $2/4$ should correspond to the partial map $0 \mapsto 0$, $4 \mapsto 2$, $8 \mapsto 4$, etc.. But we want to have $2/4 = 1/2$; so it seems that we should define an *equivalence relation* among partial maps, which relates maps which agree on a reasonable subset of the intersection of their domains.

When we abstract these ideas, we get the following: Let A be a commutative ring, and S a multiplicative submonoid in A. Let Φ denote the set of all *A-module homomorphisms* $sA \to A$ with $s \in S$, and let $\sim$ be the equivalence relation which puts $\varphi_1: s_1 A \to A$ and $\varphi_2: s_2 A \to A$ in the same equivalence class if and only if they have the same restriction to some ideal $s_3 A$ $(s_3 \in S)$ contained in both $s_1 A$ and $s_2 A$. Let A' denote the quotient set $\Phi/\sim$. We can define ring operations on this set: we add two partial maps by restricting them to a common ideal and adding their restrictions; we multiply them by composing them, again after appropriate restriction. It is not hard to verify that we get, in this way, well-defined operations on equivalence classes of partial maps, and that these make A' a ring.

We may map A into this ring by sending each $a \in A$ to the equivalence class of the map ''multiplication by a'': $A \to A$. If the set $S \subseteq A$ consists of non-zero-divisors, it is easy to verify that the image of each $s \in S$ under the above homomorphism is invertible, with inverse given by the map $sA \to A$ taking sa to a $(a \in A)$. In this situation, A' is, up to isomorphism, precisely the ring $S^{-1}A$ we are interested in. If some elements of S are zero-divisors in A, however, this does not work. (The above map $sa \mapsto a$ is not well-defined.) So to use this approach to localization, one must *first* divide A by the ideal I of all elements that are annihilated by members of S, then apply the above construction to A/I.

P.109, before third-from-last line [>]: We have just seen that under appropriate conditions, the map $A \to S^{-1}A$ is injective. There is a converse sort of result, saying that an injective ring homomorphism $A \to B$ with certain properties must be ''essentially'' this map $A \to S^{-1}A$.

Lemma II.4.c1. *Let A be an integral domain, and let S be a multiplicative subset which does not contain 0. Suppose $f: A \to B$ is a one-to-one ring homomorphism such that for every $s \in S$, $f(s)$ is invertible in B.*

Then the homomorphism $h: S^{-1}A \to B$ given by the universal property of $S^{-1}A$ is also one-to-one. Hence $S^{-1}A$ is isomorphic (*via this map h*) *to the subring $\{ f(s)^{-1} f(a) \mid s \in S, a \in A \}$ of B.*

Proof. Let a/s be a nonzero element of $S^{-1}A$; thus a is a nonzero element of A. As discussed by Lang on this page, $h(a/s) = f(s)^{-1} f(a)$. If this were zero, then multiplying by $f(s)$ we would get $f(a) = 0$, contradicting the assumption that f is one-to-one. Hence h has trivial kernel, so it is one-to-one, giving the first assertion. Hence h is an isomorphism between its domain, $S^{-1}A$, and its image, $\{h(a/s)\} = \{ f(s)^{-1} f(a)\}$, giving the second assertion. □

As Lang will note in point **2** on the next page, if we take S to be the set of all nonzero elements of A, then $S^{-1}A$ is what is known as the field of fractions of A. Hence the above Lemma shows that given any one-to-one homomorphism f of an integral domain A into a ring B in which all nonzero elements of A become invertible (for instance, any one-to-one homomorphism of A into a field B), the elements of the form $f(s)^{-1} f(a)$ in B form a subfield of B isomorphic to the field of fractions of A.

P.110, second line [=]: For ''In that case'' read ''For $S = A^*$''.

P.110, three lines before point ''**3.**'' [=]: The phrase ''From general localization properties'' is somewhat cryptic. I think what Lang means is that the localization of $A[X_1, \ldots, X_n]$ obtained by inverting all polynomials which do not have the value zero at $(b_1, \ldots, b_n)$ can be mapped homomorphically to K so as to send each X_i to b_i, by the universal property of localization (p.109), and, further, that this localization also embeds in $A(X_1, \ldots, X_n)$, so that one can map the appropriate subring of that field homomorphically to K, by the indicated substitution.

P.110, first two lines of point ''**3.**'' [=]: In a general partially ordered set, a unique maximal element need not be a *greatest* element. However, the partially ordered set of proper ideals of a ring has the property that every element is $\leq$ a maximal element. (Do you remember why?) It follows that if a ring R has a unique maximal ideal, this is a *largest* proper ideal.

P.110, end of same paragraph [>]:

Background of the terms ''local ring'' and ''localization''.

Let us think of the ring of polynomials over the complex numbers, $A = \mathbf{C}[X]$, as a ring of functions on the complex plane. We know it is a principal ideal domain, and since every polynomial in this ring is a product of linear factors (something we will prove in a later chapter, but which I assume you have seen in an undergraduate course), the irreducibles are (up to units) simply the elements $X-c$ for $c \in \mathbf{C}$. Hence the general nonzero prime ideal of A is of the form $(X-c)A$. Note that this ideal can be described as consisting of all polynomials whose value at the point c is 0.

Since A is a unique factorization domain, we see that an element $f(X)$ of this ring will be divisible by an element $g(X)$ if and only if at each $c \in \mathbf{C}$ which is a zero of the polynomial g, the polynomial f has a zero of at least as high a multiplicity as g has. Within the field of fractions of A, regarded as the ring of ''rational functions'' on $\mathbf{C}$, the subring A can be characterized as consisting of the elements which do not ''blow up'' anywhere on the plane.

Now suppose we fix a point $c \in \mathbf{C}$, and blind ourselves to everything but the ''local'' picture in the immediate neighborhood of c. With our vision thus restricted, the best approximation of the condition ''f is divisible by g'' that we can perceive is ''the order of the zero of f at c is at least the order of the zero of g at c''. Thus, whenever two such elements f and g satisfy the above condition, we will accept f/g as though it were a member of our ring. The elements so ''accepted'' indeed form a ring, consisting of all rational functions which can be written with denominator not divisible by $X-c$; equivalently, which do not ''blow up'' at c. This is the ring called $A_{(X-c)A}$ (see Lang, p.110, middle, ''... and S^{-1} is denoted $A_{\mathbf{p}}$''). Since these rings arise by considering the ''local'' behavior of rings of functions, the type of rings they exemplify are called the *local rings*.

Although the construction of adjoining inverses to a set S of elements of a ring was *first* introduced to construct fields of fractions, it is in the above context that it required more careful study; hence the general construction of rings $S^{-1}A$ from rings A has come to be called ''localization''. (Note, however, that localization with respect to a multiplicative monoid does not in general yield a local ring.)

I described the construction of $\mathbf{C}[X]_{(X-c)}$ facetiously in terms of ''blinding oneself to everything but the local picture''. But in fact, it is not a pointless exercise. It turns out that many considerations of number theory and ring theory can be best studied by considering the behavior of a ring A ''one prime at a time'', and then putting these results together; and the local rings $A_{\mathbf{p}}$ are the most convenient tools for these ''one prime at a time'' considerations.

P.111, end of §II.4 [>].

Terminological warning. There is a certain amount of confusion between terms used for rings obtained by dividing a given ring A by an ideal, and by inverting a multiplicative set of elements. The first kind of ring, A/I, is sometimes called the ''factor ring'' and sometimes the ''quotient ring'' of A by the ideal I. The former term has the disadvantage that it can also mean one of the ''factors'' A_i in a direct product decomposition $A = \Pi\, A_i$ (though to avoid ambiguity the latter can be called a ''direct factor'' of A); the term ''quotient ring'' of course invites confusion with ''ring of quotients''. A localization $S^{-1}A$ may be called a ''ring of quotients'' or a ''ring of fractions''; the former term, as noted, invites confusion with one of the terms for the construction A/I. To avoid ambiguity, we shall follow Lang in calling the construction A/I a ''factor ring''. (We have likewise called a group G/N a ''factor group''). Unfortunately, in the case of an arbitrary set S and an equivalence relation $\sim$, the term ''quotient set'' is the only standard one.

The remaining notes on this section are optional.

''Specializations'' of fields (*sketch, optional*).

Recall that I translated Lang's remark in the two sentences before point **3** on p.110, as meaning that the set of rational functions in $X_1, \dots, X_n$ that could be evaluated at a point b of K^n forms a subring of the rational function field, which may be mapped homomorphically to K so as to send each X_i to b_i. Such *partial homomorphisms* from a field E to another field K – to be precise, homomorphisms φ from a subring $E_\varphi \subseteq E$ to K, having the property that if $x \in E_\varphi$ and $\varphi(x) \neq 0$, then $x^{-1} \in E_\varphi$ as well – are studied under the name *specializations* $E \to K$. (In this situation, it is easy to verify that the ring E_φ will be a *local* subring of E.) The concept has some surprises. For instance, suppose one has a specialization $\mathbf{Q}(X, Y) \to \mathbf{Q}$ taking X to 1 and Y to 2. This information does *not* completely determine φ, for on the element $(X-1)/(Y-2)$ it may be undefined, or it may be defined and take on *any* value! One can also have specializations from a field of characteristic 0 to one of finite characteristic, while this is clearly impossible for *homomorphisms* of fields.

One way of getting uniqueness is to study specializations which have minimal domain among all specializations whose domains contain a given set of elements (e.g., the set $\{X, Y\}$ in the above discussion). Using this idea, one can show that the rational function field $\mathbf{Q}(X_1, \dots, X_n)$ has a universal property with respect to specializations, which could justify calling it the ''free field'' on $X_1, \dots, X_n$, though there is no free field in the ordinary sense. Sometimes, *maximal* specializations are also studied; these are known as *places*. The concept of specialization is also used in studying division rings; cf. P. M. Cohn's books *Free Rings and Their Relations,* 2nd ed. 1985, and *Skew Field Constructions,* 1977; or the revised versions of these books, *Free Ideal Rings and Localization in General Rings*, 2006, and *Skew Fields*, 1995. (''Skew field'' is another term for ''division ring''. ''Specialization'', incidentally, is an older general term for ''substituting in particular values'', and Lang sometimes uses it in this sense.)

Localization of noncommutative rings (sketch, optional).

In noncommutative ring theory, there are several distinct versions of the concept of localization, with varying properties. In ''universal localization'', one takes a ring R and a set of elements $\Sigma \subseteq R$, and constructs, by generators and relations, a ring $R\langle \Sigma^{-1} \rangle$ in which images of elements of Σ are all invertible, and which is universal for this property. But the study of the structure of this ring is a much more nontrivial problem than in the commutative case. E.g., an element of the form $rs^{-1}t + uv^{-1}w$ cannot, in general, be reduced to an expression involving just a single ''denominator''. To get elements of more complicated forms, such as $(rs^{-1}t + uv^{-1}w)^{-1}$, it may be necessary to *repeat* the process of adjoining inverses more than once. Finally, even if R is without zero-divisors, it may be difficult to determine whether $R\langle \Sigma^{-1} \rangle$ is or isn't the trivial ring!

However, under certain special conditions, one can construct from a ring R and a subset S a ring $S^{-1}R$ in which every element has the same simple form $s^{-1}r$ as in the commutative case. One way to find the conditions under which this can be done is to generalize the ''partial homomorphisms'' approach to localization described above. A left R-module homomorphism from a left ideal Rs to R can be thought of as right multiplication by some element $s^{-1}r$; the conditions for the sum and composite of two such maps to make sense turn out to be that for every $a \in R$ and $s \in S$, one have $Sa \cap Rs \neq \emptyset$. This is called the *left Ore condition* (after Oystein Ore), the construction of the ring $S^{-1}R$ in this case is called *Ore localization*, and when the set $S = R - \{0\}$ satisfies the left Ore condition, then R is called a *left Ore ring*, and $S^{-1}R$ turns out to be a division ring. Ore localization does not, in general, give rings with a universal property; but when it can be used, it often gives rings in which one can compute more easily than in rings arising by universal localization.

P. M. Cohn has discovered that many of the deficiencies of universal localization mentioned earlier can be overcome by inverting, not *elements* of R, but *matrices*. Cf. his books mentioned at the end of the comment on specializations, and for another approach to the material, my paper *Constructing division rings as module-theoretic direct limits*, Trans. A. M. S. **354** (2002) 2079-2114.

The fact that we are studying fields of fractions will be my excuse for mentioning a few simple but long-outstanding questions about rational numbers:

Open Question II.4.c2 (Richard Guy)**.** *Does there exist a point of the plane having rational distances from each of the vertices of the unit square,* (0,0), (0,1), (1,0), (1,1)? (= Problem D19 in Richard K. Guy, *Unsolved Problems in Number Theory*, Springer-Verlag, 1981.)

Open Questions II.4.c3 (Erdős and Graham)**.** (a) *For every positive integer* $d > 1$, *do there exist positive integers* d_1, d_2, d_3 *such that*

$$4/d = 1/d_1 + 1/d_2 + 1/d_3\,?$$

A more sweeping question is

(b) *For every positive integer* n, *is it true that all but* finitely many *of the rational numbers* n/d $(d > 0)$ *can be written* $1/d_1 + 1/d_2 + 1/d_3$ *as above?* (Cf. p.44 of P. Erdős and R. L. Graham, *Old and New Problems and Results in Combinatorial Number Theory*, Université de Genéve, 1980, and references given there.)

Re §II.5. Principal and factorial rings.

P.111, definition of **factorial ring** [~]: The standard term in English, which we shall use, is *unique factorization domain*, abbreviated UFD. (Lang almost gives this in his parenthetical alternative ''unique factorization ring''.) Recall also that we will be using the standard terms *integral domain* and *principal ideal domain* (PID) where Lang says ''entire ring'' and ''principal entire ring''.

P.112, second line [=]: By (a, b) Lang means $(a) + (b)$. I dislike the (a, b) notation, since it looks just like the symbol for an ordered pair; but it is common among number theorists and others. However, paging around this book, I don't see any other occurrences of it before Chapter IX.

P.112, fifth line after second display [=]: The conclusion ''and has a maximal element'' is by Zorn's Lemma.

P.112, proof of Theorem II.5.2 [~]:

The Ascending Chain Condition, and Theorem II.5.2.

The proof of Theorem II.5.2, that every PID is a UFD, becomes more understandable with some background. Let us note first that the hypothesis that A is a PID can be broken into two parts, which are used in very different ways. On the one hand it tells us that every ideal generated by *two* elements is principal, from which, by induction, we see that every ideal generated by finitely many elements is principal. Given this observation, the remainder of the condition that *every* ideal is principal is clearly just the statement that every ideal is finitely generated.

Now the condition that every 2-generator ideal is principal tells us that any two elements not only have a g.c.d., but a g.c.d. which can be written as a linear combination of these elements. In particular, if p is an irreducible element and a an element not divisible by p, then irreducibility of p says that p and a

have no nonunit common divisor, so their g.c.d. is 1, and we can write this as a linear combination $1 = xp + ya$. Taking the image in $A/(p)$ of this equation, we see that a becomes invertible in that ring. Thus, *every* nonzero element of $A/(p)$ is invertible, i.e., $A/(p)$ is a field. This means that (p) is a *maximal* ideal, and in particular, is *prime*; i.e., that if a product of two elements is divisible by p, one of these elements is divisible by p. This is the only consequence of the condition ''every 2-generator ideal is principal'' that we use.

The other half of the PID condition constitutes an extremely important concept of ring theory, which we characterize in the following

Lemma II.5.c1. *The following conditions on a commutative ring A are equivalent:*

(i) *Every ideal of A is finitely generated.*

(ii) *A has no infinite strictly ascending chain of ideals $\mathbf{a}_1 \subsetneq \mathbf{a}_2 \subsetneq \dots$.*

(iii) *Every nonempty set of ideals of A has a maximal element.*

Proof. (i) $\Rightarrow$ (ii): Assuming we have an infinite chain $\mathbf{a}_1 \subsetneq \mathbf{a}_2 \subsetneq \dots$, choose a finite generating set $x_1, \dots, x_n$ for its union. Each x_i lies in some $\mathbf{a}_{j_i}$; let $j = \max(j_1, \dots, j_n)$. Then clearly $\mathbf{a}_j = \mathbf{a}_{j+1}$, a contradiction.

(ii) $\Rightarrow$ (iii): Let S be a nonempty set of ideals of A. Take $\mathbf{a}_1 \in S$; if it is not maximal, choose $\mathbf{a}_2$ strictly larger; if this is not maximal, choose $\mathbf{a}_3$ strictly larger than $\mathbf{a}_2$, and so on. If this process did not terminate in finitely many steps, we would get a contradiction to (ii), so it terminates, i.e., at some step we get a maximal member of S.

(iii) $\Rightarrow$ (i): Given an ideal $\mathbf{a} \subseteq A$, let $\mathbf{b}$ be a maximal element in the set of finitely generated subideals of $\mathbf{a}$. We claim $\mathbf{b} = \mathbf{a}$. For if not, we could take $x \in \mathbf{a} - \mathbf{b}$, and then $\mathbf{b} + xA$ would be a larger finitely generated subideal of $\mathbf{a}$, contradicting our choice of $\mathbf{b}$. $\square$

A ring satisfying the equivalent conditions of the above lemma is called *Noetherian* (after Emmy Noether. Cf. Lang, introductory paragraph of §IV.4.) Condition (ii) is expressed by saying that A has *ascending chain condition*, abbreviated ACC; or more precisely, ACC on ideals. Clearly, the condition of ACC can be defined for arbitrary partially ordered sets. The analog of the lemma holds for essentially any sort of algebraic structure, with ''ideal'' replaced by any class of substructures which is closed under taking unions of chains.

Just as the fact that the natural numbers are well-ordered, i.e., satisfy *descending chain condition*, gives us the principle of mathematical induction, so one can do induction on any partially ordered set with ascending chain condition. Let me formulate this for Noetherian rings as

Lemma II.5.c2 (Principle of Noetherian induction)**.** *Let A be a Noetherian ring and P a condition on ideals, such that if $\mathbf{a}$ is an ideal and all ideals strictly larger than $\mathbf{a}$ satisfy P, then $\mathbf{a}$ satisfies P. Then all ideals of A satisfy P.*

Proof. Assuming the contrary, let S be the set of ideals of A that do not satisfy P. By the Noetherian assumption, S has a maximal element $\mathbf{a}$. Then all ideals strictly larger than $\mathbf{a}$ satisfy P, hence, by assumption, $\mathbf{a}$ too must satisfy P, a contradiction. $\square$

The effect of this principle is that in trying to prove a result about ideals in a Noetherian ring, one can safely say ''Let $\mathbf{a}$ be an ideal, and assume by Noetherian induction that all larger ideals satisfy the desired property.'' If one can deduce from this that $\mathbf{a}$ also has the desired property, then one has established that all ideals do.

The same method can be applied to any subclass of the ideals of a Noetherian ring. E.g., if we want to prove that all *principal* ideals have a property P, we may apply Noetherian induction to the condition ''either $\mathbf{a}$ is nonprincipal, or $\mathbf{a}$ satisfies P''.

Note that any PID satisfies condition (i) of Lemma II.5.c2, and so is Noetherian. Returning now to Lang's proof that every PID is a UFD, we can replace the explicit argument concerning an ascending chain of principal ideals by the reasoning, ''To prove that every nonzero nonunit $a \in A$ has a factorization into irreducibles, assume by Noetherian induction, applied to the nonzero principal ideals, that every nonunit that generates a larger ideal than (a) can be so factored. Now if a is itself irreducible, there is nothing to prove, while if there is a nontrivial factorization $a = bc$, then b and c each generate larger ideals than a, so by our inductive hypothesis, each may be factored into irreducibles. Multiplying together these factorizations, we get a factorization of a into irreducibles, as required.''

Knowing that every nonzero nonunit factors into irreducibles, and that every irreducible is prime, the proof of unique factorization is now easily completed as on the bottom of p.112.

For the reader who has, as recommended in the comments on §II.3, read the beginning of the chapter on modules, I will mention that an ideal of a commutative ring R (or a left ideal of any ring R) is the same thing as a (left) R-submodule of R, making the concept of a Noetherian ring a special case of that of a Noetherian module. Lemma II.5.c1 above is stated and proved in this broader context on the first two pages of Chapter X. (If you are interested, you should find all of §X.1 readable at this point.)

P.113, second and fourth paragraphs (''We could call ...'' and ''We observe ...'') [<]: Here Lang is talking about UFD's (''factorial rings''); though he doesn't state that assumption, he uses it.

P.113, line 3 [=]: ''We could call two elements $a, b \in A$ equivalent if there exists a unit u such that $a = bu$''. In fact, such elements a and b are standardly called *associates*. (This definition applies in any commutative ring, not necessarily a UFD.) If A is an integral domain, then elements x and y are associates if and only if $xA = yA$. (If A is not an integral domain, this is no longer true. The problem of finding a counterexample is Exercise II.5:4.)

P.113, definition of **least common multiple** [~]: Though Lang has been talking about unique factorization domains, one can give a definition of ''least common multiple'' in an arbitrary integral domain, analogous that of greatest common divisor (p.111 bottom); namely, a least common multiple of a finite family of elements means an element m which is a multiple of all the members of the family, and such that every common multiple of the family is a multiple of m. In a unique factorization domain, one sees that the element Lang describes here has this property. In a general integral domain, a finite family of elements may or may not have a least common multiple or a greatest common divisor.

P.114, end of §II.5 [>].

Sums and intersections of ideals, and g.c.d.'s and l.c.m.'s of elements.

If $\mathbf{a}$ and $\mathbf{b}$ are two ideals of a commutative ring A, we see that their sum $\mathbf{a}+\mathbf{b}$ is the least ideal containing both. There is also a greatest ideal contained in both, namely $\mathbf{a}\cap\mathbf{b}$. Note that the product ideal, $\mathbf{ab}$, is clearly contained in this intersection; it may or may not equal it. In a principal ideal domain, it is not hard to show that the product is strictly smaller than the intersection whenever $\mathbf{a}$ and $\mathbf{b}$ are generated by elements having a nontrivial common divisor. On the other hand, in *any* commutative ring A, I claim that if $\mathbf{a}$ and $\mathbf{b}$ are ideals such that $\mathbf{a} + \mathbf{b} = A$, then $\mathbf{a}\cap\mathbf{b} = \mathbf{ab}$. We have seen ''$\supseteq$''; to get ''$\subseteq$'', note that for any two ideals of a commutative ring, we have

$$(\mathbf{a}\cap\mathbf{b})(\mathbf{a}+\mathbf{b}) \;=\; (\mathbf{a}\cap\mathbf{b})\mathbf{a}+(\mathbf{a}\cap\mathbf{b})\mathbf{b} \;\subseteq\; \mathbf{ba}+\mathbf{ab} \;=\; \mathbf{ab}.$$

Substituting in $\mathbf{a} + \mathbf{b} = A$ gives the desired result. (This was stated as an ''exercise for the reader'' on p.87, third line from the bottom.)

Observe that a ''greatest common divisor'' of elements a and b of a commutative ring A means, in effect, an element c such that the ideal cA is the *smallest* principal ideal containing aA and bA; equivalently, the smallest principal ideal containing $aA + bA$. (As we have noted, given a and b, such a c may or may not exist.) In particular, if $aA + bA$ is principal (e.g., if A is a principal ideal domain), then a g.c.d. of a and b means a generator of this ideal. However, a g.c.d. may exist even if the ideal $aA + bA$ is not principal. In the polynomial ring $\mathbf{R}[x, y]$, we find that 1 is a g.c.d. of the elements x and y, though the ideal they generate consists of all polynomials with zero constant term, and so does not contain 1.

A ''least common multiple'' of a and b similarly means an element c such that cA is the greatest principal ideal contained in $aA \cap bA$. Since every ideal is a union of the principal ideals it contains, in this case we see that the only way such a greatest principal ideal can exist is if $aA \cap bA$ is itself principal.

Chapter III. Modules.

Re §III.1. Basic definitions.

P.117, bottom [>]:

The idea of a module.

The relationship between rings R and R-modules M is analogous to that between groups G and G-sets S which we studied in §I.5. Just as the concept of a group is motivated by that of a group of permutations of a set S, and the abstract concept of a group is brought back to its roots by studying sets S on which a group G acts, equivalently, sets S given with group homomorphisms $G \to \mathrm{Perm}(S)$, so the concept of a *ring* is motivated by the properties of the set of *endomorphisms of an abelian group*, and an important role in ring theory is played by abelian groups M on which a ring R acts by endomorphisms; equivalently, abelian groups M given with ring homomorphisms $R \to \mathrm{End}(M)$. These are called *left R-modules*.

In the study of actions of groups, though we set up the definition so that elements of the group were written to the left of elements of the set, there are some cases where it is more natural to write the group elements on the right. However, there is no essential difference between ''right'' and ''left'' actions, because if we have an action of one sort, we can re-encode it as an action of the opposite sort, using the observation that the group operation $(\ )^{-1}$ reverses the order of multiplication. (In formal terms $(\ )^{-1}$ is an ''anti-automorphism'' of any group G.) Thus, given a ''right action'', under which a group element g takes an element x to an element $x\cdot g$, satisfying $((x)\cdot g)\cdot h = x\cdot(gh)$, though we cannot define a left action by writing $gx =_{\text{def.}} x\cdot g$, because this definition will not behave properly with respect to composition, the definition $gx =_{\text{def.}} x\cdot g^{-1}$ will. In ring theory, on the other hand, a ring R may not have an anti-automorphism, so the concepts of *left R-module* and *right R-module* are not equivalent. However, right modules over a ring R are equivalent to left modules over a ring called R^{op} (''the opposite ring to R'') having the same elements and additive structure as R, but the reverse multiplication, $a * b = b\,a$. Thus it will suffice to discuss left modules almost exclusively in presenting general results. When we deal with modules over commutative rings, there is no distinction between left and right modules.

Since every abelian group has endomorphisms $x \mapsto nx$ for every integer n, every abelian group has a natural structure of $\mathbf{Z}$-module; indeed, $\mathbf{Z}$-modules are essentially the same thing as abelian groups. So you can think of the theory of modules as a generalization of the theory of abelian groups. Another class of modules with which you are already familiar are modules over a field, which are called *vector spaces*.

Incidentally, in defining a module, Lang referred to ''an operation of A on M (viewing A as a multiplicative monoid ...)''. Actually, he hasn't defined the concept of an action of a monoid on a set, but the definition is analogous to that of the action of a group on a set. The latter was defined as a group homomorphism $G \to \mathrm{Perm}(S)$; so the former may be defined as a monoid homomorphism $G \to \mathrm{Map}(S)$. Either concept is also equivalent to a map $G \times S \to S$ with the properties $1\,s = s$ and $(gh)s = g(hs)$ for all $g, h \in G$, $s \in M$.

P.118, end of middle block of paragraphs [>]:

Some basic constructions. Let M be a left R-module. Then you should check that (i) for any subset $N \subseteq M$, the set of elements of R which annihilate N (i.e., $\{r \in R \mid rN = 0\}$) is a left ideal of R; (ii) for any additive subgroup $N \subseteq M$, the set of elements of R carrying M into N will form a *right* ideal, and (iii) under the same assumption, the set of elements which carries N into *itself* will be a *subring*. Moreover, (iv) if N is in fact a submodule of M, then its annihilator (which we noted in (i) above was a left ideal), and the set of elements of R carrying M into N (which we noted in (ii) was a left ideal) will both be 2-sided ideals of R.

Note also in connection with (i) that if N is a singleton $\{x\}$, then the structure of the submodule $Rx = \{rx \mid r \in R\} \subseteq M$ is determined by the annihilator ideal I of x: $Rx \cong R/I$; just as the structure of the orbit Gx of an element x of a G-set is determined by the isotropy subgroup of x: $Gx \cong G/G_x$.

The above four constructions are useful ''generic'' ways of getting one-sided and two-sided ideals and subrings of R. Indeed, it is instructive to verify that *every* right ideal, left ideal, 2-sided ideal, and subring of R can be obtained in the ways indicated above, using appropriate M and N. And, just as it can be heuristically useful to regard a general subgroup H of a group G as the isotropy subgroup of an element of a G-set, so the above constructions give useful ways of thinking about ideals and subrings of a ring.

P.118, paragraph containing the bold-face words **vector space** [>]: Whenever we construct a ring R as a ring of endomorphisms of an abelian group or module M, that very construction makes M an R-module (what W.-Y. Hsiang likes to call the ''birth certificate representation'' of the given ring). Thus, since the ring of $n \times n$ matrices over K, often denoted $M_n(K)$, is constructed to model the ring of linear endomorphisms of the vector space K^n, that vector space has, as Lang notes, a structure of $M_n(K)$-module.

Likewise, since the Weyl algebra (described in my comments on §II.1) arises as the algebra of operators on a polynomial algebra $\mathbf{R}[x]$ generated by the operators X of multiplication by x, and Y of differentiation with respect of x, it follows that one can make $\mathbf{R}[x]$ into a module over this algebra by letting X and Y act in these ways.

Since I recommended reading sections III.1-III.4 before §II.3, we don't have many examples of rings yet, making it hard to give more examples of modules! At the end of §II.3 I note some further examples of modules.

P.119, last two lines, and p.120, top line [~]: If $f: M \to M'$ is a module homomorphism, then the module $M'/f(M)$ is universal (initial) among modules given with a homomorphism from M' into them whose composite with f is the zero homomorphism. This property is dual to that of the *kernel* of f, which is universal (terminal) among modules with a homomorphism into M whose composite with f is zero; hence $M'/f(M)$ is called the *cokernel* of f. We have the usual terminological alternatives of speaking loosely of the module $M'/f(M)$ as the cokernel, or using this term more precisely for the universal pair consisting of this module and the canonical map of M' into it. Lang mentions both usages, except that, as usual, he would abbreviate the pair to the map alone.

P.120, bottom, definitions of **monomorphism** and **epimorphism** [>]:

The question of what the words *monomorphism* and *epimorphism* should mean is currently a can of worms! They were originally coined (I believe by Bourbaki) as terms for ''one-one homomorphism'' and ''surjective homomorphism'' respectively. However, the early category-theorists, looking for ''category-theoretic equivalents'' of various concepts, noted that in almost all naturally occurring categories, the one-to-one homomorphisms could be characterized as those homomorphisms $f: A \to B$ such that for every object X and every pair of homomorphisms $i, j: X \to A$, one has $fi = fj \Rightarrow i = j$. They therefore named a morphism f in a general category having this property a ''monomorphism''. They then felt it natural to define an ''epimorphism'' to mean a morphism $f: A \to B$ with the dual property, namely that for every object Y and pair of morphisms $i, j: B \to Y$, one has $if = jf \Rightarrow i = j$.

Now it turns out that in about half the naturally occurring categories, the class of epimorphisms, so defined, coincides with the class of surjective homomorphisms, while in the other half it does not! However, in the latter cases, the class of epimorphisms *does* constitute a class of homomorphisms worth studying, and there has been considerable work on characterizing such maps among various sorts of algebras. Hence, ''epimorphism'' is now frequently used by algebraists in its category-theoretic sense, and I do *not* recommended using the term in the old meaning. The older meaning can conveniently be expressed by the phrase ''surjective homomorphism'' or ''onto homomorphism'', and I will do so here.

The category of modules over a ring R happens to be one of those in which epimorphisms are the same as surjective homomorphisms, so Lang's use of the term on p.120 does not contradict the category-theoretic meaning. However, I will generally use the more down-to-earth terms when there is a choice, unless the context is category-theoretic.

How should the statement that a morphism $f: X \to Y$ of algebraic objects is an epimorphism be thought of, in cases where it is not equivalent to surjectivity? Intuitively, it means that the subobject $f(X)$ ''controls'' all of Y. Depending on the sort of object we are dealing with, this may or may not entail $f(X) = Y$, i.e., surjectivity. (See Exercise I.11:1 for some positive and negative examples.) On the other hand, surjective homomorphisms are always epimorphisms. (Note that though the concepts of monomorphism and epimorphism are defined in arbitrary categories, these remarks on their relation with one-one-ness and onto-ness refer to categories of algebraic objects, and homomorphisms that are maps of underlying sets. As we have observed, one-one-ness and onto-ness are not even defined for objects of general categories.)

Although it is rare for *monomorphisms* to be anything other than one-to-one maps, here is one example where this, too, fails. Let $\mathbf{C}$ be the category of *divisible* abelian groups, that is, abelian groups which for every positive integer n satisfy $(\forall x)(\exists y)\ ny = x$. You should not find it hard to prove that in this category, the canonical map $\mathbf{Q} \to \mathbf{Q}/\mathbf{Z}$ is a monomorphism.

P.121, first paragraph [~]:

Notes on the concept of an algebra.

Lang's introduction to this concept is misleading; let me put it in context. There are many ways one can generalize the concept of *ring*. On the one hand, one can remove the associativity assumption on the multiplication, or the assumption of the existence of unit. The resulting objects are called *nonassociative rings* and *nonunital rings*. (In the former case, one frequently restricts attention to rings which, instead of associativity, satisfy some other interesting identity.)

On the other hand, one may note that in any ring, the underlying additive group structure can be looked at as a structure of $\mathbf{Z}$-module, but that in some naturally occurring examples, it has a natural structure of module over a larger ring, e.g., the rationals, the reals, or the complex numbers. One calls a ring given with a structure of A-module on its additive group, satisfying appropriate conditions, an A-*algebra*. What are these appropriate conditions? Note that in a ring R one can prove (from distributivity) that for any ring elements x and y and any integer n, one has $(nx)y = n(xy) = x(ny)$. To make the concept of A-algebra a good generalization of that of ring, one requires, likewise, that for any elements x and y of the algebra, and element $a \in A$, one have

(c19) $$(ax)y \;=\; a(xy) \;=\; x(ay).$$

(This is part of the condition in Lang, that the multiplication map of E be *bilinear*.) Note that if $a, b \in A$, $x, y \in R$, then (c19) can be used to transform $(ax)(by)$ either to $(ab)(xy)$ or to $(ba)(xy)$. Thus these must be equal. This leads to a degenerate theory unless A is assumed commutative; hence the concept of A-algebra is only defined for commutative rings A.

The two sorts of generalization of the concept of ring that we have discussed – dropping unitality or associativity on the one hand, and enhancing the underlying abelian group structure to an A-module structure on the other – are independent of one another. Thus, there is no good reason why Lang has connected the transition from rings to A-algebras with the possible removal of the conditions of associativity and/or unitality, as though rings were required to have these properties, but algebras were not.

Let us make explicit an observation that is particular to the case of associative unital algebras, and gives another way of looking at these animals. (We will use the words ''associative unital'' here, because we have been talking about other sorts of rings and algebras; but elsewhere, we will understand ''ring'' and ''algebra'' to include these conditions unless the contrary is stated.)

Lemma III.1.c1. *Let A be a commutative ring. Then the following structures are equivalent.*

(a) *An associative unital A-algebra R, i.e., an A-module R given with an associative unital bilinear multiplication.*

(b) *An associative unital ring R, together with a homomorphism f from A into the center of R.*

Sketch of Proof. Given an associative unital A-algebra R, if we regard its multiplication map as a bilinear map of abelian groups, we see that it makes R into a ring. Now let $f: A \to R$ be the map taking $a \in A$ to $a1 \in R$ (where 1 denotes the unit of the ring R, and this is multiplied by $a \in A$ using the A-module structure of R). It is easy to verify that this gives a homomorphism from A into the center of R.

Inversely, given an associative unital ring R and a homomorphism f from A into the center of R, we may make R an A-module by defining the product of $a \in A$ and $r \in R$ to mean the product $f(a)r$, evaluated in the ring R. It is easy to verify that this makes R an A-module, such that the ring multiplication of R is A-bilinear. Thus, R becomes an A-algebra, and the conditions of associativity and unitality hold in this algebra because they hold in R as a ring.

Finally, it is easy to verify that if we pass from a structure of unital associative A-algebra to a structure of ring given with a map of A into its center, and then back to an algebra structure, in the manner indicated, we get precisely the structure we started with; and likewise if we go from a structure of ring with a map of A into its center to an algebra structure and back. Thus, the constructions we have defined are inverse to one another, and give a bijection between the two sorts of structure on a set R. $\square$

(Lang sketches the above idea still more briefly in the next-to-last paragraph of p.121.)

Note that if R is commutative, the *center* of R, referred to in (b) above, is all of R. Hence we have

Corollary III.1.c2. *Let A be a commutative ring. Then the following structures are equivalent.*

(a) *A commutative associative unital A-algebra R.*

(b) *A commutative associative unital ring R, together with a homomorphism $f: A \to R$.* $\square$

There is another natural motivation for the concept of a unital associative A-algebra. If A is a commutative ring, and M is an A-module M, what structure does the set $\mathrm{End}_A(M)$ of all A-module endomorphisms of M have? Its elements can be added; they can be composed with each other, *and they can be multiplied by elements of A*. The first two of these operations satisfy the axioms for an associative unital ring, and the multiplication of this ring is bilinear with respect to the action of A. So we may think of the concept of A-algebra as an abstraction of the natural structure on these sets of A-module endomorphisms, just as the concept of ring abstracts the natural structure on the set of endomorphisms of an abelian group.

There is, unfortunately, another important technical use of the word ''algebra''. The subject-matter of the wide branch of mathematics called ''algebra'' is the properties of sets given with one or more operations – e.g., monoids, groups, G-sets for a fixed group G, rings, modules, etc.. Though many approaches to algebra, including that of this course, treat these as separate topics and introduce them one by one, the field called *Universal Algebra* or *General Algebra* studies the properties that such concepts have in common, and calls *any* such object an *algebra*. It would be very desirable if there were a name for this concept, distinct from the term used for the specific concept studied in ring-theory, but no one has come up with a good one, and the double usage is now entrenched. When there is ambiguity, one must distinguish the two concepts as ''an algebra in the sense of universal algebra'' and ''an algebra in the sense of ring theory''. Russian authors often call the former ''a universal algebra'', but this is not a good choice, because no universal property is assumed. In this course, unless otherwise stated, the word ''algebra'' will be used only in its ring-theoretic sense.

Lang does not say what is meant by a homomorphism of A-algebras. If we regard an A-algebra as an A-module with a ring structure whose multiplication is A-bilinear, then a homomorphism of A-algebras means a map which is both a ring homomorphism and an A-module homomorphism. If we regard an A-algebra as a ring with a homomorphism of A into its center, then a homomorphism of A-algebras means a ring homomorphism which makes a commuting triangle with the homomorphisms from A. These two definitions are easily shown equivalent.

What about modules over algebras? There are two ways of formulating the concept, but fortunately, they come to the same thing.

Lemma III.1.c3. *Let A be a commutative ring, and R an A-algebra. Then the following structures are equivalent:*

(a) *A left R-module M, that is, an abelian group given with a ring homomorphism $R \to \mathrm{End}(M)$, (where $\mathrm{End}(M)$ denotes the ring of abelian group endomorphisms of M).*

(b) *An A-module M given with an A-algebra homomorphism $R \to \mathrm{End}_A(M)$, (where $\mathrm{End}_A(M)$ denotes the A-algebra of A-module endomorphisms of M).*

Sketch of Proof. It is clear that a structure as in (b) gives, in particular, a structure as in (a), since $\mathrm{End}_A(M)$ is a subring of $\mathrm{End}(M)$, and an algebra homomorphism into the former is, in particular, a ring homomorphism into the latter. On the other hand, given a structure as in (a), note that the map of A into the center of R, composed with the map $R \to \mathrm{End}(M)$ gives a map $A \to \mathrm{End}(M)$, i.e., an A-module structure on A. The action of each $r \in R$ commutes with the actions of all elements of A (why?), hence each such element in fact gives an A-module endomorphism of M; so we get a structure as in (b). □

Recall that a $\mathbf{Z}$-module is the same as an abelian group. From this we easily see that a $\mathbf{Z}$-algebra is the same as a ring. Hence if we prove a result about algebras, we also have the corresponding result for rings.

P.121, **Examples** [<]: Skip this example, unless you are familiar with the subject matter. (I would love to insert here a thumbnail sketch of the concept of Lie algebra, but I think I've loaded in enough commentary on this page!)

Students who are in my Math 250A or others who are following my suggestion of reading this material immediately after §II.2 should also skip the discussion of group and monoid algebras; these will be introduced in §II.3.

P.121, next-to-last paragraph [~]: This is stated more precisely in Lemma III.1.c1 above.

Re §III.2. The group of homomorphisms.

P.122, second paragraph [>]: You should convince yourself that the calculation showing that $\mathrm{Hom}_A(X', X)$ becomes an A-module indeed works if A is commutative, but not for noncommutative A. (You might find it interesting to think about the question: Given a noncommutative ring A, is there in general *some* structure stronger than that of an abelian group that we can put in a functorial way on these hom-sets?)

Caveat: The sum of two ring homomorphisms is *not* in general a ring homomorphism. For rings R

and S, the set $\mathrm{Hom}(R, S)$ has no algebraic structure.

P.122, Proposition III.2.1 and p.123, Proposition III.2.2 [~]: Short exact sequences are a way of indicating the relation between a module, a homomorphic image of this module, and the kernel of the homomorphism. Since these are key module-theoretic constructions, when we study functors on module-categories, one of the first things we would like to know is how they behave on such sequences. We see from these two propositions that functors of the form $\mathrm{Hom}(X, -)$ and $\mathrm{Hom}(-, Y)$ each preserve most, but not all, of the exactness of short exact sequences. You should, in fact, look for examples showing that if

$$0 \to M' \to M \to M'' \to 0$$

is a short exact sequence of modules (e.g., $\mathbf{Z}$-modules, i.e., abelian groups), neither of the sequences

$$0 \to \mathrm{Hom}(X, M') \to \mathrm{Hom}(X, M) \to \mathrm{Hom}(X, M'') \to 0$$

$$0 \leftarrow \mathrm{Hom}(M', Y) \leftarrow \mathrm{Hom}(M, Y) \leftarrow \mathrm{Hom}(M'', Y) \leftarrow 0$$

need be exact, at the place where the respective proposition does not guarantee exactness.

The statements of the propositions are not limited to the cases where the given sequence is part of a short exact sequence; i.e., where the leftmost map of Proposition III.2.1 is one-to-one, and the rightmost map of Proposition III.2.2 is onto. However, you might verify that if a covariant or contravariant functor from modules to abelian groups preserves exactness of all sequences obtained by dropping the appropriate zero-module from a genuine short exact sequence, then it preserves exactness for the more general sort of exact sequence referred to in these propositions. Another useful result you might verify is that any functor from A-modules to abelian groups which takes all short exact sequences to short exact sequences in fact takes all exact sequences to exact sequences.

The fact that $\mathrm{Hom}(X, -)$ preserves exactness of sequences $0 \to Y' \to Y \to Y''$ is expressed by saying that it is a *left exact functor*; the fact that $\mathrm{Hom}(-, Y)$ preserves exactness of sequences $X' \to X \to X'' \to 0$ (though reversing the direction of arrows) is expressed by saying that it is a *right exact* (*contravariant*) *functor*.

In his proofs of these two results, Lang leaves to the reader, among other things, the ‘‘upward’’ direction; that is, the verification that *if* a sequence of the sort shown has the property that the induced sequences of hom-groups are exact for all choices of the outside object, *then* the given sequence is itself exact. The needed arguments are easy to find – once one is familiar with the following heuristic principle:

(c20) *To study the consequences of a condition holding for all morphisms of a given sort, consider a universal example.*

Let us see how to apply this in proving the ‘‘upward’’ direction in Proposition III.2.1. Lang doesn’t name the morphism $X'' \to X$ of the given sequence; let us call it μ. One of the conditions for exactness of the bottom sequence of hom-sets is that the kernel of the left-hand arrow be equal to the image of the right-hand arrow. To say that it *contains* this image for every Y is to say that for every Y, every element of the right-hand hom-set, $\mathrm{Hom}_A(X'', Y)$, when composed with μ (to map it into the middle hom-set) and then with λ (to map it into the left-hand hom-set) gives 0. So let us apply this observation to a *universal* example of a morphism from X'' into a module Y. Given no other conditions on the morphism than that it have domain X'', a universal example is the identity morphism of X''. The statement that this identity morphism, when composed with μ and then λ, gives 0 is just the condition $\mu\lambda = 0$, equivalently, $\mathrm{Im}(\lambda) \subseteq \mathrm{Ker}(\mu)$, which is indeed part of the desired statement that the original sequence be exact.

Let us now treat similarly the condition that for all Y, the kernel of the left-hand arrow *be contained in* the image of the right-hand arrow. This is a statement about every element of that kernel, so we should look for a universal instance of an element of that kernel; i.e., a universal example of a homomorphism from X into a module Y which goes to zero on composition with the map $\lambda: X' \to X$. If you draw a picture, you will see that the universal map with this property is the canonical map

$$q: X \to \mathrm{Coker}(\lambda).$$

So let us apply the exactness assumption on the bottom sequence in the case $Y = \mathrm{Coker}(\lambda)$. Since q is then in the kernel of the left-hand map of that sequence, there must be an element p in the right-hand hom-set $\mathrm{Hom}(X'', \mathrm{Coker}(\lambda))$ that maps to q in the middle hom-set, i.e., such that $q = p\mu$. Hence $\mathrm{Ker}(\mu) \subseteq \mathrm{Ker}(q) = \mathrm{Im}(\lambda)$ (the latter equality by the definition of q). Since we proved $\mathrm{Im}(\lambda) \subseteq \mathrm{Ker}(\mu)$ in the preceding paragraph, we get $\mathrm{Im}(\lambda) = \mathrm{Ker}(\mu)$, as desired.

The other parts of the ''upward'' direction of this proposition, and the proof of the same direction in the next proposition, are similar applications of (c20), which you should work out to get familiarity with this technique.

P.124, next-to-last paragraph [>]: There is no need to assume A commutative in this paragraph.

Note that, given what Lang calls a representation of R on an A-module M, the full structure one has on M can be described as one of an abelian group, together with a homomorphism $\alpha: A \to \mathrm{End}(M)$ giving the A-module structure, and a homomorphism $\rho: R \to \mathrm{End}(M)$ giving the action of R, subject to the condition that for all $a \in A$, $r \in R$, $x \in M$ we have $\alpha(a)\,\rho(r)(x) = \rho(r)\,\alpha(a)(x)$; or, writing the actions both of elements of R and of elements of A as ''multiplications'', $arx = rax$.

However, noncommutative ring theorists frequently make a slight notational change, writing endomorphisms of a *left* A-module M on the *right* of M, e.g., as $x \mapsto xh$, and defining composition of endomorphisms accordingly, so that $x(hh') = (xh)h'$. Let us denote the ring of such endomorphisms by $\mathrm{End}({}_A M)$; this will be the *opposite* of the ring Lang denotes $\mathrm{End}_A(M)$. (The *opposite ring* is defined in the comment to p.117 of Lang in this Companion.) Thus, given what Lang calls a representation of the ring R on the A-module M, if we write B for the opposite of the ring R, we can describe the map giving the representation as equivalent to a ring homomorphism $B \to \mathrm{End}({}_A M)$; in other words, a way of letting elements of B act on the *right* on the A-module M. The condition that the actions of elements of B respect the actions of elements of A now takes the form

$$(ax)b \;=\; a(xb) \qquad (a \in A,\ r \in R,\ x \in M).$$

An abelian group with actions of A and B on the left and the right respectively, and satisfying this law, is called an (A, B)-*bimodule*. (The name refers to the fact that it is composed of a *left A-module* structure and a *right B-module* structure, related by the above condition.) A major advantage of this notation is that if A and B have elements in common (for instance, if they are the same ring), then to unambiguously write the actions of both on the left would require awkward notation like $\alpha(a)\beta(b)x$; while the notation axb is unambiguous, and, in view of the above displayed equation, does not even require parentheses.

An important class of bimodules is given by the 2-sided ideals of a ring A; these are clearly (A, A)-bimodules in a natural way. (Note that Lang's notation would require calling these ''representations of A^{op} on left A-modules''.)

You will sometimes see the term (A, B)-bimodule written ''A–B–bimodule'', but I do not recommend this, since if in place of A and B one has complicated symbols, it can be unclear how the prefix is divided. We will not study bimodules (or ''representations'') in 250A, but you should be aware of the concept.

I have some further discussion of the convention of writing endomorphisms of left modules on the right and vice versa – its advantages and disadvantages – in a note at
/~gbergman/grad .hndts/left+right.ps .

Re §III.3. Direct products and sums of modules.

P.129, 7 lines from bottom [~]: Lang says an A-module generated by a subset S may be written $A\langle S\rangle$. Don't! As I have mentioned, that is noncommutative ring-theorists' notation for the *ring* generated over A by a set of elements S of an overring. For the A-module generated by S, simply write AS.

P.129, long paragraph, containing definition of submodule **generated** by a set [>]: This construction thus takes its place beside ''the submonoid of a monoid G generated by a subset $X\subseteq G$'', ''the subgroup of a group G generated by a subset $X\subseteq G$'', ''the normal subgroup of a group G generated by a subset $X\subseteq G$'', ''the subring of a ring R generated by a subset $X\subseteq R$'' and ''the right (or left, or 2-sided) ideal of a ring R generated by a subset $X\subseteq R$''. Note that in each of these cases, the set constructed is obtained as the *closure* of X under certain operations. Some properties common to such constructions are abstracted by the concept of

Closure operators.

Suppose that in each of the cases mentioned in the preceding paragraph, we denote the substructure ''generated by X'' as $\mathrm{cl}(X)$. Clearly, in each case we have

(c21) $$\mathrm{cl}(X)\ \supseteq\ X,$$

(c22) $$X\subseteq Y\ \Rightarrow\ \mathrm{cl}(X)\subseteq \mathrm{cl}(Y),$$

(c23) $$\mathrm{cl}(\mathrm{cl}(X))\ =\ \mathrm{cl}(X).$$

Operations cl with these properties are ubiquitous, not only in algebra, but in other areas of mathematics. (For instance, if S is any topological space, the operation of taking the *topological closures* of subsets of S satisfies these conditions, while if S is Euclidean n-space, the operation taking a set to its *convex hull* has the same properties.) An operator cl on subsets of any set S that satisfies (c21)-(c23) is called a *closure operator*, and a subset X of S is called *closed* with respect to cl if $\mathrm{cl}(X)=X$.

It is easy to verify that if cl is a closure operator on a set S, then the *intersection* of any family of closed subsets of S is a closed subset, and the improper subset $S\subseteq S$ is closed. Conversely, if C is a family of subsets of a set S such that the intersection of any set of elements in C is in C, and S itself is in C, then there is a unique closure operator cl on S whose closed sets are the members of C; namely, the operator taking every $X\subseteq S$ to the intersection of all members of C containing X. So closure operators on S are equivalent to sets of subsets with the above properties.

Particular sorts of closure operators satisfy additional conditions. For instance, examples arising in algebra, including all those that have come up in this course, typically satisfy the condition that the union of a *chain* of closed subsets is closed, while this is not true of topological closure. On the other hand, topological closure has the property that the union of finitely many closed sets is closed, which the algebraic examples do not.

At this point, we will not study the general concept any further, but simply put this useful definition under our belts. We will see it again in our discussions of §VI.1, §VI.14 and §XI.1.

P.129, next-to-last paragraph [~]: Lang indicates that a finitely generated module can also be called a ''finite'' module. This terminology is used by some algebraic geometers, but since it can lead to confusion, I don't recommend it, unless you are talking with those who use it regularly.

P.130, middle [>]: The concepts of linear independence for *sets* and *families* are both important, and are closely related; each can be expressed in terms of the other. A set is linearly independent if and only if the family one gets when one indexes the set by itself (using its identity map) is linearly independent as a family; a family $(x_i)_{i\in I}$ is linearly independent if and only if the set $\{x_i\mid i\in I\}$ is linearly independent, *and* the indexing map $i\mapsto x_i$ is one-to-one.

P.133, material on abelian categories [<]: I suggest skipping or skimming this, since such a brief introduction is not sufficient to make the concept useful. If you want to learn something about the subject, one old but good source is Peter Freyd's book *Abelian Categories* (Harper and Row, 1964). Rather surprisingly, Freyd's development of the theory shows that the abelian group structure that Lang and many other modern presentations of the subject assume given can in fact be deduced from simpler axioms that seem to have little to do with groups.

Re §III.4. Free modules.

P.135, second line of text [~]: Delete the assumption that S is nonempty! I will say more about this in my comments on the next section.

P.135, after proof of Theorem III.4.1 [>]:

Bilinear maps on free modules.

Note that Theorem III.4.1 characterizes *free modules* (defined in the preceding line) by a universal property parallel to the universal properties of free abelian groups and free groups. Indeed, free abelian groups are equivalent to free $\mathbf{Z}$-modules. (The *construction* of free modules on an arbitrary set will be discussed on p.137.)

The universal property of free modules is an existence and uniqueness property for module homomorphisms, also known as *linear* maps. From it, one can get a similar result for *bilinear* maps, which is very useful in constructing and studying these maps, and hence in defining algebra structures on certain free modules. This result is given in the lemma below; in the second paragraph, it is generalized further to *n-linear* maps, defined in the obvious way.

Lemma III.4.c1. *Let A be a commutative ring, and L, M and N A-modules, such that L and M are free, with bases X and Y respectively. Then for any set-map $t\colon X\times Y\to N$, there is a* unique *bilinear map $\tau\colon L\times M\to N$ whose restriction to $X\times Y\subseteq L\times M$ is t.*

More generally, the analogous result is true for n-linear maps $M_1\times\ldots\times M_n\to N$, where $M_1,\ldots,M_n$ are free on bases $X_1,\ldots,X_n$ respectively.

Proof. This could be done ''by hand'', essentially as Lang proves Theorem III.4.1, but let us note that the result is also a consequence of that theorem. For each $x\in X$, let us define a map t_x of Y into N by $t_x(y)=t(x,y)$. By Theorem III.4.1, there is a unique A-module homomorphism $\tau_x\colon M\to N$ extending t_x. Now the association $x\mapsto\tau_x$ is a set-map $X\to\mathrm{Hom}_A(M,N)$, hence there exists a unique A-module homomorphism $\tau_*\colon L\to\mathrm{Hom}_A(M,N)$ which sends each $x\in X$ to τ_x. If we define $\tau(u,v)=\tau_*(u)(v)$, we see that this is A-linear in u because τ_* is linear, and is A-linear in v because τ_* takes values in $\mathrm{Hom}_A(M,N)$, hence it is A-bilinear. From the equations uniquely characterizing the maps at each step, we find that τ is the unique bilinear map taking the desired values on elements of $X\times Y$.

The result on general multilinear maps can be gotten by induction, proving the n-linear case from the $(n-1)$-linear case just as the bilinear case was proved from the linear case above. □

P.136, start of next-to-last paragraph [>]: What Lang calls a **principal** module (using the word that is common in referring to ideals) is more often called a *cyclic* module (generalizing the use of this word in group theory).

P.136, third from last line, ''and the process can be continued'' [>]: What Lang means is that one can map a free module F_1 of finite rank onto M_1, call its kernel M_2, and if this turns out also to be finitely generated, one can map a free module F_2 of finite rank onto it, etc.. Note that in the general step, one is mapping a free module F_i onto the submodule $M_i\subseteq F_{i-1}$; hence the result can be looked at as a map $F_i\to F_{i-1}$. Because the image M_i of each of these maps is the kernel of the next, what we are getting is an exact sequence

$$\ldots\to F_3\to F_2\to F_1\to F\to M\to\{0\}.$$

As Lang notes on the next page, even when one can't get free modules of *finite* rank, one can always get such an exact sequence of free modules of possibly infinite ranks. This is called a *free resolution* of M, and is the starting point of many of the fundamental constructions of homological algebra.

You might find it interesting to look for a case where M is finitely generated, but where, taking F free of finite rank, one gets an M_1, and hence an F_1, that is not finitely generated.

P.137, second paragraph [=]: Lang already defined ''free module'' on p.135 (just before Theorem III.4.1). What he gives here is a way of constructing such a module with basis indexed by any set S. This construction can be motivated using the general heuristics for constructing universal objects that I sketched in my comment to Lang, p.74, bottom. Namely, consider an arbitrary A-module M given with an S-tuple of elements, $(f(s))_{s\in S}$. What elements of M can we name, in terms of these? Clearly, we can name all linear combinations $\Sigma\, a_s f(s)$ where the elements a_s lie in A and all but finitely many of these are zero. Moreover, given two expressions of this form, we can write down the expression of the same form which will represent the sum of the indicated elements of M, and given an expression of this form and an element $a\in A$, we can write down the expression representing the result of multiplying the indicated element of M by a. Now for particular modules M and maps f, certain distinct expressions of this form may describe the same element; but we can't see any equalities that must hold in *all* such cases; so let us take the set of all ''formal expressions'' of this sort, and define addition and scalar multiplication on these in the manner that we have seen works for actual expressions for elements of an actual A-module. We find that the result in fact forms an A-module, which we can describe as a ''module of formal expressions''. Since each such expression is uniquely determined by the S-tuple of coefficients $(a_s)_{s\in S}$, in which we have noted that almost all a_s are zero, we may more simply define our module to consist of all such S-tuples $(a_s)_{s\in S}$. As in Lang, one easily verifies that the set of S-tuples with a 1 in one place and 0's everywhere else forms a *basis*, in bijective correspondence with the given set S, so by Theorem III.4.1, we do indeed have a free module on S.

In our discussion of *free abelian groups*, we saw that all bases had the same cardinality. Is the same true for bases of modules over arbitrary rings? You might think about this. I will discuss the answer in my comments on §III.5.

P.137, discussion of projective modules [<]:

In studying free abelian groups, we noted that Lemma I.7.2 said that every surjective homomorphism of an abelian group to a free abelian group split. The same argument shows that every surjective homomorphism from an A-module to a free A-module splits, and this fact is an important tool in module theory. However, over many rings, there are modules other than free modules for which this also holds, and these may be considered ''as good as'' free modules for many purposes. They are given the name *projective modules*, and introduced here.

P.138, before **Examples** [=]: Recall that abelian groups are essentially the same as $\mathbf{Z}$-modules. Hence, let us ask, ''What are the projective $\mathbf{Z}$-modules?'', and look at this as a question about abelian groups. From condition **P3**, we see that every such module must *embed* in a free $\mathbf{Z}$-module. But we know that every subgroup of a free abelian group is free; so every projective $\mathbf{Z}$-module is free. Since we began by observing that projectivity was a property of free modules, we can conclude that the projective abelian groups are precisely the free abelian groups. In the next paragraph, Lang will mention several other classes of rings for which all projective modules are known to be free, including two large classes for which he will soon prove this, the fields and the principal ideal domains. However, for those of you who are interested, let me give a few examples of

Non-free projective modules (optional for my Math 250A).

First, a simple-minded class of examples. If A and B are any two rings, one can form the product ring $A\times B$. It is not hard to show that any module over $A\times B$ has the form $M\times N$, where M is an A-module, N is a B-module, and the product is made an $A\times B$-module in the obvious way. In particular, the free module of rank r over $A\times B$ has the form $A^r\times B^r$. It is easy to see that a module of the form $A^m\times B^n$ will be a direct summand in the free module $A^r\times B^r$ for any $r\geq\max(m,n)$; but in general, this will not be free; for instance, it will not be free if $A=B=\mathbf{Z}$ and $m\neq n$.

Now for some more interesting examples, though these require some geometric arguments.

Let A be the ring of all continuous real-valued functions on the unit circle $x^2 + y^2 = 1$. Let us think of this circle as lying in 3-dimensional space, and let F be the set of all functions that associate to each point of that circle a vector in 3-space perpendicular to the circle at that point, in a continuous fashion. Since we can multiply such a continuous vector-valued function by a continuous scalar-valued function, F becomes a module over A. In fact, we can see that F is free of rank 2, with one generator given by the function associating to every point of the circle the unit vector in the upwards direction, and the other associating to every point the unit vector in the outward normal direction. Now comes the geometry: Let us draw a ''Möbius strip'' containing this circle, let P_1 be the submodule of F consisting of those elements whose value at every point lies in (the plane determined at that point by) the Möbius strip, and let P_2 be the submodule consisting of elements whose value at every point is perpendicular to the Möbius strip. Then we see that $F = P_1 \oplus P_2$, so that P_1 and P_2 are projective modules over A. But I claim that they are not free. To show this, we first note that if they were free, they would each have to be free on a single generator. For given two elements of one of these modules, one can construct a nontrivial linear relation that they satisfy, with coefficients in A; though the argument requires a bit of (undergraduate) analysis, and I will not go into it here. On the other hand, I claim that neither P_1 nor P_2 can be generated by a single element. For the 180° twist in the construction of the Möbius strip has the consequence that every element x of either of these modules has a point where its value is the zero vector (draw a picture!); and if we take a member y of the given module which is nonzero at that point, we see that y cannot lie in the submodule Ax. So these projective modules can neither be free on one generator, nor free on more than one generator; so they are nonfree.

The above ring A is messy from an algebraic point of view, since it contains many zero-divisors. However, with a little more care one can show that this example can actually be realized over the subring of *polynomial* functions on the circle, i.e., functions which are the restrictions to the circle of polynomial functions on the plane. This ring, $\mathbf{R}[X, Y]/(X^2 + Y^2 - 1)$, can be shown to be an integral domain. (In fact, it is isomorphic to the ring of trigonometric functions on the line generated by the functions $\sin\theta$ and $\cos\theta$. That one can express the ''half-twist'' of the Möbius strip in terms of these functions is a consequence of the fact that $\tan(\theta/2)$ can be written as a rational function in $\sin\theta$ and $\cos\theta$.)

Another geometric example: Suppose we let A be the ring of continuous real-valued functions on the 2-sphere $x^2 + y^2 + z^2 = 1$, and F the A-module of all functions associating in a continuous fashion to each point of the 2-sphere an arbitrary vector in 3-space (which you should picture as drawn at that point of the sphere). Then F is free of rank 3, with a basis consisting of the functions which associate to every point of the sphere the unit vector in the x-direction, respectively the y-direction, respectively the z-direction. Now we can decompose F as $P_1 \oplus P_2$, where this time P_1 denotes the set of functions which are everywhere *tangent* to the sphere, and P_2 the set of functions everywhere *normal* to the sphere. Here, P_2 is free of rank 1, with generator the everywhere-outward-unit-normal function. One can show that if P_1 were free, it would have to be free of rank 2, and that a basis would have to consist of two elements, whose values at each point gave a basis of the tangent plane to the sphere at that point. But this would mean that each of these functions would have the property of associating to every point of the sphere a nonzero tangent vector, in a continuous fashion. Now it is a well-known result of algebraic topology that there is no such function (''You can't comb a hairy sphere''). Hence P_1 cannot be free. In this example, too, the ring of all continuous functions can be replaced by the integral domain of all polynomial functions.

Finally, one quick noncommutative example. Let R be any ring, n a positive integer, and A the ring of $n \times n$ matrices over R. Then the set P of column vectors over R of height n can be considered a left A-module. Note that if we regard A as a left module over itself, multiplication by each member of A acts separately on the n columns of members of A. From this we can see that A is isomorphic to the direct sum of n copies of P. Hence P is projective; intuitively, one may think of it as a ''free A-module of rank $1/n$''.

P.138, **Examples**, through end of section on next page [<]: You may consider that material optional reading. If you do read it, note that at several points (once in each paragraph), Lang uses ''finite'' in the sense ''finitely generated'' which he introduced on p.129, but which I have advised you not to use.

What do we get if we apply the ideas of the second and third paragraphs of this discussion to the geometric examples of nonfree projective modules we have just sketched? It is easy to see that in the ''Möbius strip'' case, P_1 and P_2 are isomorphic to each other. Calling their common isomorphism class p, and writing f for the isomorphism class of the free module of rank 1, it can be proved that all projective modules over this ring are direct sums of copies of these two modules, and that $K(A)$ is the abelian group presented by two generators and one relation, $\langle p, f \mid 2p = 2f \rangle$. (Being a finitely generated abelian group, this must be decomposable as a direct sum of cyclic subgroups. Can you find this decomposition?) We get $K_0(A)$ by dividing out by the subgroup generated by f; the result is $\langle p \mid 2p = 0\rangle \cong Z_2$. In the ''hairy sphere'' example, let me just note that because the direct sum of P_1 with the free module P_2 gives the free module F, the image in $K(A)$ (and hence in $K_0(A)$) of P_1 is the same as that of the free group of rank 2, even though P_1 is not itself free. In the $n \times n$ matrix ring example, if k is a field, one can show (cf. Exercise III.5:1) that $K(A)$ is infinite cyclic with generator $[P]$, and $K_0(A)$ is cyclic of order n.

Lang gives a second definition of $K(A)$ in the very last paragraph. It can be shown that this agrees with the first definition for a large class of rings A, but we can't go into this here. (There is an obvious homomorphism from the first group to the second; the nonobvious result is that for ''good'' A, this is an isomorphism.)

When he gives the second definition of $K(A)$, Lang notes that by modifying the class of modules considered, one gets still other groups. This is true, but one has to be careful. If, for instance, one starts with the class of countably generated A-modules, then one finds that for every module M in this class, the direct sum N of countably many copies of M is also in the class. But $N \cong N \oplus M$, hence M has zero image in this Grothendieck group; hence that group is trivial. Another example of how changing the class of groups can erase structure: If we take $A = \mathbf{Z}$, and consider the class of finite abelian groups, we find that the corresponding Grothendieck group is free of infinite rank, on the generators $[Z_p]$, where p ranges over the primes. On the other hand, in the Grothendieck group obtained from the class of finitely *generated* abelian groups, we have $[Z_p] = 0$, in view of the short exact sequence $0 \to \mathbf{Z} \to \mathbf{Z} \to Z_p \to 0$. From this it is easy to see that all finite groups have zero image in this Grothendieck group, and that it is, in fact, free of rank 1 on the generator $[\mathbf{Z}]$.

P.139, end of §III.4 [>]. *Further notes.* We can *dualize* the definition of projectivity. One calls a module I *injective* if every one-to-one map from I to a module M splits (i.e., is a split injective homomorphism, as defined in the comments on §I.7). This is equivalent to saying that whenever I occurs as a submodule of a module M, it is a direct summand in M. Free modules are not in general injective; e.g., taking $A = \mathbf{Z}$, we note that the free abelian group $\mathbf{Z}$ is a submodule of $\frac{1}{2}\mathbf{Z}$, but not a direct summand therein. Two examples of injective abelian groups (though we will not prove this) are $\mathbf{Q}$ and $\mathbf{Q}/\mathbf{Z}$. Although free modules are usually not injective, it is not hard to prove from Lemma I.8.c1(b) that for $A = \mathbf{Z}/p^n\mathbf{Z}$, free A-modules *are* injective, and from the results of the next section (or undergraduate linear algebra) that over a field, every module is injective. Basic results on injective modules over general rings are developed in Lang, §XX.4, which can be read immediately after the present section. This material is sometimes covered in 250B.

Recall that in §III.2, we saw that covariant and contravariant hom-functors $\mathrm{Hom}(X, -)$ and $\mathrm{Hom}(-, Y)$ on module categories do not generally preserve exactness of (short) exact sequences. In fact, the first functor preserves exactness of all such sequences if and only if X is a projective module (Lang's **P 4**), and the second, if and only Y is an injective module. (This is not hard to prove; Lang gets it as **I 2** on p.783.)

Re §III.5. Vector spaces.

P.139, first line of §III.5 [>]: More generally, a module over a *division ring* is called a vector space, and you will find that all the arguments used in this section work for vector spaces over division rings; so read them under that more general assumption. There *are* important results in the theory of vector spaces over fields that do not hold in this more general context; but these concern determinants of linear transformations, eigenvectors, and similar subjects, not treated in this section.

P.139, statement of Theorem III.5.1 [=]: Lang requires $V \neq 0$ in order to be consistent with his previous definition that a basis must be nonempty. Since we are not requiring this, you should drop that condition from this theorem.

P.140, Proof of Theorem III.5.2 [=]: The argument becomes simpler and more elegant if one takes as the key statement that if $v_1, \ldots, v_m$ span V and $u_1, \ldots, u_n$ are linearly independent, then $n \leq m$. Then one can drop the last seven lines on this page, in which Lang shows that the linear independence condition he is (unnecessarily) assuming on the v's carries over to the modified sequence. (Likewise, on the line after the first display on the next page, one would weaken ''is a basis of'' to ''spans''.) When this inductive argument is complete, one can conclude that if two finite subsets are *both* linearly independent and spanning sets, then their cardinalities must be equal.

P.141, second paragraph [=]: Lang leaves the infinite-basis case ''to the reader''. This is proved by the same method as for free abelian groups. You should verify that the lemma on generating sets of groups given above in the commentary to pp.41-42 also holds for modules over any ring, by the same proof. (The proper statement of that lemma is as a general result of universal algebra.) It follows that if a module is free with bases X and Y, and one of these is infinite, then they have the same cardinality.

P.141, beginning of second line of proof of Theorem III.5.3 [=]: The canonical map Lang is referring to is the map $V \to V/W$.

P.142, end of §III.5 [>]. Now read the items below.

Trouble with trivia.

Recall that when Lang defined a ''basis'' S of an A-module M on p.135, he required that S be *nonempty*. I think most mathematicians would not make such a restriction. When this condition is deleted, one finds that the empty set is a basis of the *zero module* $\{0\}$. (In verifying this, we must be careful and note that when S is empty, the phrase ''linear combination of members of S'' is not vacuous – 0 is a linear combination of members of $\varnothing$, since the empty sum is understood to be 0.) Thus, one does not have to give the statement that the zero module is free as a special proviso in the definition, as Lang does (p.135, last four words before Theorem III.4.1).

Does it matter whether we allow empty bases? Yes! If we do not, then arguments about free modules will continually require consideration of special cases. For example, under Lang's definition, his proof of Theorem III.5.3 (p.141) is incomplete: He fails to note that if f is one-to-one $\operatorname{Ker} f$ will ''not have a basis'', and likewise that if f is the zero map $\operatorname{Im} f$ ''has no basis''. Of course, the proof of the theorem in these special cases is easy, but it's nicer not to have to deal with them. The difference is all the greater in more complicated proofs, involving, for instance, infinite families of vector spaces, any of which might be zero. Moreover, if we prove a result like Theorem III.5.3 by giving separate arguments for the case where f is one-to-one or zero, this suggests that it may just be a coincidence that the same result holds in those cases as in the general case. A proof based on the correct definition makes it clear that it is not.

I think the reason Lang and some other textbook-writers specify that bases must be nonempty is the awareness that if they do not, many students will have trouble figuring out how the definition applies to empty linear combinations. But I think the best approach is for a textbook to give the *definition* without special treatment of the zero module, then to *point out*, after the definition, how this makes $\varnothing$ a basis of the zero module; and, finally, to advise any readers who have difficulty with this case to take the statement that the empty set is a basis of the zero module as an addendum to the definition, for the time being.

(Curiously, in discussing free groups in Chapter I, Lang does not require his free generating sets to be nonempty, though he does in the case of free abelian groups.)

Free modules with and without unique rank.

We can now address the question mentioned earlier, of whether a free module over a ring R can have bases of two different cardinalities. By our comment on Lang's p.141, this is only possible if both cardinalities are finite.

To understand the question computationally, note that a free left R-module of rank n is isomorphic to the ''standard'' free module R^n. Let us write the elements of the canonical basis of R^n, (1,0,0,...), (0,1,0,...), ... (0,...,0,1), as $e_1, \ldots, e_n$. A module homomorphism $f: R^m \to R^n$ is uniquely determined by the images of the m elements $e_i \in R^m$, which can be any elements $f(e_i) = \Sigma_{j \leq n}\ a_{ij} e_j$ $(a_{ij} \in R)$. Note that when we apply this homomorphism f to a general element $\Sigma_{i \leq m}\ c_i e_i \in R^m$, we get $\Sigma_{i,j}\ c_i a_{ij} e_j$, and computationally, the effect is that of multiplying a row vector $(c_1, \ldots, c_m)$ on the right by an $m \times n$ matrix $((a_{ij}))$. (Cf. the ring theorist's principle that homomorphisms of left modules should be written on the right of elements of these modules.) It is easy to verify that composition of homomorphisms among such free modules corresponds to multiplication of matrices over R. We deduce

Lemma III.5.c1. *Let R be a ring, and m, n positive integers. Then the following conditions are equivalent:*

(i) *The free left module of rank m is isomorphic to the free left module of rank n.*

(ii) *There exist an $n \times m$ matrix A and an $m \times n$ matrix B such that*

$$AB = I_n, \quad BA = I_m. \ \square$$

Now given any homomorphism of rings, $f: R \to S$, we can apply f to the entries of any family of matrices over R, and get matrices over S that satisfy any matrix relations that our original matrices did. From this observation we get

Corollary III.5.c2. (a) *If $R \to S$ is a homomorphism of rings, and m, n positive integers such that $R^m \cong R^n$ as R-modules, then $S^m \cong S^n$ as S-modules.*

(b) *If R is a ring which can be mapped homomorphically into a division ring, then any two bases of a free R-module have the same cardinality.*

In particular

(c) *If R is a nonzero commutative ring, then any two bases of a free R-module have the same cardinality.*

Proof. We get (a) by applying the given homomorphism to the entries of matrices as in point (ii) of the preceding lemma.

To see (b), let X and Y be two bases of a free R-module F. We have already seen that if either is infinite, then they have the same cardinality, so assume both finite; say X has cardinality m and Y has cardinality n. Since any two free modules on the same number of generators are isomorphic, we have $R^m \cong F \cong R^n$. Now if R has a homomorphism to a division ring D, then by part (a), this gives a D-vector-space isomorphism $D^m \cong D^n$. But by Theorem III.5.2 (which we have noted is valid for vector spaces over division algebras), this implies $m = n$, as claimed.

Finally, if R is a nonzero commutative ring, then letting I be a maximal ideal of R, we know that R/I is a field; so we can apply (b) to the canonical homomorphism $R \to R/I$ and conclude that bases of free R-modules have unique rank. $\square$

Now the result we just used, that for any maximal ideal I of a commutative ring, R/I is a field, does *not* extend to say that for any maximal ideal I of any ring, R/I is a division ring. Thus, though the results in Lang are as true for vector spaces over general division rings as for vector spaces over fields, their consequences for commutative rings are stronger than for general rings! Indeed, we shall now show that there exist noncommutative rings having free modules with nonunique rank.

To do this, we must first generalize our observation about writing homomorphisms among free modules as matrices. Note that by the universal properties of products and coproducts in any category, a morphism from a coproduct to a product, $a\colon \coprod_I X_i \to \prod_J Y_j$, is uniquely determined by the $I\times J$-tuple of morphisms a_{ij} gotten by composing a on the right with the universal morphisms (coprojections) from the X_i to $\coprod_I X_i$, and on the left with the universal morphisms (projections) from $\prod_J Y_j$ to the Y_j. In particular, since products and coproducts of finite families of modules are both direct sums, in the module case this gives a description of morphisms $a\colon \oplus_I M_i \to \oplus_J N_j$ as ''matrices'' of morphisms $a_{ij}\colon M_i \to N_j$; and it is easy to verify that these compose by formal composition of matrices. (This description uses the fact that module maps can be added as well as composed.) We can now prove

Lemma III.5.c3. (a) *Let A be any ring, M a left A-module, and m and n positive integers such that $M^m \cong M^n$ as left A-modules. Then, letting R denote the ring* $\mathrm{End}_A(M)$, *we have $R^m \cong R^n$ as left R-modules.*

In particular, as a concrete class of examples

(b) *If R is the full endomorphism ring of an infinite-dimensional vector space V, then $R^m \cong R^n$ as left R-modules for all positive integers m and n.*

Proof. (a): The given isomorphism $M^m \cong M^n$ will be described by an $m\times n$ matrix and an $n\times m$ matrix of module homomorphisms $M \to M$, i.e., elements of R, which compose to the identity matrices. Hence by Lemma III.5.5, above $R^m \cong R^n$ as left R-modules.

(b): Since any infinite cardinal α satisfies $m\alpha = n\alpha$ for all positive integers m and n, we see (by working with bijections of bases) that infinite-dimensional vector spaces V satisfy $V^m \cong V^n$, from which the assertion follows by part (a). (We could equally well have stated this result for endomorphism rings of free modules of infinite rank over an arbitrary ring.) □

The above result was easy to prove. There are more delicate results that are harder, such as that there exist rings R such that all R-modules of *even* finite positive dimension are mutually isomorphic, but are not isomorphic to free modules of *odd* dimension. Cf. P. M. Cohn, *Some remarks on the invariant basis property,* Topology, **5** (1966) 215-218.

The functorial approach (sketch, optional). You may wonder why matrices were needed to get the above results. There is in fact an approach I think more enlightening, though it involves some material we have not developed. If $f\colon R \to S$ is a ring homomorphism, then one can regard any left S-module M as a left R-module by defining $r\cdot x = f(r)x$ $(r\in R,\ x\in M)$. The resulting functor S-**Module** $\to$ R-**Module**, called ''restriction of scalars from S to R'', has a left adjoint, called ''extension of scalars from R to S''. (Lang introduces this construction in §XVI.4, though only for R and S commutative.) Now note that the composite of the above restriction of scalars functor with the forgetful functor R-**Module** $\to$ **Set** is the forgetful functor S-**Module** $\to$ **Set**. Since adjoints of composite functors are composites of adjoint functors, with order of composition reversed, we see that the free R-module functor **Set** $\to$ R-**Module**, followed by the extension of scalars functor R-**Module** $\to$ S-**Module**, gives the free S-module functor **Set** $\to$ S-**Module**. In particular, the extension of scalars functor takes the free R-module of any rank n to the free S-module of rank n. Hence if there is an isomorphism between the free R-modules of ranks m and n, there will be an isomorphism between the free S-modules of ranks m and n, as described in Corollary III.5.6(a) above. The trick of applying the ring homomorphism $R \to S$ to entries of matrices simply mimics the way this functor acts on morphisms among free modules.

The non-unique-ranks construction of Lemma III.5.c3(b) also has a functorial interpretation, this time not even requiring adjoint functors. Observe that a left A-module M has a natural right $\mathrm{End}({}_AM)$-module structure. Using this, we can make $\mathrm{Hom}_A(M,N)$ into a left $\mathrm{End}(_A M)$-module for each left module N, in a functorial manner. Clearly, when $N = M$, this module is the free module of rank 1, from which it can be deduced that, writing $R = \mathrm{End}(_A M)$, the functor $\mathrm{Hom}_A(M,-)$ takes M^n to the free left R-module of rank n. Hence if $M^m \cong M^n$ as A-modules, we get $R^m \cong R^n$ as R-modules.

Incidentally, the trick used in showing that over a commutative ring, free modules of different ranks were nonisomorphic is an instance of a method common in modern mathematics: *To show objects are nonisomorphic, find a functor taking them to objects you can already prove nonisomorphic.* We used this earlier to show free abelian groups of distinct ranks nonisomorphic. Though we did not have the theory of dimension of vector spaces available at the time, we saw that the groups one got by applying to such groups the functor $A \mapsto A/pA$ could be distinguished by their *orders*. Another standard use of the same method is in showing that the free (nonabelian) group of ranks m and n cannot be isomorphic unless $m = n$ (which was actually never shown in §I.12). To do this one applies the *abelianization* functor $G \mapsto G/G^c$. It is easy to verify by universal properties (or by Lemma II.3.c3) that the abelianization of a free group of rank n is a free abelian group of rank n; hence, since we have shown that free abelian groups have unique rank, the same is true of free groups.

Let me mention an interesting pair of open questions on bases of vector spaces.

Open Questions III.5.c4. *Let k be a field, n a positive integer, and V an n-dimensional vector space over k.*

(a) (J. Kahn) *Suppose we are given n^2 bases $\mathcal{B}_{11}, \mathcal{B}_{12}, \ldots, \mathcal{B}_{nn}$ of V (not necessarily distinct. Think of these as written in the cells of an $n\times n$ array.) Is it possible to choose from each $\mathcal{B}_{ij}$ an element b_{ij}, in such a way that for each i, $\{b_{ij} \mid j = 1,\ldots,n\}$ is a basis of V, and for each j, $\{b_{ij} \mid i = 1,\ldots,n\}$ is a basis of V? (I.e., can we choose one element from each cell so that in each row, and in each column, the chosen elements form a basis of V?)*

(b) (G.-C. Rota) *Same question, under the additional assumption that for each i, all the $\mathcal{B}_{ij}$ are the same. (So this question asks whether, given n bases $\mathcal{B}_i$ of V, we can recombine them into n bases $\mathcal{B}'_j$, each using one element of each $\mathcal{B}_i$, so that each element of each $\mathcal{B}_i$ is used in only one $\mathcal{B}'_j$.)*

(See R. Huang and G.-C. Rota, *On the relations of various conjectures on Latin squares and straightening coefficients,* Discrete Math., **128** (1994) 225-236.)

Re §III.6. The dual space and dual module.

P.142, first paragraph of §III.6 [>]: Note that the duality $E^\vee$ studied in this section is quite different from the duality $E^\wedge$ of §I.9. In the last sentence of this paragraph, Lang makes two claims which he does not justify: that the map $E \to E^{\vee\vee}$ is always injective, and that it is not always surjective. I will show both facts in the comment to be read following the proof of Theorem III.6.1.

P.143, statement of Theorem III.6.1 [=]: In the first line, delete the word ''finite''. (Lang is using the word in the sense ''finitely generated'', but it is redundant in view of the next line.) In the third line, before ''f_i is'', add the words ''*for each* i''.

P.143, first display [=]: Add, on the right, ''$= a_1 c_1 + \ldots + a_n c_n$''. Also, in the line after the second display, ''K'' should be ''A''.

P.143, end of proof of Theorem III.6.1 [>]. We can now show the two facts asserted at the bottom of p.142. Let us write $\theta\colon E \to E^{\vee\vee}$ for the map described there, let $(x_i)_{i\in I}$ be a basis of E, and let (f_i) be defined as in Theorem III.6.1.

To show θ is injective, consider an element $x \in E$, written in terms of the x_i. If $x \neq 0$, there will be some $i \in I$ such that x_i occurs with nonzero coefficient a_i in this expression. Thus, $\theta(x)(f_i) = f_i(x) = a_i \neq 0$, showing that $\theta(x) \neq 0$.

On the other hand, I claim that when A is a field and E is infinite-dimensional, θ is not surjective. We first note that the argument used in the proof of Theorem III.6.1 still shows that the family (f_i) is linearly independent (though the proof that if E has finite rank that set generates $E^\vee$, which involves taking a summation over all i, does not work in the non-finitely-generated case). Hence we can enlarge this family to a basis $\mathcal{B}$ of the vector space $E^\vee$. Let us now define $\varphi\in E^{\vee\vee}$ to be the unique linear map $E^\vee\to A$ which takes all elements of $\mathcal{B}$ to 1. If φ were of the form $\theta(x)$ $(x\in E)$, it would take each f_i to the coefficient of the basis element x_i in the expression for x, so it would take on nonzero values at only finitely many of the elements f_i. Thus, φ cannot have the form $\theta(x)$, so θ is not surjective.

Note that for any base ring A, if E is free on a basis I, then $E^\vee$ is isomorphic to an I-fold direct *product* of copies of A. Thus, our old Exercise I.7:1 effectively asked you to show that for E free, $E^\vee$ may fail to be free.

P.143, statement and proof of Corollary III.6.2 [=]: In the second line of the statement, ''V'' should be ''E''. The proof is sloppily worded. Again writing $\theta\colon E\to E^{\vee\vee}$ for the map referred to in the corollary, one should replace $\{x_1,\ \ldots,\ x_n\}$ in the second line by $\{\theta(x_1),\ \ldots,\ \theta(x_n)\}$, and ''$E = E^{\vee\vee}$'' by ''$\theta\colon E\to E^{\vee\vee}$ is an isomorphism''.

P.143, Theorem III.6.3 [=]: The assumption that U, V and W are free is much stronger than we need; we only need an assumption on U, and this assumption is that it be projective (which is weaker than free). From that condition, one sees that the given exact sequence splits, and, as we noted in our comments on split exact sequences under §I.7, this makes the sequence a ''direct sum'' of the two exact sequences $0\to W\to W\to 0\to 0$ and $0\to 0\to U\to U\to 0$. It is not hard to see that $\mathrm{Hom}(-,A)$ preserves direct sums, and therefore preserves exactness of such sequences, proving the theorem. In fact, this argument shows that the conclusion can also be strengthened, replacing $\mathrm{Hom}(-,A)$ by $\mathrm{Hom}(-,M)$ for any module M.

P.145, **Remark 2** [<]: Let me precede this with some remarks of my own.

The context in which Lang introduces this duality is unnecessarily restricted. First, the functor $\mathrm{Hom}(-,\ A)$ can be applied to arbitrary modules, not just free modules. (Indeed, in the last sentence of the first paragraph, Lang applies $^\vee$ to $E^\vee$, which may not be free if E is free of infinite rank.) Although this functor is ''uninteresting'' when applied to certain non-free modules (e.g., when $A=\mathbf{Z}$ and E is either a torsion abelian group or $\mathbf{Q}$, then $E^\vee=0$), there are other non-free modules for which it is important. When it is applied to finitely generated *projective modules*, one can show that the analogs of all the theorems of this section hold. You might also find it interesting to examine the case where $A=k[X,Y]$, a polynomial ring over a field, and E is the ideal (X,Y); in particular, to compute $E^\vee$, $E^{\vee\vee}$, and the natural map $\theta\colon E\to E^{\vee\vee}$.

Lang has taken the base-ring A to be commutative. If this is not assumed, and E is a left A-module, it turns out that the abelian group of left A-module homomorphisms $E^\vee=\mathrm{Hom}(E,\ A)$ has a natural structure of *right* A-module; and likewise, the abelian group of right-module homomorphisms from a right A-module into A has a structure of *left* A-module. (More details in my comments in §XIII.1.) Thus, one gets a duality between right and left modules, which, when these modules are finitely generated and free (or projective), still has all the properties Lang proves in the commutative situation.

In **Remark 2** on this page, Lang notes a different kind of duality, taking an R-module E to the set $E^\wedge=\mathrm{Hom}(E,\mathbf{Q}/\mathbf{Z})$ of *abelian group* homomorphisms $E\to\mathbf{Q}/\mathbf{Z}$. What he fails to mention is that this set can be made an R-module in a natural way! For instance, if E is a left R-module; then $E^\wedge$ becomes a right R-module by defining, for each $f\in E^\wedge$ and $r\in R$, the element $fr\in E^\wedge$ to be the map that acts by $fr(x)=f(rx)$. Here too, we get a natural map $E\to E^{\wedge\wedge}$. This is always one-to-one, but rarely onto. (It is onto when E is finite in the set-theoretic sense, by the results of §I.9.)

For this duality by $\mathbf{Q}/\mathbf{Z}$, and also for duality of vector spaces by the construction of this section, the failure of surjectivity of the embedding of an object E in its double dual can be considered to result from the existence of ''strange'' functionals on the dual. In some contexts it is possible to exclude these by introducing a topology on the dual, and restricting attention to continuous functionals. When one does this for the $E^\wedge$ construction, it is also natural to replace $\mathbf{Q}/\mathbf{Z}$ by $\mathbf{R}/\mathbf{Z}$. For any topological abelian group

E one then defines $E^\wedge$ to consist of all *continuous* homomorphisms $E \to \mathbf{R}/\mathbf{Z}$, one puts a natural topology on this group, and one shows that if E is *locally compact*, then so is $E^\wedge$, and the natural map $E \to E^{\wedge\wedge}$ is then an isomorphism of topological groups. (This is Pontrjagin duality, mentioned by Lang in the next paragraph; cf. Walter Rudin *Fourier Analysis on Groups*, Interscience, 1962.) One finds, inter alia, that under this duality, the compact abelian groups are the duals of the discrete abelian groups; for example, this duality takes the short exact sequence $0 \to \mathbf{Z} \to \mathbf{R} \to \mathbf{R}/\mathbf{Z} \to 0$ to itself, interchanging the right- and left-hand objects. Similarly, for a vector space E over an arbitrary field k, one can put a natural topology on $E^\vee = \mathrm{Hom}(E, k)$, such that for ''good'' E, the natural map $E \to E^{\vee\vee}$ is an isomorphism of topological vector spaces. In particular, discrete spaces are ''good'', and are dual to what are called ''linearly compact'' spaces. This is ''Lefschetz duality'', but is not so well known.

Re §III.7. Modules over principal rings.

I motivated the structure theory of finitely generated abelian groups in my comments on §I.8. Since the theory of modules over principal ideal domains closely parallels that theory, I will not repeat that discussion here. Though Lang generally leaves the non-finitely-generated case to the reader, or refers to an Appendix, it is treated in detail in those comments.

P.147, Corollary III.7.2 [>]: Lang indicates two ways of proving this. Actually, these yield different results, each stronger than the result he states. The proof he writes out shows that if E is a module over a PID which can be generated by n elements, then any submodule of E can be generated by $\leq n$ elements. The other proof shows that any submodule of a finitely generated module over any *Noetherian ring* is finitely generated (but not necessarily by fewer generators than the given module. E.g., if R is the polynomial ring $k[x, y]$ in two indeterminates over a field, and $E = R$, the free module of rank 1, then the submodule generated by x and y cannot be generated by fewer than two elements, the submodule generated by x^2, xy and y^2 cannot be generated by fewer than three elements, etc..)

P.148, first paragraph [~]: This argument shows that any finitely generated torsion-free module M over an integral domain R embeds in a free module of finite rank. When R is, in fact, a principal ideal domain, it follows that M is free, which is the fact he will use next.

P.148, Lemma III.7.4 [~]: Lang is really re-proving the fact that a free module satisfies condition **P2** (p.137), i.e., is projective.

P.149, bottom, first sentence of proof [<]: If we take a finite generating set for E, we may choose a nonzero exponent for each member of this set, and take a common multiple $a \in R$ of these elements. Thus a will annihilate all members of our generating set, hence all members of E, justifying this sentence.

P.151, statement of Theorem III.7.7 [>]: Note where Lang above the display refers to the $R/(q_i)$ as nonzero, this is equivalent to saying that the q_i are nonunits.

P.151, next-to-last line of text [=]: After ''$i = 1, \ldots, l$'' add, '', with some of these rows 'padded' with zeros on the left, if necessary, so that they all have the same length r as the longest row''.

P.153, Theorem III.7.8 [<]: Any finitely generated R-module E can be written as a homomorphic image of a free module F of finite rank, hence as a factor-module of F by a submodule M. We have just shown that E can be written as a direct sum of cyclic modules $R/(a_i R)$. Can we write F as a direct sum of free modules of rank 1, Rx_i, so that M is the sum of the submodules $Ra_i x_i$; i.e., can we ''realize'' the decomposition of Theorem III.7.7 via the given presentation of E? A positive answer is given by the next result, which is thus a strengthening of Theorem III.7.7.

Incidentally, in the statement of the theorem, the condition $M \neq 0$ is pointless. The result is clearly true for $M = 0$, and the induction with which the proof ends on the next page implicitly uses that fact, since M_1 may be 0.

Pp.153-154, sentence on bottom of one page and top of next [=]: To prove this assertion, note that because R is a principal ideal domain, the ideal $(g(x_1)) + (\lambda_1(x_1))$ can be generated by one element, say

$$(ug(x_1) + v\lambda_1(x_1)) \ = \ (g(x_1)) + (\lambda_1(x_1)).$$

But this implies

$$(ug+v\lambda_1)(M) \ \supseteq \ ((ug+v\lambda_1)(x_1)) \ = \ (g(x_1)) + (\lambda_1(x_1)) \ \supseteq \ (\lambda_1(x_1)) \ = \ \lambda_1(M)$$

(where the middle step is the preceding display). Now by maximality of $\lambda_1(M)$, the two inclusions must be equalities; the second one now shows that $g(x_1) \in (\lambda_1(x_1)) = (a_1)$, as Lang claims.

P.154, Theorem III.7.9, line 1, and preceding sentence [=]: Precisely, an ''elementary matrix'' means a matrix which differs from the identity matrix *either* by changing one off-diagonal entry 0 to a nonzero value, or by replacing one diagonal entry 1 to another unit of the ring. (Lang has a definition of the term on p.540 that leaves out the latter sort of element. I haven't checked whether that is what he really intends there; but here, elements of both sorts are needed to make the condition of generating $GL_n(R)$ plausible.)

You should verify that multiplying a matrix A on the left by an elementary matrix has the effect of a row operation on A, while multiplying on the right by an elementary has the effect of column operation. The operations of interchanging two rows and columns are not themselves of this form, but each can be expressed as the product of a small number of such operations. Thus, the assumption in the first sentence of the theorem implies that any operation of left or right multiplication by invertible matrices is equivalent to a finite number of row, respectively, column, operations.

In fact, if we delete that first sentence from the theorem, the result can still be made true by changing the phrase ''*with a finite number of row and column operations,*'' to ''*by multiplying on the right and on the left by invertible square matrices*''.

Incidentally, note that the matrix in the statement of the theorem is not assumed square.

P.155, end of §III [=]: The final paragraph is a technical remark which you can ignore.

Though the results of this section classify the structures of all finitely generated modules over PID's, if one drops the hypothesis of finite generation, one finds oneself in a very nontrivial area – even if one limits attention to torsion-free modules of finite rank (i.e., having no infinite set of linearly independent elements; equivalently, embeddable in a finite-dimensional vector space over the field of fractions of R). Such modules (over Dedekind domains, a generalization of principal ideal domains) are the subject of a book in preparation by E. Lee Lady, which can be viewed at `https://www.math.Hawaii.Edu/~lee/book`

Re §III.8. Euler-Poincaré maps.

P.156, top, italicized condition [~]: This condition is satisfied by various natural invariants – the dimension of a finite-dimensional vector space; the order of a finite group (regarded as a member of the multiplicative group of rationals), etc.. Hence one gives a name to an arbitrary function satisfying this condition, and studies such functions in the abstract.

P.156, italicized sentence near bottom [=]: Lang hasn't said what he means by ''are equivalent''; he means ''have, up to isomorphism, the same factors, with the same multiplicities, but possibly in a different order''. The Jordan-Hölder Theorem, Theorem I.3.5, is a consequence of the Schreier Refinement Theorem, which was proved on p.22 – for groups, not modules. However, the construction in the proof used only joins, intersections, and canonical isomorphisms, all of which respect module structures if these are present. Hence if we apply this proof to towers (which Lang here calls finite filtrations) of modules, regarded as abelian groups, then the isomorphisms obtained are module isomorphisms. So the Schreier Refinement Theorem also holds for modules; hence so does the Jordan-Hölder Theorem.

Re §III.9. The Snake Lemma.

P.158, statement of **Snake Lemma** [=]: The expression $f^{-1} \circ d \circ g^{-1} z''$ gives an element of N'; what Lang means, however, is the image of that element in $\mathrm{Coker}\ d'$. (One could say that there should be an additional map at the other end of that composite as well, the inclusion of $\mathrm{Ker}\ d''$ in M''. However, it is common to leave inclusion maps unwritten, but not, in general, to leave factor-maps unwritten.)

Re §III.10. Direct and inverse limits. This is an important subject, begun in §I.10. As with that section, I haven't so far had time to treat this material when I have taught Math 250, but Bjorn Poonen did cover it in Fall 2003, and gave me some errata; and I present these errata and some further observations I made in checking those errata out.

P.160, definition of **directed family of morphisms** [~]: Recall that in §I.10, Lang used ''directed family'' as short for ''inversely directed family'', the structure from which we construct inverse limits. Here, however, he begins with the concept relevant to direct limits and (in agreement with common usage) calls this a ''directed family''; so the phrase can no longer be used in the sense used previously. In the bottom paragraph of p.161, beginning ''We now reverse the arrows'', the resulting system is an inversely directed family, though he does not name it here.

P.162, fourth line of **Example**, and line below last display [=]: Change ''projective limit'' to ''inverse limit''.

Chapter IV. Polynomials.

Re §IV.1. Basic properties of polynomials in one variable.

P.174, last display and preceding line [=]: The holds by the italicized result on the lower part of p.100, ''*If we assume that at least one of the leading coefficients ...*''.

P.175, end of Proof of Theorem IV.1.2 [>].

Euclidean rings.

The technique used in proving the above theorem is called the ''Euclidean algorithm'' because it goes back to Euclid's technique for finding the greatest common divisor of two positive integers $a_1 > a_2$: subtract from a_1 that multiple of a_2 that will leave a nonnegative remainder $a_3 < a_2$; verify that a_2 and a_3 will have the same greatest common divisor (i.e., will generate the same ideal) as a_1 and a_2, and, if a_3 is again positive, repeat the process with a_2 and a_3, till one eventually gets a zero term, $a_{n+1} = 0$. Then the preceding term a_n must generate the ideal $a_1\mathbf{Z} + a_2\mathbf{Z}$. If one merely wants to prove that this ideal is principal, this argument can be shortened: Let a be a least positive element of this ideal, and prove that it must be a generator.

As the proof in Lang shows, a similar argument (due to Simon Stevin, *Arithmétique*, 1585) can be applied to polynomials, using the degree-function. In fact, there is a wide class of rings in which this method can be used, characterized in

Definition IV.1.c1. *If R is an integral domain, a* Euclidean function *on R means a set-map φ from $R - \{0\}$ to the nonnegative integers, such that*

(i) *if $a, b \in R - \{0\}$, then $\varphi(ab) \geq \varphi(b)$.*

(ii) *if $a \in R$, $b \in R - \{0\}$, then there exist $q, r \in R$ such that $a = qb + r$ and either $r = 0$ or $\varphi(r) < \varphi(b)$.*

An integral domain which admits a Euclidean function is called a Euclidean domain *or* Euclidean ring.

Mimicking the proof that $\mathbf{Z}$ and $K[X]$ are PID's, one gets

Theorem IV.1.c2. *Every Euclidean domain is a PID.* □

Another ring easily shown to be Euclidean, and hence a PID, is $\mathbf{Z}[i] \subseteq \mathbf{C}$, via the norm function $\varphi(m+ni) = m^2 + n^2$. (The conclusion of condition (ii) above is equivalent to the statement that there exists a $q \in \mathbf{Z}[i]$ whose distance from a/b is <1. To prove this, verify, more generally, that for any complex number α there is a $q \in \mathbf{Z}[i]$ whose distance from α is $\leq 1/\sqrt{2}$.)

In fact, the Euclidean algorithm is essentially the only ''easy'' way known to show that a ring is a PID. (What about using the fact that a localization $S^{-1}R$ of a PID R is a PID? Well, if R was proved a PID with the help of a Euclidean function φ, then a localization $S^{-1}R$ will also be Euclidean, via the function that takes each element x to the least of the values $\varphi(a)$ as as^{-1} runs over all expressions for x.) However, the handout ''A principal ideal domain that is not Euclidean'' shows how to get one example not of that form.

There are some generalizations of the definition of Euclidean ring, which still allow one to prove that the ring in question is a principal ideal domain. In particular, one may allow φ to have range in a general *well-ordered* set, rather than the nonnegative integers; but I don't know any cases where this generalization can be used but the standard concept cannot. Condition (i) may also be dropped from the definition; however, the class of rings admitting a function φ satisfying this weakened condition will actually be the same as the class admitting a φ that satisfies the original condition. For if φ satisfies (ii), then the function φ' defined by letting $\varphi'(a)$ be the minimum, over all nonzero $c \in R$, of $\varphi(ca)$, will satisfy both (i) and (ii).

P.175, middle, definition of **irreducible** [=]: This term was defined on p.111 (beginning of §II.5), and does not need a separate definition for polynomials. What Lang should say here is that for k a field, a polynomial $f(X)$ will be irreducible if and only if it has the stated property, since in $k[X]$, the nonzero constants are the invertible elements.

P.176, Corollary IV.1.5. [=]: Delete the final phrase ''*i.e., f induces the zero function*''. (It is true, but not the meaning of what precedes.)

P.176, next-to-last line, ''... the same function f ...'' [=]: The fact being used here is that a field k of q elements satisfies the identity $x^q = x$. This was proved in the case $k = \mathbf{Z}/p\mathbf{Z}$ in the second **Example** on p.99. You should check that the same argument works for any finite field.

P.177, Proof of Theorem IV.1.9 [<]: Actually, Lang did most of the work needed here when he proved Proposition I.4.3(vi), p.25, so he should have called on that result in this proof! By that result, it suffices to show that no U contains no more than p roots of $X^p - 1$ for any prime p; but this is immediate by the preceding theorem.

P.177, Theorem IV.1.9 [>]: Note that this shows that if $p < q$ are primes, then $\mathbf{Z}_q^*$ has at most one subgroup of order p. This is the statement which we saw, in our discussion of groups of order pq, was needed to deduce that there is no more than one isomorphism class of nonabelian groups of that order. So at last that argument is complete!

P.178, second sentence of third paragraph [>]: In fact, Lang will prove that $\mathbf{C}$ is algebraically closed in Example 5 on p.272.

P.178, third-from-last display [=]: The first formula is clear. To verify the second formula by writing f and g as sums of monomials, and computing the two sides, is tedious and unenlightening. A nicer way is to observe that the maps $(f,g) \mapsto (fg)'$ and $(f,g) \mapsto f'g + fg'$ are both *bilinear* maps of A-modules, hence they will be equal if they agree as f and g range over the spanning set of powers of X (cf. lemma in our note to p.135 of Lang). Thus, to prove that equation, it suffices to prove it in the case $f = X^m$, $g = X^n$, a quick computation.

P.179, last paragraph (excluding last partial sentence) [>]: Here properties of the pth-power map on a ring of characteristic p are obtained using the fact that certain binomial coefficients are divisible by p. In fact, this property of the binomial coefficients can be proved from a result we already know about the pth-power map on such a ring! Namely, we have seen (using properties of the group of units) that in $\mathbf{Z}/p\mathbf{Z}$, the pth power map is the identity. It follows that the polynomial $(x+1)^p - x^p - 1$ over that field, which has degree at most $p-1$, and whose coefficients are these binomial coefficients modulo p, has all p elements of this field as roots. Since the number of roots is greater than the degree, all coefficients of this polynomial must be zero.

(There are other interesting ways of showing that these coefficients are divisible by p. For instance, the νth coefficient is the number of ν-element subsets of $\{1,..., p\}$. Now the cyclic permutation $(1,..., p)$ has order p, and if ν is neither 0 nor p, it clearly acts without fixed points on the set of ν-element subsets of $\{1,..., p\}$. Hence the number of such subsets must be divisible by p.)

Pp.179-180, sentence on bottom of one page and top of next [=]: The *Frobenius endomorphism* should be defined to be the map $x \mapsto x^p$. The maps $x \mapsto x^{p^r}$ are the powers of the Frobenius. (When one considers k-algebras A for k a field of p^r elements, one sometimes calls the map $x \mapsto x^{p^r}$ the Frobenius endomorphism of A, since it, and not $x \mapsto x^p$, is an endomorphism of A as a k-algebra. But in that usage, there is a particular power of p relevant to the situation, not an arbitrary power, as Lang's wording suggests. I will always use ''Frobenius endomorphism'' to mean the map $x \mapsto x^p$.)

P.180, end of §IV.1 [>].

The Jacobian conjecture (optional).

There is a longstanding open problem on polynomial algebras, which this is a natural time to mention.

We have seen that the operation of *differentiation* of polynomials, which we learned in calculus, can be carried over formally to a function $D: A[X] \to A[X]$ for any commutative ring A, which has the algebraic properties of differentiation, though it is not defined using limits. Note, incidentally, that it is not an algebra homomorphism. Like such a homomorphism, it respects the A-module structure, but instead of the law $h(st) = h(s)h(t)$, it satisfies $D(st) = D(s)t + sD(t)$. A function from an A-algebra into itself with these properties is called a *derivation* over A.

We can likewise define *partial differentiation* operators $D_X, D_Y: A[X, Y] \to A[X, Y]$. These are each derivations over A, and they commute with one another: $D_X D_Y = D_Y D_X$.

Let us now recall that a pair (f, g) of real-valued functions on the plane $\mathbf{R}\times\mathbf{R}$ is equivalent to a single map of the plane into itself, namely $(x,y) \mapsto (f(x,y), g(x,y))$, and that if f and g are differentiable, then the information about this map analogous to the information given by the derivative of a function of one variable is given by the *Jacobian matrix*,

(c24)
$$\begin{pmatrix} D_X f & D_Y f \\ D_X g & D_Y g \end{pmatrix}.$$

In particular, this satisfies a *chain rule* saying that the Jacobian matrix of a composite of two maps is the product of the Jacobian matrices of the two maps. (If the two maps correspond to pairs of functions (f, g) and (h, i) respectively, then their composite corresponds to $(f(h, i), g(h, i))$, and the statement of the chain rule is, precisely, that the Jacobian matrix of the composite map, evaluated at a point (x, y), is the product of the Jacobian matrix of (f, g) evaluated at the image of (x, y) under (h, i), and the Jacobian matrix of (h, i) evaluated at (x, y) itself.)

This suggests that we should think of a pair (f, g) of elements of $A[X, Y]$ as the ''algebraic analog'' of a map of the plane into itself, and define the Jacobian matrix of such a pair by (c24), using our algebraically defined partial differentiation operations. If we define the ''composite'' of two such pairs (f, g) and (h, i) as the pair $(f(h, i), g(h, i))$, it is not hard to prove the chain law for these Jacobian matrices. Note that the pair (X, Y) is a neutral element for this composition, and has for Jacobian matrix the identity matrix.

There is also an algebraic interpretation of such pairs: Any A-algebra endomorphism α of the polynomial ring $A[X, Y]$ is determined by the pair of elements $(\alpha(X), \alpha(Y))$, which we can denote (f, g), and conversely (by the universal property of polynomial rings), every such pair corresponds in this way to an endomorphism. Composition of endomorphisms corresponds to substitution of one pair into another, although the order of composition is the reverse of the order we get when composing maps of the plane. (This is because there is a contravariant relation between points of the plane and functions on the plane.)

Given a pair (f, g) (equivalently, an endomorphism of $A[X, Y]$), it is not clear how to tell whether it is *invertible* under the above composition, i.e., corresponds to an A-algebra *automorphism* of $A[X, Y]$. But we can give a necessary condition: by our above observations on the chain rule, if such an endomorphism is invertible, the Jacobian matrix of the pair of functions must be invertible as a matrix of elements of $A[X, Y]$. This is equivalent to saying that the *determinant* of the Jacobian matrix is an invertible element of $A[X, Y]$.

Let us take for A a field k. Then the invertible elements of $k[X, Y]$ are the nonzero elements of k; so if an automorphism is invertible, the determinant of its Jacobian matrix will be nonzero scalar. Is the converse true? It definitely is not if k has finite characteristic p, for in this case the pair $(X^p + X,\ Y)$ has identity Jacobian matrix, but the corresponding endomorphism is not invertible. So assume k has characteristic 0. To prove the desired converse, it would suffice to prove it in the case where the determinant is 1, since any endomorphism of $k[X, Y]$ with nonzero scalar Jacobian determinant can be transformed into one with determinant 1 by composing with an invertible linear transformation. The problem thus takes the form

Open Question IV.1.c3 (Jacobian conjecture)**.** *Prove or disprove: If k is a field of characteristic 0, then any k-algebra endomorphism of $k[X, Y]$ whose Jacobian matrix has determinant 1 is an automorphism.*

We remark that when $k = \mathbf{R}$, the analogous statement for *real analytic* maps of the plane into itself is false. For let $f : \mathbf{R} \to (-1, 1)$ be a real analytic function with everywhere positive derivative. Then the map $(x, y) \mapsto (f(x),\ f'(x)^{-1} y)$ is area-preserving, hence has Jacobian determinant 1, but it carries the plane into the strip $(-1, 1) \times \mathbf{R}$, and so is not onto. If we compose this f with a locally area-preserving real analytic map that winds the strip onto an annulus, we can get an example that is also not one-to-one.

Re §IV.2. Polynomials over a factorial ring.

The goal of this and the next section is to study conditions under which polynomials can or cannot be factored, and in particular, do or do not have roots.

P.180, first paragraph of this section [>]: For an example of the concept of the ''order'' of an element at a prime, let $A = \mathbf{C}[t]$ and $p = t - a$. Then $\mathrm{ord}_p(f)$ is the order (= multiplicity) of the zero of the function f at the point a. (If f does not have a zero at a, we say ''the order of its zero there is 0''.) This is probably the source of the name. More generally, if f is a member of the field of fractions of this ring, $\mathbf{C}(t)$, called ''the field of rational functions in t'', then $\mathrm{ord}_p(f)$ is again the order of the zero of f at a if f is defined at a, while if f has a pole at a, it is the negative of the order of this pole.

In the next sentence, Lang extends this concept from elements of K to polynomials $f \in K[X]$. Note that his definition can be simplified to say that the order of $f \in A[X]$ at the prime $p \in A$ is the minimum of the orders at p of *all* the coefficients of f, zero and nonzero.

P.181, statement of Gauss's Lemma [>]:

Content, and Gauss's Lemma. To understand what the ''content'' of a polynomial $f \in K[X]$ means, think first of the case where $f \in A[X]$. In this case, it is the greatest common divisor of the coefficients of f; i.e., the ''largest'' member of A that divides f. The easiest description of the content of a general $f \in K[X]$ is gotten by bringing the coefficients of f to a common denominator, i.e., writing $f = c^{-1} g$ where $c \in A$, $g \in A[X]$; then the content of f will be c^{-1} times the greatest common divisor of the coefficients of g. (If one defines it this way, one must, of course, prove that it is unique up to multiplication by units.) Still another way of interpreting the concept is by generalizing the concept of ''divisibility'', defining $u \in K$ to ''divide'' $v \in K$ ''with respect to A'' if $v = au$ for some $a \in A$. Then the content of f is simply the greatest common divisor, in this sense, of the coefficients of f.

The concept of ''content'' is mainly a tool in the formulation and proof of Gauss's Lemma, and its consequence Theorem IV.2.3, so don't expect to see it outside this and the next section. Actually, Gauss's

Lemma is often, and, I think, most usefully, stated in a way that doesn't refer to ''content'' – though that concept, or some equivalent, is still needed in the proof. This statement, which you should verify can be deduced from the statement in Lang, is

(c25) (Gauss's Lemma, alternative formulation) If A is a UFD, with field of fractions K, and we have a polynomial $f \in A[X]$ and a factorization $f = gh$ with $g, h \in K[X]$, then there exists a nonzero element $u \in K$ such that both terms in the factorization $f = (ug)(u^{-1}h)$ lie in $A[X]$.

Re §IV.3. Criteria for irreducibility.

P.184, proof of Theorem IV.3.1 [>]: This proof can be made simpler in two ways. First, instead of reducing to the ''content 1'' case, we can use the version of Gauss's Lemma given in (c25) above to conclude that the given factorization can be taken in $A[X]$. Second, observe that if our polynomial f has factorization gh in $A[X]$, where g and h are nonunits, then since the leading terms of g and h are not divisible by p, the images of g and h in $(A/(p))[X]$ both have positive degree. But the image of their product f in that ring has the form cX^n, and the only factors of a monomial in a polynomial ring over an integral domain are monomials; hence the images of g and h are monomials, and both therefore have constant term zero. Thus, g and h have constant terms divisible by p, hence their product f has constant term divisible by p^2, contradicting the hypothesis on a_0.

Though the assumption that A is a UFD is used to reduce the proof of this theorem to showing that f has no factors of positive degree in $A[X]$, the proof of the latter fact, which is the main part of the argument, does not require that A be a UFD. Thus, we get

(c26) (Eisenstein's irreducibility theorem, variant form.) Let A be an integral domain, $f(X) = a_n X^n + ... + a_0 \in A[X]$, and $\mathbf{p}$ a prime ideal of A, such that $a_n \notin \mathbf{p}$, $a_i \in \mathbf{p}$ for all $i < n$, and $a_0 \notin \mathbf{p}^2$. Then f is not a product of two polynomials of positive degree in $A[X]$.

P.184, proof of second example [=]: Why is it true that ''It will suffice to prove that the polynomial $f(X+1)$ is irreducible over $\mathbf{Q}$''? Roughly speaking, because a factorization $f(X) = g(X)\,h(X)$ would imply a factorization $f(X+1) = g(X+1)\,h(X+1)$, so if no factorization of the latter sort exists, no factorization of the former sort does either.

The elegant way to say this is that the homomorphism $\mathbf{Q}[X] \to \mathbf{Q}[X]$ taking X to $X+1$, an instance of the evaluation homomorphisms defined on p.98, has an inverse, the evaluation homomorphism taking X to $X-1$. Hence that map is an automorphism of $\mathbf{Q}[X]$, hence preserves irreducibility.

P.184, last display [=]: The symbol '' ... '' in the numerator of the right-hand term may be deceptive if you don't stop and think about. It does not mean that the omitted coefficients are all p; rather, these coefficients are given by the Binomial Theorem.

P.186, top paragraph [=]: Hint for proving the proposition: Gauss's Lemma.

Re §IV.4. Hilbert's Theorem.

P.186, first paragraph [>]: For basic results on Noetherian rings, see the discussion of the topic in this Companion under §II.5 above. The result cited by Lang from Chapter X in the proof of Theorem IV.4.1 (the equivalence of the three conditions on p.413) is a generalization to modules of the first Lemma of those pages; but that Lemma is all that is needed in this case.

Theorem IV.4.1, here called ''Hilbert's Theorem'', is commonly known as the **Hilbert Basis Theorem**. In early usage, ''basis'' meant ''generating set'', and the theorem is equivalent to the statement that if every ideal of A has a finite generating set, so does every ideal of $A[X]$. Nowadays, in linear algebra ''basis'' only means a *linearly independent* generating set, but the old usage is fossilized in the common name for this result.

P.187, last sentence of the proof of Theorem IV.4.1 [=]: Change this to ''By induction, the polynomial $f - c_1 f_{d1} - \ldots$ lies in the ideal generated by the f_{ij}, from which we conclude that f does as well.''

P.187, Corollary IV.4.2 [=]: Note that while the X of the preceding theorem was an indeterminate, the $x_1, \ldots, x_m$ of the Corollary are arbitrary ring elements.

Re §IV.5. Partial fractions. We skip this section in 250A. However, here is one erratum:

P.187, bottom line [=]: After ''$j(p) = 0$'' add ''*and* $\alpha_p = 0$''.

Re §IV.6. Symmetric polynomials.

P.190, beginning of §IV.6 [<]: The results of §§IV.2-3 were aimed at helping us study the possible roots of a polynomial in terms of its coefficients. But one can often get additional information about a problem by turning it backward. Thus, in this section, we look at a polynomial which *is* a product of linear factors, perhaps in some extension of A, and we examine how its coefficients depend on its roots. In general, the roots might be any elements of any commutative A-algebra; but if we write down a polynomial f whose roots are distinct *indeterminates*, i.e., generators of a polynomial ring, $t_1, \ldots, t_n \in A[t_1, \ldots, t_n]$, then from the coefficients of the polynomial

$$F(X) \;=\; (X - t_1) \ldots (X - t_n) \in A[t_1, \ldots, t_n][X],$$

we can get the coefficients of any polynomial of the form $(X - a_1) \ldots (X - a_n)$ by substituting the a_i for the t_i. (Here ''substituting'' means applying the unique homomorphism of A-algebras that takes t_i to a_i. If the a_i lie in A, this will be a homomorphism $A[t_1, \ldots, t_n] \to A$; if, more generally, they lie in a commutative A-algebra B, it will be a homomorphism $A[t_1, \ldots, t_n] \to B$.) Results about the polynomial $(X - t_1) \ldots (X - t_n)$ will thus give us information about any polynomial that can be written as a product of linear factors $X - a_i$.

P.190, end of first paragraph [>]: The arguments of this section do not require us to know what the s_i look like, and Lang only shows us s_1 and s_n, but the general case is easy to describe: s_i is the sum of the $\binom{n}{i}$ monomials which are products of i distinct factors taken from $\{t_1, \ldots, t_n\}$. E.g., when $n = 4$,

$$s_2(t_1, t_2, t_3, t_4) \;=\; t_1 t_2 + t_1 t_3 + t_1 t_4 + t_2 t_3 + t_2 t_4 + t_3 t_4 .$$

P.191, lines 3-5 [>]: The idea of the ''weight'' of a polynomial is ''the degree one would expect the polynomial gotten by substituting for each X_i the symmetric function s_i (or any other polynomial of degree i) to have''. Thus, this is an ad hoc definition, specific to this context. In other contexts, one may use ''weight'' functions defined by associating a different family of weights to a family of indeterminates.

Since the degree of the zero polynomial was defined to be $-\infty$, we shall also understand this to be the weight of the zero polynomial (though Lang does not say this).

One sometimes wishes to speak of a polynomial all of whose monomials have the same weight with respect to some weight function. The formal term for such a polynomial is ''isobaric'', though one more often hears ''homogeneous with respect to the given weight function''.

P.191, 4th line of *Proof* [=]: To ''substitute $t_n = 0$'' here means to apply the homomorphism $A[t_1, \ldots, t_n] \to A[t_1, \ldots, t_{n-1}]$ taking t_n to 0 and all other t_i to themselves.

P.192, first line [=]: After ''By induction'' add ''on d''. (There are two inductions going on in this proof, one on n, and one on d.)

P.192, line 6 [=]: Where Lang writes ''each term on the right has weight $\leq d$'', he means that $g_1(X_1, \ldots, X_{n-1})$ and $X_n g_2(X_1, \ldots, X_n)$ each have weight $\leq d$. Defining $g(X_1, \ldots, X_n) = g_1(X_1, \ldots, X_{n-1}) + X_n g_2(X_1, \ldots, X_n)$, we conclude that g has weight $\leq d$, giving the conclusion of the theorem as Lang states it. However, note that if g had weight strictly less than d, then on substituting s_i for each X_i, we would get a polynomial of degree strictly less than d in the t_i's, contradicting our assumption on f. Hence the weight of g must be exactly d. (The reason we could only say about each of g_1 and $X_n g_2$ that it had weight $\leq d$ is that one or the other of them might have smaller weight;

e.g., you might work out what happens when $f(t)=s_1$, respectively $f(t)=s_n$.)

P.192, italicized statement [=]: This is equivalent to saying that the polynomial g of the theorem just proved is unique. (So you could add the word ''unique'' to the theorem, and consider this the last part of the proof.)

In the proof of this statement, Lang implicitly uses induction on n again.

P.192, end of above proof (up to **Example**) [>].

Alternative proof of Theorem IV.6.1.

The preceding proof is instructive: on substituting $t_n = 0$, one loses some information about one's symmetric polynomial, but reduces to a case in which one can inductively assume the desired result. Thus, one can represent the resulting polynomial in terms of the elementary symmetric functions in $n-1$ variables. Replacing these by the corresponding symmetric polynomials in n variables, we get polynomial which, on the one hand, is expressible in terms of the s_i's, and on the other, resembles our original polynomial to a certain extent. This turns out to mean that the difference between the original polynomial and the new one is s_n times a symmetric polynomial of lower degree, and an induction on degree then allows us to get the desired result.

Here is a different (and also instructive) way to prove the same theorem. (In stating it, I will use ''monomial'' to mean what Lang calls a ''primitive monomial'', i.e., a member of the monoid generated by the elements in question, without coefficient from the base-ring.)

The symmetric group S_n permutes the monomials in $t_1,\ldots,t_n$, and the symmetric polynomials in these indeterminates are those polynomials f such that for each *orbit* of this action, the coefficients in f of all monomials in the orbit are the same. Let us select one representative from each such orbit. To do this, we consider the monomials of each degree to be ordered *lexicographically*, i.e., we write $t_1^{d_1}\ldots t_n^{d_n} > t_1^{e_1}\ldots t_n^{e_n}$ if and only if the least i for which $d_i \neq e_i$ has $d_i > e_i$. (In other words, in comparing monomials of the same degree, we compare first the exponents of t_1, then, if these are equal, we compare the exponents of t_2, etc..) We can see that in each orbit of monomials under the action of S_n, the *greatest* monomial

(c27) $$t_1^{d_1}\ldots t_n^{d_n}$$

under this ordering will be the unique member of the orbit which satisfies

(c28) $$d_1 \geq d_2 \geq \ldots \geq d_n.$$

Moreover, it is not hard to verify that for each monomial (c27) satisfying (c28), there is a unique monomial in $s_1,\ldots,s_n$ having (c27) as its greatest term under our ordering; namely

(c29) $$s_1^{d_1-d_2}\, s_2^{d_2-d_3}\ldots s_n^{d_n}.$$

Now suppose f is any symmetric polynomial in $t_1,\ldots,t_n$. Let us look at its highest-degree monomials, and among these, find the monomial (c27) that is greatest under our ordering. Since f is symmetric, this monomial will satisfy (c28), hence if we subtract (c29) times the coefficient of this monomial, we get a symmetric polynomial whose greatest monomial of highest degree is either of smaller degree than that of f, or of the same degree, but smaller under our ordering. Repeating this process, we must sooner or later run out of monomials of the original highest degree (since there are only finitely many monomials of that degree); we then eliminate monomials of the next lower degree the same way, and so on. This process must stop when the subtraction at some step yields remainder 0. When this happens, we will have expressed f as a polynomial in $s_1,\ldots,s_n$.

The same method also yields an easy proof of the italicized assertion on p.192 following the proof of the theorem (which should really be part of the statement of the theorem): it follows from the uniqueness statement in the line above (c29) that distinct monomials in $s_1,\ldots,s_n$ involve distinct highest monomials in $t_1,\ldots,t_n$ (under our ordering). It follows that a nontrivial linear combination of monomials in

$s_1, \ldots, s_n$ can never equal zero; for if we look at the highest monomial in $t_1, \ldots, t_n$ occurring among their expansions, this can only occur in one of them, hence will not be canceled when we sum these terms.

Having seen Theorem IV.6.1 proved in two ways, you should now convince yourself that a little bit more is true than what that theorem says; namely that if $f(t) \in A[t_1, \ldots, t_n]$ is symmetric and *homogeneous* of degree d (cf. first of my comments on p.191 above), then it can be written as $g(s_1, \ldots, s_n)$ where g is homogeneous (or in purist language, isobaric) of degree d with respect to our weight function. This will be used on p.193. One way to get this result is to go back over either of the above proofs and verify that they work with ''homogeneous'' inserted everywhere. Another is to use the result as stated to get the existence of a not-necessarily-homogeneous polynomial with the stated property, then group together its monomials homogeneous of each weight getting an expression $g = g_d + g_{d-1} + \ldots + g_0$, note that since each elementary symmetric polynomial s_i is homogeneous of degree i in $t_1, \ldots, t_n$, the substitution of the s's for the X's gives an expression for f in terms of homogeneous summands $g_i(s_1, \ldots, s_n)$; but now since f was assumed homogeneous of degree d, the summands of degrees other than d must be 0, and we must have $g = g_d$, as claimed.

P.193, first paragraph [>]: Lang uses the displayed formula to define a polynomial D_f, the unique polynomial in n indeterminates such that, when one substitutes for these indeterminates the respective elementary symmetric functions in $t_1, \ldots, t_n$, one gets the product of the squares of the pairwise differences of the t_i. Now if we map $A[t_1, \ldots, t_n]$ into an arbitrary A-algebra B by sending $t_1, \ldots, t_n$ to any elements $a_1, \ldots, a_n$, the above characterization of D_f yields the following statement: Given any polynomial of degree n over an A-algebra, which is a product of linear terms, $(X - a_1) \ldots (X - a_n)$, if one substitutes for the indeterminates in D_f the coefficients of this polynomial, one gets the product of the squares of the differences of the roots a_i.

It is this property that Lang will use for the rest of the section. In fact, whenever he specifies a polynomial f, he will then use the symbol ''D'' for the element gotten by substituting the coefficients of that polynomial for the indeterminates in the above polynomial D_f. Thus, under **Quadratic case** he will take $n=2$, look at the polynomial $f(X) = X^2 + bX + c$, and understand D to mean $D(b, c)$, while under **Cubic case** he will take $n=3$, look at the polynomial $f(X) = X^3 + aX + b$ (note the absence of an X^2 term), and understand D to mean $D(0, a, b)$.

Note also that $\prod_{i<j} (t_i - t_j)^2$, in addition to being symmetric, is homogeneous of degree $n(n-1)$; hence by our earlier observations, $D(s_1, \ldots, s_n)$ will be homogeneous of weight $n(n-1)$. This property will be retained if we replace one or more of the s_i by 0; in particular, $D(0, a, b)$ is homogeneous of weight 6. Lang will use this fact, without clearly stating it, a couple of paragraphs from now.

P.193, **Cubic case** [=]: Note that by taking the coefficient of X^2 to be 0, as noted above, Lang is essentially taking $t_3 = -t_1 - t_2$. This is equivalent to looking at the images of s_1, s_2, s_3, D, etc. under the map $A[t_1, t_2, t_3] \to A[t_1, t_2]$ taking t_3 to $-t_1 - t_2$; and indeed, Lang here speaks of D as a polynomial in t_1 and t_2. (He will indicate in the top paragraph of the next page how, by a substitution, one can reduce the problem of finding the discriminant of an arbitrary cubic to the case of a cubic with no X^2 term.)

Note that in the context of the preceding pages, each symmetric polynomial s_i is homogeneous of degree i in $t_1, \ldots, t_n$, δ is homogeneous of degree $n(n-1)/2$ in those indeterminates, and hence D is homogeneous of degree $n(n-1)$. When $n=3$ and we map $k[t_1, t_2, t_3]$ onto $k[t_1, t_2]$ by sending t_3 to $-t_1 - t_2$ as described above, the latter element is homogeneous of degree 1 in t_1 and t_2, so homogeneous elements remain homogeneous of the same degrees. This allows Lang to use homogeneity arguments in determining the form of D.

Three lines after middle display on this page, for both occurrences of the word ''special'' read ''particular''.

In the second line after the next display, ''of the product'' should be ''as the product''.

Re §§IV.7-9. This is material which there is no time to cover in 250A; but I give below some clarification of material on pp.196-197 brought to my attention by David Goldberg.

P.196, 3rd line after 2nd display [=]: Where Lang speaks of ''large'' primes, I believe the point is that the product $N_0(abc)$ of the primes in question must be large, hence they can't just be among the first few primes; so the largest of them must be at least moderately large. (But not necessarily anywhere near abc.)

P.197, 2nd line after first display [=]: For ''Again the proof is immediate from'', read ''We shall show below that the above conjecture is implied by''.

P.197, 3rd line before display (1) [=]: $mv >$ should be $mn >$.

P.197, display (1) [=]: Sketch of the derivation of these formulas: In the *abc* conjecture we may bound $\max(|a|, |b|, |c|)$ below by $|a|$, and $N_0(abc)$ above by $N_0(a)N_0(b)N_0(c)$. If we apply the resulting inequality to the equation preceding (1), dropping the factors A and B because they are fixed, bounding $N_0(u^m)$ and $N_0(v^n)$ by u and v respectively, and writing the result in ''<<'' notation, we get $|u|^m << (|u|\,|v|\,N_0(k))^{1+\varepsilon}$, hence $|u|^{m-1-\varepsilon}|v|^{-1-\varepsilon} << N_0(k)^{1+\varepsilon}$. We get a similar inequality with the roles of u and v reversed, and taking a ''multiplicative linear combination'' of these which eliminates $|v|$, and solving for $|u|$, we get $|u| << N_0(k)^{(1+\varepsilon)n/(mn-(m+n)(1+\varepsilon))}$. The exponent is not the same as in Lang's (1), but given any $\varepsilon > 0$, we can find an $\varepsilon' > 0$ (independent of m and n so long as $mn/(m+n)$ is uniformly bounded above 1; and indeed the relation $mn > m+n$ leads to a bound $6/5$) such that $(1+\varepsilon')n/(mn-(m+n)(1+\varepsilon')) \leq (1+\varepsilon)n/(mn-(m+n))$. Hence, applying the above computation with that ε', we get the first formula of (1). The second formula follows by symmetry.

P.197, sentence after next-to-last display [=]: For this read, ''The Hall conjecture is essentially the case of the first formula of (1) with $m=3$, $n=2$, and $N_0(k)$ bounded by $|k| = |u^3 - v^2|$.''