Enrichment

Chapters 1–4 form an idealized undergraduate course, written in the style of a graduate text. To help those seeing abstract algebra for the first time, I have prepared this section, which contains advice, explanations and additional examples for each section in the first four chapters.

Section 1.1

When we say that the rational numbers form a group under addition, we mean that rational numbers can be added and subtracted, and the result will inevitably be rational. Similarly for the integers, the real numbers, and the complex numbers. But the integers (even the nonzero integers) do not form a group under multiplication. If a is an integer other than ± 1, there is no integer b such that $ab = 1$. The nonzero rational numbers form a group under multiplication, as do the nonzero reals and the nonzero complex numbers. Not only can we add and subtract rationals, we can multiply and divide them (if the divisor is nonzero). The rational, reals and complex numbers are examples of *fields*, which will be studied systematically in Chapter 3.

Here is what the generalized associative law is saying. To compute the product of the elements a, b, c, d and e, one way is to first compute bc, then $(bc)d$, then $a((bc)d)$, and finally $[a((bc)d)e]$. Another way is (ab), then (cd), then $(ab)(cd)$, and finally $([(ab)(cd)]e)$. All procedures give the same result, which can therefore be written as $abcde$.

Notice that the solution to Problem 6 indicates how to construct a formal proof of 1.1.4.

Section 1.2

Groups whose descriptions differ may turn out to be isomorphic, and we already have an example from the groups discussed in this section. Consider the dihedral group D_6, with elements I, R, R^2, F, RF, R^2F. Let S_3 be the group of all permutations of $\{1, 2, 3\}$. We claim that D_6 and S_3 are isomorphic. This can be seen geometrically if we view D_6 as a group of permutations of the vertices of an equilateral triangle. Since D_6 has 6 elements and there are exactly 6 permutations of 3 symbols, we must conclude that D_6 and S_3 are essentially the same. To display an isomorphism explicitly, let R correspond to the

permutation (1,2,3) and F to (2,3). Then

$$I = (1),\ R = (1,2,3),\ R^2 = (1,3,2),\ F = (2,3),\ RF = (1,2),\ R^2F = (1,3).$$

If G is a nonabelian group, then it must have an element of order 3 or more. (For example, S_3 has two elements of order 3.) In other words, if every element of G has order 1 or 2, then G is abelian. To prove this, let $a, b \in G$; we will show that $ab = ba$. We can assume with loss of generality that $a \neq 1$ and $b \neq 1$. But then $a^2 = 1$ and $b^2 = 1$, so that a is its own inverse, and similarly for b. If ab has order 1, then $ab = 1$, so a and b are inverses of each other. By uniqueness of inverses, $a = b$, hence $ab = ba$. If ab has order 2, then $abab = 1$, so $ab = b^{-1}a^{-1} = ba$.

Section 1.3

Here is another view of cosets that may be helpful. Suppose that a coded message is to be transmitted, and the message is to be represented by a code word x (an n-dimensional vector with components in some field). The allowable code words are solutions of $Ax = 0$, where A is an m by n matrix, hence the set H of code words is an abelian group under componentwise addition, a subgroup of the abelian group G of all n-dimensional vectors. (In fact, G and H are vector space, but let's ignore the additional structure.) Transmission is affected by noise, so that the received vector is of the form $z = x + y$, where y is another n-dimensional vector, called the *error vector* or *error pattern vector*. Upon receiving z, we calculate the *syndrome* $s = Az$. If s turns out to be the zero vector, we declare that no error has occurred, and the transmitted word is z. Of course our decision may be incorrect, but under suitable assumptions about the nature of the noise, our decision procedure will minimize the probability of making a mistake. Again, let's ignore this difficulty and focus on the algebraic aspects of the problem. We make the following claim:

Two vectors z_1 and z_2 have the same syndrome if and only if they lie in the same coset of H in G.

To prove this, observe that $Az_1 = Az_2$ iff $A(z_1 - z_2) = 0$ iff $z_1 - z_2 \in H$ iff $z_1 \in z_2 + H$. (We are working in an abelian group, so we use the additive notation $z_2 + H$ rather than the multiplicative notation z_2H.)

Now suppose that we agree that we are going to correct the error pattern y_1, in other words, if we receive $z = x_1 + y_1$, where x_1 is a code word, we will decode z as x_1. If we receive $z' = x_1' + y_1$, where x_1' is another code word, we decode z' as x_1'. Thus our procedure corrects y_1 regardless of the particular word transmitted. Here is a key algebraic observation:

If y_1 and y_2 are distinct vectors that lie in the same coset, it is impossible to correct both y_1 and y_2.

This holds because $y_1 = y_2 + x$ for some code word $x \neq 0$, hence $y_1 + 0 = y_2 + x$. Therefore we cannot distinguish between the following two possibilities:

1. The zero word is transmitted and the error pattern is y_1;
2. x is transmitted and the error pattern is y_2.

It follows that among all vectors in a given coset, equivalently among all vectors having the same syndrome, we can choose exactly one as a correctable error pattern. If

the underlying field has only two elements 0 and 1, then (under suitable assumptions) it is best to choose to correct the pattern of minimum weight, that is, minimum number of 1's. In particular, if the coset is the subgroup H itself, then we choose the zero vector. This agrees with our earlier proposal: if the received vector z has zero syndrome, we decode z as z itself, thus "correcting" the zero pattern, in other words, declaring that there has been no error in transmission.

For further discussion and examples, see *Information Theory* by R. B. Ash, Dover 1991, Chapter 4.

Section 1.4

Here are some intuitive ideas that may help in visualizing the various isomorphism theorems. In topology, we can turn the real interval $[0, 1]$ into a circle by gluing the endpoints together, in other words identifying 0 and 1. Something similar is happening when we form the quotient group G/N where N is a normal subgroup of G. We have identified all the elements of N, and since the identity belongs to every subgroup, we can say that we have set everything in N equal to 1 (or 0 in the abelian case). Formally, (1.4.6) gives a correspondence between the subgroup of G/N consisting of the identity alone, and the subgroup N of G.

We have already seen an example of this identification process. In (1.2.4), we started with the free group G generated by the symbols R and F, and identified all sequences satisfying the relations $R^n = I$, $F^2 = I$, and $RF = FR^{-1}$ (equivalently $RFRF = I$). Here we would like to take N to be the subgroup of G generated by R^n, F^2, and $RFRF$, but N might not be normal. We will get around this technical difficulty when we discuss generators and relations in more detail in Section 5.8.

Section 1.5

Direct products provide a good illustration of the use of the first isomorphism theorem. Suppose that $G = H \times K$; what can we say about G/H? If $(h, k) \in G$, then $(h, k) = (h, 1_K)(1_H, k)$, and intuitively we have identified $(h, 1_K)$ with the identity $(1_H, 1_K)$. What we have left is $(1_H, k)$, and it appears that G/H should be isomorphic to K. To prove this formally, define $f\colon G \to K$ by $f(h, k) = k$. Then f is an epimorphism whose kernel is $\{(h, 1_K)\colon h \in H\}$, which can be identified with H. By the first isomorphism theorem, $G/H \cong K$.

Section 2.1

Here is an interesting ring that will come up in Section 9.5 in connection with group representation theory. Let $G = \{x_1, \ldots x_m\}$ be a finite group, and let R be any ring. The *group ring* RG consists of all elements $r_1x_1 + \cdots + r_mx_m$. Addition of elements is componentwise, just as if the x_i were basis vectors of a vector space and the r_i were scalars in a field. Multiplication in RG is governed by the given multiplication in R, along

with linearity. For example,

$$(r_1x_1 + r_2x_2)(s_1x_1 + s_2x_2) = r_1s_1x_1^2 + r_1s_2x_1x_2 + r_2s_1x_2x_1 + r_2s_2x_2^2.$$

The elements x_1^2, x_1x_2, x_2x_1, and x_2^2 belong to G, which need not be abelian. The elements r_1s_1, r_1s_2, r_2s_1, and r_2s_2 belong to R, which is not necessarily commutative. Thus it is essential to keep track of the order in which elements are written.

Section 2.2

Here is some additional practice with ideals in matrix rings. If I is an ideal of $M_n(R)$, we will show that I must have the form $M_n(I_0)$ for some unique ideal I_0 of R. [$M_n(I_0)$ is the set of all n by n matrices with entries in I_0.]

We note first that for any matrix A, we have $E_{ij}AE_{kl} = a_{jk}E_{il}$. This holds because $E_{ij}A$ puts row j of A in row i, and AE_{kl} puts column k of A in column l. Thus $E_{ij}AE_{kl}$ puts a_{jk} in the il position, with zeros elsewhere.

If I is an ideal of $M_n(R)$, let I_0 be the set of all entries a_{11}, where $A = (a_{ij})$ is a matrix in I. To verify that I_0 is an ideal, observe that $(A + B)_{11} = a_{11} + b_{11}$, $ca_{11} = (cE_{11}A)_{11}$, and $a_{11}c = (AE_{11}c)_{11}$. We will show that $I = M_n(I_0)$.

If $A \in I$, set $i = l = 1$ in the basic identity involving the elementary matrices E_{ij} (see the second paragraph above) to get $a_{jk}E_{11} \in I$. Thus $a_{jk} \in I_0$ for all j and k, so $A \in M_n(I_0)$.

Conversely, let $A \in M_n(I_0)$, so that $a_{il} \in I_0$ for all i, l. By definition of I_0, there is a matrix $B \in I$ such that $b_{11} = a_{il}$. Take $j = k = 1$ in the basic identity to get $E_{i1}BE_{1l} = b_{11}E_{il} = a_{il}E_{il}$. Consequently, $a_{il}E_{il} \in I$ for all i and l. But the sum of the matrices $a_{il}E_{il}$ over all i and l is simply A, and we conclude that $A \in I$.

To prove uniqueness, suppose that $M_n(I_0) = M_n(I_1)$. If $a \in I_0$, then $aE_{11} \in M_n(I_0) = M_n(I_1)$, so $a \in I_1$. A symmetrical argument completes the proof.

Section 2.3

If a and b are relatively prime integers, then a^i and b^j are relatively prime for all positive integers i and j. Here is an analogous result for ideals. Suppose that the ideals I_1 and I_2 of the ring R are relatively prime, so that $I_1 + I_2 = R$. Let us prove that I_1^2 and I_2 are relatively prime as well. By the definitions of the sum and product of ideals, we have

$$R = RR = (I_1 + I_2)(I_1 + I_2) = I_1^2 + I_1I_2 + I_2I_1 + I_2^2 \subseteq I_1^2 + I_2 \subseteq R$$

so $R = I_1^2 + I_2$, as asserted. Similarly, we can show that $R = I_1^3 + I_2$ by considering the product of $I_1^2 + I_2$ and $I_1 + I_2$. More generally, an induction argument shows that if

$$I_1 + \cdots + I_n = R,$$

then for all positive integers $m_1, \ldots, m_n$ we have

$$I_1^{m_1} + \cdots + I_n^{m_n} = R.$$

Section 2.4

We have defined prime ideals only when the ring R is commutative, and it is natural to ask why this restriction is imposed. Suppose that we drop the hypothesis of commutativity, and define prime ideals as in (2.4.4). We can then prove that if P is a prime ideal, I and J are arbitrary ideals, and $P \supseteq IJ$, then either $P \supseteq I$ or $P \supseteq J$. [If the conclusion is false, there are elements $a \in I \setminus P$ and $b \in J \setminus P$. Then $ab \in IJ \subseteq P$, but $a \notin P$ and $b \notin P$, a contradiction.]

If we try to reverse the process and show that a proper ideal P such that $P \supseteq IJ$ implies $P \supseteq I$ or $P \supseteq J$ must be prime, we run into trouble. If ab belongs to P, then the principal ideal (ab) is contained in P. We would like to conclude that $(a)(b) \subseteq P$, so that $(a) \subseteq P$ or $(b) \subseteq P$, in other words, $a \in P$ or $b \in P$. But (ab) need not equal $(a)(b)$. For example, to express the product of the element $ar \in (a)$ and the element $sb \in (b)$ as a multiple of ab, we must invoke commutativity.

An explicit example: Let P be the zero ideal in the ring $M_n(R)$ of n by n matrices over a division ring R (see Section 2.2, exercises). Since $M_n(R)$ has no nontrivial two-sided ideals, $P \supseteq IJ$ implies $P \supseteq I$ or $P \supseteq J$. But $ab \in P$ does not imply $a \in P$ or $b \in P$, because the product of two nonzero matrices can be zero.

This example illustrates another source of difficulty. The zero ideal P is maximal, but $M_n(R)/P$ is not a division ring. Thus we cannot generalize (2.4.3) by dropping commutativity and replacing "field" by "division ring". [If R/M is a division ring, it does follow that M is a maximal ideal; the proof given in (2.4.3) works.]

Section 2.5

Let's have a brief look at polynomials in more than one variable; we will have much more to say in Chapter 8. For example, a polynomial $f(X, Y, Z)$ in 3 variables is a sum of monomials; a monomial is of the form $aX^iY^jZ^k$ where a belongs to the underlying ring R. The degree of such a monomial is $i+j+k$, and the degree of f is the maximum monomial degree. Formally, we can define $R[X, Y]$ as $(R[X])[Y]$, $R[X, Y, Z]$ as $(R[X, Y])[Z]$, etc.

Let f be a polynomial of degree n in $F[X, Y]$, where F is a field. There are many cases in which f has infinitely many roots in F. For example, consider $f(X, Y) = X + Y$ over the reals. The problem is that there is no direct extension of the division algorithm (2.5.1) to polynomials in several variables. The study of solutions to polynomial equations in more than one variable leads to algebraic geometry, which will be introduced in Chapter 8.

Section 2.6

We have shown in (2.6.8) that every principal ideal domain is a unique factorization domain. Here is an example of a UFD that is not a PID. Let $R = \mathbb{Z}[X]$, which will be shown to be a UFD in (2.9.6). Let I be the maximal ideal $< 2, X >$ (see (2.4.8)). If I is principal, then I consists of all multiples of a polynomial $f(X)$ with integer coefficients. Since $2 \in I$, we must be able to multiply $f(X) = a_0 + a_1X + \cdots + a_nX^n$ by a polynomial $g(X) = b_0 + b_1X + \cdots + b_mX^m$ and produce 2. There is no way to do this unless $f(X) = 1$ or 2. But if $f(X) = 1$ then $I = R$, a contradiction (a maximal ideal must be proper).

Thus $f(X) = 2$, but we must also be able to multiply $f(X)$ by some polynomial in $\mathbb{Z}[X]$ to produce X. This is impossible, and we conclude that I is not principal.

A faster proof that $\mathbb{Z}[X]$ is not a PID is as follows. In (2.4.8) we showed that $< 2 >$ and $< X >$ are prime ideals that are not maximal, and the result follows from (2.6.9). On the other hand, the first method produced an explicit example of an ideal that is not principal.

Section 2.7

It may be useful to look at the Gaussian integers in more detail, and identify the primes in $\mathbb{Z}[i]$. To avoid confusion, we will call a prime in $\mathbb{Z}[i]$ a *Gaussian prime* and a prime in $\mathbb{Z}$ a *rational prime*. Anticipating some terminology from algebraic number theory, we define the *norm* of the Gaussian integer $a + bi$ as $N(a + bi) = a^2 + b^2$. We will outline a sequence of results that determine exactly which Gaussian integers are prime. We use Greek letters for members of $\mathbb{Z}[i]$ and roman letters for ordinary integers.

1. α is a unit in $\mathbb{Z}[i]$ iff $N(\alpha) = 1$. Thus the only Gaussian units are ± 1 and $\pm i$.

[If $\alpha\beta = 1$, then $1 = N(1) = N(\alpha)N(\beta)$, so both $N(\alpha)$ and $N(\beta)$ must be 1.]

Let n be a positive integer.

2. If n is a Gaussian prime, then n is a rational prime not expressible as the sum of two squares.

[n is a rational prime because any factorization in $\mathbb{Z}$ is also a factorization in $\mathbb{Z}[i]$. If $n = x^2 + y^2 = (x + iy)(x - iy)$, then either $x + iy$ or $x - iy$ is a unit. By (1), $n = 1$, a contradiction.]

Now assume that n is a rational prime but not a Gaussian prime.

3. If $\alpha = a + bi$ is a nontrivial factor of n, then $gcd(a, b) = 1$.

[If the greatest common divisor d is greater than 1, then $d = n$. Thus α divides n and n divides α, so n and α are associates, a contradiction.]

4. n is a sum of two squares.

[Let $n = (a+bi)(c+di)$; since n is real we have $ad+bc = 0$, so a divides bc. By (3), a and b are relatively prime, so a divides c, say $c = ka$. Then $b(ka) = bc = -ad$, so $d = -bk$. Thus $n = ac - bd = ka^2 + kb^2 = k(a^2 + b^2)$. But $a + bi$ is a nontrivial factor of n, so $a^2+b^2 = N(a+bi) > 1$. Since n is a rational prime, we must have $k = 1$ and $n = a^2+b^2$.]

By the above results, we have:

5. If n is a positive integer, then n is a Gaussian prime if and only if n is a rational prime not expressible as the sum of two squares.

Now assume that $\alpha = a + bi$ is a Gaussian integer with both a and b nonzero. (The cases in which a or b is 0 are covered by (1) and (5).)

6. If $N(\alpha)$ is a rational prime, then α is a Gaussian prime.

[If $\alpha = \beta\gamma$ where β and γ are not units, then $N(\alpha) = N(\beta)N(\gamma)$, where $N(\beta)$ and $N(\gamma)$ are greater than 1, contradicting the hypothesis.]

Now assume that α is a Gaussian prime.

7. If $N(\alpha) = hk$ is a nontrivial factorization, so that $h > 1$ and $k > 1$, then α divides either h or k. If, say, α divides h, then so does its complex conjugate $\overline{\alpha}$.

[We have $N(\alpha) = a^2 + b^2 = (a + bi)(a - bi) = \alpha\overline{\alpha} = hk$. Since α divides the product hk, it must divide one of the factors. If $\alpha\beta = h$, take complex conjugates to conclude that $\overline{\alpha}\overline{\beta} = h$.]

8. $N(\alpha)$ is a rational prime.

[If not, then we can assume by (7) that α and $\overline{\alpha}$ divide h. If α and $\overline{\alpha}$ are not associates, then $N(\alpha) = \alpha\overline{\alpha}$ divides h, so hk divides h and therefore $k = 1$, a contradiction. If α and its conjugate are associates, then one is $\pm i$ times the other. The only way this can happen is if $\alpha = \gamma(1 \pm i)$ where γ is a unit. But then $N(\alpha) = N(\gamma)N(1 \pm i) = N(1 \pm i) = 2$, a rational prime.]

By the above results, we have:

9. If $\alpha = a + bi$ with $a \neq 0$, $b \neq 0$, then α is a Gaussian prime if and only if $N(\alpha)$ is a rational prime.

The assertions (5) and (9) give a complete description of the Gaussian primes, except that it would be nice to know when a rational prime p can be expressed as the sum of two squares. We have $2 = 1^2 + 1^2$, so 2 is not a Gaussian prime, in fact $2 = (1 + i)(1 - i)$. If p is an odd prime, then p is a sum of two squares iff $p \equiv 1 \bmod 4$, as we will prove at the beginning of Chapter 7. Thus we may restate (5) as follows: 10. If n is a positive integer, then n is a Gaussian prime iff n is a rational prime congruent to 3 mod 4.

[Note that a number congruent to 0 or 2 mod 4 must be even.]

Section 2.8

Suppose that R is an integral domain with quotient field F, and g is a ring homomorphism from R to an integral domain R'. We can then regard g as mapping R into the quotient field F' of R'. It is natural to try to extend g to a homomorphism $\overline{g}\colon F \to F'$. If $a, b \in R$ with $b \neq 0$, then $a = b(a/b)$, so we must have $g(a) = g(b)\overline{g}(a/b)$. Thus if an extension exists, it must be given by

$$\overline{g}(a/b) = g(a)[g(b)]^{-1}.$$

For this to make sense, we must have $g(b) \neq 0$ whenever $b \neq 0$, in other words, g is a monomorphism. [Note that if $x, y \in R$ and $g(x) = g(y)$, then $g(x-y) = 0$, hence $x-y = 0$, so $x = y$.] We will see in (3.1.2) that any homomorphism of fields is a monomorphism, so this condition is automatically satisfied. We can establish the existence of $\overline{g}$ by defining it as above and then showing that it is a well-defined ring homomorphism. This has already been done in Problem 8. We are in the general situation described in Problem 7, with S taken as the set of nonzero elements of R. We must check that $g(s)$ is a unit in F' for every $s \in S$, but this holds because $g(s)$ is a nonzero element of F'.

Section 2.9

Here is another useful result relating factorization over an integral domain to factorization over the quotient field. Suppose that f is a monic polynomial with integer coefficients, and that f can be factored as gh, where g and h are monic polynomials with rational coefficients. Then g and h must have integer coefficients. More generally, let D be a unique factorization domain with quotient field F, and let f be a monic polynomial in $D[X]$. If $f = gh$, with g, $h \in F[X]$, then g and h must belong to $D[X]$.

To prove this, we invoke the basic proposition (2.9.2) to produce a nonzero $\lambda \in F$ such that $\lambda g \in D[X]$ and $\lambda^{-1}h \in D[X]$. But g and h are monic, so $\lambda = 1$ and the result follows.

Let f be a cubic polynomial in $F[X]$. If f is reducible, it must have a linear factor and hence a root in F. We can check this easily if F is a finite field; just try all possibilities. A finite check also suffices when $F = \mathbb{Q}$, by the rational root test (Section 2.9, Problem 1). If g is a linear factor of f, then $f/g = h$ is quadratic. We can factor h as above, and in addition the quadratic formula is available if square roots can be extracted in F. In other words, if $a \in F$, then $b^2 = a$ for some $b \in F$.

Section 3.1

All results in this section are basic and should be studied carefully. You probably have some experience with polynomials over the rational numbers, so let's do an example with a rather different flavor. Let $F = \mathbb{F}_2$ be the field with two elements 0 and 1, and let $f \in F[X]$ be the polynomial $X^2 + X + 1$. Note that f is irreducible over F, because if f were factorable, it would have a linear factor and hence a root in F. This is impossible, as $f(0) = f(1) = 1 \neq 0$. If we adjoin a root α of f to produce an extension $F(\alpha)$, we know that f is the minimal polynomial of α over F, and that $F(\alpha)$ consists of all elements $b_0 + b_1\alpha$, with b_0 and b_1 in F. Since b_0 and b_1 take on values 0 and 1, we have constructed a field $F(\alpha)$ with 4 elements. Moreover, all nonzero elements of $F(\alpha)$ can be expressed as powers of α, as follows:

$\alpha^0 = 1$, $\alpha^1 = \alpha$, $\alpha^2 = -\alpha - 1 = 1 + \alpha$. (The last equality follows because $1 + 1 = 0$ in F.)

This is a typical computation involving finite fields, which will be studied in detail in Chapter 6.

Section 3.2

We found in Problem 3 that a splitting field for $X^4 - 2$ has degree 8 over $\mathbb{Q}$. If we make a seemingly small change and consider $f(X) = X^4 - 1$, the results are quite different. The roots of f are 1, i, -1 and $-i$. Thus $\mathbb{Q}(i)$ is the desired splitting field, and it has degree 2 over $\mathbb{Q}$ because the minimal polynomial of i over $\mathbb{Q}$ has degree 2.

A general problem suggested by this example is to describe a splitting field for $X^n - 1$ over $\mathbb{Q}$ for an arbitrary positive integer n. The splitting field is $\mathbb{Q}(\omega)$, where ω is a primitive n^{th} root of unity, for example, $\omega = e^{i2\pi/n}$. We will see in Section 6.5 that the degree of $\mathbb{Q}(\omega)$ over $\mathbb{Q}$ is $\varphi(n)$, where φ is the Euler phi function.

Section 3.3

In Problem 8 we used the existence of an algebraic closure of F to show that any set of nonconstant polynomials in $F[X]$ has a splitting field over F. Conversely, if we suppose that it is possible to find a splitting field K for an arbitrary family of polynomials over the field F, then the existence of an algebraic closure of F can be established quickly. Thus let K be a splitting field for the collection of all polynomials in $F[X]$, and let C be the algebraic closure of F in K (see (3.3.4)). Then by definition, C is an algebraic extension of F and every nonconstant polynomial in $F[X]$ splits over C. By (3.3.6), C is an algebraic closure of F.

Section 3.4

Let's have another look at Example 3.4.8 with $p = 2$ to get some additional practice with separability and inseparability. We have seen that $\sqrt{t}$ is not separable over F, in fact it is purely inseparable because its minimal polynomial $X^2 - t$ can be written as $(X - \sqrt{t})^2$. But if we adjoin a cube root of t, the resulting element $\sqrt[3]{t}$ *is* separable over F, because $X^3 - t$ has nonzero derivative, equivalently does not belong to $F[X^2]$ (see 3.4.3).

Notice also that adjoining $\sqrt{t}$ and $\sqrt[3]{t}$ is equivalent to adjoining $\sqrt[6]{t}$, in other words, $F(\sqrt{t}, \sqrt[3]{t}) = F(\sqrt[6]{t})$. To see this, first observe that if $\alpha = \sqrt[6]{t}$, then $\sqrt{t} = \alpha^3$ and $\sqrt[3]{t} = \alpha^2$. On the other hand, $(\sqrt{t}/\sqrt[3]{t})^6 = t$.

It is possible for an element α to be both separable and purely inseparable over F, but it happens if and only if α belongs to F. The minimal polynomial of α over F must have only one distinct root and no repeated roots, so $\min(\alpha, F) = X - \alpha$. But the minimal polynomial has coefficients in F (by definition), and the result follows.

Section 3.5

Suppose we wish to find the Galois group of the extension E/F, where $E = F(\alpha)$. Assume that α is algebraic over F with minimal polynomial f, and that f has n distinct roots $\alpha_1 = \alpha, \alpha_2, \ldots, \alpha_n$ in some splitting field. If $\sigma \in \mathrm{Gal}(E/F)$, then σ permutes the roots of f by (3.5.1). Given any two roots α_i and α_j, $i \neq j$, we can find an F-isomorphism that carries α_i into α_j; see (3.2.3). Do not jump to the conclusion that all permutations are allowable, and therefore $\mathrm{Gal}(E/F)$ is isomorphic to S_n. For example, we may not be able to simultaneously carry α_1 into α_2 and α_3 into α_4. Another difficulty is that the F-isomorphism carrying α_i into α_j need not be an F-automorphism of E. This suggests that normality of the extension is a key property. If E/F is the non-normal extension of Example (3.5.10), the only allowable permutation is the identity.

Section 4.1

Finitely generated algebras over a commutative ring R frequently appear in applications to algebraic number theory and algebraic geometry. We say that A is a finitely generated R-algebra if there are finitely many elements $x_1, \ldots, x_n$ in A such that every element of

A is a polynomial $f(x_1, \ldots, x_n)$ with coefficients in R. Equivalently, A is a homomorphic image of the polynomial ring $R[X_1, \ldots, X_n]$. The homomorphism is determined explicitly by mapping X_i to x_i, $i = 1, \ldots, n$. The polynomial $f(X_1, \ldots, X_n)$ is then mapped to $f(x_1, \ldots, x_n)$.

If every element is not just a polynomial in the x_i but a *linear combination* of the x_i with coefficients in R, then A is a finitely generated *module* over R. To see the difference clearly, look at the polynomial ring $R[X]$, which is a finitely generated R algebra. (In the above discussion we can take $n = 1$ and $x_1 = X$.) But if $f_1, \ldots, f_n$ are polynomials in $R[X]$ and the maximum degree of the f_i is m, there is no way to take linear combinations of the f_i and produce a polynomial of degree greater than m. Thus $R[X]$ is a not a finitely generated R-module.

Section 4.2

Here is some practice working with quotient modules. Let N be a submodule of the R-module M, and let π be the canonical map from M onto M/N, taking $x \in M$ to $x + N \in M/N$. Suppose that N_1 and N_2 are submodules of M satisfying

(a) $N_1 \leq N_2$;

(b) $N_1 \cap N = N_2 \cap N$;

(c) $\pi(N_1) = \pi(N_2)$.

Then $N_1 = N_2$.

To prove this, let $x \in N_2$. Hypothesis (c) says that $(N_1 + N)/N = (N_2 + N)/N$; we don't write N_i/N, $i = 1, 2$, because N is not necessarily a submodule of N_1 or N_2. Thus $x + N \in (N_2 + N)/N = (N_1 + N)/N$, so $x + N = y + N$ for some $y \in N_1$. By (a), $y \in N_2$, hence $x - y \in N_2 \cap N = N_1 \cap N$ by (b). Therefore $x - y$ and y both belong to N_1, and consequently so does x. We have shown that $N_2 \leq N_1$, and in view of hypothesis (a), we are finished.

Section 4.3

If M is a free R-module with basis $S = (x_i)$, then an arbitrary function f from S to an arbitrary R-module N has a unique extension to an R-homomorphism $\overline{f}\colon M \to N$; see (4.3.6).

This property characterizes free modules, in other words, if M is an R-module with a subset S satisfying the above property, then M is free with basis S. To see this, build a free module M' with basis $S' = (y_i)$ having the same cardinality as S. For example, we can take M' to be the direct sum of copies of R, as many copies as there are elements of S. Define $f\colon S \to S' \subseteq M'$ by $f(x_i) = y_i$, and let $\overline{f}$ be the unique extension of f to an R-homomorphism from M to M'. Similarly, define $g\colon S' \to S \subseteq M$ by $g(y_i) = x_i$, and let $\overline{g}$ be the unique extension of g to an R-homomorphism from M' to M. Note that $\overline{g}$ exists and is unique because M' is free. Now $\overline{g} \circ \overline{f}$ is the identity on S, so by uniqueness of extensions from S to M, $\overline{g} \circ \overline{f}$ is the identity on M. Similarly, $\overline{f} \circ \overline{g}$ is the

identity on M'. Thus M and M' are not only isomorphic, but the isomorphism we have constructed carries S into S'. It follows that M is free with basis S.

This is an illustration of the characterization of an algebraic object by a *universal mapping property*. We will see other examples in Chapter 10.

Section 4.4

Here is some practice in decoding abstract presentations. An R-module can be defined as a representation of R in an endomorphism ring of an abelian group M. What does this mean?

First of all, for each $r \in R$, we have an endomorphism f_r of the abelian group M, given by $f_r(x) = rx$, $x \in M$. To say that f_r is an endomorphism is to say that $r(x+y) = rx+ry$, $x, y \in M$, $r \in R$.

Second, the mapping $r \to f_r$ is a ring homomorphism from R to $\mathrm{End}_R(M)$. (Such a mapping is called a representation of R in $\mathrm{End}_R(M)$.) This says that $f_{r+s}(x) = f_r(x) + f_s(x)$, $f_{rs}(x) = f_r(f_s(x))$, and $f_1(x) = x$. In other words, $(r + s)x = rx + sx$, $(rs)x = r(sx)$, and $1x = x$.

Thus we have found a fancy way to write the module axioms. If you are already comfortable with the informal view of a module as a "vector space over a ring", you are less likely to be thrown off stride by the abstraction.

Section 4.5

The technique given in Problems 1–3 for finding new bases and generators is worth emphasizing. We start with a matrix A to be reduced to Smith normal form. The equations $U = AX$ give the generators U of the submodule K in terms of the basis X of the free module M. The steps in the Smith calculation are of two types:

1. Premultiplication by an elementary row matrix R. This corresponds to changing generators via $V = RU$.
2. Postmultiplication by an elementary column matrix C. This corresponds to changing bases via $Y = C^{-1}X$.

Suppose that the elementary row matrices appearing in the calculation are $R_1, \ldots, R_s$, in that order, and the elementary column matrices are $C_1, \ldots, C_t$, in that order. Then the matrices Q and P are given by

$$Q = R_s \cdots R_2 R_1, \quad P^{-1} = C_1 C_2 \cdots C_t$$

hence $P = C_t^{-1} \cdots C_2^{-1} C_1^{-1}$. The final basis for M is $Y = PX$, and the final generating set for K is $V = QU = SY$, where $S = QAP^{-1}$ is the Smith normal form (see 4.4.2).

Section 4.6

Here is a result that is used in algebraic number theory. Let G be a free abelian group of rank n, and H a subgroup of G. By the simultaneous basis theorem, there is a basis

$y_1, \dots y_n$ of G and there are positive integers $a_1, \dots a_r$, $r \leq n$, such that a_i divides a_{i+1} for all i, and $a_1 y_1, \dots, a_r y_r$ is a basis for H. We claim that the abelian group G/H is finite if and only if $r = n$, and in this case, the size of G/H is $|G/H| = a_1 a_2 \cdots a_r$.

To see this, look at the proof of (4.6.3) with R^n replaced by G and K by H. The argument shows that G/H is the direct sum of cyclic groups $\mathbb{Z}/\mathbb{Z}a_i$, $i = 1, \dots, n$, with $a_i = 0$ for $r < i \leq n$. In other words, G/H is the direct sum of r finite cyclic groups (of order $a_1, \dots, a_r$ respectively) and $n - r$ copies of $\mathbb{Z}$. The result follows.

Now assume that $r = n$, and let $x_1, \dots, x_n$ and $z_1, \dots, z_n$ be arbitrary bases for G and H respectively. Then each z_i is a linear combination of the x_i with integer coefficients; in matrix form, $z = Ax$. We claim that $|G/H|$ is the absolute value of the determinant of A. To verify this, first look at the special case $x_i = y_i$ and $z_i = a_i y_i$, $i = 1, \dots, n$. Then A is a diagonal matrix with entries a_i, and the result follows. But the special case implies the general result, because any matrix corresponding to a change of basis of G or H is unimodular, in other words, has determinant ± 1. (See Section 4.4, Problem 1.)

Section 4.7

Here is some extra practice in diagram chasing. The diagram below is commutative with exact rows.

$$\begin{array}{ccccccc} A & \xrightarrow{f} & B & \xrightarrow{g} & C & \longrightarrow & 0 \\ \downarrow{\scriptstyle t} & & \downarrow{\scriptstyle u} & & \downarrow{\scriptstyle v} & & \\ A' & \xrightarrow[f']{} & B' & \xrightarrow[g']{} & C' & \longrightarrow & 0 \end{array}$$

If t and u are isomorphisms, we will show that v is also an isomorphism. (The hypothesis on t can be weakened to surjectivity.)

Let $c' \in C'$; then $c' = g'b'$ for some $b' \in B'$. Since u is surjective, $g'b' = g'ub$ for some $b \in B$. By commutativity, $g'ub = vgb$, which proves that v is surjective.

Now assume $vc = 0$. Since g is surjective, $c = gb$ for some $b \in B$. By commutativity, $vgb = g'ub = 0$. Thus $ub \in \ker g' = \operatorname{im} f'$, so $ub = f'a'$ for some $a' \in A'$. Since t is surjective, $f'a' = f'ta$ for some $a \in A$. By commutativity, $f'ta = ufa$. We now have $ub = ufa$, so $b - fa \in \ker u$, hence $b = fa$ because u is injective. Consequently,

$$c = gb = gfa = 0$$

which proves that v is injective.