**18.** Notation as in the previous exercises.

**(a)** Prove that $\ell_1$ intersects the $x$-axis in the point $(\eta_1/2, 0)$ and that $\ell_2$ intersects the $x$-axis in the point $(\eta_2/2, 0)$.

**(b)** Prove that $C_1$ is the circle having the points $(\eta_1, -1)$ and $(0, 1)$ as diameter. Prove that $s = \eta_1'$. Similarly prove that $C_2$ is the circle having the points $(\eta_2, -1)$ and $(0, 1)$ as diameter and that $t = \eta_4'$.

**(c)** Prove that $P$ has coordinates $(\eta_1'', 0)$ and hence that the construction in the previous problem constructs the regular 17-gon by straightedge and compass.

## 14.6 GALOIS GROUPS OF POLYNOMIALS

Recall that the Galois group of a separable polynomial $f(x) \in F[x]$ is defined to be the Galois group of the splitting field of $f(x)$ over $F$.

If $K$ is a Galois extension of $F$ then $K$ is the splitting field for some separable polynomial $f(x)$ over $F$. Any automorphism $\sigma \in \mathrm{Gal}(K/F)$ maps a root of an irreducible factor of $f(x)$ to another root of the irreducible factor and $\sigma$ is uniquely determined by its action on these roots (since they generate $K$ over $F$). If we fix a labelling of the roots $\alpha_1, \ldots, \alpha_n$ of $f(x)$ we see that any $\sigma \in \mathrm{Gal}(K/F)$ defines a unique permutation of $\alpha_1, \ldots, \alpha_n$, hence defines a unique permutation of the subscripts $\{1, 2, \ldots, n\}$ (which depends on the fixed labelling of the roots). This gives an injection

$$\mathrm{Gal}(K/F) \hookrightarrow S_n$$

of the Galois group into the symmetric group on $n$ letters which is clearly a homomorphism (both group operations are composition). We may therefore think of Galois groups as subgroups of symmetric groups. Since the degree of the splitting field is the same as the order of the Galois group by the Fundamental Theorem, this explains from the group-theoretic side why the splitting field for a polynomial of degree $n$ over $F$ is of degree at most $n!$ over $F$ (Proposition 13.26).

In general, if the factorization of $f(x)$ into irreducibles is $f(x) = f_1(x) \cdots f_k(x)$ where $f_i(x)$ has degree $n_i$, $i = 1, 2, \ldots, k$, then since the Galois group permutes the roots of the irreducible factors among themselves we have $\mathrm{Gal}(K/F) \leq S_{n_1} \times \cdots \times S_{n_k}$.

If $f(x)$ is irreducible, then given any two roots of $f(x)$ there is an automorphism in the Galois group $G$ of $f(x)$ which maps the first root to the second (this follows from our extension Theorem 13.27). Such a group is said to be *transitive* on the roots, i.e., you can get from any given root to any other root by applying some element of $G$. The fact that the Galois group must be transitive on blocks of roots (namely, the roots of the irreducible factors) can often be helpful in reducing the number of possibilities for the structure of $G$ (cf. the discussion of Galois groups of polynomials of degree 4 below).

### Examples

**(1)** Consider the biquadratic extension $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over $\mathbb{Q}$, which is the splitting field of $(x^2-2)(x^2-3)$. Label the roots as $\alpha_1 = \sqrt{2}, \alpha_2 = -\sqrt{2}, \alpha_3 = \sqrt{3}$ and $\alpha_4 = -\sqrt{3}$. The elements of the Galois group are $\{1, \sigma, \tau, \sigma\tau\}$ where $\sigma$ maps $\sqrt{2}$ to $-\sqrt{2}$ and fixes $\sqrt{3}$ and $\tau$ fixes $\sqrt{2}$ and maps $\sqrt{3}$ to $-\sqrt{3}$. As permutations of the roots for this

labelling we see that $\sigma$ interchanges the first two and fixes the second two and $\tau$ fixes the first two and interchanges the second two, i.e.,

$$\sigma = (12) \qquad \text{and} \qquad \tau = (34)$$

as elements of $S_4$. Similarly, or by taking the product of these two elements, we see that

$$\sigma\tau = (12)(34) \in S_4.$$

Hence

$$\mathrm{Gal}(\mathbb{Q}(\sqrt{2}, \sqrt{3})/\mathbb{Q}) \cong \{1, (12), (34), (12)(34)\} \subset S_4$$

identifying this Galois group with the Klein-4 subgroup of $S_4$. Note that if we had changed the labelling of the roots above we would have obtained a different (isomorphic) representation of the Galois group as a subgroup of $S_4$ (for example, interchanging the second and third roots would have given the subgroup $\{1, (13), (24), (13)(24)\}$).

**(2)** The Galois group of $x^3 - 2$ acts as permutations on the three roots $\sqrt[3]{2}$, $\rho\sqrt[3]{2}$ and $\rho^2\sqrt[3]{2}$ where $\rho$ is a primitive 3rd root of unity. With this ordering, the generators $\sigma$ and $\tau$ we have defined earlier give the permutations

$$\sigma = (123) \qquad \tau = (23)$$

which gives

$$\{1, \sigma, \sigma^2, \tau, \tau\sigma, \tau\sigma^2\} = \{1, (123), (132), (23), (13), (12)\} = S_3,$$

in this case the full symmetric group on 3 letters.

Recall that every finite group is isomorphic to a subgroup of some symmetric group $S_n$. It is an open problem to determine whether every finite group appears as the Galois group for some polynomial over $\mathbb{Q}$. We have seen in the last section that every abelian group is a Galois group over $\mathbb{Q}$ (for some subfield of a cyclotomic field). We shall explicitly determine the Galois groups for polynomials of small degree ($\leq 4$) below which will in particular show that every subgroup of $S_4$ arises as a Galois group.

We first introduce some definitions and show that the "general" polynomial of degree $n$ has $S_n$ as Galois group (so the second example above should be viewed as "typical").

**Definition.** Let $x_1, x_2, \ldots, x_n$ be indeterminates. The *elementary symmetric functions* $s_1, s_2, \ldots, s_n$ are defined by

$$\begin{aligned} s_1 &= x_1 + x_2 + \cdots + x_n \\ s_2 &= x_1x_2 + x_1x_3 + \cdots + x_2x_3 + x_2x_4 + \cdots + x_{n-1}x_n \\ &\vdots \\ s_n &= x_1x_2 \cdots x_n \end{aligned}$$

i.e., the $i^{\text{th}}$ symmetric function $s_i$ of $x_1, x_2, \ldots, x_n$ is the sum of all products of the $x_j$'s taken $i$ at a time.

**Definition.** The *general polynomial of degree n* is the polynomial

$$(x - x_1)(x - x_2) \cdots (x - x_n)$$

whose roots are the indeterminates $x_1, x_2, \ldots, x_n$.

It is easy to see by induction that the coefficients of the general polynomial of degree $n$ are given by the elementary symmetric functions in the roots:

$$(x - x_1)(x - x_2) \cdots (x - x_n) = x^n - s_1 x^{n-1} + s_2 x^{n-2} + \cdots + (-1)^n s_n. \quad (14.13)$$

For any field $F$, the extension $F(x_1, x_2, \ldots, x_n)$ is then a Galois extension of the field $F(s_1, s_2, \ldots, s_n)$ since it is the splitting field of the general polynomial of degree $n$.

If $\sigma \in S_n$ is any permutation of $\{1, 2, \ldots, n\}$, then $\sigma$ acts on the rational functions in $F(x_1, x_2, \ldots, x_n)$ by permuting the subscripts of the variables $x_1, x_2, \ldots, x_n$. It is clear that this gives an automorphism of $F(x_1, x_2, \ldots, x_n)$. Identifying $\sigma \in S_n$ with this automorphism of $F(x_1, x_2, \ldots, x_n)$ identifies $S_n$ as a subgroup of $\mathrm{Aut}(F(x_1, x_2, \ldots, x_n))$.

The elementary symmetric functions $s_1, s_2, \ldots, s_n$ are fixed under any permutation of their subscripts (this is the reason they are called *symmetric*), which shows that the subfield $F(s_1, s_2, \ldots, s_n)$ is contained in the fixed field of $S_n$. By the Fundamental Theorem of Galois Theory, the fixed field of $S_n$ has index precisely $n!$ in $F(x_1, x_2, \ldots, x_n)$. Since $F(x_1, x_2, \ldots, x_n)$ is the splitting field over $F(s_1, s_2, \ldots, s_n)$ of the polynomial of degree $n$ in (13), we have

$$[F(x_1, x_2, \ldots, x_n) : F(s_1, s_2, \ldots, s_n)] \leq n!. \quad (14.14)$$

It follows that we actually have equality and that $F(s_1, s_2, \ldots, s_n)$ is precisely the fixed field of $S_n$. This proves the following result.

**Proposition 30.** The fixed field of the symmetric group $S_n$ acting on the field of rational functions in $n$ variables $F(x_1, x_2, \ldots, x_n)$ is the field of rational functions in the elementary symmetric functions $F(s_1, s_2, \ldots, s_n)$.

**Definition.** A rational function $f(x_1, x_2, \ldots, x_n)$ is called *symmetric* if it is not changed by any permutation of the variables $x_1, x_2, \ldots, x_n$.

**Corollary 31.** *(Fundamental Theorem on Symmetric Functions)* Any symmetric function in the variables $x_1, x_2, \ldots, x_n$ is a rational function in the elementary symmetric functions $s_1, s_2, \ldots, s_n$.

*Proof:* A symmetric function lies in the fixed field of $S_n$ above, hence is a rational function in $s_1, \ldots, s_n$.

This corollary explains why these are called the *elementary* symmetric functions.

*Remark:* If $f(x_1, \ldots, x_n)$ is a *polynomial* in $x_1, x_2, \ldots, x_n$ which is symmetric then it can be seen that $f$ is actually a polynomial in $s_1, s_2, \ldots, s_n$, which strengthens the statement of the corollary. It is in fact true that a symmetric polynomial whose coefficients lie in $R$, where $R$ is any commutative ring with identity, is a polynomial in the elementary symmetric functions with coefficients in $R$. A proof of this fact is implicit in the algorithm outlined in the exercises for writing a symmetric polynomial as a polynomial in the elementary symmetric functions.

**Examples**

**(1)** The expression $(x_1 - x_2)^2$ is symmetric in $x_1, x_2$. We have

$$(x_1 - x_2)^2 = (x_1 + x_2)^2 - 4x_1x_2 = s_1^2 - 4s_2,$$

a polynomial in the elementary symmetric functions.

**(2)** The polynomial $x_1^2 + x_2^2 + x_3^2$ is symmetric in $x_1, x_2, x_3$, and in this case we have

$$\begin{aligned} x_1^2 + x_2^2 + x_3^2 &= (x_1 + x_2 + x_3)^2 - 2(x_1x_2 + x_1x_3 + x_2x_3) \\ &= s_1^2 - 2s_2. \end{aligned}$$

**(3)** The polynomial $x_1^2x_2^2 + x_1^2x_3^2 + x_2^2x_3^2$ is symmetric. Since

$$\begin{aligned} (x_1x_2 + x_1x_3 + x_2x_3)^2 &= x_1^2x_2^2 + x_1^2x_3^2 + x_2^2x_3^2 + 2(x_1^2x_2x_3 + x_2^2x_1x_3 + x_3^2x_1x_2) \\ &= x_1^2x_2^2 + x_1^2x_3^2 + x_2^2x_3^2 + 2x_1x_2x_3(x_1 + x_2 + x_3) \end{aligned}$$

we have

$$x_1^2x_2^2 + x_1^2x_3^2 + x_2^2x_3^2 = s_2^2 - 2s_1s_3.$$

Suppose now we *start* with the general polynomial

$$x^n - s_1x^{n-1} + s_2x^{n-2} + \cdots + (-1)^n s_n$$

over the field $F(s_1, s_2, \ldots, s_n)$ where we view the $s_i, i = 1, 2, \ldots, n$ as indeterminates. If we define the roots of this polynomial to be $x_1, x_2, \ldots, x_n$ then the $s_i$ are precisely the elementary symmetric functions in the roots $x_1, \ldots, x_n$. Moreover, these roots are indeterminates as well in the sense that there are no polynomial relations over $F$ between them. For suppose $p(t_1, \ldots, t_n)$ is a nonzero polynomial in $n$ variables with coefficients in $F$ such that $p(x_1, \ldots, x_n) = 0$. Then the product, $\tilde{p}$, over all $\sigma$ in $S_n$ of $p(t_{\sigma(1)}, \ldots, t_{\sigma(n)})$ is a nonzero symmetric polynomial with $\tilde{p}(x_1, \ldots, x_n) = 0$. This gives a nonzero polynomial relation over $F$ among $s_1, \ldots, s_n$, a contradiction. Conversely, if the roots of a polynomial $f(x)$ are independent indeterminates over $F$, then so are the coefficients of $f(x)$ — cf. the beginning of Section 9. Thus defining the general polynomial over $F$ as having indeterminate roots or indeterminate coefficients is equivalent. From this point of view our result can be stated in the following form.

**Theorem 32.** The general polynomial

$$x^n - s_1x^{n-1} + s_2x^{n-2} + \cdots + (-1)^n s_n$$

over the field $F(s_1, s_2, \ldots, s_n)$ is separable with Galois group $S_n$.

This result says that if there are no relations among the coefficients of a polynomial of degree $n$ (which is what we mean when we say the $s_i$ are indeterminates above) then the Galois group of this polynomial over the field generated by its coefficients is the full symmetric group $S_n$. Loosely speaking, this means that the "generic" polynomial of degree $n$ will have $S_n$ as Galois group. Note, however, that over finite fields every polynomial has a *cyclic* Galois group (all extensions of finite fields are cyclic), so that "generic" polynomials in this sense do not exist. Over $\mathbb{Q}$ one can make precise the

notion of "generic" polynomial and then it is true that most polynomials have the full symmetric group as Galois group.

For $n \geq 5$ there is only one normal subgroup of $S_n$, namely the subgroup $A_n$ of index 2. Hence in general there is only one normal subfield of $F(x_1, x_2, \ldots, x_n)$ containing $F(s_1, s_2, \ldots, s_n)$ and it is an extension of degree 2.

**Definition.** Define the *discriminant* $D$ of $x_1, x_2, \ldots, x_n$ by the formula

$$D = \prod_{i<j}(x_i - x_j)^2.$$

Define the discriminant of a polynomial to be the discriminant of the roots of the polynomial.

The discriminant $D$ is a symmetric function in $x_1, \ldots, x_n$, hence is an element of $K = F(s_1, s_2, \ldots, s_n)$.

When we first defined the alternating group $A_n$ we saw that a permutation $\sigma \in S_n$ is an element of the subgroup $A_n$ if and only if $\sigma$ fixes the product

$$\sqrt{D} = \prod_{i<j}(x_i - x_j) \in \mathbb{Z}[x_1, x_2, \ldots, x_n].$$

It follows (by the Fundamental Theorem) that if $F$ has characteristic different from 2 then $\sqrt{D}$ generates the fixed field of $A_n$ and generates a quadratic extension of $K$. This proves the following proposition.

**Proposition 33.** If $\mathrm{ch}(F) \neq 2$ then the permutation $\sigma \in S_n$ is an element of $A_n$ if and only if it fixes the square root of the discriminant $D$.

We now consider the Galois groups of separable polynomials of small degree ($\leq 4$) over a field $F$ which we assume is of characteristic different from 2 and 3. Note that over $\mathbb{Q}$ or over a finite field (or, more generally, over any perfect field) the splitting field of an arbitrary polynomial $f(x)$ is the same as the splitting field for the product of the irreducible factors of $f(x)$ taken precisely once, which is a separable polynomial.

If the roots of the polynomial $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ are $\alpha_1, \alpha_2, \ldots, \alpha_n$, then the discriminant of $f(x)$ is[2]

$$D = \prod_{i<j}(\alpha_i - \alpha_j)^2.$$

Note that $D = 0$ if and only if $f(x)$ is not separable, i.e., if the roots $\alpha_1, \ldots, \alpha_n$ are not distinct. Recall that over a perfect field (e.g., $\mathbb{Q}$ or a finite field) this implies $f(x)$ is reducible since every irreducible polynomial over a perfect field is separable.

The discriminant $D$ is symmetric in the roots of $f(x)$, hence is fixed by all the automorphisms of the Galois group of $f(x)$. By the Fundamental Theorem it follows that

[2]If $f(x) = a_nx^n + \cdots + a_0$ is not monic then its discriminant is defined to be $a_n^{2n-2}$ times the $D$ defined above.

$D \in F$. The discriminant can in general be written as a polynomial in the coefficients of $f(x)$ (by Corollary 31) which are fairly complicated for larger degrees (we shall give formulas for $n \leq 4$ below). Finally, note that since

$$\sqrt{D} = \prod_{i<j}(\alpha_i - \alpha_j)$$

we have the useful fact that $\sqrt{D}$ is always contained in the splitting field for $f(x)$.

If the roots of $f(x)$ are distinct, fix some ordering of the roots and view the Galois group of $f(x)$ as a subgroup of $S_n$ as above.

**Proposition 34.** The Galois group of $f(x) \in F[x]$ is a subgroup of $A_n$ if and only if the discriminant $D \in F$ is the square of an element of $F$.

*Proof:* This is a restatement of Proposition 33 in this case. The Galois group is contained in $A_n$ if and only if every element of the Galois group fixes

$$\sqrt{D} = \prod_{i<j}(\alpha_i - \alpha_j)$$

i.e., if and only if $\sqrt{D} \in F$.

This property, together with the fact that $D = 0$ determines the presence of multiple roots, is the reason $D$ is called the *discriminant*.

## Polynomials of Degree 2

Consider the polynomial $x^2 + ax + b$ with roots $\alpha$, $\beta$. The discriminant $D$ for this polynomial is $(\alpha - \beta)^2$, which can be written as a polynomial in the elementary symmetric functions of the roots. We did this in Example 1 above:

$$D = s_1^2 - 4s_2 = (-a)^2 - 4(b) = a^2 - 4b,$$

the usual discriminant for this quadratic.

The polynomial is separable if and only if $a^2 - 4b \neq 0$. The Galois group is a subgroup of $S_2$, the cyclic group of order 2 and is trivial (i.e., $A_2$ in this case) if and only if $a^2 - 4b$ is a rational square, which completely determines the possible Galois groups.

Note that this restates results we obtained previously by explicitly solving for the roots: if the polynomial is reducible (namely $D$ is a square in $F$), then the Galois group is trivial (the splitting field is just $F$), while if the polynomial is irreducible the Galois group is isomorphic to $\mathbb{Z}/2\mathbb{Z}$ since the splitting field is the quadratic extension $F(\sqrt{D})$.

## Polynomials of degree 3

Suppose the cubic polynomial is

$$f(x) = x^3 + ax^2 + bx + c. \tag{14.15}$$

If we make the substitution $x = y - a/3$ the polynomial becomes

$$g(y) = y^3 + py + q \tag{14.16}$$

where

$$p = \frac{1}{3}(3b - a^2) \qquad q = \frac{1}{27}(2a^3 - 9ab + 27c). \tag{14.17}$$

The splitting fields for these two polynomials are the same since their roots differ by the constant $a/3 \in F$ and since the formula for the discriminant involves the *differences* of roots, we see that these two polynomials also have the *same* discriminant.

Let the roots of the polynomial in (16) be $\alpha$, $\beta$, and $\gamma$. We first compute the discriminant of this polynomial in terms of $p$ and $q$. Note that

$$g(y) = (y - \alpha)(y - \beta)(y - \gamma)$$

so that if we differentiate we have

$$D_y g(y) = (y - \alpha)(y - \beta) + (y - \alpha)(y - \gamma) + (y - \beta)(y - \gamma).$$

Then

$$\begin{aligned} D_y g(\alpha) &= (\alpha - \beta)(\alpha - \gamma) \\ D_y g(\beta) &= (\beta - \alpha)(\beta - \gamma) \\ D_y g(\gamma) &= (\gamma - \alpha)(\gamma - \beta). \end{aligned}$$

Taking the product we see that

$$D = [(\alpha - \beta)(\alpha - \gamma)(\beta - \gamma)]^2 = -D_y g(\alpha) D_y g(\beta) D_y g(\gamma).$$

Since $D_y g(y) = 3y^2 + p$, we have

$$\begin{aligned} -D &= (3\alpha^2 + p)(3\beta^2 + p)(3\gamma^2 + p) \\ &= 27\alpha^2\beta^2\gamma^2 + 9p(\alpha^2\beta^2 + \alpha^2\gamma^2 + \beta^2\gamma^2) + 3p^2(\alpha^2 + \beta^2 + \gamma^2) + p^3. \end{aligned}$$

The corresponding expressions in the elementary symmetric functions of the roots were determined in Examples 2 and 3 above. Note that here $s_1 = 0$, $s_2 = p$ and $s_3 = -q$. We obtain

$$-D = 27(-q)^2 + 9p(p^2) + 3p^2(-2p) + p^3$$

so that

$$D = -4p^3 - 27q^2. \tag{14.18}$$

This is the same as the discriminant of $f(x)$ in (15). Expressing $D$ in terms of $a, b, c$ using (17) we obtain

$$D = a^2b^2 - 4b^3 - 4a^3c - 27c^2 + 18abc \tag{14.18'}$$

### (Galois Group of a Cubic)

**a.** If the cubic polynomial $f(x)$ is reducible, then it splits either into three linear factors or into a linear factor and an irreducible quadratic. In the first case the Galois group is trivial and in the second case the Galois group is of order 2.

**b.** If the cubic polynomial $f(x)$ is irreducible then a root of $f(x)$ generates an extension of degree 3 over $F$, so the degree of the splitting field over $F$ is divisible by 3. Since the Galois group is a subgroup of $S_3$, there are only two possibilities, namely

$A_3$ or $S_3$. The Galois group is $A_3$ (i.e., cyclic of order 3) if and only if the discriminant $D$ in (18) is a square.

Explicitly, if $D$ is the square of an element of $F$, then the splitting field of the irreducible cubic $f(x)$ is obtained by adjoining any single root of $f(x)$ to $F$. The resulting field is Galois over $F$ of degree 3 with a cyclic group of order 3 as Galois group. If $D$ is not the square of an element of $F$ then the splitting field of $f(x)$ is of degree 6 over $F$, hence is the field $F(\theta, \sqrt{D})$ for any one of the roots $\theta$ of $f(x)$. This extension is Galois over $F$ with Galois group $S_3$ (generators are given by $\sigma$, which takes $\theta$ to one of the other roots of $f(x)$ and fixes $\sqrt{D}$, and $\tau$, which takes $\sqrt{D}$ to $-\sqrt{D}$ and fixes $\theta$).

We see that in both cases the splitting field for the irreducible cubic $f(x)$ is obtained by adjoining $\sqrt{D}$ and a root of $f(x)$ to $F$.

We shall give explicit formulas for the roots of (16) (*Cardano's Formulas*) in the next section after introducing the notion of a *Lagrange Resolvent*.

## Polynomials of Degree 4

Let the quartic polynomial be

$$f(x) = x^4 + ax^3 + bx^2 + cx + d$$

which under the substitution $x = y - a/4$ becomes the quartic

$$g(y) = y^4 + py^2 + qy + r$$

with

$$p = \frac{1}{8}(-3a^2 + 8b)$$
$$q = \frac{1}{8}(a^3 - 4ab + 8c)$$
$$r = \frac{1}{256}(-3a^4 + 16a^2b - 64ac + 256d).$$

Let the roots of $g(y)$ be $\alpha_1$, $\alpha_2$, $\alpha_3$, and $\alpha_4$ and let $G$ denote the Galois group for the splitting field of $g(y)$ (or of $f(x)$).

Suppose first that $g(y)$ is reducible. If $g(y)$ splits into a linear and a cubic, then $G$ is the Galois group of the cubic, which we determined above. Suppose then that $g(y)$ splits into two irreducible quadratics. Then the splitting field is the extension $F(\sqrt{D_1}, \sqrt{D_2})$ where $D_1$ and $D_2$ are the discriminants of the two quadratics. If $D_1$ and $D_2$ do not differ by a square factor then this extension is a biquadratic extension and $G$ is isomorphic to the Klein 4-subgroup of $S_4$. If $D_1$ is a square times $D_2$ then this extension is a quadratic extension and $G$ is isomorphic to $\mathbb{Z}/2\mathbb{Z}$.

We are reduced to the situation where $g(y)$ is irreducible. In this case recall that the Galois group is transitive on the roots, i.e., it is possible to get from a given root to any other root by applying some automorphism of the Galois group. Examining the possibilities we see that the only transitive subgroups of $S_4$, hence the only possibilities

for our Galois group $G$, are the groups

$S_4$, $A_4$

$D_8 = \{1, (1324), (12)(34), (1423), (13)(24), (14)(23), (12), (34)\}$ and its conjugates

$V = \{1, (12)(34), (13)(24), (14)(23)\}$

$C = \{1, (1234), (13)(24), (1432)\}$ and its conjugates.

($D_8$ is the dihedral group, a Sylow 2-subgroup of $S_4$, with 3 (isomorphic) conjugate subgroups in $S_4$, $V$ is the Klein 4-subgroup of $S_4$, normal in $S_4$, and $C$ is a cyclic group, with 3 (isomorphic) conjugates in $S_4$).

Consider the elements

$$\theta_1 = (\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4)$$
$$\theta_2 = (\alpha_1 + \alpha_3)(\alpha_2 + \alpha_4)$$
$$\theta_3 = (\alpha_1 + \alpha_4)(\alpha_2 + \alpha_3)$$

in the splitting field for $g(y)$. These elements are permuted amongst themselves by the permutations in $S_4$. The stabilizer of $\theta_1$ in $S_4$ is the dihedral group $D_8$. The stabilizers in $S_4$ of $\theta_2$ and $\theta_3$ are the conjugate dihedral subgroups of order 8. The subgroup of $S_4$ which stabilizes all three of these elements is the intersection of these subgroups, namely the Klein 4-group $V$.

Since $S_4$ merely permutes $\theta_1, \theta_2, \theta_3$ it follows that the elementary symmetric functions in the $\theta$'s are fixed by all the elements of $S_4$, hence are in $F$. An elementary computation in symmetric functions shows that these elementary symmetric functions are $2p$, $p^2 - 4r$, and $-q^2$, which shows that $\theta_1, \theta_2, \theta_3$ are the roots of

$$h(x) = x^3 - 2px^2 + (p^2 - 4r)x + q^2$$

called the *resolvent cubic* for the quartic $g(y)$. Since

$$\begin{aligned}\theta_1 - \theta_2 &= \alpha_1\alpha_3 + \alpha_2\alpha_4 - \alpha_1\alpha_2 - \alpha_3\alpha_4 \\ &= -(\alpha_1 - \alpha_4)(\alpha_2 - \alpha_3)\end{aligned}$$

and similarly

$$\theta_1 - \theta_3 = -(\alpha_1 - \alpha_3)(\alpha_2 - \alpha_4)$$
$$\theta_2 - \theta_3 = -(\alpha_1 - \alpha_2)(\alpha_3 - \alpha_4)$$

we see that the discriminant of the resolvent cubic is the *same* as the discriminant of the quartic $g(y)$, hence also as the discriminant of the quartic $f(x)$. Using our formula for the discriminant of the cubic, we can easily compute the discriminant in terms of $p, q, r$:

$$D = 16p^4r - 4p^3q^2 - 128p^2r^2 + 144pq^2r - 27q^4 + 256r^3$$

from which one can give the formula for $D$ in terms of $a, b, c, d$:

$$\begin{aligned}D = &- 128b^2d^2 - 4a^3c^3 + 16b^4d - 4b^3c^2 - 27a^4d^2 + 18abc^3 \\ &+ 144a^2bd^2 - 192acd^2 + a^2b^2c^2 - 4a^2b^3d - 6a^2c^2d \\ &+ 144bc^2d + 256d^3 - 27c^4 - 80ab^2cd + 18a^3bcd.\end{aligned}$$

The splitting field for the resolvent cubic is a subfield of the splitting field of the quartic, so the Galois group of the resolvent cubic is a quotient of $G$. Hence knowing the action of the Galois group on the roots of the resolvent cubic $h(x)$ gives information about the Galois group of $g(y)$, as follows:

### (Galois group of a quartic)

**a.** Suppose first that the resolvent cubic is irreducible. If $D$ is not a square, then $G$ is not contained in $A_4$ and the Galois group of the resolvent cubic is $S_3$, which implies that the degree of the splitting field for $g(y)$ is divisible by 6. The only possibility is then $G = S_4$.

**b.** If the resolvent cubic is irreducible and $D$ is a square, then $G$ is a subgroup of $A_4$ and 3 divides the order of $G$ (the Galois group of the resolvent cubic is $A_3$). The only possibility is $G = A_4$.

**c1.** We are left with the case where the resolvent cubic is reducible. The first possibility is that $h(x)$ has 3 roots in $F$ (i.e., splits completely). Since each of the elements $\theta_1, \theta_2, \theta_3$ is in $F$, every element of $G$ fixes all three of these elements, which means $G \subseteq V$. The only possibility is $G = V$.

**c2.** If $h(x)$ splits into a linear and a quadratic, then precisely one of $\theta_1, \theta_2, \theta_3$ is in $F$, say $\theta_1$. Then $G$ stabilizes $\theta_1$ but not $\theta_2$ and $\theta_3$, so we have $G \subseteq D_8$ and $G \not\subseteq V$. This leaves two possibilities: $G = D_8$ or $G = C$. One way to distinguish between these is to observe that $F(\sqrt{D})$ is the fixed field of the elements of $G$ in $A_4$. For the two cases being considered, we have $D_8 \cap A_4 = V$, $C \cap A_4 = \{1, (13)(24)\}$. The first group is transitive on the roots of $g(y)$, the second is not. It follows that the first case occurs if and only if $g(y)$ is irreducible over $F(\sqrt{D})$. We may therefore determine $G$ completely by factoring $g(y)$ in $F(\sqrt{D})$, and so completely determine the Galois group in all cases. (cf. the exercises following and in the next section, where it is shown that over $\mathbb{Q}$ the Galois group cannot be cyclic of degree 4 if $D$ is not the sum of two squares — so in particular if $D < 0$.)

We shall give explicit formulas for the roots of a quartic polynomial at the end of the next section.

### The Fundamental Theorem of Algebra

We end this section with two proofs of the Fundamental Theorem of Algebra. We need two facts regarding the field $\mathbb{C}$:

**(a)** Every polynomial with real coefficients of odd degree has a root in the reals. Equivalently, there are no nontrivial finite extensions of $\mathbb{R}$ of odd degree.
**(b)** Quadratic polynomials with coefficients in $\mathbb{C}$ have roots in $\mathbb{C}$. Equivalently, there are no quadratic extensions of $\mathbb{C}$.

The first result follows from the Intermediate Value Theorem in calculus, since the graph of a monic polynomial $f(x) \in \mathbb{R}[x]$ of odd degree is negative for large negative values of $x$ and positive for large positive values of $x$, hence crosses the axis somewhere. The equivalence with the second statement follows since a finite extension of $\mathbb{R}$ is a

simple extension and the minimal polynomial of a primitive element would have odd degree, hence would be both irreducible over $\mathbb{R}$ and have a root in $\mathbb{R}$, hence must be of degree 1.

The second result follows by a direct computation. By the quadratic formula it suffices to show that every complex number $\alpha = a + bi$ , $a, b \in \mathbb{R}$, has a square root in $\mathbb{C}$. Write $\alpha = re^{i\theta}$ for some $r \geq 0$ and some $\theta \in [0, 2\pi)$. Then $\sqrt{r}e^{i\theta/2}$ is a square root of $\alpha$. (Explicitly, let $c \in \mathbb{R}$ be a square root of the real number $\dfrac{a + \sqrt{a^2 + b^2}}{2}$ and let $d \in \mathbb{R}$ be a square root of the real number $\dfrac{-a + \sqrt{a^2 + b^2}}{2}$ where the signs of the two square roots are chosen so that $cd$ has the same sign as $b$. Then multiplying out we see that $(c + di)^2 = a + bi$.)

**Theorem 35.** *(Fundamental Theorem of Algebra)* Every polynomial $f(x) \in \mathbb{C}[x]$ of degree $n$ has precisely $n$ roots in $\mathbb{C}$ (counted with multiplicity). Equivalently, $\mathbb{C}$ is algebraically closed.

*Proof:* I. It suffices to prove that every polynomial $f(x) \in \mathbb{C}[x]$ has a root in $\mathbb{C}$. Let $\tau$ denote the automorphism complex conjugation. If $f(x)$ has no root in $\mathbb{C}$ then neither does the conjugate polynomial $\bar{f}(x) = \tau f(x)$ obtained by applying $\tau$ to the coefficients of $f(x)$ (since its roots are the conjugates of the roots of $f(x)$). The product $f(x)\bar{f}(x)$ has coefficients which are invariant under complex conjugation, hence has real coefficients. It suffices then to prove that a polynomial with real coefficients has a root in $\mathbb{C}$.

Suppose that $f(x)$ is a polynomial of degree $n$ with real coefficients and write $n = 2^k m$ where $m$ is odd. We prove that $f(x)$ has a root in $\mathbb{C}$ by induction on $k$. For $k = 0$, $f(x)$ has odd degree and by (a) above $f(x)$ has a root in $\mathbb{R}$ so we are done. Suppose now that $k \geq 1$. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the roots of $f(x)$ and set $K = \mathbb{R}(\alpha_1, \alpha_2, \ldots, \alpha_n, i)$. Then $K$ is a Galois extension of $\mathbb{R}$ containing $\mathbb{C}$ and the roots of $f(x)$. For any $t \in \mathbb{R}$ consider the polynomial

$$L_t = \prod_{1 \leq i < j \leq n} [x - (\alpha_i + \alpha_j + t\alpha_i\alpha_j)].$$

Any automorphism of $K/\mathbb{R}$ permutes the terms in this product so the coefficients of $L_t$ are invariant under all the elements of $\mathrm{Gal}(K/\mathbb{R})$. Hence $L_t$ is a polynomial with real coefficients. The degree of $L_t$ is

$$\frac{n(n-1)}{2} = 2^{k-1}m(2^k m - 1) = 2^{k-1}m'$$

where $m'$ is odd (since $k \geq 1$). The power of 2 in this degree is therefore less than $k$, so by induction the polynomial $L_t$ has a root in $\mathbb{C}$. Hence for each $t \in \mathbb{R}$ one of the elements $\alpha_i + \alpha_j + t\alpha_i\alpha_j$ for some $i, j$ $(1 \leq i < j \leq n)$ is an element of $\mathbb{C}$. Since there are infinitely many choices for $t$ and only finitely many values of $i$ and $j$ we see that for some $i$ and $j$ (say, $i = 1$ and $j = 2$) there are distinct real numbers $s$ and $t$ with

$$\alpha_1 + \alpha_2 + s\alpha_1\alpha_2 \in \mathbb{C} \qquad \alpha_1 + \alpha_2 + t\alpha_1\alpha_2 \in \mathbb{C}.$$

Since $s \neq t$ it follows that $a = \alpha_1 + \alpha_2 \in \mathbb{C}$ and $b = \alpha_1\alpha_2 \in \mathbb{C}$. But then $\alpha_1$ and $\alpha_2$ are the roots of the quadratic $x^2 - ax + b$ with coefficients in $\mathbb{C}$, hence are elements of $\mathbb{C}$ by (b) above, completing the proof.

II. The second proof again uses (a) and (b) above, but replaces the computations with the polynomials $L_t$ above with a simple group-theoretic argument involving the nilpotency of a Sylow 2-subgroup of the Galois group:

Let $f(x)$ be a polynomial of degree $n$ with real coefficients and let $K$ be the splitting field of $f(x)$ over $\mathbb{R}$. Then $K(i)$ is a Galois extension of $\mathbb{R}$. Let $G$ denote its Galois group and let $P_2$ denote a Sylow 2-subgroup of $G$. The fixed field of $P_2$ is an extension of $\mathbb{R}$ of odd degree, hence by (a) is trivial.

It follows that $\mathrm{Gal}(K(i)/\mathbb{C})$ is a 2-group. Since 2-groups have subgroups of all orders (recall this is true of a finite $p$-group for any prime $p$, cf. Theorem 6.1), if this group is nontrivial, there would exist a quadratic extension of $\mathbb{C}$, impossible by (b), completing the proof.

The Fundamental Theorem of Algebra was first rigorously proved by Gauss in 1816 (his doctoral dissertation in 1798 provides a proof using geometric considerations requiring some topological justification). The first proof above is essentially due to Laplace in 1795 (hence the reason for naming the polynomials $L_t$). The reason Laplace's proof was deemed unacceptable was that he assumed the existence of a splitting field for polynomials (i.e., that the roots existed *somewhere* in *some* field), which had not been established at that time. The elegant second proof is a simplification due to Artin.

## EXERCISES

1. Show that a cubic with a multiple root has a linear factor. Is the same true for quartics?
2. Determine the Galois groups of the following polynomials:
   **(a)** $x^3 - x^2 - 4$
   **(b)** $x^3 - 2x + 4$
   **(c)** $x^3 - x + 1$
   **(d)** $x^3 + x^2 - 2x - 1$.
3. Prove for any $a, b \in \mathbb{F}_{p^n}$ that if $x^3 + ax + b$ is irreducible then $-4a^3 - 27b^2$ is a square in $\mathbb{F}_{p^n}$.
4. Determine the Galois group of $x^4 - 25$.
5. Determine the Galois group of $x^4 + 4$.
6. Determine the Galois group of $x^4 + 3x^3 - 3x - 2$.
7. Determine the Galois group of $x^4 + 2x^2 + x + 3$.
8. Determine the Galois group of $x^4 + 8x + 12$.
9. Determine the Galois group of $x^4 + 4x - 1$ (cf. Exercise 19).
10. Determine the Galois group of $x^5 + x - 1$.
11. Let $F$ be an extension of $\mathbb{Q}$ of degree 4 that is not Galois over $\mathbb{Q}$. Prove that the Galois closure of $F$ has Galois group either $S_4$, $A_4$ or the dihedral group $D_8$ of order 8. Prove that the Galois group is dihedral if and only if $F$ contains a quadratic extension of $\mathbb{Q}$.
12. Prove that an extension $F$ of $\mathbb{Q}$ of degree 4 can be generated by the root of an irreducible biquadratic $x^4 + ax^2 + b$ over $\mathbb{Q}$ if and only if $F$ contains a quadratic extension of $\mathbb{Q}$.

**13.** **(a)** Let $\pm\alpha, \pm\beta$ denote the roots of the polynomial $f(x) = x^4 + ax^2 + b \in \mathbb{Z}[x]$. Prove that $f(x)$ is irreducible if and only if $\alpha^2, \alpha \pm \beta$ are not elements of $\mathbb{Q}$.[3]

**(b)** Suppose $f(x)$ is irreducible and let $G$ be the Galois group of $f(x)$. Prove that

**(i)** $G \cong V$, the Klein 4-group, if and only if $b$ is a square in $\mathbb{Q}$ if and only if $\alpha\beta \in \mathbb{Q}$ is rational.

**(ii)** $G \cong C$, the cyclic group of order 4, if and only if $b(a^2 - 4b)$ is a square in $\mathbb{Q}$ if and only if $\mathbb{Q}(\alpha\beta) = \mathbb{Q}(\alpha^2)$.

**(iii)** $G \cong D_8$, the dihedral group of order 8, if and only if $b$ and $b(a^2 - 4b)$ are not squares in $\mathbb{Q}$ if and only if $\alpha\beta \notin \mathbb{Q}(\alpha^2)$.

**14.** Prove the polynomial $x^4 - px^2 + q \in \mathbb{Q}[x]$ is irreducible for any distinct odd primes $p$ and $q$ and has as Galois group the dihedral group of order 8.[4]

**15.** Prove the polynomial $x^4 + px + p \in \mathbb{Q}[x]$ is irreducible for every prime $p$ and for $p \neq 3, 5$ has Galois group $S_4$. Prove the Galois group for $p = 3$ is dihedral of order 8 and for $p = 5$ is cyclic of order 4.[5]

**16.** Determine the Galois group over $\mathbb{Q}$ of the polynomial $x^4 + 8x^2 + 8x + 4$. Determine which of the subfields of this field are Galois over $\mathbb{Q}$ and for those which are Galois determine a polynomial $f(x) \in \mathbb{Q}[x]$ for which they are the splitting field over $\mathbb{Q}$.

**17.** Find the Galois group of $x^4 - 7$ over $\mathbb{Q}$ explicitly as a permutation group on the roots.

**18.** Let $\theta$ be a root of $x^3 - 3x + 1$. Prove that the splitting field of this polynomial is $\mathbb{Q}(\theta)$ and that the Galois group is cyclic of order 3. In particular the other roots of this polynomial can be written in the form $a + b\theta + c\theta^2$ for some $a, b, c \in \mathbb{Q}$. Determine the other roots explicitly in terms of $\theta$.

**19.** Let $f(x)$ be an irreducible polynomial of degree 4 in $\mathbb{Q}[x]$ with discriminant $D$. Let $K$ denote the splitting field of $f(x)$, viewed as a subfield of the complex numbers $\mathbb{C}$.

**(a)** Prove that $\mathbb{Q}(\sqrt{D}) \subset K$.

**(b)** Let $\tau$ denote complex conjugation and let $\tau_K$ denote the restriction of complex conjugation to $K$. Prove that $\tau_K$ is an element of $\mathrm{Gal}(K/\mathbb{Q})$ of order 1 or 2 depending on whether every element of $K$ is real or not.

**(c)** Prove that if $D < 0$ then $K$ cannot be cyclic of degree 4 over $\mathbb{Q}$ (i.e., $\mathrm{Gal}(K/\mathbb{Q})$ cannot be a cyclic group of order 4).

**(d)** Prove generally that $\mathbb{Q}(\sqrt{D})$ for squarefree $D < 0$ is not a subfield of a cyclic quartic field (cf. also Exercise 19 of Section 7).

**20.** Determine the Galois group of $(x^3 - 2)(x^3 - 3)$ over $\mathbb{Q}$. Determine all the subfields which contain $\mathbb{Q}(\rho)$ where $\rho$ is a primitive $3^{\text{rd}}$ root of unity.

**21.** Let $G \leq S_n$ be a subgroup of the symmetric group and suppose $\sigma_1, \ldots, \sigma_k$ are generators for $G$. If the function $f(x_1, x_2, \ldots, x_n)$ is fixed by the generators $\sigma_i$ show it is fixed by $G$.

**22.** (*Newton's Formulas*) Let $f(x)$ be a monic polynomial of degree $n$ with roots $\alpha_1, \ldots, \alpha_n$. Let $s_i$ be the elementary symmetric function of degree $i$ in the roots and define $s_i = 0$ for $i > n$. Let $p_i = \alpha_1^i + \cdots + \alpha_n^i$, $i \geq 0$, be the sum of the $i^{\text{th}}$ powers of the roots of $f(x)$.

---

[3] cf. the note *An Elementary Test for the Galois Group of a Quartic Polynomial*, Luise-Charlotte Kappe and Bette Warren, Amer. Math. Monthly, 96(1989), pp. 133–137.

[4] Ibid.

[5] Ibid.

Prove *Newton's Formulas*:

$$p_1 - s_1 = 0$$
$$p_2 - s_1 p_1 + 2s_2 = 0$$
$$p_3 - s_1 p_2 + s_2 p_1 - 3s_3 = 0$$
$$\vdots$$
$$p_i - s_1 p_{i-1} + s_2 p_{i-2} - \cdots + (-1)^{i-1} s_{i-1} p_1 + (-1)^i i s_i = 0$$

**23.** **(a)** If $x + y + z = 1$, $x^2 + y^2 + z^2 = 2$ and $x^3 + y^3 + z^3 = 3$, determine $x^4 + y^4 + z^4$.
**(b)** Prove generally that $x$, $y$, $z$ are not rational but that $x^n + y^n + z^n$ is rational for every positive integer $n$.

**24.** Prove that an $n \times n$ matrix $A$ over a field of characteristic 0 is nilpotent if and only if the trace of $A^k$ is 0 for all $k \geq 0$.

**25.** Prove that two $n \times n$ matrices $A$ and $B$ over a field of characteristic 0 have the same characteristic polynomial if and only if the trace of $A^k$ equals the trace of $B^k$ for all $k \geq 0$.

**26.** Use the fact that the trace of $AB$ is the same as the trace of $BA$ for any two $n \times n$ matrices $A$ and $B$ to show that $AB$ and $BA$ have the same characteristic polynomial over a field of characteristic 0 (the same result is true over a field of arbitrary characteristic).

**27.** Let $f(x)$ be a monic polynomial of degree $n$ with roots $\alpha_1, \alpha_2, \ldots, \alpha_n$.
**(a)** Show that the discriminant $D$ of $f(x)$ is the square of the Vandermonde determinant

$$\begin{vmatrix} 1 & \alpha_1 & \alpha_1^2 & \ldots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \alpha_2^2 & \ldots & \alpha_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_n & \alpha_n^2 & \ldots & \alpha_n^{n-1} \end{vmatrix} = \prod_{i>j} (\alpha_i - \alpha_j).$$

**(b)** Taking the Vandermonde matrix above, multiplying on the left by its transpose and taking the determinant show that one obtains

$$D = \begin{vmatrix} p_0 & p_1 & p_2 & \cdots & p_{n-1} \\ p_1 & p_2 & p_3 & \cdots & p_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{n-1} & p_n & p_{n+1} & \cdots & p_{2n-2} \end{vmatrix}$$

where $p_i = \alpha_1^i + \cdots + \alpha_n^i$ is the sum of the $i^{\text{th}}$ powers of the roots of $f(x)$, which can be computed in terms of the coefficients of $f(x)$ using Newton's formulas above. This gives an efficient procedure for calculating the discriminant of a polynomial.

**28.** Let $\alpha$ be a root of the irreducible polynomial $f(x) \in F[x]$ and let $K = F(\alpha)$. Let $D$ be the discriminant of $f(x)$. Prove that $D = (-1)^{n(n-1)/2} \mathrm{N}_{K/F}(f'(\alpha))$, where $f'(x) = D_x f(x)$ is the derivative of $f(x)$.

The following exercises describe the *resultant* of two polynomials and in particular provide another efficient method for calculating the discriminant of a polynomial.

**29.** Let $F$ be a field and let $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ and $g(x) = b_m x^m + b_{m-1} x^{m-1} + \cdots + b_1 x + b_0$ be two polynomials in $F[x]$.
**(a)** Prove that a necessary and sufficient condition for $f(x)$ and $g(x)$ to have a common root (or, equivalently, a common divisor in $F[x]$) is the existence of a polynomial

$a(x) \in F[x]$ of degree at most $m-1$ and a polynomial $b(x) \in F[x]$ of degree at most $n-1$ with $a(x)f(x) = b(x)g(x)$.

**(b)** Writing $a(x)$ and $b(x)$ explicitly as polynomials show that equating coefficients in the equation $a(x)f(x) = b(x)g(x)$ gives a system of $n+m$ linear equations for the coefficients of $a(x)$ and $b(x)$. Prove that this system has a nontrivial solution (hence $f(x)$ and $g(x)$ have a common zero) if and only if the determinant

$$R(f,g) = \begin{vmatrix} a_n & a_{n-1} & \cdots & a_0 & & & & \\ & a_n & a_{n-1} & \cdots & a_0 & & & \\ & & a_n & a_{n-1} & \cdots & a_0 & & \\ & & & \ddots & & & & \\ & & & & a_n & a_{n-1} & \cdots & a_0 \\ b_m & b_{m-1} & \cdots & b_0 & & & & \\ & b_m & b_{m-1} & \cdots & b_0 & & & \\ & & b_m & b_{m-1} & \cdots & b_0 & & \\ & & & \ddots & & & & \\ & & & & b_m & b_{m-1} & \cdots & b_0 \end{vmatrix}$$

is zero. Here $R(f,g)$, called the *resultant* of the two polynomials, is the determinant of an $(n+m) \times (n+m)$ matrix $R$ with $m$ rows involving the coefficients of $f(x)$ and $n$ rows involving the coefficients of $g(x)$.

**30. (a)** With notations as in the previous problem, show that we have the matrix equation

$$R \begin{pmatrix} x^{n+m-1} \\ x^{n+m-2} \\ \vdots \\ x \\ 1 \end{pmatrix} = \begin{pmatrix} x^{m-1}f(x) \\ x^{m-2}f(x) \\ \vdots \\ f(x) \\ x^{n-1}g(x) \\ x^{n-2}g(x) \\ \vdots \\ g(x) \end{pmatrix}.$$

**(b)** Let $R'$ denote the matrix of cofactors of $R$ as in Theorem 30 of Section 11.4, so $R'R = R(f,g)I$, where $I$ is the identity matrix. Multiply both sides of the matrix equation above by $R'$ and equate the bottom entry of the resulting column matrices to prove that there are polynomials $r(x), s(x) \in F[x]$ such that $R(f,g)$ is equal to $r(x)f(x) + s(x)g(x)$, i.e., the resultant of two polynomials is a linear combination (in $F[x]$) of the polynomials.

**31.** Consider $f(x)$ and $g(x)$ as general polynomials and suppose the roots of $f(x)$ are $x_1, \ldots, x_n$ and the roots of $g(x)$ are $y_1, \ldots, y_m$. The coefficients of $f(x)$ are powers of $a_n$ times the elementary symmetric functions in $x_1, x_2, \ldots, x_n$ and the coefficients of $g(x)$ are powers of $b_m$ times the elementary symmetric functions in $y_1, y_2, \ldots, y_m$.

**(a)** By expanding the determinant show that $R(f,g)$ is homogeneous of degree $m$ in the coefficients $a_i$ and homogeneous of degree $n$ in the coefficients $b_j$.

**(b)** Show that $R(f,g)$ is $a_n^m b_m^n$ times a symmetric function in $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$.

**(c)** Since $R(f,g)$ is 0 if $f(x)$ and $g(x)$ have a common root, say $x_i = y_j$, show that $R(f,g)$ is divisible by $x_i - y_j$ for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$. Conclude by

degree considerations that

$$R = a_n^m b_m^n \prod_{i=1}^{n} \prod_{j=1}^{m} (x_i - y_j).$$

**(d)** Show that the product in (c) can be also be written

$$R(f, g) = a_n^m \prod_{i=1}^{n} g(x_i) = (-1)^{nm} b_m^n \prod_{j=1}^{m} f(y_j).$$

This gives an interesting *reciprocity* between the product of $g$ evaluated at the roots of $f$ and the product of $f$ evaluated at the roots of $g$.

**32.** Consider now the special case where $g(x) = f'(x)$ is the derivative of the polynomial $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ and suppose the roots of $f(x)$ are $\alpha_1, \alpha_2, \ldots, \alpha_n$. Using the formula

$$R(f, f') = \prod_{i=1}^{n} f'(\alpha_i)$$

of the previous exercise, prove that

$$D = (-1)^{n(n-1)/2} R(f, f')$$

where $D$ is the discriminant of $f(x)$.

**33.** **(a)** Prove that the discriminant of the cyclotomic polynomial $\Phi_p(x)$ of the $p^{\text{th}}$ roots of unity for an odd prime $p$ is $(-1)^{(p-1)/2} p^{p-2}$ [One approach: use Exercise 5 of the previous section together with the determinant form for the discriminant in terms of the power sums $p_i$.]

**(b)** Prove that $\mathbb{Q}(\sqrt{(-1)^{(p-1)/2} p}) \subset \mathbb{Q}(\zeta_p)$ for $p$ an odd prime. (Cf. also Exercise 11 of Section 7.)

**34.** Use the previous exercise to prove that every quadratic extension of $\mathbb{Q}$ is contained in a cyclotomic extension (a special case of the Kronecker–Weber Theorem).

**35.** Prove that the discriminant $D$ of the polynomial $x^n + px + q$ is given by the formula $(-1)^{n(n-1)/2} n^n q^{n-1} + (-1)^{(n-1)(n-2)/2} (n-1)^{n-1} p^n$.

**36.** Prove that the discriminant of $x^n + nx^{n-1} + n(n-1)x^{n-2} + \cdots + n(n-1)\ldots(3)(2)x + n!$ is $(-1)^{n(n-1)/2}(n!)^n$.

The following exercises 37 to 43 outline two procedures for writing a symmetric function in terms of the elementary symmetric functions. Let $f(x_1, \ldots, x_n)$ be a polynomial which is symmetric in $x_1, \ldots, x_n$. Recall that the degree (sometimes called the *weight*) of the monomial $Ax_1^{a_1} x_2^{a_2} \ldots x_n^{a_n}$ $(a_i \geq 0)$ is $a_1 + a_2 + \cdots + a_n$ and that a polynomial is *homogeneous (of degree m)* if every monomial has the same degree $(m)$.

**37.** **(a)** Show that every polynomial $f(x_1, \ldots, x_n)$ can be written as a sum of homogeneous polynomials. Show that if $f(x_1, \ldots, x_n)$ is symmetric then each of these homogeneous polynomials is also symmetric.

**(b)** Show that the monomial $Bs_1^{a_1} s_2^{a_2} \ldots s_n^{a_n}$ in the elementary symmetric functions is a homogeneous polynomial in $x_1, x_2, \ldots, x_n$ of degree $a_1 + 2a_2 + \cdots + na_n$.

In writing $f(x_1, \ldots, x_n)$ as a polynomial in the symmetric functions it therefore suffices to assume that $f(x_1, \ldots, x_n)$ is homogeneous.

Recall the *lexicographic monomial order* with $x_1 > x_2 > \cdots > x_n$ defined in Section 9.6, where the nonzero monomial term with exponents $(a_1, a_2, \ldots, a_n)$ comes before the nonzero monomial term with exponents $(b_1, b_2, \ldots, b_n)$ if the initial components of the two $n$-tuples of exponents are equal and the first component where they differ has $a_i > b_i$. If $f(x_1, \ldots, x_n)$ contains the monomial $Ax_1^{a_1}x_2^{a_2}\ldots x_n^{a_n}$ then since $f(x_1, \ldots, x_n)$ is symmetric it also contains all the permuted monomials. Among these choose the lexicographically largest monomial, which therefore satisfies $a_1 \geq a_2 \geq \cdots \geq a_n \geq 0$.

**38.** **(a)** Show that the monomial $As_1^{a_1-a_2}s_2^{a_2-a_3}\cdots s_n^{a_n}$ in the elementary symmetric functions has the same lexicographic initial term.

**(b)** Show that subtracting $As_1^{a_1-a_2}s_2^{a_2-a_3}\cdots s_n^{a_n}$ from $f(x)$ yields either 0 or a symmetric polynomial of the same degree whose terms are lexicographically smaller than the terms in $f(x_1, \ldots, x_n)$.

**(c)** Show that the iteration of this procedure (lexicographic ordering, choosing the lexicographically largest term, subtracting the associated monomial in the elementary symmetric functions) terminates, expressing $f(x_1, \ldots, x_n)$ as a polynomial in the elementary symmetric functions.

**39.** Use the algorithm described in Exercise 38 to prove that a polynomial $f(x_1, \ldots, x_n)$ that is symmetric in $x_1, \ldots, x_n$ can be expressed *uniquely* as a polynomial in the elementary symmetric functions.

**40.** Use the procedure in Exercise 38 to express each of the following symmetric functions as a polynomial in the elementary symmetric functions:

**(a)** $(x_1 - x_2)^2$

**(b)** $x_1^2 + x_2^2 + x_3^2$

**(c)** $x_1^2x_2^2 + x_1^2x_3^2 + x_2^2x_3^2$.

**41.** Use the procedure in Exercise 38 to express $\sum_{i\neq j} x_i^2x_j$ as a polynomial in the elementary symmetric functions.

We now know that a symmetric polynomial $f(x_1, \ldots, x_n)$ can be written uniquely as a polynomial in the elementary symmetric functions. Using this existence and uniqueness we can describe an alternate and computationally useful method for determining the coefficients of the elementary symmetric functions in this polynomial. As in Exercise 37 we may assume that $f(x_1, \ldots, x_n)$ is homogeneous of degree $M$. Let $N$ be the maximum degree of any of the variables $x_1, \ldots, x_n$ in $f(x_1, \ldots, x_n)$.

**(a)** Determine all of the possible monomials $A_is_1^{a_1}s_2^{a_2}\cdots s_n^{a_n}$ appearing in $f(x_1, \ldots, x_n)$ from the constraints

$$a_1 + 2a_2 + \cdots + na_n = M$$
$$a_1 + a_2 + \cdots + a_n \leq N.$$

**(b)** Since $f(x_1, \ldots, x_n) = \sum A_is_1^{a_1}s_2^{a_2}\cdots s_n^{a_n}$ is a polynomial *identity*, it is valid for any substitution of values for $x_1, \ldots, x_n$. Each substitution into this equation gives a linear relation on the coefficients $A_i$ and so a sufficient number of substitutions will determine the $A_i$.

**42.** Show that the function $(x_1 + x_2 - x_3 - x_4)(x_1 + x_3 - x_2 - x_4)(x_1 + x_4 - x_2 - x_3)$ is symmetric in $x_1, x_2, x_3, x_4$ and use the preceding procedure to prove it can be expressed as a polynomial in the elementary symmetric functions as $s_1^3 - 4s_1s_2 + 8s_3$.

**43.** Express each of the following in terms of the elementary symmetric functions:

**(a)** $\sum_{i\neq j} x_i^2x_j$ **(b)** $\sum_{i,j,k \text{ distinct}} x_i^2x_jx_k$ **(c)** $\sum_{i,j,k \text{ distinct}} x_i^2x_j^2x_k^2$.

**44.** Let $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ be the roots of a quartic polynomial $f(x)$ over $\mathbb{Q}$. Show that the quantities $\alpha_1\alpha_2 + \alpha_3\alpha_4$, $\alpha_1\alpha_3 + \alpha_2\alpha_4$, and $\alpha_1\alpha_4 + \alpha_2\alpha_3$ are permuted by the Galois group of $f(x)$. Conclude that these elements are the roots of a cubic polynomial with coefficients in $\mathbb{Q}$ (also sometimes referred to as the *resolvent cubic* of $f(x)$).

**45.** If $f(x) = x^3 + px + q \in \mathbb{Z}[x]$ is irreducible, prove that its discriminant $D = -4p^3 - 27q^2$ is an integer not equal to $0, \pm 1$.

**46.** Prove that every finite group occurs as the Galois group of a field extension of the form $F(x_1, x_2, \ldots, x_n)/E$.

**47.** Let $F$ be a field of characteristic 0 in which every cubic polynomial has a root. Let $f(x)$ be an irreducible quartic polynomial over $F$ whose discriminant is a square in $F$. Determine the Galois group of $f(x)$.

**48.** This exercise determines the splitting field $K$ for the polynomial $f(x) = x^6 - 2x^3 - 2$ over $\mathbb{Q}$ (cf. also Exercise 2 of Section 8).

**(a)** Prove that $f(x)$ is irreducible over $\mathbb{Q}$ with roots the three cube roots of $1 \pm \sqrt{3}$.

**(b)** Prove that $K$ contains the field $\mathbb{Q}(\sqrt{-3})$ of 3rd roots of unity and contains $\mathbb{Q}(\sqrt{3})$, hence contains the biquadratic field $F = \mathbb{Q}(i, \sqrt{3})$. Take the product of two of the roots in (a) to prove that $K$ contains $\sqrt[3]{2}$ and conclude that $K$ is an extension of the field $L = \mathbb{Q}(\sqrt[3]{2}, i, \sqrt{3})$.

**(c)** Prove that $[L : \mathbb{Q}] = 12$ and that $K$ is obtained from $L$ by adjoining the cube root of an element in $L$, so that $[K : \mathbb{Q}] = 12$ or 36.

**(d)** Prove that if $[K : \mathbb{Q}] = 12$ then $K = \mathbb{Q}(\sqrt[3]{2}, i, \sqrt{3})$ and that $\mathrm{Gal}(K/\mathbb{Q})$ is isomorphic to the direct product of the cyclic group of order 2 and $S_3$. Prove that if $[K : \mathbb{Q}] = 12$ then there is a unique real cubic subfield in $K$, namely $\mathbb{Q}(\sqrt[3]{2})$.

**(e)** Take the quotient of the two real roots in (a) to show that $\sqrt[3]{2 + \sqrt{3}}$ and $\sqrt[3]{2 - \sqrt{3}}$ (real roots) are both elements of $K$. Show that $\alpha = \sqrt[3]{2 + \sqrt{3}} + \sqrt[3]{2 - \sqrt{3}}$ is a real root of the irreducible cubic equation $x^3 - 3x - 4$ whose discriminant is $-2^2 3^4$. Conclude that the Galois closure of $\mathbb{Q}(\alpha)$ contains $\mathbb{Q}(i)$ so in particular $\mathbb{Q}(\alpha) \neq \mathbb{Q}(\sqrt[3]{2})$.

**(f)** Conclude from (e) that $G = \mathrm{Gal}(K/\mathbb{Q})$ is of order 36. Determine all the elements of $G$ explicitly and in particular show that $G$ is isomorphic to $S_3 \times S_3$.

**49.** Prove that the Galois group over $\mathbb{Q}$ of $x^6 - 4x^3 + 1$ is isomorphic to the dihedral group of order 12. [Observe that the two real roots are inverses of each other.]

**50.** *(Criterion for the Galois Group of an Irreducible Cubic over an Arbitrary Field)* Suppose $K$ is a field and $f(x) = x^3 + ax^2 + bx + c \in K[x]$ is irreducible, so the Galois group of $f(x)$ over $K$ is either $S_3$ or $A_3$.

**(a)** Show that the Galois group of $f(x)$ is $A_3$ if and only if the resultant quadratic polynomial $g(x) = x^2 + (ab - 3c)x + (b^3 + a^3c - 6abc + 9c^2)$ has a root in $K$. [If $\alpha$, $\beta$, $\gamma$ are the roots of $f(x)$ show that the Galois group is $A_3$ if and only if the element $\theta = \alpha\beta^2 + \beta\gamma^2 + \gamma\alpha^2$ is an element of $K$ and that $\theta$ is a root of $g(x)$.] Show that the discriminant of $g(x)$ is the same as the discriminant of $f(x)$.

**(b)** $(\mathrm{ch}(K) \neq 2)$ If $K$ has characteristic different from 2 show either from (a) or directly from the definition of the discriminant that the Galois group of $f(x)$ is $A_3$ if and only if the discriminant of $f(x)$ is a square in $K$.

**(c)** $(\mathrm{ch}(K) = 2)$ If $K$ has characteristic 2 show that the discriminant of $f(x)$ is always a square. Show that $f(x)$ can be taken to be of the form $x^3 + px + q$ and that the Galois group of $f(x)$ is $A_3$ if and only if the quadratic $x^2 + qx + (p^3 + q^2)$ has a root in $K$ (equivalently, if $(p^3 + q^2)/q^2 \in K$ is in the image of the *Artin–Schreier map* $x \mapsto x^2 - x$ mapping $K$ to $K$).

**(d)** If $K = \mathbb{F}_2(t)$ where $t$ is transcendental over $\mathbb{F}_2$. Prove that the polynomials $x^3 + t^2x + t^3$, $x^3 + (t^2 + t + 1)x + (t^2 + t + 1)$, and $x^3 + (t^2 + t + 1)x + (t^3 + t^2 + t)$ have $A_3$ as Galois group while $x^3 + t^2x + t$ and $x^3 = x + t$ have $S_3$ as Galois group.

**51.** This exercise proves *Sturm's Theorem* determining the number of real roots of a polynomial $f(x) \in \mathbb{R}[x]$ in an interval $[a, b]$. The multiple roots of $f(x)$ are zeros of the g.c.d. of $f(x)$ and its derivative $f'(x)$, and it follows that to determine the real roots of $f(x)$ in $[a, b]$ we may assume that the roots of $f(x)$ are *simple*.

Apply the Euclidean algorithm to $f_0(x) = f(x)$ and its derivative $f_1(x) = f'(x)$ using the *negative* of the remainder at each stage to find a sequence of polynomials $f(x)$, $f'(x)$, $f_2(x), \ldots, f_n(x)$ with

$$f_{i-1}(x) = q_i(x)f_i(x) - f_{i+1}(x) \qquad i = 0, 1, \ldots, n-1$$

where $f_n(x) \in \mathbb{R}$ is a nonzero constant.

**(a)** Prove that consecutive polynomials $f_i(x)$, $f_{i+1}(x)$ for $i = 0, 1, \ldots, n-1$ have no common zeros. [Show that otherwise $f_{i+2}(c) = f_{i+3}(c) = \cdots = 0$, and derive a contradiction.]

**(b)** If $f_i(c) = 0$ for some $i = 0, 1, \ldots, n-1$, prove that one of the two values $f_{i-1}(c)$, $f_{i+1}(c)$ is strictly negative and the other is strictly positive.

For any real number $\alpha$, let $V(\alpha)$ denote the number of sign changes in the *Sturm sequence* of real numbers

$$f(\alpha), f'(\alpha), f_2(\alpha), \ldots, f_n(\alpha),$$

ignoring any 0's that appear (for example $-1, -2, 0, +3, -4$ has signs $--+-$ disregarding the 0, so there are 2 sign changes, the first from $-2$ to $+3$, the second from $+3$ to $-4$).

**(c)** Suppose $\alpha < \beta$ and that all the elements in the Sturm sequences for $\alpha$ and for $\beta$ are nonzero. Prove that unless $f_i(c) = 0$ for some $\alpha < c < \beta$ and some $i = 0, 1, \ldots, n-1$, then the signs of all the elements in these two Sturm sequences are the same, so in particular $V(\alpha) = V(\beta)$.

**(d)** If $f_j(c) = 0$ prove that there is a sufficiently small interval $(\alpha, \beta)$ containing $c$ so that $f_j(x)$ has no zero other than $c$ for $\alpha < x < \beta$.

**(e)** If $j \geq 1$ in (d), prove that the number of sign changes in $f_{j-1}(\alpha)$, $f_j(\alpha)$, $f_{j+1}(\alpha)$ and in $f_{j-1}(\beta)$, $f_j(\beta)$, $f_{j+1}(\beta)$ are the same. [Observe that $f_{j-1}(c)$ and $f_{j+1}(c)$ have opposite signs by (b) and $f_{j-1}(x)$ and $f_{j+1}(x)$ do not change sign in $(\alpha, \beta)$.]

**(f)** If $j = 0$ in (d) show that the number of sign changes in $f(\alpha)$, $f'(\alpha)$ is one more than the number of sign changes in $f(\beta)$, $f'(\beta)$. [If $f'(c) > 0$ then $f(x)$ is increasing at $c$, so that $f(\alpha) < 0$, $f(\beta) > 0$, and $f'(x)$ does not change sign in $(\alpha, \beta)$, so the signs change from $-+$ to $++$. Similarly if $f'(c) < 0$.]

**(g)** Prove *Sturm's Theorem*: if $f(x)$ is a polynomial with real coefficients all of whose real roots are simple then the number of real zeros of $f(x)$ in an interval $[a, b]$ where $f(a)$ and $f(b)$ are both nonzero is given by $V(a) - V(b)$. [Use (c), (e) and (f) to see that as $\alpha$ runs from $a$ to $b$ the number $V(\alpha)$ of sign changes is constant unless $\alpha$ passes through a zero of $f(x)$, in which case it decreases by precisely 1.]

**(h)** Suppose $f(x) = x^5 + px + q \in \mathbb{R}[x]$ has simple roots. Show that the sequence of polynomials above is given by $f(x)$, $5x^4 + p$, $(-4p/5)x + q$, and $-D/(256p^4)$ where $D = 256p^5 + 3125q^4$ is the discriminant of $f(x)$. Conclude for $p > 0$ that $f(x)$ has precisely one real root and for $p < 0$ that $f(x)$ has precisely 1 or 3 real roots depending on whether $D > 0$ or $D < 0$, respectively. [E.g., if $p < 0$ and $D < 0$ then at $-\infty$ the signs are $- + -+$ with 3 sign changes and at $+\infty$ the signs are $+ + ++$ with no sign changes.]

## 14.7 SOLVABLE AND RADICAL EXTENSIONS: INSOLVABILITY OF THE QUINTIC

We now investigate the question of solving for the roots of a polynomial by *radicals*, that is, in terms of the algebraic operations of addition, subtraction, multiplication, division and the extraction of $n^{\text{th}}$ roots. The quadratic formula for the roots of a polynomial of degree 2 is familiar from elementary algebra and we shall derive below similar formulas for the roots of cubic and quartic polynomials. For polynomials of degree $\geq 5$, however, we shall see that such formulas are not possible — this is Abel's Theorem on the insolvability of the general quintic. The reason for this is quite simple: we shall see that a polynomial is solvable by radicals if and only if its Galois group is a solvable group (which explains the terminology) and for $n \geq 5$ the group $S_n$ is not solvable.

We first discuss *simple radical extensions*, namely extensions obtained by adjoining to a field $F$ the $n^{\text{th}}$ root of an element $a$ in $F$. Since all the roots of the polynomial $x^n - a$ for $a \in F$ differ by factors of the $n^{\text{th}}$ roots of unity, adjoining one such root will give a Galois extension if and only if this field contains the $n^{\text{th}}$ roots of unity. Simple radical extensions are best behaved when the base field $F$ already contains the appropriate roots of unity. The symbol $\sqrt[n]{a}$ for $a \in F$ will be used to denote any root of the polynomial $x^n - a \in F[x]$.

**Definition.** The extension $K/F$ is said to be *cyclic* if it is Galois with a cyclic Galois group.

**Proposition 36.** Let $F$ be a field of characteristic not dividing $n$ which contains the $n^{\text{th}}$ roots of unity. Then the extension $F(\sqrt[n]{a})$ for $a \in F$ is cyclic over $F$ of degree dividing $n$.

*Proof:* The extension $K = F(\sqrt[n]{a})$ is Galois over $F$ if $F$ contains the $n^{\text{th}}$ roots of unity since it is the splitting field for $x^n - a$. For any $\sigma \in \text{Gal}(K/F)$, $\sigma(\sqrt[n]{a})$ is another root of this polynomial, hence $\sigma(\sqrt[n]{a}) = \zeta_\sigma \sqrt[n]{a}$ for some $n^{\text{th}}$ root of unity $\zeta_\sigma$. This gives a map

$$\begin{aligned} \text{Gal}(K/F) &\to \mu_n \\ \sigma &\mapsto \zeta_\sigma \end{aligned}$$

where $\mu_n$ denotes the group of $n^{\text{th}}$ roots of unity. Since $F$ contains $\mu_n$, every $n^{\text{th}}$ root of unity is fixed by every element of $\text{Gal}(K/F)$. Hence

$$\begin{aligned} \sigma\tau(\sqrt[n]{a}) &= \sigma(\zeta_\tau \sqrt[n]{a}) \\ &= \zeta_\tau \sigma(\sqrt[n]{a}) \\ &= \zeta_\tau \zeta_\sigma \sqrt[n]{a} = \zeta_\sigma \zeta_\tau \sqrt[n]{a} \end{aligned}$$

which shows that $\zeta_{\sigma\tau} = \zeta_\sigma \zeta_\tau$, so the map above is a homomorphism. The kernel consists precisely of the automorphisms which fix $\sqrt[n]{a}$, namely the identity. This gives an injection of $\text{Gal}(K/F)$ into the cyclic group $\mu_n$ of order $n$, which proves the proposition.

Let now $K$ be any cyclic extension of degree $n$ over a field $F$ of characteristic not dividing $n$ which contains the $n^{\text{th}}$ roots of unity. Let $\sigma$ be a generator for the cyclic group $\text{Gal}(K/F)$.

**Definition.** For $\alpha \in K$ and any $n^{\text{th}}$ root of unity $\zeta$, define the *Lagrange resolvent* $(\alpha, \zeta) \in K$ by

$$(\alpha, \zeta) = \alpha + \zeta\sigma(\alpha) + \zeta^2\sigma^2(\alpha) + \cdots + \zeta^{n-1}\sigma^{n-1}(\alpha).$$

If we apply the automorphism $\sigma$ to $(\alpha, \zeta)$ we obtain

$$\sigma(\alpha, \zeta) = \sigma\alpha + \zeta\sigma^2(\alpha) + \zeta^2\sigma^3(\alpha) + \cdots + \zeta^{n-1}\sigma^n(\alpha)$$

since $\zeta$ is an element of the base field $F$ so is fixed by $\sigma$. We have $\zeta^n = 1$ in $\mu_n$ and $\sigma^n = 1$ in $\mathrm{Gal}(K/F)$ so this can be written

$$\begin{aligned}\sigma(\alpha, \zeta) &= \sigma\alpha + \zeta\sigma^2(\alpha) + \zeta^2\sigma^3(\alpha) + \cdots + \zeta^{-1}\alpha \\ &= \zeta^{-1}(\alpha + \zeta\sigma(\alpha) + \zeta^2\sigma^2(\alpha) + \cdots + \zeta^{n-1}\sigma^{n-1}(\alpha)) \\ &= \zeta^{-1}(\alpha, \zeta). \end{aligned} \tag{14.19}$$

It follows that

$$\sigma(\alpha, \zeta)^n = (\zeta^{-1})^n(\alpha, \zeta)^n = (\alpha, \zeta)^n$$

so that $(\alpha, \zeta)^n$ is fixed by $\mathrm{Gal}(K/F)$, hence is an element of $F$ for any $\alpha \in K$.

Let $\zeta$ be a primitive $n^{\text{th}}$ root of unity. By the linear independence of the automorphisms $1, \sigma, \ldots, \sigma^{n-1}$ (Theorem 7), there is an element $\alpha \in K$ with $(\alpha, \zeta) \neq 0$. Iterating (19) we have

$$\sigma^i(\alpha, \zeta) = \zeta^{-i}(\alpha, \zeta), \qquad i = 0, 1, \ldots,$$

and it follows that $\sigma^i$ does not fix $(\alpha, \zeta)$ for any $i < n$. Hence this element cannot lie in any proper subfield of $K$, so $K = F((\alpha, \zeta))$. Since we proved $(\alpha, \zeta)^n = a \in F$ above, we have $F(\sqrt[n]{a}) = F((\alpha, \zeta)) = K$. This proves the following converse of Proposition 36.

**Proposition 37.** Any cyclic extension of degree $n$ over a field $F$ of characteristic not dividing $n$ which contains the $n^{\text{th}}$ roots of unity is of the form $F(\sqrt[n]{a})$ for some $a \in F$.

*Remark:* The two propositions above form a part of what is referred to as *Kummer theory*. A group $G$ is said to have *exponent* $n$ if $g^n = 1$ for every $g \in G$. Let $F$ be a field of characteristic not dividing $n$ which contains the $n^{\text{th}}$ roots of unity. If we take elements $a_1, \ldots, a_k \in F^\times$ then as in Proposition 36 we can see that the extension

$$F(\sqrt[n]{a_1}, \sqrt[n]{a_2}, \ldots, \sqrt[n]{a_k}) \tag{14.20}$$

is an abelian extension of $F$ whose Galois group is of exponent $n$. Conversely, any abelian extension of exponent $n$ is of this form.

Denote by $(F^\times)^n$ the subgroup of the multiplicative group $F^\times$ consisting of the $n^{\text{th}}$ powers of nonzero elements of $F$. The quotient group $F^\times/(F^\times)^n$ is an abelian group of exponent $n$. The Galois group of the extension in (20) is isomorphic to the group generated in $F^\times/(F^\times)^n$ by the elements $a_1, \ldots, a_k$ and two extensions as in (20) are equal if and only if their associated groups in $F^\times/(F^\times)^n$ are equal.

Hence the (finitely generated) subgroups of $F^\times/(F^\times)^n$ classify the abelian extensions of exponent $n$ over fields containing the $n^{\text{th}}$ roots of unity (and characteristic not

dividing $n$). Such extensions are called *Kummer extensions*.

These results generalize the case $k = 1$ above and can be proved in a similar way.

For simplicity we now consider the situation of a base field $F$ of characteristic 0. As in the previous propositions the results are valid over fields whose characteristics do not divide any of the orders of the roots that will be taken.

**Definition.**

**(1)** An element $\alpha$ which is algebraic over $F$ can be *expressed by radicals* or *solved for in terms of radicals* if $\alpha$ is an element of a field $K$ which can be obtained by a succession of simple radical extensions

$$F = K_0 \subset K_1 \subset \cdots \subset K_i \subset K_{i+1} \subset \cdots \subset K_s = K \tag{14.21}$$

where $K_{i+1} = K_i(\sqrt[n_i]{a_i})$ for some $a_i \in K_i$, $i = 0, 1, \ldots, s-1$. Here $\sqrt[n_i]{a_i}$ denotes some root of the polynomial $x^{n_i} - a_i$. Such a field $K$ will be called a *root extension* of $F$.

**(2)** A polynomial $f(x) \in F[x]$ can be *solved by radicals* if all its roots can be solved for in terms of radicals.

This gives a precise meaning to the intuitive notion that $\alpha$ is obtained by successive algebraic operations (addition, subtraction, multiplication and division) and successive root extractions. For example, the element

$$-1 + \sqrt{17} + \sqrt{2(17 - \sqrt{17})} + 2\sqrt{17 + 3\sqrt{17} - \sqrt{2(17 - \sqrt{17})} - 2\sqrt{2(17 + \sqrt{17})}}$$

encountered at the end of Section 5 (used to construct the regular 17-gon) is expressed by radicals and is contained in the field $K_4$, where

$$\begin{aligned}
K_0 &= \mathbb{Q} \\
K_1 &= K_0(\sqrt{a_0}) \qquad a_0 = 17 \\
K_2 &= K_1(\sqrt{a_1}) \qquad a_1 = 2(17 - \sqrt{17}) \\
K_3 &= K_2(\sqrt{a_2}) \qquad a_2 = 2(17 + \sqrt{17}) \\
K_4 &= K_3(\sqrt{a_3}) \qquad a_3 = 17 + 3\sqrt{17} - \sqrt{2(17 - \sqrt{17})} - 2\sqrt{2(17 + \sqrt{17})}.
\end{aligned}$$

Each of these extensions is a radical extension. The fact that no roots other than square roots are required reflects the fact that the regular 17-gon is constructible by straightedge and compass.

In considering radical extensions one may always adjoin roots of unity, since by definition the roots of unity are radicals. This is useful because then cyclic extensions become radical extensions and conversely. In particular we have:

**Lemma 38.** If $\alpha$ is contained in a root extension $K$ as in (21) above, then $\alpha$ is contained in a root extension which is Galois over $F$ and where each extension $K_{i+1}/K_i$ is cyclic.

*Proof:* Let $L$ be the Galois closure of $K$ over $F$. For any $\sigma \in \mathrm{Gal}(L/F)$ we have the chain of subfields

$$F = \sigma K_0 \subset \sigma K_1 \subset \cdots \subset \sigma K_i \subset \sigma K_{i+1} \subset \cdots \subset \sigma K_s = \sigma K$$

where $\sigma K_{i+1}/\sigma K_i$ is again a simple radical extension (since it is generated by the element $\sigma(\sqrt[n_i]{a_i})$, which is a root of the equation $x^{n_i} - \sigma(a_i)$ over $\sigma(K_i)$). It is easy to see that the composite of two root extensions is again a root extension (if $K'$ is another root extension with subfields $K_i'$, first take the composite of $K_1'$ with the fields $K_0, K_1, \ldots, K_s$, then the composite of these fields with $K_2'$, etc. so that each individual extension in this process is a simple radical extension). It follows that the composite of all the conjugate fields $\sigma(K)$ for $\sigma \in \mathrm{Gal}(L/F)$ is again a root extension. Since this field is precisely $L$, we see that $\alpha$ is contained in a Galois root extension.

We now adjoin to $F$ the $n_i$-th roots of unity for all the roots $\sqrt[n_i]{a_i}$ of the simple radical extensions in the Galois root extension $K/F$, obtaining the field $F'$, say, and then form the composite of $F'$ with the root extension:

$$F \subseteq F' = F'K_0 \subseteq F'K_1 \subseteq \cdots \subseteq F'K_i \subseteq F'K_{i+1} \subseteq \cdots \subseteq F'K_s = F'K.$$

The field $F'K$ is a Galois extension of $F$ since it is the composite of two Galois extensions. The extension from $F$ to $F' = F'K_0$ can be given as a chain of subfields with each individual extension cyclic (this is true for any abelian extension). Each extension $F'K_{i+1}/F'K_i$ is a simple radical extension and since we now have the appropriate roots of unity in the base fields, each of these individual extensions from $F'$ to $F'K$ is a cyclic extension by Proposition 36. Hence $F'K/F$ is a root extension which is Galois over $F$ with cyclic intermediate extensions, completing the proof.

Recall from Section 3.4 (cf. also Section 6.1) that a finite group $G$ is *solvable* if there exists a chain of subgroups

$$1 = G_s \leq G_{s-1} \leq \cdots \leq G_{i+1} \leq G_i \leq \cdots \leq G_0 = G \tag{14.22}$$

with $G_i/G_{i+1}$ cyclic, $i = 0, 1, \ldots, s-1$. We have proved that subgroups and quotient groups of solvable groups are solvable and that if $H \leq G$ and $G/H$ are both solvable, then $G$ is solvable.

We now prove Galois' fundamental connection between solving for the roots of polynomials in terms of radicals and the Galois group of the polynomial. We continue to work over a field $F$ of characteristic 0, but it is easy to see that the proof is valid over any field of characteristic not dividing the order of the Galois group or the orders of the radicals involved.

**Theorem 39.** The polynomial $f(x)$ can be solved by radicals if and only if its Galois group is a solvable group.

*Proof:* Suppose first that $f(x)$ can be solved by radicals. Then each root of $f(x)$ is contained in an extension as in the lemma. The composite $L$ of such extensions is

again of the same type by Proposition 21. Let $G_i$ be the subgroups corresponding to the subfields $K_i$, $i = 0, 1, \ldots, s-1$. Since

$$\text{Gal}(K_{i+1}/K_i) = G_i/G_{i+1} \qquad i = 0, 1, \ldots, s-1$$

it follows that the Galois group $G = \text{Gal}(L/F)$ is a solvable group. The field $L$ contains the splitting field of $f(x)$ so the Galois group of $f(x)$ is a quotient group of the solvable group $G$, hence is solvable.

Suppose now that the Galois group $G$ of $f(x)$ is a solvable group and let $K$ be the splitting field for $f(x)$. Taking the fixed fields of the subgroups in a chain (22) for $G$ gives a chain

$$F = K_0 \subset K_1 \subset \cdots \subset K_i \subset K_{i+1} \subset \cdots \subset K_s = K$$

where $K_{i+1}/K_i$, $i = 0, 1, \ldots, s-1$ is a cyclic extension of degree $n_i$. Let $F'$ be the cyclotomic field over $F$ of all roots of unity of order $n_i$, $i = 0, 1, \ldots, s-1$ and form the composite fields $K_i' = F'K_i$. We obtain a sequence of extensions

$$F \subseteq F' = F'K_0 \subseteq F'K_1 \subseteq \cdots \subseteq F'K_i \subseteq F'K_{i+1} \subseteq \cdots \subseteq F'K_s = F'K.$$

The extension $F'K_{i+1}/F'K_i$ is cyclic of degree dividing $n_i$, $i = 0, 1, \ldots, s-1$ (by Proposition 19). Since we now have the appropriate roots of unity in the base fields, each of these cyclic extensions is a simple radical extension by Proposition 37. Each of the roots of $f(x)$ is therefore contained in the root extension $F'K$ so that $f(x)$ can be solved by radicals.

**Corollary 40.** The general equation of degree $n$ cannot be solved by radicals for $n \geq 5$.

*Proof:* For $n \geq 5$ the group $S_n$ is not solvable as we showed in Chapter 4. The corollary follows immediately from Theorems 32 and 39.

This corollary shows that there is no formula involving radicals analogous to the quadratic formula for polynomials of degree 2 for the roots of a polynomial of degree 5. To give an example of a *specific* polynomial over $\mathbb{Q}$ of degree 5 whose roots cannot be expressed in terms of radicals we must demonstrate a polynomial of degree 5 with rational coefficients having $S_5$ (or $A_5$, which is also not solvable) as Galois group (cf. also Exercise 21, which gives a criterion for the solvability of a quintic).

**Example**

Consider the polynomial $f(x) = x^5 - 6x + 3 \in \mathbb{Q}[x]$. This polynomial is irreducible since it is Eisenstein at 3. The splitting field $K$ for this polynomial therefore has degree divisible by 5, since adjoining one root of $f(x)$ to $\mathbb{Q}$ generates an extension of degree 5. The Galois group $G$ is therefore a subgroup of $S_5$ of order divisible by 5 so contains an element of order 5. The only elements in $S_5$ of order 5 are 5-cycles, so $G$ contains a 5-cycle.

Since $f(-2) = -17$, $f(0) = 3$, $f(1) = -2$, and $f(2) = 23$ we see that $f(x)$ has a real root in each of the intervals $(-2, 0)$, $(0, 1)$ and $(1, 2)$. By the Mean Value Theorem, if there were 4 real roots then the derivative $f'(x) = 5x^4 - 6$ would have at least 3 real zeros, which it does not. Hence these are the only real roots. (This also follows easily by Descartes' rule of signs.) By the Fundamental Theorem of Algebra $f(x)$ has 5 roots in $\mathbb{C}$. Hence $f(x)$ has two complex roots which are not real. Let $\tau$ denote the automorphism of

complex conjugation in $\mathbb{C}$. Since the coefficients of $f(x)$ are real, the two complex roots must be interchanged by $\tau$ (since they are not fixed, not being real). Hence the restriction of complex conjugation to $K$ fixes three of the roots of $f(x)$ and interchanges the other two. As an element of $G$, $\tau|_K$ is therefore a transposition.

It is now a simple exercise to show that any 5-cycle together with any transposition generate all of $S_5$. It follows that $G = S_5$, so the roots of $x^5 - 6x + 3$ cannot be expressed by radicals.

As indicated in this example, a great deal of information regarding the Galois group can be obtained by understanding the *cycle types* of the automorphisms in $G$ considered as a subgroup of $S_n$. In practice this is the most efficient way of determining the Galois groups of polynomials of degrees $\geq 5$ (becoming more difficult the larger the degree, of course, if only because the possible subgroups of $S_n$ are vastly more numerous). We describe this procedure in the next section.

By Theorem 39, any polynomial of degree $n \leq 4$ can be solved by radicals, since $S_n$ is a solvable group for these $n$. For $n = 2$ this is just the familiar quadratic formula. For $n = 3$ the formula is known as *Cardano's Formula* (named for Geronimo Cardano (1501–1576)) and the formula for $n = 4$ can be reduced to this one. The formulas are valid over any field $F$ of characteristic $\neq 2, 3$, which are the characteristics dividing the orders of the radicals necessary and the orders of the possible Galois groups (which are subgroups of $S_3$ and $S_4$). For simplicity we shall derive the formulas over $\mathbb{Q}$.

## Solution of Cubic Equations by Radicals: Cardano's Formulas

From the proof of Theorem 39 and the fact that a composition series for $S_3$ as in equation (22) is given by $1 \leq A_3 \leq S_3$ we should expect that the solution of the cubic

$$f(x) = x^3 + ax^2 + bx + c$$

(or equivalently, under the substitution $x = y - a/3$,

$$g(y) = y^3 + py + q,$$

where

$$p = \frac{1}{3}(3b - a^2) \qquad q = \frac{1}{27}(2a^3 - 9ab + 27c)\ )$$

to involve adjoining the 3rd roots of unity and the formation of Lagrange resolvents involving these roots of unity.

Let $\rho$ denote a primitive 3rd root of unity, so that $\rho^2 + \rho + 1 = 0$. Let the roots of $g(y)$ be $\alpha$, $\beta$, and $\gamma$, so that

$$\alpha + \beta + \gamma = 0$$

(one of the reasons for changing from $f(x)$ to $g(x)$). Over the field $\mathbb{Q}(\sqrt{D}\,)$ where $D$ is the discriminant (computed in the last section) the Galois group of $g(y)$ is $A_3$, i.e., a cyclic group of order 3. If we adjoin $\rho$ then this extension is a radical extension of

degree 3, with generator given by a Lagrange Resolvent, as in the proof of Proposition 37. Consider therefore the elements

$$(\alpha, 1) = \alpha + \beta + \gamma = 0$$
$$\theta_1 = (\alpha, \rho) = \alpha + \rho\beta + \rho^2\gamma$$
$$\theta_2 = (\alpha, \rho^2) = \alpha + \rho^2\beta + \rho\gamma.$$

Note that the sum of these resolvents is

$$\theta_1 + \theta_2 = 3\alpha \tag{14.23}$$

since $1 + \rho + \rho^2 = 0$. Similarly

$$\rho^2\theta_1 + \rho\theta_2 = 3\beta$$
$$\rho\theta_1 + \rho^2\theta_2 = 3\gamma. \tag{14.23'}$$

We also showed in general before Proposition 37 that the cube of these resolvents must lie in $\mathbb{Q}(\sqrt{D}, \rho)$. Expanding $\theta_1^3$ we obtain

$$\alpha^3 + \beta^3 + \gamma^3 + 3\rho(\alpha^2\beta + \beta^2\gamma + \alpha\gamma^2) + 3\rho^2(\alpha\beta^2 + \beta\gamma^2 + \alpha^2\gamma) + 6\alpha\beta\gamma. \tag{14.24}$$

We have

$$\sqrt{D} = (\alpha - \beta)(\alpha - \gamma)(\beta - \gamma)$$
$$= (\alpha^2\beta + \beta^2\gamma + \alpha\gamma^2) - (\alpha\beta^2 + \beta\gamma^2 + \alpha^2\gamma).$$

Using this equation we see that (24) can be written

$$\alpha^3 + \beta^3 + \gamma^3 + 3\rho[\frac{1}{2}(S + \sqrt{D})] + 3\rho^2[\frac{1}{2}(S - \sqrt{D})] + 6\alpha\beta\gamma \tag{14.24'}$$

where for simplicity we have denoted by $S$ the expression

$$(\alpha^2\beta + \beta^2\gamma + \alpha\gamma^2) + (\alpha\beta^2 + \beta\gamma^2 + \alpha^2\gamma).$$

Since $S$ is symmetric in the roots, each of the expressions in (24') is a symmetric polynomial in $\alpha$, $\beta$ and $\gamma$, hence is a polynomial in the elementary symmetric functions $s_1 = 0$, $s_2 = p$, and $s_3 = -q$. After a short calculation one finds

$$\alpha^3 + \beta^3 + \gamma^3 = -3q \qquad S = 3q$$

so that from (24') we find ($\rho + \rho^2 = -1$ and $\rho - \rho^2 = \sqrt{-3}$)

$$\theta_1^3 = -3q + \frac{3}{2}\rho(3q + \sqrt{D}) + \frac{3}{2}\rho^2(3q - \sqrt{D}) - 6q$$
$$= \frac{-27}{2}q + \frac{3}{2}\sqrt{-3D}. \tag{14.25}$$

Similarly, we find

$$\theta_2^3 = \frac{-27}{2}q - \frac{3}{2}\sqrt{-3D}. \tag{14.25'}$$

Equations (25) and (23) essentially give the solutions of our cubic. One small point remains, however, namely the issue of extracting the cube roots of the expressions in (25) to obtain $\theta_1$ and $\theta_2$. There are 3 possible cube roots, which might suggest a total of 9 expressions in (23). This is not the case since $\theta_1$ and $\theta_2$ are not independent (adjoining one of them already gives the Galois extension containing all of the roots). A computation like the one above (but easier) shows that

$$\theta_1\theta_2 = -3p \tag{14.26}$$

showing that the choice of cube root for $\theta_1$ determines $\theta_2$. Using $D = -4p^3 - 27q^2$, we obtain Cardano's explicit formulas, as follows.

Let

$$A = \sqrt[3]{\frac{-27}{2}q + \frac{3}{2}\sqrt{-3D}}$$
$$B = \sqrt[3]{\frac{-27}{2}q - \frac{3}{2}\sqrt{-3D}}$$

where the cube roots are chosen so that $AB = -3p$. Then the roots of the equation

$$y^3 + py + q = 0$$

are

$$\alpha = \frac{A+B}{3} \quad \beta = \frac{\rho^2 A + \rho B}{3} \quad \gamma = \frac{\rho A + \rho^2 B}{3} \tag{14.27}$$

where $\rho = -\frac{1}{2} + \frac{1}{2}\sqrt{-3}$.

**Examples**

**(1)** Consider the cubic equation $x^3 - x + 1 = 0$. The discriminant of this cubic is

$$D = -4(-1)^3 - 27(1)^2 = -23$$

which is not the square of a rational number, so the Galois group for this polynomial is $S_3$. Substituting into the formulas above we have

$$A = \sqrt[3]{\frac{-27}{2} + \frac{3}{2}\sqrt{69}}$$
$$B = \sqrt[3]{\frac{-27}{2} - \frac{3}{2}\sqrt{69}}$$

where we choose $A$ to be the real cube root and then from $AB = 3$ we see that $B$ is also real. The roots of the cubic are given by (27) and we see that there is one real root and two (conjugate) complex roots (which we could have determined without solving for the roots, of course).

**(2)** Consider the equation $x^3 + x^2 - 2x - 1 = 0$. Letting $x = s - 1/3$ the equation becomes $s^3 - \frac{7}{3}s - \frac{7}{27} = 0$. Multiplying through by 27 to clear denominators and letting $y = 3s$ we see that $y$ satisfies the cubic equation

$$y^3 - 21y - 7 = 0.$$

The discriminant $D$ for this cubic is

$$D = -4(-21)^3 - 27(-7)^2 = 3^6 7^2$$

which shows that the Galois group for this (Eisenstein at 7) cubic is $A_3$. Substituting into the formulas above we have

$$A = 3\sqrt[3]{\frac{7}{2} + \frac{21}{2}\sqrt{-3}}$$

$$B = 3\sqrt[3]{\frac{7}{2} - \frac{21}{2}\sqrt{-3}}$$

and the roots of our cubics can be expressed in terms of $A$ and $B$ using the formulas above. This cubic arises from trying to express a primitive $7^{\text{th}}$ root of unity $\zeta_7$ in terms of radicals similar to the explicit formulas for the other roots of unity of small order (cf. the exercises).

In this case we have $g(-5) = -27$, $g(-1) = 13$, $g(0) = -7$ and $g(5) = 13$, so that this cubic has 3 *real* roots. The expressions above for these roots are sums of the conjugates of *complex* numbers. We shall see later that this is necessary, namely that it is impossible to solve for these real roots using only radicals involving real numbers.

A cubic with rational coefficients has either one real root and two complex conjugate imaginary roots or has three real roots. These two cases can be distinguished by the sign of the discriminant:

Suppose in the first case that the roots are $a$ and $b \pm ic$ where $a$, $b$, and $c$ are real and $c \neq 0$. Then

$$\begin{aligned}\sqrt{D} &= [a - (b + ic)][a - (b - ic)][(b + ic) - (b - ic)] \\ &= 2ic[(a - b)^2 + c^2]\end{aligned}$$

is purely imaginary, so that the discriminant $D$ is negative. Then in the formulas for $A$ and $B$ above we may choose both to be real. The first root in (27) is then real and the second two are complex conjugates.

If all three roots are real, then clearly $\sqrt{D}$ is real, so $D \geq 0$ is a nonnegative real number. If $D = 0$ then the cubic has repeated roots. For $D > 0$ (sometimes called the *Casus irreducibilis*), the formulas for the roots involve radicals of nonreal numbers, as in Example 2. We now show that for irreducible cubics this is necessary. The exercises outline the proof of the following generalization: if all the roots of the irreducible polynomial $f(x) \in \mathbb{Q}[x]$ are real and if one of these roots can be expressed by *real* radicals, then the degree of $f(x)$ is a power of 2, the Galois group of $f(x)$ is a 2-group, and the roots of $f(x)$ can be constructed by straightedge and compass.

Suppose that the irreducible cubic $f(x)$ has three real roots and that it were possible to express one of these roots by radicals involving only real numbers. Then the splitting field for the cubic would be contained in a root extension

$$\mathbb{Q} = K_0 \subset K_1 = \mathbb{Q}(\sqrt{D}) \subset \cdots \subset K_i \subset K_{i+1} \subset \cdots \subset K_s = K$$

where each field $K_i$, $i = 0, 1, \ldots, s$, is contained in the real numbers $\mathbb{R}$ and $s \geq 2$ since the quadratic extension $\mathbb{Q}(\sqrt{D})$ cannot contain the root of an irreducible cubic. We have begun this root extension with $\mathbb{Q}(\sqrt{D})$ because over this field the Galois group of the polynomial is cyclic of degree 3.

Note that for any field $F$ the extension $F(\sqrt[mn]{a})$ of $F$ can be obtained by two smaller simple radical extensions: let

$$F_1 = F(\sqrt[n]{a})$$

and let $b = \sqrt[n]{a} \in F_1$, so that

$$F(\sqrt[mn]{a}) = F_1(\sqrt[m]{b}).$$

We may therefore always assume our radical extensions are of the form $F(\sqrt[p]{a})$ where $p$ is a prime.

Suppose now that $F$ is a subfield of the real numbers $\mathbb{R}$ and let $a$ be an element of $F$. Let $p$ be a prime and let $\alpha = \sqrt[p]{a}$ denote a real $p^{\text{th}}$ root of $a$. Then $[F(\sqrt[p]{a}) : F]$ must be either 1 or $p$, as follows. The conjugates of $\alpha$ over $F$ all differ from $\alpha$ by a $p^{\text{th}}$ root of unity. It follows that the constant term of the minimal polynomial of $\alpha$ over $F$ is $\alpha^d \zeta$ where $d = [F(\sqrt[p]{a}) : F]$ is the degree of the minimal polynomial and $\zeta$ is some $p^{\text{th}}$ root of unity. Since $\alpha$ is real and $\alpha^d \zeta \in F$ is real, it follows that $\zeta = \pm 1$, so that $\alpha^d \in F$. Then, if $d \neq p$, $\alpha^d \in F$ and $\alpha^p = a \in F$ implies $\alpha \in F$, so $d = 1$.

Hence we may assume for the radical extensions above that $[K_{i+1} : K_i]$ is a prime $p_i$ and $K_{i+1} = K_i(\sqrt[p_i]{a_i})$ for some $a_i \in K_i$, $i = 0, 1, \ldots, s-1$. In other words, the original tower of real radical extensions can be refined to a tower where each of the successive radical extensions has prime degree.

If any field containing $\sqrt{D}$ contains one of the roots of $f(x)$ then it contains the splitting field for $f(x)$, hence contains all the roots of the cubic. We suppose $s$ is chosen so that $K_{s-1}$ does not contain any of the roots of the cubic.

Consider the extension $K_s/K_{s-1}$. The field $K_s$ contains all the roots of the cubic $f(x)$ and the field $K_{s-1}$ contains none of these roots. It follows that $f(x)$ is irreducible over $K_{s-1}$, so $[K_s : K_{s-1}]$ is divisible by 3. Since we have reduced to the case where this extension degree is a prime, it follows that the extension degree is precisely 3 and that the extension $K_s/K_{s-1}$ is Galois (being the splitting field of $f(x)$ over $K_{s-1}$). Since also $K_s = K_{s-1}(\sqrt[3]{a})$ for some $a \in K_{s-1}$, the Galois extension $K_s$ must also contain the other cube roots of $a$. This implies that $K_s$ contains $\rho$, a primitive $3^{\text{rd}}$ root of unity. This contradicts the assumption that $K_s$ is a subfield of $\mathbb{R}$ and shows that it is impossible to express the roots of this cubic in terms of real radicals only.

## Solution of Quartic Equations by Radicals

Consider now the case of a quartic polynomial

$$f(x) = x^4 + ax^3 + bx^2 + cx + d$$

which under the substitution $x = y - a/4$ becomes the quartic

$$g(y) = y^4 + py^2 + qy + r$$

with

$$p = \frac{1}{8}(-3a^2 + 8b)$$

$$q = \frac{1}{8}(a^3 - 4ab + 8c)$$

$$r = \frac{1}{256}(-3a^4 + 16a^2b - 64ac + 256d).$$

Let the roots of $g(y)$ be $\alpha_1$, $\alpha_2$, $\alpha_3$, and $\alpha_4$. The resolvent cubic introduced in the previous section for this quartic is

$$h(x) = x^3 - 2px^2 + (p^2 - 4r)x + q^2$$

and has roots

$$\theta_1 = (\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4)$$
$$\theta_2 = (\alpha_1 + \alpha_3)(\alpha_2 + \alpha_4)$$
$$\theta_3 = (\alpha_1 + \alpha_4)(\alpha_2 + \alpha_3).$$

The Galois group of the splitting field for $f(x)$ (or $g(y)$) over the splitting field of the resolvent cubic $h(x)$ is the Klein 4-group. Such extensions are biquadratic, which means that it is possible to solve for the roots $\alpha_1$, $\alpha_2$, $\alpha_3$, and $\alpha_4$ in terms of square roots of expressions involving the roots $\theta_1$, $\theta_2$, and $\theta_3$ of the resolvent cubic. In this case we evidently have

$$(\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4) = \theta_1 \qquad (\alpha_1 + \alpha_2) + (\alpha_3 + \alpha_4) = 0$$

which gives

$$\alpha_1 + \alpha_2 = \sqrt{-\theta_1} \qquad \alpha_3 + \alpha_4 = -\sqrt{-\theta_1}.$$

Similarly,

$$\alpha_1 + \alpha_3 = \sqrt{-\theta_2} \qquad \alpha_2 + \alpha_4 = -\sqrt{-\theta_2}$$
$$\alpha_1 + \alpha_4 = \sqrt{-\theta_3} \qquad \alpha_2 + \alpha_3 = -\sqrt{-\theta_3}.$$

An easy computation shows that $\sqrt{-\theta_1}\sqrt{-\theta_2}\sqrt{-\theta_3} = -q$, so that the choice of two of the square roots determines the third. Since $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 0$, if we add the left-hand equations above we obtain $2\alpha_1$, and similarly we may solve for the other roots of $g(y)$. We find

$$2\alpha_1 = \sqrt{-\theta_1} + \sqrt{-\theta_2} + \sqrt{-\theta_3}$$
$$2\alpha_2 = \sqrt{-\theta_1} - \sqrt{-\theta_2} - \sqrt{-\theta_3}$$
$$2\alpha_3 = -\sqrt{-\theta_1} + \sqrt{-\theta_2} - \sqrt{-\theta_3}$$
$$2\alpha_4 = -\sqrt{-\theta_1} - \sqrt{-\theta_2} + \sqrt{-\theta_3}$$

which reduces the solution of the quartic equation to the solution of the associated resolvent cubic.

## EXERCISES

**1.** Use Cardano's Formulas to solve the equation $x^3 + x^2 - 2 = 0$. In particular show that the equation has the real root

$$\frac{1}{3}(\sqrt[3]{26 + 15\sqrt{3}} + \sqrt[3]{26 - 15\sqrt{3}} - 1).$$

Show directly that the roots of this cubic are $1, -1 \pm i$. Explain this by proving that

$$\sqrt[3]{26 + 15\sqrt{3}} = 2 + \sqrt{3} \quad \sqrt[3]{26 - 15\sqrt{3}} = 2 - \sqrt{3}$$

so that

$$\sqrt[3]{26+15\sqrt{3}}+\sqrt[3]{26-15\sqrt{3}}=4.$$

2. Let $\zeta_7$ be a primitive $7^{\text{th}}$ root of unity and let $\alpha=\zeta+\zeta^{-1}$.
   (a) Show that $\zeta_7$ is a root of the quadratic $z^2-\alpha z+1$ over $\mathbb{Q}(\alpha)$.
   (b) Show using the minimal polynomial for $\zeta_7$ that $\alpha$ is a root of the cubic $x^3+x^2-2x-1$.
   (c) Use (a) and (b) together with the explicit solution of the cubic in (b) in the text to express $\zeta_7$ in terms of radicals similar to the expressions given earlier for the other roots of unity of small order. (The complicated nature of the expression explains why we did not include $\zeta_7$ earlier in our list of explicit roots of unity.)
3. Let $F$ be a field of characteristic $\neq 2$. State and prove a necessary and sufficient condition on $\alpha, \beta \in F$ so that $F(\sqrt{\alpha})=F(\sqrt{\beta})$. Use this to determine whether $\mathbb{Q}(\sqrt{1-\sqrt{2}})=\mathbb{Q}(i,\sqrt{2})$.
4. Let $K=\mathbb{Q}(\sqrt[n]{a})$, where $a\in\mathbb{Q}, a>0$ and suppose $[K:\mathbb{Q}]=n$ (i.e., $x^n-a$ is irreducible). Let $E$ be any subfield of $K$ and let $[E:\mathbb{Q}]=d$. Prove that $E=\mathbb{Q}(\sqrt[d]{a})$. [Consider $N_{K/E}(\sqrt[n]{a})\in E$ .]
5. Let $K$ be as in the previous exercise. Prove that if $n$ is odd then $K$ has no nontrivial subfields which are Galois over $\mathbb{Q}$ and if $n$ is even then the only nontrivial subfield of $K$ which is Galois over $\mathbb{Q}$ is $\mathbb{Q}(\sqrt{a})$.
6. Let $L$ be the Galois closure of $K$ in the previous two exercises (i.e., the splitting field of $x^n-a$). Prove that $[L:\mathbb{Q}]=n\varphi(n)$ or $\frac{1}{2}n\varphi(n)$. [Note that $\mathbb{Q}(\zeta_n)\cap K$ is a Galois extension of $\mathbb{Q}$.]
7. (*Kummer Generators for Cyclic Extensions*) Let $F$ be a field of characteristic not dividing $n$ containing the $n^{\text{th}}$ roots of unity and let $K$ be a cyclic extension of degree $d$ dividing $n$. Then $K=F(\sqrt[n]{a})$ for some nonzero $a\in F$. Let $\sigma$ be a generator for the cyclic group $\text{Gal}(K/F)$.
   (a) Show that $\sigma(\sqrt[n]{a})=\zeta\sqrt[n]{a}$ for some primitive $d^{\text{th}}$ root of unity $\zeta$.
   (b) Suppose $K=F(\sqrt[n]{a})=F(\sqrt[n]{b})$. Use (a) to show that $\dfrac{\sigma(\sqrt[n]{a})}{\sqrt[n]{a}}=\left(\dfrac{\sigma(\sqrt[n]{b})}{\sqrt[n]{b}}\right)^i$ for some integer $i$ relatively prime to $d$. Conclude that $\sigma$ fixes the element $\dfrac{\sqrt[n]{a}}{(\sqrt[n]{b})^i}$ so this is an element of $F$.
   (c) Prove that $K=F(\sqrt[n]{a})=F(\sqrt[n]{b})$ if and only if $a=b^ic^n$ and $b=a^jd^n$ for some $c,d\in F$, i.e., if and only if $a$ and $b$ generate the same subgroup of $F^\times$ modulo $n^{\text{th}}$ powers.
8. Let $p, q$ and $r$ be primes in $\mathbb{Z}$ with $q\neq r$. Let $\sqrt[p]{q}$ denote any root of $x^p-q$ and let $\sqrt[p]{r}$ denote any root of $x^p-r$. Prove that $\mathbb{Q}(\sqrt[p]{q})\neq\mathbb{Q}(\sqrt[p]{r})$.
9. (*Artin–Schreier Extensions*) Let $F$ be a field of characteristic $p$ and let $K$ be a cyclic extension of $F$ of degree $p$. Prove that $K=F(\alpha)$ where $\alpha$ is a root of the polynomial $x^p-x-a$ for some $a\in F$. [Note that $\text{Tr}_{K/F}(-1)=0$ since $F$ is of characteristic $p$ so that $-1=\alpha-\sigma\alpha$ for some $\alpha\in K$ where $\sigma$ is a generator of $\text{Gal}(K/F)$ by Exercise 26 of Section 2. Show that $a=\alpha^p-\alpha$ is an element of $F$.] Note that since $F$ contains the $p^{\text{th}}$ roots of unity (namely, 1) that this completes the description of all cyclic extensions of prime degree $p$ over fields containing the $p^{\text{th}}$ roots of unity in all characteristics.
10. Let $K=\mathbb{Q}(\zeta_p)$ be the cyclotomic field of $p^{\text{th}}$ roots of unity for the prime $p$ and let

$G = \mathrm{Gal}(K/\mathbb{Q})$. Let $\zeta$ denote any $p^{\text{th}}$ root of unity. Prove that $\sum_{\sigma\in G}\sigma(\zeta)$ (the trace from $K$ to $\mathbb{Q}$ of $\zeta$) is $-1$ or $p-1$ depending on whether $\zeta$ is or is not a primitive $p^{\text{th}}$ root of unity.

**11.** (*The Classical Gauss Sum*) Let $K = \mathbb{Q}(\zeta_p)$ be the cyclotomic field of $p^{\text{th}}$ roots of unity for the odd prime $p$, viewed as a subfield of $\mathbb{C}$, and let $G = \mathrm{Gal}(K/\mathbb{Q})$. Let $H$ denote the subgroup of index 2 in the cyclic group $G$. Define $\eta_0 = \sum_{\tau\in H}\tau(\zeta_p)$, $\eta_1 = \sum_{\tau\in\sigma H}\tau(\zeta_p)$, where $\sigma$ is a generator of $\mathrm{Gal}(K/\mathbb{Q})$ (the two *periods* of $\zeta_p$ with respect to $H$, i.e., the sum of the conjugates of $\zeta_p$ with respect to the two cosets of $H$ in $G$, cf. Section 5).

**(a)** Prove that $\sigma(\eta_0) = \eta_1$, $\sigma(\eta_1) = \eta_0$ and that

$$\eta_0 = \sum_{a=\text{square}} \zeta_p^a \quad , \quad \eta_1 = \sum_{b\neq\text{square}} \zeta_p^b,$$

where the sums are over the squares and nonsquares (respectively) in $(\mathbb{Z}/p\mathbb{Z})^\times$. [Observe that $H$ is the subgroup of squares in $(\mathbb{Z}/p\mathbb{Z})^\times$.]

**(b)** Prove that $\eta_0+\eta_1 = (\zeta_p, 1) = -1$ and $\eta_0-\eta_1 = (\zeta_p, -1)$ where $(\zeta_p, 1)$ and $(\zeta_p, -1)$ are two of the Lagrange resolvents of $\zeta_p$.

**(c)** Let $g = \sum_{i=0}^{p-1}\zeta_p^{i^2}$ (the classical *Gauss sum*). Prove that

$$g = (\zeta_p, -1) = \sum_{i=0}^{p-2}(-1)^i\,\sigma^i(\zeta_p).$$

**(d)** Prove that $\tau g = g$ if $\tau \in H$ and $\tau g = -g$ if $\tau \notin H$. Conclude in particular that $[\mathbb{Q}(g) : \mathbb{Q}] = 2$. Recall that complex conjugation is the automorphism $\sigma_{-1}$ on $K$ (cf. Exercise 7 of Section 5). Conclude that $\bar{g} = g$ if $-1$ is a square mod $p$ (i.e., if $p \equiv 1 \bmod 4$) and $\bar{g} = -g$ if $-1$ is not a square mod $p$ (i.e., if $p \equiv 3 \bmod 4$) where $\bar{g}$ denotes the complex conjugate of $g$.

**(e)** Prove that $g\bar{g} = p$. [The complex conjugate of a root of unity is its reciprocal. Then $\bar{g} = \sum_{j=0}^{p-2}(-1)^j\,(\sigma^j(\zeta_p))^{-1}$ gives

$$g\bar{g} = \sum_{i,j=0}^{p-2}(-1)^i(-1)^j\frac{\sigma^i(\zeta_p)}{\sigma^j(\zeta_p)} = \sum_{i,j=0}^{p-2}(-1)^{i-j}\sigma^j\left[\frac{\sigma^{i-j}(\zeta_p)}{\zeta_p}\right]$$

$$= \sum_{k=0}^{p-2}(-1)^k\sum_{j=0}^{p-2}\sigma^j\left[\frac{\sigma^k(\zeta_p)}{\zeta_p}\right]$$

where $k = i-j$. If $k = 0$ the element $\dfrac{\sigma^k(\zeta_p)}{\zeta_p}$ is 1, and if $k \neq 0$ then this is a primitive $p^{\text{th}}$ root of unity. Use the previous exercise to conclude that the inner sum is $p-1$ when $k = 0$ and is $-1$ otherwise.]

**(f)** Conclude that $g^2 = (-1)^{(p-1)/2}p$ and that $\mathbb{Q}(\sqrt{(-1)^{(p-1)/2}p}\,)$ is the unique quadratic subfield of $\mathbb{Q}(\zeta_p)$. (Cf. also Exercise 33 of Section 6.)

**12.** Let $L$ be the Galois closure of the finite extension $\mathbb{Q}(\alpha)$ of $\mathbb{Q}$. For any prime $p$ dividing the order of $\mathrm{Gal}(L/\mathbb{Q})$ prove there is a subfield $F$ of $L$ with $[L : F] = p$ and $L = F(\alpha)$.

**13.** Let $F$ be a subfield of the real numbers $\mathbb{R}$. Let $a$ be an element of $F$ and let $K = F(\sqrt[n]{a}\,)$ where $\sqrt[n]{a}$ denotes a real $n^{\text{th}}$ root of $a$. Prove that if $L$ is any Galois extension of $F$ contained in $K$ then $[L : F] \leq 2$.

**14.** This exercise shows that in general it is necessary to use complex numbers when expressing real roots in terms of radicals and generalizes the *Casus irreducibilis* of cubic equations.

Let $f(x) \in \mathbb{Q}[x]$ be an irreducible polynomial all of whose roots are real. Suppose further that one of the roots, $\alpha$, of $f(x)$ can be expressed in terms of *real* radicals (i.e., there is a root extension of real fields $\mathbb{Q} = K_0 \subset K_1 \subset \ldots \subset K_m \subset \mathbb{R}$ with $K_{i+1} = K_i(\sqrt[n_i]{a_i})$, $i = 1, 2, \ldots, m-1$, for some integers $n_i$ and some $a_i \in K_i$ and $\alpha \in K_m$). Prove that the Galois group of $f(x)$ is a 2-group. Conclude in particular that the degree of $f(x)$ is a power of 2 and that the real roots of such a polynomial can be expressed entirely in terms of real radicals if and only if these roots can be constructed by straightedge and compass. [The argument is similar to the case of cubics. Let $L \in \mathbb{R}$ be the Galois closure of $\mathbb{Q}(\alpha)$ and suppose the order of $\mathrm{Gal}(L/\mathbb{Q})$ is divisible by some odd prime $p$. Let $F$ be a subfield of $L$ with $[L : F] = p$ and $L = F(\alpha)$ (by Exercise 12) and consider the composite fields $K_i' = FK_i$, $i = 0, 1, \ldots, m$. These are again real radical extensions and by the argument in the text for the *Casus irreducibilis*, we may assume each $[K_{i+1}' : K_i']$ is a prime. Since $\alpha \notin F = FK_0$, there is an integer $s$ with $\alpha \notin K_{s-1}'$, $\alpha \in K_s'$. Since the extensions are of prime degree, we have $K_s' = K_{s-1}'(\alpha)$. Since $L = F(\alpha)$ is Galois of degree $p$, $K_s'$ is a Galois extension of $K_{s-1}'$ of degree $p$, contradicting the previous exercise.]

**15.** *('Cardano's Formulas' for a Cubic in Characteristic 2)* Suppose $f(x) = x^3 + px + q$ is an irreducible cubic over a field of characteristic 2. Let $\rho$ be a primitive 3rd root of unity and let $\theta$, $\theta'$ be the roots of the quadratic $x^2 + qx + (p^3 + q^2)$ (cf. Exercise 50 of Section 6). Let $\theta_1$ and $\theta_2$ be cube roots of $\rho q + \theta$ and $\rho q + \theta'$, respectively, where the cube roots are chosen so that $\theta_1\theta_2 = p$. Prove that the roots of $f(x)$ are given by $\alpha = \theta_1 + \theta_2$, $\beta = \rho\alpha + \theta_1$, and $\gamma = \rho\alpha + \theta_2 = \alpha + \beta$.

**16.** Let $a$ be a nonzero rational number.

**(a)** Determine when the extension $\mathbb{Q}(\sqrt{ai})$ ($i^2 = -1$) is of degree 4 over $\mathbb{Q}$.

**(b)** When $K = \mathbb{Q}(\sqrt{ai})$ is of degree 4 over $\mathbb{Q}$ show that $K$ is Galois over $\mathbb{Q}$ with the Klein 4-group as Galois group. In this case determine the quadratic extensions of $\mathbb{Q}$ contained in $K$.

**17.** Let $D \in \mathbb{Z}$ be a squarefree integer and let $a \in \mathbb{Q}$ be a nonzero rational number. Show that $\mathbb{Q}(\sqrt{a\sqrt{D}})$ cannot be a cyclic extension of degree 4 over $\mathbb{Q}$.

**18.** Let $D \in \mathbb{Z}$ be a squarefree integer and let $a \in \mathbb{Q}$ be a nonzero rational number. Prove that if $\mathbb{Q}(\sqrt{a\sqrt{D}})$ is Galois over $\mathbb{Q}$ then $D = -1$.

**19.** Let $D \in \mathbb{Z}$ be a squarefree integer and let $K = \mathbb{Q}(\sqrt{D})$.

**(a)** Prove that if $D = s^2 + t^2$ is the sum of two rational squares then there exists an extension $L/\mathbb{Q}$ containing $K$ which is Galois over $\mathbb{Q}$ with a cyclic Galois group of order 4. [Consider the extension $\mathbb{Q}(\sqrt{D + s\sqrt{D}})$.] (Note also that $D$ is the sum of two rational squares if and only if $D$ is also the sum of two integer squares, so one may assume $s$ and $t$ are integral without loss.)

**(b)** Prove conversely that if $K$ can be embedded in a cyclic extension $L$ of degree 4 as in (a) then $D$ is the sum of two squares. [One approach: (i) observe first that $L$ is quadratic over $K$, so $L = K(\sqrt{a + b\sqrt{D}})$ for some $a, b \in \mathbb{Q}$, (ii) show that $L$ contains the quadratic subfield $\mathbb{Q}(\sqrt{a^2 - b^2D})$, which must be $\mathbb{Q}(\sqrt{D})$ if $L/\mathbb{Q}$ is cyclic, and use Exercise 7.]

**(c)** Conclude in particular that $\mathbb{Q}(\sqrt{3})$ is not a subfield of any cyclic extension of degree 4 over $\mathbb{Q}$. Similarly conclude that the fields $\mathbb{Q}(\sqrt{D})$ for squarefree integers $D < 0$ are never contained in cyclic extensions of degree 4 over $\mathbb{Q}$ (this gives an alternate proof for Exercise 19, Section 6).

**20.** Let $p$ be a prime. Show that any solvable subgroup of $S_p$ of order divisible by $p$ is

contained in the normalizer of a Sylow $p$-subgroup of $S_p$ (a Frobenius group of order $p(p-1)$). Conclude that an irreducible polynomial $f(x) \in \mathbb{Q}[x]$ of degree $p$ is solvable by radicals if and only if its Galois group is contained in the Frobenius group of order $p(p-1)$. [Let $G \leq S_p$ be a solvable subgroup of order divisible by $p$. Then $G$ contains a $p$-cycle, hence is transitive on $\{1, 2, \ldots, p\}$. Let $H < G$ be the stabilizer in $G$ of the element 1, so $H$ has index $p$ in $G$. Show that $H$ contains no nontrivial normal subgroups of $G$ (note that the conjugates of $H$ are the stabilizers of the other points). Let $G^{(n-1)}$ be the last nontrivial subgroup in the derived series for $G$. Show that $H \cap G^{(n-1)} = 1$ and conclude that $|G^{(n-1)}| = p$, so that the Sylow $p$-subgroup of $G$ (which is also a Sylow $p$-subgroup in $S_p$) is normal in $G$.]

**21.** (*Criterion for the Solvability of a Quintic*) By the previous exercise, an irreducible polynomial $f(x)$ in $\mathbb{Q}[x]$ of degree 5 can be solved by radicals if and only if its Galois group (considered as a subgroup of $S_5$) is contained in the Frobenius group of order 20. It is known that this is the case if and only if an associated polynomial $g(x)$ of degree 6 has a rational root (cf. Dummit, *Solving Solvable Quintics*, Math. Comp., 57(1991), pp. 387–401). If the quintic is in the general form (where a translation is performed so that the coefficient of $x^4$ is zero)

$$f(x) = x^5 + px^3 + qx^2 + rx + s \qquad p, q, r, s \in \mathbb{Q}$$

then the associated polynomial of degree 6 is

$$\begin{aligned} g(x) = x^6 &+ 8rx^5 + (2pq^2 - 6p^2r + 40r^2 - 50qs)\,x^4 \\ &+ (-2q^4 + 21pq^2r - 40p^2r^2 + 160r^3 - 15p^2qs - 400qrs + 125ps^2)\,x^3 \\ &+ (p^2q^4 - 8q^4r + 9p^4r^2 - 136p^2r^3 + 625q^2s^2 + 400r^4 - 6p^3q^2r \\ &\quad + 76pq^2r^2 - 50pq^3s - 1400qr^2s + 500prs^2 + 90p^2qrs)\,x^2 \\ &+ (-108p^5s^2 + 32p^4r^3 - 256p^2r^4 - 3125s^4 + 512r^5 - 2pq^6 + 3q^4r^2 \\ &\quad - 58q^5s + 2750q^2rs^2 - 31p^3q^3s - 500pr^2s^2 + 19p^2q^4r \\ &\quad - 51p^3q^2r^2 + 76pq^2r^3 - 2400qr^3s - 325p^2q^2s^2 + 525p^3rs^2 \\ &\quad + 625pqs^3 + 117p^4qrs + 105pq^3rs + 260p^2qr^2s)\,x \\ &+ (q^8 + 256r^6 + 17q^4r^3 - 27p^7s^2 - 4p^6r^3 + 48p^4r^4 - 192p^2r^5 \\ &\quad + 3125p^2s^4 - 9375rs^4 - 1600qr^4s - 99p^5rs^2 - 125pq^4s^2 \\ &\quad - 124q^5rs + 3250q^2r^2s^2 - 2000pr^3s^2 - 13pq^6r + p^5q^2r^2 \\ &\quad + 65p^2q^4r^2 - 128p^3q^2r^3 - 16pq^2r^4 - 4p^5q^3s - 12p^2q^5s \\ &\quad - 150p^4q^2s^2 + 1200p^3r^2s^2 + 18p^6qrs + 12p^3q^3rs + 196p^4qr^2s \\ &\quad + 590pq^3r^2s - 160p^2qr^3s - 725p^2q^2rs^2 - 1250pqrs^3). \end{aligned}$$

In the particular case where $f(x) = x^5 + Ax + B$ this polynomial is simply

$$g(x) = x^6 + 8Ax^5 + 40A^2x^4 + 160A^3x^3 + 400A^4x^2 + (512A^5 - 3125B^4)x - 9375AB^4 + 256A^6.$$

**(a)** Use this criterion to prove that the Galois group over $\mathbb{Q}$ of the polynomial $x^5 - 5x + 12$ is the dihedral group of order 10. [Show the associated sixth degree polynomial is

$$x^6 - 40x^5 + 1000x^4 - 20000x^3 + 250000x^2 - 66400000x + 976000000$$

and has $x = 40$ as a rational root. Cf. also Exercise 35 in Section 6.]

**(b)** Use this criterion to prove that $x^5 - x - 1$ is not solvable by radicals.

## 14.8 COMPUTATION OF GALOIS GROUPS OVER $\mathbb{Q}$

In the determination of the Galois groups of polynomials of degrees $\leq 4$ in Section 6 and in the determination of the Galois group of the polynomial $x^5 - 6x + 3$ in the previous section we observed that it was possible to obtain useful information regarding the Galois group from the *cycle types* of the automorphisms as elements in $S_n$. This is very useful in computing Galois groups of polynomials over $\mathbb{Q}$ and we now briefly describe the theoretical justification.

Let $f(x)$ be a polynomial with rational coefficients. In determining the Galois group of $f(x)$ we may assume that $f(x)$ is separable and has integer coefficients. Then the discriminant $D$ of $f(x)$ is an integer and is nonzero.

For any prime $p$, consider the reduction $\overline{f}(x) \in \mathbb{F}_p[x]$ of $f(x)$ modulo $p$. If $p$ divides $D$ then the reduced polynomial $\overline{f}(x)$ has discriminant $\overline{D} = 0$ in $\mathbb{F}_p$, so is not separable.

If $p$ does not divide $D$, then $\overline{f}(x)$ is a separable polynomial over $\mathbb{F}_p$ and we can factor $\overline{f}(x)$ into distinct irreducibles

$$\overline{f}(x) = \overline{f}_1(x)\overline{f}_2(x) \cdots \overline{f}_k(x) \qquad \text{in } \mathbb{F}_p[x].$$

Let $n_i$ be the degree of $\overline{f}_i(x)$, $i = 1, 2, \ldots, k$.

The importance of this reduction is provided by the following theorem from algebraic number theory which is an elementary consequence of the study of the arithmetic in finite extensions of $\mathbb{Q}$ (and which we take for granted).

**Theorem.** For any prime $p$ not dividing the discriminant $D$ of $f(x) \in \mathbb{Z}[x]$, the Galois group over $\mathbb{F}_p$ of the reduction $\overline{f}(x) = f(x) \pmod{p}$ is permutation group isomorphic to a subgroup of the Galois group over $\mathbb{Q}$ of $f(x)$.

The meaning of the statement "permutation group isomorphic" in the theorem is that not only is the Galois group of the reduction $\overline{f}(x)$ mod $p$ of $f(x)$ isomorphic to a subgroup of the Galois group of $f(x)$ but that there is an ordering of the roots of $\overline{f}(x)$ and of $f(x)$ (depending on $p$) so that under this isomorphism the action of the corresponding automorphisms as permutations of these roots is the same. In particular there are automorphisms in the Galois group of $f(x)$ with the same cycle types as the automorphisms of $\overline{f}(x)$.

The Galois group of $\overline{f}(x)$ is a *cyclic* group since every finite extension of $\mathbb{F}_p$ is a cyclic extension. Let $\sigma$ be a generator for this Galois group over $\mathbb{F}_p$ (for example, the Frobenius automorphism). The roots of $\overline{f}_1(x)$ are permuted amongst themselves by the Galois group, and given any two of these roots there is a Galois automorphism taking the first root to the second (recall that the group is said to be *transitive* on the roots when this is the case). Similarly, the Galois group permutes the roots of each of the factors $\overline{f}_i(x)$, $i = 1, 2, \ldots, k$ transitively. Since these factors are relatively prime we also see that no root of one factor is mapped to a root of any other factor by any element of the Galois group.

View $\sigma$ as an element in $S_n$ by labelling the $n$ roots of $\overline{f}(x)$ and consider the cycle decomposition of $\sigma$, which is a product of $k$ distinct permutations since $\sigma$ permutes

the roots of each of the factors $\overline{f}_i(x)$ amongst themselves. By the observations we just made, the action of $\sigma$ on the roots of $\overline{f}_1(x)$ must be a cycle of length $n_i$ since otherwise the powers of $\sigma$ could not be transitive on the roots of $\overline{f}_1(x)$. Similarly the action of $\sigma$ on the roots of $\overline{f}_i(x)$ gives a cycle of length $n_i$, $i = 1, 2, \ldots, k$.

We see that the automorphism $\sigma$ generating the Galois group of $\overline{f}(x)$ has cycle decomposition $(n_1, n_2, \ldots, n_k)$ where $n_1, n_2, \ldots, n_k$ are the degrees of the irreducible factors of $f(x)$ reduced modulo $p$, which gives us the following result.

**Corollary 41.** For any prime $p$ not dividing the discriminant of $f(x) \in \mathbb{Z}[x]$, the Galois group of $f(x)$ over $\mathbb{Q}$ contains an element with cycle decomposition $(n_1, n_2, \ldots, n_k)$ where $n_1, n_2, \ldots, n_k$ are the degrees of the irreducible factors of $f(x)$ reduced modulo $p$.

**Example**

Consider the polynomial $x^5 - x - 1$. The discriminant of this polynomial is $2869 = 19 \cdot 151$ so we reduce at primes $\neq 19, 151$. Reducing mod 2 the polynomial $x^5 - x - 1$ factors as $(x^2 + x + 1)(x^3 + x^2 + 1) \pmod 2$ so the Galois group has a (2,3)-cycle. Cubing this element we see the Galois group contains a transposition.

Reducing mod 3 the polynomial is irreducible, as follows: $x^5 - x - 1$ has no roots mod 3 so if it were reducible mod 3 then it would have an irreducible quadratic factor, hence would have a factor in common with $x^9 - x$ (which is the product of all irreducible polynomials of degrees 1 and 2 over $\mathbb{F}_3$), hence a factor in common with either $x^4 - 1$ or $x^4 + 1$, hence a factor in common with either $x^5 - x$ or $x^5 + x$, hence a factor in common with either $-1$ or $2x + 1$ which it obviously does not. This shows both that $x^5 - x - 1$ is irreducible in $\mathbb{Z}[x]$ and that there is a 5-cycle in its Galois group.

Since $S_5$ is generated by any 5-cycle and any transposition, it follows that the Galois group of $x^5 - x - 1$ is $S_5$ (so in particular this polynomial cannot be solved by radicals, (cf. Exercise 21 of Section 7).

The arguments in the example above indicate how to construct polynomials with $S_n$ as Galois group. We use the fact that a transitive subgroup of $S_n$ containing a transposition and an $n - 1$-cycle is $S_n$. Let $f_1$ be an irreducible polynomial of degree $n$ over $\mathbb{F}_2$. Let $f_2 \in \mathbb{F}_3[x]$ be the product of an irreducible polynomial of degree 2 with irreducible polynomials of odd degree (for example, an irreducible polynomial of degree $n - 3$ and $x$ if $n$ is even and an irreducible polynomial of degree $n - 2$ if $n$ is odd). Let $f_3 \in \mathbb{F}_5[x]$ be the product of $x$ with an irreducible polynomial of degree $n - 1$. Finally, let $f(x) \in \mathbb{Z}[x]$ be any polynomial with

$$\begin{aligned} f(x) &\equiv f_1(x) \pmod 2 \\ &\equiv f_2(x) \pmod 3 \\ &\equiv f_3(x) \pmod 5. \end{aligned}$$

The reduction of $f(x)$ mod 2 shows that $f(x)$ is irreducible in $\mathbb{Z}[x]$, hence the Galois group is transitive on the $n$ roots of $f(x)$. Raising the element given by the factorization of $f(x)$ mod 3 to a suitable odd power shows the Galois group contains a transposition. The factorization mod 5 shows the Galois group contains an $n - 1$-cycle, hence the Galois group is $S_n$.

**Proposition 42.** For each $n \in \mathbb{Z}^+$ there exist infinitely many polynomials $f(x) \in \mathbb{Z}[x]$ with $S_n$ as Galois group over $\mathbb{Q}$.

There are extremely efficient algorithms for factoring polynomials $f(x) \in \mathbb{Z}[x]$ modulo $p$ (cf. Exercises 12 to 17 of Section 3), so the corollary above is an effective procedure for determining some of the cycle types of the elements of the Galois group. In using Corollary 41 some care should be taken not to assume that a *particular* cycle is an element of the Galois group. For example, one factorization might imply the existence of a (2,2) cycle, say (12)(34) and another factorization imply the existence of a transposition. One cannot conclude that the transposition is necessarily (12), however (nor (34), nor (13), etc.). The choice of (12)(34) to represent the first cycle fixes a particular ordering on the roots and this may not be the ordering with respect to which the transposition appears as (12).

Corollary 41 is particularly efficient in determining when the Galois group is large (e.g., $S_n$), since a transitive group containing sufficiently many cycle types must be $S_n$ (for example, a transitive subgroup of $S_n$ containing a transposition and an $n-1$-cycle is $S_n$, as used above). The most difficult Galois groups to determine in this way are the *small* Galois groups (e.g., a cyclic group of order $n$), since factorization after factorization will produce only elements of orders dividing $n$ and one is not sure whether there will be some $p$ yet to come producing a cycle type inconsistent with the assumption of a cyclic Galois group. If one could "compute forever" one could at least be sure of the precise distribution of cycle types among the elements of the Galois group in the following sense: suppose the Galois group $G \subseteq S_n$ has order $N$ and that there are $n_T$ elements of $G$ with cycle type $T$ (e.g., (2,2)-cycles, transpositions, etc.) so that the "density" of cycle type $T$ in $G$ is $d_T = n_T/N$. Then it is possible to define a density on the set of prime numbers (so that it makes sense to speak of "1/2" the primes, etc.) and we have the following result (which relies on the Tchebotarov Density Theorem in algebraic number theory).

**Theorem.** The density of primes $p$ for which $f(x)$ splits into type $T$ modulo $p$ is precisely $d_T$.

This says that if we knew the factorization of $f(x)$ modulo every prime we could at least determine the number of elements of $G$ with a given cycle type. Unfortunately, even this would not be sufficient to determine $G$ (up to isomorphism): it is known that there are nonisomorphic groups containing the same number of elements of all cycle types (there are two nonisomorphic groups of order 96 in $S_8$ both having cycle type distributions: 1 1-cycle, 6 (2,2)-cycles, 13 (2,2,2,2)-cycles, 32 (3,3)-cycles, 12 (4,4)-cycles, 32 (2,6)-cycles). There are infinitely many such examples (the regular representation of the elementary abelian group of order $p^3$ and for the nonabelian group of order $p^3$ of exponent $p$ give two nonisomorphic groups in $S_{p^3}$ whose nonidentity elements are all the product of $p^2$ $p$-cycles for any prime $p$).

In practice one uses the factorizations of $f(x)$ modulo small primes to get an idea of the probable Galois group (based on the previous result). One then tries to prove this is indeed the Galois group — often a difficult problem. For polynomials of small degree, definitive algorithms exist, based in part on the computation of *resolvent* polynomials.

These are analogues of the cubic resolvent used in the previous sections to determine the Galois group of quartic polynomials. These resolvent polynomials have rational coefficients and have as roots certain combinations of the roots of $f(x)$ (similar to the combinations $(\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4)$ for the cubic resolvent). One then determines the factorization of these resolvent polynomials to obtain information on the Galois group of $f(x)$ — for example the existence of a linear factor implies the Galois group lies in the stabilizer in $S_n$ of the combination of the roots of $f(x)$ chosen (for example, the dihedral group of order 8 for our resolvent cubic). It should be observed, however, that the degree of the resolvent polynomials constructed, unlike the situation of the resolvent cubic for quartic polynomials, are in general much larger than the degree of $f(x)$. The effectiveness of this computational technique also depends heavily on the explicit knowledge of the possible transitive subgroups of $S_n$. For $n = 2, 3, \ldots, 8$ the number of isomorphism classes of transitive subgroups of $S_n$ is 1, 2, 5, 5, 16, 7, 50, respectively. There is a great deal of interest in the computation of Galois groups, motivated in part by the problem of determining which groups occur as Galois groups over $\mathbb{Q}$.

We illustrate these techniques with some easier examples (from *The Computation of Galois Groups*, L. Soicher, Master's Thesis, Concordia University, Montreal, 1981).

**Examples**

**(1)** There are 5 isomorphism classes of transitive subgroups of $S_5$ given by the groups $Z_5$, $D_{10}$, $F_{20}$, the so-called Frobenius group of order 20 (the Galois group of $x^5 - 2$ with generators (1 2 3 4 5) and (2 3 5 4) in $S_5$), $A_5$ and $S_5$. The cycle type distributions for these groups are as follows:

| cycle type : | 1 | 2 | (2, 2) | 3 | (2, 3) | 4 | 5 |
|---|---|---|---|---|---|---|---|
| $Z_5$ | 1 | | | | | | 4 |
| $D_{10}$ | 1 | | 5 | | | | 4 |
| $F_{20}$ | 1 | | 5 | | | 10 | 4 |
| $A_5$ | 1 | | 15 | 20 | | | 24 |
| $S_5$ | 1 | 10 | 15 | 20 | 20 | 30 | 24. |

Given this information, the irreducibility of $x^5 - x - 1$ (giving the transitivity on the 5 roots) and the cycle type (2,3) immediately shows that the Galois group of $x^5 - x - 1$ is $S_5$.

Consider now the polynomial $x^5 + 15x + 12$. The discriminant is $2^{10}3^45^5$ so the Galois group is not contained in $A_5$. There are two possibilities: $S_5$ or $F_{20}$. One can easily determine which is more likely by factoring the polynomial modulo a number of small primes and comparing the distribution of cycle types with those in the table above. This does not *prove* the probable Galois group is actually correct. To decide which of $S_5$ and $F_{20}$ is correct one can compute the resolvent polynomial $R(x)$ of degree 15 whose roots are the distinct permutations under $S_5$ of $(\alpha_1 + \alpha_2 - \alpha_3 - \alpha_4)^2$ for 4 of the roots $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ of $f(x)$. By definition, $S_5$ is transitive on the roots of $R(x)$ and it is not difficult to check using the explicit generators for $F_{20}$ given above that $F_{20}$ is not transitive on these 15 values. It follows that $R(x)$ will be a reducible polynomial over $\mathbb{Q}$ if and only if the Galois group of the quintic is $F_{20}$. One finds that for $x^5 + 15x + 12$ the resolvent polynomial $R(x)$ factors into a polynomial of degree 5 and a polynomial of degree 10, hence the Galois group for this quintic is $F_{20}$. One

can also use Exercise 21 of the previous section (cf. Exercise 6), which is also based on the computation of a related resolvent polynomial.

**(2)** Consider the polynomial $x^7 - 14x^5 + 56x^3 - 56x + 22$. The discriminant is computed to be $2^6 7^{10}$ so the Galois group is contained in $A_7$.

Factoring the polynomial for the 42 primes not equal to 7 between 3 and 193 gives a cycle type distribution of 1 1-cycle (2.38 %), 30 (3,3)-cycles (71.43 %), 11 7-cycles (26.19 %). There are 7 isomorphism classes of transitive subgroups of $S_7$, 4 of them contained in $A_7$. Of these, one contains no (3,3)-cycles, which leaves the three possibilities $A_7$, $GL_3(\mathbb{F}_2)$, or $F_{21}$, the Frobenius group of order 21 (which has generators (1 2 3 4 5 6 7) and (2 3 5)(4 7 6) in $S_7$). The cycle type distributions for these three are as follows:

| cycle type: | 1 | 2 | (2, 2) | 3 | (2, 2, 3) | (3, 3) | (2, 4) | 5 | 7 |
|---|---|---|---|---|---|---|---|---|---|
| $F_{21}$ | 1 | | | | | 14 | | | 6 |
| $GL_3(\mathbb{F}_2)$ | 1 | | 21 | | | 56 | 42 | | 48 |
| $A_7$ | 1 | 21 | 105 | 70 | 210 | 280 | 630 | 504 | 720 |

It follows that there is a strong probability that the Galois group of this polynomial is the Frobenius group of order 21. This is actually the case (the verification requires computation of a resolvent of degree 35 and factoring it over $\mathbb{Z}$ — there are three factors, of degrees 7,7, and 21).

## EXERCISES

1. Let $p$ be a prime. Prove that the polynomial $x^4 + 1$ splits mod $p$ either into two irreducible quadratics or into 4 linear factors using Corollary 41 together with the knowledge that the Galois group of this polynomial is the Klein 4-group.
2. (Cf. Exercise 48 of Section 6).
   **(a)** Let $K$ be the splitting field of $x^6 - 2x^3 - 2$. Prove that if $[K : \mathbb{Q}] = 12$ then $K = \mathbb{Q}(\sqrt[3]{2}, i, \sqrt{3})$ and $K$ is generated over the biquadratic field $F = \mathbb{Q}(i, \sqrt{3})$ by $\alpha = \sqrt[3]{1 + \sqrt{3}}$ and by $\beta = \sqrt[3]{1 - \sqrt{3}}$. Show that if this is the case then the elements of order 3 in $\mathrm{Gal}(K/\mathbb{Q})$ lie in $\mathrm{Gal}(K/F)$. Conclude that any element of $\mathrm{Gal}(K/\mathbb{Q})$ of order 3 maps $\alpha$ to another cube root of $1 + \sqrt{3}$ and maps $\beta$ to another cube root of $1 - \sqrt{3}$ and if it is the identity on $\alpha$ or $\beta$ then it is the identity on all of $K$.
   **(b)** Show that the factorization of $f(x)$ into irreducibles over $\mathbb{F}_{13}$ is the polynomial $(x - 7)(x - 8)(x - 11)(x^3 + 3)$ and use Corollary 41 to show that $[K : \mathbb{Q}] = 36$.
   **(c)** Knowing that $G = \mathrm{Gal}(K/\mathbb{Q})$ is of order 36 determine all the elements of $G$ explicitly and in particular show that $G$ is isomorphic to $S_3 \times S_3$.
3. Prove that the Galois group of $x^5 + 20x + 16$ is $A_5$.
4. Prove that the Galois group of $x^5 + x^4 - 4x^3 - 3x^2 + 3x + 1$ is cyclic of order 5. [Show this is the minimal polynomial of $\zeta_{11} + \zeta_{11}^{-1}$.]
5. Prove that the Galois group of $x^5 + 11x + 44$ is the dihedral group $D_{10}$ (cf. Exercise 21 of Section 7).
6. Prove that the Galois group of $x^5 + 15x + 12$ is $F_{20}$, the Frobenius group of order 20 (cf. Exercise 21 of Section 7).
7. Prove that the Galois group of $x^6 + 24x - 20$ is $A_6$.
8. Prove that the Galois group of $x^7 + 7x^4 + 14x + 3$ is $A_7$.

**9.** Determine a polynomial of degree 7 whose Galois group is cyclic of order 7.

**10.** Determine the probable Galois group of $x^7 - 7x + 3$.

## 14.9 TRANSCENDENTAL EXTENSIONS, INSEPARABLE EXTENSIONS, INFINITE GALOIS GROUPS

This section collects some results on arbitrary extensions $E/F$. These results supplement those of the preceding sections and complete the basic picture of how an arbitrary (possibly infinite) extension decomposes. Since this section is primarily intended as a survey, none of the proofs are included; whenever these proofs can be easily supplied by the reader we indicate this either in the text or (with hints) in the exercises.

Throughout this section $E/F$ is an extension of fields. Recall that an element of $E$ which is not algebraic over $F$ is called transcendental over $F$. Keep in mind that extensions involving transcendentals are always of infinite degree. We generally reserve the expression "$t$ is an 'indeterminate' over $F$", when we are thinking of evaluating $t$. Field theoretically, however, the terms transcendental and indeterminate are synonymous (so that the subfield $\mathbb{Q}(\pi)$ of $\mathbb{R}$ and the field $\mathbb{Q}(t)$ are isomorphic).

**Definition.**

**(1)** A subset $\{a_1, a_2, \ldots, a_n\}$ of $E$ is called *algebraically independent* over $F$ if there is no nonzero polynomial $f(x_1, x_2, \ldots, x_n) \in F[x_1, x_2, \ldots, x_n]$ such that $f(a_1, a_2, \ldots, a_n) = 0$. An arbitrary subset $S$ of $E$ is called *algebraically independent* over $F$ if every finite subset of $S$ is algebraically independent. The elements of $S$ are called *independent transcendentals* over $F$.

**(2)** A *transcendence base* for $E/F$ is a maximal subset (with respect to inclusion) of $E$ which is algebraically independent over $F$.

Note that if $E/F$ is algebraic, the empty set is the only algebraically independent subset of $E$. In particular, elements of an algebraically independent set are necessarily transcendental. Moreover, one easily checks that $S \subseteq E$ is an algebraically independent set over $F$ if and only if each $s \in S$ is transcendental over $F(S - \{s\})$. It is also an easy exercise to see that $S$ is a transcendence base for $E/F$ if and only if $S$ is a set of algebraically independent transcendentals over $F$ and $E$ is algebraic over $F(S)$.

**Theorem.** The extension $E/F$ has a transcendence base and any two transcendence bases of $E/F$ have the same cardinality.

*Proof:* The first statement is a standard Zorn's Lemma argument. The proof of the second uses the same "Replacement Lemma" idea as was used to prove that any two bases of a vector space have the same cardinality.

**Definition.** The cardinality of a transcendence base for $E/F$ is called the *transcendence degree* of $E/F$.

Algebraic extensions are precisely the extensions of transcendence degree 0.

One special case of this theorem is when $E$ is *finitely generated* over $F$, that is, $E = F(\alpha_1, \alpha_2, \ldots, \alpha_n)$, for some (not necessarily algebraically independent) elements $\alpha_1, \ldots, \alpha_n$ of $E$. It is clear that we may renumber $\alpha_1, \ldots, \alpha_n$ so that $\alpha_1, \ldots, \alpha_m$ are independent transcendentals and $\alpha_{m+1}, \ldots, \alpha_n$ are algebraic over $F(\alpha_1, \ldots, \alpha_m)$ (so $E$ is a finite extension of the latter field). In this case $E$ is called a *function field in $m$ variables* over $F$. Such fields play a fundamental role in algebraic geometry as fields of functions on $m$-dimensional surfaces. For instance, when $F = \mathbb{C}$ and $m = 1$, these fields arise in analysis as fields of meromorphic functions on compact Riemann surfaces.

Note that if $S_1$ and $S_2$ are transcendence bases for $E/F$ it is not necessarily the case that $F(S_1) = F(S_2)$. For example, if $t$ is transcendental over $\mathbb{Q}$, $\{t\}$ and $\{t^2\}$ are both transcendence bases for $\mathbb{Q}(t)/\mathbb{Q}$ but (as we shall see shortly) $\mathbb{Q}(t^2)$ is a proper subfield of $\mathbb{Q}(t)$.

We now see that if $x_1, x_2, \ldots, x_n$ are indeterminates over $F$ and

$$f(x) = (x - x_1)(x - x_2) \cdots (x - x_n) \tag{14.28}$$

is the general polynomial of degree $n$, then the set of $n$ elementary symmetric functions $s_1, s_2, \ldots, s_n$ in the $x_i$'s are also independent transcendentals over $F$. This is because $x_1, \ldots, x_n$ is a transcendence base for $E = F(x_1, \ldots, x_n)$ over $F$ (so the transcendence degree is $n$) and $E$ is algebraic over $F(s_1, \ldots, s_n)$ (of degree $n!$). The theorem forces $s_1, \ldots, s_n$ to be a transcendence base for this extension as well (in particular, they are independent transcendentals). The general polynomial of degree $n$ over $F$ may therefore equivalently be defined by taking $a_1, \ldots, a_n$ to be any independent transcendentals (or indeterminates) and letting

$$f(x) = x^n + a_1 x^{n-1} + \cdots + a_n \tag{14.29}$$

where the roots of $f$ are denoted by $x_1, \ldots, x_n$ (and $s_i = (-1)^i a_i$).

**Definition.** An extension $E/F$ is called *purely transcendental* if it has a transcendence base $S$ such that $E = F(S)$.

In the preceding discussion, both $F(x_1, \ldots, x_n)$ and $F(s_1, \ldots, s_n)$ are purely transcendental over $F$. As an exercise (following) one can show that $\mathbb{Q}(t, \sqrt{t^3 - t})$ is not a purely transcendental extension of $\mathbb{Q}$ even though it contains no elements that are algebraic over $\mathbb{Q}$ other than those in $\mathbb{Q}$ itself (i.e., the process of decomposing a general extension into a purely transcendental extension followed by an algebraic extension cannot generally be reversed so that the algebraic piece occurs first).

If $E$ is a purely transcendental extension of $F$ of transcendence degree $n = 1$ or 2 and $L$ is an intermediate field, $F \subseteq L \subseteq E$ with the same transcendence degree, then $L$ is again a purely transcendental extension of $F$ (Lüroth ($n = 1$), Castelnuovo ($n = 2$)). This result is not true if the transcendence degree is $\geq 3$, however, although examples where $L$ fails to be purely transcendental are difficult to construct. For extensions of transcendence degree 1 the intermediate fields are described by the following theorem.

**Theorem.** Let $t$ be transcendental over $F$.

**(1)** (Lüroth) If $F \subseteq K \subseteq F(t)$, then $K = F(r)$, for some $r \in F(t)$. In particular, every nontrivial extension of $F$ contained in $F(t)$ is purely transcendental over $F$.

**(2)** If $P = P(t)$, $Q = Q(t)$ are nonzero relatively prime polynomials in $F[t]$ which are not both constant,

$$[F(t) : F(P/Q)] = \max(\deg P, \deg Q).$$

*Proof:* The proof of (2) is outlined in Exercise 18 of Section 13.2.

By part (2) of this theorem we see that $F(P/Q) = F(t)$ if and only if $P$, $Q$ are nonzero relatively prime polynomials of degree $\leq 1$ (not both constant). Thus $F(r) = F(t)$ if and only if $r = \dfrac{at+b}{ct+d}$, where $a, b, c, d \in F$ and $ad - bc \neq 0$ (called a *fractional linear transformation of* $t$). For *any* $r \in F(t) - F$ the map $t \mapsto r$ extends to an embedding of $F(t)$ into itself which is the identity on $F$. This embedding is surjective (i.e., is an automorphism of $F(t)$) precisely for the fractional linear transformations. Furthermore, the map

$$GL_2(F) \to \mathrm{Aut}(F(t)/F) \quad \text{defined by} \quad A = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \mapsto \sigma_A,$$

where $\sigma_A$ denotes the automorphism of $F(t)$ defined by mapping $t$ to $(at+b)/(ct+d)$, is a surjective homomorphism with kernel consisting of the scalar matrices. Thus

$$\mathrm{Aut}(F(t)/F) \cong PGL_2(F)$$

where $PGL_2(F) = GL_2(F)/\{\lambda I \mid \lambda \in F^\times\}$ gives the group of automorphisms of this transcendental extension (cf. Exercise 8 of Section 1).

When $\mathbb{F}$ is a finite field of order $q$, $\mathrm{Aut}(\mathbb{F}(t)/\mathbb{F}) \cong PGL_2(\mathbb{F})$ is a finite group of order $q(q-1)(q+1)$. By Corollary 11 if $K$ is the fixed field of $\mathrm{Aut}(\mathbb{F}(t)/\mathbb{F})$, then $\mathbb{F}(t)$ is Galois over $K$ with Galois group equal to $\mathrm{Aut}(\mathbb{F}(t)/\mathbb{F})$. In particular, the fixed field of $\mathrm{Aut}(\mathbb{F}(t)/\mathbb{F})$ is not $\mathbb{F}$ in this case.

This also provides further examples of the Galois correspondence which can be written out completely for small values of $q$. For instance, if $q = |\mathbb{F}| = 2$, $PGL_2(\mathbb{F})$ is nonabelian of order 6, hence is isomorphic to $S_3$, and has the following lattice of subgroups:

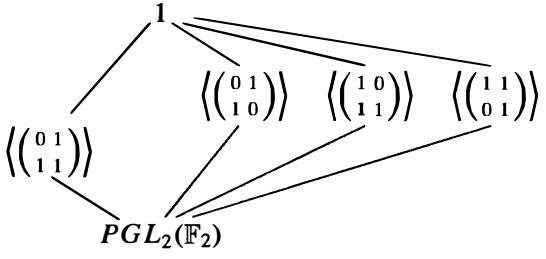



**Fig. 5**

The field $\mathbb{F}(t)$ is of degree 6 over the fixed field $K$ of $\mathrm{Aut}(\mathbb{F}(t)/\mathbb{F})$ and the lattice of subfields $K \subseteq L \subseteq \mathbb{F}(t)$ is dual to the lattice of subgroups of $S_3$. The fixed field of a

cyclic subgroup $\langle \sigma \rangle$ is easily found (via the preceding theorem) by finding a rational function $r$ in $t$ which is fixed by $\sigma$ such that $[\mathbb{F}(t) : \mathbb{F}(r)] = |\sigma|$. For example, if $\sigma : t \mapsto 1/(1+t)$, then $\sigma$ has order 3. The rational function

$$r = t + \sigma(t) + \sigma^2(t) = \frac{t^3 + t + 1}{t(t+1)}$$

is fixed by $\sigma$ and $[\mathbb{F}(t) : \mathbb{F}(r)] = 3$ (by part (2) of the theorem). Since $\mathbb{F}(r)$ is contained in the fixed field of $\langle \sigma \rangle$ and the degree of $\mathbb{F}(t)$ over the fixed field is 3, $\mathbb{F}(r)$ is the fixed field of $\langle \sigma \rangle$. In this way one can explicitly describe the lattice of all subfields of $\mathbb{F}(t)$ containing $K$ shown in Figure 6.

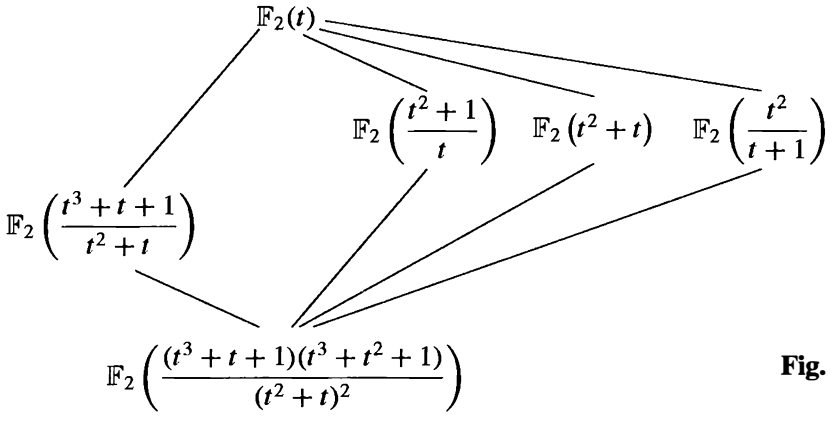



**Fig. 6**

Purely transcendental extensions of $\mathbb{Q}$ play an important role in the problem of realizing finite groups as Galois groups over $\mathbb{Q}$. We describe a deep result of Hilbert which is fundamental to this area of research. If $a_1, a_2, \ldots, a_n$ are independent indeterminates over a field $F$, we may evaluate (or *specialize*) $a_1, \ldots, a_n$ at any elements of $F$, i.e., substitute values in $F$ for the "variables" $a_1, a_2, \ldots, a_n$. If $E$ is a Galois extension of $F(a_1, \ldots, a_n)$, then $E$ is obtained as a splitting field of a polynomial whose coefficients lie in $F[a_1, \ldots, a_n]$. Any specialization of $a_1, \ldots, a_n$ into $F$ maps this polynomial into one whose coefficients lie in $F$. The specialization of $E$ is the splitting field of the resulting specialized polynomial.

**Theorem.** (Hilbert) Let $x_1, x_2, \ldots, x_n$ be independent transcendentals over $\mathbb{Q}$, let $E = \mathbb{Q}(x_1, \ldots, x_n)$ and let $G$ be a finite group of automorphisms of $E$ with fixed field $K$. If $K$ is a purely transcendental extension of $\mathbb{Q}$ with transcendence basis $a_1, a_2, \ldots, a_n$, then there are infinitely many specializations of $a_1, \ldots, a_n$ in $\mathbb{Q}$ such that $E$ specializes to a Galois extension of $\mathbb{Q}$ with Galois group isomorphic to $G$.

Hilbert's Theorem gives a sufficient condition for the specialized extension not to collapse. In general, the Galois group of the specialized extension is a subgroup of $G$ (cf. Proposition 19) and may be a proper subgroup of $G$. It is also known that the fixed

field $K$ need not always be a purely transcendental extension of $\mathbb{Q}$. An example of this occurs when $G$ is the cyclic group of order 47.

This theorem can be used to give another proof of Proposition 42:

**Corollary.** $S_n$ is a Galois group over $\mathbb{Q}$, for all $n$.

*Proof of the Corollary:* We have already proved that the fixed field of $S_n$ acting in the obvious fashion on $\mathbb{Q}(x_1, \ldots, x_n)$ is purely transcendental over $\mathbb{Q}$ (with the elementary symmetric functions as a transcendence base), so Hilbert's Theorem immediately implies the corollary.

The hypothesis that $K$ be purely transcendental over $\mathbb{Q}$ is crucial to the proof of Hilbert's Theorem. Every finite group is isomorphic to a subgroup of $S_n$ and so acts on $\mathbb{Q}(x_1, \ldots, x_n)$ for some $n$. It is not known, however, even for the subgroup $A_n$ of $S_n$ whether its fixed field under the obvious action is a purely transcendental extension of $\mathbb{Q}$ (although it is known by other means that $A_n$ is a Galois group over $\mathbb{Q}$ for all $n$). Thus there are a number of important open problems in this area of research.

One should also notice that Hilbert's Theorem does not work when the base field $\mathbb{Q}$ is replaced by an arbitrary field $F$ (suppose $F$ were algebraically closed, for instance). In particular, as noted earlier, the general polynomial $f(x)$ in Section 6 has Galois group $S_n$ over $F(a_1, \ldots, a_n)$ for any $F$, but when $F$ is a finite field, the specialized extension obtained from its splitting field is always cyclic.

We next expand on the theory of inseparable extensions described in Section 13.5. Let $p$ be a prime and let $F$ be a field of characteristic $p$.

**Definition.** An algebraic extension $E/F$ is called *purely inseparable* if for each $\alpha \in E$ the minimal polynomial of $\alpha$ over $F$ has only one distinct root.

It is easy to see that the following are equivalent:

**(1)** $E/F$ is purely inseparable
**(2)** if $\alpha \in E$ is separable over $F$, then $\alpha \in F$
**(3)** if $\alpha \in E$, then $\alpha^{p^n} \in F$ for some $n$ (depending on $\alpha$), and $m_{\alpha, F}(x) = x^{p^n} - \alpha^{p^n}$.

The following easy proposition describes composites of separable and purely inseparable extensions.

**Proposition.** If $E_1$ and $E_2$ are subfields of $E$ which are both separable (or both purely inseparable) extensions of $F$, then their composite $E_1E_2$ is separable (purely inseparable, respectively) over $F$.

*Proof:* Exercise.

One immediate consequence of this is the following result.

**Proposition.** Let $E/F$ be an algebraic extension. Then there is a unique field $E_{sep}$ with $F \subseteq E_{sep} \subseteq E$ such that $E_{sep}$ is separable over $F$ and $E$ is purely inseparable over $E_{sep}$. The field $E_{sep}$ is the set of elements of $E$ which are separable over $F$.

The degree of $E_{sep}/F$ is called the *separable degree* of $E/F$ and the degree of $E/E_{sep}$ is called the *inseparable degree* of $E/F$ (often denoted as $[E : F]_s$ and $[E : F]_i$ respectively). The product of these two degrees is the (ordinary) degree. The propositions immediately give the following corollary.

**Corollary.** Separable degrees (respectively inseparable degrees) are multiplicative.

When $E$ is generated over $F$ by the root of an irreducible polynomial $p(x) \in F[x]$ the separable and inseparable degrees of the extension $E/F$ are the same as the separable and inseparable degrees of the polynomial $p(x)$ defined in Section 13.5.

The proposition asserts that any algebraic extension may be decomposed into a separable extension followed by a purely inseparable one. Exercise 3 at the end of this section outlines an example illustrating that this decomposition cannot generally be reversed, namely an extension which is not a separable extension of a purely inseparable extension. We shall shortly state conditions on an extension under which the decomposition into separable and purely inseparable subextensions may be reversed.

We now know that an arbitrary extension $E/F$ can be decomposed into a purely transcendental extension $F(S)$ of $F$ followed by a separable extension $E_1$ of $F(S)$ followed by a purely inseparable extension $E/E_1$. In certain instances the inseparability in the algebraic extension at the "top" may be removed by a judicious choice of transcendence base:

**Proposition.** If $E$ is a finitely generated extension of a perfect field $F$, then there is a transcendence base $T$ of $E/F$ such that $E$ is a separable (algebraic) extension of $F(T)$.

A transcendence base $T$ as described in the proposition is called a *separating transcendence base.* Exercise 4 at the end of this section illustrates this with a nontrivial example.

Recall that an extension $E/F$ is *normal* if it is the splitting field of some (possibly infinite) set of polynomials in $F[x]$ (in particular, normal extensions are algebraic but not necessarily finite or separable). We previously used the synonymous term splitting field and the term normal is reintroduced here in the context of arbitrary algebraic extensions since it is used frequently in the literature, often in the context of embeddings of a field into an algebraic closure. Although the following set of equivalences can be gleaned from the preceding sections, the reader should write out a complete proof, checking that the arguments work for both infinite and inseparable extensions:

**Proposition.** Let $E/F$ be an arbitrary algebraic extension and let $\Omega$ be an algebraic closure of $E$. The following are equivalent:

**(1)** $E/F$ is a normal extension (i.e., is the splitting field over $F$ of some set of polynomials in $F[x]$)

(2) whenever $\sigma : E \to \Omega$ is an embedding such that $\sigma|_F$ is the identity, $\sigma(E) = E$
(3) whenever an irreducible polynomial $f(x) \in F[x]$ has one root in $E$, it has all its roots in $E$.

In general, any embedding of a normal extension $E/F$ into an algebraic closure of $E$ which extends the identity embedding of $F$ is an automorphism of $E$, i.e., is an element of $\mathrm{Aut}(E/F)$. Moreover, the number of such automorphisms equals the separable degree of $E/F$, provided the latter is finite:

if $E/F$ is a normal extension and $[E : F]_s$ is finite, $\quad |\mathrm{Aut}(E/F)| = [E : F]_s$.

If $[E : F]_s$ is infinite we shall see shortly that $|\mathrm{Aut}(E/F)|$ is also infinite but need not be of the same cardinality.

If $E/F$ is a normal extension whose separable degree is finite, let $E_0$ be the fixed field of $\mathrm{Aut}(E/F)$. By Corollary 11, $E/E_0$ is a (separable) Galois extension whose degree equals $|\mathrm{Aut}(E/F)|$. It follows that $E_0/F$ must be purely inseparable (of degree equal to $[E : F]_i$), i.e., the separable and purely inseparable pieces of the extension may be reversed for normal extensions. More precisely, we easily obtain the following proposition.

**Proposition.** If $E/F$ is normal with $[E : F]_s < \infty$, then $E = E_{sep}E_{pi}$, where $E_{pi}$ is a purely inseparable extension of $F$ ($E_{pi}$ consists of all purely inseparable elements of $E$ over $F$) and $E_{sep} \cap E_{pi} = F$.

Finally, we mention how Galois Theory generalizes to infinite extensions.

**Definition.** An extension $E/F$ is called *Galois* if it is algebraic, normal and separable. In this case $\mathrm{Aut}(E/F)$ is called the *Galois group* of the extension and is denoted by $\mathrm{Gal}(E/F)$.

For infinite extensions there need not be a bijection between the set of all subgroups of the Galois group and the set of all subfields of $E$ containing $F$, as the following example illustrates.

Let $E$ be the subfield of $\mathbb{R}$ obtained by adjoining to $\mathbb{Q}$ all square roots of positive rational numbers. One easily sees that $E$ may also be described as the splitting field of the set of polynomials $x^2 - p$, where $p$ runs over all primes in $\mathbb{Z}^+$. Note that $E$ is a (countably) infinite Galois extension of $\mathbb{Q}$. Since every automorphism $\sigma$ of $E$ is determined by its action on the square roots of the primes and $\sigma$ either fixes or negates each of these, $\sigma^2$ is the identity automorphism. It follows that $\mathrm{Aut}(E)$ is an infinite elementary abelian 2-group. Thus $\mathrm{Aut}(E)$ is an infinite dimensional vector space over $\mathbb{F}_2$. By an exercise in the section on dual spaces (Section 11.3) the number of nonzero homomorphisms of $\mathrm{Aut}(E)$ into $\mathbb{F}_2$ is uncountable, whence their kernels (which are subspaces of co-dimension 1) are uncountable in number (and distinct). Thus $\mathrm{Aut}(E)$ has *uncountably* many subgroups of index 2, whereas $\mathbb{Q}$ has only a *countable* number of quadratic extensions.

The basic problem is that many (most) subgroups of $\mathrm{Gal}(E/F)$ do not correspond (in a bijective fashion) to subfields of $E$ containing $F$. In order to pick out the "right"

set of subgroups of $\mathrm{Gal}(E/F)$ we must introduce a topology on this group (called the Krull topology). The axioms for the collection of (topologically) closed subsets of a topological space are precisely the bookkeeping devices which single out the relevant subgroups (these are listed in Section 15.2). Galois theory for finite extensions force certain subgroups of finite index to be closed sets and these in turn determine the topology on the entire group (as we might expect since every extension of $F$ inside $E$ is a composite of finite extensions). Moreover, the Galois group of $E/F$ is the inverse limit of the collection of finite groups $\mathrm{Gal}(K/F)$, where $K$ runs over all finite Galois extensions of $F$ contained in $E$ (cf. Exercise 10, Section 7.6).

**Theorem.** (Krull) Let $E/F$ be a Galois extension with Galois group $G$. Topologize $G$ by taking as a base for the closed sets the subgroups of $G$ which are the fixing subgroups of the finite extensions of $F$ in $E$, together with all left and right cosets of these subgroups. Then with this ("Krull") topology the closed subgroups of $G$ correspond bijectively with the subfields of $E$ containing $F$ and the corresponding lattices are dual. Closed normal subgroups of $G$ correspond to normal extensions of $F$ in $E$.

One important area of current research is to describe (as a topological group) the Galois group of certain field extensions such as $\overline{F}/F$, where $\overline{F}$ is the algebraic closure of $F$. Little is known about the latter group when $F = \mathbb{Q}$ (in particular, its normal subgroups of finite index, i.e., which finite groups occur as Galois groups over $\mathbb{Q}$, are not known). If $E$ is the algebraic closure of the finite field $\mathbb{F}_p$, the Galois group of this extension is the topologically cyclic group $\widehat{\mathbb{Z}}$ with the Frobenius automorphism as a topological generator. The group $\widehat{\mathbb{Z}}$ is an uncountable group (in particular, is not isomorphic to $\mathbb{Z}$) with the property that every closed subgroup of finite index is normal with cyclic quotient. Note that $\widehat{\mathbb{Z}}$ must also have nontrivial infinite closed subgroups (unlike $\mathbb{Z}$) since $E$ contains proper subfields which are infinite over $\mathbb{F}_p$ (such as the composite of all extensions of $\mathbb{F}_p$ of $q$-power degree, for any prime $q$ — this Galois extension of $\mathbb{F}_p$ has Galois group $\mathbb{Z}_q$, the $q$-adic integers, as described in Exercise 11 of Section 7.6).

## EXERCISES

1. Prove that every purely inseparable extension is normal.
2. Let $p$ be a prime and let $K = \mathbb{F}_p(x, y)$ with $x$ and $y$ independent transcendentals over $\mathbb{F}_p$. Let $F = \mathbb{F}_p(x^p - x, y^p - x)$.
   (a) Prove that $[K : F] = p^2$ and the separable degree and inseparable degree of $K/F$ are both equal to $p$.
   (b) Prove that there is a subfield $E$ of $K$ containing $F$ which is purely inseparable over $F$ of degree $p$ (so then $K$ is a separable extension of $E$ of degree $p$). [Let $s = x^p - x \in F$ and $t = y^p - x \in F$ and consider $s - t$.]
3. Let $p$ be an odd prime, let $s$ and $t$ be independent transcendentals over $\mathbb{F}_p$, and let $F$ be the field $\mathbb{F}_p(s, t)$. Let $\beta$ be a root of $x^2 - sx + t = 0$ and let $\alpha$ be a root of $x^p - \beta = 0$ (in some algebraic closure of $F$). Set $E = F(\beta)$ and $K = F(\alpha)$.
   (a) Prove that $E$ is a Galois extension of $F$ of degree 2 and that $K$ is a purely inseparable extension of $E$ of degree $p$.

**(b)** Prove that $K$ is not a normal extension of $F$. [If it were, conjugate $\beta$ over $F$ to show that $K$ would contain a $p^{\text{th}}$ root of $s$ and then also a $p^{\text{th}}$ root of $t$, so $[K : F] \geq p^2$, a contradiction.]

**(c)** Prove that there is no field $K_0$ such that $F \subseteq K_0 \subseteq K$ with $K_0/F$ purely inseparable and $K/K_0$ separable. [If there were such a field, use Exercise 1 and the fact that the composite of two normal extensions is again normal to show that $K$ would be a normal extension of $F$.]

**4.** Under the notation of the previous exercise prove that $\alpha, s$ is a separating transcendence base for $K$ over $\mathbb{F}_p$.

**5.** Let $p$ be a prime, let $t$ be transcendental over $\mathbb{F}_p$ and let $K$ be obtained by adjoining to $\mathbb{F}_p(t)$ all $p$-power roots of $t$. Prove that $K$ has transcendence degree 1 over $\mathbb{F}_p$ and has no separating transcendence base.

**6.** Show that if $t$ is transcendental over $\mathbb{Q}$ then $\mathbb{Q}(t, \sqrt{t^3 - t})$ is not a purely transcendental extension of $\mathbb{Q}$. (This is an example of what is called an *elliptic* function field.)

**7.** Let $k$ be the field with 4 elements, $t$ a transcendental over $k$, $F = k(t^4 + t)$ and $K = k(t)$.

**(a)** Show that $[K : F] = 4$.

**(b)** Show that $K$ is separable over $F$.

**(c)** Show that $K$ is Galois over $F$.

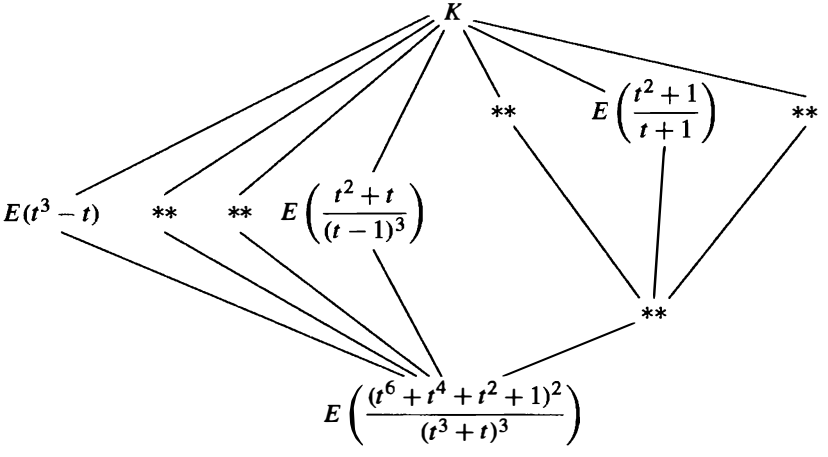
**(d)** Describe the lattice of subgroups of the Galois group and the corresponding lattice of subfields of $K$, giving each subfield in the form $k(r)$, for some rational function $r$.

**8.** Let $p$ be an odd prime, $k$ an algebraically closed field of characteristic $p$ and let $t$ be transcendental over $k$. Suppose $F$ is a degree 2 field extension of $k(t)$. Show that $F$ can be written in the form $k(t, y)$, for some $y \in F$ with $y^2 \in k(t)$ and $y$ transcendental over $k$. If $y^2 = 4t^3 - t - 1$, find $[F : k(y)]$ and describe $k(t) \cap k(y)$ as $k(r)$, for some $r \in k(t)$.

**9.** Let $t$ be transcendental over $\mathbb{F}_3$, let $K = \mathbb{F}_3(t)$, let $G = \text{Aut}(K/\mathbb{F}_3)$ and let $F$ be the fixed field of $G$.

**(a)** Prove $G \cong S_4$ and deduce that there is a unique field $E$ with $F \subseteq E \subseteq K$ and $[E : F] = 2$. [Recall that $G \cong PGL_2(\mathbb{F}_3)$; show that $GL_2(\mathbb{F}_3)$ permutes the 4 lines in a 2-dimensional vector space over $\mathbb{F}_3$ and the kernel of this permutation representation is the scalar matrices.]

**(b)** Complete the description of the lattice of subfields of $K$ containing $E$:




Give each subfield in the form $E(r)$ for some rational function $r$. (The lattice of

subgroups of $A_4$ appears in Section 3.5).

**10.** Prove that a purely transcendental proper extension of a field is never algebraically closed.

**11.** Let $S$ be a set of independent transcendentals over a field $F$ and let $\Omega$ be an algebraic closure of $F(S)$. Prove that any permutation on $S$ extends to an element of $\mathrm{Aut}(F(S)/F)$. Prove that any such automorphism of $F(S)$ extends to an automorphism of $\Omega$. Deduce that $\mathbb{C}$ has infinitely many automorphisms.

**12.** Let $K$ be a subfield of $\mathbb{C}$ maximal with respect to the property "$\sqrt{2} \notin K$."
- **(a)** Show such a field $K$ exists.
- **(b)** Show that $\mathbb{C}$ is algebraic over $K$.
- **(c)** Prove that every finite extension of $K$ in $\mathbb{C}$ is Galois with Galois group a cyclic 2-group.
- **(d)** Deduce that $[\mathbb{C} : K]$ is countable (and not finite).

**13.** Let $K$ be the fixed field in $\mathbb{C}$ of an automorphism of $\mathbb{C}$. Prove that every finite extension of $K$ in $\mathbb{C}$ is cyclic.

**14.** Let $K_n$ be the splitting field of $(x^2-p_1)(x^2-p_2)\cdots(x^2-p_n)$ over $\mathbb{Q}$, where $p_1, \ldots, p_n$ are the first $n$ primes. Prove that the Galois group of $K_n/\mathbb{Q}$ is an elementary abelian 2-group of order $2^n$.

**15.** Let $K_0 = \mathbb{Q}$ and for $n \geq 0$ define the field $K_{n+1}$ as the extension of $K_n$ obtained by adjoining to $K_n$ all roots of all cubic polynomials over $K_n$. Let $K$ be the union of the subfields $K_n$, $n \geq 0$. Prove that $K$ is a Galois extension of $\mathbb{Q}$. Prove that every cubic polynomial over $K$ splits completely over $K$. Prove that there are nontrivial algebraic extensions of $K$.

**16.** Let $F$ be the composite of all the splitting fields of irreducible cubics over $\mathbb{Q}$. Prove that $F$ does not contain all quadratic extensions of $\mathbb{Q}$.

**17.** Let $K_0 = \mathbb{Q}$ and for $n \geq 0$ define the field $K_{n+1}$ as the extension of $K_n$ obtained by adjoining to $K_n$ all radicals of elements in $K_n$. Let $K$ be the union of the subfields $K_n$, $n \geq 0$. Prove that $K$ is a Galois extension of $\mathbb{Q}$. Prove that there are no nontrivial solvable Galois extensions of $K$. Prove that there are nontrivial Galois extensions of $K$.

**18.** Let $F_0 = \mathbb{Q}$ and for $n \geq 0$ define the field $F_{n+1}$ as the extension of $F_n$ obtained by adjoining to $F_n$ all real radicals of elements in $F_n$. Let $F$ be the union of the subfields $F_n$, $n \geq 0$. Let $K^+$ be the fixed field of complex conjugation restricted to the field $K$ in the previous exercise (the maximal real subfield of $K$). Prove that $F \neq K^+$.

**19.** This exercise proves that if $K/F$ is a Galois extension of fields, then $\mathrm{Gal}(K/F)$ is isomorphic to $\varprojlim \mathrm{Gal}(L/F)$, where the inverse limit is taken over all the finite Galois extensions $L$ of $F$ contained in $K$.
- **(a)** Show that $K$ is the union of the fields $L$.
- **(b)** Prove that the map $\varphi : \mathrm{Gal}(K/F) \to \varprojlim \mathrm{Gal}(L/F)$ defined by mapping $\sigma$ in $\mathrm{Gal}(K/F)$ to $(\ldots, \sigma|_L, \ldots)$, where $\sigma|_L$ is the restriction of $\sigma$ to $L$, is a homomorphism.
- **(c)** Show that $\varphi$ is injective.
- **(d)** If $(\ldots, \sigma_L, \ldots) \in \varprojlim \mathrm{Gal}(L/F)$, define $\sigma \in \mathrm{Gal}(K/F)$ by $\sigma(\alpha) = \sigma_L(\alpha)$ if $\alpha \in L$. Prove that $\sigma$ is a well defined automorphism and deduce that $\varphi$ is surjective.

# Part V

# INTRODUCTION TO COMMUTATIVE RINGS, ALGEBRAIC GEOMETRY, AND HOMOLOGICAL ALGEBRA

In this part of the book we continue the study of rings and modules, concentrating first on commutative rings. The topic of commutative algebra, which is of interest in its own right, is also a basic foundation for other areas of algebra. To indicate some of the importance of the algebraic topics introduced, we parallel the development of the ring theory in Chapter 15 with an introduction to affine algebraic geometry. Each section first presents the basic algebraic theory and then follows with an application of those ideas to geometry together with an indication of computational methods using the theory of Gröbner bases from Chapter 9. The purpose here is twofold: the first is to present an application of algebraic techniques in the important branch of mathematics called Algebraic Geometry, and the second is to indicate some of the motivations for the algebraic concepts introduced from their origins in geometric questions.

This connection of geometry and algebra shows a rich interplay between these two areas of mathematics and demonstrates again how results and structures in one circle of mathematical ideas provide insights into another.

In Chapter 16 we continue with some of the fundamental structures involving commutative rings, culminating with Dedekind Domains and a structure theorem for modules over such rings which is a generalization of the structure theorem for modules over P.I.D.s in Chapter 12.

In Chapter 17 we describe some of the basic techniques of "homological algebra," which continues with some of the questions raised by the failure of exactness of some of the sequences considered in Chapter 10. The cohomology of groups in this chapter is intended to serve both as a more in-depth application of homological algebra to see its uses in practice, and as a relatively self contained exposition of this important topic.

CHAPTER 15

# Commutative Rings and Algebraic Geometry

Throughout this chapter $R$ will denote a commutative ring with $1 \neq 0$.

## 15.1 NOETHERIAN RINGS AND AFFINE ALGEBRAIC SETS

In this section we study Noetherian rings in greater detail. These are a natural generalization of Principal Ideal Domains and were introduced briefly in Chapter 12. Note that when $R$ is considered as a left module over itself, its $R$-submodules are precisely its ideals, so the definition in Section 1 of Chapter 12 may be phrased in the following form:

**Definition.** A commutative ring $R$ is said to be *Noetherian* or to satisfy the *ascending chain condition on ideals* (or *A.C.C. on ideals*) if there is no infinite increasing chain of ideals in $R$, i.e., whenever $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ is an increasing chain of ideals of $R$, then there is a positive integer $m$ such that $I_k = I_m$ for all $k \geq m$.

**Proposition 1.** If $I$ is an ideal of the Noetherian ring $R$, then the quotient $R/I$ is a Noetherian ring. Any homomorphic image of a Noetherian ring is Noetherian.

*Proof:* If $R$ is a ring and $I$ is an ideal in $R$, then any infinite ascending chain of ideals in the quotient $R/I$ would correspond by the Lattice Isomorphism Theorem to an infinite ascending chain of ideals in $R$. This gives the first statement, and the second follows by the first Isomorphism Theorem.

**Theorem 2.** The following are equivalent:

**(1)** $R$ is a Noetherian ring.

**(2)** Every nonempty set of ideals of $R$ contains a maximal element under inclusion.

**(3)** Every ideal of $R$ is finitely generated.

*Proof:* The proof is identical to that of Theorem 1 in Section 12.1 in the special case where the $R$-module $M$ is $R$ itself (and submodules are ideals).

**Examples**

Every Principal Ideal Domain is Noetherian since it satisfies condition (3) of Theorem 2. In particular, $\mathbb{Z}$, the polynomial ring $k[x]$ where $k$ is a field, and the Gaussian integers $\mathbb{Z}[i]$, are Noetherian rings. The ring $\mathbb{Z}[x_1, x_2, \ldots]$ is not Noetherian since the ideal $(x_1, x_2, \ldots)$ cannot be generated by any finite set (any finite set of generators involves only finitely many of the $x_i$). Exercise 33(d) in Section 7.4 shows that the ring of continuous real valued functions on $[0, 1]$ is not Noetherian.

A Noetherian ring may have arbitrarily long ascending chains of ideals and may have infinitely long descending chains of ideals. For example, $\mathbb{Z}$ has the infinite descending chain

$$(2) \supset (4) \supset (8) \supset \cdots$$

i.e., a Noetherian ring need not satisfy the *descending chain condition on ideals (D.C.C.)*. We shall see, however, that a commutative ring satisfying D.C.C. on ideals necessarily also satisfies A.C.C., i.e., is Noetherian; such rings are called *Artinian* and are studied in Chapter 16.

The following theorem and its corollary, which we record here for completeness, were proved in Section 9.6 (Theorem 21 and Corollary 22, respectively).

**Theorem 3.** *(Hilbert's Basis Theorem)* If $R$ is a Noetherian ring then so is the polynomial ring $R[x]$.

Note that Hilbert's Basis Theorem shows how larger Noetherian rings may be built from existing ones in a manner analogous to Theorem 7 of Section 9.3 (which proved that if $R$ is a U.F.D., then so is $R[x]$).

**Corollary 4.** The polynomial ring $k[x_1, x_2, \ldots, x_n]$ with coefficients from a field $k$ is a Noetherian ring.

Let $k$ be a field. Recall that a ring $R$ is a *$k$-algebra* if $k$ is contained in the center of $R$ and the identity of $k$ is the identity of $R$.

**Definition.**

**(1)** The ring $R$ is a *finitely generated $k$-algebra* if $R$ is generated as a ring by $k$ together with some finite set $r_1, r_2, \ldots, r_n$ of elements of $R$.

**(2)** Let $R$ and $S$ be $k$-algebras. A map $\psi : R \to S$ is a *$k$-algebra homomorphism* if $\psi$ is a ring homomorphism that is the identity on $k$.

If $R$ is a $k$-algebra then $R$ is both a ring and a vector space over $k$, and it is important to distinguish the sense in which elements of $R$ are generators for $R$. For example, the polynomial ring $k[x_1, \ldots, x_n]$ in a finite number of variables over $k$ is a finitely generated $k$-algebra since $x_1, \ldots, x_n$ are ring generators, but for $n > 0$ this ring is an *infinite* dimensional vector space over $k$.

**Corollary 5.** The ring $R$ is a finitely generated $k$-algebra if and only if there is some surjective $k$-algebra homomorphism

$$\varphi : k[x_1, x_2, \ldots, x_n] \to R$$

from the polynomial ring in a finite number of variables onto $R$ that is the identity map on $k$. Any finitely generated $k$-algebra is therefore Noetherian.

*Proof:* If $R$ is generated as a $k$-algebra by $r_1, \ldots, r_n$, then we may define the map $\varphi : k[x_1, \ldots, x_n] \to R$ by $\varphi(x_i) = r_i$ for all $i$ and $\varphi(a) = a$ for all $a \in k$. Then $\varphi$ extends uniquely to a surjective ring homomorphism. Conversely, given a surjective homomorphism $\varphi$, the images of $x_1, \ldots, x_n$ under $\varphi$ then generate $R$ as a $k$-algebra, proving that $R$ is finitely generated. Since $k[x_1, \ldots, x_n]$ is Noetherian by the previous corollary, any finitely generated $k$-algebra is therefore the quotient of a Noetherian ring, hence also Noetherian by Proposition 1.

**Example**

Suppose the $k$-algebra $R$ is finite dimensional as a vector space over $k$, for example when $R = k[x]/(f(x))$, where $f$ is any nonzero polynomial in $k[x]$. Then in particular $R$ is a finitely generated $k$-algebra since a vector space basis also generates $R$ as a ring. In this case since ideals are also $k$-subspaces any ascending or descending chain of ideals has at most $\dim_k R + 1$ distinct terms, hence $R$ satisfies both A.C.C. and D.C.C. on ideals.

The basic idea behind "algebraic geometry" is to equate geometric questions with algebraic questions involving ideals in rings such as $k[x_1, \ldots, x_n]$. The Noetherian nature of these rings reduces many questions to consideration of finitely many algebraic equations (and this was in turn one of the main original motivations for Hilbert's Basis Theorem). We first consider the principal geometric object, the notion of an "algebraic set" of points.

## Affine Algebraic Sets

Recall that the set $\mathbb{A}^n$ of $n$-tuples of elements of the field $k$ is called *affine $n$-space over $k$* (cf. Section 10.1). If $x_1, x_2, \ldots, x_n$ are independent variables over $k$, then the polynomials $f$ in $k[x_1, x_2, \ldots, x_n]$ can be viewed as $k$-valued functions $f : \mathbb{A}^n \to k$ on $\mathbb{A}^n$ by evaluating $f$ at the points in $\mathbb{A}^n$:

$$f : (a_1, a_2, \ldots, a_n) \mapsto f(a_1, a_2, \ldots, a_n) \in k.$$

This gives a ring of $k$-valued functions on $\mathbb{A}^n$, denoted by $k[\mathbb{A}^n]$ and called the *coordinate ring of* $\mathbb{A}^n$. For instance, when $k = \mathbb{R}$ and $n = 2$, the coordinate ring of Euclidean 2-space $\mathbb{R}^2$ is denoted by $\mathbb{R}[\mathbb{A}^2]$ and is the ring of polynomials in two variables, say $x$ and $y$, acting as real valued functions on $\mathbb{R}^2$ (the usual "coordinate functions").

Each subset $S$ of functions in the coordinate ring $k[\mathbb{A}^n]$ determines a subset $\mathcal{Z}(S)$ of affine space, namely the set of points where all functions in $S$ are simultaneously zero:

$$\mathcal{Z}(S) = \{(a_1, a_2, \ldots, a_n) \in \mathbb{A}^n \mid f(a_1, a_2, \ldots, a_n) = 0 \text{ for all } f \in S\},$$

where $\mathcal{Z}(\emptyset) = \mathbb{A}^n$.

**Definition.** A subset $V$ of $\mathbb{A}^n$ is called an *affine algebraic set* (or just an algebraic set) if $V$ is the set of common zeros of some set $S$ of polynomials, i.e., if $V = \mathcal{Z}(S)$ for some $S \subseteq k[\mathbb{A}^n]$. In this case $V = \mathcal{Z}(S)$ is called the *locus of* $S$ in $\mathbb{A}^n$.

If $S = \{f\}$ or $\{f_1, \ldots, f_m\}$ we shall simply write $\mathcal{Z}(f)$ or $\mathcal{Z}(f_1, \ldots, f_m)$ for $\mathcal{Z}(S)$ and call it the locus of $f$ or $f_1, \ldots, f_m$, respectively. Note that the locus of a single polynomial of the form $f - g$ is the same as the solutions in affine $n$-space of the equation $f = g$, so affine algebraic sets are the solution sets to systems of polynomial equations, and as a result occur frequently in mathematics.

**Examples**

**(1)** If $n = 1$ then the locus of a single polynomial $f \in k[x]$ is the set of roots of $f$ in $k$. The algebraic sets in $\mathbb{A}^1$ are $\emptyset$, any finite set, and $k$ (cf. the exercises).

**(2)** The one point subsets of $\mathbb{A}^n$ for any $n$ are affine algebraic since $\{(a_1, a_2, \ldots, a_n)\}$ is $\mathcal{Z}(x_1 - a_1,\ x_2 - a_2,\ \ldots,\ x_n - a_n)$. More generally, any finite subset of $\mathbb{A}^n$ is an affine algebraic set.

**(3)** One may define lines, planes, etc. in $\mathbb{A}^n$ — these are *linear algebraic sets*, the loci of sets of linear (degree 1) polynomials of $k[x_1, \ldots, x_n]$. For example, a line in $\mathbb{A}^2$ is defined by an equation $ax + by = c$ (which is the locus of the polynomial $f(x, y) = ax + by - c \in k[x, y]$). A line in $\mathbb{A}^3$ is the locus of two linear polynomials of $k[x, y, z]$ that are not multiples of each other. In particular, the coordinate axes, coordinate planes, etc. in $\mathbb{A}^n$ are all affine algebraic sets. For instance, the $x$-axis in $\mathbb{A}^3$ is the zero set $\mathcal{Z}(y, z)$ and the $x,y$ plane is the zero set $\mathcal{Z}(z)$.

**(4)** In general the algebraic set $\mathcal{Z}(f)$ of a nonconstant polynomial $f$ is called a *hypersurface* in $\mathbb{A}^n$. Conic sections are familiar algebraic sets in the Euclidean plane $\mathbb{R}^2$. For example, the locus of $y - x^2$ is the parabola $y = x^2$, the locus of $x^2 + y^2 - 1$ is the unit circle, and $\mathcal{Z}(xy - 1)$ is the hyperbola $y = 1/x$. The $x$- and $y$-axes are the algebraic sets $\mathcal{Z}(y)$ and $\mathcal{Z}(x)$ respectively. Likewise, quadric surfaces such as the ellipsoid defined by the equation $x^2 + \dfrac{y^2}{4} + \dfrac{z^2}{9} = 1$ are affine algebraic sets in $\mathbb{R}^3$.

We leave as exercises the straightforward verification of the following properties of affine algebraic sets. Let $S$ and $T$ be subsets of $k[\mathbb{A}^n]$.

**(1)** If $S \subseteq T$ then $\mathcal{Z}(T) \subseteq \mathcal{Z}(S)$ (i.e., $\mathcal{Z}$ is inclusion reversing or *contravariant*).

**(2)** $\mathcal{Z}(S) = \mathcal{Z}(I)$, where $I = (S)$ is the ideal in $k[\mathbb{A}^n]$ generated by the subset $S$.

**(3)** The intersection of two affine algebraic sets is again an affine algebraic set, in fact $\mathcal{Z}(S) \cap \mathcal{Z}(T) = \mathcal{Z}(S \cup T)$. More generally an arbitrary intersection of affine algebraic sets is an algebraic set: if $\{S_j\}$ is any collection of subsets of $k[\mathbb{A}^n]$, then

$$\cap \mathcal{Z}(S_j) = \mathcal{Z}(\cup S_j).$$

**(4)** The union of two affine algebraic sets is again an affine algebraic set, in fact $\mathcal{Z}(I) \cup \mathcal{Z}(J) = \mathcal{Z}(IJ)$, where $I$ and $J$ are ideals and $IJ$ is their product.

**(5)** $\mathcal{Z}(0) = \mathbb{A}^n$ and $\mathcal{Z}(1) = \emptyset$ (here 0 and 1 denote constant functions).

By (2), every affine algebraic set is the algebraic set corresponding to an *ideal* of the coordinate ring. Thus we may consider

$$\mathcal{Z} : \{\text{ideals of } k[\mathbb{A}^n]\} \to \{\text{affine algebraic sets in } \mathbb{A}^n\}.$$

Since every ideal $I$ in the Noetherian ring $k[x_1, x_2, \ldots, x_n]$ is finitely generated, say $I = (f_1, f_2, \ldots, f_q)$, it follows from (3) that $\mathcal{Z}(I) = \mathcal{Z}(f_1) \cap \mathcal{Z}(f_2) \cap \cdots \cap \mathcal{Z}(f_q)$, i.e., *each affine algebraic set is the intersection of a finite number of hypersurfaces in* $\mathbb{A}^n$. Note that this "geometric" property in affine $n$-space is a consequence of an "algebraic" property of the corresponding coordinate ring (namely, Hilbert's Basis Theorem).

If $V$ is an algebraic set in affine $n$-space, then there may be many ideals $I$ such that $V = \mathcal{Z}(I)$. For example, in affine 2-space over $\mathbb{R}$ the $y$-axis is the locus of the ideal $(x)$ of $\mathbb{R}[x, y]$, and also the locus of $(x^2)$, $(x^3)$, etc. More generally, the zeros of any polynomial are the same as the zeros of all its positive powers, and it follows that $\mathcal{Z}(I) = \mathcal{Z}(I^k)$ for all $k \geq 1$. We shall study the relationship between ideals that determine the same affine algebraic set in the next section when we discuss radicals of ideals.

While the ideal whose locus determines a particular algebraic set $V$ is not unique, there is a unique largest ideal that determines $V$, given by the set of *all* polynomials that vanish on $V$. In general, for any subset $A$ of $\mathbb{A}^n$ define

$$\mathcal{I}(A) = \{f \in k[x_1, \ldots, x_n] \mid f(a_1, a_2, \ldots, a_n) = 0 \text{ for all } (a_1, a_2, \ldots, a_n) \in A\}.$$

It is immediate that $\mathcal{I}(A)$ is an *ideal*, and is the unique largest ideal of functions that are identically zero on $A$. This defines a correspondence

$$\mathcal{I} : \{\text{subsets in } \mathbb{A}^n\} \to \{\text{ideals of } k[\mathbb{A}^n]\}.$$

**Examples**

**(1)** In the Euclidean plane, $\mathcal{I}$(the $x$-axis) is the ideal generated by $y$ in the coordinate ring $\mathbb{R}[x, y]$.

**(2)** Over any field $k$, the ideal of functions vanishing at $(a_1, a_2, \ldots, a_n) \in \mathbb{A}^n$ is a maximal ideal since it is the kernel of the surjective ring homomorphism from $k[x_1, \ldots, x_n]$ to the field $k$ given by evaluation at $(a_1, a_2, \ldots, a_n)$. It follows that

$$\mathcal{I}((a_1, a_2, \ldots, a_n)) = (x_1 - a_1,\ x_2 - a_2,\ \ldots,\ x_n - a_n).$$

**(3)** Let $V = \mathcal{Z}(x^3 - y^2)$ in $\mathbb{A}^2$. If $(a, b) \in \mathbb{A}^2$ is an element of $V$ then $a^3 = b^2$. If $a \neq 0$, then also $b \neq 0$ and we can write $a = (b/a)^2$, $b = (b/a)^3$. It follows that $V$ is the set $\{(a^2, a^3) \mid a \in k\}$. For any polynomial $f(x, y) \in k[x, y]$ we can write $f(x, y) = f_0(x) + f_1(x)y + (x^3 - y^2)g(x, y)$. For $f(x, y) \in \mathcal{I}(V)$, i.e., $f(a^2, a^3) = 0$ for all $a \in k$, it follows that $f_0(a^2) + f_1(a^2)a^3 = 0$ for all $a \in k$. If $f_0(x) = a_r x^r + \cdots + a_0$ and $f_1(x) = b_s x^s + \cdots + b_0$ then

$$f_0(x^2) + x^3 f_1(x^2) = (a_r x^{2r} + \cdots + a_0) + (b_s x^{2s+3} + \cdots + b_0 x^3)$$

and this polynomial is 0 for every $a \in k$. If $k$ is infinite, this polynomial has infinitely many zeros, which can happen only if all of the coefficients are zero. The coefficients of the terms of even degree are the coefficients of $f_0(x)$ and the coefficients of the terms of odd degree are the coefficients of $f_1(x)$, so it follows that $f_0(x)$ and $f_1(x)$ are both 0. It follows that $f(x, y) = (x^3 - y^2)g(x, y)$, and so

$$\mathcal{I}(V) = (x^3 - y^2) \subset k[x, y].$$

If $k$ is finite, however, there may be elements in $\mathcal{I}(V)$ not lying in the ideal $(x^3 - y^2)$. For example, if $k = \mathbb{F}_2$, then $V$ is simply the set $\{(0, 0), (1, 1)\}$ and so $\mathcal{I}(V)$ contains the polynomial $x(x - 1)$ (cf. Exercise 15).

The following properties of the map $\mathcal{I}$ are very easy exercises. Let $A$ and $B$ be subsets of $\mathbb{A}^n$.

**(6)** If $A \subseteq B$ then $\mathcal{I}(B) \subseteq \mathcal{I}(A)$ (i.e., $\mathcal{I}$ is also *contravariant*).
**(7)** $\mathcal{I}(A \cup B) = \mathcal{I}(A) \cap \mathcal{I}(B)$.
**(8)** $\mathcal{I}(\emptyset) = k[x_1, \ldots, x_n]$ and, if $k$ is infinite, $\mathcal{I}(\mathbb{A}^n) = 0$.

Moreover, there are easily verified relations between the maps $\mathcal{Z}$ and $\mathcal{I}$:

**(9)** If $A$ is any subset of $\mathbb{A}^n$ then $A \subseteq \mathcal{Z}(\mathcal{I}(A))$, and if $I$ is any ideal then $I \subseteq \mathcal{I}(\mathcal{Z}(I))$.
**(10)** If $V = \mathcal{Z}(I)$ is an affine algebraic set then $V = \mathcal{Z}(\mathcal{I}(V))$, and if $I = \mathcal{I}(A)$ then $\mathcal{I}(\mathcal{Z}(I)) = I$, i.e., $\mathcal{Z}(\mathcal{I}(\mathcal{Z}(I))) = \mathcal{Z}(I)$ and $\mathcal{I}(\mathcal{Z}(\mathcal{I}(A))) = \mathcal{I}(A)$.

The last relation shows that the maps $\mathcal{Z}$ and $\mathcal{I}$ act as inverses of each other provided one restricts to the collection of affine algebraic sets $V = \mathcal{Z}(I)$ in $\mathbb{A}^n$ and to the set of ideals in $k[\mathbb{A}^n]$ of the form $\mathcal{I}(V)$. In the case where the field $k$ is algebraically closed we shall (in the following two sections) characterize those ideals $I$ that are of the form $\mathcal{I}(V)$ for some affine algebraic set $V$ in terms of purely ring-theoretic properties of the ideal $I$ (this is the famous "Zeros Theorem" of Hilbert, cf. Theorem 32).

**Definition.** If $V \subseteq \mathbb{A}^n$ is an affine algebraic set the quotient ring $k[\mathbb{A}^n]/\mathcal{I}(V)$ is called the *coordinate ring of* $V$, and is denoted by $k[V]$.

Note that for $V = \mathbb{A}^n$ and $k$ infinite we have $\mathcal{I}(V) = 0$, so this definition extends the previous terminology. The polynomials in $k[\mathbb{A}^n]$ define $k$-valued functions on $V$ simply by restricting these functions on $\mathbb{A}^n$ to the subset $V$. Two such polynomial functions $f$ and $g$ define the *same* function on $V$ if and only if $f - g$ is identically 0 on $V$, which is to say that $f - g \in \mathcal{I}(V)$. Hence the cosets $\overline{f} = f + \mathcal{I}(V)$ giving the elements of the quotient $k[V]$ are precisely the restrictions to $V$ of ordinary polynomial functions $f$ from $\mathbb{A}^n$ to $k$ (which helps to explain the notation $k[V]$). If $x_i$ denotes the $i^{\text{th}}$ coordinate function on $\mathbb{A}^n$ (projecting an $n$-tuple onto its $i^{\text{th}}$ component), then the restriction $\overline{x_i}$ of $x_i$ to $V$ (which also just gives the $i^{\text{th}}$ component of the elements in $V$ viewed as a subset of $\mathbb{A}^n$) is an element of $k[V]$, and $k[V]$ is finitely generated as a $k$-algebra by $\overline{x_1}, \ldots, \overline{x_n}$ (although this need not be a minimal generating set).

**Example**

If $V = \mathcal{Z}(xy - 1)$ is the hyperbola $y = 1/x$ in $\mathbb{R}^2$, then $\mathbb{R}[V] = \mathbb{R}[x, y]/(xy - 1)$. The polynomials $f(x, y) = x$ (the $x$-coordinate function) and $g(x, y) = x + (xy - 1)$, which are different functions on $\mathbb{R}^2$, define the same function on the subset $V$. On the point $(1/2, 2) \in V$, for example, both give the value $1/2$. In the quotient ring $\mathbb{R}[V]$ we have $\overline{x}\,\overline{y} = 1$, so $\mathbb{R}[V] \cong \mathbb{R}[x, 1/x]$. For any function $\overline{f} \in \mathbb{R}[V]$ and any $(a, b) \in V$ we have $\overline{f}(a, b) = f(a, 1/a)$ for any polynomial $f \in k[x, y]$ mapping to $\overline{f}$ in the quotient.

Suppose now that $V \subseteq \mathbb{A}^n$ and $W \subseteq \mathbb{A}^m$ are two affine algebraic sets. Since $V$ and $W$ are defined by the vanishing of polynomials, the most natural algebraic maps between $V$ and $W$ are those defined by polynomials:

**Definition.** A map $\varphi : V \to W$ is called a *morphism* (or *polynomial map* or *regular map*) of algebraic sets if there are polynomials $\varphi_1, \ldots, \varphi_m \in k[x_1, x_2, \ldots, x_n]$ such that

$$\varphi((a_1, \ldots, a_n)) = (\varphi_1(a_1, \ldots, a_n), \ldots, \varphi_m(a_1, \ldots, a_n))$$

for all $(a_1, \ldots, a_n) \in V$. The map $\varphi : V \to W$ is an *isomorphism* of algebraic sets if there is a morphism $\psi : W \to V$ with $\varphi \circ \psi = 1_W$ and $\psi \circ \varphi = 1_V$.

Note that in general $\varphi_1, \varphi_2, \ldots, \varphi_m$ are not uniquely defined. For example, both $f = x$ and $g = x + (xy - 1)$ in the example above define the same morphism from $V = \mathcal{Z}(xy - 1)$ to $W = \mathbb{A}^1$.

Suppose $F$ is a polynomial in $k[x_1, \ldots, x_m]$. Then $F \circ \varphi = F(\varphi_1, \varphi_2, \ldots, \varphi_m)$ is a polynomial in $k[x_1, \ldots, x_n]$ since $\varphi_1, \varphi_2, \ldots, \varphi_m$ are polynomials in $x_1, \ldots, x_n$. If $F \in \mathcal{I}(W)$, then $F \circ \varphi((a_1, a_2, \ldots, a_n)) = 0$ for every $(a_1, a_2, \ldots, a_n) \in V$ since $\varphi((a_1, a_2, \ldots, a_n)) \in W$. Thus $F \circ \varphi \in \mathcal{I}(V)$. It follows that $\varphi$ induces a well defined map from the quotient ring $k[x_1, \ldots, x_m]/\mathcal{I}(W)$ to the quotient ring $k[x_1, \ldots, x_n]/\mathcal{I}(V)$:

$$\begin{aligned} \tilde{\varphi} : k[W] &\to k[V] \\ f &\mapsto f \circ \varphi \end{aligned}$$

where $f \circ \varphi$ is given by $F \circ \varphi + \mathcal{I}(V)$ for any polynomial $F = F(x_1, \ldots, x_m)$ with $f = F + \mathcal{I}(W)$. It is easy to check that $\tilde{\varphi}$ is a $k$-algebra homomorphism (for example, $\tilde{\varphi}(f + g) = (f + g) \circ \varphi = f \circ \varphi + g \circ \varphi = \tilde{\varphi}(f) + \tilde{\varphi}(g)$ shows that $\tilde{\varphi}$ is additive). Note also the contravariant nature of $\tilde{\varphi}$: the morphism from $V$ to $W$ induces a $k$-algebra homomorphism from $k[W]$ to $k[V]$.

Suppose conversely that $\Phi$ is any $k$-algebra homomorphism from the coordinate ring $k[W] = k[x_1, \ldots, x_m]/\mathcal{I}(W)$ to $k[V] = k[x_1, \ldots, x_n]/\mathcal{I}(V)$. Let $F_i$ be a representative in $k[x_1, \ldots, x_n]$ for the image under $\Phi$ of $\bar{x}_i \in k[W]$ (i.e., $\Phi(x_i \bmod \mathcal{I}(W))$ is $F_i \bmod \mathcal{I}(V)$). Then $\varphi = (F_1, \ldots, F_m)$ defines a polynomial map from $\mathbb{A}^n$ to $\mathbb{A}^m$, and in fact $\varphi$ is a morphism from $V$ to $W$. To see this it suffices to check that $\varphi$ maps a point of $V$ to a point of $W$ since by definition $\varphi$ is already defined by polynomials. If $g \in \mathcal{I}(W) \subset k[x_1, \ldots, x_m]$, then in $k[W]$ we have

$$g(x_1 + \mathcal{I}(W), \ldots, x_m + \mathcal{I}(W)) = g(x_1, \ldots, x_m) + \mathcal{I}(W) = \mathcal{I}(W) = 0 \in k[W],$$

and so

$$\Phi(g(x_1 + \mathcal{I}(W), \ldots, x_m + \mathcal{I}(W))) = 0 \in k[V].$$

Since $\Phi$ is a $k$-algebra homomorphism, it follows that

$$g(\Phi(x_1 + \mathcal{I}(W)), \ldots, \Phi(x_m + \mathcal{I}(W)) = 0 \in k[V].$$

By definition, $\Phi(x_i + \mathcal{I}(W)) = F_i \bmod \mathcal{I}(V)$, so

$$g(F_1 \bmod \mathcal{I}(V), \ldots, F_m \bmod \mathcal{I}(V)) = 0 \in k[V],$$

i.e.,

$$g(F_1, \ldots, F_m) \in \mathcal{I}(V).$$

It follows that $g(F_1(a_1, \ldots, a_n), \ldots, F_m(a_1, \ldots, a_n)) = 0$ for every $(a_1, \ldots, a_n)$ in $V$. This shows that if $(a_1, \ldots, a_n) \in V$, then every polynomial in $\mathcal{I}(W)$ vanishes

on $\varphi(a_1, \ldots, a_n)$. By property (10) of the maps $\mathcal{Z}$ and $\mathcal{I}$ above, this means that $\varphi(a_1, \ldots, a_n) \in \mathcal{Z}(\mathcal{I}(W)) = W$, which proves that $\varphi$ maps a point in $V$ to a point in $W$. It follows that $\varphi = (F_1, \ldots, F_m)$ is a morphism from $V$ to $W$. Since the $F_i$ are well defined modulo $\mathcal{I}(V)$, this morphism from $V$ to $W$ does not depend on the choice of the $F_i$. Furthermore, the morphism $\varphi$ induces the original $k$-algebra homomorphism $\Phi$ from $k[W]$ to $k[V]$, i.e., $\widetilde{\varphi} = \Phi$, since both homomorphisms take the value $F_i + \mathcal{I}(V)$ on $x_i + \mathcal{I}(W) \in k[W]$. This proves the first two statements in the following theorem.

**Theorem 6.** Let $V \subseteq \mathbb{A}^n$ and $W \subseteq \mathbb{A}^m$ be affine algebraic sets. Then there is a bijective correspondence

$$\left\{ \begin{array}{c} \text{morphisms from } V \text{ to } W \\ \text{as algebraic sets} \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} k\text{-algebra homomorphisms} \\ \text{from } k[W] \text{ to } k[V] \end{array} \right\}.$$

More precisely,

**(1)** Every morphism $\varphi : V \to W$ induces an associated $k$-algebra homomorphism $\widetilde{\varphi} : k[W] \to k[V]$ defined by $\widetilde{\varphi}(f) = f \circ \varphi$.

**(2)** Every $k$-algebra homomorphism $\Phi : k[W] \to k[V]$ is induced by a unique morphism $\varphi : V \to W$, i.e., $\Phi = \widetilde{\varphi}$.

**(3)** If $\varphi : V \to W$ and $\psi : W \to U$ are morphisms of affine algebraic sets, then $\widetilde{\psi \circ \varphi} = \widetilde{\varphi} \circ \widetilde{\psi} : k[U] \to k[V]$.

**(4)** The morphism $\varphi : V \to W$ is an isomorphism if and only if $\widetilde{\varphi} : k[W] \to k[V]$ is a $k$-algebra isomorphism.

*Proof:* The proof of (3) is left as an exercise and (4) is then immediate.

**Example**

For any infinite field $k$ let $V = \mathbb{A}^1$ and let $W = \mathcal{Z}(x^3 - y^2) = \{(a^2, a^3) \mid a \in k\}$. The map $\varphi : V \to W$ defined by $\varphi(a) = (a^2, a^3)$ is a morphism from $V$ to $W$. Note that $\varphi$ is a bijection. The coordinate rings are $k[V] = k[x]$ and $k[W] = k[x, y]/(x^3 - y^2)$ (by the computations in a previous example — it is at this point we need $k$ to be infinite) and the associated $k$-algebra homomorphism of coordinate rings is determined by

$$\begin{aligned} \widetilde{\varphi} : k[W] &\longrightarrow k[V] \\ x &\mapsto x^2 \\ y &\mapsto x^3. \end{aligned}$$

The image of $\widetilde{\varphi}$ is the subalgebra $k[x^2, x^3] = k + x^2 k[x]$ of $k[x]$, so in particular $\widetilde{\varphi}$ is not surjective. Hence $\widetilde{\varphi}$ is not an isomorphism of coordinate rings, and it follows that $\varphi$ is not an isomorphism of algebraic sets, even though the morphism $\varphi$ is a bijective map. The inverse map is given by $\psi(0, 0) = 0$ and $\psi(a, b) = b/a$ for $b \neq 0$, and this cannot be achieved by a polynomial map.

The bijection in Theorem 6 gives a translation from maps between two geometrically defined algebraic sets $V$ and $W$ into algebraic maps between their coordinate rings. It also allows us to define a morphism intrinsically in terms of $V$ and $W$ without explicit reference to the ambient affine spaces containing them:

**Corollary 7.** Suppose $\varphi : V \to W$ is a map of affine algebraic sets. Then $\varphi$ is a morphism if and only if for every $f \in k[W]$ the composite map $f \circ \varphi$ is an element of $k[V]$ (as a $k$-valued function on $V$). When $\varphi$ is a morphism, $\varphi(v) = w$ with $v \in V$ and $w \in W$ if and only if $\widetilde{\varphi}^{-1}(\mathcal{I}(\{v\})) = \mathcal{I}(\{w\})$.

*Proof:* We first prove that if $\varphi$ is any map from $V$ to $W$ such that $\widetilde{\varphi}$ is a $k$-algebra homomorphism then $\varphi(v) = w$ if and only if $\widetilde{\varphi}^{-1}(\mathcal{I}(\{v\})) = \mathcal{I}(\{w\})$, which will in particular establish the second statement. Note that $\varphi(v) = w$ if and only if every polynomial $f$ vanishing at $w$ vanishes at $\varphi(v)$ (by property (10) above: $\{w\} = \mathcal{Z}(\mathcal{I}(\{w\}))$). Since $f$ vanishes at $\varphi(v)$ if and only if $\widetilde{\varphi}(f)$ vanishes at $v$, this is equivalent to the statement that $\widetilde{\varphi}(f) \in \mathcal{I}(\{v\})$ for every $f \in \mathcal{I}(\{w\})$, i.e., $\widetilde{\varphi}(\mathcal{I}(\{w\})) \subseteq \mathcal{I}(\{v\})$, or $\mathcal{I}(\{w\}) \subseteq \widetilde{\varphi}^{-1}(\mathcal{I}(\{v\}))$. Since both $\mathcal{I}(\{w\})$ and $\mathcal{I}(\{v\})$ are maximal ideals, this is equivalent to $\widetilde{\varphi}^{-1}(\mathcal{I}(\{v\})) = \mathcal{I}(\{w\})$.

We now prove the first statement. If $\varphi$ is a morphism, then $f \circ \varphi \in k[V]$ for every $f \in k[W]$. For the converse, observe first that composition with any map $\varphi : V \to W$ defines a $k$-algebra homomorphism $\widetilde{\varphi}$ from the $k$-algebra of $k$-valued functions on $W$ to the $k$-algebra of $k$-valued functions on $V$ (this is immediate from the pointwise definition of the addition and multiplication of functions). If $f \circ \varphi \in k[V]$ for every $f \in k[W]$, then $\widetilde{\varphi}$ is a $k$-algebra homomorphism from $k[W]$ to $k[V]$, so by the proposition, $\widetilde{\varphi} = \widetilde{\Phi}$ for a unique morphism $\Phi : V \to W$. Also, since $\widetilde{\varphi}$ is a $k$-algebra homomorphism from $k[W]$ to $k[V]$ it follows by what we have already shown that $\varphi(v) = w$ if and only if $\widetilde{\varphi}^{-1}(\mathcal{I}(\{v\})) = \mathcal{I}(\{w\})$. Because $\widetilde{\varphi} = \widetilde{\Phi}$, this is equivalent to $\widetilde{\Phi}^{-1}(\mathcal{I}(\{v\})) = \mathcal{I}(\{w\})$, and so $\Phi(v) = w$. Hence $\varphi$ and $\Phi$ define the same map on $V$ and so $\varphi$ is a morphism, completing the proof.

Corollary 7 and the last part of Theorem 6 show that the isomorphism type of the coordinate ring of $V$ (as a $k$-algebra) does not depend on the embedding of $V$ in a particular affine $n$-space.

## Computations in Affine Algebraic Sets and $k$-algebras

The theory of Gröbner bases developed in Section 9.6 is very useful in computations involving affine algebraic sets, for example in computing in the coordinate rings $k[\mathbb{A}^n]/\mathcal{I}(V)$. When $n > 1$ it can be difficult to describe the elements in this quotient ring explicitly. By Theorem 23 in Section 9.6, each polynomial $f$ in $k[\mathbb{A}^n]$ has a unique remainder after general polynomial division by the elements in a Gröbner basis for $\mathcal{I}(V)$, and this remainder therefore serves as a unique representative for the coset $\bar{f}$ of $f$ in the quotient $k[\mathbb{A}^n]/\mathcal{I}(V)$.

### Examples

**(1)** In the example $W = \mathcal{Z}(x^3 - y^2)$ above, we showed $I = \mathcal{I}(W) = (x^3 - y^2)$ for any infinite field $k$ and so $k[W] = k[x, y]/(x^3 - y^2)$. Here $x^3 - y^2$ gives a Gröbner basis for $I$ with respect to the lexicographic monomial ordering with $y > x$, so every polynomial $f = f(x, y)$ can be written uniquely in the form $f(x, y) = f_0(x) + f_1(x)y + f_I$ with $f_0(x), f_1(x) \in k[x]$ and $f_I \in I$. Then $f_0(x) + f_1(x)y$ gives a unique representative for $\bar{f}$ in $k[W]$. With respect to the lexicographic monomial ordering with $x > y$,

$x^3 - y^2$ is again a Gröbner basis for $I$, but now the remainder representing $\bar{f}$ in $k[W]$ is of the form $h_0(y) + h_1(y)x + h_2(y)x^2$.

**(2)** Let $V = \mathcal{Z}(xz+y^2+z^2, xy-xz+yz-2z^2) \subset \mathbb{C}^3$ and $W = \mathcal{Z}(u^3-uv^2+v^3) \subset \mathbb{C}^2$. We shall show later that $I = \mathcal{I}(V) = (xz+y^2+z^2, xy-xz+yz-2z^2) \subset \mathbb{C}[x, y, z]$ and $J = \mathcal{I}(W) = (u^3 - uv^2 + v^3) \subset \mathbb{C}[u, v]$. In this case $u^3 - uv^2 + v^3$ gives a Gröbner basis for $J$ for the lexicographic monomial ordering with $u > v$ similar to the previous example. The situation for $I$ is more complicated. With respect to the lexicographic monomial ordering with $x > y > z$ the reduced Gröbner basis for $I$ is given by

$$g_1 = xy + y^2 + yz - z^2, \qquad g_2 = xz + y^2 + z^2, \qquad g_3 = y^3 - y^2z + z^3.$$

Unique representatives for $\mathbb{C}[V] = \mathbb{C}[x, y, z]/(x^2 + xz + y^2, 2x^2 - xy + xz - yz)$ are given by the remainders after general polynomial division by $\{g_1, g_2, g_3\}$.

We saw already in Section 9.6 that Gröbner bases and elimination theory can be used in the explicit computation of affine algebraic sets $\mathcal{Z}(S)$, or, equivalently, in explicitly solving systems of algebraic equations. The same theory can be used to determine explicitly a set of generators for the image and kernel of a $k$-algebra homomorphism

$$\Phi : k[y_1, \ldots, y_m]/J \longrightarrow k[x_1, \ldots, x_n]/I$$

where $I$ and $J$ are ideals. In the particular case when $I = \mathcal{I}(V)$ and $J = \mathcal{I}(W)$ are the ideals associated to affine algebraic sets $V \subseteq \mathbb{A}^n$ and $W \subseteq \mathbb{A}^m$ then by Theorem 6, the $k$-algebra homomorphism $\Phi$ corresponds to a morphism from $V$ to $W$, and we shall apply the results here to affine algebraic sets in Section 3.

For $1 \leq i \leq m$, let $\varphi_i \in k[x_1, \ldots, x_n]$ be any polynomial representing the coset $\Phi(\bar{y}_i)$, where as usual we use a bar to denote the coset of an element in a quotient. The polynomials $\varphi_1, \ldots, \varphi_n$ are unique up to elements of $I$. Then the image of a coset $f(y_1, \ldots, y_m) + J$ under $\Phi$ is the coset $f(\varphi_1, \ldots, \varphi_m) + I$. Given any $\varphi_1, \ldots, \varphi_n$, the map sending $y_i$ to $\varphi_i$ induces a $k$-algebra homomorphism $\Phi$ if and only if $f(y_1, \ldots, y_m) \in I$ for every $f \in J$, a condition which can be checked on a set of generators for $J$.

**Proposition 8.** With notation as above, let $R = k[y_1, \ldots, y_m, x_1, \ldots, x_n]$ and let $\mathcal{A}$ be the ideal generated by $y_1 - \varphi_1, \ldots, y_m - \varphi_m$ together with generators for $I$. Let $G$ be the reduced Gröbner basis of $\mathcal{A}$ with respect to the lexicographic monomial ordering $x_1 > \cdots > x_n > y_1 > \cdots > y_m$. Then

**(a)** The kernel of $\Phi$ is $\mathcal{A} \cap k[y_1, \ldots, y_m]$ modulo $J$. The elements of $G$ in $k[y_1, \ldots, y_m]$ (taken modulo $J$) generate $\ker \Phi$.

**(b)** If $f \in k[x_1, \ldots, x_n]$, then $\bar{f}$ is in the image of $\Phi$ if and only if the remainder after general polynomial division of $f$ by the elements in $G$ is an element $h \in k[y_1, \ldots, y_m]$, in which case $\Phi(\bar{h}) = \bar{f}$.

*Proof:* If we show $\ker \Phi = \mathcal{A} \cap k[y_1, \ldots, y_m]$ modulo $J$ then (a) follows by Proposition 30 in Section 9.6. Suppose first that $f \in \mathcal{A} \cap k[y_1, \ldots, y_m]$. If $f_1, \ldots, f_s$ are generators for $I$ in $k[x_1, \ldots, x_n]$, then

$$f(y_1, \ldots, y_m) = \sum_{i=1}^{n} a_i(y_i - \varphi_i) + \sum_{j=1}^{s} b_i f_i$$

as polynomials in $R$, where $a_1, \ldots, a_n, b_1, \ldots, b_s \in R$. Substituting $y_i = \varphi_i$ we see that $f(\varphi_1, \ldots, \varphi_m)$ is an element of $I$. Since $\Phi(\bar{f}) = f(\varphi_1, \ldots, \varphi_m)$ modulo $I$, it follows that $f$ represents a coset in the kernel of $\Phi$. Conversely, suppose $f \in k[y_1, \ldots, y_m]$ represents an element in $\ker \Phi$. Then $f(\varphi_1, \ldots, \varphi_m) \in I$ (in $k[x_1, \ldots, x_n]$) and so also $f(\varphi_1, \ldots, \varphi_m) \in \mathcal{A}$ (in $R$). Since $y_i - \varphi_i \in \mathcal{A}$,

$$f(y_1, \ldots, y_m) \equiv f(\varphi_1, \ldots, \varphi_m) \equiv 0 \bmod \mathcal{A}$$

so $f \in \mathcal{A} \cap k[y_1, \ldots, y_m]$.

For (b), suppose first that $f \in k[x_1, \ldots, x_n]$ represents an element in the image of $\Phi$, i.e., $\bar{f} = \Phi(\bar{h})$ for some polynomial $h \in k[y_1, \ldots, y_m]$. Then

$$f(x_1, \ldots, x_n) - h(\varphi_1, \ldots, \varphi_m) \in I$$

as polynomials in $k[x_1, \ldots, x_n]$, and so $f(x_1, \ldots, x_n) - h(\varphi_1, \ldots, \varphi_m) \in \mathcal{A}$ as polynomials in $R$. As before, since each $y_i - \varphi_i \in \mathcal{A}$ it follows that

$$f(x_1, \ldots, x_n) - h(y_1, \ldots, y_m) \in \mathcal{A}.$$

Then $f(x_1, \ldots, x_n)$ and $h(y_1, \ldots, y_m)$ leave the same remainder after general polynomial division by the elements in $G$. Since $x_1 > \cdots > x_n > y_1 > \cdots > y_m$, the remainder of $h(y_1, \ldots, y_m)$ is again a polynomial $h_0$ only involving $y_1, \ldots, y_m$. Note also that $h - h_0 \in \mathcal{A} \cap k[y_1, \ldots, y_m]$ so $\bar{h}$ and $\overline{h_0}$ differ by an element in $\ker \Phi$ by (a), so $\Phi(\overline{h_0}) = \Phi(\bar{h}) = \bar{f}$. For the converse, if $f$ leaves the remainder $h \in k[y_1, \ldots, y_m]$ after general polynomial division by the elements in $G$ then $f(x_1, \ldots, x_n) - h(y_1, \ldots, y_m) \in \mathcal{A}$, i.e.,

$$f(x_1, \ldots, x_n) - h(y_1, \ldots, y_m) = \sum_{i=1}^{n} a_i(y_i - \varphi_i) + \sum_{j=1}^{s} b_j f_j$$

as polynomials in $R$, where $a_1, \ldots, a_n, b_1, \ldots, b_s \in R$. Substituting $y_i = \varphi_i$ we obtain

$$f(x_1, \ldots, x_n) - h(\varphi_1, \ldots, \varphi_m) \in I$$

as polynomials in $x_1, \ldots x_n$, and so $\bar{f} = \Phi(\bar{h})$.

It follows in particular from Proposition 8 that $\Phi$ will be a surjective homomorphism if and only if for each $i = 1, 2, \ldots, n$, dividing $x_i$ by the elements in the Gröbner basis $G$ leaves a remainder $h_i$ in $k[y_1, \ldots, y_m]$. In particular, $x_n - h_n$ leaves a remainder of 0. But this means the leading term of some element $g_n$ in $G$ divides the leading term of $x_n - h_n$ and since $x_1 > \cdots > x_n > y_1 > \cdots > y_m$ by the choice of the ordering, the leading term of $x_n - h_n$ is just $x_n$. It follows that $LT(g_n) = x_n$ and so $g_n = x_n - h_{n,0} \in G$ for some $h_{n,0} \in k[y_1, \ldots, y_m]$ (in fact $h_{n,0}$ is the remainder of $h_n$ after division by the elements in $G$). Next, since $x_{n-1} - h_{n-1}$ leaves a remainder of 0, there is an element $g_{n-1}$ in $G$ whose leading term is $x_{n-1}$. Since $G$ is a reduced Gröbner basis and $g_n \in G$ , the leading term of $g_n$, i.e., $x_n$, does not divide any of the terms in $g_{n-1}$ and it follows that $g_{n-1} = x_{n-1} - h_{n-1,0} \in G$ for some $h_{n-1,0} \in k[y_1, \ldots, y_m]$. Proceeding in a similar fashion we obtain the following corollary, showing that whether $\Phi$ is surjective can be seen immediately from the elements in the reduced Gröbner basis.

**Corollary 9.** The map $\Phi$ is surjective if and only if for each $i$, $1 \leq i \leq n$, the reduced Gröbner basis $G$ contains a polynomial $x_i - h_i$ where $h_i \in k[y_1, \ldots, y_m]$.

**Examples**

**(1)** Let $\Phi : \mathbb{Q}[u, v] \to \mathbb{Q}[x]$ be defined by $\Phi(u) = x^2 + x$ and $\Phi(v) = x^3$. The reduced Gröbner basis $G$ for the ideal $\mathcal{A} = (u - x^2 - x, v - x^3)$ with respect to the lexicographic monomial ordering $x > u > v$ is

$$g_1 = x^2 + x - u, \qquad g_3 = vx - x - u^2 + u + 2v,$$
$$g_2 = ux + x - u - v, \qquad g_4 = u^3 - 3uv - v^2 - v.$$

The kernel of $\Phi$ is the ideal generated by $G \cap \mathbb{Q}[u, v] = \{g_4\}$. By Corollary 9, we see that $\Phi$ is not surjective. The remainder after general polynomial division of $x^4$ by $\{g_1, g_s, g_3, g_4\}$ is $x + u^2 - u - 2v \notin \mathbb{Q}[u, v]$, so $x^4$ is not in the image of $\Phi$. The remainder of $x^5 + x$ is $-u^2 + uv + u + 2v \in \mathbb{Q}[u, v]$ so $x^5 + x = \Phi(-u^2 + uv + u + 2v)$ is in the image of $\Phi$, as a quick check will confirm.

**(2)** Let $V = \mathcal{Z}(I) \subset \mathbb{C}^3$ and $W = \mathcal{Z}(J) \subset \mathbb{C}^2$ where $I = (xz + y^2 + z^2, xy - xz + yz - 2z^2)$ and $J = (u^3 - uv^2 + v^3)$ as in Example 2 following Corollary 7. Then the map $\varphi : V \to W$ defined by $\varphi((a, b, c)) = (c, b)$ is a morphism from $V$ to $W$. To see this, we must check that $(c, b) \in W$ if $(a, b, c) \in V$. Equivalently, by Theorem 6, we must check that the map

$$\widetilde{\varphi} : \mathbb{C}[u, v]/(u^3 - uv^2 + v^3) \longrightarrow \mathbb{C}[x, y, z]/(xz + y^2 + z^2, xy - xz + yz - 2z^2)$$

induced by mapping $u$ to $z$ and $v$ to $y$ is a $\mathbb{C}$-algebra homomorphism. This in turn is equivalent to verifying that $f = z^3 - zy^2 + y^3$ is an element of the ideal $I$. In this case $f$ is actually an element in the reduced Gröbner basis for $I$:

$$xy + y^2 + yz - z^2, \qquad xz + y^2 + z^2, \qquad y^3 - y^2z + z^3,$$

so certainly $f \in I$. (Note that dividing $f$ by the original two generators for $I$ leaves the nonzero remainder $f$ itself, from which it is much less clear that $f \in I$, so it is important to use a Gröbner basis when working in coordinate rings.)

**(3)** In the previous example, let $\mathcal{A} = (u - z, v - y, xz + y^2 + z^2, xy - xz + yz - 2z^2) \subset \mathbb{C}[u, v, x, y, z]$ as in Proposition 8. With respect to the lexicographic monomial ordering $x > y > z > u > v$ the reduced Gröbner basis $G$ for $\mathcal{A}$ is

$$xu + u^2 + v^2, \quad xv - u^2 + uv + v^2, \quad y - v, \quad z - u, \quad u^3 - uv^2 + v^3.$$

By Proposition 8, we see that $\ker \widetilde{\varphi}$ is generated by $u^3 - uv^2 + v^3 \equiv 0 \bmod J$, so $\widetilde{\varphi}$ is injective. Since there is no element of the form $x - h(u, v)$ in $G$, $\widetilde{\varphi}$ is not surjective (in fact $x$ is not in the image).

As a final example, we use the determination of the kernel of $k$-algebra homomorphisms to compute minimal polynomials of elements in simple algebraic field extensions.

**Proposition 10.** Suppose $\alpha$ is a root of the irreducible polynomial $p(x) \in k[x]$ and $\beta \in k(\alpha)$, say $\beta = f(\alpha)$ for the polynomial $f \in k[x]$. Let $G$ be the reduced Gröbner basis for the ideal $(p, y - f)$ in $k[x, y]$ for the lexicographic monomial ordering $x > y$. Then the minimal polynomial of $\beta$ over $k$ is the monic polynomial in $G \cap k[y]$.

*Proof:* The kernel of the $k$-algebra homomorphism $k[y] \to k[x]/(p) \cong k(\alpha)$ defined by mapping $y$ first to $f$ and then to $\beta$ is the principal ideal generated by the minimal polynomial of $\beta$ in $k[y]$, and the result follows by Proposition 8.

**Example**

Take $k = \mathbb{Q}$, and let $\beta = 1+\sqrt[3]{2}+3\sqrt[3]{4} \in \mathbb{Q}(\sqrt[3]{2})$. Then the ideal $(x^3-2, y-(1+x+3x^2))$ in $\mathbb{Q}[x, y]$ has reduced Gröbner basis $\{53x - 3y^2 + 7y + 32, y^3 - 3y^2 - 15y - 93\}$ for the lexicographic monomial ordering $x > y$, so the minimal polynomial for $\beta$ is $y^3 - 3y^2 - 15y - 93$.

## EXERCISES

Let $R$ be a commutative ring with $1 \neq 0$ and let $k$ be a field.

1. Prove the converse to Hilbert's Basis Theorem: if the polynomial ring $R[x]$ is Noetherian, then $R$ is Noetherian.
2. Show that each of the following rings are not Noetherian by exhibiting an explicit infinite increasing chain of ideals:
   **(a)** the ring of continuous real valued functions on $[0, 1]$,
   **(b)** the ring of all functions from any infinite set $X$ to $\mathbb{Z}/2\mathbb{Z}$.
3. Prove that the field $k(x)$ of rational functions over $k$ in the variable $x$ is not a finitely generated $k$-algebra. (Recall that $k(x)$ is the field of fractions of the polynomial ring $k[x]$. Note that $k(x)$ *is* a finitely generated *field extension* over $k$.)
4. Prove that if $R$ is Noetherian, then so is the ring $R[[x]]$ of formal power series in the variable $x$ with coefficients from $R$ (cf. Exercise 3, Section 7.2). [Mimic the proof of Hilbert's Basis Theorem.]
5. (*Fitting's Lemma*) Suppose $M$ is a Noetherian $R$-module and $\varphi : M \to M$ is an $R$-module endomorphism of $M$. Prove that $\ker(\varphi^n) \cap \text{image}(\varphi^n) = 0$ for $n$ sufficiently large. Show that if $\varphi$ is surjective, then $\varphi$ is an isomorphism. [Observe that $\ker(\varphi) \subseteq \ker(\varphi^2) \subseteq \ldots$.]
6. Suppose that $0 \longrightarrow M' \longrightarrow M \longrightarrow M'' \longrightarrow 0$ is an exact sequence of $R$-modules. Prove that $M$ is a Noetherian $R$-module if and only if $M'$ and $M''$ are Noetherian $R$-modules.
7. Prove that submodules, quotient modules, and finite direct sums of Noetherian $R$-modules are again Noetherian $R$-modules.
8. If $R$ is a Noetherian ring, prove that $M$ is a Noetherian $R$-module if and only if $M$ is a finitely generated $R$-module. (Thus any submodule of a finitely generated module over a Noetherian ring is also finitely generated.)
9. For $k$ a field show that any subring of the polynomial ring $k[x]$ containing $k$ is Noetherian. Give an example to show such subrings need not be U.F.D.s. [If $k \subset R \subseteq k[x]$ and $y \in R - k$ show that $k[x]$ is a finitely generated $k[y]$-module; then use the previous two exercises. For the second, consider $k[x^2, x^3]$.]
10. Prove that the subring $k[x, x^2y, x^3y^2, \ldots, x^iy^{i-1}, \ldots]$ of the polynomial ring $k[x, y]$ is *not* a Noetherian ring, hence not a finitely generated $k$-algebra. (Thus subrings of Noetherian rings need not be Noetherian and subalgebras of finitely generated $k$-algebras need not be finitely generated.)
11. Suppose $R$ is a commutative ring in which all the prime ideals are finitely generated. This exercise proves that $R$ is Noetherian.

**(a)** Prove that if the collection of ideals of $R$ that are not finitely generated is nonempty, then it contains a maximal element $I$, and that $R/I$ is a Noetherian ring.

**(b)** Prove that there are finitely generated ideals $J_1$ and $J_2$ containing $I$ with $J_1J_2 \subseteq I$ and that $J_1J_2$ is finitely generated. [Observe that $I$ is not a prime ideal.]

**(c)** Prove that $I/J_1J_2$ is a finitely generated $R/I$-submodule of $J_1/J_1J_2$. [Use Exercise 8.]

**(d)** Show that (c) implies the contradiction that $I$ would be finitely generated over $R$ and deduce that $R$ is Noetherian.

**12.** Suppose $R$ is a Noetherian ring and $S$ is a finitely generated $R$-algebra. If $T \subseteq S$ is an $R$-algebra such that $S$ is a finitely generated $T$-*module*, prove that $T$ is a finitely generated $R$-algebra. [If $s_1, \ldots, s_n$ generate $S$ as an $R$-algebra, and $s_1', \ldots, s_m'$ generate $S$ as a $T$-module, show that the elements $s_i$ and $s_j's_k'$ can be written as finite $T$-linear combinations of the $s_i'$. If $T_0$ is the $R$-subalgebra generated by the coefficients of these linear combinations, show $S$ (hence $T_0$) is finitely generated (by the $s_i'$) as a $T_0$-module, and conclude that $T$ is finitely generated as an $R$-algebra.]

**13.** Verify properties (1) to (10) of the maps $\mathcal{Z}$ and $\mathcal{I}$.

**14.** Show that the affine algebraic sets in $\mathbb{A}^1$ over any field $k$ are $\emptyset$, $k$, and finite subsets of $k$.

**15.** If $k = \mathbb{F}_2$ and $V = \{(0, 0), (1, 1)\} \subset \mathbb{A}^2$, show that $\mathcal{I}(V)$ is the product ideal $\mathfrak{m}_1\mathfrak{m}_2$ where $\mathfrak{m}_1 = (x, y)$ and $\mathfrak{m}_2 = (x - 1, y - 1)$.

**16.** Suppose that $V$ is a finite algebraic set in $\mathbb{A}^n$. If $V$ has $m$ points, prove that $k[V]$ is isomorphic as a $k$-algebra to $k^m$. [Use the Chinese Remainder Theorem.]

**17.** If $k$ is a finite field show that every subset of $\mathbb{A}^n$ is an affine algebraic set.

**18.** If $k = \mathbb{F}_q$ is the finite field with $q$ elements show that $\mathcal{I}(\mathbb{A}^1) = (x^q - x) \subset k[x]$.

**19.** For each nonconstant $f \in k[x]$ describe $\mathcal{Z}(f) \subseteq \mathbb{A}^1$ in terms of the unique factorization of $f$ in $k[x]$, and then use this to describe $\mathcal{I}(\mathcal{Z}(f))$. Deduce that $\mathcal{I}(\mathcal{Z}(f)) = (f)$ if and only if $f$ is the product of distinct linear factors in $k[x]$.

**20.** If $f$ and $g$ are irreducible polynomials in $k[x, y]$ that are not associates (do not divide each other), show that $\mathcal{Z}((f, g))$ is either $\emptyset$ or a finite set in $\mathbb{A}^2$. [If $(f, g) \neq (1)$, show $(f, g)$ contains a nonzero polynomial in $k[x]$ (and similarly a nonzero polynomial in $k[y]$) by letting $R = k[x]$, $F = k(x)$, and applying Gauss's Lemma to show $f$ and $g$ are relatively prime in $F[y]$.]

**21.** Identify each $2 \times 2$ matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with entries from $k$ with the point $(a, b, c, d)$ in $\mathbb{A}^4$. Show that the group $SL_2(k)$ of matrices of determinant 1 is an algebraic set in $\mathbb{A}^4$.

**22.** Prove that $SL_n(k)$ is an affine algebraic set in $\mathbb{A}^{n^2}$. [Generalize the preceding exercise.]

**23.** Let $V$ be any line in $\mathbb{R}^2$ (the zero set of any nonzero linear polynomial $ax + by - c$). Prove that $\mathbb{R}[V]$ is isomorphic as an $\mathbb{R}$-algebra to the polynomial ring $\mathbb{R}[x]$, and give the corresponding isomorphism from $\mathbb{A}^1$ to $V$.

**24.** Let $V = \mathcal{Z}(xy - z) \subseteq \mathbb{A}^3$. Prove that $V$ is isomorphic to $\mathbb{A}^2$ and provide an explicit isomorphism $\varphi$ and associated $k$-algebra isomorphism $\tilde{\varphi}$ from $k[V]$ to $k[\mathbb{A}^2]$, along with their inverses. Is $V = \mathcal{Z}(xy - z^2)$ isomorphic to $\mathbb{A}^2$?

**25.** Suppose $V \subseteq \mathbb{A}^n$ is an affine algebraic set and $f \in k[V]$. The *graph* of $f$ is the collection of points $\{(a_1, \ldots, a_n, f(a_1, \ldots, a_n))\}$ in $\mathbb{A}^{n+1}$. Prove that the graph of $f$ is an affine algebraic set isomorphic to $V$. [The morphism in one direction maps $(a_1, \ldots, a_n)$ to $(a_1, \ldots, a_n, f(a_1, \ldots, a_n))$.]

**26.** Let $V = \mathcal{Z}(xz - y^2, yz - x^3, z^2 - x^2y) \subseteq \mathbb{A}^3$.

**(a)** Prove that the map $\varphi : \mathbb{A}^1 \to V$ defined by $\varphi(t) = (t^3, t^4, t^5)$ is a surjective morphism. [For the surjectivity, if $(x, y, z) \neq (0, 0, 0)$, let $t = y/x$.]

**(b)** Describe the corresponding $k$-algebra homomorphism $\tilde{\varphi} : k[V] \to k[\mathbb{A}^1]$ explicitly.

**(c)** Prove that $\varphi$ is not an isomorphism.

**27.** Suppose $\varphi : V \to W$ is a morphism of affine algebraic sets. If $W'$ is an affine algebraic subset of $W$ prove that the preimage $V' = \varphi^{-1}(W')$ of $W'$ in $V$ is an affine algebraic subset of $V$. If $W' = \mathcal{Z}(I)$ show that $V' = \mathcal{Z}(\tilde{\varphi}(I))$ for the corresponding morphism $\tilde{\varphi} : k[W] \to k[V]$.

**28.** Prove that if $V$ and $W$ are affine algebraic sets, then so is $V \times W$ and $k[V \times W] \cong k[V] \otimes_k k[W]$.

The following seven exercises introduce the notion of the *associated primes* of an $R$-module $M$. Cf. also Exercises 30–40 in Section 4 and Exercises 25–30 in Section 5.

**Definition.** A prime ideal $P$ of $R$ is said to be *associated* to the $R$-module $M$ (sometimes called an *assassin* for $M$) if $P$ is the annihilator of some element $m$ of $M$, i.e., if $M$ contains a submodule $Rm$ isomorphic to $R/P$. The collection of associated primes for $M$ is denoted $\mathrm{Ass}_R(M)$.

When $M = I$ is an ideal in $R$, it is customary to abuse the terminology and refer instead to the elements of $\mathrm{Ass}_R(R/I)$ (rather than the less interesting collection $\mathrm{Ass}_R(I)$) as the *primes associated to* $I$. (Cf. Exercises 28–29 in Section 5.)

**29.** If $R = \mathbb{Z}$ and $M = \mathbb{Z}/n\mathbb{Z}$, show that $\mathrm{Ass}_R(M)$ consists of the prime ideals $(p)$ for the prime divisors $p$ of $n$.

**30.** If $M$ is the union of some collection of submodules $M_i$, prove that $\mathrm{Ass}_R(M)$ is the union of the collection $\mathrm{Ass}_R(M_i)$.

**31.** Suppose that $\mathrm{Ann}(m) = P$, i.e., that $Rm \cong R/P$. Prove that if $0 \neq m' \in Rm$ then $\mathrm{Ann}(m') = P$. Deduce that $\mathrm{Ass}_R(R/P) = \{P\}$. [Observe that $R/P$ is an integral domain.]

**32.** Suppose that $M$ is an $R$-module and that $P$ is a maximal element in the collection of ideals of the form $\mathrm{Ann}(m)$, for $m \in M$. Prove that $P$ is a prime ideal. [If $P = \mathrm{Ann}(m)$ and $ab \in P$, show that $bm \neq 0$ implies $\mathrm{Ann}(m) \subseteq \mathrm{Ann}(bm)$ and use the maximality of $P$ to deduce that $a \in \mathrm{Ann}(bm) = P$.]

**33.** Suppose $R$ is a Noetherian ring and $M \neq 0$ is an $R$-module. Prove that $\mathrm{Ass}_R(M) \neq \emptyset$. [Use Exercise 32.]

**34.** If $L$ is a submodule of $M$ with quotient $N \cong M/L$, prove that there are containments $\mathrm{Ass}_R(N) \subseteq \mathrm{Ass}_R(M) \subseteq \mathrm{Ass}_R(L) \cup \mathrm{Ass}_R(N)$, and show that both containments can be proper. [If $Rm \cong R/P$, show that $Rm \cap L = 0$ implies $P \in \mathrm{Ass}_R(N)$ and if $Rm \cap L \neq 0$ then $P \in \mathrm{Ass}_R(L)$ (by Exercise 31). For the second statement, consider $n\mathbb{Z} \subset \mathbb{Z}$.]

**35.** Suppose $M$ is an $R$-module and let $\mathcal{S}$ be a subset of the prime ideals in $\mathrm{Ass}_R(M)$. Prove there is a submodule $N$ of $M$ with $\mathrm{Ass}_R(N) = \mathcal{S}$ and $\mathrm{Ass}_R(M/N) = \mathrm{Ass}_R(M) - \mathcal{S}$. [Consider the collection of submodules $N'$ of $M$ with $\mathrm{Ass}_R(N') \subseteq \mathcal{S}$. Use Exercise 30 and Zorn's Lemma to show that there is a maximal submodule $N$ subject to $\mathrm{Ass}_R(N) \subseteq \mathcal{S}$. If $P \in \mathrm{Ass}_R(M/N)$, there is a submodule $M'/N \cong R/P$. Use the previous exercise to show that $\mathrm{Ass}_R(M') \subseteq \mathrm{Ass}_R(N) \cup \{P\}$ and then use maximality of $N$ to show $P \in \mathrm{Ass}_R(M) - \mathcal{S}$, so that $\mathrm{Ass}_R(M/N) \subseteq \mathrm{Ass}_R(M) - \mathcal{S}$ and $\mathrm{Ass}_R(N) \subseteq \mathcal{S}$. Use the previous exercise again to conclude that equality holds in each.]

Suppose $M$ is a finitely generated module over the commutative ring $R$ with generators $m_1, \ldots, m_n$. The *Fitting ideal* $\mathcal{F}_R(M)$ (of level 0) of $M$ (also called a *determinant ideal*) is the ideal in $R$ generated by the determinants of all $n \times n$ matrices $A = (r_{ij})$ where $r_{ij} \in R$ and $r_{i1}m_1 + \cdots + r_{in}m_n = 0$ in $M$, i.e., the rows of $A$ consist of the coefficients in $R$ of relations among the generators $m_i$ ($A$ is called an $n \times n$ "relations matrix" for $M$). The following five exercises outline some of the properties of the Fitting ideal.

**36.** **(a)** Show that the Fitting ideal of $M$ is also the ideal in $R$ generated by all the $n \times n$ minors of all $p \times n$ matrices $A = (r_{ij})$ for $p \geq 1$ whose rows consist of the coefficients in $R$ of relations among the generators $m_i$.

**(b)** Let $A$ be a fixed $p \times n$ matrix as in (a) and let $A'$ be a $p \times n$ matrix obtained from $A$ by any elementary row or column operation. Show that the ideal in $R$ generated by all the $n \times n$ minors of $A$ is the same as the ideal in $R$ generated by all the $n \times n$ minors of $A'$.

**37.** Suppose $m_1, \ldots, m_n$ and $m'_1, \ldots, m'_{n'}$ are two sets of $R$-module generators for $M$. Let $\mathcal{F}$ denote the Fitting ideal for $M$ computed using the generators $m_1, \ldots, m_n$ and let $\mathcal{F}'$ denote the Fitting ideal for $M$ computed using the generators $m_1, \ldots, m_n, m'_1, \ldots, m'_{n'}$.

**(a)** Show that $m'_s = a_{s'1}m_1 + \cdots + a_{s'n}m_n$ for some $a_{s'1}, \ldots, a_{s'n} \in R$, and deduce that $(-a_{s'1}, \ldots, -a_{s'n}, 0, \ldots, 0, 1, 0, \ldots 0)$ is a relation among $m_1, \ldots, m_n, m'_1, \ldots, m'_{n'}$.

**(b)** If $A = (r_{ij})$ is an $n \times n$ matrix whose rows are the coefficients of relations among $m_1, \ldots, m_n$ show that $\det A = \det A'$ where $A'$ is an $(n+n') \times (n+n')$ matrix whose rows are the coefficients of relations among $m_1, \ldots, m_n, m'_1, \ldots, m'_{n'}$. Deduce that $\mathcal{F} \subseteq \mathcal{F}'$. [Use (a) to find a block upper triangular $A'$ having $A$ in the upper left block and the $n' \times n'$ identity matrix in the lower right block.]

**(c)** Prove that $\mathcal{F}' \subseteq \mathcal{F}$ and conclude that $\mathcal{F}' = \mathcal{F}$. [Use the previous exercise.]

**(d)** Deduce from (c) that the Fitting ideal $\mathcal{F}_R(M)$ of $M$ is an invariant of $M$ that does not depend on the choice of generators for $M$ used to compute it.

**38.** All modules in this exercise are assumed finitely generated.

**(a)** If $M$ can be generated by $n$ elements prove that $\mathrm{Ann}(M)^n \subseteq \mathcal{F}_R(M) \subseteq \mathrm{Ann}(M)$, where $\mathrm{Ann}(M)$ is the annihilator of $M$ in $R$. [If $A$ is an $n \times n$ relations matrix for $M$, then $AX = 0$, where $X$ is the column matrix whose entries are $m_1, \ldots, m_n$. Multiply by the adjoint of $A$ to deduce that $\det A$ annihilates $M$.]

**(b)** If $M = M_1 \times M_2$ is the direct product of the $R$-modules $M_1$ and $M_2$ prove that $\mathcal{F}_R(M) = \mathcal{F}_R(M_1)\mathcal{F}_R(M_2)$.

**(c)** If $M = (R/I_1) \times \cdots \times (R/I_n)$ is the direct product of cyclic $R$-modules for ideals $I_i$ in $R$ prove that $\mathcal{F}_R(M) = I_1 I_2 \ldots I_n$.

**(d)** If $R = \mathbb{Z}$ and $M$ is a finitely generated abelian group show that $\mathcal{F}_{\mathbb{Z}}(M) = 0$ if $M$ is infinite and $\mathcal{F}_{\mathbb{Z}}(M) = |M|\mathbb{Z}$ if $M$ is finite.

**(e)** If $I$ is an ideal in $R$ prove that the image of $\mathcal{F}_R(M)$ in the quotient $R/I$ is $\mathcal{F}_{R/I}(M/IM)$.

**(f)** Prove that $\mathcal{F}_R(M/IM) \subseteq (\mathcal{F}_R(M), I) \subseteq R$.

**(g)** If $\varphi : M \to M'$ is a surjective $R$-module homomorphism prove $\mathcal{F}_R(M) \subseteq \mathcal{F}_R(M')$.

**(h)** If $0 \to L \to M \to N \to 0$ is a short exact sequence of $R$-modules, prove that $\mathcal{F}_R(L)\mathcal{F}_R(N) \subseteq \mathcal{F}_R(M)$.

**(i)** Suppose $R$ is the polynomial ring $k[x, y, z]$ over the field $k$. Let $M = R/(x, y^2, yz, z^2)$ and let $L$ be the submodule $(x, y, z)/(x, y^2, yz, z^2)$ of $M$. Prove that $\mathcal{F}_R(M)$ is $(x, y^2, yz, z^2)$ and $\mathcal{F}_R(L)$ is $(x, y, z)^2$. (This shows that in general the Fitting ideal of a submodule $L$ of $M$ need not contain the Fitting ideal for $M$.)

**39.** Suppose $M$ is an $R$-module and that $\varphi : R^n \to M$ is a surjective $R$-module homomorphism (i.e., $M$ can be generated by $n$ elements). Let $L = \ker \varphi$. Prove that the image of the

$R$-module homomorphism from $\bigwedge^n(L) \to \bigwedge^n(R^n) \cong R$ induced by the inclusion of $L$ in $R^n$ is the Fitting ideal $\mathcal{F}_R(M)$.

**40.** Suppose $R$ and $S$ are commutative rings, $\varphi : R \to S$ is a ring homomorphism, $M$ is a finitely generated $R$-module, and $M' = S \otimes_R M$ is the $S$-module obtained by extending scalars from $R$ to $S$. Prove that the Fitting ideal $\mathcal{F}_S(M')$ for $M'$ over $S$ is the extension to $S$ of the Fitting ideal $\mathcal{F}_R(M)$ for $M$ over $R$.

The following two exercises indicate how the remainder in Theorem 23 of Chapter 9 can be used to effect computations in quotients of polynomial rings.

**41.** Suppose $\{g_1, \ldots, g_m\}$ is a Gröbner basis for the ideal $I$ in $k[x_1, \ldots, x_n]$. Prove that the monomials $m$ not divisible by any $LT(g_i)$, $1 \le i \le m$, give a $k$-vector space basis for the quotient $k[x_1, \ldots, x_n]/I$.

**42.** Let $I = (x^3y - xy^2 + 1, x^2y^2 - y^3 - 1)$ as in Example 1 following Proposition 9.26.

**(a)** Use the previous exercise to show that $\{1, y, y^2, y^3\}$ is a basis for the $k$-vector space $k[x, y]/I$.

**(b)** Compute the $4 \times 4$ multiplication table for the basis vectors in (a).

**43.** Suppose $K[x_1, \ldots, x_n]$ is a polynomial ring in $n$ variables over a field $K$ and $k$ is a subfield of $K$. If $I$ is an ideal in $k[x_1, \ldots, x_n]$, let $I'$ be the ideal generated by $I$ in $K[x_1, \ldots, x_n]$.

**(a)** If $G$ is a Gröbner basis for the ideal $I$ in $k[x_1, \ldots, x_n]$ with respect to some monomial ordering, show that $G$ is also a Gröbner basis for the ideal $I'$ in $K[x_1, \ldots, x_n]$ with respect to the same monomial ordering. [Use Buchberger's Criterion.]

**(b)** Prove that the dimension of the quotient $k[x_1, \ldots, x_n]/I$ as a vector space over $k$ is the same as the dimension of the quotient $K[x_1, \ldots, x_n]/I'$ as a vector space over $K$. [One method: use (a) and Exercise 41.]

**(c)** Prove that $I = k[x_1, \ldots, x_n]$ if and only if $I' = K[x_1, \ldots, x_n]$.

**44.** Let $V = \mathcal{Z}(x^3 - x^2z - y^2z)$ and $W = \mathcal{Z}(x^2 + y^2 - z^2)$ in $\mathbb{C}^3$. Then $\mathcal{I}(V) = (x^3 - x^2z - y^2z)$ and $\mathcal{I}(W) = (x^2 + y^2 - z^2)$ in $\mathbb{C}[x, y, z]$ (cf. Exercise 23 in Section 3). Show that $\varphi((a, b, c)) = (a^2c - b^2c, 2abc, -a^3)$ defines a morphism from $V$ to $W$.

**45.** Let $V = \mathcal{Z}(x^3 + y^3 + 7z^3) \subset \mathbb{C}^3$. Then $\mathcal{I}(V) = (x^3 + y^3 + 7z^3)$ in $\mathbb{C}[x, y, z]$ (cf. Exercise 24 in Section 3).

**(a)** Show that

$$\tilde{\varphi}(x) = x(y^3 - 7z^3), \qquad \tilde{\varphi}(y) = y(7z^3 - x^3), \qquad \tilde{\varphi}(z) = z(x^3 - y^3)$$

defines a $\mathbb{C}$-algebra homomorphism from $k[V]$ to itself.

**(b)** Let $\varphi : V \to V$ be the morphism corresponding to $\tilde{\varphi}$. Observe that $(-2, 1, 1) \in V$ and compute $\varphi((-2, 1, 1)) \in V$.

**(c)** Prove there are infinitely many points $(a, b, c)$ on $V$ with $a, b, c \in \mathbb{Z}$ and the greatest common divisor of $a$, $b$, and $c$ is 1.

**46.** Let $V = \mathcal{Z}(xz + y^2 + z^2, xy - xz + yz - 2z^2) \subset \mathbb{C}^3$ and $W = \mathcal{Z}(u^3 - uv^2 + v^3) \subset \mathbb{C}^2$ as in Example 2 following Corollary 9. Show that the map $\varphi((a, b)) = (-2a^2 + ab, ab - b^2, a^2 - ab)$ defines a morphism from $W$ to $V$. Show the corresponding $\mathbb{C}$-algebra homomorphism from $k[V]$ to $k[W]$ has a kernel generated by $x^2 - 3y^2 + yz$.

**47.** Define $\Phi : \mathbb{Q}[u, v, w] \to \mathbb{Q}[x, y]$ by $\Phi(u) = x^2 + y$, $\Phi(v) = x + y^2$, and $\Phi(w) = x - y$. Show that neither $x$ nor $y$ is in the image of $\Phi$. Show that $f = 2x^3 - 4xy - 2y^3 - 4y$ is in the image of $\Phi$ and find a polynomial in $\mathbb{Q}[u, v, w]$ mapping to $f$. Show that $\ker \Phi$ is the ideal generated by

$$u^2 - 2uv - 2uw^2 + 4uw + v^2 - 2vw^2 - 4vw + w^4 + 3w^2.$$

**48.** Suppose $\alpha$ is a root of the irreducible polynomial $p(x) \in k[x]$ and $\beta = f(\alpha)/g(\alpha)$ with polynomials $f(x), g(x) \in k[x]$ where $g(\alpha) \neq 0$.

**(a)** Show $ag + bp = 1$ for some polynomials $a, b \in k[x]$ and show $\beta = h(\alpha)$ where $h = af$.

**(b)** Show that the ideals $(p, y - h)$ and $(p, gy - f)$ are equal in $k[x, y]$.

**(c)** Conclude that the minimal polynomial for $\beta$ is the monic polynomial in $G \cap k[y]$ where $G$ is the reduced Gröbner basis for the ideal $(p, gy - f)$ in $k[x, y]$ for the lexicographic monomial ordering $x > y$.

**(d)** Find the minimal polynomial over $\mathbb{Q}$ of $(3 - \sqrt[3]{2} + \sqrt[3]{4})/(1 + 3\sqrt[3]{2} - 3\sqrt[3]{4})$.

## 15.2 RADICALS AND AFFINE VARIETIES

Since the zeros of a polynomial $f$ are the same as the zeros of the powers $f^2, f^3, \ldots$ in general there are many different ideals in the ring $k[x_1, x_2, \ldots, x_n]$ whose zero locus define the same algebraic set $V$ in affine $n$-space. This leads to the notion of the radical of an ideal, which can be defined in any commutative ring:

**Definition.** Let $I$ be an ideal in a commutative ring $R$.

**(1)** The *radical* of $I$, denoted by $\operatorname{rad} I$, is the collection of elements in $R$ some power of which lie in $I$, i.e.,

$$\operatorname{rad} I = \{a \in R \mid a^k \in I \text{ for some } k \geq 1\}.$$

**(2)** The radical of the zero ideal is called the *nilradical* of $R$.

**(3)** An ideal $I$ is called a *radical* ideal if $I = \operatorname{rad} I$.

Note that $a \in R$ is in the nilradical of $R$ if and only if some power of $a$ is 0, so the nilradical of $R$ is the set of all nilpotent elements of $R$.

**Proposition 11.** Let $I$ be an ideal in the commutative ring $R$. Then $\operatorname{rad} I$ is an ideal containing $I$, and $(\operatorname{rad} I)/I$ is the nilradical of $R/I$. In particular, $R/I$ has no nilpotent elements if and only if $I = \operatorname{rad} I$ is a radical ideal.

*Proof:* It is clear that $I \subseteq \operatorname{rad} I$. By definition, the nilradical of $R/I$ consists of the elements in the quotient some power of which is 0. Under the Lattice Isomorphism Theorem for rings this collection of elements corresponds to the elements of $R$ some power of which lie in $I$, i.e., $\operatorname{rad} I$. It is therefore sufficient to prove that the nilradical $N$ of any commutative ring $R$ is an ideal. Since $0 \in N$, $N \neq \emptyset$. If $a \in N$ and $r \in R$, then since $a^n = 0$ for some $n \geq 1$, the commutativity of $R$ implies that $(ra)^n = r^n a^n = 0$, so $ra \in N$. It remains to see that if $a, b \in N$ then $a + b \in N$. Suppose $a^n = 0$ and $b^m = 0$. Since the Binomial Theorem holds in the commutative ring $R$ (cf. Exercise 25 in Section 7.3),

$$(a + b)^{n+m} = \sum_{i=0}^{n+m} r_i a^i b^{n+m-i}$$

for some ring elements $r_i$ (the binomial coefficients in $R$). For each term in this sum either $i \geq n$ (in which case $a^i = 0$) or $n + m - i \geq m$, (in which case $b^{n+m-i} = 0$). Hence $(a + b)^{n+m} = 0$, which shows that $a + b$ is nilpotent, i.e., $a + b \in N$.

**Proposition 12.** The radical of a proper ideal $I$ is the intersection of all prime ideals containing $I$. In particular, the nilradical is the intersection of all the prime ideals in $R$.

*Proof:* Passing to $R/I$, Proposition 11 shows that it suffices to prove this result for $I = 0$, and in this case the statement is that the nilradical $N$ of $R$ is the intersection of all the prime ideals in $R$. Let $N'$ denote the intersection of all the prime ideals in $R$.

Let $a$ be any nilpotent element in $R$ and let $P$ be any prime ideal. Since $a^k = 0$ for some $k$, there is a smallest positive power $n$ such that $a^n \in P$. Then the product $a^{n-1}a \in P$, and since $P$ is prime, either $a^{n-1} \in P$ or $a \in P$. The former contradicts the minimality of $n$, and so $a \in P$. Since $P$ was arbitrary, $a \in N'$, which shows that $N \subseteq N'$.

We prove the reverse containment $N' \subseteq N$ by showing that if $a \notin N$, then $a \notin N'$. If $a$ is an element of $R$ not contained in $N$, let $\mathcal{S}$ be the family of all proper ideals not containing any positive power of $a$. The collection $\mathcal{S}$ is not empty since $0 \in \mathcal{S}$. Also, if $a^k$ is not contained in any ideal in the chain $I_1 \subseteq I_2 \subseteq \cdots$, then $a^k$ is also not contained in the union of these ideals, which shows that chains in $\mathcal{S}$ have upper bounds. By Zorn's Lemma, $\mathcal{S}$ has a maximal element, $P$. The ideal $P$ must in fact be a prime ideal, as follows. Suppose for some $x$ and $y$ not contained in $P$, the product $xy$ is an element of $P$. By the maximality of $P$, $a^n \in (x) + P$ and $a^m \in (y) + P$ for some positive integers $n$ and $m$. Then $a^{n+m} \in (xy) + P = P$ contradicting the fact that $P$ is an element of $\mathcal{S}$. This shows that $P$ is indeed a prime ideal not containing $a$, and hence $a \notin N'$, completing the proof.

Note that in Noetherian rings, Theorem 2 can be used to circumvent the appeal to Zorn's Lemma in the preceding proof.

**Corollary 13.** Prime (and hence also maximal) ideals are radical.

*Proof:* If $P$ is a prime ideal, then $P$ is clearly the intersection of all the prime ideals containing $P$, so $P = \text{rad } P$ by the proposition.

**Examples**

**(1)** In the ring of integers $\mathbb{Z}$, the ideal $(a)$ is a radical ideal if and only if $a$ is square-free or zero. More generally, if $a = p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r}$ with $a_i \geq 1$ for all $i$, is the prime factorization of the positive integer $a$, then $\text{rad}(a) = (p_1 p_2 \cdots p_r)$. For instance, $\text{rad}(180) = (30)$. Note that $(p_1), (p_2), \ldots, (p_r)$ are precisely the prime ideals containing the ideal $(a)$ and that their intersection is the ideal $(p_1 p_2 \cdots p_r)$. More generally, in any U.F.D. $R$, $\text{rad}(a) = (p_1 p_2 \cdots p_r)$ if $a = p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r}$ is the unique factorization of $a$ into distinct irreducibles.

**(2)** The ideal $(x^3 - y^2)$ in $k[x, y]$ is a prime ideal (Exercise 14, Section 9.1), hence is radical.

**(3)** If $l_1, \ldots, l_m$ are linear polynomials in $k[x_1, x_2, \ldots, x_n]$ then $I = (l_1, \ldots, l_m)$ is either $k[x_1, x_2, \ldots, x_n]$ or a prime ideal, hence $I$ is a radical ideal.

**Proposition 14.** If $R$ is a Noetherian ring then for any ideal $I$ some positive power of $\text{rad } I$ is contained in $I$. In particular, the nilradical, $N$, of a Noetherian ring is a nilpotent ideal: $N^k = 0$ for some $k \geq 1$.

*Proof:* For any ideal $I$, the ideal rad $I$ is finitely generated since $R$ is Noetherian. If $a_1, \ldots, a_m$ are generators of rad $I$, then by definition of the radical, for each $i$ we have $a_i^{k_i} \in I$ for some positive integer $k_i$. Let $k$ be the maximum of all the $k_i$. Then the ideal $(\text{rad } I)^{km}$ is generated by elements of the form $a_1^{d_1} a_2^{d_2} \cdots a_m^{d_m}$ where $d_1 + \cdots + d_m = km$, and each of these elements has at least one factor $a_i^{d_i}$ with $d_i \geq k$. Then $a_i^{d_i} \in I$, hence each generator of $(\text{rad } I)^{km}$ lies in $I$, and so $(\text{rad } I)^{km} \subseteq I$.

## The Zariski Topology

We saw in the preceding section that if we restrict to the set of ideals $I$ of $k[\mathbb{A}^n]$ arising as the ideals associated with some algebraic set $V$, i.e., with $I = \mathcal{I}(V)$, then the maps $\mathcal{Z}$ (from such ideals to algebraic sets) and $\mathcal{I}$ (from algebraic sets to ideals) are inverses of each other: $\mathcal{Z}(\mathcal{I}(V)) = V$ and $\mathcal{I}(\mathcal{Z}(I)) = I$. The elements of the ring $k[\mathbb{A}^n]/\mathcal{I}(V)$ give $k$-valued functions on $V$ and, since $k$ has no nilpotent elements, powers of nonzero functions are also nonzero functions. Put another way, the ring $k[\mathbb{A}^n]/\mathcal{I}(V)$ has no nilpotent elements, so by Proposition 11, the ideal $\mathcal{I}(V)$ is always a radical ideal.

For arbitrary fields $k$, it is in general not true that every radical ideal is the ideal of some algebraic set, i.e., of the form $\mathcal{I}(V)$ for some algebraic set $V$. For example, the ideal $(x^2 + 1)$ in $\mathbb{R}[x]$ is maximal, hence is a radical ideal (by Corollary 13), but is not the ideal of any algebraic set — if it were, then $x^2 + 1$ would have to vanish on that set, but $x^2 + 1$ has no zeros in $\mathbb{R}$. A similar construction works for any field $k$ that is not algebraically closed — there exists an irreducible polynomial $p(x)$ of degree at least 2 in $k[x]$, which then generates the maximal (hence radical) ideal $(p(x))$ in $k[x]$ that has no zeros in $k$. It is perhaps surprising that the presence of polynomials in one variable that have no zeros is the *only* obstruction to a radical ideal (in *any* number of variables) not being of the form $\mathcal{I}(V)$. This is shown by the next theorem, which provides a fundamental connection between "geometry" and "algebra" and shows that over an *algebraically closed* field (such as $\mathbb{C}$) every radical ideal is of the form $\mathcal{I}(V)$. Over these fields the "geometrically defined" ideals $I = \mathcal{I}(V)$ are therefore the same as the radical ideals, which is a "purely algebraic" property of the ideal $I$ (namely that $I = \text{rad } I$).

**Theorem.** *(Hilbert's Nullstellensatz)* Let $E$ be an algebraically closed field. Then $\mathcal{I}(\mathcal{Z}(I)) = \text{rad } I$ for every ideal $I$ of $E[x_1, x_2, \ldots, x_n]$. Moreover, the maps $\mathcal{Z}$ and $\mathcal{I}$ in the correspondence

$$\{\text{affine algebraic sets}\} \underset{\mathcal{Z}}{\overset{\mathcal{I}}{\rightleftarrows}} \{\text{radical ideals}\}$$

are bijections that are inverses of each other.

*Proof:* This will be proved in the next section (cf. Theorem 32).

**Example**

The maps $\mathcal{I}$ and $\mathcal{Z}$ in the Nullstellensatz are defined over any field $k$, and as mentioned are not bijections if $k$ is not algebraically closed. For any field $k$, however, the map $\mathcal{Z}$ is always surjective and the map $\mathcal{I}$ is always injective (cf. Exercise 9).

One particular consequence of the Nullstellensatz is that for any *proper* ideal $I$ we have $\mathcal{Z}(I) \neq \emptyset$ since $\operatorname{rad} I \neq k[\mathbb{A}^n]$. Hence there always exists at least one common zero ("nullstellen" in German) for all the polynomials contained in a proper ideal (over an algebraically closed field).

We next see that the affine algebraic sets define a topology on affine $n$-space. Recall that a *topological space* is any set $X$ together with a collection of subsets $\mathcal{T}$ of $X$, called the *closed sets* in $X$, satisfying the following axioms:

**(i)** an arbitrary intersection of closed sets is closed: if $S_i \in \mathcal{T}$ for $i$ in any index set, then $\cap S_i \in \mathcal{T}$,

**(ii)** a finite union of closed sets is closed: if $S_1, \ldots, S_q \in \mathcal{T}$ then $S_1 \cup \cdots \cup S_q \in \mathcal{T}$, and

**(iii)** the empty set and the whole space are closed: $\emptyset, X \in \mathcal{T}$.

A subset $U$ of $X$ is called *open* if its complement, $X - U$, is closed (i.e., $X - U \in \mathcal{T}$). The axioms for a topological space are often (equivalently) phrased in terms of the collection of open sets in $X$.

There are many examples of topological spaces, and a wealth of books on topology. A fixed set $X$ may have a number of different topologies on it, and the collections of closed sets need not be related in these different structures. On any set $X$ there are always at least two topologies: the so-called discrete topology in which every subset of $X$ is closed (i.e., $\mathcal{T}$ is the collection of *all* subsets of $X$), and the so-called trivial topology in which the only closed sets are $\emptyset$ and $X$ required by axiom (iii).

Suppose now that $X = \mathbb{A}^n$ is affine $n$-space over an arbitrary field $k$. Then the collection $\mathcal{T}$ consisting of all the affine algebraic sets in $\mathbb{A}^n$ satisfies the three axioms for a topological space — these are precisely properties (3), (4) and (5) of algebraic sets in the preceding section. It follows that these sets can be taken to be the closed sets in a topology on $\mathbb{A}^n$:

**Definition.** The *Zariski topology* on affine $n$-space over an arbitrary field $k$ is the topology in which the closed sets are the affine algebraic sets in $\mathbb{A}^n$.

The Zariski topology is quite "coarse" in the sense that there are "relatively few" closed (or open) sets. For example, for the Zariski topology on $\mathbb{A}^1$ the only closed sets are $\emptyset$, $k$ and the finite sets (cf. Exercise 14 in Section 1), and so the nonempty open sets are the complements of finite sets. If $k$ is an infinite field it follows that in the Zariski topology any two nonempty open sets in $\mathbb{A}^1$ have nonempty intersection. In the language of point-set topology, the Zariski topology is always $T_1$ (points are closed sets), but for infinite fields the Zariski topology is never $T_2$ (Hausdorff), i.e., two distinct points never belong to two disjoint open sets (cf. the exercises). For example, when $k = \mathbb{R}$, a nonempty Zariski open set is just the real line $\mathbb{R}$ with some finite number of points removed, and any two such sets have (infinitely many) points in common. Note also that the Zariski open (respectively, closed) sets in $\mathbb{R}$ are also open (respectively, closed) sets with respect to the usual Euclidean topology. The converse is not true; for example the interval $[0,1]$ is closed in the Euclidean topology but is not closed in the Zariski topology. In this sense the Euclidean topology on $\mathbb{R}$ is much "finer"; there are

many more open sets in the Euclidean topology, in fact the collection of Euclidean open (respectively, closed) sets properly contains the collection of Zariski open (respectively, closed) sets.

The Zariski topology on $\mathbb{A}^n$ is defined so that the affine algebraic subsets of $\mathbb{A}^n$ are closed. In other words, the topology is defined by the zero sets of the ideals in the coordinate ring of $\mathbb{A}^n$. A similar definition can be used to define a Zariski topology on *any* algebraic set $V$ in $\mathbb{A}^n$, as follows. If $k[V]$ is the coordinate ring of $V$, then the distinct elements of $k[V]$ define distinct $k$-valued functions on $V$ and there is a natural way of defining

$$\mathcal{Z} : \{\text{ideals in } k[V]\} \longrightarrow \{\text{algebraic subsets of } V\}$$
$$\mathcal{I} : \{\text{subsets of } V\} \longrightarrow \{\text{ideals in } k[V]\}$$

just as for the case $V = \mathbb{A}^n$. For example, if $\overline{J}$ is an ideal in $k[V]$, then $\mathcal{Z}(\overline{J})$ is the set of elements in $V$ that are common zeros of all the functions in the ideal $\overline{J}$. It is easy to verify that the resulting zero sets in $V$ satisfy the three axioms for a topological space, defining a *Zariski topology on* $V$, where the closed sets are the algebraic subsets, $\mathcal{Z}(\overline{J})$, for any ideal $\overline{J}$ of $k[V]$. By the Lattice Isomorphism Theorem, the ideals of $k[V]$ are the ideals of $k[x_1, \ldots, x_n]$ that contain $\mathcal{I}(V)$ taken mod $\mathcal{I}(V)$. If $J$ is the complete preimage in $k[x_1, \ldots, x_n]$ of $\overline{J}$, then the locus of $J$ in $\mathbb{A}^n$ is the same as the locus of $\overline{J}$ in $V$. It follows that this definition of the Zariski topology on $V$ is just the *subspace topology* for $V \subseteq \mathbb{A}^n$. (Recall that in a topological space $X$, the closed sets with respect to the subspace topology of a subspace $Y$ are defined to be the sets $C \cap Y$, where $C$ is a closed set in $X$.) The advantage to the definition of the Zariski topology on $V$ above is that it is defined intrinsically in terms of the coordinate ring $k[V]$ of $V$, and since the isomorphism type of $k[V]$ does not depend on the affine space $\mathbb{A}^n$ containing $V$, the Zariski topology on $V$ also depends only on $V$ and not on the ambient affine space in which $V$ may be embedded.

If $V$ and $W$ are two affine algebraic spaces, then since a morphism $\varphi : V \to W$ is defined by polynomial functions, it is easy to see that $\varphi$ is *continuous* with respect to the Zariski topologies on $V$ and $W$ (cf. Exercise 27 in Section 1, which shows that the inverse image of a Zariski closed set under a morphism is Zariski closed). In fact the Zariski topology is the coarsest topology in which points are closed and for which polynomial maps are continuous. There exist maps that are continuous with respect to the Zariski topology that are not morphisms, however (cf. Exercise 17).

We have the usual topological notions of closure and density with respect to the Zariski topology.

**Definition.** For any subset $A$ of $\mathbb{A}^n$, the *Zariski closure* of $A$ is the smallest algebraic set containing $A$. If $A \subseteq V$ for an algebraic set $V$ then $A$ is *Zariski dense* in $V$ if the Zariski closure of $A$ is $V$.

For example, if $k = \mathbb{R}$, the algebraic sets in $\mathbb{A}^1$ are $\emptyset$, $\mathbb{R}$, and finite subsets of $\mathbb{R}$ by Exercise 14 in Section 1. The Zariski closure of any infinite set $A$ of real numbers is then all of $\mathbb{A}^1$ and $A$ is Zariski dense in $\mathbb{A}^1$.

**Proposition 15.** The Zariski closure of a subset $A$ in $\mathbb{A}^n$ is $\mathcal{Z}(\mathcal{I}(A))$.

*Proof:* Certainly $A \subseteq \mathcal{Z}(\mathcal{I}(A))$. Suppose $V$ is any algebraic set containing $A$: $A \subseteq V$. Then $\mathcal{I}(V) \subseteq \mathcal{I}(A)$ and $\mathcal{Z}(\mathcal{I}(A)) \subseteq \mathcal{Z}(\mathcal{I}(V)) = V$, so $\mathcal{Z}(\mathcal{I}(A))$ is the smallest algebraic set containing $A$.

If $\varphi : V \to W$ is a morphism of algebraic sets, the image $\varphi(V)$ of $V$ need not be an algebraic subset of $W$, i.e., need not be Zariski closed in $W$. For example the projection of the hyperbola $V = \mathcal{Z}(xy - 1)$ in $\mathbb{R}^2$ onto the $x$-axis has image $\mathbb{R}^1 - \{0\}$, which as we have just seen is not an affine algebraic set.

The next result shows that the Zariski closure of the image of a morphism is determined by the kernel of the associated $k$-algebra homomorphism.

**Proposition 16.** Suppose $\varphi : V \to W$ is a morphism of algebraic sets and $\widetilde{\varphi} : k[W] \to k[V]$ is the associated $k$-algebra homomorphism of coordinate rings. Then

**(1)** The kernel of $\widetilde{\varphi}$ is $\mathcal{I}(\varphi(V))$.

**(2)** The Zariski closure of $\varphi(V)$ is the zero set in $W$ of $\ker \widetilde{\varphi}$. In particular, the homomorphism $\widetilde{\varphi}$ is injective if and only if $\varphi(V)$ is Zariski dense in $W$.

*Proof:* Since $\widetilde{\varphi} = f \circ \varphi$, we have $\widetilde{\varphi}(f) = 0$ if and only if $(f \circ \varphi)(P) = 0$ for all $P \in V$, i.e., $f(Q) = 0$ for all $Q = \varphi(P) \in \varphi(V)$, which is the statement that $f \in \mathcal{I}(\varphi(V))$, proving the first statement. Since the Zariski closure of $\varphi(V)$ is the zero set of $\mathcal{I}(\varphi(V))$ by the previous proposition, the first statement in (2) follows.

If $\widetilde{\varphi}$ is injective then the Zariski closure of $\varphi(V)$ is $\mathcal{Z}(0) = W$ and so $\varphi(V)$ is Zariski dense. Conversely, suppose $\varphi(V)$ is Zariski dense in $W$, i.e., $\mathcal{Z}(\mathcal{I}(\varphi(V))) = W$. Then $\mathcal{I}(\varphi(V)) = \mathcal{I}(\mathcal{Z}(\mathcal{I}(\varphi(V)))) = \mathcal{I}(W) = 0$ and so $\ker \widetilde{\varphi} = 0$.

By Proposition 16 the ideal of polynomials defining the Zariski closure of the image of a morphism $\varphi$ is the kernel of the corresponding $k$-algebra homomorphism $\widetilde{\varphi}$ in Theorem 6. Proposition 8(1) allows us to compute this kernel using Gröbner bases.

**Example: (Implicitization)**

A morphism $\varphi : \mathbb{A}^n \to \mathbb{A}^m$ is just a map

$$\varphi((a_1, a_2, \ldots, a_n)) = (\varphi_1(a_1, a_2, \ldots, a_n), \ldots, \varphi_m(a_1, a_2, \ldots, a_n))$$

where $\varphi_i$ is a polynomial. If $k$ is an infinite field, then $\mathcal{I}(\mathbb{A}^m)$ and $\mathcal{I}(\mathbb{A}^n)$ are both 0, so we may write $k[\mathbb{A}^m] = k[y_1, \ldots, y_m]$ and $k[\mathbb{A}^n] = k[x_1, \ldots, x_n]$. The $k$-algebra homomorphism $\widetilde{\varphi} : k[\mathbb{A}^m] \to k[\mathbb{A}^n]$ corresponding to $\varphi$ is then defined by mapping $y_i$ to $\varphi_i = \varphi_i(x_1, \ldots, x_n)$. The image $\varphi(\mathbb{A}^n)$ consists of the set of points $(b_1, \ldots, b_m)$ with

$$\begin{aligned} b_1 &= \varphi_1(a_1, a_2, \ldots, a_n) \\ b_2 &= \varphi_2(a_1, a_2, \ldots, a_n) \\ &\vdots \\ b_m &= \varphi_m(a_1, a_2, \ldots, a_n) \end{aligned}$$

where $a_i \in k$. This is the collection of points in $\mathbb{A}^m$ *parametrized* by the functions $\varphi_1, \ldots, \varphi_m$ (with the $a_i$ as parameters). In general such a parametrized collection of points

is not an algebraic set. Finding the equations for the smallest algebraic set containing these points is referred to as *implicitization*, since it amounts to finding a ('smallest') collection of equations satisfied by the $b_i$ (the 'implicit' algebraic relations).

By Proposition 16, this algebraic set is the Zariski closure of $\varphi(\mathbb{A}^n)$ and is the zero set of $\ker\widetilde{\varphi}$. By Proposition 8 this kernel is given by $\mathcal{A} \cap k[y_1, \ldots, y_m]$, where $\mathcal{A}$ is the ideal in $k[x_1, \ldots, x_n, y_1, \ldots, y_m]$ generated by the polynomials $y_1 - \varphi_1, \ldots, y_m - \varphi_m$. If we compute the reduced Gröbner basis $G$ for $\mathcal{A}$ with respect to the lexicographic monomial ordering $x_1 > \cdots > x_n > y_1 > \cdots > y_m$, then the polynomials of $G$ lying in $k[y_1, \ldots, y_m]$ generate $\ker\tilde{\pi}$. The zero set of these polynomials defines the Zariski closure of $\varphi(\mathbb{A}^n)$ and therefore give the implicitization.

For an explicit example, consider the points $A = \{(a^2, a^3) \mid a \in \mathbb{R}\}$ in $\mathbb{R}^2$. Using coordinates $x, y$ for $\mathbb{R}^2$ and $t$ for $\mathbb{R}^1$, the ideal $\mathcal{A}$ in $\mathbb{R}[x, y, z, t]$ is $(x - t^2, y - t^3)$. The only element of the reduced Gröbner basis for $\mathcal{A}$ for the ordering $t > x > y$ lying in $\mathbb{R}[x, y]$ is $x^3 - y^2$, so $\mathcal{Z}(x^3 - y^2)$ is the smallest algebraic set in $\mathbb{R}^2$ containing $A$.

**Example: (Projections of Algebraic Sets)**

Suppose $V \subseteq \mathbb{A}^n$ is an algebraic set and $m < n$. Let $\pi : V \to \mathbb{A}^m$ be the morphism projecting onto the first $m$ coordinates:

$$\pi((a_1, a_2, \ldots, a_n)) = (a_1, a_2, \ldots, a_m).$$

If we use coordinates $x_1, \ldots, x_n$ in $k[V]$ and coordinates $y_1, \ldots, y_m$ in $k[\mathbb{A}^m]$, the $k$-algebra homomorphism corresponding to $\pi$ is given by the map

$$\tilde{\pi} : k[y_1, \ldots, y_m] \longrightarrow k[x_1, \ldots, x_n]/\mathcal{I}(V)$$
$$y_i \longmapsto x_i.$$

Suppose $V = \mathcal{Z}(I)$ and $I = (f_1, \ldots, f_s)$. The Zariski closure of $\pi(V)$ is the zero set of $\ker\tilde{\pi} = \mathcal{A} \cap k[y_1, \ldots, y_m]$ where $\mathcal{A}$ is the ideal in $k[x_1, \ldots, x_n, y_1, \ldots, y_m]$ generated by the polynomials $y_1 - x_1, \ldots, y_m - x_m$ together with a set of generators for $\mathcal{I}(V)$. The polynomials involving only $y_1, \ldots, y_m$ in the reduced Gröbner basis $G$ for $\mathcal{A}$ with respect to the lexicographic monomial ordering $x_1 > \cdots > x_n > y_1 > \cdots > y_m$ are generators for the Zariski closure of $\pi(V)$.

If $k$ is algebraically closed we can actually do better with the help of the Nullstellensatz, which gives $\mathcal{I}(V) = \operatorname{rad} I$. Then it is straightforward to see that we obtain the same zero set if in the ideal $\mathcal{A}$ we replace the generators for $\mathcal{I}(V)$ by the generators $f_1, \ldots, f_s$ of $I$ (cf. Exercise 46).

For an explicit example, consider projection onto the first two coordinates of $V = \mathcal{Z}(xy - z^2, xz - y, x^2 - z)$ in $\mathbb{C}^3$. Using $u, v$ as coordinates in $\mathbb{C}^2$, we find the reduced Gröbner basis $G$ for the ideal $(u - x, v - y, xy - z^2, xz - y, x^2 - z)$ for the ordering $x > y > z > u > v$ contains only the polynomial $u^3 - v$ in $\mathbb{C}[u, v]$. The smallest algebraic set containing $\pi(V)$ is then the cubic $v = u^3$.

## Affine Varieties

We next consider the question of whether an algebraic set can be decomposed into smaller algebraic sets and the corresponding algebraic formulation in terms of its coordinate ring.

**Definition.** A nonempty affine algebraic set $V$ is called *irreducible* if it cannot be written as $V = V_1 \cup V_2$, where $V_1$ and $V_2$ are proper algebraic sets in $V$. An irreducible affine algebraic set is called an affine *variety*.

Equivalently, an algebraic set (which is a closed set in the Zariski topology) is irreducible if it cannot be written as the union of two proper, closed subsets.

**Proposition 17.**

**(1)** The affine algebraic set $V$ is irreducible if and only if $\mathcal{I}(V)$ is a prime ideal.

**(2)** Every nonempty affine algebraic set $V$ may be written uniquely in the form

$$V = V_1 \cup V_2 \cup \cdots \cup V_q$$

where each $V_i$ is irreducible, and $V_i \not\subseteq V_j$ for all $j \neq i$ (i.e., the decomposition is "minimal" or "irredundant").

*Proof:* Let $I = \mathcal{I}(V)$ and suppose first that $V = V_1 \cup V_2$ is reducible, where $V_1$ and $V_2$ are proper closed subsets. Since $V_1 \neq V$, there is some function $f_1$ that vanishes on $V_1$ but not on $V$, i.e., $f_1 \in \mathcal{I}(V_1) - I$. Similarly, there is a function $f_2 \in \mathcal{I}(V_2) - I$. Then $f_1 f_2$ vanishes on $V_1 \cup V_2 = V$, so $f_1 f_2 \in I$ which shows that $I$ is not a prime ideal. Conversely, if $I$ is not a prime ideal, there exists $f_1, f_2 \in k[\mathbb{A}^n]$ such that $f_1 f_2 \in I$ but neither $f_1$ nor $f_2$ belongs to $I$. Let $V_1 = \mathcal{Z}(f_1) \cap V$ and $V_2 = \mathcal{Z}(f_2) \cap V$. Since the intersection of closed sets is closed, $V_1$ and $V_2$ are algebraic sets. Since neither $f_1$ nor $f_2$ vanishes on $V$, both $V_1$ and $V_2$ are proper subsets of $V$. Because $f_1 f_2 \in I$, $V \subseteq \mathcal{Z}(f_1 f_2) = \mathcal{Z}(f_1) \cup \mathcal{Z}(f_2)$, and so $V$ is reducible. This proves (1).

To prove (2), let $\mathcal{S}$ be the collection of nonempty algebraic sets that cannot be written as a finite union of irreducible algebraic sets, and suppose by way of contradiction that $\mathcal{S} \neq \emptyset$. Let $I_0$ be a maximal element of the corresponding set of ideals, $\{\mathcal{I}(V) \mid V \in \mathcal{S}\}$, which exists (by Theorem 2) since $k[\mathbb{A}^n]$ is Noetherian. Then $V_0 = \mathcal{Z}(I_0)$ is a *minimal* element of $\mathcal{S}$. Since $V_0 \in \mathcal{S}$, it cannot be irreducible by the definition of $\mathcal{S}$. On the other hand, if $V_0 = V_1 \cup V_2$ for some proper, closed subsets $V_1, V_2$ of $V_0$, then by the minimality of $V_0$ both $V_1$ and $V_2$ may be written as finite unions of irreducible algebraic sets. Then $V_0$ may be written as a finite union of irreducible algebraic sets, a contradiction. This proves $\mathcal{S} = \emptyset$, i.e., every affine algebraic set has a decomposition into affine varieties.

To prove uniqueness, suppose $V$ has two decompositions into affine varieties (where redundant terms have been removed from each decomposition):

$$V = V_1 \cup V_2 \cup \cdots \cup V_r = U_1 \cup U_2 \cup \cdots \cup U_s.$$

Then $V_1$ is contained in the union of the $U_i$. Since $V_1 \cap U_i$ is an algebraic set for each $i$, we obtain a decomposition of $V_1$ into algebraic subsets:

$$V_1 = (V_1 \cap U_1) \cup (V_1 \cap U_2) \cup \cdots \cup (V_1 \cap U_s).$$

Since $V_1$ is irreducible, we must have $V_1 = V_1 \cap U_j$ for some $j$, i.e., $V_1 \subseteq U_j$. By the symmetric argument we have $U_j \subseteq V_{j'}$ for some $j'$. Thus $V_1 \subseteq V_{j'}$, so $j' = 1$ and $V_1 = U_j$. Applying a similar argument for each $V_i$ it follows that $r = s$ and that $\{V_1, \ldots, V_r\} = \{U_1, \ldots, U_s\}$. This completes the proof.

**Corollary 18.** An affine algebraic set $V$ is a variety if and only if its coordinate ring $k[V]$ is an integral domain.

*Proof:* This follows immediately since $\mathcal{I}(V)$ is a prime ideal if and only if the quotient $k[V] = k[\mathbb{A}^n]/\mathcal{I}(V)$ is an integral domain (Proposition 13 of Chapter 7).

**Definition.** If $V$ is a variety, then the field of fractions of the integral domain $k[V]$ is called the field of *rational functions* on $V$ and is denoted by $k(V)$. The *dimension* of a variety $V$, denoted dim $V$, is defined to be the transcendence degree of $k(V)$ over $k$.

**Examples**

**(1)** Single points in $\mathbb{A}^n$ are affine varieties since their corresponding ideals in $k[\mathbb{A}^n]$ are maximal ideals. The coordinate ring of a point is isomorphic to $k$, which is also the field of rational functions. The dimension of a single point is 0. Any finite set is the union of its single point subsets, and this is its unique decomposition into affine subvarieties.

**(2)** The $x$-axis in $\mathbb{R}^2$ is irreducible since it has coordinate ring $\mathbb{R}[x, y]/(y) \cong \mathbb{R}[x]$, which is an integral domain. Similarly, the $y$-axis and, more generally, lines in $\mathbb{R}^2$ are also irreducible (cf. Exercise 23 in Section 1). Linear sets in $\mathbb{R}^n$ are affine varieties. The field of rational functions on the $x$-axis is the quotient field $\mathbb{R}(x)$ of $\mathbb{R}[x]$, which is why $\mathbb{R}(x)$ is called a rational function field. The dimension of the $x$-axis (or, more generally, any line) is 1.

**(3)** The union of the $x$ and $y$ axes in $\mathbb{R}^2$, namely $\mathcal{Z}(xy)$, is not a variety: $\mathcal{Z}(xy) = \mathcal{Z}(x) \cup \mathcal{Z}(y)$ is its unique decomposition into subvarieties. The corresponding coordinate ring $\mathbb{R}[x, y]/(xy)$ contains zero divisors.

**(4)** The hyperbola $xy = 1$ in $\mathbb{R}^2$ is a variety since we saw in Section 1 that its coordinate ring is the integral domain $\mathbb{R}[x, 1/x]$. Note that the two disjoint branches of the hyperbola (defined by $x > 0$ and $x < 0$) are not subvarieties (cf. also Exercises 12–13).

**(5)** If $V = \mathcal{Z}(l_1, l_2, \ldots, l_m)$ is the zero set of *linear* polynomials $l_1, \ldots, l_m$ in $k[x_1, \ldots, x_m]$ and $V \neq \emptyset$, then $V$ is an affine variety (called a *linear variety*). Note that determining whether $V \neq \emptyset$ is a linear algebra problem.

We end this section with some general ring-theoretic results that were originally motivated by their connection with decomposition questions in geometry.

## Primary Decomposition of Ideals in Noetherian Rings

The second statement in Proposition 17 shows that any ideal of the form $\mathcal{I}(V)$ in $k[\mathbb{A}^n]$ may be written uniquely as a finite intersection of prime ideals, and by Hilbert's Nullstellensatz this applies in particular to all radical ideals when $k$ is algebraically closed. In a large class of commutative rings (including all Noetherian rings) every ideal has a *primary decomposition*, which is a similar decomposition but allows ideals that are analogous to "prime powers" (but see the examples below). This decomposition can be considered as a generalization of the factorization of an integer $n \in \mathbb{Z}$ into the product of prime powers. We shall be primarily concerned with the case of Noetherian rings.

**Definition.** A proper ideal $Q$ in the commutative ring $R$ is called *primary* if whenever $ab \in Q$ and $a \notin Q$, then $b^n \in Q$ for some positive integer $n$. Equivalently, if $ab \in Q$ and $a \notin Q$, then $b \in \text{rad } Q$.

Some of the basic properties of primary ideals are given in the following proposition.

**Proposition 19.** Let $R$ be a commutative ring with 1.

**(1)** Prime ideals are primary.

**(2)** The ideal $Q$ is primary if and only if every zero divisor in $R/Q$ is nilpotent.

**(3)** If $Q$ is primary then rad $Q$ is a prime ideal, and is the unique smallest prime ideal containing $Q$.

**(4)** If $Q$ is an ideal whose radical is a maximal ideal, then $Q$ is a primary ideal.

**(5)** Suppose $M$ is a maximal ideal and $Q$ is an ideal with $M^n \subseteq Q \subseteq M$ for some $n \geq 1$. Then $Q$ is a primary ideal with rad $Q = M$.

*Proof:* The first two statements are immediate from the definition of a primary ideal. For (3), suppose $ab \in$ rad $Q$. Then $a^m b^m = (ab)^m \in Q$, and since $Q$ is primary, either $a^m \in Q$, in which case $a \in$ rad $Q$, or $(b^m)^n \in Q$ for some positive integer $n$, in which case $b \in$ rad $Q$. This proves that rad $Q$ is a prime ideal, and it follows that rad $Q$ is the smallest prime ideal containing $Q$ (Proposition 12).

To prove (4) we pass to the quotient ring $R/Q$; by (2), it suffices to show that every zero divisor in this quotient ring is nilpotent. We are reduced to the situation where $Q = (0)$ and $M =$ rad $Q =$ rad(0), which is the nilradical, is a maximal ideal. Since the nilradical is contained in every prime ideal (Proposition 12), it follows that $M$ is the unique prime ideal, so also the unique maximal ideal. If $d$ were a zero divisor, then the ideal $(d)$ would be a proper ideal, hence contained in a maximal ideal. This implies that $d \in M$, hence every zero divisor is indeed nilpotent.

Finally, suppose $M^n \subseteq Q \subseteq M$ for some $n \geq 1$ where $M$ is a maximal ideal. Then $Q \subseteq M$ so rad $Q \subseteq$ rad $M = M$. Conversely, $M^n \subseteq Q$ shows that $M \subseteq$ rad $Q$, so rad $Q = M$ is a maximal ideal, and $Q$ is primary by (4).

**Definition.** If $Q$ is a primary ideal, then the prime ideal $P =$ rad $Q$ is called the *associated prime* to $Q$, and $Q$ is said to *belong* to $P$ (or to be *$P$-primary*).

It is easy to check that a finite intersection of $P$-primary ideals is again a $P$-primary ideal (cf. the exercises).

**Examples**

**(1)** The primary ideals in $\mathbb{Z}$ are 0 and the ideals $(p^m)$ for $p$ a prime and $m \geq 1$.

**(2)** For any field $k$, the ideal $(x)$ in $k[x, y]$ is primary since it is a prime ideal. For any $n \geq 1$, the ideal $(x, y)^n$ is primary since it is a power of the maximal ideal $(x, y)$.

**(3)** The ideal $Q = (x^2, y)$ in the polynomial ring $k[x, y]$ is primary since we have $(x, y)^2 \subseteq (x^2, y) \subseteq (x, y)$. Similarly, $Q' = (4, x)$ in $\mathbb{Z}[x]$ is a $(2, x)$-primary ideal.

**(4)** Primary ideals need not be powers of prime ideals. For example, the primary ideal $Q$ in the previous example is not the power of a prime ideal, as follows. If $(x^2, y) = P^k$ for some prime ideal $P$ and some $k \geq 1$, then $x^2, y \in P^k \subseteq P$ so $x, y \in P$. Then $P = (x, y)$, and since $y \notin (x, y)^2$, it would follow that $k = 1$ and $Q = (x, y)$. Since $x \notin (x^2, y)$, this is impossible.

**(5)** If $R$ is Noetherian, and $Q$ is a primary ideal belonging to the prime ideal $P$, then

$$P^m \subseteq Q \subseteq P$$

for some $m \geq 1$ by Proposition 14. If $P$ is a maximal ideal, then the last statement in Proposition 19 shows that the converse also holds. This is not necessarily true if $P$

is a prime ideal that is *not maximal*. For example, consider the ideal $I = (x^2, xy)$ in $k[x, y]$. Then $(x^2) \subset I \subset (x)$, and $(x)$ is a prime ideal, but $I$ is not primary: $xy \in I$ and $x \notin I$, but no positive power of $y$ is an element of $I$. This example also shows that an ideal whose radical is prime (but not maximal as in (4) of the proposition) is not necessarily primary.

**(6)** Powers of prime ideals need not be primary. For example, consider the quotient ring $R = \mathbb{R}[x, y, z]/(xy - z^2)$, the coordinate ring of the cone $z^2 = xy$ in $\mathbb{R}^3$, and let $P = (\bar{x}, \bar{z})$ be the ideal generated by $\bar{x}$ and $\bar{z}$ in $R$. This is a prime ideal in $R$ since the quotient is $R/(\bar{x}, \bar{z}) \cong \mathbb{R}[x, y, z]/(x, z) \cong \mathbb{R}[y]$ (because $(xy - z^2) \subset (x, z)$). The ideal

$$P^2 = (\bar{x}^2, \bar{x}\bar{z}, \bar{z}^2) = (\bar{x}^2, \bar{x}\bar{z}, \bar{x}\bar{y}) = \bar{x}(\bar{x}, \bar{y}, \bar{z}),$$

however, is not primary: $\bar{x}\bar{y} = \bar{z}^2 \in P^2$, but $\bar{x} \notin P^2$, and no power of $\bar{y}$ is in $P^2$. Note that $P^2$ is another example of an ideal that is not primary whose radical is prime.

**(7)** Suppose $R$ is a U.F.D. If $\pi$ is an irreducible element of $R$ then it is easy to see that the powers $(\pi^n)$ for $n = 1, 2, \ldots$ are $(\pi)$-primary ideals. Conversely, suppose $Q$ is a $(\pi)$-primary ideal, and let $n$ be the largest integer with $Q \subseteq (\pi^n)$ (such an integer exists since, for example, $\pi^k \in Q$ for some $k \geq 1$, so $n \leq k$). If $q$ is an element of $Q$ not contained in $(\pi^{n+1})$, then $q = r\pi^n$ for some $r \in R$ and $r \notin (\pi)$. Since $r \notin (\pi)$ and $Q$ is $(\pi)$-primary, it follows that $\pi^n \in Q$. This shows that $Q = (\pi^n)$.

In the examples above, the ideal $(x^2, xy)$ in $k[x, y]$ is not a primary ideal, but it can be written as the intersection of primary ideals: $(x^2, xy) = (x) \cap (x, y)^2$.

**Definition.**

**(1)** An ideal $I$ in $R$ has a *primary decomposition* if it may be written as a finite intersection of primary ideals:

$$I = \bigcap_{i=1}^{m} Q_i \qquad Q_i \text{ a primary ideal.}$$

**(2)** The primary decomposition above is *minimal* and the $Q_i$ are called the *primary components of* $I$ if

- **(a)** no primary ideal contains the intersection of the remaining primary ideals, i.e., $Q_i \not\supseteq \cap_{j \neq i} Q_j$ for all $i$, and
- **(b)** the associated prime ideals are all distinct: $\operatorname{rad} Q_i \neq \operatorname{rad} Q_j$ for $i \neq j$.

We now prove that in a Noetherian ring every proper ideal has a minimal primary decomposition. This result is often called the Lasker–Noether Decomposition Theorem, since it was first proved for polynomial rings by the chess master Emanuel Lasker and the proof was later greatly simplified and generalized by Emmy Noether.

**Definition.** A proper ideal $I$ in the commutative ring $R$ is said to be *irreducible* if $I$ cannot be written nontrivially as the intersection of two other ideals, i.e., if $I = J \cap K$ with ideals $J$, $K$ implies that $I = J$ or $I = K$.

It is easy to see that a prime ideal is irreducible (see Exercise 11 in Section 7.4). The ideal $(x, y)^2$ in $k[x, y]$ in Example 2 earlier shows that primary ideals need not

be irreducible since it is the intersection of the ideals $(x) + (x, y)^2 = (x, y^2)$ and $(y)+(x, y)^2 = (y, x^2)$. In a Noetherian ring, however, irreducible ideals are necessarily primary:

**Proposition 20.** Let $R$ be a Noetherian ring. Then

**(1)** every irreducible ideal is primary, and

**(2)** every proper ideal in $R$ is a finite intersection of irreducible ideals.

*Proof:* To prove (1) let $Q$ be an irreducible ideal and suppose that $ab \in Q$ and $b \notin Q$. It is easy to check that for any fixed $n$ the set of elements $x \in R$ with $a^n x \in Q$ is an ideal, $A_n$, in $R$. Clearly $A_1 \subseteq A_2 \subseteq \ldots$ and since $R$ is Noetherian this ascending chain of ideals must stabilize, i.e., $A_n = A_{n+1} = \ldots$ for some $n > 0$. Consider the two ideals $I = (a^n) + Q$ and $J = (b) + Q$ of $R$, each containing $Q$. If $y \in I \cap J$ then $y = a^n z + q$ for some $z \in R$ and $q \in Q$. Since $ab \in Q$, it follows that $aJ \subseteq Q$, and in particular $ay \in Q$. Then $a^{n+1}z = ay - aq \in Q$, so $z \in A_{n+1} = A_n$. But $z \in A_n$ means that $a^n z \in Q$, so $y \in Q$. It follows that $I \cap J = Q$. Since $Q$ is irreducible and $(b) + Q \neq Q$ (since $b \notin Q$), we must have $a^n \in Q$, which shows that $Q$ is primary.

The proof of (2) is the same as the proof of the second statement in Proposition 17. Let $\mathcal{S}$ be the collection of ideals of $R$ that cannot be written as a finite intersection of irreducible ideals. If $\mathcal{S}$ is not empty, then since $R$ is Noetherian, there is a maximal element $I$ in $\mathcal{S}$. Then $I$ is not itself irreducible, so $I = J \cap K$ for some ideals $J$ and $K$ distinct from $I$. Then $I \subset J$ and $I \subset K$ and the maximality of $I$ implies that neither $J$ nor $K$ is in $\mathcal{S}$. But this means that both $J$ and $K$ can be written as finite intersections of irreducible ideals, hence the same would be true for $I$. This is a contradiction, so $\mathcal{S} = \emptyset$, which completes the proof of the proposition.

It is immediate from the previous proposition that in a Noetherian ring every proper ideal has a primary decomposition. If any of the primary ideals in this decomposition contains the intersection of the remaining primary ideals, then we may simply remove this ideal since this will not change the intersection. Hence we may assume the decomposition satisfies (a) in the definition of a minimal decomposition. Since a finite intersection of $P$-primary ideals is again $P$-primary (Exercise 31), replacing the primary ideals in the decomposition with the intersections of all those primary ideals belonging to the same prime, we may also assume the decomposition satisfies (b) in the definition of a minimal decomposition. This proves the first statement of the following:

**Theorem 21.** *(Primary Decomposition Theorem)* Let $R$ be a Noetherian ring. Then every proper ideal $I$ in $R$ has a minimal primary decomposition. If

$$I = \bigcap_{i=1}^{m} Q_i = \bigcap_{i=1}^{n} Q_i'$$

are two minimal primary decompositions for $I$ then the sets of associated primes in the two decompositions are the same:

$$\{\operatorname{rad} Q_1, \operatorname{rad} Q_2, \ldots, \operatorname{rad} Q_m\} = \{\operatorname{rad} Q_1', \operatorname{rad} Q_2', \ldots, \operatorname{rad} Q_n'\}.$$

Moreover, the primary components $Q_i$ belonging to the minimal elements in this set of associated primes are uniquely determined by $I$.

*Proof:* The proof of the uniqueness of the set of associated primes is outlined in the exercises, and the proof of the uniqueness of the primary components associated to the minimal primes will be given in Section 4.

**Definition.** If $I$ is an ideal in the Noetherian ring $R$ then the associated prime ideals in any primary decomposition of $I$ are called the *associated prime ideals of* $I$. If an associated prime ideal $P$ of $I$ does not contain any other associated prime ideal of $I$ then $P$ is called an *isolated prime ideal*; the remaining associated prime ideals of $I$ are called *embedded prime ideals*.

The prime ideals associated to an ideal $I$ provide a great deal of information about the ideal $I$ (cf. for example Exercises 41 and 43):

**Corollary 22.** Let $I$ be a proper ideal in the Noetherian ring $R$.

**(1)** A prime ideal $P$ contains the ideal $I$ if and only if $P$ contains one of the associated primes of $I$, hence if and only if $P$ contains one of the isolated primes of $I$, i.e., the isolated primes of $I$ are precisely the minimal elements in the set of all prime ideals containing $I$. In particular, there are only finitely many minimal elements among the prime ideals containing $I$.

**(2)** The radical of $I$ is the intersection of the associated primes of $I$, hence also the intersection of the isolated primes of $I$.

**(3)** There are prime ideals $P_1, \ldots, P_n$ (not necessarily distinct) containing $I$ such that $P_1 P_2 \cdots P_n \subseteq I$.

*Proof:* The first statement in (1) is an exercise (cf. Exercise 37), and the remainder of (1) follows. Then (2) follows from (1) and Proposition 12, and (3) follows from (2) and Proposition 14.

The last statement in Theorem 21 states that not only the isolated primes, but also the primary components belonging to the isolated primes, are uniquely determined by $I$. In general the primary decomposition of an ideal $I$ is itself not unique.

**Examples**

**(1)** Let $I = (x^2, xy)$ in $\mathbb{R}[x, y]$. Then

$$(x^2, xy) = (x) \cap (x, y)^2 = (x) \cap (x^2, y)$$

are two minimal primary decompositions for $I$. The associated primes for $I$ are $(x)$ and $\text{rad}((x, y)^2) = \text{rad}((x^2, y)) = (x, y)$. The prime $(x)$ is the only isolated prime since $(x) \subset (x, y)$, and $(x, y)$ is an embedded prime. A prime ideal $P$ contains $I$ if and only if $P$ contains $(x)$. The $(x)$-primary component of $I$ corresponding to this isolated prime is just $(x)$ and occurs in both primary decompositions; the $(x, y)$-primary component of $I$ corresponding to this embedded prime is not uniquely determined — it is $(x, y)^2$ in the first decomposition and is $(x^2, y)$ in the second. The radical of $I$ is the isolated prime $(x)$.

This example illustrates the origin of the terminology: in general the irreducible components of the algebraic space $\mathcal{Z}(I)$ defined by $I$ are the zero sets of the isolated primes for $I$, and the zero sets of the embedded primes are irreducible subspaces of

these components (so are "embedded" in the irreducible components). In this example, $\mathcal{Z}(I)$ is the set of points with $x^2 = xy = 0$, which is just the $y$-axis in $\mathbb{R}^2$. There is only one irreducible component of this algebraic space (namely the $y$-axis), which is the locus for the isolated prime $(x)$. The locus for the embedded prime $(x, y)$ is the origin $(0, 0)$, which is an irreducible subspace embedded in the $y$-axis.

**(2)** Suppose $R$ is a U.F.D. If $a = {p_1}^{e_1} \cdots {p_m}^{e_m}$ is the unique factorization into distinct prime powers of the element $a \in R$, then $(a) = (p_1)^{e_1} \cap \cdots \cap (p_m)^{e_m}$ is the minimal primary decomposition of the principal ideal $(a)$. The associated primes to $(a)$ are $(p_1), \ldots, (p_m)$ and are all isolated. The primary decomposition of ideals is a generalization of the factorization of elements into prime powers. See also Exercise 44 for a characterization of U.F.D.s in terms of minimal primary decompositions.

For any Noetherian ring, an ideal $I$ is radical if and only if the primary components of a minimal primary decomposition of $I$ are all *prime* ideals (in which case this primary decomposition is unique), cf. Exercise 43. This generalizes the observation made previously that Proposition 17 together with Hilbert's Nullstellensatz shows that any radical ideal in $k[\mathbb{A}^n]$ may be written uniquely as a finite intersection of prime ideals when the field $k$ is algebraically closed — this is the algebraic statement that an algebraic set can be decomposed uniquely into the union of irreducible algebraic sets.

## EXERCISES

1. Prove (3) of Corollary 22 directly by considering the collection $\mathcal{S}$ of ideals that do not contain a finite product of prime ideals. [If $I$ is a maximal element in $\mathcal{S}$, show that since $I$ is not prime there are ideals $J$, $K$ properly containing $I$ (hence not in $\mathcal{S}$) with $JK \subseteq I$.]
2. Let $I$ and $J$ be ideals in the ring $R$. Prove the following statements:
   **(a)** If $I^k \subseteq J$ for some $k \geq 1$ then $\operatorname{rad} I \subseteq \operatorname{rad} J$.
   **(b)** If $I^k \subseteq J \subseteq I$ for some $k \geq 1$ then $\operatorname{rad} I = \operatorname{rad} J$.
   **(c)** $\operatorname{rad}(IJ) = \operatorname{rad}(I \cap J) = \operatorname{rad} I \cap \operatorname{rad} J$.
   **(d)** $\operatorname{rad}(\operatorname{rad} I) = \operatorname{rad} I$.
   **(e)** $\operatorname{rad} I + \operatorname{rad} J \subseteq \operatorname{rad}(I + J)$ and $\operatorname{rad}(I + J) = \operatorname{rad}(\operatorname{rad} I + \operatorname{rad} J)$.
3. Prove that the intersection of two radical ideals is again a radical ideal.
4. Let $I = \mathfrak{m}_1\mathfrak{m}_2$ be the product of the ideals $\mathfrak{m}_1 = (x, y)$ and $\mathfrak{m}_2 = (x-1, y-1)$ in $\mathbb{F}_2[x, y]$. Prove that $I$ is a radical ideal. Prove that the ideal $(x^3 - y^2)$ is a radical ideal in $\mathbb{F}_2[x, y]$.
5. If $I = (xy, (x - y)z) \subset k[x, y, z]$ prove that $\operatorname{rad} I = (xy, xz, yz)$. For this ideal prove directly that $\mathcal{Z}(I) = \mathcal{Z}(\operatorname{rad} I)$, that $\mathcal{Z}(I)$ is not irreducible, and that $\operatorname{rad} I$ is not prime.
6. Give an example to show that over a field $k$ that is not algebraically closed the containment $I \subseteq \mathcal{I}(\mathcal{Z}(I))$ can be proper even when $I$ is a radical ideal.
7. Suppose $R$ and $S$ are rings and $\varphi : R \to S$ is a ring homomorphism. If $I$ is an ideal of $R$ show that $\varphi(\operatorname{rad} I) \subseteq \operatorname{rad}(\varphi(I))$. If in addition $\varphi$ is surjective and $I$ contains the kernel of $\varphi$ show that $\varphi(\operatorname{rad} I) = \operatorname{rad}(\varphi(I))$.
8. Suppose the prime ideal $P$ contains the ideal $I$. Prove that $P$ contains the radical of $I$.
9. Prove that for any field $k$ the map $\mathcal{Z}$ in the Nullstellensatz is always surjective and the map $\mathcal{I}$ in the Nullstellensatz is always injective. [Use property (10) of the maps $\mathcal{Z}$ and $\mathcal{I}$ in Section 1.] Give examples (over a field $k$ that is not algebraically closed) where $\mathcal{Z}$ is not injective and $\mathcal{I}$ is not surjective.

**10.** Prove that for $k$ a finite field the Zariski topology is the same as the discrete topology: every subset is closed (and open).

**11.** Let $V$ be a variety in $\mathbb{A}^n$ and let $U_1$ and $U_2$ be two subsets of $\mathbb{A}^n$ that are open in the Zariski topology. Prove that if $V \cap U_1 \neq \emptyset$ and $V \cap U_2 \neq \emptyset$ then $V \cap U_1 \cap U_2 \neq \emptyset$. Conclude that *any* nonempty open subset of a variety is *everywhere dense* in the Zariski topology (i.e., its closure is all of $V$).

**12.** Use the fact that nonempty open sets of an affine variety are everywhere dense to prove that an affine variety is connected in the Zariski topology. (A topological space is *connected* if it is not the union of two disjoint, proper, open subsets.)

**13.** Prove that the affine algebraic set $V$ is connected in the Zariski topology if and only if $k[V]$ is not a direct sum of two nonzero ideals. Deduce from this that a variety is connected in the Zariski topology.

**14.** Prove that if $k$ is an infinite field, then the varieties in $\mathbb{A}^1$ are the empty set, the whole space, and the one point subsets. What are the varieties in $\mathbb{A}^1$ in the case of a finite field $k$?

**15.** Suppose $V$ is a hypersurface in $\mathbb{A}^n$ and $\mathcal{I}(V) = (f)$ for some nonconstant polynomial $f \in k[x_1, x_2, \ldots, x_n]$. Prove that $V$ is a variety if and only if $f$ is irreducible.

**16.** Suppose $V \subseteq \mathbb{A}^n$ is an affine variety and $f \in k[V]$. Prove that the *graph* of $f$ (cf. Exercise 25 in Section 1) is an affine variety.

**17.** Prove that any permutation of the elements of a field $k$ is a continuous map from $\mathbb{A}^1$ to itself in the Zariski topology on $\mathbb{A}^1$. Deduce that if $k$ is an infinite field, there are Zariski continuous maps from $\mathbb{A}^1$ to itself that are not polynomials.

**18.** Let $V$ be an affine algebraic set in $\mathbb{A}^n$ over $k = \mathbb{C}$.

- **(a)** Prove that morphisms of algebraic sets over $\mathbb{C}$ are continuous in the Euclidean topology (the topology on $\mathbb{C}^n$ obtained by identifying $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ with its usual Euclidean topology).
- **(b)** Prove that $V$ is a closed set in the Euclidean topology on $\mathbb{C}^n$ (so the Zariski closed sets of $\mathbb{A}^n$ over $\mathbb{C}$ are also Euclidean closed).
- **(c)** Give an example of a set that is closed in the Euclidean topology but is not closed in the Zariski topology, i.e., is not an affine algebraic set (so the Euclidean topology is "finer" than the Zariski topology).

**19.** Give an example of an injective $k$-algebra homomorphism $\tilde{\varphi} : k[W] \to k[V]$ whose associated morphism $\varphi : V \to W$ is not surjective.

**20.** Suppose $\varphi : V \to W$ is a surjective morphism of affine algebraic sets. Prove that if $V$ is a variety then $W$ is a variety.

**21.** Let $V$ be an algebraic set in $\mathbb{A}^n$ and let $f \in k[V]$. Define $V_f = \{v \in V \mid f(v) \neq 0\}$.

- **(a)** Show that $V_f$ is a Zariski open set in $V$ (called a *principal open set* in $V$).
- **(b)** Let $J$ be the ideal in $k[x_1, \ldots, x_n, x_{n+1}]$ generated by $\mathcal{I}(V)$ and $x_{n+1} f - 1$, and let $W = \mathcal{Z}(J) \subseteq \mathbb{A}^{n+1}$. Show that $J = \mathcal{I}(W)$ and that the map $\pi : \mathbb{A}^{n+1} \to \mathbb{A}^n$ by projection onto the first $n$ coordinates is a Zariski continuous bijection from $W$ onto $V_f$ (so the principal *open* set $V_f$ in $V$ may be embedded as a *closed* set in some (larger) affine space).
- **(c)** If $U$ is any open set in $V$ show that $U = V_{f_1} \cup \cdots \cup V_{f_m}$ for some $f_1, \ldots, f_m \in k[V]$. (This shows that the principal open sets form a *base* for the Zariski topology.)

**22.** Prove that $GL_n(k)$ is an open affine algebraic set in $\mathbb{A}^{n^2}$ and can be embedded as a closed affine algebraic set in $\mathbb{A}^{n^2+1}$. In particular, deduce that the set $k^\times$ of nonzero elements in

$\mathbb{A}^1$ embeds into $\mathbb{A}^2$ as the hyperbola $xy = 1$. [Use the preceding exercise.]

**23.** Show that if $k$ is infinite then $\{(a, a^2, a^3) \mid a \in k\} \subset \mathbb{A}^3$ is an affine algebraic variety. If $k$ is finite show that this set is always reducible.

**24.** Let $V = \mathcal{Z}(xz - y^2, yz - x^3, z^2 - x^2y) \subset \mathbb{A}^3$. Show that if $k$ is infinite then $V$ is an affine variety. [Use Exercise 26 of Section 1 and Exercise 20.]

**25.** Suppose $f(x) = x^3 + ax^2 + bx + c$ is an irreducible cubic in $\mathbb{Q}[x]$ of discriminant $D$. Let $I = (x + y + z + a, xy + xz + yz - b, xyz + c)$ in $\mathbb{Q}[x, y, z]$.

**(a)** Prove that $I$ is a prime ideal if and only if $D$ is not a square in $\mathbb{Q}$, in which case $I$ is a maximal ideal and $\mathbb{Q}[x, y, z]/I$ is a splitting field for $f(x)$ over $\mathbb{Q}$.

**(b)** If $D = r^2$, prove that the primary decomposition of $I$ is $I = Q_+ \cap Q_-$ where $Q_\pm = (I, (x - y)(x - z)(y - z) \pm r)$. Prove $Q_+$ and $Q_-$ are maximal ideals, and $\mathbb{Q}[x, y, z]$ modulo $Q_+$ or $Q_-$ is a splitting field for $f(x)$ over $\mathbb{Q}$.

**26.** A topological space $X$ is called *quasicompact* if whenever any collection of closed subsets $V_i$ of $X$ has empty intersection, then some finite number of these also has empty intersection, i.e.,

$$\text{whenever } \bigcap_i V_i = \emptyset \text{ there exists } V_{i_1}, V_{i_2}, \ldots, V_{i_N} \text{ such that } \bigcap_{t=1}^{N} V_{i_t} = \emptyset.$$

Prove that every affine algebraic set is quasicompact. [Translate the definition into a property of ideals in $k[x_1, \ldots, x_n]$.] (A quasicompact and Hausdorff space is called *compact*.)

**27.** When $k$ is an infinite field prove that the Zariski topology on $k^2$ is not the same as taking the Zariski topology on $k$ and then forming the product topology on $k \times k$. [By Exercise 14 of Section 1, in the product topology on $k \times k$ the Zariski closed sets in $k \times k$ are finite unions of sets of the form $\{a\} \times \{b\}$, $\{a\} \times k$ and $k \times \{b\}$, for any $a, b \in k$.]

**28.** Prove that each of the following rings have infinitely many minimal prime ideals, and that (0) is not the intersection of any finite number of these (so (0) does not have a primary decomposition in these rings):

**(a)** the infinite direct product ring $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \times \cdots$ (which is a Boolean ring, cf. Exercise 23 in Section 7.4).

**(b)** $k[x_1, x_2, \ldots]/(x_1x_2, x_3x_4, \ldots, x_{2i-1}x_{2i}, \ldots)$, where $x_1, x_2, \ldots$ are independent variables over the field $k$.

**29.** Suppose that $A$ and $B$ are ideals with $AB \subseteq Q$ for a primary ideal $Q$. Prove that if $A \not\subseteq Q$ then $B \subset \operatorname{rad} Q$.

**30.** Let $Q$ be a $P$-primary ideal and suppose $A$ is an ideal not contained in $Q$. Define $A' = \{r \in R \mid rA \subseteq Q\}$ to be the elements of $R$ that when multiplied by elements of $A$ give elements of $Q$. Prove that $A'$ is a $P$-primary ideal.

**31.** Prove that if $Q_1$ and $Q_2$ are primary ideals belonging to the same prime ideal $P$, then $Q_1 \cap Q_2$ is a primary ideal belonging to $P$. Conclude that a finite intersection of $P$-primary ideals is again $P$-primary.

**32.** Prove that if $Q_1$ and $Q_2$ are primary ideals belonging to the same *maximal* ideal $M$, then $Q_1 + Q_2$ and $Q_1Q_2$ are primary ideals belonging to $M$. Conclude that finite sums and finite products of $M$-primary ideals are again $M$-primary.

**33.** Let $I = (x^2, xy, xz, yz)$ in $k[x, y, z]$. Prove that a primary decomposition of $I$ is $I = (x, y) \cap (x, z) \cap (x, y, z)^2$, determine the isolated and embedded primes of $I$, and find $\operatorname{rad} I$.

**34.** Suppose $\varphi : R \to S$ is a surjective ring homomorphism. Prove that an ideal $Q$ in $R$ containing the kernel of $\varphi$ is primary if and only if $\varphi(Q)$ is primary in $S$, and when this is

the case the prime associated to $\varphi(Q)$ is the image $\varphi(P)$ of the prime $P$ associated to $Q$.

**35.** Suppose $\varphi : R \to S$ is a ring homomorphism.

**(a)** Suppose $I$ is an ideal of $R$ containing $\ker \varphi$ with minimal primary decomposition $I = Q_1 \cap \cdots \cap Q_m$ with $\operatorname{rad} Q_i = P_i$. If $\varphi$ is a surjective homomorphism prove that $\varphi(I) = \varphi(Q_1) \cap \cdots \cap \varphi(Q_m)$, where $\operatorname{rad} \varphi(Q_i)$ is given by $\varphi(P_i)$, is a minimal primary decomposition of $\varphi(I)$. [Use the previous exercise.]

**(b)** Suppose $I$ is an ideal of $S$ with minimal primary decomposition $I = Q_1 \cap \cdots \cap Q_m$ with $\operatorname{rad} Q_i = P_i$. Prove that $\varphi^{-1}(I) = \varphi^{-1}(Q_1) \cap \cdots \cap \varphi^{-1}(Q_m)$, where $\operatorname{rad} \varphi^{-1}(Q_i)$ is given by $\varphi^{-1}(P_i)$, is a primary decomposition of $\varphi^{-1}(I)$, and is minimal if $\varphi$ is surjective.

**36.** Let $I = (xy, x - yz)$ in $k[x, y, z]$. Prove that $(x, z) \cap (y^2, x - yz)$ is a minimal primary decomposition of $I$. [Consider the ring homomorphism $\varphi : k[x, y, z] \to k[y, z]$ given by mapping $x$ to $yz$, $y$ to $y$, and $z$ to $z$ and use the previous exercise.]

**37.** Prove that a prime ideal $P$ contains the ideal $I$ if and only if $P$ contains one of the associated primes of a minimal primary decomposition of $I$. [Use Exercise 3 and Exercise 11 in Section 7.4.]

**38.** Show that every associated prime ideal for a radical ideal is isolated. [Suppose that $P_2 = \operatorname{rad} Q_2 \subseteq P_1 = \operatorname{rad} Q_1$ in the decomposition of Theorem 21 for the radical ideal $I$. Show that if $a \in Q_2 \cap \cdots \cap Q_m \subseteq P_2$ then $a^n \in I$ for some $n \geq 1$, conclude that $a \in Q_1$ and derive a contradiction to the minimality of the primary decomposition.]

**39.** Fix an element $a$ in the ring $R$. For any ideal $I$ in the ring $R$ let $I_a = \{r \in R \mid ar \in I\}$.

**(a)** Prove that $I_a$ is an ideal and $I_a = R$ if and only if $a \in I$.

**(b)** Prove that $(I \cap J)_a = I_a \cap J_a$ for ideals $I$ and $J$.

**(c)** Suppose that $Q$ is a $P$-primary ideal and that $a \notin Q$. Prove that $Q_a$ is a $P$-primary ideal and that $Q_a = Q$ if $a \notin P$.

**40.** With notation as in the previous exercise, suppose $I = Q_1 \cap \cdots \cap Q_m$ is a minimal primary decomposition of the ideal $I$ and let $P_i$ be the prime ideal associated to $Q_i$.

**(a)** Prove that $I_a = (Q_1)_a \cap \cdots \cap (Q_m)_a$ and that $\operatorname{rad}(I_a) = \operatorname{rad}((Q_1)_a) \cap \cdots \cap \operatorname{rad}((Q_m)_a)$.

**(b)** Prove that $\operatorname{rad}(I_a)$ is the intersection of the prime ideals $P_i$ for which $a \notin Q_i$. [Use the previous exercise.]

**(c)** Prove that if $\operatorname{rad}(I_a)$ is a prime ideal then $\operatorname{rad}(I_a) = P_j$ for some $j$. [Use the fact that prime ideals are irreducible.]

**(d)** For each $i = 1, \ldots, m$, prove that $\operatorname{rad}(I_a) = P_i$ for some $a \in R$. [Show there exists an $a \in R$ with $a \notin Q_i$ but $a \in Q_j$ for all $j \neq i$.]

**(e)** Show from (c) and (d) that the associated primes for a minimal primary decomposition are precisely the collection of prime ideals among the ideals $\operatorname{rad}(I_a)$ for $a \in R$, and conclude that they are uniquely determined by $I$ independent of the minimal primary decomposition.

**41.** Let $P_1, \ldots, P_m$ be the associated prime ideals of the ideal (0) in the Noetherian ring $R$.

**(a)** Show that $P_1 \cap \cdots \cap P_m$ is the collection of nilpotent elements in $R$. [Apply Corollary 22 to $I = (0)$.]

**(b)** Show that $P_1 \cup \cdots \cup P_m$ is the collection of zero divisors in $R$. [Let $I = (0)$ in the previous exercise and show that the set of zero divisors is given by the set $\cup_{a \in R - \{0\}} (0)_a = \cup_{a \in R - \{0\}} \operatorname{rad}((0)_a)$.]

**42.** Suppose $R$ is a Noetherian ring. Prove that $R$ is either an integral domain, has nonzero nilpotent elements, or has at least two minimal prime ideals. [Use the previous exercise.]

**43.** Prove that the ideal $I$ in the Noetherian ring $R$ is radical if and only if the primary compo-

nents of a minimal primary decomposition are all prime ideals, and conclude that in this case the minimal primary decomposition is unique. [If $I = Q_1 \cap \cdots \cap Q_m$ is radical with $Q_i$ a $P_i$-primary component of a minimal decomposition, show that if $a \in P_1 \cap \cdots \cap P_m$ then some power of $a$ is in $I$, hence $a \in I$ since $I$ is radical. Deduce that $I = P_1 \cap \cdots \cap P_m$ and show that this is also a minimal primary decomposition, i.e., for any $i$ there exists $b$ with $b \notin P_i$, but $b \in P_j$ for $j \neq i$. If $a \in P_i$, show that $ab \in Q_i$, and that $a \in Q_i$. Conclude that $Q_i = P_i$.]

**44.** Prove that a Noetherian integral domain $R$ is a U.F.D. if and only if for every $a \in R$ the isolated primes associated to the principal ideal $(a)$ are principal ideals. [See Example 2 following Corollary 22. To prove $R$ is a U.F.D., show that an irreducible $a \in R$ is prime and then follow the proof of Theorem 14 in Section 8.3.]

**45.** Let $R$ be the ring of all real valued functions on the open interval $(-1, 1)$ that have derivatives of all orders (the ring of $C^\infty$ functions). Let

$$F(x) = \begin{cases} e^{-1/x^4} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

(you may assume $F \in R$ and $F^{(n)}(0) = 0$ for all $n \geq 0$). Let $(F)$ be the principal ideal generated by $F$ and let $A = \text{rad}((F))$. Let $M$ be the (maximal) ideal of all functions in $R$ that are zero at $x = 0$ and let $P = \cap_{n=1}^{\infty} M^n$.

**(a)** Prove that $M = (x)$ is the ideal generated by the function $x$ in $R$ and that $M^n = (x^n)$ consists of the functions whose first $n - 1$ derivatives vanish at the origin.

**(b)** Prove that $R$ is not Noetherian (compare Exercise 33 in Section 7.4). [One approach is the following: Let $G(x)$ be the function that is 0 for $x < 0$ and is equal to $F(x)$ for $x \geq 0$. Let $I_n$ be the ideal of functions in $R$ vanishing for all $x \leq 1/n$. Use translates of $G(x)$ to show that $I_1 \subset I_2 \subset I_3 \subset \cdots$ is an infinite ascending chain.]

**(c)** Prove that $P$ consists of the functions all of whose derivatives are zero at $x = 0$ (i.e., the functions whose associated Taylor series at $x = 0$ is identically zero), and that $P$ is a prime ideal.

**(d)** Prove that $F \in P$ and deduce that $A \subsetneq P$.

**(e)** Prove that $A \neq P$. [Let $G(x) = e^{-1/x^2}$ when $x \neq 0$ and $G(0) = 0$. Show that $G \in P$ but $G \notin A$.]

**(f)** Show that there is a prime ideal $Q$ containing $(F)$ with $Q \neq P, M$. Prove that $Q \subset P$ i.e., there are nonzero prime ideals properly contained in $P$.

**46.** Let $\mathcal{A}$ be any ideal in $R = k[x_1, \ldots, x_n, y_1, \ldots, y_m]$.

**(a)** Show that $\text{rad}(\mathcal{A} \cap k[y_1, \ldots, y_m]) = \text{rad}\, \mathcal{A} \cap k[y_1, \ldots, y_m]$.

**(b)** Suppose $(f_1, \ldots, f_s)$ is an ideal in $k[x_1, \ldots, x_n]$. Let $F_1, \ldots, F_t$ be generators for the radical of $(f_1, \ldots, f_s)$, computed in $k[x_1, \ldots, x_n]$. Suppose $J$ is an ideal in $R$ and let $\mathcal{A} = J + (f_1, \ldots, f_s)$, $\mathcal{B} = J + (F_1, \ldots, F_t)$ as ideals in $R$. Prove that $\text{rad}\, \mathcal{A} = \text{rad}\, \mathcal{B}$.

**(c)** Conclude from (a) and (b) that $\mathcal{A} = (y_1 - x_1, \ldots, y_m - x_m, f_1, \ldots, f_s) \cap k[y_1, \ldots, y_m]$ and $\mathcal{B} = (y_1 - x_1, \ldots, y_m - x_m, F_1, \ldots, F_t) \cap k[y_1, \ldots, y_m]$ have the same zero sets over an algebraically closed field $k$. [Use Hilbert's Nullstellensatz.]

**47.** Determine the Zariski closure in $\mathbb{C}^3$ of the points on the curve $\{(a^2, a^3, a^4) \mid a \in \mathbb{C}\}$.

**48.** Show that $\mathcal{Z}(x^3 - xyz + z^2)$ is the smallest algebraic set in $\mathbb{R}^3$ containing the points $\{(st, s + t, s^2t) \mid s, t \in \mathbb{R}\}$.

**49.** Show that $\mathcal{Z}(x^3z^2 - 3xy^2z^2 - y^6 - z^4)$ is the smallest algebraic set in $\mathbb{R}^3$ containing the points $\{(s^2 + t^2, st, s^3) \mid s, t \in \mathbb{R}\}$.

**50.** Find equations defining the Zariski closure of the set of points $\{(s^4, s^3t, st^3, t^4) \mid s, t \in \mathbb{R}\}$.

**51.** Show that $V = \mathcal{Z}(x^2 - y^2z)$ (the *Whitney umbrella surface*) is the smallest algebraic set in $\mathbb{R}^3$ containing the points $S = \{(st, s, t^2) \mid s, t \in \mathbb{R}\}$. Show that $S$ is not Zariski closed in $V$ (the missing points explain the name for the surface). Do the same over $\mathbb{C}$, but show that in this case $S = V$ is closed.

**52.** Let $V = \mathcal{Z}(xz^2 - w^3, xw^2 - y^4, y^4z^2 - w^5) \subset \mathbb{C}^4$. Determine the Zariski closure of the image of $V$ under the projection $\pi((x, y, z, w)) = (x, y, z)$.

**53.** Let $V = \mathcal{Z}(xy - 1)$ in $\mathbb{A}^2$ and let $S$ be the projection of $V$ onto the $x$-axis in $\mathbb{A}^1$.

**(a)** If $k = \mathbb{R}$, show that $\mathcal{I}(V) = (xy - 1) \subset \mathbb{R}[x, y]$ and that $(u - x, xy - 1) \cap \mathbb{R}[u] = 0$ in $\mathbb{R}[x, y, u]$. Use Propositions 8 and 16 to conclude that the Zariski closure of $S$ is $\mathbb{A}^1$ and show that $S$ is not itself closed.

**(b)** If $k = \mathbb{F}_3$, show that $\mathcal{I}(V) = (xy - 1, x^3 - x, y^3 - y) \subset \mathbb{F}_3[x, y]$ and that $(u - x, xy - 1, x^3 - x, y^3 - y) \cap \mathbb{F}_3[u] = (u^2 - 1)$ in $\mathbb{F}_3[x, y, u]$. Use Propositions 8 and 16 to conclude that $S$ is Zariski closed in $\mathbb{A}^1$.

**54.** Recall the *ideal quotient* $(I : J) = \{r \in R \mid rJ \in I\}$ of two ideals $I$, $J$ in a ring $R$ (cf. Exercise 34 *ff.* in Section 9.6). Clearly $I \subseteq (I : J)$.

**(a)** Show that $\mathcal{Z}(I) - \mathcal{Z}(J)$, the set of elements of $\mathcal{Z}(I)$ not lying in $\mathcal{Z}(J)$, is contained in $\mathcal{Z}((I : J))$ and conclude that the Zariski closure of $\mathcal{Z}(I) - \mathcal{Z}(J)$ is contained in $\mathcal{Z}((I : J))$.

**(b)** Show that if $k$ is algebraically closed and $I$ is a radical ideal then $\mathcal{Z}((I : J))$ is precisely the Zariski closure of $\mathcal{Z}(I) - \mathcal{Z}(J)$.

**(c)** Show that if $V$ and $W$ are affine algebraic sets then $(\mathcal{I}(V) : \mathcal{I}(W)) = \mathcal{I}(V - W)$.

## 15.3 INTEGRAL EXTENSIONS AND HILBERT'S NULLSTELLENSATZ

In this section we consider the important concept of an integral extension of rings, which is a generalization to rings of algebraic extensions of fields. This leads to the definition of the "integers" in finite extensions of $\mathbb{Q}$ (the basic subject of the branch of mathematics called algebraic number theory) and is also related to the existence of tangent lines for algebraic curves.

**Definition.** Suppose $R$ is a subring of the commutative ring $S$ with $1 = 1_S \in R$.

**(1)** An element $s \in S$ is *integral over* $R$ if $s$ is the root of a monic polynomial in $R[x]$.

**(2)** The ring $S$ is an *integral extension of* $R$ or just *integral over* $R$ if every $s \in S$ is integral over $R$.

**(3)** The *integral closure* of $R$ in $S$ is the set of elements of $S$ that are integral over $R$.

**(4)** The ring $R$ is said to be *integrally closed in* $S$ if $R$ is equal to its integral closure in $S$. The integral closure of an integral domain $R$ in its field of fractions is called the *normalization of* $R$. An integral domain is called *integrally closed* or *normal* if it is integrally closed in its field of fractions.

Before giving some examples of integral extensions we prove some basic properties of integral elements analogous to those of algebraic elements over fields.

**Proposition 23.** Let $R$ be a subring of the commutative ring $S$ with $1 \in R$ and let $s \in S$. Then the following are equivalent:

**(1)** $s$ is integral over $R$,

**(2)** $R[s]$ is a finitely generated $R$-module (where $R[s]$ is the ring of all $R$-linear combinations of powers of $s$), and

**(3)** $s \in T$ for some subring $T$, $R \subseteq T \subseteq S$, that is a finitely generated $R$-module.

*Proof:* Suppose first that (1) holds and let $s$ be a root of the monic polynomial $x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in R[x]$. Then

$$s^n = -(a_{n-1}s^{n-1} + a_{n-2}s^{n-2} + \cdots + a_0)$$

and so $s^n$, and then all higher powers of $s$, can be expressed as $R$-linear combinations of $s^{n-1}, \ldots, s, 1$. Hence $R[s] = R1 + Rs + \cdots + Rs^{n-1}$ is finitely generated as an $R$-module, which gives (2).

If (2) holds, then (3) holds with $T = R[s]$.

Suppose that (3) holds and let $v_1, v_2, \ldots, v_n$ be a finite generating set for $T$. Then for $i = 1, 2, \ldots, n$ the element $sv_i$ is an element of $T$ since $T$ is a ring, and so can be written as $R$-linear combinations of $v_1, \ldots, v_n$:

$$sv_i = \sum_{j=1}^{n} a_{ij}v_j,$$

i.e.,

$$0 = \sum_{j=1}^{n} (\delta_{ij}s - a_{ij})v_j \qquad i = 1, 2, \ldots, n$$

where $\delta_{ij}$ is the Kronecker delta. If $B$ is the $n \times n$ matrix whose $i, j$ entry is $\delta_{ij}s - a_{ij}$, and $v$ is the $n \times 1$ column vector whose entries are $v_1, \ldots, v_n$, then these equations are simply $Bv = 0$. It follows from Cramer's Rule that $(\det B)v_i = 0$ for all $i$ (cf. Exercise 3, Section 11.4). Since $1 \in T$ is an $R$-linear combination of $v_1, \ldots, v_n$, it follows that $\det B = 0$. But $B = sI - A$, where $A$ is the matrix $(a_{ij})$. Thus $s$ is a root of the monic polynomial $\det(xI - A) \in R[x]$ (the characteristic polynomial of $A$), and so $s$ is a root of a monic polynomial with coefficients in $R$, which gives (1), completing the proof.

**Corollary 24.** Let $R \subseteq S$ be as in Proposition 23 and let $s, t \in S$.

**(1)** If $s$ and $t$ are integral over $R$ then so are $s \pm t$ and $st$.

**(2)** The integral closure of $R$ in $S$ is a subring of $S$ containing $R$.

**(3)** Integrality is transitive: let $S$ be a subring of $T$; if $T$ is integral over $S$ and $S$ is integral over $R$, then $T$ is integral over $R$.

*Proof:* Let $s$ and $t$ be integral over $R$. By Proposition 23 both $R[s]$ and $R[t]$ are finitely generated $R$-modules, say

$$R[s] = Rs_1 + Rs_2 + \cdots + Rs_n$$
$$R[t] = Rt_1 + Rt_2 + \cdots + Rt_m.$$

Then

$$R[s, t] = Rs_1t_1 + \cdots + Rs_it_j + \cdots + Rs_nt_m$$

is a ring containing $s \pm t$ and $st$ that is also a finitely generated $R$-module. Hence $s \pm t$ and $st$ are also integral over $R$, which proves (1) and also (2).

To prove (3), let $t \in T$. Since $t$ is integral over $S$, it is the root of some monic polynomial $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in S[x]$. Since $a_i \in S$ is integral over $R$, each ring $R[a_i]$ is a finitely generated $R$-module and so the ring $R_1 = R[a_0, a_1, \ldots, a_{n-1}]$ is also a finitely generated $R$-module. Since the monic polynomial $p(x)$ has its coefficients in $R_1$, $t$ is integral over $R_1$ and it follows that the ring $R_1[t] = R[a_0, a_1, \ldots, a_{n-1}, t]$ is a finitely generated $R$-module. By the proposition, this means that $t$ is integral over $R$, which gives (3).

The second statement in Corollary 24 shows that taking the elements of $S$ that are integral over $R$ gives a (possibly larger) subring of $S$, and the last statement in the corollary shows that the process of taking the integral closure stops after one step:

**Corollary 25.** Let $R$ be a subring of the commutative ring $S$ with $1 \in R$. Then the integral closure of $R$ in $S$ is integrally closed in $S$.

**Examples**

**(1)** If $R$ and $S$ are fields then $S$ is integral over $R$ if and only if $S$ is algebraic over $R$ — if $s \in S$ is a root of the polynomial $p(x)$ with coefficients in $R$ then it is a root of the monic polynomial obtained by dividing by the (nonzero) leading coefficient of $p(x)$.

**(2)** Suppose $S$ is an integral extension of $R$ and $I$ is an ideal in $S$. Then $S/I$ is an integral ring extension of $R/(R \cap I)$ (reducing the monic polynomial over $R$ satisfied by $s \in S$ modulo $I$ gives a monic polynomial satisfied by $\bar{s} \in S/I$ over $R/(R \cap I)$).

**(3)** If $R$ is a U.F.D. then $R$ is integrally closed, as follows. Suppose $a/b$ is an element in the field of fractions of $R$ (with $b \neq 0$ and $a$ and $b$ having no common factors) and satisfies $(a/b)^n + r_{n-1}(a/b)^{n-1} + \cdots + r_1(a/b) + r_0 = 0$ with $r_0, \ldots, r_{n-1} \in R$. Then

$$a^n = b(-r_{n-1}a^{n-1} - \cdots - r_1ab^{n-2} - r_0b^{n-1})$$

shows that any irreducible element dividing $b$ divides $a^n$, hence divides $a$. Since $a/b$ is in lowest terms, this shows that $b$ must be a unit, i.e., $a/b \in R$.

**(4)** The polynomial ring $k[x, y]$ over the field $k$ is integrally closed in its fraction field $k(x, y)$ by example (3) above. The ideal $(x^2 - y^3)$ is prime (cf. Exercise 14, Section 9.1), so the quotient ring $R = k[x, y]/(x^2 - y^3) = k[\bar{x}, \bar{y}]$ is an integral domain. This domain is not integrally closed, however, since $\bar{x}/\bar{y}$ is an element of the fraction field of $R$ that is integral over $R$ (since $(\bar{x}/\bar{y})^3 - \bar{x} = 0$), but is not an element of $R$. In particular, $R$ is not a U.F.D. by the previous example.

We next consider the behavior of ideals in integral ring extensions.

**Definition.** Let $\varphi : R \to S$ be a homomorphism of commutative rings.

**(a)** If $I$ is an ideal in $R$ then the *extension* of $I$ to $S$ is the ideal $\varphi(I)S$ of $S$ generated by the image of $I$.

**(b)** If $J$ is an ideal of $S$, then the *contraction* in $R$ of $J$ is the ideal $\varphi^{-1}(J)$.

In the special case where $R$ is a subring of $S$ and $\varphi$ is the natural injection, the extension of $I \subseteq R$ is the ideal $IS$ in $S$ and the contraction of $J \subseteq S$ is the ideal $J \cap R$ of $R$.

It is immediate from the definition that

**(1)** $I \subseteq IS \cap R$, more generally, $I$ is contained in the contraction of its extension to $S$, and

**(2)** $(J \cap R)\, S \subseteq J$, more generally, $J$ contains the extension of its contraction in $R$.

In general equality need not hold in either situation (cf. the exercises).

If $Q$ is a prime ideal in $S$, then its contraction is prime in $R$ (although the contraction of a maximal ideal need not be maximal). On the other hand, if $P$ is a prime ideal in $R$, its extension need not be prime (or even proper) in $S$; moreover, it is not generally true that $P$ is the contraction of a prime ideal of $S$ (cf. the exercises). For integral ring extensions, however, the situation is more controlled:

**Theorem 26.** Let $R$ be a subring of the commutative ring $S$ with $1 \in R$ and suppose that $S$ integral over $R$.

**(1)** Assume that $S$ is an integral domain. Then $R$ is a field if and only if $S$ is a field.

**(2)** Let $P$ be a prime ideal in $R$. Then there is a prime ideal $Q$ in $S$ with $P = Q \cap R$. Moreover, $P$ is maximal if and only if $Q$ is maximal.

**(3)** *(The Going-up Theorem)* Let $P_1 \subseteq P_2 \subseteq \cdots \subseteq P_n$ be a chain of prime ideals in $R$ and suppose there are prime ideals $Q_1 \subseteq Q_2 \subseteq \cdots \subseteq Q_m$ of $S$ with $P_i = Q_i \cap R$, $1 \leq i \leq m$ and $m < n$. Then the ascending chain of ideals can be completed: there are prime ideals $Q_{m+1} \subseteq \cdots \subseteq Q_n$ in $S$ such that $P_i = Q_i \cap R$ for all $i$.

**(4)** *(The Going-down Theorem)* Assume that $S$ is an integral domain and $R$ is integrally closed in $S$. Let $P_1 \supseteq P_2 \supseteq \cdots \supseteq P_n$ be a chain of prime ideals in $R$ and suppose there are prime ideals $Q_1 \supseteq Q_2 \supseteq \cdots \supseteq Q_m$ of $S$ with $P_i = Q_i \cap R$, $1 \leq i \leq m$ and $m < n$. Then the descending chain of ideals can be completed: there are prime ideals $Q_{m+1} \supseteq \cdots \supseteq Q_n$ in $S$ such that $P_i = Q_i \cap R$ for all $i$.

*Proof:* To prove (1) assume first that $R$ is a field and let $s$ be a nonzero element of $S$. Then $s$ is integral over $R$, so

$$s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0 = 0$$

for some $a_0, a_1, \ldots, a_{n-1}$ in $R$. Since $S$ is an integral domain, we may assume $a_0 \neq 0$ (otherwise cancel factors of $s$). Then

$$s(s^{n-1} + a_{n-1}s^{n-2} + \cdots + a_1) = -a_0$$

and since $(-1/a_0) \in R$, this shows that $(-1/a_0)(s^{n-1} + a_{n-1}s^{n-2} + \cdots + a_1)$ is an inverse for $s$ in $S$, so $S$ is a field. Conversely, suppose $S$ is a field and $r$ is a nonzero element of $R$. Since $r^{-1} \in S$ is integral over $R$ we have

$$r^{-m} + a_{m-1}r^{-m+1} + \cdots + a_1 r^{-1} + a_0 = 0$$

for some $a_0, \ldots, a_{m-1} \in R$. Then $r^{-1} = -(a_{m-1} + \cdots + a_1 r^{m-2} + a_0 r^{m-1}) \in R$, so $R$ is a field.

The proof of the first statement in (2) is given in Corollary 50. For the second statement, observe that the integral domain $S/Q$ is an integral extension of $R/P$ (Example 2 following Corollary 25). By (1), $S/Q$ is a field if and only if $R/P$ is a field, i.e., $Q$ is maximal if and only if $P$ is maximal.

To prove (3), it suffices by induction to prove that if $P_1 \subseteq P_2$ and $Q_1$ is a prime of $S$ with $Q_1 \cap R = P_1$ then there is a prime $Q_2$ of $S$ with $Q_1 \subseteq Q_2$ and $Q_2 \cap R = P_2$. Since $\overline{S} = S/Q_1$ is an integral extension of $\overline{R} = R/P_1$, the first part of (2) shows that there exists a prime $\overline{Q_2}$ of $\overline{S}$ with $\overline{Q_2} \cap \overline{R} = P_2/P_1$. Then the preimage $Q_2$ of $\overline{Q_2}$ in $S$ is a prime ideal containing $Q_1$ with $Q_2 \cap R = P_2$.

The proof of (4) is outlined in Exercise 24 in Section 4.

**Corollary 27.** Suppose $R$ is a subring of the ring $S$ with $1 \in R$ and assume $S$ is integral and finitely generated (as a ring) over $R$. If $P$ is a maximal ideal in $R$ then there is a nonzero and finite number of maximal ideals $Q$ of $S$ with $Q \cap R = P$.

*Proof:* There exists at least one maximal ideal $Q$ lying over $P$ by (2) of the theorem, so we must see why there are only finitely many such maximal ideals in $S$. If $Q$ is a maximal ideal of $S$ with $Q \cap R = P$ then $S/Q$ is a field containing the field $R/P$. To prove that there are only finitely many possible $Q$ it suffices to prove that there are only finitely many homomorphisms from $S$ to a field containing $R/P$ that extend the homomorphism from $R$ to $R/P$. Let $S = R[s_1, \ldots, s_n]$, where the elements $s_i$ are integral over $R$ by assumption, and let $p_i(x)$ be a monic polynomial with coefficients in $R$ satisfied by $s_i$. If $Q$ is a maximal ideal of $S$ then $S/Q = (R/P)[\bar{s}_1, \ldots, \bar{s}_n]$ is the field extension of the field $R/P$ with generators $\bar{s}_1, \ldots, \bar{s}_n$. The element $\bar{s}_i$ is a root of the monic polynomial $\bar{p}_i(x)$ with coefficients in $R/P$ obtained by reducing the coefficients of $p_i(x)$ mod $P$. There are only a finite number of possible roots of this monic polynomial (in a fixed algebraic closure of $R/P$), and so only finitely many possible field extensions of the form $(R/P)[\bar{s}_1, \ldots, \bar{s}_n]$, which proves the corollary.

## Algebraic Integers

We can use the concept of an integral ring extension to define the "integers" in extension fields of the rational numbers $\mathbb{Q}$:

**Definition.** Let $K$ be an extension field of $\mathbb{Q}$.

**(1)** An element $\alpha \in K$ is called an *algebraic integer* if $\alpha$ is integral over $\mathbb{Z}$, i.e., if $\alpha$ is the root of some monic polynomial with coefficients in $\mathbb{Z}$.

**(2)** The integral closure of $\mathbb{Z}$ in $K$ is called the *ring of integers* of $K$, and is denoted by $\mathcal{O}_K$.

An algebraic integer is clearly algebraic over $\mathbb{Q}$, so the ring of all algebraic integers is the ring of integers in $\overline{\mathbb{Q}}$, an algebraic closure of $\mathbb{Q}$. Examples of algebraic integers include $\sqrt{2}$, $\sqrt{-1}$, $\sqrt[3]{5}$, etc. since these elements are certainly roots of monic polynomials with coefficients in $\mathbb{Z}$. The definition of an algebraic integer $\alpha$ is that $\alpha$ be a root

of *some* monic polynomial in $\mathbb{Z}[x]$, a condition which seems difficult to check. The next proposition gives a simple criterion for $\alpha$ to be an algebraic integer in terms of the minimal polynomial for $\alpha$.

**Proposition 28.** An element $\alpha$ in some field extension of $\mathbb{Q}$ is an algebraic integer if and only if $\alpha$ is algebraic over $\mathbb{Q}$ and its minimal polynomial $m_{\alpha,\mathbb{Q}}(x)$ has integer coefficients. In particular, the algebraic integers in $\mathbb{Q}$ are the integers $\mathbb{Z}$, i.e., $\mathcal{O}_{\mathbb{Q}} = \mathbb{Z}$.

*Proof:* If $\alpha$ is algebraic over $\mathbb{Q}$ with $m_{\alpha,\mathbb{Q}}(x) \in \mathbb{Z}[x]$, then by definition $\alpha$ is integral over $\mathbb{Z}$. Conversely, assume $\alpha$ is integral over $\mathbb{Z}$, and let $f(x)$ be a monic polynomial in $\mathbb{Z}[x]$ of minimum degree having $\alpha$ as a root. If $f$ were reducible in $\mathbb{Q}[x]$, then by Gauss' Lemma $f(x) = g(x)h(x)$ for some monic polynomials $g(x)$, $h(x)$ in $\mathbb{Z}[x]$ of degree smaller than the degree of $f$. But then $\alpha$ would be a root of either $g$ or $h$, contradicting the minimality of $f$. Hence $f$ is irreducible in $\mathbb{Q}[x]$, so $f(x) = m_{\alpha,\mathbb{Q}}(x)$ and so the minimal polynomial for $\alpha$ has coefficients in $\mathbb{Z}$. Finally, the minimal polynomial of $\alpha = a/b \in \mathbb{Q}$ ($a/b$ reduced to lowest terms and $b > 0$) is $bx - a$, which is monic if and only if $b = 1$, so $\alpha \in \mathbb{Q}$ is an algebraic integer if and only if $\alpha \in \mathbb{Z}$.

Because the integers $\mathbb{Z}$ are the algebraic integers in $\mathbb{Q}$, for emphasis (and clarity) the elements of $\mathbb{Z}$ are sometimes referred to as the "rational integers" to distinguish them from the "integers" in extensions of finite degree over $\mathbb{Q}$ (called *number fields*). The next result gives some of the basic structure of the ring of integers in a general number field.

**Theorem 29.** Let $K$ be a number field of degree $n$ over $\mathbb{Q}$.

**(1)** The ring $\mathcal{O}_K$ of integers in $K$ is a Noetherian ring and is a free $\mathbb{Z}$-module of rank $n$.

**(2)** For every $\beta \in K$ there is some nonzero $d \in \mathbb{Z}$ such that $d\beta$ is an algebraic integer. In particular, $K$ is the field of fractions of $\mathcal{O}_K$.

**(3)** If $\beta_1, \beta_2, \ldots, \beta_n$ is any $\mathbb{Q}$-basis of $K$, then there is an integer $d$ such that $d\beta_1, d\beta_2, \ldots, d\beta_n$ is a basis for a free $\mathbb{Z}$-submodule of $\mathcal{O}_K$ of rank $n$. Any basis of the $\mathbb{Z}$-module $\mathcal{O}_K$ is also a basis for $K$ as a vector space over $\mathbb{Q}$.

*Proof:* Note first that any $\mathbb{Z}$-linear dependence relation among elements in $\mathcal{O}_K$ is a $\mathbb{Q}$-linear dependence relation in $K$, and multiplying a $\mathbb{Q}$-linear dependence relation of elements of $\mathcal{O}_K$ in $K$ by a common denominator for the coefficients yields a $\mathbb{Z}$-linear dependence relation in $\mathcal{O}_K$. Let $\beta$ be any element of $K$ and let $x^k + a_{k-1}x^{k-1} + \cdots + a_0$ be the minimal polynomial of $\beta$ over $\mathbb{Q}$. If $d$ is a common denominator for the coefficients, then multiplying through by $d^k$ shows that

$$(d\beta)^k + da_{k-1}(d\beta)^{k-1} + \cdots + d^{k-1}a_1(d\beta) + d^k a_0 = 0,$$

and $d^k a_0, d^{k-1}a_1, \ldots, da_{k-1} \in \mathbb{Z}$. Hence $d\beta$ is an algebraic integer, which proves the first part of (2) and then the second statement in (2) follows immediately.

If $\beta_1, \ldots, \beta_n$ are a $\mathbb{Q}$-basis for $K$ over $\mathbb{Q}$, then there is a nonzero integer $d$ such that $d\beta_1, \ldots, d\beta_n$ all lie in $\mathcal{O}_K$. These elements are still linearly independent over $\mathbb{Q}$, so in particular are independent over $\mathbb{Z}$, hence generate a free submodule of $\mathcal{O}_K$ of rank $n$,

which proves the first statement in (3).

Since $\mathcal{O}_K$ is a subring of the field $K$, it is a torsion free $\mathbb{Z}$-module. If $\mathcal{O}_K$ were contained in some finitely generated $\mathbb{Z}$-module it would follow that $\mathcal{O}_K$ is also finitely generated over $\mathbb{Z}$, hence is a free $\mathbb{Z}$-module. If $L$ is the Galois closure of $K$, then $\mathcal{O}_K \subseteq \mathcal{O}_L$ and so it suffices to see that $\mathcal{O}_L$ is contained in a finitely generated $\mathbb{Z}$-module. Let $\alpha_1, \ldots, \alpha_m$ be a $\mathbb{Q}$-basis for $L$. Multiplying by an integer $d \in \mathbb{Z}$, if necessary, we may assume that each $\alpha_i$ is an algebraic integer, i.e., $\alpha_1, \ldots, \alpha_m \in \mathcal{O}_L$. For each fixed $\theta \neq 0$ in $L$, the map

$$T_\theta : L \to \mathbb{Q} \quad \text{defined by} \quad T_\theta(\alpha) = \mathrm{Tr}_{L/\mathbb{Q}}(\theta\alpha)$$

(where $\mathrm{Tr}_{L/\mathbb{Q}}$ denotes the trace map from $L$ to $\mathbb{Q}$, cf. Exercise 18 in Section 14.2) is a $\mathbb{Q}$-linear transformation from $L$ to $\mathbb{Q}$. This linear transformation is nonzero because $T_\theta(\theta^{-1}) = \mathrm{Tr}_{L/\mathbb{Q}}(1) = m$. It follows that the map from $L$ to $\mathrm{Hom}_{\mathbb{Q}}(L, \mathbb{Q})$ mapping $\theta$ to $T_\theta$ is an injective homomorphism of vector spaces over $\mathbb{Q}$. Since both spaces have the same dimension over $\mathbb{Q}$, the map is an isomorphism. Put another way, every linear functional on $L$ is of the form $T_\theta$ for some $\theta \in L$. In particular, there are elements $\alpha_1', \ldots, \alpha_m'$ in $L$ whose corresponding linear transformations $T_{\alpha_i'}$ give the dual basis of $\alpha_1, \ldots, \alpha_m$, i.e.,

$$\mathrm{Tr}_{L/\mathbb{Q}}(\alpha_i'\alpha_j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases}$$

Since $\alpha_1', \ldots, \alpha_m'$ are linearly independent, they give a basis for $L$ over $\mathbb{Q}$. Hence every element $\beta \in \mathcal{O}_L$ can be written

$$\beta = a_1\alpha_1' + \cdots + a_i\alpha_i' + \cdots + a_m\alpha_m'$$

with $a_1, \ldots, a_m \in \mathbb{Q}$. Multiplying by $\alpha_j$ and taking the trace shows that

$$\mathrm{Tr}_{L/\mathbb{Q}}(\beta\alpha_j) = a_1\mathrm{Tr}_{L/\mathbb{Q}}(\alpha_1'\alpha_j) + \cdots + a_i\mathrm{Tr}_{L/\mathbb{Q}}(\alpha_i'\alpha_j) + \cdots + a_m\mathrm{Tr}_{L/\mathbb{Q}}(\alpha_m'\alpha_j) = a_j.$$

But $\beta$ and $\alpha_j$ are both elements of $\mathcal{O}_L$, so also $\beta\alpha_j$ is an element of $\mathcal{O}_L$, and this implies that $a_j = \mathrm{Tr}_{L/\mathbb{Q}}(\beta\alpha_j)$ is an element of $\mathbb{Z}$ (cf. Exercise 18(d) of Section 14.2). It follows that

$$\mathcal{O}_L \subseteq \mathbb{Z}\alpha_1' + \cdots + \mathbb{Z}\alpha_m'$$

so that $\mathcal{O}_L$ is contained in a finitely generated $\mathbb{Z}$-module, proving that $\mathcal{O}_K$ (and also $\mathcal{O}_L$) is a free $\mathbb{Z}$-module.

Since $K$ has dimension $n$ as a vector space over $\mathbb{Q}$, it follows that $\mathcal{O}_K$ is a free $\mathbb{Z}$-module of rank at most $n$ (by Theorem 5 of Section 12.1). Because $\mathcal{O}_K$ also contains a free $\mathbb{Z}$-submodule of rank $n$, it follows that the $\mathbb{Z}$-rank of $\mathcal{O}_K$ is precisely $n$, proving (1), and then the second statement in (3) follows by the remarks on $\mathbb{Z}$-linear and $\mathbb{Q}$-linear dependence relations.

Finally, any ideal $I$ in $\mathcal{O}_K$ is a $\mathbb{Z}$-submodule of a free $\mathbb{Z}$-module of rank $n$, so is a free $\mathbb{Z}$-module of rank at most $n$, and a set of $\mathbb{Z}$-module generators for $I$ is also a set of $\mathcal{O}_K$-generators. Hence every ideal of $\mathcal{O}_K$ can be generated by at most $n$ elements, which implies that $\mathcal{O}_K$ is a Noetherian ring and completes the proof.

**Definition.** An *integral basis* for the number field $K$ is a basis of the ring of integers in $K$ considered as a free $\mathbb{Z}$-module of rank $[K : \mathbb{Q}]$.

If $P$ is a nonzero prime ideal in the ring of integers $\mathcal{O}_K$ of a number field $K$ then $P \cap \mathbb{Z}$ is a prime ideal in $\mathbb{Z}$. If $\alpha \in P$, then the constant term of the minimal polynomial for $\alpha$ over $\mathbb{Q}$ is then an element in $P \cap \mathbb{Z}$, which shows that $P \cap \mathbb{Z} = p\mathbb{Z}$ is also a nonzero prime ideal in $\mathbb{Z}$. By Theorem 26, every prime ideal $(p)$ in $\mathbb{Z}$ arises in this way. Since $p\mathbb{Z}$ is a maximal ideal, it also follows from (2) in Theorem 26 that *nonzero prime ideals in $\mathcal{O}_K$ are maximal*, and then by Corollary 27, there are finitely many prime ideals $P$ in $\mathcal{O}_K$ with $P \cap \mathbb{Z} = p\mathbb{Z}$. We shall see later (Corollary 16 in Section 16.3) that *every nonzero ideal in the ring of integers of a number field can be written uniquely as the product of prime ideals*, and in the case of the ideal $p\mathcal{O}_K$ the distinct prime factors are precisely the finitely many ideals $P$ in $\mathcal{O}_K$ with $P \cap \mathbb{Z} = p\mathbb{Z}$. This property replaces the unique factorization of *elements* in $\mathcal{O}_K$ into primes (which need not hold since $\mathcal{O}_K$ *need not be a U.F.D.*). We shall also see that primary ideals in $\mathcal{O}_K$ are powers of prime ideals (in fact this is equivalent to the unique factorization of ideals of $\mathcal{O}_K$ into products of prime ideals, cf. the exercises).

**Example: (The Ring of Integers in Quadratic Extensions of $\mathbb{Q}$)**

If $K$ is a quadratic extension of $\mathbb{Q}$ then $K = \mathbb{Q}(\sqrt{D})$ for some squarefree integer $D$. Then

$$\mathcal{O}_{\mathbb{Q}(\sqrt{D})} = \mathbb{Z}[\omega] = \mathbb{Z} \cdot 1 + \mathbb{Z} \cdot \omega,$$

with integral basis $1, \omega$, where

$$\omega = \begin{cases} \sqrt{D}, & \text{if } D \equiv 2, 3 \bmod 4 \\ \dfrac{1+\sqrt{D}}{2}, & \text{if } D \equiv 1 \bmod 4. \end{cases}$$

This is the quadratic integer ring introduced in Section 7.1. Since $\omega$ satisfies $\omega^2 - D = 0$ (respectively, $\omega^2 - \omega + (1 - D)/4$) for $D \equiv 2, 3 \bmod 4$ (respectively, $D \equiv 1 \bmod 4$), it follows that $\omega$ is an algebraic integer in $K$ and so $\mathbb{Z}[\omega] \subseteq \mathcal{O}_K$. To prove that this is the full ring of integers in $K$, let $\alpha = a + b\sqrt{D}$ with $a, b \in \mathbb{Q}$, and suppose that $\alpha$ is an algebraic integer. If $b = 0$, then $\alpha \in \mathbb{Q}$ and so $a \in \mathbb{Z}$. If $b \neq 0$, the minimal polynomial of $\alpha$ is $x^2 - 2ax + (a^2 - b^2 D)$. Then Proposition 28 shows that $2a$ and $a^2 - b^2 D$ are elements of $\mathbb{Z}$. Then $4(a^2 - b^2 D) = (2a)^2 - (2b)^2 D \in \mathbb{Z}$, hence $4b^2 D \in \mathbb{Z}$. Since $D$ is squarefree it follows that $2b$ is an integer. Write $a = x/2$ and $b = y/2$ for some integers $x, y$. Since $a^2 - b^2 D$ is an integer, $x^2 - y^2 D \equiv 0 \pmod 4$. Since 0 and 1 are the only squares mod 4 and $D$ is not divisible by 4, it is easy to check that the only possibilities are the following:

(i) $D \equiv 2$ or $3 \pmod 4$ and $x, y$ are both even, or

(ii) $D \equiv 1 \pmod 4$ and $x, y$ are both even or both odd.

In case (i), $a, b \in \mathbb{Z}$ and $\alpha \in \mathbb{Z}[\omega]$. In case (ii), $a + b\sqrt{D} = r + s\omega$ where $r = (x - y)/2$ and $s = y$ are both integers, so again $\alpha \in \mathbb{Z}[\omega]$.

**Example: (The Ring of Integers in Cyclotomic Fields)**

The ring of integers in the cyclotomic field $\mathbb{Q}(\zeta_n)$ of $n^{\text{th}}$ roots of unity is $\mathbb{Z}[\zeta_n]$, where $\zeta_n$ is any primitive $n^{\text{th}}$ root of 1. The elements $1, \zeta_n, \ldots, \zeta_n^{\varphi(n)-1}$ are an integral basis. It is clear that $\zeta_n$ is an algebraic integer since it is a root of $x^n - 1$, so the ring $\mathbb{Z}[\zeta_n]$ is contained in the ring of integers. The proof that this is the full ring of algebraic integers in $\mathbb{Q}(\zeta_n)$ involves techniques from algebraic number theory beyond the scope of the material here.

## Noether's Normalization Lemma and Hilbert's Nullstellensatz

We now apply some of the techniques from the algebraic theory of integral ring extensions to affine geometry.

**Definition.** If $k$ is a field the elements $y_1, y_2, \ldots, y_q$ in some $k$-algebra are called *algebraically independent* over $k$ if there is no nonzero polynomial $p$ in $q$ variables over $k$ such that $p(y_1, y_2, \ldots, y_q) = 0$.

Thus $y_1, y_2, \ldots, y_q$ are algebraically independent if and only if the $k$-algebra homomorphism from the polynomial ring $k[x_1, \ldots, x_q]$ to $k[y_1, \ldots, y_q]$ defined by $x_i \mapsto y_i$ is an isomorphism. Elements in a field extension of $k$ are algebraically independent if and only if they are independent transcendentals over $k$.

**Theorem 30.** *(Noether's Normalization Lemma)* Let $k$ be a field and suppose that $A = k[r_1, r_2, \ldots, r_m]$ is a finitely generated $k$-algebra. Then for some $q$, $0 \leq q \leq m$, there are algebraically independent elements $y_1, y_2, \ldots, y_q \in A$ such that $A$ is integral over $k[y_1, y_2, \ldots, y_q]$.

*Proof:* Proceed by induction on $m$. If $r_1, \ldots, r_m$ are algebraically independent over $k$ then take $y_i = r_i$, $i = 1, \ldots, m$. Otherwise, there exists $f(x_1, \ldots, x_m) \in k[x_1, \ldots, x_m]$ such that $f(r_1, \ldots, r_m) = 0$. The polynomial $f$ is a sum of monomials of the form $ax_1^{e_1}x_2^{e_2} \cdots x_m^{e_m}$, where the degree of this monomial is $e_1 + \cdots + e_m$ and the degree, $d$, of $f$ is the maximum of the degrees of its monomials. Renumbering the variables if necessary, we may assume that $f$ is a nonconstant polynomial in $x_m$ with coefficients in the ring $k[x_1, x_2, \ldots, x_{m-1}]$. We now perform a change of variables that transforms (or "normalizes") $f$ into a *monic* polynomial in $x_m$ with coefficients from a subring of $A$ which is generated over $k$ by $m - 1$ elements, at which point we shall be able to apply induction.

Define integers $\alpha_i = (1 + d)^i$ and new variables $X_i = x_i - x_m^{\alpha_i}$ for $1 \leq i \leq m - 1$. Let

$$g(X_1, X_2, \ldots, X_{m-1}, x_m) = f(X_1 + x_m^{\alpha_1}, X_2 + x_m^{\alpha_2}, \ldots, X_{m-1} + x_m^{\alpha_{m-1}}, x_m),$$

so $g \in k[X_1, \ldots, X_{m-1}, x_m]$. Each monomial term of $f$ contributes a single term of the form a constant times $x_m^e$ to $g$. It is also easy to check that the choice of $\alpha_i$ ensures that distinct monomials in $f$ give different values of $e$ (for example by viewing the degrees of the monomials in the new variables as integers expressed in base $b = d + 1$). If $N$ is the highest power of $x_m$ that occurs, then it follows that

$$g = cx_m^N + \sum_{i=0}^{N-1} h_i(X_1, \ldots, X_{m-1})x_m^i$$

for some nonzero $c \in k$. If now $s_i = r_i - r_m^{\alpha_i}$ then

$$\frac{1}{c}g(s_1, s_2, \ldots, s_{m-1}, r_m) = \frac{1}{c}f(r_1, r_2, \ldots, r_{m-1}, r_m) = 0,$$

which shows that $r_m$ is integral over $B = k[s_1, \ldots, s_{m-1}]$. Each $r_i$ for $1 \leq i \leq m - 1$ is integral over $B[r_m]$ since $r_i$ is a root of the monic polynomial $x - s_i - r_m^{\alpha_i}$, so $A$ is

integral over $B[r_m]$. By transitivity of integrality, $A$ is therefore integral over $B$. Since $B$ is a $k$-algebra generated by $m-1$ elements, induction completes the proof.

A more "geometric" interpretation of Noether's Normalization Lemma is indicated in Exercise 15. We next use the Normalization Lemma to prove that if $k$ is an algebraically closed field then the maximal ideals of the polynomial ring $k[x_1, x_2, \ldots, x_n]$ are of the form $(x_1-a_1, \ldots, x_n-a_n)$ for some $a_1, \ldots, a_n \in k$. Viewing $k[x_1, x_2, \ldots, x_n]$ as the ring of polynomial functions on $\mathbb{A}^n$, this says that the maximal ideals correspond to the kernels of evaluation maps at points of $\mathbb{A}^n$ — similar to the corresponding result for rings of continuous functions on a compact set (cf. Exercises 33, 34 in Section 7.4).

**Theorem 31.** *(Hilbert's Nullstellensatz — Weak Form)* Let $k$ be an algebraically closed field. Then $M$ is a maximal ideal in the polynomial ring $k[x_1, x_2, \ldots, x_n]$ if and only if $M = (x_1 - a_1, \ldots, x_n - a_n)$ for some $a_1, \ldots, a_n \in k$. Equivalently, the maps $\mathcal{Z}$ and $\mathcal{I}$ give a bijective correspondence

$$\{\text{points in } \mathbb{A}^n\} \underset{\mathcal{Z}}{\overset{\mathcal{I}}{\rightleftarrows}} \{\text{maximal ideals in } k[\mathbb{A}^n]\}.$$

Moreover, if $I$ is any proper ideal in $k[x_1, x_2, \ldots, x_n]$ then $\mathcal{Z}(I) \neq \emptyset$.

*Proof:* Certainly $(x_1 - a_1, \ldots, x_n - a_n)$ is a maximal ideal in $k[x_1, x_2, \ldots, x_n]$. Conversely, for any maximal ideal $M$ in $k[x_1, x_2, \ldots, x_n]$, let $E = k[x_1, x_2, \ldots, x_n]/M$. Then $E$ is a field containing $k$ that is finitely generated over $k$ (by $\bar{x}_i, \ldots, \bar{x}_n$). By Noether's Normalization Lemma, $E$ is integral over a polynomial ring $k[y_1, \ldots, y_q]$. Then $k[y_1, \ldots, y_q]$ is a field by Theorem 26(1), and since a polynomial ring in one or more variables is never a field, it follows that $q = 0$. Hence $E$ is integral over $k$, so $E$ is algebraic over $k$. Because $k$ is algebraically closed, $E = k$, i.e., $\bar{x}_i \in k$ for $1 \leq i \leq n$. Hence for $i = 1, \ldots, n$ there is some $a_i \in k$ such that $x_i - a_i \in M$. This means that the maximal ideal $(x_1 - a_1, \ldots, x_n - a_n)$ is contained in $M$, so $M = (x_1 - a_1, \ldots, x_n - a_n)$. Finally, if $I$ is any nonzero ideal in $k[x_1, x_2, \ldots, x_n]$ then $I$ is contained in a maximal ideal $M = (x_1 - a_1, \ldots, x_n - a_n)$, and so $(a_1, \ldots, a_n) \in \mathcal{Z}(I)$.

**Theorem 32.** *(Hilbert's Nullstellensatz)* Let $k$ be an algebraically closed field. Then $\mathcal{I}(\mathcal{Z}(I)) = \operatorname{rad} I$ for every ideal $I$ of $k[x_1, x_2, \ldots, x_n]$. Moreover, the maps $\mathcal{Z}$ and $\mathcal{I}$ define inverse bijections

$$\{\text{affine algebraic sets}\} \underset{\mathcal{Z}}{\overset{\mathcal{I}}{\rightleftarrows}} \{\text{radical ideals}\}.$$

*Proof:* Since $\operatorname{rad} I \subseteq \mathcal{I}(\mathcal{Z}(I))$ it remains to prove the reverse inclusion. By Hilbert's Basis Theorem, $I = (f_1, f_2, \ldots, f_m)$. Let $g \in \mathcal{I}(\mathcal{Z}(I))$. Introduce a new variable $x_{n+1}$ and consider the ideal $I'$ generated by $f_1, \ldots, f_m$ and $x_{n+1}g - 1$ in $k[x_1, \ldots, x_n, x_{n+1}]$. At any point of $\mathbb{A}^{n+1}$ where $f_1, \ldots, f_m$ vanish the polynomial $g$ also vanishes since $g \in \mathcal{I}(\mathcal{Z}(I))$, so that $x_{n+1}g - 1$ is nonzero. Hence $\mathcal{Z}(I') = \emptyset$ in $\mathbb{A}^{n+1}$. By the Weak Form of the Nullstellensatz, $I'$ cannot be a proper ideal, i.e., $1 \in I'$. Write

$$1 = a_1 f_1 + \cdots + a_m f_m + a_{m+1}(x_{n+1}g - 1) \qquad \text{for some } a_i \in k[x_1, \ldots, x_{n+1}].$$

Letting $y = 1/x_{n+1}$ and multiplying by a high power of $y$ in this equation shows that

$$y^N = c_1 f_1 + \cdots + c_m f_m + c_{m+1}(g - y) \qquad \text{for some } c_i \in k[x_1, \ldots, x_n, y].$$

Substituting $g$ for $y$ in this polynomial equation shows that $g^N \in I$ (in $k[x_1, \ldots, x_n]$), i.e., $g \in \text{rad } I$. Hence $\mathcal{I}(\mathcal{Z}(I)) \subseteq \text{rad } I$ and so $\mathcal{I}(\mathcal{Z}(I)) = \text{rad } I$, completing the proof.

It follows directly from Proposition 12 and Theorem 26(2) that if $S$ is an integral extension of $R$ with $1 \in R$ and if $I$ is an ideal of $R$, then

$$(\text{rad}_S\, IS) \cap R = \text{rad}_R\, I$$

where $IS$ is the ideal generated by $I$ in $S$, and the subscript indicates the ring in which the radicals are being computed. This has the following geometric interpretation.

**Corollary 33.** *(Variant of Hilbert's Nullstellensatz)* If $k$ is any field with algebraic closure $\bar{k}$ and $I$ is an ideal in $k[x_1, x_2, \ldots, x_n]$, then $\mathcal{I}_k(\mathcal{Z}_{\bar{k}}(I)) = \text{rad } I$, where $\mathcal{Z}_{\bar{k}}(I)$ is the zero set in $\bar{k}^n$ of the polynomials in $I$ and $\mathcal{I}_k(\mathcal{Z}_{\bar{k}}(I))$ is the ideal of polynomials in $k[x_1, x_2, \ldots, x_n]$ vanishing at all the points in $\mathcal{Z}_{\bar{k}}(I)$. In particular, $I = (1)$ if and only if there are no common zeros in $\bar{k}^n$ of the polynomials in $I$.

*Proof:* Since $\bar{k}[x_1, x_2, \ldots, x_n]$ is an integral extension of $k[x_1, x_2, \ldots, x_n]$ (generated by the integral elements $\bar{k}$), the corollary follows immediately from Theorem 32 and the remarks on radicals above.

From the Nullstellensatz we now have a dictionary between geometric and ring-theoretic objects over the algebraically closed field $k$:

| Geometry | Algebra |
|---|---|
| affine algebraic set $V$ | coordinate ring $k[V]$ |
| points of $V$ | maximal ideals of $k[V]$ |
| affine algebraic subsets in $V$ | radical ideals of $k[V]$ |
| subvarieties in $V$ | prime ideals in $k[V]$ |
| morphism $\varphi : V \to W$ | $k$-algebra homomorphism $\widetilde{\varphi} : k[W] \to k[V]$ |

## Computing Radicals

There are algorithms for computing radicals and primary decompositions in polynomial rings using Gröbner bases. While they are relatively elementary, they are somewhat technical and so we limit our discussion here to some preliminary results.

For hypersurfaces $V = \mathcal{Z}(f)$ defined by a single polynomial $f \in k[x_1, \ldots, x_n]$, determining $\mathcal{I}(V) = \text{rad}(f)$ is straightforward. Since $k[x_1, \ldots, x_n]$ is a U.F.D., $f$ factors uniquely as the product of powers of nonassociate irreducibles: $f = p_1^{a_1} \cdots p_s^{a_s}$ and then $\text{rad}(f)$ is generated by $p_1 \cdots p_s$ (the 'squarefree part' of $f$).

**Example**

Suppose $W = \mathcal{Z}(J)$ with $J = (u^3 - uv^2 + v^3) \in \mathbb{Q}[u, v]$. The polynomial $x^3 - x + 1$ is irreducible over $\mathbb{Q}$, so $f = u^3 - uv^2 + v^3$ is irreducible in $\mathbb{Q}[u, v]$. Hence $\operatorname{rad} J = J$ and $\mathcal{I}(W) = J$.

For nonprincipal ideals $I$, determining $\operatorname{rad} I$ is more complicated. The following proposition (based on Hilbert's Nullstellensatz) gives a criterion determining when an element is contained in $\operatorname{rad} I$.

**Proposition 34.** Suppose $k$ is any field. If $I = (f_1, \ldots, f_s)$ is a proper ideal in $k[x_1, \ldots, x_n]$, then $f \in \operatorname{rad} I$ if and only if $(f_1, \ldots, f_s, 1 - yf) = k[x_1, \ldots, x_n, y]$.

*Proof:* By Corollary 33, $(f_1, \ldots, f_s, 1 - yf) = k[x_1, \ldots, x_n, y]$ if and only if the equations

$$1 - yf(x_1, \ldots, x_n) = 0, \quad f_1(x_1, \ldots, x_n) = 0, \quad \ldots, \quad f_s(x_1, \ldots, x_n) = 0$$

have no common zero over the algebraic closure $\bar{k}$ of $k$. For a given $(a_1, \ldots, a_n) \in \bar{k}^n$, the equation $1 - yf(a_1, \ldots, a_n) = 0$ has a solution $y$ unless $f(a_1, \ldots, a_n) = 0$. Hence, the system of equations has no common zero if and only if for every $(a_1, \ldots, a_n) \in \bar{k}^n$ with $f_1(a_1, \ldots, a_n) = \cdots = f_s(a_1, \ldots, a_n) = 0$ we also have $f(a_1, \ldots, a_n) = 0$. Equivalently, if $(a_1, \ldots, a_n) \in \mathcal{Z}_{\bar{k}}(I)$, then also $f(a_1, \ldots, a_n) = 0$, i.e., we have $f \in \mathcal{I}_k(\mathcal{Z}_{\bar{k}}(I)) = \operatorname{rad} I$, by Corollary 33.

Since the reduced Gröbner basis (with respect to any fixed monomial ordering) for an ideal is unique, we immediately obtain the following algorithmic method for determining when a polynomial lies in the radical of an ideal.

**Corollary 35.** Suppose $I = (f_1, \ldots, f_s)$ in $k[x_1, \ldots, x_n]$. Then $f \in \operatorname{rad} I$ if and only if $\{1\}$ is the reduced Gröbner basis for the ideal $(f_1, \ldots, f_s, 1 - yf)$ in $k[x_1, \ldots, x_n, y]$ with respect to any monomial ordering.

**Example**

Consider $I = (x^2 - y^2, xy)$ in $k[x, y]$. The reduced Gröbner basis for $(x^2 - y^2, xy, 1 - tx)$ in $k[x, y, t]$ with respect to the order $x > y > t$ is $\{1\}$, showing $x \in \operatorname{rad}(I)$. To determine the smallest power of $x$ lying in $I$, we find that the ideal $(x^2 - y^2, xy, x^3)$ in $k[x, y]$ has the same reduced Gröbner basis as $I$ (namely $\{x^2 - y^2, xy, y^3\}$), but $(x^2 - y^2, x^2, xy)$ has basis $\{x^2, xy, y^2\}$. It follows that $x^3 \in I$ and $x^2 \notin I$ (alternatively, $x^3$ leaves a nonzero remainder after general polynomial division by $\{x^2 - y^2, xy, y^3\}$, but $x^3$ has a remainder of 0). By a similar computation (or by symmetry), $y \in \operatorname{rad} I$, with $y^3 \in I$ but $y^2 \notin I$. Since $(x, y) \subseteq \operatorname{rad} I$, it follows that $\operatorname{rad} I = (x, y)$.

Some additional results for computing radicals are presented in the exercises.

## EXERCISES

Let $R$ be a subring of the commutative ring $S$ with $1 \in R$.

1. Use the fact that a U.F.D. is integrally closed to prove that the Gaussian integers, $\mathbb{Z}[i]$, is the ring of integers in $\mathbb{Q}(i)$.
2. Suppose $k$ is a field and let $t = \bar{x}/\bar{y}$ in the field of fractions of the integral domain $R = k[x, y]/(x^2 - y^3)$. Prove that $K = k(t)$ is the fraction field of $R$ and $k[t]$ is the integral closure of $R$ in $K$.
3. Suppose $k$ is a field and $i$ and $j$ are relatively prime positive integers. Find the normalization of the integral domain $R = k[x, y]/(x^i - y^j)$ (cf. Exercise 14, Section 9.1).
4. Suppose $k$ is a field and let $P$ be the ideal $(y^2 - x^3 - x^2)$ in the polynomial ring $k[x, y]$. Prove that $P$ is a prime ideal and find the normalization of the integral domain $R = k[x, y]/P$. [To prove $P$ is prime, show that $y^2 - x^3 - x^2$ is irreducible in the U.F.D. $k[x, y]$. Then consider $t = \bar{y}/\bar{x} \in R$.]
5. If $R$ is an integral domain with field of fractions $F$, show that $F$ is a finitely generated $R$-module if and only if $R = F$.
6. For each of the following give specific rings $R \subseteq S$ and explicit ideals in these rings that exhibit the specified relation:
   **(a)** an ideal $I$ of $R$ such that $I \neq SI \cap R$ (so the contraction of the extension of an ideal $I$ need not equal $I$)
   **(b)** a prime ideal $P$ of $R$ such that there is no prime ideal $Q$ of $S$ with $P = Q \cap R$
   **(c)** a maximal ideal $M$ of $S$ such that $M \cap R$ is not maximal in $R$
   **(d)** a prime ideal $P$ of $R$ whose extension $PS$ to $S$ is not a prime ideal in $S$
   **(e)** an ideal $J$ of $S$ such that $J \neq (J \cap R)S$ (so the extension of the contraction of an ideal $J$ need not equal $J$).
7. Let $\mathcal{O}_K$ be the ring of integers in a number field $K$.
   **(a)** Suppose that every nonzero ideal $I$ of $\mathcal{O}_K$ can be written as the product of powers of prime ideals. Prove that an ideal $Q$ of $\mathcal{O}_K$ is $P$-primary if and only if $Q = P^m$ for some $m \geq 1$. [Show first that since nonzero primes in $\mathcal{O}_K$ are maximal that ${P_1}^{m_1} \subseteq {P_2}^{m_2}$ for distinct nonzero primes $P_1, P_2$ implies $P_1 = P_2$.]
   **(b)** Suppose that an ideal $Q$ of $\mathcal{O}_K$ is $P$-primary if and only if $Q = P^m$ for some $m \geq 1$. Assuming all of Theorem 21, prove that every nonzero ideal $I$ of $\mathcal{O}_K$ can be written uniquely as the product of powers of prime ideals. [Prove that ${P_1}^{m_1}$ and ${P_2}^{m_2}$ are comaximal ideals if $P_1$ and $P_2$ are distinct nonzero prime ideals and use the Chinese Remainder Theorem.]
8. Prove that if $s_1, \ldots, s_n \in S$ are integral over $R$, then the ring $R[s_1, \ldots, s_n]$ is a finitely generated $R$-module.
9. Suppose that $S$ is integral over $R$ and that $P$ is a prime ideal in $R$. Prove that every element $s$ in the ideal $PS$ generated by $P$ in $S$ satisfies an equation $s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0 = 0$ where the coefficients $a_0, a_1, \ldots, a_{n-1}$ are elements of $P$. [If $s = p_1 s_1 + \cdots + p_m s_m \in PS$, show that $T = R[s_1, \ldots, s_m]$ satisfies the hypotheses in Proposition 23(3). Follow the proof in Proposition 23 that $s$ is integral, noting that $s \in PT$ so that the $a_{ij}$ are elements of $P$.]
10. Prove the following generalization of Proposition 28: Suppose $R$ is an integrally closed integral domain with field of fractions $k$ and $\alpha$ is an element of an extension field $K$ of $k$. Show that $\alpha$ is integral over $R$ if and only if $\alpha$ is algebraic over $k$ and the minimal polynomial $m_{\alpha,k}(x)$ for $\alpha$ over $k$ has coefficients in $R$. [If $\alpha$ is integral prove the conjugates

of $\alpha$, i.e., the roots of $m_{\alpha,k}(x)$, are also integral, so the elementary symmetric functions of the conjugates are elements of $k$ that are integral over $R$.]

**11.** Suppose $R$ is an integrally closed integral domain with field of fractions $k$ and $p(x) \in R[x]$ is a monic polynomial. Show that if $p(x) = a(x)b(x)$ with monic polynomials $a(x), b(x) \in k[x]$ then $a(x), b(x) \in R[x]$ (compare to Gauss' Lemma, Proposition 5, Section 9.3). [See the previous exercise.]

**12.** Suppose $S$ is an integral domain that is integral over a ring $R$ as in the previous exercise. If $P$ is a prime ideal in $R$, let $s$ be any element in the ideal $PS$ generated by $P$ in $S$. Prove that, with the exception of the leading term, the coefficients of the minimal polynomial $m_{s,k}(x)$ for $s$ over $k$ are elements of $P$. [By Exercise 10, $m_{s,k}(x) \in R[x]$. Exercise 9 shows that $s$ is a root of a monic polynomial $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ with $a_0, \ldots, a_{n-1} \in P$. Use the previous exercise to show that $p(x) = m_{s,k}(x)b(x)$ with $b(x)$ in $R[x]$, and consider this equation in the integral domain $(R/P)[x]$. ]

The next two exercises extend Exercise 6 in Section 7.5 by characterizing fields that are not fields of fractions of any of their proper subrings.

**13.** Let $K$ be a field of characteristic 0 and let $A$ be a subring of $K$ maximal with respect to $1/2 \notin A$. (Such $A$ exists by Zorn's Lemma.) Let $F$ be the field of fractions of $A$ in $K$.
- **(a)** Show that $K$ is algebraic over $F$. [If $t$ is transcendental over $F$, show that $1/2 \notin A[t]$.]
- **(b)** Show that $A$ is integrally closed in $K$. [Show that $1/2$ is not in the integral closure of $A$ in $K$.]
- **(c)** Deduce from (a) and (b) that $K = F$.

**14.** Show that a field $K$ is the field of fractions of some proper subring of $K$ if and only if $K$ is not a subfield of the algebraic closure of a finite field. [If $K$ contains $t$ transcendental over $\mathbb{F}_p$ argue as in the preceding exercise with $1/t$ in place of $1/2$ to show that $K$ is the quotient field of some proper subring.]

The next exercise gives a "geometric" interpretation of Noether's Normalization Lemma, showing that every affine algebraic set is a *finite covering* of some affine $n$-space.

**15.** Let $V$ be an affine algebraic set over an algebraically closed field $k$. Prove that for some $n$ there is a surjective morphism from $V$ onto $\mathbb{A}^n$ with finite fibers, and that if $V$ is a variety, then $n$ can be taken to be the dimension of $V$. [By Noether's Normalization Lemma the finitely generated $k$-algebra $S = k[V]$ contains a polynomial subalgebra $R = k[x_1, x_2, \ldots, x_n]$ such that $S$ is integral over $R$. Apply Theorem 6 to the inclusion of $R$ in $S$ to obtain a morphism $\varphi$ from $V$ to $\mathbb{A}^n$. To see that $\varphi$ is surjective with finite fibers, apply Corollary 27 to the maximal ideal $(x_1 - a_1, \ldots, x_n - a_n)$ of $R$ corresponding to a point $(a_1, \ldots, a_n)$ of $\mathbb{A}^n$.]

**16.** Let $V$ be an affine algebraic set in $\mathbb{C}^n$. Prove that $V$ is compact in the Euclidean topology (i.e., closed and bounded) if and only if it is finite. [Use Exercise 18 in Section 2, the previous exercise, and the behavior of compact sets with respect to continuous functions.]

**17.** Let $R$ be a subring of the commutative ring $S$ with $1_S \in R$ and suppose that $S$ is integral over $R$. This exercise proves that $R$ and $S$ have the same *Krull dimension*, cf. Section 16.1.
- **(a)** If $P_1 \subset P_2 \subset \cdots \subset P_n$ is a chain of distinct prime ideals in $R$ prove that there is a chain $Q_1 \subset Q_2 \subset \cdots \subset Q_n$ of distinct prime ideals in $S$ with $Q_i \cap R = P_i$.
- **(b)** Prove conversely that if $Q_1 \subset Q_2 \subset \cdots \subset Q_n$ is a chain of distinct prime ideals in $S$ and $P_i = Q_i \cap R$ then $P_1 \subset P_2 \subset \cdots \subset P_n$ is a chain of distinct prime ideals in $R$. [To prove the $P_i$ are distinct, pass to a quotient and reduce the problem to showing that if $Q$ is a nonzero prime ideal in the integral domain $S$ then $Q \cap R$ is a nonzero prime

ideal in $R$. In this case, if $s \in Q$ is nonzero, show that the constant coefficient of a polynomial of minimal degree in $R[x]$ satisfied by $s$ is a nonzero element in $Q \cap R$.]

**18.** Let $V = \mathcal{Z}(I)$ and $W = \mathcal{Z}(J)$ where $I$ is the ideal $(uv + v) \subset \mathbb{C}[u, v]$ and $J$ is the ideal $(-2y - y^2 + 2z + z^2, 2x - yz - z^2) \subset \mathbb{C}[x, y, z]$.

**(a)** Show that $I$ and $J$ are prime ideals. Conclude that $I = \mathcal{I}(V)$ and $J = \mathcal{I}(W)$ and that $V$ and $W$ are varieties.

**(b)** Show that the map $\varphi : V \to W$ defined by $\varphi((a_1, a_2)) = (a_1^2 + a_2, a_1 + a_2, a_1 - a_2)$ is an isomorphism.

**19.** Let $I = (x^3 + y^3 + z^3, x^2 + y^2 + z^2, (x + y + z)^3) \subset k[x, y, z]$. Use Gröbner bases to show that $x, y, z \in \operatorname{rad} I$ if $\operatorname{ch}(k) \neq 2, 3$.

**20.** Let $I = (x^3 + y^3 + z^3, xy + xz + yz, xyz) \subset k[x, y, z]$. Use Gröbner bases to show that $x, y, z \in \operatorname{rad} I$.

**21.** Let $I = (x^4 + y^4 + z^4, x + y + z) \subset k[x, y, z]$.

**(a)** Use Gröbner bases to show that $xy + xz + yz \in \operatorname{rad} I$ if $\operatorname{ch}(k) \neq 2$ and determine the smallest power of $xy + xz + yz$ contained in $I$. Show that none of $x$, $y$ or $z$ is contained in $\operatorname{rad} I$.

**(b)** If $J = (x^4 + y^4 + z^4, x + y + z, xy + xz + yz)$ show that the reduced Gröbner basis of $J$ relative to the lexicographic ordering $x > y > z$ is $\{x + y + z, y^2 + yz + z^2\}$. Deduce that $k[x, y, z]/J \cong k[y, z]/(y^2 + yz + z^2)$ and that $J$ is radical if $\operatorname{ch}(k) \neq 3$.

**(c)** If $\operatorname{ch}(k) \neq 2, 3$, show that $\operatorname{rad} I = J$.

**(d)** If $\operatorname{ch}(k) = 3$, show that $\operatorname{rad} I = (x - y, y - z)$.

**(e)** If $\operatorname{ch}(k) = 2$, show that $I = (x + y + z)$ is a prime, hence radical, ideal.

**22.** Let $I = (x^2y + z^3, x + y^3 - z, 2y^4z - yz^2 - z^3) \subset k[x, y, z]$. Use Gröbner bases to show that $x, y, z \in \operatorname{rad} I$ and conclude that $\operatorname{rad} I = (x, y, z)$. Show that $x^9, y^7, z^9$ are the smallest powers of $x, y, z$, respectively, lying in $I$.

**23.** Let $V = \mathcal{Z}(x^3 - x^2z - y^2z)$ and $W = \mathcal{Z}(x^2 + y^2 - z^2)$ in $\mathbb{C}^3$. Show that $\mathcal{I}(V) = (x^3 - x^2z - y^2z)$ and $\mathcal{I}(W) = (x^2 + y^2 - z^2)$ in $\mathbb{C}[x, y, z]$.

**24.** Let $V = \mathcal{Z}(x^3 + y^3 + 7z^3) \subset \mathbb{C}^3$. Show that $\mathcal{I}(V) = (x^3 + y^3 + 7z^3)$ in $\mathbb{C}[x, y, z]$.

**25.** Let $I = (xz + y^2 + z^2, xy - xz + yz - 2z^2)$ and let $K = I + (x^2 - 3y^2 + yz) \subset \mathbb{C}[x, y, z]$.

**(a)** By Exercise 46 in Section 1, there is an injective $\mathbb{C}$-algebra homomorphism from $\mathbb{C}[x, y, z]/K$ to $\mathbb{C}[u, v]/(u^3 - uv^2 + v^3)$. Use this together with the example preceding Proposition 34 to prove that $K$ is a radical ideal and deduce that $\operatorname{rad} I \subseteq K$.

**(b)** Show that $\operatorname{rad} I \subseteq (y, z)$.

**(c)** Show that $K \cap (y, z) = I$ and deduce that $I$ is radical, so that $\mathcal{I}(V) = I$ if $V = \mathcal{Z}(I)$.

**(d)** Show that $y(x^2 - 3y^2 + yz)$ and $z(x^2 - 3y^2 + yz)$ are elements of $I$ but none of $y$, $z$, or $x^2 - 3y^2 + yz$ is contained in $I$.

**26.** Let $I$ be an ideal in $k[x_1, \ldots, x_n]$. Prove that the following are equivalent (an ideal satisfying any of these conditions is called a *zero-dimensional ideal* because of (d)):

**(a)** The quotient $k[x_1, \ldots, x_n]/I$ has finite dimension as a vector space over $k$.

**(b)** $I \cap k[x_i] \neq 0$ for each $i = 1, 2, \ldots, n$.

**(c)** If $G$ is any reduced Gröbner basis for $I$ then for each $i = 1, \ldots, n$, there is a $g_i \in G$ with leading term $x_i^{n_i}$ for some $n_i \geq 1$.

**(d)** The set of common zeros $\mathcal{Z}_{\bar{k}}(I)$ of the polynomials in $I$ in an algebraic closure $\bar{k}$ of $k$ is finite.

[For (a) implies (b) use the injection $k[x_i]/(I \cap k[x_i]) \hookrightarrow k[x_1, \ldots, x_n]/I$. For (b) implies (c) note some $LT(g_i)$ divides the leading term of a generator for $I \cap k[x_i]$. For (c) implies (a)

use Exercise 37 in Section 9.6. Show (b) implies (d). For (d) implies (b) show the product $m_{a_1,k}(x_i) \dots m_{a_N,k}(x_i)$ of the minimal polynomials of the $i^{\text{th}}$ coordinates $a_1, \dots, a_N$ of the points in $\mathcal{Z}_{\bar{k}}(I)$ is a nonzero polynomial in $\mathcal{I}(\mathcal{Z}_{\bar{k}}(I))$ and apply Corollary 33.]

**27.** Let $I$ be a zero-dimensional ideal in $k[x_1, \dots, x_n]$ and let $I'$ be the ideal generated by $I$ in $\bar{k}[x_1, \dots, x_n]$ where $\bar{k}$ is the algebraic closure of $k$. Let $\mathcal{Z}(I)$ be the zero set of $I$ in $k^n$ and let $\mathcal{Z}_{\bar{k}}(I)$ be the zero set of $I$ (equivalently, of $I'$) in $\bar{k}^n$.

**(a)** Prove that $|\mathcal{Z}_{\bar{k}}(I)| = \dim_{\bar{k}} \bar{k}[x_1, \dots, x_n]/\operatorname{rad} I'$. [Show that $\operatorname{rad} I'$ is the product of the maximal ideals corresponding to the points in $V_{\bar{k}}$ and use the Chinese Remainder Theorem.]

**(b)** Show $|\mathcal{Z}(I)| \leq \dim_k k[x_1, \dots, x_n]/I$. [One approach: use Exercise 43 in Section 1 and observe that $\dim_{\bar{k}} \bar{k}[x_1, \dots, x_n]/\operatorname{rad} I' \leq \dim_{\bar{k}} \bar{k}[x_1, \dots, x_n]/I'$.]

**28.** Suppose $I$ is a zero-dimensional ideal in $k[x_1, \dots, x_n]$, and suppose $I \cap k[x_i]$ is generated by the nonzero polynomial $h_i$ (cf. Exercise 26). Let $r_i$ be the product of the irreducible factors of $h_i$ (the 'squarefree part' of $h_i$).

**(a)** Prove that $I + (r_1, \dots, r_n) \subseteq \operatorname{rad} I$.

**(b)** *(Radicals of zero-dimensional ideals for perfect fields)* If $k$ is a perfect field, prove that $\operatorname{rad} I = I + (r_1, \dots, r_n)$. [Use induction on $n$. Write $r_1 = p_1 \dots p_t$ with distinct irreducibles $p_i$ in $k[x_1]$. If $J = I + (r_1, \dots, r_n)$ show that $J = J_1 \cap \dots \cap J_t$ where $J_t = J + (p_t)$. Show for each $i$ that reduction modulo $p_i$ induces an isomorphism $k[x_1, \dots, x_n]/J_i \cong K[x_2, \dots, x_n]/J_i'$ where $K$ is the extension field $k[x]/(p_i)$ and $J_i' \subseteq K[x_2, \dots, x_n]$ is the reduction of the ideal $J_i$ modulo $(p_i)$. Use Exercise 11 of Section 13.5 to show that the image of $r_j$ in $J_i' \cap K[x_j]$ remains a nonzero squarefree polynomial for each $j = 2, \dots, n$ since $k$ is perfect. Conclude by induction that $J_i'$ is a radical ideal. Deduce that $J_i$ is a radical ideal, and finally that $J$ is a radical ideal.]

**(c)** Find the radicals of $(x^7 + x + y^3, x^4 + y^3 + y)$, $(x^3 - xy^2 + x, x^2y + y^3)$, and $(x^4 + y^3, x^3 - xy + y^2)$ in $\mathbb{Q}[x, y]$ and of $(x^2 + y^2z, x^2y^2 + z^3, y^2 + z^2)$ in $\mathbb{Q}[x, y, z]$.

**(d)** Let $k = \mathbb{F}_p(t)$. Show that $I = (x^p + t, y^p - t)$ is a zero-dimensional ideal in $k[x, y]$ such that both $I \cap k[x]$ and $I \cap k[y]$ contain nonzero squarefree polynomials, but that $I$ is not a radical ideal (so the result in (b) need not hold if $k$ is not perfect). [Show that $x + y \in \operatorname{rad} I$ but $x + y \notin I$.]

## 15.4 LOCALIZATION

The idea of "localization at a prime" in a ring is an extremely powerful and pervasive tool in algebra for isolating the behavior of the ideals in a ring. It is an algebraic analogue of the familiar idea of localizing at a point when considering questions of, for example, the differentiability of a function $f(x)$ on the real line. In fact one of the important applications (and also one of the original motivations for the development) of this technique is to translate such "local" properties in the geometry of affine algebraic spaces to corresponding properties of their coordinate rings.

We first consider a very general construction of "rings of fractions." Let $D$ be a multiplicatively closed subset of $R$ containing 1 (i.e., $1 \in D$ and $ab \in D$ if $a, b \in D$). The next result constructs a new ring $D^{-1}R$ which is the "smallest" ring in which the elements of $D$ become units. This generalizes the construction of rings of fractions in Section 7.5 by allowing $D$ to contain zero or zero divisors, and so in this case $R$ need not embed as a subring of $D^{-1}R$.

**Theorem 36.** Let $R$ be a commutative ring with 1 and let $D$ be a multiplicatively closed subset of $R$ containing 1. Then there is a commutative ring $D^{-1}R$ and a ring homomorphism $\pi : R \to D^{-1}R$ satisfying the following universal property: for any homomorphism $\psi : R \to S$ of commutative rings that sends 1 to 1 such that $\psi(d)$ is a unit in $S$ for every $d \in D$, there is a unique homomorphism $\Psi : D^{-1}R \to S$ such that $\Psi \circ \pi = \psi$.

*Proof:* The proof is very similar to the proof of Theorem 15 in Section 7.5. In this case we define a relation on $R \times D$ by

$$(r, d) \sim (s, e) \quad \text{if and only if} \quad x(er - ds) = 0 \quad \text{for some } x \in D.$$

This relation is clearly reflexive and symmetric. If $(r, d) \sim (s, e)$ and $(s, e) \sim (t, f)$ then $x(er - ds) = 0$ and $y(fs - et) = 0$ for some $x, y \in D$. Multiplying the first equation by $fy$ and the second by $dx$ and adding gives $exy(fr - dt) = 0$. Since $D$ is closed under multiplication, $(r, d) \sim (t, f)$ and so $\sim$ is transitive.

Let $r/d$ denote the equivalence class of $(r, d)$ under $\sim$ and let $D^{-1}R$ be the set of these equivalence classes. Define addition and multiplication in $D^{-1}R$ by

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \qquad \text{and} \qquad \frac{a}{b} \times \frac{c}{d} = \frac{ac}{bd}.$$

It is an exercise to check that these operations are well defined and make $D^{-1}R$ into a commutative ring with $1 = 1/1$. For each $d \in D$, $d/1$ is a unit in $D^{-1}R$ (even in the degenerate case when $D^{-1}R$ is the zero ring).

Finally, define $\pi : R \to D^{-1}R$ by $\pi(r) = r/1$. It follows easily that $\pi$ is a ring homomorphism. Suppose that $\psi : R \to S$ is a homomorphism of commutative rings that sends 1 to 1 such that $\psi(d)$ is a unit in $S$ for every $d \in D$. Define

$$\Psi : D^{-1}R \to S \qquad \text{by} \qquad \Psi\left(\frac{r}{d}\right) = \psi(r)\psi(d)^{-1}.$$

This map is well defined because if $r/d = s/e$ then $x(er - ds) = 0$ for some $x \in D$. Then $\psi(x)(\psi(er) - \psi(ds)) = 0$ in $S$, so $\psi(er) - \psi(ds) = 0$ since $\psi(x)$ is a unit in $S$, and therefore $\psi(r)\psi(d)^{-1} = \psi(s)\psi(e)^{-1}$. It is immediate that $\Psi$ is a ring homomorphism and $\Psi \circ \pi = \psi$.

Finally, $\Psi$ is unique because every element of $D^{-1}R$ can be written as a product $(r/1)(d/1)^{-1}$. The value of $\Psi$ on each element of the form $x/1$ is uniquely determined by $\psi$, namely $\Psi(x/1) = \Psi(\pi(x)) = \psi(x)$. Since $\Psi$ is a ring homomorphism, its value on $u^{-1}$ for any unit $u$ is uniquely determined by $\Psi(u)$. Thus $\Psi$ is uniquely determined on every element of $D^{-1}R$, completing the proof.

**Corollary 37.** In the notation of Theorem 36,

**(1)** $\ker \pi = \{r \in R \mid xr = 0 \text{ for some } x \in D\}$; in particular, $\pi : R \to D^{-1}R$ is an injection if and only if $D$ contains no zero divisors of $R$, and

**(2)** $D^{-1}R = 0$ if and only if $0 \in D$, hence if and only if $D$ contains nilpotent elements.

*Proof:* By definition, we have $\pi(r) = 0$ if and only if $(r, 1) \sim (0, 1)$, i.e., if and only if $xr = 0$ for some $x \in D$, which is (1). For (2), note that $D^{-1}R = 0$ if and only

if the 1 of this ring is zero, i.e., $(1, 1) \sim (0, 1)$. This occurs if and only if $x1 = 0$ for some $x \in D$, i.e., if and only if $0 \in D$.

**Definition.** The ring $D^{-1}R$ is called the *ring of fractions of $R$ with respect to $D$* or the *localization of $R$ at $D$*.

**Examples**

**(1)** Let $R$ be an integral domain and let $D = R - \{0\}$. Then $D^{-1}R$ is the field of fractions, $Q$, of $R$ described in Section 7.5. More generally, if $D$ is any multiplicatively closed subset of $R - \{0\}$, then $D^{-1}R$ is the subring of $Q$ consisting of elements $r/d$ with $r \in R$ and $d \in D$.

**(2)** Let $R$ be any commutative ring with 1 and let $f$ be any element of $R$. Let $D$ be the multiplicative set $\{f^n \mid n \geq 0\}$ of nonnegative powers of $f$ in $R$. Define $R_f = D^{-1}R$. Note that $R_f = 0$ if and only if $f$ is nilpotent. If $f$ is not nilpotent, then $f$ becomes a unit in $R_f$. It is not difficult to see that

$$R_f \cong R[x]/(xf - 1),$$

where $R[x]$ is the polynomial ring in the variable $x$ (cf. the exercises). Note also that $R_f$ and $R_{f^n}$ are naturally isomorphic for any $n \geq 1$ since both $f$ and $f^n$ are units in both rings. If $f$ is a zero divisor then $\pi : R \to R_f$ does not embed $R$ into $R_f$. For example, let $R = k[x, y]/(xy)$, and take $f = x$. Then $x$ is a unit in $R_x$ and $y$ is mapped to 0 by the first part of the corollary (explicitly: $y = xy/x = 0$ in $R_x$). In this case $\pi(R) = k[x] \subset R_f = k[x, x^{-1}]$.

**(3)** (*Localizing at a Prime*) Let $P$ be a prime ideal in any ring $R$ and let $D = R - P$. By definition of a prime ideal $D$ is multiplicatively closed. Passing to the ring $D^{-1}R$ in this case is called *localizing $R$ at $P$* and the ring $D^{-1}R$ is denoted by $R_P$. Every element of $R$ not in $P$ becomes a unit in $R_P$. For example, if $R = \mathbb{Z}$ and $P = (p)$ is a prime ideal, then

$$\mathbb{Z}_{(p)} = \{\frac{a}{b} \in \mathbb{Q} \mid p \nmid b\} \subseteq \mathbb{Q}$$

and every integer $b$ not divisible by $p$ is a unit.

**(4)** If $V$ is any nonempty set and $k$ is a field, let $R$ be any ring of $k$-valued functions on $V$ containing the constant functions (for instance, the ring of all continuous real valued functions on the closed interval $[0, 1]$). For any $a \in V$ let $M_a$ be the ideal of functions in $R$ that vanish at $a$. Then $M_a$ is the kernel of the ring homomorphism from $R$ to the field $k$ given by evaluating each function in $R$ at $a$. Since $R$ contains the constant functions, evaluation is surjective and so $M_a$ is a maximal (hence also prime) ideal. The localization of $R$ at this prime ideal is then

$$R_{M_a} = \left\{\frac{f}{g} \mid f, g \in R,\ g(a) \neq 0\right\}.$$

Each function in $R_{M_a}$ can then be evaluated at $a$ by $(f/g)(a) = f(a)/g(a)$, and this value does not depend on the choice of representative for the class $f/g$, so $R_{M_a}$ becomes a ring of $k$-valued "rational functions" defined at $a$.

We next consider extensions and contractions of ideals with respect to the map $\pi : R \to D^{-1}R$ in Theorem 36. To ease some of the notation, if $I$ is an ideal of $R$, let ${}^eI$ denote the extension of $I$ to $D^{-1}R$ (instead of the more cumbersome $D^{-1}R\,\pi(I)$), and if $J$ is an ideal of $D^{-1}R$, let ${}^cJ$ denote the contraction of $J$ to $R$.

If $I$ is an ideal of $R$ then it is easy to see that every element of ${}^eI$ can be written in the form $a/d$ for some $a \in I$ and $d \in D$, so the extension of $I$ to $D^{-1}R$ is also frequently denoted by $D^{-1}I$.

**Proposition 38.** In the preceding notation we have

**(1)** For any ideal $J$ of $D^{-1}R$ we have $J = {}^e({}^cJ)$. In particular, every ideal of $D^{-1}R$ is the extension of some ideal of $R$, and distinct ideals of $D^{-1}R$ have distinct contractions in $R$.

**(2)** For any ideal $I$ of $R$ we have

$$ {}^c({}^eI) = \{r \in R \mid dr \in I \text{ for some } d \in D\}. $$

Also, ${}^eI = D^{-1}R$ if and only if $I \cap D \neq \emptyset$.

**(3)** Extension and contraction give a bijective correspondence

$$ \left\{ \begin{array}{c} \text{prime ideals } P \text{ of } R \\ \text{with } P \cap D = \emptyset \end{array} \right\} \underset{c}{\overset{e}{\rightleftarrows}} \left\{ \text{prime ideals of } D^{-1}R \right\}. $$

**(4)** If $R$ is Noetherian (or Artinian) then $D^{-1}R$ is Noetherian (Artinian, respectively).

*Proof:* We always have ${}^e({}^cJ) \subseteq J$. For the reverse inclusion let $a/d \in J$. Then $a/1 = d(a/d) \in J$, and so $a \in \pi^{-1}(J) = {}^cJ$. Thus $a/1 \in {}^e({}^cJ)$, so we also have $(a/1)(1/d) = a/d \in {}^e({}^cJ)$, hence $J = {}^e({}^cJ)$. This proves the first statement in (1) and the second statement follows immediately.

Let $I' = \{r \in R \mid dr \in I \text{ for some } d \in D\}$. We first show $I' \subseteq {}^c({}^eI)$. If $r \in I'$ then there is some $d \in D$ such that $dr = a \in I$. Then $r/1 = a/d \in {}^eI$, so $r \in {}^c({}^eI)$. To show the reverse containment ${}^c({}^eI) \subseteq I'$, let $r \in {}^c({}^eI)$ so that $r/1 = a/d$ for some $a \in I$ and $d \in D$. Then $x(dr - a) = 0$ for some $x \in D$, so $xdr = xa \in I$, and because $xd \in D$ it follows that $r \in I'$. This proves the first assertion of (2). Now ${}^eI = D^{-1}R$ if and only if $1/1 \in {}^eI$, if and only if $1 \in {}^c({}^eI) = I'$. The second assertion of (2) then follows from the definition of $I'$.

To prove (3) observe first that if $Q$ is a prime ideal in $D^{-1}R$, then its preimage under any homomorphism sending 1 to 1 is a prime ideal (cf. Exercise 13, Section 7.4), so $c$ maps prime ideals of $D^{-1}R$ to prime ideals of $R$ disjoint from $D$. In the reverse direction, let $P$ be a prime ideal of $R$ disjoint from $D$ and let $Q = {}^eP$ and suppose $(a/d_1)(b/d_2) \in Q$. Then $(ab)/(d_1d_2) \in Q$, so $ab/(d_1d_2) = c/d$ for some $c \in P$ and $d \in D$. Then $x(dab - d_1d_2c) = 0$ for some $x \in D$. Since $c \in P$ we have $xdab \in P$, and since $P$ is a prime ideal disjoint from $D$ we have $ab \in P$. Since $P$ is prime, either $a \in P$ or $b \in P$, hence $a/d_1$ or $b/d_2$ is in $Q$. This proves $Q$ is a prime ideal and shows that $e$ maps prime ideals of $R$ disjoint from $D$ to prime ideals of $D^{-1}R$. Finally, it follows immediately from (2) that $P = {}^c({}^eP)$ for every prime ideal of $R$ disjoint from $D$. Thus $c$ and $e$ are inverse correspondences, hence are bijections between these sets of prime ideals. This establishes (3).

By (1) every ascending (respectively, descending) chain of distinct ideals in $D^{-1}R$ contracts to an ascending (respectively, descending) chain of distinct ideals in $R$, giving (4) and completing the proof.

Because $1 \in D$, first localizing the ideal $I$ and then contracting that localization as in (2) results in an ideal in $R$ containing $I$: $I \subseteq {}^c({}^eI)$.

**Definition.** Suppose $R$ is a commutative ring with 1 and $D$ is a multiplicatively closed subset containing 1. The *saturation* of the ideal $I$ in $R$ with respect to $D$ is the ideal ${}^c({}^eI)$ in $R$, where contraction and extension are computed with respect to $\pi : R \mapsto D^{-1}R$. If $I = {}^c({}^eI)$ then $I$ is said to be *saturated* with respect to $D$.

Loosely speaking, (2) of Proposition 38 shows that the saturation of $I$ consists of elements of $R$ that would lie in $I$ if we allowed denominators from $D$. The ideal is saturated with respect to $D$ if we don't obtain any additional elements even if we allow denominators from $D$.

We can apply our results on localization to give an algorithm for determining whether an ideal $P$ in the polynomial ring $k[x_1, \ldots, x_n]$ with coefficients in the field $k$ is prime. The basic idea is to use the fact that $k[x_1, \ldots, x_i] = k[x_1, \ldots, x_{i-1}][x_i]$ to consider inductively whether the ideals $P_i = P \cap k[x_1, \ldots, x_i]$ are prime.

In general, suppose $R$ is a commutative ring. If $P$ is a prime ideal in $R[x]$ then $P \cap R$ is a prime ideal in $R$ and so $S = R/(P \cap R)$ is an integral domain. Let $F$ denote its quotient field. We then have two natural ring homomorphisms:

$$R[x] \longrightarrow (R/P \cap R)[x] = S[x] \longrightarrow F[x]$$

where the first is the natural projection homomorphism and the second is the natural inclusion induced by $S \subseteq F$. Note that $F[x]$ is the localization of $S[x]$ with respect to the multiplicatively closed set $D = S - \{0\}$. The next proposition shows that the image of $P$ under the first homomorphism is a prime ideal in $S[x]$ that is saturated with respect to $D$ and extends to a prime ideal in $F[x]$, and that, conversely, we can determine whether an ideal is prime in $R[x]$ by these properties.

**Proposition 39.** Suppose $R$ is a commutative ring with 1 and $I$ is an ideal in $R[x]$. Then $I$ is a prime ideal in $R[x]$ if and only if

- **i.** $J = I \cap R$ is a prime ideal in $R$, i.e., $S = R/J$ is an integral domain, and
- **ii.** if $\overline{I}$ is the image of $I$ in $S[x]$ then $\overline{I}F[x]$ is a prime ideal in $F[x]$ satisfying $\overline{I}F[x] \cap S[x] = \overline{I}$.

*Proof:* Suppose $I$ is a prime ideal in $R[x]$, so that $J = I \cap R$ is a prime ideal in $R$ and $S = R/J$ is an integral domain. By Proposition 2 in Chapter 9, the kernel of the reduction homomorphism $R[x] \mapsto S[x] = (R/J)[x]$ is $J[x]$, which is contained in $I[x]$, so we have a ring isomorphism $R[x]/I \cong S[x]/\overline{I}$. Since $R[x]/I$ is an integral domain, it follows that $\overline{I}$ is a prime ideal in the integral domain $S[x]$. The elements of $\overline{I} \cap S$ are the images of the elements in $R \cap I$, so $\overline{I} \cap S = 0$. Since the ring $F[x]$ is the localization of $S[x]$ with respect to the multiplicatively closed set $S - \{0\}$, condition (ii) follows by Proposition 38(3).

Conversely, if $I$ is not prime, then either $J$ is not prime in $R$ or $J$ is prime in $R$ but $\overline{I}$ is not prime in $S[x]$. In the latter case either $\overline{I}F[x]$ is not prime in $F[x]$ or, again

by Proposition 38(3), $\overline{I}$ is not saturated. Thus, if $I$ is not prime, either (i) or (ii) fails, completing the proof.

Since $F[x]$ is a Euclidean Domain, the ideal $\overline{I}F[x] = (h(x))$ in Proposition 39 is principal, and is prime if and only if $h(x)$ is either 0 or is irreducible in $F[x]$. Suppose $h(x)$ is an element in $I$ whose image in $S[x]$ has leading coefficient $a \in S$. The next proposition shows that $a$ gives a bound on the denominators necessary for the saturation $\overline{I}F[x] \cap S[x]$ and can be used to compute this saturation.

**Proposition 40.** Let $S$ be an integral domain with fraction field $F$ and let $A$ be a nonzero ideal in $S[x]$. Suppose $AF[x] = (h(x))$ where $h(x)$ is a polynomial in $S[x]$ with leading coefficient $a \in S$. Let $S_a$ be the localization of $S$ with respect to the powers of $a$. Then

**(1)** $AF[x] \cap S[x] = AS_a[x] \cap S[x]$, and

**(2)** if $\mathcal{A}$ denotes the ideal generated by $A$ and $1 - at$ in the polynomial ring $S[x, t]$, then $AS_a[x] \cap S[x] = \mathcal{A} \cap S[x]$.

*Proof:* We first show $AF[x] \cap S_a[x] = AS_a[x]$. Since $S_a \subseteq F$, the containment $AS_a[x] \subseteq AF[x] \cap S_a[x]$ is immediate. Suppose now that $f(x) \in AF[x] \cap S_a[x]$. If the leading term of $f(x)$ is $sx^N$ and the leading term of $h(x)$ is $ax^m$, then since $AF[x] = (h(x))$ we have $N \geq m$. Then the polynomial $f(x) - (s/a)x^{N-m}h(x)$ is again in $AF[x] \cap S_a[x]$ and is of lower degree than $f(x)$. Iterating, we see that $f(x)$ can be written as a polynomial in $S_a[x]$ times $h(x)$, so $f(x) \in AS_a[x]$. Intersecting both sides of $AF[x] \cap S_a[x] = AS_a[x]$ with $S[x]$ gives the first statement in the proposition.

To prove the second statement, suppose first that $f(x) \in \mathcal{A} \cap S[x]$. Then we can write $f(x) = f_1(x, t)b(x) + f_2(x, t)(1 - at)$ for some polynomials $b(x) \in A$ and $f_1, f_2 \in S[x, t]$. Substituting $t = 1/a$ gives $f(x) = f_1(x, 1/a)b(x)$, and since $f_1(x, 1/a) \in S_a[x]$, we obtain $f(x) \in AS_a[x] \cap S[x]$. Conversely, suppose that $f(x) = b(x)g(x) \in S[x]$ where $g(x) \in S_a(x)$ and $b(x) \in A$. If $a^N$ is the largest power of $a$ appearing in the denominators of the coefficients of $g(x)$ then $a^N g(x) \in S[x]$. Writing $f(x) = (at)^N f(x) + (1 - (at)^N)f(x) = b(x)t^N(a^N g(x)) + (1 - (at)^N)f(x)$ we see that $f(x) \in \mathcal{A} \cap S[x]$, giving the reverse containment and completing the proof.

Suppose now that $P$ is an ideal in $k[x_1, \ldots, x_n]$. Let $P_i$ for $i = 1, \ldots, n$ be the intersection of $P$ with $k[x_1, \ldots, x_i]$. We use Propositions 39 and 40 to determine inductively whether $P_1, P_2, \ldots, P_n = P$ are prime ideals in their respective polynomial rings.

The ideal $P_1$ will be prime in the Euclidean Domain $k[x_1]$ if and only if it is 0 or is generated by an irreducible polynomial. Suppose now that $i \geq 2$ and we have already proved that $P_{i-1}$ is a prime ideal in $k[x_1, \ldots, x_{i-1}]$, so that the quotient ring $S = k[x_1, \ldots, x_{i-1}]/P_{i-1}$ is an integral domain. If $F$ denotes the quotient field of $S$, then by Proposition 39, $P_i$ is a prime ideal in $k[x_1, \ldots, x_i]$ if and only if its image in $(k[x_1, \ldots, x_{i-1}]/P_{i-1})[x_i] = S[x_i]$ is a saturated ideal whose extension to the Euclidean Domain $F[x_i]$ is a prime ideal. Suppose $h(x_i) \in S[x_i]$ is a generator for this ideal and $a$ is the leading coefficient of $h(x_i)$. Then $(h(x_i))$ is a prime ideal in $F[x_i]$ if and only if

$h(x_i) = 0$ or $h(x_i)$ is an irreducible polynomial. By Proposition 40, the image of $P_i$ in $S[x_i]$ will be saturated if and only if it equals $\mathcal{A} \cap S[x_i]$ where $\mathcal{A}$ is the ideal generated by $P_i$ and $1 - at$ in $S[x_i, t]$. This latter condition can be checked in $k[x_1, \ldots, x_i, t]$: it is equivalent to checking that the intersection of the ideal generated by $P_i$ and $1 - at$ in $k[x_1, \ldots, x_i, t]$ with $k[x_1, \ldots, x_i]$ is just $P_i$ (cf. Exercise 3).

Combining these observations with our results on Gröbner bases from Chapter 9 we obtain the following algorithm for determining whether the ideal $P$ in $k[x_1, \ldots, x_n]$ is prime (or, equivalently, whether the associated affine algebraic set is a variety).

## Algorithm for Determining when an Ideal in $k[x_1, \ldots, x_n]$ is Prime

**(1)** Compute the reduced Gröbner basis $G = \{g_1, \ldots, g_m\}$ for $P$ with respect to the lexicographic monomial ordering $x_n > \cdots > x_1$.

By Proposition 29 in Section 9.6 the elements of $G$ lying in $k[x_1, \ldots, x_i]$ will be the reduced Gröbner basis $\{g_1, \ldots, g_{m_i}\}$ for $P_i = P \cap k[x_1, \ldots, x_i]$.

**(2)** Determine whether $P_1$ is a prime ideal in $k[x_1]$ by checking that $P_1 = 0$ or the nonzero generator of $P_1$ is irreducible in $k[x_1]$.

For each $i \geq 2$, suppose $P_{i-1}$ has been determined to be a prime ideal in $k[x_1, \ldots, x_{i-1}]$ (otherwise, $P$ is not a prime ideal in $k[x_1, \ldots, x_n]$). Let $S = k[x_1, \ldots, x_{i-1}]/P_{i-1}$ and let $F$ be the fraction field of $S$. Apply steps (3) and (4) to determine whether $P_i$ is a prime ideal in $k[x_1, \ldots, x_i]$.

**(3)** If $m_i = m_{i-1}$ then $P_i$ maps to the zero ideal in $S[x_i]$, hence is prime. Otherwise the image of $P_i$ in $S[x_i]$ and in $F[x_i]$ is a nonzero ideal, and is generated by the images of $g_{m_{i-1}+1}, \ldots, g_{m_i}$. Apply the Euclidean algorithm in $F[x_i]$ to these generators to find an element $h(x_i)$ in $P_i$ whose image in $F[x_i]$ generates the image of $P_i$ in $F[x_i]$. Determine whether $h(x_i)$ is irreducible in $F[x_i]$—if not then $P_i$ and $P$ are not prime ideals.

(Note that after applying the Euclidean algorithm to the generators of the image of $P_i$ in $F[x_i]$ we can multiply by a single element of $S$ to 'clear denominators' in each equation so that all remainders (and in particular the last nonzero remainder $h(x_i)$) will be elements in the image of $P_i$.)

**(4)** Let $a \in k[x_1, \ldots, x_{i-1}]$ be the leading coefficient of $h(x_i)$ (as a polynomial in $x_i$). Compute the reduced Gröbner basis in $k[x_1, \ldots, x_i, t]$ for the ideal generated by $P_i$ and $1 - at$ with respect to the lexicographic monomial ordering $t > x_i > \cdots > x_1$. Determine whether the elements of this reduced basis that lie in $k[x_1, \ldots, x_i]$ are $\{g_1, \ldots, g_{m_i}\}$—if so, then $P_i$ is a prime ideal in $k[x_1, \ldots, x_i]$ and if not then $P_i$ and $P$ are not prime ideals.

Finally, we note that similar ideas (together with some minor modifications to extend results on Gröbner bases to polynomial rings $R[x_1, \ldots, x_n]$ with coefficients in an integral domain $R$) can be used to provide algorithms for determining when an ideal in, for example, $\mathbb{Z}[x_1, \ldots, x_n]$ is prime.

## Examples

**(1)** Consider the ideal $P = (xz - y^2, yz - x^3, z^2 - x^2y)$ in $k[x, y, z]$ for any infinite field $k$. It follows from Exercise 26 in Section 1 that $P$ is a prime ideal since there is an injection of $k[x, y, z]/P$ into the integral domain $k[\mathbb{A}^1]$ (cf. Exercise 24 in Section 2). Here we prove $P \subset \mathbb{Q}[x, y, z]$ is prime using the ideas in this section. The reduced Gröbner basis for $P$ with respect to the lexicographic monomial ordering $x > y > z$ is $\{x^3 - yz, x^2y - z^2, xy^3 - z^3, xz - y^2, y^5 - z^4\}$. Hence $P_1 = P \cap \mathbb{Q}[z] = (0)$, and $P_2 \cap \mathbb{Q}[y, z] = (y^5 - z^4)$. Since $P_1 = 0$, the ideal $P_1$ is prime in $\mathbb{Q}[z]$.

We next check $P_2$ is prime in $\mathbb{Q}[y, z]$, which can be done directly (cf. Exercise 4 or Exercise 14 in Section 9.1). In this case $S = \mathbb{Q}[z]$ and $F = \mathbb{Q}(z)$. The image of $P_2$ in $F[y]$ is generated by $h(y) = y^5 - z^4$, which is irreducible in $\mathbb{Q}(z)[y]$. The leading coefficient of $h(y)$ is 1, and the reduced Gröbner basis for $(y^5 - z^4, 1 - t)$ in $\mathbb{Q}[y, z, t]$ with respect to the lexicographic monomial ordering $t > y > z$ is $\{y^5 - z^4, 1 - t\}$. The element in the reduced Gröbner basis for $P_2$ is the only element of this basis lying in $\mathbb{Q}[y, z]$ so $P_2$ is a prime ideal in $\mathbb{Q}[y, z]$.

We now use the fact that $P_2$ is prime to prove that $P$ is prime. In this case $S$ is the integral domain $\mathbb{Q}[y, z]/P_2 = \mathbb{Q}[y, z]/(y^5 - z^4)$ with quotient field $F$ given by

$$S = \mathbb{Q}[\bar{z}] + \mathbb{Q}[\bar{z}]\bar{y} + \mathbb{Q}[\bar{z}]\bar{y}^2 + \mathbb{Q}[\bar{z}]\bar{y}^3 + \mathbb{Q}[\bar{z}]\bar{y}^4$$

$$F = \mathbb{Q}(\bar{z}) + \mathbb{Q}(\bar{z})\bar{y} + \mathbb{Q}(\bar{z})\bar{y}^2 + \mathbb{Q}(\bar{z})\bar{y}^3 + \mathbb{Q}(\bar{z})\bar{y}^4$$

where $\bar{y}^5 = \bar{z}^4$. The image of $P$ in $S[x]$ is the ideal $\overline{P}$ generated by the elements $g_1 = x^3 - \bar{y}\bar{z}$, $g_2 = \bar{y}x^2 - \bar{z}^2$, $g_3 = \bar{y}^3x - \bar{z}^3$, $g_4 = \bar{z}x - \bar{y}^2$, and $\bar{y}^5 - \bar{z}^4 = 0$.

The greatest common divisor in $F[x]$ of $g_1, g_2, g_3, g_4$ generating the image of $P$ in $F[x]$ is the irreducible polynomial $x - \bar{y}^2/\bar{z}$. The polynomial $h(x) = zx - y^2$ in $P$ has image generating the same ideal in $F[x]$, so we may take $a = z$ in (4) of the algorithm. The reduced Gröbner basis for $(xz - y^2, yz - x^3, z^2 - x^2y, 1 - zt)$ with respect to the lexicographic monomial ordering $t > x > y > z$ consists of the reduced Gröbner basis for $P$ together with the elements $ty^2 - x$ and $tz - 1$ involving $t$, so $P$ is a prime ideal in $\mathbb{Q}[x, y, z]$.

**(2)** Consider the ideal $P = (xz - y^3, xy - z^2)$ in $\mathbb{Q}[x, y, z]$, with reduced Gröbner basis for the lexicographic monomial ordering $x > y > z$ given by $\{xy - z^2, xz - y^3, y^4 - z^3\}$. Here $P_1 = 0$ and $P_2 = P \cap \mathbb{Q}[y, z] = (y^4 - z^3)$ are prime ideals as in Example 1. In this case $S = \mathbb{Q}[y, z]/P_2$ is given by

$$S = \mathbb{Q}[\bar{z}] + \mathbb{Q}[\bar{z}]\bar{y} + \mathbb{Q}[\bar{z}]\bar{y}^2 + \mathbb{Q}[\bar{z}]\bar{y}^3$$

with $\bar{y}^4 = \bar{z}^3$, with quotient field $F$ similar to the previous example, and $\overline{P} = (g_1, g_2)$ in $S[x]$ where $g_1 = \bar{y}x - \bar{z}^2$ and $g_2 = \bar{z}x - \bar{y}^3$. The extension of $\overline{P}$ to $F[x]$ is generated by the irreducible polynomial $\bar{y}x - \bar{z}^2$, and $h(x) = yx - z^2$ is an element of $P$ having the same image in $F[x]$, with leading coefficient $a = y$. The reduced Gröbner basis for the ideal $(xz - y^3, xy - z^2, 1 - yt)$ in $\mathbb{Q}[x, y, z, t]$ using the lexicographic ordering $t > x > y > z$ is $\{x^2 - y^2z, xy - z^2, xz - y^3, y^4 - z^3, ty - 1, tz^2 - x\}$, containing the element $x^2 - y^2z$ not in the reduced Gröbner basis for $P$, so $P$ is *not* a prime ideal in $\mathbb{Q}[x, y, z]$. This computation not only shows $P$ is not a prime ideal, it does so by explicitly showing the image of $P$ in $S[x]$ is not saturated using the localization $S_a$. The computation of $a = y$ allows us to find an explicit pair of elements not in $P$ whose product is in $P$: $f = x^2 - y^2z \notin P$ and $y \notin P$, but some power of $y$ times $f$ lies in $P$. In this case a quick computation verifies that $yf \in P$.

## Localizations of Modules

Suppose now that $M$ is an $R$-module and $D$ is a multiplicatively closed subset of $R$ containing 1 as above. Then the ideas used in the construction of $D^{-1}R$ can be used to construct a $D^{-1}R$-module $D^{-1}M$ from $M$ in a similar fashion, as follows. Define the relation on $D \times M$ by

$$(d, m) \sim (e, n) \quad \text{if and only if} \quad x(dn - em) = 0 \quad \text{for some } x \in D,$$

which is easily checked to be an equivalence relation. Let $m/d$ denote the equivalence class of $(d, m)$ and let $D^{-1}M$ denote the set of equivalence classes. It is then straightforward to verify that the operations

$$\frac{m}{d} + \frac{n}{e} = \frac{em + dn}{de} \qquad \text{and} \qquad \left(\frac{r}{d}\right)\left(\frac{m}{e}\right) = \frac{rm}{de}$$

are well defined and give $D^{-1}M$ the structure of a $D^{-1}R$-module.

**Definition.** The $D^{-1}R$-module $D^{-1}M$ is called the *module of fractions of $M$ with respect to $D$* or the *localization of $M$ at $D$*.

Note that the localization $D^{-1}M$ is also an $R$-module (since each $r \in R$ acts by $r/1$ on $D^{-1}M$), and there is an $R$-module homomorphism

$$\pi : M \to D^{-1}M \quad \text{defined by} \quad \pi(m) = \frac{m}{1}.$$

It follows directly from the definition of the equivalence relation that

$$\ker \pi = \{m \in M \mid dm = 0 \text{ for some } d \in D\}.$$

The homomorphism $\pi$ has a universal property analogous to that in Theorem 36. Suppose $N$ is an $R$-module with the property that left multiplication on $N$ by $d$ is a bijection of $N$ for every $d \in D$. If $\psi : M \to N$ is any $R$-module homomorphism then there is a unique $R$-module homomorphism $\Psi : D^{-1}M \to N$ such that $\Psi \circ \pi = \psi$.

If $M$ and $N$ are $R$-modules and $\varphi : M \to N$ is an $R$-module homomorphism, then for any multiplicative set $D$ in $R$ it is easy to check that there is an induced $D^{-1}R$-module homomorphism from $D^{-1}M$ to $D^{-1}N$ defined by mapping $m/d$ to $\varphi(m)/d$.

The next result shows that the localization of $M$ at $D$ is related to the tensor product.

**Proposition 41.** Let $D$ be a multiplicatively closed subset of $R$ containing 1 and let $M$ be an $R$-module. Then $D^{-1}M \cong D^{-1}R \otimes_R M$ as $D^{-1}R$-modules, i.e., $D^{-1}M$ is the $D^{-1}R$-module obtained by extension of scalars from the $R$-module $M$.

*Proof:* The map from $D^{-1}R \times M$ to $D^{-1}M$ defined by mapping $(r/d, m)$ to $rm/d$ is well defined and $R$-balanced, so induces a homomorphism from $D^{-1}R \otimes_R M$ to $D^{-1}M$. The map sending $m/d$ to $(1/d) \otimes m$ gives a well defined inverse homomorphism (if $m/d = m'/d'$ in $D^{-1}M$ then $x(d'm - dm') = 0$ for some $x \in D$, and then $(1/d) \otimes m$ can be written as $(1/xd'd) \otimes (xd'm) = (1/xd'd) \otimes (xdm') = (1/d') \otimes m'$). Hence $D^{-1}M$ is isomorphic to $D^{-1}R \otimes_R M$ as an $R$-module since these inverse isomorphisms are also $D^{-1}R$-module homomorphisms.

Localizing a ring $R$ or an $R$-module $M$ at $D$ behaves very well with respect to algebraic operations on rings and modules, as the following proposition shows:

**Proposition 42.** Let $R$ be a commutative ring with 1 and let $D^{-1}R$ be its localization with respect to the multiplicatively closed subset $D$ of $R$ containing 1.

**(1)** Localization commutes with finite sums and intersections of ideals: If $I$ and $J$ are ideals of $R$, then

$$D^{-1}(I+J) = D^{-1}(I) + D^{-1}(J) \quad \text{and} \quad D^{-1}(I \cap J) = D^{-1}(I) \cap D^{-1}(J).$$

Localization commutes with quotients:

$$D^{-1}R \,/\, D^{-1}I \cong D^{-1}(R/I),$$

(where the localization on the right is with respect to the image of $D$ in the quotient $R/I$).

**(2)** Localization commutes with taking radicals: If $N$ is the nilradical of $R$, then $D^{-1}N$ is the nilradical of $D^{-1}R$. If $I$ is an ideal in $R$, then $\text{rad}(D^{-1}I)$ is $D^{-1}(\text{rad}\, I)$.

**(3)** Primary ideals correspond to primary ideals in the correspondence (3) of Proposition 38. More precisely, suppose $Q$ is a $P$-primary ideal in $R$. If $D \cap P \neq \emptyset$ then $D^{-1}Q = D^{-1}R$. If $D \cap P = \emptyset$ then $D^{-1}P$ is a prime ideal, the extension $D^{-1}Q$ of $Q$ is a $D^{-1}P$-primary ideal in $D^{-1}R$, and the contraction back to $R$ of $D^{-1}Q$ is $Q$.

**(4)** Localization commutes with finite sums, intersections and quotients of modules: If $L$ and $N$ are submodules of the $R$-module $M$, then

**(a)** $D^{-1}(L+N) = D^{-1}L + D^{-1}N$ and $D^{-1}(L \cap N) = D^{-1}L \cap D^{-1}N$,

**(b)** $D^{-1}N$ is a submodule of $D^{-1}M$ and $D^{-1}M \,/\, D^{-1}N = D^{-1}(M/N)$.

**(5)** Localization commutes with finite direct sums of modules: If $M$ and $N$ are $R$-modules, then $D^{-1}(M \oplus N) \cong D^{-1}M \oplus D^{-1}N$.

**(6)** Localization is exact (i.e., $D^{-1}R$ is a flat $R$-module): If $0 \to L \xrightarrow{\psi} M \xrightarrow{\varphi} N \to 0$ is a short exact sequence of $R$-modules, then the induced sequence $0 \to D^{-1}L \xrightarrow{\psi'} D^{-1}M \xrightarrow{\varphi'} D^{-1}N \to 0$ of $D^{-1}R$-modules is also exact.

*Proof:* We first prove (6). Suppose that $0 \to L \xrightarrow{\psi} M \xrightarrow{\varphi} N \to 0$ is a short exact sequence of $R$-modules. Every element of $D^{-1}N$ is of the form $n/d$ for some $n \in N$ and $d \in D$. Since $\varphi$ is surjective, $n = \varphi(m)$ for some $m \in M$, so $\varphi'(m/d) = \varphi(m)/d = n/d$ and $\varphi' : D^{-1}M \to D^{-1}N$ is surjective. If $m/d$ is in the kernel of $\varphi'$ then $d_1\varphi(m) = 0$ for some $d_1 \in D$. Then $\varphi(d_1 m) = 0$ implies $d_1 m = \psi(l)$ for some $l \in L$ by the exactness of the original sequence at $M$, so $m/d = d_1 m/(d_1 d) = \psi(l)/(d_1 d) = \psi'(l/(d_1 d))$ and $\ker(\varphi') \subseteq \text{image}(\psi')$. If $\psi(l)/d \in \text{image}(\psi')$ then $\varphi'(\psi(l)/d) = \varphi(\psi(l))/d = 0$, which shows the reverse inclusion $\text{image}(\psi') \subseteq \ker(\varphi')$, and we have exactness of the induced sequence at $D^{-1}M$. Finally, suppose $\psi'(l/d) = 0$. Then $d_2\psi(l) = 0$ for some $d_2 \in D$, i.e., $\psi(d_2 l) = 0$, so $d_2 l = 0$ by the injectivity of $\psi$. Hence $l/d = d_2 l/(d_2 d) = 0$ and $\psi'$ is injective. This proves that the sequence $0 \to D^{-1}L \xrightarrow{\psi'} D^{-1}M \xrightarrow{\varphi'} D^{-1}N \to 0$ is exact.

To prove the first statement in (1), note that $(i+j)/d = i/d + j/d$ for $i \in I$, $j \in J$ and $d \in D$ shows $D^{-1}(I+J) \subseteq D^{-1}(I) + D^{-1}(J)$; and $i/d_1 + j/d_2 = (d_2 i + d_1 j)/(d_1 d_2)$ for $i \in I$, $j \in J$ and $d_1, d_2 \in D$ shows $D^{-1}(I) + D^{-1}(J) \subseteq D^{-1}(I+J)$. For the second statement, the inclusion $D^{-1}(I \cap J) \subseteq D^{-1}(I) \cap D^{-1}(J)$ is immediate. If

$a/d \in D^{-1}(I) \cap D^{-1}(J)$ then $d_1 a \in I$ and $d_2 a \in J$ for some $d_1, d_2 \in D$. Then $d_1 d_2 a \in I \cap J$ and $a/d = (d_1 d_2 a)/(d_1 d_2 d)$ gives the inclusion $D^{-1}(I) \cap D^{-1}(J) \subseteq D^{-1}(I \cap J)$. The last statement in (1) follows by applying (6) to the exact sequence $0 \to I \xrightarrow{\psi} R \xrightarrow{\varphi} R/I \to 0$.

To prove (2), suppose first that $a \in \operatorname{rad} I$, so that $a^n \in I$ for some $n \geq 1$. Then $(a/d)^n = a^n/d^n \in D^{-1}I$ so $D^{-1}(\operatorname{rad} I) \subseteq \operatorname{rad}(D^{-1}I)$. Conversely, if $a/d \in \operatorname{rad}(D^{-1}I)$ then $(a/d)^n \in D^{-1}I$ for some $n \geq 1$, i.e., $d_1 a^n \in I$ for some $d_1 \in D$. Hence $(d_1 a)^n = d_1^{n-1}(d_1 a^n) \in I$, so $d_1 a \in \operatorname{rad} I$ and then $a/d = d_1 a/(d_1 d) \in D^{-1}(\operatorname{rad} I)$ shows that $\operatorname{rad}(D^{-1}I) \subseteq D^{-1}(\operatorname{rad} I)$. This proves the second statement in (2), and the first statement follows by applying this to the ideal $I = (0)$.

For (3), note first that $D \cap P = \emptyset$ if and only if $D \cap Q = \emptyset$ (one inclusion is obvious and the other follows since $d \in D \cap P$ implies $d^n \in D \cap Q$ for some $n$). The statement for $D \cap P \neq \emptyset$ and the fact that $D^{-1}P$ is a prime ideal for $D \cap P = \emptyset$ were proved in Proposition 38. To see that $D^{-1}Q$ is a primary ideal in $D^{-1}R$, suppose that $(a/d_1)(b/d_2) \in D^{-1}Q$ and $a/d_1 \notin D^{-1}Q$. Then there is some element $d \in D$ so that $dab \in Q$, and since $a \notin Q$ and $Q$ is primary, we have $(db)^n \in Q$ for some $n \geq 1$. Then $(b/d_2)^n = d^n b^n/(d^n d_2^n) \in D^{-1}Q$, so that $D^{-1}Q$ is primary. The radical of $D^{-1}Q$ is $D^{-1}P$ by (2). Finally, by (2) of Proposition 38, the contraction of $D^{-1}Q$ is an ideal of $R$ containing $Q$ and consists precisely of the elements $r \in R$ with $dr \in Q$ for some $d \in D$. Since $Q$ is $P$-primary, the definition of primary implies that if $dr \in Q$ and $d \notin P$, then $r \in Q$, hence the contraction of $D^{-1}Q$ is $Q$.

The proof of (4) is essentially the same as the proof of (1) and is left as an exercise.

It is easy to see that if the exact sequence $0 \to L \xrightarrow{\psi} M \xrightarrow{\varphi} N \to 0$ of $R$-modules splits, then the exact sequence $0 \to D^{-1}L \xrightarrow{\psi'} D^{-1}M \xrightarrow{\varphi'} D^{-1}N \to 0$ of $D^{-1}R$-modules also splits, which gives (5).

Proposition 38 shows that localizing at the multiplicatively closed set $D$ emphasizes the ideals of $R$ not containing any elements of $D$ since the other ideals of $R$ become trivial when extended to $D^{-1}R$. The following proposition provides a more precise statement in terms of the effect of localization on primary decomposition of ideals.

**Proposition 43.** Let $R$ be a Noetherian ring and let

$$I = Q_1 \cap \cdots \cap Q_m$$

be a minimal primary decomposition of the proper ideal $I$, where $Q_i$ is a $P_i$-primary ideal. Suppose $D$ is a multiplicatively closed set of $R$ containing 1 and the primary ideals $Q_1, \ldots, Q_m$ are numbered so that $D \cap P_i = \emptyset$ for $1 \leq i \leq t$ and $D \cap P_i \neq \emptyset$ for $t+1 \leq i \leq m$. Then

$$D^{-1}I = D^{-1}Q_1 \cap \cdots \cap D^{-1}Q_t$$

is a minimal primary decomposition of $D^{-1}I$ in $D^{-1}R$ and $D^{-1}Q_i$ is a $D^{-1}P_i$-primary ideal. Further, the contraction of $D^{-1}Q_i$ back to $R$ is $Q_i$ for $1 \leq i \leq t$ and

$$^c(D^{-1}I) = Q_1 \cap \cdots \cap Q_t$$

is a minimal primary decomposition of the contraction of $D^{-1}I$ back to $R$.

*Proof:* By (3) of Proposition 42, $D^{-1}Q_i = D^{-1}R$ for $t+1 \leq i \leq m$, and $D^{-1}Q_i$ is a $D^{-1}P_i$-primary ideal with pullback $Q_i$ for $1 \leq i \leq t$. By (1) of the same proposition, $D^{-1}I = D^{-1}Q_1 \cap \cdots \cap D^{-1}Q_t$, and (3) shows that this is a primary decomposition. Contracting to $R$ shows that $^c(D^{-1}I) = Q_1 \cap \cdots \cap Q_t$, which also implies that the decompositions are minimal.

In particular we can finish the proof of Theorem 21:

**Corollary 44.** The primary ideals belonging to the isolated primes in a minimal primary decomposition of $I$ are uniquely defined by $I$.

*Proof:* Let $P$ be a minimal element in the set $\{P_1, \ldots, P_m\}$ of primes belonging to $I$, and take $D = R - P$ in Proposition 43. Then $D \cap P_i = \emptyset$ only for $P = P_i$, so the contraction of the localization of $I$ at $D$ is precisely the primary ideal $Q$ belonging to the minimal prime $P$. Since the prime ideals $\{P_1, \ldots, P_m\}$ of primes belonging to $I$ are uniquely determined by $I$, it follows that the primary ideals $Q$ belonging to the isolated primes of $I$ are also uniquely determined by $I$.

The effect of isolating in on certain prime ideals by localization is particularly precise in the case of localizing at a prime $P$ (considered in Example 3 following Corollary 37 above). We first recall the definition of an important type of ring (cf. Exercises 37–39 in Section 7.4).

**Definition.** A commutative ring with 1 that has a unique maximal ideal is called a *local ring.*

**Proposition 45.** Let $R$ be a commutative ring with 1. Then the following are equivalent:

**(1)** $R$ is a local ring with unique maximal ideal $M$

**(2)** if $M$ is the set of elements of $R$ that are not units, then $M$ is an ideal

**(3)** there is a maximal ideal $M$ of $R$ such that every element $1 + m$ with $m \in M$ is a unit in $R$.

*Proof:* If $a \in R$ then the ideal $(a)$ is either $R$, in which case $a$ is a unit, or is a proper ideal, in which case $(a)$ is contained in a maximal ideal (Proposition 11 of Section 7.4). It follows that if $R$ is a local ring and $M$ is its unique maximal ideal then every $a \notin M$ is a unit, so $M$ consists precisely of the set of nonunits in $R$, showing that (1) implies (2). It also follows that if the set $M$ of nonunits in $R$ is an ideal then this ideal must be the unique maximal ideal in $R$, so that (2) implies (1).

Suppose now that (3) is satisfied. If $a$ is an element of $R$ not contained in the maximal ideal $M$, then $(a) + M = R$, so that $ab + m = 1$ for some $b \in R$ and $m \in M$. Then $ab = 1 - m$ is a unit by assumption, so $a$ is also a unit. This shows that $M$ is the unique maximal ideal in $R$, so (3) implies (1). Conversely, if $R$ is a local ring, then $1 + m \notin M$ for any $m \in M$, so $1 + m$ is a unit, so (1) implies (3).

**Proposition 46.** For any commutative ring $R$ with 1, let $R_P$ be the localization of $R$ at the prime ideal $P$ and let ${}^eP$ be the extension of $P$ to $R_P$.

**(1)** The ring $R_P$ is a local ring with unique maximal ideal ${}^eP$. The contraction of ${}^eP$ to $R$ is $P$, i.e., ${}^c({}^eP) = P$, and the map from $R$ to $R_P$ induces an injection of the integral domain $R/P$ into $R_P/{}^eP$. The quotient $R_P/{}^eP$ is a field and is isomorphic to the fraction field of the integral domain $R/P$.

**(2)** If $R$ is an integral domain, then $R_P$ is an integral domain. The ring $R$ injects into the local ring $R_P$, and, identifying $R$ with its image in $R_P$, the unique maximal ideal of $R_P$ is $PR_P$.

**(3)** The prime ideals in $R_P$ are in bijective correspondence with the prime ideals of $R$ contained in $P$.

**(4)** If $P$ is a minimal nonzero prime ideal of $R$ then $R_P$ has a unique nonzero prime ideal.

**(5)** If $P = M$ is a maximal ideal and $I$ is any $M$-primary ideal of $R$ then $R_M/{}^eI \cong R/I$. In particular, $R_M/{}^eM \cong R/M$ and $({}^eM)/({}^eM)^n \cong M/M^n$ for all $n \geq 1$.

*Proof:* If $P'$ is a prime ideal of $R$, then $P' \cap (R - P) = \emptyset$ if and only if $P' \subseteq P$, so (3) is immediate from (3) in Proposition 38, and (4) follows. Since ${}^eP \neq R_P$ by (2) of Proposition 38, it follows from (3) that $R_P$ is a local ring with unique maximal ideal ${}^eP$, which proves the first statement in (1).

By Proposition 38(2) the contraction ${}^c({}^eP)$ is the set $\{r \in R \mid dr \in P \text{ for some } d \in R - P\}$, and since $P$ is prime, $dr \in P$ with $d \notin P$ implies $r \in P$. This shows that ${}^c({}^eP) = P$, which is the second statement in (1).

The kernel of the map from $R$ to $R_P/{}^eP$ is ${}^c({}^eP) = P$, so the induced map from $R/P$ into $R_P/{}^eP$ is injective. The quotient $R_P/{}^eP$ is a field by the first part of (1), so there is an induced homomorphism from the fraction field of the integral domain $R/P$ into $R_P/{}^eP$. The universal property of the localization $R_P$ shows there is an inverse homomorphism from $R_P/{}^eP$ to the fraction field of $R/P$ (since every element of $R$ not in $P$ maps to a unit in $R/P$). It follows that $R_P/{}^eP$ is isomorphic to the fraction field of $R/P$.

If $R$ is an integral domain, then $R - P$ has no zero divisors, so $R$ injects into $R_P$ by Corollary 37; if $R$ is identified with its image in $R_P$ then ${}^eP = PR_P$, so (2) follows.

To prove (5), by Proposition 42(1) we may pass to the quotient $R/I$ and so reduce to the case $I = 0$. In this case the maximal ideal $P = M$ in $R$ is the nilradical of $R$, hence is the unique maximal ideal of $R$. By Proposition 45 every element of $R - M$ is a unit, so $R_P = R$, and each of the statements in (5) follows immediately, completing the proof of the proposition.

### Example

The results of (5) of the proposition are not true in general if $P$ is a prime ideal that is not maximal. For example, $P = (0)$ in $R = \mathbb{Z}$ has $R/P = \mathbb{Z}$ and $R_P/PR_P = \mathbb{Q}$; in this case $(PR_P)/(PR_P)^n \cong P/P^n = 0$ for all $n \geq 1$ (cf. the exercises).

**Definition.** Let $M$ be an $R$-module, let $P$ be a prime ideal of $R$ and set $D = R - P$. The $R_P$-module $D^{-1}M$ is called the *localization of $M$ at $P$*, and is denoted by $M_P$.

By Proposition 41, $M_P$ can also be identified with the tensor product $R_P \otimes_R M$. When $R$ is an integral domain and $P = (0)$, then $M_{(0)}$ is a module over the field of fractions $F$ of $R$, i.e., is a vector space over $F$.

The element $m/1$ is zero in $M_P$ if and only if $rm = 0$ for some $r \in R - P$, so localizing at $P$ annihilates the $P'$-torsion elements of $M$ for primes $P'$ not contained in $P$. In particular, *localizing at (0) over an integral domain annihilates the torsion subgroup of $M$.*

**Definition.** If $R$ is an integral domain, then the *rank* of the $R$-module $M$ is the dimension of the localization $M_{(0)}$ as a vector space over the field of fractions of $R$.

It is easy to see that this definition of rank agrees with the notion of rank introduced in Chapter 12.

**Example**

Let $R = \mathbb{Z}$ and let $\mathbb{Z}_{(p)}$ be the localization of $\mathbb{Z}$ at the nonzero prime ideal $(p)$. Any abelian group $M$ is a $\mathbb{Z}$-module so we may localize $M$ at $(p)$ by forming $M_{(p)}$. This abelian group is the same as the quotient of $M$ with respect to the subgroup of elements whose order is finite and not divisible by $p$. If $M$ is a finite (or, more generally, torsion) abelian group, then $M_{(p)}$ is a $p$-group, and is the Sylow $p$-subgroup or $p$-primary component of $M$. The localization $M_{(0)}$ of $M$ at (0) is the trivial group. For a specific example, let $M = \mathbb{Z}/6\mathbb{Z}$ be the cyclic group of order 6, considered as a $\mathbb{Z}$-module. Then the localization of $M$ at $p = 2$ is $\mathbb{Z}/2\mathbb{Z}$, at $p = 3$ is $\mathbb{Z}/3\mathbb{Z}$, and reduces to 0 at all other prime ideals of $\mathbb{Z}$.

Localization of a module $M$ at a prime $P$ in general produces a simpler module $M_P$ whose properties are easier to determine. It is then of interest to translate these "local" properties of $M_P$ back into "global" information about the module $M$ itself. For example, the most basic question of whether a module $M$ is 0 can be answered locally:

**Proposition 47.** Let $M$ be an $R$-module. Then the following are equivalent:

**(1)** $M = 0$,

**(2)** $M_P = 0$ for all prime ideals $P$ of $R$, and

**(3)** $M_{\mathfrak{m}} = 0$ for all maximal ideals $\mathfrak{m}$ of $R$.

*Proof:* The implications (1) implies (2) implies (3) are obvious, so it remains to prove that (3) implies (1). Suppose $m$ is a nonzero element in $M$, and consider the annihilator $I$ of $m$ in $R$, i.e., the ideal of elements $r \in R$ with $rm = 0$. Since $m$ is nonzero $I$ is a proper ideal in $R$. Let $\mathfrak{m}$ be a maximal ideal of $R$ containing $I$ and consider the element $m/1$ in the corresponding localization $M_{\mathfrak{m}}$ of $M$. If this element were 0, then $rm = 0$ for some $r \in R - \mathfrak{m}$. But then $r$ would be an element in $I$ not contained in $\mathfrak{m}$, a contradiction. Hence $M_{\mathfrak{m}} \neq 0$, which proves that (3) implies (1).

It is not in general true that a property shared by all of the localizations of a module $M$ is also shared by $M$. For example, all of the localizations of a ring $R$ can be integral domains without $R$ itself being an integral domain (for example, $\mathbb{Z}/6\mathbb{Z}$ above). Nevertheless, a great deal of information *can* be ascertained from studying the various possible localizations, and this is what makes this technique so useful. If $R$ is an integral

domain, for example, then each of the localizations $R_P$ can be considered as a subring of the fraction field $F$ of $R$ that contains $R$; the next proposition shows that the elements of $R$ are the only elements of $F$ contained in every localization.

**Proposition 48.** Let $R$ be an integral domain. Then $R$ is the intersection of the localizations of $R$: $R = \cap_P R_P$. In fact, $R = \cap_{\mathfrak{m}} R_{\mathfrak{m}}$ is the intersection of the localizations of $R$ at the maximal ideals $\mathfrak{m}$ of $R$.

*Proof:* As mentioned, $R \subseteq \cap_{\mathfrak{m}} R_{\mathfrak{m}}$. Suppose now that $a$ is an element of the fraction field $F$ of $R$ that is contained in $R_{\mathfrak{m}}$ for every maximal ideal $\mathfrak{m}$ of $R$, and consider

$$I_a = \{d \in R \mid da \in R\}.$$

It is easy to check that $I$ is an ideal of $R$, and that $a \in R$ if and only if $1 \in I_a$, i.e., $I_a = R$. Suppose that $I_a \neq R$. Then there is a maximal ideal $\mathfrak{m}$ containing $I_a$, and since $a \in R_{\mathfrak{m}}$ we have $a = r/d$ for some $r \in R$ and $d \in R - \mathfrak{m}$. But then $d \in I_a$ and $d \notin \mathfrak{m}$, a contradiction. Hence $a \in R$, so $\cap_{\mathfrak{m}} R_{\mathfrak{m}} \subseteq R$, and we have proved the second assertion in the proposition. The first is then immediate.

Another important property of a ring $R$ that can be detected locally is normality:

**Proposition 49.** Let $R$ be an integral domain. Then the following are equivalent:

**(1)** $R$ is normal, i.e., $R$ is integrally closed (in its field of fractions)

**(2)** $R_P$ is normal for all prime ideals $P$ of $R$

**(3)** $R_{\mathfrak{m}}$ is normal for all maximal ideals $\mathfrak{m}$ of $R$.

*Proof:* Let $F$ be the field of fractions of $R$, so all of the various localizations of $R$ may be considered as subrings of $F$.

Assume first that $R$ is integrally closed and suppose $y \in F$ is integral over $R_P$. Then $y$ is a root of a monic polynomial of degree $n$ with coefficients of the form $a_i/d_i$ for some $d_i \notin P$. The element $y' = y(d_0 d_1 \cdots d_{n-1})^n$ is then a root of a monic polynomial of degree $n$ with coefficients from $R$, i.e., $y'$ is integral over $R$. Since $R$ is assumed normal, this implies $y' \in R$, and so $y = y'/(d_0 \cdots d_{n-1}) \in R_P$, which proves that (1) implies (2). The implication (2) implies (3) is trivial. Suppose now that $R_{\mathfrak{m}}$ is normal for all maximal ideals $\mathfrak{m}$ of $R$ and let $y$ be an element of $F$ that is integral over $R$. Since $R \subseteq R_{\mathfrak{m}}$, $y$ is in particular also integral over $R_{\mathfrak{m}}$ and so $y \in R_{\mathfrak{m}}$ for every maximal ideal by assumption. Then $y \in R$ by the previous proposition, which proves that (3) implies (1).

We now may easily prove the first part of the Going–up Theorem (cf. Section 3) that was used in the proof of Corollary 27.

**Corollary 50.** Let $R$ be a subring of the commutative ring $S$ with $1 \in R$, and assume that $S$ is integral over $R$. If $P$ is a prime ideal in $R$, then there is a prime ideal $Q$ of $S$ with $P = Q \cap R$.

*Proof:* Let $D = R - P$ so that $D$ is a multiplicatively closed subset of both $R$ and $S$. Then the following diagram commutes:

$$\begin{array}{ccc} R & \xrightarrow{\pi} & D^{-1}R = R_P \\ \iota\downarrow & & \downarrow\iota \\ S & \xrightarrow{\pi} & D^{-1}S \end{array}$$

where the vertical maps are inclusions. It is easy to see that $D^{-1}S$ is integral over $R_P$ (Exercise 20). Let $\mathfrak{m}$ be any maximal ideal of $D^{-1}S$. Then $\mathfrak{m} \cap R_P$ is a maximal ideal in $R_P$ by the second statement in Theorem 26(2) (note that the first part of Theorem 26(2) was not used in the proof of the second statement). By Proposition 38(1), $\mathfrak{m} \cap R_P$ is the extension of $P$ to the local ring $R_P$, and the contraction of this ideal to $R$ is just $P$. Put another way, the preimage of $\mathfrak{m}$ by the maps along the top and right of the diagram above is $P$. If $Q \subset S$ denotes the preimage of $\mathfrak{m}$ by the map along the bottom of the diagram, then $Q$ is a prime ideal by Proposition 38(3). Since $Q \cap R$ is the pullback of $Q$ by the map along the left of the diagram above, the commutativity of the diagram shows that $Q \cap R = P$.

## Local Rings of Affine Algebraic Varieties

For the remainder of this section, let $k$ be an algebraically closed field and let $V$ be an affine variety over $k$ with coordinate ring $k[V]$. Then $k[V]$ is an integral domain, so we may form its field of fractions:

$$k(V) = \{f/g \mid f, g \in k[V],\ g \neq 0\}.$$

The elements of $k(V)$ are called *rational functions* on $V$ and $k(V)$ is called the *field of rational functions* on $V$. When $k[V]$ is a Unique Factorization Domain there is an essentially unique representative for $f/g$ that is in "lowest terms," but in general each fraction $f/g \in k(V)$ has many representations as a ratio of two elements of $k[V]$. Since $k[V]$ is an integral domain, $f/g = f_1/g_1$ if and only if $fg_1 = f_1g$.

The elements of $k[V]$ can be considered as $k$-valued functions on $V$, and if the denominator doesn't vanish the same is true for an element of $k(V)$ (which helps to explain the terminology for this field). Since the same element of $k(V)$ may be written in the form $f/g$ in several ways, we make the following definition:

**Definition.** We say $f/g$ is *regular at* $v$ or *defined at the point* $v \in V$ if there is some $f_1, g_1 \in k[V]$ with $f/g = f_1/g_1$ and $g_1(v) \neq 0$.

If $f_2, g_2$ is another such pair with $g_2(v) \neq 0$, then $f_1(v)/g_1(v) = f_2(v)/g_2(v)$ as elements of $k$, so whenever $f/g$ is regular at $v$ there is a well defined way of specifying its value in $k$ at $v$.

**Example**

The variety $V = \mathcal{Z}(xz - yw)$ in $\mathbb{A}^4$ has coordinate ring $k[V] = k[x, y, z, w]/(xz - yw)$. Consider the element $f = \bar{x}/\bar{y}$ in the quotient field $k(V)$ of $k[V]$. Since $\bar{x}\bar{z} = \bar{y}\bar{w}$ in $k[V]$, the element $f$ can also be written as $\bar{w}/\bar{z}$. From the first expression for $f$ it follows that $f$

is regular at all points of $V$ where $\bar{y} \neq 0$, and from the second expression it follows that $f$ is regular at all points of $V$ where $\bar{z} \neq 0$. It is not too difficult to show that these are all the points of $V$ where $f$ is regular. Furthermore, there is no single expression $f = a/b$ for $f$ with $a, b \in k[V]$ such that $b(v) \neq 0$ for every $v$ where $f$ is regular (cf. Exercise 25).

If $f/g \in k(V)$ is regular at the point $v$, say $f/g = f_1/g_1$ with $g_1(v) \neq 0$, then $f/g$ is also regular at all the points $v$ in the Zariski open neighborhood $V_{g_1}$ of $v$ where $g_1 \neq 0$. As a $k$-valued function on $V$ this means that if $f/g$ is defined at $v$, then it is also defined in a (Zariski open) neighborhood of $v$. Since any nonempty open set of an affine variety is Zariski dense (cf. Exercise 11 in Section 2), we see that every rational function on $V$ is defined at a dense set of points in $V$ (so "almost everywhere" in a suitable sense). Also, each pair $f_1/g_1$ and $f_2/g_2$ representing $f/g$ agree as functions on the open neighborhood $V_{g_1} \cap V_{g_2}$ of $v$, but the "size" of this neighborhood depends on $g_1$ and $g_2$ — there is in general not a common open neighborhood of $v$ where *all* representatives of $f/g$ with nonzero denominator at $v$ are simultaneously defined.

If $v$ is a fixed point in $V$, then a rational function $f/g$ is regular at $v$ if and only if $f/g = f_1/g_1$ for some $f_1, g_1 \in k[V]$ with $g_1 \notin \mathcal{I}(v)$, the ideal of functions on $V$ that are zero at $v$. This means that the set of rational functions that are defined at $v$ is the same as the localization of $k[V]$ at the maximal ideal $\mathcal{I}(v)$:

**Definition.** For each point $v \in V$ the collection of rational functions on $V$ that are defined at $v$,

$$\mathcal{O}_{v,V} = \{f/g \in k(V) \mid f/g \text{ is regular at } v\},$$

is called the *local ring of* $V$ *at* $v$. Equivalently, the local ring of $V$ at $v$ is the localization of $k[V]$ at the maximal ideal $\mathcal{I}(v)$.

In particular, $\mathcal{O}_{v,V}$ is a local ring with unique maximal ideal $\mathfrak{m}_{v,V}$, where

$$\mathfrak{m}_{v,V} = \{f/g \in \mathcal{O}_{v,V} \mid f/g = f_1/g_1 \text{ with } f_1(v) = 0,\ g_1(v) \neq 0\}$$

is the set of rational functions on $V$ that are defined and equal to 0 at $v$. Since $\mathcal{O}_{v,V}$ is a localization of the Noetherian integral domain $k[V]$ at a prime ideal, $\mathcal{O}_{v,V}$ is also a Noetherian integral domain. Note also that $\mathcal{O}_{v,V}/\mathfrak{m}_{v,V} \cong k[V]/\mathcal{I}(v) \cong k$ by Proposition 46(5).

Recall that the polynomial maps from $V$ to $k$ are also referred to as the *regular* maps of $V$ to $k$. This is because these are precisely the rational functions on $V$ that are regular everywhere:

**Proposition 51.** If $V$ is an affine variety over an algebraically closed field $k$ then the rational functions on $V$ that are regular at all points of $V$ are precisely the polynomial functions $k[V]$.

*Proof:* This follows from Proposition 48, which shows that the intersection (in $k(V)$) of all of the localizations of $k[V]$ at the maximal ideals of $k[V]$ is precisely $k[V]$.

Since the maximal ideals of $k[V]$ are in bijective correspondence with the points of $V$, the fact that the local ring $\mathcal{O}_{v,V}$ is the same as the localization of $k[V]$ at the maximal ideal corresponding to $v$ shows that $\mathcal{O}_{v,V}$ depends intrinsically on the ring $k[V]$ and is independent of the embedding of $V$ in a particular affine space.

Suppose $\varphi : V \to W$ is a morphism of affine varieties with associated $k$-algebra homomorphism $\widetilde{\varphi} : k[W] \to k[V]$. If $v \in V$ is mapped to $w \in W$ by $\varphi$, then it is straightforward to show that $\widetilde{\varphi}$ induces a homomorphism (also denoted by $\widetilde{\varphi}$) between the corresponding local rings:

$$\widetilde{\varphi} : \mathcal{O}_{w,W} \to \mathcal{O}_{v,V} \quad \text{where} \quad \widetilde{\varphi}(h/k) = \widetilde{\varphi}(h)/\widetilde{\varphi}(k),$$

and that under this homomorphism, $\widetilde{\varphi}^{-1}(\mathfrak{m}_{v,V}) = \mathfrak{m}_{w,W}$ (a homomorphism of local rings having this property is called a *local homomorphism*). Note that $\widetilde{\varphi}$ does not in general extend to a field homomorphism from *all* of $k(W)$ into $k(V)$ since elements of $k[W]$ lying in the kernel of $\widetilde{\varphi}$ do not map to invertible elements in $k(V)$. It is also easy to check that if $\psi \circ \varphi$ is a composition of morphisms then on the local rings $\widetilde{\psi \circ \varphi} = \widetilde{\varphi} \circ \widetilde{\psi}$.

The local ring $\mathcal{O}_{v,V}$ can be used to provide an algebraic definition of the "smoothness" (in the sense of the existence of tangents) of $V$ at $v$, as we now indicate. Suppose first that $V = \mathcal{Z}(f)$ is the hypersurface variety in $\mathbb{A}^n$ defined by the zeros of an irreducible polynomial $f$ in $k[x_1, \ldots, x_n]$. For any point $v = (v_1, \ldots, v_n)$ on $V$ let $D_v(f)(x_1, \ldots, x_n)$ be the linear polynomial:

$$D_v(f)(x_1, \ldots, x_n) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(v)\, x_i,$$

where the partial derivative of $f$ with respect to $x_i$ is given by the usual formal rule for the derivative of a polynomial in $x_i$ (with all other variables considered constant). The polynomial $D_v(f)(x_1 - v_1, \ldots, x_n - v_n)$ is the first order Taylor polynomial of the function $f$ at $v$, so gives the best linear approximation to $f(x_1, \ldots, x_n) \in k[x_1, \ldots, x_n]$ at $v$. It follows that if $\mathbf{T}$ is the linear variety $\mathcal{Z}(D_v(f)(x_1, \ldots, x_n))$ consisting of those points where $D_v(f)$ is zero, then the translate $v + \mathbf{T}$ is "tangent" to the hypersurface $\mathcal{Z}(f)$ at $v$.

**Example**

Suppose $f = x^2 - y \in k[x, y]$, so that $V = \mathcal{Z}(f)$ is just the parabola $y = x^2$. We have $\partial f/\partial x = 2x$ and $\partial f/\partial y = -1$, which at $v = (3, 9)$ are equal to 6 and $-1$, respectively. Then

$$D_{(3,9)}(f)(x, y) = 6x - y,$$

and the corresponding linear variety $\mathbf{T}$ is the line $y = 6x$ through the origin. The translate $(3, 9) + \mathbf{T}$ is the usual tangent line to the parabola at $(3, 9)$. The Taylor expansion of $x^2 - y$ at $(3, 9)$ is $x^2 - y = [\,6(x - 3) - (y - 9)\,] + (x - 3)^2$. The first order terms are $D_{(3,9)}(f)(x - 3, y - 9)$ and give the best linear approximation to $x^2 - y$ near (3,9).

It is straightforward to extend these notions to any affine variety $V$ in $\mathbb{A}^n$.

**Definition.** Define the *tangent space to $V$ at $v$* to be the linear variety

$$\mathbb{T}_{v,V} = \mathcal{Z}(\{D_v(f)(x_1, \ldots, x_n) \mid f \in \mathcal{I}(V)\}).$$

The formal partial derivatives are $k$-linear and obey the usual product rule for derivatives, so the tangent space may be computed from the generators for $\mathcal{I}(V)$:

$$\text{if} \quad \mathcal{I}(V) = (f_1, f_2, \ldots, f_m) \qquad \text{then} \quad \mathbb{T}_{v,V} = \bigcap_{i=1}^{m} \mathcal{Z}(D_v(f_i)).$$

Note that $\mathbb{T}_{v,V}$ is an intersection of vector spaces, so is a vector subspace of $k^n$.

This definition of the tangent space $\mathbb{T}_{v,V}$, while making apparent the connection with tangents to the variety $V$, seems to depend on the embedding of $V$ in $\mathbb{A}^n$. In fact the tangent space can be defined entirely in terms of the local ring $\mathcal{O}_{v,V}$, as the next proposition proves.

**Proposition 52.** Let $V$ be an affine variety over the algebraically closed field $k$ and let $v$ be a point on $V$ with local ring $\mathcal{O}_{v,V}$ and corresponding maximal ideal $\mathfrak{m}_{v,V}$. Then there is a $k$-vector space isomorphism

$$(\mathbb{T}_{v,V})^* \cong \mathfrak{m}_{v,V}/\mathfrak{m}_{v,V}^2$$

where $(\mathbb{T}_{v,V})^*$ denotes the vector space dual (cf. Section 11.3) of the tangent space $\mathbb{T}_{v,V}$ to $V$ at $v$.

*Proof:* Let $(k^n)^*$ denote the $n$-dimensional vector space dual to $k^n$. Since each $D_v(f)$ is a linear function, $D_v$ is a linear transformation from $k[x_1, \ldots, x_n]$ to $(k^n)^*$.

Let $M_v$ be the maximal ideal in $k[x_1, \ldots, x_n]$ generated by the set $x_i - v_i$ for $1 \leq i \leq n$. The image $M_v/\mathcal{I}(V)$ of $M_v$ in $k[V]$ is the ideal $\mathcal{I}(v)$ of functions on $V$ that are zero at $v$ and $\mathcal{I}(v)^2 = M_v^2 + \mathcal{I}(V)$. Then $\mathcal{O}_{v,V}$ is the localization of $k[V]$ at $\mathcal{I}(v)$; and identifying $\mathcal{I}(v)$ with its image in $\mathcal{O}_{v,V}$ we have $\mathfrak{m}_{v,V} = \mathcal{I}(v)\mathcal{O}_{v,V}$ (Proposition 46(2)). By definition of $D_v$ we have $D_v(x_i - v_i) = x_i$, and since these linear functions form a basis of $(k^n)^*$, it follows that $D_v$ maps $M_v$ surjectively onto $(k^n)^*$. The kernel of $D_v$ consists of the elements of $k[x_1, \ldots, x_n]$ whose Taylor expansion at $v$ starts in degree at least 2 and these are just the elements in $M_v^2$. Hence $D_v$ defines an isomorphism

$$D_v : M_v/M_v^2 \xrightarrow{\sim} (k^n)^*.$$

The tangent space $\mathbb{T}_{v,V}$ is a vector subspace of $k^n$, so every linear function on $k^n$ restricts to a linear function on $\mathbb{T}_{v,V}$. Composing $D_v$ with this restriction map gives a linear transformation

$$D : M_v \xrightarrow{D_v} (k^n)^* \xrightarrow{\text{res}} (\mathbb{T}_{v,V})^*$$

which is surjective since the individual maps are each surjective. We have already seen that $\mathcal{I}(v)^2 = M_v^2 + \mathcal{I}(V)$, so $\mathcal{I}(v)/\mathcal{I}(v)^2 \cong M_v/(M_v^2 + \mathcal{I}(V))$. It follows by Proposition 46(5) that $\mathfrak{m}_{v,V}/\mathfrak{m}_{v,V}^2 \cong \mathcal{I}(v)/\mathcal{I}(v)^2$. To prove the proposition it is therefore sufficient to show that $\ker D = M_v^2 + \mathcal{I}(V)$, since then

$$\mathfrak{m}_{v,V}/\mathfrak{m}_{v,V}^2 \cong M_v/(M_v^2 + \mathcal{I}(V)) = M_v/\ker D \cong (\mathbb{T}_{v,V})^*.$$

The polynomial $f$ is in $\ker D$ if and only if $D_v(f)$ is zero on $\mathbb{T}_{v,V}$, i.e., if and only if the linear term of the Taylor polynomial of $f$ expanded about $v$ lies in $\mathcal{I}(\mathbb{T}_{v,V})$. Since the linear terms of the functions in $\mathcal{I}(V)$ generate the ideal $\mathcal{I}(\mathbb{T}_{v,V})$, it follows that $f$ is in $\ker D$ if and only if $f - g$ has zero linear term for some $g$ in $\mathcal{I}(V)$. But this is equivalent to $f \in \mathcal{I}(V) + M_v^2$, so $\ker D = \mathcal{I}(V) + M_v^2$, completing the proof of the proposition.

Recall that the *dimension* of a variety $V$ is by definition the transcendence degree of the field $k(V)$ over $k$. Since each local ring $\mathcal{O}_{v,V}$ has $k(V)$ as its field of fractions, the dimension of $V$ is determined by the transcendence degree over $k$ of the field of fractions of any of its local rings.

**Definition.** We say $V$ is *nonsingular* at the point $v \in V$ (or $v$ is a *nonsingular point* of $V$) if the dimension of the $k$-vector space $\mathbb{T}_{v,V}$ is $\dim V$. Equivalently (by Proposition 52), $v$ is a nonsingular point of $V$ if $\dim_k(\mathfrak{m}_{v,V}/\mathfrak{m}_{v,V}^2) = \dim V$. Otherwise the point $v$ is called a *singular point.* The variety $V$ is *nonsingular* or *smooth* if it is nonsingular at every point.

The geometric picture is that at a nonsingular point $v$ there are as many independent tangents as one would expect: a tangent line on a curve, a tangent plane on a surface, etc.

Whether a variety $V$ is nonsingular at a point $v$ can be determined from properties of the local ring $\mathcal{O}_{v,V}$, namely whether $\dim_k(\mathfrak{m}_{v,V}/\mathfrak{m}_{v,V}^2) = \dim \mathcal{O}_{v,V}$. A local ring having this property is said to be a *regular local ring.* In particular, the notion of singularity does not depend on the embedding of $V$ in a specific affine space. This algebraic interpretation can be used to *define* smoothness for abstract algebraic varieties, where the geometric intuition of tangent planes to surfaces (for example) is not as obvious.

If $f_1, \ldots, f_m$ are generators for $\mathcal{I}(V)$ defining $V$ in $\mathbb{A}^n$, then the dimension of $V$ can be determined from a Gröbner basis for $\mathcal{I}(V)$ (cf. Exercise 29). Determining the dimension of the tangent space $\mathbb{T}_{v,V}$ as a vector space over $k$ is a linear algebra problem: this vector space is the set of solutions of the $m$ linear equations $D_v(f_i)(x_1, \ldots, x_n) = 0$. If $r$ is the rank of the $m \times n$ matrix of coefficients $\partial f_i/\partial x_j(v)$ of this system of equations, then $\mathbb{T}_{v,V}$ is a vector space of dimension $n-r$. Using this it is not too difficult to establish the following:

1. We have $\dim V \leq \dim_k(\mathbb{T}_{v,V}) \leq n$ for every point $v$ in $V \subseteq \mathbb{A}^n$.

2. The set of singular points of $V$ is a proper Zariski closed subset of $V$. The set of nonsingular points of $V$ is a nonempty open subset of $V$; in particular the nonsingular points of $V$ are dense in $V$ (so "most" points of $V$ are nonsingular).

We also state without proof the following result which further relates the local geometry of $V$ to the algebraic properties of the local rings of $V$:

3. If $v$ is a nonsingular point, then the local ring $\mathcal{O}_{v,V}$ is a Unique Factorization Domain; in particular, $\mathcal{O}_{v,V}$ is integrally closed (cf. Example 3 following Corollary 25).

The variety $V$ is said to be *factorial* if $\mathcal{O}_{v,V}$ is a U.F.D. for every point $v \in V$, and is said to be a *normal* variety if $\mathcal{O}_{v,V}$ is integrally closed for every $v \in V$ (which by Proposition 49 is equivalent to $k[V]$ being integrally closed). By (3) above we have

$$\textit{smooth varieties} \subseteq \textit{factorial varieties} \subseteq \textit{normal varieties.}$$

In general each of the above containments is proper. In the case when $V$ has dimension 1, i.e., $V$ is an *affine curve*, however, these three properties are in fact equivalent: we shall prove later that an irreducible affine curve is smooth if and only if it is normal or factorial (cf. Corollary 13 in Section 16.2). It follows that over an algebraically closed field $k$,

*an irreducible affine curve $C$ is smooth if and only if $k[C]$ is integrally closed.*

For any irreducible affine curve $C$ the integral closure, $S$, of $k[V]$ in $k(V)$ is also the coordinate ring of an irreducible affine curve $\widetilde{C}$. Then $S$ is integral over $k[V]$ and, by Theorem 30 and Corollary 27 it follows that there is a morphism from the smooth curve $\widetilde{C}$ onto $C$ that has finite fibers. The curve $\widetilde{C}$ is called the *normalization* or the *nonsingular model* of $C$, and one can show that it is unique up to isomorphism. Note how the existence of a smooth curve mapping finitely to $C$ (a problem in "geometry") is solved by the existence of integral closures in ring extensions (a problem in "algebra").

We shall give another characterization of smoothness for irreducible affine curves at the end of Section 16.2.

## EXERCISES

As usual $R$ is a commutative ring with 1 and $D$ is a multiplicatively closed set in $R$.

**1.** Suppose $M$ is a finitely generated $R$-module. Prove that $D^{-1}M = 0$ if and only if $dM = 0$ for some $d \in D$.

**2.** Let $I$ be an ideal in $R$, let $D$ be a multiplicatively closed subset of $R$ with ring of fractions $D^{-1}R$, and let ${}^c({}^eI) = R$ be the saturation of $I$ with respect to $D$.
- **(a)** Prove that ${}^c({}^eI) = R$ if and only if ${}^eI = D^{-1}R$ if and only if $I \cap D \neq \emptyset$.
- **(b)** Prove that $I = {}^c({}^eI)$ is saturated if and only if for every $d \in D$, if $da \in I$ then $a \in I$.
- **(c)** Prove that extension and contraction define inverse bijections between the ideals of $R$ saturated with respect to $D$ and the ideals of $D^{-1}R$.
- **(d)** Let $I = (2x, 3y) \subset \mathbb{Z}[x, y]$. Show the saturation of $I$ with respect to $\mathbb{Z} - \{0\}$ is $(x, y)$.

**3.** If $I$ is an ideal in the commutative ring $R$ let $\varphi : R[x_1, \ldots, x_n] \cong (R/I)[x_1, \ldots, x_n]$ be the ring homomorphism with kernel $I[x_1, \ldots, x_n]$ given by reducing coefficients modulo $I$. If $\overline{A}$ is an ideal in $(R/I)[x_1, \ldots, x_n]$, let $A$ denote the inverse image of $\overline{A}$ under $\varphi$.
- **(a)** For any $i \geq 1$ show that the inverse image under $\varphi$ of the subring $(R/I)[x_1, \ldots, x_i]$ is $R[x_1, \ldots, x_i] + I[x_1, \ldots, x_n]$.
- **(b)** Prove that $\varphi(A \cap R[x_1, \ldots, x_i]) = \overline{A} \cap (R/I)[x_1, \ldots, x_i]$

**4.** Let $f = y^5 - z^4$, viewed as a polynomial in $y$ with coefficients in $\mathbb{Q}[z]$.
- **(a)** Prove that $f$ has no roots in $\mathbb{Q}[z]$.
- **(b)** Suppose $f = (y^2 + ay + b)(y^3 + cy^2 + dy + e)$. Show that $a, b, c, d, e$ satisfy the system of equations

$$a + c = 0, \quad ac + b + d = 0, \quad ad + bc + e = 0, \quad ae + bd = 0, \quad be - z^4 = 0.$$

Deduce that $e^5 = z^{12}$ and conclude that $f$ is irreducible in $\mathbb{Q}[y, z]$. [Use elimination.]

5. Suppose $R$ is a U.F.D. with field of fractions $F$ and $p \in R[x]$ is a monic polynomial.
   **(a)** Show that the ideal $pR[x]$ generated by $p$ in $R[x]$ is prime if and only if the ideal $pF[x]$ generated by $p$ in $F[x]$ is prime. [Use Gauss' Lemma.]
   **(b)** Show that $pR[x]$ is saturated, i.e., that $pF[x] \cap R[x] = pR[x]$.
6. Show that $I = (y^3 - xz, xy^2 - z^2)$ is not a prime ideal in $\mathbb{Q}[x, y, z]$ and find explicit elements $a, b \in \mathbb{Q}[x, y, z]$ with $ab \in I$ but $a \notin I$ and $b \notin I$.
7. Show that $P = (y^3 - xz, xy^2 - z^2, x^2 - yz)$ is a prime ideal in $\mathbb{Q}[x, y, z]$.
8. Show that $P = (x^2 - yz, w^2 - x^4 z)$ is a prime ideal in $\mathbb{Q}[x, y, z, w]$.
9. Show that $P = (xz^2 - w^3, xw^2 - y^4, y^4z^2 - w^5)$ is a prime ideal in $\mathbb{Q}[x, y, z, w]$.
10. Show that $I = (xy - w^3, y^2 - zw)$ is not a prime ideal in $\mathbb{Q}[x, y, z, w]$ and find $a, b$ with $ab \in I$ but $a, b \notin I$.
11. Let $R_P$ be the localization of $R$ at the prime $P$. Prove that if $Q$ is a $P$-primary ideal of $R$ then $Q = {}^c({}^eQ)$ with respect to the extension and contraction of $Q$ to $R_P$. Show the same result holds if $Q$ is $P'$-primary for some prime $P'$ contained in $P$.
12. Let $R = \mathbb{R}[x, y, z]/(xy - z^2)$, let $P = (\bar{x}, \bar{z})$ be the prime ideal generated by the images of $x$ and $y$ in $R$, and let $R_P$ be the localization of $R$ at $P$. Prove that $P^2 R_P \cap R = (\bar{x})$ and is strictly larger than $P^2$.
13. Prove that if $N$ and $N'$ are two $R$-submodules of an $R$-module $M$ with $N_P = N'_P$ in the localization $M_P$ for every prime ideal $P$ of $R$ (or just for every maximal ideal) then $N = N'$.
14. Suppose $\varphi : M \to N$ is an $R$-module homomorphism. Prove that $\varphi$ is injective (respectively, surjective) if and only if the induced $R_P$-module homomorphism $\varphi : M_P \to N_P$ is injective (respectively, surjective) for every prime ideal $P$ of $R$ (or just for every maximal ideal of $R$).
15. Let $R = \mathbb{Z}[\sqrt{-5}]$ be the ring of integers in the quadratic field $\mathbb{Q}(\sqrt{-5})$ and let $I$ be the prime ideal $(2, 1 + \sqrt{-5})$ of $R$ generated by 2 and $1 + \sqrt{-5}$ (cf. Exercise 5, Section 8.2). Recall that every nonzero prime ideal $P$ of $R$ contains a prime $p \in \mathbb{Z}$.
   **(a)** If $P$ is a prime ideal of $R$ not containing 2 prove that $I_P = R_P$.
   **(b)** If $P$ is a prime ideal of $R$ containing 2 prove that $P = I$ and that $I_P = (1+\sqrt{-5})R_P$.
   **(c)** Prove that $I_P \cong R_P$ as $R_P$-modules for every prime ideal $P$ of $R$ but that $I$ and $R$ are not isomorphic $R$-modules. (This example shows that it is important in Exercise 14 to be *given* the $R$-module homomorphism $\varphi$.) [Observe that $I \cong R$ as $R$-modules if and only if $I$ is a *principal* ideal.]
16. Prove that localization commutes with tensor products: there is a unique isomorphism of $D^{-1}R$-modules $\varphi : (D^{-1}M) \otimes_{D^{-1}R} (D^{-1}N) \cong D^{-1}(M \otimes_R N)$ with $\varphi((m/d) \otimes (n/d')) = (m \otimes n)/dd'$ for any $R$-modules $M$, $N$, and multiplicatively closed set $D$ in $R$.
17. Prove that the $R$-module $A$ is a flat $R$-module if and only if $A_P$ is a flat $R_P$-module for every prime ideal $P$ of $R$ (or just for every maximal ideal of $R$). [Use Proposition 41, Exercises 14 and 16, and the exactness properties of localization.]
18. In the notation of Example 2 following Corollary 37, prove that $R_f \cong R[x]/(fx - 1)$ if $f$ is not nilpotent in $R$. [Show that the map $\varphi : R[x] \to R_f$ defined by $\varphi(r) = r/1$ and $\varphi(x) = 1/f$ gives a surjective ring homomorphism and the universal property in Theorem 36 gives an inverse.]
19. Prove that if $R$ is an integrally closed integral domain and $D$ is any multiplicatively closed subset of $R$ containing 1, then $D^{-1}R$ is integrally closed.

**20.** Suppose that $R$ is a subring of the ring $S$ with $1 \in R$ and that $S$ is integral over $R$. If $D$ is any multiplicatively closed subset of $R$, prove that $D^{-1}S$ is integral over $D^{-1}R$.

**21.** Suppose $\varphi : R \to S$ is a ring homomorphism and $D'$ is a multiplicatively closed subset of $S$. Let $D = \varphi^{-1}(D')$. Prove that $D$ is a multiplicatively closed subset of $R$ and that the map $\varphi' : D^{-1}R \to D'^{-1}S$ given by $\varphi'(r/d) = \varphi(r)/\varphi(d)$ is a ring homomorphism.

**22.** Suppose $P \subseteq Q$ are prime ideals in $R$ and let $R_Q$ be the localization of $R$ at $Q$. Prove that the localization $R_P$ is isomorphic to the localization of $R_Q$ at the prime ideal $PR_Q$ (cf. the preceding exercise).

**23.** Let $\varphi : A \to B$ be a homomorphism of commutative rings with $\varphi(1_A) = 1_B$, and let $P$ be a prime ideal of $A$. Let contraction and extension of ideals with respect to $\varphi$ be denoted by superscripts $c$ and $e$ respectively. Prove that $P$ is the contraction of a prime ideal in $B$ if and only if $P = (P^e)^c$. [Localize $B$ at $\varphi(A - P)$.]

**24.** *(The Going-down Theorem)* Let $S$ be an integral domain, let $R$ be an integrally closed subring of $S$ containing $1_S$, and let $k$ be the field of fractions of $R$. Suppose that $P_2 \subseteq P_1$ are prime ideals in $R$ and that $Q_1$ is a prime ideal in $S$ with $Q_1 \cap R = P_1$. Let $S_{Q_1}$ be the localization of $S$ at $Q_1$.

**(a)** Show that $P_2 \subseteq P_2 S_{Q_1} \cap R$.

**(b)** Suppose that $a \in P_2 S_{Q_1} \cap R$ and write $a = s/d$ with $s \in P_2 S$ and $d \in S$, $d \notin Q_1$. If the minimal polynomial of $s$ over $k$ is $x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ with $a_0, \ldots, a_{n-1} \in P_2$ (cf. Exercise 12 in Section 3) show that the minimal polynomial of $d$ over $k$ is $x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0$ where $b_i = a_i/a^{n-i}$ and conclude that $b_i \in R$. [Use Exercise 10 in Section 3.]

**(c)** Show that $a \in P_2$ and conclude that $P_2 S_{Q_1} \cap R = P_2$. [Show $a \notin P_2$ implies $b_i \in P_2$ for $i = 0, 1, \ldots, n-1$, which would imply $d^n \in P_2 S \subseteq P_1 S \subseteq Q_1$ and so $d \in Q_1$.]

**(d)** Prove that $P_2 S_{Q_1}$ is contained in a prime ideal $P$ of $S_{Q_1}$ with $P \cap R = P_2$. [Use (c) and the previous exercise for $\varphi : R \to S_{Q_1}$.]

**(e)** Let $Q_2 = P \cap S$. Prove that $Q_2 \subseteq Q_1$ and that $Q_2 \cap R = P_2$.

**(f)** Use induction together with the previous result to prove the Going-down Theorem: Theorem 26(4).

**25.** Let $k$ be an algebraically closed field and let $V = \mathcal{Z}(xz - yw) \subset \mathbb{A}^4$. Prove that the set of points $v$ where $f = \bar{x}/\bar{y} \in k(V)$ is regular is precisely the set of points $(x, y, z, w)$ where $y \neq 0$ or $z \neq 0$. [If $f = \bar{a}/\bar{b}$ show that $ay - bx \in (xz - yw)$ as polynomials in $k[x, y, z, w]$ and conclude that $b \in (y, z)$.] Prove that there is no function $a/b \in k(V)$ with $b(v) \neq 0$ for every $v$ where $f$ is regular.

**26.** *(Differentials of Morphisms)* Let $\varphi : V \to W$ be a morphism of affine varieties over the algebraically closed field $k$ and suppose $\varphi(v) = w$.

**(a)** Show that $\varphi$ induces a linear map from the $k$-vector space $M_w/M_w^2$ to the $k$-vector space $M_v/M_v^2$, and use this to show that $\varphi$ induces a linear map $d\varphi$ (called the *differential* of $\varphi$) from the $k$-vector space $\mathbb{T}_{v,V}$ to the $k$-vector space $\mathbb{T}_{w,W}$.

**(b)** Prove that if $V \subseteq \mathbb{A}^n$, $W \subseteq \mathbb{A}^m$ and $\varphi = (F_1(x_1, \ldots, x_n), \ldots, F_m(x_1, \ldots, x_n))$ then $d\varphi : \mathbb{T}_{v,V} \to \mathbb{T}_{w,W}$ is given explicitly by

$$(d\varphi)(a_1, \ldots, a_n) = (D_v(F_1)(a_1, \ldots, a_n), \ldots, D_v(F_m)(a_1, \ldots, a_n)).$$

[If $g = g(y_1, \ldots, y_m)$ show that the chain rule implies

$$\frac{\partial(g \circ \varphi)}{\partial x_i}(v) = \sum_{j=1}^{m} \frac{\partial g}{\partial y_j}(w) \frac{\partial F_j}{\partial x_i}(v),$$

so that $D_v(g \circ \varphi)(a_1, \ldots, a_n) = D_w(g)(b_1, \ldots, b_m)$ where $b_j = D_v(F_j)(a_1, \ldots, a_n)$. Then use the fact that $g \circ \varphi \in \mathcal{I}(V)$ if $g \in \mathcal{I}(W)$.]

**(c)** If $\psi : U \to V$ is another morphism with $\psi(u) = v$, prove that the associated $d(\varphi \circ \psi) : \mathbb{T}_{u,U} \to \mathbb{T}_{w,W}$ is the same as $d\varphi \circ d\psi$.

**(d)** Prove that if $\varphi$ is an isomorphism then $d\varphi$ is a vector space isomorphism from $\mathbb{T}_{v,V}$ to $\mathbb{T}_{w,W}$ for every $\varphi(v) = w$.

**27.** Let $V = \mathbb{A}^1$ and $W = \mathcal{Z}(xz - y^2, yz - x^3, z^2 - x^2y) \subset \mathbb{A}^3$. Let $\varphi : V \to W$ be the surjective morphism $\varphi(t) = (t^3, t^4, t^5)$ (cf. Exercise 26 in Section 1). For each $t \in \mathbb{A}^1$ describe the differential $d\varphi : \mathbb{T}_{t,\mathbb{A}^1} \to \mathbb{T}_{(t^3,t^4,t^5),W}$ in the previous exercise explicitly; in particular prove that $d\varphi$ is an isomorphism of vector spaces for all $t \neq 0$ and is the zero map for $t = 0$. Use this to prove that $V$ and $W$ are not isomorphic.

**28.** If $k$ is a field, the quotient $k[x]/(x^2)$ is called the *ring of dual numbers* over $k$. If $V$ is an affine algebraic set over $k$, show that a $k$-algebra homomorphism from $k[V]$ to $k[x]/(x^2)$ is equivalent to specifying a point $v \in V$ with $\mathcal{O}_{v,V}/\mathfrak{m}_{v,V} = k$ (called a *$k$-rational point* of $V$) together with an element in the tangent space $\mathbb{T}_{v,V}$ of $V$ at $v$.

**29.** (*Computing the dimension of a variety*) Let $P$ be a prime ideal in $k[x_1, \ldots, x_n]$, set $P_0 = 0$ and let $P_i = P \cap k[x_1, \ldots, x_i]$. Define the varieties $V_i = \mathcal{Z}(P_i) \subseteq \mathbb{A}^i$ with $V_0$ the zero dimensional variety consisting of a single point and coordinate ring $k$.

**(a)** Show that $\dim V_{i-1} \leq \dim V_i \leq \dim V_{i-1} + 1$. [First exhibit an injection from $k[V_{i-1}]$ into $k[V_i]$; then show that $k[V_i]$ is a $k$-algebra generated by $k[V_{i-1}]$ and one additional generator.]

**(b)** If the ideal generated by $P_{i-1}$ in $k[x_1, \ldots, x_i]$ equals $P_i$, show that $V_i \cong V_{i-1} \times \mathbb{A}^1$ and deduce that $\dim V_i = \dim V_{i-1} + 1$.

**(c)** If the ideal generated by $P_{i-1}$ in $k[x_1, \ldots, x_i]$ is properly contained in $P_i$, show that $\dim V_i = \dim V_{i-1}$.

**(d)** Show that $\dim V$ equals the number of $i \in \{1, 2, \ldots, n\}$ such that the ideal generated by $P_{i-1}$ in $k[x_1, \ldots, x_i]$ equals the ideal $P_i$. Deduce that if $G$ is the reduced Gröbner basis for $P$ with respect to the lexicographic monomial ordering $x_n > \cdots > x_1$ and $G_i = G \cap k[x_1, \ldots, x_i]$ where $G_0 = \emptyset$, and $N$ is the number of $i$ with $G_i \neq G_{i-1}$ for $1 \leq i \leq n$, then $\dim V = n - N$.

The following eleven exercises introduce the notion of the *support* of an $R$-module $M$ and its relation to the associated primes of $M$. Cf. also Exercises 29 to 35 in Section 1 and Exercises 25 to 30 in Section 5.

**Definition.** If $M$ is an $R$-module, then the set of prime ideals $P$ of $R$ for which the localization $M_P$ is nonzero is called the *support* of $M$, denoted $\mathrm{Supp}(M)$.

**30.** Prove that $M = 0$ if and only if $\mathrm{Supp}(M) = \emptyset$. [Use Proposition 47.]

**31.** If $0 \to L \to M \to N \to 0$ is an exact sequence of $R$-modules, prove that the localization $M_P$ is nonzero if and only if one of the localizations $N_P$ and $L_P$ is nonzero and deduce that $\mathrm{Supp}(M) = \mathrm{Supp}(L) \cup \mathrm{Supp}(N)$. In particular, if $M = M_1 \oplus \cdots \oplus M_n$ prove that $\mathrm{Supp}(M) = \mathrm{Supp}(M_1) \cup \cdots \cup \mathrm{Supp}(M_n)$.

**32.** Suppose $P \subseteq Q$ are prime ideals in $R$ and that $M$ is an $R$-module. Prove that the localization of the $R$-module $M_Q$ at $P$ is the localization $M_P$, i.e., $(M_Q)_P = M_P$. [Argue directly, or use Proposition 41 and the associativity of the tensor product.]

**33.** Suppose $P \subseteq Q$ are prime ideals in $R$ and that $M$ is an $R$-module. Prove that if $P \in \mathrm{Supp}(M)$ then $Q \in \mathrm{Supp}(M)$. [Use the previous exercise.]

**34.** **(a)** Suppose $M = Rm$ is a cyclic $R$-module. Prove that $M_P = 0$ if and only if there is

an element $r \in R$, $r \notin P$ with $rm = 0$. Deduce that $P \in \text{Supp}(M)$ if and only if $P$ contains the annihilator of $m$ in $R$ (cf. Exercise 10 in Section 10.1).

**(b)** If $M = Rm_1 + \cdots + Rm_n$ is a finitely generated $R$-module prove that $P \in \text{Supp}(M)$ if and only if $P$ is contained in $\text{Supp}(Rm_i)$ for some $i = 1, \ldots, n$. [Use Proposition 42.] Deduce that $P \in \text{Supp}(M)$ if and only if $P$ contains the annihilator $\text{Ann}(M)$ of $M$ in $R$. [Note $\text{Ann}(M) = \cap_{i=1}^{n} \text{Ann}(Rm_i)$, then use (a) and Exercise 11 of Section 7.4.]

**35.** Suppose $P$ is a prime ideal of $R$ with $P \cap D = \emptyset$. Prove that if $P \in \text{Ass}_R(M)$ then $D^{-1}P \in \text{Ass}_{D^{-1}R}(D^{-1}M)$. [Use Proposition 38(3) and Proposition 42.]

**36.** Suppose $D^{-1}P \in \text{Ass}_{D^{-1}R}(D^{-1}M)$ where $P = (a_1, \ldots, a_n)$ is a finitely generated prime ideal in $R$ with $P \cap D = \emptyset$.

**(a)** Suppose $m/d \in D^{-1}M$ has annihilator $D^{-1}P$ in $D^{-1}R$. Show that $d_i a_i m = 0 \in R$ for some $d_1, \ldots, d_n \in D$.

**(b)** Let $d' = d_1 d_2 \ldots d_n$. Show that $P = \text{Ann}(d'm)$ and conclude that $P \in \text{Ass}_R(M)$. [The inclusion $P \subseteq \text{Ann}(d'm)$ is immediate. For the reverse inclusion, show that $b \in \text{Ann}(d'm)$ implies that $b/1$ annihilates $m/d$ in $D^{-1}M$, hence $b/1 \in D^{-1}P$, and conclude $b \in P$.]

**37.** Suppose $M$ is a module over the Noetherian ring $R$. Use the previous two exercises to show that under the bijection of Proposition 38(3) the prime ideals $P$ of $\text{Ass}_R(M)$ with $P \cap D = \emptyset$ correspond bijectively with the prime ideals of $\text{Ass}_{D^{-1}R}(D^{-1}M)$.

**38.** Suppose $M$ is a module over the Noetherian ring $R$ and $D$ is a multiplicatively closed subset of $R$. Let $\mathcal{S}$ be the subset of prime ideals $P$ in $\text{Ass}_R(M)$ with $P \cap D \neq \emptyset$. This exercise proves that the kernel $N$ of the localization map $M \to D^{-1}M$ is the unique submodule $N$ of $M$ with $\text{Ass}_R(N) = \mathcal{S}$ and $\text{Ass}_R(M/N) = \text{Ass}_R(M) - \mathcal{S}$.

**(a)** If $N'$ is a submodule of $M$ with $\text{Ass}_R(N') = \mathcal{S}$ and $\text{Ass}_R(M/N') = \text{Ass}_R(M) - \mathcal{S}$ as in Exercise 35 in Section 1, prove that the diagram

$$\begin{array}{ccc} M & \xrightarrow{\pi} & M/N' \\ {\scriptstyle\varphi}\downarrow & & \downarrow{\scriptstyle\varphi'} \\ D^{-1}M & \xrightarrow{\pi'} & D^{-1}(M/N') \end{array}$$

is commutative, where $\pi$ and $\pi'$ are the natural projections (cf. Proposition 42(6)) and $\varphi$, $\varphi'$ are the localization homomorphisms.

**(b)** Show that $\text{Ass}_{D^{-1}R}(D^{-1}N') = \emptyset$ and conclude that $D^{-1}N' = 0$ and that $\pi'$ is injective. [Use the previous exercise, the definition of $\mathcal{S}$, and Exercise 34 in Section 1.]

**(c)** If $x$ is the kernel $K$ of $\varphi'$ show that $\text{Ann}(x) \cap D \neq \emptyset$ and that $\text{Ass}_R(K) \subseteq \mathcal{S}$. Show that $\text{Ass}_R(K) \subseteq \text{Ass}_R(M/N')$ implies that $\text{Ass}_R(K) = \emptyset$, and deduce that $K = 0$.

**(d)** Prove $\varphi$ and $\pi$ have the same kernel, i.e., $N = N'$, and this submodule of $M$ is unique.

The next two exercises establish a fundamental relation between the sets $\text{Ass}_R(M)$ and $\text{Supp}(M)$ of prime ideals related to the $R$-module $M$.

**39.** Prove that $\text{Ass}_R(M) \subseteq \text{Supp}(M)$. [If $Rm \cong R/P$ use Proposition 42(4) and Proposition 46(1) to show that $0 \neq (Rm)_P \subseteq M_P$.]

**40.** Suppose that $R$ is Noetherian and $M$ is an $R$-module.

**(a)** If $P \in \text{Supp}(M)$ prove that $P$ contains a prime ideal $Q$ with $Q \in \text{Ass}_R(M)$.

**(b)** If $P$ is a minimal prime in $\text{Supp}(M)$, show that $P \in \text{Ass}_R(M)$. [Use Exercise 33 in Section 1 to show that $\text{Ass}_{R_P}(M_P) \neq \emptyset$ and then use Exercise 37.]

**(c)** Conclude that $\text{Ass}_R(M) \subseteq \text{Supp}(M)$ and that these two sets have the same minimal elements.

## 15.5 THE PRIME SPECTRUM OF A RING

Throughout this section the term "ring" will mean commutative ring with 1 and all ring homomorphisms $\varphi : R \to S$ will be assumed to map $1_R$ to $1_S$.

We have seen that most of the geometric properties of affine algebraic sets $V$ over $k$ can be translated into algebraic properties of the associated coordinate rings $k[V]$ of $k$-valued functions on $V$. For example, the morphisms from $V$ to $W$ correspond to $k$-algebra ring homomorphisms from $k[W]$ to $k[V]$. When the field $k$ is an algebraically closed field this translation is particularly precise: Hilbert's Nullstellensatz establishes a bijection between the points $v$ of $V$ and the maximal ideals $M = \mathcal{I}(v)$ of $k[V]$, and if $\varphi : V \to W$ is a morphism then $\varphi(v) \in W$ corresponds to the maximal ideal $\widetilde{\varphi}^{-1}(M)$ in $k[W]$. In this development we have generally started with geometric properties of the affine algebraic sets and then seen that many of the algebraic properties common to the associated coordinate rings can be defined for arbitrary commutative rings. Suppose now we try to reverse this, namely start with a general commutative ring as the algebraic object and attempt to define a corresponding "geometric" object by analogy with $k[V]$ and $V$.

Given a commutative ring $R$, perhaps the most natural analogy with $k[V]$ and $V$ would suggest defining the collection of maximal ideals $M$ of $R$ as the "points" of the associated geometric object. Under this definition, if $\widetilde{\varphi} : R' \to R$ is a ring homomorphism, then $\widetilde{\varphi}^{-1}(M)$ should correspond to the maximal ideal $M$. Unfortunately, the inverse image of a maximal ideal by a ring homomorphism in general need not be a maximal ideal. Since the inverse image of a *prime* ideal under a ring homomorphism (that maps 1 to 1) *is* prime, this suggests that a better definition might include the prime ideals of $R$. This leads to the following:

**Definition.** Let $R$ be a commutative ring with 1. The *spectrum* or *prime spectrum* of $R$, denoted Spec $R$, is the set of all prime ideals of $R$. The set of all maximal ideals of $R$, denoted mSpec $R$, is called the *maximal spectrum* of $R$.

**Examples**

**(1)** If $R$ is a field then Spec $R$ = mSpec $R = \{(0)\}$.

**(2)** The points in Spec $\mathbb{Z}$ are the prime ideal (0) and the prime ideals $(p)$ where $p > 0$ is a prime, and mSpec $\mathbb{Z}$ consists of all the prime ideals of Spec $\mathbb{Z}$ except (0).

**(3)** The elements of Spec $\mathbb{Z}[x]$ are the following:

- **(a)** (0)
- **(b)** $(p)$ where $p$ is a prime in $\mathbb{Z}$
- **(c)** $(f)$ where $f \neq 1$ is a polynomial of content 1 (i.e., the g.c.d. of its coefficients is equal to 1) that is irreducible in $\mathbb{Q}[x]$
- **(d)** $(p, g)$ where $p$ is a prime in $\mathbb{Z}$ and $g$ is a monic polynomial that is irreducible mod $p$.

The elements of mSpec $\mathbb{Z}[x]$ are the primes in (d) above.

In the analogy with $k[V]$ and $V$ when $k$ is algebraically closed, the elements $f \in k[V]$ are functions on $V$ with values in $k$, obtained by evaluating $f$ at the point $v$ in $V$. Note that "evaluation at $v$" defines a homomorphism from $k[V]$ to $k$ with kernel $\mathcal{I}(v)$, and that the value of $f$ at $v$ is the element of $k$ representing $f$ in the quotient

$k[V]/\mathcal{I}(v) \cong k$. Put another way, the value of $f \in k[V]$ at $v \in V$ can be viewed as the element $\bar{f} \in k[V]/\mathcal{I}(v) \cong k$. A similar definition can be made in general:

**Definition.** If $f \in R$ then the *value* of $f$ at the point $P \in \operatorname{Spec} R$ is the element $f(P) = \bar{f} \in R/P$.

Note that the values of $f$ at different points $P$ in general lie in *different* integral domains. Note also that in general $f \in R$ is not uniquely determined by its values, rather $f$ is determined only up to an element in the nilradical of $R$ (cf. Exercise 3).

There are analogues of the maps $\mathcal{Z}$ and $\mathcal{I}$ and also for the Zariski topology. For any subset $A$ of $R$ define

$$\mathcal{Z}(A) = \{P \in X \mid A \subseteq P\} \subseteq \operatorname{Spec} R,$$

the collection of prime ideals containing $A$. It is immediate that $\mathcal{Z}(A) = \mathcal{Z}(I)$, where $I = (A)$ is the ideal generated by $A$ so there is no loss simply in considering $\mathcal{Z}(I)$ where $I$ is an ideal of $R$. Note that, by definition, $P \in \mathcal{Z}(I)$ if and only if $I \subseteq P$, which occurs if and only if $f \in P$ for every $f \in I$. Viewing $f \in R$ as a function on $\operatorname{Spec} R$ as above, this says that $P \in \mathcal{Z}(I)$ if and only if $f(P) = f \bmod P = 0 \in R/P$ for all $f \in I$. In this sense, $\mathcal{Z}(I)$ consists of the points in $\operatorname{Spec} R$ at which all the functions in $I$ have the value 0.

For any subset $Y$ of $\operatorname{Spec} R$ define

$$\mathcal{I}(Y) = \bigcap_{P \in Y} P,$$

the intersection of the prime ideals in $Y$.

**Proposition 53.** Let $R$ be a commutative ring with 1. The maps $\mathcal{Z}$ and $\mathcal{I}$ between $R$ and $\operatorname{Spec} R$ defined above satisfy

**(1)** for any ideal $I$ of $R$, $\mathcal{Z}(I) = \mathcal{Z}(\operatorname{rad}(I)) = \mathcal{Z}(\mathcal{I}(\mathcal{Z}(I)))$, and $\mathcal{I}(\mathcal{Z}(I)) = \operatorname{rad} I$,

**(2)** for any ideals $I, J$ of $R$, $\mathcal{Z}(I \cap J) = \mathcal{Z}(IJ) = \mathcal{Z}(I) \cup \mathcal{Z}(J)$, and

**(3)** if $\{I_j\}$ is an arbitrary collection of ideals of $R$, then $\mathcal{Z}(\cup I_j) = \cap \mathcal{Z}(I_j)$.

*Proof:* If $P$ is a prime ideal containing the ideal $I$ then $P$ contains $\operatorname{rad} I$ (Exercise 8, Section 2), which implies $\mathcal{Z}(I) = \mathcal{Z}(\operatorname{rad}(I))$. Since $\operatorname{rad} I$ is the intersection of all the prime ideals containing $I$ (Proposition 12), the definition of $\mathcal{I}(I)$ gives $\mathcal{Z}(\operatorname{rad}(I)) = \mathcal{Z}(\mathcal{I}(I))$. Similarly,

$$\mathcal{I}(\mathcal{Z}(I)) = \bigcap_{P \in \mathcal{Z}(I)} P = \bigcap_{I \subseteq P} P = \operatorname{rad} I,$$

which completes the proof of (1). It is immediate that $\mathcal{Z}(I \cap J) = \mathcal{Z}(I) \cup \mathcal{Z}(J)$. Suppose the prime ideal $P$ contains $IJ$. If $P$ does not contain $I$ then there is some element $i \in I$ with $i \notin P$. Since $iJ \subseteq P$, it follows that $J \subseteq P$. This proves $\mathcal{Z}(IJ) = \mathcal{Z}(I) \cup \mathcal{Z}(J)$ and completes the proof of (2). The proof of (3) is immediate.

The first statement in the proposition shows that every set $\mathcal{Z}(I)$ in $\operatorname{Spec} R$ occurs for some *radical* ideal $I$, and since $\mathcal{I}(\mathcal{Z}(I)) = \operatorname{rad} I$, this radical ideal is unique.

The second two statements in the proposition show that the collection

$$\mathcal{T} = \{\mathcal{Z}(I) \mid I \text{ is an ideal of } R\}$$

satisfies the three axioms for the closed sets of a topology on Spec $R$ as in Section 2.

**Definition.** The topology on Spec $R$ defined by the closed sets $\mathcal{Z}(I)$ for the ideals $I$ of $R$ is called the *Zariski topology* on Spec $R$.

By definition, the closure in the Zariski topology of the singleton set $\{P\}$ in Spec $R$ consists of all the prime ideals of $R$ that contain $P$. In particular, a point $P$ in Spec $R$ is closed in the Zariski topology if and only if the prime ideal $P$ is not contained in any other prime ideals of $R$, i.e., if and only if $P$ is a maximal ideal (so the Zariski topology on Spec $R$ is not generally Hausdorff). These points are given a name:

**Definition.** The maximal ideals of $R$ are called the *closed points* in Spec $R$.

In terms of the terminology above, the points in Spec $R$ that are closed in the Zariski topology are precisely the points in mSpec $R$.

A closed subset of a topological space is *irreducible* if it is not the union of two proper closed subsets, or, equivalently, if every nonempty open set is dense. Arguments similar to those used to prove Proposition 17 show that the closed subset $Y = \mathcal{Z}(I)$ in Spec $R$ is irreducible if and only if $\mathcal{I}(Y) = \text{rad } I$ is prime (cf. Exercise 16).

The following proposition summarizes some of these results:

**Proposition 54.** The maps $\mathcal{Z}$ and $\mathcal{I}$ define inverse bijections

$$\{\text{Zariski closed subsets of Spec } R\} \underset{\mathcal{Z}}{\overset{\mathcal{I}}{\rightleftarrows}} \{\text{radical ideals of } R\}.$$

Under this correspondence the closed points in Spec $R$ correspond to the maximal ideals in $R$, and the irreducible subsets of Spec $R$ correspond to the prime ideals in $R$.

**Examples**

**(1)** If $X = \text{Spec } \mathbb{Z}$ then $X$ is irreducible and the nonzero primes give closed points in $X$. The point (0) is not a closed point, in fact the closure of (0) is all of $X$, i.e., (0) is *dense* in Spec $\mathbb{Z}$. For this reason the element (0) is called a *generic point* in Spec $\mathbb{Z}$.

Since every ideal of $\mathbb{Z}$ is principal, the Zariski closed sets in Spec $\mathbb{Z}$ are $\emptyset$, Spec $\mathbb{Z}$ and any finite set of nonzero prime ideals in $\mathbb{Z}$.

**(2)** Suppose $X = \text{Spec } \mathbb{Z}[x]$ as in Example 3 previously. For each integer prime $p$ the Zariski closure of the element $(p) \in X$ consists of the maximal ideals $(p, g)$ of type (d). Likewise for each $\mathbb{Q}$-irreducible polynomial $f$ of type (c), the Zariski closure of the element $(f)$ is the collection of prime ideals of type (d) where $g$ is some divisor of $f$ in $\mathbb{Z}/p\mathbb{Z}[x]$.

**Example: (Affine $k$-algebras)**

Suppose $R = k[V]$ is the coordinate ring of some affine algebraic set $V \subseteq \mathbb{A}^n$ over an algebraically closed field $k$. Then $R = k[x_1, \ldots, x_n]/\mathcal{I}(V)$ where $\mathcal{I}(V)$ is a radical ideal in $k[x_1, \ldots, x_n]$. In particular $R$ is a finitely generated $k$-algebra and since $\mathcal{I}(V)$ is radical, $R$ contains no nonzero nilpotent elements.

**Definition.** A finitely generated algebra over an algebraically closed field $k$ having no nonzero nilpotent elements is called an *affine $k$-algebra.*

If $R$ is an affine $k$-algebra, then by Corollary 5 there is a surjective $k$-algebra homomorphism $\pi : k[x_1, \ldots, x_n] \to R$ whose kernel $I = \ker \pi$ must be a radical ideal since $R$ has no nonzero nilpotent elements. Let $V = \mathcal{Z}(I) \subseteq \mathbb{A}^n$. Then $R \cong k[x_1, \ldots, x_n]/I = k[V]$ is the coordinate ring of an affine algebraic set over $k$. Hence *affine $k$-algebras are precisely the rings arising as the rings of functions on affine algebraic sets over algebraically closed fields.*

By the Nullstellensatz, the points of mSpec $R$ are in bijective correspondence with $V$, and the points of Spec $R$ are in bijective correspondence with the subvarieties of $V$. By Theorem 6, morphisms between two affine algebraic sets correspond bijectively with ($k$-algebra) homomorphisms of affine $k$-algebras. In the language of categories these results show that over an algebraically closed field $k$ there is an equivalence of categories

$$\left\{ \begin{array}{c} \text{affine algebraic sets} \\ \text{morphisms of algebraic sets} \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \text{affine } k\text{-algebras} \\ k\text{-algebra homomorphisms} \end{array} \right\}.$$

The map from left to right sends the affine algebraic set $V$ to its coordinate ring $k[V]$. The map from right to left sends the affine $k$-algebra $R$ to mSpec $R$. The pair (mSpec $R$, $R$) is sometimes called the *canonical model* of the affine $k$-algebra $R$.

Over an algebraically closed field $k$, a $k$-algebra homomorphism $\varphi : R \to S$ between two affine $k$-algebras as in the previous example has the property (by the Nullstellensatz) that the inverse image of a maximal ideal in $S$ is a maximal ideal in $R$. As previously mentioned, one reason for considering Spec $R$ rather than just mSpec $R$ for more general rings is that inverse images of maximal ideals under ring homomorphisms are not in general maximal ideals. When $R$ is an affine $k$-algebra corresponding to an affine algebraic set $V$, the space Spec $R$ contains not only the "geometric points" of $V$ (in the form of the closed points in Spec $R$), but also the non-closed points corresponding to all of the subvarieties of $V$ (in the form of the non-closed points in Spec $R$, i.e., the prime ideals $P$ of $R$ that are not maximal).

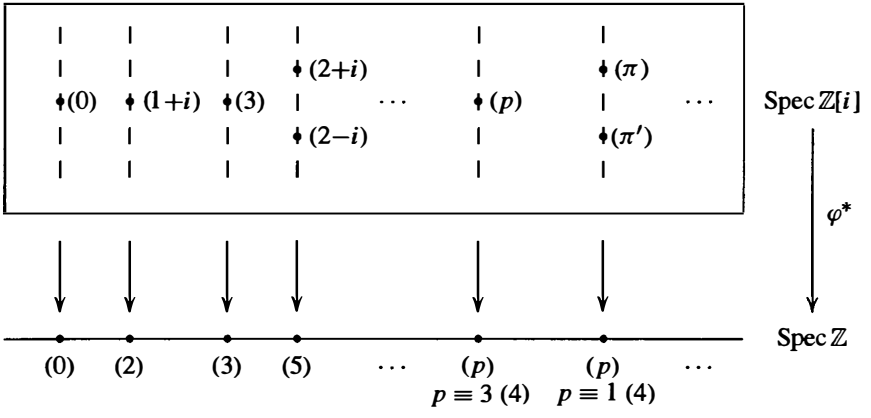
In general, if $\varphi : R \to S$ is a ring homomorphism mapping $1_R$ to $1_S$ and $P$ is a prime ideal in $S$ then $\varphi^{-1}(P)$ is a prime ideal in $R$. This defines a map $\varphi^* : \text{Spec } S \to \text{Spec } R$ with $\varphi^*(P) = \varphi^{-1}(P)$. If $\mathcal{Z}(I) \subseteq \text{Spec } R$ is a Zariski closed subset of Spec $R$, then it is easy to show that $(\varphi^*)^{-1}(\mathcal{Z}(I))$ is the Zariski closed subset $\mathcal{Z}(\varphi(I)S)$ defined by the ideal generated by $\varphi(I)$ in $S$. Since the inverse image of a closed subset in Spec $R$ is a closed subset in Spec $S$, the induced map $\varphi^*$ is continuous in the Zariski topology. This proves the following proposition.

**Proposition 55.** Every ring homomorphism $\varphi : R \to S$ mapping $1_R$ to $1_S$ induces a map $\varphi^* : \text{Spec } S \to \text{Spec } R$ that is continuous with respect to the Zariski topologies on Spec $R$ and Spec $S$.

While the generalization from affine algebraic sets to Spec $R$ for general rings $R$ has made matters slightly more complicated, there are (at least) two very important benefits gained by this more general setting. The first is that Spec $R$ can be considered even for commutative rings $R$ containing nilpotent elements; the second is that Spec $R$ need not be a $k$-algebra for any field $k$, and even when it is, the field $k$ need not be algebraically closed. The fact that many of the properties found in the situation of affine $k$-algebras hold in more general settings then allows the application of "geometric" ideas to these situations (for example, to Spec $R$ when $R$ is finite).

**Examples**

**(1)** The natural inclusion $\varphi : \mathbb{Z} \to \mathbb{Z}[i]$ induces a map $\varphi^* : \operatorname{Spec}\mathbb{Z}[i] \to \operatorname{Spec}\mathbb{Z}$. The fiber of $\varphi^*$ over the nonzero prime $P$ in $\mathbb{Z}$ consists of the prime ideals of $\mathbb{Z}[i]$ containing $P$. If $P = (p)$ where $p = 2$ or $p$ is a prime congruent to 3 mod 4, then there is only one element in this fiber; if $p$ is a prime congruent to 1 mod 4, then there are two elements in the fiber: the primes $(\pi)$ and $(\pi')$ where $p = \pi\pi'$ in $\mathbb{Z}[i]$, cf. Proposition 18 in Section 8.3. This can be represented pictorially in the following figure:



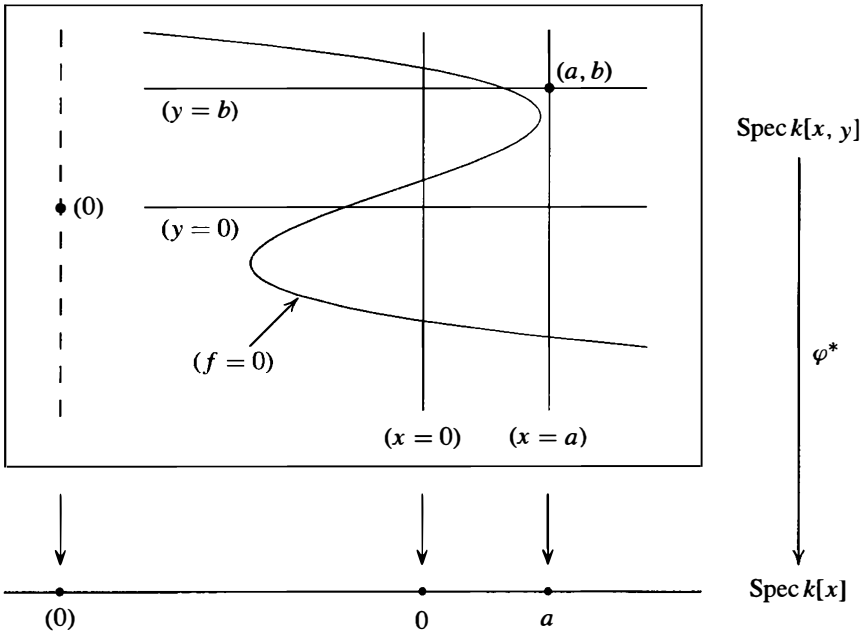

**(2)** If $k$ is an algebraically closed field then Spec $k[x]$ consists of (0) and the ideals $(x-a)$ for $a \in k$; the natural inclusion $\varphi : k[x] \to k[x, y]$ induces the Zariski continuous map $\varphi^* : \operatorname{Spec} k[x, y] \to \operatorname{Spec} k[x]$. The elements of Spec $k[x, y]$ are

**(a)** (0),
**(b)** $(f)$ where $f$ is an irreducible polynomial in $k[x, y]$, and
**(c)** $(x - a, y - b)$ with $a, b \in k$

(cf. Exercise 4). The prime (0) is Zariski dense in Spec $k[x, y]$; the Zariski closure of the primes in (b) consists of the primes $(x - a, y - b)$ in (c) with $f(a, b) = 0$; the closed points, i.e., the elements of mSpec $k[x, y]$, are the primes in (c).

By the Nullstellensatz, each prime ideal $P$ in Spec $k[x, y]$ is uniquely determined by the corresponding zero set $\mathcal{Z}(P)$. The prime (0) $\in k[x, y]$ corresponds to $\mathbb{A}^2$. The prime $(f)$ corresponds to the points where $f(x, y) = 0$, and $P = (f)$ is the intersection of all the maximal ideals containing $P$. The maximal ideal $(x - a, y - b)$ corresponds to the point $(a, b) \in \mathbb{A}^2$. Fibered over Spec $k[x]$ by the map $\varphi^*$ these primes can be pictured geometrically as in the diagram on the following page.

In this diagram, the prime $(x - a)$ in Spec $k[x]$ is identified with the element $a \in k$. The prime $(x) \in \operatorname{Spec} k[x, y]$ corresponds to the points in $\mathbb{A}^2$ with $x = 0$, i.e.,



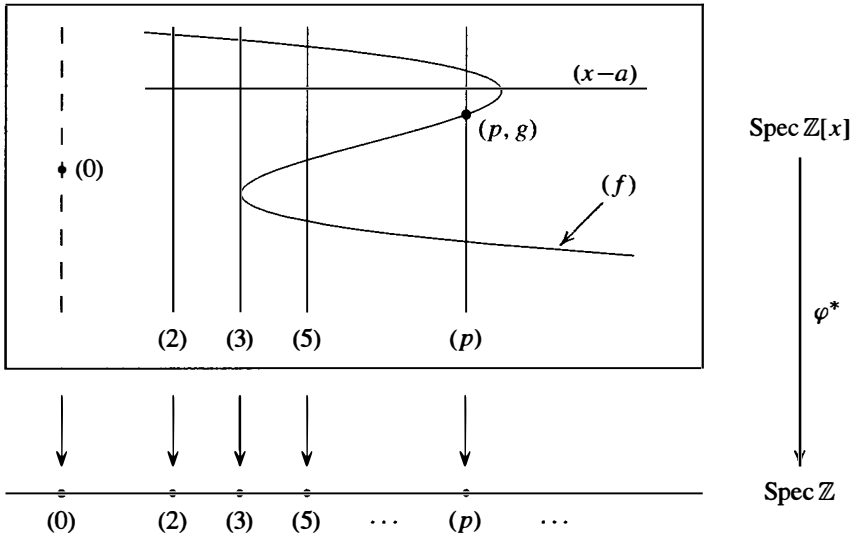

with the $y$-axis in $\mathbb{A}^2$; the prime $(y) \in \operatorname{Spec} k[x, y]$ similarly corresponds to the $x$-axis. The prime $(f) \in \operatorname{Spec} k[x, y]$ corresponds to the irreducible curve $f(x, y) = 0$ in $\mathbb{A}^2$; the points $(a, b) \in \mathbb{A}^2$ lying on this curve correspond to the maximal ideals $(x - a, y - b) \in \operatorname{Spec} k[x, y]$ containing $(f)$. The closed point $(x - a, y - b) \in \operatorname{Spec} k[x, y]$ corresponds to the "geometric point" $(a, b) \in \mathbb{A}^2$.

Note that $\operatorname{Spec} k[x, y]$ captures all of the geometry of algebraic sets in $\mathbb{A}^2$: every algebraic set in $\mathbb{A}^2$ is the finite union of some subset of the irreducible algebraic sets corresponding to the elements of $\operatorname{Spec} k[x, y]$ pictured above. With the exception of the everywhere dense point (0), the "geometric" picture of $\operatorname{Spec} k[x, y]$ is precisely the usual geometry of the affine plane $\mathbb{A}^2$. When $k$ is not algebraically closed the situation is slightly more complicated, but the picture is similar, cf. Exercise 4.

**(3)** The situation for $\operatorname{Spec} \mathbb{Z}[x]$, viewed as fibered over $\operatorname{Spec} \mathbb{Z}$ by the natural inclusion $\mathbb{Z} \to \mathbb{Z}[x]$ is very similar to the situation of $\operatorname{Spec} k[x, y]$ in the previous example. The elements of $\operatorname{Spec} \mathbb{Z}[x]$ were discussed in Example 2 following Proposition 54 and can be pictured as in the diagram on the following page.

The element (0) is Zariski dense in $\operatorname{Spec} \mathbb{Z}[x]$. The closure of $(p)$ consists of $(p)$ and all the closed points $(p, g)$ where $g$ is a monic polynomial in $\mathbb{Z}[x]$ that is irreducible mod $p$. The closure of $(f)$ consists of $(f)$ together with the maximal ideals $(p, g)$ that contain $(f)$, which is the same as saying that the image of $f$ in the quotient $\mathbb{Z}[x]/(p, g)$ is 0, i.e., the irreducible polynomial $g$ is a factor of $f$ mod $p$. The closed points, $\operatorname{mSpec} \mathbb{Z}[x]$, are the maximal ideals $(p, g)$.

Note that the maximal ideals $(p, g)$ containing $(f)$ are precisely the closed points in $\operatorname{mSpec} \mathbb{Z}[x]$ in the diagram above where the "function" $f$ on $\operatorname{Spec} \mathbb{Z}[x]$ (taking the prime $P$ to $f(P) = f \bmod P \in \mathbb{Z}[x]/P$) is zero. For example, the polynomial $f = x^3 - 4x^2 + x - 9 \in \mathbb{Z}[x]$ fits the diagram above: $f$ is irreducible in $\mathbb{Z}[x]$, and




over $\mathbb{F}_p$ factors into irreducibles as follows:

$$f \equiv x^3 + x + 1 \bmod 2$$
$$f \equiv x(x+1)^2 \bmod 3$$
$$f \equiv (x+1)(x+2)(x+3) \bmod 5.$$

There is one point in the fiber over (2) intersecting $(f)$, namely the closed point $(2, x^3 + x + 1)$. There are two closed points in the fiber over (3) given by $(3, x)$ and $(3, x+1)$ (with some "multiplicity" at the latter point). Over (5) there are three closed points: $(5, x+1)$, $(5, x+2)$, and $(5, x+3)$. For the diagram above, the prime $p$ might be $p = 53$, since this is the first prime $p$ greater than 5 for which this polynomial has three irreducible factors mod $p$. Note that while the prime $(f)$ is drawn as a smooth curve in this diagram to emphasize the geometric similarity with the structure of Spec $k[x, y]$ in the previous example, the fibers above the primes in Spec $\mathbb{Z}$ are discrete, so some care should be exercised. For example, since $f$ factors as $(x+2)(x^2+x+6)$ mod 7, the intersection of $(f)$ with the fiber above (7) contains only the two points $(7, x+2)$ and $(7, x^2+x+6)$, each with multiplicity one.

The possible number of closed points in $(f)$ lying in a fiber over $(p) \in$ Spec $\mathbb{Z}$ is controlled by the Galois group of the polynomial $f$ over $\mathbb{Q}$ (cf. Section 14.8). For example, $f = x^4 + 1$ has one closed point in the fiber above (2) and either two or four closed points in a fiber above $(p)$ for $p$ odd (cf. Exercise 8).

The space Spec $R$ together with its Zariski topology gives a geometric generalization for arbitrary commutative rings of the points in a variety $V$. We now consider the question of generalizing the ring of rational functions on $V$.

When $V$ is a variety over the algebraically closed field $k$ the elements in the quotient field $k(V)$ of the coordinate ring $k[V]$ define the rational functions on $V$. Each element $\alpha$ in $k(V)$ can in general be written as a quotient $a/f$ of elements $a, f \in k[V]$ in many different ways. The set of points $U$ at which $\alpha$ is regular is an open subset of $V$; by definition, it consists of all the points $v \in V$ where $\alpha$ can be represented by

some quotient $a/f$ with $f(v) \neq 0$, and then the representative $a/f$ defines an element in the local ring $\mathcal{O}_{v,V}$. Note also that the same representative $a/f$ defines $\alpha$ not only at $v$, but also at all the other points where $f$ is nonzero, namely on the open subset $V_f = \{w \in V \mid f(w) \neq 0\}$ of $V$. These open sets $V_f$ (called principal open sets, cf. Exercise 21 in Section 2) for the various possible representatives $a/f$ for $\alpha$ give an open cover of $U$. The example of the function $\alpha = \bar{x}/\bar{y}$ for $V = \mathcal{Z}(xz - yw) \subset \mathbb{A}^4$ preceding Proposition 51 shows that in general a single representative for $\alpha$ does not suffice to determine all of $U$ — for this example, $U = V_{\bar{y}} \cup V_{\bar{z}}$, and $U$ is not covered by any single $V_f$ (cf. Exercise 25 of Section 4).

This interpretation of rational functions as functions that are regular on open subsets of $V$ can be generalized to Spec $R$. We first define the analogues $X_f$ in $X = \text{Spec } R$ of the sets $V_f$ and establish their basic properties.

**Definition.** For any $f \in R$ let $X_f$ denote the collection of prime ideals in $X = \text{Spec } R$ that do not contain $f$. Equivalently, $X_f$ is the set of points of Spec $R$ at which the value of $f \in R$ is nonzero. The set $X_f$ is called a *principal* (or *basic) open set* in Spec $R$.

Since $X_f$ is the complement of the Zariski closed set $\mathcal{Z}(f)$ it is indeed an open set in Spec $R$ as the name implies. Some basic properties of the principal open sets are indicated in the next proposition. Recall that a map between topological spaces is a *homeomorphism* if it is continuous and bijective with continuous inverse.

**Proposition 56.** Let $f \in R$ and let $X_f$ be the corresponding principal open set in $X = \text{Spec } R$. Then

**(1)** $X_f = X$ if and only if $f$ is a unit, and $X_f = \emptyset$ if and only if $f$ is nilpotent,

**(2)** $X_f \cap X_g = X_{fg}$,

**(3)** $X_f \subseteq X_{g_1} \cup \cdots \cup X_{g_n}$ if and only if $f \in \text{rad}(g_1, \ldots, g_n)$; in particular $X_f = X_g$ if and only if $\text{rad}(f) = \text{rad}(g)$,

**(4)** the principal open sets form a basis for the Zariski topology on Spec $R$, i.e., every Zariski open set in $X$ is the union of some collection of principal open sets $X_f$,

**(5)** the natural map from $R$ to $R_f$ induces a homeomorphism from Spec $R_f$ to $X_f$, where $R_f$ is the localization of $R$ at $f$,

**(6)** the spectrum of any ring is quasicompact (i.e., every open cover has a finite subcover); in particular, $X_f$ is quasicompact, and

**(7)** if $\varphi : R \to S$ is any homomorphism of rings (with $\varphi(1_R) = 1_S$) then under the induced map $\varphi^* : Y = \text{Spec } S \to \text{Spec } R$ the full preimage of the principal open set $X_f$ in $X$ is the principal open set $Y_{\varphi(f)}$ in $Y$.

*Proof:* Parts (1), (2) and (7) are left as easy exercises. For (3), observe that, by definition, $X_{g_1} \cup \cdots \cup X_{g_n}$ consists of the primes $P$ not containing at least one of $g_1, \ldots, g_n$. Hence $X_{g_1} \cup \cdots \cup X_{g_n}$ is the complement of the closed set $\mathcal{Z}((g_1, \ldots, g_n))$ consisting of the primes $P$ that contain the ideal generated by $g_1, \ldots, g_n$. If $(g_1, \ldots, g_n) = R$ then $X_{g_1} \cup \cdots \cup X_{g_n} = X$ and there is nothing to prove. Otherwise, $X_f \subseteq X_{g_1} \cup \cdots \cup X_{g_n}$ if and only if every prime $P$ with $f \notin P$ also satisfies $P \notin \mathcal{Z}((g_1, \ldots, g_n))$. This latter condition is equivalent to the statement that if the prime $P$ contains the ideal

$(g_1, \ldots, g_n)$ then $P$ also contains $f$, i.e., $f$ is contained in the intersection of all the prime ideals $P$ containing $(g_1, \ldots, g_n)$. Since this intersection is $\text{rad}(g_1, \ldots, g_n)$ by Proposition 12, this proves (3).

If $U = X - \mathcal{Z}(I)$ is a Zariski open subset of $X$, then $U$ is the union of the sets $X_f$ with $f \in I$, which proves (4).

The natural ring homomorphism from $R$ to the localization $R_f$ establishes a bijection between the prime ideals in $R_f$ and the prime ideals in $R$ not containing $(f)$ (Proposition 38). The corresponding Zariski continuous map from Spec $R_f$ to Spec $R$ is therefore continuous and bijective. Since every ideal of $R_f$ is the extension of some ideal of $R$ (cf. Proposition 38(1)), it follows that the inverse map is also continuous, which proves (5).

In (6), every open set is the union of principal open sets by (4), so it suffices to prove that if $X$ is covered by principal open sets $X_{g_i}$ (for $i$ in some index set $\mathcal{J}$) then $X$ is a finite union of some of the $X_{g_i}$. If the ideal $I$ generated by the $g_i$ were a proper ideal in $R$, then $I$ would be contained in some maximal ideal $P$. But in this case the element $P$ in $X = \text{Spec } R$ would not be contained in any principal open set $X_{g_i}$, contradicting the assumption that $X$ is covered by the $X_{g_i}$. Hence $I = R$ and so $1 \in R$ can be written as a finite sum $1 = a_1 g_{i_1} + \cdots + a_n g_{i_n}$ with $i_1, \ldots, i_n \in \mathcal{J}$. Consider the finite union $X_{g_1} \cup \cdots \cup X_{g_n}$. Any point $P$ in $X$ not contained in this union would be a prime in $R$ that contains $g_{i_1}, \ldots, g_{i_n}$, hence would contain 1, a contradiction. It follows that $X = X_{g_1} \cup \cdots \cup X_{g_n}$ as needed. The second part of (6) follows from (5).

We now define an analogue for $X = \text{Spec } R$ of the rational functions on a variety $V$. As we observed, for the variety $V$ a rational function $\alpha \in k(V)$ is a regular function on some open set $U$. At each point $v \in U$ there is a representative $a/f$ for $\alpha$ with $f(v) \neq 0$, and this representative is an element in the localization $\mathcal{O}_{v,V} = k[V]_{\mathcal{I}(v)}$. In this way the regular function $\alpha$ on $U$ can be considered as a function from $U$ to the disjoint union of these localizations: the point $v \in U$ is mapped to the representative $a/f \in k[V]_{\mathcal{I}(v)}$. Furthermore the same representative can be used simultaneously not only at $v$ but on the whole Zariski neighborhood $V_f$ of $v$ (so, "locally near $v$," $\alpha$ is given by a single quotient of elements from $k[V]$). Note that $a/f$ is an element in the localization $k[V]_f$, which is contained in each of the localizations $k[V]_{\mathcal{I}(w)}$ for $w \in V_f$.

We now generalize this to Spec $R$ by considering the collection of functions $s$ from the Zariski open subset $U$ of Spec $R$ to the disjoint union of the localizations $R_P$ for $P \in U$ such that $s(P) \in R_P$ and such that $s$ is given locally by quotients of elements of $R$. More precisely:

**Definition.** Suppose $U$ is a Zariski open subset of Spec $R$. If $U = \emptyset$, define $\mathcal{O}(U) = 0$. Otherwise, define $\mathcal{O}(U)$ to be the set of functions $s : U \to \bigsqcup_{Q \in U} R_Q$ from $U$ to the disjoint union of the localizations $R_Q$ for $Q \in U$ with the following two properties:

**(1)** $s(Q) \in R_Q$ for every $Q \in U$, and

**(2)** for every $P \in U$ there is an open neighborhood $X_f \subseteq U$ of $P$ in $U$ and an element $a/f^n$ in the localization $R_f$ defining $s$ on $X_f$, i.e., $s(Q) = a/f^n \in R_Q$ for every $Q \in X_f$.

If $s, t$ are elements in $\mathcal{O}(U)$ then $s + t$ and $st$ are also elements in $\mathcal{O}(U)$ (cf. Exercise 18), so each $\mathcal{O}(U)$ is a ring. Also, every $a \in R$ gives an element in $\mathcal{O}(U)$

defined by $s(Q) = a \in R_Q$, and in particular $1 \in R$ gives an identity for the ring $\mathcal{O}(U)$. If $U'$ is an open subset of $U$, then there is a natural restriction map from $\mathcal{O}(U)$ to $\mathcal{O}(U')$ which is a homomorphism of rings (cf. Exercise 19).

**Definition.** Let $R$ be a commutative ring with 1, and let $X = \operatorname{Spec} R$.

**(1)** The collection of rings $\mathcal{O}(U)$ for the Zariski open sets of $X$ together with the restriction maps $\mathcal{O}(U) \to \mathcal{O}(U')$ for $U' \subseteq U$ is called the *structure sheaf* on $X$, and is denoted simply by $\mathcal{O}$ (or $\mathcal{O}_X$).

**(2)** The elements $s$ of $\mathcal{O}(U)$ are called the *sections of* $\mathcal{O}$ over $U$. The elements of $\mathcal{O}(X)$ are called the *global sections* of $\mathcal{O}$.

The next proposition generalizes the result of Proposition 51 that the only rational functions on a variety $V$ that are regular everywhere are the elements of the coordinate ring $k[V]$.

**Proposition 57.** Let $X = \operatorname{Spec} R$ and let $\mathcal{O} = \mathcal{O}_X$ be its structure sheaf. The global sections of $\mathcal{O}$ are the elements of $R$, i.e., $\mathcal{O}(X) \cong R$. More generally, if $X_f$ is a principal open set in $X$ for some $f \in R$, then $\mathcal{O}(X_f)$ is isomorphic to the localization $R_f$.

*Proof:* Suppose that $a/f^n$ is an element of the localization $R_f$. Then the map defined by $s(Q) = a/f^n \in R_Q$ for $Q \in X_f$ gives an element in $\mathcal{O}(X_f)$, and it is immediate that the resulting map $\psi$ from $R_f$ to $\mathcal{O}(X_f)$ is a ring homomorphism. Suppose that $a/f^n = b/f^m$ in $R_Q$ for every $Q \in X_f$, i.e., $g(af^m - bf^n) = 0$ in $R$ for some $g \notin Q$. If $I$ is the ideal in $R$ of elements $r \in R$ with $r(af^m - bf^n) = 0$, it follows from $g \in I$ that $I$ is not contained in $Q$ for any $Q \in X_f$. Put another way, every prime ideal of $R$ containing $I$ also contains $f$. Hence $f$ is contained in the intersection of all the prime ideals of $R$ containing $I$, which is to say that $f \in \operatorname{rad} I$. Then $f^N \in I$ for some integer $N \geq 0$, and so $f^N(af^m - bf^n) = 0$ in $R$. But this shows that $a/f^n = b/f^m$ in $R_f$ and so the map $\psi$ is injective. Suppose now that $s \in \mathcal{O}(X_f)$. Then by definition $X_f$ can be covered by principal open sets $X_{g_i}$ on which $s(Q) = a_i/g_i^{n_i} \in R_Q$ for every $Q \in X_{g_i}$. By (6) of Proposition 56, we may take a finite number of the $g_i$ and then by taking different $a_i$ we may assume all the $n_i$ are equal (since $a_i/g_i^{n_i} = (a_i g_i^{n-n_i})/g_i^n$ if $n$ is the maximum of the $n_i$). Since $s(Q) = a_i/g_i^n = a_j/g_j^n$ in $R_Q$ for all $Q \in X_{g_i g_j} = X_{g_i} \cap X_{g_j}$, the injectivity of $\psi$ (applied to $R_{g_i g_j}$) shows that $a_i/g_i^n = a_j/g_j^n$ in $R_{g_i g_j}$. This means that $g_i g_j{}^N(a_i g_j^n - a_j g_i^n) = 0$, i.e.,

$$a_i g_i^N g_j{}^{n+N} = a_j g_i^{n+N} g_j{}^N$$

in $R$ for some $N \geq 0$, and we may assume $N$ sufficiently large that this holds for every $i$ and $j$. Since $X_f$ is the union of the $X_{g_i} = X_{g_i^{n+N}}$, $f$ is contained in the radical of the ideal generated by the $g_i^n$ by (3) of Proposition 56, say

$$f^M = \sum_i b_i g_i^{n+N}$$

for some $M \geq 1$ and $b_i \in R$. Define $a = \sum b_i a_i g_i^N \in R$. Then

$$g_j^N a_j f^M = \sum_i b_i (a_j g_i^{n+N} g_j{}^N) = \sum_i b_i (a_i g_i^N g_j{}^{n+N}) = g_j{}^{n+N} a.$$

It follows that $a/f^M = a_j/g_j^n$ in $R_{g_j}$, and so the element in $\mathcal{O}(X_f)$ defined by $a/f^M$ in $R_f$ agrees with $s$ on every $X_{g_j}$, and so on all of $X_f$ since these open sets cover $X_f$. Hence the map $\psi$ gives an isomorphism $R_f \cong \mathcal{O}(X_f)$. Taking $f = 1$ gives $R \cong \mathcal{O}(X)$, completing the proof.

In the case of affine varieties $V$ the local ring $\mathcal{O}_{v,V}$ at the point $v \in V$ is the collection of all the rational functions in $k(V)$ that are defined at $v$. Put another way, $\mathcal{O}_{v,V}$ is the union of the rings of regular functions on $U$ for the open sets $U$ containing $P$, where this union takes place in the function field $k(V)$ of $V$. In the more general case of $X = \operatorname{Spec} R$, the rings $\mathcal{O}(U)$ for the open sets containing $P \in \operatorname{Spec} R$ are not contained in such an obvious common ring. In this case we proceed by considering the collection of pairs $(s, U)$ with $U$ an open set of $X$ containing $P$ and $s \in \mathcal{O}(U)$. We identify two pairs $(s, U)$ and $(s', U')$ if there is an open set $U'' \subseteq U \cap U'$ containing $P$ on which $s$ and $s'$ restrict to the same element of $\mathcal{O}(U'')$. In the situation of affine varieties, this says that two functions defined in Zariski neighborhoods of the point $v$ define the same regular function at $v$ if they agree in some common neighborhood of $v$. The collection of equivalence classes of pairs $(s, U)$ defines the *direct limit* of the rings $\mathcal{O}(U)$, and is denoted $\varinjlim \mathcal{O}(U)$ (cf. Exercise 8 in Section 7.6).

**Definition.** If $P \in X = \operatorname{Spec} R$, then the direct limit, $\varinjlim \mathcal{O}(U)$, of the rings $\mathcal{O}(U)$ for the open sets $U$ of $X$ containing $P$ is called the *stalk* of the structure sheaf at $P$, and is denoted $\mathcal{O}_P$.

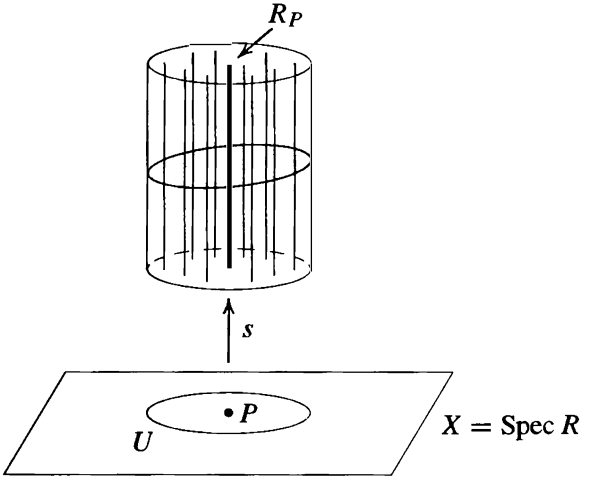
**Proposition 58.** Let $X = \operatorname{Spec} R$ and let $\mathcal{O} = \mathcal{O}_X$ be its structure sheaf. The stalk of $\mathcal{O}$ at the point $P \in X$ is isomorphic to the localization $R_P$ of $R$ at $P$: $\mathcal{O}_P \cong R_P$. In particular, the stalk $\mathcal{O}_P$ is a local ring.

*Proof:* If $(s, U)$ represents an element in the stalk $\mathcal{O}_P$, then $s(P)$ is an element of the localization $R_P$. By the definition of the direct limit, this element does not depend on the choice of representative $(s, U)$, and so gives a well defined ring homomorphism $\varphi$ from $\mathcal{O}_P$ to $R_P$. If $a, f \in R$ with $f \notin P$, then the map $s(Q) = a/f \in R_Q$ defines an element in $\mathcal{O}(X_f)$. Then the class of $(s, X_f)$ in the stalk $\mathcal{O}_P$ is mapped to $a/f$ in $R_P$ by $\varphi$, so $\varphi$ is a surjective map. To see that $\varphi$ is also injective, suppose that the classes of $(s, U)$ and $(s', U')$ in $\mathcal{O}_P$ satisfy $s(P) = s'(P)$ in $R_P$. By definition of $\mathcal{O}(U)$, $s = a/g^n$ on $X_g$ for some $g \notin P$. Similarly, $s' = b/(g')^m$ on $X_{g'}$ for some $g' \notin P$. Since $a/g^n = b/(g')^m$ in $R_P$, there is some $h \notin P$ with $h(a(g')^m - bg^n) = 0$ in $R$. If $Q \in X_{gg'h} = X_g \cap X_{g'} \cap X_h$ this last equality shows that $a/g^n = b/(g')^m$ in $R_Q$, so that $s$ and $s'$ agree when restricted to $X_{gg'h}$. By definition of the direct limit, $(s, U)$ and $(s', U')$ define the same element in the stalk $\mathcal{O}_P$, which proves that $\varphi$ is injective and establishes the proposition.

Proposition 58 shows that the algebraically defined localization $R_P$ for $P \in \operatorname{Spec} R$ plays the role of the local ring $\mathcal{O}_{v,V}$ of regular functions at $v$ for the affine variety $V$. If $\mathfrak{m}_P$ denotes the maximal ideal $PR_P$ in $R_P$ and $k(P) = R_P/\mathfrak{m}_P$ denotes the corresponding quotient field (which by Proposition 46(1) is also the fraction field of $R/P$), then the *tangent space* at $P$ is defined to be the $k(P)$-vector space dual of $\mathfrak{m}_P/\mathfrak{m}_P^2$.

This is an algebraic definition that generalizes the definition of the tangent space $\mathbb{T}_{v,V}$ to a variety $V$ at a point $v$ (by Proposition 52). This can now be used to define what it means for a point in Spec $R$ to be nonsingular: the point $P \in \text{Spec } R$ is *nonsingular* or *smooth* if the local ring $R_P$ is what is called a "regular local ring" (cf. Section 16.2).

Proposition 58 also suggests a nice geometric view of the structure sheaf on Spec $R$. If we view each point $P \in \text{Spec } R$ as having the local ring $R_P$ above it, then above the open set $U$ in $X = \text{Spec } R$ is a "sheaf" (in the sense of a "bundle") of these "stalks" (in the sense of a "stalk of wheat"), which helps explain some of the terminology. A section $s$ in the structure sheaf $\mathcal{O}(U)$ is a map from $U$ to this bundle of stalks. The image of $U$ under such a section $s$ is indicated by the shaded region in the following figure.




**Definition.** Let $R$ be a commutative ring with 1. The pair $(\text{Spec } R, \mathcal{O}_{\text{Spec } R})$, consisting of the space Spec $R$ with the Zariski topology together with the structure sheaf $\mathcal{O}_{\text{Spec } R}$, is called an *affine scheme.*

The notion of an affine scheme gives a completely algebraic generalization of the geometry of affine algebraic sets valid for arbitrary commutative rings, and is the starting point for modern algebraic geometry.

**Examples**

**(1)** If $F$ is any field then $X = \text{Spec } F = \{(0)\}$. In this case there are only two open sets $X$ and $\emptyset$, both of which are principal open sets: $X = X_1$ and $\emptyset = X_0$. The global sections are $\mathcal{O}(X) = F$. There is only one stalk: $\mathcal{O}_{(0)} = F_0 = F$.

**(2)** If $R = \mathbb{Z}$ then because $R$ is a P.I.D. every open set in $X = \text{Spec } \mathbb{Z}$ is principal open:

$$X_n = \{(p) \mid p \nmid n\} \qquad \text{and}$$

$$\mathcal{O}(X_n) = \mathbb{Z}_n = \mathbb{Z}[1/n] = \{a/b \in \mathbb{Q} \mid \text{if the prime } p \mid b \text{ then } p \mid n\}.$$

For nonzero $p$ the stalk at $(p)$ is the local ring $\mathbb{Z}_{(p)}$, and the stalk at (0) is $\mathbb{Q}$. All the restriction maps as well as the maps from sections to stalks are the natural inclusions.

**(3)** For a general integral domain $R$ with quotient field $F$ the stalks and sections are

$$\mathcal{O}(U) = \{a/b \in F \mid b \notin P \text{ for all } P \in U\}$$
$$\mathcal{O}_P = R_P = \{a/b \in F \mid b \notin P\}$$

where the stalk at (0) is $F$, i.e., $\mathcal{O}_{(0)} = F$. Again, the restriction maps and the maps to the stalks are all inclusions.

**(4)** For the local ring $R = \mathbb{Z}_{(2)} = \{a/b \in \mathbb{Q} \mid b \text{ odd}\}$ we have $\operatorname{Spec} R = \{(0), (2)\}$ with (2) the only closed point and $\{(0)\} = X_2$ a principal open set. The sections $\mathcal{O}(\{(0)\})$ are $R_2 = \mathbb{Q}$, and the stalks are $\mathcal{O}_{(0)} = R_{(0)} = \mathbb{Q}$ and $\mathcal{O}_{(2)} = R_{(2)} = R$.

We next consider the relationship of the affine schemes corresponding to rings $R$ and $S$ with respect to a ring homomorphism from $R$ to $S$.

Suppose that $\varphi : R \to S$ is a ring homomorphism. We have already seen in Proposition 56(7) that there is an induced continuous map $\varphi^*$ from $Y = \operatorname{Spec} S$ to $X = \operatorname{Spec} R$ and that under this map the full preimage of the principal open set $X_g$ for $g \in R$ is the principal open set $Y_{\varphi(g)}$. It follows that $\varphi$ also induces a map on corresponding sections, as follows. Let $Q' \in Y$ be any element in $\operatorname{Spec} S$ and let $Q = \varphi^*(Q') = \varphi^{-1}(Q') \in X$ be the corresponding element in $\operatorname{Spec} R$. If $U$ is a Zariski open set in $X$ containing $Q$, then $U' = (\varphi^*)^{-1}(U)$ is a Zariski open set in $Y$ containing $Q'$. Note that $\varphi$ induces a natural ring homomorphism, $\varphi_Q$ say, from the localization $R_Q$ to the localization $S_{Q'}$ defined by $\varphi_Q(a/f) = \varphi(a)/\varphi(f) \in S_{Q'}$ for $f \notin Q$. Let $s \in \mathcal{O}_X(U)$ be a section of the structure sheaf of $X$ given locally in the neighborhood $X_g$ of $P \in X$ by $a/g^n$. It is easy to check that the composite

$$s' : U' \xrightarrow{\varphi^*} U \xrightarrow{s} \bigsqcup_{Q \in U} R_Q \xrightarrow{\varphi} \bigsqcup_{Q' \in U} S_{Q'}$$

defines a map given locally in the neighborhood $Y_{\varphi(g)}$ by the element $\varphi(a)/\varphi(g)^n$, so that $s' \in \mathcal{O}_Y(U')$ is a section of the structure sheaf of $Y$. It is then straightforward to check that the resulting map $\varphi^\# : \mathcal{O}_X(U) \to \mathcal{O}_Y(U')$ is a ring homomorphism (mapping $1 \in \mathcal{O}_X(U)$ to $1 \in \mathcal{O}_Y(U')$) that is compatible with the restriction maps on $\mathcal{O}_X$ and $\mathcal{O}_Y$ (cf. Exercise 20). It also follows that there is an induced ring homomorphism on the stalks: $\varphi^\# : \mathcal{O}_{X,P} \to \mathcal{O}_{Y,P'}$ for any point $P' \in \operatorname{Spec} S$ and corresponding point $P = \varphi^*(P') \in \operatorname{Spec} R$. Under the isomorphism in Proposition 58, the homomorphism $\varphi^\#$ from $R_P \cong \mathcal{O}_{X,P}$ to $S_{P'} \cong \mathcal{O}_{Y,P'}$ is just the natural ring homomorphism $\varphi_P$ on the localizations induced by the homomorphism $\varphi$. In particular, the inverse image under $\varphi^\#$ of the maximal ideal in the local ring $\mathcal{O}_{Y,P'}$ is the maximal ideal in the local ring $\mathcal{O}_{X,P}$.

**Definition.** Suppose $(\operatorname{Spec} R, \mathcal{O}_{\operatorname{Spec} R})$ and $(\operatorname{Spec} S, \mathcal{O}_{\operatorname{Spec} S})$ are two affine schemes. A *morphism of affine schemes* from $(\operatorname{Spec} S, \mathcal{O}_{\operatorname{Spec} S})$ to $(\operatorname{Spec} R, \mathcal{O}_{\operatorname{Spec} R})$ is a pair $(\varphi^*, \varphi^\#)$ such that

**(1)** $\varphi^* : \operatorname{Spec} S \to \operatorname{Spec} R$ is Zariski continuous,

**(2)** there are ring homomorphisms $\varphi^\# : \mathcal{O}(U) \to \mathcal{O}(\varphi^{*-1}(U))$ for every Zariski open subset $U$ in $\operatorname{Spec} R$ that commute with the restriction maps, and

**(3)** if $P' \in \operatorname{Spec} S$ with corresponding point $P = \varphi^*(P) \in \operatorname{Spec} R$, then under the induced homomorphism on stalks $\varphi^\# : \mathcal{O}_{\operatorname{Spec} R, P} \to \mathcal{O}_{\operatorname{Spec} S, P'}$ the preimage of the maximal ideal of $\mathcal{O}_{\operatorname{Spec} S, P'}$ is the maximal ideal of $\mathcal{O}_{\operatorname{Spec} R, P}$.

A homomorphism $\psi : A \to B$ from the local ring $A$ to the local ring $B$ with the property that the preimage of the maximal ideal of $B$ is the maximal ideal of $A$ is called a *local homomorphism* of local rings. The third condition in the definition is then the statement that the induced homomorphism on stalks is required to be a local homomorphism.

With this terminology, the discussion preceding the definition shows that a ring homomorphism $\varphi : R \to S$ induces a morphism of affine schemes from $(\operatorname{Spec} S, \mathcal{O}_{\operatorname{Spec} S})$ to $(\operatorname{Spec} R, \mathcal{O}_{\operatorname{Spec} R})$.

Conversely, suppose $(\varphi^*, \varphi^\#)$ is a morphism of affine schemes from $(\operatorname{Spec} S, \mathcal{O}_{\operatorname{Spec} S})$ to $(\operatorname{Spec} R, \mathcal{O}_{\operatorname{Spec} R})$. Then in particular, for $U = \operatorname{Spec} R$, $(\varphi^*)^{-1}(U) = \operatorname{Spec} S$, so by assumption there is a ring homomorphism $\varphi^\# : \mathcal{O}_{\operatorname{Spec} R}(\operatorname{Spec} R) \to \mathcal{O}_{\operatorname{Spec} S}(\operatorname{Spec} S)$ defined on the global sections. By Proposition 57, we have $\mathcal{O}_{\operatorname{Spec} R}(\operatorname{Spec} R) \cong R$ and $\mathcal{O}_{\operatorname{Spec} S}(\operatorname{Spec} S) \cong S$ as rings. Composing with these isomorphisms shows that $\varphi^\#$ gives a ring homomorphism $\varphi : R \to S$. By Proposition 58 we have a local homomorphism $\varphi^\# : R_P \to S_{P'}$, and by the compatibility with the restriction homomorphisms it follows that the diagram

$$\begin{array}{ccc} R & \xrightarrow{\varphi} & S \\ \downarrow & & \downarrow \\ R_P & \xrightarrow{\varphi^\#} & S_{P'} \end{array}$$

commutes, where the two vertical maps are the natural localization homomorphisms. Since $\varphi^\#$ is assumed to be a local homomorphism, $(\varphi^\#)^{-1}(P'S_{P'}) = PR_P$, from which it follows that $\varphi^{-1}(P') = P$. Hence the continuous map from Spec $S$ to Spec $R$ induced by $\varphi$ is the same as $\varphi^*$, and it follows easily that $\varphi$ also induces the homomorphism $\varphi^\#$. This shows that there is a ring homomorphism $\varphi : R \to S$ inducing both $\varphi^*$ and $\varphi^\#$ as before.

We summarize this in the following proposition:

**Theorem 59.** Every ring homomorphism $\varphi : R \to S$ induces a morphism

$$(\varphi^*, \varphi^\#) : (\operatorname{Spec} S, \mathcal{O}_{\operatorname{Spec} S}) \to (\operatorname{Spec} R, \mathcal{O}_{\operatorname{Spec} R})$$

of affine schemes. Conversely, every morphism of affine schemes arises from such a ring homomorphism $\varphi$.

Theorem 59 is the analogue for Spec $R$ of Theorem 6, which converted geometric questions relating to affine algebraic sets to algebraic questions for their coordinate rings.

The condition that the homomorphism on stalks be a local homomorphism in the definition of a morphism of affine schemes is necessary: a continuous map on the spectra together with a set of compatible ring homomorphisms on sections (hence also on stalks) is not sufficient to force these maps to come from a ring homomorphism.

**Example**

Let $R = \mathbb{Z}_{(2)}$ and $S = \mathbb{Q}$ as in the preceding set of examples. Define $\varphi^* : \operatorname{Spec} \mathbb{Q} \to \operatorname{Spec} \mathbb{Z}_{(2)}$ by $\varphi^*((0)) = (2)$ (which is Zariski continuous). Define $\varphi^\# : \mathcal{O}(\operatorname{Spec} R) \to \mathcal{O}(\operatorname{Spec} S)$ to be the inclusion map $\mathbb{Z}_{(2)} \hookrightarrow \mathbb{Q}$ and define $\varphi^\#$ for all other $U \subseteq \operatorname{Spec} R$ simply to be the zero map. It is straightforward to check that these homomorphisms commute with the restriction maps. This family of maps does *not* arise from a ring homomorphism, however, because on the stalks for $(0) \in \operatorname{Spec} S$ and $\varphi^*((0)) = (2) \in \operatorname{Spec} R$ the induced homomorphism

$$\varphi^\# : \mathcal{O}_{\operatorname{Spec} R,(2)} \hookrightarrow \mathcal{O}_{\operatorname{Spec} S,(0)}$$

is the injection $\mathbb{Z}_{(2)} \hookrightarrow \mathbb{Q}$, which is not a *local* homomorphism (the inverse image of (0) is (0) and not the maximal ideal $2\mathbb{Z}_{(2)}$).

The proof of Theorem 59 shows that a morphism $(\varphi^*, \varphi^\#)$ of affine schemes necessarily comes from the ring homomorphism defined by $\varphi^\#$ on global sections. In this example, the homomorphism on global sections is the inclusion map of $R$ into $S$. The inclusion map from $R$ to $S$ defines a map from Spec $S$ to Spec $R$ that maps $(0) \in \operatorname{Spec} S$ to $(0) \in \operatorname{Spec} R$ and not to $(2) \in \operatorname{Spec} R$, so this map does not agree with the original map $\varphi^*$.

The previous example shows that the converse in Theorem 59 would not be true without the third (local homomorphism) condition in the definition of a morphism of affine schemes. As a result, Theorem 59 shows that the appropriate place to view affine schemes is in the category of *locally ringed spaces*. Roughly speaking, a locally ringed space is a topological space $X$ together with a collection of rings $\mathcal{O}(U)$ for each open subset of $X$ (with a compatible set of homomorphisms from $\mathcal{O}(U)$ to $\mathcal{O}(U')$ if $U' \subseteq U$ and with some local conditions on the sections) such that the stalks $\mathcal{O}_P = \varinjlim \mathcal{O}(U)$ for $P \in U$ are local rings. The morphisms in this category are continuous maps between the topological spaces together with ring homomorphisms between corresponding $\mathcal{O}(U)$ with precisely the same conditions as imposed in the definition of a morphism of affine schemes.

A *scheme* is a locally ringed space in which each point lies in a neighborhood isomorphic to an affine scheme (with some compatibility conditions between such neighborhoods), and is a fundamental object of study in modern algebraic geometry. The affine schemes considered here form the building blocks that are "glued together" to define general schemes in the same way that ordinary Euclidean spaces form the building blocks that are "glued together" to define manifolds in analysis.

## EXERCISES

All rings are assumed commutative with identity, and all ring homomorphisms are assumed to map identities to identities.

1. If $N$ is the nilradical of $R$, prove that Spec $R$ and Spec $R/N$ are homeomorphic. [Show that the natural homomorphism from $R$ to $R/N$ induces a Zariski continuous isomorphism from Spec $R/N$ to Spec $R$.]
2. Let $I$ be an ideal in the ring $R$. Prove that the continuous map from Spec $R/I$ to Spec $R$ induced by the canonical projection homomorphism $R \to R/I$ maps Spec $R/I$ homeomorphically onto the closed set $\mathcal{Z}(I)$ in Spec $R$.

3. Prove that two elements $f, g \in R$ have the same values at all elements $P$ in Spec $R$ if and only if $f - g$ is contained in the nilradical of $R$. In particular, prove that an element in an affine $k$-algebra is uniquely determined by its values.

4. Let $k$ be an arbitrary field, not necessarily algebraically closed. Prove that the prime ideals in $k[x, y]$ (i.e., the elements of Spec $k[x, y]$) are
   **(i)** $(0)$,
   **(ii)** $(f)$ where $f$ is an irreducible polynomial in $k[x, y]$, and
   **(iii)** $(p(x), g(x, y))$ where $p(x)$ is an irreducible polynomial in $k[x]$ and $g(x, y)$ is an irreducible polynomial in $k[x, y]$ that is irreducible modulo $p(x)$, i.e., $g(x, y)$ remains irreducible in the quotient $k[x, y]/(p(x))$.

   Prove that mSpec $k[x, y]$ consists of the primes in (iii). [Use Exercise 20 in Section 1.]

5. Let $\mathfrak{m} = (p(x), g(x, y))$ be a maximal ideal in $k[x, y]$ as in the previous exercise. Show that $K = k[x, y]/\mathfrak{m}$ is an algebraic field extension of $k$, so that $k[x, y]$ can also be viewed as a subring of $K[x, y]$. If $x, y$ are mapped to $\alpha, \beta \in K$, respectively, under the canonical homomorphism $k[x, y] \to k[x, y]/\mathfrak{m}$, prove that $\mathfrak{m} = k[x, y] \cap (x - \alpha, y - \beta) \subseteq K[x, y]$.

6. Describe the elements in Spec $\mathbb{R}[x]$ and Spec $\mathbb{C}[x]$. Describe the elements in Spec $\mathbb{Z}_{(2)}[x]$ where $\mathbb{Z}_{(2)} = \{a/b \in \mathbb{Q} \mid b \text{ is odd}\}$ is the localization of $\mathbb{Z}$ at the prime (2).

7. Let $(f) = (x^5 + x + 1)$ in Spec $\mathbb{Z}[x]$ viewed as fibered over Spec $\mathbb{Z}$ as in Example 3 following Proposition 55. Show that there are two closed points in the fiber over (2), three closed points in the fiber over (5), four closed points in the fiber over (19), and five closed points in the fiber over (211).

8. Let $(f) = (x^4 + 1)$ in Spec $\mathbb{Z}[x]$ viewed as fibered over Spec $\mathbb{Z}$ as in Example 3 following Proposition 55. Prove that there is one closed point in the fiber over (2), four closed points in the fiber over $p$ for $p$ odd, $p \equiv 1 \bmod 8$, and two closed points in the fiber over $p$ for all other odd primes $p$ (cf. Corollary 16 in Section 3 of Chapter 14).

9. Prove that the elements in the fiber over $(p)$ of the Zariski continuous map from Spec $\mathbb{Z}[x]$ to Spec $\mathbb{Z}$ are homeomorphic with the elements in Spec$(\mathbb{Z}[x] \otimes_{\mathbb{Z}} \mathbb{F}_p)$.

10. Let $X =$ Spec $R$ and let $X_f$ be the principal open set corresponding to $f \in R$. Prove that $X_f \cap X_g = X_{fg}$. Prove that $X_f = X$ if and only if $f$ is a unit in $R$, and that $X_f = \emptyset$ if and only if $f$ is nilpotent.

11. If $X_f$ and $X_g$ are principal open sets in $X =$ Spec $R$, prove that the open set $X_f \cup X_g$ is the complement of the closed set $\mathcal{Z}(I)$ where $I = (f, g)$ is the ideal in $R$ generated by $f$ and $g$.

12. Prove that a Zariski open subset $U$ of $X =$ Spec $R$ is quasicompact if and only if $U$ is a finite union of principal open subsets. Give an example of a ring $R$, a Zariski open subset $U$ of Spec $R$, and a Zariski open covering of $U$ that cannot be reduced to a finite subcovering.

13. Let $\varphi : R \to S$ be a homomorphism of rings. Prove that under the induced map $\varphi^*$ from $Y =$ Spec $S$ to $X =$ Spec $R$ the full preimage of the principal open set $X_f$ in $X$ is the principal open set $Y_{\varphi(f)}$ in $Y$.

14. Suppose that $R = R_1 \times R_2$ is the direct product of the rings $R_1$ and $R_2$. Prove that $X =$ Spec $R$ is the disjoint union of open subspaces $X_1$, $X_2$ (which are therefore also closed), where $X_1$ is homeomorphic to Spec $R_1$ and $X_2$ is homeomorphic to Spec $R_2$.

15. Prove that $X =$ Spec $R$ is not connected if and only if $R$ is the direct product of two nonzero rings if and only if $R$ contains an idempotent $e$ with $e \neq 0, 1$ (cf. the previous exercise).

**16.** Prove that $X = \operatorname{Spec} R$ is irreducible (i.e., any two nonempty open subsets have a nontrivial intersection) if and only if $X_f \cap X_g \neq \emptyset$ for any two nonempty principal open sets $X_f$ and $X_g$. Deduce that $X = \operatorname{Spec} R$ is irreducible if and only if the nilradical of $R$ is a prime ideal. [Use Exercise 10.]

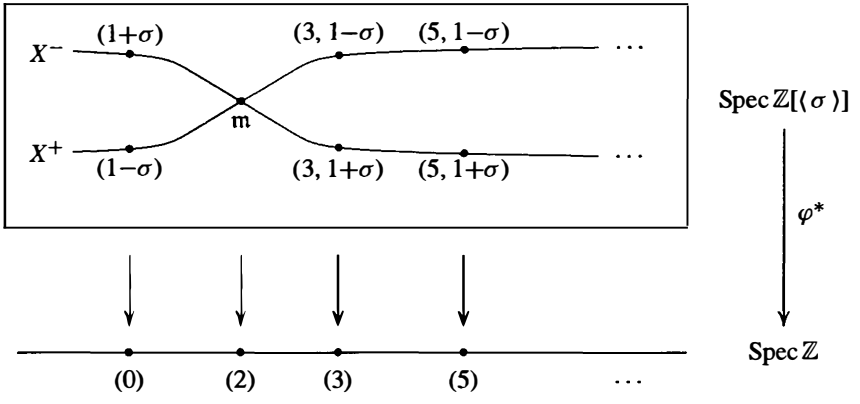
**17.** Let $G = \langle \sigma \rangle$ be a group of order 2, let $R = \mathbb{Z}[G] = \{a + b\sigma \mid a, b \in \mathbb{Z}\}$ be the corresponding group ring, and let $X = \operatorname{Spec} R$.

**(a)** Prove that the nilradical of $R$ is (0) but is not a prime ideal. Prove that $X = X^+ \cup X^-$ where $X^+ = \mathcal{Z}(1 - \sigma)$ and $X^- = \mathcal{Z}(1 + \sigma)$. [Use $(1 + \sigma)(1 - \sigma) = 0$.]

**(b)** Prove that the homomorphism $\mathbb{Z}[G] \to \mathbb{Z}$ defined by mapping $\sigma$ to 1 induces a homeomorphism of $X^+$ with $\operatorname{Spec} \mathbb{Z}$, and the homomorphism mapping $\sigma$ to $-1$ induces a homeomorphism of $X^-$ with $\operatorname{Spec} \mathbb{Z}$.

**(c)** Prove that $X^+ \cap X^-$ consists of the single element $\mathfrak{m} = (1 + \sigma, 1 - \sigma) = (2, 1 - \sigma)$ and that this is a closed point in $X$.

**(d)** Show that $(1 - \sigma)$ and $(1 + \sigma)$ are the unique non-closed points in $X$, with closures $X^+$ and $X^-$, respectively. Describe the closed points, $\operatorname{mSpec} R$, in $X$ and prove that $\operatorname{Spec} \mathbb{Z}[\langle \sigma \rangle]$ can be pictured as follows:




**18.** Let $\mathcal{O}$ be the structure sheaf on $X = \operatorname{Spec} R$, let $U$ be an open set in $X$, and suppose $s, t \in \mathcal{O}(U)$. If $s = a/f_1^n$ on $X_{f_1}$ and $t = b/f_2^m$ on $X_{f_2}$, show that

$$st = (abf_1^m f_2^n)/(f_1 f_2)^{n+m} \quad \text{and} \quad s + t = (af_1^m f_2^{m+n} + bf_1^{m+n} f_2^n)/(f_1 f_2)^{n+m}$$

on $X_{f_1 f_2}$. Deduce that $\mathcal{O}(U)$ is a commutative ring with identity.

**19.** Let $\mathcal{O}$ be the structure sheaf on $X = \operatorname{Spec} R$, let $V \subseteq U$ be open sets in $X$, and let $s \in \mathcal{O}(U)$. Suppose $P \in V$ and that $s = a/f^n$ on $X_f \subseteq U$.

**(a)** Show that there is a principal open set $X_{f'} \subseteq V \cap X_f$ containing $P$.

**(b)** Show that $(f')^m = bf$ for some $b \in R$.

**(c)** Show that $s = (ab^n)/(f')^{mn}$ on $X_{f'}$ and conclude that restricting $s$ to $V$ gives a well defined ring homomorphism from $\mathcal{O}(U)$ to $\mathcal{O}(V)$.

**20.** Let $\varphi : R \to S$ be a homomorphism of rings, let $X = \operatorname{Spec} R$, $Y = \operatorname{Spec} S$, and let $V \subseteq U$ be Zariski open subsets of $X$. Set $V' = (\varphi^*)^{-1}(V)$ and $U' = (\varphi^*)^{-1}(U)$, the corresponding Zariski open subsets of $Y$ with respect to the continuous map $\varphi^* : Y \to X$ induced by $\varphi$. Prove that the induced map $\varphi^\# : \mathcal{O}_X(U) \to \mathcal{O}_Y(U')$ on sections is a ring homomorphism. Prove that $V' \subseteq U'$ and that $\varphi^\#$ is compatible with restriction i.e., that

the diagram

$$\begin{array}{ccc} \mathcal{O}_X(U) & \xrightarrow{\varphi^\#} & \mathcal{O}_Y(U') \\ \downarrow & & \downarrow \\ \mathcal{O}_X(V) & \xrightarrow{\varphi^\#} & \mathcal{O}_Y(V') \end{array}$$

is commutative, where the vertical maps are the restriction homomorphisms.

**21.** Suppose $D$ is a multiplicatively closed subset of $R$. Show that the localization homomorphism $R \to D^{-1}R$ induces a homeomorphism from $\text{Spec}(D^{-1}R)$ to the collection of prime ideals $P$ of $R$ with $P \cap D = \emptyset$.

**22.** Show that $\text{Spec } k[x, y]/(xy)$ is connected but is the union of two proper closed subsets each homeomorphic to $\text{Spec } k[x]$, hence is not irreducible (cf. Exercise 16).

**23.** For each of the following rings $R$ exhibit the elements of $\text{Spec } R$, the open sets $U$ in $\text{Spec } R$, the sections $\mathcal{O}(U)$ of the structure sheaf for $\text{Spec } R$ for each open $U$, and the stalks $\mathcal{O}_P$ at each point $P \in \text{Spec } R$:
**(a)** $\mathbb{Z}/4\mathbb{Z}$ **(b)** $\mathbb{Z}/6\mathbb{Z}$ **(c)** $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$ **(d)** $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.

**24.** **(a)** If every ideal of $R$ is principal, show every open set in $\text{Spec } R$ is a principal open set.
**(b)** Show that if $R = \mathbb{Z}[x]/(4, x^2)$ then $R$ contains a nonprincipal ideal, but every open set in $\text{Spec } R$ is a principal open set.

**25.** **(a)** If $M$ is an $R$-module prove that $\text{Supp}(M)$ is a Zariski closed subset of $\text{Spec } R$. [Use Exercise 33 of Section 4.]
**(b)** If $M$ is a finitely generated $R$-module prove that $\text{Supp}(M) = \mathcal{Z}(\text{Ann}(M)) \subseteq \text{Spec } R$. [Use Exercise 34 of Section 4.]

**26.** Suppose $M$ is a finitely generated module over the Noetherian ring $R$.
**(a)** Prove that there are finitely many minimal primes $*P_1, \ldots, P_n$ containing $\text{Ann}(M)$. [Use Corollary 22.]
**(b)** Prove that $\{P_1, \ldots, P_n\}$ is also the set of minimal primes in $\text{Ass}_R(M)$ and that $\text{Supp}(M)$ is the union of the Zariski closed sets $\mathcal{Z}(P_1), \ldots, \mathcal{Z}(P_n)$ in $\text{Spec } R$. [Use the previous exercise and Exercise 40 in Section 4.]

The previous exercise gives a geometric view of a finitely generated module $M$ over a Noetherian ring $R$: over each point $P$ in $\text{Spec } R$ is the localization $M_P$ (the *stalk* over $P$). The stalk is nonzero precisely over the points in the Zariski closed subsets $\mathcal{Z}(P_1), \ldots, \mathcal{Z}(P_n)$ where the $P_i$ are the minimal primes in $\text{Ass}_R(M)$. These ideas lead to the notion of the *(coherent) module sheaf on* $\text{Spec } R$ *associated to* $M$ (with a picture similar to that of the structure sheaf following Proposition 58), which is a powerful tool in modern algebraic geometry.

**27.** Let $R = k[x, y]$ and let $M$ be the ideal $(x, y)$ in $R$. Prove that $\text{Supp}(M) = \text{Spec } R$ and $\text{Ass}_R(M) = \{0\}$.

The next two exercises show that the associated primes for an ideal $I$ in a Noetherian ring $R$ in the sense of primary decomposition are the associated primes for $I$ in the sense of $\text{Ass}_R(R/I)$.

**28.** This exercise proves that the ideal $Q$ in a Noetherian ring $R$ is $P$-primary if and only if $\text{Ass}_R(R/Q) = \{P\}$.
**(a)** Suppose $Q$ is a $P$-primary ideal and let $M$ be the $R$-module $R/Q$. If $0 \neq m \in M$, show that $Q \subseteq \text{Ann}(m) \subseteq P$ and that $\text{rad Ann}(m) = P$. Deduce that if $\text{Ann}(m)$ is a prime ideal then it is equal to $P$ and hence that $\text{Ass}_R(R/Q) = \{P\}$. [Use Exercise 33 in Section 1.]

**(b)** For any ideal $Q$ of $R$, let $0 \neq M \subseteq R/Q$. Prove that the radical of $\text{Ann}(M)$ is the intersection of the prime ideals in $\text{Supp}(M)$. [Use Proposition 12 and Exercise 25.]

**(c)** For $M$ as in (b), prove that the radical of $\text{Ann}M$ is also the intersection of the prime ideals in $\text{Ass}_R(M)$. [Use Exercise 26(b).]

**(d)** If $Q$ is an ideal of $R$ with $\text{Ass}_R(R/Q) = \{P\}$ prove that $\text{rad}\, Q = P$. [Use the fact that $Q = \text{Ann}(R/Q)$ and (c).]

**(e)** If $Q$ is an ideal of $R$ with $\text{Ass}_R(R/Q) = \{P\}$ prove that $Q$ is $P$-primary. [If $ab \in Q$ with $a \notin Q$ consider $0 \neq M = (Ra + Q)/Q \subseteq R/Q$ and show that $b$ is contained in $\text{Ann}M \subseteq \text{rad}\,\text{Ann}(M)$. Use Exercises 33–34 in Section 1, to show that $\text{Ass}_R(M) = \{P\}$, then use (c) to show that $\text{rad}\,\text{Ann}(M) = P$, and conclude finally that $b \in P$.]

**29.** Suppose $I = Q_1 \cap \cdots \cap Q_n$ is a minimal primary decomposition of the ideal $I$ in the Noetherian ring $R$ with $P_i = \text{rad}\, Q_i$, $i = 1, \ldots, n$. This exercise proves that $\text{Ass}_R(R/I) = \{P_1, \ldots, P_n\}$.

**(a)** Prove that the natural projection homomorphisms induce an injection of $R/I$ into $R/Q_1 \oplus \cdots \oplus R/Q_n$ and deduce that $\text{Ass}_R(R/I) \subseteq \{P_1, \ldots, P_n\}$. [Use Exercise 34 in Section 1 and the previous exercise.]

**(b)** Let $Q_i' = \cap_{j \neq i} Q_j$. Show that the minimality of the decomposition implies that $0 \neq Q_i'/I = (Q_i' + Q_i)/Q_i \subseteq R/Q_i$. Deduce that $\text{Ass}_R(Q_i'/I) = \{P_i\}$. [Use Exercises 33–34 in Section 1 and the previous exercise.] Deduce that $\{P_i\} \in \text{Ass}_R(R/I)$, so that $\text{Ass}_R(R/I) = \{P_1, \ldots, P_n\}$. [Use $Q_i'/I \subseteq R/I$ and Exercise 34 in Section 1.]

**30.** Let $I$ be the ideal $(x^2, xy, xz, yz)$ in $R = k[x, y, z]$. Prove that $\text{Ass}_R(R/I)$ consists of the primes $\{(x, y), (x, z), (x, y, z)\}$.

**31.** *(Spec for Quadratic Integer Rings)* Let $R$ be the ring of integers in the quadratic field $K = \mathbb{Q}(\sqrt{D})$ where $D$ is a squarefree integer and let $P$ be a nonzero prime ideal in $R$. This exercise shows how the prime ideals in $R$ are determined explicitly from the primes $(p)$ in $\mathbb{Z}$, giving in particular a description of Spec $R$ fibered over Spec $\mathbb{Z}$.

As in the discussion and example following Theorem 29, we have $R = \mathbb{Z}[\omega]$ where $\omega = \sqrt{D}$ if $D \equiv 2, 3 \bmod 4$ (respectively, $\omega = (1+\sqrt{D})/2$ if $D \equiv 1 \bmod 4$), with minimal polynomial $m_\omega(x) = x^2 - D$ (respectively, $m_\omega(x) = x^2 - x + (1-D)/4$), and $P \cap \mathbb{Z} = p\mathbb{Z}$ is a nonzero prime ideal of $\mathbb{Z}$.

**(a)** For any prime $p$ in $\mathbb{Z}$ show that $R/pR \cong \mathbb{Z}[x]/(p, m_\omega(x)) \cong \mathbb{F}_p[x]/(\overline{m}_\omega(x))$ as rings, where $\overline{m}_\omega(x)$ is the reduction of $m_\omega(x)$ modulo $p$. Deduce that there is a prime ideal $P$ in $R$ with $P \cap \mathbb{Z} = (p)$ (this gives an alternate proof of Theorem 26(2) in this case).

**(b)** Use the isomorphism in (a) to prove that $P$ is determined explicitly by the factorization of $m_\omega(x)$ modulo $p$:

**(i)** If $\overline{m}_\omega(x) \equiv (x - a)^2 \bmod p$ where $a \in \mathbb{Z}$ then $P = (p, \omega - a)$ and $pR = P^2$. Show that this case occurs only for the finitely many primes $p$ dividing the discriminant of $m_\omega(x)$.

**(ii)** If $\overline{m}_\omega(x) \equiv (x - a)(x - b) \bmod p$ with integers $a, b \in \mathbb{Z}$ that are distinct modulo $p$ then $P$ is either $P_1 = (p, \omega - a)$ or $P_2 = (p, \omega - b)$ and $P_1$, $P_2$ are distinct prime ideals in $R$ with $pR = P_1 P_2$.

**(iii)** If $\overline{m}_\omega(x)$ is irreducible modulo $p$ then $P = pR$.

**(c)** Show that the picture for Spec $R$ over Spec $\mathbb{Z}$ for any $D$ is similar to that for the case $R = \mathbb{Z}[i]$ when $D = -1$: there is precisely one nonclosed point $(0) \in$ Spec $R$ over $(0) \in$ Spec $\mathbb{Z}$, precisely one closed point $P \in$ Spec $R$ over each of the primes $(p)$ in Spec $\mathbb{Z}$ in (i) (called *ramified* primes) and over the primes in (iii) (called *inert* primes), and precisely two closed points over the primes in (ii) (called *split* primes).

CHAPTER 16

# Artinian Rings, Discrete Valuation Rings, and Dedekind Domains

Throughout this chapter $R$ will denote a commutative ring with $1 \neq 0$.

## 16.1 ARTINIAN RINGS

In this section we shall study the basic theory of commutative rings that satisfy the descending chain condition (D.C.C.) on ideals, the Artinian rings (named after E. Artin). While one might at first expect that these rings have properties analogous to those for the commutative rings satisfying the ascending chain condition (the Noetherian rings), in fact this is not the case. The structure of Artinian rings is very restricted; for example an Artinian ring is necessarily also Noetherian (Theorem 3). Noncommutative Artinian rings play a central role in Representation Theory (cf. Chapters 18 and 19).

**Definition.** For any commutative ring $R$ the *Krull dimension* (or simply the *dimension*) of $R$ is the maximum possible length of a chain $P_0 \subset P_1 \subset P_2 \subset \cdots \subset P_n$ of distinct prime ideals in $R$. The dimension of $R$ is said to be infinite if $R$ has arbitrarily long chains of distinct prime ideals.

A ring with finite dimension must satisfy both the ascending and descending chain conditions on prime ideals (although not necessarily on all ideals). A field has dimension 0 and a Principal Ideal Domain that is not a field has dimension 1.

We shall see shortly that rings with D.C.C. on ideals always have dimension 0 (i.e., primes are maximal). If $R$ is an integral domain that is also a finitely generated $k$–algebra over a field $k$, then the dimension of $R$ is equal to the transcendence degree over $k$ of the field of fractions of $R$ (cf. Exercise 11). In particular, the Krull dimension agrees with the definition introduced earlier for the dimension of an affine variety. The advantage of the definition above is that it does not refer to any $k$–algebra structure and applies to arbitrary commutative rings $R$.

**Definition.** The *Jacobson radical* of $R$ is the intersection of all maximal ideals of $R$ and is denoted by Jac $R$.

The Jacobson radical is analogous to the Frattini subgroup of a group, and it enjoys some corresponding properties (cf. Exercise 24 in Section 6.1):

**Proposition 1.** Let $\mathcal{J}$ be the Jacobson radical of the commutative ring $R$.

**(1)** If $I$ is a proper ideal of $R$, then so is $(I, \mathcal{J})$, the ideal generated by $I$ and $\mathcal{J}$.

**(2)** The Jacobson radical contains the nilradical of $R$: $\operatorname{rad} 0 \subseteq \operatorname{Jac} R$.

**(3)** An element $x$ belongs to $\mathcal{J}$ if and only if $1 - rx$ is a unit for all $r \in R$.

**(4)** *(Nakayama's Lemma)* If $M$ is any finitely generated $R$-module and $\mathcal{J}M = M$, then $M = 0$.

*Proof:* If $I$ is a proper ideal in $R$, then $I \subseteq M$ for some maximal ideal $M$. Since $\mathcal{J} \subseteq M$, also $(I, \mathcal{J}) \subseteq M$, which proves (1).

Part (2) follows from the definitions of the two radicals and Proposition 12 in Section 15.2 since maximal ideals are prime.

Suppose $1 - rx$ is not a unit and let $M$ be a maximal ideal containing $1 - rx$. Since $1 \notin M, rx \notin M$, so $x$ cannot belong to $\mathcal{J}$ because $\mathcal{J} \subseteq M$. Conversely, suppose $x \notin \mathcal{J}$, i.e., there is a maximal ideal $M$ with $x \notin M$. Then $R = (x, M)$, hence $1 = rx + y$ for some $y \in M$. Thus $1 - rx = y \in M$ and so $1 - rx$ is not a unit, which proves (3).

To prove (4), assume $M \neq 0$ and let $n$ be the smallest integer such that $M$ is generated by $n$ elements, say $m_1, \ldots, m_n$. Since $M = \mathcal{J}M$ we have

$$m_n = r_1 m_1 + r_2 m_2 + \cdots + r_n m_n \qquad \text{for some } r_1, r_2, \ldots, r_n \in \mathcal{J}.$$

Thus $(1 - r_n)m_n = r_1 m_1 + \cdots + r_{n-1} m_{n-1}$. By (3), $1 - r_n$ is a unit, so $m_n$ lies in the module generated by $m_1, \ldots, m_{n-1}$, contradicting the minimality of $n$. Hence $M = 0$, completing the proof.

**Definition.** A commutative ring $R$ is said to be *Artinian* or to satisfy the *descending chain condition on ideals* (or *D.C.C. on ideals*) if there is no infinite decreasing chain of ideals in $R$, i.e., whenever $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$ is a decreasing chain of ideals of $R$, then there is a positive integer $m$ such that $I_k = I_m$ for all $k \geq m$. Similarly, an $R$-module $M$ is said to be Artinian if it satisfies D.C.C. on submodules.

It is immediate from the Lattice Isomorphism Theorem that every quotient $R/I$ of an Artinian ring $R$ by an ideal $I$ is again an Artinian ring.

The following result for Artinian rings is parallel to results in Theorem 15.2. The proof is completely analogous, and so is left as an exercise.

**Proposition 2.** The following are equivalent:

**(1)** $R$ is an Artinian ring.

**(2)** Every nonempty set of ideals of $R$ contains a minimal element under inclusion.

The next result gives the main structure theorem for Artinian rings.

**Theorem 3.** Let $R$ be an Artinian ring.

**(1)** There are only finitely many maximal ideals in $R$.

**(2)** The quotient $R/(\text{Jac } R)$ is a direct product of a finite number of fields. More precisely, if $M_1, \ldots, M_n$ are the finitely many maximal ideals in $R$ then

$$R/(\text{Jac } R) \cong k_1 \times \cdots \times k_n,$$

where $k_i$ is the field $R/M_i$ for $1 \leq i \leq n$.

**(3)** Every prime ideal of $R$ is maximal, i.e., $R$ has Krull dimension 0. The Jacobson radical of $R$ equals the nilradical of $R$ and is a nilpotent ideal: $(\text{Jac } R)^m = 0$ for some $m \geq 1$.

**(4)** The ring $R$ is isomorphic to the direct product of a finite number of Artinian local rings.

**(5)** Every Artinian ring is Noetherian.

*Proof:* To prove (1), let $\mathcal{S}$ be the set of all ideals of $R$ that are the intersection of a finite number of maximal ideals. By Proposition 2, $\mathcal{S}$ has a minimal element, say $M_1 \cap M_2 \cap \cdots \cap M_n$. Then for any maximal ideal $M$ we have

$$M \cap M_1 \cap M_2 \cap \cdots \cap M_n = M_1 \cap M_2 \cap \cdots \cap M_n,$$

so $M \supseteq M_1 \cap M_2 \cap \cdots \cap M_n$. By Exercise 11 in Section 7.4, $M \supseteq M_i$ for some $i$. Thus $M = M_i$ and so $M_1, \ldots, M_n$ are all the maximal ideals of $R$.

The proof of (2) is immediate from the Chinese Remainder Theorem (Section 7.6) applied to $M_1, \ldots, M_n$, since these maximal ideals are clearly pairwise comaximal and their intersection is Jac $R$.

For (3), we first prove $\mathcal{J} = \text{Jac } R$ is nilpotent. By D.C.C. there is some $m > 0$ such that $\mathcal{J}^m = \mathcal{J}^{m+i}$ for all positive $i$. By way of contradiction assume $\mathcal{J}^m \neq 0$. Let $\mathcal{S}$ be the set of proper ideals $I$ such that $I\mathcal{J}^m \neq 0$, so $\mathcal{J} \in \mathcal{S}$. Let $I_0$ be a minimal element of $\mathcal{S}$. There is some $x \in I_0$ such that $x\mathcal{J}^m \neq 0$, so by minimality we must have $I_0 = (x)$. But now $((x)\mathcal{J})\mathcal{J}^m = x\mathcal{J}^{m+1} = x\mathcal{J}^m$, so it follows by minimality of $(x)$ that $(x) = (x)\mathcal{J}$. By Nakayama's Lemma above, $(x) = 0$, a contradiction. This proves Jac $R$ is nilpotent.

Since Jac $R$ is nilpotent, in particular Jac $R \subseteq$ rad 0, so these two ideals are equal by the second statement in Proposition 1.

Every prime ideal $P$ in $R$ contains the nilradical of $R$, hence contains Jac $R$ by what has already been proved,. The image of $P$ is a prime ideal in the quotient ring $R/(\text{Jac } R) = k_1 \times \cdots \times k_n$. But in a direct product of rings $R_1 \times R_2$ (where each $R_i$ has a 1) every ideal is of the form $I_1 \times I_2$, where $I_j$ is an ideal of $R_j$ for $j = 1, 2$ (cf. Exercise 3 in Section 7.6). It follows that a prime ideal in $k_1 \times \cdots \times k_n$ consists of the elements that are 0 in one of the components. In particular, such a prime ideal is also a maximal ideal in $k_1 \times \cdots \times k_n$ and it follows that $P$ was a maximal ideal in $R$, which finishes the proof of (3).

Let $M_1, \ldots, M_n$ be all the distinct maximal ideals of $R$ and let $(\text{Jac } R)^m = 0$ as in (3). Then

$$\prod_{i=1}^{n} M_i^m \subseteq \left( \prod_{i=1}^{n} M_i \right)^m \subseteq (\text{Jac } R)^m = 0.$$

By the Chinese Remainder Theorem it follows that

$$R \cong (R/M_1^m) \times (R/M_2^m) \times \cdots \times (R/M_n^m),$$

and each $R/M_i^m$ is an Artinian ring with unique maximal ideal $M_i/M_i^m$, proving (4).

To prove (5), it suffices by (4) to prove that an Artinian local ring is Noetherian, so assume $R$ is Artinian with unique maximal ideal $M$. In this case we have $M = \text{Jac}\, R$, so $M^m = (\text{Jac}\, R)^m = 0$ for some positive $m$. Then $R \cong R/M^m$, and in this case it is an exercise to see that $R/M^m$ is Noetherian if and only if it is Artinian (cf. Exercise 8).

**Corollary 4.** The ring $R$ is Artinian if and only if $R$ is Noetherian and has Krull dimension 0.

*Proof:* The forward implication was proved in Theorem 3. Suppose now that $R$ is Noetherian and that $R$ has Krull dimension 0, i.e., that prime ideals of $R$ are maximal. Since $R$ is Noetherian, by Corollary 22(3) in Section 15.2, the ideal $(0) = P_1 \cdots P_n$ is the product of (not necessarily distinct) prime ideals, and these prime ideals are then maximal since $R$ has dimension 0. By the Chinese Remainder Theorem, $R$ is isomorphic to the direct product of a finite number of Noetherian rings of the form $R/M^m$ where $M$ is a maximal ideal in $R$. As in the proof of (5) of the theorem, $R/M^m$ is Artinian, and it follows that $R$ is Artinian.

**Examples**

**(1)** Let $n > 1$ be an integer. Since the ring $R = \mathbb{Z}/n\mathbb{Z}$ is finite, it is Artinian. If $n = p_1^{a_1} p_2^{a_2} \cdots p_s^{a_s}$ is the unique factorization of $n$ into distinct prime powers, then

$$\mathbb{Z}/n\mathbb{Z} \cong (\mathbb{Z}/p_1^{a_1}\mathbb{Z}) \times (\mathbb{Z}/p_2^{a_2}\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p_s^{a_s}\mathbb{Z}).$$

Each $\mathbb{Z}/p_i^{a_i}\mathbb{Z}$ is an Artinian local ring with unique maximal ideal $(p_i)/(p_i^{a_i})$, so this is the decomposition of $\mathbb{Z}/n\mathbb{Z}$ given by Theorem 3(4). The Jacobson radical of $R$ is the ideal generated by $p_1 p_2 \cdots p_s$, the squarefree part of $n$ and $R/(\text{Jac}\, R) \cong (\mathbb{Z}/p_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p_s\mathbb{Z})$ is a direct product of fields. The ideals generated by $p_i$ for $i = 1, \ldots, s$ are the maximal ideals of $R$.

**(2)** For any field $k$, a $k$-algebra $R$ that is finite dimensional as a vector space over $k$ is Artinian because ideals in $R$ are in particular $k$-subspaces of $R$, hence the length of any chain of ideals in $R$ is bounded by $\dim_k R$.

**(3)** Suppose $f$ is a nonzero polynomial in $k[x]$ where $k$ is a field. Then the quotient ring $R = k[x]/(f(x))$ is Artinian by the previous example. The decomposition of $R$ as a direct product of Artinian local rings is given by

$$k[x]/(f(x)) \cong k[x]/(f_1(x)^{a_1}) \times \cdots \times k[x]/(f_s(x)^{a_s})$$

where $f(x) = f_1(x)^{a_1} \cdots f_s(x)^{a_s}$ is the factorization of $f(x)$ into powers of distinct irreducibles in $k[x]$ (cf. Proposition 16 in Section 9.5). The Jacobson radical of $R$ is the ideal generated by the squarefree part of $f(x)$ and the maximal ideals of $R$ are the ideals generated by the irreducible factors $f_i(x)$ for $i = 1, \ldots, s$ similar to Example 1.

## EXERCISES

Let $R$ be a commutative ring with 1 and let $\mathcal{J}$ be its Jacobson radical.

1. Suppose $R$ is an Artinian ring and $I$ is an ideal in $R$. Prove that $R/I$ is also Artinian.
2. Show that every finite commutative ring with 1 is Artinian.
3. Prove that an integral domain of Krull dimension 0 is a field.
4. Prove that an Artinian integral domain is a field.
5. Suppose $I$ is a nilpotent ideal in $R$ and $M = IM$ for some $R$-module $M$. Prove that $M = 0$.
6. Suppose that $0 \longrightarrow M' \longrightarrow M \longrightarrow M'' \longrightarrow 0$ is an exact sequence of $R$-modules. Prove that $M$ is an Artinian $R$-module if and only if $M'$ and $M''$ are Artinian $R$-modules.
7. Suppose $R = F$ is a field. Prove that an $R$-module $M$ is Artinian if and only if it is Noetherian if and only if $M$ is a finite dimensional vector space over $F$.
8. Let $M$ be a maximal ideal of the ring $R$ and suppose that $M^n = 0$ for some $n \geq 1$. Prove that $R$ is Noetherian if and only if $R$ is Artinian. [Observe the each successive quotient $M^i/M^{i+1}, i = 0, \ldots, n-1$ in the filtration $R \supseteq M \supseteq \cdots \supseteq M^{n-1} \supseteq M^n = 0$ is a module over the field $F = R/M$. Then use the previous two exercises and Exercise 6 of Section 15.1.]
9. Let $M$ be a finitely generated $R$-module. Prove that if $x_1, \ldots, x_n$ are elements of $M$ whose images in $M/\mathcal{J}M$ generate $M/\mathcal{J}M$, then they generate $M$. Deduce that if $R$ is Noetherian and the images of $a_1, \ldots, a_n$ in $\mathcal{J}/\mathcal{J}^2$ generate $\mathcal{J}/\mathcal{J}^2$, then $\mathcal{J} = (a_1, \ldots, a_n)$. [Let $N$ be the submodule generated by $x_1, \ldots, x_n$ and apply Nakayama's Lemma to the module $A = M/N$.]
10. Let $R = \mathbb{Z}_{(2)}$ be the localization of $\mathbb{Z}$ at the prime ideal (2). Prove that Jac $R = (2)$ is the ideal generated by 2. If $M = \mathbb{Q}$, prove that $M/2M$ is a finitely generated $R$-module but that $M$ is not finitely generated over $R$. Why doesn't this contradict the previous exercise? [Note the hypotheses in Nakayama's Lemma.]
11. Let $V$ be an affine variety over a field $k$ and let $R = k[V]$ be its coordinate ring. Let $d_t(R)$ denote the transcendence degree of the field of fractions $k(V)$ over $k$, and let $d_p(R)$ be the Krull dimension of $R$ defined in terms of chains of prime ideals. This exercise shows $d_t(R) = d_p(R)$. By Noether's Normalization Lemma there is a polynomial subring $R_1 = k[y_1, \ldots, y_m]$ of $R$ such that $R$ is integral over $R_1$.
    (a) Show that $d_t(R_1) = d_t(R) = m$ and that $d_p(R_1) = d_p(R)$. Deduce that we may assume $R = R_1$. [Use the Going-up and Going-down Theorems (cf. Theorem 26, Section 15.3) to prove the second equality.]
    (b) When $R = R_1$ show that $d_p(R) \geq d_t(R)$ by exhibiting an explicit chain of prime ideals of length $m$.
    (c) When $R = R_1$ show that any nonzero prime ideal of $R$ contains an element $f$ such that $R(f)$ is transcendental over $R$ of transcendence degree 1. Use induction to show that $d_p(R) \leq d_t(R)$, and deduce that $d_p(R) = d_t(R)$.
12. Let $R$ be a Noetherian local ring with maximal ideal $M$.
    (a) The quotient $M/M^2$ is a module (i.e., vector space) over the field $R/M$. Prove that $d = \dim_{R/M}(M/M^2)$ is finite.
    (b) Prove that $M$ can be generated as an ideal in $R$ by $d$ elements and by no fewer. [Use Exercise 9.]
    (c) Let $R = k[x_1, \ldots, x_n]_{(x_1, \ldots, x_n)}$ be the localization of the polynomial ring $k[x_1, \ldots, x_n]$ over the field $k$ at the maximal ideal $(x_1, \ldots, x_n)$, and let $M$ be the maximal ideal in

$R$. Prove that $\dim_{R/M}(M/M^2) = n = \dim R$. [Cf. the previous exercise.]

It can be shown that $\dim_{R/M}(M/M^2) \geq \dim R$ for any Noetherian local ring $R$ with maximal ideal $M$. A Noetherian local ring $R$ is called a *regular local ring* if $\dim_{R/M}(M/M^2) = \dim R$. It is a fact that a regular local ring is necessarily an integral domain and is also integrally closed.

**13.** If $R$ is a Noetherian ring, prove that the Zariski topology on Spec $R$ is discrete (i.e., every subset is Zariski open and also Zariski closed) if and only if $R$ is Artinian.

**14.** Suppose $I$ is the ideal $(x_1, x_2^2, x_3^3, \dots)$ in the polynomial ring $k[x_1, x_2, x_3, \dots]$ where $k$ is a field and let $R$ be the quotient ring $k[x_1, x_2, x_3, \dots]/I$. Prove that the image of the ideal $(x_1, x_2, x_3, \dots)$ in $R$ is the unique prime ideal in $R$ but is not finitely generated. Deduce that $R$ is a local ring of Krull dimension 0 but is not Artinian.

## 16.2 DISCRETE VALUATION RINGS

In the previous section we showed that the Artinian rings are the Noetherian rings having Krull dimension 0. We now consider the easiest Noetherian rings of dimension 1, the Discrete Valuation Rings first introduced in Section 8.1:

**Definition.**

**(1)** A *discrete valuation* on a field $K$ is a function $v : K^\times \to \mathbb{Z}$ satisfying

**(i)** $v$ is surjective,

**(ii)** $v(xy) = v(x) + v(y)$ for all $x, y \in K^\times$,

**(iii)** $v(x + y) \geq \min\{v(x), v(y)\}$ for all $x, y \in K^\times$ with $x + y \neq 0$.

The subring $\{x \in K \mid v(x) \geq 0\} \cup \{0\}$ is called the *valuation ring* of $v$.

**(2)** An integral domain $R$ is called a *Discrete Valuation Ring* (D.V.R.) if $R$ is the valuation ring of a discrete valuation $v$ on the field of fractions of $R$.

The valuation $v$ is often extended to all of $K$ by defining $v(0) = +\infty$, in which case (ii) and (iii) hold for all $a, b \in K$.

**Examples**

**(1)** The localization $\mathbb{Z}_{(p)}$ of $\mathbb{Z}$ at any nonzero prime ideal $(p)$ is a D.V.R. with respect to the discrete valuation $v_p$ on $\mathbb{Q}$ defined as follows (cf. Exercise 27, Section 7.1). Every element $a/b \in \mathbb{Q}^\times$ can be written uniquely in the form $p^n(a_1/b_1)$ where $n \in \mathbb{Z}$, $a_1/b_1 \in \mathbb{Q}^\times$ and both $a_1$ and $b_1$ are relatively prime to $p$. Define

$$v_p\left(\frac{a}{b}\right) = v_p\left(p^n \frac{a_1}{b_1}\right) = n.$$

One easily checks that the axioms for a D.V.R. are satisfied. We call $v_p$ the *p-adic valuation* on $\mathbb{Q}$. The corresponding valuation ring is the set of rational numbers with $n \geq 0$ together with 0, i.e., the rational numbers $a/b$ where $b$ is not divisible by $p$, which is $\mathbb{Z}_{(p)}$.

**(2)** For any field $F$, let $f$ be an irreducible polynomial in $F[x]$. Every nonzero element in the field $F(x)$ can be written uniquely in the form $f^n(a/b)$ where $n \in \mathbb{Z}$, $a/b \in F[x]^\times$ and both $a$ and $b$ are relatively prime to $f$. Then

$$v_f\left(f^n \frac{a}{b}\right) = n$$

defines a valuation on $F(x)$ and the corresponding valuation ring is the localization $F[x]_f$ of $F[x]$ at $f$ consisting of the rational functions in $F(x)$ whose denominator is not divisible by $f$. When $f = x - \alpha$ is a polynomial of degree 1 in $F[x]$, the valuation $v_f$ gives the *order of the zero* (if $n \geq 0$) or *pole* (if $n < 0$) of the element in $F(x)$ at $x = \alpha$.

**(3)** The ring of formal Laurent series $F((x))$ with coefficients in the field $F$ has a discrete valuation $v$ defined by

$$v\left(\sum_{i \geq n}^{\infty} a_i x^i\right) = n$$

(cf. Exercise 5, Section 7.2). The corresponding D.V.R. is the ring $F[[x]]$ of power series in $x$ with coefficients in $F$.

Note that $v(1) = v(1) + v(1)$ implies that $v(1) = 0$, so every Discrete Valuation Ring $R$ is a ring with identity $1 \neq 0$. Since $R$ is a subring of a field by definition, $R$ is in particular an integral domain. It is easy to see that a D.V.R. is a Euclidean Domain (cf. Example 4 in Section 8.1), so in particular is also a P.I.D. and a U.F.D. In fact the factorization and ideal structure of a D.V.R. is very simple, as the next proposition shows.

**Proposition 5.** Suppose $R$ is a Discrete Valuation Ring with respect to the valuation $v$, and let $t$ be any element of $R$ with $v(t) = 1$. Then

**(1)** A nonzero element $u \in R$ is a unit if and only if $v(u) = 0$.

**(2)** Every nonzero element $r \in R$ can be written in the form $r = ut^n$ for some unit $u \in R$ and some $n \geq 0$. Every nonzero element $x$ in the field of fractions of $R$ can be written in the form $x = ut^n$ for some unit $u \in R$ and some $n \in \mathbb{Z}$.

**(3)** Every nonzero ideal of $R$ is a principal ideal of the form $(t^n)$ for some $n \geq 0$. In particular, $R$ is a Noetherian ring.

*Proof:* If $u$ is a unit, then $uv = 1$ for some $v \in R$ and then $v(u) + v(v) = v(uv) = 1$ with $v(u) \geq 0$ and $v(v) \geq 0$ shows that $v(u) = 0$. Conversely, if $u$ is nonzero and $v(u) = 0$ then $u^{-1} \in K$ satisfies $v(u^{-1}) + v(u) = v(1) = 0$. Hence $v(u^{-1}) = 0$ and $u^{-1} \in R$, so $u$ is a unit. This proves (1).

For (2), note that if $v(x) = n$ then $v(xt^{-n}) = 0$, so $xt^{-n} = u$ is a unit in $R$ by (1). Hence $x = ut^n$, where $x \in R$ if and only if $n = v(x) \geq 0$.

If $I$ is a nonzero ideal in $R$, let $r \in I$ be an element with $v(r)$ minimal. If $v(r) = n$, then $r$ differs from $t^n$ by a unit by (2), so $t^n \in I$ and $(t^n) \subseteq I$. If now $a$ is any nonzero element of $I$, then $v(a) \geq n$ by choice of $n$. Then $v(at^{-n}) \geq 0$ and so $at^{-n} \in R$, which shows that $a \in (t^n)$. Hence $I = (t^n)$, proving the first statement in (3). It is then clear that ascending chains of ideals in $R$ are finite, proving that $R$ is Noetherian and completing the proof.

**Definition.** If $R$ is a D.V.R. with valuation $v$, then an element $t$ of $R$ with $v(t) = 1$ is called a *uniformizing* (or *local*) *parameter* for $R$.

**Corollary 6.** Let $R$ be a Discrete Valuation Ring.

**(1)** The ring $R$ is an integrally closed local ring with unique maximal ideal given by the elements with strictly positive valuation: $M = \{r \in R \mid v(r) > 0\}$. Every nonzero ideal in $R$ is of the form $M^n$ for some integer $n \geq 0$.

**(2)** The only prime ideals of $R$ are $M$ and 0, i.e., Spec $R = \{0, M\}$. In particular, a D.V.R. has Krull dimension 1.

*Proof:* Any U.F.D. is integrally closed in its fraction field (Example 3 in Section 15.3), so $R$ is integrally closed. The remainder of the statements follow immediately from the description of the ideals of $R$ in Proposition 5.

The definition of a Discrete Valuation Ring is extremely explicit in terms of a valuation on the fraction field, and as a result it appears that it might be difficult to recognize whether a given ring $R$ is a D.V.R. from purely "internal" algebraic properties of $R$. In fact, the ring-theoretic properties in Proposition 5 and Corollary 6 characterize Discrete Valuation Rings. The following theorem gives several alternate algebraic descriptions of Discrete Valuation Rings in which there is no explicit mention of the valuation.

**Theorem 7.** The following properties of a ring $R$ are equivalent:

**(1)** $R$ is a Discrete Valuation Ring,

**(2)** $R$ is a P.I.D. with a unique maximal ideal $P \neq 0$,

**(3)** $R$ is a U.F.D. with a unique (up to associates) irreducible element $t$,

**(4)** $R$ is a Noetherian integral domain that is also a local ring whose unique maximal ideal is nonzero and principal,

**(5)** $R$ is a Noetherian, integrally closed, integral domain that is also a local ring of Krull dimension 1 i.e., $R$ has a unique nonzero prime ideal: Spec $R = \{0, M\}$.

*Proof:* That (1) implies each of the other properties was proved above.

If (2) holds then (3) is immediate since irreducible elements generate prime ideals in a U.F.D. (Proposition 12, Section 8.3).

If (3) holds, then every nonzero element in $R$ can be written uniquely in the form $ut^n$ for some unit $u$ and some $n \geq 0$. Then every nonzero element in the fraction field of $R$ can be written uniquely in the form $ut^n$ for some unit $u$ and some $n \in \mathbb{Z}$. It is now straightforward to check that the map $v(ut^n) = n$ is a discrete valuation on the field of fractions of $R$, and $R$ is the valuation ring of $v$, and (1) holds.

Suppose (4) holds, let $M = (t)$ be the unique maximal ideal of $R$, and let $M_0 = \bigcap_{i=1}^{\infty} M^i$. Then $M_0 = MM_0$, and since $R$ is Noetherian $M_0$ is finitely generated. By hypothesis $M = \text{Jac } R$, so by Nakayama's Lemma $M_0 = 0$. If $I$ is any proper, nonzero ideal of $R$ then there is some $n \geq 0$ such that $I \subseteq M^n$ but $I \not\subseteq M^{n+1}$. Let $a \in I - M^{n+1}$ and write $a = t^n u$ for some $u \in R$. Then $u \notin M$, and so $u$ is a unit in the local ring $R$. Thus $(a) = (t^n) = M^n$ for every $a \in I - M^{n+1}$. This shows that $I = (t^n)$, and so every ideal of $R$ is principal, which shows that (2) holds.

We have shown that (1), (2), (3) and (4) are equivalent, and that each of these implies (5). To complete the proof we show that (5) implies (4), which amounts to showing that the ideal $M$ in (5) is a principal ideal. Since $0 \neq M = \text{Jac } R$ and $M$ is

finitely generated because $R$ is Noetherian, by Nakayama's Lemma (Proposition 1(4)), $M \neq M^2$. Let $t \in M - M^2$. We argue that $M = (t)$. By Proposition 12 in Section 15.2, the assumption that $M$ is the unique nonzero prime ideal in $R$ implies that $M = \text{rad}\,(t)$, and then Proposition 14 in Section 15.2 implies that some power of $M$ is contained in $(t)$. Proceeding by way of contradiction, assume $(t) \neq M$, so that $M^n \subseteq (t)$ but $M^{n-1} \not\subseteq (t)$ for some $n \geq 2$. Then there is an element $x \in M^{n-1} - (t)$ such that $xM \subseteq (t)$. Note that $t \neq 0$ so $y = x/t$ belongs to the field of fractions of $R$. Also, $y \notin R$ because $x = ty \notin (t)$. However, by choice of $x$ we have $yM \subseteq R$, and then one checks that $yM$ is an ideal in $R$. If $yM = R$ then $1 = ym$ for some $m \in M$. This leads to a contradiction because we would then have $t = xm \in M^2$, contrary to the choice of $t$. Thus $yM$ is a proper ideal, hence is contained in the unique maximal ideal of $R$, namely $yM \subseteq M$. Now $M$ is a finitely generated $R$-module on which $y$ acts by left multiplication as an $R$-module homomorphism. By the same (determinant) method as in the proof of Proposition 23 in Section 15.3 there is a monic polynomial $p$ with coefficients in $R$ such that $p(y)m = 0$ for all $m \in M$. Since $p(y)$ is an element of a field containing $R$ and $M$, we must have $p(y) = 0$. Hence $y$ is integral over $R$. Since $R$ is integrally closed by assumption, it follows that $y \in R$, a contradiction. Hence $M = (t)$ is principal, so (5) implies (4), completing the proof of the theorem.

**Corollary 8.** If $R$ is any Noetherian, integrally closed, integral domain and $P$ is a minimal nonzero prime ideal of $R$, then the localization $R_P$ of $R$ at $P$ is a Discrete Valuation Ring.

*Proof:* By results in Section 15.4, the localization $R_P$ is a Noetherian (Proposition 38(4)), integrally closed (Proposition 49), integral domain (Proposition 46(2)), that is a local ring with unique nonzero prime ideal (Proposition 46(4)), so $R_P$ satisfies (5) in the theorem.

**Examples**

**(1)** If $R$ is any Principal Ideal Domain then every localization $R_P$ of $R$ at a nonzero prime ideal $P = (p)$ is a Discrete Valuation Ring. This follows immediately from Corollary 8 since $R$ is integrally closed (being a U.F.D., cf. Example 3 in Section 15.3) and nonzero prime ideals in a P.I.D. are maximal (Proposition 8.7). Note that the quotient field $K$ of $R_P$ is the same as the quotient field of $R$, so each nonzero prime $p$ in $R$ produces a valuation $v_p$ on $K$, given by the formula

$$v\left(p^n \frac{a}{b}\right) = n$$

where $a$ and $b$ are elements of $R$ not divisible by $p$. This generalizes both Examples 1 and 2 above.

**(2)** The ring $\mathbb{Z}_p$ of $p$-adic integers is a Discrete Valuation Ring since it is a P.I.D. with unique maximal ideal $p\mathbb{Z}_p$ (cf. Exercise 11, Section 7.6). The fraction field of $\mathbb{Z}_p$ is called the *field of p-adic numbers* and is denoted $\mathbb{Q}_p$. The element $p$ is a uniformizing parameter for $\mathbb{Z}_p$, so every nonzero element in $\mathbb{Q}_p$ can be written uniquely in the form $p^n u$ for some $n \in \mathbb{Z}$ and unit $u \in \mathbb{Z}_p^\times$, (where $u = a_0 + a_1 p + a_2 p^2 + \ldots$ with $0 < a_0 < p$ as in Exercise 11(c), Section 7.6). The corresponding *p-adic valuation* $v_p$ on $\mathbb{Q}_p$ is then given by $v_p(p^n u) = n$.

A discrete valuation $v$ on a field $K$ defines an associated *metric* (or "distance function"), $d_v$, on $K$ as follows: fix any real number $\beta > 1$ (the actual value of $\beta$ does not matter for verifying the axioms of a metric), and for all $a, b \in K$ define

$$d_v(a, b) = ||a - b||_v \qquad \text{where} \qquad ||a||_v = \beta^{-v(a)}$$

and where we set $d_v(a, a) = 0$. It is easy to check that $d_v$ satisfies the three axioms for a metric:

**(i)** $d_v(a, b) \geq 0$, with equality holding if and only if $a = b$,
**(ii)** $d_v(a, b) = d_v(b, a)$, i.e., $d_v$ is symmetric,
**(iii)** $d_v(a, b) \leq d_v(a, c) + d_v(c, b)$, for all $a, b, c \in K$, i.e., $d_v$ satisfies the "triangle inequality."

The triangle inequality is a consequence of axiom (iii) of the discrete valuation. Indeed, a stronger version of the triangle inequality holds:

**(iii)′** $d_v(a, b) \leq \max\{d_v(a, c), d_v(c, b)\}$, for all $a, b, c \in K$.

For this reason $d_v$ is sometimes called an *ultrametric*. One may now use Cauchy sequences to form the *completion* of $K$ with respect to $d_v$, denoted by $K_v$, in the same way that the real numbers $\mathbb{R}$ are constructed from the rational numbers $\mathbb{Q}$. It is not difficult to show that $K_v$ is also a field with a discrete valuation that agrees with $v$ on the dense subset $K$ of $K_v$.

**Examples**

**(1)** Consider the $p$-adic valuation $v_p$ on $\mathbb{Q}$ and take $\beta = p$. Write $||a||_p$ for $||a||_{v_p}$, so that for $a, b$ relatively prime to $p$,

$$||p^n \frac{a}{b}||_p = p^{-n}.$$

Note that integers (or rational numbers) have small $p$-adic absolute value if they are divisible by a large power of $p$. For example, the sequence $1, p, p^2, p^3, \ldots$ converges to zero in the $p$-adic metric.

It is not too difficult to see that the completion of $\mathbb{Q}$ with respect to the $p$-adic metric is the field $\mathbb{Q}_p$ of $p$-adic numbers, and the completion of $\mathbb{Z}$ is the ring $\mathbb{Z}_p$ of $p$-adic integers. One way to see this is to check that each element $a$ of the completion may be represented as a *p-adic Laurent series*:

$$a = \sum_{n=n_0}^{\infty} a_i p^i \qquad \text{where } n_0 \in \mathbb{Z} \text{ and } a_i \in \{0, 1, \ldots, p-1\} \text{ for all } i,$$

and then use Example 2 previously. In terms of this expansion, the $p$-adic valuation is given by $v_p(a) = n_0$ (when $a_{n_0} \neq 0$).

**(2)** In a similar way, the completion of $F(x)$ with respect to the valuation $v_x$ in Example 2 at the beginning of this section gives the field $F((x))$ with corresponding valuation ring $F[[x]]$ in Example 3 in the same set of examples.

The completion of a field $K$ with respect to a discrete valuation $v$ is a field $K_v$ in which the elements can be easily described in terms of a uniformizing parameter. In addition, $K_v$ is a topological space where the topology is defined by the metric $d_v$. Furthermore, Cauchy sequences of elements in $K_v$ converge to elements of $K_v$ (i.e., $K_v$

is *complete* in the $v$-adic topology). This is similar to the situation of the completion $\mathbb{R}$ of $\mathbb{Q}$ with respect to the usual Euclidean metric. This allows the application of ideas from analysis to the study of such rings, and is an important tool in the study of algebraic number fields and in algebraic geometry.

## Fractional Ideals

We complete our discussion of Discrete Valuation Rings by giving another characterization of D.V.R.s in terms of "fractional ideals," which can be defined for any integral domain:

**Definition.** For any integral domain $R$ with fraction field $K$, a *fractional ideal* of $R$ is an $R$-submodule $A$ of $K$ such that $dA \subseteq R$ for some nonzero $d \in R$ (equivalently, a submodule of the form $d^{-1}I$ for some nonzero $d \in R$ and ideal $I$ of $R$).

The equivalence of these two definitions follows from the observation that $dA$ is an $R$-submodule (i.e., an ideal) of $R$.

The notion of a fractional ideal in $K$ depends on the ring $R$. Loosely speaking, a fractional ideal is an ideal of $R$ up to a fixed "denominator" $d$. The ideals of $R$ are also fractional ideals of $R$ (with denominator $d = 1$) and are the fractional ideals that are contained in $R$. For clarity these are occasionally called the *integral ideals* of $R$. When $R$ is a Noetherian integral domain, a fractional ideal of $R$ is the same as a finitely generated $R$-submodule of $K$ (cf. Exercise 6).

For any $x \in K$ the (cyclic) $R$-module $Rx = \{rx \mid r \in R\}$ is called the *principal fractional ideal* generated by $x$.

If $A$ and $B$ are fractional ideals, their product, $AB$, is defined to be the set of all finite sums of elements of the form $ab$ where $a \in A$ and $b \in B$. If $A = d^{-1}I$ and $B = (d')^{-1}J$ for ideals $I$, $J$ in $R$ and nonzero $d$, $d' \in R$, then $AB = (dd')^{-1}IJ$ where $IJ$ is the usual product ideal. In particular, this shows that the product of two fractional ideals is a fractional ideal.

**Definition.** The fractional ideal $A$ is said to be *invertible* if there exists a fractional ideal $B$ with $AB = R$, in which case $B$ is called the *inverse* of $A$ and denoted $A^{-1}$.

If $A$ is an invertible fractional ideal, the fractional ideal $B$ with $AB = R$ is unique: $AB = AC = R$ implies $B = B(AC) = (BA)C = C$.

**Proposition 9.** Let $R$ be an integral domain and let $A$ be a fractional ideal of $R$.

**(1)** If $A$ is a nonzero principal fractional ideal then $A$ is invertible.

**(2)** If $A$ is nonzero then the set $A' = \{x \in K \mid xA \subseteq R\}$ is a fractional ideal of $R$. In general we have $AA' \subseteq R$ and $AA' = R$ if and only if $A$ is invertible, in which case $A^{-1} = A'$.

**(3)** If $A$ is an invertible fractional ideal of $R$ then $A$ is finitely generated.

**(4)** The set of invertible fractional ideals is an abelian group under multiplication with identity $R$. The set of nonzero principal fractional ideals is a subgroup of the invertible fractional ideals.

*Proof:* If $A = xR$ is a nonzero principal fractional ideal, then taking $B = x^{-1}R$ shows that $A$ is invertible, proving (1).

One easily sees that $A'$ is an $R$-submodule of $K$. If $A$ is a nonzero fractional ideal there is some nonzero element $d \in R$ such that $dA \subseteq R$, so $A$ contains nonzero elements of $R$. Let $a$ be any nonzero element of $A$ contained in $R$. Then by definition of $A'$ we have $aA' \subseteq R$, so $A'$ is a fractional ideal. Also by definition, $AA' \subseteq R$. If $AA' = R$ then $A$ is invertible with inverse $A^{-1} = A'$. Conversely, if $AB = R$, then $B \subseteq A'$ by definition of $A'$. Then $R = AB \subseteq AA' \subseteq R$, showing that $AA' = R$, proving (2).

If $A$ is invertible, then $AA' = R$ by (2) and so $1 = a_1a_1' + \cdots + a_na_n'$ for some $a_1, \ldots, a_n \in A$ and $a_1', \ldots, a_n' \in A'$. If $a \in A$, then $a = (aa_1')a_1 + \cdots + (aa_n')a_n$, where each $aa_i' \in R$ by definition of $A'$. It follows that $A$ is generated over $R$ by $a_1, \ldots, a_n$ and so $A$ is finitely generated, proving (3).

Finally, it is clear that the product of two invertible fractional ideals is again invertible. This product is commutative, associative, and $RA = A$ for any fractional ideal. The inverse of an invertible fractional ideal is an invertible fractional ideal by definition, proving the first statement in (4). The second statement in (4) is immediate since the product of $xR$ and $yR$ is $(xy)R$ and the inverse of $xR$ is $x^{-1}R$.

**Definition.** If $R$ is an integral domain, then the quotient of the group of invertible fractional ideals of $R$ by the subgroup of nonzero principal fractional ideals of $R$ is called the *class group* of $R$. The order of the class group of $R$ is called the *class number* of $R$.

The class group of $R$ is the trivial group and the class number of $R$ is 1 if and only if $R$ is a P.I.D. The class group of $R$ measures how close the ideals of $R$ are to being principal.

Whether a fractional ideal $A$ of $R$ is invertible is also related to whether $A$ is *projective* as an $R$-module. Recall that an $R$-module $M$ is projective over $R$ if and only if $M$ is a direct summand of a free module (Proposition 30, Section 10.5). Equivalently, $M$ is projective if and only if there is a free $R$-module $F$ and $R$-module homomorphisms $f : F \to M$ and $g : M \to F$ with $f \circ g = 1$ (Proposition 25, Section 10.5).

**Proposition 10.** Let $R$ be an integral domain with fraction field $K$ and let $A$ be a nonzero fractional ideal of $R$. Then $A$ is invertible if and only if $A$ is a projective $R$-module.

*Proof:* Assume first that $A$ is invertible, so $\sum_{i=1}^n a_ia_i' = 1$ for some $a_i \in A$ and $a_i' \in A'$ as in (2) of Proposition 9. Let $F$ be the free $R$-module on $y_1, \ldots, y_n$. Define $f : F \to A$ by $f(\sum_{i=1}^n r_iy_i) = \sum_{i=1}^n r_ia_i$ and $g : A \to F$ by $f(c) = \sum_{i=1}^n (ca_i')y_i$. It is immediate that both $f$ and $g$ are $R$-module homomorphisms (note that $ca_i' \in R$ by definition of $A'$). Since

$$(f \circ g)(c) = f\left(\sum_{i=1}^n (ca_i')y_i\right) = \sum_{i=1}^n (ca_i')a_i = c\left(\sum_{i=1}^n a_ia_i'\right) = c,$$

so $f \circ g = 1$ and $A$ is a direct summand of $F$, hence is projective.

Conversely, suppose that $A$ is nonzero and projective, so there is a free $R$-module $F$ and $R$-homomorphisms $f : F \to A$ and $g : A \to F$ with $f \circ g = 1$. Fix any $0 \neq a \in A$ and suppose $g(a) = \sum_{i=1}^{n} \tilde{a}_i y_i$ where $\tilde{a}_i \in R$ and $y_1, \ldots, y_n$ is part of a set of free generators for $F$. Define $a_i = f(y_i)$ and $a_i' = \tilde{a}_i/a \in K$ for $i = 1, \ldots, n$. For any $b \in A$ we have $bg(a) = ag(b) = g(ab)$ since $g$ is an $R$-module homomorphism. Write $g(b) = \sum_{i=1}^{n} \tilde{b}_i y_i + \sum_{j \in \mathcal{J}} \tilde{b}_j y_j$ where $\{y_j\}$ for $j \in \mathcal{J}$ are the remaining elements in the set of free generators for $F$. Then

$$\sum_{i=1}^{n} (b\tilde{a}_i) y_i = \sum_{i=1}^{n} (a\tilde{b}_i) y_i + \sum_{j \in \mathcal{J}} (a\tilde{b}_j) y_j \, .$$

We may equate coefficients of the elements in the free $R$-module basis for $F$ in this equation and it follows that $g(b) = \sum_{i=1}^{n} \tilde{b}_i y_i$ where $\tilde{b}_i \in R$ and that $b\tilde{a}_i = a\tilde{b}_i$ for $i = 1, \ldots, n$. In particular, it follows from the definition of $a_i'$ that $ba_i' = b(\tilde{a}_i/a) = \tilde{b}_i$ is an element of $R$ for every element $b$ of $A$. This shows that $a_i' \in A'$ for $i = 1, \ldots, n$. Since $f \circ g = 1$, we have

$$a = f \circ g(a) = f\left(\sum_{i=1}^{n} \tilde{a}_i y_i\right) = \sum_{i=1}^{n} \tilde{a}_i a_i = \sum_{i=1}^{n} (aa_i') a_i = a\left(\sum_{i=1}^{n} a_i a_i'\right) ,$$

and so $\sum_{i=1}^{n} a_i a_i' = 1$. It follows that $AA' = R$ and so $A$ is invertible by Proposition 9, completing the proof.

The next result shows that if the integral domain $R$ is also a local ring, then whether fractional ideals are invertible determines whether $R$ is a D.V.R.

**Proposition 11.** Suppose the integral domain $R$ is a local ring that is not a field. Then $R$ is a Discrete Valuation Ring if and only if every nonzero fractional ideal of $R$ is invertible.

*Proof:* If $R$ is a D.V.R. with uniformizing parameter $t$, then by Proposition 5 every nonzero ideal of $R$ is of the form $(t^n)$ for some $n \geq 0$ and every element $d$ in $R$ can be written in the form $ut^m$ for some unit $u \in R$ and some $m \geq 0$. It follows that every nonzero fractional ideal of $R$ is of the form $t^N R$ for some $N \in \mathbb{Z}$, so is a principal fractional ideal and hence invertible by the previous proposition.

Conversely, suppose that every nonzero fractional ideal of $R$ is invertible. Then every nonzero ideal of $R$ is finitely generated by (3) of Proposition 9, so $R$ is Noetherian. Let $M$ be the unique maximal ideal of $R$. If $M = M^2$ then $M = 0$ by Nakayama's Lemma, and then $R$ would be a field, contrary to hypothesis. Hence there is an element $t$ with $t \in M - M^2$. By assumption $M$ is invertible, and since $t \in M$, the fractional ideal $tM^{-1}$ is a nonzero ideal in $R$. If $tM^{-1} \subseteq M$, then $t \in M^2$, contrary to the choice of $t$. Hence $tM^{-1} = R$, so $(t) = M$, and $M$ is a nonzero principal ideal. It follows by the equivalent condition 4 of Theorem 7 that $R$ is a D.V.R., completing the proof.

We end this section with an application to algebraic geometry.

## Nonsingularity and Local Rings of Affine Plane Curves

Let $k$ be an algebraically closed field and let $C$ be an irreducible affine *curve* over $k$. In other words, $C$ is an affine algebraic set whose coordinate ring $k[C]$ is an integral domain and whose field of rational functions $k(C)$ has transcendence degree 1 over $k$ (cf. Section 15.4).

Recall that, by definition, the point $v$ on $C$ is nonsingular if $\mathfrak{m}_{v,C}/\mathfrak{m}_{v,C}^2$ is a 1-dimensional vector space over $k$, where $\mathfrak{m}_{v,C}$ is the unique maximal ideal in the local ring $\mathcal{O}_{v,C}$ of rational functions on $C$ defined at $v$.

**Proposition 12.** Let $v$ be a point on the irreducible affine curve $C$ over $k$. Then $C$ is nonsingular at $v$ if and only if the local ring $\mathcal{O}_{v,C}$ is a Discrete Valuation Ring.

*Proof:* Suppose first that $v$ is nonsingular. Then $\dim_k(\mathfrak{m}_{v,C}/\mathfrak{m}_{v,C}^2) = 1$, and since $\mathcal{O}_{v,C}$ is Noetherian, it follows from Exercise 12 in Section 1 that $\mathfrak{m}_{v,C}$ is principal. Hence $\mathcal{O}_{v,C}$ is a D.V.R. by Theorem 7(4). Conversely, suppose $\mathcal{O}_{v,C}$ is a D.V.R. and $t$ is a uniformizing element for $\mathcal{O}_{v,C}$. Then every element in $\mathfrak{m}_{v,C}$ can be written uniquely in the form $at$ for some $a$ in $\mathcal{O}_{v,C}$. The map from $\mathfrak{m}_{v,C}$ to $\mathcal{O}_{v,C}/\mathfrak{m}_{v,C}$ defined by mapping $at$ to $a \bmod \mathfrak{m}_{v,C}$ is easily checked to be a surjective $\mathcal{O}_{v,C}$-module homomorphism with kernel $\mathfrak{m}_{v,C}^2$. Hence $\mathfrak{m}_{v,C}/\mathfrak{m}_{v,C}^2$ is isomorphic as an $\mathcal{O}_{v,C}/\mathfrak{m}_{v,C}$-module to $\mathcal{O}_{v,C}/\mathfrak{m}_{v,C}$. Since $\mathcal{O}_{v,C}/\mathfrak{m}_{v,C} \cong k$ (Proposition 46(5) in Section 15.4), it follows that $\dim_k(\mathfrak{m}_{v,C}/\mathfrak{m}_{v,C}^2) = 1$, and so $v$ is a nonsingular point on $C$.

**Definition.** If $v$ is a nonsingular point on $C$ with corresponding discrete valuation $\nu_v$ defined on $k(C)$, then $\nu_v(f) = n$ for $f \in k(V)$ is the *order of zero of $f$ at $v$* (if $n \geq 0$) or the *order of the pole of $f$ at $v$* (if $n < 0$).

Using the criterion for nonsingularity for points on curves in Proposition 12 we can prove a result first mentioned in Section 15.4:

**Corollary 13.** An irreducible affine curve $C$ over an algebraically closed field $k$ is smooth if and only if its coordinate ring $k[C]$ is integrally closed.

*Proof:* The curve $C$ is smooth if and only if every localization $\mathcal{O}_{v,C}$ is a D.V.R. Since $k[C]$ has Krull dimension 1 (Exercise 11 in Section 1), the same is true for each $\mathcal{O}_{v,C}$. It then follows by Theorem 7(5) that every localization $\mathcal{O}_{v,C}$ is a D.V.R. if and only if $\mathcal{O}_{v,C}$ is integrally closed. By Proposition 49 in Section 15.4, this in turn is equivalent to the statement that $k[C]$ is integrally closed, which proves the corollary.

### EXERCISES

1. Suppose $R$ is a Discrete Valuation Ring with respect to the valuation $\nu$ on the fraction field $K$ of $R$. If $x, y \in K$ with $\nu(x) < \nu(y)$ prove that $\nu(x + y) = \min(\nu(x), \nu(y))$. [Note that $x + y = x(1 + y/x)$.]
2. Suppose $R$ is a Discrete Valuation Ring with unique maximal ideal $M$ and quotient $F = R/M$. For any $n \geq 0$ show that $M^n/M^{n+1}$ is a vector space over $F$ and that $\dim_F(M^n/M^{n+1}) = 1$.

3. Suppose $R$ is an integral domain that is also a local ring whose unique maximal ideal $M = (t)$ is nonzero and principal, and suppose that $\cap_{n\geq 1}(t^n) = 0$. Prove that $R$ is a Discrete Valuation Ring. [Show that every nonzero element in $R$ can be written in the form $ut^n$ for some unit $u \in R$ and some $n \geq 0$.]

4. Suppose $R$ is a Noetherian local ring whose unique maximal ideal $M = (t)$ is principal. Prove that either $R$ is a Discrete Valuation Ring or $t^n = 0$ for some $n \geq 0$. In the latter case show that $R$ is Artinian.

5. Suppose that $R$ is a Noetherian integral domain that is also a local ring of Krull dimension 1. Let $M$ be the unique maximal ideal of $R$ and let $F = R/M$, so that $M/M^2$ is a vector space over $F$.
   **(a)** Prove that if $\dim_F(M/M^2) = 1$ then $R$ is a Discrete Valuation Ring.
   **(b)** If every nonzero ideal of $R$ is a power of $M$ prove that $R$ is a Discrete Valuation Ring.

6. Let $R$ be an integral domain with fraction field $K$. Prove that every finitely generated $R$-submodule of $K$ is a fractional ideal of $R$. If $R$ is Noetherian, prove that $A$ is a fractional ideal of $R$ if and only if $R$ is a finitely generated $R$-submodule of $K$.

7. If $R$ is an integral domain and $A$ is a fractional ideal of $R$, prove that if $A$ is projective then $A$ is finitely generated. Conclude that every integral domain that is not Noetherian contains an ideal that is not projective.

8. Suppose $R$ is a Noetherian integral domain that is also a local ring with nonzero maximal ideal $M$. Prove that $R$ is a D.V.R. if and only if the only $M$-primary ideals in $R$ are the powers of $M$.

9. Let $C = \mathcal{Z}(xz - y^2, yz - x^3, z^2 - x^2y) \subset \mathbb{A}^3$ over the algebraically closed field $k$. If $v = (0, 0, 0) \in C$, prove that $\dim_k(\mathfrak{m}_{v,C}/\mathfrak{m}_{v,C}^2) = 3$ so that $v$ is singular on $C$. Conclude that $k[C]$ is not integrally closed in $k(C)$ and determine its integral closure. [cf. Exercise 27, Section 15.4.]

## 16.3 DEDEKIND DOMAINS

In the previous section we showed that Discrete Valuation Rings are the local rings that are integrally closed Noetherian integral domains of Krull dimension 1. In this section we consider the effect of relaxing the condition that the ring be a local ring:

**Definition.** A *Dedekind Domain* is a Noetherian, integrally closed, integral domain of Krull dimension 1.

Equivalently, $R$ is a Dedekind Domain if $R$ is a Noetherian, integrally closed, integral domain that is not a field in which every nonzero prime ideal is maximal.

The first result shows that Dedekind Domains are a generalization of the class of Principal Ideal Domains. We shall see later (Theorem 22) that there is a structure theorem for finitely generated modules over a Dedekind Domain extending the corresponding result for P.I.D.s proved in Section 12.1.

**Proposition 14.**
**(1)** Every Principal Ideal Domain is a Dedekind Domain.
**(2)** The ring of integers in an algebraic number field is a Dedekind Domain.

*Proof:* A P.I.D. is clearly Noetherian, is integrally closed since it is a U.F.D. (Example 3, Section 15.3), and nonzero prime ideals are maximal (Proposition 7 in Section 8.2), which proves (1). Let $\mathcal{O}_K$ be the ring of integers in the number field $K$, i.e., the integral closure of $\mathbb{Z}$ in $K$. Then Corollary 25 in Section 15.3 shows that $\mathcal{O}_K$ is integrally closed, $\mathcal{O}_K$ is Noetherian by Theorem 29 in Section 15.3, and the fact that nonzero prime ideals in $\mathcal{O}_K$ are maximal was proved in the discussion following the same theorem. This proves (2).

The following theorem gives a number of important equivalent characterizations of Dedekind Domains. Recall that the basic properties of fractional ideals were developed in the previous section.

**Theorem 15.** Suppose $R$ is an integral domain with fraction field $K \neq R$. The following are equivalent conditions for $R$ to be a Dedekind Domain:

**(1)** The ring $R$ is Noetherian, integrally closed, and every nonzero prime ideal is maximal.

**(2)** The ring $R$ is Noetherian and for each nonzero prime $P$ of $R$ the localization $R_P$ is a Discrete Valuation Ring.

**(3)** Every nonzero fractional ideal of $R$ in $K$ is invertible.

**(4)** Every nonzero fractional ideal of $R$ in $K$ is a projective $R$-module.

**(5)** Every nonzero proper ideal $I$ of $R$ can be written as a finite product of prime ideals: $I = P_1 P_2 \cdots P_n$ (not necessarily distinct).

When the condition in (5) holds, the set of primes $\{P_1, \ldots, P_n\}$ is uniquely determined and so every nonzero proper ideal $I$ of $R$ can be written uniquely (up to order) as a product of powers of prime ideals.

*Proof:* If $R$ satisfies (1), then $R_P$ is a D.V.R. by Corollary 8, so (1) implies (2). Conversely, assume each $R_P$ is a D.V.R. Then $R$ is integrally closed by Proposition 49 in Section 15.4 and every nonzero prime ideal is maximal by Proposition 46(3) in Section 15.4, so (2) implies (1).

Suppose now that (1) is satisfied and that $A$ is a nonzero fractional ideal of $R$. Let $A' = \{x \in K \mid xA \subseteq R\}$ as in Proposition 9. For any prime ideal $P$ of $R$ the behavior of $R$-modules under localization shows that $(AA')_P = A_P(A')_P = A_P(A_P)'$ (cf. Exercise 4). Since $R_P$ is a D.V.R. by what has already been shown, $A_P(A_P)' = R_P$ by Proposition 11. Hence $(AA')_P = R_P$ for all nonzero primes $P$ of $R$, so $AA' = R$ (Exercise 13 in Section 15.4), and $A$ is invertible, showing (1) implies (3). Conversely, suppose every nonzero fractional ideal of $R$ is invertible. Then every ideal in $R$ is finitely generated by Proposition 9(3), so $R$ is Noetherian. Every localization $R_P$ of $R$ at a nonzero prime $P$ is a local ring in which the nonzero fractional ideals are invertible (cf. Exercise 4), hence is a D.V.R. by Proposition 11. Hence (3) implies (2) and so (1), (2) and (3) are equivalent. The equivalence of these with (4) is given by Proposition 10.

Suppose now that (1) is satisfied, and let $I$ be any nonzero proper ideal in $R$. Since $R$ is Noetherian, $I$ has a minimal primary decomposition $I = Q_1 \cap \cdots \cap Q_n$ as in Theorem 21 of Section 15.2. The associated primes $P_i = \text{rad } Q_i$ for $i = 1, \ldots, n$ are all distinct, and since primes are maximal in $R$ by hypothesis, the associated primes are all pairwise comaximal, and it follows easily that the same is true for the $Q_i$ (Exercise

5). It follows that $Q_1 \cap \cdots \cap Q_n = Q_1 \cdots Q_n$ (Theorem 17 in Section 7.6) so that $I$ is the product of primary ideals. The $P$-primary ideals of $R$ correspond bijectively with the $PR_P$-primary ideals in the localization $R_P$ (Proposition 42(3) in Section 15.4), and since $R_P$ is a D.V.R. (because (1) implies (2)), it follows from Corollary 6 that if $Q$ is a $P$-primary ideal in $R$ then $Q = P^m$ for some integer $m \geq 1$. Applying this to $Q_i$, $i = 1, \ldots, n$ shows that $I$ is the product of powers of prime ideals, which gives the first implication in (5).

Conversely, suppose that all the nonzero proper ideals of $R$ can be written as a product of prime ideals. We first show for any integral domain that a factorization of an ideal into *invertible* prime ideals is unique, i.e., if $P_1 \cdots P_n = \tilde{P}_1 \cdots \tilde{P}_m$ are two factorizations of $I$ into invertible prime ideals then $n = m$ and the two sets of primes $\{P_1, \ldots, P_n\}$ and $\{\tilde{P}_1, \ldots, \tilde{P}_m\}$ are equal. Suppose $\tilde{P}_1$ is a minimal element in the set $\{\tilde{P}_1, \ldots, \tilde{P}_m\}$. Since $P_1 \cdots P_n \subseteq \tilde{P}_1$, the prime ideal $\tilde{P}_1$ contains one of the primes $P_1, \ldots, P_n$, say $P_1 \subseteq \tilde{P}_1$. Similarly $P_1$ contains $\tilde{P}_i$ for some $i = 1, \ldots, m$. Then $\tilde{P}_i \subseteq P_1 \subseteq \tilde{P}_1$ and by the minimality of $\tilde{P}_1$ it follows that $\tilde{P}_i = P_1 = \tilde{P}_1$, so the factorization becomes $P_1 P_2 \cdots P_n = P_1 \tilde{P}_2 \cdots \tilde{P}_m$. Since $P_1$ is invertible, multiplying by the inverse ideal shows that $P_2 \cdots P_n = \tilde{P}_2 \cdots \tilde{P}_m$ and an easy induction finishes the proof. In particular, the uniqueness statement in (5) now follows from the first statement in (5) since in a Dedekind domain every fractional ideal, in particular every prime ideal of $R$, is invertible.

We next show that *invertible* primes in $R$ are maximal. Suppose then that $P$ is an invertible prime ideal in $R$ and take $a \in R, a \notin P$. We want to show that $P + aR = R$. By assumption, the two ideals $P + aR$ and $P + a^2R$ can be written as a product of prime ideals, say $P + aR = P_1 \cdots P_n$ and $P + a^2R = \tilde{P}_1 \cdots \tilde{P}_m$. Note that $P \subseteq P_i$ for $i = 1, \ldots, n$ and also $P \subseteq \tilde{P}_j$ for $j = 1, \ldots, m$. In the quotient $R/P$, which is an integral domain, we have the factorization $(\bar{a}) = (P_1/P) \cdots (P_n/P)$, and each $P_i/P$ is a prime ideal in $R/P$. Since the product is a principal ideal, each $P_i/P$ is also an invertible $R/P$-ideal (cf. Exercise 2). Similarly, $(\bar{a}^2) = (\tilde{P}_1/P) \cdots (\tilde{P}_m/P)$ is a factorization into a product of invertible prime ideals. Then $(\bar{a})^2 = (P_1/P)^2 \cdots (P_n/P)^2 = (\tilde{P}_1/P) \cdots (\tilde{P}_m/P)$ give two factorizations into a product of invertible prime ideals in the integral domain $R/P$, so by the uniqueness result in the previous paragraph, $m = 2n$ and $\{P_1/P, P_1/P, \ldots, P_n/P, P_n/P\} = \{\tilde{P}_1/P, \ldots, \tilde{P}_m/P\}$. It follows that the set of primes $\tilde{P}_1, \ldots, \tilde{P}_m$ in $R$ consists of the primes $P_1, \ldots, P_n$, each repeated twice. This shows that $P + a^2R = (P + aR)^2$. Since $P \subseteq P + a^2R$ and $(P + aR)^2 \subseteq P^2 + aR$, we have $P \subseteq P^2 + aR$, so every element $x$ in $P$ can be written in the form $x = y + az$ where $y \in P^2$ and $z \in R$. Then $az = x - y \in P$ and since $a \notin P$, we have $z \in P$, which shows that $P \subseteq P^2 + aP$. Clearly $P^2 + aP \subseteq P$ and so $P = P^2 + aP = P(P + aR)$. Since $P$ is assumed invertible, it follows that $R = P + aR$ for any $a \in R - P$, which proves that $P$ is a maximal ideal.

We now show that every nonzero prime ideal is invertible. If $P$ is a nonzero prime ideal, let $a$ be any nonzero element in $P$. By assumption, $Ra = P_1 \cdots P_n$ can be written as a product of prime ideals, and $P_1, \ldots, P_n$ are invertible since their product is principal (by Exercise 2 again). Since $P_1 \cdots P_n = Ra \subseteq P$, the prime ideal $P$ contains $P_i$ for some $1 \leq i \leq n$. Since $P_i$ is maximal by the previous paragraph, it follows that

$P = P_i$ is invertible.

Finally, since every nonzero proper ideal of $R$ is a product of prime ideals, it follows that every nonzero ideal of $R$ is invertible, and since every fractional ideal of $R$ is of the form $(d^{-1})I$ for some ideal in $R$, also every fractional ideal of $R$ is invertible. This proves that (5) implies (3), and complete the proof of the theorem.

The following corollary follows immediately from Proposition 14:

**Corollary 16.** If $\mathcal{O}_K$ is the ring of integers in an algebraic number field $K$ then every nonzero ideal $I$ in $\mathcal{O}_K$ can be written uniquely as the product of powers of distinct prime ideals:

$$I = P_1^{e_1} P_2^{e_2} \cdots P_n^{e_n},$$

where $P_1, \ldots, P_n$ are distinct prime ideals and $e_i \geq 1$ for $i = 1, \ldots, n$.

*Remark:* The development of Dedekind Domains given here reverses the historical development. As mentioned in Section 9.3, the unique factorization of nonzero *ideals* into a product of prime *ideals* replaces the failure of unique factorization of nonzero *elements* into products of prime *elements* in rings of integers of number fields. This property of rings of integers in Corollary 16 is what led originally to the definition of an ideal, and Dedekind originally defined what we now call Dedekind Domains by property 5 in Theorem 15. It was Noether who observed that they can also be characterized by property (1), which we have taken as the initial definition of a Dedekind Domain.

The unique factorization into prime ideals in Dedekind Domains can be used to explicitly define the valuations $v_P$ on $R$ with respect to which the valuation rings are the localizations $R_P$ in Theorem 15(2) (cf. Exercise 6). We now indicate how unique factorization for ideals can be used to define a divisibility theory for ideals similar to the divisibility of integers in $\mathbb{Z}$.

**Definition.** If $A$ and $B$ are ideals in the integral domain $R$ then $B$ is said to *divide* $A$ (and $A$ is *divisible by* $B$) if there is an ideal $C$ in $R$ with $A = BC$.

If $B$ divides $A$ then certainly $A \subseteq B$. If $R$ is a Dedekind Domain, the converse is true: $A \subseteq B$ implies $C = AB^{-1} \subseteq BB^{-1} = R$ so $C$ is an ideal in $R$ with $BC = A$.

We can also define the notion of the *greatest common divisor* $(A, B)$ of two ideals $A$ and $B$: $(A, B)$ divides both $A$ and $B$ and any ideal dividing both $A$ and $B$ divides $(A, B)$. The second statement in the next proposition shows that this greatest common divisor always exists for integral ideals in a Dedekind Domain and gives a formula for it similar to the formula for the greatest common divisor of two integers.

**Proposition 17.** Suppose $R$ is a Dedekind Domain and $A$, $B$ are two nonzero ideals in $R$, with prime ideal factorizations $A = P_1^{e_1} \cdots P_n^{e_n}$ and $B = P_1^{f_1} \cdots P_n^{f_n}$ (where $e_i, f_i \geq 0$ for $i = 1, \ldots, n$). Then

**(1)** $A \subseteq B$ if and only if $B$ divides $A$ (i.e., "to contain is to divide") if and only if $f_i \leq e_i$ for $i = 1, \ldots, n$,

**(2)** $A + B = (A, B) = P_1^{\min(e_1, f_1)} \cdots P_n^{\min(e_n, f_n)}$, so in particular $A$ and $B$ are relatively prime, $A + B = R$, if and only if they have no prime ideal factors in common.

*Proof:* We proved the first statement in (1) above. If each $f_i \leq e_i$, then taking $C = P_1^{e_1 - f_1} \cdots P_n^{e_n - f_n} \subseteq R$ shows that $B$ divides $A$. Conversely, if $B$ divides $A$, then writing $C$ as a product of prime ideals in $A = BC$ shows that $f_i \leq e_i$ for all $i$, which proves all of (1). Since $A + B$ is the smallest ideal containing both $A$ and $B$, (2) now follows from (1).

**Proposition 18.** *(Chinese Remainder Theorem)* Suppose $R$ is a Dedekind Domain, $P_1, P_2, \ldots, P_n$ are distinct prime ideals in $R$ and $a_i \geq 0$ are integers, $i = 1, \ldots, n$. Then

$$R/P_1^{a_1} \cdots P_n^{a_n} \cong R/P_1^{a_1} \times R/P_2^{a_2} \times \cdots \times R/P_n^{a_n}.$$

Equivalently, for any elements $r_1, r_2, \ldots, r_n \in R$ there exists an element $r \in R$, unique up to an element in $P_1^{a_1} \cdots P_n^{a_n}$, with

$$r \equiv r_1 \bmod P_1^{a_1}, \quad r \equiv r_2 \bmod P_2^{a_2}, \quad \ldots, \quad r \equiv r_n \bmod P_n^{a_n}.$$

*Proof:* This is immediate from Theorem 17 in Section 7.6 since the previous proposition shows that the $P_i^{a_i}$ are pairwise comaximal ideals.

**Corollary 19.** Suppose $I$ is an ideal in the Dedekind Domain $R$. Then

**(1)** there is an ideal $J$ of $R$ relatively prime to $I$ such that the product $IJ = (a)$ is a principal ideal,

**(2)** if $I$ is nonzero then every ideal in the quotient $R/I$ is principal; equivalently, if $I_1$ is an ideal of $R$ containing $I$ then $I_1 = I + Rb$ for some $b \in R$, and

**(3)** every ideal in $R$ can be generated by two elements; in fact if $I$ is nonzero and $0 \neq a \in I$ then $I = Ra + Rb$ for some $b \in I$.

*Proof:* Suppose $I = P_1^{e_1} \cdots P_n^{e_n}$ is the prime ideal factorization of $I$ in $R$. For each $i = 1, \ldots, n$, let $r_i$ be an element of $P_i^{e_i} - P_i^{e_i + 1}$. By the proposition, there is an element $a \in R$ with $a \equiv r_i \bmod P_i^{e_i + 1}$ for all $i$. Hence $a \in P_i^{e_i} - P_i^{e_i + 1}$ for all $i$, so the power of $P_i$ in prime ideal factorization of $(a)$ is precisely $e_i$ by (1) of Proposition 17:

$$(a) = P_1^{e_1} \cdots P_n^{e_n} P_{n+1}^{e_{n+1}} \cdots P_m^{e_m}$$

for some prime ideals $P_{n+1}, \ldots, P_m$ distinct from $P_1, \ldots, P_n$. Letting $J = P_{n+1}^{e_{n+1}} \cdots P_m^{e_m}$ gives (1). For (2), by the Chinese Remainder Theorem it suffices to prove that every ideal in $R/P^m$ is principal in the case of a power of a prime ideal $P$, and this is immediate since $R/P^m \cong R_P/P^m R_P$ and the localization $R_P$ is a P.I.D. Finally, (3) follows from (2) by taking $I = Ra$.

The first statement in Corollary 19 shows that there is an integral ideal $J$ relatively prime to $I$ lying in the inverse class of $I$ in the class group of $R$. One can even impose additional conditions on $J$, cf. Exercise 11.

**Corollary 20.** If $R$ is a Dedekind Domain then $R$ is a P.I.D. (i.e., $R$ has class number 1) if and only if $R$ is a U.F.D.

*Proof:* Every P.I.D. is a U.F.D., so suppose that $R$ is a U.F.D. and let $P$ be any prime ideal in $R$. Then $P = Ra + Rb$ for some $a \neq 0$ and $b$ in $R$ by Corollary 19. We have $(a') \subseteq P$ for one of the irreducible factors $a'$ of $a$ since their product is an element in the prime $P$, and then $P$ divides $(a')$ in $R$ by Proposition 17(1). It follows that $P = (a')$ is principal since $(a')$ is a prime ideal (Proposition 12 in Section 8.3). Since every ideal in $R$ is a product of prime ideals, every ideal of $R$ is principal, i.e., $R$ is a P.I.D.

Corollary 20 shows that the class number of a Dedekind domain $R$ gives a measure of the failure of unique factorization of elements. It is a fundamental result in algebraic number theory that the class number of the ring of integers of an algebraic number field is finite. For general Dedekind Domains, however, the class number need not be finite. In fact, for any abelian group $A$ (finite or infinite) there is a Dedekind Domain whose class group is isomorphic to $A$.

## Modules over Dedekind Domains and the Fundamental Theorem of Finitely Generated Modules

We turn next to the consideration of modules over Dedekind Domains $R$. Every fractional ideal of $R$ is an $R$-module and the first statement in the following proposition shows that two fractional ideals of $R$ are isomorphic as $R$-modules if and only if they represent the same element in the class group of $R$.

**Proposition 21.** Let $R$ be a Dedekind Domain with fraction field $K$.

**(1)** Suppose $I$ and $J$ are two fractional ideals of $R$. Then $I \cong J$ as $R$-modules if and only if $I$ and $J$ differ by a nonzero principal ideal: $I = (a)J$ for some $0 \neq a \in K$.

**(2)** More generally, suppose $I_1, I_2, \ldots, I_n$ and $J_1, J_2, \ldots, J_m$ are nonzero fractional ideals in the fraction field $K$ of the Dedekind Domain $R$. Then

$$I_1 \oplus I_2 \oplus \cdots \oplus I_n \cong J_1 \oplus J_2 \oplus \cdots \oplus J_m$$

as $R$-modules if and only if $n = m$ and the product ideals $I_1 I_2 \cdots I_n$ and $J_1 J_2 \cdots J_n$ differ by a principal ideal:

$$I_1 I_2 \cdots I_n = (a) J_1 J_2 \cdots J_n$$

for some $0 \neq a \in K$.

**(3)** In particular,

$$I_1 \oplus I_2 \oplus \cdots \oplus I_n \cong \underbrace{R \oplus \cdots \oplus R}_{n-1 \text{ factors}} \oplus (I_1 I_2 \cdots I_n)$$

and $R^n \oplus I \cong R^n \oplus J$ if and only if $I$ and $J$ differ by a principal ideal: $I = (a)J$, $a \in K$.

*Proof:* Multiplication by $0 \neq a \in K$ gives an $R$-module isomorphism from $J$ to $(a)J$, so if $I = (a)J$ we have $I \cong J$ as $R$-modules. For the converse, observe that we

may assume $J \neq 0$ and then $I \cong J$ implies $R \cong J^{-1}I$. But this says that $J^{-1}I = aR$ is principal (with generator $a$ given by the image of $1 \in R$), i.e., $I = (a)J$, proving (1).

We next show that for any nonzero fractional ideals $I$ and $J$ that $I \oplus J \cong R \oplus IJ$. Replacing $I$ and $J$ by isomorphic $R$-modules $aI$ and $bJ$, if necessary, we may assume that $I$ and $J$ are integral ideals that are relatively prime (cf. Exercise 12), so that $I + J = R$ and $I \cap J = IJ$. It is easy to see that the map from $I \oplus J$ to $I + J = R$ defined by mapping $(x, y)$ to $x + y$ is a surjective $R$-module homomorphism with kernel $I \cap J = IJ$, so we have an exact sequence

$$0 \longrightarrow IJ \longrightarrow I \oplus J \longrightarrow R \longrightarrow 0$$

of $R$-modules. This sequence splits since $R$ is free, so $I \oplus J \cong R \oplus IJ$, as claimed.

The first statement in (3) now follows by induction, and combining this statement with (1) shows that if $I_1 \cdots I_n = (a)J_1 \cdots J_n$ for some nonzero $a \in K$ then $I_1 \oplus \cdots \oplus I_n$ is isomorphic to $J_1 \oplus \cdots \oplus J_n$. This proves the "if" statement in (2). It remains to prove the "only if" statement in (2) since the corresponding statement in (3) is a special case. So suppose $I_1 \oplus I_2 \oplus \cdots \oplus I_n \cong J_1 \oplus J_2 \oplus \cdots \oplus J_m$ as $R$-modules.

Since $I \otimes_R K$ is the localization of the ideal $I$ in $K$ (cf. Proposition 41 in Section 15.4) it follows that $I \otimes_R K \cong K$ for any nonzero fractional ideal $I$ of $K$. Since tensor products commute with direct sums, $(I_1 \oplus \cdots \oplus I_n) \otimes_R K \cong K^n$ is an $n$-dimensional vector space over $K$. Similarly, $J_1 \oplus \cdots \oplus J_m \otimes_R K \cong K^m$, from which it follows that $n = m$.

Note that replacing $I_1$ by the isomorphic fractional ideal $a_1^{-1}I_1$ for any nonzero element $a_1 \in I_1$ does not effect the validity of the statements in (2). Hence we may assume $I_1$ contains $R$, and similarly we may assume that each of the fractional ideals in (2) contains $R$. Let $\varphi$ denote the $R$-module isomorphism from $I_1 \oplus \cdots \oplus I_n$ to $J_1 \oplus \cdots \oplus J_n$. For $i = 1, 2, \ldots, n$ define

$$\varphi((0, \ldots, 0, 1, 0, \ldots, 0)) = (a_{1,i}, a_{2,i}, \ldots, a_{n,i}) \in J_1 \oplus J_2 \oplus \cdots \oplus J_n$$

where $1 \in I_i$ on the left hand side occurs in position $i$. Since $\varphi$ is an $R$-module homomorphism it follows that

$$J_j = a_{j,1}I_1 + a_{j,2}I_2 + \cdots + a_{j,i}I_i + \cdots + a_{j,n}I_n$$

for each $j = 1, 2, \ldots, n$. Taking the product of these ideals for $j = 1, 2, \ldots, n$ it follows that

$$(a_{j_1,1}a_{j_2,2} \cdots a_{j_n,n})I_1I_2 \cdots I_n \subseteq J_1J_2 \cdots J_n$$

for any permutation $\{j_1, j_2, \ldots, j_n\}$ of $\{1, 2, \ldots, n\}$. Hence

$$dI_1I_2 \cdots I_n \subseteq J_1J_2 \cdots J_n$$

where $d$ is the determinant of the matrix $(a_{i,j})$, since the determinant is the sum of terms $\epsilon(\sigma)a_{1.\sigma(1)} \cdots a_{n.\sigma(n)}$ where $\epsilon(\sigma)$ is the sign of the permutation $\sigma$ of $\{1, 2, \ldots, n\}$. Similarly, for $j = 1, \ldots, n$, define

$$\varphi^{-1}((0, \ldots, 0, 1, 0, \ldots, 0)) = (b_{1,j}, b_{2,j}, \ldots, b_{n,j}) \in I_1 \oplus I_2 \oplus \cdots \oplus I_n$$

where $1 \in J_j$ on the left hand side occurs in position $j$. The product of the two matrices $(a_{i,j})$ and $(b_{i,j})$ is just the identity matrix, so $d \neq 0$ and the determinant of the matrix $(b_{i,j})$ is $d^{-1}$. As above we have

$$d^{-1}J_1J_2 \cdots J_n \subseteq I_1I_2 \cdots I_n,$$

which shows that $I_1 I_2 \cdots I_n = (a) J_1 J_2 \cdots J_n$, where $0 \neq a = d^{-1} \in K$, completing the proof of the proposition.

We now consider finitely generated modules over Dedekind Domains and prove a structure theorem for such modules extending the results in Chapter 12 for finitely generated modules over P.I.D.s.

Recall that the *rank* of $M$ is the maximal number of $R$-linearly independent elements in $M$, or, equivalently, the dimension of $M \otimes_R K$ as a $K$-vector space, where $K$ is the fraction field of $R$ (cf. Exercises 1–4, 20 in Section 12.1).

**Theorem 22.** Suppose $M$ is a finitely generated module over the Dedekind Domain $R$. Let $n \geq 0$ denote the rank of $M$ and let $\mathrm{Tor}(M)$ be the torsion submodule of $M$. Then

$$M \cong \underbrace{R \oplus R \oplus \cdots \oplus R \oplus I}_{n \text{ factors}} \oplus \mathrm{Tor}(M)$$

for some ideal $I$ of $R$, and

$$\mathrm{Tor}(M) \cong R/P_1^{e_1} \times R/P_2^{e_2} \times \cdots \times R/P_s^{e_s}$$

for some $s \geq 0$ and powers $P_i^{e_i}$, $e_1 \geq 1$, of (not necessarily distinct) prime ideals. The ideals $P_i^{e_i}$ for $i = 1, \ldots, s$ are unique and the ideal $I$ is unique up to multiplication by a principal ideal.

*Proof:* Suppose first that $M$ is a finitely generated torsion free module over $R$, i.e., $\mathrm{Tor}(M) = 0$. Then the natural $R$-module homomorphism from $M$ to $M \otimes_R K$ is injective, so we may view $M$ as an $R$-submodule of the vector space $M \otimes_R K$. If $M$ has rank $n$ over $R$, then $M \otimes_R K$ is a vector space over $K$ of dimension $n$. Let $x_1, \ldots, x_n$ be a basis for $M \otimes_R K$ over $K$ and let $m_1, \ldots, m_s$ be $R$-module generators for $M$. Each $m_i$, $i = 1, \ldots, s$ can be written as a $K$-linear combination of $x_1, \ldots, x_n$. Let $0 \neq d \in R$ be a common denominator for all the coefficients in $K$ of these linear combinations, and set $y_i = x_i/d$, $i = 1, \ldots, n$. Then

$$M \subseteq Ry_1 + \cdots + Ry_n \subset Kx_1 + \cdots + Kx_n$$

which shows that $M$ is contained in a *free* $R$-submodule of rank $n$ and every element $m$ in $M$ can be written uniquely in the form

$$m = a_1 y_1 + \cdots + a_n y_n$$

with $a_1, \ldots, a_n \in R$. The map $\varphi : M \to R$ defined by $\varphi(a_1 y_1 + \cdots + a_n y_n) = a_n$ is an $R$-module homomorphism, so we have an exact sequence

$$0 \longrightarrow \ker \varphi \longrightarrow M \xrightarrow{\varphi} I_1 \longrightarrow 0$$

where $I_1$ is the image of $\varphi$ in $R$, hence is an ideal in $R$. The submodule $\ker \varphi$ is also a torsion free $R$-module whose rank is at most $n - 1$ (since it is contained in $Ry_1 + \cdots + Ry_{n-1}$), and it follows by comparing ranks that $I_1$ is nonzero and that $\ker \varphi$ has rank precisely $n - 1$. By (4) of Theorem 15, $I_1$ is a projective $R$-module, so this sequence splits:

$$M \cong I_1 \oplus (\ker \varphi).$$

By induction on the rank, we see that a finitely generated torsion free $R$-module is isomorphic to the direct sum of $n$ nonzero ideals of $R$:

$$M \cong I_1 \oplus I_2 \oplus \cdots \oplus I_n.$$

Since $I_1, \ldots, I_n$ are each projective $R$-modules, it follows that any finitely generated torsion free $R$-module is projective.

If now $M$ is any finitely generated $R$-module, the quotient $M/\mathrm{Tor}(M)$ is finitely generated and torsion free, hence projective by what was just proved. The exact sequence

$$0 \longrightarrow \mathrm{Tor}(M) \longrightarrow M \longrightarrow M/\mathrm{Tor}(M) \longrightarrow 0$$

therefore splits, and so

$$M \cong \mathrm{Tor}(M) \oplus (M/\mathrm{Tor}(M)).$$

By the results in the previous paragraph $M/\mathrm{Tor}(M)$ is isomorphic to a direct sum of $n$ nonzero ideals of $R$, and by Proposition 21 we obtain

$$M \cong \underbrace{R \oplus R \oplus \cdots \oplus R \oplus I}_{n \text{ factors}} \oplus \mathrm{Tor}(M)$$

for some ideal $I$ of $R$. The uniqueness statement regarding the ideal $I$ is also immediate from the uniqueness statement in Proposition 21(3).

It remains to prove the statements regarding the torsion submodule $\mathrm{Tor}(M)$. Suppose then that $N$ is a finitely generated torsion $R$-module. Let $I = \mathrm{Ann}(N)$ be the annihilator of $N$ in $R$ and suppose $I = P_1^{e_1} \cdots P_t^{e_t}$ is the prime ideal factorization of $I$ in $R$, where $P_1, \ldots, P_t$ are distinct prime ideals. Then $N$ is a module over $R/I$, and

$$R/I \cong R/P_1^{e_1} \times R/P_2^{e_2} \times \cdots \times R/P_t^{e_t}.$$

It follows that

$$N \cong (N/P_1^{e_1}N) \times (N/P_2^{e_2}N) \times \cdots \times (N/P_t^{e_t}N)$$

as $R$-modules. Each $N/P^eN$ is a finitely generated module over $R/P^e \cong R_P/P^eR_P$ where $R_P$ is the localization of $R$ at the prime $P$, i.e., is a finitely generated module over $R_P$ that is annihilated by $P^eR_P$. Since $R$ is a Dedekind Domain, each $R_P$ is a P.I.D. (even a D.V.R.), so we may apply the Fundamental Theorem for Finitely Generated Modules over a P.I.D. to see that each $N/P^eN$ is isomorphic as an $R_P$-module to a direct sum of finitely many modules of the form $R_P/P^fR_P$ where $f \leq e$. It follows that each $N/P^eN$ is isomorphic as an $R$-module to a direct sum of finitely many modules of the form $R/P^fR$ where $f \leq e$. This proves that $N$ is isomorphic to the direct sum of finitely many modules of the form $R/P_i^{f_i}$ for various prime ideals $P_i$. Hence $\mathrm{Tor}(M)$ can be decomposed into a direct sum as in the statement in the theorem.

Finally, it remains to prove that the ideals $P_i^{e_i}$ for $i = 1, \ldots, s$ in the decomposition of $\mathrm{Tor}(M)$ are unique. This is similar to the uniqueness argument in the proof of Theorem 10 in Section 12.1 (cf. also Exercises 11–12 in Section 12.1): for any prime ideal $P$ of $R$, the quotient $P^{i-1}M/P^iM$ is a vector space over the field $R/P$ and the difference $\dim_{R/P} P^{i-1}M/P^iM - \dim_{R/P} P^iM/P^{i+1}M$ is the number of direct summands of $M$ isomorphic to $R/P^i$, hence is uniquely determined by $M$. This concludes the proof of the theorem.

If $M$ is a finitely generated module over the Dedekind Domain $R$ as in Theorem 22, then the isomorphism type of $M$ as an $R$-module is determined by the *rank* $n$, the prime powers $P_i^{e_i}$ for $i = 1, \ldots, s$ (called the *elementary divisors* of $M$, and the class of the ideal $I$ in the class group of $R$ (called the *Steinitz class* of $M$). Note that a P.I.D. is the same as a Dedekind Domain whose class number is 1, in which case every nonzero ideal $I$ of $R$ is isomorphic as an $R$-module simply to $R$. In this case, Theorem 22 reduces to the elementary divisor form of the structure theorem for finitely generated modules over P.I.D.s in Chapter 12. There is also an invariant factor version of the description of the torsion $R$-modules in Theorem 22 (cf. Exercise 14).

The next result extends the characterization of finitely generated projective modules over P.I.D.s (Exercise 21 in Section 12.1) to Dedekind Domains.

**Corollary 23.** A finitely generated module over a Dedekind Domain is projective if and only if it is torsion free.

*Proof:* We showed that a finitely generated torsion free $R$-module is projective in the proof of Theorem 22, so by the decomposition of $M$ in Theorem 22, $M$ is projective if and only if $\text{Tor}(M)$ is projective (cf. Exercise 3 in Section 10.5). To complete the proof it suffices to show that no nonzero torsion $R$-module is projective, which is left as an exercise (cf. Exercise 15).

## EXERCISES

1. If $R$ is an integral domain, show that every fractional ideal of $R$ is invertible if and only if every integral ideal of $R$ is invertible.
2. Suppose $R$ is an integral domain with fraction field $K$ and $A_1, A_2, \ldots, A_n$ are fractional ideals of $R$ whose product is a nonzero principal fractional ideal: $A_1 A_2 \cdots A_n = Rx$ for some $0 \neq x \in K$. For each $i = 1, \ldots, n$ prove that $A_i$ is an invertible fractional ideal with inverse $(x^{-1})A_1 \cdots A_{i-1} A_{i+1} \cdots A_n$.
3. Suppose $R$ is an integral domain with fraction field $K$ and $P$ is a nonzero prime ideal in $R$. Show that the fractional ideals of $R_P$ in $K$ are the $R_P$-modules of the form $AR_P$ where $A$ is a fractional ideal of $R$.
4. Suppose $R$ is an integral domain with fraction field $K$ and $A$ is a fractional ideal of $R$ in $K$. Let $A' = \{x \in K \mid xA \subseteq R\}$ as in Proposition 9.
   **(a)** For any prime ideal $P$ in $R$ prove that the localization $(A')_P$ of $A'$ at $P$ is a fractional ideal of $R_P$ in $K$.
   **(b)** If $A$ is a finitely generated $R$-module, prove that $(A')_P = (A_P)'$ where $(A_P)'$ is the fractional $R_P$ ideal $\{x \in K \mid xA_P \subseteq R_P\}$ corresponding to the localization $A_P$.
5. If $Q_1$ is a $P_1$-primary ideal and $Q_2$ is a $P_2$-primary ideal where $P_1$ and $P_2$ are comaximal ideals in a Noetherian ring $R$, prove that $Q_1$ and $Q_2$ are also comaximal. [Use Proposition 14 in Section 15.2.]
6. Suppose $R$ is a Dedekind Domain with fraction field $K$.
   **(a)** Prove that every nonzero fractional ideal of $R$ in $K$ can be written uniquely as the product of distinct prime powers $P_1^{a_1} \cdots P_n^{a_n}$ where the $a_i$ are nonzero integers, possibly negative.

**(b)** If $0 \neq x \in K$, let $P^{v_P(x)}$ be the power of the prime $P$ in the factorization of the principal ideal $(x)$ as in (a) (where $v_P(x) = 0$ if $P$ is not one of the primes occurring). Prove $v_P$ is a valuation on $K$ with valuation ring $R_P$, the localization of $R$ at $P$.

**7.** Suppose $R$ is a Noetherian integral domain that is not a field. Prove that $R$ is a Dedekind Domain if and only if for every maximal ideal $M$ of $R$ there are no ideals $I$ of $R$ with $M^2 \subset I \subset M$. [Use Exercise 12 in Section 1 and Theorems 7 and 15.]

**8.** Suppose $R$ is a Noetherian integral domain with Krull dimension 1. Prove that every nonzero ideal $I$ in $R$ can be written uniquely as a product of primary ideals whose radicals are all distinct. [Cf. the proof of Theorem 15. Use the uniqueness of the primary components belonging to the isolated primes in a minimal primary decomposition (Theorem 21 in Section 15.2).]

**9.** Suppose $R$ is an integral domain. Prove that $R_P$ is a D.V.R. for every nonzero prime ideal $P$ if and only if $R_M$ is a D.V.R. for every nonzero maximal ideal.

**10.** Suppose $R$ is a Noetherian integral domain that is not a field. Prove that $R$ is a Dedekind Domain if and only if nonzero primes $M$ are maximal and every $M$-primary ideal is a power of $M$.

**11.** If $I$ and $J$ are nonzero ideals in the Dedekind Domain $R$ show there exists an integral ideal $I_1$ in $R$ that is relatively prime to both $I$ and $J$ such that $I_1 I$ is a principal ideal in $R$.

**12.** If $I$ and $J$ are nonzero fractional ideals for the Dedekind Domain $R$ prove there are elements $\alpha, \beta \in K$ such that $\alpha I$ and $\beta J$ are nonzero integral ideals in $R$ are relatively prime.

**13.** Suppose $I$ and $J$ are nonzero ideals in the Dedekind Domain $R$. Prove that there is an ideal $I_1 \cong I$ that is relatively prime to $J$. [Use Corollary 19 to find an ideal $I_2$ with $I_2 I = (a)$ and $(I_2, J) = R$. If $I_2 = P_1^{e_1} \cdots P_n^{e_n}$, choose $b \in R$ with $b \in P_i^{e_i} - P_i^{e_i+1}$ and $b \equiv 1 \bmod P$ for every prime $P$ dividing $J$. Show that $(b) = I_2 I_1$ for some ideal $I_1$ and consider $(a)I_1$ to prove that $I_1 \cong I$.]

**14.** Prove that every finitely generated torsion module over a Dedekind Domain $R$ is isomorphic to a direct sum $R/I_1 \oplus R/I_2 \oplus \cdots \oplus R/I_n$ with unique nonzero ideals $I_1, \ldots, I_n$ of $R$ satisfying $I_1 \subseteq I_2 \subseteq \cdots \subseteq I_n$ (called the *invariant factors* of $M$). [cf. Section 12.1.]

**15.** If $P$ is a nonzero prime ideal in the Dedekind Domain $R$ prove that $R/P^n$ is not a projective $R$-module for any $n \geq 1$. [Consider the exact sequence $0 \to P^n/P^{n+1} \to R/P^{n+1} \to R/P^n \to 0$.] Conclude that if $M \neq 0$ is a finitely generated torsion $R$-module then $M$ is not projective. [cf. Exercise 3, Section 10.5.]

**16.** Prove that the class number of the Dedekind Domain $R$ is 1 if and only if every finitely generated projective $R$-module is free.

**17.** Suppose $R$ is a Dedekind Domain.

**(a)** Show that $I \sim J$ if and only if $I \cong J$ as $R$-modules defines an equivalence relation on the set of nonzero fractional ideals of $R$. Let $C(R)$ be the corresponding set of $R$-module isomorphism classes and let $[I] \in C(R)$ denote the equivalence class containing the fractional ideal $I$ of $R$.

**(b)** Show that the multiplication $[I][J] = [I \oplus J]$ gives a well defined binary operation with respect to which $C(R)$ is an abelian group with identity $1 = [R]$.

**(c)** Prove that the abelian group $C(R)$ in (b) is isomorphic to the class group of $R$.

**18.** If $R$ is a Dedekind Domain and $I$ is any nonzero ideal, prove that $R/I$ contains only finitely many ideals. In particular, show that $R/I$ is an Artinian ring.

**19.** Suppose $I$ is a nonzero fractional ideal in the Dedekind Domain $R$. Explicitly exhibit $I$ as a direct summand of a free $R$-module to show that $I$ is projective. [Consider $I \oplus I^{-1}$

and use Proposition 21.]

**20.** Suppose $I$ and $J$ are two nonzero fractional ideals in the Dedekind Domain $R$ and that $I^n = J^n$ for some $n \neq 0$. Prove that $I = J$.

**21.** Suppose $K$ is an algebraic number field and $\mathcal{O}_K$ is the ring of integers in $K$. If $P$ is a nonzero prime ideal in $\mathcal{O}_K$ prove that $P = (p, \pi)$ for some prime $p \in \mathbb{Z}$ and algebraic integer $\pi \in \mathcal{O}_K$.

**22.** Suppose $K = \mathbb{Q}(\sqrt{D})$ is a quadratic extension of $\mathbb{Q}$ where $D$ is a squarefree integer and $\mathcal{O}_K$ is the ring of integers in $K$.

**(a)** Prove that $|\mathcal{O}_K/(p)| = p^2$. [Observe that $\mathcal{O}_K \cong \mathbb{Z}^2$ as an abelian group.]

**(b)** Use Corollary 16 to show that there are 3 possibilities for the prime ideal factorization of $(p)$ in $\mathcal{O}_K$:
(i) $(p) = P$ is a prime ideal with $|\mathcal{O}_K/P| = p^2$,
(ii) $(p) = P_1P_2$ with distinct prime ideals $P_1$, $P_2$ and $|\mathcal{O}_K/P_1| = |\mathcal{O}_K/P_2| = p$,
(iii) $(p) = P^2$ for some prime ideal $P$ with $|\mathcal{O}_K/P| = p$.

(In cases (i), (ii), and (iii) the prime $p$ is said to be *inert*, *split*, or *ramified* in $\mathcal{O}_K$, respectively. The set of ramified primes is finite: the primes $p$ dividing $D$ if $D \equiv 1, 2 \bmod 4$; $p = 2$ and the primes $p$ dividing $D$ if $D \equiv 3 \bmod 4$. Cf. Exercise 31 in Section 15.5.)

**(c)** Determine the prime ideal factorizations of the primes $p = 2, 3, 5, 7, 11$ in the ring of integers $\mathcal{O}_K = \mathbb{Z}[\sqrt{-5}]$ of $K = \mathbb{Q}(\sqrt{-5})$.

**23.** Let $\mathcal{O}$ be the ring of integers in the algebraic closure $\overline{\mathbb{Q}}$ of $\mathbb{Q}$.

**(a)** Show that the infinite sequence of ideals in $\mathcal{O}$ $(2) \subseteq (\sqrt{2}) \subseteq (\sqrt[4]{2}) \subseteq (\sqrt[8]{2}) \subseteq \cdots$ is strictly increasing, and so $\mathcal{O}$ is not Noetherian.

**(b)** Show that $\mathcal{O}$ has Krull dimension 1. [Use Theorem 26 in Section 15.3.]

**(c)** Let $K$ be a number field and let $I$ be any ideal in $\mathcal{O}_K$. Show that there is some finite extension $L$ of $K$ such that $I$ becomes principal when extended to $\mathcal{O}_L$, i.e., the ideal $I\mathcal{O}_L$ is principal (where $L$ depends on $I$)—you may use the theorem that the class group of $K$ is a finite group. [cf. Exercise 20.]

**(d)** Prove that $\mathcal{O}$ is a Bezout Domain (cf. Section 8.1).

**24.** Suppose $F$ and $K$ are algebraic number fields with $\mathbb{Q} \subseteq F \subseteq K$, with rings of integers $\mathcal{O}_F$ and $\mathcal{O}_K$, respectively. Since $\mathcal{O}_F \subseteq \mathcal{O}_K$, the ring $\mathcal{O}_K$ is naturally a module over $\mathcal{O}_F$.

**(a)** Prove $\mathcal{O}_K$ is a torsion free $\mathcal{O}_F$-module of rank $n = [K : F]$. [Compute ranks over $\mathbb{Z}$.] If $\mathcal{O}_K$ is *free* over $\mathcal{O}_F$ then $\mathcal{O}_K$ is said to have a *relative integral basis* over $\mathcal{O}_F$.

**(b)** Prove that if $F$ has class number 1 then $\mathcal{O}_K$ has a relative integral basis over $\mathcal{O}_F$.

If $K = \mathbb{Q}(\sqrt{-5}, \sqrt{2})$ then the ring of integers $\mathcal{O}_K$ is given by

$$\mathcal{O}_K = \mathbb{Z} + \mathbb{Z}\sqrt{-5} + \mathbb{Z}\sqrt{-10} + \mathbb{Z}\omega \qquad \text{where } \omega = (\sqrt{-10} + \sqrt{2})/2.$$

**(c)** If $F_1 = \mathbb{Q}(\sqrt{2})$ prove that $\mathcal{O}_K$ has a relative integral basis over $\mathcal{O}_{F_1}$ and find an explicit basis $\{\alpha, \beta\}$: $\mathcal{O}_K = \mathcal{O}_{F_1} \cdot \alpha + \mathcal{O}_{F_1} \cdot \beta$.

**(d)** If $F_2 = \mathbb{Q}(\sqrt{-5})$, show that $P_3 = (3, 1 + \sqrt{-5}) = (3, 5 - \sqrt{-5})$ is a prime ideal of $\mathcal{O}_{F_2}$ that is not principal and that $\mathcal{O}_K = \mathcal{O}_{F_2} \cdot 1 + (1/3)P_3 \cdot \omega$. [Check that $\sqrt{-10} = -(5 - \sqrt{-5})\omega/3$.] Conclude that the Steinitz class of $\mathcal{O}_K$ as a module over $\mathcal{O}_{F_2}$ is the nontrivial class of $P_3$ in the class group of $\mathcal{O}_{F_2}$ and so there is no relative integral basis of $\mathcal{O}_K$ over $\mathcal{O}_{F_2}$.

**(e)** Determine whether $\mathcal{O}_K$ has a relative integral basis over the ring of integers of the remaining quadratic subfield $F_3 = \mathbb{Q}(\sqrt{-10})$ of $K$.

**25.** Suppose $C$ is a nonsingular irreducible affine curve over an algebraically closed field $k$. Prove that the coordinate ring $k[C]$ is a Dedekind Domain.

CHAPTER 17

# Introduction to Homological Algebra and Group Cohomology

Let $R$ be a ring with 1. In Section 10.5 we saw that a short exact sequence

$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0 \tag{17.1}$$

of $R$-modules gives rise to an exact sequence of abelian groups

$$0 \longrightarrow \mathrm{Hom}_R(N, D) \xrightarrow{\varphi'} \mathrm{Hom}_R(M, D) \xrightarrow{\psi'} \mathrm{Hom}_R(L, D) \tag{17.2}$$

for any $R$-module $D$ and that the homomorphism $\psi'$ is in general not surjective so this sequence cannot always be extended to a short exact sequence. Equivalently, homomorphisms from $L$ to $D$ cannot in general be lifted to homomorphisms from $M$ into $D$. In this chapter we introduce some of the techniques of "homological algebra," which provide a method of extending some exact sequences in a natural way. For the situation above one obtains an infinite exact sequence involving the "cohomology groups" $\mathrm{Ext}^n_R(\_\_, D)$ (cf. Theorem 8), and these groups provide a measure of the set of homomorphisms from $L$ into $D$ that cannot be extended to $M$. We then consider the analogous questions for the other two functors considered in Section 10.5, namely taking homomorphisms *from* $D$ into the terms of the sequence (1) and tensoring the sequence (1) with $D$.

In the subsequent sections we concentrate on an important special case of this general type of homological construction—the "cohomology of finite groups." We make explicit the computations in this case and indicate some applications of these techniques to establish some new results in group theory. In this sense, Sections 2–4 may be considered as an explicit "example" illustrating some uses of the general theory in Section 1.

Cohomology and homology groups occur in many areas of mathematics. The formal notions of homology and cohomology groups and the general area of homological algebra arose from algebraic topology around the middle of the 20th century in the study of the relation between the higher homotopy groups and the fundamental group of a topological space, although the study of certain specific cohomology groups, such as Schur's work on group extensions (described in Section 4), predates this by half a century. As with much of algebra, the ideas common to a number of different areas were abstracted into general theories. Much of the language of homology and cohomology reflects its topological origins: homology groups, chains, cycles, boundaries, etc.

In this section we describe some general terminology and results in homological algebra leading to the so called Long Exact Sequence in Cohomology. We then define certain (cohomology) groups associated to the sequence (2) and apply the general homological results to obtain a long exact sequence extending this sequence at the right end. We then indicate the corresponding development for sequences obtained by taking homomorphisms from $D$ to the terms in (1) or by tensoring the terms with $D$.

We begin with a generalization of the notion of an exact sequence, namely a sequence of abelian group homomorphisms where successive maps compose to zero (i.e., the image of one map is contained in the kernel of the next):

**Definition.** Let $\mathcal{C}$ be a sequence of abelian group homomorphisms:

$$0 \longrightarrow C^0 \xrightarrow{d_1} C^1 \longrightarrow \cdots \longrightarrow C^{n-1} \xrightarrow{d_n} C^n \xrightarrow{d_{n+1}} \cdots. \tag{17.3}$$

**(1)** The sequence $\mathcal{C}$ is called a *cochain complex* if the composition of any two successive maps is zero: $d_{n+1} \circ d_n = 0$ for all $n$.

**(2)** If $\mathcal{C}$ is a cochain complex, its $n^{\text{th}}$ *cohomology group* is the quotient group $\ker d_{n+1}/\text{image } d_n$, and is denoted by $H^n(\mathcal{C})$.

There is a completely analogous "dual" version in which the homomorphisms are between groups in *decreasing* order, in which case the sequence corresponding to (3) is written $\cdots \xrightarrow{d_{n+1}} C_n \xrightarrow{d_n} \cdots \xrightarrow{d_1} C_0 \to 0$. Then if the composition of any two successive homomorphisms is zero, the complex is called a *chain complex*, and its *homology groups* are defined as $H_n(C) = \ker d_n/\text{image } d_{n+1}$. For chain complexes the notation is often chosen so that the indices appear as subscripts and are decreasing, whereas for cochain complexes the indices are superscripts and are increasing. We shall instead use a uniform notation for the maps on both, since it will be clear from the context whether we are dealing with a chain or a cochain complex.

Chain complexes were the first to arise in topological settings, with cochain complexes soon following. With our applications in Section 2 in mind, we shall concentrate on cochains and cohomology, although all of the general results in this section have similar statements for chains and homology. We shall also be interested in the situation where each $C^n$ is an $R$-module and the homomorphisms $d_n$ are $R$-module homomorphisms (referred to simply as a *complex of $R$-modules*), in which case the groups $H^n(C)$ are also $R$-modules.

Note that if $\mathcal{C}$ is a cochain (respectively, chain) complex then $\mathcal{C}$ is an exact sequence if and only if all its cohomology (respectively, homology) groups are zero. Thus the $n^{\text{th}}$ cohomology (respectively, homology) group measures the failure of exactness of a complex at the $n^{\text{th}}$ stage.

**Definition.** Let $\mathcal{A} = \{A^n\}$ and $\mathcal{B} = \{B^n\}$ be cochain complexes. A *homomorphism of complexes* $\alpha : \mathcal{A} \to \mathcal{B}$ is a set of homomorphisms $\alpha_n : A^n \to B^n$ such that for every $n$ the following diagram commutes:

$$\begin{array}{ccccccc} \cdots & \longrightarrow & A^n & \longrightarrow & A^{n+1} & \longrightarrow & \cdots \\ & & \downarrow{\scriptstyle \alpha_n} & & \downarrow{\scriptstyle \alpha_{n+1}} & & \\ \cdots & \longrightarrow & B^n & \longrightarrow & B^{n+1} & \longrightarrow & \cdots \end{array} \tag{17.4}$$

**Proposition 1.** A homomorphism $\alpha : \mathcal{A} \to \mathcal{B}$ of cochain complexes induces group homomorphisms from $H^n(\mathcal{A})$ to $H^n(\mathcal{B})$ for $n \geq 0$ on their respective cohomology groups.

*Proof:* It is an easy exercise to show that the commutativity of (4) implies that the images and kernels at each stage of the maps in the first row are mapped to the corresponding images and kernels for the maps in the second row, thus giving a well defined map on the respective quotient (cohomology) groups.

**Definition.** Let $\mathcal{A} = \{A^n\}$, $\mathcal{B} = \{B^n\}$ and $\mathcal{C} = \{C^n\}$ be cochain complexes. A *short exact sequence* of complexes $0 \to \mathcal{A} \xrightarrow{\alpha} \mathcal{B} \xrightarrow{\beta} \mathcal{C} \to 0$ is a sequence of homomorphisms of complexes such that $0 \to A^n \xrightarrow{\alpha_n} B^n \xrightarrow{\beta_n} C^n \to 0$ is short exact for every $n$.

One of the main features of cochain complexes is that they lead to long exact sequences in cohomology, which is our first main result:

**Theorem 2.** *(The Long Exact Sequence in Cohomology)* Let $0 \to \mathcal{A} \xrightarrow{\alpha} \mathcal{B} \xrightarrow{\beta} \mathcal{C} \to 0$ be a short exact sequence of cochain complexes. Then there is a long exact sequence of cohomology groups:

$$\begin{aligned} 0 \to H^0(\mathcal{A}) \to H^0(\mathcal{B}) \to H^0(\mathcal{C}) \xrightarrow{\delta_0} H^1(\mathcal{A}) \\ \to H^1(\mathcal{B}) \to H^1(\mathcal{C}) \xrightarrow{\delta_1} H^2(\mathcal{A}) \to \cdots \end{aligned} \tag{17.5}$$

where the maps between cohomology groups at each level are those in Proposition 1. The maps $\delta_n$ are called *connecting homomorphisms.*

*Proof:* The details of this proof are somewhat lengthy. For each $n$ the verification that the sequence $H^n(\mathcal{A}) \to H^n(\mathcal{B}) \to H^n(\mathcal{C})$ is exact is a straightforward check of the definition of exactness of each map, similar to the proof of Theorem 33 in Section 10.5. The construction of a connecting homomorphism $\delta_n$ is outlined in Exercise 2. Some work is then needed to show that $\delta_n$ is a homomorphism, and that the sequence is exact at $\delta_n$.

One immediate consequence of the existence of the long exact sequence in Theorem 2 is the fact that if any two of the cochain complexes $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ are exact, then so is the third (cf. Exercise 6).

## Homomorphisms and the Groups $Ext_R^n(A,B)$

To apply Theorem 2 to analyze the sequence (2), we try to produce a cochain complex whose first few cohomology groups in the long exact sequence (5) agree with the terms in (2). To do this we introduce the notion of a "resolution" of an $R$-module:

**Definition.** Let $A$ be any $R$-module. A *projective resolution* of $A$ is an exact sequence

$$\cdots \longrightarrow P_n \xrightarrow{d_n} P_{n-1} \longrightarrow \cdots \xrightarrow{d_1} P_0 \xrightarrow{\epsilon} A \longrightarrow 0 \tag{17.6}$$

such that each $P_i$ is a projective $R$-module.

Every $R$-module has a projective resolution: Let $P_0$ be any free (hence projective) $R$-module on a set of generators of $A$ and define an $R$-module homomorphism $\epsilon$ from $P_0$ onto $A$ by Theorem 6 in Chapter 10. This begins the resolution $\epsilon : P_0 \to A \to 0$. The surjectivity of $\epsilon$ ensures that this sequence is exact. Next let $K_0 = \ker \epsilon$ and let $P_1$ be any free module mapping onto the submodule $K_0$ of $P_0$; this gives the second stage $P_1 \to P_0 \to A$ which, by construction, is also exact. We can continue this way, taking at the $n^{\text{th}}$ stage a free $R$-module $P_{n+1}$ that maps surjectively onto the submodule $\ker d_n$ of $P_n$, obtaining in fact a *free* resolution of $A$.

One of the reasons that *projective* modules are used in the resolution of $A$ is that this makes it possible to lift various maps (cf. the proof of Proposition 4 following, for instance).

In general a projective resolution is infinite in length, but if $A$ is itself projective, then it has a very simple projective resolution of finite length, namely $0 \to A \xrightarrow{1} A \to 0$ given by the identity map from $A$ to itself.

Given the projective resolution (6), we may form a related sequence by taking homomorphisms of each of the terms into $D$, keeping in mind that this reverses the direction of the homomorphisms. This yields the sequence

$$\begin{aligned} 0 \longrightarrow &\mathrm{Hom}_R(A, D) \xrightarrow{\epsilon} \mathrm{Hom}_R(P_0, D) \xrightarrow{d_1} \mathrm{Hom}_R(P_1, D) \xrightarrow{d_2} \cdots \\ &\cdots \xrightarrow{d_{n-1}} \mathrm{Hom}_R(P_{n-1}, D) \xrightarrow{d_n} \mathrm{Hom}_R(P_n, D) \xrightarrow{d_{n+1}} \cdots \end{aligned} \tag{17.7}$$

where to simplify notation we have denoted the induced maps from $\mathrm{Hom}_R(P_{n-1}, D)$ to $\mathrm{Hom}_R(P_n, D)$ for $n \geq 1$ again by $d_n$ and similarly for the map induced by $\epsilon$ (cf. Section 10.5). This sequence is not necessarily exact, however it *is* a cochain complex (this is part of the proof of Theorem 33 in Section 10.5). The corresponding cohomology groups have a special name.

**Definition.** Let $A$ and $D$ be a $R$-modules. For any projective resolution of $A$ as in (6) let $d_n : \mathrm{Hom}_R(P_{n-1}, D) \to \mathrm{Hom}_R(P_n, D)$ for all $n \geq 1$ as in (7). Define

$$\mathrm{Ext}_R^n(A, D) = \ker d_{n+1} / \text{image } d_n$$

where $\mathrm{Ext}_R^0(A, D) = \ker d_1$. The group $\mathrm{Ext}_R^n(A, D)$ is called the $n^{\text{th}}$ *cohomology group derived from the functor* $\mathrm{Hom}_R(\_\_, D)$. When $R = \mathbb{Z}$ the group $\mathrm{Ext}_{\mathbb{Z}}^n(A, D)$ is also denoted simply $\mathrm{Ext}^n(A, D)$.