Algebra & Geometry

A First Course on Varieties

Emily Clader and Dustin Ross

December 12, 2023

Contents

Notation and Conventions

- R and S denote rings.
- K denotes a field, which is assumed to be algebraically closed after Chapter 1.
- All rings are commutative with unity, denoted 1.
- It is not assumed that $1 \neq 0$ for general rings, but it is assumed that $1 \neq 0$ for integral domains (and fields).
- All ring homomorphisms $\varphi : R \to S$ satisfy $\varphi(1) = 1$.
- $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ denote the sets of integers, rational numbers, real numbers, and complex numbers, respectively.
- $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ denotes the set of natural numbers.

Chapter 0

Polynomial Rings

LEARNING OBJECTIVES FOR CHAPTER 0

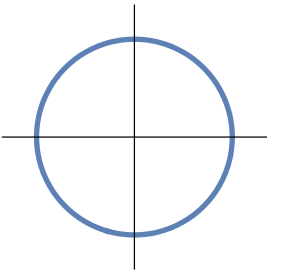
- Review the concepts of polynomials and polynomial rings.
- Review the notions of ideals and quotients, especially in the context of polynomial rings.
- Become familiar with properties of polynomials over fields, such as the existence and uniqueness of irreducible factorizations.
- Develop tools to determine if a given polynomial is irreducible.

Algebraic geometry studies solutions of polynomial equations by building a dictionary between the geometry of the solution sets and the algebra of the defining polynomial equations. In this preliminary chapter, we develop the algebraic notions of polynomial rings that are prerequisite to the study of algebraic geometry. The chapter culminates with a proof of the important fact that every polynomial over a field factors uniquely into irreducible polynomials.

The reader is assumed to have taken a first course in ring theory and to be familiar with the notions of rings, integral domains, and fields. However, knowledge beyond the most fundamental definitions and results is not expected. Surely, this chapter will read more quickly for those students with a more robust algebraic background, while a student newer to abstract algebra may choose to devote a significant amount of time to mastering the contents of these pages.

The purpose of this chapter is to establish the algebraic foundation on which the rest of the book is built, focusing on the fundamental properties of polynomial rings that will be most useful for later developments. As such, the choice has often been made to forego generality for the sake of brevity; for example, we will never consider noncommutative rings, so every ideal will be a two-sided ideal. Nearly all of the examples in this chapter illustrate concepts in the specific setting of polynomial rings, and the authors hope that the intuition developed in this setting might help the interested student study these concepts more generally.

Section 0.1 Polynomials



Polynomials and their solutions are some of the first objects we encounter in our mathematical lives. For example, you may even remember the first time you learned that the solutions in $\mathbb{R}^2$ of the two-variable polynomial equation

$$x^2 + y^2 - 1 = 0$$

describe the unit circle. However, if our goal is to study the unit circle, then there are many other polynomials that one might choose to describe it; for example, the unit circle is also equal to the solutions of either

$$2x^2 + 2y^2 - 2 = 0 \quad \text{or} \quad x^4 + 2x^2y^2 + y^4 - 1 = 0.$$

This leads to a natural question: is there a *best* polynomial that describes the unit circle? The answer proposed by algebraic geometry is, in some sense, the most egalitarian: all of the polynomials that describe the unit circle are equally important, and we should study them together as a *set*. What does it mean, then, to study a *set* of polynomials?

As we will soon learn, the set of polynomials describing the unit circle is much more than just a subset of polynomials, it has important algebraic structure that reflects the geometry of the circle. To be able to describe this algebraic structure in this example and beyond, we must first establish precise notation and terminology regarding the set of polynomials and their algebraic structure. To begin our formal discussion of polynomials, we start with the notion of a monomial.

0.1 **DEFINITION** *Monomials*

A *monomial* in the variables $x_1, \ldots, x_n$ is an expression of the form

$$x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n},$$

where $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ is the *exponent vector*. Two monomials are equal if and only if they have the same exponent vector.

When there are only a few variables, they are often represented with distinct letters such as x, y, and z.

The variables $x_1, \ldots, x_n$ should be viewed as formal symbols, and their role is simply to serve as placeholders for $\alpha_1, \ldots, \alpha_n$. The data of a monomial in $x_1, \ldots, x_n$ is equivalent to the data of an exponent vector $(\alpha_1, \ldots, \alpha_n)$; however, placing each α_i as the exponent of x_i will prove useful when multiplication of monomials is defined below. Variables that appear with an exponent of 0 are typically omitted; for example,

$$x^2y^3z^0 = x^2y^3 \quad \text{and} \quad x^0y^0z^0 = 1.$$

As in the case of $f(x,y) = x^2 + y^2 - 1$, polynomials are built by taking linear combinations of monomials. In the most general setting, the coefficients of these linear combinations belong to an arbitrary ring R, as in the following definition.

0.2 **DEFINITION** *Polynomials over R*

A *polynomial* in the variables $x_1, \ldots, x_n$ over R is an expression of the form

$$f = f(x_1, \ldots, x_n) = \sum_{\alpha \in \mathbb{N}^n} a_\alpha x^\alpha,$$

where $a_\alpha \in R$ for each $\alpha \in \mathbb{N}^n$ and $a_\alpha = 0$ for all but finitely many α. Two polynomials

$$f = \sum_\alpha a_\alpha x^\alpha \quad \text{and} \quad g = \sum_\alpha b_\alpha x^\alpha$$

are equal if and only if $a_\alpha = b_\alpha$ for all $\alpha \in \mathbb{N}^n$.

0.3 **EXAMPLE** Polynomials

Polynomials are typically written as finite sums, omitting all summands that have a coefficient of zero.

The following are two examples of polynomials in the variables x and y over the ring of integers $\mathbb{Z}$:

$$f = xy^2 + 3xy + 2 \quad \text{and} \quad g = -xy + 4.$$

The polynomials f and g can also be viewed as having coefficients in $\mathbb{Q}$, $\mathbb{R}$, or any other ring containing $\mathbb{Z}$. Observe that we can create new polynomials from f and g by adding them and multiplying them in the familiar way:

$$f + g = xy^2 + 2xy + 6,$$
$$fg = -x^2y^3 - 3x^2y^2 + 4xy^2 + 10xy + 8.$$

As the reader is encouraged to verify in Exercise 0.1.1, the operations of addition and multiplication, formalized in the next definition, endow the set of polynomials with the structure of a ring.

0.4 **DEFINITION** *Polynomial rings*

The *polynomial ring* $R[x_1, \ldots, x_n]$ is the set of all polynomials in variables $x_1, \ldots, x_n$ over R. Polynomial addition and multiplication are defined by

$$\Big(\sum_\alpha a_\alpha x^\alpha\Big) + \Big(\sum_\alpha b_\alpha x^\alpha\Big) = \sum_\alpha (a_\alpha + b_\alpha) x^\alpha$$

and

$$\Big(\sum_\alpha a_\alpha x^\alpha\Big)\Big(\sum_\alpha b_\alpha x^\alpha\Big) = \sum_\alpha \Big(\sum_{\alpha_1 + \alpha_2 = \alpha} a_{\alpha_1} b_{\alpha_2}\Big) x^\alpha.$$

The additive identity $0 \in R[x_1, \ldots, x_n]$ is the polynomial for which $a_\alpha = 0$ for all α, and the multiplicative identity $1 \in R[x_1, \ldots, x_n]$ is the polynomial for which

$$a_\alpha = \begin{cases} 1 & \text{if } (\alpha_1, \ldots, \alpha_n) = (0, \ldots, 0) \\ 0 & \text{if } (\alpha_1, \ldots, \alpha_n) \neq (0, \ldots, 0). \end{cases}$$

When working with polynomials rings, it can be useful to leverage their recursive nature in order to be able to use proofs by induction on the number of variables. The next result is somewhat self-evident, but we state it carefully as it will be used often.

0.5 PROPOSITION *Recursive nature of polynomial rings*

There is a canonical isomorphism of rings

$$R[x_1,\dots,x_n] = R[x_1,\dots,x_{n-1}][x_n],$$

where the right-hand side is the ring of polynomials in one variable x_n with coefficients in the ring $R[x_1,\dots,x_{n-1}]$.

The next example illustrates the main idea behind Proposition 0.5.

0.6 EXAMPLE $\mathbb{Z}[x,y]$ versus $\mathbb{Z}[x][y]$

Consider the polynomial

$$f = x^3y^2 + xy^2 - 2xy - x + 1 \in \mathbb{Z}[x,y].$$

We can view f as an element of $\mathbb{Z}[x][y]$ by grouping all terms that have the same exponent in y. In doing so, we write

$$f = (x^3 + x)y^2 + (-2x)y + (-x + 1) \in \mathbb{Z}[x][y].$$

As a polynomial in y, the coefficients of f are $x^3 + x$, $-2x$, and $-x + 1$, all of which are elements of $\mathbb{Z}[x]$.

PROOF OF PROPOSITION 0.5

To prove the proposition, we describe the canonical ring isomorphism, which, as illustrated in Example 0.6, simply groups all terms of a polynomial in $R[x_1,\dots,x_n]$ for which x_n appears with the same exponent. To make this more precise, consider the following function:

When two rings are isomorphic, there will often be a multitude of possible isomorphisms. The word canonical *means that there is one natural choice among all possible isomorphisms. The symbol $\cong$ is used to denote isomorphisms, while $=$ is used for canonical isomorphisms.*

$$\varphi : R[x_1,\dots,x_n] \to R[x_1,\dots,x_{n-1}][x_n]$$
$$\sum_{\alpha} a_\alpha x_1^{\alpha_1}\cdots x_n^{\alpha_n} \mapsto \sum_{d\geq 0}\left(\sum_{\alpha\in\mathbb{N}^{n-1}} a_\alpha x_1^{\alpha_1}\cdots x_{n-1}^{\alpha_{n-1}}\right)x_n^d.$$

The verification that φ is a ring isomorphism is Exercise 0.1.3. □

The following result concerning polynomial rings over integral domains is a first application of Proposition 0.5.

0.7 PROPOSITION *Polynomials over integral domains*

If R is an integral domain, then $R[x_1, \ldots, x_n]$ is an integral domain.

PROOF We proceed by induction on the number of variables.

(Base case) Let $f, g \in R[x_1]$ be nonzero. We must prove that $fg \neq 0$. Write

$$f = a_d x_1^d + a_{d-1} x_1^{d-1} \cdots + a_0 \quad \text{and} \quad g = b_e x_1^e + b_{e-1} x_1^{e-1} \cdots + b_0,$$

where $a_d \neq 0$ and $b_e \neq 0$. By definition of multiplication in $R[x_1]$,

$$fg = (a_d b_e) x_1^{d+e} + (a_d b_{e-1} + a_{d-1} b_e) x_1^{d+e-1} + \cdots + a_0 b_0.$$

Since R is an integral domain, $a_d b_e \neq 0$. Since fg has at least one nonzero coefficient, we conclude that $fg \neq 0$.

(Induction step) Assume $S = R[x_1, \ldots, x_{n-1}]$ is an integral domain; we must show that $R[x_1, \ldots, x_n]$ is an integral domain. By Proposition 0.5,

$$R[x_1, \ldots, x_n] = S[x_n].$$

Since S is an integral domain, the argument used in the base case immediately implies that $S[x_n]$, and thus $R[x_1, \ldots, x_n]$, is an integral domain. □

The numbers d and e appearing in the proof of Proposition 0.7 are important attributes of the polynomials f and g, called their degrees. The next definition generalizes the notion of degree to polynomials with any number of variables.

0.8 DEFINITION *Monomial and polynomial degree*

The *degree* of a monomial $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ is

$$\deg(x^\alpha) = \alpha_1 + \cdots + \alpha_n \in \mathbb{N}.$$

The *degree* of a nonzero polynomial $f = \sum a_\alpha x^\alpha \in R[x_1, \ldots, x_n]$ is

$$\deg(f) = \max\{\deg(x^\alpha) \mid a_\alpha \neq 0\}.$$

0.9 EXAMPLE Degree

The monomials x^2yz, z^4, x, and 1 have degrees 4, 4, 1, and 0, respectively, and

$$\deg(x^2yz + z^4 + x + 1) = 4.$$

The reader may have noticed that the zero polynomial has not been assigned a degree, which is intentional. One of the most useful properties of degree is additivity, described in the next result, which fails for any choice of $\deg(0) \in \mathbb{N}$.

0.10 **PROPOSITION** *Additivity of degree*

If R is an integral domain and $f, g \in R[x_1, \ldots, x_n]$ are nonzero, then

$$\deg(fg) = \deg(f) + \deg(g).$$

PROOF Let $f, g \in R[x_1, \ldots, x_n]$ be nonzero polynomials of degree d and e, respectively. Write

$$f = f_d + f_{d-1} + \cdots + f_0 \quad \text{and} \quad g = g_e + g_{e-1} + \cdots + g_0,$$

where f_i comprises all terms in f of degree i, and similar for g_j. By assumption, $f_d \neq 0$ and $g_e \neq 0$. Some time reflecting should convince the reader that degree is additive on monomials, so the highest degree monomials that could possibly appear with nonzero coefficient in fg have degree $d + e$, and any such monomial must arise in the product $f_d g_e$. Since $R[x_1, \ldots, x_n]$ is an integral domain, we see that $f_d g_e \neq 0$, from which we conclude that $\deg(fg) = d + e = \deg(f) + \deg(g)$. □

Exercises for Section 0.1

0.1.1 Prove that addition and multiplication of polynomials endows $R[x_1, \ldots, x_n]$ the structure of a ring (commutative with unity).

0.1.2 Group the terms of the polynomial

$$f = xyz^2 + xyz + z^3 + x^2z^2 + yz^2 + z + x + 1 \in R[x, y, z]$$

to view it as an element of $R[x, y][z]$, $R[x][y, z]$, and $R[y, z][x]$.

0.1.3 Prove that the function

$$\varphi : R[x_1, \ldots, x_n] \to R[x_1, \ldots, x_{n-1}][x_n]$$

defined in the proof of Proposition 0.5 is a ring isomorphism.

0.1.4 Prove that R is canonically isomorphic to a subring of $R[x_1, \ldots, x_n]$. (The polynomials that lie in R are called *constant polynomials*.)

0.1.5 Let $a = (a_1, \ldots, a_n) \in R^n$. Prove that there is a unique ring homomorphism

$$\varphi_a : R[x_1, \ldots, x_n] \to R$$

such that $\varphi_a(r) = r$ for all $r \in R$ and $\varphi_a(x_i) = a_i$ for all i. (This homomorphism is called *evaluation at a* and is usually written $\varphi_a(f) = f(a_1, \ldots, a_n)$.)

0.1.6 Show that Propositions 0.7 and 0.10 fail if R is not an integral domain.

0.1.7 Show that Proposition 0.10 cannot be extended to any choice of $\deg(0) \in \mathbb{N}$.

0.1.8 Prove that $\deg(f + g) \leq \max\{\deg(f), \deg(g)\}$.

Section 0.2 Irreducible polynomials

Algebraic geometry is primarily interested in polynomial rings $K[x_1, \ldots, x_n]$ with coefficients in a field K. In order to study these polynomial rings, it is useful to have a good understanding of the "atomic" elements—those polynomials that cannot be factored in a nontrivial way—and how each polynomial decomposes into the atomic ones. As motivation for these ideas, it is instructive to turn to the more familiar case of the ring of integers $\mathbb{Z}$, where the atomic elements are the *prime* numbers.

Recall that an integer $p \in \mathbb{Z}_{\geq 2}$ is *prime* if for all $m, n \in \mathbb{Z}$,

$$p = mn \implies m = \pm 1 \text{ or } n = \pm 1. \tag{0.11}$$

In other words, a prime integer is one that cannot be factored in a nontrivial way. One of the central results in number theory (and in all of mathematics, for that matter) is the existence and uniqueness of prime factorization: every integer $n \in \mathbb{Z}_{\geq 2}$ can be written as a product of prime numbers in a unique way, up to reordering the factors.

We would like to study these ideas more generally, especially in the context of polynomial rings. To do so, we begin with the ring-theoretic definition of "atomic," including the notion of a *unit*, which generalizes the $\pm 1 \in \mathbb{Z}$ appearing in (0.11).

0.12 **DEFINITION** *Units and irreducible elements*

An element $u \in R$ is called a *unit* if it has a multiplicative inverse. The set of units in R is denoted $R^* \subseteq R$. An element $p \in R$ is called *irreducible* if it is neither zero nor a unit, and

$$p = ab \implies a \in R^* \text{ or } b \in R^*.$$

An element is *reducible* if it is neither zero, a unit, nor irreducible.

In other words, a nonzero element is irreducible if it cannot be factored into a product of two nonunits. In the case of polynomial rings $K[x_1, \ldots, x_n]$, the units are the nonzero constant polynomials (Exercise 0.2.2):

The distinction between units and nonunits is necessary because every element factors if we allow units:

$$a = u(u^{-1}a).$$

$$K[x_1, \ldots, x_n]^* = K^* = K \setminus \{0\}.$$

It follows that a polynomial $f \in K[x_1, \ldots, x_n]$ is irreducible if and only if

(i) f is nonconstant, and

(ii) f cannot be written as a product of two nonconstant polynomials.

0.13 **EXAMPLE** Linear polynomials in $K[x_1, \ldots, x_n]$ are irreducible.

We say that a polynomial $f \in K[x_1, \ldots, x_n]$ is *linear* if $\deg(f) = 1$. If f is linear and $f = gh$, then by additivity of degree,

$$1 = \deg(f) = \deg(g) + \deg(h).$$

It follows that either $\deg(g) = 0$ or $\deg(h) = 0$, implying that g or h is constant.

0.14 **EXAMPLE** $y - x^2$ is irreducible in $K[x, y]$.

Suppose

$$y - x^2 = gh$$

for some $g, h \in K[x, y]$. As an element of $K[x][y]$, $y - x^2$ has degree 1, implying that (as polynomials in y) one of g or h must also have degree 1 and the other must have degree 0. Without loss of generality, we can write

$$g = ay + b \quad \text{and} \quad h = c$$

where $a, b, c \in K[x]$. This implies that $y - x^2 = acy + bc$. By matching coefficients of y, we see that $ac = 1$, which implies that $h = c \in K^*$. Thus, $y - x^2$ is irreducible.

0.15 **EXAMPLE** $x^2 + 1 \in \mathbb{C}[x]$ versus $x^2 + 1 \in \mathbb{R}[x]$

In $\mathbb{C}[x]$, we have a factorization

$$x^2 + 1 = (x - \mathrm{i})(x + \mathrm{i}),$$

which shows that $x^2 + 1$ is reducible in $\mathbb{C}[x]$. In $\mathbb{R}[x]$, on the other hand, it is not possible to factor $x^2 + 1$ into two linear factors (Exercise 0.2.3), implying that $x^2 + 1$ is irreducible in $\mathbb{R}[x]$. This example illustrates how the behavior of polynomial rings heavily depends on the choice of coefficient field K.

Since prime factorization is such a fundamental property of the ring of integers, it is useful to have a generalization of this property to the setting of integral domains. The following definition captures the essence of unique prime factorization in $\mathbb{Z}$.

0.16 **DEFINITION** *Unique factorization domain*

An integral domain R is called a *factorization domain* (*FD*) if

(i) for every nonzero, nonunit $a \in R$, there exist irreducible elements $p_1, \ldots, p_\ell \in R$ such that $a = p_1 \cdots p_\ell$.

It is called a *unique factorization domain* (*UFD*) if, in addition,

(ii) whenever $p_1 \cdots p_\ell = q_1 \cdots q_m$ for some irreducible elements p_i and q_j, then $\ell = m$ and, after possibly reordering, there exist units u_i such that $p_i = u_i q_i$ for all i.

Unique prime factorization in the ring of integers implies that $\mathbb{Z}$ is a UFD. One of the fundamental properties of $K[x_1, \ldots, x_n]$ is that it is also a UFD, as we will see over the course of this chapter.

To prove that $K[x_1, \ldots, x_n]$ is a UFD, we must prove that irreducible factorizations exist and that they are unique. In the current section, we content ourselves with proving existence of irreducible factorizations.

There are many examples of rings that are not UFDs. We direct the reader to Exercise 0.3.15 for one such example.

0.17 PROPOSITION *$K[x_1,\dots,x_n]$ is a FD*

If $f \in K[x_1,\dots,x_n]$ is a nonconstant polynomial, then there exist irreducible polynomials $p_1,\dots,p_\ell$ such that $f = p_1 \cdots p_\ell$.

PROOF We proceed by induction on the degree of f.

(Base case) If $\deg(f) = 1$, then f is irreducible by Example 0.13. In particular, f has an irreducible factorization (with $\ell = 1$).

(Induction step) Assume that every polynomial of degree less than d can be factored into irreducible polynomials, and suppose $f \in K[x_1,\dots,x_n]$ has degree d. If f is irreducible, then f has an irreducible factorization with $\ell = 1$. If f is not irreducible, then $f = gh$ with $\deg(g) < d$ and $\deg(h) < d$. By the induction hypothesis, there are irreducible factorizations

$$g = p_1 \cdots p_\ell \quad \text{and} \quad h = q_1 \cdots q_k.$$

Thus, f admits an irreducible factorization

$$f = p_1 \cdots p_\ell \cdot q_1 \cdots q_k. \qquad \square$$

0.18 EXAMPLE An irreducible factorization

It follows from Examples 0.13 and 0.15 that

$$x^2 + 1 = (x - \mathrm{i})(x + \mathrm{i})$$

is an irreducible factorization of $x^2 + 1 \in \mathbb{C}[x]$.

For inspiration on how one might prove uniqueness of irreducible factorizations, we return to the familiar case of $\mathbb{Z}$. The key to proving uniqueness of prime factorizations in the integers is Euclid's Lemma, which says that $p \in \mathbb{Z}_{\geq 2}$ is prime if and only if, for all $m, n \in \mathbb{Z}$,

$$p \mid mn \implies p \mid m \text{ or } p \mid n.$$

This second characterization of prime integers naturally generalizes to rings.

We use the standard notation $a \mid b$ as shorthand for "a divides b," which means that $b = ca$ for some c.

0.19 DEFINITION *Prime element*

An element $p \in R$ is *prime* if it is neither zero nor a unit and, for all $a, b \in R$,

$$p \mid ab \implies p \mid a \text{ or } p \mid b.$$

As we will see below, the question of whether irreducible factorizations are unique can be reduced to proving the ring-theoretic analogue of Euclid's Lemma. More specifically, given an integral domain R, Euclid's Lemma translates to a statement equating prime elements in R with irreducible elements. The following result verifies one of the implications: in integral domains, all primes are irreducible.

0.20 PROPOSITION *Prime implies irreducible*

In an integral domain, every prime element is irreducible.

PROOF Let R be an integral domain and let $p \in R$ be prime. Toward proving that p is irreducible, suppose

$$p = ab \tag{0.21}$$

for some $a, b \in R$; we must prove that $a \in R^*$ or $b \in R^*$. Notice that (0.21) implies, in particular, that $p \mid ab$. By primeness of p, we have $p \mid a$ or $p \mid b$. Without loss of generality, assume $p \mid a$ and write $a = pc$ for some $c \in R$. Substituting this expression into (0.21), we see that

$$p = pcb.$$

Since R is an integral domain and $p \neq 0$, we can cancel p from both sides to obtain $1 = cb$, implying that $b \in R^*$. □

The converse of Proposition 0.20 is not true in general. In fact, the converse is, in some sense, equivalent to uniqueness of factorizations, which is the content of the next result.

Look ahead to Exercise 0.3.15 for an example where the converse of Proposition 0.20 fails.

0.22 PROPOSITION *FDs versus UFDs*

Let R be a factorization domain. Then R is a unique factorization domain if and only if every irreducible element of R is prime.

PROOF We prove both implications.

($\Rightarrow$) Suppose that R is a UFD and let $p \in R$ be irreducible. Towards proving that p is prime, suppose that $p \mid ab$ for some $a, b \in R$, and choose $c \in R$ such that $ab = pc$. Since R is a FD, everything admits irreducible factorizations:

$$a = p_1 \cdots p_k, \quad b = q_1 \cdots q_\ell, \quad \text{and} \quad c = r_1 \cdots r_m.$$

Thus, we have two irreducible factorizations

$$p_1 \cdots p_k \cdot q_1 \cdots q_\ell = p \cdot r_1 \cdots r_m.$$

Since R is a UFD, there exists a unit u such that $p = up_i$ or $p = uq_j$ for some i or j. It follows that $p \mid a$ or $p \mid b$.

($\Leftarrow$) Suppose that every irreducible element of R is prime, and let

$$p_1 \cdots p_\ell = q_1 \cdots q_m \tag{0.23}$$

be two irreducible (hence, prime) factorizations. Assume without loss of generality that $\ell \leq m$. Since p_1 is prime and $p_1 \mid q_1 \cdots q_m$, it follows from Exercise 0.2.7 that $p_1 \mid q_j$ for some j. After possibly reordering, assume $j = 1$ and write $q_1 = u_1 p_1$

for some $u_1 \in R$. Since q_1 is irreducible, u_1 or p_1 must be a unit, but p_1 cannot be a unit because it is irreducible. Thus, u_1 is a unit. Canceling p_1 from both sides of (0.23), we obtain

$$p_2 \cdots p_\ell = u_1 q_2 \cdots q_m.$$

Since $p_2 \nmid u_1$, we can repeat the above argument to see that, after possibly reordering, there is a unit $u_2 \in R$ such that $q_2 = u_2 p_2$ and

$$p_3 \cdots p_\ell = u_1 u_2 q_3 \cdots q_m.$$

Continuing this process for ℓ steps, we see that there are units $u_1, \ldots, u_\ell$ such that $q_i = u_i p_i$ and

$$1 = u_1 \cdots u_\ell q_{\ell+1} \cdots q_m.$$

Since each q_j is irreducible, and thus not a unit, we conclude that $\ell = m$, finishing the proof. □

Since we already know that $K[x_1, \ldots, x_n]$ is a FD, Proposition 0.22 provides a strategy for proving that $K[x_1, \ldots, x_n]$ is a UFD: it suffices to show that every irreducible polynomial is prime. In order to accomplish this, we need a more robust algebraic foundation upon which we can work with these ideas. In order to build this foundation, we first turn to a discussion of ideals and quotients (Section 0.3), prime and maximal ideals (Section 0.4), and the special case of single-variable polynomials (Section 0.5). We will then return to the question of unique factorization of polynomials in Section 0.6.

Exercises for Section 0.2

0.2.1 Prove that the set of units $R^* \subseteq R$ is a group under multiplication.

0.2.2 Prove that $K[x_1, \ldots, x_n]^* = K^* = K \setminus \{0\}$.

0.2.3 Using the fact that $a^2 + 1 \neq 0$ for any real number $a \in \mathbb{R}$, prove that $x^2 + 1$ does not factor into linear terms.

0.2.4 Let $f \in K[x]$ be a polynomial of degree 2 or 3. Prove that f is irreducible if and only if there does not exist an element $a \in K$ such that $f(a) = 0$.

0.2.5 Show that the previous problem fails for polynomials of degree 4 by giving an explicit example of a reducible polynomial in $\mathbb{R}[x]$ that has no zeros.

0.2.6 Using degree arguments, prove that $x^2 + y^2 - 1$ is irreducible in $\mathbb{C}[x, y]$.

0.2.7 Suppose $p \in R$ is prime. If $p \mid a_1 \cdots a_n$, prove that $p \mid a_i$ for some i.

0.2.8 (a) Describe the units in $\mathbb{Z}[x]$.
(b) Give an example of a linear polynomial in $\mathbb{Z}[x]$ that is reducible.

0.2.9 Explain why every field is a UFD.

Section 0.3 Ideals and quotients

One of the central constructions in ring theory is that of taking quotients by ideals. In this section, we review the quotient construction, along with the most fundamental result regarding quotients—the First Isomorphism Theorem. Ideals and quotient rings are standard topics in a first course in ring theory, so the proofs in this section are left to the exercises. However, we include a number of instructive examples to illustrate how to work with ideals and quotients in the context of polynomial rings.

Our discussion of quotient rings begins with the notion of an ideal.

0.24 **DEFINITION** *Ideals*

An *ideal* of R is a nonempty subset $I \subseteq R$ satisfying two properties:

(i) for all $a, b \in I$, $a - b \in I$, and

(ii) for all $a \in I$ and $r \in R$, $ra \in I$.

In words, a nonempty subset of a ring is an ideal if (i) it is closed under subtraction and (ii) it absorbs multiplication. In many contexts, the easiest way to describe an ideal is by specifying a *generating set*. This method of describing ideals is made precise in the next definition.

0.25 **DEFINITION** *Generating sets of ideals*

If $A \subseteq R$ is a subset, then the *ideal generated by* A is the set of all R-linear combinations of elements of A:

$$\langle A \rangle = \left\{ \sum_{i=1}^{n} r_i a_i \;\middle|\; n \in \mathbb{N},\ r_1, \ldots, r_n \in R,\ a_1, \ldots, a_n \in A \right\} \subseteq R.$$

An ideal that can be generated by a single element is called *principal*.

The reader is encouraged to verify that the set $\langle A \rangle$ is, in fact, an ideal. Moreover, it is the smallest ideal of R that contains the set A (Exercise 0.3.4).

By definition, a principal ideal $\langle a \rangle$ consists of all multiples of its generator:

$$\langle a \rangle = \{ra \mid r \in R\} = \{b \in R \mid a \,|\, b\}.$$

Principal ideals are especially nice, but not all ideals in polynomial rings are generated by a single polynomial. The next example illustrates one such nonprincipal ideal (see Exercise 0.3.5).

0.26 **EXAMPLE** An ideal in $R[x_1, \ldots, x_n]$

Let $I \subseteq R[x_1, \ldots, x_n]$ be the subset consisting of all polynomials whose constant term is zero. The set of such polynomials is closed under subtraction and absorbs multiplication, so I is an ideal. Moreover, a polynomial is in I if and only if you can factor out at least one variable from each term. Regarding generators, this implies that $I = \langle x_1, \ldots, x_n \rangle$.

Given an ideal $I \subseteq R$, define a relation on R by

$$r \sim s \iff r - s \in I.$$

Using Condition (i) in Definition 0.24, it can be shown that $\sim$ is an equivalence relation (Exercise 0.3.6). The equivalence class of an element $r \in R$ under this equivalence relation is called a *coset*, denoted

$$[r] = r + I = \{s \in R \mid s \sim r\} \subseteq R.$$

We typically prefer the notation $[r]$ when the ideal I is understood from context, but use the notation $r + I$ when it is useful to emphasize the role of I. Notice that

$$[r] = [s] \iff r - s \in I.$$

0.27 **EXAMPLE** Cosets

Consider the principal ideal $I = \langle x \rangle \subseteq R[x]$. Notice that $[x+2] = [x^2+2]$ because

$$(x+2) - (x^2+2) = x - x^2 \in \langle x \rangle.$$

More generally, $[f(x)] = [g(x)]$ if and only if $f(0) = g(0) \in R$. In other words, the collection of cosets is in natural bijection with the ring R via the identification

$$[f(x)] \longmapsto f(0) \in R.$$

In the previous example, we saw that the collection of cosets is in bijection with the coefficient ring R, and can therefore be given the structure of a ring. The next definition describes how to endow the set of cosets with a ring structure for any ideal.

0.28 **DEFINITION** *Quotient rings*

Let $I \subseteq R$ be an ideal.The *quotient ring* R/I is the set of cosets

$$R/I = \{[r] \mid r \in R\}.$$

Coset addition and multiplication are defined by

$$[r] + [s] = [r+s] \quad \text{and} \quad [r][s] = [rs].$$

It is not obvious that addition and multiplication in R/I are well-defined. In particular, since different elements can be chosen to represent the same coset, it is necessary to verify that the operations are independent of the choice of coset representatives. This verification follows from Conditions (i) and (ii) in Definition 0.24; we leave the computation to the reader (Exercise 0.3.8).

In the quotient R/I, the additive identity is $0 = [0]$ and the multiplicative identity is $1 = [1]$. Since $[a] = 0$ if and only if $a \in I$, the quotienting process can be thought of as "setting elements of I equal to zero."

Notice that, for any ideal $I \subseteq R$, there is a ring homomorphism

$$\begin{aligned} \pi : R &\to R/I \\ a &\mapsto [a]. \end{aligned}$$

This homomorphism is called the *quotient homomorphism*.

0.29 **EXAMPLE** Quotient ring computations

Consider the principal ideal $\langle y - x^2 \rangle \subseteq R[x,y]$. By definition of addition and multiplication in the quotient ring, we see that

$$[y] - [x]^2 = [y - x^2] = 0 \in \frac{R[x,y]}{\langle y - x^2 \rangle}.$$

In particular, this implies that $[y] = [x]^2$. Taking this logic a step farther, we see, for example, that

$$[y]^2 = [x]^4 \quad \text{and} \quad [x]^2[y]^3 = [x]^8.$$

In general, for any polynomial $f(x,y) \in R[x,y]$, observe that

$$[f(x,y)] = f([x],[y]) = f([x],[x]^2) = [f(x,x^2)].$$

In other words, when we form the quotient by the ideal $\langle y - x^2 \rangle$, we are able to treat $y - x^2$ as the zero element and replace every occurrence of y with x^2. In particular, every element of the quotient can be represented in the variable x alone. As we will see in Example 0.32 below, the quotient ring in this example is isomorphic to $R[x]$.

The most important application of the quotient construction is that it provides a tool for turning homomorphisms into isomorphisms. The proof of the three points in the following fundamental result are left to the reader (Exercise 0.3.9).

0.30 **THEOREM** *First Isomorphism Theorem for rings*

If $\varphi : R \to S$ is a ring homomorphism, then

(i) $\operatorname{im}(\varphi) = \{s \in S \mid s = \varphi(r) \text{ for some } r \in R\}$ is a subring of S,

(ii) $\ker(\varphi) = \{r \in R \mid \varphi(r) = 0\}$ is an ideal of R, and

(iii) the function

$$\begin{aligned} [\varphi] : \frac{R}{\ker(\varphi)} &\to \operatorname{im}(\varphi) \\ [r] &\mapsto [\varphi(r)] \end{aligned}$$

is a well-defined ring isomorphism.

We close this section with a few detailed examples that demonstrate applications of the first isomorphism theorem in the context of polynomial rings. For more examples, we direct the reader to the exercises.

0.31 EXAMPLE $\langle x_1, \ldots, x_n \rangle \subseteq R[x_1, \ldots, x_n]$

Consider the ring homomorphism

$$\begin{aligned} \varphi : R[x_1, \ldots, x_n] &\to R \\ f(x_1, \ldots, x_n) &\mapsto f(0, \ldots, 0). \end{aligned}$$

Notice that $f(0, \ldots, 0)$ is simply the constant term of f. Some time reflecting should convince the reader that φ is surjective and the kernel of φ consists of all polynomials with vanishing constant term. Thus, by Example 0.26,

$$\ker(\varphi) = \langle x_1, \ldots, x_n \rangle \subseteq R[x_1, \ldots, x_n],$$

and by the First Isomorphism Theorem, we conclude that $[\varphi]$ is an isomorphism:

$$\frac{R[x_1, \ldots, x_n]}{\langle x_1, \ldots, x_n \rangle} \cong R.$$

Using similar arguments, this example can be generalized in a number of ways. See, for example, Exercises 0.3.11 and 0.3.12.

0.32 EXAMPLE $\langle y - x^2 \rangle \subseteq R[x, y]$

This example verifies that

$$\frac{R[x, y]}{\langle y - x^2 \rangle} \cong R[x].$$

Based on the computations in Example 0.29, this should make sense: we can replace every occurrence of y with x^2 and write every coset in terms of x alone. To make this argument precise using the first isomorphism theorem, it suffices to construct a surjective homomorphism $\varphi : R[x, y] \to R[x]$ with kernel $\langle y - x^2 \rangle$.

Define

$$\begin{aligned} \varphi : R[x, y] &\to R[x] \\ f(x, y) &\mapsto f(x, x^2). \end{aligned}$$

It is straightforward to convince oneself that φ is a surjective ring homomorphism. Thus, it remains to prove that $\langle y - x^2 \rangle = \ker(\varphi)$. We prove both inclusions.

($\subseteq$) Suppose $f(x, y) \in \langle y - x^2 \rangle$. This means that there exists $g(x, y) \in R[x, y]$ such that

$$f(x, y) = (y - x^2) g(x, y).$$

Evaluating φ at $f(x, y)$, we see that

$$\varphi(f(x, y)) = (x^2 - x^2) g(x, x^2) = 0,$$

so $f(x, y) \in \ker(\varphi)$.

($\supseteq$) Suppose $f(x, y) \in \ker(\varphi)$. By the computations carried out in Example 0.29, we see that $[f(x, y)] = [f(x, x^2)]$ in the quotient ring $R[x, y] / \langle y - x^2 \rangle$, which means that

$$f(x, y) - f(x, x^2) \in \langle y - x^2 \rangle. \tag{0.33}$$

Moreover, since $f(x,y) \in \ker(\varphi)$, we know that

$$0 = \varphi(f(x,y)) = f(x,x^2) \in R[x] \subseteq R[x,y]. \tag{0.34}$$

Applying (0.34) and (0.33), we conclude that

$$f(x,y) = f(x,y) - f(x,x^2) \in \langle y - x^2 \rangle.$$

Exercise 0.3.13 provides a useful generalization of this result.

Exercises for Section 0.3

0.3.1 Let $I \subseteq R$ be an ideal. Prove that $I = R$ if and only if I contains a unit.

0.3.2 Prove that ideals are closed under addition.

0.3.3 Prove that the only ideals of a field K are $\{0\}$ and K.

0.3.4 Let $A \subseteq R$ be a subset. Prove the following.

(a) The set $\langle A \rangle$ is an ideal of R.
(b) If $I \subseteq R$ is any ideal containing A, then $\langle A \rangle \subseteq I$.

0.3.5 Prove that $\langle x_1, \ldots, x_n \rangle \subseteq K[x_1, \ldots, x_n]$ is not a principal ideal if $n > 1$.

0.3.6 Let $I \subseteq R$ be an ideal and consider the relation on R given by

$$r \sim s \iff r - s \in I.$$

(a) Prove that $\sim$ is *reflexive*: $r \sim r$ for all $r \in R$.
(b) Prove that $\sim$ is *symmetric*: $r \sim s$ if and only if $s \sim r$.
(c) Prove that $\sim$ is *transitive*: if $r \sim s$ and $s \sim t$, then $r \sim t$.

Thus, $\sim$ is an equivalence relation.

0.3.7 Prove that $r + I = \{r + a \mid a \in I\}$, justifying the notation.

0.3.8 Let $I \subseteq R$ be an ideal and let $r_1, r_2, s \in R$. Prove the following.

(a) If $r_1 \sim r_2$, then $r_1 + s \sim r_2 + s$.
(b) If $r_1 \sim r_2$, then $r_1 s \sim r_2 s$.

Thus, addition and multiplication in the quotient ring R/I are well-defined.

0.3.9 Prove the first isomorphism theorem for rings.

0.3.10 Let $\varphi : R \to S$ be a ring homomorphism.

(a) If $I \subseteq S$ is an ideal, prove that $\varphi^{-1}(I) \subseteq R$ is an ideal.
(b) If φ is surjective and $I \subseteq R$ is an ideal, prove that $\varphi(I) \subseteq S$ is an ideal.
(c) Give an example of a nonsurjective ring homomorphism $\varphi : R \to S$ and an ideal $I \subseteq R$ such that $\varphi(I)$ is not an ideal.

0.3.11 Prove that

$$\frac{R[x_1,\ldots,x_n]}{\langle x_{k+1},\ldots,x_n\rangle} \cong R[x_1,\ldots,x_k].$$

0.3.12 Let $a_1,\ldots,a_n$ be elements of R. Prove that

$$\frac{R[x_1,\ldots,x_n]}{\langle x_1-a_1,\ldots,x_n-a_n\rangle} \cong R.$$

0.3.13 Let $f_1,\ldots,f_k \in R[x_1,\ldots,x_n]$ and $g \in R[x_1,\ldots,x_{n-1}]$. Consider the ring homomorphism

$$\begin{aligned}\varphi : R[x_1,\ldots,x_n] &\to R[x_1,\ldots,x_{n-1}]\\ f(x_1,\ldots,x_{n-1},x_n) &\mapsto f(x_1,\ldots,x_{n-1},g).\end{aligned}$$

Use the First Isomorphism Theorem to prove that

$$\frac{R[x_1,\ldots,x_n]}{\langle f_1,\ldots,f_k,x_n-g\rangle} \cong \frac{R[x_1,\ldots,x_{n-1}]}{\langle \varphi(f_1),\ldots,\varphi(f_k)\rangle}.$$

(Notice that Examples 0.31 and 0.32 and Exercise 0.3.11 and 0.3.12 all follow from the $k=0$ case of this result. It essentially says that quotienting by x_n-g is equivalent to replacing all occurances of x_n with g.)

0.3.14 Prove that every element of the quotient ring

$$\frac{R[x,y]}{\langle x^2+y^2-1\rangle}$$

can be represented uniquely by a polynomial of the form $f(x)+yg(x)$ where $f(x),g(x)\in R[x]$.

0.3.15 Consider the quotient ring

$$R=\frac{K[x,y]}{\langle x^2-y^3\rangle}.$$

(a) Prove that every element of R can be represented by a polynomial of the form $f(y)+xg(y)$ where $f(y),g(y)\in K[y]$.
(b) Prove that R is an integral domain.
(c) Prove that $[x],[y]\in R$ are irreducible.
(d) Prove that $[x],[y]\in R$ are not prime.
(e) Find an element in R that has two distinct irreducible factorizations.

Section 0.4 Prime and maximal ideals

Given an ideal $I \subseteq R$, how do the ring-theoretic properties of R/I translate to properties of the ideal I? For example, if R/I is an integral domain or a field, what does this tell us about I? In this section, we discuss these questions through the introduction of two special types of ideals—prime and maximal ideals—both of which are central in the study of ring theory. We provide several examples of prime and maximal ideals in the context of polynomial rings, and we close with an application of how these notions can be used to study irreducible factorizations.

0.35 **DEFINITION** *Prime and maximal ideals*

An ideal $I \subseteq R$ is *prime* if $I \neq R$ and, for all $a, b \in R$,

$$ab \in I \Longrightarrow a \in I \quad \text{or} \quad b \in I.$$

An ideal $I \subseteq R$ is *maximal* if $I \neq R$ and there does not exist an ideal $J \subseteq R$ such that

$$I \subsetneq J \subsetneq R.$$

0.36 **EXAMPLE** Eponymous example of prime ideals

It follows from the definitions (Exercise 0.4.1) that a nonzero element $p \in R$ is prime if and only if $\langle p \rangle \subseteq R$ is a prime ideal. In particular, $\langle 2 \rangle$, $\langle 3 \rangle$, $\langle 5 \rangle$, and $\langle 7 \rangle$ are all examples of prime ideals in $\mathbb{Z}$, but $\langle 1 \rangle$, $\langle 4 \rangle$, $\langle 6 \rangle$, and $\langle 8 \rangle$ are not.

0.37 **EXAMPLE** $\langle x \rangle \subseteq K[x]$ is maximal

To prove that $\langle x \rangle$ is maximal, suppose $J \subseteq K[x]$ is an ideal and $\langle x \rangle \subsetneq J$; we prove that $J = K[x]$.

Since $\langle x \rangle$ consists of all polynomials without a constant term, J must contain at least one polynomial with a nonzero constant term:

$$f = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \in J \quad \text{and} \quad a_0 \neq 0.$$

Since $\langle x \rangle$ is a subset of J, all polynomials without a constant term are elements of J. This implies that

$$g = a_1 x + a_2 x^2 + \cdots + a_n x^n \in J.$$

Since ideals are closed under subtraction, $a_0 = f - g \in J$. Since a_0 is a unit in $K[x_1, \ldots, x_n]$, this implies that $J = K[x_1, \ldots, x_n]$.

The next result is a useful characterization of prime and maximal ideals in terms of the ring-theoretic properties of their quotients.

0.38 PROPOSITION *Quotients by prime and maximal ideals*

Let $I \subseteq R$ be an ideal.

(i) The ideal I is prime if and only if R/I is an integral domain.

(ii) The ideal I is maximal if and only if R/I is a field.

PROOF We prove the forward implication for each property, and we leave the reverse implications to Exercises 0.4.4 and 0.4.5.

We prove the forward implication in (i) by contrapositive. Assume that R/I is not an integral domain, so that there exist nonzero cosets $[a], [b] \in R/I$ such that $[a][b] = 0$. By definition of the quotient ring, this implies that $a, b \notin I$ but $ab \in I$. In other words, I is not a prime ideal.

To prove the forward implication of (ii), suppose I is maximal and let $[a] \in R/I$ be a nonzero coset, meaning that $a \notin I$. We must show that $[a]$ has a multiplicative inverse. Consider the ideal generated by I and a:

$$J = \langle I, a \rangle.$$

Since $a \notin I$, $I \subsetneq J$. By maximality of I, this implies that $J = R$, so $1 \in J$. By definition of generating sets of ideals, we can write 1 as

$$1 = r_1 b_1 + \cdots + r_n b_n + sa$$

for some $b_1, \ldots, b_n \in I$ and $r_1, \ldots, r_n, s \in R$. Since I is itself an ideal, this implies that $1 = b + sa$ where $b = r_1 b_1 + \cdots + r_n b_n \in I$ and $s \in R$. Therefore, in the quotient ring R/I,

$$1 = [1] = [b + sa] = [b] + [s][a] = [s][a].$$

Thus, $[a]$ has a multiplicative inverse. □

Since every field is an integral domain, we obtain the following immediate consequence of Proposition 0.38.

0.39 COROLLARY *Maximal implies prime*

Every maximal ideal is a prime ideal.

0.40 EXAMPLE $\langle x^2 \rangle \subseteq R[x]$ is not prime

Notice that

$$[x] \in \frac{R[x]}{\langle x^2 \rangle}$$

is a nonzero element of the quotient ring, but $[x][x] = [x^2] = 0$. Thus, $[x]$ is a zero divisor. Since the quotient ring $R[x]/\langle x^2 \rangle$ contains a zero divisor, it is not an integral domain. From Proposition 0.38, we conclude that $\langle x^2 \rangle$ is not a prime ideal. In particular, this implies that $\langle x^2 \rangle$ is not maximal, which can also be verified directly:

$$\langle x^2 \rangle \subsetneq \langle x \rangle \subsetneq K[x].$$

0.41 EXAMPLE $\langle y - x^2 \rangle \subseteq K[x, y]$ is prime but not maximal

By Example 0.32,

$$K[x, y]/\langle y - x^2 \rangle \cong K[x].$$

Since $K[x]$ is an integral domain, but not a field, we conclude that $\langle y - x^2 \rangle$ is a prime ideal, but not a maximal ideal.

0.42 EXAMPLE $\langle x_1, \ldots, x_n \rangle \subseteq K[x_1, \ldots, x_n]$ is maximal

By Example 0.31,

$$K[x_1, \ldots, x_n]/\langle x_1, \ldots, x_n \rangle \cong K.$$

Since K is a field, we conclude that $\langle x_1, \ldots, x_n \rangle \subseteq K[x_1, \ldots, x_n]$ is a maximal ideal.

0.43 EXAMPLE $\langle x^2 + 1 \rangle \subseteq \mathbb{R}[x]$ versus $\langle x^2 + 1 \rangle \subseteq \mathbb{C}[x]$

Consider the quotient

$$\frac{\mathbb{R}[x]}{\langle x^2 + 1 \rangle}.$$

Observe that $[x]$ satisfies $[x]^2 = -[1] = -1$. In other words, $[x]$ is a square root of -1. We know of another ring that has a square root of -1; namely, the field of complex numbers $\mathbb{C}$.

Consider the function

$$\begin{aligned} \varphi : \mathbb{R}[x] &\to \mathbb{C} \\ f(x) &\mapsto f(\mathrm{i}). \end{aligned}$$

It can be shown (Exercise 0.4.6) that φ is a surjective ring homomorphism and that $\ker(\varphi) = \langle x^2 + 1 \rangle$. Thus, by the First Isomorphism Theorem, we conclude that the quotient ring is isomorphic to the field of complex numbers:

$$\frac{\mathbb{R}[x]}{\langle x^2 + 1 \rangle} \cong \mathbb{C}.$$

Therefore, $\langle x^2 + 1 \rangle \subseteq \mathbb{R}[x]$ is a maximal ideal.

Now, consider the quotient ring

$$\frac{\mathbb{C}[x]}{\langle x^2 + 1 \rangle}.$$

In this case, neither of the elements $[x - \mathrm{i}]$ nor $[x + \mathrm{i}]$ is zero, but their product is:

$$[x - \mathrm{i}][x + \mathrm{i}] = [x^2 + 1] = 0 \in \frac{\mathbb{C}[x]}{\langle x^2 + 1 \rangle}.$$

Thus, the quotient ring over $\mathbb{C}$ contains zero-divisors—it is not an integral domain. Therefore, $\langle x^2 - 1 \rangle \subseteq \mathbb{C}[x]$ is neither prime nor maximal.

We close this section with an application that illustrates how the notions of prime and maximal ideals can be used to help study questions concerning irreducible and prime elements. We begin with a definition of a particularly nice type of ring.

0.44 **DEFINITION** *Principal ideal domain*

An integral domain R is called a *principal ideal domain (PID)* if every ideal in R is principal.

For example, the ring of integers $\mathbb{Z}$ is a PID, as the reader is encouraged to verify in Exercise 0.4.7. As Exercise 0.3.5 shows, the polynomial ring $K[x_1, \ldots, x_n]$ is not a PID for $n > 1$ because $\langle x_1, \ldots, x_n \rangle$ is not a principal ideal. However, as we will see in the next section, the single-variable polynomial ring $K[x]$ is a PID.

The next result uses the notions of prime and maximal ideals to prove that every PID is a UFD. In particular, along with Exercise 0.4.7, this provides a self-contained proof of the uniqueness of prime factorization in $\mathbb{Z}$.

0.45 **PROPOSITION** *PIDs are UFDs*

Every principal ideal domain is a unique factorization domain.

PROOF Suppose that R is a PID. We begin by proving that R is a FD. Suppose, towards a contradiction, that there exists a nonzero, nonunit $a \in R$ that does not factor as a product of irreducible elements. This implies that a is not irreducible so we can factor it as $a = a_1 b_1$ where neither a_1 nor b_1 is a unit. If both a_1 and b_1 factored as products of irreducible elements, then so would a. Therefore, without loss of generality, assume a_1 does not factor as a product of irreducible elements and write $a_1 = a_2 b_2$ where neither a_2 nor b_2 is a unit. As before, we can assume, without loss of generality, that a_2 does not factor as a product of irreducible elements.

Continuing the above process, we recursively construct a sequence

$$(a = a_0, a_1, a_2, a_3, \ldots)$$

where, for every $i \geq 0$, $a_i = a_{i+1} b_{i+1}$ for some nonunit b_{i+1}. It follows from Exercise 0.4.8 that the ideals $\langle a_i \rangle$ fit into a chain of strict containment:

$$\langle a_0 \rangle \subsetneq \langle a_1 \rangle \subsetneq \langle a_2 \rangle \subsetneq \langle a_3 \rangle \subsetneq \ldots$$

The union of the above ideals is an ideal by Exercise 0.4.9. Since R is a PID, choose $c \in R$ such that

$$\langle c \rangle = \bigcup_{i=0}^{\infty} \langle a_i \rangle.$$

Then c must be in $\langle a_n \rangle$ for some n, which implies that $\langle c \rangle = \langle a_k \rangle$ for all $k \geq n$. This contradicts the strict containment $\langle a_n \rangle \subsetneq \langle a_{n+1} \rangle$.

Now, to prove that R is a UFD, it suffices, by Proposition 0.22 to prove that every irreducible element of R is prime. Let $p \in R$ be irreducible. To prove that p is prime, it suffices to prove that $\langle p \rangle$ is maximal. Towards this end, suppose $I \subseteq R$ is an ideal such that $\langle p \rangle \subseteq I \subseteq R$. We must prove that either $I = \langle p \rangle$ or $I = R$.

Since R is a PID, $I = \langle r \rangle$ for some $r \in R$. Since $p \in \langle p \rangle \subseteq \langle r \rangle$, it follows that $p = rs$ for some $s \in R$. By the irreducibility of p, either r or s is a unit. The reader should take a moment to verify that r being a unit implies that $I = R$ and s being a unit implies that $I = \langle p \rangle$. □

Exercises for Section 0.4

0.4.1 Prove that a nonzero element $p \in R$ is prime if and only if the principal ideal $\langle p \rangle \subseteq R$ is a prime ideal.

0.4.2 Prove that $\langle y, y - x^2 \rangle \subseteq K[x,y]$ is not prime, even though both generators are prime elements.

0.4.3 Prove that the zero ideal $\langle 0 \rangle \subseteq R$ is prime if and only if R is an integral domain.

0.4.4 Let $I \subseteq R$ be an ideal such that R/I is an integral domain. Prove that I is a prime ideal.

0.4.5 Let $I \subseteq R$ be an ideal such that R/I is a field. Prove that I is a maximal ideal.

0.4.6 Prove that

$$\frac{\mathbb{R}[x]}{\langle x^2 + 1 \rangle} \cong \mathbb{C}.$$

0.4.7 Prove that $\mathbb{Z}$ is a PID. (**Hint:** If $I \subseteq \mathbb{Z}$ is an ideal, let n be the smallest positive integer in I. Prove that $I = \langle n \rangle$.)

0.4.8 Let R be an integral domain. If $a = bc$ and $c \notin R^*$, prove that $\langle a \rangle \subsetneq \langle b \rangle$.

0.4.9 Let $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ be an ascending chain of ideals of a ring R. Prove that

$$I = \bigcup_{k=1}^{\infty} I_k$$

is an ideal of R.

0.4.10 Let R be a PID. Prove that every nonzero prime ideal in R is maximal.

Section 0.5 Single-variable polynomials

In this section, we consider the ring of single-variable polynomial rings with coefficients in a field. We introduce a number of fundamental results, concluding with the fact that $K[x]$ is a UFD, which will serve as the starting place to prove that $K[x_1, \ldots, x_n]$ is a UFD. The results in this section are direct consequences of the *polynomial division algorithm*, which is the polynomial analogue of the long division algorithm we learn in grade school.

0.46 **THEOREM** *Polynomial division algorithm*

For any $f, g \in K[x]$ with $g \neq 0$, there exist unique polynomials $q, r \in K[x]$ such that

$$f = qg + r$$

with $r = 0$ or $\deg(r) < \deg(g)$.

The polynomials q and r satisfying the conditions in the division algorithm are called the *quotient* and *remainder* of f divided by g. Notice that the remainder is zero if and only if $g \mid f$.

Before presenting a proof of the polynomial division algorithm, we work through a detailed example that illustrates the step-by-step process for computing the quotient and remainder. The reader is encourage to work out several additional examples in Exercise 0.5.1.

0.47 **EXAMPLE** Polynomial long division

Consider polynomials $f = x^3 + x^2 + 1$ and $g = x - 2$ in $\mathbb{Q}[x]$. We compute the quotient and remainder of f divided by g.

Step 1: Subtract the unique multiple of g that cancels the leading term of f:

$$f - x^2 g = 3x^2 + 1.$$

Step 2: Subtract another multiple of g to cancel the leading term of $3x^2 + 1$:

$$f - x^2 g - 3xg = -6x + 1.$$

Step 3: Subtract another multiple of g to cancel the leading term of $-6x + 1$:

$$f - x^2 g - 3xg + 6g = -11.$$

Final step: Since the polynomial -11 has degree strictly smaller than g, we stop here. Rearranging terms, we see that

$$f = qg + r$$

where $q = x^2 + 3x - 6$ and $r = -11$.

Notice that each step decreased the degree of the polynomial appearing in the right-hand side, ensuring that the process would eventually terminate.

PROOF OF THEOREM 0.46 We begin by proving that quotients and remainders exist, then we prove that they are unique.

Fix a nonzero polynomial g. In order to show that quotients and remainders exist for any f divided by this particular g, we proceed by induction on $\deg(f)$ (in the case where $f = 0$, set $q = r = 0$).

(Base case) Suppose $\deg(f) = 0$. If $\deg(g) > 0$, set $q = 0$ and $r = f$. If $\deg(g) = 0$, then g is a nonzero constant. Since K is a field, g has a multiplicative inverse. Set $q = fg^{-1}$ and $r = 0$. The reader can directly verify that these choices of q and r satisfy the conditions in the division algorithm.

(Induction step) Let $f \in K[x]$ be a polynomial of degree $d > 0$. If $\deg(g) > d$, set $q = 0$ and $r = f$. If $\deg(g) \leq d$, set $k = \deg(g)$ and let a_d and b_k be the leading coefficients of f and g, respectively. By construction, the polynomial

$$\tilde{f} = f - a_d b_k^{-1} x^{d-k} g$$

is zero or $\deg(\tilde{f}) < d$. If $\tilde{f} = 0$, set $q = a_d b_k^{-1} x^{d-k}$ and $r = 0$. Otherwise, by the induction hypothesis, choose $\tilde{q}$ and $\tilde{r}$ such that $\tilde{r} = 0$ or $\deg(\tilde{r}) < \deg(g)$ and

$$\tilde{f} = \tilde{q}g + \tilde{r},$$

and set $q = \tilde{q} + a_d b_k^{-1} x^{d-k}$ and $r = \tilde{r}$. In each case, the reader can check that q and r satisfy the conditions in the division algorithm, completing the induction step and the proof of existence.

It remains to prove uniqueness. If q, r and $\tilde{q}, \tilde{r}$ both satisfy the conclusion of the division algorithm, then

$$f = qg + r = \tilde{q}g + \tilde{r} \Longrightarrow g(\tilde{q} - q) = (r - \tilde{r}).$$

By assumption, either $r - \tilde{r} = 0$ or $\deg(r - \tilde{r}) < \deg(g)$. In the latter case, additivity of degree implies that $\deg(\tilde{q} - q) < 0$, a contradiction. Therefore, it must be the case that $r = \tilde{r}$. Since $g \neq 0$ and $K[x]$ is an integral domain, it then follows that $q = \tilde{q}$. □

With the division algorithm in hand, we now prove a slew of important consequences. The first two applications concern zeroes of single-variable polynomials.

0.48 **COROLLARY** *Factor theorem*

If $f(x) \in K[x]$ and $a \in K$, then $f(a) = 0$ if and only if $(x - a) \mid f(x)$.

PROOF Using the division algorithm, write

$$f(x) = (x - a)q + r$$

for some $q, r \in K[x]$ such that $r = 0$ or $\deg(r) < \deg(x - a) = 1$. In either case, the remainder must be a constant: $r \in K$. Evaluating at $x = a$, we see that

$$f(a) = (a - a)q(a) + r = r.$$

The result then follows from the observation that $(x - a) \mid f(x)$ if and only if the remainder of $f(x)$ divided by $x - a$, which we just proved is $f(a)$, is zero. □

0.49 EXAMPLE $x^n - 1 \in K[x]$

Consider the polynomial $f(x) = x^n - 1 \in K[x]$. Since $f(1) = 1^n - 1 = 0$, Corollary 0.48 implies that $(x-1) \mid (x^n - 1)$. Indeed, by multiplying out the right-hand side, one checks that

$$x^n - 1 = (x-1)(x^{n-1} + x^{n-2} + \cdots + x + 1).$$

0.50 COROLLARY *Finite zeroes theorem*

If $f(x) \in K[x]$ is a nonzero polynomial of degree d, then there are at most d values $a \in K$ such that $f(a) = 0$.

PROOF We proceed by induction on d.

(Base case) Suppose $d = 0$. Then $f = b$ for some nonzero constant $b \in K$. Thus, $f(a) = b \neq 0$ for all $a \in K$, so f does not have any zeroes.

(Induction step) Let f be a polynomial of degree $d > 0$. If f does not have any zeroes, then we are done. If f has at least one zero $a \in K$, then Corollary 0.48 implies that

$$f = (x-a)g$$

for some $g \in K[x]$. By additivity of degree, $\deg(g) = d - 1$, so the induction hypothesis implies that g has at most $d-1$ zeroes. Since every zero of f other than a must also be a zero of g, we conclude that f has at most d zeroes. □

0.51 EXAMPLE $x^n - 1 \in \mathbb{C}[x]$

It may be helpful for this example to recall Euler's formula:

$$e^{i\theta} = \cos(\theta) + i\sin(\theta).$$

By Corollary 0.50, the polynomial $x^n - 1 \in \mathbb{C}[x]$ has at most n zeroes. For $j = 1, \ldots, n$, consider the complex number

$$a_j = e^{\frac{2\pi i}{n} j}.$$

Since

$$(a_j)^n = e^{2\pi i j} = 1^j = 1,$$

we see that a_j is a zero of $x^n - 1$ for every j. Since $\{a_1, \ldots, a_n\}$ is a set of n distinct zeroes, we conclude that these must be all of the zeroes of $x^n - 1$.

The following result is another important consequence of the division algorithm.

0.52 COROLLARY *$K[x]$ is a PID*

The single-variable polynomial ring $K[x]$ is a principal ideal domain.

PROOF Let $I \subseteq K[x]$ be an ideal and define the set

$$S = \{\deg(f) \mid f \in I \text{ and } f \neq 0\} \subseteq \mathbb{N}.$$

If $S = \emptyset$, then I is the zero ideal $\langle 0 \rangle$, thus principal. If $S \neq \emptyset$, then, by the well-ordering principle, S contains a least element, call it d. Let $f \in I$ be a nonzero element such that $\deg(f) = d$. To prove that I is principal, we show that $I = \langle f \rangle$.

Since $f \in I$, then we obtain the inclusion $\langle f \rangle \subseteq I$ for free. To prove the other inclusion, suppose $g \in I$. Applying the division algorithm, we have

$$g = qf + r$$

with $r = 0$ or $\deg(r) < \deg(f) = d$. Since $f, g \in I$, it follows that $r \in I$. If $r \neq 0$, then $\deg(r) \in S$ and $\deg(r) < d$, contradicting that $d = \min(S)$. Thus, $r = 0$, from which we conclude that $g \in \langle f \rangle$. □

We close this section with the important result that $K[x]$ is a UFD, which is now an immediate consequence of Proposition 0.45.

0.53 **COROLLARY** *$K[x]$ is a UFD*

The single-variable polynomial ring $K[x]$ is a unique factorization domain.

Exercises for Section 0.5

0.5.1 Compute the quotient and remainder for the following pairs in $\mathbb{Q}[x]$.

(a) $f = 2x^3 + 7x^2 + 2x + 9$, $g = 2x + 3$

(b) $f = 3x^3 - 2x^2 + 5$, $g = x^2 - 1$

(c) $f = x^3 + 3x^2 - 4x - 12$, $g = x^2 + x - 6$

0.5.2 Consider $f = x^3 - x^2 + x - 1 \in \mathbb{Q}[x]$.

(a) Use Corollary 0.48 to show that $x - 1$ divides f.

(b) Compute the quotient of f divided by $x - 1$.

0.5.3 Give an example to show that the polynomial division algorithm fails in $\mathbb{Z}[x]$.

0.5.4 Prove that Corollary 0.48 holds in $R[x]$ for any ring R.

0.5.5 Prove that Corollary 0.50 holds in $R[x]$ if and only if R is an integral domain.

0.5.6 Give an example of a ring R and a nonzero polynomial $f \in R[x]$ with infinitely many zeroes.

0.5.7 Compute the unique irreducible factorization of $x^n - 1 \in \mathbb{C}[x]$.

Section 0.6 Unique factorization in polynomial rings

In this section, we conclude the proof that $K[x_1, \ldots, x_n]$ is a unique factorization domain. By Propositions 0.17 and 0.22, all that remains to be proved is the following analogue of Euclid's Lemma.

0.54 **PROPOSITION** *Euclid's Lemma for polynomials*

Every irreducible polynomial in $K[x_1, \ldots, x_n]$ is prime.

The proof of Proposition 0.54 involves an induction argument on the number of variables, starting with the base case of $K[x]$. Because the proof is rather involved, we start with a brief overview of the main ideas, then we develop each of those ideas in turn, finally merging them into a formal proof at the end of this section.

To motivate the ideas in the section, recall that we can view the polynomial ring $K[x_1, \ldots, x_n]$ as a polynomial ring in $n-1$ variables:

$$K[x_1, \ldots, x_n] = R[x_1, \ldots, x_{n-1}],$$

where $R = K[x_n]$. One of the important new ideas that we introduce in this section is that of the *fraction field*, which associates a field $\mathrm{Frac}(R)$ to any integral domain R, along with a canonical inclusion $R \subseteq \mathrm{Frac}(R)$. In particular, if $K' = \mathrm{Frac}(K[x_n])$, we obtain an inclusion

$$K[x_1, \ldots, x_n] \subseteq K'[x_1, \ldots, x_{n-1}].$$

Using this inclusion, we can begin to see an induction argument coming together. In particular, if our induction hypothesis is that irreducible polynomials in $n-1$ variables over any field are prime, then we can proceed to prove that irreducible polynomials in n variables are prime using the following two steps (Lemmas 0.59 and 0.60, respectively):

1. Prove that every irreducible polynomial in $K[x_1, \ldots, x_n]$ remains irreducible in $K'[x_1, \ldots, x_{n-1}]$ (hence prime, by the induction hypothesis).
2. Prove that every irreducible polynomial in $K[x_1, \ldots, x_n]$ that happens to be prime in $K'[x_1, \ldots, x_{n-1}]$ is also prime in $K[x_1, \ldots, x_n]$.

Now that we have outlined the road ahead, we begin in earnest by developing the notions of fraction fields, starting with the definition of fractions.

0.55 **PROPOSITION/DEFINITION** *Fractions*

Let R be an integral domain. A *fraction* of elements in R is an expression of the form a/b where $a, b \in R$ and $b \neq 0$. Equality of fractions is defined by

$$\frac{a}{b} = \frac{c}{d} \iff ad = bc \in R.$$

Equality of fractions is an equivalence relation, and the set of equivalence classes is denoted $\mathrm{Frac}(R)$.

That equality of fractions is, in fact, an equivalence relation is verified in Exercise 0.6.1. We now consider a few familiar examples of the fraction construction.

0.56 **EXAMPLE** $\mathbb{Q} = \text{Frac}(\mathbb{Z})$

While the definition of equality of fractions might be confusing at first glance, it is modeled on the familiar way that rational numbers are constructed from the integers. In particular, as we learn in grade school, to check an equality of rational numbers, such as

$$\frac{3}{4} = \frac{6}{8},$$

we cross multiply: $3 \cdot 8 = 4 \cdot 6 \in \mathbb{Z}$.

0.57 **EXAMPLE** Rational functions

As the polynomial rings $K[x_1, \ldots, x_n]$ play such a central role in our story, fractions of polynomials have a special name: they are called *rational functions*. We denote the set of rational functions using the following notation:

$$K(x_1, \ldots, x_n) = \text{Frac}(K[x_1, \ldots, x_n]).$$

Consider, for example, the following two elements of $K(x, y)$:

$$\frac{2x^2 + x - 2xy - y}{x^2 - y^2} \quad \text{and} \quad \frac{2x + 1}{x + y}.$$

In fact, these rational functions are equal because, as the reader can verify,

$$(2x^2 + x - 2xy - y)(x + y) = (2x + 1)(x^2 - y^2).$$

Another way to view this equality is by canceling like factors in the numerator and denominator:

$$\frac{2x^2 + x - 2xy - y}{x^2 - y^2} = \frac{(x - y)(2x + 1)}{(x - y)(x + y)} = \frac{2x + 1}{x + y}.$$

The word function *here is standard but misleading; a rational function should not necessarily be thought of as a* function*, per se, with a domain, a range, and a rule. Rather, it is simply a formal quotient of polynomials.*

In fact, the set of all fractions is more than just a set, it forms a field under the familiar operations of addition and multiplication of quotients. Since the same fraction can be represented in multiple ways, it needs to be verified that the operations are well-defined, meaning that they are independent of the choice of representatives. In addition, one should verify that, with these operations, the set of fractions satisfies the field axioms. We formalize the addition and multiplication below, and leave the necessary verifications to Exercise 0.6.2.

0.58 PROPOSITION/DEFINITION *Fraction field*

Let R be an integral domain. The operations of addition and multiplication defined by

$$\frac{a}{b}+\frac{c}{d}=\frac{ad+bc}{bd} \quad \text{and} \quad \frac{a}{b}\cdot\frac{c}{d}=\frac{ac}{bd}.$$

are well-defined on equivalence classes of fractions and endow $\mathrm{Frac}(R)$ with the structure of a field, called the *fraction field* of R.

One of the most important properties of the fraction field is that it canonically contains R as a subring (Exercise 0.6.3):

$$R=\{a/1 \mid a\in R\}\subseteq \mathrm{Frac}(R).$$

In particular, $K[x_n]\subseteq K(x_n)$, and thus, $K[x_1,\dots,x_n]\subseteq K(x_n)[x_1,\dots,x_{n-1}]$. Moreover, given any element $f\in K(x_n)[x_1,\dots,x_{n-1}]$, we can always find a polynomial $r\in K[x_n]$ such that $rf\in K[x_1,\dots,x_n]$ (Exercise 0.6.5). Multiplying f by such an r is called *clearing the denominators in* f, and is used frequently in the proofs of this section.

We are now ready to prove the two lemmas required for Proposition 0.54.

0.59 LEMMA

If f is irreducible in $K[x_1,\dots,x_n]$ and $f\notin K[x_n]$, then f is irreducible as an element of $K(x_n)[x_1,\dots,x_{n-1}]$.

PROOF We prove the contrapositive. Assume that $f\in K[x_1,\dots,x_n]$ is reducible in $K(x_n)[x_1,\dots,x_{n-1}]$; our goal is to prove that f is reducible in $K[x_1,\dots,x_n]$. Since f is reducible in $K(x_n)[x_1,\dots,x_{n-1}]$, we can write $f=gh$ where neither g nor h is an element of $K(x_n)$. By clearing the denominators in both g and h, we can write

$$rf=g_0h_0,$$

where $r\in K[x_n]$ and $g_0,h_0\in K[x_1,\dots,x_n]\setminus K[x_n]$. Since the single-variable polynomial ring $K[x_n]$ is a UFD (Corollary 0.53), write $r=p_1\cdots p_\ell$ where each p_i is irreducible, and thus prime. We have

$$p_1\cdots p_\ell f=g_0h_0.$$

Since p_1 is prime in $K[x_n]$, it follows from Exercise 0.6.6 that p_n is also prime in $K[x_n][x_1,\dots,x_{n-1}]=K[x_1,\dots,x_n]$. Therefore, by definition of primeness, $p_1\mid g_0$ or $p_1\mid h_0$. Without loss of generality, suppose $p_1\mid g_0$ and write $g_0=p_1g_1$ and $h_1=h_0$ so that

$$p_2\cdots p_\ell f=g_1h_1.$$

Repeating the above procedure with $p_2,\dots,p_\ell$, we conclude that $f=g_\ell h_\ell$. Since neither g_0 nor h_0 were elements of $K[x_n]$, it follows that neither g_ℓ nor h_ℓ are elements of K. Thus, f is reducible in $K[x_1,\dots,x_n]$. □

0.60 LEMMA

If $f \in K[x_1, \ldots, x_n]$ is irreducible as an element of $K[x_1, \ldots, x_n]$ and prime as an element of $K(x_n)[x_1, \ldots, x_{n-1}]$, then f is prime in $K[x_1, \ldots, x_n]$.

PROOF Suppose that $f \in K[x_1, \ldots, x_n]$ is irreducible, and that f is prime in $K(x_n)[x_1, \ldots, x_{n-1}]$. To prove that f is prime in $K[x_1, \ldots, x_n]$, suppose $f \mid gh$ in $K[x_1, \ldots, x_n]$. Since f is prime in $K(x_n)[x_1, \ldots, x_{n-1}]$, we know that $f \mid g$ or $f \mid h$ in $K(x_n)[x_1, \ldots, x_{n-1}]$. Without loss of generality, assume $f \mid g$ and write

$$af = g$$

for some $a \in K(x_n)[x_1, \ldots, x_{n-1}]$. By clearing denominators in a, write

$$a_0 f = rg \in K[x_1, \ldots, x_n]$$

for some $a_0 \in K[x_1, \ldots, x_n]$ and $r \in K[x_n]$. As in the proof of Lemma 0.59, let $r = p_1 \cdots p_\ell$ be a prime factorization of r, from which it follows that p_1 must divide a_0 or f. Since f is irreducible in $K[x_1, \ldots, x_n]$, the only way p_1 could divide f is if they differed by a constant, which would imply that $f \in K[x_n]$. This would mean that f is a unit in $K(x_n)[x_1, \ldots, x_{n-1}]$, which contradicts the assumption that f is prime in $K(x_n)[x_1, \ldots, x_{n-1}]$. Thus, p_1 must divide a_0. Write $a_0 = a_1 p_1$ so that

$$a_1 f = p_2 \cdots p_\ell g.$$

Repeating this procedure for $p_2, \ldots, p_\ell$, we conclude that

$$a_\ell f = g \in K[x_1, \ldots, x_n].$$

Therefore, $f \mid g$ in $K[x_1, \ldots, x_n]$ and it follows that f is prime. □

Combining the previous two lemmas, we now prove Proposition 0.54.

PROOF OF PROPOSITION 0.54 Proceeding by induction on n, the base case $n = 1$ is Corollary 0.53. To prove the induction step, suppose that, for some $n \geq 2$ and for any field K, every irreducible polynomial in $K[x_1, \ldots, x_{n-1}]$ is prime. Let K be a field and let $f \in K[x_1, \ldots, x_n]$ be irreducible; we must show that f is prime.

Since f is not a unit, it has positive degree in at least one variable; suppose without loss of generality that it has positive degree in a variable other than x_n. Using that f is irreducible and not an element of $K[x_n]$, Lemma 0.59 implies that f is irreducible as an element of $K(x_n)[x_1, \ldots, x_{n-1}]$. Because $K(x_n)$ is a field, the induction hypothesis implies that f is prime in $K(x_n)[x_1, \ldots, x_{n-1}]$. Therefore, by applying Lemma 0.60, we conclude that f is prime in $K[x_1, \ldots, x_n]$. □

As an immediate consequence of Propositions 0.17, 0.22, and 0.54, we now conclude that $K[x_1, \ldots, x_n]$ is a unique factorization domain. For ease of reference, we close this section with the precise statement of this result.

It can be proved, more generally, that $R[x_1, \ldots, x_n]$ is a UFD whenever R is a UFD. This level of generality is not necessary for our purposes.

0.61 **THEOREM** *$K[x_1,\dots,x_n]$ is a UFD*

If $f \in K[x_1,\dots,x_n]$ is nonconstant, then there exist irreducible polynomials $p_1,\dots,p_\ell \in K[x_1,\dots,x_n]$ such that

$$f = p_1 \cdots p_\ell.$$

If $f = q_1 \cdots q_m$ is another irreducible factorization, then $\ell = m$ and, after possibly reordering terms, q_i is a constant multiple of p_i for every i.

Exercises for Section 0.6

0.6.1 Prove that equality of fractions is an equivalence relation.

0.6.2 Let R be an integral domain and $a/b, c/d, r/s \in \mathrm{Frac}(R)$ with $a/b = c/d$.

(a) Prove that

$$\frac{a}{b} + \frac{r}{s} = \frac{c}{d} + \frac{r}{s} \quad \text{and} \quad \frac{a}{b} \cdot \frac{r}{s} = \frac{a}{b} \cdot \frac{r}{s}.$$

(b) Prove that $\mathrm{Frac}(R)$ satisfies the field axioms.

0.6.3 Let R be an integral domain.

(a) Prove that the function

$$\begin{aligned} \varphi : R &\to \mathrm{Frac}(R) \\ a &\mapsto a/1 \end{aligned}$$

is an injective ring homomorphism.

(b) Let K be a field with $R \subseteq K$. Prove that $\mathrm{Frac}(R) \subseteq K$.

0.6.4 Suppose that R is not an integral domain. Explain what might go wrong if we try to construct the fraction field of R.

0.6.5 Let $f \in K(x_n)[x_1,\dots,x_{n-1}]$. Prove that there is a polynomial $r \in K[x_n]$ such that $rf \in K[x_1,\dots,x_n]$.

0.6.6 Let $a \in R$ and consider the surjective homomorphism

$$\pi : R[x_1,\dots,x_n] \to (R/\langle a\rangle)[x_1,\dots,x_n].$$

(a) Prove that $\ker(\pi) = \langle a\rangle \subseteq R[x_1,\dots,x_n]$ and conclude that

$$\frac{R[x_1,\dots,x_n]}{\langle a\rangle} \cong (R/\langle a\rangle)[x_1,\dots,x_n].$$

(b) Prove that a is prime in R if and only if a is prime in $R[x_1,\dots,x_n]$.

Section 0.7 Irreducibility criteria

In order to have a large bank of concrete examples in algebraic geometry, it is useful to have methods at our disposal for studying specific polynomials. For example, if we have a particular polynomial in mind, such as

$$x^2 + y^2 + z^2 - 1 \in \mathbb{R}[x, y, z],$$

it might be helpful to be able to determine quickly whether or not this polynomial is irreducible. In this final section of the chapter, we discuss two criteria for determining whether a given polynomial is irreducible. The first result follows quickly from our prior developments and we leave its proof to Exercise 0.7.1.

0.62 **PROPOSITION** *Characterization of irreducible polynomials*

Let $f \in K[x_1, \ldots, x_n]$. The following are equivalent.

1. f is irreducible;
2. f is prime;
3. $\langle f \rangle$ is a prime ideal;
4. $K[x_1, \ldots, x_n]/\langle f \rangle$ is an integral domain.

0.63 **EXAMPLE** $y - x^2$ is irreducible in $K[x, y]$, revisited

We have already proved directly that $y - x^2$ is irreducible. In light of Proposition 0.62, this can be seen from the fact that $K[x]$ is an integral domain and

$$\frac{K[x, y]}{\langle y - x^2 \rangle} \cong K[x].$$

It is not always helpful to apply Proposition 0.62 in practice, because the problem of showing that an ideal is prime or that a quotient is an integral domain is typically just as difficult as showing directly that a polynomial is irreducible—the proposition translates the problem but does not necessarily simplify it.

The next test, called Eisenstein's Criterion, while a bit more complicated to state, is much more useful in practice, as will be illustrated in the subsequent examples.

0.64 **PROPOSITION** *Eisenstein's Criterion*

Let R be an integral domain and $f = a_n x^n + \cdots + a_1 x + a_0 \in R[x]$ a polynomial that satisfies the following conditions.

1. There does not exist a nonunit $b \in R$ such that $b \mid f$.
2. There exists a prime element $p \in R$ such that
   - $p \mid a_i$ for $i < n$,
   - $p^2 \nmid a_0$.

Then f is irreducible.

Unlike Proposition 0.62, Eisenstein's Criterion is not an if-and-only-if statement. In particular, Eisenstein's Criterion can never be used to determine whether a single-variable polynomial $f \in K[x]$ is irreducible, simply because a field K does not contain any prime elements. Before proving Eisenstein's Criterion, we provide a few example applications to demonstrate how to use it in the context of multi-variable polynomials over fields. We include a number of further examples in the exercises.

0.65 **EXAMPLE** $f = x^2 + y^2 - 1$ is irreducible in $\mathbb{R}[x, y]$

If we view f as an element of $R[y]$ where $R = \mathbb{R}[x]$, then

$$f = a_2 y^2 + a_1 y + a_0$$

where $a_2 = 1$, $a_1 = 0$, and $a_0 = x^2 - 1 = (x-1)(x+1)$. Notice that these coefficients do not have any common nonconstant divisors in $\mathbb{R}[x]$, which verifies the first condition in Eisenstein's Criterion. Since

$$\frac{\mathbb{R}[x]}{\langle x-1 \rangle} \cong \mathbb{R},$$

is an integral domain, $p = x - 1$ is prime in $\mathbb{R}[x]$ and satisfies the second condition of Eisenstein's Criterion. Therefore, we conclude that f is an irreducible polynomial.

Working over $\mathbb{R}$ is not essential for these two examples; the same argument works for any field for which $1 \neq -1$ (when $\mathrm{char}(K) \neq 2$).

0.66 **EXAMPLE** $f = x^2 + y^2 + z^2 - 1$ is irreducible in $\mathbb{R}[x, y, z]$

As in the previous example, write

$$f = a_2 z^2 + a_1 z + a_0$$

where $a_2 = 1$, $a_1 = 0$ and $a_0 = x^2 + y^2 - 1$. These coefficients do not have any common nonconstant divisors in $\mathbb{R}[x, y]$, verifying the first condition in Eisentein's Criterion. Set $p = x^2 + y^2 - 1$, which is irreducible by the previous example, and thus prime because $\mathbb{R}[x, y]$ is a UFD. The polynomial p satisfies the second condition of Eisenstein's Criterion, from which we conclude that f is irreducible.

0.67 **EXAMPLE** $f = x^2y^2 + yz^2 + x^3z^2 \in K[x, y, z]$ is irreducible

Write

$$f = a_3 x^3 + a_2 x^2 + a_1 x + a_0 \in R[x]$$

where $R = K[y, z]$ and $a_3 = z^2$, $a_2 = y^2$, $a_1 = 0$, and $a_0 = yz^2$. Notice that these coefficients do not have any common nonconstant divisors in $K[y, z]$, which verifies the first condition in Eisenstein's Criterion. Let $p = y \in R$. Since

$$\frac{R}{\langle p \rangle} = \frac{K[y, z]}{\langle y \rangle} \cong K[z]$$

is an integral domain, we see that $p = y$ is prime in R. Because p satisfies the second condition of Eisenstein's Criterion, we conclude that f is irreducible.

PROOF OF PROPOSITION 0.64 Let f and p be as in the statement of the proposition, and suppose $f = gh$ for some $g, h \in R[x]$. We must show that g or h is a unit in R. By the first condition, f is not divisible by any nonunits in R, so it suffices to prove that either g or h is an element of R.

Toward a contradiction, suppose $g = b_\ell x^\ell + \cdots + b_0$ and $h = c_k x^k + \cdots + c_0$ both have positive degree. Consider the ring homomorphism

$$\begin{aligned} \varphi : R[x] &\to (R/\langle p \rangle)[x] \\ \sum_{n \geq 0} a_n x^n &\mapsto \sum_{n \geq 0} [a_n] x^n. \end{aligned}$$

By our assumptions on f, it follows that

$$\varphi(f) = [a_n] x^n = \varphi(g)\varphi(h),$$

with $[a_n] \neq 0$. Since $R/\langle p \rangle$ is an integral domain, $\varphi(g)$ and $\varphi(h)$ must each consist of a single nonzero term (Exercise 0.7.2), and by additivity of degree, it follows that $\varphi(g) = [b_\ell] x^\ell$ and $\varphi(h) = [c_k] x^k$. In particular, this implies that $p \mid b_0$ and $p \mid c_0$, so that $p^2 \mid b_0 c_0 = a_0$, which contradicts the assumptions on p. □

Exercises for Section 0.7

0.7.1 Prove Proposition 0.62 by citing relevant results from previous sections.

0.7.2 Suppose R is an integral domain and $g, h \in R[x]$ such that $gh = ax^n$ has a single nonzero term. Prove that each of g and h have a single nonzero term.

0.7.3 Prove that $wx - yz \in K[w, x, y, z]$ is irreducible.

0.7.4 Prove that $xyz + x^2z^2 + yz^3 + x \in K[x, y, z]$ is irreducible.

0.7.5 Assume $\text{char}(K) \neq 2$ and $n \geq 2$. Prove that $x_1^2 + \cdots + x_n^2 - 1$ is irreducible in $K[x_1, \ldots, x_n]$.

0.7.6 Assume $n \geq 3$ and $m \geq 1$. Prove that $x_1^m + \cdots + x_n^m \in \mathbb{C}[x_1, \ldots, x_n]$ is irreducible.

Part I

Affine varieties

Chapter 1

Varieties and Ideals

LEARNING OBJECTIVES FOR CHAPTER 1

- Acquaint ourselves with affine space $\mathbb{A}^n_K$.
- Describe how to use the $\mathcal{V}$- and $\mathcal{I}$-operators to move between subsets of $K[x_1, \ldots, x_n]$ and subsets of $\mathbb{A}^n_K$.
- Become familiar with the notions of affine varieties in $\mathbb{A}^n_K$ and radical ideals in $K[x_1, \ldots, x_n]$.
- State the Nullstellensatz and use it to describe the bijection between affine varieties and radical ideals.

Algebraic geometry is, at its heart, a dictionary for translating between different languages: the language of algebra and the language of geometry. As in any dual-language dictionary, this involves translation in both directions. Given an algebraic object, such as a polynomial, we produce a geometric object by determining the vanishing set of the polynomial. Conversely, given a geometric object, we produce an algebraic object by determining the set of all polynomials that vanish on the given geometric set.

In this chapter, we begin our study of algebraic geometry in earnest by making these two operations precise by way of the $\mathcal{V}$- and $\mathcal{I}$-operators. Crucially, we find that these operators are not surjective. Not every geometric set is the vanishing set of some collection of polynomials, for example; those geometric sets that are obtained in this way are called *affine varieties*. Conversely, not every set of polynomials is obtainable by starting from a geometric set and calculating the polynomials that vanish on it; the study of algebraic sets obtained in this way will lead us to develop the algebraic notion of a *radical ideal*.

The chapter culminates with the statement of a result that might properly be termed the "Fundamental Theorem of Algebraic Geometry," though it instead goes by the German name *Nullstellensatz*. Under one key hypothesis—that the ground field is *algebraically closed*—the Nullstellensatz asserts that when one restricts attention to affine varieties on the geometric side and to radical ideals on the algebraic side, the $\mathcal{V}$- and $\mathcal{I}$-operators provide a true dictionary—a bijection, to put it mathematically—between algebra and geometry.

Section 1.1 The $\mathcal{V}$-operator

In order to study the vanishing sets of polynomials, we must begin by specifying where those polynomials and their solutions live. Choose a field K, referred to as the *ground field*, and consider the polynomial ring $K[x_1, \ldots, x_n]$. Elements of this ring, in addition to being abstract polynomials, can also be used to define functions that take elements of K^n as input and output elements of K. For example, the polynomial

$$f = 2xy + 4z^2 \in \mathbb{R}[x, y, z]$$

defines a function from $\mathbb{R}^3$ to $\mathbb{R}$. If one inputs the element $(1, -1, 3)$, then the output of f is the single real number

$$f(1, -1, 3) = 2 \cdot 1 \cdot (-1) + 4 \cdot 3^2 = 34.$$

The domain of a polynomial function is referred to as *affine space*.

1.1 **DEFINITION** *Affine space*

The *n-dimensional affine space over K*, denoted $\mathbb{A}^n_K$, is the set of n-tuples of elements of K:

$$\mathbb{A}^n_K = \{(a_1, \ldots, a_n) \mid a_i \in K\}.$$

As a set, $\mathbb{A}^n_K$ is the same as the vector space K^n. So why give it a new name and a new notation? When we write K^n, we are viewing this set as an algebraic object with addition and scalar multiplication operations—that is, as a vector space. When we write $\mathbb{A}^n_K$, on the other hand, we forget the algebraic structure on this set: we view its elements not as vectors that can be added to one another, but rather as inputs to polynomial functions. In particular, the element $(0, \ldots, 0)$ is very special in the vector space K^n, since it is the additive identity, but it is essentially the same as any other element in the affine space $\mathbb{A}^n_K$.

While elements of K^n are typically referred to as vectors, elements of $\mathbb{A}^n_K$ are called "points" to highlight their geometric significance.

Often, when the ground field K is understood from context, we simply write $\mathbb{A}^n$ instead of $\mathbb{A}^n_K$.

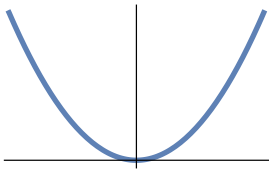
To distinguish between polynomials and their evaluations, we use letters at the end of the alphabet (x, y, and z) to denote variables, and letters at the beginning of the alphabet (a, b, and c) to denote elements of K. So, for example, $f = x^2y$ denotes an element of the ring $K[x, y]$, whereas $f(a, b) = a^2b$ denotes the element of K obtained by evaluating f at the point $(a, b) \in \mathbb{A}^2$. If $f \in K[x_1, \ldots, x_n]$ satisfies

$$f(a_1, \ldots, a_n) = 0$$

for some point $(a_1, \ldots, a_n) \in \mathbb{A}^n$, we say that f *vanishes* at $(a_1, \ldots, a_n)$.

Given a polynomial $f \in K[x_1, \ldots, x_n]$, the set of all points at which f vanishes is a subset of $\mathbb{A}^n$. For example, if

$$f = y - x^2 \in \mathbb{R}[x, y],$$



then f vanishes at the point $(0,0)$, as well as at the point $(1,1)$, the point $(2,4)$, the point $(-1,1)$, and so on. The set of all points in $\mathbb{A}^2_{\mathbb{R}}$ at which $f = y - x^2$ vanishes forms the familiar parabola. This is our first taste of algebraic geometry: an algebraic object (the element $y - x^2$ of the ring $\mathbb{R}[x,y]$) led us to a geometric object (the parabola).

More generally, one can study the set of points in $\mathbb{A}^n$ at which every element of a (possibly infinite) set of polynomials vanishes.

1.2 DEFINITION *Vanishing set*

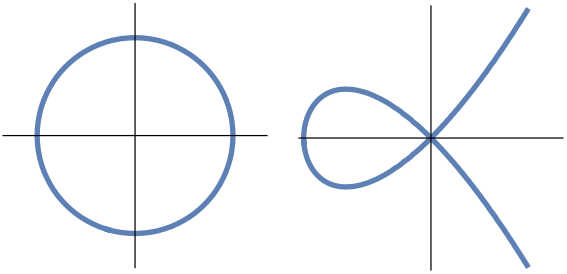
Let $\mathcal{S} \subseteq K[x_1,\ldots,x_n]$ be a set of polynomials. The *vanishing set of* $\mathcal{S}$ is

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

It is common to say that points of $\mathcal{V}(\mathcal{S})$ are *solutions* of the *polynomials* in $\mathcal{S}$. When $\mathcal{S} = \{f_1,\ldots,f_r\}$ is finite, we write $\mathcal{V}(f_1,\ldots,f_r)$ instead of $\mathcal{V}(\{f_1,\ldots,f_r\})$.

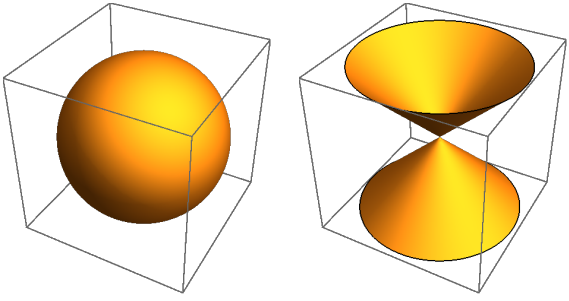
1.3 EXAMPLE Curves in $\mathbb{A}^2_{\mathbb{R}}$

Consider $\mathcal{S} = \{y - x^2\} \subseteq K[x,y]$. Then $\mathcal{V}(\mathcal{S}) = \{(a,a^2) \mid a \in K\} \subseteq \mathbb{A}^2$. When $K = \mathbb{R}$, this is the parabola above. Similarly, $\mathcal{V}(x^2+y^2-1) \subseteq \mathbb{A}^2_{\mathbb{R}}$ and $\mathcal{V}(y^2-x^3-x^2) \subseteq \mathbb{A}^2_{\mathbb{R}}$ are the planar curves depicted below.



1.4 EXAMPLE Surfaces in $\mathbb{A}^3_{\mathbb{R}}$

The vanishing sets $\mathcal{V}(x^2+y^2+z^2-1) \subseteq \mathbb{A}^3_{\mathbb{R}}$ and $\mathcal{V}(x^2+y^2-z^2) \subseteq \mathbb{A}^3_{\mathbb{R}}$ are the unit sphere and the cone, respectively, pictured in the following images.



1.5 EXAMPLE The coordinate axes

If $\mathcal{S} = \{xy\} \subseteq K[x,y]$, then $\mathcal{V}(\mathcal{S}) = \{(a,b) \in \mathbb{A}^2 \mid ab = 0\} \subseteq \mathbb{A}^2$. Since K is a field, $ab = 0$ if and only if either $a = 0$ or $b = 0$ (or both), so $\mathcal{V}(\mathcal{S})$ is the union of the points where $a = 0$ and those where $b = 0$. When $K = \mathbb{R}$, this is the union of the x-axis and the y-axis in the real plane.

1.6 EXAMPLE Single points

Let $\mathcal{S} = \{x,y\} \subseteq K[x,y]$. Then

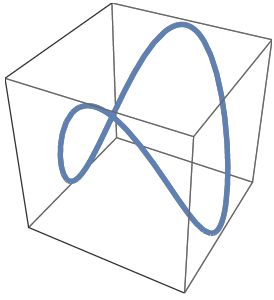
$$\mathcal{V}(\mathcal{S}) = \{(a,b) \in \mathbb{A}^2 \mid a = 0 \text{ and } b = 0\} = \{(0,0)\} \subseteq \mathbb{A}^2.$$

That is, $\mathcal{V}(\mathcal{S})$ consists of a single point: the origin.

Similarly, if $\mathcal{S} = \{x - i, y - (1+i), z - 5\} \subseteq \mathbb{C}[x,y,z]$, then

$$\mathcal{V}(\mathcal{S}) = \{(i, 1+i, 5)\} \subseteq \mathbb{A}^3_{\mathbb{C}},$$

which, again, is a single point.



1.7 EXAMPLE A curve in $\mathbb{A}^3$

The vanishing set of

$$\mathcal{S} = \{x^2 + y^2 - 1,\ x^2 - y^2 - z\} \subseteq \mathbb{R}[x,y,z]$$

consists of all points $(a,b,c) \in \mathbb{A}^3_{\mathbb{R}}$ where

$$a^2 + b^2 - 1 = 0 = a^2 - b^2 - c.$$

The images above have ground field $\mathbb{R}$. Over $\mathbb{R}$, we have a geometric intuition; for example, we have an idea about what it means to be a "curve" or "surface." Algebraic geometry aims to make this intuition precise for general fields.

Given that we are now viewing polynomials as functions $\mathbb{A}^n \to K$, it is worth pointing out that there is a subtle but important difference between *polynomials* and *polynomial functions*. In particular, it is a somewhat unsettling fact that different polynomials in $K[x_1, \ldots, x_n]$ can give rise to the same function $\mathbb{A}^n \to K$.

For example, let $K = \mathbb{F}_2 = \{0,1\}$, the field with two elements; recall that addition and multiplication in this field are both carried out modulo 2. Let

$$f(x) = 1 + x + x^2 \quad \text{and} \quad g(x) = 1. \tag{1.8}$$

As elements of $\mathbb{F}_2[x]$, these polynomials are not equal, because they have different coefficients on the monomial x as well as on the monomial x^2. However, viewing them as functions $\mathbb{A}^1 \to \mathbb{F}_2$, we see that

$$f(0) = 1 + 0 + 0^2 = 1 = g(0) \quad \text{and} \quad f(1) = 1 + 1 + 1^2 = 1 = g(1).$$

Thus, since f and g give the same output for every input in their domain, they are equal as functions, even though they are different as polynomials.

To describe the difference between polynomials and their corresponding functions more generally, let $K[\mathbb{A}^n]$ denote the set of polynomial functions $\mathbb{A}^n \to K$. That is, an element of $K[\mathbb{A}^n]$ is a function of the form

$$(a_1, \ldots, a_n) \mapsto f(a_1, \ldots, a_n)$$

for some $f \in K[x_1, \ldots, x_n]$. Some time reflecting should convince the reader that $K[\mathbb{A}^n]$ is a ring under addition and multiplication of functions to K, and taking a polynomial to its corresponding function defines a surjective ring homomorphism $K[x_1, \ldots, x_n] \to K[\mathbb{A}^n]$. This homomorphism fails to be injective exactly when different polynomials give rise to the same function. For example,

$$\mathbb{F}_2[x] \to \mathbb{F}_2[\mathbb{A}^1]$$

is not an injection, because the different polynomials $f, g \in \mathbb{F}_2[x]$ defined in (1.8) give rise to the same polynomial function in $\mathbb{F}_2[\mathbb{A}^1]$.

The next result shows that the above phenomenon is unique to finite fields, so we need not worry about it in what follows as long as we assume that K is infinite.

1.9 **PROPOSITION** *Polynomials versus functions*

The ring homomorphism $K[x_1, \ldots, x_n] \to K[\mathbb{A}^n]$ that takes polynomials to their corresponding functions is an isomorphism if and only if K is infinite.

PROOF First, assume $K = \{a_1, \ldots, a_n\}$ is finite, and consider the nonzero polynomial

$$f = \prod_{i=1}^{n} (x_1 - a_i) \in K[x_1] \subseteq K[x_1, \ldots, x_n].$$

Plugging in any value of K for x_1 produces a factor of zero, so f defines the zero function $\mathbb{A}^n \to K$. This implies that $K[x_1, \ldots, x_n] \to K[\mathbb{A}^n]$ is not an injection. Therefore, if $K[x_1, \ldots, x_n] \to K[\mathbb{A}^n]$ is an isomorphism, then K must be infinite.

Conversely, suppose K is infinite. By definition, $K[x_1, \ldots, x_n] \to K[\mathbb{A}^n]$ is surjective. Thus, it remains to prove injectivity, or equivalently, that the kernel is the zero polynomial. We accomplish this by induction on n.

(Base case) Suppose $h \in K[x]$ defines the zero function $\mathbb{A}^1 \to K$. Since K is an infinite field, this implies that h has infinitely many zeros. By Corollary 0.50, nonzero polynomials have finitely many zeros, so h must be the zero polynomial.

(Induction step) Suppose $h \in K[x_1, \ldots, x_n]$ defines the zero function, and write

$$h = \sum_{i=0}^{m} h_i x_n^i$$

where $h_i \in K[x_1, \ldots, x_{n-1}]$. For each $(a_1, \ldots, a_{n-1}) \in \mathbb{A}^{n-1}$, the single-variable polynomial $h(a_1, \ldots, a_{n-1}, x_n) \in K[x_n]$ defines the zero function $\mathbb{A}^1 \to K$. Thus, by the argument in the base case, $h(a_1, \ldots, a_{n-1}, x_n)$ is the zero polynomial. In particular, this implies that $h_i(a_1, \ldots, a_{n-1}) = 0$ for all i. Since this is true for every $(a_1, \ldots, a_{n-1}) \in \mathbb{A}^{n-1}$, it follows that $h_i : \mathbb{A}^{n-1} \to K$ is the zero function for all i. By the induction hypothesis, $h_i \in K[x_1, \ldots, x_{n-1}]$ is the zero polynomial for every i, and it follows that h is the zero polynomial. □

Exercises for Section 1.1

1.1.1 Sketch the following vanishing sets:

(a) $\mathcal{V}(x^2-1) \subseteq \mathbb{A}^1_{\mathbb{R}}$;
(b) $\mathcal{V}(x^2-y^2) \subseteq \mathbb{A}^2_{\mathbb{R}}$;
(c) $\mathcal{V}(y-x^2, y-x) \subseteq \mathbb{A}^2_{\mathbb{R}}$;
(d) $\mathcal{V}(x^2+y^2+z^2-1, z) \subseteq \mathbb{A}^3_{\mathbb{R}}$;
(e) $\mathcal{V}(x^2+y^2-z^2, x) \subseteq \mathbb{A}^3_{\mathbb{R}}$;
(f) $\mathcal{V}(x^2+y^2-z^2, z) \subseteq \mathbb{A}^3_{\mathbb{R}}$;
(g) $\mathcal{V}(xy-y^2, x^2-xy-x+y) \subseteq \mathbb{A}^2_{\mathbb{R}}$;

1.1.2 Express the following sets as $\mathcal{V}(\mathcal{S})$ for some $\mathcal{S}$:

(a) $\{0, \pi, -1\} \subseteq \mathbb{A}^1_{\mathbb{R}}$;
(b) the x-axis in $\mathbb{A}^3_{\mathbb{R}}$;
(c) $\{(4,-1,3)\} \subseteq \mathbb{A}^3_{\mathbb{R}}$;
(d) $\{(-1,0),(1,0)\} \subseteq \mathbb{A}^2_{\mathbb{R}}$;
(e) $\{(a,a,a) \mid a \in \mathbb{R}\} \subseteq \mathbb{A}^3_{\mathbb{R}}$;
(f) $\{(\cos(a), \sin(a), \cos^2(a) - \sin^2(a)) \mid a \in [0, 2\pi)\} \subseteq \mathbb{A}^3_{\mathbb{R}}$.

1.1.3 Show that the origin $\{(0,0)\} \subseteq \mathbb{A}^2_K$ can be defined by a single equation if $K = \mathbb{R}$, but not if $K = \mathbb{C}$.

1.1.4 Let $f, g \in K[x_1, \ldots, x_n]$. Prove that

(a) $\mathcal{V}(fg) = \mathcal{V}(f) \cup \mathcal{V}(g)$;
(b) $\mathcal{V}(f, g) = \mathcal{V}(f) \cap \mathcal{V}(g)$.

1.1.5 Let

$$X = \mathcal{V}(x^2 - yz, xz - x) \subseteq \mathbb{A}^3.$$

(a) Prove that

$$X = \mathcal{V}(x, y) \cup \mathcal{V}(x, z) \cup \mathcal{V}(x^2 - y, z - 1).$$

(b) Use part (a) to sketch X in the case where $K = \mathbb{R}$.

1.1.6 Prove that every finite set in $\mathbb{A}^n_K$ can be expressed as the vanishing set of a collection of n polynomials. (**Hint**: Use induction on n.)

Section 1.2 Affine varieties

Not every subset of $\mathbb{A}^n$ is the vanishing set of some collection of polynomials. For example, let $X \subseteq \mathbb{A}^1_{\mathbb{R}}$ be the set of all nonzero real numbers:

$$X = \{a \in \mathbb{R} \mid a \neq 0\} \subsetneq \mathbb{A}^1_{\mathbb{R}}.$$

Let's show that X cannot be realized as the vanishing of a set of polynomials. Suppose, toward a contradiction, that $X = \mathcal{V}(\mathcal{S})$ for some set $\mathcal{S} \subseteq \mathbb{R}[x]$. Then $\mathcal{S}$ must contain at least one nonzero element, since if $\mathcal{S}$ were either $\emptyset$ or $\{0\}$, its vanishing set would be all of $\mathbb{A}^1_{\mathbb{R}}$. Let $f \in \mathcal{S}$ be any nonzero element. Since $X = \mathcal{V}(\mathcal{S})$, we have $f(a) = 0$ for all $a \in \mathbb{R} \setminus \{0\}$. This means that f is a nonzero single-variable polynomial with infinitely many zeros, contradicting Corollary 0.50.

In light of examples such as this, we give a name to those special subsets of affine space that can be defined as the vanishing of a set of polynomials.

1.10 **DEFINITION** *Affine variety*

A subset $X \subseteq \mathbb{A}^n$ is called an *affine variety* if $X = \mathcal{V}(\mathcal{S})$ for some subset $\mathcal{S} \subseteq K[x_1, \ldots, x_n]$.

We have already met a number of affine varieties in the examples presented in Section 1.1. Additionally, for each affine variety in Section 1.1, we specified a set of polynomials that realized it as a vanishing set. Two more examples that are perhaps more basic than any of the previous ones, but nevertheless crucial, are the empty set and the entirety of affine space.

1.11 **EXAMPLE** The empty set and the entirety of affine space

The constant polynomial $1 \in K[x_1, \ldots, x_n]$ does not vanish at any point, so

$$\mathcal{V}(1) = \emptyset \subseteq \mathbb{A}^n$$

The zero polynomial $0 \in K[x_1, \ldots, x_n]$ vanishes at every point, so

$$\mathcal{V}(0) = \mathbb{A}^n.$$

Thus, both $\emptyset$ and $\mathbb{A}^n$ are affine varieties.

It is quite easy, and not very enlightening, to come up with an endless list of examples of affine varieties by simply writing down sets of polynomials and considering their vanishing sets. A more subtle task is to understand what makes affine varieties special among subsets of affine space. In other words, what sorts of subsets of affine space are not affine varieties? We have already seen one example: the set $\mathbb{R} \setminus \{0\}$ is not an affine variety in $\mathbb{A}^1_{\mathbb{R}}$. In fact, the only affine varieties in $\mathbb{A}^1$ are $\emptyset$, finite collections of points, and all of $\mathbb{A}^1$ (Exercise 1.2.1).

In $\mathbb{A}^2$, on the other hand, an infinite proper subset can certainly be an affine variety; the parabola in $\mathbb{A}^2_{\mathbb{R}}$ is an example. What a proper subset of Euclidean space cannot have, however, if it is to be an affine variety, is a nonempty interior. The following example illustrates this phenomenon.

1.12 EXAMPLE A solid square is not an affine variety

Let X be the filled-in square in $\mathbb{A}^2_{\mathbb{R}}$ defined as

$$X = \{(a,b) \in \mathbb{A}^2_{\mathbb{R}} \mid -1 \leq a \leq 1 \text{ and } -1 \leq b \leq 1\}.$$

Suppose, toward a contradiction, that $X = \mathcal{V}(\mathcal{S})$ for some set $\mathcal{S} \subseteq \mathbb{R}[x,y]$. Let $f \in \mathcal{S}$. We will argue that f is the zero polynomial. Write

$$f = \sum_{i=0}^{m} f_i y^i$$

where $f_i \in \mathbb{R}[x]$ for each i. Since $f \in \mathcal{S}$, it vanishes at all values of X. In other words, for any value $a \in [-1,1]$, the single variable polynomial

$$f(a,y) = \sum_{i=0}^{m} f_i(a) y^i \in \mathbb{R}[y]$$

vanishes at all values $b \in [-1,1]$. Since a nonzero single-variable polynomial can only have finitely many zeros, it follows that $f(a,y)$ must be the zero polynomial, implying that $f_i(a) = 0$ for all i. In other words, we have argued that the single-variable polynomials $f_i \in \mathbb{R}[x]$ vanish at all values $a \in [-1,1]$. Again, using the fact that nonzero single-variable polynomials have finitely many zeros, this implies that f_i is the zero polynomial for each i, so f is the zero polynomial. This argument shows that the only polynomial that can be in $\mathcal{S}$ is the zero polynomial, from which it follows that $X = \mathcal{V}(\mathcal{S}) = \mathbb{A}^2_{\mathbb{R}}$, a contradiction.

Students with a background in topology are encouraged to prove, more generally, that if one gives $\mathbb{A}^n_{\mathbb{R}}$ the Euclidean topology, then the only affine variety that has a nonempty topological interior is the entirety of $\mathbb{A}^n_{\mathbb{R}}$ (Exercise 1.2.8).

1.13 EXAMPLE The graph of e^x in $\mathbb{A}^2_{\mathbb{R}}$ is not an affine variety

Let X be the graph of the exponential function on the real numbers:

$$X = \{(a,b) \in \mathbb{A}^2_{\mathbb{R}} \mid b = e^a\} \subseteq \mathbb{A}^2_{\mathbb{R}}. \tag{1.14}$$

A careful proof that X is not an affine variety is outlined in Exercise 1.2.7. It should seem reasonable that X is not an affine variety: it is defined by the vanishing of the expression $y - e^x$, which is not a polynomial in x and y. One must be careful with this sort of reasoning, however. The expression $\sin^2(x) + \cos^2(x) + y$ is not a polynomial, either, but the set

$$\{(a,b) \in \mathbb{A}^2_{\mathbb{R}} \mid \sin^2(a) + \cos^2(a) + b = 0\} \subseteq \mathbb{A}^2_{\mathbb{R}}$$

is the same as the set

$$\{(a,b) \in \mathbb{A}^2_{\mathbb{R}} \mid 1 + b = 0\} = \mathcal{V}(1+y) \subseteq \mathbb{A}^2_{\mathbb{R}},$$

so it is an affine variety.

It is certainly possible for the same affine variety to arise as $\mathcal{V}(\mathcal{S})$ for different sets $\mathcal{S}$. For example,

$$\mathcal{V}(x,y) = \mathcal{V}(x+y, x-y) = \{(0,0)\} \subseteq \mathbb{A}^2,$$

as the reader can readily verify. In fact, the set $\mathcal{S}$ can be replaced by the entire ideal $\langle S\rangle \subseteq K[s_0,\ldots,x_n]$ that is generated by $\mathcal{S}$ without affecting its vanishing set.

1.15 **PROPOSITION** *Affine varieties are defined by ideals*

If $\mathcal{S} \subseteq K[x_1,\ldots,x_n]$ is a set of polynomials, then

$$\mathcal{V}(\mathcal{S}) = \mathcal{V}(\langle S\rangle).$$

PROOF Exercise 1.2.2. □

To put the discussion of this section schematically, we view the $\mathcal{V}$-operator as a function

$$\mathcal{V} : \{\text{subsets of } K[x_1,\ldots,x_n]\} \longrightarrow \{\text{subsets of } \mathbb{A}^n\}.$$

This function is not surjective, since not every subset of $\mathbb{A}^n$ is an affine variety, but it becomes surjective, by definition, if we restrict the codomain to affine varieties:

$$\mathcal{V} : \{\text{subsets of } K[x_1,\ldots,x_n]\} \longrightarrow \{\text{affine varieties in } \mathbb{A}^n\}.$$

The $\mathcal{V}$-operator is also not injective, because different subsets of $K[x_1,\ldots,x_n]$ can define the same affine variety. As a first pass toward making it bijective, Proposition 1.15 shows that we can restrict the domain to ideals and maintain a surjective function:

$$\mathcal{V} : \{\text{ideals of } K[x_1,\ldots,x_n]\} \longrightarrow \{\text{affine varieties in } \mathbb{A}^n\}.$$

In fact, this operator is still not injective (see Exercise 1.2.3). It will take a further restriction on the domain and an assumption on the ground field K in order to finally obtain a bijective version of the $\mathcal{V}$-operator.

Exercises for Section 1.2

1.2.1 Prove that the only affine varieties in $\mathbb{A}^1$ are $\emptyset$, finite collections of points, and all of $\mathbb{A}^1$.

1.2.2 Prove Proposition 1.15.

1.2.3 Give an example of different ideals $I, J \in K[x]$ such that $\mathcal{V}(I) = \mathcal{V}(J)$.

1.2.4 Prove that the set

$$X = \{(a, a^2, a^3) \mid a \in K\} \subseteq \mathbb{A}^3$$

is an affine variety by finding a set of polynomials $\mathcal{S}$ for which $X = \mathcal{V}(\mathcal{S})$. (The variety X is called the *affine twisted cubic curve*.)

1.2.5 Let K be an infinite field. A *plane* in $\mathbb{A}^3_K$ is an affine variety P that can be defined as the vanishing set of a nonconstant linear polynomial:

$$P = \mathcal{V}(Ax + By + Cz + D),$$

where $A, B, C, D \in K$ and A, B, C are not all zero. Prove the following.

(a) The affine twisted cubic of Exercise 1.2.4 is not contained in any plane.

(b) For any degree-two polynomials $f_1, f_2, f_3 \in K[x]$, the set

$$X = \{(f_1(a), f_2(a), f_3(a)) \mid a \in K\} \subseteq \mathbb{A}^3$$

is contained in at least one plane. (**Hint**: Use linear algebra.)

1.2.6 Let K be an infinite field. Prove that the set

$$X = \{(a, ab) \mid a, b \in K\} \subseteq \mathbb{A}^2$$

is not an affine variety.

1.2.7 Prove that the vanishing set X of the expression $y - e^x$ inside $\mathbb{A}^2_{\mathbb{R}}$ is not an affine variety using the following steps.

(a) Suppose, toward a contradiction, that $X = \mathcal{V}(\mathcal{S})$ for some $\mathcal{S} \subseteq \mathbb{R}[x, y]$. Explain why there exists a nonzero polynomial $f \in \mathbb{R}[x, y]$ such that $f(a, e^a) = 0$ for all $a \in \mathbb{R}$.

(b) For any polynomial f as above, write

$$f(x, y) = p_0(x) + p_1(x)y + p_2(x)y^2 + \cdots + p_d(x)y^d,$$

where $p_d \neq 0 \in \mathbb{R}[x]$. Show that

$$\frac{p_0(a)}{e^{da}} + \frac{p_1(a)}{e^{(d-1)a}} + \frac{p_2(a)}{e^{(d-2)a}} + \cdots + \frac{p_{d-1}(a)}{e^a} + p_d(a) = 0$$

for all $a \in \mathbb{R}$.

(c) Take the limit of the above expression as $a \to \infty$ to conclude that

$$\lim_{a \to \infty} p_d(a) = 0.$$

By arguing that $\lim_{a \to \infty} g(a) \neq 0$ for any nonzero $g \in \mathbb{R}[x]$, deduce a contradiction.

1.2.8 (For students with some knowledge of topology) View $\mathbb{A}^n_{\mathbb{R}} = \mathbb{R}^n$ as a topological space with the Euclidean topology. Let $X \subseteq \mathbb{A}^n_{\mathbb{R}}$ be any affine variety other than $\mathbb{A}^n_{\mathbb{R}}$ itself. Prove that, as a topological subspace of $\mathbb{A}^n_{\mathbb{R}}$, the interior of X is empty. (The same result also holds, by a similar proof, with $\mathbb{R}$ replaced by $\mathbb{C}$.)

Section 1.3 The $\mathcal{I}$-operator

In Section 1.1, we learned how to associate, to any subset of $K[x_1, \ldots, x_n]$, a subset of $\mathbb{A}^n$ via the $\mathcal{V}$-operator. In this section, we reverse the procedure, describing an operator that associates a subset of $K[x_1, \ldots, x_n]$ to any subset of $\mathbb{A}^n$.

1.16 **DEFINITION** *Vanishing ideal*

Let $X \subseteq \mathbb{A}^n$ be a subset. The *vanishing ideal* of X is defined by

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

In other words, the vanishing ideal of X is the set of all polynomials that vanish on all of X. As the name suggests, the set $\mathcal{I}(X)$ is more than just a subset of the polynomial ring $K[x_1, \ldots, x_n]$, it is an ideal (see Exercise 1.3.1).

1.17 **EXAMPLE** Polynomials vanishing at $(0,0)$

Let $X = \{(0,0)\} \subseteq \mathbb{A}^2$. A polynomial $f \in K[x,y]$ vanishes at $(0,0)$ if and only if the constant term of f is zero. As explained in Example 0.26, the set of all such polynomials comprises the ideal $\langle x, y\rangle$. Thus, $\mathcal{I}(X) = \langle x, y\rangle$.

1.18 **EXAMPLE** Vanishing ideals in one variable

Let $X = \{1, 3\} \subseteq \mathbb{A}^1_{\mathbb{R}}$. By Corollary 0.48, a polynomial $f \in \mathbb{R}[x]$ vanishes at 1 if and only if $x - 1$ divides f, and f vanishes at 3 if and only if $x - 3$ divides f. Since $x - 1$ and $x - 3$ are irreducible, it follows from unique factorization in $\mathbb{R}[x]$ that

$$\mathcal{I}(X) = \{f \in \mathbb{R}[x] \mid (x-1)(x-3) \text{ divides } f\} = \langle x^2 - 4x + 3\rangle.$$

A similar procedure computes the vanishing ideal of any $X \subseteq \mathbb{A}^1$ (Exercise 1.3.3).

1.19 **EXAMPLE** Vanishing ideal of the parabola

Let K be an infinite field and let X be the affine variety defined by

$$X = \mathcal{V}(y - x^2) = \{(a, a^2) \mid a \in K\} \subseteq \mathbb{A}^2.$$

Which polynomials vanish at every point of X? Certainly any polynomial of the form $(y - x^2)f(x,y)$ vanishes at every point of X, so $\langle y - x^2\rangle \subseteq \mathcal{I}(X)$. Let's show that the reverse containment also holds.

Let $f \in \mathcal{I}(X)$. The arguments in Example 0.29 show that $[f(x,y)] = [f(x,x^2)]$ in the quotient ring $K[x,y]/\langle y - x^2\rangle$. In particular, this means that

$$f(x,y) - f(x,x^2) \in \langle y - x^2\rangle. \tag{1.20}$$

Using our assumption that f vanishes on $X = \{(a, a^2) \mid a \in K\}$, we see that $f(a, a^2) = 0$ for every $a \in K$, which implies that $f(x,x^2) \in K[x]$ is a single-variable polynomial with infinitely many zeros. This is only possible if $f(x,x^2)$ is the zero polynomial. Substituting $f(x,x^2) = 0$ into Equation (1.20), we see that $f \in \langle y - x^2\rangle$, implying that $\mathcal{I}(X) \subseteq \langle y - x^2\rangle$.

Having argued both inclusions, we conclude that $\mathcal{I}(X) = \langle y - x^2\rangle$.

Even in the simple case of the parabola in Example 1.19, it was already somewhat involved to show that the vanishing ideal was $\langle y - x^2 \rangle$. Indeed, computing the vanishing ideal of affine varieties in general is a nontrivial task that usually requires ad hoc methods in each case (see Exercises 1.3.3–1.3.8 for more examples of such computations). As we will see in Section 1.5, the Nullstellensatz greatly simplifies the task of computing vanishing ideals, and it is partly for this reason that it deserves the status of the "Fundamental Theorem of Algebraic Geometry."

Now that we have introduced both the $\mathcal{V}$- and $\mathcal{I}$-operators, which pass between subsets of $K[x_1, \ldots, x_n]$ and subsets of $\mathbb{A}^n$, it is natural to wonder to what extent these operators are inverse to each other. The next result provides a first answer, showing that they are not generally inverse, but that they become inverse upon restricting to vanishing ideals and affine varieties.

1.21 **PROPOSITION** *Composing $\mathcal{V}$- and $\mathcal{I}$-operators*

1. Let $X \subseteq \mathbb{A}^n$. Then
$$\mathcal{V}(\mathcal{I}(X)) \supseteq X,$$
with equality if and only if $X = \mathcal{V}(\mathcal{S})$ for some $\mathcal{S} \subseteq K[x_1, \ldots, x_n]$.
2. Let $\mathcal{S} \subseteq K[x_1, \ldots, x_n]$. Then
$$\mathcal{I}(\mathcal{V}(\mathcal{S})) \supseteq \mathcal{S},$$
with equality if and only if $\mathcal{S} = \mathcal{I}(X)$ for some $X \subseteq \mathbb{A}^n$.

PROOF We prove Part 1 and leave Part 2 to Exercise 1.3.2.

Let $a = (a_1, \ldots, a_n) \in X$. To prove that $a \in \mathcal{V}(\mathcal{I}(X))$, consider any polynomial $f \in \mathcal{I}(X)$. By definition of the vanishing ideal, we have that $f(b) = 0$ for all $b \in X$. In particular, $f(a) = 0$. Thus, we have proved that, for every $f \in \mathcal{I}(X)$, $f(a) = 0$. This implies that $a \in \mathcal{V}(\mathcal{I}(X))$.

If equality holds, then X is the vanishing set of $\mathcal{S} = \mathcal{I}(X)$, proving one direction of the if-and-only-if statement. To prove the converse, suppose that $X = \mathcal{V}(\mathcal{S})$ for some set $\mathcal{S} \subseteq K[x_1, \ldots, x_n]$; we must prove that $\mathcal{V}(\mathcal{I}(X)) \subseteq X$. It is equivalent to prove that $a \notin X$ implies $a \notin \mathcal{V}(\mathcal{I}(X))$, so suppose the former. Since $X = \mathcal{V}(\mathcal{S})$ and $a \notin X$, there exists some $f \in \mathcal{S}$ such that $f(a) \neq 0$. Since $f \in \mathcal{S}$ and $X = \mathcal{V}(\mathcal{S})$, f vanishes on all of X, implying that $f \in \mathcal{I}(X)$. Because $f \in \mathcal{I}(X)$ and $f(a) \neq 0$, we conclude that $a \notin \mathcal{V}(\mathcal{I}(X))$. □

If $X \subseteq \mathbb{A}^n$ is an affine variety, Part 1 of the proposition says that there is distinguished ideal for which X is the vanishing set, namely $\mathcal{I}(X)$.

If $X \subseteq \mathbb{A}^n$ is a subset, then the set $\mathcal{V}(\mathcal{I}(X))$ appearing in the first part of Proposition 1.21 has another interpretation: it is the smallest affine variety containing X (Exercise 1.3.10). In this sense, it is analogous to the ideal generated by a set $\mathcal{S}$, which is the smallest ideal containing $\mathcal{S}$; we might even call $\mathcal{V}(\mathcal{I}(X))$ the "affine variety generated by X" and denote it by V_X. With this notation, the last part of Proposition 1.21 becomes analogous to Proposition 1.15:

$$\mathcal{I}(X) = \mathcal{I}(V_X).$$

Another important comment is that the containments in Proposition 1.21 can certainly be strict. For example, if X is not an affine variety, then $X \neq \mathcal{V}(\mathcal{S})$ for any $\mathcal{S}$, so $X \neq \mathcal{V}(\mathcal{I}(X))$. Similarly, if $\mathcal{S}$ is not an ideal, then it cannot be the case that $\mathcal{S} = \mathcal{I}(\mathcal{V}(\mathcal{S}))$, because the latter is an ideal. In fact, even when $\mathcal{S}$ *is* an ideal, equality still need not hold, as illustrated in the next example.

1.22 **EXAMPLE** $\mathcal{I}(\mathcal{V}(I)) \neq I$

Let $I = \langle x^2 \rangle \subseteq K[x]$. Then

$$\mathcal{V}(I) = \{a \in K \mid a^2 = 0\}.$$

Since K is a field, $a^2 = 0$ if and only if $a = 0$. Thus, $\mathcal{V}(I) = \{0\}$. The same reasoning as in Example 1.18 shows that $\mathcal{I}(\{0\}) = \langle x \rangle$, so

$$\mathcal{I}(\mathcal{V}(I)) = \mathcal{I}(\{0\}) = \langle x \rangle.$$

The vanishing ideal $\mathcal{I}(\mathcal{V}(I)) = \langle x \rangle$ contains but is not equal to $I = \langle x^2 \rangle$; for example, $x \in \langle x \rangle$ but $x \notin \langle x^2 \rangle$. As a consequence of the second part of Proposition 1.21, we conclude that I is not the vanishing ideal of any $X \subseteq \mathbb{A}^1$.

At the end of the Section 1.2, we saw that the $\mathcal{V}$-operator gives a surjection

$$\mathcal{V} : \{\text{ideals in } K[x_1, \ldots, x_n]\} \longrightarrow \{\text{affine varieties in } \mathbb{A}^n\}.$$

It follows from Proposition 1.21 that the $\mathcal{V}$-operator becomes a bijection (with inverse the $\mathcal{I}$-operator) if we restrict the domain to the set of *vanishing ideals*—those ideals that arise as vanishing ideals of some set. Our goal then, if we want to understand the dictionary between algebra and geometry, is to obtain a better understanding of the vanishing ideals in $K[x_1, \ldots, x_n]$. Motivated by the observations in Example 1.22, we take a first step in this direction in the next section, where we introduce the algebraic notion of a radical ideal.

Exercises for Section 1.3

1.3.1 For any subset $X \subseteq \mathbb{A}^n$, prove that $\mathcal{I}(X) \subseteq K[x_1, \ldots, x_n]$ is an ideal.

1.3.2 Prove Proposition 1.21, Part 2.

1.3.3 Let $X \subseteq \mathbb{A}^1$.

(a) Suppose that $X = \{a_1, \ldots, a_r\}$ is finite. Use unique factorization in $K[x]$ to prove that

$$\mathcal{I}(X) = \langle (x - a_1)(x - a_2) \cdots (x - a_r) \rangle \subseteq K[x].$$

(b) Suppose that X is infinite. Prove that $\mathcal{I}(X) = \langle 0 \rangle$.

1.3.4 Compute the vanishing ideal of $\mathcal{V}(x^2 + 1) \subseteq \mathbb{A}^1$

(a) over $\mathbb{R}$, and

(b) over $\mathbb{C}$.

1.3.5 Let $X = \{(a_1, \ldots, a_n)\} \in \mathbb{A}^n$ be a single point. Prove that

$$\mathcal{I}(X) = \langle x_1 - a_1, \ldots, x_n - a_n \rangle.$$

1.3.6 Let $X = \mathcal{V}(x^2 + y^2 - 1) \subseteq \mathbb{A}^2_{\mathbb{R}}$ be the unit circle. This exercise proves that $\mathcal{I}(X) = \langle x^2 + y^2 - 1 \rangle$.

(a) Prove that $\mathcal{I}(X) \supseteq \langle x^2 + y^2 - 1 \rangle$.

(b) Prove that $\mathcal{I}(X) \subseteq \langle x^2 + y^2 - 1 \rangle$, possibly using the following proof outline. Suppose $f \in \mathcal{I}(X)$.

i. Prove that

$$f - g_1 - yg_2 \in \langle x^2 + y^2 - 1 \rangle$$

for some $g_1, g_2 \in \mathbb{R}[x]$.

ii. Using that $f \in \mathcal{I}(X)$, prove that

$$g_1(a)^2 = (1 - a^2)g_2(a)^2$$

for all $a \in [-1, 1]$, and conclude that $g_1(x)^2 = (1 - x^2)g_2(x)^2$.

iii. Use unique factorization to prove that g_1 and g_2 are both the zero polynomial and conclude that $f \in \langle x^2 + y^2 - 1 \rangle$.

1.3.7 Let $X = \mathcal{V}(x_1^2 + \cdots + x_n^2 - 1) \subseteq \mathbb{A}^n_{\mathbb{R}}$ be the unit n-sphere. Generalize the previous exercise to prove that $\mathcal{I}(X) = \langle x_1^2 + \cdots + x_n^2 - 1 \rangle$.

1.3.8 Let K be an infinite field with $1 \neq -1$ and let $X = \mathcal{V}(x^2 + y^2 - 1) \subseteq \mathbb{A}^2_K$.

(a) For any $a \in K$ with $a^2 \neq -1$, prove that

$$\left(\frac{a^2 - 1}{a^2 + 1}, \frac{2a}{a^2 + 1} \right) \in X.$$

(b) Prove that there are infinitely many values $a \in K$ such that $(a, b) \in X$ for some $b \in K$.

(c) Generalize the result of Exercise 1.3.6 to base field K.

1.3.9 Let $K = \mathbb{F}_2$, the finite field with two elements. Let

$$X = \mathcal{V}(y - x^2) \subseteq \mathbb{A}^2_{\mathbb{F}_2}.$$

Prove that $\mathcal{I}(X) \neq \langle y - x^2 \rangle$, in contrast to the analogous case over $\mathbb{R}$ studied in Example 1.19. What is $\mathcal{I}(X)$?

1.3.10 Let $X \subseteq \mathbb{A}^n$ be any subset.

(a) Prove that $\mathcal{V}(\mathcal{I}(X))$ is the smallest affine variety containing X, in the following sense: if $Y \subseteq \mathbb{A}^n$ is any affine variety and $X \subseteq Y$, then $\mathcal{V}(\mathcal{I}(X)) \subseteq Y$.

(b) Demonstrate part (a) by choosing any set $X \subseteq \mathbb{A}^2_{\mathbb{R}}$ that is not an affine variety and calculating $\mathcal{V}(\mathcal{I}(X))$.

Section 1.4 Radical ideals

In the previous section, we learned that not every ideal in $K[x_1, \ldots, x_n]$ arises as a vanishing ideal of some subset in $\mathbb{A}^n$. In particular, we noticed in Example 1.22 that the ideal $\langle x^2 \rangle \subseteq K[x]$ is not a vanishing ideal. How, then, can we recognize whether a given ideal is $\mathcal{I}(X)$ for some X? We investigate one important property of vanishing ideals in this section: vanishing ideals are *radical*.

1.23 **DEFINITION** *Radical ideal*

Let R be a ring. An ideal $I \subseteq R$ is *radical* if, for all $a \in R$,

$$a^m \in I \text{ for some integer } m > 0 \Longrightarrow a \in I.$$

In other words, an ideal is radical if it is closed under taking roots: whenever a power of an element is in the ideal, the element itself must be in the ideal. Notice that the ideal $\langle x^2 \rangle \subseteq K[x]$ from Example 1.22 is not radical, because $x^2 \in \langle x^2 \rangle$ but $x \notin \langle x^2 \rangle$. The next result says that the property of radical-ness is an attribute of all vanishing ideals.

1.24 **PROPOSITION** *Vanishing ideals are radical*

For any set $X \subseteq \mathbb{A}^n$, the vanishing ideal $\mathcal{I}(X)$ is a radical ideal.

In particular, Proposition 1.24 and the observation that $\langle x^2 \rangle$ is not a radical ideal together imply that $\langle x^2 \rangle$ is not a vanishing ideal, as we observed in Example 1.22.

PROOF OF PROPOSITION 1.24 Let $X \subseteq \mathbb{A}^n$ be a subset and suppose there exists a positive integer m such that $f^m \in \mathcal{I}(X)$, or in other words, such that $f^m(a) = 0$ for all $a \in X$. Then

$$0 = f^m(a) = \big(f(a)\big)^m \in K.$$

Since K is a field, it has no zero divisors; in particular, $\big(f(a)\big)^m = 0$ if and only if $f(a) = 0$. Thus, $f(a) = 0$ for all $a \in X$, so $f \in \mathcal{I}(X)$. □

Proposition 1.24 suggests, in particular, that radical ideals play a central role in algebraic geometry. Consequently, we devote the rest of this section to collecting some of the fundamental notions pertaining to radical ideals.

By definition, an ideal fails to be radical if it is not closed under taking roots. You might suspect, then, that you could construct a radical ideal simply by adding in the missing roots. The next definition makes this construction precise.

1.25 **DEFINITION** *Radical of an ideal*

Let $I \subseteq R$ be an ideal. The *radical* of I is

$$\sqrt{I} = \{a \in R \mid a^m \in I \text{ for some } m > 0\}.$$

This process of "adding in the missing roots" indeed yields a radical ideal, as the next result shows.

1.26 PROPOSITION *Radicals are radical*

If $I \subseteq R$ is an ideal, then $\sqrt{I}$ is a radical ideal.

PROOF The proof that $\sqrt{I}$ is an ideal is Exercise 1.4.2. To prove that $\sqrt{I}$ is radical, assume $a^m \in \sqrt{I}$ for some integer $m > 0$; we must prove that $a \in \sqrt{I}$. By the definition of $\sqrt{I}$, we have

$$a^n \in \sqrt{I} \Longrightarrow (a^n)^m \in I \text{ for some integer } m > 0.$$

In other words, $a^{(nm)} \in I$, from which it follows that $a \in \sqrt{I}$. □

If I is a radical ideal, then it is already closed under taking roots, so $\sqrt{I} = I$. Conversely, if $\sqrt{I} = I$, then Proposition 1.26 implies that I is a radical ideal. Thus, we have proved the following useful characterization of radical ideals.

1.27 COROLLARY *Characterization of radical ideals*

The ideal $I \subseteq R$ is radical if and only if $I = \sqrt{I}$.

We now provide several examples to help familiarize ourselves with radicals.

1.28 EXAMPLE

If $R = K[x]$ and $I = \langle x^2 \rangle$, then we have already seen that I is not radical. If we want to enlarge I by adding all possible roots of elements in I, what should we add? Let's show that

$$\sqrt{I} = \langle x \rangle.$$

To see this, first note that $x \in \sqrt{I}$ because $x^2 \in I$. Since $\sqrt{I}$ is an ideal, the fact that $x \in \sqrt{I}$ then implies that $\langle x \rangle \subseteq \sqrt{I}$, proving one containment.

Conversely, suppose that $f \notin \langle x \rangle$. This means that f has a nonzero constant term. It follows that f^m has a nonzero constant term for all integers $m > 0$, so $f^m \notin I$. Hence, $f \notin \sqrt{I}$, which proves by contrapositive that $\sqrt{I} \subseteq \langle x \rangle$.

1.29 EXAMPLE

If $R = \mathbb{Z}$ and $I = \langle 12 \rangle$, then I is not radical because $6^2 \in I$ but $6 \notin I$. In fact,

$$\sqrt{I} = \langle 6 \rangle.$$

To prove the containment $\sqrt{I} \subseteq \langle 6 \rangle$, let $r \in \sqrt{I}$. Then $r^m = 12k$ for some positive integer m and some integer k. Since $2 \,|\, r^m$, it follows from Euclid's Lemma that $2 \,|\, r$. The same reasoning shows that $3 \,|\, r$. Since r is divisible by both 2 and 3, it is divisible by 6, which means that $r \in \langle 6 \rangle$, proving one containment.

Conversely, if $r \in \langle 6 \rangle$, then $r = 6k$ for some integer k. Thus,

$$r^2 = 36k^2 = 12(3k^2),$$

so $r^2 \in I$. This implies that $r \in \sqrt{I}$, verifying that $\langle 6 \rangle \subseteq \sqrt{I}$.

Comparing the radical ideal $\langle 6\rangle$ and the nonradical ideal $\langle 12\rangle$ in the previous example, we see that 6 is *square-free*—in other words, it is not divisible by the square of a prime—but 12 is not square-free; it is divisible by 2^2. Arguments similar to those in Exampe 1.29 show that $\langle n\rangle \subseteq \mathbb{Z}$ is radical if and only if n is square-free.

In the context of algebraic geometry, it would be useful to be able to determine when an ideal $I \subseteq K[x_1,\dots,x_n]$ is radical. For general ideals, this can be quite difficult. However, the situation is analogous to the case of the integers when I is a principal ideal. To state the result precisely, let $f \in K[x_1,\dots,x_n]$, and consider an irreducible factorization:

$$f = p_1 \cdots p_m.$$

Because of irreducibility, $p_i \mid p_j$ if and only if they differ by a constant. By collecting all of the terms that differ by a constant, we can write

$$f = a q_1^{k_1} \cdots q_\ell^{k_\ell} \tag{1.30}$$

where $a \in K$ is a constant, each q_i is irreducible, and $q_i \nmid q_j$ whenever $i \neq j$. We say that (1.30) is a *distinct irreducible factorization* of f and that the q_i are the *distinct irreducible factors* of f. It follows from unique factorization that the distinct irreducible factors are unique up to reordering and multiplying by constants.

The next result describes the radical of a principal ideal in terms of these factors.

1.31 **PROPOSITION** *Radicals of principal ideals*

If $f \in K[x_1,\dots,x_n]$ has distinct irreducible factors $q_1,\dots,q_\ell$, then

$$\sqrt{\langle f\rangle} = \langle q_1 \cdots q_\ell\rangle. \tag{1.32}$$

In particular, $\langle f\rangle$ is radical if and only if f is not divisible by the square of a nonconstant polynomial.

PROOF We prove both inclusions in the equality (1.32) and leave the deduction of the if-and-only-if assertion to Exercise 1.4.3.

($\subseteq$) Suppose $g \in \sqrt{\langle f\rangle}$. Then $g^m = hf$ for some $m > 0$ and $h \in K[x_1,\dots,x_n]$. In particular, $q_i \mid g^m$ for every i. By uniqueness of distinct irreducible factors, each q_i must be one of the distinct irreducible factors of g^m. Since the distinct irreducible factors of g^m are the same as those of g, then each q_i must be one of the distinct irreducible factors of g. It follows that $q_1 \cdots q_\ell \mid g$, so $g \in \langle q_1 \cdots q_\ell\rangle$.

($\supseteq$) Suppose $g \in \langle q_1 \cdots q_\ell\rangle$ and write $g = hq_1 \cdots q_\ell$ for some polynomial h. Regarding the distinct irreducible factorization

$$f = a q_1^{k_1} \cdots q_\ell^{k_\ell},$$

set $m = \max\{k_1,\dots,k_\ell\}$. It follows that

$$g^m = h^m q_1^m \cdots q_\ell^m = \left(h^m q_1^{m-k_1} \cdots q_\ell^{m-k_\ell} a^{-1}\right) f \in \langle f\rangle,$$

which implies that $g \in \sqrt{\langle f\rangle}$. □

1.33 **EXAMPLE** $\langle x^2+y^2\rangle \subseteq \mathbb{C}[x,y]$ is radical

The irreducible factorization of x^2+y^2 in $\mathbb{C}[x,y]$ is

$$x^2+y^2 = (x-iy)(x+iy).$$

Since the irreducible factors are distinct, x^2+y^2 is square free. Thus, the ideal $\langle x^2+y^2\rangle \subseteq \mathbb{C}[x,y]$ is radical.

1.34 **EXAMPLE** A radical ideal that is not a vanishing ideal

Over the real numbers, the polynomial $x^2+y^2 \in \mathbb{R}[x,y]$ is irreducible. Thus, by Proposition 1.31, the ideal $\langle x^2+y^2\rangle$ is radical. However, since the origin is the only point at which x^2+y^2 vanishes, it follows that

$$\mathcal{I}(\mathcal{V}(\langle x^2+y^2\rangle)) = \mathcal{I}(\{(0,0)\}) = \langle x,y\rangle \neq \langle x^2+y^2\rangle.$$

Thus, Proposition 1.21 implies that $\langle x^2+y^2\rangle$ is not a vanishing ideal. In particular, the converse of Proposition 1.24 does not hold over $\mathbb{R}$: radical ideals need not be vanishing ideals.

In our study of rings, we have now had the opportunity to meet three special types of ideals: maximal, prime, and radical. We already know that maximal ideals are prime. The next result adds radical ideals to this hierarchy.

1.35 **PROPOSITION** *Prime and maximal ideals are radical*

Every prime ideal is a radical ideal.

PROOF Toward proving the contrapositive, assume that I is not a radical ideal. Choose an element $a \notin I$ such that $a^m \in I$ for some $m > 1$. If m_0 is the smallest integer greater than 1 such that $a^{m_0} \in I$, then neither a nor a^{m_0-1} are elements of I, but their product a^{m_0} is an element of I. Thus, I is not prime. □

Schematically, for any ring R, we have the following hierarchy of ideals:

$$\{\text{ideals}\} \supseteq \{\text{radical ideals}\} \supseteq \{\text{prime ideals}\} \supseteq \{\text{maximal ideals}\}.$$

While these inclusions are not strict for every ring, they are strict for multivariable polynomial rings over fields (Exercise 1.4.4).

To conclude this section, we return to our overarching goal of obtaining a bijection between algebraic objects and geometric objects. Since $\mathcal{V}(\mathcal{I}(X)) = X$ for all affine varieties $X \subseteq \mathbb{A}^n$ (Proposition 1.21) and $\mathcal{I}(X)$ is radical (Proposition 1.24), the $\mathcal{V}$-operator remains a surjection upon restricting the domain:

$$\mathcal{V} : \{\text{radical ideals in } K[x_1,\ldots,x_n]\} \longrightarrow \{\text{affine varieties in } \mathbb{A}^n\}. \tag{1.36}$$

One might be so optimistic as to hope that (1.36) is our sought-after bijection between algebraic objects and geometric objects. Unfortunately, Example 1.34 provides a counterexample: $\langle x^2+y^2\rangle$ and $\langle x,y\rangle$ are distinct radical ideals in $\mathbb{R}[x,y]$

with the same vanishing set $\{(0,0)\} \in \mathbb{A}^2_{\mathbb{R}}$. Nonetheless, if we make an additional assumption on the ground field K—that it is *algebraically closed*—then (1.36) is, indeed, the bijection we desire. This result is a consequence of the Nullstellensatz, to which we turn in the next section.

Exercises for Section 1.4

1.4.1 Determine which of the following ideals are radical. For those that are not radical, compute their radical.

(a) $\langle 4 \rangle \subseteq \mathbb{Z}$
(b) $\langle 6 \rangle \subseteq \mathbb{Z}$
(c) $\langle 18 \rangle \subseteq \mathbb{Z}$
(d) $\langle x^2 + y^2 - 1 \rangle \subseteq \mathbb{R}[x,y]$
(e) $\langle x^2, y^3 \rangle \subseteq \mathbb{R}[x,y]$
(f) $\langle y - x^2, y \rangle \subseteq \mathbb{R}[x,y]$
(g) $\langle x^2 - y^2 \rangle \subseteq \mathbb{R}[x,y]$

1.4.2 Let $I \subseteq R$ be an ideal. Prove that $\sqrt{I} \subseteq R$ is an ideal. (**Hint**: Use the binomial theorem to prove that $\sqrt{I}$ is closed under addition.)

1.4.3 Prove that a principal ideal $\langle f \rangle \subseteq K[x_1, \ldots, x_n]$ is radical if and only if f is not divisible by the square of a nonconstant polynomial.

1.4.4 (a) Give an example of an ideal $I \subseteq K[x,y]$ that is not radical.
(b) Give an example of a radical ideal $I \subseteq K[x,y]$ that is not prime.
(c) Give an example of a prime ideal $I \subseteq K[x,y]$ that is not maximal.

1.4.5 Prove, by example, that an ideal $\langle f_1, f_2 \rangle$ need not be radical even if f_1 and f_2 are both square-free.

1.4.6 Prove that the set of all zero-divisors in a ring R is a radical ideal.

1.4.7 Let R be a ring and let $I \subseteq R$ an ideal. Prove that $\sqrt{I} = R$ if and only if $I = R$.

1.4.8 Let R be a ring and let $I, J \subseteq R$ be ideals. Prove that

$$\sqrt{I \cap J} = \sqrt{I} \cap \sqrt{J}.$$

1.4.9 Let R be a ring and $I \subseteq R$ an ideal. Prove that $\sqrt{I}$ is the intersection of all prime ideals in R that contain I.

Section 1.5 The Nullstellensatz

"Nullstellensatz" is a German word composed of "Nullstellen" (zeroes) and "Satz" (theorem).

In this section, we state the theorem that forms the foundation of much of algebraic geometry: the *Nullstellensatz*. This result allows us to set up a powerful correspondence between affine varieties and radical ideals in polynomial rings, which is the backbone of our dictionary between the worlds of geometry and algebra.

To state the theorem, we require a key assumption on the ground field.

1.37 **DEFINITION** *Algebraically closed*

A field K is said to be *algebraically closed* if any nonconstant polynomial in $K[x]$ has at least one zero in K.

For example, the fields $\mathbb{Q}$ and $\mathbb{R}$ are not algebraically closed because x^2+1 does not have any rational or real zeroes. However, x^2+1 does have zeroes in $\mathbb{C}$, namely i and $-\mathrm{i}$. In fact, every nonconstant polynomial in $\mathbb{C}[x]$ has at least one zero, which is the statement of the Fundamental Theorem of Algebra.

1.38 **THEOREM** *Fundamental Theorem of Algebra*

The field $\mathbb{C}$ is algebraically closed.

Although the Fundamental Theorem of Algebra is often taught at an early stage, it is by no means obvious. There are many proofs, including arguments via Galois Theory, complex analysis, and topology. None of these falls within the scope of this book, and the fact that $\mathbb{C}$ is algebraically closed is not necessary for the logical development of the material. The Fundamental Theorem of Algebra is introduced here simply to emphasize that there is at least one familiar and concrete example of an algebraically closed field, and we encourage the reader to accept it without proof.

As we observed in Section 1.1, much of algebraic geometry is more straightforward when the ground field is infinite, because one no longer needs to draw a distinction between polynomials and the functions they define. One advantage of working with algebraically closed fields is that they are automatically infinite.

1.39 **PROPOSITION** *Algebraically closed fields are infinite*

Let K be an algebraically closed field. Then K is infinite.

PROOF Exercise 1.5.1. □

In particular, none of the finite fields $\mathbb{F}_p$ for any prime $p \in \mathbb{Z}$ are algebraically closed. What fields are algebraically closed, then, besides $\mathbb{C}$? There are perhaps no other familiar examples of algebraically closed fields, but there is a procedure by which one can construct, from any field K, an *algebraic closure* $\overline{K}$, which is the smallest algebraically closed field in which K is contained. The most familiar

application of this construction says that $\overline{\mathbb{R}} = \mathbb{C}$. Applying this procedure to any field at all yields a host of new examples of algebraically closed fields, albeit not particularly familiar ones: $\overline{\mathbb{Q}}$, $\overline{\mathbb{F}_p}$, and so on. Readers unfamiliar with this material are encouraged, whenever K is assumed to be algebraically closed, to think of the case $K = \mathbb{C}$.

Having discussed what it means for a field to be algebraically closed, we can now state the Nullstellensatz, the proof of which is deferred to Chapter 5.

1.40 **THEOREM** *Nullstellensatz*

Let K be an algebraically closed field. Then, for any ideal $I \subseteq K[x_1, \ldots, x_n]$,

$$\mathcal{I}(\mathcal{V}(I)) = \sqrt{I}.$$

The containment $\sqrt{I} \subseteq \mathcal{I}(\mathcal{V}(I))$ is true over any field, and can be proved directly from the definitions (Exercise 1.5.3). The other inclusion requires a good deal of work, for which K being algebraically closed is essential.

The Nullstellensatz helps us answer the motivating question posed at the end of Section 1.2: on what domain and codomain does the $\mathcal{V}$-operator become a bijection? Over algebraically closed fields, the answer is that $\mathcal{V}$ is a bijection between radical ideals and affine varieties. This is the first key instance of the precise dictionary between algebra and geometry.

1.41 **COROLLARY** *Radical ideals and affine varieties*

If K is algebraically closed, then

$$\mathcal{V} : \{\text{radical ideals of } K[x_1, \ldots, x_n]\} \longrightarrow \{\text{affine varieties in } \mathbb{A}^n\}$$

is a bijection with inverse $\mathcal{I}$.

PROOF Since a function is bijective if and only if it has an inverse, it suffices to prove that $\mathcal{I}$ is the inverse of $\mathcal{V}$. In other words, we must show that, for any affine variety $X \subseteq \mathbb{A}^n$,

$$\mathcal{V}(\mathcal{I}(X)) = X \tag{1.42}$$

and, for any radical ideal $I \subseteq K[x_1, \ldots, x_n]$,

$$\mathcal{I}(\mathcal{V}(I)) = I. \tag{1.43}$$

The equality (1.42) is the first part of Proposition 1.21, so it is true without any assumptions on K. To prove (1.43), notice that

$$\mathcal{I}(\mathcal{V}(I)) = \sqrt{I} = I,$$

where the first equality is the Nullstellensatz and the second is the characterization of radical ideals as those ideals that are equal to their radical (Corollary 1.27). □

In addition to providing a bijection between affine varieties and radical ideals, the Nullstellensatz is also a useful tool for computing vanishing ideals. In particular, it is often much easier to determine whether an ideal is radical than it is to determine whether it is a vanishing ideal. The following examples illustrate this point.

1.44 **EXAMPLE** $\mathcal{I}(\mathcal{V}(y-x^2)) = \langle y-x^2\rangle$

Let K be algebraically closed and consider the affine variety

$$X = \mathcal{V}(y-x^2) \subseteq \mathbb{A}^2.$$

Since every algebraically closed field is infinite, we know from Example 1.19 that $\mathcal{I}(X) = \langle y-x^2\rangle$. However, the argument in that example was somewhat involved and special to the particular polynomial $y-x^2$. For algebraically closed fields, there is a much simpler argument that applies to all irreducible polynomials.

In particular, knowing that $y-x^2 \in K[x,y]$ is an irreducible polynomial, it follows that $\langle y-x^2\rangle \subseteq K[x,y]$ is a prime ideal (Proposition 0.62), and therefore radical (Proposition 1.35). Thus, by the Nullstellensatz,

$$\mathcal{I}(X) = \mathcal{I}(\mathcal{V}(y-x^2)) = \sqrt{\langle y-x^2\rangle} = \langle y-x^2\rangle.$$

1.45 **EXAMPLE** $\mathcal{I}(\mathcal{V}(f)) = \langle f\rangle$ when f is square-free

Generalizing the previous example, we see that, whenever K is algebraically closed and $f \in K[x_1,\ldots,x_n]$ is not divisible by the square of a nonzero polynomial, we have

$$\mathcal{I}(\mathcal{V}(f)) = \sqrt{\langle f\rangle} = \langle f\rangle.$$

The first equality is the Nullstellensatz and the second follows from Proposition 1.31. In particular, if f is irreducible, then the vanishing ideal of $\mathcal{V}(f)$ is simply $\langle f\rangle$.

Notice that the conclusion of this example fails when K is not algebraically closed. For instance, consider the irreducible polynomial $x^2+1 \in \mathbb{R}[x]$. Since this polynomial does not have any zeros,

$$\mathcal{I}(\mathcal{V}(x^2+1)) = \mathcal{I}(\emptyset) = K[x] \neq \langle x^2+1\rangle.$$

The bijection between radical ideals and affine varieties merely scratches the surface of the rich dictionary that we will continue to build between algebra and geometry. To draw an analogy with languages, the bijection in Corollary 1.41 should be thought of as a translation of nouns between two languages; such a translation might allow us to have very simple conversations, but if we want to take full advantage of the richness of language, we should also translate the verbs, the adjectives, the adverbs, and so on.

Over the course of the next three chapters, we will continue to build the dictionary between algebra and geometry. As we do so, we will have the opportunity to introduce a number of new algebraic notions that are useful along the way. In Chapter 5, once we have developed a more robust algebraic foundation and a fuller appreciation of the dictionary between algebra and geometry, we will return to give the proof of the Nullstellensatz.

ON OUR ASSUMPTIONS REGARDING THE GROUND FIELD K

The Nullstellensatz is the backbone of algebraic geometry, and as such, **we assume for the remainder of this book, unless otherwise stated, that K is an algebraically closed field.** Many of the definitions and results that we develop remain valid over general fields. Others, however, require slight modifications, and some are just outright wrong in the non-algebraically-closed case. To help the reader appreciate our assumptions, we regularly turn to the setting of $K = \mathbb{R}$ to illustrate nonexamples of results where being algebraically closed is essential.

Even though the central results of algebraic geometry do not hold over the field of real numbers, as they are not algebraically closed, much of our geometric intuition for affine varieties arises from viewing solutions of polynomials over $\mathbb{R}$. Indeed, every geometric image of a vanishing set in Section 1.1 depicts an affine variety over $\mathbb{R}$. As algebraic geometers, it is important to develop the skill of using our knowledge and intuition over $\mathbb{R}$ as a source of insight, while at the same time not being misled by phenomena that may occur in that special setting as a result of the fact that $\mathbb{R}$ is not algebraically closed.

As we move forward, even though our ground field will always be assumed to be algebraically closed, we will continue to discuss and depict examples of varieties by looking at their solutions over $\mathbb{R}$, and we will continue to use familiar words from our years of experience working with these sets. For example, we refer to the variety $\mathcal{V}(y - x^2) \subseteq \mathbb{A}^2$ as a *parabola* and $\mathcal{V}(x^2 + y^2 + z^2 - 1) \subseteq \mathbb{A}^3$ as the *unit sphere*, even though, over general fields, these varieties may not closely resemble the geometric picture in our mind that the words *parabola* and *sphere* connote. Since $\mathbb{R}$ is a subset of the algebraically closed field $\mathbb{C}$, the reader is welcome to assume $K = \mathbb{C}$ throughout, in which case the images over $\mathbb{R}$ depicted in the examples are a subset of the full solution set over $\mathbb{C}$. The images do not give us the whole picture, but they at least provide a glimpse into the nature of the variety.

Another important attribute of a field is its *characteristic*. Recall that the characteristic of K, denoted $\mathrm{Char}(K)$, is the smallest positive integer n such that

$$\underbrace{1 + \cdots + 1}_{n} = 0 \in K.$$

If no such n exists, as is the case for $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, then we say that the field has characteristic 0. All of the results in this book hold for a general algebraically closed field. However, in many examples, we often want to avoid a finite list of characteristics because they might exhibit unusual behavior with particular types of polynomials. For example, when $\mathrm{Char}(K) \neq 2$, the polynomial $x^2 + y^2 - 1$ is irreducible, but if $1 + 1 = 0 \in K$, then

$$x^2 + y^2 - 1 = (x + y + 1)^2.$$

Rather than mentioning the exceptional characteristics, we often assume for simplicity in specific examples that $K = \mathbb{C}$, even though the examples usually extend to algebraically closed fields with only finitely many exceptions on the characteristic.

Exercises for Section 1.5

1.5.1 Prove that any algebraically closed field is infinite. (**Hint**: If $K = \{a_1, \ldots, a_r\}$ is finite, can you construct a polynomial in $K[x]$ with no zeroes in K?)

1.5.2 Let K be algebraically closed and let $f \in K[x]$ be a polynomial of degree d. Prove that there exist $a_0, a_1, \ldots, a_d \in K$, not necessarily distinct, such that

$$f = a_0(x - a_1)\ldots(x - a_d).$$

1.5.3 Let K be any field and $I \subseteq K[x_1, \ldots, x_n]$. Prove that one inclusion of the Nullstellensatz holds without assuming that K is algebraically closed:

$$\sqrt{I} \subseteq \mathcal{I}(\mathcal{V}(I)).$$

1.5.4 Prove that the Nullstellensatz fails for any field that is not algebraically closed.

1.5.5 Let K be algebraically closed and let $I \subseteq K[x_1, \ldots, x_n]$ be an ideal. Assuming the Nullstellensatz, prove that $\mathcal{V}(I) = \emptyset$ if and only if $I = K[x_1, \ldots, x_n]$. (This result is often called the *Weak Nullstellensatz*.)

1.5.6 Assuming the Nullstellensatz, calculate $\mathcal{I}(X)$ for the following varieties X:

(a) $X = \mathcal{V}(x^2 - y^3, x^2 + y^3) \subseteq \mathbb{A}^2_{\mathbb{C}}$;

(b) $X = \mathcal{V}(x) \cup \mathcal{V}(y - z) \subseteq \mathbb{A}^3_{\mathbb{C}}$;

(c) $X = \mathcal{V}((x^2y^2 + yz^2 + x^3z^2)(x^2 + y^2 + 1)) \subseteq \mathbb{A}^3_{\mathbb{C}}$.

1.5.7 Let K be algebraically closed. For each $a \in K$, let

$$X_a := \mathcal{V}(y - x^2, y - a) \subseteq \mathbb{A}^2_K.$$

(a) For what values of a do we have $\mathcal{I}(X_a) = \langle y - x^2, y - a\rangle$? When this is not the case, what is $\mathcal{I}(X_a)$?

(b) Draw a picture, over the (not algebraically closed) field $K = \mathbb{R}$, of the affine varieties X_a for several representative values of K. Can you explain, geometrically, the difference between the values of a for which the equality in part (a) holds and the values of t for which it does not hold?

1.5.8 Assume $\mathrm{Char}(K) = 2$. Prove that $x^2 + y^2 - 1 = (x + y + 1)^2 \in K[x]$. More generally, assume $\mathrm{Char}(K) = p$ and show that

$$\Big(\sum_{i=1}^{m} f_i\Big)^p = \sum_{i=1}^{m} f_i^p.$$

Chapter 2

Irreducibility of Affine Varieties

LEARNING OBJECTIVES FOR CHAPTER 2

- Investigate inclusions, intersections, and unions of affine varieties.
- Prove that every affine variety can be written as the vanishing set of a finite set of polynomials.
- Learn what it means for an affine variety to be irreducible and how affine varieties decompose into irreducible affine varieties.
- Compute irreducible decompositions in a number of examples.
- Refine the dictionary between radical ideals and affine varieties.

When studying the integers, a key tool is the existence of prime factorizations. There is an analogue when studying polynomials (factorization into irreducibles) or when studying finite abelian groups (decomposition as a direct sum of cyclic groups). These settings all demonstrate the way in which one can understand a class of mathematical objects by specifying the atomic, indecomposable objects as well as how a general object decomposes into its atomic pieces. In this chapter, we apply this philosophy to affine varieties, introducing the notion of an *irreducible* affine variety and describing how every affine variety decomposes uniquely as a finite union of irreducible affine varieties—its *irreducible components*.

In order to get there, it is necessary to lay some preliminary groundwork. First, since our goal is to prove that every affine variety can be written as a *union* of its irreducible components, we need a general understanding of how affine varieties behave with respect to set-theoretic notions like inclusions, intersections, and unions, which we discuss in Section 2.1. Furthermore, just as an integer can be factored into primes by a process of repeated factorization that eventually terminates, we need to be sure that the analogous process for affine varieties cannot produce an infinitely nested chain of smaller varieties. This condition translates to a purely algebraic property of $K[x_1, \ldots, x_n]$—it is a Noetherian ring—which is the topic of Section 2.2. Once the groundwork has been laid, we introduce the notion of irreducibility in Section 2.3 and we prove that every affine variety uniquely decomposes as a finite union of irreducible affine varieties in Section 2.4.

Section 2.1 Inclusions, intersections, and unions

In this section, we discuss the ways in which the $\mathcal{V}$- and $\mathcal{I}$-operators interact with set-theoretic notions, beginning with their behavior with respect to inclusions.

2.1 PROPOSITION *$\mathcal{V}$ and $\mathcal{I}$ are inclusion-reversing*

Let $\mathcal{S}, \mathcal{T} \subseteq K[x_1, \ldots, x_n]$ and $X, Y \subseteq \mathbb{A}^n$ be subsets.

1. If $\mathcal{S} \subseteq \mathcal{T}$, then $\mathcal{V}(\mathcal{S}) \supseteq \mathcal{V}(\mathcal{T})$.
2. If $X \subseteq Y$, then $\mathcal{I}(X) \supseteq \mathcal{I}(Y)$.

Furthermore, if X and Y are affine varieties, then

$$X \subseteq Y \text{ if and only if } \mathcal{I}(X) \supseteq \mathcal{I}(Y).$$

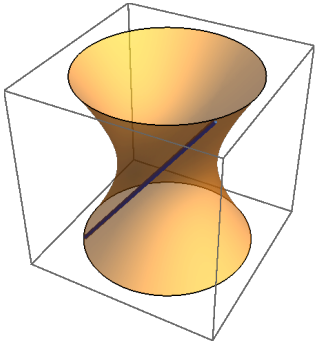
In words, the first item says that a larger set of polynomials has fewer common solutions than a smaller one, while the second item says that a larger set of points in $\mathbb{A}^n$ has fewer polynomials that vanish on it than a smaller one. The reader is encouraged to take a moment to convince themselves of these statements on an intuitive level before attempting a formal proof.

PROOF OF PROPOSITION 2.1 Items 1 and 2 are left to Exercises 2.1.1 and 2.1.2, respectively, where it is also shown that the converse of each of these statements can fail. For the final if-and-only-if statement, the "only-if" direction is the statement of Item 2, so it remains to prove the "if" direction. Assume, then, that $\mathcal{I}(X) \supseteq \mathcal{I}(Y)$, which implies by Item 1 that $\mathcal{V}(\mathcal{I}(X)) \subseteq \mathcal{V}(\mathcal{I}(Y))$. Using the assumption that X and Y are affine varieties, we apply Proposition 1.21 to see that $\mathcal{V}(\mathcal{I}(X)) = X$ and $\mathcal{V}(\mathcal{I}(Y)) = Y$, from which we conclude that $X \subseteq Y$. □

2.2 EXAMPLE A line on a hyperboloid

Consider the ideals

$$I = \langle x^2 + y^2 - z^2 - 1 \rangle \text{ and } J = \langle x - z, y - 1 \rangle.$$



Then $\mathcal{V}(I)$ is the one-sheeted hyperboloid depicted to the right over $\mathbb{R}$, and the variety $\mathcal{V}(J)$ is the line contained on the hyperboloid, whose points are of the form $\{(a, 1, a) \mid a \in K\}$. The containment $\mathcal{V}(I) \supseteq \mathcal{V}(J)$ follows from the containment of ideals $I \subseteq J$, which is verified by noting that the generator of I lies in J:

$$x^2 + y^2 - z^2 - 1 = (x + z)(x - z) + (y + 1)(y - 1) \in J.$$

Having discussed inclusions, we now turn our attention to intersections and unions. Is the intersection or union of a set of affine varieties itself an affine variety? If so, and if we happen to know a set of defining equations for the original collection of varieties, can we find defining equations for the intersection and union? We explore these questions, beginning with a familiar example.

2.3 **EXAMPLE** Intersection and union of coordinate axes

Consider the affine varieties $\mathcal{V}(x) \subseteq \mathbb{A}^2$ and $\mathcal{V}(y) \subseteq \mathbb{A}^2$, which are the y-axis and the x-axis, respectively. The intersection of the two coordinate axes is simply the origin, which we saw in Example 1.6 is defined by the vanishing of the set $\{x, y\}$. Thus,

$$\mathcal{V}(x) \cap \mathcal{V}(y) = \mathcal{V}(x, y). \tag{2.4}$$

Their union, on the other hand, is the affine variety of Example 1.5, which is defined by the vanishing of the polynomial xy. Thus,

$$\mathcal{V}(x) \cup \mathcal{V}(y) = \mathcal{V}(xy). \tag{2.5}$$

If we interpret Equations (2.4) and (2.5) in terms of ideals, then the ideal $\langle xy \rangle$ appearing in Equation (2.5) is the intersection of the ideals $\langle x \rangle$ and $\langle y \rangle$. The ideal $\langle x, y \rangle$ in Equation (2.4) is not quite the union of $\langle x \rangle$ and $\langle y \rangle$—since this union is not an ideal—but is the ideal generated by the union (Exercise 2.1.3). In this way, Example 2.3 illustrates the following general result.

2.6 **PROPOSITION** *Intersections and unions of vanishing sets*

For any ideals $I, J \subseteq K[x_1, \ldots, x_n]$,

$$\mathcal{V}(I) \cap \mathcal{V}(J) = \mathcal{V}(I \cup J),$$

$$\mathcal{V}(I) \cup \mathcal{V}(J) = \mathcal{V}(I \cap J).$$

PROOF We prove the second equality and leave the first to Exercise 2.1.4.

($\subseteq$): Since $I \cap J \subseteq I$, Proposition 2.1 implies that $\mathcal{V}(I) \subseteq \mathcal{V}(I \cap J)$. By the same token, we have $\mathcal{V}(J) \subseteq \mathcal{V}(I \cap J)$. Taken together, we conclude that

$$\mathcal{V}(I) \cup \mathcal{V}(J) \subseteq \mathcal{V}(I \cap J).$$

($\supseteq$): Suppose $a = (a_1, \ldots, a_n) \notin \mathcal{V}(I) \cup \mathcal{V}(J)$. Since $a \notin \mathcal{V}(I)$, there exists $f \in I$ such that $f(a) \neq 0$. Similarly, there exists $g \in J$ such that $g(a) \neq 0$. Because ideals absorb multiplication, the product fg lies in I and J, so $fg \in I \cap J$. Since

$$(fg)(a) = f(a)g(a) \neq 0,$$

we conclude that $a \notin \mathcal{V}(I \cap J)$, completing the proof. □

Since every affine variety X is the vanishing set of some ideal (namely, the ideal $\mathcal{I}(X)$), Proposition 2.6 implies that the intersection and union of any two affine varieties is, itself, an affine variety. Each of the equations in Proposition 2.6 has a downside, however. In the first equation, the issue is that the union $I \cup J$ of ideals is not, in general, an ideal (Exercise 2.1.5). In the second equation, although $I \cap J$ is an ideal, it can be inconvenient to work with in practice; for example, if one knows generators for I and J, it is not obvious how to deduce generators for $I \cap J$.

Both of these issues can be rectified by rephrasing Proposition 2.6 in terms of the following pair of algebraic operations on ideals.

2.7 **DEFINITION** *Sums and products of ideals*

Let I and J be ideals in a ring R. The *sum* of I and J is the ideal

$$I + J = \{r + s \mid r \in I, s \in J\},$$

and the *product* of I and J is the ideal

$$I \cdot J = \left\{ \sum_{i=1}^{m} r_i s_i \;\middle|\; r_i \in I, s_i \in J \right\}.$$

The definition of the sum of two ideals is what you might expect: it is the set consisting of pairwise sums, which happens to be an ideal. The product, however, requires an additional step: since the set of pairwise products is not closed under addition, one needs to include all finite sums of pairwise products in order to obtain an ideal. The verification that the sum and product of two ideals are, in fact, ideals is left to Exercise 2.1.7, where a number of other useful properties are developed.

An important aspect of working with sums and products of ideals is that, if we have generators for I and J, say $I = \langle a_1, \ldots, a_k \rangle$ and $J = \langle b_1, \ldots, b_\ell \rangle$, then we can immediately write down generators for the sum and product ideals (Exercise 2.1.7):

$$I + J = \langle a_1, \ldots, a_k, b_1, \ldots, b_\ell \rangle,$$
$$I \cdot J = \langle a_i b_j \mid i = 1, \ldots, k \text{ and } j = 1, \ldots, \ell \rangle.$$

For example, in the ring $K[x, y]$, we have

$$\langle x \rangle + \langle y \rangle = \langle x, y \rangle,$$
$$\langle x \rangle \cdot \langle y \rangle = \langle xy \rangle.$$

Utilizing sums and products, we have the following modification of Proposition 2.6.

2.8 **PROPOSITION** *Intersections and unions revisited*

For any ideals $I, J \subseteq K[x_1, \ldots, x_n]$,

$$\mathcal{V}(I) \cap \mathcal{V}(J) = \mathcal{V}(I + J),$$
$$\mathcal{V}(I) \cup \mathcal{V}(J) = \mathcal{V}(I \cdot J).$$

PROOF For the first equation, we need only observe that $I + J$ is the ideal generated by $I \cup J$ (Exercise 2.1.7), and then the first equation of Proposition 2.8 follows from the first equation of Proposition 2.6.

To prove the second equation, notice that $I \cdot J$ is contained in both I and J (Exercise 2.1.7). Therefore, the proof that $\mathcal{V}(I) \cup \mathcal{V}(J) \subseteq \mathcal{V}(I \cdot J)$ carries over verbatim from Proposition 2.6. Similarly, because $fg \in I \cdot J$ whenever $f \in I$ and $g \in J$, the proof that $\mathcal{V}(I) \cup \mathcal{V}(J) \supseteq \mathcal{V}(I \cdot J)$ also applies unchanged. □

2.9 **EXAMPLE** Intersection and union via ideals

Consider the affine varieties $\mathcal{V}(x,y)$ and $\mathcal{V}(x-y)$ in $\mathbb{A}^2$, which are the origin and a line through the origin. This example computes their intersection and union algebraically, verifying what one would expect.

Applying Proposition 2.8 and the description of the ideal sum in terms of generators, we have

$$\mathcal{V}(x,y) \cap \mathcal{V}(x-y) = \mathcal{V}(\langle x,y\rangle + \langle x-y\rangle) = \mathcal{V}(x,y,x-y).$$

Note that $\langle x,y,x-y\rangle = \langle x,y\rangle$, since $x-y$ is already in the ideal that x and y generate. Thus, the above can be expressed as

$$\mathcal{V}(x,y) \cap \mathcal{V}(x-y) = \mathcal{V}(x,y),$$

which captures the geometric observation that the intersection of these two affine varieties is the origin.

As for their union, Proposition 2.8 implies that

$$\mathcal{V}(x,y) \cup \mathcal{V}(x-y) = \mathcal{V}(\langle x,y\rangle \cdot \langle x-y\rangle) = \mathcal{V}(x(x-y),y(x-y)).$$

An element $(a,b) \in \mathcal{V}(x(x-y),y(x-y))$ must satisfy the equations

$$a(a-b) = 0 \quad \text{and} \quad b(a-b) = 0.$$

The first of these implies that either $a = 0$ or $a = b$. In case $a = 0$, the second equation implies that $b^2 = 0$ and hence $b = 0$, and in case $a = b$, the second equation is automatically satisfied. In this way, one confirms that

$$\mathcal{V}(x^2 - xy, xy - y^2) = \{(a,b) \in \mathbb{A}^2 \mid a = b\}.$$

In other words, we have verified algebraically that the union of the origin and the line $y = x$ is, as expected, just the line.

Thus far, we have considered only pairwise unions and intersections, but the astute reader may realize that everything generalizes to unions and intersections of finitely many affine varieties $\mathcal{V}(I_1),\ldots,\mathcal{V}(I_k)$. In fact, intersections can be pushed even further, to collections of infinitely many affine varieties $\mathcal{V}(I_1),\mathcal{V}(I_2),\mathcal{V}(I_3),\ldots$ or even collections of uncountably many affine varieties.

Notationally, in order to speak of arbitrary collections of ideals, we consider sets $\{I_\alpha\}_{\alpha\in A}$, where A is an arbitrary set (the *indexing set*) and $I_\alpha \subseteq K[x_1,\ldots,x_n]$ is an ideal for each $\alpha \in A$. For example, if $A = \{1,2,3\}$, this would be a collection $\{I_1,I_2,I_3\}$. If $A = \mathbb{N}$, it would be a collection $\{I_0,I_1,I_2,\ldots\}$ of countably-infinitely many ideals. We could even have $A = \mathbb{R}$, meaning the collection contains not just ideals $I_0,I_1,I_2,\ldots$ but also ideals I_{-1}, $I_{1/2}$, $I_{\sqrt{2}}$, I_π, and so on.

With this notation established, the general result is the following.

2.10 **PROPOSITION** *General intersections and unions*

For any collection $\{I_\alpha\}_{\alpha\in A}$ of ideals $I_\alpha \subseteq K[x_1,\dots,x_n]$,

$$\bigcap_{\alpha\in A} \mathcal{V}(I_\alpha) = \mathcal{V}\Big(\bigcup_{\alpha\in A} I_\alpha\Big).$$

For any finite collection $\{I_1,\dots,I_k\}$ of ideals $I_i \subseteq K[x_1,\dots,x_n]$,

$$\bigcup_{i=1}^{k} \mathcal{V}(I_i) = \mathcal{V}\Big(\bigcap_{i=1}^{k} I_i\Big).$$

Proposition 2.10 can also be stated in terms of ideal sums and products, but a bit of care must be taken in defining infinite sums of ideals.

PROOF The proof mimics the proof of Proposition 2.6. The reason finiteness is required in the second equation is that the product fg that appears in the proof of Proposition 2.6 is replaced here by a product of one f_i from each I_i, and infinite products of polynomials are not polynomials. □

Finiteness is essential in order for the union of affine varieties to be an affine variety. For example, $\mathbb{Z} \subseteq \mathbb{C} = \mathbb{A}^1_{\mathbb{C}}$ is an infinite union of its points, each of which is an affine variety, but we know that $\mathbb{Z} \subseteq \mathbb{A}^1_{\mathbb{C}}$ is not an affine variety, because it is an infinite proper subset.

Proposition 2.10 implies that arbitrary intersections and finite unions of affine varieties are affine varieties. Readers familiar with topology may recognize these conditions: along with the property that $\emptyset$ and $\mathbb{A}^n$ are affine varieties, these form the defining conditions on the closed sets of a topology, so their complements form the open sets. This topology on $\mathbb{A}^n$ is called the *Zariski topology*, named in honor of Oscar Zariski (1899–1986), who made foundational contributions to modern algebraic geometry by placing the classical Italian approach, in which he was trained, on a more rigorous algebraic footing.

Though familiarity with topology will not be assumed in this book, the terminology of Zariski open and closed sets permeates throughout algebraic geometry, so we present the definition here for future reference.

2.11 **DEFINITION** *Zariski topology on* $\mathbb{A}^n$

A subset $X \subseteq \mathbb{A}^n$ is called *Zariski-closed* if X is an affine variety.

A subset $U \subseteq \mathbb{A}^n$ is called *Zariski-open* if $\mathbb{A}^n \setminus U$ is an affine variety.

The adjective Zariski *distinguishes this topology from other natural topologies on* $\mathbb{A}^n_K$*, such as the Euclidean topology for* $K = \mathbb{R}$ *or* $\mathbb{C}$.

The interested reader with a background in topology is directed to Exercise 2.1.9 to explore some basic properties of the Zariski topology and how it compares to the Euclidean topology.

Exercises for Section 2.1

2.1.1 Let $\mathcal{S}, \mathcal{T} \subseteq K[x_1, \ldots, x_n]$ be subsets.

(a) Prove that $\mathcal{S} \subseteq \mathcal{T}$ implies that $\mathcal{V}(\mathcal{S}) \supseteq \mathcal{V}(\mathcal{T})$.

(b) Prove, by example, that the converse of (a) can fail.

2.1.2 Let $X, Y \subseteq \mathbb{A}^n$ be subsets.

(a) Prove that $X \subseteq Y$ implies that $\mathcal{I}(X) \supseteq \mathcal{I}(Y)$.

(b) Prove, by example, that the converse of (a) can fail.

2.1.3 This exercise concerns the ideals $\langle x \rangle \subseteq K[x, y]$ and $\langle y \rangle \subseteq K[x, y]$.

(a) Prove that $\langle x \rangle \cap \langle y \rangle = \langle xy \rangle$.

(b) Prove that $\big\langle \langle x \rangle \cup \langle y \rangle \big\rangle = \langle x, y \rangle$.

2.1.4 Complete the proof of Proposition 2.6 by proving that

$$\mathcal{V}(I) \cap \mathcal{V}(J) = \mathcal{V}(I \cup J)$$

for any ideals $I, J \subseteq K[x_1, \ldots, x_n]$.

2.1.5 Prove that a union of two ideals in a ring is an ideal if and only if one of the ideals is contained in the other.

2.1.6 Let I and J be ideals of a ring R. Prove, by example, that $\{ab \mid a \in I,\ b \in J\}$ is not necessarily an ideal of R.

2.1.7 Let I and J be ideals of a ring R.

(a) Prove that $I + J$ and $I \cdot J$ are both ideals.

(b) Suppose that $I = \langle a_1, \ldots, a_k \rangle$ and $J = \langle b_1, \ldots, b_\ell \rangle$. Prove that

$$I + J = \langle a_1, \ldots, a_k, b_1, \ldots, b_\ell \rangle$$

and

$$I \cdot J = \big\langle \{a_i b_j \mid i = 1, \ldots, k,\ j = 1, \ldots, \ell\} \big\rangle .$$

(c) Prove that $I + J$ is the ideal generated by $I \cup J$.

(d) Prove that

$$I \cdot J \subseteq I \cap J.$$

2.1.8 Assume that K is infinite. Prove that any two nonempty Zariski-open sets in $\mathbb{A}^n_K$ have nonempty intersection. (For students with some background in topology, this says that the Zariski topology on $\mathbb{A}^n_K$ is not Hausdorff.)

2.1.9 (For students with some background in topology) Compare the Zariski topology on $\mathbb{A}^n_{\mathbb{R}} = \mathbb{R}^n$ to the Euclidean topology (induced by the Euclidean metric). Is one of these topologies coarser than the other?

Section 2.2 Finite generation

The notion of a vanishing set $\mathcal{V}(\mathcal{S})$ makes sense whether $\mathcal{S}$ is finite or infinite, but often an infinite set can be replaced by a finite one without affecting the corresponding vanishing set. For example, the ideal $\langle y - x^2 \rangle \subseteq K[x,y]$ contains infinitely many polynomials, but $\mathcal{V}(\langle y - x^2 \rangle)$ is equal to $\mathcal{V}(y - x^2)$, the vanishing set of just a single polynomial.

A natural question, then, is whether any affine variety can be expressed as $\mathcal{V}(\mathcal{S})$ for a *finite* set $\mathcal{S}$. The answer to this question is yes, and the algebraic proof of this fact is the goal of this section. We begin with a bit of algebraic terminology.

2.12 **DEFINITION** *Finitely-generated ideals*

An ideal I of a ring R is said to be *finitely-generated* if $I = \langle r_1, \ldots, r_k \rangle$ for finitely many elements $r_1, \ldots, r_k \in R$.

The ideal $\langle y - x^2 \rangle \subseteq K[x,y]$, for example, is finitely-generated, as is the ideal $\langle x, y \rangle \subseteq K[x,y]$. In fact, one must look to a ring that is rather less familiar to find an example of an ideal that is not finitely-generated.

2.13 **EXAMPLE** An ideal that is not finitely-generated

Let $R = K[x_1, x_2, x_3, \ldots]$ be a polynomial ring in infinitely many variables. Explicitly, monomials in this ring are expressions of the form

$$x^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3} \cdots,$$

where $\alpha = (\alpha_1, \alpha_2, \alpha_3, \ldots)$ is an exponent vector satisfying

(i) $\alpha_i \in \mathbb{N}$ for all i,

(ii) $\alpha_i = 0$ for all but finitely many i.

An element of R is defined to be a finite K-linear combination of monomials:

$$f = \sum_\alpha a_\alpha x^\alpha,$$

where the sum is over all exponent vectors α satisfying (i) and (ii), the coefficients a_α are elements of K, and $a_\alpha = 0$ for all but finitely many α.

In R, the ideal generated by all of the variables,

$$I = \langle x_1, x_2, x_3, \ldots \rangle,$$

is not finitely-generated, as the reader is encouraged to verify (Exercise 2.2.1).

Despite the previous example, many of the rings with which we are familiar have the property that all of their ideals are finitely-generated. These rings are given a special name, in honor of Emmy Noether (1882–1935). In addition to her pioneering work in abstract algebra, Noether also guided the development of modern physics by discovering the connection between symmetries and conservations laws.

2.14 **DEFINITION** *Noetherian ring*

A ring is said to be *Noetherian* if all of its ideals are finitely-generated.

Our goal is to prove that $K[x_1, \ldots, x_n]$ is Noetherian. First, we begin with a few examples that we already know to be Noetherian.

2.15 **EXAMPLE** Fields are Noetherian

If K is any field, then the only ideals of K are $\{0\}$ and K. Both of these are finitely-generated, because $\{0\} = \langle 0 \rangle$ and $K = \langle 1 \rangle$. Thus, K is Noetherian.

2.16 **EXAMPLE** PIDs are Noetherian

By definition, every ideal in a principal ideal domain is generated by a single element, and is thus finitely-generated. Therefore, the rings $\mathbb{Z}$ and $K[x]$ are both examples of Noetherian rings.

There is an alternative way to characterize what it means for a ring to be Noetherian, using nested chains of ideals. Although this second characterization is not as easy to state, it can be very useful in practice.

2.17 **PROPOSITION** *The ascending chain condition*

A ring R is Noetherian if and only if, given any ideals $I_1, I_2, I_3, \ldots$ of R with

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots,$$

there exists a natural number k such that $I_d = I_k$ for all $d \geq k$.

In other words, Proposition 2.17 says that Noetherian rings are characterized by the *ascending chain condition*: every ascending chain of ideals must eventually stabilize. It is not satisfied for the (non-Noetherian) ring $R = K[x_1, x_2, x_3, \ldots]$; for example, the chain of ideals

$$\langle x_1 \rangle \subsetneq \langle x_1, x_2 \rangle \subsetneq \langle x_1, x_2, x_3 \rangle \subsetneq \cdots$$

continues to grow at each step.

PROOF OF PROPOSITION 2.17 We prove both implications.

($\Rightarrow$) Suppose R is Noetherian, and let

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$$

be an ascending chain of ideals of R. Consider the union of these nested ideals

$$I = \bigcup_{k=1}^{\infty} I_k,$$

which is an ideal of R by Exercise 0.4.9. Since R is Noetherian, $I = \langle a_1, \ldots, a_r \rangle$ for some $a_1, \ldots, a_r \in R$. Each a_i is in the union of the I_k, so it must lie in at least

one of them; say $a_i \in I_{k_i}$. Since the ideals are nested, it follows that $a_i \in I_d$ for all $d \geq k_i$. In particular, if we set $k = \max\{k_1, \ldots, k_r\}$, then $\{a_1, \ldots, a_r\} \subseteq I_d$ for all $d \geq k$, implying that $I = \langle a_1, \ldots, a_r \rangle \subseteq I_d$ for all $d \geq k$. However, since I is the union of the I_k, we also have $I \supseteq I_d$, from which we conclude that $I = I_d$ for all $d \geq k$, verifying the ascending chain condition.

($\Leftarrow$) We prove this direction by proving the contrapositive. Suppose R is not Noetherian, and choose an ideal I of R that is not finitely-generated. Choose any element $a_1 \in I$ and define $I_1 = \langle a_1 \rangle$. Then $I_1 \subseteq I$, but since I is not finitely-generated, the inclusion must be strict. Thus, choose an element $a_2 \in I \setminus I_1$ and define $I_2 = \langle a_1, a_2 \rangle$. We now have

$$I_1 \subsetneq I_2 \subsetneq I;$$

the first inclusion is strict because $a_2 \notin I_1$, and the second inclusion is strict because I is not finitely-generated. We can continue this process indefinitely by choosing $a_{n+1} \in I \setminus \langle a_1, \ldots, a_n \rangle$ and defining $I_{n+1} = \langle a_1, \ldots, a_{n+1} \rangle$. This process recursively produces an ascending chain of ideals that never stabilizes, so the ascending chain condition fails. □

The word "basis" is a somewhat outdated artifact: in Hilbert's time, a set of ideal generators was referred to as a basis. This terminology persists today, to some extent—for example, in the term "Gröbner basis"—but is relatively uncommon.

Equipped with the ascending chain characterization of the Noetherian property, our next objective is to show that the polynomial rings $K[x_1, \ldots, x_n]$ are Noetherian. The proof uses induction, adding one variable at a time. The induction step follows from the following key algebraic result that goes by the name of *Hilbert's Basis Theorem*, in honor of David Hilbert (1862–1943), a prolific mathematician who was the first to prove this result in 1890 as part of his work on invariant theory.

2.18 **THEOREM** *Hilbert's Basis Theorem*

If R is a Noetherian ring, then $R[x]$ is a Noetherian ring.

PROOF Suppose that R is a Noetherian ring and, toward a contradiction, suppose that $R[x]$ is not Noetherian. Let $I \subseteq R[x]$ be an ideal that is not finitely-generated. Define an infinite sequence of polynomials in $R[x]$ by the following recursion:

1. Choose $f_1 \in I$ to be a nonzero polynomial of minimum degree.
2. Having chosen $f_1, \ldots, f_j \in I$, choose $f_{j+1} \in I \setminus \langle f_1, \ldots, f_j \rangle$ to be a nonzero polynomial of minimum degree.

It cannot be the case that $\deg(f_j) > \deg(f_{j+1})$ as this would contradict the choice of f_j having minimum degree in $I \setminus \langle f_1, \ldots, f_{j-1} \rangle$. Therefore, the degrees of the polynomials in the sequence $(f_1, f_2, f_3, \ldots)$ are nondecreasing.

For each j, let a_j be the leading coefficient of f_j and consider the following ascending chain of ideals in R:

$$\langle a_1 \rangle \subseteq \langle a_1, a_2 \rangle \subseteq \langle a_1, a_2, a_3 \rangle \subseteq \cdots.$$

Since R is Noetherian, this chain must eventually stabilize; suppose it stabilizes at the kth step. Then $a_{k+1} \in \langle a_1, \ldots, a_k \rangle$. Choose elements $r_1, \ldots, r_k \in R$ such that

$$a_{k+1} = r_1 a_1 + \cdots + r_k a_k.$$

Using the fact that $\deg(f_i) \leq \deg(f_{k+1})$ for all $i < k+1$, define the polynomial

$$g = x^{\deg(f_{k+1})-\deg(f_1)} r_1 f_1 + \cdots + x^{\deg(f_{k+1})-\deg(f_k)} r_k f_k \in R[x].$$

By design, g has the same leading coefficient as f_{k+1}, which implies that

$$\deg(f_{k+1} - g) < \deg(f_{k+1}).$$

Since f_{k+1} and g are both elements of I, it follows that $f_{k+1} - g \in I$. However, since g is an element of $\langle f_1, \ldots, f_k \rangle$ and f_{k+1} is not, it follows that $f_{k+1} - g$ cannot be in $\langle f_1, \ldots, f_k \rangle$. Thus, $f_{k+1} - g \in I \setminus \langle f_1, \ldots, f_k \rangle$ and $\deg(f_{k+1} - g) < \deg(f_{k+1})$, contradicting the minimality of degree in the choice of f_{k+1}. □

This finally brings us to the following fundamental property of polynomial rings.

2.19 COROLLARY *$K[x_1, \ldots, x_n]$ is Noetherian*

For any field K, the polynomial ring $K[x_1, \ldots, x_n]$ is Noetherian.

PROOF The proof is by induction on n.

(Base case) As was noted in Example 2.15, any field K is Noetherian, proving the base case $n = 0$.

(Induction step) Suppose $K[x_1, \ldots, x_{n-1}]$ is Noetherian. Using the canonical isomorphism

$$K[x_1, \ldots, x_n] = K[x_1, \ldots, x_{n-1}][x_n]$$

and applying Hilbert's Basis Theorem for $R = K[x_1, \ldots, x_{n-1}]$, we conclude that $K[x_1, \ldots, x_n]$ is Noetherian. □

The geometric interpretation of the fact that $K[x_1, \ldots, x_n]$ is Noetherian is that any affine variety can be defined by the vanishing of finitely many polynomials. In fact, a somewhat stronger statement is true.

2.20 COROLLARY *Affine varieties are finitely-generated*

If $\mathcal{S} \subseteq K[x_1, \ldots, x_n]$ is any subset, then there is a finite subset $\mathcal{T} \subseteq \mathcal{S}$ such that

$$\mathcal{V}(\mathcal{S}) = \mathcal{V}(\mathcal{T}).$$

PROOF Let $\mathcal{S} \subseteq K[x_1, \ldots, x_n]$ be a (possibly infinite) set and let $I_{\mathcal{S}}$ be the ideal generated by $\mathcal{S}$. By Corollary 2.19, there exist elements $f_1, \ldots, f_k \in K[x_1, \ldots, x_n]$ such that

$$I_{\mathcal{S}} = \langle f_1, \ldots, f_k \rangle.$$

This result is stronger than simply saying that $\mathcal{V}(\mathcal{S})$ can be defined by a finite set of polynomials; it is also asserting that the finite set can be taken to be a subset of $\mathcal{S}$.

The polynomials $f_1, \ldots, f_k$ may not themselves belong to the set $\mathcal{S}$, but by definition of $I_{\mathcal{S}}$, we can write each f_i as

$$f_i = \sum_{j=1}^{\ell_i} g_{i,j} h_{i,j} \tag{2.21}$$

where $g_{i,j} \in K[x_1, \ldots, x_n]$ and $h_{i,j} \in \mathcal{S}$. Define the finite subset

$$\mathcal{T} = \{h_{i,j} \mid 1 \leq i \leq k,\ 1 \leq j \leq \ell_i\} \subseteq \mathcal{S}.$$

Equation (2.21) implies $f_i \in I_{\mathcal{T}}$ for all i, so $I_{\mathcal{S}} = \langle f_1, \ldots, f_k \rangle \subseteq I_{\mathcal{T}}$. Conversely, since $\mathcal{T} \subseteq \mathcal{S}$, we obtain the other inclusion $I_{\mathcal{T}} \subseteq I_{\mathcal{S}}$. Thus, $I_{\mathcal{S}} = I_{\mathcal{T}}$, and we conclude that

$$\mathcal{V}(\mathcal{S}) = \mathcal{V}(I_{\mathcal{S}}) = \mathcal{V}(I_{\mathcal{T}}) = \mathcal{V}(\mathcal{T}). \qquad \square$$

Exercises for Section 2.2

2.2.1 Prove that $\langle x_1, x_2, x_3, \ldots \rangle \subseteq K[x_1, x_2, x_3, \ldots]$ is not finitely-generated.

2.2.2 Because $\mathbb{Z}$ is a principal ideal domain (thus, Noetherian), Proposition 2.17 implies that any ascending chain of ideals $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ in $\mathbb{Z}$ must terminate. Explain this phenomenon concretely: namely, if you express each $I_i = \langle a_i \rangle$ for some integer a_i, what is the relationship between a_i and a_{i+1}? Why must there exist a natural number k such that $a_k = a_{k+1} = a_{k+2} = \cdots$?

2.2.3 Let R be a Noetherian ring and let $I \subseteq R$ be an ideal. Prove that R/I is Noetherian.

2.2.4 Prove that every Noetherian ring is a factorization domain.

2.2.5 (This exercise shows that a subring of a Noetherian ring need not be Noetherian.) Consider the ring homomorphism

$$\begin{aligned} \varphi : K[x_1, x_2, x_3, \ldots] &\to K[y, z] \\ f(x_1, x_2, x_3, \ldots) &\mapsto f(yz, yz^2, yz^3, \ldots). \end{aligned}$$

Prove that φ is injective and conclude that $K[y, z]$ has a non-Noetherian subring.

2.2.6 Let R be a Noetherian ring and let $\varphi : R \to R$ be a ring homomorphism. Prove that φ is an isomorphism if and only if φ is surjective. (**Hint**: Consider the ideals $I_1 = \ker(\varphi)$, $I_2 = \ker(\varphi \circ \varphi)$, $I_3 = \ker(\varphi \circ \varphi \circ \varphi)$, and so on.)

2.2.7 Suppose

$$\mathbb{A}^n = \bigcup_{\alpha \in A} U_\alpha$$

where each $U_\alpha \subseteq \mathbb{A}^n$ is Zariski open. Prove that there is a finite subset $B \subseteq A$ such that

$$\mathbb{A}^n = \bigcup_{\alpha \in B} U_\alpha.$$

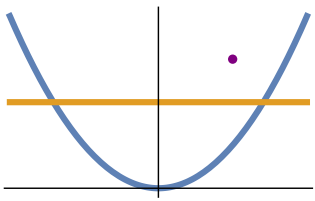
(This shows that the Zariski topology on $\mathbb{A}^n$ is compact.)

Section 2.3 Irreducible affine varieties

We now come to the heart of this chapter and a central concept in algebraic geometry: the notion of irreducibility. To motivate the idea, consider the affine varieties

$$X_1 = \mathcal{V}(y - x^2), \quad X_2 = \mathcal{V}(y - 2), \quad \text{and} \quad X_3 = \mathcal{V}(x - 1, y - 3).$$

Over $\mathbb{R}$, the affine variety $X = X_1 \cup X_2 \cup X_3$ is shown to the right. Imagine now that you were given this image without being told that X was a union of three affine varieties. You could probably still tell, visually, that X was equal to such a union, and by studying the picture carefully you might even be able to determine the varieties. These varieties are the atomic pieces, or irreducible components, of X.



As this discussion suggests, the way we decompose affine varieties into their constituent pieces is by breaking them up into a finite union of smaller affine varieties. As such, the atomic ones are those that cannot be written as a union of two smaller affine varieties. We make this notion precise in the next definition.

2.22 **DEFINITION** *Reducible and irreducible affine variety*

An affine variety $X \subseteq \mathbb{A}^n$ is *reducible* if $X = X_1 \cup X_2$ for some affine varieties $X_1, X_2 \subsetneq X$, and X is *irreducible* if it is neither empty nor reducible.

2.23 **EXAMPLE**

The affine variety $X = \mathcal{V}(x^2 - y^2) \subseteq \mathbb{A}^2$ is reducible. To see this, notice that

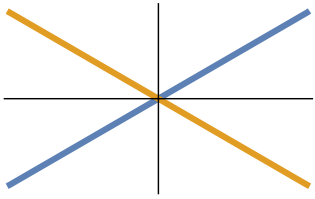
$$X = \mathcal{V}((x+y)(x-y)) = \mathcal{V}(x+y) \cup \mathcal{V}(x-y),$$

where the second equality follows from Proposition 2.8. Therefore, the two affine varieties

$$X_1 = \mathcal{V}(x+y) \subsetneq X$$

and

$$X_2 = \mathcal{V}(x-y) \subsetneq X$$



satisfy $X = X_1 \cup X_2$. Visually, X_1 and X_2 are the two lines that constitute X in the image to the right. In fact, as we will see in the next section, the two varieties X_1 and X_2 are the unique irreducible components of X.

2.24 **EXAMPLE**

By contrast, the parabola $X = \mathcal{V}(y - x^2) \subseteq \mathbb{A}^2$ is irreducible. This assertion should be geometrically believable: unlike the affine variety of Example 2.23, the parabola consists of just a single "piece." While this intuition is not yet a proof, the irreducibility of X will follow from Proposition 2.25 below.

As the above examples illustrate, proving that an affine variety X is reducible is straightforward: one must simply find a pair of affine varieties $X_1 \subsetneq X$ and $X_2 \subsetneq X$ whose union is X. It is less clear how to prove that an affine variety is irreducible. The following algebraic characterization of irreducibility provides a key tool.

2.25 PROPOSITION *Irreducibility algebraically*

An affine variety $X \subseteq \mathbb{A}^n$ is irreducible if and only if $\mathcal{I}(X)$ is a prime ideal.

PROOF We prove both implications by proving their contrapositives.

($\Leftarrow$) Suppose that X is reducible and choose affine varieties $X_1, X_2 \subsetneq X$ such that

$$X = X_1 \cup X_2.$$

Since $X_1 \subsetneq X$, it follows from Proposition 2.1 that $\mathcal{I}(X_1) \supsetneq \mathcal{I}(X)$. Thus, there exists $f \in \mathcal{I}(X_1)$ with $f \notin \mathcal{I}(X)$, and similarly, there exists $g \in \mathcal{I}(X_2)$ with $g \notin \mathcal{I}(X)$. For any $a \in X$, we either have $a \in X_1$ (and hence $f(a) = 0$) or $a \in X_2$ (and hence $g(a) = 0$), proving that $(fg)(a) = f(a)g(a) = 0$, so $fg \in \mathcal{I}(X)$. We have thus argued the existence of a pair of elements $f, g \in K[x_1, \ldots, x_n]$ with

$$f \notin \mathcal{I}(X), \quad g \notin \mathcal{I}(X), \quad \text{and} \quad fg \in \mathcal{I}(X), \tag{2.26}$$

which proves that $\mathcal{I}(X)$ is not prime.

($\Rightarrow$) Suppose that $\mathcal{I}(X)$ is not prime and choose $f, g \in K[x_1, \ldots, x_n]$ satisfying the conditions in (2.26). Define

$$X_1 = \mathcal{V}(f) \cap X \quad \text{and} \quad X_2 = \mathcal{V}(g) \cap X.$$

Proposition 2.6 implies that X_1 and X_2 are both affine varieties, and both are contained in X. Furthermore, the containments must be strict; if $X = X_1$, for example, then $X = \mathcal{V}(f) \cap X$, which means that $X \subseteq \mathcal{V}(f)$. If this were the case, then $f(a) = 0$ for all $a \in X$, meaning that $f \in \mathcal{I}(X)$, contradicting our assumptions.

By construction, we have $X_1 \cup X_2 \subseteq X$, but the other containment also holds. To see this, let $a \in X$. The fact that $fg \in \mathcal{I}(X)$ implies that $fg(a) = f(a)g(a) = 0$. It follows that either $f(a) = 0$ or $g(a) = 0$, implying that either $a \in \mathcal{V}(f)$ or $a \in \mathcal{V}(g)$. Since $a \in X$ by assumption, we conclude that either $a \in \mathcal{V}(f) \cap X = X_1$ or $a \in V(g) \cap X = X_2$, so $a \in X_1 \cup X_2$. We have thus found two affine varieties $X_1, X_2 \subsetneq X$ such that $X = X_1 \cup X_2$, so X is reducible. □

2.27 EXAMPLE The parabola is irreducible

Consider the affine variety $X = \mathcal{V}(y - x^2) \subseteq \mathbb{A}^2$, whose vanishing ideal we computed in Example 1.19 to be $\mathcal{I}(X) = \langle y - x^2 \rangle$. Since $y - x^2$ is irreducible, $\mathcal{I}(X)$ is prime (Proposition 0.62), which proves that X is irreducible.

It is often the case that an affine variety is described in terms of defining equations, or equivalently, an ideal $I = \langle f_1, \ldots, f_k \rangle$. However, this defining ideal may not be equal to the vanishing ideal. This raises the question: given an ideal I, is there a way to determine if $\mathcal{V}(I)$ is irreducible, without a priori knowledge of the vanishing ideal? The Nullstellensatz provides the following useful answer.

2.28 PROPOSITION *Irreducibility of $\mathcal{V}(I)$*

If $I \subseteq K[x_1, \ldots, x_n]$ is a prime ideal, then $\mathcal{V}(I)$ is irreducible. In particular, if $f \in K[x_1, \ldots, x_n]$ is an irreducible polynomial, then $\mathcal{V}(f)$ is irreducible.

PROOF Suppose that I is a prime ideal. Then I is radical by Proposition 1.35. Thus, by the Nullstellensatz,

$$\mathcal{I}(\mathcal{V}(I)) = \sqrt{I} = I.$$

Since $\mathcal{I}(\mathcal{V}(I))$ is prime, we conclude from Proposition 2.25 that $\mathcal{V}(I)$ is irreducible.

To prove the second assertion, assume that f is an irreducible polynomial. Then $\langle f \rangle$ is a prime ideal by Proposition 0.62. Therefore, the first statement in the proposition implies that $\mathcal{V}(f) = \mathcal{V}(\langle f \rangle)$ is irreducible. □

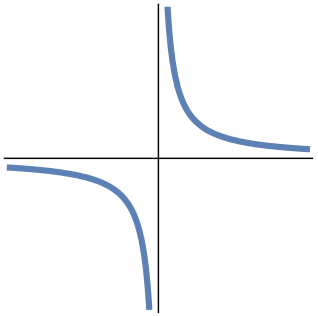
We point out that Proposition 2.28 fails when K is not algebraically closed. For example, $x^2 + 1 \in \mathbb{R}[x]$ is irreducible but

$$\mathcal{V}(x^2 + 1) = \emptyset \in \mathbb{A}^1_{\mathbb{R}}$$

and, by definition, the empty set is not irreducible. For an example of an irreducible polynomial (over $\mathbb{R}$) that defines a *nonempty* reducible variety, see Exercise 2.3.6.

As we have previously mentioned, it is often useful to use our intuition over $\mathbb{R}$ to glean information about varieties more generally. For example, it should be somewhat intuitively clear that the parabola is irreducible over $\mathbb{R}$ because it is comprised of a single "piece," and this intuition extends to more general fields: the variety $\mathcal{V}(y - x^2)$ is irreducible over any infinite field K. However, one should be careful with this sort of reasoning over $\mathbb{R}$, as the next example illustrates.

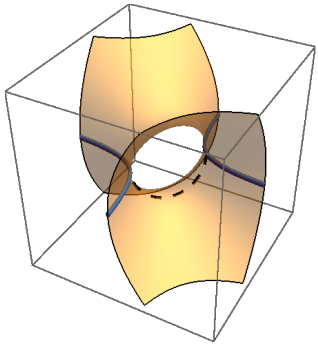
2.29 EXAMPLE The hyperbola is irreducible



Consider the hyperbola $X = \mathcal{V}(xy - 1) \subseteq \mathbb{A}^2$, which is pictured to the right over $\mathbb{R}$. At a glance, our geometric intuition tells us that X consists of two "pieces," one in the first quadrant and one in the third. It would be natural to guess, then, that X is reducible. To the contrary, X is actually irreducible. Over an algebraically closed field, this follows from the fact that $xy - 1$ is irreducible. However, X can also be shown to be irreducible over any infinite field, including $\mathbb{R}$ (Exercises 2.3.7).

How, then, did our intuition fail us in this example? The answer is that the solutions over $\mathbb{R}$ do not capture the entire picture. If we expand our view and consider the zeros of $xy - 1$ over the algebraic closure of $\mathbb{R}$, namely $\mathbb{C}$, then we see that the two "pieces" are actually connected to each other via complex solutions. For example, we can get from the point $(1, 1)$ in the upper piece to the point $(-1, -1)$ in the lower piece by walking along the set of complex solutions

$$\{(e^{\pi i t}, e^{-\pi i t}) \mid t \in [0, 1]\}.$$



Motivated by this observation, we can attempt to draw $\mathcal{V}(xy-1) \subseteq \mathbb{A}^2_{\mathbb{C}}$. This is a bit challenging because $\mathbb{A}^2_{\mathbb{C}}$ is 4-dimensional over the real numbers: $\mathbb{A}^2_{\mathbb{C}} = \mathbb{C}^2 = \mathbb{R}^4$. However, by mapping down to $\mathbb{R}^3$, one can show (Exercise 2.3.8) that the complex solutions can be identified with the surface to the right, where we have included a depiction of the hyperbola and the points connecting $(1,1)$ to $(-1,-1)$. Thus, we see that the complex picture is much more consistent with what we might expect an irreducible variety to look like intuitively: it is comprised of just a single "piece."

This example illustrates that, while it is important in algebraic geometry to use our geometric intuition over $\mathbb{R}$, we may not see the whole picture when we do so, and this intuition should only be trusted insofar as it can be justified algebraically.

A special type of irreducible affine variety is one that consists of a single point: $X = \{a\} \subseteq \mathbb{A}^n$. Since single points are minimal among varieties (with respect to inclusion), then the inclusion-reversing nature of the $\mathcal{V}$- and $\mathcal{I}$-operators suggest that their vanishing ideals should be maximal ideals, which, as we know, are special types of prime ideals. This is true, and even more can be said.

2.30 **PROPOSITION** *Single points and maximal ideals*

Let $I \subseteq K[x_1, \ldots, x_n]$ be an ideal. The following are equivalent:

(i) $I = \mathcal{I}(\{a\})$ for some point $a = (a_1, \ldots, a_n) \in \mathbb{A}^n$;

(ii) $I = \langle x_1 - a_1, \ldots, x_n - a_n \rangle$ for some $a_1, \ldots, a_n \in K$;

(iii) $K[x_1, \ldots, x_n]/I \cong K$;

(iv) I is a maximal ideal.

PROOF That (i) implies (ii) is Exercise 1.3.5. That (ii) implies (iii) is Exercise 0.3.12. That (iii) implies (iv) follows from Proposition 0.38. Thus, it remains to prove that (iv) implies (i). Toward proving the contrapositive, suppose that X is not a single point; we must prove that $\mathcal{I}(X)$ is not maximal. There are two cases to consider: $X = \emptyset$ or X has more than one point. If $X = \emptyset$, then

$$\mathcal{I}(X) = K[x_1, \ldots, x_n],$$

which is not a maximal ideal. If, on the other hand, X has more than one point, let $a \in X$ be any point and notice that

$$\emptyset \subsetneq \{a\} \subsetneq X$$

is a strict containment of affine varieties. Applying Proposition 2.1, we see that

$$K[x_1, \ldots, x_n] \supsetneq \mathcal{I}(\{a\}) \supsetneq \mathcal{I}(X),$$

is a strict containment of ideals, showing that $\mathcal{I}(X)$ is not maximal. □

Notice that the equivalence of (iii) and (iv) is a purely algebraic statement that does not hold over non-algebraically-closed fields; for example, the ideal $\langle x^2+1\rangle$ is maximal in $\mathbb{R}[x]$ even though

$$\mathbb{R}[x]/\langle x^2+1\rangle \cong \mathbb{C} \ncong \mathbb{R}.$$

This observation reflects that $\langle x^2+1\rangle \subseteq \mathbb{R}[x]$ is not a vanishing ideal.

We close this section on irreducibility by describing a refined dictionary between ideals and varieties. As we have already seen, the Nullstellensatz implies that the $\mathcal{V}$-operator is a bijection between radical ideals in $K[x_1,\ldots,x_n]$ and affine varieties in $\mathbb{A}^n$, with inverse given by the $\mathcal{I}$-operator (Corollary 1.41). We now introduce a refinement of this bijection that adds prime and maximal ideals to the mix.

2.31 **PROPOSITION** *Refined dictionary between ideals and varieties*

The $\mathcal{V}$- and $\mathcal{I}$-operators are inverse, inclusion-reversing bijections that translate between the following hierarchies of ideals and varieties:

$$\begin{array}{ccc}
\{\text{radical ideals in } K[x_1,\ldots,x_n]\} & \longleftrightarrow & \{\text{affine varieties in } \mathbb{A}^n\} \\
\cup\mathrel{|} & & \cup\mathrel{|} \\
\{\text{prime ideals in } K[x_1,\ldots,x_n]\} & \longleftrightarrow & \{\text{irreducible varieties in } \mathbb{A}^n\} \\
\cup\mathrel{|} & & \cup\mathrel{|} \\
\{\text{maximal ideals in } K[x_1,\ldots,x_n]\} & \longleftrightarrow & \{\text{points in } \mathbb{A}^n\}.
\end{array}$$

PROOF If I is a radical ideal and $X = \mathcal{V}(I)$, then the Nullstellensatz implies that $I = \mathcal{I}(X)$. Thus, that the bijection between radical ideals and affine varieties is inclusion-reversing is the final statement in Proposition 2.1. To show that $\mathcal{V}$ and $\mathcal{I}$ restrict to a bijection between prime ideals and irreducible varieties, it suffices to observe that $I = \mathcal{I}(X)$ is prime if and only if $X = \mathcal{V}(I)$ is irreducible (Proposition 2.25). To show that $\mathcal{V}$ and $\mathcal{I}$ restrict to a bijection btetween maximal ideals and single points, it suffices to observe that $I = \mathcal{I}(X)$ is maximal if and only if $X = \mathcal{V}(I)$ is a single point (Proposition 2.30). □

Exercises for Section 2.3

2.3.1 Prove that affine space is irreducible over any infinite field.

2.3.2 Prove that an affine variety over a finite field is irreducible if and only if it consists of a single point.

2.3.3 Prove that $\mathcal{V}(xy) \subseteq \mathbb{A}^2$ is reducible.

2.3.4 Let X be an irreducible affine variety, and suppose that

$$X = \bigcup_{i=1}^{r} X_i,$$

where each X_i is an affine variety. Prove that $X = X_i$ for some i.

2.3.5 Prove that the affine variety $\mathcal{V}(x^2+y^2) \subseteq \mathbb{A}^2$ is irreducible over $\mathbb{R}$ but reducible over $\mathbb{C}$.

2.3.6 Consider the function $f = (x^2+1)(x^2-1)^2 + y^2 \in \mathbb{R}[x,y]$.

(a) Use Eisenstein's criterion to prove that f is irreducible.

(b) Prove that $\mathcal{V}(f)$ is reducible, consisting of two distinct points.

2.3.7 Let K be an infinite field. This exercise proves the irreducibility of

$$X = \mathcal{V}(xy-1) \subseteq \mathbb{A}^2.$$

(a) Use properties of $\mathcal{I}$ and $\mathcal{V}$ to prove that $\mathcal{I}(X) \supseteq \langle xy-1 \rangle$.

(b) Prove that $\mathcal{I}(X) \subseteq \langle xy-1 \rangle$, possibly using the following proof outline.

i. Let $f \in \mathcal{I}(X)$. Prove that

$$y^k f - g \in \langle xy-1 \rangle$$

for some $k \in \mathbb{N}$ and $g \in K[x]$.

ii. Using that $f \in \mathcal{I}(X)$, prove that g is the zero polynomial.

iii. Using that $y^k f \in \langle xy-1 \rangle$, prove that $f \in \langle xy-1 \rangle$.

(c) Prove that $\langle xy-1 \rangle$ is a prime ideal.

2.3.8 The surface pictured in Example 2.3.8 is the real affine variety

$$Y = \mathcal{V}(uv-w^2) \subseteq \mathbb{A}^3_{\mathbb{R}}.$$

Let $X = \mathcal{V}(xy-1) \subseteq \mathbb{A}^2_{\mathbb{C}}$.

(a) Prove that $X = \{(re^{i\theta}, \frac{1}{r}e^{-i\theta}) \mid r \in \mathbb{R}_{>0},\ 0 \leq \theta < 2\pi\}$.

(b) Prove that the function $F : X \to \mathbb{A}^3_{\mathbb{R}}$ defined by taking $(re^{i\theta}, \frac{1}{r}e^{-i\theta})$ to

$$\left(\frac{r^2-1}{2r} + \frac{r^2+1}{2r}\cos\theta, \frac{1-r^2}{2r} + \frac{r^2+1}{2r}\cos\theta, \frac{r^2+1}{2r}\sin\theta \right)$$

is a bijection onto Y.

Section 2.4 Irreducible decompositions

In the previous section, we were introduced to the notion of irreducibility for affine varieties. In this section, we prove the fundamental fact that every affine variety is the finite union of a unique set of irreducible affine varieties. This is one of the most important consequences of the algebraic fact that $K[x_1, \ldots, x_n]$ is Noetherian.

If X is reducible, then we can write X as a union of affine varieties $X_1, X_2 \subsetneq X$, which may themselves be reducible. By further decomposing X_1 and X_2, one can split X into a union comprised of more and more affine varieties, stopping only when the constituent pieces are irreducible. This process is analogous to the way in which one gradually factors an integer into primes or factors a polynomial into irreducibles. The following proposition says that, just like for prime factorizations of integers or irreducible factorizations of polynomials, this process eventually terminates, and the decomposition obtained in this way is unique.

2.32 **PROPOSITION/DEFINITION** *Irreducible decomposition*

Let $X \subseteq \mathbb{A}^n$ be a nonempty affine variety. Then there exist irreducible affine varieties $X_1, \ldots, X_r \subseteq X$ such that $X_i \not\subseteq X_j$ for any $i \neq j$ and

$$X = \bigcup_{i=1}^{r} X_i. \tag{2.33}$$

Moreover, the affine varieties $X_1, \ldots, X_r$ are unique up to reordering; we call these the *irreducible components* of X, and refer to (2.33) as the *irreducible decomposition* of X.

We should stress here that both the finiteness of the number of irreducible components and the fact that $X_i \not\subseteq X_j$ for all $i \neq j$ are crucial features in order for the irreducible decomposition to be unique. To see why, consider the parabola $X = \mathcal{V}(y - x^2) \subseteq \mathbb{A}^2$. Because X is already irreducible, its irreducible decomposition has just a single component. On the other hand, if we did not insist on the finiteness of the number of X_i, then expressing X as the union of all of its points,

$$X = \bigcup_{p \in X} \{p\},$$

would be a different irreducible decomposition. If we did not insist that $X_i \not\subseteq X_j$ for all $i \neq j$, then we could obtain a different irreducible decomposition as, for example,

$$X = \{(0,0)\} \cup X.$$

PROOF OF PROPOSITION 2.32 We must prove existence and uniqueness.

(**Existence**) Suppose, toward a contradiction, that $X \subseteq \mathbb{A}^n$ is a nonempty affine variety that does not have a finite irreducible decomposition. In particular, this implies that X is not irreducible, so write

$$X = X_1 \cup X_1'$$

where $X_1, X_1' \subsetneq X$. If both X_1 and X_1' have finite irreducible decompositions, then the union of these would be a finite irreducible decomposition of X, which goes against our supposition. Thus, it must be the case that either X_1 or X_1' does not have a finite irreducible decomposition. Without loss of generality, suppose X_1 does not have a finite irreducible decomposition, and write

$$X_1 = X_2 \cup X_2'$$

where $X_2, X_2' \subsetneq X$. Again, since X_1 does not have a finite irreducible decomposition, then at least one of X_2 or X_2' does have a finite irreducible decomposition. Suppose that X_2 does not have a finite irreducible decomposition, and write

$$X_2 = X_3 \cup X_3'$$

where $X_3, X_3' \subsetneq X$. Continuing in this way, we construct an infinite chain of nested affine varieties in $\mathbb{A}^n$:

$$X \supsetneq X_1 \supsetneq X_2 \supsetneq X_3 \supsetneq \cdots.$$

By Proposition 2.1, this yields an infinite chain of nested ideals in $K[x_1, \ldots, x_n]$:

$$\mathcal{I}(X) \subsetneq \mathcal{I}(X_1) \subsetneq \mathcal{I}(X_2) \subsetneq \mathcal{I}(X_3) \subsetneq \cdots,$$

which contradicts that $K[x_1, \ldots, x_n]$ is Noetherian (Corollary 2.19). The contradiction implies that every affine variety $X \subseteq \mathbb{A}^n$ must have at least one finite irreducible decomposition.

(**Uniqueness**) Let

$$X = \bigcup_{i=1}^{r} X_i = \bigcup_{j=1}^{s} Y_j$$

be two irreducible decompositions of X. We must prove that $r = s$ and that, after possibly reordering, we have $X_i = Y_i$ for each i. Without loss of generality, assume that $r \geq s$.

Since $X_1 \subseteq X$, we have

$$X_1 = X_1 \cap X = X_1 \cap \left(\bigcup_{j=1}^{s} Y_j \right) = \bigcup_{j=1}^{s} (X_1 \cap Y_j).$$

Using that X_1 is irreducible, it follows (Exercise 2.3.4) that $X_1 = X_1 \cap Y_j$ for some j. Reordering $Y_1, \ldots, Y_s$, assume that $X_1 = X_1 \cap Y_1$, which implies that $X_1 \subseteq Y_1$.

By the same token, since $Y_1 \subseteq X$, we have

$$Y_1 = Y_1 \cap X = Y_1 \cap \left(\bigcup_{i=1}^{r} X_i \right) = \bigcup_{i=1}^{s} (Y_1 \cap X_i),$$

so $Y_1 = Y_1 \cap X_i$ for some i, and $Y_1 \subseteq X_i$. It follows that $X_1 \subseteq Y_1 \subseteq X_i$. Since, by the definition of an irreducible decomposition, $X_1 \not\subseteq X_i$ for any $i \neq 1$, it must be the case that $i = 1$, and the containment $X_1 \subseteq Y_1 \subseteq X_1$ implies $X_1 = Y_1$.

Repeating this argument with X_2 in place of X_1 shows that $X_2 = Y_j$ for some j. Since $X_2 \neq X_1$, it cannot be the case that $j = 1$. Thus, after reordering $Y_2, \ldots, Y_s$, we may assume that $X_2 = Y_2$. We can continue in this way, showing that $X_i = Y_i$ for each $i \in \{1, \ldots, r\}$. In particular, this proves that $r \leq s$. Since we assumed that $r \geq s$, we conclude that $r = s$ and $X_i = Y_i$ for all i. □

2.34 **EXAMPLE** Irreducible components of an intersection

Consider the affine variety

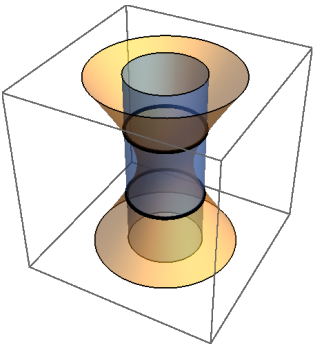
$$X = \mathcal{V}(2x^2 + 2y^2 - z^2 - 1, x^2 + y^2 - 1) \subseteq \mathbb{A}^3.$$

Notice that $X = Y_1 \cap Y_2$ where

$$Y_1 = \mathcal{V}(2x^2 + 2y^2 - z^2 - 1)$$

and

$$Y_2 = \mathcal{V}(x^2 + y^2 - 1).$$



To gain some intuition for the variety X, let us consider the picture for $K = \mathbb{R}$. In that case, Y_1 is a one-sheeted hyperboloid and Y_2 is a circular cylinder, depicted to the right. From this image, we can see that their intersection, X, consists of two circles. One might naturally guess that these two circles are the irreducible components of X, so let us take this intuition and verify it algebraically.

If $(a, b, c) \in X$, then the coordinates satisfy

$$2a^2 + 2b^2 - c^2 = 1 \quad \text{and} \quad a^2 + b^2 = 1.$$

Subtracting twice the second equation from the first, we see that these equations are satisfied if and only if

$$c^2 = 1 \quad \text{and} \quad a^2 + b^2 = 1.$$

Since $c^2 = 1$ if and only if $c = \pm 1$, we can then see that

$$\begin{aligned} X &= \{(a, b, 1) \mid a^2 + b^2 = 1\} \cup \{(a, b, -1) \mid a^2 + b^2 = 1\}. \\ &= \underbrace{\mathcal{V}(x^2 + y^2 - 1, z - 1)}_{X_1} \cup \underbrace{\mathcal{V}(x^2 + y^2 - 1, z + 1)}_{X_2} \end{aligned}$$

Over the real numbers, X_1 and X_2 are precisely the circles depicted above.

To prove that X_1 and X_2 are, in fact, the irreducible components of X, it remains to prove that they are each irreducible. While this fact can be proved over $\mathbb{R}$, it is much simpler to prove over an algebraically closed field, like $\mathbb{C}$. In particular, over $\mathbb{C}$, it suffices (by Proposition 2.28) to observe that

$$\langle x^2 + y^2 - 1, z \pm 1 \rangle \subseteq \mathbb{C}[x, y, z]$$

is a prime ideal, which the reader is encouraged to verify (see Exercises 2.4.2).

In general, it is not easy to determine the irreducible components of a variety, especially if the variety is described by many polynomials in a lot of variables, making it impossible to draw a picture and use our geometric intuition. However, in the special case that the variety is defined by a single polynomial f, the irreducible decomposition of $\mathcal{V}(f)$ is closely related to the irreducible factorization of f.

2.35 PROPOSITION *Irreducible decomposition of $\mathcal{V}(f)$*

If $f \in K[x_1, \ldots, x_n]$ has distinct irreducible factors $q_1, \ldots, q_m$, then the irreducible decomposition of $\mathcal{V}(f)$ is

$$\mathcal{V}(f) = \mathcal{V}(q_1) \cup \cdots \cup \mathcal{V}(q_m).$$

PROOF From the Nullstellensatz, it follows that

$$\mathcal{V}(f) = \mathcal{V}\Big(\sqrt{\langle f \rangle}\Big).$$

Applying Proposition 1.31, we see that $\sqrt{\langle f \rangle} = \langle q_1 \cdots q_m \rangle$, so

$$\mathcal{V}(f) = \mathcal{V}(q_1 \cdots q_m).$$

Since $q_1 \cdots q_m$ vanishes at a point if and only if one of the q_i vanishes at that point, we have

$$\mathcal{V}(f) = \mathcal{V}(q_1) \cup \cdots \cup \mathcal{V}(q_m). \tag{2.36}$$

Since each q_i is irreducible, Proposition 2.28 implies that each $\mathcal{V}(q_i)$ is irreducible.

To finish the proof, we must verify that $\mathcal{V}(q_i) \not\subseteq \mathcal{V}(q_j)$ for any $i \neq j$. By definition of distinct irreducible factors, we know that $q_i \nmid q_j$ for any $i \neq j$. This implies that $\langle q_i \rangle \not\supseteq \langle q_j \rangle$ for any $i \neq j$. Since each q_i is irreducible, then $\langle q_i \rangle$ is a prime ideal, and the Nullstellensatz implies that

$$\mathcal{I}(\mathcal{V}(q_i)) = \mathcal{I}(\mathcal{V}(\langle q_i \rangle)) = \sqrt{\langle q_i \rangle} = \langle q_i \rangle.$$

Thus, since $\mathcal{I}(\mathcal{V}(q_i)) \not\supseteq \mathcal{I}(\mathcal{V}(q_j))$ for any $i \neq j$, the final statement in Proposition 2.1 implies that $\mathcal{V}(q_i) \not\subseteq \mathcal{V}(q_j)$ for any $i \neq j$. □

Not every ideal in $K[x_1, \ldots, x_n]$ is generated by a single polynomial, and one might naturally wonder how to compute an irreducible decomposition of $\mathcal{V}(I)$ for nonprincipal ideals I. While this is a difficult task to do by hand, it is accomplishable with the aid of a computer. In particular, given a set of generators $I = \langle f_1, \ldots, f_k \rangle$, there are effective algorithms utilizing Gröbner bases for computing the irreducible decomposition of $\mathcal{V}(I)$. We will not describe these computational tools, focusing instead on the more theoretical aspects of the dictionary between algebra in geometry.

For an introductory treatment of algebraic geometry that focuses on the more computational and algorithmic aspects of the theory, see the book of Cox, Little, and O'Shea [].

Exercises for Section 2.4

2.4.1 Let K be an infinite field of characteristic not equal to 2. Prove that the irreducible decomposition of $\mathcal{V}(x^2 - y^2) \subseteq \mathbb{A}^2$ is

$$\mathcal{V}(x^2 - y^2) = \mathcal{V}(x + y) \cup \mathcal{V}(x - y).$$

What changes if the characteristic is equal to 2?

2.4.2 Prove that

$$\frac{\mathbb{C}[x,y,z]}{\langle x^2+y^2-1, z-1\rangle} \cong \frac{\mathbb{C}[x,y]}{\langle x^2+y^2-1\rangle}.$$

Conclude that $\langle x^2+y^2-1, z-1\rangle$ is prime.

2.4.3 Let K be any field and consider a nonconstant polynomial $f \in K[x]$. Describe the irreducible decomposition of $\mathcal{V}(f) \subseteq \mathbb{A}^n$ in terms of the irreducible factorization of f. Which irreducible factors of f are relevant and which are not?

2.4.4 What are the irreducible components of

$$\mathcal{V}(x^2+y^2+z^2-2z, x^2+y^2-z^2) \subseteq \mathbb{A}^3_{\mathbb{R}}?$$

(**Hint:** Graph the surfaces to see how they intersect.)

2.4.5 Calculate the irreducible decomposition of

$$\mathcal{V}(x^2+y^2+z^2-2z,\ x^2+y^2-z^2) \subseteq \mathbb{A}^3_{\mathbb{C}}.$$

(**Hint:** There are three components. How does the complex picture differ from the real picture in the previous exercise?)

2.4.6 Calculate the irreducible decomposition of

$$\mathcal{V}(xy+z,\ x^2-x+y^2+yz) \subseteq \mathbb{A}^3_{\mathbb{C}}.$$

(**Hint:** There are two components.)

Chapter 3

Coordinate Rings

LEARNING OBJECTIVES FOR CHAPTER 3

- Become acquainted with the coordinate ring $K[X]$ of an affine variety X, both in terms of polynomial functions and as a quotient ring.
- Become familiar with K-algebras, and identify finitely-generated K-algebras with quotients of polynomial rings.
- Determine whether a ring is reduced, and using quotients, whether an ideal is radical.
- Characterize coordinate rings algebraically as those rings that are finitely-generated reduced K-algebras.
- Explore, in specific examples, how to find an affine variety X whose coordinate ring is a given finitely-generated reduced K-algebra A.

The work we did in Chapter 1 gives us one key method for moving back and forth between the worlds of algebra and geometry, using the $\mathcal{V}$- and $\mathcal{I}$-operators as inverse bijections between affine varieties in $\mathbb{A}^n$ and radical ideals in $K[x_1, \ldots, x_n]$. But on the algebraic side, ideals play a special role that we have not yet invoked: they are precisely the subsets of $K[x_1, \ldots, x_n]$ by which one can take a quotient to produce a ring. If $X \subseteq \mathbb{A}^n$ is an affine variety, then, how should we interpret the quotient of $K[x_1, \ldots, x_n]$ by $\mathcal{I}(X)$ in terms of the affine variety X?

The answer, as we will see in this chapter, is that this quotient is naturally isomorphic to the *coordinate ring* of X, a ring whose elements are *polynomial functions* from X to the ground field K. Once we define these objects precisely in Section 3.1, we will have a new way to pass from the world of geometry to the world of algebra:

$$\begin{aligned} \{\text{affine varieties}\} &\to \{\text{rings}\} \\ X &\mapsto K[X]. \end{aligned}$$

As always, then, we ask whether this association is a two-way dictionary. It is not, at the outset, because not every ring is $K[X]$ for some X. Our search for an algebraic characterization of the rings that arise as coordinate rings will lead us to define the notion of a *K-algebra*—a special class of rings into which polynomial rings and their quotients fall—and to study their key algebraic properties. The culminating result of the chapter is that precisely when a ring is a *finitely-generated reduced K-algebra* is it the coordinate ring of some affine variety.

Section 3.1 Polynomial functions on affine varieties

We learned in Section 1.1 that each abstract polynomial $f \in K[x_1, \ldots, x_n]$ can be used to define a function $\mathbb{A}^n \to K$, obtained by mapping $(a_1, \ldots, a_n) \in \mathbb{A}^n$ to $f(a_1, \ldots, a_n) \in K$. If $X \subseteq \mathbb{A}^n$ is an affine variety, then we can restrict the domain of such a polynomial function to X, yielding a new function

$$\begin{aligned} f|_X : X &\to K \\ (a_1, \ldots, a_n) &\mapsto f(a_1, \ldots, a_n). \end{aligned}$$

A function $F : X \to K$ that arises in this way is referred to as a *polynomial function*.

3.1 **DEFINITION** *Polynomial function*

Let $X \subseteq \mathbb{A}^n$ be an affine variety. A *polynomial function on* X is a function $F : X \to K$ such that $F = f|_X$ for some $f \in K[x_1, \ldots, x_n]$.

3.2 **EXAMPLE** Polynomial functions on the parabola

Let $X = \mathcal{V}(y - x^2) \subseteq \mathbb{A}^2$. The function

$$\begin{aligned} F : X &\to K \\ (a, b) &\mapsto a + b \end{aligned}$$

is a polynomial function, since

$$F = f|_X \quad \text{where} \quad f = x + y \in K[x, y].$$

Note that $f = x + y$ is not the only polynomial that gives rise to F. For example, since $a^2 = b$ for all $(a, b) \in X$, it follows that $F = g|_X$ where $g = x + x^2$ and $F = h|_X$ where $h = x + 2y - x^2$.

3.3 **EXAMPLE** Coordinate functions

Let $X \subseteq \mathbb{A}^n$ be an affine variety. Then, for each $i \in \{1, \ldots, n\}$, the ith *coordinate function* on X is the function

$$\begin{aligned} C_i : X &\to K \\ (a_1, \ldots, a_n) &\mapsto a_i. \end{aligned}$$

The coordinate functions are polynomial because C_i is the restriction of the function associated to the polynomial $x_i \in K[x_1, \ldots, x_n]$.

3.4 **EXAMPLE** The empty function

If $X = \emptyset \subseteq \mathbb{A}^n$, then there is only one function $F : \emptyset \to K$, the *empty function*. Moreover, upon restricting the domain to the empty set, every function $\mathbb{A}^n \to K$ gives rise to the empty function. In particular, this implies that the empty function is the unique polynomial function on the affine variety $\emptyset \subseteq \mathbb{A}^n$.

It should be clear from this discussion that one can concoct polynomial functions on $X \subseteq \mathbb{A}^n$ simply by choosing any polynomial $f \in K[x_1, \ldots, x_n]$, considering the corresponding function on $\mathbb{A}^n$, and then restricting its domain to X. Why, then, do we define polynomial functions on X in what appears to be the opposite way: starting from F and then searching for an f that restricts to it?

The primary reason we take this approach is that, for a given polynomial function $F : X \to K$, there may be many polynomials $f \in K[x_1, \ldots, x_n]$ such that $F = f|_X$, and we do not wish to view these as different polynomial functions on X. This is already apparent in Example 3.2, where the distinct polynomials $f, g, h \in K[x, y]$ all define the same function $F : X \to K$. It is not the polynomials $f \in K[x_1, \ldots, x_n]$ that are important here; rather, it is the function $F : X \to K$ that we intend to study.

Starting from F has its drawbacks, however, because depending on how the definition of F is presented, it may not be immediately clear whether it is the restriction of a polynomial. The next example illustrates this phenomenon, and serves as a caution against making quick judgments about polynomiality.

3.5 **EXAMPLE** A nonobviously polynomial function

Let $X = \mathcal{V}(xy - 1) \subseteq \mathbb{A}^2$. Since $a \neq 0$ for any $(a, b) \in X$, we can consider the function defined by

$$\begin{aligned} F : X &\to K \\ (a, b) &\mapsto \frac{1}{a}. \end{aligned}$$

The output $\frac{1}{a}$ is not a polynomial in a and b, which may lead one to guess that F is not a polynomial function. However, the fact that $ab - 1 = 0$ for all $(a, b) \in X$ means that $\frac{1}{a} = b$, and therefore, $F = f|_X$ where $f = y \in K[x, y]$.

More generally, any function $F : X \to K$ of the form

$$F(a, b) = \frac{f(a, b)}{a^j b^k}$$

with $f \in K[x, y]$ and $j, k \in \mathbb{N}$ is a polynomial function on X (Exercise 3.1.3).

The set of polynomial functions on X can be endowed with the structure of a ring by adding and multiplying functions in the usual way:

$$\begin{aligned} (F + G)(a_1, \ldots, a_n) &= F(a_1, \ldots, a_n) + G(a_1, \ldots, a_n), \\ (F \cdot G)(a_1, \ldots, a_n) &= F(a_1, \ldots, a_n) \cdot G(a_1, \ldots, a_n). \end{aligned}$$

Thus, starting with an affine variety X, we can produce a ring associated to it. This ring is central to the study of algebraic geometry.

3.6 **DEFINITION** *Coordinate ring*

Let $X \subseteq \mathbb{A}^n$ be an affine variety. The *coordinate ring* of X, denoted $K[X]$, is the ring of all polynomial functions on X.

The additive identity $0 \in K[X]$ is the constant function that takes the value $0 \in K$ for all $a \in X$, and the multiplicative identity is the constant function that takes the value $1 \in K$ for all $a \in X$. These functions arise from the polynomials $0, 1 \in K[x_1, \ldots, x_n]$, respectively: $0 = 0|_X$ and $1 = 1|_X$.

Given an affine variety X, can we compute $K[X]$? In other words, can we identify $K[X]$ with a more familiar ring? The next result provides a step in this direction by presenting the coordinate ring as a quotient.

3.7 **PROPOSITION** *The coordinate ring as a quotient*

If $X \subseteq \mathbb{A}^n$ is an affine variety, then there is a canonical ring isomorphism

$$K[X] = K[x_1, \ldots, x_n]/\mathcal{I}(X).$$

PROOF By the First Isomorphism Theorem, it suffices to find a canonical surjective homomorphism $\varphi : K[x_1, \ldots, x_n] \to K[X]$ whose kernel is $\mathcal{I}(X)$. Define φ by

$$\varphi(f) = f|_X.$$

Noting that $(f+g)|_X = f|_X + g|_X$ and $(f \cdot g)|_X = f|_X \cdot g|_X$, we see that φ is a ring homomorphism. By definition, every polynomial function on X arises from some polynomial $f \in K[x_1, \ldots, x_n]$, so φ is surjective. Finally, $f \in \ker \varphi$ if and only if $f|_X = 0$, which is the same as saying that $f \in \mathcal{I}(X)$. This shows that $\ker \varphi = \mathcal{I}(X)$. □

3.8 **EXAMPLE** Coordinate ring of affine space

Proposition 1.9 says that $K[\mathbb{A}^n] = K[x_1, \ldots, x_n]$ if and only if K is infinite.

3.9 **EXAMPLE** Coordinate ring of the parabola

In Example 3.2, where $X = \mathcal{V}(y - x^2) \subseteq \mathbb{A}^2$, we saw that the three polynomials

$$f = x + y, \quad g = x + x^2, \quad \text{and} \quad h = x - x^2 + 2y$$

all give rise to the same polynomial function $F : X \to K$. This reflects the fact that $[f] = [g] = [h]$ in the quotient ring

$$K[X] = \frac{K[x,y]}{\mathcal{I}(X)} = \frac{K[x,y]}{\langle y - x^2 \rangle},$$

which we can readily check:

$$f - g = y - x^2 \in \langle y - x^2 \rangle \quad \text{and} \quad g - h = 2x^2 - 2y \in \langle y - x^2 \rangle.$$

For an affine variety $X \subseteq \mathbb{A}^n$, Definition 3.6 and Proposition 3.7 provide two different characterizations of the coordinate ring $K[X]$. It is important to keep both interpretations in mind: elements of the coordinate ring should be simultaneously viewed as functions $F : X \to K$ and as equivalence classes of polynomials in

$K[x_1, \ldots, x_n]$. The canonical isomorphism of Proposition 3.7 identifies the polynomial function $f|_X$ with the equivalence class $[f]$ for any $f \in K[x_1, \ldots, x_n]$.

As advertised in the introduction to this chapter, we have now introduced a new way of passing from geometry to algebra, by associating to an affine variety X its coordinate ring $K[X]$. In keeping with our philosophy that the passage from geometry to algebra should be a two-way dictionary, we now ask whether the passage from an affine variety to its coordinate ring can be reversed. More precisely, given a ring R, does there exist an affine variety X such that $R = K[X]$? The answer, we will find, is affirmative if R is a *reduced finitely-generated K-algebra*, and the next three sections are devoted to defining and studying these terms.

Exercises for Section 3.1

3.1.1 Describe the ring of polynomial functions on the empty set. What is 0? 1?

3.1.2 Let $X = \mathcal{V}(x^2 + y^2 - 2z^2) \subseteq \mathbb{A}^3$. List three distinct elements of $K[x, y, z]$ that restrict to the same polynomial function in $K[X]$, and list two elements of $K[x, y, z]$ that restrict to different polynomial functions in $K[X]$.

3.1.3 Let $X = \mathcal{V}(xy - 1) \subseteq \mathbb{A}^2$. Prove that any function $F : X \to K$ of the form

$$F(a, b) = \frac{f(a, b)}{a^j b^k}$$

with $f \in K[x, y]$ and $j, k \in \mathbb{N}$ is a polynomial function.

3.1.4 Let $X \subseteq \mathbb{A}^n$ be an affine variety.

(a) Explain why $K[X]$ is an integral domain if and only if X is irreducible.

(b) As an illustration of part (a), let $X = \mathcal{V}(x^2 - xy) \subseteq \mathbb{A}^2$. Prove that X is not irreducible by finding two affine varieties $X_1 \subsetneq X$ and $X_2 \subsetneq X$ such that $X = X_1 \cup X_2$. Then, verify that $K[X]$ is not an integral domain by finding two nonzero functions in $K[X]$ whose product is zero.

3.1.5 Let $X = \{p\} \subseteq \mathbb{A}^n$ be a single point. Prove that the function $\varphi : K[X] \to K$ defined by $\varphi(F) = F(p)$ is a ring isomorphism.

3.1.6 Let $p_1, \ldots, p_m \in \mathbb{A}^n$ be points in $\mathbb{A}^n$, and let $X = \{p_1, \ldots, p_m\}$. Prove that the function

$$\begin{aligned} \varphi : K[X] &\to \overbrace{K \oplus \cdots \oplus K}^{m} \\ \varphi(F) &= (F(p_1), \ldots, F(p_m)) \end{aligned}$$

is a ring isomorphism. Recall that addition and multiplication on direct sum are defined component-wise:

$$\begin{aligned} (r_1, \ldots, r_m) + (s_1, \ldots, s_m) &= (r_1 + s_1, \ldots, r_m + s_m) \\ (r_1, \ldots, r_m) \cdot (s_1, \ldots, s_m) &= (r_1 \cdot s_1, \ldots, r_m \cdot s_m). \end{aligned}$$

3.1.7 (a) Give an example of an infinite affine variety $X \subseteq \mathbb{A}^3$ such that the three coordinate functions are all the same polynomial function.

(b) Prove that the solution to part (a) is unique.

Section 3.2 K-algebras

What types of rings arise as $K[X]$ for some affine variety X? The first answer to this question is given by considering the special role played by the ground field K.

To motivate our discussion, consider the case $X = \mathbb{A}^1$, for which

$$K[\mathbb{A}^1] = K[x].$$

As we have discussed at length, $K[x]$ is a ring. However, it is more than just an ordinary ring; it also has the structure of a vector space over K, since along with being able to add and multiply polynomials, we can also multiply polynomials by scalars in K (and the two operations of addition and scalar multiplication satisfy the usual vector space axioms). Unlike the vector spaces one typically studies in a first linear algebra course, $K[x]$ is infinite-dimensional, with a basis given by

$$\mathcal{B} = \{1, x, x^2, x^3, \dots\}.$$

Nonetheless, just like the more familiar finite-dimensional vector spaces, every element of $K[x]$ can be written uniquely as a linear combination of elements in $\mathcal{B}$.

More generally, every coordinate ring has this same enhanced structure—it is simultaneously a ring and a vector space over K—and this structure naturally sets coordinate rings apart from more general rings. In this section, we develop the algebraic foundations of K-algebras, which formalize this structure.

3.10 **DEFINITION** *K-algebra*

A K-algebra is a ring A together with a *scalar multiplication* function

$$\begin{aligned} K \times A &\to A \\ (r, a) &\mapsto r \cdot a \end{aligned}$$

that is related to the ring structure by the following axioms:

1. $r \cdot (a + b) = r \cdot a + r \cdot b$ for all $r \in K$ and all $a, b \in A$;
2. $(r + s) \cdot a = r \cdot a + s \cdot a$ for all $r, s \in K$ and all $a \in A$;
3. $(rs) \cdot a = r \cdot (s \cdot a)$ for all $r, s \in K$ and all $a \in A$;
4. $1 \cdot a = a$ for all $a \in A$, where 1 is the unity in K;
5. $r \cdot (ab) = (r \cdot a)b = a(r \cdot b)$ for all $r \in K$ and all $a, b \in A$.

To help parse the axioms in the Definition 3.10, the products within K and A have been written by concatenating the elements, reserving the symbol "$\cdot$" for scalar multiplication.

The first four axioms stipulate that A forms a vector space over K and the fifth specifies how scalar multiplication interacts with multiplication in A. It follows from the axioms, and is true of vector spaces in general, that $0 \cdot a = 0$ and $(-1) \cdot a = -a$ (Exercise 3.2.1).

3.11 **EXAMPLE** Polynomial rings

The prototypical example of a K-algebra, especially from the perspective of algebraic geometry, is the polynomial ring $K[x_1, \ldots, x_n]$. Indeed, along with being able to add and multiply polynomials, we can also multiply a polynomial by a scalar in K, and the axioms in Definition 3.10 are straightforward to verify.

3.12 **EXAMPLE** Coordinate rings

Let X be an affine variety. The coordinate ring $K[X]$ forms a K-algebra. For any $F \in K[X]$ and $r \in K$, we define $r \cdot F \in K[X]$ to be the function given by

$$(r \cdot F)(a_1, \ldots, a_n) = r \cdot F(a_1, \ldots, a_n),$$

where the multiplication on the right-hand side is the usual multiplication in K. To check that $r \cdot F$ is in fact an element of $K[X]$, notice that $F = f|_X$ for some $f \in K[x_1, \ldots, x_n]$, and it follows that $r \cdot F = (r \cdot f)|_X$. Since $r \cdot f \in K[x_1, \ldots, x_n]$, we see that $r \cdot F \in K[X]$. The axioms in Definition 3.10 are again readily verified.

3.13 **EXAMPLE** Extension rings of K

If A is any ring that contains K as a subring, then A is naturally a K-algebra where scalar multiplication is the usual ring multiplication in A. In fact, given our assumptions (rings are commutative with unity), every nontrivial K-algebra arises in this way. More precisely, given any K-algebra $A \neq \{0\}$, there is a canonical inclusion $K \to A$, and viewing K as a subring of A under this inclusion, scalar multiplication is identified with the usual multiplication in A (Exercise 3.2.5). In particular, every nontrivial K-algebra canonically contains a copy of K.

3.14 **EXAMPLE** Nonexamples of K-algebras

By Example 3.13, any nonzero ring that does not contain K is not a K-algebra. For example, since $\mathbb{Z}$ does not contain a field, it is not a K-algebra for any field K.

Our development of K-algebras is not complete until we specify the appropriate morphisms between them. Given that a K-algebra is an enhanced ring, with an additional scalar multiplication operation, a K-algebra homomorphism is an enhanced ring homomorphism that preserves scalar multiplication.

3.15 **DEFINITION** *Homomorphism of K-algebras*

Let A and B be K-algebras. A *K-algebra homomorphism* $\varphi : A \to B$ is a ring homomorphism for which

$$\varphi(r \cdot a) = r \cdot \varphi(a)$$

for all $r \in K$ and $a \in A$. We say that φ is an *isomorphism of K-algebras* and write $A \cong B$ if φ has an inverse that is also a K-algebra homomorphism.

Being an isomorphism appears to be stronger than being a bijection—not only should an inverse function exist, but it must also be a K-algebra homomorphism. However, as is conveniently the case for groups, rings, and fields, if φ is a bijective homomorphism, then its inverse is automatically a homomorphism (Exercise 3.2.2).

3.16 **EXAMPLE** *K-algebra homomorphisms from polynomial rings*

Consider the evaluation function

$$\begin{aligned} \varphi : \mathbb{R}[x,y] &\to \mathbb{R} \\ f(x,y) &\mapsto f(2,3). \end{aligned}$$

Some time reflecting should convince the reader that φ is an $\mathbb{R}$-algebra homomorphism. In addition, knowing that φ is an $\mathbb{R}$-algebra homomorphism, we can also see that it is completely determined by the image of x and y. For example, once we know that $\varphi(x) = 2$ and $\varphi(y) = 3$, then using that φ is a ring homomorphism that preserves scalar multiplication, we obtain

$$\varphi(5x^2y + 2y + xy) = 5(2)^2(3) + 2(3) + (2)(3) = 72.$$

More generally, for any K-algebra A and subset $\{a_1, \ldots, a_n\} \subseteq A$, there is a unique K-algebra homomorphism

$$\varphi : K[x_1, \ldots, x_n] \to A$$

satisfying $\varphi(x_i) = a_i$ for all i (Exercise 3.2.3). This shows that a K-algebra homomorphism $K[x_1, \ldots, x_n] \to A$ is equivalent to a choice of $a_1, \ldots, a_n \in A$.

Just like for groups and rings, there is a First Isomorphism Theorem for K-algebras, which is a fundamental tool for proving that two K-algebras are isomorphic. In order to state it, we must first define K-algebra quotients and subalgebras.

Since a K-algebra A is a ring, we already know that we can form the quotient ring A/I for any ideal $I \subseteq A$. The next result shows that the quotient ring A/I naturally inherits a K-algebra structure from A.

3.17 **PROPOSITION** *Quotient K-algebras*

Let A be a K-algebra. If $I \subseteq A$ is an ideal, then the quotient ring A/I is a K-algebra, in which scalar multiplication is defined by

$$r \cdot [a] = [r \cdot a]. \tag{3.1}$$

PROOF We must check that the scalar multiplication given by (3.1) is well-defined. Toward this end, the key point is that ideals are automatically closed under scalar multiplication: if $a \in I$ and $r \in K$, then

$$r \cdot a = r \cdot (1a) = (r \cdot 1)a,$$

where 1 is the unity in A. Since $r \cdot 1 \in A$ and $a \in I$, the absorbing property of ideals implies that $(r \cdot 1)a \in I$. Thus, $r \cdot a \in I$, verifying that I is closed under scalar multiplication.

From here, to see that scalar multiplication is well-defined, suppose $[a] = [b]$. Then $a - b \in I$, and using the fact that I absorbs scalar multiplication, we have $r \cdot a - r \cdot b = r \cdot (a - b) \in I$, which tells us that $[r \cdot a] = [r \cdot b]$.

Since I is an ideal, we already know that A/I is a ring, so all that remains to be checked is the five axioms of Definition 3.10. These all follow readily from the validity of the corresponding axioms for A. □

We now provide the final ingredient required for the First Isomorphism Theorem.

3.18 **DEFINITION** *Subalgebra*

Let A be a K-algebra. A *subalgebra* $B \subseteq A$ is a subring for which $r \cdot b \in B$ for all $r \in K$ and $b \in B$.

In other words, a subset of a K-algebra is a subalgebra if it is both a subring and closed under scalar multiplication, thereby forming a K-algebra in its own right. A natural example of a subalgebra is the K-algebra $K[x]$ as a subalgebra of $K[x, y]$.

3.19 **THEOREM** *First Isomorphism Theorem for K-algebras*

If $\varphi : A \to B$ is a K-algebra homomorphism, then

(i) $\mathrm{im}(\varphi)$ is a subalgebra of B,

(ii) $\ker(\varphi)$ is an ideal of A, and

(iii) the function

$$[\varphi] : \frac{A}{\ker(\varphi)} \to \mathrm{im}(\varphi)$$
$$[a] \mapsto \varphi(a)$$

is a well-defined isomorphism of K-algebras.

PROOF Exercise 3.2.7. □

3.20 **EXAMPLE** $\mathbb{C}$ as a quotient $\mathbb{R}$-algebra

Consider $\mathbb{C}$ as an $\mathbb{R}$-algebra, where scalar multiplication is the usual multiplication, and let φ be the function

$$\varphi : \mathbb{R}[x] \to \mathbb{C}$$
$$\varphi(f) = f(\mathrm{i}).$$

One can check (Exercise 3.2.8) that φ is a surjective $\mathbb{R}$-algebra homomorphism with kernel $\langle x^2 + 1 \rangle$. Thus, we obtain an $\mathbb{R}$-algebra isomorphism

$$\mathbb{C} \cong \mathbb{R}[x]/\langle x^2 + 1 \rangle.$$

As vector spaces over $\mathbb{R}$, the set of complex numbers $\mathbb{C}$ has a basis $\{1, \mathrm{i}\}$ and the quotient $\mathbb{R}[x]/\langle x^2 + 1 \rangle$ has basis $\{1, [x]\}$. The isomorphism $[\varphi]$ identifies 1 with 1 and $[x]$ with i, which is motivated by the fact that $[x]^2 = \mathrm{i}^2 = -1 \in \mathbb{R}$.

Returning to our motivating example of coordinate rings, we now use our knowledge of K-algebras to strengthen Proposition 3.7 to the setting of K-algebras. Let $X \subseteq \mathbb{A}^n$ be an affine variety and consider the ring homomorphism

$$\begin{aligned} K[x_1, \ldots, x_n] &\to K[X] \\ f &\mapsto f|_X. \end{aligned}$$

Since $(r \cdot f)|_X = r \cdot (f|_X)$, this is a K-algebra homomorphism. Exactly as in the proof of Proposition 3.7, it is surjective with kernel $\mathcal{I}(X)$. Thus, we obtain the following result from the First Isomorphism Theorem.

3.21 **PROPOSITION** *The coordinate ring as a quotient K-algebra*

If $X \subseteq \mathbb{A}^n$ is an affine variety, then there is a canonical K-algebra isomorphism

$$K[X] = K[x_1, \ldots, x_n]/\mathcal{I}(X).$$

Exercises for Section 3.2

3.2.1 Let A be a K-algebra. Prove that $0 \cdot a = 0$ and $(-1) \cdot a = -a$ for all $a \in A$.

3.2.2 Let $\varphi : A \to B$ be a homomorphism of K-algebras. Suppose that φ is a bijection, so that φ has an inverse function $\varphi^{-1} : B \to A$. Prove that this inverse function is a homomorphism of K-algebras.

3.2.3 Let A be a K-algebra and $\{a_1, \ldots, a_n\} \subseteq A$ a subset. Prove that there is a unique K-algebra homomorphism $\varphi : K[x_1, \ldots, x_n] \to A$ that satisfies $\varphi(x_i) = a_i$.

3.2.4 Give an example of a ring homomorphism $\varphi : \mathbb{C} \to \mathbb{C}$ that is not a $\mathbb{C}$-algebra homomorphism.

3.2.5 Let $A \neq \{0\}$ be a K-algebra. Prove that the function $\varphi : K \to A$ defined by $\varphi(r) = r \cdot 1$ is an injective ring homomorphism. Viewing K as a subring of A via this homomorphism, prove that scalar multiplication is identified with the usual multiplication in A.

3.2.6 Let A and B be rings containing K, endowed with the K-algebra structure of Example 3.13. Prove that any K-algebra homomorphism $\varphi : A \to B$ must be the identity on K.

3.2.7 Prove the First Isomorphism Theorem for K-algebras.

3.2.8 Fill in the details of the proof in Example 3.20 that $\mathbb{C} \cong \mathbb{R}[x]/\langle x^2 + 1 \rangle$ as $\mathbb{R}$-algebras.

Section 3.3 Generators of K-algebras

Our ongoing task, recall, is to determine precisely which rings arise as coordinate rings. The previous section shows that, for a ring R to be a coordinate ring, it must be a K-algebra, and Proposition 3.21 refines this statement: R must be a quotient of a polynomial ring. But it may not be immediately obvious whether a given ring is isomorphic to such a quotient; we saw in Example 3.20, for instance, that $\mathbb{C}$ is isomorphic as an $\mathbb{R}$-algebra to a quotient of $\mathbb{R}[x]$, despite not being initially presented as such.

The goal of this section is to characterize exactly which K-algebras arise as quotients of polynomial rings. One of the key ingredients in order to do this is the notion of K-algebra generators.

3.22 **DEFINITION** *Polynomial combination, generators*

Let A be a K-algebra and let $\mathcal{S} \subseteq A$ be a subset. A *polynomial combination of* $\mathcal{S}$ is an element of A of the form

$$f(a_1, \ldots, a_n)$$

for some polynomial $f \in K[x_1, \ldots, x_n]$ and $a_1, \ldots, a_n \in \mathcal{S}$. The set of all polynomial combinations of $\mathcal{S}$ is called the *subalgebra of A generated by $\mathcal{S}$*, and it is denoted $K[\mathcal{S}]$.

If $\mathcal{S} = \{a_1, \ldots, a_n\}$ is a finite set, we omit the set brackets and write $K[\mathcal{S}] = K[a_1, \ldots, a_n]$.

The reader is encouraged to check that $K[\mathcal{S}]$ is, indeed a subalgebra of A, and it is the smallest subalgebra that contains $\mathcal{S}$ (Exercise 3.3.1). An alternative way to think of $K[\mathcal{S}]$ is that it consists of all elements in A that can be obtained from elements of $\mathcal{S}$ using the operations of addition, multiplication, and scalar multiplication by elements of K.

3.23 **EXAMPLE** Subalgebras of $K[x, y]$

Consider the K-algebra $A = K[x, y]$. If we set $a = x$, then we see that the subalgebra generated by a is the collection of all polynomials in x:

$$K[a] = K[x] \subseteq K[x, y].$$

On the other hand, if $b = x + y$, then $K[b]$ is the collection of all polynomials of the form

$$\sum_{i=0}^{k} r_i (x + y)^i$$

where $k \in \mathbb{N}$ and $r_i \in K$ for all i. Taking the generators a and b together, we see that $K[a, b]$ contains both $x = a$ and $y = b - a$, from which we conclude that $K[a, b]$ is the set of all polynomials in x and y:

$$K[a, b] = K[x, y].$$

3.24 **EXAMPLE** Generators for $K[x,y]/\langle xy-1\rangle$

Consider the K-algebra

$$A = \frac{K[x,y]}{\langle xy-1\rangle}.$$

Any element of A is of the form $[f(x,y)]$ for some $f(x,y) \in K[x,y]$. By definition of coset arithmetic, we have

$$[f(x,y)] = f([x],[y]).$$

Thus, any element of A can be written as a polynomial expression in $a = [x]$ and $b = [y]$. This implies that A is generated as an algebra by a and b:

$$A = K[a,b].$$

While the notation $K[a,b]$ in Example 3.24 is reminiscent of that for a polynomial ring, A is *not* the same thing as the ring of polynomials in variables a and b. In particular, there is a relation between the generators:

$$ab = 1.$$

Here and throughout, we use letters at the end of the alphabet, such as x and y, to denote variables in polynomial rings. By definition, these variables do not have any relations among themselves, meaning that two polynomials are equal if and only if they have the same coefficients. On the other hand, we use letters at the beginning of the alphabet, such as a and b, to denote generators of K-algebras, which may satisfy relations; in other words, it's possible for $f(a,b) = g(a,b)$ even if f and g are different polynomials.

In the previous two examples, the entire K-algebra could be generated by finitely many elements. We capture this by saying that they are *finitely-generated* K-algebras.

3.25 **DEFINITION** *Finitely-generated*

Let A be a K-algebra. We say that A is *finitely-generated* if there exist $a_1,\dots,a_n \in A$ such that $A = K[a_1,\dots,a_n]$.

For example, the polynomial ring $K[x_1\dots,x_n]$ is a finitely-generated K-algebra, simply by taking $a_i = x_i$ for all $i = 1,\dots,n$. More generally, the quotient ring $K[x_1,\dots,x_n]/I$ is finitely-generated for any ideal I, as we can take $a_i = [x_i]$ for all $i = 1,\dots,n$, generalizing Example 3.24. In fact, up to isomorphism, these are the only examples of finitely-generated K-algebras, as we now verify.

3.26 **PROPOSITION** *Characterization of finitely-generated K-algebras*

Let A be a K-algebra. Then A is finitely-generated if and only if there is an isomorphism of K-algebras

$$A \cong K[x_1,\dots,x_n]/I$$

for some $n \geq 0$ and some ideal $I \subseteq K[x_1,\dots,x_n]$.

PROOF Suppose that A is finitely-generated. By definition, this means that there exist $a_1, \ldots, a_n \in A$ such that $A = K[a_1, \ldots, a_n]$. Define a function

$$\begin{aligned} \varphi : K[x_1, \ldots, x_n] &\to A \\ f(x_1, \ldots, x_n) &\mapsto f(a_1, \ldots, a_n). \end{aligned}$$

It is straightforward to check that φ is a K-algebra homomorphism, and the fact that $A = K[a_1, \ldots, a_n]$ is equivalent to the statement that φ is surjective. Letting $I = \ker(\varphi)$, the First Isomorphism Theorem implies that

$$A \cong K[x_1, \ldots, x_n]/I.$$

Conversely, suppose that there exists an isomorphism

$$\psi : K[x_1, \ldots, x_n]/I \to A,$$

and let $a_i = \psi([x_i])$ for $i = 1, \ldots, n$. We aim to show that $A = K[a_1, \ldots, a_n]$. Suppose $a \in A$; we must show that a is a polynomial expression in $a_1, \ldots, a_n$. Since ψ is surjective, there exists $[f] \in K[x_1, \ldots, x_n]/I$ such that $\psi([f]) = a$. Then

$$a = \psi([f(x_1, \ldots, x_n)]) = \psi(f([x_1], \ldots, [x_n])) = f(\psi([x_1]), \ldots, \psi([x_n])),$$

where the second equality follows from arithmetic of cosets and the third from the assumption that ψ is a K-algebra homomorphism. Since $a_i = \psi([x_i])$, we see that a is a polynomial expression in $a_1, \ldots, a_n$, and we conclude that A is finitely-generated. □

An isomorphism of the form in Proposition 3.26 is often referred to as a *presentation* of the K-algebra A. The images of $x_1, \ldots, x_n$ are the *generators* of the presentation and the polynomials in I are the *relations* of the presentation.

By Proposition 3.21, coordinate rings of affine varieties are canonically isomorphic to quotients of polynomial rings. Thus, we obtain the next result as a consequence of Proposition 3.26.

For a few examples of algebras that are not finitely-generated, have a look at Exercises 3.3.7 - 3.3.9.

3.27 **COROLLARY** *$K[X]$ is finitely-generated*

Let $X \subseteq \mathbb{A}^n$ be an affine variety. Then the coordinate ring $K[X]$ is a finitely-generated K-algebra.

More explicitly, the proof of Proposition 3.26 shows that $K[X]$ can be generated as a K-algebra by the elements $[x_1], \ldots, [x_n]$ in the canonical isomorphism

$$K[x_1, \ldots, x_n]/\mathcal{I}(X) = K[X]$$

of Proposition 3.21. These images are the n coordinate functions

$$\begin{aligned} C_i : X &\to K \\ (a_1, \ldots, a_n) &\mapsto a_i. \end{aligned}$$

Thus, $K[X]$ is the K-algebra generated by the coordinate functions; this is the reason we call it the "coordinate ring" of X.

Exercises for Section 3.3

3.3.1 Let A be a K-algebra and $\mathcal{S} \subseteq A$ be a subset.

(a) Prove that $K[\mathcal{S}]$ is a subalgebra of A.

(b) Prove that $K[\mathcal{S}]$ is contained in every subalgebra of A that contains $\mathcal{S}$.

3.3.2 Give two different examples of three elements that generate $K[x,y,z]$.

3.3.3 Consider the K-algebra

$$A = K[x,y,z]/\langle y - x^2, z\rangle.$$

Find an element $a \in A$ such that $A = K[a]$.

3.3.4 Consider the K-algebra

$$A = K[x,y]/\langle xy - 1\rangle.$$

Prove that $A \neq K[a]$ for any $a \in A$.

3.3.5 Suppose that A and B be K-algebras such that $A = K[a_1,\ldots,a_n]$ for some $a_1,\ldots,a_n \in A$. Prove that, if $\varphi,\psi : A \to B$ are K-algebra homomorphisms such that $\varphi(a_i) = \psi(a_i)$ for all $i = 1,\ldots,n$, then $\varphi(a) = \psi(a)$ for all $a \in A$.

3.3.6 Consider $\mathbb{R}$ as a $\mathbb{Q}$-algebra.

(a) The subalgebra $\mathbb{Q}[\sqrt{2}] \subseteq \mathbb{R}$ is finitely-generated, so Proposition 3.26 implies that $\mathbb{Q}[\sqrt{2}] \cong \mathbb{Q}[x_1,\ldots,x_n]/I$ for some n and I. Find an explicit n and I for which this is the case.

(b) Repeat part (a) for the subalgebra $\mathbb{Q}[\sqrt[3]{2}] \subseteq \mathbb{R}$.

(c) Repeat part (a) for the subalgebra $\mathbb{Q}[\pi] \subseteq \mathbb{R}$.

3.3.7 Prove that $K[x_1,x_2,x_3,\ldots]$ is not a finitely-generated K-algebra. (The ring $K[x_1,x_2,x_3,\ldots]$ was defined in Example 2.13.)

3.3.8 Prove that any finitely-generated $\mathbb{Q}$-algebra is countable and conclude that $\mathbb{R}$ is not a finitely-generated $\mathbb{Q}$-algebra.

3.3.9 This exercise shows that a subalgebra of a finitely-generated algebra need not be finitely-generated. Consider the subalgebra

$$A = K[x, xy, xy^2, xy^3, xy^4, \ldots] \subseteq K[x,y].$$

Prove that A is not a finitely-generated K-algebra.

Section 3.4 Nilpotents and reduced rings

We learned in Corollary 3.27 that the coordinate ring $K[X]$ of an affine variety X is a finitely-generated K-algebra. This follows from the interpretation of $K[X]$ as a quotient:

$$K[X] = K[x_1, \ldots, x_n]/\mathcal{I}(X).$$

However, there is still one important aspect of this quotient that we have not yet taken into account: $K[X]$ is not just a quotient by an arbitrary ideal, it is a quotient by a *radical* ideal. What, then, does the fact that $\mathcal{I}(X)$ is radical imply about the algebraic properties of the coordinate ring $K[X]$? The answer, as it turns out, can be phrased in terms the following definition.

3.28 **DEFINITION** *Nilpotents and reduced rings*

Let R be a ring. An element $a \in R$ is *nilpotent* if there exists a natural number $m \geq 1$ such that $a^m = 0$. We say that R is *reduced* if it has no nonzero nilpotent elements.

3.29 **EXAMPLE** Reduced rings

Any integral domain is necessarily reduced, since a nonzero nilpotent element would be a zero divisor. Not all reduced rings are integral domains, however. For example, the quotient ring

$$\frac{K[x,y]}{\langle xy \rangle}$$

is not an integral domain, because $[x]$ and $[y]$ are zero divisors, but it is reduced. To see that this ring is reduced, suppose that $[f]^m = 0$; we must prove that $[f] = 0$. Since

$$[f^m] = [f]^m = 0 \in \frac{K[x,y]}{\langle xy \rangle},$$

we know that xy divides f^m, implying that x and y both divide f^m. Since x and y are irreducible in $K[x,y]$, and thus prime, it follows that x and y both divide f. Therefore, xy divides f, and $[f] = 0$. Thus, the ring does not contain any nonzero nilpotents, so it is reduced.

That every integral domain is reduced, but not vice versa, is a manifestation of the fact that every prime ideal is radical, but not vice versa (see Proposition 3.31).

3.30 **EXAMPLE** A nonreduced ring

The quotient ring

$$\frac{K[x]}{\langle x^2 \rangle}$$

is not reduced, because it has a nonzero nilpotent:

$$[x] \neq 0 \quad \text{satisfies} \quad [x]^2 = 0.$$

The next result is a quotient characterization of radical ideals, analogous to the quotient characterizations of prime and maximal ideals.

3.31 **PROPOSITION** *Quotients by radical ideals*

An ideal $I \subseteq R$ is radical if and only if R/I is reduced.

PROOF Exercise 3.4.2. □

We can now give a complete algebraic characterization of the type of rings that arise as coordinate rings of affine varieties over K; they are finitely-generated K-algebras that are reduced as rings.

3.32 **PROPOSITION** *Characterization of coordinate rings*

If $X \subseteq \mathbb{A}^n$ is an affine variety, then the coordinate ring $K[X]$ is a finitely-generated reduced K-algebra. Conversely, if A is a finitely-generated reduced K-algebra, then $A \cong K[X]$ for some affine variety $X \subseteq \mathbb{A}^n$.

PROOF Suppose X is an affine variety. By Proposition 3.21,

$$K[X] = K[x_1, \ldots, x_n]/\mathcal{I}(X).$$

Thus, $K[X]$ is finitely-generated by Proposition 3.26 and reduced by Proposition 3.31 and the fact that $\mathcal{I}(X)$ is a radical ideal.

Conversely, suppose that A is a finitely-generated reduced K-algebra. By Proposition 3.26, we can write

$$A \cong K[x_1, \ldots, x_n]/I$$

for some n and I, and by Proposition 3.31, we know that I is a radical ideal. Define $X = \mathcal{V}(I) \subseteq \mathbb{A}^n$. By the Nullstellensatz,

$$\mathcal{I}(X) = \mathcal{I}(\mathcal{V}(I)) = \sqrt{I} = I.$$

It then follows from Proposition 3.21 that $A \cong K[X]$. □

To make Proposition 3.32 effective, we should be able to produce, given a finitely-generated reduced K-algebra A, an affine variety X for which $K[X] \cong A$. In the next example, we illustrate how to carry out this procedure in practice.

3.33 **EXAMPLE** Determining X from $K[X]$

Let A be the subalgebra

$$A = K[u^2, uv, v^2] \subseteq K[u, v].$$

Then A is manifestly finitely-generated, as it is generated by the three elements u^2, uv, and v^2, and it is reduced, because it is a subalgebra of the reduced K-algebra $K[u, v]$. Thus, there should exist an affine variety X such that $K[X] \cong A$.

To find X, let's give the three generators names,

$$x = u^2, \quad y = uv, \quad \text{and} \quad z = v^2.$$

Notice that these three generators satisfy the relation

$$xz - y^2 = (u^2)(v^2) - (uv)^2 = 0.$$

Consider the affine variety defined by this relation:

$$X = \mathcal{V}(xz - y^2) \subseteq \mathbb{A}^3.$$

In the rest of this example, we prove that $K[X] \cong A$.

First, we observe that the polynomial $xz - y^2$ is irreducible (using, for example, Eisenstein's criterion), so the ideal $\langle xz - y^2 \rangle$ is prime and hence radical. It follows from the Nullstellensatz that

$$\mathcal{I}(X) = \mathcal{I}(\mathcal{V}(xz - y^2)) = \sqrt{\langle xz - y^2 \rangle} = \langle xz - y^2 \rangle,$$

so

$$K[X] \cong \frac{K[x,y,z]}{\langle xz - y^2 \rangle}.$$

What remains to be shown is that

$$\frac{K[x,y,z]}{\langle xz - y^2 \rangle} \cong A. \tag{3.34}$$

To see this, define a K-algebra homomorphism $\varphi : K[x,y,z] \to A$ by

$$\varphi(f) = f(u^2, uv, v^2).$$

Since the three generators u^2, uv, v^2 of A are all in the image of φ, it follows that φ is surjective. Therefore, the sought-after isomorphism in (3.34) follows from the First Isomorphism Theorem for K-algebras if we can prove that $\ker(\varphi) = \langle xz - y^2 \rangle$.

Since

$$\varphi(xz - y^2) = (u^2)(v^2) - (uv)^2 = 0,$$

every element of $\langle xz - y^2 \rangle$ is sent to 0 by φ, so $\ker(\varphi) \supseteq \langle xz - y^2 \rangle$. To prove the other inclusion, suppose $f \in \ker(\varphi)$ and consider the coset

$$[f] \in \frac{K[x,y,z]}{\langle xz - y^2 \rangle}.$$

By repeated use of the equation $[y^2] = [xz]$, we see that

$$[f] = [g(x,z) + yh(x,z)]$$

for some polynomials $g, h \in K[x,z]$. In other words,

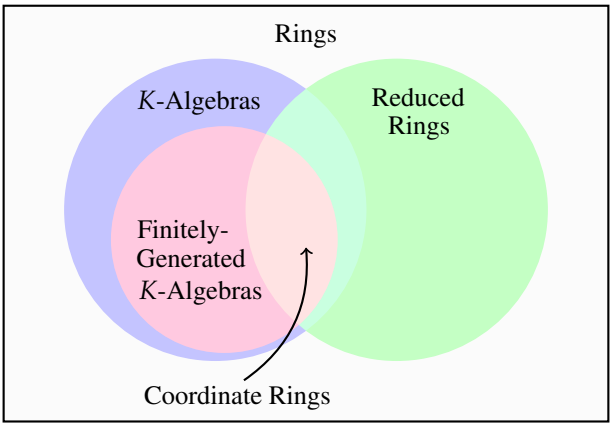
$$f = g(x,z) + y \cdot h(x,z) + \ell(x,y)(xz - y^2)$$

for some $\ell \in K[x,z]$. Applying φ, we obtain

$$0 = \varphi(f) = g(u^2, v^2) + uv \cdot h(u^2, v^2) \in K[u,v].$$

Since the term $g(u^2, v^2)$ is a polynomial with only even powers of both u and v and the term $uv \cdot h(u^2, v^2)$ is a polynomial with only odd powers of u and v, there can be no cancellation between these two terms. Therefore, we must have $g = h = 0$, implying that $f \in \langle xz - y^2 \rangle$. Thus, $\ker(\varphi) \subseteq \langle xz - y^2 \rangle$, finishing the argument.

The following diagram depicts the developments of this chapter. In particular, in the category of rings, we have pinned down exactly which rings arise as coordinate rings over K: they must be K-algebras that are both finitely-generated and reduced.



Now that we have an algebraic language in which we can characterize and discuss coordinate rings, our goal in the next chapter is to investigate what the coordinate ring $K[X]$ tells us about the affine variety X. As we will see, coordinate rings know essentially everything about their corresponding affine variety, which is a powerful tool, allowing us to bring all the tools of ring theory to bear on the study of affine varieties.

Exercises for Section 3.4

3.4.1 Prove that any subring of a reduced ring is reduced.

3.4.2 Prove Proposition 3.31.

3.4.3 Recall that the *direct sum* of rings R and S is the ring

$$R \oplus S = \{(r,s) \mid r \in R,\ s \in S\},$$

with addition and multiplication defined componentwise:

$$(r_1,s_1) + (r_2,s_2) = (r_1 + r_2, s_1 + s_2),$$
$$(r_1,s_1)(r_2,s_2) = (r_1 r_2, s_1 s_2).$$

(a) Prove that, if R and S are reduced, then $R \oplus S$ is reduced.

(b) Prove that the K-algebra $A = K[x]/\langle x^2 \rangle$ is isomorphic to $K \oplus K$ as a K-vector space but not as a ring.

3.4.4 Let R be a reduced ring. Prove that $R[x]$ is reduced, and conclude, by induction, that $R[x_1, \ldots, x_n]$ is reduced for any n.

3.4.5 Let $A = K[u^2, u^3] \subseteq K[u]$.

(a) Explain how you know that A is finitely-generated and reduced.

(b) Find an affine variety X such that $K[X] \cong A$, and prove your answer.

3.4.6 Let $A = K[u+w, v+w] \subseteq K[u,v,w]$. Find an affine variety X such that $K[X] \cong A$, and prove your answer.

3.4.7 Let A be the K-vector space with basis

$$\{x^i \mid i \geq 0\} \cup \{x^i y \mid i \geq 0\}.$$

Define a (commutative) product on the elements of this basis by

$$\begin{aligned} x^i \cdot x^j &= x^{i+j} \\ x^i \cdot x^j y &= x^{i+j} y \\ x^i y \cdot x^j y &= x^{i+j+3}, \end{aligned}$$

and extend this product to all elements of A by linearity in K. One can prove that A is a finitely-generated reduced K-algebra, and hence there should exist an affine variety X such that $K[X] \cong A$. Find such an X, and prove your answer.

3.4.8 Let $a_1, \ldots, a_m \in K[x_1, \ldots, x_n]$ and consider the subalgebra

$$A = K[a_1, \ldots, a_m] \subseteq K[x_1, \ldots, x_n].$$

Let $I \subseteq K[y_1, \ldots, y_m]$ be the *ideal of relations* of $a_1, \ldots, a_m$:

$$f \in I \iff f(a_1, \ldots, a_m) = 0 \in K[x_1, \ldots, x_n].$$

Define $X = \mathcal{V}(I) \subseteq \mathbb{A}^m$. Prove that $K[X] \cong A$.

3.4.9 Give an explicit example of a ring that belongs in each region of the Venn diagram presented at the end of this section.

Chapter 4

Polynomial Maps

LEARNING OBJECTIVES FOR CHAPTER 4

- Become familiar with the notion of polynomial maps between affine varieties and the notion of isomorphism.
- Learn about pullback homomorphisms and compute pullbacks in concrete examples.
- Use the bijectivity of pulling back to prove that affine varieties are isomorphic if and only if their coordinate rings are isomorphic.
- Become familiar with the equivalence between algebra and geometry as an intrinsic (as opposed to extrinsic) statement.

The previous chapter provides us with an association

$$\begin{aligned}\{\text{affine varieties}\} &\to \{\text{finitely-generated reduced } K\text{-algebras}\}\\ X &\mapsto K[X]\end{aligned}$$

and confirms that it is surjective. But is it injective—that is, if $K[X] = K[Y]$, is it necessarily the case that $X = Y$?

This question is more subtle than it might first appear. To answer it, one must decide whether two K-algebras are "the same"; is $K[x]$ the same as $K[y]$, for example? The literal answer is no, but the reader would be forgiven for finding this answer unsatisfying, given how conditioned we are to viewing isomorphic rings as identical. To capture this intuition that isomorphism is "sameness," we might instead ask whether, if $K[X] \cong K[Y]$, it is necessarily the case that $X \cong Y$?

We do not yet have a notion, however, of what it means for two affine varieties to be "isomorphic," or even how a "morphism" between affine varieties should be defined. This is something one should do whenever a new type of mathematical object—groups, rings, topological spaces, et cetera—is introduced: ask which maps between those objects preserve their relevant structure. In the context of groups, the relevant maps are group homomorphisms, while for rings they are ring homomorphisms, and for topological spaces they are continuous maps. Because algebraic geometry is concerned with polynomials, it comes as no surprise that the relevant maps between affine varieties are *polynomial maps*, which we define in this chapter.

Equipped with the definition of polynomial maps, we can make sense of what it means to say that affine varieties X and Y are *isomorphic*, and we can prove that $X \cong Y$ if and only if $K[X] \cong K[Y]$. This is the goal of the chapter and the heart of our dictionary: the equivalence of algebra and geometry.

Section 4.1 Polynomial maps between affine varieties

In the previous chapter, we were introduced to the coordinate ring $K[X]$ of an affine variety $X \subseteq \mathbb{A}^m$, whose elements are polynomial functions $F : X \to K$. Identifying K with $\mathbb{A}^1$, we can view such functions as a special case of maps between affine varieties (between X and $\mathbb{A}^1$), and our first goal is to extend the definition to allow for maps from any affine variety to any other.

There is a deliberate distinction in terminology between the words function *and* map. *A* function *takes values in the ground field whereas a* map *takes values in an affine variety.*

Let $X \subseteq \mathbb{A}^m$ be an affine variety. Since every element of the coordinate ring $K[X]$ is a function $X \to K$, we see that n elements $F_1, \ldots, F_n \in K[X]$ give rise to a map $X \to \mathbb{A}^n$ defined by $a \in X \mapsto (F_1(a), \ldots, F_n(a)) \in \mathbb{A}^n$. If $Y \subseteq \mathbb{A}^n$ is an affine variety and the image of the map $X \to \mathbb{A}^n$ happens to lie in Y, then we obtain a map $X \to Y$. Maps that arise from polynomial functions in this way are called *polynomial maps*.

4.1 **DEFINITION** *Polynomial map between affine varieties*

Let $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$ be affine varieties. A map $F : X \to Y$ is said to be a *polynomial map* if there exist $F_1, \ldots, F_n \in K[X]$ such that, for every $a \in X$,

$$F(a) = (F_1(a), \ldots, F_n(a)).$$

In particular, a polynomial function on X is the same thing as a polynomial map $F : X \to \mathbb{A}^1$. In our first example, we consider a polynomial map whose target is the affine space $\mathbb{A}^3$.

4.2 **EXAMPLE** Polynomial maps to affine space

Consider the parabola $X = \mathcal{V}(x_2 - x_1^2) \subseteq \mathbb{A}^2$. (We have named the variables x_1 and x_2, rather than x and y, in preparation for the next example.) Then

$$\begin{aligned} F : X &\to \mathbb{A}^3 \\ (a_1, a_2) &\mapsto (a_1 - a_1^2, a_1 + a_2, a_1^2 - a_2^2) \end{aligned}$$

is a polynomial map, since its three component functions

$$F_1(a_1, a_2) = a_1 - a_1^2, \quad F_2(a_1, a_2) = a_1 + a_2, \quad \text{and} \quad F_3(a_1, a_2) = a_1^2 - a_2^2$$

arise from the polynomials

$$f_1 = x_1 - x_1^2, \quad f_2 = x_1 + x_2, \quad \text{and} \quad f_3 = x_1^2 - x_2^2,$$

respectively, and are thus elements of $K[X]$. Notice that the polynomials

$$g_1 = x_1 - x_2, \quad g_2 = x_1 + x_1^2, \quad \text{and} \quad g_3 = x_2 - x_2^2$$

give rise to the same polynomial map $F : X \to \mathbb{A}^3$, because $[f_i] = [g_i] \in K[X]$.

If $X \subseteq \mathbb{A}^m$ is an affine variety, it is straightforward to produce polynomial maps $F : X \to \mathbb{A}^n$ whose codomain is an affine space. In particular, any choice of polynomials $f_1, \ldots, f_n \in K[x_1, \ldots, x_m]$ defines the coordinate functions $F_1, \ldots, F_n$ of such a map, by setting

$$F_i = [f_i] \in \frac{K[x_1, \ldots, x_m]}{\mathcal{I}(X)} = K[X],$$

and another choice $g_1, \ldots, g_n$ of polynomials produces the same polynomial map if and only if $f_i - g_i \in \mathcal{I}(X)$ for every i.

On the other hand, if $Y \subsetneq \mathbb{A}^n$ is an affine variety other than affine space itself, then not every choice of $F_1, \ldots, F_n \in K[X]$ gives a map $F : X \to Y$. In particular, in order to ensure that the image of every point of X is a point in Y, we must require that for every $a \in X$, the point

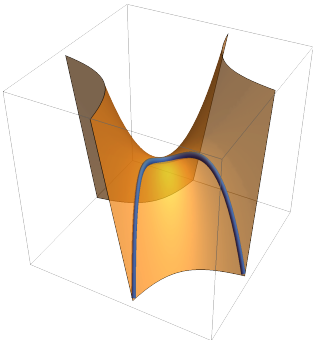
$$(F_1(a), \ldots, F_n(a)) \in \mathbb{A}^n$$

actually lies in Y, meaning that it is a solution of the defining polynomials of Y. More concretely, if $Y = \mathcal{V}(\mathcal{S})$ where $\mathcal{S} \subseteq K[x_1, \ldots, x_n]$, then we must check that, for every $a \in X$ and every $g \in \mathcal{S}$,

$$g(F_1(a), \ldots, F_n(a)) = 0.$$

4.3 EXAMPLE A polynomial map to an affine variety

As above, let $X = \mathcal{V}(x_2 - x_1^2) \subseteq \mathbb{A}^2$, but now let $Y = \mathcal{V}(y_1 y_2 - y_3) \subseteq \mathbb{A}^3$. Over the real numbers, Y is the one-sheeted hyperboloid depicted to the right, and the function



$$\begin{aligned} F : X &\to Y \\ (a_1, a_2) &\mapsto (a_1 - a_1^2, a_1 + a_2, a_1^2 - a_2^2) \end{aligned}$$

of Example 4.2 is a polynomial map from X to Y, whose image is illustrated as the curve on the hyperboloid. We have already checked that the three component functions of F are polynomial, but we now must also confirm that the image actually lies in Y. This is equivalent to the claim that $F(a_1, a_2)$ satisfies the defining equation $y_1 y_2 - y_3$ of Y, or in other words that

$$F_1(a_1, a_2) F_2(a_1, a_2) - F_3(a_1, a_2) = 0$$

whenever $(a_1, a_2) \in X$. To check this, we simply substitute in the expressions for F_1, F_2, F_3 and rearrange:

$$(a_1 - a_2^2)(a_1 + a_2) - (a_1^2 - a_2^2) = (a_1 + a_2)(a_2 - a_1^2) = (a_1 + a_2) \cdot 0 = 0,$$

where the second equality follows from $(a_1, a_2) \in X$, implying that $a_2 - a_1^2 = 0$.

To distinguish between coordinates on X and on Y, we often use subscripts, as in the above example: $x_1, x_2, \ldots$ for X and $y_1, y_2, \ldots$ for Y.

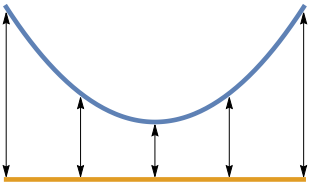
From the algebraic context, the reader is already familiar with the notion that some homomorphisms are isomorphisms, and that isomorphic objects share all of their relevant properties; isomorphic groups, for example, share all group-theoretic properties. Similarly, *isomorphisms of affine varieties* allow us to talk about what it means for affine varieties to be essentially the same.

4.4 DEFINITION *Isomorphism of affine varietes*

Let $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$ be affine varieties. A polynomial map $F : X \to Y$ is said to be an *isomorphism* if it has an inverse function $F^{-1} : Y \to X$ that is also a polynomial map. If such an isomorphism exists, we say that X and Y are *isomorphic* and write $X \cong Y$.

4.5 EXAMPLE The parabola is isomorphic to $\mathbb{A}^1$

Let $X = \mathcal{V}(x_2 - x_1^2) \subseteq \mathbb{A}^2$. Notice that the two maps F and G defined by



$$F : X \to \mathbb{A}^1 \qquad G : \mathbb{A}^1 \to X$$
$$(a_1, a_2) \mapsto a_1 \qquad b \mapsto (b, b^2)$$

are both polynomial maps. Furthermore, it is straightforward to check that $F \circ G$ is the identity on $\mathbb{A}^1$ and, using the fact that $a_2 = a_1^2$ for every $(a_1, a_2) \in X$, it can also be checked that $G \circ F$ is the identity on X. Thus, $X \cong \mathbb{A}^1$. In the figure above, we have depicted the isomorphisms between the affine line and the parabola over $\mathbb{R}$, where F is the downward map and G is the upward map.

4.6 EXAMPLE Isomorphic projections

Generalizing the previous example, suppose that $X = \mathcal{V}(x_m - g) \subseteq \mathbb{A}^m$ where $g \in K[x_1, \ldots, x_{m-1}]$. Consider the two maps

$$F : X \to \mathbb{A}^{m-1}$$
$$(a_1, \ldots, a_m) \mapsto (a_1, \ldots, a_{m-1})$$

and

$$G : \mathbb{A}^{m-1} \to X$$
$$(b_1, \ldots, b_{m-1}) \mapsto (b_1, \ldots, b_{m-1}, g(b_1, \ldots, b_{m-1})).$$

Both F and G are polynomial maps, and using the fact that $a_m = g(a_1, \ldots, a_{m-1})$ for every $(a_1, \ldots, a_m) \in X$, it follows that they are inverse to each other. Thus, we conclude that $X \cong \mathbb{A}^{m-1}$. See Exercise 4.1.4 for a further generalization of this example.

4.7 EXAMPLE Translations are isomorphisms

Given $c = (c_1, \ldots, c_m) \in \mathbb{A}^m$, consider the translation map

$$\begin{aligned} T_c : \mathbb{A}^m &\to \mathbb{A}^m \\ (a_1, \ldots, a_m) &\mapsto (a_1 + c_1, \ldots, a_m + c_m). \end{aligned}$$

Then for any affine variety $X \subseteq \mathbb{A}^m$, the translation $T_c(X)$ is also an affine variety; to see this, notice that a polynomial $f(x_1, \ldots, x_m)$ vanishes on $T_c(X)$ if and only if $f(x_1 + c_1, \ldots, x_m + c_m)$ vanishes on X, so

$$T_c(X) = \mathcal{V}\Big(\{ f \in K[x_1, \ldots, x_n] \mid f(x_1 + c_1, \ldots, x_m + c_m) \in \mathcal{I}(X) \} \Big).$$

The map $T_c : X \to T_c(X)$ is manifestly polynomial, and in fact, it is an isomorphism. To prove this, it suffices to notice that it has a polynomial inverse defined by translating back by the point $(-c_1, \ldots, -c_m) \in \mathbb{A}^m$.

Exercise 4.1.6 generalizes this example to compositions of translations with invertible linear maps; such compositions are called *affine linear transformations*.

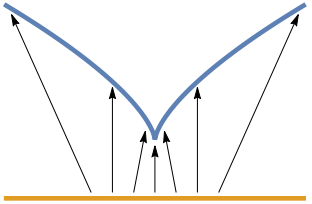
In order for a map of affine varieties to be an isomorphism, it must be bijective, because this is necessary for an inverse function to exist. However, *not every bijective polynomial map of affine varieties is an isomorphism*, because an inverse function, even if it exists, need not be a polynomial map. The next example illustrates this phenomenon.

Geometry and algebra differ here: in algebra, if a homomorphism (of groups, rings, algebras, et cetera) has an inverse function, then that inverse function is automatically a homomorphism. See Exercise 3.2.2.

4.8 EXAMPLE Bijective polynomial maps need not be isomorphisms

Consider $X = \mathcal{V}(x^2 - y^3) \subseteq \mathbb{A}^2$. Then

$$\begin{aligned} F : \mathbb{A}^1 &\to X \\ a &\mapsto (a^3, a^2) \end{aligned}$$



is the polynomial map depicted to the right. (Note that for every $a \in \mathbb{A}^1$, the point $(a^3, a^2) \in \mathbb{A}^2$ indeed satisfies the equation $x^2 - y^3 = 0$.) Moreover, the reader can verify that an inverse to F is given by

$$\begin{aligned} G : X &\to \mathbb{A}^1 \\ (b, c) &\mapsto \begin{cases} b/c & \text{if } c \neq 0 \\ 0 & \text{if } c = 0. \end{cases} \end{aligned}$$

However, G is not a polynomial map. That G involves a quotient of its inputs certainly hints at its non-polynomiality, but, as we saw in Example 3.5, more care is required to be sure that G is not polynomial. In the next section, we develop the necessary tools to prove that $X \ncong \mathbb{A}^1$, from which it follows that G *cannot* be a polynomial map. The "kink" in X is a visual clue that $X \ncong \mathbb{A}^1$.

To prove that $X \cong Y$, the task that needs to be accomplished is somewhat straightforward: we must find an isomorphism between them. But proving that $X \ncong Y$ is quite a bit more subtle. How can we rigorously prove the nonexistence of any isomorphism? In the context of algebra, we have quite a few tools for doing so: proving that two rings are not isomorphic involves finding a ring-theoretic property (like being an integral domain or a UFD) that one has but the other does not.

Therefore, if we want to be able to detect when affine varieties are not isomorphic, our goal should be to prove that $X \cong Y$ implies $K[X] \cong K[Y]$. Once we have accomplished this, then an algebraic proof that $K[X] \ncong K[Y]$ would imply that $X \ncong Y$, allowing us to import the methods of algebra to detect when affine varieties are not isomorphic. Thus, our goal is to develop a procedure for converting isomorphisms of affine varieties to isomorphisms of their corresponding coordinate rings. This procedure is the *pullback* and is the topic of the next section.

Exercises for Section 4.1

4.1.1 Let $X = \mathcal{V}(y^2 - z^2 + xy - z, z^2 - x^3y^2) \subseteq \mathbb{A}^3$. Prove that $F(a) = (1, a, a)$ defines a polynomial map $F : \mathbb{A}^1 \to X$. Is F an isomorphism?

4.1.2 Let $X = \mathcal{V}(x^2 + y^2 - 1) \subseteq \mathbb{A}^2$ and let $Y = \mathcal{V}(u^2 + v^2 - 2) \subseteq \mathbb{A}^2$. Prove that $F(a, b) = (a + b, a - b)$ defines a polynomial map $F : X \to Y$. Is F an isomorphism?

4.1.3 Let $X = \mathcal{V}(y^2 - x^3 - x^2) \subseteq \mathbb{A}^2$. Prove that $F(a) = (a^2 - 1, a^3 - a)$ defines a polynomial map $F : \mathbb{A}^1 \to X$. Is F an isomorphism? (**Hint**: Draw a picture.)

4.1.4 Let $X = \mathcal{V}(f_1, \ldots, f_k, x_m - g) \subseteq \mathbb{A}^m$ where $g \in K[x_1, \ldots, x_{m-1}]$. For each $i = 1, \ldots, k$, define

$$\tilde{f}_i = f_i(x_1, \ldots, x_{m-1}, g) \in K[x_1, \ldots, x_{m-1}],$$

and set $Y = \mathcal{V}(\tilde{f}_1, \ldots, \tilde{f}_k) \subseteq \mathbb{A}^{m-1}$. Prove that $X \cong Y$.

(This shows that if one of the defining equations of an affine variety is linear in one of the variables, then it can be replaced with an isomorphic affine variety defined by fewer equations in fewer variables.)

4.1.5 This exercise shows that the image of a polynomial map may or may not be an affine variety.

(a) Prove that the image of the polynomial map

$$\begin{aligned} F : \mathbb{A}^1 &\to \mathbb{A}^3 \\ a &\mapsto (a, a^2, a^3) \end{aligned}$$

is an affine variety.

(b) Prove that the image of the polynomial map

$$\begin{aligned} G : \mathbb{A}^2 &\to \mathbb{A}^2 \\ (a, b) &\mapsto (a, ab) \end{aligned}$$

is not an affine variety.

4.1.6 Let M be an invertible $m \times m$ matrix with coefficients in K and $c \in \mathbb{A}^m$. Identifying $\mathbb{A}^m$ with the vector space K^m and M with a linear transformation $\varphi_M : K^m \to K^m$, define the function

$$\begin{aligned} F : \mathbb{A}^m &\to \mathbb{A}^m \\ a &\mapsto \varphi_M(a) + c. \end{aligned}$$

Prove the following.

(a) If $X \subseteq \mathbb{A}^m$ is an affine variety, then $F(X) \subseteq \mathbb{A}^m$ is an affine variety.

(b) If $X \subseteq \mathbb{A}^m$ is an affine variety, then $X \cong F(X)$.

4.1.7 Let $X = \mathcal{V}(\ell_1, \ldots, \ell_k) \subseteq \mathbb{A}^n$ where each ℓ_i is a linear polynomial:

$$\ell_i = a_{i1}x_1 + \cdots + a_{in}x_n + b_i \in K[x_1, \ldots, x_n].$$

Let $M = (a_{ij})$ be the $k \times n$ matrix of linear coefficients. Assuming that X is nonempty, prove that

$$X \cong \mathbb{A}^{n - \mathrm{rk}(M)}.$$

(**Hint:** Use the Rank-Nullity Theorem.)

Section 4.2 Pullback homomorphisms

A collection of mathematical objects together with their structure-preserving maps is, loosely speaking, the definition of a category.

Polynomial maps are the structure-preserving maps between varieties in the same way that homomorphisms (of groups, rings, or K-algebras) are the structure-preserving maps between algebraic objects. And just as one can move from geometry to algebra by sending X to $K[X]$, there is a passage from geometry to algebra given by sending a polynomial map between affine varieties to a corresponding K-algebra homomorphism between coordinate rings. This passage is accomplished utilizing the notion of the pullback homomorphism.

4.9 **DEFINITION** *Pullback homomorphism*

Let $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$ be affine varieties, and let $F : X \to Y$ be a polynomial map. The *pullback homomorphism* induced by F is

$$\begin{aligned} F^* : K[Y] &\to K[X] \\ F^*(G) &= G \circ F. \end{aligned}$$

Notice that pulling back changes the direction of the map:

$F : X \to Y \Rightarrow F^* : K[Y] \to K[X]$.

In order for the definition of the pullback homomorphism to make sense, one must verify that, for every $G \in K[Y]$, the composition $G \circ F$ is an element of $K[X]$. Since $G : Y \to K$ and $F : X \to Y$, the definition of compositions implies that $G \circ F$ is, indeed, a function from X to K; schematically:

$$X \xrightarrow{F} Y \xrightarrow{G} K \implies X \xrightarrow{G \circ F} K.$$

The fact that $G \circ F$ is, moreover, a polynomial function follows from the fact that compositions of polynomial functions are polynomial functions (Exercise 4.2.4).

4.10 **EXAMPLE** Pullback homomorphism

Let $X = \mathcal{V}(x_2 - x_1^2) \subseteq \mathbb{A}^2$ and let $Y = \mathcal{V}(y_1 y_2 - y_3) \subseteq \mathbb{A}^3$. Consider the polynomial map $F : X \to Y$ of Example 4.3:

$$F(a_1, a_2) = (a_1 - a_1^2, a_1 + a_2, a_1^2 - a_2^2).$$

Consider the function $G \in K[Y]$ defined by $G(b_1, b_2, b_3) = b_1^2 - b_2 b_3$. Pulling back by F, we obtain the polynomial function $F^*(G) \in K[X]$ defined by

$$(F^*G)(a_1, a_2) = (G \circ F)(a_1, a_2) = (a_1 - a_1^2)^2 - (a_1 + a_2)(a_1^2 - a_2^2).$$

Similarly, pulling back $H \in K[Y]$ defined by $H(b_1, b_2, b_3) = b_1 + b_2 - b_3$, we obtain the polynomial function $F^*(H) \in K[X]$ defined by

$$(F^*H)(a_1, a_2) = (H \circ F)(a_1, a_2) = (a_1 - a_1^2) + (a_1 + a_2) - (a_1^2 - a_2^2).$$

As seen in the previous example, once we have chosen polynomial expressions for $F = (F_1, \ldots, F_n)$ and G, then we obtain a polynomial expression for $F^*(G)$ simply by composing the polynomial expressions for F and G. To expand on this observation, suppose that $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$. Then

$$K[X] = \frac{K[x_1, \ldots, x_m]}{\mathcal{I}(X)} \quad \text{and} \quad K[Y] = \frac{K[y_1, \ldots, y_n]}{\mathcal{I}(Y)}.$$

If we choose polynomials $f_1, \ldots, f_n \in K[x_1, \ldots, x_m]$ and $g \in K[y_1, \ldots, y_m]$ such that $F_i = [f_i]$ and $G = [g]$, then for every $a \in X$ and $b = (b_1, \ldots, b_n) \in Y$,

$$F(a) = (f_1(a), \ldots, f_n(a)) \quad \text{and} \quad G(b_1, \ldots, b_n) = g(b_1, \ldots, b_n).$$

This implies that

$$F^*(G)(a) = G(F(a)) = g(f_1(a), \ldots, f_n(a)).$$

Thus, $F^*(G) = [g(f_1, \ldots, f_n)]$, where $g(f_1, \ldots, f_n) \in K[x_1, \ldots, x_m]$ is the polynomial obtained from g by replacing y_i with $f_i(x_1, \ldots, x_m)$.

4.11 EXAMPLE Pullback homomorphism, revisited

In the same setting as Example 4.10, the component functions of F arise from the polynomials

$$f_1(x_1, x_2) = x_1 - x_1^2, \quad f_2(x_1, x_2) = x_1 + x_2, \quad \text{and} \quad f_3(x_1, x_2) = x_1^2 - x_2^2.$$

The polynomial function $G \in K[Y]$ arises from the polynomial

$$g = y_1^2 - y_2 y_3.$$

As can be seen by visual inspection, the pullback function $F^*(G)$ arises from the polynomial

$$g(f_1, f_2, f_3) = (x_1 - x_1^2)^2 - (x_1 + x_2)(x_1^2 - x_2^2).$$

As the name suggests, the pullback homomorphism is more than just a function; it is a homomorphism of K-algebras, as we now justify.

4.12 PROPOSITION *F^* is a homomorphism*

If $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$ are affine varieties and $F : X \to Y$ is a polynomial map, then

$$F^* : K[Y] \to K[X]$$

is a homomorphism of K-algebras.

PROOF To show that F^* is a K-algebra homomorphism, we must check that it preserves addition, multiplication, and scalar multiplication. We prove the first of these and leave the other two to Exercise 4.2.6.

Students of linear algebra may recognize the pullback as a polynomial generalization of the dual of a linear map; see Exercise 4.2.10.

To see that F^* respects addition, let $G_1, G_2 \in K[Y]$ be two polynomial functions. Evaluating $F^*(G_1 + G_2)$ at any value $a \in X$, we obtain

$$\begin{aligned}
\big(F^*(G_1+G_2)\big)(a) &= \big((G_1+G_2)\circ F\big)(a) && \text{(definition of pullback)}\\
&= (G_1+G_2)(F(a)) && \text{(definition of composition)}\\
&= G_1(F(a)) + G_2(F(a)) && \text{(definition of } + \text{ in } K[Y])\\
&= (G_1\circ F)(a) + (G_2\circ F)(a) && \text{(definition of composition)}\\
&= \big((G_1\circ F) + (G_2\circ F)\big)(a) && \text{(definition of } + \text{ in } K[X])\\
&= (F^*G_1 + F^*G_2)(a). && \text{(definition of pullback)}
\end{aligned}$$

Thus, $F^*(G_1+G_2) = F^*G_1 + F^*G_2$, verifying that F^* preserves addition. □

Recall that our motivation for introducing the pullback homomorphism was to equip ourselves with algebraic tools for determining whether or not two affine varieties are isomorphic. Since the definition of "isomorphism" (in any category) requires checking that the composition of two morphisms is the identity, a preliminary result toward this objective is to prove that pullbacks behave well with respect to compositions and the identity function.

4.13 PROPOSITION *Pullbacks preserve compositions and the identity*

Let $X \subseteq \mathbb{A}^\ell$, $Y \subseteq \mathbb{A}^m$, and $Z \subseteq \mathbb{A}^n$ be affine varieties.

1. If $F : X \to Y$ and $G : Y \to Z$ are polynomial maps, then
$$(G\circ F)^* = F^*\circ G^*.$$
2. The pullback of the identity function is the identity function:
$$(\mathrm{id}_X)^* = \mathrm{id}_{K[X]}.$$

PROOF To prove the first statement, suppose that $H \in K[Z]$. Using associativity of compositions, we then compute

In the language of category theory, these properties of the pullback go by the name "functoriality."

$$(G\circ F)^*(H) = H\circ(G\circ F) = (H\circ G)\circ F = F^*(H\circ G) = F^*(G^*(H)),$$

which shows that $(G\circ F)^* = F^*\circ G^*$. For the second statement, let $H \in K[X]$. Then

$$(\mathrm{id}_X)^*(H) = H\circ \mathrm{id}_X = H,$$

so $(\mathrm{id}_X)^*$ is indeed the identity function on $K[X]$. □

Our task of using coordinate rings to detect whether affine varieties are isomorphic now has its first resolution.

4.14 COROLLARY *Pullbacks of isomorphisms are isomorphisms*

Let $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$ be affine varieties. If $F : X \to Y$ is an isomorphism, then $F^* : K[Y] \to K[X]$ is an isomorphism.

PROOF Let $F : X \to Y$ be an isomorphism with inverse $F^{-1} : Y \to X$. To prove that F^* is an isomorphism, it suffices to prove that F^* and $(F^{-1})^*$ are inverse to each other. We verify this using Proposition 4.13:

$$F^* \circ (F^{-1})^* = (F^{-1} \circ F)^* = (\mathrm{id}_X)^* = \mathrm{id}_{K[X]}$$

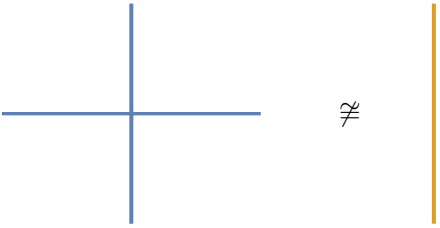
and

$$(F^{-1})^* \circ F^* = (F \circ F^{-1})^* = (\mathrm{id}_Y)^* = \mathrm{id}_{K[Y]}. \qquad \square$$

The converse of Corollary 4.14 is also true, but it requires some additional work to prove, and we defer this discussion until the next section. In the meantime, let us take a look at a few example applications of this result.

4.15 EXAMPLE $\mathcal{V}(xy) \ncong \mathbb{A}^1$

Consider the affine variety $X = \mathcal{V}(xy) \subseteq \mathbb{A}^2$. Let us prove that X is not isomorphic to $\mathbb{A}^1$. Intuitively, this should be somewhat clear from the depiction of the two varieties below. In particular, X consists of two affine lines meeting at a point, which certainly looks quite different than a single affine line.



To make this intuition precise, notice that $X = \mathcal{V}(x) \cup \mathcal{V}(y)$ is a reducible affine variety. Thus, $\mathcal{I}(X)$ is not a prime ideal, implying that

$$K[X] = \frac{K[x,y]}{\mathcal{I}(X)}$$

is not an integral domain. Since $K[\mathbb{A}^1] = K[z]$ is a single-variable polynomial ring, it is an integral domain. Thus, given that the property of being an integral domain is preserved under isomorphism, we see that $K[X] \ncong K[\mathbb{A}^1]$, and we conclude from the contrapositive of Corollary 4.14 that $X \ncong \mathbb{A}^1$.

The previous example quickly generalizes: if one affine variety is irreducible and another is not, then they cannot be isomorphic, because one of their coordinate rings is an integral domain and the other is not. The next example, on the other hand, illustrates an example of proving that two *irreducible* affine varieties are not isomorphic, which can be trickier. This example also concludes the discussion that we began in Example 4.8, that bijective polynomial maps between affine varieties are not necessarily isomorphisms.

4.16 EXAMPLE $V(x^2 - y^3) \ncong \mathbb{A}^1$

As in Example 4.8, let $X = V(x^2 - y^3) \subseteq \mathbb{A}^2$. In order to show that the polynomial bijection $F : \mathbb{A}^1 \to X$ defined by $f(a) = (a^3, a^2)$ does not have a polynomial inverse, we prove that $X \ncong \mathbb{A}^1$. To do so, we analyze the coordinate rings. Since $x^2 - y^3$ is irreducible, the vanishing ideal of X is

$$\mathcal{I}(X) = \langle x^2 - y^3 \rangle.$$

Thus,

$$K[X] = \frac{K[x, y]}{\langle x^2 - y^3 \rangle}.$$

In order to prove that $X \ncong \mathbb{A}^1$, it suffices to find one ring-theoretic property of the single-variable polynomial ring $K[\mathbb{A}^1] = K[z]$ that is not satisfied by $K[X]$. An example of such a property is that $K[X]$ is not a UFD. Indeed, it can be shown (Exercise 0.3.15) that $[x]$ and $[y]$ are distinct irreducible elements of $K[X]$, so the equality

$$[x]^2 = [y]^3$$

expresses the same element in two inequivalent ways as a product of irreducibles.

In this section, we have discussed a way to associate a K-algebra homomorphism to every polynomial map. In particular, we now have an association from the category of affine varieties to the category of K-algebras that is defined on objects and morphisms by

$$X \longmapsto K[X]$$
$$(F : X \to Y) \longmapsto (F^* : K[Y] \to K[X]).$$

Along with the conditions in Proposition 4.13, such an association is called a *functor* between these categories. In the next section, we show that the association of morphisms is invertible: every K-algebra homomorphism between coordinate rings arises from a unique polynomial map between the corresponding affine varieties. This is a key step in proving the converse to Corollary 4.14.

Exercises for Section 4.2

4.2.1 Let $F : \mathbb{A}^2 \to \mathbb{A}^2$ be the polynomial map defined by

$$F(a_1, a_2) = (a_1^2 + 2a_1a_2, a_1 - a_2).$$

Calculate $F^*(G)$, where $G \in K[\mathbb{A}^2]$ is the function defined by

$$G(b_1, b_2) = b_1 - b_2^2.$$

4.2.2 Let $Y = \mathcal{V}(y_3 - y_1^3) \subseteq \mathbb{A}^3$, and let $F : \mathbb{A}^1 \to Y$ be the polynomial map defined by

$$F(a) = (a, a^2, a^3).$$

Calculate $F^*([g]) \in K[\mathbb{A}^1]$, where $g = y_1y_2 + y_3^2 \in K[y_1, y_2, y_3]$.

4.2.3 Let $X, Y \subseteq \mathbb{A}^n$ be affine varieties with $X \subseteq Y$, and let $F : X \to Y$ be the inclusion. Describe $F^* : K[Y] \to K[X]$.

4.2.4 Let X and Y be affine varieties, and let $F : X \to Y$ be a function. Prove the following: if F is a polynomial map, then $G \circ F \in K[X]$ for any $G \in K[Y]$.

4.2.5 Let X and Y be affine varieties, and let $F : X \to Y$ be a function. Prove the following: if $F \circ G \in K[X]$ for all $G \in K[Y]$, then F is a polynomial map. (This is the converse of the previous problem.)

4.2.6 Let $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$ be affine varieties and $F : X \to Y$ a polynomial map. Prove that $F^* : K[Y] \to K[X]$ preserves ring multiplication and scalar multiplication.

4.2.7 For each of the following pairs of affine varieties X and Y, decide whether $X \cong Y$ and prove your answer.

(a) $X = \mathbb{A}^1$ and $Y = \mathbb{A}^2$;
(b) $X = \mathcal{V}(x_1, x_2) \subseteq \mathbb{A}^2$ and $Y = \mathcal{V}(y_1 - y_2, y_1^2 - y_2) \subseteq \mathbb{A}^2$;
(c) $X = \mathcal{V}(x_3 - x_1^2 + x_1x_2) \subseteq \mathbb{A}^3$ and $Y = \mathbb{A}^2$;
(d) $X = \mathcal{V}(x_1x_2) \subseteq \mathbb{A}^2$ and $Y = \mathcal{V}(y_1^2 - y_2^2) \subseteq \mathbb{A}^2$;

4.2.8 Prove that $\mathcal{V}(xy - 1) \subseteq \mathbb{A}^2$ is not isomorphic to $\mathbb{A}^1$.

4.2.9 Let $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$ be affine varieties, and let $F : X \to Y$ be a polynomial map.

(a) Prove that if F is surjective, then F^* is injective.
(b) Based on part (a), we might hope that if F is injective, then F^* is surjective. Prove that this is false by showing that the polynomial map

$$\begin{aligned} F : \mathcal{V}(xy - 1) &\to \mathbb{A}^1 \\ F(a, b) &= a \end{aligned}$$

is injective but F^* is not surjective.

4.2.10 (For students with some knowledge of linear algebra) Let V and W be K-vector spaces, and let $F : V \to W$ be a linear map. Choose bases for V and W in order to fix isomorphisms $V \cong K^m$ and $W \cong K^n$; via these isomorphisms, we can identify V and W with affine spaces $\mathbb{A}^m_K$ and $\mathbb{A}^n_K$, and hence as affine varieties. Denoting by $V^\vee$ and $W^\vee$ the dual vector spaces, explain why

$$V^\vee \subseteq K[V] \quad \text{and} \quad W^\vee \subseteq K[W],$$

and verify that $F^*|_{W^\vee}$ coincides with the dual map $F^\vee : W^\vee \to V^\vee$.

Section 4.3 Pulling back is a bijection

If $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$ are affine varieties, then the association that takes a polynomial map to its pullback provides a function between two sets of morphisms:

$$\{\text{polynomial maps } X \to Y\} \to \{K\text{-algebra homomorphisms } K[Y] \to K[X]\}$$
$$F \mapsto F^*.$$

The main result of this section is that this function is a bijection.

4.17 PROPOSITION *Pulling back is a bijection*

Let $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$ be affine varieties. The correspondence $F \mapsto F^*$ is a bijection between the set of polynomial maps $X \to Y$ and the set of K-algebra homomorphisms $K[Y] \to K[X]$.

In the language of category theory, the bijection of Proposition 4.17 says that the functor taking X to $K[X]$ and F to F^* is fully faithful.

To prove Proposition 4.17, it suffices to produce an inverse to the procedure that takes F to F^*. That is, given any K-algebra homomorphism $\varphi : K[Y] \to K[X]$, it suffices to show that there is a unique polynomial map $F : X \to Y$ such that $F^* = \varphi$. The proof of this statement can be notationally hard to follow, so we begin with a concrete example.

4.18 EXAMPLE Inverting the pullback

Consider the affine varieties of Example 4.3:

$$X = \mathcal{V}(x_2 - x_1^2) \subseteq \mathbb{A}^2 \quad \text{and} \quad Y = \mathcal{V}(y_1y_2 - y_3) \subseteq \mathbb{A}^3.$$

Using that $x_2 - x_1^2$ and $y_1y_2 - y_3$ are both irreducible, we compute

$$K[X] = \frac{K[x_1, x_2]}{\langle x_2 - x_1^2 \rangle} \quad \text{and} \quad K[Y] = \frac{K[y_1, y_2, y_3]}{\langle y_1y_2 - y_3 \rangle}.$$

Let us consider a K-algebra homomorphism $\varphi : K[Y] \to K[X]$. Such a homomorphism is determined by sending each generator $[y_i]$ to $[f_i]$ for some polynomial $f_i \in K[x_1, x_2]$; for example, consider the homomorphism

$$\begin{aligned} \varphi : K[Y] &\to K[X] \\ \varphi([y_1]) &= [x_1 + x_2] \\ \varphi([y_2]) &= [x_1] \\ \varphi([y_3]) &= [x_1x_2 + x_2]. \end{aligned}$$

The image of any element $[g] \in K[Y]$ is determined by the fact that φ is a K-algebra homomorphism; for example,

$$\varphi([y_1^2 + y_2 - y_2y_3]) = [(x_1 + x_2)^2 + x_1 - x_1(x_1x_2 + x_2)],$$

and more generally, for any $g \in K[y_1, y_2, y_3]$,

$$\varphi([g]) = [g(x_1 + x_2, x_1, x_1x_2 + x_2)].$$

Not any choice of the three polynomials f_1, f_2, f_3 would have given a well-defined K-algebra homomorphism; it must be the case that $[0] = [y_1y_2 - y_3]$ is sent to $[0] \in K[X]$, or in other words that $f_1f_2 - f_3$ lies in $\langle x_2 - x_1^2 \rangle$. This is indeed the case for our particular choice of f_1, f_2, f_3:

$$(x_1 + x_2)x_1 - (x_1x_2 + x_2) = x_1^2 - x_2 \in \langle x_2 - x_1^2 \rangle. \tag{4.1}$$

Having described φ, can we find a polynomial map $F : X \to Y$ for which $F^* = \varphi$? Such a map sends elements of $X \subseteq \mathbb{A}^2$ to elements of $Y \subseteq \mathbb{A}^3$, so it is defined by three polynomials in two variables. Can you think of any candidates for three such polynomials? There is a natural choice: the three polynomials f_1, f_2, f_3 that were used to define φ. In other words, consider the function

$$\begin{aligned} F : X &\to Y \\ F(a_1, a_2) &= (a_1 + a_2,\ a_1,\ a_1a_2 + a_2). \end{aligned}$$

While F is manifestly a polynomial map, we should confirm that it indeed sends elements of X to elements of Y, or in other words that $F(a_1, a_2)$ satisfies the defining equation of Y. Explicitly, we verify that

$$(a_1 + a_2)a_1 - (a_1a_2 + a_2) = a_1^2 - a_2 = 0, \tag{4.2}$$

where the last equality is because $(a_1, a_2) \in X = \mathcal{V}(x_2 - x_1^2)$. Note the similarity in equations (4.1) and (4.2): what was needed in order to verify that φ was well-defined was precisely what was needed in order to verify that F mapped X to Y.

Finally, to confirm that $F^* = \varphi$, it suffices to check that these two homomorphisms agree on the generators $[y_1], [y_2], [y_3]$ of $K[Y]$. Viewing $[y_i]$ as a function $Y \to K$, it is simply the coordinate function

$$\begin{aligned} [y_i] : Y &\to K \\ [y_i](b_1, b_2, b_3) &= b_i. \end{aligned}$$

Thus, by definition of the pullback,

$$F^*([y_1])(a_1, a_2) = ([y_1] \circ F)(a_1, a_2) = [y_1](a_1 + a_2, a_1, a_1a_2 + a_2) = a_1 + a_2.$$

In other words,

$$F^*([y_1]) = [x_1 + x_2],$$

implying that $F^*([y_1]) = \varphi([y_1])$. Similarly, F^* agrees with φ on the other two generators, so we have successfully constructed a polynomial map F for which $F^* = \varphi$. Moreover, the construction essentially illustrates the uniqueness of F: the requirement that $F^*([y_i]) = \varphi([y_i])$ determines the ith component function of F, and these component functions uniquely determine F.

Generalizing the ideas of this example, we are ready to prove Proposition 4.17.

PROOF OF PROPOSITION 4.17 Let $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$ be affine varieties with

$$K[X] = \frac{K[x_1, \ldots, x_m]}{\mathcal{I}(X)} \quad \text{and} \quad K[Y] = \frac{K[y_1, \ldots, y_n]}{\mathcal{I}(Y)}.$$

In order to prove the proposition, we show that, for every K-algebra homomorphism $\varphi : K[Y] \to K[X]$, there exists a unique polynomial map $F : X \to Y$ with $F^* = \varphi$.

(**Existence**) Suppose $\varphi : K[Y] \to K[X]$ is a K-algebra homomorphism. Following the procedure in the example, let $F_i = \varphi([y_i]) \in K[X]$, and consider the polynomial map

$$F = (F_1, \ldots, F_n) : X \to \mathbb{A}^n.$$

As in the example, we must verify that the image of F lies in Y and therefore gives rise to a polynomial map $F : X \to Y$. To prove this, suppose that $a \in X$; we must show that

$$(F_1(a), \ldots, F_n(a)) \in Y.$$

Since $Y = \mathcal{V}(\mathcal{I}(Y))$, it suffices to check that, for all $h \in \mathcal{I}(Y)$,

$$h(F_1(a), \ldots, F_n(a)) = 0.$$

Let $h \in \mathcal{I}(Y)$. Then $[h] = 0 \in K[Y]$, so $\varphi([h]) = 0 \in K[X]$, since φ is a homomorphism. But this implies that

$$\varphi([h(y_1, \ldots, y_n)]) = h(\varphi([y_1]), \ldots, \varphi([y_n])) = h(F_1, \ldots, F_n)$$

is the zero function on X, implying that $h(F_1(a), \ldots, F_n(a)) = 0$. Thus, F is indeed a polynomial map from X to Y.

It remains to show that $F^* = \varphi$. As in the example, for every $i = 1, \ldots, n$,

$$F^*([y_i]) = [y_i] \circ (F_1, \ldots, F_n) = F_i = \varphi([y_i]).$$

Thus, F^* and φ agree on the generators $[y_i]$, implying that $F^* = \varphi$.

(**Uniqueness**) Suppose $F, G : X \to Y$ are polynomial maps with $F^* = G^*$. We must show that $F = G$. By definition,

$$F = (F_1, \ldots, F_n) \quad \text{and} \quad G = (G_1, \ldots, G_n)$$

where $F_i, G_i \in K[X]$ are polynomial functions on X. Evaluating F^* and G^* on $[y_i]$, we have

$$F^*([y_i]) = [y_i] \circ (F_1, \ldots, F_n) = F_i \in K[X],$$

and, similarly, $G^*([y_i]) = G_i \in K[X]$. Since $F^* = G^*$, it follows that

$$F_i = F^*([y_i]) = G^*([y_i]) = G_i$$

for all $i = 1, \ldots, n$, so $F = G$. □

The payoff for the work undertaken to prove Proposition 4.17 is that we can now precisely detect whether two affine varieties are isomorphic by studying their coordinate rings.

4.19 COROLLARY *Coordinate rings detect isomorphism*

Let $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$ be affine varieties. Then

$$X \cong Y \iff K[X] \cong K[Y].$$

PROOF The forward implication is the content of Corollary 4.14. To prove the other direction, suppose that $\varphi : K[Y] \to K[X]$ is an isomorphism with inverse φ^{-1}. By Theorem 4.17, there exist polynomial maps $F : X \to Y$ and $G : Y \to X$ such that $F^* = \varphi$ and $G^* = \varphi^{-1}$. We claim that F and G are inverse to one another. Indeed, by Proposition 4.13,

$$(F \circ G)^* = G^* \circ F^* = \varphi^{-1} \circ \varphi = \mathrm{id}_{K[Y]} = (\mathrm{id}_Y)^*.$$

Since $F \circ G$ and id_Y have the same pullback, the bijectivity of Theorem 4.17 implies that $F \circ G = \mathrm{id}_Y$. Similarly, $G \circ F = \mathrm{id}_X$. Since F and G are inverse polynomial maps, we conclude that $F : X \to Y$ is an isomorphism and $F^* = \varphi$. □

Exercises for Section 4.3

4.3.1 Suppose that $m \leq n$ and consider the natural injection

$$\varphi : K[x_1, \ldots, x_m] \to K[x_1, \ldots, x_n].$$

Describe the corresponding polynomial map $F : \mathbb{A}^n \to \mathbb{A}^m$.

4.3.2 Suppose that $m \leq n$ and consider the surjection

$$\begin{aligned} \varphi : K[x_1, \ldots, x_n] &\to K[x_1, \ldots, x_m] \\ f(x_1, \ldots, x_n) &\mapsto f(x_1, \ldots, x_m, 0, \ldots, 0). \end{aligned}$$

Describe the corresponding polynomial map $F : \mathbb{A}^m \to \mathbb{A}^n$.

4.3.3 Consider the homomorphism of K-algebras

$$\varphi : K[x, y] \to K[t]$$

defined by

$$\varphi(f) = f(t + 1, t^2 + t).$$

For which affine varieties X and Y and which polynomial map $F : X \to Y$ do we have $\varphi = F^*$?

4.3.4 Let $X = \mathcal{V}(x^2 + y^2 + z^2 - 1) \subseteq \mathbb{A}^3_{\mathbb{R}}$, for which

$$K[X] = \frac{K[x, y, z]}{\langle x^2 + y^2 + z^2 - 1 \rangle}.$$

Consider the homomorphism of K-algebras

$$\begin{aligned} \varphi : K[u, v] &\to K[X] \\ \varphi(f) &= [f(x + y + z, xyz)]. \end{aligned}$$

For which polynomial map $F : X \to \mathbb{A}^2$ do we have $\varphi = F^*$?

4.3.5 Let $f_1, f_2, f_3 \in K[w]$ and let $X = \mathcal{V}(x + y - z) \subseteq \mathbb{A}^3$.

(a) Under what conditions on f_1, f_2, and f_3 does

$$\begin{aligned} F : \mathbb{A}^1 &\to X \\ F(a) &= (f_1(a), f_2(a), f_3(a)) \end{aligned}$$

give a well-defined polynomial map to X? Give an explicit example of $f_1, f_2, f_3 \in K[w]$ for which this is the case, and an explicit example for which it is not the case.

(b) Under what conditions on f_1, f_2, and f_3 does there exist a well-defined K-algebra homomorphism

$$\varphi : \frac{K[x, y, z]}{\langle x + y - z \rangle} \to K[w]$$

defined on the generators by

$$\begin{aligned} \varphi([x]) &= f_1(w) \\ \varphi([y]) &= f_2(w) \\ \varphi([z]) &= f_3(w)? \end{aligned}$$

Give an explicit example of $f_1, f_2, f_3 \in K[w]$ for which this is the case, and an explicit example for which it is not the case.

(c) Using parts (a) and (b), explicitly describe the bijection between the set of polynomial maps $\mathbb{A}^1 \to X$ and the set of K-algebra homomorphisms $K[X] \to K[\mathbb{A}^1]$.

4.3.6 Let $X \subseteq \mathbb{A}^m$ be an affine variety, and let $Y \subseteq \mathbb{A}^n$ be a single point. There is only one possible polynomial map $X \to Y$, so by Theorem 4.17, there is only one possible K-algebra homomorphism $K[Y] \to K[X]$. What is $K[Y]$, and what is the one K-algebra homomorphism $K[Y] \to K[X]$?

4.3.7 Suppose that $X \subseteq A^m$ is an irreducible affine variety and that $Y \subseteq \mathbb{A}^n$ consists of two distinct points. Prove that there are exactly two polynomial maps $F : X \to Y$. Describe the two maps explicitly.

4.3.8 If $X \subseteq \mathbb{A}^n$ is any affine variety, then the set of polynomial maps $X \to \mathbb{A}^1$ is precisely $K[X]$, so Theorem 4.17 implies there is a bijection between $K[X]$ and the set of K-algebra homomorphism $K[\mathbb{A}^1] \to K[X]$. Describe this bijection explicitly.

Section 4.4 The equivalence of algebra & geometry

Combining the results of this chapter and the previous one, we now prove that the passage from X to $K[X]$ truly is a dictionary between affine varieties and finitely-generated reduced K-algebras, where we view objects on both sides as "the same" if they are isomorphic. In the terminology of *isomorphism classes*—the equivalence classes under the equivalence relation of being isomorphic (which is an equivalence relation in the setting of K-algebras, of affine varieties, or more generally, in any category)—the results we have proven lead to the following theorem.

4.20 **THEOREM** *Equivalence of algebra & geometry*

The association $X \mapsto K[X]$ induces a bijection

$$\left\{\begin{array}{c}\text{isomorphism classes of}\\ \text{affine varieties}\end{array}\right\} \longrightarrow \left\{\begin{array}{c}\text{isomorphism classes of}\\ \text{finitely-generated reduced } K\text{-algebras}\end{array}\right\}.$$

PROOF Recalling that each coordinate ring is a finitely-generated reduced K-algebra (Proposition 3.32), we can view the association $X \to K[X]$ as a function

$$\{\text{affine varieties}\} \longrightarrow \left\{\begin{array}{c}\text{isomorphism classes of}\\ \text{finitely-generated reduced } K\text{-algebras}\end{array}\right\}.$$

To see that this function is well-defined on isomorphism classes of affine varieties, we notice that $K[X] \cong K[Y]$ whenever $X \cong Y$—this is one direction of Corollary 4.19. To see that this function is injective on isomorphism classes, we notice that $X \cong Y$ whenever $K[X] \cong K[Y]$—this is the other direction of Corollary 4.19. Finally, to justify surjectivity, notice that every finitely-generated reduced K-algebra is the coordinate ring of some affine variety (Proposition 3.32). □

In category-theoretic language, the bijection of Theorem 4.20 reflects an equivalence of categories *between affine varieties and finitely-generated reduced K-algebras.*

Put more loosely, Theorem 4.20 asserts that all of the geometric information about the affine variety X is encoded in the K-algebra $K[X]$. But perhaps we should be a bit more careful: the particular affine space $\mathbb{A}^n$ in which X lives is "geometric information" about X, and yet this information cannot be recovered from the isomorphism class of $K[X]$. For instance, the coordinate rings of $\mathcal{V}(y-x^2) \subseteq \mathbb{A}^2$ and $\mathcal{V}(y-x^2,z) \subseteq \mathbb{A}^3$ are members of the same isomorphism class despite arising from affine varieties in different ambient affine spaces.

The reason the ambient affine space of X cannot be recovered from the isomorphism class of $K[X]$ is that isomorphic affine varieties can live in different affine spaces. Such a property of affine varieties—or indeed, of any mathematical objects—that is not preserved under isomorphism can be thought of as a "coincidental" property, one that depends on some extraneous choice. By contrast, a property preserved by isomorphisms is one that pertains to the object's "essence." To make these ideas precise, it is useful to have the following definition.

4.21 **DEFINITION** *Intrinsic/extrinsic property*

Let $\mathcal{C}$ be a class of mathematical objects with a notion of isomorphism. A property $\mathcal{P}$ is said to be *intrinsic* if, whenever two objects are isomorphic, one of them has property $\mathcal{P}$ if and only if the other has property $\mathcal{P}$. A property that is not intrinsic is said to be *extrinsic*.

More precisely, $\mathcal{C}$ is a category, and $\mathcal{P}$ is a subset of the objects of $\mathcal{C}$.

The property of being inside $\mathbb{A}^2$, for instance, is an extrinsic property on the class of affine varieties, since $\mathcal{V}(y-x^2) \subseteq \mathbb{A}^2$ has it but the isomorphic variety $\mathcal{V}(y-x^2,z) \subseteq \mathbb{A}^3$ does not. Here are some further examples of intrinsic and extrinsic properties, both in the context of algebra and geometry.

4.22 **EXAMPLE** Intrinsic properties of rings

Being reduced is an intrinsic property of rings. To prove this, let $\varphi : R \to S$ be a ring isomorphism; we must show that R is reduced if and only if S is reduced. Assume that R is not reduced. Then there exists $a \in R$ such that $a \neq 0$ and $a^m = 0$ for some $m \geq 1$. Using standard properties of ring isomorphisms, we see that $\varphi(a) \neq 0$ and

$$\varphi(a)^m = \varphi(a^m) = \varphi(0) = 0.$$

Thus, $\varphi(a)$ is a nonzero nilpotent, showing that S is not reduced. The same proof applied to $\varphi^{-1} : S \to R$ shows that, if S is not reduced, then R is not reduced.

Similar arguments show that being an integral domain, a UFD, a PID, or a field are all intrinsic properties of rings (Exercise 4.4.1).

4.23 **EXAMPLE** Number of generators is extrinsic

Every finitely-generated K-algebra is isomorphic to a quotient $K[x_1,\ldots,x_n]/I$, but the number n of generators is extrinsic. For example, the K-algebras

$$\frac{K[x,y]}{\langle y-x^2\rangle} \quad \text{and} \quad \frac{K[x,y,z]}{\langle y-x^2,z\rangle}$$

are isomorphic, even though the first has two generators and the second has three. On the other hand, the *minimal* number of generators is an intrinsic property.

4.24 **EXAMPLE** Irreducibility is intrinsic

The property of being an irreducible affine variety is intrinsic. This can be proved directly using the definition of polynomial maps and irreducibility (Exercise 4.4.2), but we can also prove it using our dictionary between geometry and algebra. To do so, suppose $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$ are affine varieties and $X \cong Y$. Then

$$\begin{aligned} X \text{ is irreducible} &\iff \mathcal{I}(X) \text{ is radical} && \text{(Proposition 2.25)} \\ &\iff K[X] \text{ is an integral domain} && \text{(Propositions 3.21 and 0.38).} \end{aligned}$$

By assumption, $X \cong Y$, and therefore $K[X] \cong K[Y]$ (Corollary 4.19). Since being an integral domain is an intrinsic property of K-algebras, we conclude that X is irreducible if and only if Y is irreducible. Thus, irreducibility is an intrinsic property.

An intrinsic property $\mathcal{P}$ of mathematical objects of class $\mathcal{C}$ can be viewed as a subset of the set of isomorphism classes, consisting of those isomorphism classes in which one (and hence every) representative has property $\mathcal{P}$. In particular, an intrinsic property of affine varieties—such as irreducibility—can be viewed as a subset of the set of isomorphism classes of affine varieties. Via the bijection of Theorem 4.20, this can be identified with a subset of the set of isomorphism classes of finitely-generated reduced K-algebras, which can then be viewed as an intrinsic property of finitely-generated reduced K-algebras. Which algebraic property is it? In the case of irreducibility, the answer is that irreducibility of affine varieties corresponds to the property of a finitely-generated reduced K-algebra being an integral domain.

More generally, we can now ask a very broad question: given an intrinsic geometric property or construction applicable to affine varieties, what is its manifestation in the category of K-algebras? Or, conversely, given an intrinsic algebraic property or construction applicable to K-algebras, what is its manifestation in the category of affine varieties? Both algebra and geometry are illuminated by these questions, and specific examples of such phenomena form the backbone of the algebraic geometry to come.

Exercises for Section 4.4

4.4.1 Prove that the following are intrinsic properties of rings:

(a) Being an integral domain;
(b) Being a field;
(c) Being a principal ideal domain;
(d) Being a unique factorization domain.

4.4.2 Let $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$ be affine varieties, and let $F : X \to Y$ be an isomorphism.

(a) Prove that, if a subset $X_1 \subseteq X$ is an affine variety, then $F(X_1) \subseteq Y$ is an affine variety.
(b) Prove directly from Definition 2.22 that irreducibility is intrinsic.

4.4.3 Prove that the number of irreducible components of an affine variety is an intrinsic property.

4.4.4 Prove that being a finite set is an intrinsic property of affine varieties. Describe the corresponding intrinsic property of finitely-generated reduced K-algebras.

4.4.5 We say that two morphisms $F_1 : A_1 \to B_1$ and $F_2 : A_2 \to B_2$ are *isomorphic* if there exist isomorphisms $G_1 : A_1 \to A_2$ and $G_2 : B_1 \to B_2$ such that

$$F_1 = G_2^{-1} \circ F_2 \circ G_1.$$

(a) Prove that isomorphism is an equivalence relation on the set of morphisms.
(b) Prove that there is a bijection between isomorphism classes of polynomial maps between affine varieties and isomorphism classes of K-algebra homomorphisms between finitely-generated reduced K-algebras.

4.4.6 (a) Let $Y \subseteq \mathbb{A}^n$ be an affine variety and let $X \subseteq Y$ be a subset. We say that X is *dense* in Y if there does not exist an affine variety $Z \subseteq \mathbb{A}^n$ such that $X \subseteq Z \subsetneq Y$. Prove that X is dense in Y if and only if the only polynomial function $G \in K[Y]$ that vanishes on X is the zero function.

(b) Let $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$ be affine varieties. We say that a polynomial map $F : X \to Y$ is *dominant* if $F(X)$ is dense in Y. Prove that a polynomial map is dominant if and only if its pullback is injective.

4.4.7 Let $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$ be affine varieties. We say that a polynomial map $F : X \to Y$ is a *closed embedding* if there exists an affine variety $Z \subseteq Y$ such that $F(X) = Z$ and the induced map $F : X \to Z$ is an isomorphism. Prove that a polynomial map is a closed embedding if and only if its pullback is surjective.

Chapter 5

Proof of the Nullstellensatz

LEARNING OBJECTIVES FOR CHAPTER 5

- Generalize the notion of vector spaces (over fields) to modules (over rings).
- Extend the notion of algebras over fields to algebras over rings.
- Explore the difference between finitely-generated algebras and finitely-generated modules.
- Determine, via the concept of integrality, when a finitely-generated algebra is in fact a finitely-generated module.
- Investigate the general structure of finitely-generated K-algebras via Noether normalization.
- Use Noether normalization to prove the Nullstellensatz.

Now that we have collected, in the form of the equivalence of algebra and geometry, some evidence of the power of the Nullstellensatz, the time has come to prove it. The journey to a proof of the Nullstellensatz necessitates a rather long interlude into purely algebraic material, including a tour of R-modules and R-algebras, culminating with the Noether Normalization Theorem in Section 5.4.

In addition to being a key step in the proof of the Nullstellensatz, the Noether Normalization Theorem is a powerful result of independent interest about the structure of finitely-generated K-algebras. It says that, while such an algebra certainly need not be finitely-generated as a K-vector space (for example, the polynomial ring $K[x_1, \ldots, x_n]$ is not), it can always be expressed as a finitely-generated "vector space" over a subalgebra that is isomorphic to a polynomial ring. We put the word "vector space" in quotes here because these scalars do not form a field, and hence, in order to make sense of Noether Normalization, we must generalize the definition of a vector space to allow scalars from a general ring. These generalized vector spaces, which we introduce in Section 5.1, are referred to as *modules*.

While the basic definition of a module is no different from that of a vector space, the theory in this setting leads to a number of new ideas, the most important of which is the notion of integrality and its relationship to finite generation. The first four sections of this chapter are devoted to the development of the theory of R-modules and R-algebras, culminating in Section 5.4 with a proof of Noether Normalization. In Section 5.5, we receive the payoff for this work: the proof of the Nullstellensatz, and thus, a complete justification of the equivalence of algebra and geometry.

Section 5.1 Modules

To motivate the concept of modules, we begin with a discussion of the algebraic structure of a few familiar coordinate rings. Consider

$$K[x,y] \quad \text{and} \quad K[x,y]/\langle x^2+y^2-1\rangle,$$

which are the coordinate rings of the affine plane and the unit circle, respectively. The affine plane certainly does not feel like it should be isomorphic to the unit circle, suggesting that there must be some algebraic property that we can use to distinguish between these K-algebras. Our aim is to describe such a property using ideas from linear algebra.

To start, consider these coordinate rings as vector spaces. Notice that every element of $K[x,y]$ can be written uniquely as a K-linear combination of elements

$$\mathcal{B} = \{x^i y^j \mid i,j \in \mathbb{N}\},$$

which tells us that $K[x,y]$ is an infinite-dimensional vector space over K with basis $\mathcal{B}$. If we consider the ring $K[x,y]/\langle x^2+y^2-1\rangle$, on the other hand, then we can repeatedly use the relation $[y^2] = [1-x^2]$ to write every element uniquely as

$$[f(x)+g(x)y]$$

for some $f,g \in K[x]$. In other words, $K[x,y]/\langle x^2+y^2-1\rangle$ is also an infinite-dimensional vector space over K, but it has a basis given by the smaller set

$$\mathcal{B}' = \{[x^i y^j] \mid i \in \mathbb{N},\ j \in \{0,1\}\}.$$

Even though $\mathcal{B}'$ can be viewed as a proper subset of $\mathcal{B}$, both $\mathcal{B}$ and $\mathcal{B}'$ are countably infinite, which implies that these two vector spaces are, in fact, isomorphic. Since the two coordinate rings are isomorphic as vector spaces, we see that the theory of vector spaces alone is not enough to distinguish between them.

However, if we allow ourselves to enlarge our "scalars," replacing K with the ring $R = K[x]$, then we notice that every element of $K[x,y]$ can be written uniquely as an R-linear combination of elements of the infinite set

$$\mathcal{S} = \{1, y, y^2, y^3, \dots\},$$

whereas, for the ring $K[x,y]/\langle x^2+y^2-1\rangle$, in order to write every element as an R-linear combination, we only require the two-element set

$$\mathcal{S}' = \{[1],[y]\}.$$

In other words, if we pretend for a moment that $R = K[x]$ is a field (it's not!), then we have observed that $K[x,y]$ is an infinite-dimensional "vector space" over R while $K[x,y]/\langle x^2+y^2-1\rangle$ is finite-dimensional. Thus, we have succeeded in finding a distinguishing property between these two coordinate rings.

To make this hypothetical argument a reality, we require an extension of the notion of vector spaces to the setting where the scalars are allowed to be taken to be a ring but not necessarily a field, a setting that is captured by the important algebraic concept of modules. We begin our discussion of modules in this section with the definition and some basic notions. As always, R denotes a ring, and all rings are assumed to be commutative with unity.

5.1 DEFINITION *R-module*

An R-module is an abelian group M (with operation denoted $+$) together with a *scalar multiplication* function

$$R \times M \to M$$
$$(r, a) \mapsto r \cdot a$$

satisfying the following axioms:

1. $r \cdot (a + b) = r \cdot a + r \cdot b$ for all $r \in R$ and all $a, b \in M$;
2. $(r + s) \cdot a = r \cdot a + s \cdot a$ for all $r, s \in R$ and all $a, b \in M$;
3. $(rs) \cdot a = r \cdot (s \cdot a)$ for all $r, s \in R$ and all $a \in M$;
4. $1 \cdot a = a$ for all $a \in M$, where $1 \in R$ is the multiplicative identity.

When $R = K$ is a field, Definition 5.1 is nothing more than the definition of a vector space over K. Many, but not all, of the notions of vector spaces naturally generalize to the module setting. Let us begin our discussion of modules with several examples that will be helpful to keep in mind.

5.2 EXAMPLE R^n is an R-module

The standard example of a vector space is K^n, and this generalizes to the R-module setting. More specifically, consider the Cartesian product

$$R^n = \{(a_1, \ldots, a_n) \mid a_i \in R \text{ for each } i\}.$$

The set R^n is naturally an R-module, with addition and scalar multiplication defined exactly as in the vector space setting:

$$(a_1, \ldots, a_n) + (b_1, \ldots, b_n) = (a_1 + b_1, \ldots, a_n + b_n),$$
$$r \cdot (a_1, \ldots, a_n) = (ra_1, \ldots, ra_n).$$

5.3 EXAMPLE Polynomial rings are modules

The polynomial ring $R[x_1, \ldots, x_n]$ is an R-module, with addition and scalar multiplication defined in the usual way:

$$\Big(\sum_\alpha b_\alpha x^\alpha\Big) + \Big(\sum_\alpha c_\alpha x^\alpha\Big) = \sum_\alpha (b_\alpha + c_\alpha) x^\alpha \quad \text{and} \quad r \cdot \Big(\sum_\alpha b_\alpha x^\alpha\Big) = \sum_\alpha (rb_\alpha) x^\alpha.$$

The module axioms are a straightforward consequence of the ring axioms.

5.4 EXAMPLE Extension rings

Generalizing the previous example, if $R \subseteq S$ is a subring, then S can be viewed as an R-module, where for $r \in R$ and $s \in S$, we define $r \cdot s$ by the ring multiplication inside S. The module axioms, again, are a consequence of the ring axioms.

As a special case that arose in the discussion at the beginning of this section, we can consider $M = K[x, y]$ as a module over the subring $R = K[x]$. That is, elements in $K[x, y]$ can be added as usual and any element in $K[x, y]$ can be multiplied by a "scalar" in $K[x]$.

5.5 EXAMPLE Abelian groups are $\mathbb{Z}$-modules

Let M be any abelian group. Then M can be viewed as a $\mathbb{Z}$-module, where for $n \in \mathbb{Z}$ and $a \in M$ the scalar multiplication is defined by

$$n \cdot a = \begin{cases} \underbrace{a + \cdots + a}_{n \text{ times}} & \text{if } n > 0 \\ 0 & \text{if } n = 0 \\ \underbrace{(-a) + \cdots + (-a)}_{-n \text{ times}} & \text{if } n < 0. \end{cases}$$

In fact, one can check from the module axioms that this is the only way to define scalar multiplication that makes M into a $\mathbb{Z}$-module; see Exercise 5.1.10. It follows that every abelian group is a $\mathbb{Z}$-module in a *canonical* way.

Conversely, every $\mathbb{Z}$-module is an abelian group by forgetting scalar multiplication. Thus, $\mathbb{Z}$-modules and abelian groups are really two different names for the same thing.

Of course, our discussion of R-modules is not complete without introducing the relevant notion of morphisms between them. Given that a module is an abelian group with the additional structure of scalar multiplication, it is natural to define a module homomorphism as a group homomorphism that preserves scalar multiplication.

5.6 DEFINITION *Homomorphisms of R-modules*

Let M and N be R-modules. An *R-module homomorphism* $\varphi : M \to N$ is a group homomorphism for which

$$\varphi(r \cdot a) = r \cdot \varphi(a)$$

for all $r \in R$ and $a \in M$. We say that φ is an *isomorphism of R-modules* and write $M \cong N$ if φ has an inverse that is also an R-module homomorphism.

In other words, an R-module homomorphism satisfies

$$\varphi(r \cdot a) = r \cdot \varphi(a) \quad \text{and} \quad \varphi(a + b) = \varphi(a) + \varphi(b)$$

for all $a, b \in M$ and $r \in R$. In particular, if $R = K$ is a field, so that M and N are K-vector spaces, then a K-module homomorphism is precisely the same thing as a *linear map* of vectors spaces over K.

Another important module-theoretic notion is that of a submodule.

5.7 DEFINITION *Submodule*

Let M be an R-module. A *submodule* $N \subseteq M$ is a subgroup for which $r \cdot a \in N$ for all $r \in R$ and $a \in N$.

It can be checked (using that R has unity) that a subset $N \subseteq M$ is a submodule if and only if it is closed under the two operations:

$$a + b \in N \quad \text{and} \quad r \cdot a \in N$$

for all $a, b \in N$ and $r \in R$. Thus, the notion of submodules naturally generalizes the notion of linear subspaces from the study of vector spaces (see Exercise 5.1.6).

Given an R-module M and a submodule $N \subseteq M$, we can form the group quotient M/N, and this group quotient naturally inherits the structure of an R-module, with scalar multiplication defined by

$$r \cdot [a] = [r \cdot a].$$

The reader is encouraged to check that scalar multiplication is well-defined and that the quotient M/N satisfies the R-module axioms (Exercise 5.1.7). Just like for groups, rings, and K-algebras, there is a version of the First Isomorphism Theorem for R-modules; this is the content of Exercise 5.1.8.

In the definition of an R-module, we started with an additive abelian group. However, in many cases relevant to us, such as the setting of polynomial rings and their quotients, the additive abelian group will also have a multiplicative structure that endows it with the structure of a ring. In this case, we call the resulting structure an *R-algebra*, made precise in the following definition.

5.8 **DEFINITION** *R-algebra*

An *R-algebra* is a ring A together with a scalar multiplication function

$$\begin{aligned} R \times A &\to A \\ (r, a) &\mapsto r \cdot a \end{aligned}$$

satisfying the four axioms of an R-module as well as

$$r \cdot (ab) = (r \cdot a)b = a(r \cdot b)$$

for all $r \in R$ and all $a, b \in A$.

The reader should notice that the definition of an R-algebra is not new—after replacing R with K, it is identical to the definition of a K-algebra from Section 3.2. In fact, most of the concepts we discussed concerning K-algebras—such as homomorphisms, subalgebras, ideals, quotients, the First Isomorphism Theorem, and generators—carry over verbatim to the R-algebra setting, and we do not restate them here. As was the case for K-algebras, the prototypical R-algebra is the polynomial ring $R[x_1, \ldots, x_n]$.

In contrast to the setting of K-algebras, however, we gain some flexibility in our perspective now that we do not require our scalars to form a field. For example, even if our motivation is to study polynomials over a field, we now have the ability to view one of the variables as a "scalar" and write

$$K[x_1, \ldots, x_n] = R[x_1, \ldots, x_{n-1}] \quad \text{where} \quad R = K[x_n].$$

This opens up the possibility of proving assertions concerning K-algebras by using induction arguments in the more general R-algebra setting. It is essentially for this reason that the setting of R-algebras is the correct level of algebraic generality that we require for our development of algebraic geometry.

Given an R-algebra, there is a unique underlying R-module obtained by forgetting the multiplicative structure. On the other hand, if you start with an R-module, then there are typically many ways to put a multiplicative structure on it to endow it with the structure of an R-algebra. We illustrate this in the next example.

5.9 EXAMPLE Different R-algebras with the same underlying R-module

Consider the R-module $M = R^2$. A natural way to make M into an R-algebra is to define multiplication componentwise:

$$(a,b) \cdot (c,d) = (ac, bd).$$

However, this is not the only way that we can make M into an R-algebra; another way is given by defining multiplication as follows:

$$(a,b) \cdot (c,d) = (ac, ad + bc).$$

While this second multiplication might feel a bit strange at first glance, the resulting R-algebra is actually isomorphic to the familiar quotient $R[x]/\langle x^2 \rangle$, as the reader is encouraged to verify in Exercise 5.1.9.

In the next section, we turn to a discussion of module generators, which allows us to generalize the important notion of finite-dimensionality from linear algebra to the module setting.

Exercises for Section 5.1

5.1.1 Let $M = K[x,y]$. Find at least three different rings R for which M is an R-module.

5.1.2 Give an example of a ring R, an R-module M, and a subgroup $N \subseteq M$ that is not an R-module.

5.1.3 Prove that R is an $R[x]$-module under the scalar multiplication defined by

$$(r_0 + r_1 x + r_2 x^2 + \cdots + r_n x^n) \cdot a = r_0 a.$$

5.1.4 Let R be a ring and M an R-module. Use the module axioms to prove that

$$0 \cdot a = 0 \text{ for all } a \in M$$

and

$$r \cdot 0 = 0 \text{ for all } r \in R.$$

5.1.5 Prove that an R-module homomorphism is an isomorphism if and only if it is bijective.

5.1.6 Let M be an R-module and $N \subseteq M$ a subset. Prove that N is a submodule if and only if

$$a + b \in N \quad \text{and} \quad r \cdot a \in N \quad \text{for all} \quad a, b \in N \quad \text{and} \quad r \in R.$$

5.1.7 Let M be an R-module, $N \subseteq M$ a submodule, and M/N the group quotient.

(a) Suppose that $[a_1] = [a_2] \in M/N$. Prove that

$$[r \cdot a_1] = [r \cdot a_2]$$

Conclude that scalar multiplication is well-defined in M/N.

(b) Prove that M/N satisfies the R-module axioms.

5.1.8 Let $\varphi : M \to N$ be a homomorphism of R-modules.

(a) Prove that $\ker(\varphi)$ is a submodule of M.

(b) Prove that $\operatorname{im}(\varphi)$ is a submodule of N.

(c) Prove that the function

$$\begin{aligned} [\varphi] : M/\ker(\varphi) &\to \operatorname{im}(\varphi) \\ [a] &\mapsto \varphi(a) \end{aligned}$$

is a well-defined isomorphism of R-modules.

5.1.9 Let A be the R-algebra defined by endowing R^2 with the multiplication

$$(a, b) \cdot (c, d) = (ac, ad + bc).$$

Prove that $A \cong R[x]/\langle x^2 \rangle$.

5.1.10 Let M be an abelian group. Prove that the only definition of scalar multiplication that makes M into a $\mathbb{Z}$-module is the one given in Example 5.5.

Section 5.2 Module generators

As we learned in the previous section, a module is an algebraic structure that generalizes vector spaces to the setting where the scalars form a ring but not necessarily a field. In this section, we generalize the important vector space concept of finite-dimensionality to the module setting. The key notions we require for this generalization are those of linear combinations and generators.

5.10 **DEFINITION** *Linear combination, generators*

Let M be an R-module and let $\mathcal{S} \subseteq M$ be a subset. A *linear combination of* $\mathcal{S}$ is an element of M of the form

$$r_1 a_1 + \cdots + r_n a_n$$

for some $n \in \mathbb{N}$, $r_i \in R$, and $a_i \in \mathcal{S}$. The set of all linear combinations of $\mathcal{S}$ is called the *submodule of M generated by $\mathcal{S}$*, and it is denoted $R\mathcal{S}$.

It is a worthwhile exercise to verify that $R\mathcal{S}$ is, in fact, a submodule of M, and that it is the smallest submodule of M that contains the set $\mathcal{S}$ (Exercise 5.2.1). Let us consider a few examples.

5.11 **EXAMPLE** Submodules of $R[x]$

Consider $R[x]$ as an R-module. Then the submodule generated by $\{x^2, x^3\}$ is

$$R\{x^2, x^3\} = \{ax^2 + bx^3 \mid a, b \in R\} \subseteq R[x].$$

In other words, it consists of polynomials whose only potentially nonzero coefficients occur in the x^2 and x^3 terms. Similarly,

$$R\{1, x^2, x^4, x^6, \dots\} = \left\{ \sum_{i=0}^{n} a_i x^{2i} \mid n \in \mathbb{N},\ a_i \in R \right\} \subseteq R[x]$$

consists of polynomials whose nonzero coefficients occur with even powers of x.

5.12 **EXAMPLE** Submodules of $\mathbb{Z}$

Consider $\mathbb{Z}$ as a $\mathbb{Z}$-module. Then the submodule generated by $\{4, 6\}$ is

$$\mathbb{Z}\{4, 6\} = \{a \cdot 4 + b \cdot 6 \mid a, b \in \mathbb{Z}\}.$$

Noting that $2 = (-1) \cdot 4 + 1 \cdot 6 \in \mathbb{Z}\{4, 6\}$, it is not too hard to see that every even integer can be obtained as a linear combination of 4 and 6, which proves that $2\mathbb{Z} \subseteq \mathbb{Z}\{4, 6\}$. On the other hand, every linear combination of 4 and 6 is even, so $\mathbb{Z}\{4, 6\} \subseteq 2\mathbb{Z}$. Taken together, we have proved that

$$\mathbb{Z}\{4, 6\} = 2\mathbb{Z}.$$

If instead, we consider the submodule generated by 2 and 3, we see that

$$1 = (-1) \cdot 2 + 1 \cdot 3 \in \mathbb{Z}\{2, 3\},$$

which implies that $\mathbb{Z}\{2, 3\} = \mathbb{Z}$.

5.13 **EXAMPLE** The coordinate ring of the unit circle as a $K[x]$-module

Consider the coordinate ring of the unit circle $X = \mathcal{V}(x^2 + y^2 - 1) \subseteq \mathbb{A}^2$:

$$K[X] = \frac{K[x,y]}{\langle x^2 + y^2 - 1 \rangle}.$$

We can view $K[X]$ as a $K[x]$-module in a natural way by defining

$$f \cdot [g] = [fg]$$

for any $f \in K[x]$ and $[g] \in K[X]$. By repeated use of the equation $[y^2] = [1 - x^2]$, every element of $K[X]$ can be written in the form

$$[f_1(x) + f_2(x)y] = f_1(x) \cdot [1] + f_2(x) \cdot [y],$$

which shows that $K[X] = K[x]\{[1], [y]\}$.

We are especially interested in whether a module can be generated by a finite set, generalizing the concept of finite-dimensionality from the study of vector spaces.

5.14 **DEFINITION** *Finitely-generated module*

We say that M is a *finitely-generated R-module* if there exist $a_1, \ldots, a_n \in M$ such that

$$M = R\{a_1, \ldots, a_n\}.$$

Example 5.13 shows that the coordinate ring of the unit circle is a finitely-generated $K[x]$-module, generated by $[1]$ and $[y]$. The next example illustrates a familiar module that is not finitely-generated.

5.15 **EXAMPLE** $R[x]$ is not a finitely-generated R-module

Consider the polynomial ring $R[x]$. To prove that $R[x]$ is not finitely-generated, suppose $f_1, \ldots, f_n \in R[x]$ is any finite set of polynomials and consider the submodule

$$R\{f_1, \ldots, f_n\} \subseteq R[x].$$

We must prove that this submodule is not all of $R[x]$. To do so, let d be the maximum degree of the polynomials $f_1, \ldots, f_n$. Then any linear combination of these polynomials must have degree bounded above by d. In particular, $x^{d+1} \notin R\{f_1, \ldots, f_n\}$.

Further differences between modules and vector spaces are discussed in Exercise 5.2.7.

In many ways, modules behave like vector spaces, but it is important to note their key differences. The reader might recall a standard result in linear algebra that says that every finitely-generated vector space over K is isomorphic to K^n for some n. In the module setting, this is not the case; for example, given a nontrivial finite group M, we may view it as a $\mathbb{Z}$-module (Example 5.5), and it is finitely-generated because it is generated by all of its elements. However, it is not the case that $M \cong \mathbb{Z}^n$ for any n because $1 < |M| < \infty$, but $\mathbb{Z}^n$ is either infinite (if $n > 0$) or has a single element (if $n = 0$).

In practice, most of the modules in this book will arise naturally with a multiplicative operation, giving them the structure of an algebra. Given an R-algebra, we can talk about its module properties, which pertain to just addition and scalar multiplication (in other words, *linear* algebra), or we can talk about its algebra properties, which also include the multiplication operation (in other words, *polynomial* algebra). We contrast these two perspectives in the next example.

5.16 **EXAMPLE** Submodule versus subalgebra generated by a set

In Example 5.11, we saw that the submodule of $R[x]$ generated by x^2 and x^3 is

$$R\{x^2, x^3\} = \{ax^2 + bx^3 \mid a, b \in R\}.$$

To contrast this with the algebra setting, let's consider the subalgebra generated by these same two elements. As defined in Section 3.3, the subalgebra $R[x^2, x^3]$ consists of all polynomial combinations of x^2 and x^3, so, in addition to containing the linear combinations as above, it contains additional elements, such as

$$(x^2)^2 = x^4, \quad x^2 \cdot x^3 = x^5 \quad \text{and} \quad (x^3)^2 = x^6.$$

In fact, one can show (Exercise 5.2.2) that $R[x^2, x^3]$ consists of all polynomials in $R[x]$ in which the linear coefficient is zero:

$$R[x^2, x^3] = R\{1, x^2, x^3, x^4, \dots\}.$$

Thus, we see that the subalgebra generated by x^2 and x^3 is much larger than the submodule; it is not even finitely-generated as a module.

If A is an R-algebra and $\mathcal{S} \subseteq A$ is a subset, then

$$R\mathcal{S} \subseteq R[\mathcal{S}],$$

simply because every linear combination is a special type of polynomial combination. It follows that, if A is finitely-generated as a module, then it must be finitely-generated as an algebra. On the other hand, given a finitely-generated algebra, it is usually not the case that it is finitely-generated as a module; the polynomial ring $R[x_1, \dots, x_n]$, for example, is finitely-generated as an algebra by $x_1, \dots, x_n$, but not finitely-generated as a module. Thus, being finitely-generated in the module sense is much more restrictive than being finitely-generated in the algebra sense.

In light of this, it is useful to ask whether a given finitely-generated algebra is finitely-generated as a module. For example, if we consider $\mathbb{R}$ as a $\mathbb{Z}$-module and choose a real number $a \in \mathbb{R}$, then the subalgebra $\mathbb{Z}[a]$ is finitely-generated as a $\mathbb{Z}$-algebra. We ask: Is it finitely-generated as a module? In the next two examples, we investigate this question for two different values of a.

5.17 **EXAMPLE** $\mathbb{Z}[\sqrt{2}]$ is a finitely-generated $\mathbb{Z}$-module

Consider the real numbers $\mathbb{R}$ as a $\mathbb{Z}$-algebra, and let us investigate the finitely-generated subalgebra

$$\mathbb{Z}[\sqrt{2}] \subseteq \mathbb{R}.$$

Elements of $\mathbb{Z}[\sqrt{2}]$ are those real numbers that can be obtained as a polynomial combination $f(\sqrt{2})$ for some $f \in \mathbb{Z}[x]$. Consider, for example, the polynomial

$$f = 3 + 5x + 4x^2 - x^3.$$

Then, by definition, $f(\sqrt{2}) \in \mathbb{Z}[\sqrt{2}]$. Simplifying, we see that

$$\begin{aligned} f(\sqrt{2}) &= 3 + 5\sqrt{2} + 4(\sqrt{2})^2 - (\sqrt{2})^3 \\ &= 3 + 5\sqrt{2} + 4 \cdot 2 - 2\sqrt{2} \\ &= 11 + 3\sqrt{2}. \end{aligned}$$

Thus, the polynomial combination $f(\sqrt{2})$ is the same as evaluating the linear polynomial $11 + 3x$ at $\sqrt{2}$. Generalizing this same trick to any polynomial, it can be shown that $\mathbb{Z}[\sqrt{2}]$ is a finitely-generated $\mathbb{Z}$-module (Exercise 5.2.4):

$$\mathbb{Z}[\sqrt{2}] = \mathbb{Z}\{1, \sqrt{2}\}.$$

5.18 **EXAMPLE** $\mathbb{Z}[1/2]$ is not a finitely-generated $\mathbb{Z}$-module

Consider again the real numbers $\mathbb{R}$ as a $\mathbb{Z}$-algebra and let us investigate the finitely-generated subalgebra

$$\mathbb{Z}[1/2] \subseteq \mathbb{R}.$$

This subalgebra consists of polynomial combinations of $1/2$, so taking, for example, $f = 3 + 5x + 4x^2 - x^3$, we see that

$$f(1/2) = 3 + 5(1/2) + 4(1/4) - (1/8) = 75/8 \in \mathbb{Z}[1/2].$$

Notice that, for a polynomial f of degree d, the largest power of 2 that will appear in a denominator of one of the terms in $f(1/2)$ is 2^d. This implies that, upon combining the terms and writing the rational number $f(1/2)$ as a reduced fraction, the denominator will not be divisible by 2^{d+1}. We now use this observation to prove that $\mathbb{Z}[1/2]$ is not finitely-generated as a $\mathbb{Z}$-module.

Suppose, toward a contradiction, that $\mathbb{Z}[1/2] = \mathbb{Z}\{a_1, \ldots, a_n\}$. Since each a_i is an element of $\mathbb{Z}[1/2]$, we know that $a_i = f_i(1/2)$ for some polynomial f_i in $\mathbb{Z}[x]$. Let d be the maximum degree of the f_i. Then, upon writing each a_i as a reduced fraction, none of the denominators is divisible by 2^{d+1}. Since taking $\mathbb{Z}$-linear combinations will never introduce additional powers of 2 in the denominators, after reducing, this proves that

$$\frac{1}{2^{d+1}} \notin \mathbb{Z}\{a_1, \ldots, a_n\}.$$

However, since $1/2^{d+1} = f(1/2)$ for $f = x^{d+1} \in \mathbb{Z}[x]$, it follows that

$$\frac{1}{2^{d+1}} \in \mathbb{Z}[1/2],$$

proving that

$$\mathbb{Z}\{a_1, \ldots, a_n\} \neq \mathbb{Z}[1/2],$$

a contradiction.

Let us pause to ponder the previous two examples. In both examples, we considered a $\mathbb{Z}$-algebra generated by a single real number. We might expect these two algebras to be very similar, but one of them turned out to be a finitely-generated module while the other did not. What, then, is the distinction between the numbers $\sqrt{2}$ and $1/2$ that led to this very different behavior? In the next section, we answer this question by giving a general criterion for determining whether a finitely-generated algebra is actually finitely-generated as a module.

Exercises for Section 5.2

5.2.1 Let M be an R-module and $\mathcal{S} \subseteq M$ a subset. Prove the following.

(a) The set $R\mathcal{S}$ is a submodule of M.

(b) If $N \subseteq M$ is any submodule containing $\mathcal{S}$, then $R\mathcal{S} \subseteq N$.

5.2.2 Prove that

$$R[x^2, x^3] = R\{1, x^2, x^3, x^4, \dots, \}.$$

5.2.3 Consider $\mathbb{Z}$ as a $\mathbb{Z}$-module and let $a, b \in \mathbb{Z}$. Prove that

$$\mathbb{Z}\{a, b\} = \gcd(a, b)\mathbb{Z}.$$

5.2.4 Prove that $\mathbb{Z}[\sqrt{2}] = \mathbb{Z}\{1, \sqrt{2}\}$.

5.2.5 Prove that $\mathbb{Z}[\pi]$ is not finitely-generated as a $\mathbb{Z}$-module.

5.2.6 Let $R \subseteq S \subseteq T$ be rings. Prove that if S is a finitely-generated R-module and T is a finitely-generated S-module, then T is a finitely-generated R-module.

5.2.7 Recall from Example 5.12 that $\mathbb{Z}$ is generated as a $\mathbb{Z}$-module by the set $\{2, 3\}$.

(a) Prove that $\{2, 3\}$ is a minimal generating set, in the sense that no proper subset of $\{2, 3\}$ generates $\mathbb{Z}$ as a $\mathbb{Z}$-module.

(b) Prove that, although every element of $\mathbb{Z}$ can be expressed as $r_1 \cdot 2 + r_2 \cdot 3$ for some $r_1, r_2 \in \mathbb{Z}$, this expression is not unique.

(c) Prove that $\{1\}$ also generates $\mathbb{Z}$ as a $\mathbb{Z}$-module, and that it is a minimal generating set.

(This exercise highlights two differences between R-modules and vector spaces. First, if V is a vector space, then a minimal generating set for V is necessarily a *basis*, meaning a set in terms of which every element of V can be expressed uniquely. By parts (a) and (b), this is not the case for R-modules. Second, if V is a vector space, then every minimal generating set has the same size. Parts (a) and (c) show that this is not the case for R-modules.)

Section 5.3 Integrality

Consider a ring inclusion $R \subseteq S$, and let us view S as an R-module. We ask the question: Under what conditions is S a finitely-generated R-module? At the end of the last section, we studied two examples of this setup:

1. $R = \mathbb{Z}$ and $S = \mathbb{Z}[\sqrt{2}]$, and
2. $R = \mathbb{Z}$ and $S = \mathbb{Z}[1/2]$.

In the first case, we observed that $\mathbb{Z}[\sqrt{2}]$ is, in fact, a finitely-generated $\mathbb{Z}$-module, whereas in the second case, we argued that $\mathbb{Z}[1/2]$ is not. Looking back at those examples, one major difference we see between $\sqrt{2}$ and $1/2$ is that taking powers of $\sqrt{2}$ eventually brings us to an element of $\mathbb{Z}$, while taking powers of $1/2$ never brings us back to $\mathbb{Z}$. Indeed, the fact that $(\sqrt{2})^2 = 2$ is what allowed us to reduce all polynomial expressions in $\sqrt{2}$ to linear polynomials.

The goal of this section is to formalize the above observation for general rings. The key new concept for this discussion is the notion of *integrality*. For the following definition, recall that a polynomial is *monic* if its leading coefficient is one; that is, a monic polynomial in $R[x]$ has the form

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

for some $a_0, \ldots, a_{n-1} \in R$.

5.19 **DEFINITION** *Algebraic and integral elements*

Let $R \subseteq S$ be rings, and let $a \in S$. We say that a is *algebraic over* R if there exists a polynomial $f \in R[x]$ such that $f(a) = 0 \in S$. If, moreover, there exists a monic polynomial $f \in R[x]$ such that $f(a) = 0$, then we say that a is *integral over* R.

If R is a field, then an element is algebraic if and only if it is integral, since we can simply divide any polynomial by its leading term to obtain a monic polynomial. In more general rings, where division may not make sense, being integral is stronger than being algebraic.

Let us consider several examples in the setting where $R = \mathbb{Z}$ and $S = \mathbb{R}$.

5.20 **EXAMPLE** $\sqrt{2}$ is integral over $\mathbb{Z}$

Since $a = \sqrt{2}$ is a root of the monic polynomial

$$x^2 - 2 \in \mathbb{Z}[x],$$

we see that $\sqrt{2}$ is integral over $\mathbb{Z}$.

5.21 **EXAMPLE** π is not algebraic over $\mathbb{Z}$

At some point in your mathematical journey, you may have learned that π is a *transcendental number*, which implies that it does not satisfy any polynomial equations over $\mathbb{Z}$. Thus, π is not algebraic over $\mathbb{Z}$.

5.22 **EXAMPLE** $1/2$ is algebraic but not integral over $\mathbb{Z}$

The element $a = 1/2$ is algebraic over $\mathbb{Z}$, since it is a root of the polynomial

$$g(x) = 2x - 1 \in \mathbb{Z}[x],$$

but it is not integral. This is not immediately obvious; although g is not monic, one might still hope that a monic polynomial with $1/2$ as a root exists. But if

$$h(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}[x]$$

were such a polynomial, then multiplying both sides of the equation $h(1/2) = 0$ by 2^{n-1} would yield

$$\frac{1}{2} + a_{n-1} + a_{n-2} \cdot 2 + a_{n-3} \cdot 2^2 + \cdots + a_1 \cdot 2^{n-2} + a_0 \cdot 2^{n-1} = 0.$$

Moving everything but the first term to the right-hand side, we have expressed $1/2$ as sum of integers, which is impossible, proving that $1/2$ is not integral.

More generally, a rational number $a \in \mathbb{Q}$ is integral over $\mathbb{Z}$ if and only if $a \in \mathbb{Z}$ (Exercise 5.3.2).

It may happen that every element of S is algebraic, or even integral, over R. When this occurs, we use the following terminology.

5.23 **DEFINITION** *Algebraic and integral extensions*

Let $R \subseteq S$ be rings. We say that S is *algebraic over R* (respectively, *integral over R*) if every element of S is algebraic (respectively, integral) over R.

5.24 **EXAMPLE** $\mathbb{Z}[\sqrt{2}]$ is integral over $\mathbb{Z}$

In order to prove that $\mathbb{Z}[\sqrt{2}]$ is integral over $\mathbb{Z}$, we must show that *every* element of $\mathbb{Z}[\sqrt{2}]$ is integral over $\mathbb{Z}$. To do this, first recall (Example 5.17) that

$$\mathbb{Z}[\sqrt{2}] = \{r + s\sqrt{2} \mid r, s \in \mathbb{Z}\}.$$

Thus, given an element of $\mathbb{Z}[\sqrt{2}]$ we can write it as $r + s\sqrt{2}$ for some $r, s \in \mathbb{Z}$. Squaring, we obtain the equation

$$(r + s\sqrt{2})^2 = (r^2 + 2s^2) + (2rs)\sqrt{2}.$$

Rearranging and squaring again yields the equation

$$\Big((r + s\sqrt{2})^2 - (r^2 + 2s^2)\Big)^2 = (2rs)^2 \cdot 2.$$

This last equation implies that $r + s\sqrt{2}$ is a root of the monic polynomial

$$f(x) = \big(x^2 - (r^2 + 2s^2)\big)^2 - (2rs)^2 \cdot 2 \in \mathbb{Z}[x].$$

Thus, every element of $\mathbb{Z}[\sqrt{2}]$ is integral over $\mathbb{Z}$.

5.25 **EXAMPLE** $\mathbb{R}$ is not algebraic over $\mathbb{Z}$

Since the real numbers contain transcendental numbers, such as π, we conclude that $\mathbb{R}$ is not algebraic over $\mathbb{Z}$.

5.26 **EXAMPLE** $\mathbb{Q}$ is algebraic but not integral over $\mathbb{Z}$

The ring $\mathbb{Q}$ is algebraic over the subring $\mathbb{Z}$ because any element $a = p/q \in \mathbb{Q}$ is a root of a polynomial

$$f(x) = qx - p \in \mathbb{Z}[x].$$

However, $\mathbb{Q}$ is not integral over $\mathbb{Z}$, by Example 5.22.

We are now ready to return to our goal of determining when a finitely-generated R-algebra is finitely-generated as an R-module, which is closely related to the question of integrality. In particular, the next result tells us that a finitely-generated R-algebra is finitely-generated as an R-module if and only if it is integral over R. Moreover, in order to check integrality, it suffices to check that the algebra generators are integral over R.

5.27 **THEOREM** *Finite generation and integrality*

Let $R \subseteq S$ be rings with $S = R[a_1, \ldots, a_n]$. The following are equivalent:

(i) S is a finitely-generated R-module;

(ii) S is integral over R;

(iii) a_i is integral over R for each $i = 1, \ldots, n$.

Before we begin the proof, we mention that the arguments involve certain manipulations of matrices whose entries come from the ring R, and the reader is unlikely to have previously worked with matrices in this generality. All such manipulations (matrix-vector products, for example, or determinants of matrices) are defined by the same formulas that define them in the more familiar setting where the entries come from $\mathbb{R}$ or some other field. These definitions make sense with entries in any ring because they involve only sums and products of element. We assume the reader has some familiarity with matrix computations.

PROOF OF THEOREM 5.27 We prove (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i).

(i) $\Rightarrow$ **(ii)**: Suppose that S is a finitely-generated R-module, which means that there exist $v_1, \ldots, v_m \in S$ such that $S = R\{v_1, \ldots, v_m\}$. We must prove that an arbitrary element $b \in S$ is integral over R. To do so, first multiply b by each of the module generators and express the product as a linear combination of these generators:

$$bv_i = c_{i1}v_1 + c_{i2}v_2 + \cdots + c_{im}v_m,$$

where $c_{ij} \in R$ for $i, j = 1, \ldots, m$. Moving all the terms of each of these equations

to the left-hand side, we have a system of equations

$$\begin{aligned}(b-c_{11})v_1+(-c_{12})v_2+\cdots+(-c_{1m})v_m&=0\\(-c_{21})v_1+(b-c_{22})v_2+\cdots+(-c_{2m})v_m&=0\\&\vdots\\(-c_{m1})v_1+(-c_{12})v_2+\cdots+(b-c_{mm})v_m&=0.\end{aligned}$$

Such a system is more conveniently expressed in matrix-vector form: if C is the matrix whose (i,j) entry is c_{ij}, then we have

$$(bI-C)\cdot\begin{pmatrix}v_1\\\vdots\\v_m\end{pmatrix}=\begin{pmatrix}0\\\vdots\\0\end{pmatrix}, \tag{5.1}$$

where I is the $m\times m$ identity matrix.

At this point, we appeal to *Cramer's Rule*, a result about matrices that the reader may have seen when studying linear algebra. We state it here and direct the reader to Exercise 5.3.6 for a proof that assumes a few basic properties of determinants.

5.28 **LEMMA** *Cramer's Rule*

Let A be an $m\times m$ matrix with entries in a ring S. Let $\vec{v},\vec{w}\in S^m$, which we view as column vectors, and suppose that

$$A\vec{v}=\vec{w}.$$

Then, for all $i=1,\ldots,m$, we have

$$\det(A)\cdot v_i=\det(A_i),$$

where A_i is the matrix obtained from A by replacing the ith column by $\vec{w}$.

Equipped with this tool, we apply it to the matrix-vector equation (5.1) to obtain

$$\det(bI-C)\cdot v_i=0 \tag{5.2}$$

for all i; notice, here, that the right-hand side is zero because it is the determinant of a matrix with a column of zeroes. Recalling that v_i are generators for S as an R-module, we can express the element $1\in S$ as a linear combination of $v_1,\ldots,v_m$:

$$1=d_1v_1+\cdots+d_mv_m.$$

Multiplying both sides by $\det(bI-C)$ and applying (5.2) yields

$$\det(bI-C)=0.$$

This implies that b is a root of the polynomial

$$f(x)=\det(xI-C)\in R[x].$$

Notice that the leading term of $f(x) = \det(xI - C)$ as a polynomial in x is the product of the diagonal entries:

$$f(x) = \prod_{i=1}^{m}(x - c_{ii}) + \text{lower-order terms.}$$

From this, we see that $f(x)$ is monic, proving that b is integral over R.

(ii) $\Rightarrow$ (iii): If S is integral over R, then every element of S is integral over R, so in particular, each a_i is integral.

(iii) $\Rightarrow$ (i): Suppose that each a_i is integral over R, so there exist monic polynomials $f_1, \ldots, f_n \in R[x]$ such that $f_i(a_i) = 0$ for each i. Let $d_i > 0$ denote the degree of f_i. Our aim is to prove that S is generated as an R-module by the finite set

$$\mathcal{T} = \{a_1^{k_1} \cdots a_n^{k_n} \mid 0 \leq k_i < d_i\}.$$

The first step in proving that $S = R\mathcal{T}$ is to prove that, for every $i = 1, \ldots, n$,

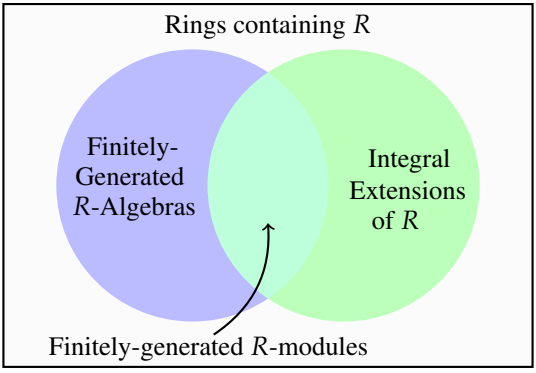
$$R[a_i] = R\{a_i^{k_i} \mid 0 \leq k_i < d_i\}. \tag{5.3}$$

This step follows from an induction argument, using the relation $f_i(a_i) = 0$ to reduce the degree of polynomial expressions in a_i (Exercise 5.3.7), similarly to the argument used for $\mathbb{Z}[\sqrt{2}]$. Once (5.3) is established, it then follows that, for any $\ell_i \geq 0$, we can write $a_i^{\ell_i} \in R[a_i]$ as an R-linear combination of $\{a_i^{k_i} \mid 0 \leq k_i < d_i\}$. Multiplying these linear combinations together, and expanding, we then see that, for any $\ell_1, \ldots, \ell_n \geq 0$, the element

$$a_1^{\ell_1} \cdots a_n^{\ell_n} \in S \tag{5.4}$$

can be written as an R-linear combination of elements in $\mathcal{T}$. Since every element of $S = R[a_1, \ldots, a_n]$ can be written as an R-linear combination of expressions of the form (5.4), we conclude that every element of S can be written as an R-linear combination of elements in $\mathcal{T}$, proving that $S = R\mathcal{T}$, as desired. □

In general, for a ring extension $R \subseteq S$, the ring S need not be finitely-generated, either as an R-module or as an R-algebra, nor does it need to be integral. However, Theorem 5.27 tells us that those extensions that are both finitely-generated as R-algebras and are integral over R are exactly the same as those that are finitely-generated as R-modules, which we illustrate in the following diagram.



Exercises for Section 5.3

5.3.1 Is the element $\frac{\sqrt{2}}{2} \in \mathbb{R}$ algebraic over $\mathbb{Z}$? Is it integral over $\mathbb{Z}$?

5.3.2 Prove that $a \in \mathbb{Q}$ is integral over $\mathbb{Z}$ if and only if $a \in \mathbb{Z}$.

5.3.3 Prove that $\mathbb{C}$ is integral (or equivalently, algebraic) over $\mathbb{R}$. (**Hint**: Mimic Example 5.24.)

5.3.4 For each of the following rings, let $a = [x]$ be the coset of x. Determine whether each ring is integral over the subring $R[a]$?

(a) $R[x,y]$
(b) $R[x,y]/\langle y^2 - x^2 \rangle$
(c) $R[x,y]/\langle xy \rangle$
(d) $R[x,y,z]/\langle y^2 - x^3, z^3 - x^4 \rangle$

5.3.5 Since, by Example 5.26, $\mathbb{Q}$ is not integral over $\mathbb{Z}$, it follows from Theorem 5.27 that $\mathbb{Q}$ is not a finitely-generated $\mathbb{Z}$-module. Verify this directly by showing that, for any $a_1, \ldots, a_n \in \mathbb{Q}$, there is a strict containment

$$\mathbb{Z}\{a_1, \ldots, a_n\} = \{r_1 a_1 + \cdots + r_n a_n \mid r_1, \ldots, r_n \in \mathbb{Z}\} \subsetneq \mathbb{Q}.$$

5.3.6 This exercise proves Cramer's rule, assuming some basic familiarity with matrix multiplication and determinants. Let A be an $m \times m$ matrix with entries in a ring S and let $\vec{v}, \vec{w} \in R^m$ be vectors such that $A\vec{v} = \vec{w}$.

(a) Let I_i denote the matrix obtained from the $m \times m$ identity matrix I by replacing the ith column by $\vec{v}$. Prove that

$$AI_i = A_i,$$

where A_i is the obtained from A by replacing the ith column by $\vec{w}$.

(b) Use your favorite method for computing determinants to prove that

$$\det(I_i) = v_i,$$

then use the multiplicativity of determinants to conclude that

$$\det(A) \cdot v_i = \det(A_i).$$

5.3.7 Let $R \subseteq S$ be rings and let $a \in S$. Suppose that there is a degree d monic polynomial $f \in R[x]$ such that $f(a) = 0$. Prove that

$$R[a] = R\{1, a, a^2, \ldots, a^{d-1}\}.$$

Where does your proof fail if a is algebraic, but not integral, over R?

5.3.8 Let $R \subseteq S$ be rings. Prove that the set $\{s \in S \mid s \text{ is integral over } R\}$ is a subring of S containing R. (This is called the *integral closure* of R in S.)

(**Hint:** Use Theorem 5.27.)

5.3.9 Let $R = \mathbb{Z}$. Describe examples of rings S that lie in every region of the Venn diagram from this section.

Section 5.4 Noether normalization

In the last three sections, we developed a number of module-theoretic notions, and we now bring those developments to bear on the particular type of algebraic objects that are of interest in algebraic geometry: finitely-generated K-algebras.

The Noether Normalization Theorem is a structural result about all finitely-generated K-algebras. To motivate the statement, recall that the prototype of a finitely-generated K-algebra is the polynomial ring

$$K[x_1, \ldots, x_d].$$

Not every finitely-generated K-algebra is isomorphic to a polynomial ring, and the Noether Normalization Theorem seeks to answer the question: How closely can we "approximate" finitely-generated K-algebras with polynomial rings? As we will see, the answer is that, given a finitely-generated K-algebra A, we can always find a subalgebra $B \subseteq A$ such that

1. $B \cong K[x_1, \ldots, x_d]$ for some d, and
2. A is a finitely-generated B-module or, equivalently, A is integral over B.

In other words, the Noether Normalization Theorem ensures that every finitely-generated K-algebra is finitely-generated as a module (the smallest kind of ring extension) over a polynomial ring (the simplest type of K-algebra).

Before stating and proving the Noether Normalization Theorem, we pause to introduce the notion of algebraic independence, which generalizes linear independence to the setting of polynomial algebra and will help us discuss a criterion for when a subalgebra is isomorphic to a polynomial ring.

5.29 **DEFINITION** *Algebraically (in)dependent*

Let A be an R-algebra. We say that $\mathcal{S} \subseteq A$ is *algebraically dependent over* R if there exists $a_1, \ldots, a_d \in \mathcal{S}$ and a nonzero polynomial $f \in R[x_1, \ldots, x_d]$ such that

$$f(a_1, \ldots, a_d) = 0.$$

If no such $a_1, \ldots, a_d$ and f exist, we say that $\mathcal{S}$ is *algebraically independent*.

Algebraically independent elements generate subalgebras that are isomorphic to polynomial rings, as described in the next result, which is an application of the First Isomorphism Theorem (Exercise 5.4.1).

5.30 **PROPOSITION** *Algebraically independent generators*

Let A be an R-algebra and let $a_1, \ldots, a_d \in A$. Then $\{a_1, \ldots, a_d\}$ is algebraically independent over R if and only if the evaluation map

$$\begin{aligned} \varphi : R[x_1, \ldots, x_d] &\to R[a_1, \ldots, a_d] \subseteq A \\ f(x_1, \ldots, x_d) &\mapsto f(a_1, \ldots, a_d) \end{aligned}$$

is an R-algebra isomorphism.

In other words, Proposition 5.30 asserts that, given an R-algebra A, finding a subalgebra B that is isomorphic to a polynomial ring is equivalent to finding a subset of algebraically independent elements. This idea connects our motivation to the following statement of the Noether Normalization Theorem, valid for any field K.

5.31 **THEOREM** *Noether Normalization Theorem*

If A is a finitely-generated K-algebra, then there exists an algebraically independent subset $\{a_1, \ldots, a_d\} \subseteq A$ such that A is integral over $K[a_1, \ldots, a_d]$.

In particular, although it may not be possible to find algebraically independent elements $a_1, \ldots, a_d$ that generate all of A (since A may not be isomorphic to a polynomial ring), the fact that A is integral over $K[a_1, \ldots, a_d]$, and thus a finitely-generated $K[a_1, \ldots, a_d]$-module, says that these elements generate a polynomial ring that is "as close as possible to A."

The Noether Normalization Theorem motivates the following definition.

5.32 **DEFINITION** *Noether basis*

Let A be a finitely-generated K-algebra. A subset $\{a_1, \ldots, a_d\} \subseteq A$ is a *Noether basis* of A over K if

(i) $a_1, \ldots, a_d$ are algebraically independent, and

(ii) A is integral over $K[a_1, \ldots, a_d]$.

In this language, the Noether Normalization Theorem asserts that every finitely-generated K-algebra contains a Noether basis. Note that, by Theorem 5.27, the second condition in Definition 5.32 is equivalent to A being a finitely-generated module over the subring $K[a_1, \ldots, a_d]$.

5.33 **EXAMPLE** A Noether basis for $K[x,y]/\langle x^2+y^2-1\rangle$

Consider the K-algebra

$$A = \frac{K[x,y]}{\langle x^2+y^2-1\rangle}.$$

Notice that A is generated by $[x]$ and $[y]$, but these two elements are not algebraically independent because

$$[x]^2 + [y]^2 - 1 = [x^2+y^2-1] = 0.$$

Let $a = [x] \in A$. We prove that $\{a\}$ is a Noether basis of A.

First, we observe that a is algebraically independent over K. To prove this, we must show that $f(a) \neq 0$ for any nonzero single-variable polynomial f. Notice that, for any such f, we have $f(x) \notin \langle x^2+y^2-1\rangle$. Therefore,

$$f(a) = [f(x)] \neq 0 \in \frac{K[x,y]}{\langle x^2+y^2-1\rangle},$$

as claimed.

Next, notice that $K[a] \neq A$, because we have no way of writing $[y]$ as a polynomial expression in $a = [x]$. However, we have already seen in Example 5.13 that A is generated by $[1]$ and $[y]$ as a $K[a]$-module:

$$A = K[a]\{[1], [y]\}.$$

This proves that A is a finitely-generated $K[a]$-module, and thus it is integral over $K[a]$. Given that a is algebraically independent over K and A is integral over $K[a]$, we conclude that $\{a\}$ is a Noether basis.

This example readily generalizes to the coordinate ring of the unit n-sphere

$$A = \frac{K[x_1, \ldots, x_n]}{\langle x_1^2 + \cdots + x_n^2 - 1 \rangle}$$

to show that $a_1 = [x_1], \ldots, a_{n-1} = [x_{n-1}]$ is a Noether basis (Exercise 5.4.2).

5.34 **EXAMPLE** A Noether basis for $K[x, y]/\langle xy - 1 \rangle$

Consider the K-algebra

$$A = \frac{K[x, y]}{\langle xy - 1 \rangle}.$$

As in Example 5.33, A is generated by $[x]$ and $[y]$, but these are not algebraically independent, so they do not form a Noether basis. One might naturally guess, then, that either of the single elements $[x]$ or $[y]$ would form a Noether basis. However, this is not the case. For example, while $[x]$ is certainly algebraically independent over K, it can be checked that A is *not* integral over $K[x]$ (Exercise 5.4.3).

Even though neither $[x]$ nor $[y]$ alone form a Noether basis for A, the Noether Normalization Theorem guarantees that a Noether basis must exist. In this case, a Noether basis is given by the element $a = [x + y]$. A proof of this assertion is outlined in Exercise 5.4.4.

Now that we have seen a few examples, let us turn toward a proof, which uses a clever induction argument that crucially relies on Theorem 5.27.

PROOF OF THEOREM 5.31 Any finitely-generated K-algebra can, by definition, be expressed as $A = K[b_1, \ldots, b_n]$ for some $b_1, \ldots, b_n \in A$, and we prove the theorem by induction on the number n of generators.

Base Case: If $n = 0$, then $A = K$ and the empty set is a Noether basis.

Induction Step: Suppose that Noether bases exist for all K-algebras with fewer than n generators, and let $A = K[b_1, \ldots, b_n]$. If $b_1, \ldots, b_n$ are algebraically independent over K, then $\{b_1, \ldots, b_n\}$ is a Noether basis and we are done. Assume, then, that $b_1, \ldots, b_n$ are not algebraically independent over K, meaning that there exists a nonzero polynomial $f \in K[x_1, \ldots x_n]$ such that

$$f(b_1, \ldots, b_n) = 0.$$

We will manipulate the polynomial f to make it monic in b_1. Let N be a positive integer greater than the maximum exponent appearing on any variable in any

monomial of f, and define new elements

$$\begin{aligned}
\widetilde{b}_2 &= b_2 - b_1^N \\
\widetilde{b}_3 &= b_3 - b_1^{N^2} \\
\widetilde{b}_4 &= b_4 - b_1^{N^3} \\
&\vdots \\
\widetilde{b}_n &= b_n - b_1^{N^{n-1}},
\end{aligned} \tag{5.5}$$

so that

$$0 = f(b_1, \ldots, b_n) = f\big(b_1,\ \widetilde{b}_2 + b_1^N,\ \widetilde{b}_3 + b_1^{N^2},\ \ldots,\ \widetilde{b}_n + b_1^{N^{n-1}}\big).$$

In terms of $b_1, \tilde{b}_2, \ldots, \tilde{b}_n$, a monomial $c \cdot b_1^{a_1} \cdots b_n^{a_n}$ becomes

$$c \cdot b_1^{a_1} \big(\widetilde{b}_2 + b_1^N\big)^{a_2} \big(\widetilde{b}_3 + b_1^{N^2}\big)^{a_3} \cdots \big(\widetilde{b}_n + b_1^{N^{n-1}}\big)^{a_n}. \tag{5.6}$$

Collecting terms with the same power of b_1, we can rearrange (5.6) to

$$c \cdot b_1^{a_1 + a_2 N + a_3 N^2 + \cdots + a_n N^{n-1}} + \text{lower-degree terms in } b_1.$$

Because N is larger than $a_1, \ldots, a_n$, the number $a_1 + a_2 N + a_3 N^2 + \cdots + a_n N^{n-1}$ uniquely determines the numbers $a_1, \ldots, a_n$. (This statement is the "uniqueness of base-N expansions"; for example, if $N = 10$, it is the statement that the digits of a number are uniquely determined by the number itself—see Exercise 5.4.7.) In particular, if

$$x_1^{a_1} \cdots x_n^{a_n} \quad \text{and} \quad x_1^{a_1'} \cdots x_n^{a_n'}$$

are two different monomials of f, then we cannot have

$$a_1 + a_2 N + a_3 N^2 + \cdots + a_n N^{n-1} = a_1' + a_2' N + a_3' N^2 + \cdots + a_n' N^{n-1}.$$

It follows that every monomial of f, after evaluating at

$$(b_1, \ldots, b_n) = \big(b_1,\ \widetilde{b}_2 + b_1^N,\ \widetilde{b}_3 + b_1^{N^2},\ \ldots,\ \widetilde{b}_n + b_1^{N^{n-1}}\big),$$

has a different highest power of b_1. Let $c \cdot x_1^{a_1} \cdots x_n^{a_n}$ be the unique nonzero term in f whose highest power of b_1 is the largest. Then

$$c^{-1} \cdot f(x, \widetilde{b}_2 + x^N, \widetilde{b}_3 + x^{N^2}, \ldots, \widetilde{b}_n + x^{N^{n-1}}) \in K[\widetilde{b}_2, \ldots, \widetilde{b}_n][x]$$

is a monic polynomial that vanishes when we set $x = b_1$, which means that b_1 is integral over $K[\widetilde{b}_2, \ldots, \widetilde{b}_n]$. It follows from Theorem 5.27 that $K[\widetilde{b}_2, \ldots, \widetilde{b}_n][b_1]$ is integral over $K[\widetilde{b}_2, \ldots, \widetilde{b}_n]$. But since we can freely convert between polynomial expressions in $b_1, b_2, \ldots, b_n$ and $b_1, \widetilde{b}_2, \ldots, \widetilde{b}_n$, we have

$$K[\widetilde{b}_2, \ldots, \widetilde{b}_n][b_1] = K[b_1, \widetilde{b}_2, \ldots, \widetilde{b}_n] = K[b_1, b_2, \ldots, b_n] = A.$$

Thus, we have proven that A is integral over $K[\widetilde{b}_2, \ldots, \widetilde{b}_n]$. By the induction hypothesis, there exists a Noether basis $\{a_1, \ldots, a_d\} \subseteq K[\widetilde{b}_2, \ldots, \widetilde{b}_n]$. It follows that each of the following rings is integral over the one that precedes it:

$$K[a_1, \ldots, a_d] \subseteq K[\widetilde{b}_2, \ldots, \widetilde{b}_n] \subseteq A.$$

By Theorem 5.27, this means that each is finitely-generated as a module over the one that precedes it. By Exercise 5.2.6, we conclude that A is finitely-generated as a module, and thus integral, over $K[a_1, \ldots, a_d]$. Therefore, $\{a_1, \ldots, a_d\}$ is a Noether basis of A, concluding the induction argument. □

The Noether Normalization Theorem is the culmination of the last four sections of algebraic developments, and the payoffs for all of this hard work will be plentiful. Most immediately, as we will see in the next section, Noether normalization can be used to prove the Nullstellensatz, and therefore, the equivalence of algebra and geometry. However, the payoff does not end there; Noether normalization is also closely connected to the notion of dimension, a concept we will introduce in the next chapter. As we'll see, if X is an irreducible affine variety, the number of elements in any Noether basis is the *dimension* of X.

Exercises for Section 5.4

5.4.1 Prove Proposition 5.30.

5.4.2 Consider the K-algebra

$$A = \frac{K[x_1, \ldots, x_n]}{\langle x_1^2 + \cdots + x_n^2 - 1 \rangle}.$$

and let $a_i = [x_i]$ for all $i = 1, \ldots, n-1$.

(a) Prove that $a_1, \ldots, a_{n-1}$ are algebraically independent.

(b) Prove that A is integral over $K[a_1, \ldots, a_{n-1}]$.

5.4.3 Let $A = K[x, y]/\langle xy - 1 \rangle$ and let $a = [x]$.

(a) Prove that a is algebraically independent over K.

(b) Prove that A is *not* integral over $K[a]$.

5.4.4 Let $A = K[x, y]/\langle xy - 1 \rangle$ and let $a = [x + y]$.

(a) Prove that a is algebraically independent over K.

(b) Prove that A is integral over $K[a]$.

5.4.5 Let A be a K-algebra. Prove that every Noether basis of A is a maximal algebraically independent set. (To say that $a_1, \ldots, a_n$ is a *maximal algebraically independent set* over K is to say that $a_1, \ldots, a_d, a_{d+1}$ is algebraically dependent over K for all $a_{d+1} \in L$.)

5.4.6 Show by example that the converse of Exercise 5.4.5 is false: that is, if A is a finitely-generated K-algebra and $a_1, \ldots, a_d$ is a maximal algebraically independent set, it need not be the case that A is integral over $K[a_1, \ldots, a_d]$.

5.4.7 Let N be a positive integer and set $[N] = \{0, 1, \ldots, N-1\}$. For $k \geq 0$, consider the function

$$\begin{aligned} \psi_n : [N]^{k+1} &\to \mathbb{N} \\ (a_0, \ldots, a_k) &\mapsto a_0 + a_1 N + a_2 N^2 + \cdots + a_k N^k \end{aligned}$$

(a) If $N = 10$, how would you describe the number $\psi_k(a_0, \ldots, a_k)$?

(b) Prove that ψ_k is injective for all $k \geq 0$. (In other words, the number $a_0 + a_1 N + a_2 N^2 + \cdots + a_k N^k$ uniquely determines $a_0, \ldots, a_k$.)

Section 5.5 Proof of the Nullstellensatz

We have now built up the necessary algebraic tools to discuss a proof of the Nullstellensatz and thus a complete justification of the equivalence of algebra and geometry that was developed in Chapters 1 – 4. We begin with two useful lemmas that will allow us to prove a seemingly weaker version of the Nullstellensatz. Then, we prove that—surprisingly—this weak version in fact implies the full Nullstellensatz itself.

5.35 LEMMA *Algebraic over algebraically closed fields*

Let $K \subseteq L$ be fields such that L is algebraic over K. If K is algebraically closed, then $L = K$.

The name "algebraically closed" reflects the fact that any attempt to adjoin an element to K that is algebraic over K produces only elements that are already there, so K is "closed under adjoining algebraic elements."

PROOF Suppose, toward a contradiction, that $a \in L \setminus K$. Then, since a is algebraic over K, there exists a polynomial $f \in K[x]$ such that $f(a) = 0$. There may be many such polynomials, but let f be one of minimum possible degree. Because K is algebraically closed, there exists $b \in K$ such that $f(b) = 0$, so Corollary 0.48 tells us that we can factor $f(x)$ as

$$f(x) = (x - b)g(x)$$

for some $g \in K[x]$ with $\deg(g) < \deg(f)$. Evaluating both sides at $x = a$ yields

$$0 = (a - b)g(a).$$

We have assumed that $a \notin K$ and $b \in K$, so $a \neq b$. Hence, the above is only possible if $g(a) = 0$, which contradicts the assumption that f is a minimum-degree polynomial in $K[x]$ that vanishes at a. □

The next result is the primary consequence of the Noether Normalization Theorem that we require in this section. It was first proved by Zariski, though by a different method than what we present here; it is valid for all fields K.

5.36 LEMMA *Zariski's Lemma*

Let $K \subseteq L$ be fields. If L is a finitely-generated K-algebra, then L is algebraic over K.

PROOF Let $K \subseteq L$ be fields such that L is a finitely-generated K-algebra. By the Noether Normalization Theorem, there exist elements $a_1, \ldots, a_d \in L$ that are algebraically independent over K such that L is integral, and thus algebraic, over $K[a_1, \ldots, a_d]$. To show that L is algebraic over K, it suffices to prove that $d = 0$.

Toward a contradiction, suppose that $d > 0$. Then, since L is a field, there exists $a_1^{-1} \in L$. Because L is integral over $K[a_1, \ldots, a_d]$, the element a_1^{-1} is a solution to a monic polynomial with coefficients in $K[a_1, \ldots, a_d]$:

$$(a_1^{-1})^m + c_1(a_1^{-1})^{m-1} + \cdots + c_{m-1}(a_1^{-1}) + c_m = 0,$$

where $c_1, \ldots, c_m \in K[a_1, \ldots, a_d]$. Multiply both sides by a_1^m to obtain

$$1 + c_1 a_1 + \cdots + c_{m-1} a_1^{m-1} + c_m a_1^m = 0.$$

This is a polynomial equation relating $a_1, \ldots, a_d$, contradicting the algebraic independence of these elements. The contradiction implies that $d = 0$, as desired. □

The following result, which we will prove from the previous two lemmas, is commonly referred to as the "Weak Nullstellensatz." This name is a bit of a misnomer because, as we will see below and in Exercise 5.5.1, the weak Nullstellensatz is logically equivalent to the Nullstellensatz.

5.37 PROPOSITION *Weak Nullstellensatz*

Let K be algebraically closed. For any proper ideal $I \subsetneq K[x_1, \ldots, x_n]$, we have $\mathcal{V}(I) \neq \emptyset$.

PROOF Suppose that K is algebraically closed, and let $I \subsetneq K[x_1, \ldots, x_n]$ be a proper ideal. By Exercise 5.5.3 below, there exists a maximal ideal J of $K[x_1, \ldots, x_n]$ containing I. Since $I \subseteq J$ implies $\mathcal{V}(J) \subseteq \mathcal{V}(I)$, it suffices to prove that $\mathcal{V}(J) \neq \emptyset$.

Since J is maximal, the quotient ring

$$L = K[x_1, \ldots, x_n]/J$$

is a field. Moreover, L is a finitely-generated K-algebra (generated by $[x_1], \ldots, [x_n]$) and hence Zariski's Lemma implies that L is algebraic over K. But K is algebraically closed, so Lemma 5.35 implies that $L = K$.

As $L = K$, there is a K-algebra isomorphism $\varphi : L \to K$. Define $a_i = \varphi([x_i])$. Using that φ is a K-algebra homomorphism, we see that

$$\varphi([x_i - a_i]) = \varphi([x_i]) - \varphi([a_i]) = a_i - a_i = 0.$$

Since φ is injective, this implies that that $[x_i - a_i] = 0$ for all i. In other words,

$$\langle x_1 - a_1, \ldots, x_n - a_n \rangle \subseteq J.$$

Since $\langle x_1 - a_1, \ldots, x_n - a_n \rangle$ is a maximal ideal and J is a proper ideal, this implies that $J = \langle x_1 - a_1, \ldots, x_n - a_n \rangle$. It follows that

$$\mathcal{V}(J) = \{(a_1, \ldots, a_n)\} \neq \emptyset,$$

as required. □

The weak Nullstellensatz is a rather blunt tool; given a proper ideal I, it only ensures that $\mathcal{V}(I)$ is nonempty. The full Nullstellensatz, on the other hand, is very precise—it tells us exactly which polynomials vanish on $\mathcal{V}(I)$. We now prove that the full Nullstellensatz is actually implied by the weak Nullstellensatz.

5.38 THEOREM *Nullstellensatz recalled*

Let K be an algebraically closed field. Then, for any ideal $J \subseteq K[x_1, \ldots, x_n]$, we have

$$\mathcal{I}(\mathcal{V}(J)) = \sqrt{J}.$$

PROOF The inclusion

$$\mathcal{I}(\mathcal{V}(J)) \supseteq \sqrt{J}$$

holds over any field and follows directly from the definitions; see Exercise 1.5.3. We must prove, then, that $\mathcal{I}(\mathcal{V}(J)) \subseteq \sqrt{J}$. Since $K[x_1,\dots,x_n]$ is Noetherian, we write $J = \langle f_1,\dots,f_m\rangle$.

Let $g \in \mathcal{I}(\mathcal{V}(J))$. Introduce a new variable x_{n+1} and consider the ideal

$$J' = \langle f_1,\dots,f_m,h\rangle \subseteq K[x_1,\dots,x_n,x_{n+1}],$$

where

$$h(x_1,\dots,x_{n+1}) = 1 - x_{n+1}g(x_1,\dots,x_n).$$

The trick of adding an extra variable is a clever device for bringing the Weak Nullstellensatz to bear on the strong Nullstellensatz.

We claim that $\mathcal{V}(J') = \emptyset \subseteq \mathbb{A}^{n+1}$. To see this, suppose toward a contradiction that $a = (a_1,\dots,a_n,a_{n+1}) \in \mathcal{V}(J')$. Then $f_i(a) = 0$ for all i, but since f_i only involves the variables $x_1,\dots,x_n$, this is equivalent to the statement that $f_i(a_1,\dots,a_n) = 0$ for all i. That is, $(a_1,\dots,a_n) \in \mathcal{V}(J)$. Since $g \in \mathcal{I}(\mathcal{V}(J))$, it follows that $g(a_1,\dots,a_n) = 0$, and hence

$$h(a) = 1 - a_{n+1}g(a_1,\dots,a_n) = 1 \neq 0.$$

Since $h \in J'$, this contradicts the fact that $a \in \mathcal{V}(J')$.

We have thus shown that $\mathcal{V}(J') = \emptyset$, so the Weak Nullstellensatz implies that $J' = K[x_1,\dots,x_{n+1}]$. In particular, the constant polynomial 1 can be expressed in terms of the generators of J':

$$1 = q_1f_1 + \cdots + q_mf_m + r(1 - x_{n+1}g), \tag{5.7}$$

where $q_1,\dots,q_m,r \in K[x_1,\dots,x_{n+1}]$. We would like to isolate g in this equation, but doing so necessitates division by x_{n+1}. To make sense of this, consider the larger ring $K(x_{n+1})[x_1,\dots,x_n]$, where we allow rational functions in x_{n+1} (we were introduced to this ring in Section 0.6). The equation (5.7) still holds in this larger ring, and now we can divide both sides by a sufficiently high power of x_{n+1} so that no positive powers of x_{n+1} appear:

$$x_{n+1}^{-k} = \widetilde{q}_1f_1 + \cdots + \widetilde{q}_mf_m + \widetilde{r}(x_{n+1}^{-1} - g), \tag{5.8}$$

where $\widetilde{q}_1,\dots,\widetilde{q}_m,\widetilde{r} \in K[x_1,\dots,x_n,x_{n+1}^{-1}] \subseteq K(x_{n+1})[x_1,\dots,x_n]$.

Now that we have an equation in the polynomial ring $K[x_1,\dots,x_n,x_{n+1}^{-1}]$, we can make the substitution $x_{n+1}^{-1} = g \in K[x_1,\dots,x_n]$, which yields the equation

$$g^k = \widehat{q}_1f_1 + \cdots + \widehat{q}_mf_m,$$

where $\widehat{q}_i \in K[x_1,\dots,x_n]$ is obtained from $\widetilde{q}_i$ by setting $x_{n+1}^{-1} = g$. This last equation shows that $g^k \in J$, so $g \in \sqrt{J}$, and the proof is complete. □

The last step of this proof, where we set $x_{n+1}^{-1} = g$, often strikes readers as suspicious on a first pass; why can we simply choose a value for x_{n+1}^{-1}? We stress that the reason this is valid is that x_{n+1}^{-1} is simply a variable like any other in the ring $K[x_1, \ldots, x_n, x_{n+1}^{-1}]$. As an illustration of the procedure, let us carry it out explicitly in a small example.

5.39 **EXAMPLE** An illustration of the proof of the Nullstellensatz

Let $J = \langle x^2, y \rangle \subseteq K[x, y]$, and let $g = x + y$. Denoting the variable x_{n+1} by z in this case, equation (5.7) reads

$$1 = q_1(x, y, z) \cdot x^2 + q_2(x, y, z) \cdot y + r(x, y, z) \cdot \big(1 - z(x + y)\big),$$

and it is straightforward to check that this equation holds for the following choice of q_1, q_2, and r:

$$1 = z^2 \cdot x^2 + \big(z^2(2x + y)\big) \cdot y + \big(1 + z(x + y)\big) \cdot \big(1 - z(x + y)\big).$$

Dividing both sides by z^2 eliminates all positive powers of z, yielding the equation

$$z^{-2} = 1 \cdot x^2 + (2x + y) \cdot y + \big(z^{-1} + (x + y)\big) \cdot \big(z^{-1} - (x + y)\big)$$

in $K[x, y, z^{-1}]$. Setting $z^{-1} = g = x + y$ results in the equation

$$(x + y)^2 = 1 \cdot x^2 + (2x + y) \cdot y,$$

which is manifestly true in $K[x, y]$ and illustrates that $g^2 \in J$.

Exercises for Section 5.5

5.5.1 Prove that the Nullstellensatz implies the Weak Nullstellensatz.

5.5.2 Give an example to show that the Weak Nullstellensatz can fail if K is not algebraically closed. In the notation of Proposition 5.37, what are I, J, and L in your example? Discuss why your example is consistent with Zariski's Lemma even though it is inconsistent with the Weak Nullstellensatz.

5.5.3 Let $J \subseteq K[x_1, \ldots, x_n]$ be an ideal with $J \neq K[x_1, \ldots, x_n]$. Prove that there exists a maximal ideal containing J. (**Hint**: Use that $K[x_1, \ldots, x_n]$, being Noetherian, satisfies the ascending chain condition.)

5.5.4 Let K be an algebraically closed field, and let $I \subsetneq K[x_1, \ldots, x_n]$ be a proper ideal. By Exercise 5.5.3, there exists a maximal ideal $J \subseteq K[x_1, \ldots, x_n]$ containing I. On the other hand, recall from Proposition 2.31 that the Nullstellensatz yields a bijection between maximal ideals of $K[x_1, \ldots, x_n]$ and points of $\mathbb{A}^n$.

(a) Interpret the statements "I is a proper ideal" and "J is a maximal ideal containing I" in terms of $\mathcal{V}(I)$ and/or $\mathcal{V}(J)$.

(b) Using this interpretation in terms of varieties, find a maximal ideal containing $I = \langle y^2 - x^3 - x^2 \rangle \subseteq K[x,y]$.

5.5.5 Let $J = \langle x^2 + y^2 - 1, y - 1 \rangle \subseteq K[x,y]$ and let $g = x \in K[x,y]$.

(a) Calculate $\mathcal{V}(J) \subseteq \mathbb{A}^2$, and confirm (without citing the Nullstellensatz) that $g \in \mathcal{I}(\mathcal{V}(J))$ but $g \notin J$.

(b) Let $f_1 = x^2 + y^2 - 1$ and $f_2 = y - 1$ be the generators of J, and let $J' = \langle f_1, f_2, 1 - zx \rangle$, as in the proof of the Nullstellensatz. Find $q_1, q_2, r \in K[x,y,z]$ such that equation (5.7) holds.

(c) Set $z = 1/g$ in the equation from part (b) and clear denominators to deduce an equation that exhibits $g \in \sqrt{J}$.

(d) You have just verified that the ideal $J = \langle x^2 + y^2 - 1, y - 1 \rangle$ is not radical. On the other hand,

$$J = \langle x^2 + y^2 - 1 \rangle + \langle y - 1 \rangle,$$

which implies that

$$\mathcal{V}(J) = \mathcal{V}(x^2 + y^2 - 1) \cap \mathcal{V}(y - 1).$$

Draw a picture of $\mathcal{V}(x^2 + y^2 - 1)$ and $\mathcal{V}(y - 1)$ over the real numbers. Do you have a guess about what geometric feature of these varieties is responsible for $\langle x^2 + y^2 - 1 \rangle + \langle y - 1 \rangle$ not being radical?

5.5.6 Prove that a field K is algebraically closed if and only if every maximal ideal $J \subseteq K[x_1, \ldots, x_n]$ has the form

$$J = \langle x_1 - a_1, \ldots, x_n - a_n \rangle$$

for some $a_1, \ldots, a_n \in K$.

Chapter 6

Dimension of Affine Varieties

LEARNING OBJECTIVES FOR CHAPTER 6

- Explore the notion of dimension and how it relates to algebraic independence in coordinate rings.
- Become acquainted with function fields of affine varieties and compute them in several examples.
- Learn to measure field extensions using transcendence bases.
- Define and study dimension of affine varieties.
- Prove the Fundamental Theorem of Dimension Theory as an application of Noether normalization.

Given a set of polynomial equations, arguably the most fundamental geometric question one could ask about its solution set is how "big" it is. This is the question on which we aim to shed light in this chapter by introducing the important concept of *dimension* of affine varieties.

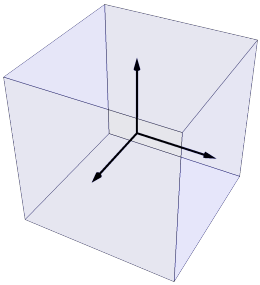
Intuitively, dimension measures the freedom to be able to move within a set. So how do we measure the freedom to move within an affine variety? As we will see, the key to answering this question lies within the coordinate ring. One of the motivating ideas of this chapter is that *the dimension of an affine variety is the maximum number of algebraically independent elements in its coordinate ring*.

This motivating idea might remind the reader of their knowledge of dimension from linear algebra, where the dimension of a (finite-dimensional) vector space can be defined as the maximum number of linearly independent elements. Indeed, there are many analogies between the ideas developed in this chapter and the ideas concerning dimension in linear algebra. In particular, the definition and properties of *transcendence bases*, introduced in Section 6.3, will closely parallel ideas concerning bases of vector space. In order to obtain a robust theory of transcendence bases, one requires working with fields, instead of rings, which is why Section 6.2 is devoted to the notion of *function fields* of irreducible affine varieties.

Once the groundwork regarding function fields and transcendence bases has been laid, Section 6.5 defines and describes some basic properties of dimension. For example, we see that $\dim(\mathbb{A}^n) = n$ and, if $X \subseteq Y$, then $\dim(X) \leq \dim(Y)$. The ideas of this chapter then culminate in Section 6.6 with what we call the Fundamental Theorem of Dimension Theory. This result, which is central in algebraic geometry, essentially says that each polynomial equation in a set decreases the dimension of the solution set by at most one.

Section 6.1 Motivating ideas

Dimension is a notion with which we are all intimately familiar; for example, at a young age, most of us probably came to grips with the understanding that we live in a three-dimensional world. The way in which we understand those three dimensions is that we have three basic directions in which we can move: forward-backward, left-right, and up-down. Of course, there are more than just these three basic directions, but every other direction is a combination of just these three.



For many of us, linear algebra is the first mathematical subject in which we learn a precise definition of dimension, and it naturally captures the notions of "basic" and "combination" alluded to in the previous paragraph. Recall that a vector space V over a field K is said to have *dimension* n if it has a basis of size n. Importantly, one must check that any two bases of V have the same size in order to make sure that this definition is well-defined. One way to view basis vectors is as the basic directions of movement inside V. In the image above, we have depicted the three standard basis vectors in $\mathbb{R}^3$: any direction of movement in $\mathbb{R}^3$ can be uniquely expressed as a linear combination of these.

Note the important role the field K plays in vector space dimension: the usual plane is two-dimensional as an $\mathbb{R}$-vector space but only one-dimensional as a $\mathbb{C}$-vector space.

The idea that dimension measures our freedom to move within a set is the intuition that will guide our definition of dimension in algebraic geometry. We further motivate our development with the following familiar non-linear example.

6.1 **EXAMPLE** Dimension of the sphere

Consider the sphere $X = \mathcal{V}(x^2 + y^2 + z^2 - 1) \subseteq \mathbb{A}^3$. Over the real numbers, this is simply the familiar unit sphere, and we can even imagine the unit sphere as a model for the surface of the Earth.

We naturally view the surface of our planet as having two dimensions because, at any location, we have two basic directions in which we can move; when we are not at one of the poles, we call these directions north-south and east-west. But how do we make our real-world intuition of these two dimensions of freedom algebraically precise?

One way to argue is the following: if we choose two of the coordinates of a point in X, say $x = a$ and $y = b$, for some $a, b \in K$, then the defining equation for X tells us that the third coordinate is constrained by these choices: we must have

$$z^2 = 1 - a^2 - b^2.$$

In particular, there are at most two possible solutions for the third coordinate.

In other words, in order to describe a point on X, we have two variables of freedom—we can choose the x- and y-coordinates without constraint—whereupon the third coordinate is then determined up to finitely-many values. This reflects the algebraic fact that the subset $\{[x],[y]\} \subseteq K[X]$ is algebraically independent over K—fixing the value of one does not constrain the another—whereas $\{[x],[y],[z]\}$ is algebraically dependent over K, because

$$[x]^2 + [y]^2 + [z]^2 - 1 = [x^2 + y^2 + z^2 - 1] = 0 \in K[X].$$

The previous example suggests that our "freedom to move" within X is measured by algebraically independent functions: anytime we have algebraically independent coordinate functions in $K[X]$, then we can choose the values of those coordinates freely, without constraint. Since dimension should measure our maximum freedom of movement, we arrive at the following motivating idea.

6.2 **KEY IDEA** *Dimension of an affine variety*

The *dimension* of an affine variety X should be the maximum number of elements in $K[X]$ that are algebraically independent over K.

In fact, one could take this key idea as the definition of dimension; the reason we do not is because it is not particularly easy to work with. For starters, it is not even clear from this description whether or not dimension is finite: given an affine variety X, how do we know that it is not possible to find larger and larger sets of algebraically independent functions in $K[X]$? In order to argue that this cannot happen, we need to lay some groundwork first. Developing this groundwork carefully will require some work on our part, undertaken in Sections 6.2 – 6.4, before we finally present a more robust definition of dimension in Section 6.5 and show that it is equivalent to the description above (see Corollary 6.34).

In the meantime, we can keep Key Idea 6.2 in the back of our minds as motivation and turn to a discussion of our aspirations for dimension. What are the essential properties that we should expect in a notion of dimension for affine varieties? We begin by listing those properties, which we call the *axioms* for dimension.

6.3 **DEFINITION** *Axioms for dimension of affine varieties*

We say that a function

$$D : \left\{ \begin{array}{c} \text{isomorphism classes of} \\ \text{nonempty affine varieties} \end{array} \right\} \longrightarrow \mathbb{N}$$

is a *dimension function* if it satisfies the following properties:

1. $D(X) = 0$ if X consists of a single point;
2. $D(X_1 \cup \cdots \cup X_n) = \max\{D(X_1), \ldots, D(X_n)\}$;
3. If $X \subseteq \mathbb{A}^n$ is irreducible and $f \in K[x_1, \ldots, x_n]$ is such that $X \cap \mathcal{V}(f)$ is neither empty nor all of X, then $D(X \cap \mathcal{V}(f)) = D(X) - 1$.

Let us briefly interpret these axioms. First of all, since two isomorphic affine varieties share the same essential properties, we should expect their dimensions to be the same. This is the reason that we should view dimension as a function from the set of isomorphism classes of affine varieties.

The empty variety was omitted from the domain of D. By convention, we define $\dim(\emptyset) = -1$.

It should be somewhat clear from our intuition why we require Axiom 1: there is not a lot of freedom to move within a single point, so we should expect the dimension of a point to be zero.

Regarding Axiom 2, consider as an example the union of a line X_1 and a plane X_2 in $\mathbb{A}^3$. If the line is contained in the plane, then their union is simply equal to the plane, so $\dim(X_1 \cup X_2) = \dim(X_2)$, the maximum of the two individual dimensions. If the line is not contained in the plane, however, then our intuition for "freedom to move" breaks down: the number of directions in which you can move within $X_1 \cup X_2$ depends on the point at which you stand. To resolve this ambiguity, we simply declare that the larger dimension trumps the smaller. More generally, we will find that dimension is only readily definable for irreducible varieties, and we will define the dimension of reducible variety as the maximum dimension of its irreducible components.

Finally, for Axiom 3, let us parse the statement further by writing X as the vanishing $\mathcal{V}(f_1, \ldots, f_m)$. Then

$$X \cap \mathcal{V}(f) = \mathcal{V}(f_1, \ldots, f_m, f).$$

In other words, Axiom 3 is essentially saying that if we impose one additional equation in our vanishing set, then the dimension should go down by exactly one. It is hard to overstate how important this property of dimension is in algebraic geometry, which is why, in this text, we dub this property the *Fundamental Theorem of Dimension Theory*.

In the vector space setting, Axiom 3 is a consequence of the Rank-Nullity Theorem (see Exercise 6.1.5)

We note that the assumptions in the hypothesis of Axiom 3 are all necessary. We assume that X is irreducible because, if not, then it could consist of two irreducible components of the same dimension and intersecting with $\mathcal{V}(f)$ might just pick out one of these components. For example, in $\mathbb{A}^2$, the variety $X = \mathcal{V}(xy)$ is the union of the two axes and intersecting with $\mathcal{V}(x)$ simply picks out the y-axis. We assume that $X \cap \mathcal{V}(f)$ is neither empty nor all of X to avoid the situations where the dimension leaps all the way down to -1 or stays at $\dim(X)$, both of which violate the conclusion of Axiom 3.

One might naturally ask why we do not include more axioms for dimension. For example, it might be natural for us to assume that $\mathbb{A}^n$ has dimension n and, more generally, that the dimension of an affine variety defined by linear equations is equal to its dimensions as a vector space. The reason we do not assume more is because the short list of axioms listed in Definition 6.3 already uniquely determines the dimension of all affine varieties, as stated in the next result.

6.4 PROPOSITION *There exists at most one dimension function*

If D_1 and D_2 are both dimension functions, then $D_1(X) = D_2(X)$ for all affine varieties X.

PROOF A proof is outlined in Exercise 6.1.8 □

In light of Proposition 6.4, the path ahead is clear: our goal is to define a function

$$\dim : \left\{ \begin{array}{c} \text{isomorphism classes of} \\ \text{nonempty affine varieties} \end{array} \right\} \longrightarrow \mathbb{N}.$$

and to prove that it satisfies Axioms 1 – 3 of Definition 6.3. The next two sections are devoted to developing the algebraic tools we require in order to define dimension, including a robust understanding of algebraic (in)dependence. Once we have given a rigorous definition of dimension in Section 6.5, Axioms 1 and 2 will be rather straightforward to prove; in fact, they can already be proved using Key Idea 6.2 (see Exercises 6.1.1, 6.1.3, and 6.1.4). Proving Axiom 3, on the other hand, is quite involved, with a key step coming from the Noether Normalization Theorem. A slightly strengthened form of Axiom 3 will be proved in Section 6.6 as the Fundamental Theorem of Dimension Theory.

Exercises for Section 6.1

6.1.1 Assuming Key Idea 6.2, prove that the dimension of a single point is zero.

6.1.2 Assuming Key Idea 6.2, prove that $\dim(\mathbb{A}^n) \geq n$.

6.1.3 Let $X, Y \subseteq \mathbb{A}^n$ be affine varieties with $X \subseteq Y$. Assume that, using Key Idea 6.2, $\dim(X) = d_1$ and $\dim(Y) = d_2$.

(a) Let $f_1, \ldots, f_m \in K[x_1, \ldots, x_n]$ be polynomials such that

$$f_1|_X, \ldots, f_m|_X \in K[X]$$

are algebraically independent over K. Prove that

$$f_1|_Y, \ldots, f_m|_Y \in K[Y]$$

are algebraically independent over K.

(b) Prove that $d_1 \leq d_2$.

6.1.4 Let $X_1, \ldots, X_m \subseteq \mathbb{A}^n$ be affine varieties and assume that, using Key Idea 6.2, $\dim(X_i) = d_i$. Let $X = X_1 \cup \cdots \cup X_m$.

(a) Let $f_1, \ldots, f_\ell \in K[x_1, \ldots, x_n]$ be polynomials such that

$$f_1|_X, \ldots, f_\ell|_X \in K[X]$$

are algebraically independent over K. Prove that

$$f_1|_{X_i}, \ldots, f_\ell|_{X_i} \in K[X_i]$$

are algebraically independent over K for some i.

(b) Prove that

$$\dim(X_1 \cup \cdots \cup X_m) = \max\{d_1, \ldots, d_m\}.$$

(**Hint:** One inequality uses (a) and the other uses the previous exercise.)

6.1.5 Let $\ell_1, \ldots, \ell_m$ be linear homogeneous polynomials in n variables and consider the linear subspace $V \subseteq K^n$ defined by their vanishing:

$$V = \mathcal{V}(\ell_1, \ldots, \ell_m) \subseteq \mathbb{A}^n = K^n.$$

(a) Interpret V as the kernel of a specific matrix M, and use the Rank-Nullity Theorem to write the dimension of V in terms of n and $\mathrm{rk}(M)$.

(b) Let ℓ be another homogeneous linear equation and consider the subspace

$$W = V \cap \mathcal{V}(\ell)$$

Interpret W as the kernel of a specific matrix M'. How are M and M' related?

(c) Using the relationship between M and M', describe how the dimensions of V and W related? How does this compare to Axiom 3 in Definition 6.3?

(**Hint:** For (c), there are two possible cases to consider, what are they?)

6.1.6 Let D be a dimension function. Prove that $D(\mathbb{A}^n) = n$.

6.1.7 Let $X \subseteq \mathbb{A}^n$ be an irreducible affine variety that is not a single point. Prove that there exists $f \in K[x_1, \ldots, x_n]$ such that

$$\emptyset \subsetneq X \cap \mathcal{V}(f) \subsetneq X.$$

(This exercise is useful for the next one.)

6.1.8 Let D_1 and D_2 be dimension functions.

(a) Prove that $D_1(X) = 0$ if and only if X is a finite union of points. Conclude that $D_1(X) = D_2(X)$ whenever $D_1(X) = 0$.

(b) Assume that $m \geq 1$ and that $D_1(X) = D_2(X)$ when $D_1(X) < m$. Prove that $D_1(X) = D_2(X)$ for all X with $D_1(X) = m$.

(c) Combine the above arguments into an inductive proof of Proposition 6.4.

6.1.9 This exercise proves that a dimension function does not necessarily exist when K is not algebraically closed. Let $K = \mathbb{R}$ and, toward a contradiction, assume that a dimension function D exists.

(a) Prove that $D(\mathbb{A}^2) = 2$ and $D(\{(0,0)\}) = 0$.

(b) Prove that $\{(0,0)\} = \mathcal{V}(f) \cap \mathbb{A}^2$ for some f.

(c) Argue that parts (a) and (b) contradict Axiom (iii).

Section 6.2 Function fields

As we learned in the last section, the dimension of an affine variety should be a measure of the maximum number of algebraically independent elements in its coordinate ring. However, working with algebraic (in)dependence in K-algebras can be rather difficult, and it is much easier to work with these notions in the context of field extensions of K. Therefore, in this section, we study a way of passing from an irreducible affine variety X to a corresponding field $K(X)$, its *function field*.

6.5 **DEFINITION** *Function field*

Let X be an irreducible affine variety. The *function field* of X, denoted $K(X)$ is the fraction field of its coordinate ring:

$$K(X) = \operatorname{Frac}(K[X]).$$

The elements of $K(X)$ are called *rational functions* on X.

Fraction fields were defined in Section 0.6, where we learned that elements of $K(X)$ are ratios of the form f/g with $f, g \in K[X]$ and $g \neq 0$. Two fractions f_1/g_1 and f_2/g_2 are equal if and only if $f_1g_2 = f_2g_1 \in K[X]$, and the operations of addition and multiplication are defined in the usual way:

$$\frac{f_1}{g_1} + \frac{f_2}{g_2} = \frac{f_1g_2 + f_2g_1}{g_1g_2} \quad \text{and} \quad \frac{f_1}{g_1} \cdot \frac{f_2}{g_2} = \frac{f_1f_2}{g_1g_2}.$$

Since fraction fields are only defined for integral domains (Exercise 0.6.4), function fields are only defined for irreducible affine varieties. Let us turn to a few examples.

6.6 **EXAMPLE** Function field of affine space

Since the coordinate ring of $\mathbb{A}^n$ is $K[x_1, \ldots, x_n]$, the function field is the field of rational functions

$$K(\mathbb{A}^n) = K(x_1, \ldots, x_n),$$

which we already encountered in Section 0.6.

6.7 **EXAMPLE** Function field of $\mathcal{V}(x^2 - y^3)$

Let $X = \mathcal{V}(x^2 - y^3) \subseteq \mathbb{A}^2$. Computing the coordinate ring, we have

$$K[X] = K[x,y]/\langle x^2 - y^3 \rangle \quad \Longrightarrow \quad K(X) = \operatorname{Frac}(K[x,y]/\langle x^2 - y^3 \rangle).$$

This looks like a rather complicated field; however, we can identify it with a more familiar one. Consider the K-algebra homomorphism

$$\begin{aligned} \varphi : K[X] &\to K[t] \\ [f(x,y)] &\mapsto f(t^3, t^2). \end{aligned}$$

This is the pullback of the polynomial map defined in Example 4.8. Notice that φ is not an isomorphism, because the polynomial t is not in the image. If we pass to

function fields, then we can extend φ to a field homomorphism

$$\begin{aligned}\overline{\varphi} : K(X) &\to K(t)\\ \frac{[f(x,y)]}{[g(x,y)]} &\mapsto \frac{f(t^3,t^2)}{g(t^3,t^2)}.\end{aligned}$$

Now that we are allowed to divide, we see that t actually lies in the image of $\overline{\varphi}$:

$$t = \frac{\varphi([x])}{\varphi([y])} = \overline{\varphi}\Big(\frac{[x]}{[y]}\Big).$$

This allows us to invert $\overline{\varphi}$:

$$\overline{\varphi}^{-1}\Big(\frac{f(t)}{g(t)}\Big) = \frac{f([x]/[y])}{g([x]/[y])} = \frac{[y^{\deg(f)+\deg(g)}f(x/y)]}{[y^{\deg(f)+\deg(g)}g(x/y)]},$$

where the second equality simply serves to clear denominators in f and g so that the image is explicitly expressed as an element of $K(X)$. It follows that $K(X) \cong K(t)$. We encourage the reader to check the details of this argument in Exercise 6.2.3.

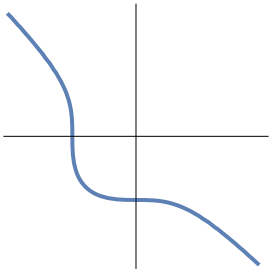
What we can conclude from this example is that, even though X and $\mathbb{A}^1$ are not isomorphic, their fraction fields are, because they are both isomorphic to $K(t)$. Thus, even though the coordinate ring knows everything about the isomorphism class of an affine variety, some of that information is lost upon passing to the function field.

When two affine varieties have isomorphic function fields, we say that the varieties are birationally equivalent. *This is an important notion in algebraic geometry, but it does not play a central role in this text.*

In the previous two examples, both of the function fields we considered were isomorphic to a field of rational functions $K(x_1,\dots,x_d)$ for some d. If an irreducible variety has a function field isomorphic to a field of rational functions, we say that the variety is *rational*. It turns out that rational varieties are actually quite special—it is a general fact beyond the scope of this text that "most" varieties are not rational. In particular, for a "random" polynomial $f \in K[x_1,\dots,x_n]$ of sufficiently large degree, $\mathcal{V}(f)$ will not be rational. The next example illustrates, perhaps, the simplest example of a non-rational variety.

6.8 EXAMPLE $\mathcal{V}(x^3+y^3+1)$ is not rational

Let $K = \mathbb{C}$ and let $X = \mathcal{V}(x^3+y^3+1) \subseteq \mathbb{A}^2$. The real points of X are depicted in the image to the right. We argue that X is not rational. This argument hinges on the fact that, if $f_1, f_2, f_3 \in K[x_1,\dots,x_n]$ are nonzero such that (i) no two of them share a common irreducible factor and (ii) $f_1^3 + f_2^3 + f_3^3 = 0$, then it must be the case that f_1, f_2, and f_3 are all constant (see Exercise 6.2.4 for a proof).



To argue that X is not rational, first notice that x^3+y^3+1 is irreducible, so

$$K[X] = K[x,y]/\langle x^3+y^3+1\rangle \quad\Longrightarrow\quad K(X) = \mathrm{Frac}(K[x,y]/\langle x^3+y^3+1\rangle).$$

Suppose that

$$\varphi : K(X) \to K(x_1,\ldots,x_d)$$

is a field homomorphism for some $d > 0$; we prove that $\mathrm{im}(\varphi) \subseteq K$, so φ cannot be an isomorphism. Notice that $\varphi([x]) = f/g$ and $\varphi([y]) = h/k$ for some polynomials $f,g,h,k \in K[x_1,\ldots,x_d]$ with $g,k \neq 0$. Using that $K[x_1,\ldots,x_n]$ is a UFD, we may factor each polynomial into irreducibles and reduce the quotients; in other words, we can assume that neither f and g nor h and k share common irreducible factors. As $[x^3+y^3+1] = 0 \in K(X)$, we obtain the relation

$$0 = \varphi([x^3+y^3+1]) = \left(\frac{f}{g}\right)^3 + \left(\frac{h}{k}\right)^3 + 1.$$

Clearing denominators, we obtain the relation

$$(fk)^3 + (gh)^3 + (gk)^3 = 0. \tag{6.1}$$

Equation (6.1) tells us that $g^3 \mid (fk)^3$, but since f and g are assumed not to have any common irreducible factors, it follows that $g^3 \mid k^3$. A parallel argument tells us that $k^3 \mid g^3$. Thus, since $g^3 \mid k^3$ and $k^3 \mid g^3$, it follows that $g^3 = ak^3$ for some constant a. Substituting this into Equation (6.1) and canceling the nonzero factor of k^3, we have

$$f^3 + (ah)^3 + g^3 = 0. \tag{6.2}$$

Notice that the irreducible factors of g are the same as those of k, and thus distinct from both f and h, by assumption. Furthermore, f and h cannot have an irreducible factor in common, or else, by Equation (6.2), this would also be an irreducible factor of g, contradicting our assumptions regarding f and g. Therefore, no two of the terms in Equation (6.2) share a common irreducible factor, and we may apply Exercise 6.2.4 to conclude that f, g, h, and k are all constant.

The argument above shows that $K(X)$ is not isomorphic to $K(x_1,\ldots,x_d)$ for any $d > 0$, but what if $d = 0$? Using that X has more than one point, we know that $K \subsetneq K[X]$, implying that $K \subsetneq K(X)$ and hence $K(X)$ is not isomorphic to K. Taken together with the previous argument, we conclude that X is not rational.

We conclude the section with a brief comment on terminology: even though elements of $K(X)$ are called *rational functions*, they are not actually functions on X. In particular, given an element $f/g \in K(X)$ with $g \neq 0$, it is possible that $g(a) = 0$ for some (but not all) values $a \in X$. Thus, f/g only defines a function on the subset of X where $g \neq 0$. It is common to use dashed arrows for rational functions to remind the reader that they are not defined on the entire domain:

$$\frac{f}{g} : X \dashrightarrow K.$$

If we restrict the domain to the complement of $\mathcal{V}(g)$, then every rational function gives rise to an actual function that assigns a value to each element of the domain:

$$\frac{f}{g} : X \setminus \mathcal{V}(g) \longrightarrow K.$$

Exercises for Section 6.2

6.2.1 Let $X = \mathcal{V}(xy - 1) \subseteq \mathbb{A}^2$. Prove that $K(X) \cong K(t)$.

6.2.2 Let R and S be integral domains and $\varphi : R \to S$ an injective ring homomorphism. Prove that there exists a well-defined field homomorphism

$$\begin{aligned} \overline{\varphi} : \operatorname{Frac}(R) &\to \operatorname{Frac}(S) \\ \frac{a}{b} &\mapsto \frac{\varphi(a)}{\varphi(b)}. \end{aligned}$$

6.2.3 Let φ and $\overline{\varphi}$ be the homomorphisms of Example 6.7.

(a) Prove that φ is injective, and use this to prove that $\overline{\varphi}$ is well-defined.

(b) Prove that $\overline{\varphi}$ and $\overline{\varphi}^{-1}$ are, in fact, inverse field homomorphisms.

6.2.4 Assume that $\operatorname{Char}(K) \neq 3$, and let $f_1, f_2, f_3 \in K[x_1, \ldots, x_n]$ be nonzero polynomials such that (i) no two of them share a common irreducible factor and (ii) $f_1^3 + f_2^3 + f_3^3 = 0$. This exercise outlines a proof of the fact that f_1, f_2, and f_3 must all be constant.

(a) Toward a contradiction, assume that f_1 is nonconstant in x_1 and let f_i' denote the derivative of f_i with respect to x_1. Prove that

$$f_1^2(f_1'f_3 - f_1f_3') = f_2^2(f_2f_3' - f_2'f_3).$$

(**Hint:** Start by differentiating $f_1^3 + f_2^3 + f_3^3 = 0$ using the chain rule.)

(b) Prove that $f_1f_3' - f_1'f_3 \neq 0$ and conclude that $f_2f_3' - f_2'f_3 \neq 0$. (**Hint:** Apply the quotient rule to differentiate f_1/f_3.)

(c) Prove that $f_1^2 \mid (f_2f_3' - f_2'f_3)$ and conclude that

$$2\deg_{x_1}(f_1) \leq \deg_{x_1}(f_2) + \deg_{x_1}(f_3) - 1.$$

(d) Repeating the above with f_1, f_2, and f_3 permuted in all three ways, you obtain three inequalities. Add these inequalities to find a contradiction.

6.2.5 Let $K = \mathbb{C}$ and let $X = \mathcal{V}(x^2 + y^2 - 1) \subseteq \mathbb{A}^2$. Prove that

$$\varphi : K(X) \to K(t)$$

defined by

$$\varphi\Big(\frac{[f(x,y)]}{[g(x,y)]}\Big) = \frac{f\big(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\big)}{g\big(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\big)}$$

is a well-defined field isomorphism.

Section 6.3 Transcendence bases

As we discussed in Section 6.1, the dimension of an affine variety X should be equal to the maximum number of algebraically independent elements in $K[X]$. As it turns out, studying algebraic (in)dependence is more straightforward in the field setting, which is why, in Section 6.2, we introduced the function field $K(X)$, associated to an irreducible affine variety. Given an irreducible affine variety, there is a natural set of inclusions

$$K \subseteq K[X] \subseteq K(X),$$

and algebraically independent elements in $K[X]$ remain algebraically independent in $K(X)$. With this motivation, we now turn to studying algebraic (in)dependence in the setting of field extensions of the form $K \subseteq K(X)$.

To help us discuss field extensions, we introduce the notion of field generators.

6.9 **DEFINITION** *Field generators*

Let $K \subseteq L$ be fields and let $\mathcal{S} \subseteq L$ be a subset. The *field extension of K generated by $\mathcal{S}$* is

$$K(\mathcal{S}) = \left\{ ab^{-1} : a, b \in K[\mathcal{S}],\ b \neq 0 \right\} \subseteq L.$$

We say that L is a *finitely-generated field extension of K* if $L = K(\mathcal{S})$ for a finite set $\mathcal{S} \subseteq L$.

Notice that $K(\mathcal{S})$ is the smallest subfield of L that contains both K and $\mathcal{S}$ (Exercise 6.3.1), and it is canonically isomorphic to $\mathrm{Frac}(K[\mathcal{S}])$. In addition, we have (Exercise 6.3.2)

$$K(\mathcal{S}_1)(\mathcal{S}_2) = K(\mathcal{S}_1 \cup \mathcal{S}_2).$$

When $\mathcal{S} = \{a_1, \ldots, a_n\}$ is finite, we write $K(\mathcal{S}) = K(a_1, \ldots, a_n)$.

6.10 **EXAMPLE** Fields of rational functions

For our purposes, the prototype of a field extension of K is the field of rational functions

$$K \subseteq K(x_1, \ldots, x_n).$$

As the notation suggests, $K(x_1, \ldots, x_n)$ is finitely-generated by $x_1, \ldots, x_n$.

6.11 **EXAMPLE** Function fields

Let $X \subseteq \mathbb{A}^n$ be an irreducible affine variety with coordinate ring $K[X]$ and function field $K(X)$. The coordinate functions $[x_1], \ldots, [x_n] \in K[X]$ can naturally be viewed as elements of $K(X)$ and every element of $K(X)$ can be written as

$$\frac{[f(x_1, \ldots, x_n)]}{[g(x_1, \ldots, x_n)]} = \frac{f([x_1], \ldots, [x_n])}{g([x_1], \ldots, [x_n])},$$

for some $f, g \in K[x_1, \ldots, x_n]$ with $[g] \neq 0$. By definition, the final expression lies in $K([x_1], \ldots, [x_n])$, implying that $K(X)$ is finitely-generated by $[x_1], \ldots, [x_n]$.

We now aim to define the key notion of transcendence bases, which is an analogue in the field extension setting of vector space bases. Recall that a basis of a vector space must satisfy two conditions: (i) it must be linearly independent and (ii) it must span the vector space. The first condition places an upper bound on the size of a basis—if you have too many elements in a vector space, then they will be linearly dependent—and the second condition places a lower bound on the size of a basis—if you have too few elements, then they will not span the entire vector space. In the same way, transcendence bases are required to satisfy two properties and, as we will see in the next section, the first places an upper bound on the size of transcendence bases and the second places a lower bound.

6.12 **DEFINITION** *Transcendence basis*

Let $K \subseteq L$ be fields. A subset $\mathcal{S} \subseteq L$ is a *transcendence basis* of L over K if it satisfies the following two conditions:

(i) $\mathcal{S}$ is algebraically independent over K, and

(ii) L is algebraic over $K(\mathcal{S})$.

By clearing denominators, (ii) is equivalent to the seemingly stronger condition that L is algebraic over the K-algebra $K[\mathcal{S}]$ (Exercise 6.3.4).

Before discussing concrete examples of transcendence bases, it will be helpful to become familiar with a few properties regarding algebraic field extensions like the one appearing in condition (ii) of Definition 6.12. The next result, which is the field analogue of Theorem 5.27, is the key tool that we require.

6.13 **PROPOSITION** *Extending fields by algebraic elements*

Let $K \subseteq L$ be fields and let $a_1, \ldots, a_n \in L$. The following are equivalent:

(i) Each a_i is algebraic over K.

(ii) $K[a_1, \ldots, a_n]$ is a finite-dimensional vector space over K.

(iii) $K[a_1, \ldots, a_n]$ is algebraic over K.

(iv) $K[a_1, \ldots, a_n] = K(a_1, \ldots, a_n)$.

PROOF The equivalence of conditions (i) – (iii) is a special case of Theorem 5.27 when $R = K$ is a field. In this setting, algebraicity is equivalent to integrality, and finitely-generated modules are the same as finite-dimensional vector spaces.

To prove that (iv) is equivalent to the other three, it suffices to prove that (iv) is equivalent to (iii). That (iv) implies (iii) is essentially the statement of Zariski's Lemma (Lemma 5.36). More specifically, if $K[a_1, \ldots, a_n] = K(a_1, \ldots, a_n)$, then $K[a_1, \ldots, a_n]$ is a field containing K that is finitely-generated as a K-algebra, from which Zariski's Lemma tells us that $K[a_1, \ldots, a_n]$ is algebraic over K.

It remains to prove that (iii) implies (iv). That $K[a_1, \ldots, a_n] \subseteq K(a_1, \ldots, a_n)$ follows from the definition of generators, so let us focus on the other inclusion. Using that $K(a_1, \ldots, a_n)$ is the smallest subfield of L containing K and $a_1, \ldots, a_n$,

this inclusion follows if we can prove that $K[a_1,\dots,a_n]$ is a field. To do so, let $a \in K[a_1,\dots,a_n]$ be nonzero; we must prove that a has a multiplicative inverse. Using the assumption that $K[a_1,\dots,a_n]$ is algebraic over K, we know that there is a nontrivial algebraic relation of the form

$$c_d a^d + \cdots + c_1 a + c_0 = 0 \quad \Longrightarrow \quad c_d a^d + \cdots + c_i a^i = 0,$$

where $c_0,\dots,c_d \in K$ and c_i is the first nonzero coefficient. Solving for a^i, we obtain

$$a^i = -c_i^{-1}(c_d a^d + \cdots + c_{i+1}a^{i+1}).$$

Cancelling the factor of a^i from both sides and factoring out a on the right, we conclude that a has an inverse:

$$1 = \big[-c_i^{-1}(c_d a^{d-i-1} + \cdots + c_{i+1})\big]a. \qquad \square$$

The one condition in Proposition 6.13 that is not a special case of Theorem 5.27 is condition (iv). We illustrate this condition in the following example.

6.14 EXAMPLE $\mathbb{Q}(\sqrt{2}) = \mathbb{Q}[\sqrt{2}]$

Consider $\sqrt{2} \in \mathbb{R}$. Since $\sqrt{2}$ is a root of the polynomial $x^2 - 2 \in \mathbb{Q}[x]$, we see that $\sqrt{2}$ is algebraic over $\mathbb{Q}$. Thus, Proposition 6.13 asserts that $\mathbb{Q}(\sqrt{2}) = \mathbb{Q}[\sqrt{2}]$, which we now verify.

Using the argument of Example 5.17, we compute that $\mathbb{Q}[\sqrt{2}] = \mathbb{Q}\{1, \sqrt{2}\}$. It then follows from the definition of field generators that every element of $\mathbb{Q}(\sqrt{2})$ has the form

$$\frac{a + b\sqrt{2}}{c + d\sqrt{2}}$$

where $a, b, c, d \in \mathbb{Q}$ with c and d not both zero. We can "rationalize" the quotient:

$$\frac{a + b\sqrt{2}}{c + d\sqrt{2}} = \frac{(a + b\sqrt{2})(c - d\sqrt{2})}{(c + d\sqrt{2})(c - d\sqrt{2})} = \frac{(ac - 2bd) + (ad + bc)\sqrt{2}}{c^2 - 2d^2} = r + q\sqrt{2},$$

where

$$r = \frac{ac - 2bd}{c^2 - 2d^2} \in \mathbb{Q} \quad \text{and} \quad q = \frac{ad + bc}{c^2 - 2d^2} \in \mathbb{Q}.$$

This shows that every element of $\mathbb{Q}(\sqrt{2})$ is actually an element of $\mathbb{Q}[\sqrt{2}]$. Since the other equality follows from the definitions, we conclude that $\mathbb{Q}(\sqrt{2}) = \mathbb{Q}[\sqrt{2}]$.

One of the most important consequences of Proposition 6.13 is the following. The failure of this corollary in the ring setting—see Example 6.16 below—is the essential reason why studying algebraic (in)dependence is easier in the field setting than it is in the ring setting.

6.15 COROLLARY *Algebraic over algebraic is algebraic*

Let $K \subseteq L \subseteq M$ be finitely-generated field extensions. If L is algebraic over K and M is algebraic over L, then M is algebraic over K.

PROOF By Proposition 6.13, the claim is equivalent to proving that, if L is a finite-dimensional vector space over K and M is a finite-dimensional vector space over L, then M is a finite-dimensional vector space over K. This follows from the observation (see Exercise 5.2.6) that, if $L = K\{a_1, \ldots, a_m\}$ and $M = L\{b_1, \ldots, b_n\}$, then

$$M = K\{a_i b_j \mid 1 \leq i \leq m,\ 1 \leq j \leq n\}.$$ □

The next example illustrates the failure of Corollary 6.15 in the ring setting.

6.16 **EXAMPLE** Corollary 6.15 fails for ring extensions

Consider the ring extensions

$$\mathbb{C} \subseteq \frac{\mathbb{C}[x]}{\langle x^2 - 1\rangle} \subseteq \frac{\mathbb{C}[x,y]}{\langle x^2 - 1, (x-1)y\rangle}.$$

We show that the second ring is algebraic over the first and the third is algebraic over the second, but that the third ring is not algebraic over the first.

Beginning with the first extension, notice that, by using the relation $[x^2] = 1$, any element of $\mathbb{C}[x]/\langle x^2 - 1\rangle$ can be written as $[ax + b]$ for some $a, b \in \mathbb{C}$. A direct computation shows that such an element $[ax + b]$ is a root of the polynomial

$$f(z) = z^2 - 2bz + b^2 - a^2 \in \mathbb{C}[z],$$

proving that the first extension is algebraic.

To show that the second extension is algebraic, notice that, by using the relations $[x^2] = 1$ and $[xy] = [y]$, any element of $\mathbb{C}[x,y]/\langle x^2 - 1, (x-1)y\rangle$ can be written as $[ax + b + yg(y)]$ for some $a, b \in \mathbb{C}$ and $g(y) \in \mathbb{C}[y]$. A direct computation shows that such an element $[ax + b + yg(y)]$ is a root of the polynomial

$$f(z) = [x-1]z + [ax - bx + b - a] \in \big(\mathbb{C}[x]/\langle x^2 - 1\rangle\big)[z],$$

proving that the second extension is also algebraic.

However, even though both extensions are algebraic, the extension

$$\mathbb{C} \subseteq \frac{\mathbb{C}[x,y]}{\langle x^2 - 1, (x-1)y\rangle}$$

is not algebraic. To convince ourselves of this, it is enough to argue that $[y]$ is not algebraic over $\mathbb{C}$. Indeed, every nonzero element of $\langle x^2 - 1, (x-1)y\rangle \subseteq \mathbb{C}[x,y]$ is necessarily a multiple of $x - 1$, and therefore, given any nonzero polynomial $f(z) \in \mathbb{C}[z]$, it will never be the case that

$$f([y]) = [f(y)] = 0 \in \frac{\mathbb{C}[x,y]}{\langle x^2 - 1, (x-1)y\rangle}.$$

While Corollary 6.15 fails over rings in general, by passing to fraction fields, it can be proved in the special setting of integral domains (Exercise 6.3.5).

We now close this section with several concrete examples to illustrate the notion of transcendence bases. The examples range from the empty example to a long-standing open problem concerning transcendental numbers in $\mathbb{R}$.

6.17 **EXAMPLE** Transcendence bases of algebraic extensions

A field extension $K \subseteq L$ has an empty transcendence basis if and only if L is algebraic over K. To verify this, first notice that $\emptyset \subseteq L$ is algebraically independent for vacuous reasons (since $\emptyset$ does not contain any elements, there cannot be an algebraic relation among them). Thus, by definition, $\emptyset \subseteq L$ is a transcendence basis if and only if L is algebraic over $K(\emptyset) = K$.

6.18 **EXAMPLE** Transcendence basis for the unit sphere

Let $K = \mathbb{C}$ and let $X = \mathcal{V}(x^2 + y^2 + z^2 - 1) \subseteq \mathbb{A}^3$. We verify that $\{[x], [y]\}$ is a transcendence basis of the field extension $\mathbb{C} \subseteq \mathbb{C}(X)$. To do this, we have two conditions to check.

First, we verify that $\{[x], [y]\}$ is algebraically independent over $\mathbb{C}$. Suppose that there is a polynomial relation $0 = f([x], [y])$. We must prove that $f = 0$. Since $0 = f([x], [y]) = [f(x, y)] \in \mathbb{C}[X]$, we have

$$f(x, y) \in \mathcal{I}(X) = \langle x^2 + y^2 + z^2 - 1 \rangle.$$

However, since every nonzero element of $\langle x^2 + y^2 + z^2 - 1 \rangle$ has positive degree in z and $f(x, y)$ does not, this implies that $f(x, y)$ must be the zero polynomial, and we conclude that $\{[x], [y]\}$ is algebraically independent over $\mathbb{C}$.

In order to prove that $\mathbb{C}(X)$ is algebraic over $\mathbb{C}([x], [y])$, it suffices, by Proposition 6.13, to prove that $[z]$ is algebraic over $\mathbb{C}([x], [y])$. This follows from the observation that $[z]$ is a zero of the polynomial

$$g(w) = w^2 + ([x]^2 + [y]^2 - 1) \in \mathbb{C}([x], [y])[w].$$

More generally, if $f \in K[x_1, \ldots, x_n]$ is an irreducible polynomial not contained in $K[x_1, \ldots, x_{n-1}]$, an analogous argument shows that $\{[x_1], \ldots, [x_{n-1}]\}$ is a transcendence basis of

$$\mathrm{Frac}(K[x_1, \ldots, x_n] / \langle f \rangle).$$

6.19 **EXAMPLE** π and e

It is a long-standing open problem to determine whether π and e are algebraically independent over $\mathbb{Q}$. If they are algebraically independent, then $\{\pi, e\}$ is a transcendence basis of $\mathbb{Q}(\pi, e)$ over $\mathbb{Q}$; otherwise, either $\{\pi\}$ or $\{e\}$ would suffice.

Now that we have seen a few examples of transcendence bases, some natural questions arise. Do transcendence bases exist for all field extensions? If so, what can be said about the size of transcendence bases? In the next section, we prove that, if $K \subseteq L$ is a finitely-generated field extension, then finite transcendence bases exist, and they all have the same size. This will allow us to define the transcendence degree of such a field extension as the size of any transcendence basis, similarly to the way one can define the dimension of a vector space. The notion of transcendence degree will be the key to making a rigorous definition of dimension of affine varieties in Section 6.5.

Exercises for Section 6.3

6.3.1 Let $K \subseteq L$ be a field extension and let $\mathcal{S} \subseteq L$ be a subset.

(a) Prove that $K(\mathcal{S})$ is a subfield of L.

(b) If L' is a subfield of L that contains K and $\mathcal{S}$, prove that $K(\mathcal{S}) \subseteq L'$.

6.3.2 Let $K \subseteq L$ be a field extension and let $\mathcal{S}_1, \mathcal{S}_2 \subseteq L$ be subsets. Prove that

$$K(\mathcal{S}_1)(\mathcal{S}_2) = K(\mathcal{S}_1 \cup \mathcal{S}_2).$$

6.3.3 Let A be a finitely-generated K-algebra that is also an integral domain and let $\mathcal{S} \subseteq A$ be a set of generators. Prove that $\mathcal{S}$ generates $\mathrm{Frac}(A)$ as a field extension of K.

6.3.4 Let $K \subseteq L$ be a field extension and let $\mathcal{S} \subseteq L$ be a set. Prove that L is algebraic over $K(\mathcal{S})$ if and only if L is algebraic over $K[\mathcal{S}]$.

6.3.5 Let K be a field and suppose that $A \subseteq B$ are finitely-generated K-algebras such that A is algebraic over K and B is algebraic over A. Assuming that A and B are both integral domains, prove that B is algebraic over K.

(**Hint:** Use fraction fields.)

6.3.6 Let $f \in K[x_1, \ldots, x_n]$ be an irreducible polynomial that is not an element of $K[x_1, \ldots, x_{n-1}]$ and consider the irreducible affine variety $X = \mathcal{V}(f)$. Prove that $\{[x_1], \ldots, [x_{n-1}]\}$ is a transcendence basis of $K(X)$ over K.

Section 6.4 Transcendence degree

In the last section, we familiarized ourselves with transcendence bases. In many ways, transcendence bases of field extensions are similar to bases of vector spaces. Importantly, bases of vector spaces provide a way of measuring how big a vector space is: all bases have the same size and the dimension of a vector space is the size of any basis. In this section, we develop the analogous notions for transcendence bases. In particular, we show that every finitely-generated field extension has a finite transcendence basis and that all such bases have the same size. This allows us to define the *transcendence degree* of a field extension $K \subseteq L$, which is a measure of the size of L relative to K.

We begin by proving the existence of finite transcendence bases.

6.20 **PROPOSITION** *Existence of finite transcendence bases*

Let $K \subseteq L$ be fields and $a_1, \ldots, a_n \in L$. If L is algebraic over $K(a_1, \ldots, a_n)$, then $\{a_1, \ldots, a_n\}$ contains a transcendence basis of L over K. In particular, finite transcendence bases exist for all finitely-generated field extensions.

PROOF We prove the result by induction on n. If $n = 0$, then L is algebraic over K and $\emptyset$ is a transcendence basis (Example 6.17), proving the base case.

Suppose that the result is true for all $b_1, \ldots, b_k \in L$ with $k < n$, and let $a_1, \ldots, a_n \in L$ be such that L is algebraic over $K(a_1, \ldots, a_n)$. If $\{a_1, \ldots, a_n\}$ is algebraically independent, then $\{a_1, \ldots, a_n\}$ is a transcendence basis of L over K, and we're done. Otherwise, there must exist some polynomial relation among the elements $a_1, \ldots, a_n$. Let $f \in K[x_1, \ldots, x_n]$ be a nonzero polynomial such that $f(a_1, \ldots, a_n) = 0$. This polynomial relation must depend nontrivially on at least one of the a_i; assume without loss of generality that it depends on a_n. Then a_n is a solution of

$$f(a_1, \ldots, a_{n-1}, x_n) \in K[a_1, \ldots, a_{n-1}][x_n] \subseteq K(a_1, \ldots, a_{n-1})[x_n]$$

showing that a_n is algebraic over $K(a_1, \ldots, a_{n-1})$. Thus, by Proposition 6.13,

$$K(a_1, \ldots, a_{n-1}) \subseteq K(a_1, \ldots, a_n)$$

is an algebraic extension. By the induction hypothesis, $\{a_1, \ldots, a_{n-1}\}$ contains a transcendence basis of $K(a_1, \ldots, a_{n-1})$. Up to relabeling, we can assume this transcendence basis is $\{a_1, \ldots, a_k\}$ for some $k < n$. Then $\{a_1, \ldots, a_k\}$ is algebraically independent and

$$K(a_1, \ldots, a_k) \subseteq K(a_1, \ldots, a_{n-1})$$

is an algebraic extension. We now have a sequence of algebraic extensions

$$K(a_1, \ldots, a_k) \subseteq K(a_1, \ldots, a_{n-1}) \subseteq K(a_1, \ldots, a_n) \subseteq L.$$

By Corollary 6.15, we conclude that $K(a_1, \ldots, a_k) \subseteq L$ is an algebraic extension. Along with the fact that $\{a_1, \ldots, a_k\}$ is algebraically independent, this tells us that $\{a_1, \ldots, a_k\}$ is a transcendence basis of L over K, completing the induction step.

The final statement in the proposition follows from the fact that, if L is finitely-generated over K, then L is equal to—and thus algebraic over—$K(a_1,\ldots,a_n)$ for any set of generators $a_1,\ldots,a_n \in L$. □

With a little more work and the Axiom of Choice, one can prove that (possibly infinite) transcendence bases exist for all field extensions.

Our next task is to prove that every transcendence basis has the same size. The next lemma, named after the strategy used in the proof, is the key result that we require.

6.21 LEMMA *The exchange lemma*

Let $K \subseteq L$ be fields and let $a_1,\ldots,a_n \in L$ be elements such that L is algebraic over $K(a_1,\ldots,a_n)$. If a set $\mathcal{S} \subseteq L$ is algebraically independent over K, then $\mathcal{S}$ is finite and $|\mathcal{S}| \leq n$.

PROOF Toward a contradiction, suppose that $|\mathcal{S}| > n$. To prove the result, we recursively "exchange" elements in $\{a_1,\ldots,a_n\}$ for elements of $\mathcal{S}$. In other words, at the completion of the kth step in this process, we will have chosen distinct elements $b_1,\ldots,b_k \in \mathcal{S}$ such that L is algebraic over $K(b_1,\ldots,b_k,a_{k+1},\ldots,a_n)$. We now describe this recursive process, starting with the first step.

(Step 1) Choose an element $b_1 \in \mathcal{S}$. Since L is algebraic over $K(a_1,\ldots,a_n)$, it follows that L is algebraic over $K[a_1,\ldots,a_n]$ (Exercise 6.3.4). Thus, we can choose a polynomial $f(x,a_1,\ldots,a_n) \in K[a_1,\ldots,a_n][x]$ with

$$f(b_1,a_1,\ldots,a_n) = 0.$$

Since $b_1 \in \mathcal{S}$ and $\mathcal{S}$ is algebraically independent over K, it follows that f depends on at least one of the a_i—without loss of generality, a_1. Then we can view a_1 as a root of polynomial $f(b_1,x,a_2,\ldots,a_n) \in K[b_1,a_2,\ldots,a_n][x]$. This implies that a_1 is algebraic over $K[b_1,a_2,\ldots,a_n]$ and, thus, also algebraic over the larger ring $K(b_1,a_2,\ldots,a_n)$. By Proposition 6.13, we obtain an algebraic field extension

$$\begin{aligned} K(b_1,a_2,\ldots,a_n) &\subseteq K(b_1,a_2,\ldots,a_n)[a_1] \\ &= K(b_1,a_2,\ldots,a_n)(a_1) \\ &= K(b_1,a_1,a_2,\ldots,a_n). \end{aligned}$$

In addition, since L is algebraic over $K(a_1,\ldots,a_n)$, it is also algebraic over the larger field $K(b_1,a_1,\ldots,a_n)$. Thus, we see that both of the extensions in the chain

$$K(b_1,a_2,\ldots,a_n) \subseteq K(b_1,a_1,a_2,\ldots,a_n) \subseteq L$$

are algebraic extensions. Applying Corollary 6.15 to this chain of extensions, we see that L is algebraic over $K(b_1,a_2,\ldots,a_n)$, completing Step 1.

(Step k) Suppose that, for some $k \leq n$ we have chosen $b_1,\ldots,b_{k-1} \in \mathcal{S}$ such that L is algebraic over $K(b_1,\ldots,b_{k-1},a_k,\ldots,a_n)$. Since $|\mathcal{S}| > n \geq k$, we may choose another element $b_k \in \mathcal{S} \setminus \{b_1,\ldots,b_{k-1}\}$, and there must be a polynomial relation $f(b_1,\ldots,b_k,a_k,\ldots,a_n) = 0$. Since $b_1,\ldots,b_k$ are elements of the algebraically independent set $\mathcal{S}$, this relation must depend on at least one of the a_i;

without loss of generality, suppose it depends on a_k. Then a_k is algebraic over $K(b_1,\dots,b_k,a_{k+1},\dots,a_n)$ and thus, both extensions in the chain

$$K(b_1,\dots,b_k,a_{k+1},\dots,a_n) \subseteq K(b_1,\dots,b_k,a_k,\dots,a_n) \subseteq L$$

are algebraic extensions. Therefore, L is algebraic over $K(b_1,\dots,b_k,a_{k+1},\dots,a_n)$, which completes Step k.

After completing the nth step in the recursion described above, we obtain a subset $\{b_1,\dots,b_n\} \subsetneq \mathcal{S}$ such that L is algebraic over $K(b_1,\dots,b_n)$. Thus, any $b \in \mathcal{S} \setminus \{b_1,\dots,b_n\} \subseteq L$ is algebraic over $K[b_1,\dots,b_n]$, which leads to a contradiction of the algebraic independence of $\mathcal{S}$. □

The exchange lemma allows us to deduce the following important result.

6.22 **THEOREM** *Transcendence bases all have the same size*

If L is a finitely-generated field extension of K, then every transcendence basis of L over K has the same finite size.

PROOF Suppose that L is a finitely-generated field extension of K. By Proposition 6.20, we know that a finite transcendence basis exists, so choose one such transcendence basis $\{a_1,\dots,a_n\}$. Let $\mathcal{S}$ be any other transcendence basis. By definition, L is algebraic over $K(a_1,\dots,a_n)$ and $\mathcal{S}$ is algebraically independent. Thus, by the exchange lemma, $\mathcal{S}$ is finite and $|\mathcal{S}| \leq n$. Write $\mathcal{S} = \{b_1,\dots,b_m\}$ with $m \leq n$. To finish the proof, we must prove that $m \geq n$. Since $\{a_1,\dots,a_n\}$ and $\{b_1,\dots,b_m\}$ are both transcendence bases, L is algebraic over $K(b_1,\dots,b_m)$ and $\{a_1,\dots,a_n\}$ is algebraically independent. Thus, by the exchange lemma again, we see that $m \geq n$. □

We now come to the following central definition, which will be the key to defining and studying the dimension of affine varieties in the remainder of this chapter.

6.23 **DEFINITION** *Transcendence degree*

Let $K \subseteq L$ be a finitely-generated field extension. The *transcendence degree of L over K*, denoted $\operatorname{trdeg}_K(L)$, is the size of any transcendence basis of L over K.

We revisit the examples from the previous section.

6.24 **EXAMPLE** Transcendence degree of algebraic extensions

Recall from Example 6.17 that $K \subseteq L$ is algebraic if and only if $\emptyset$ is a transcendence basis. In other words, we see that $K \subseteq L$ is algebraic if and only if $\operatorname{trdeg}_K(L) = 0$. Thus, algebraic field extensions are the smallest type of field extensions.

6.25 **EXAMPLE** Transcendence degree and the unit sphere

Let $K = \mathbb{C}$ and let $X = \mathcal{V}(x^2+y^2+z^2-1) \subseteq \mathbb{A}^3$. As we saw in Example 6.18, the two-element set $\{[x],[y]\} \subseteq K(X)$ is a transcendence basis over K. Thus, $\operatorname{trdeg}_K K(X) = 2$.

6.26 **EXAMPLE** Transcendence degree of $\mathbb{R}(\pi, e)$

To rephrase the open problem of Example 6.19, it is currently unknown whether $\operatorname{trdeg}_{\mathbb{R}}\mathbb{R}(\pi, e)$ is one or two.

To conclude this section, we list a few useful properties for bounding transcendence degree. Notice that Proposition 6.20 provides a tool for computing transcendence bases and transcendence degree "from above": every set $\{a_1, \ldots, a_n\} \subseteq L$ such that L is algebraic over $K(a_1, \ldots, a_n)$ can be trimmed down to a transcendence basis of L. In particular, any set of generators of L can be trimmed down to a transcendence basis. The next result provides a tool for computing transcendence bases and transcendence degree "from below": if $\{a_1, \ldots, a_n\} \subseteq L$ is an algebraically independent set, then it can be built up to a transcendence basis of L.

6.27 **PROPOSITION** *Independent sets are contained in bases*

Let $K \subseteq L$ be a finitely-generated field extension and suppose that $a_1, \ldots, a_n \in L$ are algebraically independent. Then $\{a_1, \ldots, a_n\}$ is contained in a transcendence basis of L over K.

PROOF Exercise 6.4.3 □

As a consequence of Propositions 6.20 and 6.27, we obtain the following result that provides a tool for finding upper and lower bounds on transcendence degree.

6.28 **COROLLARY** *Bounding transcendence degree*

Let $K \subseteq L$ be a finitely-generated field extension and let $a_1, \ldots, a_n \in L$.

1. If L is algebraic over $K(a_1, \ldots, a_n)$, then $\operatorname{trdeg}_K(L) \leq n$.
2. If $a_1, \ldots, a_n \in L$ are algebraically independent, then $\operatorname{trdeg}_K(L) \geq n$.

Exercises for Section 6.4

6.4.1 Suppose that $L \subseteq M$ are fields that are finitely-generated over K. Prove that $\operatorname{trdeg}_K(L) \leq \operatorname{trdeg}_K(M)$.

6.4.2 Compute the transcendence degree of $\mathbb{Q}(\pi, \mathrm{i}, \sqrt{2})$ over $\mathbb{Q}$.

6.4.3 Prove Proposition 6.27.

6.4.4 Suppose that $\operatorname{trdeg}_K(L) = d$.

(a) Prove that $\{a_1, \ldots, a_d\} \subseteq L$ is a transcendence basis of L over K if and only if $a_1, \ldots, a_d$ are algebraically independent.

(b) Prove that $\{a_1, \ldots, a_d\} \subseteq L$ is a transcendence basis of L over K if and only if L is algebraic over $K(a_1, \ldots, a_d)$.

6.4.5 Suppose that L is a field that is also a finitely-generated K-algebra.

(a) Prove that $\mathrm{trdeg}_K(L)$ is the maximum number of algebraically independent elements in L.

(b) Prove that $\mathrm{trdeg}_K(L)$ is the minimum number of elements of L that generate a field over which L is algebraic.

6.4.6 Let A be a finitely-generated K-algebra that is also an integral domain. Prove that any Noether basis of $A \subseteq \mathrm{Frac}(A)$ is a transcendence basis of $\mathrm{Frac}(A)$.

6.4.7 Using the notions of vector space dimension, span, and linear independence, state the linear algebra analogues of all of the results in this section, as well as Exercises 6.4.4 and 6.4.5.

Section 6.5 Dimension: definition and first properties

Now that we have introduced the concept of transcendence degree, we are ready to state the formal definition of dimension.

6.29 **DEFINITION** *Dimension of an affine variety*

Let X be an irreducible affine variety. The *dimension* of X is defined by

$$\dim_K(X) = \operatorname{trdeg}_K(K(X)).$$

The dimension of a nonempty reducible variety is the maximum dimension of its irreducible components. The empty set has dimension -1.

When the field K is clear from context, we omit it and write $\dim(X)$.

There are many terms regarding dimension that are commonly used. For example, varieties of dimension one and two are called *curves* and *surfaces*, respectively, while varieties of dimension n are often called *n-folds*. The *codimension* of an affine variety $X \subseteq \mathbb{A}^n$ is defined by

$$\operatorname{codim}(X) = n - \dim(X).$$

More generally, if $X \subseteq Y \subseteq \mathbb{A}^n$ are affine varieties, then the codimension of X in Y is defined by

$$\operatorname{codim}_Y(X) = \dim(Y) - \dim(X).$$

Let us compute dimension in a few concrete examples.

6.30 **EXAMPLE** Affine space $\mathbb{A}^n$ has dimension n

As we saw in Example 6.6, the function field of $\mathbb{A}^n$ is the field of rational functions:

$$K(\mathbb{A}^n) = K(x_1, \ldots, x_n).$$

Since $x_1, \ldots, x_n \in K(\mathbb{A}^n)$ are algebraically independent and generate $K(\mathbb{A}^n)$, they form a transcendence basis. Thus,

$$\dim(\mathbb{A}^n) = \operatorname{trdeg}_K(K(\mathbb{A}^n)) = n.$$

6.31 **EXAMPLE** The unit sphere has dimension 2

Let $K = \mathbb{C}$ and let $X = \mathcal{V}(x^2 + y^2 + z^2 - 1) \subseteq \mathbb{A}^3$. As we saw in Example 6.18, the two elements

$$[x], [y] \in \mathbb{C}(X)$$

form a transcendence basis. Thus,

$$\dim(X) = \operatorname{trdeg}_{\mathbb{C}}(\mathbb{C}(X)) = 2.$$

More generally (Exercise 6.5.4), if $f \in K[x_1, \ldots, x_n]$ is irreducible and $X = \mathcal{V}(f)$, then $\dim(X) = n - 1$.

We required the function field in order to ensure that transcendence degree was well-defined; however, now that we have laid the groundwork, we can discuss dimension purely in terms of coordinate rings. The next result describes how we can use the coordinate ring of possibly reducible varieties to bound their dimension.

6.32 PROPOSITION *Polynomial functions and dimension*

Let X be a nonempty affine variety and let $F_1, \ldots, F_d \in K[X]$.

1. If $F_1, \ldots, F_d$ are algebraically independent, then $d \leq \dim(X)$.
2. If $K[X]$ is integral over $K[F_1, \ldots, F_d]$, then $\dim(X) \leq d$.

PROOF We prove the first statement and leave the second as Exercise 6.5.5. Suppose that $F_1, \ldots, F_d \in K[X]$ are algebraically independent and consider the irreducible decomposition $X = X_1 \cup \cdots \cup X_m$. Restricting each F_i to X_j, we obtain functions $F_i|_{X_j} \in K[X_j]$. We claim that $F_1|_{X_j}, \ldots, F_d|_{X_j}$ are algebraically independent for at least one j. To see why, suppose this were not the case. Then, for each $j = 1, \ldots, m$, there would be a nonzero polynomial $g_j \in K[z_1, \ldots, z_d]$ such that

$$g_j(F_1|_{X_j}, \ldots, F_d|_{X_j}) = 0 \in K[X_j].$$

But if X is the union of the $X_1, \ldots, X_m$ and $g_j(F_1, \ldots, F_d)$ vanishes when restricted to X_j for $j = 1, \ldots, m$, then their product would vanish on X:

$$\prod_{j=1}^{m} g_j(F_1, \ldots, F_d) = 0 \in K[X].$$

However, this contradicts the assumption that $F_1, \ldots, F_d \in K[X]$ are algebraically independent.

Thus, for some j, the functions

$$F_1|_{X_j}, \ldots, F_d|_{X_j} \in K[X_j] \subseteq K(X_j).$$

are algebraically independent. It then follows that

$$\dim(X) \geq \dim(X_j) = \operatorname{trdeg}_K(K(X_j)) \geq d,$$

where the first inequality and the equality follow from the definition of dimension, while the second inequality follows from Corollary 6.28, Part 2. □

As a consequence of Proposition 6.32, we have the following result, which says that every Noether basis of a coordinate ring has the same size, equal to the dimension of the affine variety.

6.33 COROLLARY *Noether bases and dimension*

If X is a nonempty affine variety and $\{F_1, \ldots, F_d\} \subseteq K[X]$ is a Noether basis, then $d = \dim(X)$.

PROOF By definition of Noether basis, $F_1, \ldots, F_d$ are algebraically independent over K and $K[X]$ is integral over $K[F_1, \ldots, F_d]$. Thus, the two parts of Proposition 6.32 imply that $d \leq \dim(X)$ and $\dim(X) \geq d$, from which we conclude that $d = \dim(X)$. □

We can also tie our formal discussion of dimension back to the characterization of dimension given in Key Idea 6.2, which is the first part of the next result.

6.34 **COROLLARY** *Characterizations of dimension*

Let X be a nonempty affine variety.

1. The dimension of X is the maximum number of algebraically independent functions in $K[X]$.
2. The dimension of X is the minimum number of $F_1, \ldots, F_d \in K[X]$ such that $K[X]$ is integral over $K[F_1, \ldots, F_d]$.

PROOF We prove the first part and leave the second to Exercise 6.5.6.

If $\mathcal{S} \subseteq K[X]$ is algebraically independent, then the first part of Proposition 6.32 tells us that $\mathcal{S}$ is finite and $|\mathcal{S}| \leq \dim(X)$. Thus, the maximum number of algebraically independent functions in $K[X]$ is bounded above by $\dim(X)$. On the other hand, Noether normalization tells us that $K[X]$ contains a Noether basis, which, by Corollary 6.33, has size $\dim(X)$. This implies that $K[X]$ contains at least one algebraically independent set of size $\dim(X)$, and we conclude that the maximum number of algebraically independent elements in $K[X]$ is equal to $\dim(X)$. □

We now have a number of ways to think about the dimension of an affine variety, and we can use these ideas to prove that dimension satisfies certain natural properties. The first property regards inclusions of affine varieties. Since dimension measures the size of an affine variety, we should certainly expect that $X \subseteq Y$ implies $\dim(X) \leq \dim(Y)$. Less obvious, though, is the fact that this implication becomes strict when we restrict to irreducible affine varieties. In other words, if $Y \subseteq \mathbb{A}^n$ is an irreducible affine variety, then the only affine varieties that are strictly contained in Y must have strictly smaller dimension. This is the content of the next result.

6.35 **PROPOSITION** *Dimension and inclusions*

If $X, Y \subseteq \mathbb{A}^n$ are affine varieties with $X \subseteq Y$, then $\dim(X) \leq \dim(Y)$. Furthermore, if Y is irreducible and $X \subsetneq Y$, then $\dim(X) < \dim(Y)$.

PROOF Let $X, Y \subseteq \mathbb{A}^n$ be affine varieties with $X \subseteq Y$, and define $d = \dim(X)$. By Corollary 6.34, there exists an algebraically independent subset of $K[X]$ of size d. Let $f_1, \ldots, f_d \in K[x_1, \ldots, x_n]$ be polynomials such that $\{f_1|_X, \ldots, f_d|_X\}$ is an algebraically independent subset of $K[X]$. Since the functions $f_1, \ldots, f_d$ do not satisfy an algebraic relation when restricted to X, then they certainly do not satisfy an algebraic relation when restricted to the larger set Y (the reader is encouraged to pause and convince themselves of this assertion). Thus, $\{f_1|_Y, \ldots, f_d|_Y\}$ is an algebraically independent set in $K[Y]$, and by Corollary 6.34, we conclude that $\dim(Y) \geq d = \dim(X)$.

To prove the second statement, suppose, in addition, that Y is irreducible and that $X \subsetneq Y$. Our aim is to prove that $\dim(X) < \dim(Y)$. Toward a contradiction, assume that $\dim(Y) = \dim(X) = d$. Using the assumption that $X \subsetneq Y$, choose a polynomial $f \in K[x_1, \ldots, x_n]$ such that $f|_Y \neq 0 \in K[Y]$ but $f|_X = 0 \in K[X]$. Since the maximum number of algebraically independent elements in $K[Y]$ is $\dim(Y) = d$, the set $\{f_1|_Y, \ldots, f_d|_Y, f|_Y\}$ cannot be algebraically independent. Thus, we can choose a nontrivial polynomial

$$g(z) \in (K[f_1|_Y, \ldots, f_d|_Y])[z]$$

such that $g(f|_Y) = 0$. Without loss of generality, we may assume that g is such a polynomial of smallest possible degree. Writing the relation term-by-term, we have

$$0 = g_k(f_1|_Y, \ldots, f_d|_Y)(f|_Y)^k + \cdots + g_1(f_1|_Y, \ldots, f_d|_Y)f|_Y + g_0(f_1|_Y, \ldots, f_d|_Y)$$

Notice that g_0 is not the zero polynomial, since otherwise we could cancel one factor of $f|_Y$ (using the fact that Y is irreducible so $K[Y]$ is an integral domain), producing a polynomial of smaller degree that vanishes when evaluated at $f|_Y$. Upon restricting to X, this implies that g_0 is a nonzero polynomial relation among $f_1|_X, \ldots, f_d|_X$, contradicting that these are algebraically independent elements of $K[Y]$. □

The second statement of Proposition 6.35 fails without the assumption that Y is irreducible. For example, the x-axis $\mathcal{V}(y) \subseteq \mathbb{A}^2$ is strictly contained in the union of the axes $\mathcal{V}(xy) \subseteq \mathbb{A}^2$, but both varieties have dimension one.

We now have a rigorous definition of dimension, and it is not too difficult to prove that it satisfies the first two axioms of Definition 6.3 (see Exercises 6.5.1, 6.5.2, and 6.5.3). Therefore, it remains to prove Axiom 3, which is the content of the remaining section in this chapter.

Exercises for Section 6.5

6.5.1 Prove that dimension is an intrinsic property. In other words, prove that two isomorphic affine varieties have the same dimension.

6.5.2 Prove that the dimension of a single point is zero.

6.5.3 Let $X_1, \ldots, X_m \subseteq \mathbb{A}^n$. Use uniqueness of irreducible decompositions to prove that

$$\dim(X_1 \cup \cdots \cup X_m) = \max\{\dim(X_1), \ldots, \dim(X_m)\}.$$

6.5.4 Let $f \in K[x_1, \ldots, x_n]$ be irreducible and consider $X = \mathcal{V}(f) \subseteq \mathbb{A}^n$. Prove that $\dim(X) = n - 1$.

6.5.5 Let X be a nonempty affine variety and let $F_1, \ldots, F_d \in K[X]$ such that $K[X]$ is integral over $K[F_1, \ldots, F_d]$. Prove that $\dim(X) \leq d$.

6.5.6 Prove that $\dim(X)$ is the minimum number of functions $F_1, \ldots, F_d \in K[X]$ such that $K[X]$ is integral over $K[F_1, \ldots, F_d]$.

Section 6.6 The Fundamental Theorem

We now arrive at the technical heart of dimension theory: the Fundamental Theorem of Dimension Theory. This result says that, if $X \subseteq \mathbb{A}^n$ is an irreducible affine variety and $f \in K[x_1, \ldots, x_n]$, then either $X \cap \mathcal{V}(f)$ is a trivial intersection (empty or all of X) or $\dim(X \cap \mathcal{V}(f)) = \dim(X) - 1$.

The proof of this result is rather involved, and it requires some fortitude on the part of the reader. It draws on many of the ideas introduced in prior chapters, with a key step coming from Noether Normalization. In addition, it requires several new ideas regarding minimal polynomials and their relation to determinants of certain linear transformations of algebraic field extensions. We begin with a discussion of the new ideas, which are captured in Lemmas 6.36, 6.38, and 6.39, all of which will then be used in the proof of the Fundamental Theorem (Theorems 6.42).

6.36 LEMMA/DEFINITION *Minimal polynomial*

Let $L \subseteq M$ be a field extension. For any nonzero $a \in M$ that is algebraic over L, there exists a unique irreducible monic polynomial $\mu_a \in L[x]$ such that $\mu_a(a) = 0 \in M$. We call μ_a the *minimal polynomial of a over L*.

PROOF Let $L \subseteq M$ be a field extension and $a \in M$ algebraic over L. By algebraicity, there exists a nontrivial polynomial in $L[x]$ that vanishes at a. We may find such a polynomial of minimal possible degree, and by dividing by the leading coefficient, we can even find one that is monic. Let μ_a be a monic polynomial of minimal degree that vanishes at a.

To prove the lemma, we prove that every monic irreducible polynomial that vanishes at a is equal to μ_a. Suppose that $g \in L[x]$ is a monic irreducible polynomial that vanishes at a. By the division algorithm,

$$g = q\mu_a + r$$

where $r = 0$ or $\deg(r) < \deg(\mu_a)$. Evaluating at a and using $g(a) = \mu_a(a) = 0$, we see that $r(a) = 0$. Thus, the minimality of the degree of μ_a implies that $r = 0$. Therefore, $g = q\mu_a$, from which the irreducibility of g implies that q is a unit. Since both g and μ_a are monic, we must have $q = 1$, and we conclude that $g = \mu_a$. □

6.37 EXAMPLE Minimal polynomial of $\sqrt{2}$ over $\mathbb{Q}$

The minimal polynomial of $\sqrt{2}$ over $\mathbb{Q}$ is $x^2 - 2 \in \mathbb{Q}[x]$. To prove this, we simply observe that $x^2 - 2$ is monic, irreducible, and vanishes at $\sqrt{2}$, from which Lemma 6.36 implies that it is the minimal polynomial of $\sqrt{2}$ over $\mathbb{Q}$.

Minimal polynomials play a key role in the proof of Theorem 6.42. In that setting, we will have a Noether basis $\{F_1, \ldots, F_d\}$ of a coordinate ring $K[X]$ (where X is irreducible) and we will be considering the field extension

$$K(F_1, \ldots, F_d) \subseteq K(X).$$

The key result we require, which follows from the next lemma, is that, for any $F \in K[X] \subseteq K(X)$, the coefficients of μ_F lie in $K[F_1, \ldots, F_d] \subseteq K(F_1, \ldots, F_d)$.

6.38 LEMMA *Minimal polynomials and Noether bases*

Let A be a finitely-generated K-algebra that is also an integral domain and let $\{a_1, \ldots, a_d\}$ be a Noether basis of A over K.

1. The field extension $K(a_1, \ldots, a_d) \subseteq \mathrm{Frac}(A)$ is algebraic.
2. For any nonzero $a \in A \subseteq \mathrm{Frac}(A)$, the minimal polynomial μ_a of a over $K(a_1, \ldots, a_d)$ satisfies

$$\mu_a \in K[a_1, \ldots, a_d][x].$$

By definition, the minimal polynomial μ_a is an element of $K(a_1, \ldots, a_d)[x]$, meaning that the coefficients are rational functions in the Noether basis. The assertion of the second part of the lemma is that the coefficients of the minimal polynomial are actually *polynomials* in the Noether basis.

PROOF OF LEMMA 6.38 To prove that $K(a_1, \ldots, a_d) \subseteq \mathrm{Frac}(A)$ is algebraic, let $a/b \in \mathrm{Frac}(A)$. By the definition of Noether basis, A is integral and thus algebraic over $K[a_1, \ldots, a_d]$. Therefore, we may choose a polynomial $f \in K[a_1, \ldots, a_d][x]$ such that $f(a) = 0$. Write

$$f = f_d x^d + f_{d-1} x^{d-1} + \cdots + f_1 x + f_0$$

and define

$$g = (f_d b^d) x^d + (f_{d-1} b^{d-1}) x^{d-1} + \cdots + (f_1 b) x + f_0.$$

Then $g \in K[a_1, \ldots, a_d][x]$ and $g(a/b) = f(a) = 0$, proving that $\mathrm{Frac}(A)$ is algebraic over $K[a_1, \ldots, a_d]$ and thus algebraic over $K(a_1, \ldots, a_d)$.

To prove Part 2, let $a \in A$. Again using that A is integral over $K[a_1, \ldots, a_d]$, we can choose a monic polynomial $f \in K[a_1, \ldots, a_d][x]$ such that $f(a) = 0$. Furthermore, we can assume that f has minimal degree among the monic polynomials that vanish at a, in which case f must also be irreducible. By the definition of Noether basis, $a_1, \ldots, a_d$ are algebraically independent, so f is an irreducible element of the multivariable polynomial ring

$$f \in K[a_1, \ldots, a_d][x] = K[a_1, \ldots, a_d, x]$$

By repeated use of Proposition 0.59, the monic polynomial f remains irreducible in the larger ring $K(a_1, \ldots, a_d)[x]$, and since it vanishes at a, it must be equal to the minimal polynomial of a. Therefore, $\mu_a = f \in K[a_1, \ldots, a_d][x]$. □

To set up the final lemma required for the proof of Theorem 6.42, let $L \subseteq M$ be a finitely-generated algebraic extension, which, by Proposition 6.13, implies that M is a finite-dimensional vector space over L. For any element $a \in M$, define

$$\begin{aligned} T_a : M &\to M \\ b &\mapsto ab. \end{aligned}$$

Notice that this function is a linear transformation of M as a vector space over L. More precisely, for any $b_1, b_2 \in M$ and any $c \in L$, we check that

$$T_a(b_1 + cb_2) = a(b_1 + cb_2) = ab_1 + cab_2 = T_a(b_1) + cT_a(b_2).$$

As $T_a : M \to M$ is L-linear, it has a well-defined determinant $\det(T_a) \in L$, which can be computed by picking a basis of M over L, writing the linear transformation as a matrix, and computing the determinant of the matrix using any of the usual formulas for determinants. Importantly, the determinant is independent of the basis. The next lemma relates this determinant to the minimal polynomial of a.

6.39 LEMMA *Determinants of multiplication transformations*

Let $L \subseteq M$ be a finitely-generated algebraic field extension. Then, for any nonzero $a \in M$ with minimal polynomial $\mu_a(x) \in L[x]$, we have

$$\det(T_a) = \pm\mu_a(0)^\ell$$

for some positive integer ℓ.

PROOF Assume that $L \subseteq M$ is an algebraic extension, or equivalently, that M is a finite-dimensional vector space over L. Given any nonzero $a \in M$ with minimal polynomial

$$\mu_a(x) = x^d + \mu_{a,d-1}x^{d-1} + \cdots + \mu_{a,1}x + \mu_{a,0},$$

it follows from Exercise 6.6.1 that we can find a basis for M as an L-vector space of the form

$$\{b_1, b_1a, \ldots, b_1a^{d-1}, b_2, b_2a, \ldots, b_2a^{d-1}, \ldots, b_\ell, b_\ell a, \ldots, b_\ell a^{d-1}\}.$$

Writing the linear transformation T_a as a matrix in terms of this basis, T_a is a block diagonal matrix

$$T_a = \begin{pmatrix} T'_a & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & T'_a \end{pmatrix}$$

where

$$T'_a = \begin{pmatrix} 0 & 0 & \cdots & 0 & -\mu_{a,0} \\ 1 & 0 & \cdots & 0 & -\mu_{a,1} \\ 0 & 1 & \cdots & 0 & -\mu_{a,2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -\mu_{a,d-1} \end{pmatrix}.$$

Computing determinants in the standard way, we have

$$\det(T_a) = \det(T'_a)^\ell = ((-1)^{d-1}\mu_{a,0})^\ell = \pm\mu_a(0)^\ell. \qquad \square$$

6.40 EXAMPLE $\det(T_a)$ for $a \in L$

Let $L \subseteq M$ be an algebraic field extension and suppose that $a \in L$. Then the minimal polynomial of a is $\mu_a = x - a$. Therefore, Lemma 6.39 implies that $\det(T_a) = \pm a^\ell$ for some positive integer ℓ. This conclusion can also be argued directly using the fact that, in any basis of M as a vector space over L, we have $T_a = aI$ where I is the identity matrix. Thus, $\det(T_a) = \det(aI) = a^\ell$, where ℓ is the dimension of M as an L-vector space.

6.41 **EXAMPLE** $\det(T_a)$ with $a = \sqrt{2} \in \mathbb{Q}(\sqrt{2}, \sqrt{3})$

Consider $a = \sqrt{2} \in \mathbb{Q}(\sqrt{2}, \sqrt{3})$. A basis for $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ as a vector space over $\mathbb{Q}$ is given by

$$\{1, \sqrt{2}, \sqrt{3}, \sqrt{2}\sqrt{3}\}$$

In terms of this basis, we can write

$$T_a = \begin{pmatrix} 0 & 2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

From this matrix expression, we compute $\det(T_a) = (-2)^2 = \mu_a(0)^2$.

We are now prepared to prove the main result of this section. The following result is a restatement of Axiom 3 in Definition 6.3 and thus, completes the proof that the notion of dimension developed in this chapter is the unique dimension function on the set of isomorphism classes of affine varieties. We call this the "weak" version of the fundamental theorem because we will formulate a stronger version below.

6.42 **THEOREM** *Weak Fundamental Theorem of Dimension Theory*

If $X \subseteq \mathbb{A}^n$ is an irreducible affine variety and $f \in K[x_1, \ldots, x_n]$ is a polynomial such that $X \cap \mathcal{V}(f)$ is neither empty nor all of X, then

$$\dim(X \cap \mathcal{V}(f)) = \dim(X) - 1.$$

PROOF Let $X \subseteq \mathbb{A}^n$ be an irreducible affine variety and $f \in K[x_1, \ldots, x_n]$ such that $X \cap \mathcal{V}(f)$ is neither empty nor all of X. For convenience, let us define $Y = X \cap \mathcal{V}(f)$ and set $F = f|_X \in K[X]$. We aim to prove that

$$\dim(Y) = \dim(X) - 1.$$

Since Y is an affine variety strictly contained in the irreducible affine variety X, Proposition 6.35 tells us that $\dim(Y) \leq \dim(X) - 1$. Thus, it remains to prove

$$\dim(Y) \geq \dim(X) - 1.$$

By Noether normalization, we may choose a Noether basis $F_1, \ldots, F_d \in K[X]$, and Corollary 6.33 implies that $\dim(X) = d$. Let μ_F be the minimal polynomial of $F \in K[X] \subseteq K(X)$ over $K(F_1, \ldots, F_d)$. By Lemma 6.38,

$$\mu_F \in K[F_1, \ldots, F_d][x].$$

Write

$$\mu_F(x) = x^k + \mu_{F,k-1}x^{k-1} + \cdots + \mu_{F,1}x + \mu_{F,0}.$$

Since $\mu_F(F) = 0$ and $F|_Y = 0$, we see that $\mu_{F,0}|_Y = 0$. Notice that $\mu_{F,0}$ is neither zero nor a unit; if it were zero, then $\mu_F(x)$ would be a reducible polynomial, and if it were a unit, then $\mu_{F,0}|_Y \neq 0$. The following claim is central to the proof.

Claim: If $G \in K[F_1, \ldots, F_d] \subseteq K[X]$, then

$$G|_Y = 0 \quad \text{if and only if} \quad G \in \sqrt{\langle \mu_{F,0} \rangle} \subseteq K[F_1, \ldots, F_d].$$

($\Rightarrow$): Suppose that $G \in K[F_1, \ldots, F_d]$ and $G|_Y = 0$. Since $Y = X \cap \mathcal{V}(f)$, it follows from the Nullstellensatz (Exercise 6.6.2) that

$$G \in \sqrt{\langle f|_X \rangle} = \sqrt{\langle F \rangle} \subseteq K[X].$$

Thus, $G^m = HF$ for some $H \in K[X]$. Considering the algebraic field extension $K(F_1, \ldots, F_d) \subseteq K(X)$, we see that there exist positive integers ℓ_1, ℓ_2, ℓ_3 such that

$$\begin{aligned} G^{m\ell_1} &= \det(T_{G^m}) \\ &= \det(T_{HF}) \\ &= \det(T_H)\det(T_F) \\ &= \pm \mu_{H,0}^{\ell_2} \mu_{F,0}^{\ell_3}. \end{aligned}$$

The first equality follows from $G^m \in K[F_1, \ldots, F_d]$ (as in Example 6.40), the second from $G^m = HF$, the third from multiplicativity of determinants and the fact that $T_{HF} = T_H T_F$, and the fourth from Lemma 6.39. Since $H \in K[X]$, Lemma 6.38 implies that $\mu_{H,0} \in K[F_1, \ldots, F_d]$. Thus, the equation

$$G^{m\ell_1} = \pm \mu_{H,0}^{\ell_2} \mu_{F,0}^{\ell_3}$$

lives in the polynomial ring $K[F_1, \ldots, F_d]$, so $G \in \sqrt{\langle \mu_{F,0} \rangle} \subseteq K[F_1, \ldots, F_d]$.

($\Leftarrow$): If $G^m = H\mu_{F,0}$ for some positive integer m and some $H \in K[F_1, \ldots, F_d]$, then $\mu_{F,0}|_Y = 0$ implies that $G^m|_Y = 0$, from which it follows that $G|_Y = 0$.

Having proved the claim, we now prove the theorem. Consider the restriction

$$\begin{aligned} \varphi : K[F_1, \ldots, F_d] &\to K[Y] \\ G &\mapsto G|_Y. \end{aligned}$$

The claim implies that $\ker(\varphi) = \sqrt{\langle \mu_{F,0} \rangle}$. Since $F_1, \ldots, F_d$ are algebraically independent, $K[F_1, \ldots, F_d]$ is a polynomial ring, so Proposition 1.31 implies that $\sqrt{\langle \mu_{F,0} \rangle} = \langle Q \rangle$ where Q is the product of the distinct irreducible factors of $\mu_{F,0}$. Thus, by the First Isomorphism Theorem, we obtain an injection

$$\begin{aligned} [\varphi] : \frac{K[F_1, \ldots, F_d]}{\ker(\varphi) = \langle Q \rangle} &\to K[Y] \\ [G] &\mapsto G|_Y. \end{aligned}$$

Since $\mu_{F,0}$ is not a unit, then neither is Q, so it must depend on at least one of the generators; without loss of generality, assume that it depends on F_d. Then $[F_1], \ldots, [F_{d-1}]$ are algebraically independent in the domain of $[\varphi]$, and by injectivity, it follows that $F_1|_Y, \ldots, F_{d-1}|_Y$ are algebraically independent in $K[Y]$. Thus, Proposition 6.32 implies that $\dim(Y) \geq d - 1$, as desired. $\square$

As mentioned before the statement of Theorem 6.42, we can actually prove a stronger statement of the fundamental theorem. Theorem 6.42 asserts that *at least one* of the irreducible components of $X \cap \mathcal{V}(f)$ has dimension $\dim(X) - 1$. However, it actually turns out that *every* irreducible component of $X \cap \mathcal{V}(f)$ has dimension $\dim(X) - 1$, as we now verify.

6.43 THEOREM *Strong Fundamental Theorem of Dimension*

If $X \subseteq \mathbb{A}^n$ is an irreducible affine variety and $f \in K[x_1, \ldots, x_n]$ is a polynomial such that $X \cap \mathcal{V}(f)$ is neither empty nor all of X, then every irreducible component of $X \cap \mathcal{V}(f)$ has dimension $\dim(X) - 1$.

PROOF As in the proof of Theorem 6.42, set $Y = X \cap \mathcal{V}(f)$. Let Y' be any irreducible component of Y and let Y'' be the union of the other irreducible components. Since $Y'' \subsetneq Y' \cup Y''$, we may choose a polynomial $g \in K[x_1, \ldots, x_n]$ such that $g|_{Y''} = 0$ but $g|_{Y'} \neq 0$. Choose defining equations $X = \mathcal{V}(f_1, \ldots, f_m)$, and define two new varieties in $\mathbb{A}^{n+1}$ by

$$\widetilde{X} = \mathcal{V}(f_1, \ldots, f_m, x_{n+1}g - 1) \quad \text{and} \quad \widetilde{Y} = \mathcal{V}(f_1, \ldots, f_m, f, x_{n+1}g - 1).$$

Notice that $\widetilde{Y} = \widetilde{X} \cap \mathcal{V}(f)$.

It can be shown (see Exercise 6.6.3) that $\widetilde{X}$ and $\widetilde{Y}$ are irreducible with

$$K(\widetilde{X}) \cong K(X) \quad \text{and} \quad K(\widetilde{Y}) \cong K(Y').$$

Since dimension is defined only in terms of the function field, it follows that

$$\dim(\widetilde{X}) = \dim(X) \quad \text{and} \quad \dim(\widetilde{Y}) = \dim(Y').$$

Since $\emptyset \subsetneq Y' \subsetneq X$, we see that

$$-1 < \dim(Y') = \dim(\widetilde{Y}) < \dim(X) = \dim(\widetilde{X}).$$

This implies that $\widetilde{Y} = \widetilde{X} \cap \mathcal{V}(f)$ is a nonempty, proper subset of $\widetilde{X}$. Therefore, the hypotheses of Theorem 6.42 are all met with respect to $\widetilde{X}$ and $\widetilde{Y}$, and we conclude that

$$\dim(Y') = \dim(\widetilde{Y}) = \dim(\widetilde{X}) - 1 = \dim(X) - 1. \qquad \square$$

By repeatedly applying Theorem 6.43, one obtains the following concrete application, which says that the codimension of an affine variety is bounded above by the number of defining equations.

6.44 COROLLARY *Defining equations and dimension*

If $X = \mathcal{V}(f_1, \ldots, f_k) \subseteq \mathbb{A}^n$ is nonempty, then $\dim(X) \geq n - k$.

PROOF Exercise 6.6.4 $\square$

In proving the previous corollary, the reader will recognize why the strong form of the Fundamental Theorem of Dimension is so much more useful than the weak.

Theorem 6.43 also implies the following alternative characterization of dimension, which, in many algebraic geometry textbooks, is taken as the definition.

6.45 COROLLARY *Chains of inclusions and dimension*

If $X \subseteq \mathbb{A}^n$ is an affine variety, then $\dim(X)$ is equal to the maximum d such that there exist irreducible affine varieties $X_0, \ldots, X_d \subseteq \mathbb{A}^n$ with

$$X_0 \subsetneq X_1 \subsetneq \cdots \subsetneq X_d \subseteq X.$$

PROOF Exercise 6.6.6 □

In concluding this chapter, we mention that it is totally reasonable to study dimension of algebraic varieties over fields that are not algebraically closed, such as $\mathbb{R}$. In this setting, the Fundamental Theorem of Dimension Theory fails wildly (see Exercise 6.6.7, for example). However, the results discussed in Section 6.5 continue to hold, because they did not require the Nullstellensatz. Surprisingly, even though our proof of Corollary 6.45 required the Nullstellensatz, one can actually circumvent the Nullstellensatz and prove that this characterization is equal to the transcendence degree definition over general fields. For this, and more general results on dimension theory, we direct the reader to a more advanced text on commutative algebra.

Exercises for Section 6.6

6.6.1 Let $L \subseteq M$ be an algebraic extension and let a be an element of M whose minimal polynomial has degree d.

(a) Prove that $\{1, a, \ldots, a^{d-1}\}$ is a basis of $L(a)$ as a vector space over L.

(b) If $b_1, \ldots, b_\ell \subseteq M$ is a basis for M as a vector space over $L(a)$, prove that $\{b_i a^j \mid 1 \leq i \leq \ell,\ 0 \leq j \leq d-1\}$ is a basis of M as a vector space over L.

(c) Given a basis of the form in Part (b), prove that the matrix associated to the linear transformation $T_a : M \to M$ is block diagonal of the form given in the proof of Lemma 6.39.

6.6.2 Let $X \subseteq \mathbb{A}^n$ be an affine variety and $f \in K[x_1, \ldots, x_n]$. Use the Nullstellensatz to prove that $F \in K[X]$ vanishes on $X \cap \mathcal{V}(f)$ if and only if

$$F \in \sqrt{\langle f|_X \rangle}.$$

6.6.3 Let $X = \mathcal{V}(f_1, \ldots, f_m) \subseteq \mathbb{A}^n$ and let X' be any irreducible component of X. Let $g \in K[x_1, \ldots, x_n]$ be a polynomial that vanishes on every component of X except X', and define

$$\tilde{X} = \mathcal{V}(f_1, \ldots, f_m, x_{n+1}g - 1) \subseteq \mathbb{A}^{n+1}.$$

This exercise proves that $\tilde{X}$ is irreducible and that $K(\tilde{X}) = K(X')$.

(a) Consider the homomorphism

$$\begin{aligned} \varphi : K[x_1, \ldots, x_n, x_{n+1}] &\to K(X') \\ h(x_1, \ldots, x_n, x_{n+1}) &\mapsto [h(x_1, \ldots, x_n, g^{-1})]. \end{aligned}$$

Prove that $\ker(\varphi) = \mathcal{I}(\widetilde{X})$.

(b) Applying the First Isomorphism Theorem, Part (a) implies that the homomorphism $[\varphi] : K[\widetilde{X}] \to K(X')$ is an injection. Use this to explain why $\widetilde{X}$ is irreducible.

(c) Use $[\varphi]$ to prove that $K(\widetilde{X}) \cong K(X')$.

6.6.4 Prove Corollary 6.44.

6.6.5 Suppose that $f, g \in K[x, y, z]$ have at least one common zero. Prove that they have infinitely many common zeros.

6.6.6 Prove Corollary 6.45.

6.6.7 Prove that every affine variety over $\mathbb{R}$ can be realized as the vanishing of a single polynomial. In particular, over $\mathbb{R}$, the dimension of a variety has nothing to do with the number of defining equations.

6.6.8 Let X be an affine variety. Prove that the dimension of X is equal to the maximum d such that there exist prime ideals $P_0, \ldots, P_d \subseteq K[X]$ such that

$$P_0 \subsetneq P_1 \subsetneq \cdots \subsetneq P_d \subsetneq K[X].$$

(The supremum of lengths of chains of prime ideals in a ring is called the *Krull dimension* of the ring, and this exercise proves that the dimension of an affine variety is equal to the Krull dimension of its coordinate ring.)

Chapter 7

Tangent Spaces and Smoothness

LEARNING OBJECTIVES FOR CHAPTER 7

- Build intuition for linearizations and tangent spaces.
- Develop tools for computing linearizations and tangent spaces.
- Understand the intrinsic nature of tangent spaces.
- Explore the relationship between the dimension of a variety and the dimensions of its tangent spaces.
- Determine the smooth and singular points of affine varieties.

There is a very good chance that the pictures of affine varieties that we have depicted in this book have taken you back to your days of calculus. You probably recall computing derivatives and tangent spaces in your calculus class, and you may also remember that derivatives and tangent spaces provide a way to characterize where graphs have "singularites." For example, the fact that the derivative of the absolute value function $f(x) = |x|$ is undefined at $x = 0$ corresponds to the geometric observation that the graph has a "corner" over $x = 0$. Our goal in this chapter is to introduce the "calculus" of algebraic geometry; in particular, we aim to develop the notions of linearizations and tangent spaces and to use them to give a precise meaning of singular points of affine varieties.

We begin this chapter by defining tangent spaces in Section 7.1. Given an affine variety $X \subseteq \mathbb{A}^n$ and a point $a \in X$, the tangent space T_aX is a vector subspace of K^n whose elements can geometrically be viewed as vectors that are tangent to X at a. As we discuss in Section 7.2, viewing T_aX as an algebraic object (a vector space) is especially important because it allows us to interpret tangent spaces intrinsically in terms of the coordinate ring $K[X]$. In particular, this implies that isomorphic affine varieties have isomorphic tangent spaces at corresponding points. The dimension of the tangent space T_aX as a vector space is always bounded below by the dimension of X, as we prove in Section 7.3, and the special points at which the tangent space has larger dimension are the singular points of the variety. We close this chapter by discussing properties of smooth and singular points in Section 7.4.

Section 7.1 Linearizations and tangent spaces

Tangent lines and tangent planes play a starring role in most single- and multi-variable calculus classes. In these classes, one typically starts with a definition of the derivative in terms of limits, then uses the limit definition to derive the standard rules for differentiation. For example, one of the standard differentiation rules states that the derivative of x^k is kx^{k-1} for any $k \in \mathbb{N}$. Since limits do not make sense over a general ground field K, the starting point for our discussion of tangency in algebraic geometry is with the differentiation rules for polynomials.

We assume that the reader is familiar with the standard rules for differentiating multi-variable polynomials, and we extend these formulas to any ring R. In other words, if

$$f = a_m x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + a_0 \in R[x],$$

then the derivative of f with respect to x is defined by

$$\frac{\partial f}{\partial x} = ma_m x^{m-1} + (m-1)a_{m-1}x^{m-2} + \cdots + a_1 \in R[x],$$

where ma_m refers to adding a_m to itself m times. If $f \in \mathbb{R}[x_1, \ldots, x_n]$ is a multi-variable polynomial, then we can define a partial derivative $\frac{\partial f}{\partial x_i}$ for each variable x_i by differentiating f as an element of $R'[x_i]$, where

$$R' = R[x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n].$$

Using these partial derivatives, the key definition required for our development of tangent spaces is the following.

7.1 **DEFINITION** *Linearization of a polynomial at a point*

If $f \in K[x_1, \ldots, x_n]$ and $a = (a_1, \ldots, a_n) \in \mathbb{A}^n$, then the *linearization of f at a* is defined by

$$L_a f = f(a) + \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}(a) \cdot (x_i - a_i) \right) \in K[x_1, \ldots, x_n].$$

We note that $\frac{\partial f}{\partial x_i} \in K[x_1, \ldots, x_n]$ is a polynomial and $\frac{\partial f}{\partial x_i}(a) \in K$ denotes the evaluation of that polynomial at a, so $L_a f \in K[x_1, \ldots, x_n]$ is a linear polynomial. In multi-variable calculus (in other words, when $K = \mathbb{R}$), the linearization of f at a is introduced as the linear function that most closely approximates f near a, and it is sometimes called the *linear approximation of f at a*.

The linearization of an affine variety at a point is defined as the vanishing of the linearizations of all polynomials in the vanishing ideal.

7.2 **DEFINITION** *Linearization of an affine variety at a point*

Let $X \subseteq \mathbb{A}^n$ be an affine variety and $a \in X$. The *linearization of X at a* is the affine variety

$$L_a X = \mathcal{V}\big(\{L_a f \mid f \in \mathcal{I}(X)\}\big).$$

The reader is encouraged to check (Exercise 7.1.1) that, if $\mathcal{I}(X) = \langle f_1, \ldots, f_m \rangle$, then

$$L_a X = \mathcal{V}(L_a f_1, \ldots, L_a f_m).$$

Since every vanishing ideal is finitely-generated, this implies that the linearization can always be defined by a finite set of linear polynomials. We also note that the initial term $f_i(a)$ in the definition of each $L_a f_i$ vanishes, since $f_i \in \mathcal{I}(X)$ and $a \in X$.

Let us consider a few examples of linearizations.

7.3 EXAMPLE Linearizations of a parabola

Consider the parabola $X = \mathcal{V}(y - x^2) \subseteq \mathbb{A}^2$, for which $\mathcal{I}(X) = \langle y - x^2 \rangle$. To compute the linearization at the origin, we must compute the linearization of the generator $f = y - x^2$ at $a = (0,0)$. From the definition, we have

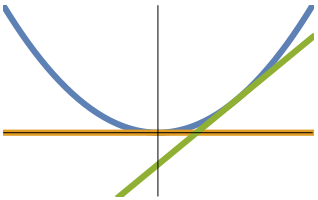
$$L_a f = 0(x-0) + 1(y-0) = y.$$

Thus, the linearization of the parabola at the origin is the x-axis: $L_a X = \mathcal{V}(y)$.

If we consider the point $b = (1,1)$, on the other hand, we compute

$$L_b f = -2(x-1) + 1(y-1) = -2x + y + 1.$$

From this, we see that the linearization of the parabola at $(1,1)$ is $L_b X = \mathcal{V}(-2x + y + 1)$.



We have depicted both of these linearizations in the image to the right, which is consistent with our intuition from calculus.

7.4 EXAMPLE Linearizations of a sphere

Consider the sphere $Y = \mathcal{V}(x^2 + y^2 + z^2 - 1) \subseteq \mathbb{A}^3_{\mathbb{C}}$, for which $\mathcal{I}(Y)$ is generated by $g = x^2 + y^2 + z^2 - 1$. To compute the linearization at $a = (0,0,-1) \in Y$, we start by computing the linearization of the generator:

$$L_a g = 0(x-0) + 0(y-0) + -2(z+1) = -2(z+1).$$

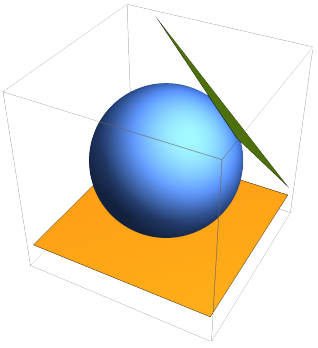
From this, we see that $L_a Y = \mathcal{V}(-2(z+1)) = \mathcal{V}(z+1)$, which is the plane containing a that is parallel to the xy-plane.

If, on the other hand, we consider the point $b = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}) \in Y$, then the linearization of the generator is

$$L_b g = \frac{2}{\sqrt{3}}(x + y + z - \sqrt{3}).$$

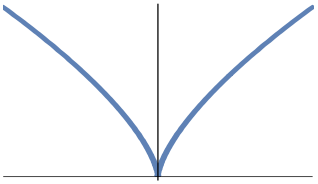
Thus, we have

$$L_b Y = \mathcal{V}\big(x + y + z - \sqrt{3}\big).$$



These linearizations are depicted over the real numbers in the image, and are again consistent with our intuition from calculus.

7.5 EXAMPLE Linearizations of a cusp



Consider the variety $X = \mathcal{V}(x^2 - y^3) \subseteq \mathbb{A}^2$, pictured to the right over $\mathbb{R}$. As you can see, this curve has a "cusp" at the origin, whereas it looks "smooth" at all other points. This geometric observation is reflected in the linearization: as the reader is encouraged to check in Exercise 7.1.3, the linearization of X at the origin is two-dimensional (all of $\mathbb{A}^2$), whereas the linearization at any other point is one-dimensional. We will see in Section 7.4 that the dimension of the linearization is a way of detecting the "singular" points in a variety such as the cusp in X.

We now use the notion of linearizations to define tangent spaces.

7.6 DEFINITION *Tangent vector and tangent space at a point*

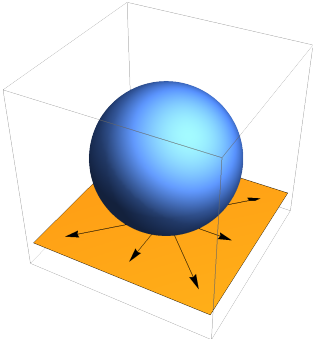
Let $X \subseteq \mathbb{A}^n$ be an affine variety and $a = (a_1, \ldots, a_n) \in X$. For any $b = (b_1, \ldots, b_n) \in L_a X$, the *tangent vector associated to* b is defined by

$$\overrightarrow{ab} = (b_1 - a_1, \ldots, b_n - a_n) \in K^n.$$

The *tangent space of* X *at* a is the collection of tangent vectors:

$$T_a X = \{\overrightarrow{ab} \mid b \in L_a X\} \subseteq K^n.$$

It is common to view the tangent vector $\overrightarrow{ab}$ geometrically as an arrow from a to b. Returning to the unit sphere from Example 7.4 and taking $a = (0, 0, -1)$, several examples of tangent vectors are drawn in the image to the right.



At this point the reader may (rightfully) be confused about the distinction between $L_a X$ and $T_a X$—these two objects feel very similar. In fact, the natural function

$$\begin{aligned} L_a X &\to T_a X \\ b &\mapsto \overrightarrow{ab} \end{aligned}$$

is a bijection that identifies $L_a X$ with $T_a X$ as sets. Why, then, do we choose to give these two similar objects different names and notation? The reason is that we view $L_a X$ as a geometric object—an affine variety consisting of points in $\mathbb{A}^n$—while we view $T_a X$ as an algebraic object, consisting of vectors in K^n. In fact, as we will see below, the set $T_a X$ is a vector subspace of K^n, and this key fact allows us to use standard tools from linear algebra to study tangent spaces.

To prove that the tangent space is a vector subspace of K^n, we first discuss a reinterpretion of $T_a X$ using gradient vectors.

7.7 DEFINITION *Gradient vector*

For any $a \in \mathbb{A}^n$ and $f \in K[x_1, \ldots, x_n]$ the *gradient of f at a* is the vector

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a)\right) \in K^n.$$

The next result characterizes tangent spaces in terms of gradient vectors, similarly to how tangent planes are computed in multivariable calculus.

7.8 PROPOSITION *Characterization of tangent spaces*

Let $X \subseteq \mathbb{A}^n$ be an affine variety and $a \in X$. Then

$$T_a X = \{\vec{v} \in K^n \mid \nabla f(a) \cdot \vec{v} = 0 \text{ for all } f \in \mathcal{I}(X)\}.$$

The "$\cdot$" appearing in Proposition 7.8 is the standard dot product.

PROOF Let $\vec{v} = (v_1, \ldots, v_n) \in K^n$. By the definition of $T_a X$, we see that $\vec{v} \in T_a X$ if and only if $\vec{v} = \overrightarrow{ab}$ for some $b \in L_a X$. Unwinding this, it is equivalent to the requirement that

$$b := (v_1 + a_1, \ldots, v_n + a_n) \in L_a X.$$

By the definition of $L_a X$, we have that $b \in L_a X$ if and only if, for all $f \in \mathcal{I}(X)$,

$$0 = L_a f(b) = \sum_{i=1}^{n} \left[\frac{\partial f}{\partial x_i}(a)\right](b_i - a_i) = \nabla f(a) \cdot \vec{v}.$$

Thus, $\vec{v} \in T_a X$ if and only if $\nabla f(a) \cdot \vec{v} = 0$ for all $f \in \mathcal{I}(X)$. □

As in the case of $L_a X$, it is straightforward to verify that, if we have a finite set of generators $\mathcal{I}(X) = \langle f_1, \ldots, f_m \rangle$, then the vanishing of Proposition 7.8 need only be checked on the generators:

$$T_a X = \{\vec{v} \in K^n \mid \nabla f_i(a) \cdot \vec{v} = 0 \text{ for all } i = 1, \ldots, m\}.$$

As a consequence of Proposition 7.8, we now prove that $T_a X$ is a vector space.

7.9 COROLLARY *The tangent space is a vector space*

Let $X \subseteq \mathbb{A}^n$ be an affine variety and $a \in X$. The tangent space $T_a X$ is a vector subspace of K^n.

PROOF Suppose $\vec{v}, \vec{w} \in T_a X$ and $r \in K$. To check that the tangent space is a vector subspace, we must check that $\vec{v} + r\vec{w} \in T_a X$. To accomplish this, Proposition 7.8 says that we must check that $\nabla f(a)(\vec{v} + r\vec{w}) = 0$ for all $f \in \mathcal{I}(X)$. Let $f \in \mathcal{I}(X)$. Then

$$\nabla f(a) \cdot (\vec{v} + r\vec{w}) = \nabla f(a) \cdot \vec{v} + r \nabla f(a) \cdot \vec{w} = 0$$

where the first equality follows from the linearity of the dot product and the second equality follows from Proposition 7.8 and the assumption that $\vec{v}, \vec{w} \in T_aX$. Therefore, by Proposition 7.8, we see that $\vec{v} + r\vec{w} \in T_aX$, concluding the proof. □

Starting with an affine variety $X \subseteq \mathbb{A}^n$ and a point $a \in X$, we have now defined a vector space $T_aX \subseteq K^n$. However, the definition of T_aX we have given is extrinsic: it depends heavily on the inclusion $X \subseteq \mathbb{A}^n$. In the next section, our aim is to give an intrinsic characterization of T_aX that depends only on the K-algebra $K[X]$ and the maximal ideal $I_a \subseteq K[X]$ comprised of polynomial functions on X that vanish at a.

Exercises for Section 7.1

7.1.1 Let $X \subseteq \mathbb{A}^n$ be an affine variety with $a \in X$ and $\mathcal{I}(X) = \langle f_1, \ldots, f_m \rangle$.

(a) Prove that
$$L_aX = \mathcal{V}(L_af_1, \ldots, L_af_m).$$

(b) Prove that
$$T_aX = \{\vec{v} \in K^n \mid \nabla f_i(a) \cdot \vec{v} = 0 \text{ for all } i = 1, \ldots, m\}$$

7.1.2 Give an example of an affine variety $X = \mathcal{V}(f) \subseteq \mathbb{A}^n$ such that
$$L_aX \neq \mathcal{V}(L_af).$$
(In other words, when computing linearizations, it does not suffice to use defining polynomials; rather, one requires generators for the vanishing ideal.)

7.1.3 Let $X = \mathcal{V}(x^2 - y^3) \subseteq \mathbb{A}^2$. Prove that the linearization at $(0,0)$ is two-dimensional and that the linearization at any other point is one-dimensional.

7.1.4 Let $X = \mathcal{V}(x^2 + y^2 - z) \subseteq \mathbb{A}^3$.

(a) Draw a picture of X over $\mathbb{R}$.

(b) Prove that the linearization at $(0,0,0)$ is three-dimensional and that the linearization at any other point of X is two-dimensional.

7.1.5 For any affine variety $X \subseteq \mathbb{A}^n$ and $a \in X$, prove that
$$\dim(T_aX) = \dim(L_aX),$$
where the left-hand side is the dimension as a vector space and the right-hand side is the dimension as an affine variety.

Section 7.2 Tangent spaces from coordinate rings

In this section, our aim is to reinterpret the tangent space T_aX purely in terms of the coordinate ring $K[X]$. One of the most important consequences of this intrinsic description of the tangent space is that it will imply that tangent spaces are preserved under isomorphism. In other words, if $F : X \to Y$ is an isomorphism and $a \in X$, then we can conclude that T_aX and $T_{F(a)}Y$ are isomorphic vector spaces.

In order to start our intrinsic description of the tangent space, we first require an intrinsic way of thinking about a point $a \in X$. Given $a \in X$, define

$$I_a = \{F \in K[X] \mid F(a) = 0\} \subseteq K[X].$$

From the definition, we see that I_a is a subset of $K[X]$. However, more is true: I_a is a maximal ideal (Exercise 7.2.1).

In the quotient description of $K[X]$, the definition of I_a can be made quite concrete. To see this, suppose that $X \subseteq \mathbb{A}^n$, so that $K[X] = K[x_1, \ldots, x_n]/\mathcal{I}(X)$. Let $a = (a_1, \ldots, a_n) \in X$. Then, given $F = [f] \in K[X]$ where $f \in K[x_1, \ldots, x_n]$, we have $F \in I_a$ if and only if $f(a) = 0$. Since every polynomial that vanishes at a can be written as

$$f = (x_1 - a_1)f_1 + \cdots + (x_n - a_n)f_n$$

for some polynomials $f_1, \ldots, f_n \in K[x_1, \ldots, x_n]$, we see that

$$I_a = \langle [x_1 - a_1], \ldots, [x_n - a_n] \rangle \subseteq \frac{K[x_1, \ldots, x_n]}{\mathcal{I}(X)}. \tag{7.1}$$

Our aim in this section is to prove the following result, which characterizes the tangent space T_aX in terms of $I_a \subseteq K[X]$.

7.10 THEOREM $I_a/I_a^2 = (T_aX)^\vee$

Let $X \subseteq \mathbb{A}^n$ be an affine variety and let $a \in X$. Then there is a canonical vector space isomorphism

$$I_a/I_a^2 = (T_aX)^\vee.$$

Before proving this theorem, we pause to parse the statement by recalling some important concepts from linear algebra.

First of all, if V is a vector space over K, then $V^\vee$ denotes the dual space: an element of $V^\vee$ is a linear map $\varphi : V \to K$. Since linear maps can be added and multiplied by scalars, the dual space is, itself, a vector space (of the same dimension as V, if V is finite-dimensional). Therefore, the right-hand side of the equality in Theorem 7.10 is the vector space of linear maps $\varphi : T_aX \to K$.

To parse the left-hand side of Theorem 7.10, first observe that I_a can naturally be viewed as a vector space: adding two functions that vanish at a results in a function that vanishes at a, and multiplying a function that vanishes at a by a scalar results in a function that vanishes at a. The ideal $I_a^2 = I_a \cdot I_a$ is the ideal product of I_a with itself. For example, using the generators for I_a in Equation (7.1), it follows that

$$I_a^2 = \langle [(x_i - a_i)(x_j - a_j)] \mid 1 \leq i, j \leq n \rangle \subseteq I_a. \tag{7.2}$$

Notice that I_a^2 is a vector subspace of I_a, so it makes sense to take a vector space quotient I_a/I_a^2. The reader is encouraged to convince themselves (Exercise 7.2.2) that, in terms of the generators in Equation (7.1), the vector space quotient is spanned by $[x_1 - a_1], \ldots, [x_n - a_n]$:

$$I_a/I_a^2 = K\{[x_1 - a_1], \ldots, [x_n - a_n]\}.$$

In particular, this implies that I_a/I_a^2 is a finite-dimensional vector space, and Theorem 7.10 is the assertion that there is a canonical isomorphism between I_a/I_a^2 and $(T_aX)^\vee$ as finite-dimensional vector spaces.

PROOF OF THEOREM 7.10 We define a canonical surjective linear map

$$\varphi : I_a \to (T_aX)^\vee$$

and prove that $\ker(\varphi) = I_a^2$. To define φ, suppose that $F \in I_a$ and write $F = [f]$ for some $f \in K[x_1, \ldots, x_n]$. We define $\varphi(F)$ to be the linear map

$$\begin{aligned} \varphi(F) : T_aX &\to K \\ \vec{v} &\mapsto \nabla f(a) \cdot \vec{v}. \end{aligned}$$

Of course, we should be worried that this definition depends on the choice of representative f. However, due to Proposition 7.8, it follows (Exercise 7.2.3) that

$$[f] = [g] \Longrightarrow \nabla f(a) \cdot \vec{v} = \nabla g(a) \cdot \vec{v} \quad \text{for all} \quad \vec{v} \in T_aX,$$

so $\varphi(F) \in (T_aX)^\vee$ is independent of the choice of representative f. To check that φ is linear, we require that $\varphi(F + rG) = \varphi(F) + r\varphi(G)$ for any $F, G \in K[X]$ and $r \in K$; this follows from linearity of derivatives (Exercise 7.2.4). Thus, it remains to prove that φ is surjective and that $\ker(\varphi) = I_a^2$.

To prove that φ is surjective, let $\rho : T_aX \to K$ be a linear map. Define

$$f = \rho(\vec{e}_1)(x_1 - a_1) + \cdots + \rho(\vec{e}_n)(x_n - a_n) \in K[x_1, \ldots, x_n],$$

where $\vec{e}_1, \ldots, \vec{e}_n$ are the standard basis vectors of K^n, and notice that $[f] \in I_a$. We claim that $\varphi([f]) = 0$. Unraveling the definitions, we see that $\nabla f(a) \cdot \vec{e}_i = \rho(\vec{e}_i)$ for all $i = 1, \ldots, n$. Since a linear map on $T_aX \subseteq K^n$ is uniquely determined by its values on the standard basis vectors, we have $\nabla f(a) \cdot \vec{v} = \rho(\vec{v})$ for all $\vec{v} \in T_aX$ and we conclude that $\varphi([f]) = \rho$. Thus, φ is surjective.

To prove that $\ker(\varphi) = I_a^2$, first suppose that $F \in I_a^2$. By Equation (7.2), it follows that $F = [f]$ where f has the form

$$f = \sum_{i,j=1}^{n} (x_i - a_i)(x_j - a_j) f_{i,j} \quad \text{for some} \quad f_{i,j} \in K[x_1, \ldots, x_n].$$

Using the product rule to compute partial derivatives of f, one can calculate that $\nabla f(a) = \vec{0}$, from which it follows that $F \in \ker(\varphi)$. Thus, $I_a^2 \subseteq \ker(\varphi)$.

To prove the other inclusion, suppose that $F \in \ker(\varphi)$ and write $F = [f]$ for some $f \in K[x_1, \ldots, x_n]$. The assumption that $F \in \ker(\varphi)$ means that

$$\nabla f(a) \cdot \vec{v} = 0 \quad \text{for all} \quad \vec{v} \in T_aX.$$

Choosing generators $\mathcal{I}(X) = \langle g_1, \ldots, g_m \rangle$, Proposition 7.8 and a standard result in linear algebra (Exercise 7.2.5) then imply that

$$\nabla f(a) = \sum_{i=1}^{m} a_i \nabla g_i(a)$$

for some values $a_1, \ldots, a_m \in K$. Since gradients act linearly on polynomials and since $\mathcal{I}(X)$ is closed under taking linear combinations, we then see that

$$\nabla f(a) = \nabla g(a) \quad \text{for some} \quad g = \sum_{i=1}^{m} a_i g_i \in \mathcal{I}(X).$$

Using that $F = [f]$ is in the domain of φ and thus $[f] \in I_a$, we can write

$$f = \sum_{i=1}^{n} b_i(x_i - a_i) + \sum_{i,j=1}^{n} f_{ij}(x_i - a_i)(x_j - a_j).$$

Writing a similar expression for g, the equality $\nabla f(a) = (b_1, \ldots, b_n) = \nabla g(a)$ implies that

$$f - g = \sum_{i,j=1}^{n} (f_{ij} - g_{ij})(x_i - a_i)(x_j - a_j).$$

Therefore, since $g \in \mathcal{I}(X)$, we conclude that $[f] = [f - g]$, from which it follows that

$$F = [f - g] \in \left\langle [(x_i - a_i)(x_j - a_j)] \mid 1 \leq i, j \leq n \right\rangle = I_a^2,$$

finishing the proof. □

A very important fact about duals of finite-dimensional vector spaces is that there is a canonical isomorphism

$$V = (V^\vee)^\vee \tag{7.3}$$

that takes a vector $\vec{v} \in V$ to the linear map $V^\vee \to K$ that sends $\varphi \in V^\vee$ to $\varphi(\vec{v})$. If the reader has never pondered this fact, we encourage them to take some moments to reflect on this isomorphism; in particular, it is a useful exercise to verify that it is both injective and surjective. Taking duals of both sides in Theorem 7.10 and applying (7.3), we arrive at the following result.

7.11 COROLLARY $T_a X = (I_a / I_a^2)^\vee$

Let $X \subseteq \mathbb{A}^n$ be an affine variety and let $a \in X$. Then there is a canonical vector space isomorphism

$$T_a X = (I_a / I_a^2)^\vee.$$

The importance of Corollary 7.11 is that it leads to a completely intrinsic interpretation of the tangent space. We spell this out carefully in the next result.

7.12 COROLLARY *Tangent spaces are intrinsic*

Let X and Y be affine varieties, $F : X \to Y$ an isomorphism, and $a \in X$. Then F induces a vector space isomorphism

$$T_a X \cong T_{F(a)} Y.$$

PROOF By Exercise 7.2.6, the pullback $F^* : K[Y] \to K[X]$ identifies the maximal ideals $I_{F(a)}$ and I_a:

$$F^*(I_{F(a)}) = I_a,$$

from which it also follows that

$$F^*(I_{F(a)}^2) = I_a^2.$$

Therefore, F^* induces a vector space isomorphism $I_{F(a)}/I_{F(a)}^2 \cong I_a/I_a^2$, and taking duals gives an isomorphism $T_a X \cong T_{F(a)} Y$. □

Exercises for Section 7.2

7.2.1 Let X be an affine variety and $a \in X$ a point. Prove that $I_a \subseteq K[X]$ is a maximal ideal.

7.2.2 Let $X \subseteq \mathbb{A}^n$ be an affine variety and $a \in X$. Prove that

$$I_a/I_a^2 = K\{[x_1 - a_1], \ldots, [x_n - a_n]\}.$$

7.2.3 Let $X \subseteq \mathbb{A}^n$ be an affine variety and $a \in X$. Prove that if $[f] = [g] \in K[X]$, then $\nabla f(a) \cdot \vec{v} = \nabla g(a) \cdot \vec{v}$ for all $\vec{v} \in T_a X$.

7.2.4 Let φ be defined as in the proof of Theorem 7.10. For any $F, G \in K[X]$ and $r \in K$, use linearity of derivatives to prove that

$$\varphi(F + rG) = \varphi(F) + r\varphi(G).$$

7.2.5 Let $\vec{v}_1, \ldots, \vec{v}_m \in K^n$ and let $V \subseteq K^n$ be the linear subspace defined by

$$V = \{\vec{v} \in K^n \mid \vec{v}_i \cdot \vec{v} = 0 \text{ for all } i = 1, \ldots, m\}.$$

Suppose that $\vec{w} \in K^n$ satisfies $\vec{w} \cdot \vec{v} = 0$ for all $\vec{v} \in V$. Prove that $\vec{w}$ is in the span of $\vec{v}_1, \ldots, \vec{v}_m$. (**Hint:** Consider the matrix M with rows $\vec{v}_1, \ldots, \vec{v}_m$. and the matrix M' obtained from M by appending the row $\vec{w}$. Use a rank-nullity argument to prove that $\mathrm{rk}(M) = \mathrm{rk}(M')$.)

7.2.6 Let $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$ be affine varieties, $F : X \to Y$ an isomorphism, and $a \in X$. Prove that $F^* : K[Y] \to K[X]$ identifies the maximal ideals $I_{F(a)}$ and I_a:

$$F^*(I_{F(a)}) = I_a.$$

Section 7.3 Tangent spaces and dimension

In Section 7.1, we saw several examples of linearizations and tangent spaces. One thing you may have observed is that the dimension of the tangent space was sometimes bigger than, but never smaller than, the dimension of the variety itself. That the dimension of an irreducible variety gives a lower bound for the dimension of the tangent space at any point on that variety is the main result of this section.

7.13 **PROPOSITION** *Lower bound on tangent space dimension*

Let $X \subseteq \mathbb{A}^n$ be an irreducible affine variety and $a \in X$. Then

$$\dim(T_aX) \geq \dim(X).$$

We note that the dimension appearing in the left-hand side of the inequality is the dimension of $T_aX \subseteq K^n$ as a vector space, while the dimension in the right-hand side is the dimension of $X \subseteq \mathbb{A}^n$ as an affine variety. If one prefers, they may interpret both sides of the inequality in Proposition 7.13 as dimensions of affine varieties in $\mathbb{A}^n$ by noting (Exercise 7.1.5) that $\dim(T_aX) = \dim(L_aX)$.

Before proving Proposition 7.13, we first prove a stronger biconditional result in the zero-dimensional setting.

7.14 **LEMMA** *Zero-dimensional tangent spaces*

If $X \subseteq \mathbb{A}^n$ is an irreducible affine variety and $a \in X$, then $\dim(X) = 0$ if and only if $\dim(T_aX) = 0$.

PROOF If $\dim(X) = 0$, then X consists of the single point a and it can be checked from the definitions that $T_aX = \{0\}$ (Exercise 7.3.1).

To prove the other direction, suppose that $\dim(T_aX) = 0$ and consider the maximal ideal $I_a \subseteq K[X]$ comprised of polynomial functions that vanish at a. Our aim is to show that I_a is the zero ideal, from which it follows that the zero ideal is maximal in $K[X]$, so $K[X]$ is a field, from which the Nullstellensatz implies that X is a single point.

Since a finite-dimensional vector space and its dual have the same dimension, Theorem 7.10 and the assumption that $\dim(T_aX) = 0$ imply that $I_a = I_a^2$. Choose generators

$$I_a = \langle F_1, \ldots, F_m \rangle.$$

Since $I_a = I_a^2$, we can view each F_i as an element of I_a^2, allowing us to find equations of the form

$$F_i = G_{i,1}F_1 + \cdots + G_{i,m}F_m,$$

where $G_{i,j} \in I_a$ for each $i, j = 1, \ldots, m$. Thus, we obtain a system of linear equations

$$(I - G) \cdot \begin{pmatrix} F_1 \\ \vdots \\ F_m \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

where G is the matrix with entries $G_{i,j}$. An application of Cramer's rule then implies that

$$\det(I - G)F_i = 0 \in K[X]$$

for all $i = 1, \ldots, m$, where I is the $m \times m$ identity matrix. Notice that $\det(I - G)$ is a polynomial function on X, and because $G_{i,j}(a) = 0$ for all i, j, we see that $\det(I - G)$ takes value 1 at a. Thus, $\det(I - G)$ is not the zero function on X, and because $K[X]$ is an integral domain (here, we use the assumption that X is irreducible), we conclude that $F_i = 0$ for all i. Thus, I_a is the zero ideal, concluding the proof. □

We now prove Proposition 7.13. The proof uses induction on $\dim(X)$, where the induction step (which requires Lemma 7.14) is accomplished by slicing both X and $L_a X$ with a hyperplane—a variety defined by a single linear equation—and applying the Fundamental Theorem of Dimension Theory.

PROOF OF PROPOSITION 7.13 Given an irreducible affine variety $X \subseteq \mathbb{A}^n$ and a point $a \in X$, we prove that $\dim(T_a X) \geq \dim(X)$ by induction on $\dim(X)$.

Base case: If $\dim(X) = 0$, then the fact that vector space dimensions are non-negative implies $\dim(T_a X) \geq \dim(X)$. (In fact, by Lemma 7.14, we know more: $\dim(T_a X) = \dim(X)$ in this case.)

Induction step: Let $X \subseteq \mathbb{A}^n$ be an irreducible affine variety of positive dimension, and suppose that the inequality of the proposition holds for all irreducible affine varieties of dimension $\dim(X) - 1$. Let $a = (a_1, \ldots, a_n) \in X$.

Since we have assumed that $\dim(X) > 0$, the "if" direction of Lemma 7.14 implies that $T_a X \supsetneq \{0\}$, which is equivalent to $L_a X \supsetneq \{a\}$. Thus, we can choose a point $b \in L_a X \setminus \{a\}$. The two points $a, b \in \mathbb{A}^n$ must differ in at least one coordinate; without loss of generality, assume that they differ in the first coordinate and define the hyperplane

$$H = \mathcal{V}(x_1 - a_1) \subseteq \mathbb{A}^n.$$

Set $Y = X \cap H$. Notice first that Y cannot be all of X. Indeed, if $Y = X$, then $X \subseteq H$, which would imply that $L_a X \subseteq H$ (Exercises 7.3.2 and 7.3.3). But H was chosen specifically so that $b \in L_a X$ but $b \notin H$, so $L_a X \not\subseteq H$. Thus, we indeed have $Y \neq X$. Moreover, since $a \in Y$, we conclude that

$$\emptyset \subsetneq Y \subsetneq X, \tag{7.15}$$

from which the Fundamental Theorem of Dimension Theory implies that every irreducible component of Y has dimension $\dim(X) - 1$.

Let Z be an irreducible component of Y that contains a. Since $Z \subseteq X$ and $Z \subseteq H$, we have that $L_a Z \subseteq L_a X$ (Exercise 7.3.2) and $L_a Z \subseteq H$ (Exercise 7.3.3). This implies that

$$L_a Z \subseteq L_a X \cap H \subsetneq L_a X, \tag{7.16}$$

from which it follows that $T_a Z \subsetneq T_a X$, so $\dim(T_a Z) < \dim(T_a X)$.

Putting everything together and using the induction hypothesis on Z, we have that

$$\dim(X) - 1 = \dim(Z) \leq \dim(T_a Z) \leq \dim(T_a X) - 1,$$

from which it follows that $\dim(X) \leq \dim(T_a X)$, completing the proof. □

While Proposition 7.13 gives a lower bound for the dimension of T_aX, there does not exist an upper bound. To illustrate this point, the next example describes a method for constructing affine curves containing a point at which the tangent space has arbitrarily high dimension.

7.17 EXAMPLE Small varieties with big tangent spaces

Fix $n \geq 1$ and consider the affine variety

$$X_n = \mathcal{V}(x_2^n - x_1^{n+1}, x_3^n - x_1^{n+2}, \ldots, x_n^n - x_1^{2n-1}) \subseteq \mathbb{A}^n.$$

Note that X_2 is the cusp of Example 7.5. It can be shown (Exercise 7.3.4) that X_n is a one-dimensional irreducible affine variety and that its points take the form

$$X_n = \{(b^n, b^{n+1}, \ldots, b^{2n-1}) \mid b \in \mathbb{A}^1\}.$$

Let $a = (0, \ldots, 0) \in X$; we argue that $L_aX_n = \mathbb{A}^n$, implying that $T_aX_n = K^n$.

Recall that $L_aX_n = \mathcal{V}(L_af \mid f \in \mathcal{I}(X))$. So consider a polynomial

$$f \in \mathcal{I}(X_n) \subseteq K[x_1, \ldots, x_n].$$

Since f vanishes at $a = (0, \ldots, 0) \in X_n$, it has a vanishing constant term and we can write

$$f = \sum_{i=1}^{n} c_i x_i + g$$

where $c_i \in K$ and the terms in g all have degree at least two. Evaluating at a point $(b^n, \ldots, b^{2n-1}) \in X_n$, we have

$$0 = f(b^n, \ldots, b^{2n-1}) = \sum_{i=1}^{n} c_i b^{n+i-1} + g(b^n, \ldots, b^{2n-1}).$$

Since this relation holds for all $b \in \mathbb{A}^1$, it follows that the associated polynomial is the zero polynomial:

$$0 = \sum_{i=1}^{n} c_i y^{n+i-1} + g(y^n, \ldots, y^{2n-1}) \in K[y].$$

Since the terms in $\sum_{i=1}^{n} c_i y^{n+i-1}$ all have distinct powers of y that are less than $2n$ whereas the terms in $g(y^n, \ldots, y^{2n-1})$ all have degree at least $2n$, it follows that $c_i = 0$ for all i. Therefore, the linear terms in f vanish, so $L_af = 0$. Since this holds for all $f \in \mathcal{I}(X_n)$, we conclude that $L_aX_n = \mathcal{V}(0) = \mathbb{A}^n$ and $T_aX_n = K^n$.

The Whitney Embedding Theorem states that any n-dimensional smooth manifold can be viewed as a submanifold of $\mathbb{R}^{2n-1}$.

The previous example has an interesting consequence about embedding affine varieties that can be contrasted with the Whitney Embedding Theorem in the study of smooth manifolds. Since the tangent space of X_n at the origin is n-dimensional, then any affine variety isomorphic to X_n will also have a tangent space that is n-dimensional, which implies that the curve X_n cannot be isomorphic to any affine variety in $\mathbb{A}^m$ with $m < n$. Thus, for any $m > 0$, this shows that there exist one-dimensional affine varieties that cannot be embedded in $\mathbb{A}^m$.

Exercises for Section 7.3

7.3.1 If $X = \{a\} \subseteq \mathbb{A}^n$ is a single point, prove that $T_aX = \{0\} \subseteq K^n$.

7.3.2 Let $X \subseteq Y \subseteq \mathbb{A}^n$ be affine varieties. For any $a \in X$, prove that

$$L_aX \subseteq L_aY.$$

Conclude that $T_aX \subseteq T_aY$.

7.3.3 Prove that, if $X \subseteq \mathbb{A}^n$ is a linear variety, then $L_aX = X$ for any $a \in X$.

7.3.4 Let X_n be the affine variety defined in Example 7.17.

(a) Prove that

$$X_n = \{(b^n, b^{n+1}, \ldots, b^{2n-1}) \mid b \in \mathbb{A}^1\}.$$

(b) Prove that $\dim(X_n) = 1$. (**Hint:** Prove that $\{[x_1]\}$ is a Noether basis.)

(c) Prove that X_n is irreducible. (**Hint:** Part (a) may be helpful.)

7.3.5 This exercise illustrates some strange behavior of tangent spaces that occurs over the real numbers. Consider the irreducible real affine variety

$$X = \mathcal{V}(y^2 - x(x+1)^2) \subseteq \mathbb{A}^2_{\mathbb{R}}.$$

(a) Prove that $(-1,0)$ is an isolated point of X. In other words, prove that $(-1,0) \in X$ and you can find a circular disk D of some positive radius centered at $(-1,0)$ such that $D \cap X = \{(-1,0)\}$.

(b) Given that $(-1,0)$ is an isolated point of X, you might expect that $\dim(T_{(-1,0)}X) = 0$. To the contrary, show that $\dim(T_{(-1,0)}X) = 2$.

(c) As we have seen in previous examples, the real solutions of polynomial equations do not typically see the whole picture; often one needs to look at the complex solutions. If we consider the complex variety

$$X_{\mathbb{C}} = \mathcal{V}(y^2 - x(x+1)^2) \subseteq \mathbb{A}^2_{\mathbb{C}},$$

what do you think is happening at the point $(-1,0) \in X_{\mathbb{C}}$ that helps explain your answer to part (b)?

Section 7.4 Smooth and singular points

One of the most important aspects of tangent spaces is that they detect when affine varieties have "kinks" or "cusps." For instance, we saw in Example 7.5 that the "cusp" point in the variety $\mathcal{V}(x^2 - y^3) \subseteq \mathbb{A}^2$ is special in that the tangent space at this point is two-dimensional, while the tangent space at all other points is only one-dimensional. Such special points in a variety where the dimension of the tangent space jumps are called singular points, as we make precise in the next definition.

7.18 DEFINITION *Smooth and singular points*

Let X be an irreducible affine variety and let $a \in X$. We say that X is *smooth at* a if $\dim(T_aX) = \dim(X)$; otherwise, we say that X is *singular at* a. We say that X is *smooth* if it is smooth at every point $a \in X$; otherwise, we say that X is *singular*.

By Proposition 7.13, singular points are points $a \in X$ for which

$$\dim(T_aX) > \dim(X).$$

One can also define smoothness of reducible varieties, but because different components can have different dimensions, the dimension of X should be replaced with its *local dimension at* a, which is the maximum dimension of all irreducible components that contain a. For simplicity, we restrict our focus to irreducible varieties throughout our discussion of smoothness.

7.19 EXAMPLE Affine space is smooth

Since $\mathcal{I}(\mathbb{A}^n) = \{0\} \subseteq K[x_1, \ldots, x_n]$, it follows that, for any $a \in \mathbb{A}^n$, we have

$$L_a\mathbb{A}^n = \mathcal{V}(L_a0) = \mathcal{V}(0) = \mathbb{A}^n,$$

which implies that $T_a\mathbb{A}^n = K^n$. Therefore,

$$\dim(T_a\mathbb{A}^n) = \dim(K^n) = n = \dim(\mathbb{A}^n),$$

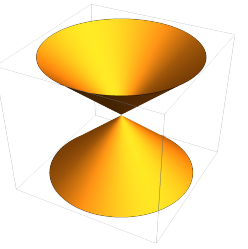
showing that $\mathbb{A}^n$ is smooth at all points.

7.20 EXAMPLE The cusp is singular at the origin

Consider the variety $X = \mathcal{V}(x^2 - y^3) \subseteq \mathbb{A}^2$ of Example 7.5. By Exercise 7.1.3, the tangent space is two-dimensional at $(0,0)$ and one-dimensional at all other points of X. Since X is itself a one-dimensional variety, this shows that X is singular at the origin but smooth at all other points.

7.21 EXAMPLE The cone is singular at the origin

Let $X = \mathcal{V}(x^2 + y^2 - z^2) \subseteq \mathbb{A}^3$, which is the cone whose real points are depicted to the right. It can be shown that X is singular at the origin, where the surface is pinched down to a point, but smooth at all other points (Exercise 7.4.4).



Given that the tangent spaces of an affine variety $X \subseteq \mathbb{A}^n$ are vector subspaces of K^n, one might rightfully expect that tools from linear algebra can be used to determine the singular points of a variety. The key object we require to import linear algebra tools into studying tangent spaces and singularities is the Jacobian matrix, which simply organizes the partial derivatives of a finite collection of polynomials.

7.22 **DEFINITION** *Jacobian matrix*

Given polynomials $f_1, \ldots, f_m \in K[x_1, \ldots, x_n]$, the *Jacobian matrix of* $f_1, \ldots, f_m$ is the $m \times n$ matrix

$$\mathrm{Jac}_{f_1,\ldots,f_m} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

Notice that the entries in the matrix $\mathrm{Jac}_{f_1,\ldots,f_m}$ are all elements of $K[x_1, \ldots, x_n]$, so it makes sense to evaluate them at any point $a \in \mathbb{A}^n$, obtaining an $m \times n$ matrix

$$\mathrm{Jac}_{f_1,\ldots,f_m}(a) \in K^{mn}.$$

The next result interprets the tangent space at a as the kernel of this matrix, which leads to a determinantal characterization of the singular points of a variety.

7.23 **PROPOSITION** *Jacobian criterion for smoothness*

If $X \subseteq \mathbb{A}^n$ is an irreducible affine variety with $\mathcal{I}(X) = \langle f_1, \ldots, f_m \rangle$, then

$$T_a X = \ker(\mathrm{Jac}_{f_1,\ldots,f_m}(a)).$$

Consequently, X is singular at $a \in X$ if and only if

$$\mathrm{rk}(\mathrm{Jac}_{f_1,\ldots,f_m}(a)) < \mathrm{codim}(X).$$

PROOF Notice that the rows of $\mathrm{Jac}_{f_1,\ldots,f_m}(a)$ are the gradients of the functions $f_1, \ldots, f_m$:

$$\mathrm{Jac}_{f_1,\ldots,f_m}(a) = \begin{pmatrix} \nabla f_1(a) \\ \vdots \\ \nabla f_m(a) \end{pmatrix}.$$

By definition of matrix multiplication, we have that $\vec{v} \in \ker(\mathrm{Jac}_{f_1,\ldots,f_m}(a))$ if and only if $\nabla f_i(a) \cdot \vec{v} = 0$ for all $i = 1, \ldots, m$. Therefore, Proposition 7.8 and the comments immediately following its proof imply that

$$T_a X = \ker(\mathrm{Jac}_{f_1,\ldots,f_m}(a)).$$

The Rank-Nullity Theorem then implies that

$$\dim(T_a X) = n - \mathrm{rk}(\mathrm{Jac}_{f_1,\ldots,f_m}(a)).$$

By Proposition 7.13, it follows that

$$\operatorname{rk}(\operatorname{Jac}_{f_1,\dots,f_m}(a)) \leq \operatorname{codim}(X),$$

and the definition of smooth and singular points then implies that X is singular at a if and only if equality fails. □

The Jacobian criterion is especially simple in the case of hypersurfaces.

7.24 **EXAMPLE** Jacobian criterion for hypersurfaces

Let $X = \mathcal{V}(f)$ where $f \in K[x_1,\dots,x_n]$ is irreducible, so that $\mathcal{I}(X) = \langle f \rangle$. Then $\operatorname{codim}(X) = 1$ and the Jacobian criterion says that X is singular at a if and only if the $1 \times n$ matrix

$$\begin{pmatrix} \dfrac{\partial f}{\partial x_1}(a) & \cdots & \dfrac{\partial f}{\partial x_n}(a) \end{pmatrix}$$

has rank zero. Thus, X is singular at a if and only if

$$\frac{\partial f}{\partial x_1}(a) = \cdots = \frac{\partial f}{\partial x_n}(a) = 0.$$

By Proposition 7.23, the singular points of a variety are characterized as the points where the Jacobian drops rank. A convenient attribute of this characterization is that the points at which the Jacobian drops rank are, themselves, the solutions of a system of polynomial equations. This leads to the following important consequence.

7.25 **COROLLARY** *Singular points are closed*

If $X \subseteq \mathbb{A}^n$ is an irreducible affine variety and $\operatorname{Sing}(X) \subseteq X$ is the set of singular points of X, then $\operatorname{Sing}(X)$ is also an affine variety.

Before presenting the proof, we first recall that an $r \times r$ *minor* of a matrix M is the determinant of an $r \times r$ matrix obtained by removing some subset of rows and columns from M, and an important linear algebra result states that $\operatorname{rk}(M) < r$ if and only if all $r \times r$ minors of M vanish.

PROOF OF COROLLARY 7.25 Choose a generating set $\mathcal{I}(X) = \langle f_1,\dots,f_m \rangle$ and, for simplicity, set $r = \operatorname{codim}(X)$. By Proposition 7.23, a point $a \in \mathbb{A}^n$ is a singular point of X if and only if $f_i(a) = 0$ for all $i = 1,\dots,m$ and

$$\operatorname{rk}(\operatorname{Jac}_{f_1,\dots,f_m}(a)) < r.$$

Thus, $\operatorname{Sing}(X)$ is the vanishing set of the polynomials $f_1,\dots,f_m \in K[x_1,\dots,x_n]$ together with the $r \times r$ minors of $\operatorname{Jac}_{f_1,\dots,f_m}$, each of which is a polynomial in $K[x_1,\dots,x_n]$. □

Knowing that $\operatorname{Sing}(X)$ is an affine variety allows us to study $\operatorname{Sing}(X)$ using all of the tools in our affine variety toolkit. For instance, the next example uses this fact to show that "almost all points" (in a dimension-theoretic sense) of an irreducible hypersurface are smooth.

7.26 **EXAMPLE** Generic smoothness of irreducible hypersurfaces

Let $X = \mathcal{V}(f) \subseteq \mathbb{A}^n$ where $f \in K[x_1, \ldots, x_n]$ is irreducible. We claim that X has at least one smooth point. To justify this, suppose to the contrary that every point of X is singular. By the Jacobian criterion, this implies that

$$\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \in K[x_1, \ldots, x_n]$$

all vanish on X, so they are elements of the vanishing ideal $\mathcal{I}(X) = \langle f \rangle$. Since f is irreducible, it cannot be constant; without loss of generality, suppose that f has positive degree in x_1. Then $\frac{\partial f}{\partial x_1}$ is not the zero polynomial and has degree strictly less than f, contradicting that it is an element of $\langle f \rangle$.

Thus, X has at least one smooth point, from which it follows that $\mathrm{Sing}(X) \subsetneq X$. Knowing that both $\mathrm{Sing}(X)$ and X are affine varieties, and that X is irreducible, it then follows from Proposition 6.35 that $\dim(\mathrm{Sing}(X)) < \dim(X)$. In other words, since the dimension of $\mathrm{Sing}(X)$ is strictly smaller than X, this says that the singular points are very special points of X: almost all points of X are smooth, a property that is often referred to as *generic smoothness*.

We note that generic smoothness holds for general affine varieties and follows from the fact that every affine variety has at least one smooth point. A proof of the latter fact is slightly beyond the scope of our current discussion.

Exercises for Section 7.4

7.4.1 Consider the affine variety $X = \mathcal{V}(y^2 - x^3 - x^2) \subseteq \mathbb{A}^2_{\mathbb{C}}$.

(a) Draw a picture of the real points of X and make a conjecture about where X might be singular.

(b) Use the Jacobian criterion to determine all points where X is singular.

7.4.2 Let $f \in K[x_1, \ldots, x_n]$ be a square-free polynomial and consider the affine variety $X = \mathcal{V}(x_{n+1}^2 - f) \in \mathbb{A}^{n+1}$.

(a) Prove that X is irreducible.

(b) Prove that X is smooth.

7.4.3 Prove that the complex unit sphere

$$X = \mathcal{V}(x_1^2 + \cdots + x_n^2 - 1) \subseteq \mathbb{A}^n_{\mathbb{C}}$$

is smooth.

7.4.4 Prove that the cone $\mathcal{V}(x^2 + y^2 - z^2) \subseteq \mathbb{A}^3$ is singular at the origin and smooth at all other points.

7.4.5 Consider the curve $X = \mathcal{V}(y - x^2, z - x^3) \subseteq \mathbb{A}^3$. Prove that X is smooth using two different methods:

(a) by proving that X is isomorphic to $\mathbb{A}^1$; and

(b) by applying the Jacobian criterion.

Chapter 8

Products

LEARNING OBJECTIVES FOR CHAPTER 8

- Prove that the product of affine varieties is an affine variety, and determine defining equations of $X \times Y$ from those of X and Y.
- Describe the elements of a tensor product of R-modules, and use the tensor product rules to detect when two elements are equal.
- Identify, in several examples, a tensor product of R-modules with a more familiar module.
- Prove that a tensor product of two polynomial rings is isomorphic as an algebra to a polynomial ring.
- Compute the coordinate ring of $X \times Y$ as a tensor product of the coordinate rings of X and Y.

A common way in which to build new mathematical objects from old is to take Cartesian products. The product of two topological spaces, for example, is a topological space when equipped with the product topology, and the product of two groups is a group when equipped with the componentwise operation. In this chapter, we seek to understand the Cartesian product of two affine varieties.

It takes a small bit of work to convince oneself that the product of affine varieties is itself an affine variety, but after carrying this out, the equivalence of algebra and geometry invites a natural question: what is the algebraic operation on coordinate rings that corresponds to the geometric operation of Cartesian product? More precisely, how is the K-algebra $K[X \times Y]$ related to $K[X]$ and $K[Y]$?

The answer to this question is that $K[X \times Y]$ is the *tensor product* of $K[X]$ and $K[Y]$; in the special case where $X = Y = \mathbb{A}^1$, this is the statement that $K[x,y]$ is the tensor product of $K[x]$ and $K[y]$. Because tensor products are likely to be unfamiliar—or if not unfamiliar, then intimidating—to many readers, we do not assume any prior knowledge of them. Instead, we start from the goal of defining an algebraic operation that combines $K[x]$ and $K[y]$ to produce $K[x,y]$, and we build up the definition of the tensor product slowly to suit that goal. After studying the definition and properties of tensor products, we will be prepared to prove that

$$K[X \times Y] = K[X] \otimes_K K[Y]$$

for any affine varieties X and Y. As an application, we prove in the final section of this chapter that the product of two smooth affine varieties of dimension d and e is, itself, a smooth affine variety of dimension $d + e$.

Section 8.1 The product of varieties is a variety

Given affine varieties $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$, the Cartesian product $X \times Y$ is

$$X \times Y := \{(a,b) \mid a \in X,\ b \in Y\} \subseteq \mathbb{A}^m \times \mathbb{A}^n.$$

On the other hand, $\mathbb{A}^m \times \mathbb{A}^n$ is canonically identified with $\mathbb{A}^{m+n}$, so we can view

$$X \times Y \subseteq \mathbb{A}^{m+n}.$$

In what follows, we use different variable sets for $\mathbb{A}^m$ and $\mathbb{A}^n$ to avoid confusing the two.

Our first goal is to prove that, in this affine space, $X \times Y$ is itself an affine variety.

8.1 **PROPOSITION** *The product of varieties is a variety*

The product of affine varieties is a variety. Specifically, if

$$X = \mathcal{V}(f_1,\dots,f_r) \subseteq \mathbb{A}^m \text{ and } Y = \mathcal{V}(g_1,\dots,g_s) \subseteq \mathbb{A}^n,$$

where $f_1,\dots,f_r \in K[x_1,\dots,x_m]$ and $g_1,\dots,g_s \in K[y_1,\dots,y_n]$, then

$$X \times Y = \mathcal{V}(f_1,\dots,f_r,g_1,\dots,g_s) \subseteq \mathbb{A}^{m+n},$$

in which we view both $f_1,\dots,f_r$ and $g_1,\dots,g_s$ as elements of the larger polynomial ring $K[x_1,\dots,x_m,y_1,\dots,y_n]$.

PROOF A point $(a,b) = (a_1,\dots,a_n,b_1,\dots,b_m) \in \mathbb{A}^{m+n}$ lies in $X \times Y$ if and only if $a \in X$ and $b \in Y$, which, by the definitions of X and Y as vanishing sets, is true if and only if

$$f_1(a) = \cdots = f_r(a) = 0$$

and

$$g_1(b) = \cdots = g_s(b) = 0.$$

If we now view $f_1,\dots,f_r$ as elements of $K[x_1,\dots,x_m,y_1,\dots,y_n]$ that happen to involve only the first m variables, then $f_i(a,b) = f_i(a)$. Similarly, $g_j(a,b) = g_j(b)$. Thus, $(a,b) \in X \times Y$ if and only if

$$f_1(a,b) = \cdots = f_r(a,b) = g_1(a,b) = \cdots = g_s(a,b) = 0,$$

which says precisely that $X \times Y = \mathcal{V}(f_1,\dots,f_r,g_1,\dots,g_s)$. □

8.2 **EXAMPLE** A parabola in $\mathbb{A}^3$

Let $X = \mathcal{V}(y - x^2) \subseteq \mathbb{A}^2$ and $Y = \mathcal{V}(z-1) \subseteq \mathbb{A}^1$. Then

$$X \times Y = \mathcal{V}(y - x^2, z - 1) \subseteq \mathbb{A}^3,$$

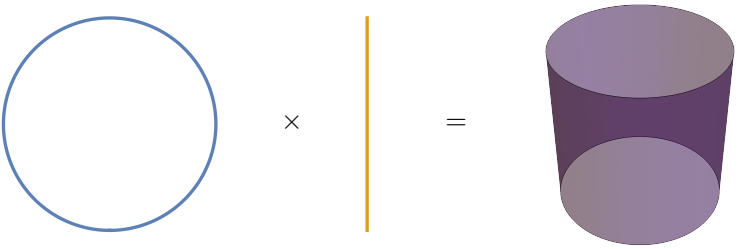
which is a parabola contained in the plane that is parallel to the xy-plane and at a height of 1.

8.3 **EXAMPLE** A cylinder

Let $X = \mathcal{V}(x^2 + y^2 - 1) \subseteq \mathbb{A}^2$ and $Y = \mathbb{A}^1$. Then

$$X \times Y = \mathcal{V}(x^2 + y^2 - 1) \subseteq \mathbb{A}^3,$$

which is vertical cylinder depicted below.



Passing from the world of geometry to the world of algebra, we are met with the next natural question, the one that motivates the rest of this chapter: can we compute the coordinate ring $K[X \times Y]$ in terms of the two coordinate rings $K[X]$ and $K[Y]$?

The best possible answer to this question would be to specify an algebraic operation $\star$ such that

$$K[X] \star K[Y] = K[X \times Y]$$

for all affine varieties X and Y. Whatever $\star$ might be, a special case would necessarily be that

$$K[x] \star K[y] = K[x, y],$$

coming from taking $X = Y = \mathbb{A}^1$ and therefore $X \times Y = \mathbb{A}^2$.

The reader might search her mind at this moment for any ways she currently knows to combine two K-algebras A and B to produce a new K-algebra $A \star B$. The first such operation that many students learn is the direct sum $A \oplus B$. This is not what we seek, though; the natural ring structure on $A \oplus B$ is componentwise,

$$\begin{aligned}(a_1, b_1) + (a_2, b_2) &= (a_1 + a_2, b_1 + b_2),\\ (a_1, b_1) \cdot (a_2, b_2) &= (a_1 a_2, b_1 b_2),\end{aligned}$$

and it is straightforward to show (Exercise 8.1.5) that this never produces an integral domain. Thus, $K[x, y]$ cannot be the same as $K[x] \oplus K[y]$. See Exercise 8.1.6 for another perspective on why the direct sum is not the right operation for this purpose.

The correct choice for the operation $\star$ turns out to be the *tensor product* of K-algebras. In the next three sections, we define tensor products of any modules and study their key properties before returning to geometry to prove that the tensor product indeed captures the coordinate ring of a product of affine varieties.

Exercises for Section 8.1

8.1.1 Let $X = \mathcal{V}(x^2 - 1) \subseteq \mathbb{A}^1$ and let $Y = \mathcal{V}(y - z) \subseteq \mathbb{A}^2$. Draw a picture over $\mathbb{R}$ of $X \times Y \subseteq \mathbb{A}^3$.

8.1.2 Draw a picture over $\mathbb{R}$ of the affine variety

$$Z = \mathcal{V}(x + z^2 - z, x) \subseteq \mathbb{A}^3 = \mathbb{A}^2 \times \mathbb{A}^1.$$

Is $Z = X \times Y$ for affine varieties $X \subseteq \mathbb{A}^2$ and $Y \subseteq \mathbb{A}^3$? If so, what are X and Y? If not, how can you tell?

8.1.3 Draw a picture over $\mathbb{R}$ of the affine variety

$$Z = \mathcal{V}(xy, yz, xz - x, z^2 - z) \subseteq \mathbb{A}^3 = \mathbb{A}^2 \times \mathbb{A}^1.$$

Is $Z = X \times Y$ for affine varieties $X \subseteq \mathbb{A}^2$ and $Y \subseteq \mathbb{A}^3$? If so, what are X and Y? If not, how can you tell?

8.1.4 Let $I \subseteq K[x_1, \ldots, x_n]$ and $J \subseteq K[y_1, \ldots, y_m]$ be nonzero ideals. Let

$$I' \subseteq K[x_1, \ldots, x_n, y_1, \ldots, y_m]$$

consist of the same elements as I, but viewed as polynomials in this larger variable set, and let $J' \subseteq K[x_1, \ldots, x_n, y_1, \ldots, y_m]$ be defined analogously.

(a) Prove that I' and J' are not ideals.

(b) Prove that $\mathcal{V}(I) \times \mathcal{V}(J) = \mathcal{V}(\langle I' \rangle + \langle J' \rangle)$, where $\langle I' \rangle$ and $\langle J' \rangle$ are the ideals in $K[x_1, \ldots, x_n, y_1, \ldots, y_m]$ generated by the sets I' and J'.

8.1.5 Prove that, if A and B are any rings besides $\{0\}$, then the ring $A \oplus B$ with componentwise addition and multiplication is not an integral domain.

8.1.6 This problem provides a further perspective on the difference between $K[x, y]$ and $K[x] \oplus K[y]$.

(a) Prove that, under the natural inclusions

$$K[x] \subseteq K[x, y] \text{ and } K[y] \subseteq K[x, y],$$

any element $h \in K[x, y]$ can be expressed as

$$h = f_1 g_1 + f_2 g_2 + \cdots + f_k g_k$$

for some $f_1, \ldots, f_k \in K[x]$ and $g_1, \ldots, g_k \in K[y]$.

(b) Prove that, under the natural inclusions

$$K[x] \subseteq K[x] \oplus K[y] \text{ and } K[y] \subseteq K[x] \oplus K[y],$$

any element $h \in K[x] \oplus K[y]$ can be expressed as

$$h = f + g$$

for some $f \in K[x]$ and $g \in K[y]$.

Section 8.2 Tensor products of modules

The goal of this section is to define the *tensor product* $M \otimes_R N$, a new R-module built from R-modules M and N. The model situation is when $M = R[x]$ and $N = R[y]$, in which case we will have

$$R[x] \otimes_R R[y] = R[x,y].$$

In order to understand how to generalize this example, we focus on one key property of the two-variable polynomial ring, which we have already mentioned in Exercise 8.1.6: every element $h \in R[x,y]$ can be expressed (non-uniquely) as a sum

$$h = f_1 g_1 + f_2 g_2 + \cdots + f_k g_k$$

for $f_1, \ldots, f_k \in R[x]$ and $g_1, \ldots, g_k \in R[y]$. For instance, $3x^2y + 2x + 4xy^2$ can be written as a sum of three terms

$$(3x^2) \cdot y + (2x) \cdot 1 + (4x) \cdot y^2.$$

Now let M and N be any R-modules. A "product" $f_i g_i$ in which $f_i \in M$ and $g_i \in N$ can be understood, formally, as an element of $M \times N$, so the first step toward defining $M \otimes_R N$ is to construct a module in which it makes sense to add such products together. This is a special case of the more general notion of *formal linear combinations*.

8.4 **DEFINITION** *Formal $\mathbb{Z}$-linear combination*

Let S be any set. A *formal $\mathbb{Z}$-linear combination* of elements of S is an expression of the form

$$\sum_{s \in S} a_s \cdot s,$$

where $a_s \in \mathbb{Z}$ for all $s \in S$ and $a_s = 0$ for all but finitely many choices of s. We consider two formal $\mathbb{Z}$-linear combinations equal if and only if all of their coefficients agree; that is,

$$\sum_{s \in S} a_s \cdot s = \sum_{s \in S} b_s \cdot s$$

if and only if $a_s = b_s$ for all $s \in S$.

We stress here that the elements of S in a formal linear combination are nothing but symbols that record the information of their coefficients, analogously to the way the variables $x_1, \ldots, x_n$ in a monomial $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ are nothing but symbols that record the information of their exponents $\alpha_1, \ldots, \alpha_n$. In particular, the data of a formal linear combination of elements of S is equivalent to the data of a function

$$f : S \to \mathbb{Z}$$
$$f(s) = a_s,$$

in which $f(s) = 0$ for all but finitely many $s \in S$. We only choose to express this information as a linear combination for notational convenience.

8.5 EXAMPLE Formal linear combinations of elements of a finite set

Let $S = \{\diamondsuit, \heartsuit, \clubsuit, \spadesuit\}$. Then

$$3\diamondsuit - 2\heartsuit + 1\clubsuit + 7\spadesuit$$

and

$$0\diamondsuit + 4\heartsuit - 30\clubsuit + 0\spadesuit$$

are examples of formal linear combinations of elements of S.

We have chosen a strange set S here to emphasize that the elements of S need not have any structure, nor any meaning beyond their role as placeholders for their coefficients.

8.6 EXAMPLE Formal linear combinations of elements of an infinite set

Let $S = \{x_1, x_2, x_3, \ldots\}$. Taking the convention of omitting a summand $a_s \cdot s$ when $a_s = 0$, formal linear combinations of elements of S are *finite* sums like

$$7x_2 + 4x_5 - 2x_6$$

or

$$3x_1 + 5x_{17} - 12x_{100} - 9x_{120}.$$

Even though S may not have any structure whatsoever, the set of all formal linear combinations of elements of S forms a group.

8.7 DEFINITION *Free abelian group*

Let S be any set. The set of all formal linear combinations of elements of S is called the *free abelian group on* S and is denoted $\mathbb{Z}S$. It is an abelian group under the operation

$$\sum_{s\in S} a_s \cdot s + \sum_{s\in S} b_s \cdot s = \sum_{s\in S} (a_s + b_s) \cdot s.$$

For instance, adding the formal linear combinations in Example 8.5 gives

$$\big(3\diamondsuit - 2\heartsuit + 1\clubsuit + 7\spadesuit\big) + \big(0\diamondsuit + 4\heartsuit - 30\clubsuit + 0\spadesuit\big) = 3\diamondsuit + 2\heartsuit - 29\clubsuit + 7\spadesuit,$$

essentially a process of "combining like terms." In fact, it is not difficult to show (Exercise 8.2.1) that

$$\mathbb{Z}\{\diamondsuit, \heartsuit, \clubsuit, \spadesuit\} \cong \mathbb{Z}^4,$$

via the isomorphism

$$a\diamondsuit + b\heartsuit + c\clubsuit + d\spadesuit \mapsto (a, b, c, d).$$

The reader can likely extrapolate from here that

$$\mathbb{Z}S \cong \mathbb{Z}^n$$

for any finite set S of cardinality n. Since further details on free abelian groups are not needed for our development of the tensor product, we leave the proof of this statement—as well as more on the case where S is infinite—to the exercises.

Returning to our goal of defining $M \otimes_R N$, let M and N again be any R-modules. We can now consider the free abelian group

$$\mathbb{Z}(M \times N) = \left\{ a_1 \cdot (m_1, n_1) + \cdots + a_k \cdot (m_k, n_k) \;\middle|\; \begin{smallmatrix} a_1,\ldots,a_k \in \mathbb{Z}, \\ m_1,\ldots,m_k \in M, \\ n_1,\ldots,n_k \in N \end{smallmatrix} \right\},$$

where we again take the convention of omitting terms from a formal linear combination if their coefficient is zero. For example, a typical element of $\mathbb{Z}(R[x] \times R[y])$ might be

$$3 \cdot (x^2, y) + 2 \cdot (x, 1) + 4 \cdot (x, y^2).$$

Replacing commas with multiplication, the above begins to look a lot like

$$3x^2y + 2x + 4xy^2 \in R[x, y],$$

so the reader may become hopeful that $\mathbb{Z}(R[x] \times R[y])$ is isomorphic to $R[x, y]$.

Alas, this is not yet the case. While there is a map from $\mathbb{Z}(R[x] \times R[y])$ to $R[x, y]$ defined by

$$a_1 \cdot (f_1, g_1) + \cdots + a_k \cdot (f_k, g_k) \mapsto a_1 f_1 g_1 + \cdots + a_k f_k g_k,$$

this map is not injective. In particular, the three formal linear combinations

$$\begin{aligned} &3 \cdot (x^2, y) + 2 \cdot (x, 1) + 4 \cdot (x, y^2), \\ &1 \cdot (3x^2, y) + 1 \cdot (2x, 1) + 1 \cdot (4x, y^2), \\ &1 \cdot (x^2, 3y) + 1 \cdot (x, 2) + 1 \cdot (x, 4y^2) \end{aligned}$$

are all different element of $\mathbb{Z}(R[x] \times R[y])$, but they all map to the same element $3x^2y + 2x + 4xy^2$ of $R[x, y]$. Another source of non-injectivity comes from elements like

$$1 \cdot (x^2, y) + 1 \cdot (x^2, 1) \quad \text{and} \quad 1 \cdot (x^2, y + 1),$$

which are different elements of $\mathbb{Z}(R[x] \times R[y])$ that map to the same element

$$x^2y + x^2 = x^2(y + 1)$$

of $R[x, y]$.

By taking a quotient of $\mathbb{Z}(R[x] \times R[y])$ by a certain subgroup, however, we can equate elements whose images in $R[x, y]$ are the same. This, at last, will give us the definition of the tensor product. We return to the general context of R-modules to state the definition.

8.8 **DEFINITION** *Tensor product of R-modules*

Let M and N be R-modules. Then the *tensor product* of M and N is

$$M \otimes_R N = \frac{\mathbb{Z}(M \times N)}{H},$$

where H is the subgroup of $\mathbb{Z}(M \times N)$ generated by all elements of the following three forms:

$$\begin{aligned}
&(m, n_1 + n_2) - (m, n_1) - (m, n_2), \text{ where } m \in M,\ n_1, n_2 \in N;\\
&(m_1 + m_2, n) - (m_1, n) - (m_2, n), \text{ where } m_1, m_2 \in M,\ n \in N;\\
&(rm, n) - (m, rn), \text{ where } r \in R,\ m \in M,\ n \in N.
\end{aligned}$$

The tensor product is an R-module, with addition defined from the addition in $\mathbb{Z}(M \times N)$ and scalar multiplication defined by

$$r \cdot [a_1(m_1, n_1) + \cdots + a_k(m_k, n_k)] = [a_1(rm_1, n_1) + \cdots + a_k(rm_k, n_k)].$$

We denote the coset in $M \otimes_R N$ of an element $a_1(m_1, n_1) + \cdots + a_k(m_k, n_k)$ of $\mathbb{Z}(M \times N)$ by

$$a_1 m_1 \otimes n_1 + \cdots + a_k m_k \otimes n_k. \tag{8.9}$$

The quotient by H ensures that, in $M \otimes_R N$, the following equations hold for all $m, m_1, m_2 \in M$, all $n, n_1, n_2 \in N$, and all $r \in R$:

$$\begin{aligned}
m \otimes (n_1 + n_2) &= m \otimes n_1 + m \otimes n_2\\
(m_1 + m_2) \otimes n &= m_1 \otimes n + m_2 \otimes n\\
(rm) \otimes n &= m \otimes (rn).
\end{aligned}$$

We sometimes refer to these equations as the "tensor product relations."

8.10 **EXAMPLE** Elements of $R[x] \otimes_R R[y]$

In the tensor product $R[x] \otimes_R R[y]$, an example of an element is

$$3x \otimes y + x \otimes y^2 + 2x^2 \otimes 1.$$

By the tensor product relations, this is the same as the element

$$x \otimes (3y) + x \otimes y^2 + 2x^2 \otimes 1,$$

and also the same as

$$x \otimes (3y + y^2) + 2x^2 \otimes 1,$$

as well as many others. Note that, if $\otimes$ is replaced by multiplication in $R[x, y]$, then the above two-variable polynomials are all equal; this is the motivation for defining the tensor product relations the way we do.

8.11 EXAMPLE Elements of $\mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_4$

Consider the tensor product of the two finite abelian groups ($\mathbb{Z}$-modules) $\mathbb{Z}_2$ and $\mathbb{Z}_4$. An example of an element in $\mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_4$ might be

$$1 \otimes 2 + 1 \otimes 3. \tag{8.12}$$

This can be re-written in various ways using the tensor product relations; for example, the first summand equals

$$1 \otimes (2 \cdot 1) = (2 \cdot 1) \otimes 1 = 0 \otimes 1$$

and the second summand equals

$$1 \otimes (3 \cdot 1) = (3 \cdot 1) \otimes 1 = 1 \otimes 1,$$

so the element in (8.12) is the same as

$$0 \otimes 1 + 1 \otimes 1 = (0 + 1) \otimes 1 = 1 \otimes 1.$$

In Example 8.14 of the next section, we give a complete description of the elements of $\mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_4$, but the motivated reader might try to describe them now.

Two comments are in order at this point. First, there is an apparent ambiguity in the notation (8.9): an expression like $2m \otimes n$ might mean

$$(2m) \otimes n,$$

i.e., the equivalence class of the element $(m + m, n) \in \mathbb{Z}(M \times N)$, or it might mean

$$2 \cdot m \otimes n,$$

i.e., the sum of the equivalence class of $(m, n) \in \mathbb{Z}(M \times N)$ with itself. The ambiguity is resolved by the tensor product relations, however, since they imply

$$(2m) \otimes n = (m + m) \otimes n = m \otimes n + m \otimes n = 2 \cdot m \otimes n.$$

As a result, we are justified in writing $2m \otimes n$ without any clarifying parentheses.

Second, an arbitrary element of $M \otimes_R N$ can be expressed as

$$m_1 \otimes n_1 + \cdots + m_k \otimes n_k,$$

in which $m_1, \ldots, m_k \in M$ and $n_1, \ldots, n_k \in N$. Indeed, by definition, an element of $M \otimes_R N$ has the form (8.9) for some $a_1, \ldots, a_k \in \mathbb{Z}$, but these coefficients can be absorbed into $m_1, \ldots, m_k$ by the argument in the previous paragraph. We sometimes refer to elements of $M \otimes_R N$ of the form $m \otimes n$ as *simple tensors*, so another way to say what we have just said is that every element of $M \otimes_R N$ is a sum of simple tensors.

Beware! A common mistake is to think that every element of $M \otimes_R N$ is of the form $m \otimes n$, which is false.

Equipped now with the definition of the tensor product, have we achieved our goal? In particular, is it true that

$$R[x] \otimes_R R[y] \cong R[x,y]?$$

We at least have a candidate for an isomorphism:

$$\varphi : R[x] \otimes_R R[y] \to R[x,y]$$
$$\varphi\Big(f_1(x) \otimes g_1(y) + \cdots + f_k(x) \otimes g_k(y)\Big) = f_1(x)g_1(y) + \cdots + f_k(x)g_k(y).$$

This matches with our intuition from earlier in the chapter; for example,

$$\varphi(3x^2 \otimes y + 2x \otimes 1 + 4x \otimes y^2) = 3x^2y + 2x + 4xy^2.$$

But since the domain of φ is a quotient, it is not at all clear that φ is well-defined. To prove this, we need to develop some further theory of tensor products.

Exercises for Section 8.2

8.2.1 Prove that the function

$$\varphi : \mathbb{Z}\{\diamondsuit, \heartsuit, \clubsuit, \spadesuit\} \to \mathbb{Z}^4$$
$$\varphi(a\diamondsuit + b\heartsuit + c\clubsuit + d\spadesuit) = (a,b,c,d)$$

is an isomorphism of groups.

8.2.2 This exercise shows that $\mathbb{Z}S$ is a free $\mathbb{Z}$-module for any set S.

(a) Suppose that S is a finite set with n elements. Extending Exercise 8.2.1, exhibit an isomorphism of groups $\mathbb{Z}S \cong \mathbb{Z}^n$.

(b) Now let S be any set, not necessarily finite. Prove that the elements

$$\{1 \cdot s \mid s \in S\} \subseteq \mathbb{Z}S$$

are linearly independent over $\mathbb{Z}$, and that $\mathbb{Z}S$ is generated by these elements.

8.2.3 Use the tensor product relations to express the element

$$x^2 \otimes y^3 + x \otimes (2y^3) \in \mathbb{R}[x] \otimes_{\mathbb{R}} \mathbb{R}[y]$$

in at least two other ways.

8.2.4 Let $a, b \geq 2$ be integers. Prove that every element of $\mathbb{Z}_a \otimes_{\mathbb{Z}} \mathbb{Z}_b$ is a simple tensor.

8.2.5 Carefully describe all of the elements of $\mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_5$. Which ones are equal to which other ones? What can you say about the group $\mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_5$?

Section 8.3 First properties of tensor products

We begin this section with a basic but useful property of tensor products, which the reader might expect given that our intuition has thus far been built from thinking of $\otimes$ as capturing multiplication in a two-variable polynomial ring.

8.13 PROPOSITION *Tensoring with zero*

Let M and N be R-modules. Then in $M \otimes_R N$, we have

$$m \otimes 0 = 0$$

for all $m \in M$ and

$$0 \otimes n = 0$$

for all $n \in N$. (Here, the "0" on the right-hand side of both equations denotes the additive identity in the module $M \otimes_R N$.)

PROOF By the tensor product relations,

$$m \otimes 0 = m \otimes (0+0) = m \otimes 0 + m \otimes 0.$$

Subtracting $m \otimes 0$ from both sides yields $m \otimes 0 = 0$. The proof of the second equation is similar. □

8.14 EXAMPLE $\mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_4 \cong \mathbb{Z}_2$

An arbitrary element of the tensor product $\mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_4$ is

$$m_1 \otimes n_1 + \cdots + m_k \otimes n_k,$$

where $m_1, \ldots, m_k \in \mathbb{Z}_2$ and $n_1, \ldots, n_k \in \mathbb{Z}_4$. By Proposition 8.13, any summands in which $m_i = 0$ vanish, so it suffices to assume that $m_i = 1$ for all i. But

$$1 \otimes n_1 + \cdots + 1 \otimes n_k = 1 \otimes (n_1 + \cdots + n_k),$$

so in fact, any element of $\mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_4$ is equal to $1 \otimes n$ for some $n \in \mathbb{Z}_4$. This leaves four possibilities:

$$1 \otimes 0, \ 1 \otimes 1, \ 1 \otimes 2, \text{ and } 1 \otimes 3.$$

The first and third of these are in fact equal, since $1 \otimes 0 = 0$ and also

$$1 \otimes 2 = 1 \otimes (2 \cdot 1) = (2 \cdot 1) \otimes 1 = 0 \otimes 1 = 0.$$

Similarly, the elements $1 \otimes 1$ and $1 \otimes 3$ are equal, because

$$1 \otimes 3 = 1 \otimes (3 \cdot 1) = (3 \cdot 1) \otimes 1 = 1 \otimes 1.$$

Thus, we have

$$\mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_4 = \{0, 1 \otimes 1\}.$$

Assuming these two elements are distinct—which will follow from Proposition 8.20 below—this shows that $\mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_4 \cong \mathbb{Z}_2$.

8.15 EXAMPLE $\mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_3 = \{0\}$

The same argument as the previous example shows that any element of $\mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_3$ is equal to $1 \otimes n$ for some $n \in \mathbb{Z}_3$, leaving three possibilities:

$$1 \otimes 0,\ 1 \otimes 1,\ \text{and}\ 1 \otimes 2.$$

In fact, however, all three of these are equal to $0 \in \mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_3$. To see this, note that

$$1 \otimes 1 = 1 \otimes (4 \cdot 1) = (4 \cdot 1) \otimes 1 = 0 \otimes 1 = 0$$

and

$$1 \otimes 2 = 1 \otimes (2 \cdot 1) = (2 \cdot 1) \otimes 1 = 0 \otimes 1 = 0.$$

Thus, $\mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_3 = \{0\}$.

8.16 EXAMPLE $\mathbb{Z}_n \otimes_{\mathbb{Z}} \mathbb{Q} = \{0\}$

Let n be any natural number, and consider the tensor product $\mathbb{Z}_n \otimes_{\mathbb{Z}} \mathbb{Q}$. An arbitrary element of this tensor product is

$$a_1 \otimes b_1 + \cdots + a_k \otimes b_k$$

with $a_1, \ldots, a_k \in \mathbb{Z}_n$ and $b_1, \ldots, b_k \in \mathbb{Q}$. For each of these summands, however, we have

$$a_i \otimes b_i = a_i \otimes \left(n \cdot \frac{b_i}{n} \right) = (n \cdot a_i) \otimes \frac{b_i}{n} = 0 \otimes \frac{b_i}{n} = 0,$$

so there are no nonzero elements in $\mathbb{Z}_n \otimes_{\mathbb{Z}} \mathbb{Q}$.

Proposition 8.13 is useful for showing that an element of a tensor product is zero, but—as we saw already in Example 8.14—proving that an element is *not* zero is another matter. For this, we need a way to detect whether an element of $\mathbb{Z}(M \times N)$ lies in the subgroup H given in Definition 8.8. In fact, this issue is really the heart of the matter in our development of tensor products. Recall that our overarching goal is to prove that

$$R[x] \otimes_R R[y] \cong R[x, y],$$

and while we have a candidate isomorphism $\varphi : R[x] \otimes_R R[y] \to R[x, y]$, we must prove that φ is well-defined. This amounts to showing that elements of $R[x] \otimes_R R[y]$ that are equal to zero—or in other words, elements of $\mathbb{Z}(R[x] \times R[y])$ that lie in H—are sent to zero by φ. Knowing what lies in H is the crux.

To resolve this issue, we briefly detour into the terminology of bilinearity.

8.17 **DEFINITION** *R-bilinear*

Let M, N, and L be R-modules. A function

$$\varphi : M \times N \to L$$

is *R-bilinear* if

- $\varphi(m_1 + m_2, n) = \varphi(m_1, n) + \varphi(m_2, n)$ for all $m_1, m_2 \in M,\ n \in N$;
- $\varphi(rm, n) = r\varphi(m, n)$ for all $r \in R,\ m \in M,\ n \in N$;
- $\varphi(m, n_1 + n_2) = \varphi(m, n_1) = \varphi(m, n_2)$ for all $m \in M,\ n_1, n_2 \in N$;
- $\varphi(m, rn) = r\varphi(m, n)$ for all $r \in R, m \in M,\ n \in N$.

In other words, if either the first input or the second input to an R-bilinear map φ is held fixed, then φ is a homomorphism of R-modules in the other input.

8.18 **EXAMPLE** Multiplication is bilinear

Considering $\mathbb{R}$ as an $\mathbb{R}$-module, the map

$$\begin{aligned} \varphi : \mathbb{R} \times \mathbb{R} &\to \mathbb{R} \\ \varphi(m, n) &= mn \end{aligned}$$

is $\mathbb{R}$-bilinear by the distributive, associative, and commutative properties of multiplication in $\mathbb{R}$. More generally, if A is an R-algebra, then the map

$$\begin{aligned} \varphi : A \times A &\to A \\ \varphi(m, n) &= mn \end{aligned}$$

is R-bilinear by the algebra axioms.

8.19 **EXAMPLE** Products of linear maps are bilinear

Considering $\mathbb{R}^2$, $\mathbb{R}^3$, and $\mathbb{R}$ as $\mathbb{R}$-modules, the map

$$\begin{aligned} \varphi : \mathbb{R}^2 \times \mathbb{R}^3 &\to \mathbb{R} \\ \varphi\big((x, y), (u, v, w)\big) &= (2x + 3y) \cdot (5u - v + w) \end{aligned}$$

is $\mathbb{R}$-bilinear. More generally, if M and N are R-modules and A is an R-algebra, and if $\varphi_1 : M \to A$ and $\varphi_2 : N \to A$ are homomorphisms of R-modules, then

$$\begin{aligned} \varphi : M \times N &\to A \\ \varphi(m, n) &= \varphi_1(m)\varphi_2(n) \end{aligned}$$

is R-bilinear (Exercise 8.3.2).

The definition of R-bilinearity likely reminds the reader of the tensor product relations, so it may come as no surprise that R-bilinearity is the key to determining which elements of $M \otimes_R N$ are zero. The following proposition makes this precise.

8.20 PROPOSITION *Defining property of tensor products*

Let M, N, and L be R-modules. If $\varphi : M \times N \to L$ is R-bilinear, then the function

$$\begin{aligned} \widehat{\varphi} &: M \otimes_R N \to L \\ \widehat{\varphi}(m_1 \otimes n_1 + \cdots + m_k \otimes n_k) &= \varphi(m_1, n_1) + \cdots + \varphi(m_k, n_k) \end{aligned}$$

is a well-defined homomorphism of R-modules.

Before proving the proposition, let us illustrate it in a few examples.

8.21 EXAMPLE Mapping $\mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_4$ to $\mathbb{Z}_2$

To define a homomorphism from $\mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_4$ to $\mathbb{Z}_2$, start with

$$\begin{aligned} \varphi : \mathbb{Z}_2 \times \mathbb{Z}_4 &\to \mathbb{Z}_2 \\ \varphi(a, b) &= a \cdot \pi(b), \end{aligned}$$

where $\pi : \mathbb{Z}_4 \to \mathbb{Z}_2$ is the map that reduces inputs modulo 2. The reader is encouraged to convince themselves directly of the bilinearity of φ, though it also follows from Example 8.19 since π is a $\mathbb{Z}$-module homomorphism. Given this bilinearity, φ induces

$$\begin{aligned} \widehat{\varphi} &: \mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_4 \to \mathbb{Z}_2 \\ \widehat{\varphi}(a_1 \otimes b_1 + \cdots + a_k \otimes b_k) &= a_1\pi(b_1) + \cdots + a_k\pi(b_k), \end{aligned}$$

which by Proposition 8.20 is a well-defined homomorphism of $\mathbb{Z}$-modules. The fact that $\widehat{\varphi}$ is well-defined is illustrated by the fact that equivalent elements of $\mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_4$ are mapped to the same element of $\mathbb{Z}_2$; for example, we saw in Example 8.14 that

$$1 \otimes 1 = 1 \otimes 3 \in \mathbb{Z}_2 \otimes_{\mathbb{Z}} \mathbb{Z}_4,$$

and the reader can readily verify that

$$\widehat{\varphi}(1 \otimes 1) = \widehat{\varphi}(1 \otimes 3) = 1 \in \mathbb{Z}_2.$$

Note, furthermore, that the fact that $\widehat{\varphi}(1 \otimes 1) \neq 0$ implies $1 \otimes 1 \neq 0$, resolving the issue raised in Example 8.14. In fact, since $\mathbb{Z}_2$ has only two elements, this proves that $\widehat{\varphi}$ is an isomorphism.

8.22 EXAMPLE Mapping $R[x] \otimes R[y]$ to $R[x, y]$

To consider an example that is especially relevant for us, let us define a homomorphism from $R[x] \otimes_R R[y]$ to $R[x, y]$ by starting with the following R-bilinear map:

$$\begin{aligned} \varphi : R[x] \times R[y] &\to R[x, y] \\ \varphi(f, g) &= f \cdot g. \end{aligned}$$

Again, the reader is encouraged to check the R-bilinearity of φ directly, but it is also a special case of Exampe 8.19. The induced homomorphism

$$\widehat{\varphi} : R[x] \otimes_R R[y] \to R[x,y]$$

given by Proposition 8.20 is defined by

$$\widehat{\varphi}(f_1 \otimes g_1 + \cdots + f_k \otimes g_k) = f_1 \cdot g_1 + \cdots + f_k \cdot g_k.$$

For example,

$$\widehat{\varphi}(x^2 \otimes y + x^2 \otimes 1) = x^2 y + x^2,$$

while

$$\widehat{\varphi}(x^2 \otimes (y+1)) = x^2(y+1).$$

The equality of these two images is consistent with the equality

$$x^2 \otimes y + x^2 \otimes 1 = x^2 \otimes (y+1) \in R[x] \otimes_R R[y].$$

and illustrates the fact that $\widehat{\varphi}$ is well-defined.

With these examples in mind, let us turn to the proof of the proposition in general.

PROOF OF PROPOSITION 8.20 First, extend the function $\varphi : M \times N \to L$ to a function $\widetilde{\varphi} : \mathbb{Z}(M \times N) \to L$ by defining it to act linearly on formal linear combinations. That is, define

$$\widetilde{\varphi} : \mathbb{Z}(M \times N) \to L$$
$$\widetilde{\varphi}\big(a_1(m_1,n_1) + \cdots + a_k(m_k,n_k)\big) = a_1 \cdot \varphi(m_1,n_1) + \cdots + a_k \cdot \varphi(m_k,n_k).$$

This function is automatically well-defined (since its domain has no relations), and some moment's reflecting should convince the reader that it is a homomorphism of additive groups.

The claim, now, is that $\widetilde{\varphi}(\alpha) = 0$ for all α in the subgroup H specified by Definition 8.8. If this is the case, then $\widetilde{\varphi}$ induces a well-defined homomorphism of additive groups

$$\frac{\mathbb{Z}(M \times N)}{H} \to L$$
$$[a_1(m_1,n_1) + \cdots + a_k(m_k,n_k)] \mapsto a_1 \cdot \varphi(m_1,n_1) + \cdots + a_k \cdot \varphi(m_k,n_k),$$

and this is precisely $\widehat{\varphi}$.

To prove the claim, let $\alpha \in H$. Then α is a sum of the generators of H listed in Definition 8.8, so since $\widetilde{\varphi}$ is a homomorphism of additive groups, it suffices to prove that $\widetilde{\varphi}$ sends each of these generators to zero. This is indeed the case; for example,

$$\widetilde{\varphi}\big((m, n_1+n_2) - (m,n_1) - (m,n_2)\big) = \varphi(m,n_1+n_2) - \varphi(m,n_1) - \varphi(m,n_2),$$

which equals zero by the definition of R-bilinearity. A similar argument applies to the second type of generator of H, whereas for the third type of generator, we have

$$\widetilde{\varphi}\big((rm,n) - (m,rn)\big) = \varphi(rm,n) - \varphi(m,rn) = r\varphi(m,n) - r\varphi(m,n) = 0,$$

again by the definition of R-bilinearity. It follows that $\widetilde{\varphi}$ sends every generator of H to zero, so $\widetilde{\varphi}(\alpha) = 0$.

We now know that $\widehat{\varphi}$ is a well-defined homomorphism of additive groups, so the only thing left to check is that it respects scalar multiplication. This is a straightforward unwinding of the definitions that we leave to Exercise 8.3.3. □

The stage is now set to prove that our candidate isomorphism from $R[x] \otimes_R R[y]$ to $R[x,y]$ is an isomorphism after all. We conclude this story in the next section.

Exercises for Section 8.3

8.3.1 Prove that the map

$$\varphi : R[x] \times R[y] \to R[x,y]$$
$$\varphi\big(f(x), g(y)\big) = f(x) \cdot g(y)$$

is R-bilinear.

8.3.2 Let M and N be R-modules, let A be an R-algebra, and let $\varphi_1 : M \to A$ and $\varphi_2 : N \to A$ be homomorphisms of R-modules. Prove that the map

$$\varphi : M \times N \to A$$
$$\varphi(m,n) = \varphi_1(m)\varphi_2(n)$$

is R-bilinear.

8.3.3 Prove that, if $\varphi : M \times N \to L$ is R-bilinear, then the function $\widehat{\varphi}$ of Proposition 8.20 respects scalar multiplication; that is,

$$\widehat{\varphi}(r \cdot (m_1 \otimes n_1 + \cdots + m_k \otimes n_k)) = r \cdot \widehat{\varphi}(m_1 \otimes n_1 + \cdots + m_k \otimes n_k)$$

for all $r \in R$, $m_1, \ldots, m_k \in M$, and $n_1, \ldots, n_k \in N$.

8.3.4 Prove that, for any natural numbers a and b,

$$\mathbb{Z}_a \otimes_{\mathbb{Z}} \mathbb{Z}_b \cong \mathbb{Z}_{\gcd(a,b)}.$$

(You may wish to cite the result of Exercise 8.2.4.)

8.3.5 Let V be a K-vector space with basis $\{e_i\}_{i \in I}$ and let W be a K-vector space with basis $\{f_j\}_{j \in J}$.

(a) Prove that the elements

$$\{e_i \otimes f_j \mid i \in I,\ j \in J\}$$

form a basis of $V \otimes_K W$.

(b) Assuming V and W are finite-dimensional, what does part (a) tell you about the relationship between $\dim(V \otimes_K W)$, $\dim(V)$, and $\dim(W)$?

(c) Use part (a) to argue that $K[x] \otimes_K K[y] \cong K[x,y]$ as K-vector spaces by giving a bijection between bases of each.

8.3.6 Let M and N be any R-modules. Prove that $M \otimes_R N \cong N \otimes_R M$.

8.3.7 Let M be any R-module. Prove that $M \otimes_R R \cong M$ as R-modules by showing that the function

$$\sigma : M \to M \otimes_R R$$
$$\sigma(m) = m \otimes 1$$

is an isomorphism.

8.3.8 Let M, M', and N be R-modules. Prove that

$$(M \oplus M') \otimes_R N \cong (M \otimes_R N) \oplus (M' \otimes_R N).$$

8.3.9 (a) To what familiar $\mathbb{R}$-vector space is $\mathbb{R}^n \otimes_{\mathbb{R}} \mathbb{C}$ isomorphic? Prove your answer.

(b) To what familiar $\mathbb{R}$-vector space is $\mathbb{R}[x] \otimes_{\mathbb{R}} \mathbb{C}$ isomorphic? Prove your answer.

8.3.10 Let R and S be rings with $R \subseteq S$, where we view S as an R-module in the usual way, and let M be an R-module. Prove that $M \otimes_R S$ (which, by construction, is an R-module) in fact has the structure of an S-module, where the scalar multiplication is given by

$$s \cdot (m_1 \otimes s_1 + \cdots + m_k \otimes s_k) := m_1 \otimes (ss_1) + \cdots + m_k \otimes (ss_k).$$

(The passage from M to $M \otimes_R S$ is called *extension of scalars* and is illustrated by the two examples in Exercise 8.3.9)

Section 8.4 Tensor products of algebras

Our goal in introducing tensor products was to construct an operation that combines $K[x]$ and $K[y]$ to produce $K[x,y]$, and we can now prove that $\otimes$ achieves this goal.

8.23 PROPOSITION $R[x] \otimes_R R[y] \cong R[x,y]$

The map

$$\begin{aligned} R[x] \otimes_R R[y] &\to R[x,y] \\ f_1 \otimes g_1 + \cdots + f_k \otimes g_k &\mapsto f_1 \cdot g_1 + \cdots + f_k \cdot g_k \end{aligned}$$

is a well-defined isomorphism of R-modules.

PROOF As we saw in Example 8.22, the map

$$\varphi : R[x] \times R[y] \to R[x,y]$$

given by $\varphi(f,g) = f \cdot g$ is R-bilinear. Thus, by Proposition 8.20, it induces a well-defined homomorphism of R-modules

$$\widehat{\varphi} : R[x] \otimes_R R[y] \to R[x,y],$$

which is precisely the map in the statement of Proposition 8.23. To prove that $\widehat{\varphi}$ is an isomorphism, note that any element of $R[x,y]$ can be written uniquely as

$$\sum_{i,j} r_{ij} x^i y^j,$$

where the sum is over all pairs of non-negative integers i,j and all but finitely many of the coefficients $r_{ij} \in R$ are zero. By the same token, any element of $R[x] \otimes_R R[y]$ can be expressed, using the tensor product relations, as

$$\sum_{i,j} (r_{ij} x^i) \otimes y^j;$$

see Exercise 8.4.1. Thus, we can define a function

$$\begin{aligned} \psi : R[x,y] &\to R[x] \otimes_R R[y] \\ \psi\left(\sum_{i,j} r_{ij} x^i y^j\right) &= \sum_{i,j} (r_{ij} x^i) \otimes y^j, \end{aligned}$$

which is a homomorphism of R-modules and inverse to $\widehat{\varphi}$. □

This argument is a restatement, in the more general context of free R-modules, of the argument for K-vector spaces in Exercise 8.3.5.

We have now proven that, when $X = Y = \mathbb{A}^1$, we have

$$K[X] \otimes_K K[Y] \cong K[X \times Y]$$

as K-modules. There is still one last piece of the puzzle missing, however: the right-hand side of the above isomorphism is a K-algebra, whereas the left-hand side is thus

far only a K-module. To upgrade our isomorphism to the level of K-algebras, we must explain how to give an algebra structure to the tensor product of two algebras.

8.24 **DEFINITION** *Tensor product of R-algebras*

Let A and B be R-algebras. Their *tensor product* is the R-module $A \otimes_R B$, equipped with the R-algebra structure given by setting

$$(m \otimes n)(m' \otimes n') = (mm') \otimes (nn')$$

and extending by linearity.

The stipulation that we "extend by linearity" is nothing but the standard distributivity of multiplication. For example,

$$(x \otimes y^2 + x^3 \otimes y)(1 \otimes y + 2x^2 \otimes y^4) = x \otimes y^3 + 2x^3 \otimes y^6 + x^3 \otimes y^2 + 2x^5 \otimes y^5$$

in $R[x] \otimes_R R[y]$.

Given that elements of $A \otimes_R B$ can be expressed in multiple ways, it is not immediately clear that the multiplication in Definition 8.24 is well-defined. For instance, we have

$$(5x) \otimes (y + y^2) = x \otimes 5y + x \otimes 5y^2,$$

but suppose we multiply both the left-hand side and the right-hand side by $x^2 \otimes y$; do we get the same answer? Indeed we do: on one hand, we get

$$\Big((5x) \otimes (y + y^2)\Big)\Big(x^2 \otimes y\Big) = (5x^3) \otimes (y^2 + y^3),$$

while on the other hand, we get

$$\Big(x \otimes 5y + x \otimes 5y^2\Big)\Big(x^2 \otimes y\Big) = x^3 \otimes 5y^2 + x^3 \otimes 5y^3,$$

and these two expressions are easily seen to be equal under the tensor product rules.

To check that the product in Definition 8.24 is well-defined in general, it is helpful to use what we already know about constructing well-defined maps out of tensor products. Toward this end, fix $\beta \in A \otimes_R B$ and define a function

$$\mu : A \times B \to A \otimes_R B$$

by sending $(m, n) \in A \times B$ to the product of $m \otimes n$ with β:

$$\mu(m, n) = (m \otimes n)\beta.$$

It is straightforward to check that μ is R-bilinear; for example,

$$\begin{aligned} \mu(m_1 + m_2, n) &= \big((m_1 + m_2) \otimes n\big)\beta \\ &= \big(m_1 \otimes n + m_2 \otimes n\big)\beta \\ &= (m_1 \otimes n)\beta + (m_2 \otimes n)\beta \\ &= \mu(m_1, n) + \mu(m_2, n). \end{aligned}$$

It follows that μ induces a well-defined map

$$A \otimes_R B \to A \otimes_R B,$$

and this is precisely the map $\alpha \mapsto \alpha\beta$. Thus, multiplication by any $\beta \in A \otimes_R B$ is indeed well-defined.

One should also verify that this multiplication makes $A \otimes_R B$ into an R-algebra, meaning that it satisfies the axioms of both an R-module and a ring and that these structures are compatible via the equation

$$r \cdot (\alpha\beta) = (r \cdot \alpha)\beta = \alpha(r \cdot \beta) \tag{8.25}$$

for all $r \in R$ and all $\alpha, \beta \in A \otimes_R B$. The fact that $A \otimes_R B$ is an R-module follows from the results of Section 8.2, while the fact that it is a ring follows from associativity (which, in turn, follows from associativity of the product in A and B) and distributivity (which holds by definition, since we have defined the product by extending linearly). The compatibility (8.25) is a consequence of the tensor product relations; we leave the details to Exercise 8.4.2.

Equipped with this algebra structure, a stronger version of Proposition 8.23 is now available.

8.26 **PROPOSITION** *$R[x] \otimes_R R[y] \cong R[x,y]$ as algebras*

The map

$$\begin{aligned} R[x] \otimes_R R[y] &\to R[x,y] \\ f_1 \otimes g_1 + \cdots + f_k \otimes g_k &\mapsto f_1 \cdot g_1 + \cdots + f_k \cdot g_k \end{aligned}$$

is an isomorphism R-algebras.

PROOF Proposition 8.23 shows that this map is an isomorphism of R-modules, and by Exercise 8.4.3, it respects multiplication. □

We can now at last conclude that

$$K[\mathbb{A}^1] \otimes_K K[\mathbb{A}^1] \cong K[\mathbb{A}^1 \times \mathbb{A}^1]$$

as K-algebras, so it is meaningful to ask: given any affine varieties X and Y, do we have an isomorphism of K-algebras

$$K[X] \otimes_K K[Y] \cong K[X \times Y]?$$

The goal of the next section is to prove that this is indeed the case.

Exercises for Section 8.4

8.4.1 Prove that any element of $R[x] \otimes_R R[y]$ can be expressed, via the tensor product relations, as

$$\sum_{i,j} (r_{ij} x^i) \otimes y^j,$$

in which $r_{ij} \in R$.

8.4.2 Prove that

$$r \cdot (\alpha\beta) = (r \cdot \alpha)\beta = \alpha(r \cdot \beta)$$

for all $r \in R$ and all $\alpha, \beta \in R[x] \otimes_R R[y]$.

8.4.3 Prove that the map $\widehat{\varphi}$ of Proposition 8.23 satisfies

$$\widehat{\varphi}(\alpha\beta) = \widehat{\varphi}(\alpha)\widehat{\varphi}(\beta)$$

for all $\alpha, \beta \in R[x] \otimes_R R[y]$.

8.4.4 Generalize the argument of Proposition 8.26 to prove that

$$R[x_1, \ldots, x_n] \otimes_R R[y_1, \ldots, y_m] \cong R[x_1, \ldots, x_n, y_1, \ldots, y_m]$$

as R-algebras.

8.4.5 Prove that, for any $a, b \in K$, there is an isomorphism of K-algebras

$$\frac{K[x]}{\langle x - a \rangle} \otimes_K \frac{K[y]}{\langle y - b \rangle} \cong \frac{K[x, y]}{\langle x - a, y - b \rangle}.$$

To what more familiar K-algebra are both sides isomorphic?

8.4.6 Let R and S be rings with $R \subseteq S$.

(a) Viewing S as an R-module in the usual way, prove that there is an isomorphism of R-algebras

$$R[x] \otimes_R S \cong S[x].$$

(b) Prove that the isomorphism of part (a) also holds as S-algebras, where the left-hand side is viewed as an S-algebra via Exercise 8.3.10.

Section 8.5 The coordinate ring of a product

We proved in Section 8.1 that the product of affine varieties is a variety, and we are now ready to compute the coordinate ring of such a product in terms of the coordinate ring of the two factors.

8.27 **THEOREM** *The coordinate ring of a product*

For any affine varieties X and Y, we have

$$K[X \times Y] = K[X] \otimes_K K[Y]$$

as K-algebras.

Before we prove this theorem in general, let us refocus on the goal by returning to a specific example.

8.28 **EXAMPLE** The coordinate ring of a parabola in $\mathbb{A}^3$

As in Example 8.2, let $X = \mathcal{V}(y - x^2) \subseteq \mathbb{A}^2$ and let $Y = \mathcal{V}(z-1) \subseteq \mathbb{A}^1$, so that

$$X \times Y = \mathcal{V}(y - x^2, z - 1) \subseteq \mathbb{A}^3,$$

a parabola in the $z = 1$ plane of $\mathbb{A}^3$. One can prove (directly, or via the Nullstellensatz in the case where K is algebraically closed) that

$$K[X \times Y] = \frac{K[x,y,z]}{\langle y - x^2, z - 1\rangle}.$$

The reader is encouraged to verify that this quotient is isomorphic as a K-algebra to $K[x]$. On the other hand, the ideals $\langle y - x^2\rangle \subseteq K[x,y]$ and $\langle z-1\rangle \subseteq K[z]$ are also radical, so

$$K[X] = \frac{K[x,y]}{\langle y - x^2\rangle} \cong K[x]$$

and

$$K[Y] = \frac{K[z]}{\langle z-1\rangle} \cong K.$$

Thus, by the result of Exercise 8.3.7, we have

$$K[X] \otimes_K K[Y] \cong K[x] \otimes_K K \cong K[x],$$

so indeed, $K[X \times Y]$ and $K[X] \otimes_K K[Y]$ are isomorphic.

PROOF OF THEOREM 8.27 Let $X \subseteq \mathbb{A}^n$ and $Y \subseteq \mathbb{A}^m$. To define the isomorphism of Theorem 8.27, consider the function

$$\begin{aligned} \varphi : K[X] \times K[Y] &\to K[X \times Y] \\ \varphi(F,G) &= F \times G, \end{aligned}$$

where $F \times G$ is the function on $X \times Y$ given by

$$(F \times G)(a,b) = F(a)G(b).$$

The reader can readily convince themselves that, if F and G are polynomial functions, then $F \times G$ is indeed also a polynomial function. Furthermore, φ is K-bilinear, so it induces a homomorphism of K-modules

$$\hat{\varphi} : K[X] \otimes_K K[Y] \to K[X \times Y].$$

To know that $\hat{\varphi}$ is a homomorphism of K-algebras, we should also check that it respects multiplication. This follows from the observation that, for $F, F' \in K[X]$ and $G, G' \in K[Y]$, one has

$$\big((FF') \times (GG')\big) = \big(F \times G\big)\big(F' \times G'\big) \in K[X \times Y]$$

(Exercise 8.5.3). This says that

$$\hat{\varphi}\big((F \otimes G)(F' \otimes G')\big) = \hat{\varphi}(F \otimes G)\hat{\varphi}(F' \otimes G'), \tag{8.29}$$

so $\hat{\varphi}$ respects multiplication of simple tensors. More generally, the fact that $\hat{\varphi}(\alpha\beta) = \hat{\varphi}(\alpha)\hat{\varphi}(\beta)$ for any α and β in $K[X] \otimes_K K[Y]$ follows from (8.29) by distributivity.

Recall that a simple tensor is an element of $M \otimes_R N$ of the form $m \otimes n$, and any element of $M \otimes_R N$ is a sum of simple tensors.

Thus, $\hat{\varphi}$ is a homomorphism of K-algebras, and it remains to prove that it is a bijection.

(**Surjectivity**) Choose any $H \in K[X \times Y]$. Then H is the restriction to $X \times Y$ of a polynomial function $h \in K[x_1, \ldots, x_n, y_1, \ldots, y_m]$. By breaking h into a sum of monomials, one sees that

$$h(x,y) = \sum_{i=1}^{\ell} f_i(x) g_i(y)$$

for polynomials $f_1, \ldots, f_\ell \in K[x_1, \ldots, x_n]$ and $g_1, \ldots, g_\ell \in K[y_1, \ldots, y_n]$. Let $F_1, \ldots, F_\ell \in K[X]$ be the polynomial functions defined by restricting $f_1, \ldots, f_\ell$, and similarly $G_1, \ldots, G_\ell \in K[Y]$. Then

$$\hat{\varphi}\left(\sum_{i=1}^{\ell} F_i \otimes G_i\right) = H,$$

as one sees by evaluating both sides on an arbitrary point $(a,b) \in X \times Y$.

(**Injectivity**) Suppose that

$$H = \sum_{i=1}^{\ell} F_i \otimes G_i \in K[X] \otimes_K K[Y]$$

and $\hat{\varphi}(H) = 0$. Using the tensor product relations, one can always ensure that $G_1, \ldots, G_\ell$ are linearly independent over K; the general proof of this statement is Exercise 8.5.4, but as an illustrative example, suppose that

$$H = F_1 \otimes G_1 + F_2 \otimes G_2 + F_3 \otimes G_3$$

in which we have a linear dependence like $G_3 = G_1 + 2G_2$. Then the reader can easily check that

$$H = (F_1 + F_3) \otimes G_1 + (F_2 + 2F_3) \otimes G_2.$$

If G_1 and G_2 are linearly independent, then we are done, whereas if they are not, the process can be repeated to combine the two summands of H into one.

Assume, then, that $G_1, \ldots, G_\ell$ are linearly independent over K. Our goal is to prove that the fact that $\hat{\varphi}(H) = 0$ implies $H = 0$, and to do this, it suffices to show that $F_i = 0$ for all i. Toward a contradiction, suppose that $F_i \neq 0$ for some i; without loss of generality, we may assume that $F_1 \neq 0$. This means that $F_1(a) \neq 0$ for some $a \in X$.

Fix such a point $a \in X$, and consider the polynomial function

$$\sum_{i=1}^{\ell} F_i(a) G_i \in K[Y].$$

This function sends any $b \in Y$ to

$$\sum_{i=1}^{\ell} F_i(a) G_i(b) = \hat{\varphi}(H)(a,b) = 0,$$

where the second equality follows from our assumption that $\hat{\varphi}(H) = 0$. Thus,

$$\sum_{i=1}^{\ell} F_i(a) G_i = 0 \in K[Y].$$

This is a nontrivial linear dependence among $G_1, \ldots, G_\ell$, since the coefficient $F_1(a)$ is nonzero, so we have arrived at a contradiction. It follows that $F_i = 0$ for all i and hence $H = 0$, completing the proof that $\hat{\varphi}$ is an isomorphism. □

8.30 **EXAMPLE** Polynomial functions on a parabola in $\mathbb{A}^3$

Returning once more to the variety $X \times Y$ of Example 8.28, an example of a polynomial function $H \in K[X \times Y]$ might be

$$H : X \times Y \to K$$
$$H(a,b,c) = ac + b^2,$$

which is the restriction of the polynomial $h(x,y,z) = xz + y^2$. If $F_1, F_2 \in K[X]$ are the restrictions of the polynomials $f_1(x,y) = x$ and $f_2(x,y) = y^2$, respectively, and $G_1, G_2 \in K[Y]$ are the restrictions of the polynomials $g_1(z) = z$ and $g_2(z) = 1$, respectively, then we have

$$\hat{\varphi}(F_1 \otimes G_1 + F_2 \otimes G_2) = H.$$

Of course, recalling that $X \times Y = \mathcal{V}(y - x^2, z - 1)$, we could equally well express H as

$$H(a,b,c) = a + a^4,$$

in which case we would see that

$$\hat{\varphi}((F_1 + F_1^4) \otimes 1) = H.$$

This corresponds to the fact that

$$F_1 \otimes G_1 + F_2 \otimes G_2 = (F_1 + F_1^4) \otimes 1 \in K[X] \otimes_K K[Y],$$

as one can verify from the defining equations of X and Y together with the tensor product rules.

In the next, and final, section of the chapter, we use Theorem 8.27 to prove various properties of products pertaining to irreducibility, dimension, and smoothness.

Exercises for Section 8.5

8.5.1 What is the coordinate ring of the cylinder $\mathcal{V}(x^2 + y^2 - 1) \subseteq \mathbb{A}^3$? Express your answer both as a tensor product and as a quotient of $K[x, y, z]$.

8.5.2 Let $X \subseteq \mathbb{A}^n$ be any affine variety, and let $Y \subseteq \mathbb{A}^m$ be a finite set consisting of r points.

(a) Describe $X \times Y \subseteq \mathbb{A}^n \times \mathbb{A}^m$ geometrically.

(b) Prove that $K[Y] \cong K^r$, and deduce—using the results of Exercises 8.3.8 and 8.3.7—that $K[X \times Y]$ is isomorphic to the direct sum of r copies of $K[X]$.

(c) Explain the relationship between the geometric statement in part (a) and the algebraic statement in part (b).

8.5.3 Let X and Y be affine varieties, and let $F, F' \in K[X]$ and $G, G' \in K[Y]$. Prove that

$$\big((FF') \times (GG')\big) = \big(F \times G\big)\big(F' \times G'\big)$$

in $K[X \times Y]$.

8.5.4 Let M and N be modules over a field R. Prove that any element of $M \otimes_R N$ can be expressed as

$$m_1 \otimes n_1 + \cdots + m_\ell \otimes n_\ell$$

in which $n_1, \ldots, n_\ell$ are linearly independent over R. (**Hint**: Any element of $M \otimes_R N$ can be represented as a sum of simple tensors; choose such a representation with the minimum number of summands.)

8.5.5 Let X and Y be affine varieties, and let $p : X \times Y \to X$ be the projection map $p(a, b) = a$. Since p is surjective, $p^* : K[X] \to K[X \times Y]$ is an injective homomorphism of K-algebras by Exercise 4.2.9. Describe p^* explicitly as a homomorphism

$$K[X] \to K[X] \otimes_K K[Y].$$

Section 8.6 Attributes of products

In this section, we study attributes, such as dimension, irreducibility, and smoothness, of products of affine varieties. We begin with a discussion of dimension. One should expect that $\dim(X \times Y) = \dim(X) + \dim(Y)$, either by analogy to the dimension of vector spaces or by viewing dimension intuitively as the number of independent directions in which a point can move. The first result of this section says that this expectation is indeed satisfied.

8.31 **PROPOSITION** *The dimension of a product*

If X and Y are nonempty affine varieties, then

$$\dim(X \times Y) = \dim(X) + \dim(Y).$$

PROOF Suppose that $\dim(X) = d$ and $\dim(Y) = e$. By Theorem 5.31, Noether bases exist for each of the K-algebras $K[X]$ and $K[Y]$, and by Corollary 6.33, any such Noether bases have size d and e, respectively. Since the subalgebra generated by a Noether basis is isomorphic to a polynomial ring, this implies that $K[X]$ is module-finite over a subalgebra isomorphic to $K[\mathbb{A}^d]$, and similarly for $K[Y]$. In other words,

$$K[X] \cong K[\mathbb{A}^d]\{a_1, \ldots, a_m\} \quad \text{and} \quad K[Y] \cong K[\mathbb{A}^e]\{b_1, \ldots, b_n\}$$

for some $a_1, \ldots, a_m \in K[X]$ and $b_1, \ldots, b_n \in K[Y]$. By Theorem 8.27, we have

$$K[X \times Y] = K[X] \otimes_K K[Y],$$

implying that $K[X \times Y]$ contains a subalgebra isomorphic to

$$K[\mathbb{A}^d] \otimes_K K[\mathbb{A}^e] = K[\mathbb{A}^{d+e}]$$

and what remains to show is that $K[X \times Y]$ is module-finite over this subalgebra.

Any element of $K[X]$ can be written as $F_1 a_1 + \cdots + F_m a_m$ for some elements $F_1, \ldots, F_m \in K[\mathbb{A}^d]$, and any element of $K[Y]$ can be written as $G_1 b_1 + \cdots + G_n b_n$ for some $G_1, \ldots, G_n \in K[\mathbb{A}^e]$. From this we see that any simple tensor in $K[X \times Y]$ can be written as

$$\begin{aligned}(F_1 a_1 + \cdots + F_m a_m) \otimes (G_1 b_1 + \cdots + G_n b_n) &= \sum_{i,j} F_i a_i \otimes G_j b_j \\ &= \sum_{i,j} (F_i \otimes G_j)(a_i \otimes b_j)\end{aligned}$$

where $F_i \otimes G_j \in K[\mathbb{A}^{d+e}]$. Since every simple tensor in $K[X \times Y]$ can be written as a $K[\mathbb{A}^{d+e}]$-linear combination of simple tensors of the form $a_i \otimes b_j$, it then follows that any element of $K[X \times Y]$ can be thus written. Therefore, we have proved that

$$K[X \times Y] \cong K[\mathbb{A}^{d+e}]\{a_i \otimes b_j \mid i = 1, \ldots, m,\ b = 1, \ldots, n\}.$$

Since $K[X \times Y]$ is module-finite over a polynomial ring with $d + e$ variables, we conclude that $\dim(X \times Y) = d + e = \dim(X) + \dim(Y)$. □

We now turn toward a discussion of irreducibility. In fact, this discussion is necessary in order to set the stage for results on smoothness, since we have defined smoothness only in the setting of irreducible affine varieties. Thus, in order for it to be sensible to show that a product of smooth affine varieties is smooth, we must first confirm that a product of irreducible varieties is irreducible.

8.32 **PROPOSITION** *Products of irreducible varieties are irreducible*

If X and Y are irreducible affine varieties, then $X \times Y$ is irreducible.

8.33 **EXAMPLE** A parabola in $\mathbb{A}^3$ is irreducible

If $X = \mathcal{V}(y - x^2) \subseteq \mathbb{A}^2$ and $Y = \mathcal{V}(z-1) \subseteq \mathbb{A}^1$, then both X and Y are irreducible, and the product

$$X \times Y = \mathcal{V}(y - x^2, z - 1) \subseteq \mathbb{A}^3$$

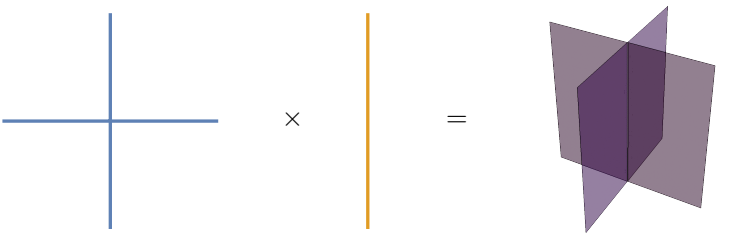
is a parabola in $\mathbb{A}^3$. The irreducibility of $X \times Y$ follows from the isomorphism $X \times Y \cong X$, which is simply a restriction of the vertical projection $\mathbb{A}^3 \to \mathbb{A}^2$.

8.34 **EXAMPLE** A reducible product

Let $X = \mathcal{V}(xy) \subseteq \mathbb{A}^2$ and let $Y = \mathbb{A}^1$. Then $X = \mathcal{V}(x) \cup \mathcal{V}(y)$ is reducible, as is the product:

$$X \times Y = \mathcal{V}(x) \cup \mathcal{V}(y) \subseteq \mathbb{A}^3;$$

see the image below.



While this example is rather simplified by the fact that Y is an entire affine space, it contains the kernel of the proof of Proposition 8.32: if Y is irreducible, then a decomposition of $X \times Y$ as a union of two affine varieties can only come from a similar decomposition of X.

PROOF OF PROPOSITION 8.32 Suppose that $X \subseteq \mathbb{A}^n$ and $Y \subseteq \mathbb{A}^m$ are irreducible. To prove that $X \times Y$ is irreducible, we must show that one can only have

$$X \times Y = Z_1 \cup Z_2 \tag{8.35}$$

for affine varieties $Z_1, Z_2 \subseteq X \times Y \subseteq \mathbb{A}^{m+n}$ if either $Z_1 = X \times Y$ or $Z_2 = X \times Y$.

Suppose that (8.35) holds, and for each point $a \in X$, consider the set

$$\{a\} \times Y \subseteq X \times Y.$$

For example, if $X \times Y$ is the cylinder of Example 8.3 or the intersecting planes of Example 8.34, then the sets $\{a\} \times Y$ are the vertical lines in the product. Each set $\{a\} \times Y$ is an affine variety (because both $\{a\}$ and Y are affine varieties), and they are all isomorphic to Y; therefore, since Y is irreducible, $\{a\} \times Y$ is irreducible for all $a \in X$. Intersecting both sides of (8.35) with $\{a\} \times Y$ yields

$$\{a\} \times Y = \Big(\big(\{a\} \times Y\big) \cap Z_1\Big) \cup \Big(\big(\{a\} \times Y\big) \cup Z_2\Big).$$

Given that $\{a\} \times Y$ is irreducible, it must be the case that either

$$\big(\{a\} \times Y\big) \cap Z_1 = \{a\} \times Y \quad \text{or} \quad \big(\{a\} \times Y\big) \cap Z_2 = \{a\} \times Y.$$

In other words, for any $a \in X$, either

$$\{a\} \times Y \subseteq Z_1 \quad \text{or} \quad \{a\} \times Y \subseteq Z_2.$$

Thus, if we define two sets

$$X_1 = \big\{a \in X \mid \{a\} \times Y \subseteq Z_1\big\} \quad \text{and} \quad X_2 = \big\{a \in X \mid \{a\} \times Y \subseteq Z_2\big\},$$

then

$$X_1 \cup X_2 = X. \tag{8.36}$$

So far, we only know that X_1 and X_2 are subsets of X, but in fact, they are affine varieties themselves. To see this, recall that Z_1 is an affine variety, so we have

$$Z_1 = \mathcal{V}(f_1, \ldots, f_r)$$

for some polynomials $f_1, \ldots, f_r \in K[x_1, \ldots, x_n, y_1, \ldots, y_m]$. For any $b \in Y$, fixing the y-coordinates of these polynomials at b and letting the x-coordinates vary yields an affine variety

$$(X_1)_b = \mathcal{V}\big(f_1(x, b), \ldots, f_r(x, b)\big) \subseteq \mathbb{A}^n.$$

In other words,

$$\begin{aligned}(X_1)_b &= \{a \in \mathbb{A}^n \mid f_1(a, b) = \cdots = f_r(a, b) = 0\} \\ &= \{a \in \mathbb{A}^n \mid (a, b) \in Z_1\}.\end{aligned}$$

Note that

$$X_1 = \{a \in X \mid (a, b) \in Z_1 \text{ for all } b \in Y\} = X \cap \bigcap_{b \in Y} (X_1)_b,$$

so X_1 is an intersection of affine varieties and is thus an affine variety. The same argument applies to X_2.

Recalling equation (8.36), we have now expressed X as a union of two affine varieties. Given that X is irreducible, this is only possible if either $X_1 = X$ or $X_2 = X$. The case that $X_1 = X$ means that for all $a \in X$,

$$\{a\} \times Y \subseteq Z_1,$$

or in other words, that for all $a \in X$ and all $b \in Y$,

$$(a, b) \in Z_1.$$

That is, $X \times Y = Z_1$. Similarly, the case that $X_2 = X$ means that $X \times Y = Z_2$. We have thus shown that either $X \times Y = Z_1$ or $X \times Y = Z_2$, and this completes the proof that $X \times Y$ is irreducible. □

The stage is now set for the final topic of this section: a discussion of smoothness of products. The following result describing linearizations and tangent spaces of products is the crucial step toward showing that a product of smooth affine varieties is smooth.

8.37 **PROPOSITION** *Tangent spaces of products*

If $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$ are affine varieties with $a \in X$ and $b \in Y$, then

$$L_{(a,b)}(X \times Y) = L_a X \times L_b Y \quad \text{and} \quad T_{(a,b)}(X \times Y) = T_a X \oplus T_b Y.$$

PROOF Suppose that $X = \mathcal{V}(f_1, \ldots, f_k)$ and $Y = \mathcal{V}(g_1, \ldots, g_\ell)$. By Proposition 8.1, we have

$$X \times Y = \mathcal{V}(f_1, \ldots, f_k, g_1, \ldots, g_\ell).$$

As we saw in Section 7.1,

$$\begin{aligned} L_a X &= \mathcal{V}(L_a f_1, \ldots, L_a f_k), \\ L_b X &= \mathcal{V}(L_b g_1, \ldots, L_b g_\ell), \quad \text{and} \\ L_{(a,b)}(X \times Y) &= \mathcal{V}(L_a f_1, \ldots, L_a f_k, L_b g_1, \ldots, L_b g_\ell). \end{aligned}$$

Thus, by Proposition 8.1 again, we conclude that $L_{(a,b)}(X \times Y) = L_a X \times L_b Y$.

To prove the statement regarding tangent spaces, recall from Section 7.1 that

$$T_a X = \left\{\overrightarrow{aa'} \mid a' \in L_a X\right\} \subseteq K^m \quad \text{and} \quad T_a Y = \left\{\overrightarrow{bb'} \mid b' \in L_a X\right\} \subseteq K^n.$$

Therefore, the vector space $T_a X \oplus T_b Y \subseteq K^m \oplus K^n = K^{m+n}$ consists of vectors of the form

$$\left(\overrightarrow{aa'}, \overrightarrow{bb'}\right)$$

with $a' \in L_a X$ and $b' \in L_b Y$. By definition of the vector between two points of affine space (Section 7.1), it follows that

$$\left(\overrightarrow{aa'}, \overrightarrow{bb'}\right) = \overrightarrow{(a,b)(a',b')}.$$

Thus,

$$T_a X \oplus T_b Y = \left\{\overrightarrow{(a,b)(a',b')} \;\middle|\; (a', b') \in L_a X \times L_b Y\right\}.$$

Since $L_aX \times L_bY = L_{(a,b)}(X \times Y)$ (by the first part of the proposition), we then conclude that

$$T_aX \oplus T_bY = \left\{ \overrightarrow{(a,b)(a',b')} \mid (a',b') \in L_{(a,b)}(X \times Y) \right\} = T_{(a,b)}(X \times Y). \quad \square$$

8.38 COROLLARY *Products and smoothness*

If X and Y are smooth irreducible varieties, then so is $X \times Y$.

PROOF Suppose that X and Y are smooth irreducible varieties. Given any $(a,b) \in X \times Y$, we then have

$$\begin{aligned} \dim(X \times Y) &= \dim(X) + \dim(Y) \\ &= \dim(T_aX) + \dim(T_bY) \\ &= \dim(T_aX \oplus T_bY) \\ &= \dim(T_{(a,b)}(X \times Y)), \end{aligned}$$

where the first equality is Proposition 8.31, the second uses the assumption that both X and Y are smooth, the third is the fact that vector space dimension is additive over direct sums, and the fourth is Proposition 8.37. By definition of smoothness, this implies that $X \times Y$ is smooth. $\square$

Exercises for Section 8.6

8.6.1 Let $X = \mathcal{V}(xy) \subseteq \mathbb{A}^2$ and let $Y = \mathcal{V}(z^2 - 1) \subseteq \mathbb{A}^1$. Calculate the irreducible decomposition of $X \times Y \subseteq \mathbb{A}^3$.

8.6.2 Does the converse of Proposition 8.32 hold? Prove or give a counterexample.

8.6.3 State and prove a result that describes the irreducible components of $X \times Y$ in terms of the irreducible components of X and Y.

8.6.4 Write an alternative proof of Proposition 8.31 by arguing that, for any irreducible affine varieties X and Y, we have

(a) $K(X \times Y) = K(X) \otimes_K K(Y)$, and

(b) $\operatorname{trdeg}_K(K(X) \otimes_K K(Y)) = \operatorname{trdeg}_K K(X) + \operatorname{trdeg}_K K(Y)$.

8.6.5 Suppose that K is algebraically closed, and let A and B be finitely-generated reduced K-algebras.

(a) Combine Proposition 8.32 with other results you have learned to show that if A and B are integral domains, then $A \otimes_K B$ is an integral domain.

(b) Prove that the result of part (a) can fail if K is not algebraically closed by arguing that

$$\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} \cong \mathbb{C} \oplus \mathbb{C}$$

as $\mathbb{R}$-algebras, and although $\mathbb{C}$ is an integral domain, $\mathbb{C} \oplus \mathbb{C}$ is not.

8.6.6 Does the converse of Corollary 8.38 hold? More specifically, if X and Y are irreducible and $X \times Y$ is smooth, does this imply that both X and Y are smooth? Prove or give a counterexample.

Part II

Projective varieties

Chapter 9

Projective Varieties

LEARNING OBJECTIVES FOR CHAPTER 9

- Become acquainted with projective space from various perspectives.
- Understand what it means for a polynomial to vanish at a point of projective space.
- Become acquainted with the projective $\mathcal{V}$- and $\mathcal{I}$-operators, and the relationship between them.
- Calculate the affine restrictions of a projective variety and the projective closure of an affine variety.
- Familiarize ourselves with the projective Nullstellensatz.

Up until this point in the book, all of the varieties that we have studied have lived inside of affine space $\mathbb{A}^n$. There is a larger ambient space, however, in which the notion of "variety" also make sense, known as *projective space* and denoted $\mathbb{P}^n$. The goal of this chapter is to define $\mathbb{P}^n$ and the *projective varieties* one obtains as vanishing sets of polynomials inside $\mathbb{P}^n$.

The motivation for this generalization comes from the desire to make uniform statements in settings where a statement about affine varieties has unavoidable exceptions. A key example of this phenomenon is the statement that, in $\mathbb{A}^2$, any pair of lines must intersect—with the unavoidable exception of parallel lines. Projective space $\mathbb{P}^2$ can be viewed as the result of adding "points at infinity" to $\mathbb{A}^2$, where each line in $\mathbb{A}^2$ meets a particular point at infinity dictated by the line's slope. With the addition of these extra points, we find in $\mathbb{P}^2$ that every pair of lines intersects, without exception. This is a special case of a beautiful result known as Bézout's Theorem, which states that a pair of curves in $\mathbb{P}^2$, defined by polynomials of degrees r and s, intersect in $r \cdot s$ points when counted appropriately. While the corresponding statement in $\mathbb{A}^2$ is often true, one can easily find exceptions: the parabola $\mathcal{V}(y - x^2)$ and the vertical line $\mathcal{V}(x)$ intersect in only a single point, for example. From the perspective of $\mathbb{P}^2$, this exception again occurs because there is an additional intersection "at infinity" that is hidden when one restricts attention to $\mathbb{A}^2$.

These observations in plane geometry led algebraic geometers to ultimately understand $\mathbb{P}^n$, and not $\mathbb{A}^n$, as the most natural ambient space in which to work. While the definition of projective space can be difficult to digest on a first pass, the elegance and uniformity that it will lend to our study of algebraic geometry will certainly be worth the effort.

Section 9.1 Projective space

Part of what makes the study of projective varieties challenging on a first encounter—but also what makes it rich and interesting—is the multitude of different ways in which one can define projective space. We will present three different perspectives on projective space, beginning with the one that is the most computationally useful.

9.1 **DEFINITION** *Projective space, first perspective*

Let $n \in \mathbb{N}$. The n-dimensional *projective space* over K, denoted $\mathbb{P}^n_K$ or simply $\mathbb{P}^n$, is the set

$$\mathbb{P}^n = \frac{K^{n+1} \setminus \{(0,0,\ldots,0)\}}{\sim},$$

where $\sim$ is the equivalence relation given by

$$(a_0, a_1, \ldots, a_n) \sim (b_0, b_1, \ldots, b_n)$$

$$\Longleftrightarrow$$

$$(\lambda a_0, \lambda a_1, \ldots, \lambda a_n) = (b_0, b_1, \ldots, b_n) \text{ for some } \lambda \in K \setminus \{0\}.$$

We denote the equivalence class of $(a_0, a_1, \ldots, a_n)$ by $[a_0 : a_1 : \cdots : a_n]$.

Note that the term "dimension" in this context should be taken, for now, as nothing more than an indication of the number of coordinates; since $\mathbb{P}^n$ is neither a vector space nor an affine variety, it cannot be meaningfully given a dimension in any of the contexts in which that term has been used thus far in this book. Nevertheless, our use of the term "n-dimensional" may make somewhat more sense after the following examples.

9.2 **EXAMPLE** 0-dimensional projective space

An element of $\mathbb{P}^0$ is an equivalence class $[a]$, where $a \in K \setminus \{0\}$ and $[a] = [b]$ if $\lambda a = b$. In particular, taking $\lambda = 1/a$ shows that $[a] = [1]$ for any $a \in K \setminus \{0\}$, so $\mathbb{P}^0$ has just a single element:

$$\mathbb{P}^0 = \{[1]\}.$$

9.3 **EXAMPLE** 1-dimensional projective space

Elements of $\mathbb{P}^1$ are of the form $[a_0 : a_1]$, where $a_0, a_1 \in K$ are not both zero. For instance, $[1 : 2] \in \mathbb{P}^1$, and scaling both coordinates by the same $\lambda \in K \setminus \{0\}$ yields different representations of the same element:

$$[1 : 2] = [2 : 4] = [3 : 6] = [-1 : -2] = \cdots.$$

It is instructive to divide the elements of $\mathbb{P}^1$ into two types: those whose first coordinate is nonzero and those whose first coordinate is zero. Consider an element of the first type, such as $[3 : 7] \in \mathbb{P}^1$. Scaling both coordinates by $1/3$ shows that

$$[3 : 7] = [1 : 7/3] \in \mathbb{P}^1,$$

and similarly, any element of $\mathbb{P}^1$ with nonzero first coordinate is equal to $[1 : b]$ for some $b \in K$. On the other hand, an element of $\mathbb{P}^1$ whose first coordinate is zero is always equal to $[0 : 1]$; for instance, scaling both coordinates by $1/4$ shows that $[0 : 4] = [0 : 1] \in \mathbb{P}^1$.

The conclusion that we arrive at, then, is that

$$\mathbb{P}^1 = \{[1 : b] \mid b \in K\} \sqcup \{[0 : 1]\}. \tag{9.4}$$

We use the symbol $\sqcup$ for disjoint unions*; in other words, $A = B \sqcup C$ means $A = B \cup C$ and $B \cap C = \emptyset$.*

Some reflection should convince the reader that two elements $[1 : b]$ and $[1 : b']$ with $b \neq b'$ cannot be equal to one another in $\mathbb{P}^1$. As a result, there is a natural bijection between the elements of the form $[1 : b] \in \mathbb{P}^1$ and the elements of $\mathbb{A}^1$ given by

$$\begin{aligned} \{[1 : b] \mid b \in K\} &\to \mathbb{A}^1 \\ [1 : b] &\mapsto b. \end{aligned}$$

Under this bijection, the decomposition (9.4) can be viewed as

$$\mathbb{P}^1 = \mathbb{A}^1 \sqcup \{[0 : 1]\}. \tag{9.5}$$

9.6 EXAMPLE 2-dimensional projective space

As above, the elements of $\mathbb{P}^2$ can be divided into two types, depending on whether their first coordinate is nonzero or zero, and those with nonzero first coordinate can be rescaled to the form $[1 : b_1 : b_2]$. Thus,

$$\mathbb{P}^2 = \{[1 : b_1 : b_2] \mid b_1, b_2 \in K\} \sqcup \{[0 : a_1 : a_2] \mid a_1, a_2 \in K \text{ not both zero}\}.$$

Also analogously to the previous example, elements of the first type are in natural bijection with $\mathbb{A}^2$:

$$\begin{aligned} \{[1 : b_1 : b_2] \mid b_1, b_2 \in K\} &\to \mathbb{A}^2 \\ [1 : b_1 : b_2] &\mapsto (b_1, b_2). \end{aligned}$$

Now there is not just a single element with first coordinate zero, however, but many; for example, $[0 : 0 : 1] \neq [0 : 1 : 1]$. In fact, elements of $\mathbb{P}^2$ with first coordinate zero are in natural bijection with a projective space of one dimension lower:

$$\begin{aligned} \{[0 : b_1 : b_2] \mid b_1, b_2 \in K \text{ not both zero}\} &\to \mathbb{P}^1 \\ [0 : b_1 : b_2] &\mapsto [b_1 : b_2]. \end{aligned}$$

Under these two bijections, we have shown that

$$\mathbb{P}^2 = \mathbb{A}^2 \sqcup \mathbb{P}^1. \tag{9.7}$$

The decompositions (9.5) and (9.7) can be generalized to any n, and doing so brings us to our second perspective on projective space.

9.8 **PROPOSITION** *Projective space, second perspective*

For any $n \geq 1$, there is a natural bijection

$$\mathbb{P}^n = \mathbb{A}^n \sqcup \mathbb{P}^{n-1}.$$

The elements of $\mathbb{P}^{n-1}$ inside $\mathbb{P}^n$ are referred to as *points at infinity* in $\mathbb{P}^n$.

The proof of this bijection is the content of Exercise 9.1.3; the key point, as we saw in the cases of $\mathbb{P}^1$ and $\mathbb{P}^2$ above, is that elements of $\mathbb{P}^n$ with nonzero first coordinate correspond to elements of $\mathbb{A}^n$, whereas elements with first coordinate zero correspond to elements of $\mathbb{P}^{n-1}$.

Why, though, the terminology "points at infinity"? To understand this, consider the case of $\mathbb{P}^1_{\mathbb{R}}$. Under the decomposition

$$\mathbb{P}^1_{\mathbb{R}} = \mathbb{A}^1_{\mathbb{R}} \sqcup \{[0:1]\},$$

the elements $1, 2, 3, \ldots \in \mathbb{A}^1_{\mathbb{R}}$ correspond to the following elements of $\mathbb{P}^1_{\mathbb{R}}$:

$$[1:1],\ [1:2],\ [1:3], \ldots \in \mathbb{P}^1_{\mathbb{R}}.$$

By rescaling, though, these can be re-expressed as

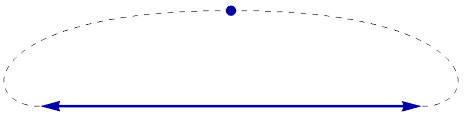
$$\left[\tfrac{1}{1}:1\right],\ \left[\tfrac{1}{2}:1\right],\ \left[\tfrac{1}{3}:1\right], \ldots \in \mathbb{P}^1_{\mathbb{R}}.$$

Thus, as $n \in \mathbb{A}^1_{\mathbb{R}}$ approaches ∞, the corresponding element $[\frac{1}{n} : 1] \in \mathbb{P}^1_{\mathbb{R}}$ approaches $[0:1]$. This is why we refer to $[0:1]$ as the *point at infinity*, writing

Our use of "limits" in $\mathbb{P}^n$ is merely intuitive, for now, since a topology is needed to make limits precise.

$$\mathbb{P}^1_{\mathbb{R}} = \mathbb{A}^1_{\mathbb{R}} \sqcup \{\infty\}.$$

Note that the points $-n \in \mathbb{A}^1_{\mathbb{R}}$, corresponding to $[1 : -n] \in \mathbb{P}^1_{\mathbb{R}}$, also approach $[0:1]$ as $n \to \infty$. Thus, visually, it is illustrative to depict $\mathbb{P}^1_{\mathbb{R}}$ as a loop: as we go infinitely far in either direction of $\mathbb{A}^1_{\mathbb{R}}$, we end up at the same point $[0:1]$.



For $n > 1$, it becomes more difficult to give a visual representation of $\mathbb{P}^n$, but the same perspective still holds. For example, we have

$$\mathbb{P}^2_{\mathbb{R}} = \mathbb{A}^2_{\mathbb{R}} \sqcup \{\text{points at infinity}\}.$$

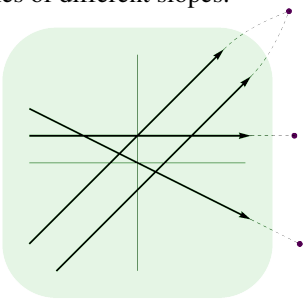
In this space, we can "walk toward infinity" along any line in $\mathbb{A}^2_{\mathbb{R}}$. To do so, consider the sequence of points $(1, m), (2, 2m), (3, 3m), \ldots$ that radiate outward along the line $y = mx$. These correspond in $\mathbb{P}^2_{\mathbb{R}}$ to

$$[1:1:m],\ [1:2:2m],\ [1:3:3m], \ldots \in \mathbb{P}^2_{\mathbb{R}},$$

which can be re-scaled to equal

$$\left[\tfrac{1}{1} : 1 : m\right],\ \left[\tfrac{1}{2} : 1 : m\right],\ \left[\tfrac{1}{3} : 1 : m\right], \ldots \in \mathbb{P}^2_{\mathbb{R}}$$

and thus approach $[0 : 1 : m]$. This limit is a "point at infinity" in $\mathbb{P}^2_{\mathbb{R}}$, since it is a point with first coordinate zero. Below, we have depicted $\mathbb{A}^2_{\mathbb{R}}$ along with several points at infinity in $\mathbb{P}^2_{\mathbb{R}}$ that are approached along lines of different slopes.



We see now why $\mathbb{P}^2_{\mathbb{R}}$ has many points at infinity whereas $\mathbb{P}^1_{\mathbb{R}}$ had just one: in $\mathbb{P}^2_{\mathbb{R}}$, the point at infinity that we approach by walking outward along a line depends on the slope of that line. In fact, the idea of "following a line to the point at infinity to which it leads" can be made precise as a bijection between points at infinity in $\mathbb{P}^2_{\mathbb{R}}$ and lines through the origin in $\mathbb{A}^2_{\mathbb{R}}$. This is a special case of a more general phenomenon, which we now state.

9.9 PROPOSITION *Points at infinity are slopes of lines*

For any $n \geq 1$, there is a natural bijection

$$\{\text{points at infinity in } \mathbb{P}^n\} = \{\text{lines through } (0,\ldots,0) \text{ in } \mathbb{A}^n\}. \tag{9.10}$$

PROOF A line through $(0,\ldots,0) \in \mathbb{A}^n$ is, by definition, a set of points of the form

$$L = \{(a_1 b, a_2 b, \ldots, a_n b) \mid b \in K\},$$

where $a_i \in K$ are fixed and at least one a_i is nonzero. The bijection (9.10), then, is given by associating to the line

$$L = \{(a_1 b, a_2 b, \ldots, a_n b) \mid b \in K\}$$

the point at infinity

$$[0 : a_1 : a_2 : \cdots : a_n] \in \mathbb{P}^n,$$

which (as the reader is encouraged to verify intuitively) should be viewed as the point toward which an outward-radiating sequence of points on L tends. □

Given that the points at infinity in $\mathbb{P}^n$ are also in bijection with $\mathbb{P}^{n-1}$, Proposition 9.9 can be viewed in another light: it gives us our third perspective on projective space, which is often taken as an alternative definition of the space itself.

9.11 COROLLARY *Projective space, third perspective*

For any $n \in \mathbb{N}$, there is a natural bijection

$$\mathbb{P}^n = \{\text{lines through } (0,\ldots,0) \text{ in } \mathbb{A}^{n+1}\}.$$

Tracing through the bijection of Proposition 9.9 explains how to match up the first and third perspectives with one another: an element $[a_0 : a_1 : \cdots : a_n] \in \mathbb{P}^n$ is equivalent to a point at infinity $[0 : a_0 : a_1 : \cdots : a_n] \in \mathbb{P}^{n+1}$, which corresponds to a line

$$L = \{(a_0 b, a_1 b, \ldots, a_n b) \mid b \in K\}$$

through the origin in $\mathbb{A}^{n+1}$.

With these ideas combined, an element of $\mathbb{P}^n$ can be viewed in three different ways: as an equivalence class $[a_0 : a_1 : \cdots : a_n]$, as a point either in $\mathbb{A}^n$ or at infinity, or as a line through the origin in $\mathbb{A}^{n+1}$. Moving fluidly between these perspectives as the context dictates is one of the skills that the reader will develop as we explore projective space and—beginning in the next section—the analogue in projective space of all we know about varieties.

Exercises for Section 9.1

9.1.1 Let $[2 : 1 : 3] \in \mathbb{P}^2$. Prove that

$$[2 : 1 : 3] = [6 : 3 : 9]$$

but that

$$[2 : 1 : 3] \neq [6 : 4 : 12].$$

In general, which $[a : b : c] \in \mathbb{P}^2$ satisfy $[2 : 1 : 3] = [a : b : c]$?

9.1.2 Prove that $[a_0 : \cdots : a_n] = [b_0 : \cdots : b_n] \in \mathbb{P}^n$ if and only if

$$a_i b_j = a_j b_i \quad \text{for all} \quad i, j.$$

9.1.3 Prove that there is a natural bijection

$$\mathbb{P}^n = \mathbb{A}^n \sqcup \mathbb{P}^{n-1}$$

in three steps:

(a) Let

$$U = \{[a_0 : a_1 : \cdots : a_n] \in \mathbb{P}^n \mid a_0 \neq 0\}.$$

and

$$V = \{[a_0 : a_1 : \cdots : a_n] \in \mathbb{P}^n \mid a_0 = 0\}.$$

Prove that $\mathbb{P}^n = U \sqcup V$.

(b) Prove that there is a natural bijection between U and $\mathbb{A}^n$.

(c) Prove that there is a natural bijection between V and $\mathbb{P}^{n-1}$.

9.1.4 Prove that there is a natural bijection

$$\mathbb{P}^n = \mathbb{A}^n \sqcup \mathbb{A}^{n-1} \sqcup \mathbb{A}^{n-2} \sqcup \cdots \sqcup \mathbb{A}^1 \sqcup \mathbb{A}^0.$$

9.1.5 Let $[2 : 1 : 3] \in \mathbb{P}^2$. Describe this point as

(a) an element of $\mathbb{A}^2 \sqcup \{\text{points at infinity}\}$,

(b) a line through $(0, 0, 0)$ in $\mathbb{A}^3$.

9.1.6 Repeat Problem 9.1.5 for the point $[0 : 1 : 3] \in \mathbb{P}^2$. Which lines through $(0, 0, 0) \in \mathbb{A}^3$ correspond to points at infinity in $\mathbb{P}^2$?

Section 9.2 The projective $\mathcal{V}$-operator

Just like an affine variety, a projective variety is defined as the common vanishing of a set of polynomials. Taking the perspective on $\mathbb{P}^n$ given in Definition 9.1, the inputs to those polynomials are tuples $(a_0, \ldots, a_n)$. However, since Definition 9.1 involves an equivalence relation, a polynomial might vanish on one representative but not on another, so it's not immediately clear what we mean when we say that a polynomial "vanishes" at a point of projective space.

For instance, suppose we consider points $[a_0 : a_1] \in \mathbb{P}^1$ as inputs to the two-variable polynomial

$$f = x_0^2 - x_1.$$

The point $[2 : 4]$ would seem to be in the vanishing set of f, since

$$f(2,4) = 2^2 - 4 = 0.$$

On the other hand, however, we see that $[2 : 4] = [4 : 8]$, and

$$f(4,8) = 4^2 - 8 = 8 \neq 0.$$

Thus, the question of whether f vanishes at the point $[2 : 4] = [4 : 8]$ does not seem to have a well-defined answer. The solution to this discrepancy is simply to declare that a polynomial "vanishes" at a point of projective space only if it vanishes when evaluated at *every* representative of the point.

9.12 **DEFINITION** *Projective vanishing*

Let $f \in K[x_0, \ldots, x_n]$ be a polynomial and let $a \in \mathbb{P}^n$ be a point. We say that f *vanishes at* a and write $f(a) = 0$ if

$$f(a_0, \ldots, a_n) = 0$$

for every representative $a = [a_0 : \cdots : a_n]$.

For example, the polynomial $f(x_0, x_1) = x_0^2 - x_1$ does *not* vanish at the point $[2 : 4] \in \mathbb{P}^1$ because it does not vanish when evaluated at the equivalent representative $a = [4 : 8]$. A priori, checking that a polynomial vanishes at *every* representative of a point seems to be an arduous task—after all, there are infinitely-many representatives for any point. However, this task can be simplified with the introduction of homogeneous polynomials.

9.13 **DEFINITION** *Homogeneous polynomial*

A polynomial $f \in K[x_0, \ldots, x_n]$ is *homogeneous of degree* d if every nonzero term of f has degree d.

For example, the polynomial $f = x_0^2 - x_1$ is inhomogeneous, because it has a nonzero term of degree one and another of degree two, while the polynomial $g = x_0^2 - 2x_0x_1$ is homogeneous of degree 2. The zero polynomial is vacuously homogeneous of every degree, since it does not have any nonzero terms.

In the context of studying vanishing within projective space, the importance of working with homogeneous polynomials is the following result.

9.14 **LEMMA** *Projective vanishing of homogeneous polynomials*

Let $f \in K[x_0, \ldots, x_n]$ be a homogeneous polynomial and let

$$[a_0 : \cdots : a_n] = [b_0 : \cdots : b_n] \in \mathbb{P}^n.$$

Then

$$f(a_0, \ldots, a_n) = 0 \iff f(b_0, \ldots, b_n) = 0.$$

In other words, when working with homogeneous polynomials, vanishing of a polynomial can be verified by checking vanishing at a single representative. For example, consider the homogeneous polynomial $g = x_0^2 - 2x_0x_1$. To check that g vanishes at the point $[2 : 1] \in \mathbb{P}^1$, it suffices to verify vanishing at one representative:

$$g(2, 1) = 2^2 - 2 \cdot 2 \cdot 1 = 0.$$

If we replace $[2 : 1]$ by the alternative representative $[4 : 2]$ (or any other representative for this point of $\mathbb{P}^1$), the vanishing persists:

$$g(4, 2) = 4^2 - 2 \cdot 4 \cdot 2 = 0.$$

PROOF OF LEMMA 9.14 The key observation we need is that f is homogeneous of degree d if and only if

$$f(\lambda a_0, \ldots, \lambda a_n) = \lambda^d f(a_0, \ldots, a_n) \tag{9.15}$$

for all $\lambda, a_0, \ldots, a_n \in K$; see Exercise 9.2.1. Given this, suppose that

$$[a_0 : \cdots : a_n] = [b_0 : \cdots : b_n] \in \mathbb{P}^n.$$

By the definition of the equivalence relation on $\mathbb{P}^n$, there exists a nonzero $\lambda \in K$ such that

$$(\lambda a_0, \ldots, \lambda a_n) = (b_0, \ldots, b_n).$$

From (9.15) we then see that $f(b_0, \ldots, b_n)$ and $f(a_0, \ldots, a_n)$ differ by the nonzero scalar multiple λ^d, so one vanishes if and only if the other does. □

Lemma 9.14 shows that we can readily determine whether a homogeneous polynomial vanishes at a point of projective space, simply by checking a single representative, but what about the vanishing of an inhomogeneous polynomial? To address this question, we introduce the homogeneous components of a polynomial.

9.16 **DEFINITION** *Homogeneous components*

Given a polynomial $f \in K[x_0, \ldots, x_n]$, the *dth homogeneous component of* f, denoted f_d, is the sum of all nonzero terms of f of degree d. If f does not have any nonzero terms of degree d, then $f_d = 0$.

For example, the nonzero homogeneous components of the polynomial

$$f = x^4z + x^2y + xyz + x + y + 5$$

are

$$f_5 = x^4z, \quad f_3 = x^2y + xyz, \quad f_1 = x + y, \quad f_0 = 5.$$

The next result describes the vanishing of a polynomial at a point of projective space in terms of the vanishing of its homogeneous components.

9.17 LEMMA *Projective vanishing and homogeneous components*

Let $f \in K[x_0, \ldots, x_n]$ be a polynomial and $a \in \mathbb{P}^n$ a point. Then f vanishes at a if and only if every homogeneous component of f vanishes at a.

In other words, in order to determine whether a general polynomial vanishes at (every representative of) a point of $\mathbb{P}^n$, Lemmas 9.14 and 9.17 together imply that it suffices to check whether each homogeneous component vanishes at a single representative of that point.

PROOF OF LEMMA 9.17 Let $f \in K[x_0, \ldots, x_n]$ be a polynomial of degree d, which can be written as a sum of its homogeneous components of degree $\leq d$:

$$f = \sum_{k=0}^{d} f_k. \tag{9.18}$$

If each f_k vanishes at every representative of a point a, then it follows from (9.18) that f vanishes at every representative of a.

Conversely, assume that f vanishes at every representative of a. Choose one particular representative $a = [a_0 : \cdots : a_n]$. Then, for any $\lambda \in K \setminus \{0\}$, we have

$$\begin{aligned} 0 = f(\lambda a_0, \ldots, \lambda a_n) &= \sum_{k=0}^{d} f_k(\lambda a_0, \ldots, \lambda a_n) \\ &= \sum_{k=0}^{d} \lambda^k f_k(a_0, \ldots, a_n), \end{aligned}$$

where the second equality uses (9.15). In other words, the single variable polynomial

$$\sum_{k=0}^{d} x^k f_k(a_0, \ldots, a_n) \in K[x]$$

vanishes at infinitely many values of K, so it must be the zero polynomial, implying that $f_k(a_0, \ldots, a_n) = 0$ for all k. Since f_k is homogeneous, Lemma 9.14 then implies that f_k vanishes at every representative of a. □

With a better understanding of what it means for polynomials to vanish at points of projective space, we now come to the natural definition of a *projective variety*.

9.19 **DEFINITION** *Projective $\mathcal{V}$-operator*

Let $\mathcal{S} \subseteq K[x_0, x_1, \ldots, x_n]$ be a set of polynomials. The *projective vanishing set* of $\mathcal{S}$ is

$$\mathcal{V}_{\mathbb{P}}(\mathcal{S}) = \{a \in \mathbb{P}^n \mid f(a) = 0 \text{ for all } f \in \mathcal{S}\}.$$

We say that a subset $X \subseteq \mathbb{P}^n$ is a *projective variety* if $X = \mathcal{V}_{\mathbb{P}}(\mathcal{S})$ for some set $\mathcal{S} \subseteq K[x_0, \ldots, x_n]$.

We often omit the subscript and write simply $\mathcal{V}(\mathcal{S}) \subseteq \mathbb{P}^n$ when it is clear from context that we are working in projective space, as opposed to affine space.

Lemma 9.17 implies that every projective variety can be described by a set of homogeneous polynomials—simply replace the inhomogeneous polynomials in $\mathcal{S}$ with their homogeneous components. Because of this, it is common in practice to describe projective varieties using only homogeneous polynomials, as in the following examples.

9.20 **EXAMPLE** $\emptyset$ and $\mathbb{P}^n$ are projective varieties

As in the affine case, we have $\mathcal{V}(1) = \emptyset$ and $\mathcal{V}(0) = \mathbb{P}^n$, so $\emptyset$ and $\mathbb{P}^n$ are projective varieties.

9.21 **EXAMPLE** Projective varieties in $\mathbb{P}^1$

In $\mathbb{P}^1$, let

$$X = \mathcal{V}(2x_0 - x_1) = \{[a_0 : a_1] \in \mathbb{P}^1 \mid 2a_0 - a_1 = 0\}.$$

A point $[a_0 : a_1]$ in X cannot have $a_0 = 0$, since then the equation $2a_0 - a_1 = 0$ would force that $a_1 = 0$, as well. Thus, we have

$$X = \{[a : 2a] \in \mathbb{P}^1 \mid a \in K \setminus \{0\}\} = \{[1 : 2]\},$$

since multiplying both coordinates by a^{-1} shows that $[a : 2a] = [1 : 2]$ for any a. More generally (in perfect analogy to the situation for $\mathbb{A}^1$), any projective variety in $\mathbb{P}^1$ is either $\emptyset$, $\mathbb{P}^1$, or a finite set of points (Exercise 9.2.2).

9.22 **EXAMPLE** A line in $\mathbb{P}^2$

In $\mathbb{P}^2$, consider the projective variety

$$X = \mathcal{V}(x_0 + x_1 - x_2) = \{[a_0 : a_1 : a_2] \in \mathbb{P}^2 \mid a_0 + a_1 - a_2 = 0\}.$$

A point $[a_0 : a_1 : a_2]$ in X cannot have $a_0 = a_1 = 0$, since then the defining equation would force that $a_2 = 0$, as well. It follows that

$$X = \{[a_0 : a_1 : a_0 + a_1] \in \mathbb{P}^2 \mid a_0, a_1 \in K \text{ not both } 0\},$$

and from here it is not difficult to see that the points of $\mathcal{V}(x_0 + x_1 - x_2)$ are in bijection with $\mathbb{P}^1$.

The geometry of this example mirrors the affine case, in which the vanishing of a degree-1 polynomial in $\mathbb{A}^2$ is a line isomorphic to $\mathbb{A}^1$. We refer to X as a *line* in $\mathbb{P}^2$ to emphasize this analogy, but this terminology should be taken with a grain of salt: as we saw in the previous section, $\mathbb{P}^1$ does not look like our familiar notion of a "number line," even over the real numbers.

As a first step toward utilizing the algebraic structure of polynomial rings to study projective varieties, we note that every projective variety can be defined by an ideal, a result that is parallel to Proposition 1.15 in the affine setting.

9.23 **PROPOSITION** *Projective varieties are defined by ideals*

If $\mathcal{S} \subseteq K[x_0, x_1, \ldots, x_n]$ is a set of polynomials, then

$$\mathcal{V}_{\mathbb{P}}(\mathcal{S}) = \mathcal{V}_{\mathbb{P}}(\langle \mathcal{S} \rangle).$$

PROOF Exercise 9.2.5 □

As in the affine case, knowing that any projective variety can be defined by an ideal allows us to leverage the algebraic structure of polynomial rings to deduce that every projective variety can be defined by finitely many polynomials. Moreover, in the projective case we get a little more: utilizing Lemma 9.17, it follows that every projective variety is the vanishing of a finite set of *homogeneous* polynomials.

9.24 **COROLLARY** *Projective varieties are finitely-generated*

Any projective variety $X \subseteq \mathbb{P}^n$ is of the form $X = \mathcal{V}_{\mathbb{P}}(f_1, \ldots, f_k)$ where $f_1, \ldots, f_k \in K[x_0, x_1, \ldots, x_n]$ are homogeneous polynomials.

PROOF Exercise 9.2.6. □

At this point, we can begin to see the utility of working algebraically with defining ideals of projective varieties, rather than merely sets of polynomials. In particular, it is through passing to an ideal and using Hilbert's Basis Theorem for polynomial rings that one proves Corollary 9.24. Just like in the affine setting, while there may be many defining ideals for a single projective variety, there is always one distinguished ideal among all of its defining ideals—the vanishing ideal. In the next section, we turn to a discussion of vanishing ideals in the projective setting.

Exercises for Section 9.2

9.2.1 Let $f \in K[x_0, x_1, \ldots, x_n]$ be a nonzero polynomial. Prove that f is homogeneous of degree d if and only if, for any $\lambda, a_0, a_1, \ldots, a_n \in K$,

$$f(\lambda a_0, \lambda a_1, \ldots, \lambda a_n) = \lambda^d f(a_0, a_1 \ldots, a_n).$$

9.2.2 Prove that every projective variety in $\mathbb{P}^1$ is either $\emptyset$, $\mathbb{P}^1$, or a finite set of points.

9.2.3 Let

$$X = \mathcal{V}(x_0x_1 - x_2^2) \subseteq \mathbb{P}^2.$$

(a) Prove that an element $[a_0 : a_1 : a_2] \in X$ cannot have $a_0 = a_1 = 0$, and from here, describe a bijection between X and $\mathbb{P}^1$.

(b) Let $U = \{[a_0 : a_1 : a_2] \in \mathbb{P}^2 \mid a_0 \neq 0\}$, which, by the results of the previous section, is in natural bijection with $\mathbb{A}^2$. Prove that $X \cap U$ is identified by this bijection with an affine variety in $\mathbb{A}^2$. What is that affine variety?

(c) Compute all points of $X \setminus U$, and describe how these points are approached by points in the affine variety you found in (b).

9.2.4 Prove that finite unions and arbitrary intersections of projective varieties are, themselves, projective varieties.

9.2.5 Prove Proposition 9.23.

9.2.6 Prove Corollary 9.24.

Section 9.3 The projective $\mathcal{I}$-operator

The projective $\mathcal{V}$-operator allows us to pass from collections of polynomials to subsets of projective space, and we now turn to the projective $\mathcal{I}$-operator, which moves us in the opposite direction. The definition of $\mathcal{I}_{\mathbb{P}}$ is as one might expect.

9.25 DEFINITION *Projective $\mathcal{I}$-operator*

Let $X \subseteq \mathbb{P}^n$ be a subset. The *vanishing ideal* of X is

$$\mathcal{I}_{\mathbb{P}}(X) = \{f \in K[x_0, \ldots, x_n] \mid f(a) = 0 \text{ for all } a \in X\}.$$

We say that a subset of $K[x_0, \ldots, x_n]$ is a *projective vanishing ideal* if it is of the form $\mathcal{I}_{\mathbb{P}}(X)$ for some $X \subseteq \mathbb{P}^n$.

As for the $\mathcal{V}_{\mathbb{P}}$-operator, we often write $\mathcal{I}(X)$ when it is clear from context whether we are working in affine or projective space.

Recalling Definition 9.12, when we say that $f(a) = 0$, we are asserting that f vanishes at *every* representative of the point $a \in \mathbb{P}^n$. For instance, if $[1:0] \in X$, then we can say for certain that the polynomial $f = x_0 - 1$ is *not* an element of $\mathcal{I}(X)$—even though $f(1,0) = 0$, notice that f does not vanish when evaluated at the representative $[2:0] = [1;0]$. The next example elaborates further on this discussion.

9.26 EXAMPLE Vanishing ideal of a point in $\mathbb{P}^1$

Let $X = \{[1:0]\} \subseteq \mathbb{P}^1$. Then $f = x_1$ is in $\mathcal{I}(X)$, since any representative of the point $[1:0] \in X$ is of the form $[a:0]$ for some a, and $f(a,0) = 0$ for any choice of a. More generally, we see that $\langle x_1 \rangle \subseteq \mathcal{I}(X)$, and in fact, we claim that there is equality: $\mathcal{I}(X) = \langle x_1 \rangle$.

To prove the remaining inclusion, let $f \in \mathcal{I}(X)$. Then $f(a,0) = 0$ for all nonzero $a \in K$. It follows that $f(x_0,0)$ is a single-variable polynomial with infinitely many zeroes, so it must be the zero polynomial. Write f as an element of $(K[x_0])[x_1]$:

$$f = \sum_{d \geq 0} f_d(x_0) x_1^d.$$

Using $f_0(x_0) = f(x_0,0) = 0$, we conclude that

$$f = x_1 \sum_{d \geq 1} f_d(x_0) x_1^{d-1} \in \langle x_1 \rangle.$$

9.27 EXAMPLE Vanishing ideal of a line in $\mathbb{P}^2$

If X is the line $\mathcal{V}(x_0 + x_1 - x_2) \subseteq \mathbb{P}^2$ of Example 9.22, then

$$\mathcal{I}(X) = \langle x_0 + x_1 - x_2 \rangle.$$

The fact that every element of $\langle x_0 + x_1 - x_2 \rangle$ vanishes at every point of X is essentially immediate, while the reverse inclusion is the content of Exercise 9.3.8.

As in the affine case, projective vanishing ideals are, in fact, ideals, and moreover, they are readily seen to be radical ideals. In the projective setting, though, we get even more. In particular, Lemma 9.17 implies that, for every $f \in \mathcal{I}(X)$, every homogeneous component of f must also be an element of $\mathcal{I}(X)$. This attribute of $\mathcal{I}(X)$ is the defining property of what it means to be a *homogeneous ideal*.

9.28 DEFINITION *Homogeneous ideals in polynomial rings*

An ideal $I \subseteq K[x_0, \ldots, x_n]$ is *homogeneous* if, for every $f \in I$, every homogeneous component of f is also in I.

While the above definition of homogeneous ideals is directly motivated by our discussion of vanishing ideals, the following result offers an important alternative characterization of homogeneous ideals that is, perhaps, more straightforward, and that can be quite useful in practice.

9.29 PROPOSITION *Characterizing homogeneous ideals*

An ideal $I \subseteq K[x_0, \ldots, x_n]$ is homogeneous if and only if it admits a set of homogeneous generators.

PROOF First, suppose that I is a homogeneous ideal, and let $\mathcal{S} \subseteq I$ be the subset consisting of all homogeneous polynomials in I. We claim that $I = \langle \mathcal{S} \rangle$. The inclusion $\langle \mathcal{S} \rangle \subseteq I$ is immediate, since $\mathcal{S} \subseteq I$ and I is an ideal. Conversely, suppose that $f \in I$. Then we can express f as a sum of nonzero homogeneous components f_k, and the fact that I is a homogeneous set means that $f_k \in I$ for each k. Given that f_k is homogeneous, it follows that $f_k \in \mathcal{S}$. Therefore, f is a sum of elements of $\mathcal{S}$, so $f \in \langle \mathcal{S} \rangle$.

Conversely, suppose $I = \langle \mathcal{S} \rangle$, where $\mathcal{S}$ is a set of homogeneous polynomials. To prove that I is a homogeneous set, let $f \in I$. The fact that $I = \langle \mathcal{S} \rangle$ means that

$$f = \sum_{i=1}^{m} g_i h_i$$

for some $g_i \in K[x_0, \ldots, x_n]$ and $h_i \in \mathcal{S}$; in particular, h_i is homogeneous of some degree d_i. For each k, we have

$$f_k = \sum_{i=1}^{m} (g_i h_i)_k,$$

where, in the right-hand side, $(g_i h_i)_k$ denotes the kth homogeneous component of the polynomial $g_i h_i$. By Exercise 9.3.1, we can rewrite the summands as

$$(g_i h_i)_k = \begin{cases} (g_i)_{k-d_i} \cdot h_i & \text{if } d_i \leq k \\ 0 & \text{if } d_i > k, \end{cases}$$

where $(g_i)_{k-d_i}$ is the $(k-d_i)$th homogeneous component of g_i. This implies that $(g_i h_i)_k \in I$, so f_k is a sum of elements of I and hence $f_k \in I$. Thus, I is a homogeneous ideal and the proof is complete. □

Having established an understanding of homogeneous ideals, we now return to the primary topic of this section. The next result summarizes the most important algebraic attributes of vanishing ideals.

9.30 **PROPOSITION** *$\mathcal{I}_{\mathbb{P}}(X)$ is a homogeneous radical ideal*

If $X \subseteq \mathbb{P}^n$ is any subset, then $\mathcal{I}_{\mathbb{P}}(X) \subseteq K[x_0, x_1, \ldots, x_n]$ is a homogeneous radical ideal.

PROOF The fact that $\mathcal{I}_{\mathbb{P}}(X)$ is a radical ideal follows from the exact same argument as in the affine case (Proposition 1.24), as the reader is encouraged to verify. That $\mathcal{I}(X)$ is homogeneous follows directly from Lemma 9.17. □

As in the affine case, our primary reason for defining vanishing ideals is to have a distinguished defining ideal for any projective variety. That the vanishing ideal serves this role is verified in the third item of the next result, which is just one of a number of important properties relating the projective $\mathcal{V}$- and $\mathcal{I}$-operators.

9.31 **PROPOSITION** *Basic properties of $\mathcal{V}_{\mathbb{P}}$ and $\mathcal{I}_{\mathbb{P}}$*

Let $\mathcal{S}, \mathcal{T} \subseteq K[x_0, x_1, \ldots, x_n]$ and $X, Y \subseteq \mathbb{P}^n$ be subsets.

1. If $\mathcal{S} \subseteq \mathcal{T}$, then $\mathcal{V}_{\mathbb{P}}(\mathcal{S}) \supseteq \mathcal{V}_{\mathbb{P}}(\mathcal{T})$.
2. If $X \subseteq Y$, then $\mathcal{I}_{\mathbb{P}}(X) \supseteq \mathcal{I}_{\mathbb{P}}(Y)$.
3. $\mathcal{V}_{\mathbb{P}}(\mathcal{I}_{\mathbb{P}}(X)) \supseteq X$, with equality if and only if X is a projective variety.
4. $\mathcal{I}_{\mathbb{P}}(\mathcal{V}_{\mathbb{P}}(\mathcal{S})) \supseteq \mathcal{S}$, with equality if and only if $\mathcal{S}$ is a projective vanishing ideal.

PROOF The proofs of these statements are analogous to those of their affine counterparts (Propositions 2.1 and 1.21), as the reader is encouraged to verify. □

Continuing to parallel the affine situation, we recall that, in the affine case, the relationship between the $\mathcal{V}$- and $\mathcal{I}$-operators was leveraged to prove the existence and uniqueness of irreducible decompositions. We now state the projective analogue, starting with the natural definition of an irreducible projective variety, which carries over verbatim from the affine case.

9.32 **DEFINITION** *Irreducible projective variety*

A projective variety $X \subseteq \mathbb{P}^n$ is *reducible* if $X = X_1 \cup X_2$ for some projective varieties $X_1, X_2 \subsetneq X$, and X is *irreducible* if it is neither empty nor reducible.

As one might expect, irreducible decompositions always exist and are unique in the projective setting, and the proof of this fact is parallel to the affine situation.

9.33 PROPOSITION/DEFINITION *Irreducible decomposition*

Let $X \subseteq \mathbb{P}^n$ be a nonempty projective variety. Then there exist irreducible projective varieties $X_1, \ldots, X_r \subseteq X$ such that $X_i \not\subseteq X_j$ for any $i \neq j$ and

$$X = \bigcup_{i=1}^{r} X_i. \tag{9.34}$$

Moreover, the projective varieties $X_1, \ldots, X_r$ are unique up to reordering; we call these the *irreducible components* of X, and refer to (9.34) as the *irreducible decomposition* of X.

PROOF The proof, which uses the relationship between the projective $\mathcal{V}$- and $\mathcal{I}$-operators, along with the Noetherian property of $K[x_0, \ldots, x_n]$, is analogous to that of the affine statement (Proposition/Definition 2.32), as the reader is encouraged to verify. □

At this point, the structural parallels between projective varieties and affine varieties have begun to emerge, and indeed, many of the results in the projective case have proofs that are identical, or at least analogous, to the affine case. However, our geometric intuition for projective varieties is still lacking; after all, how can we draw pictures of projective varieties when $\mathbb{P}^n$ is so difficult to visualize, even over the real numbers, for $n \geq 2$? The key to answering this question lies in two techniques for moving between the projective and affine settings, to which we devote the next two sections.

Exercises for Section 9.3

9.3.1 Let $g \in K[x_0, x_1, \ldots, x_n]$ be any polynomial and let $h \in [x_0, x_1, \ldots, x_n]$ be homogeneous of degree d. Prove that, for each $k \geq 0$, we have

$$(gh)_k = \begin{cases} g_{d-k} \cdot h & \text{if } d \leq k \\ 0 & \text{if } d > k. \end{cases}$$

9.3.2 Which of the following ideals are homogeneous? Prove your answers.

(a) $\langle x + 1 \rangle \subseteq K[x, y]$
(b) $\langle x + y \rangle \subseteq K[x, y]$
(c) $\langle x^2 + y, x^2 - y \rangle \subseteq K[x, y]$
(d) $\langle x^2 + y^2, xy^2 + y^3 + x^2, y^2 - x^2 \rangle \subseteq K[x, y]$

9.3.3 Let $X \subseteq \mathbb{P}^n$ be a projective variety. Prove that X is irreducible if and only if $\mathcal{I}_{\mathbb{P}}(X)$ is a prime ideal.

9.3.4 Prove that a subset $X \subseteq \mathbb{P}^n$ consists of a single point if and only if $\mathcal{I}_{\mathbb{P}}(X)$ is a maximal ideal.

9.3.5 Review the proof of Proposition 1.24 to make sure that every step can be carried out in the projective setting, thereby proving that projective vanishing ideals are radical ideals.

9.3.6 Review the proofs of Propositions 2.1 and 1.21 to make sure that every step can be carried out in the projective setting, thereby proving Proposition 9.31.

9.3.7 Review the proof of Proposition/Definition 2.32 to make sure that every step can be carried out in the projective setting, thereby proving Proposition/Definition 9.33.

9.3.8 Let $X = \mathcal{V}(x_0 + x_1 - x_2) \subseteq \mathbb{P}^2$. We prove that $\mathcal{I}(X) = \langle x_0 + x_1 - x_2 \rangle$. One inclusion is immediate, as discussed in Example 9.27. For the reverse inclusion, proceed as follows:

(a) Let $f \in \mathcal{I}(X)$, and assume that f is homogeneous. Prove that for any $a_0 \neq 0$ and any a_1, we have

$$f(a_0, a_1, a_0 + a_1) = 0.$$

(b) Conclude that for any $a_0 \neq 0$, the polynomial $f(a_0, x_1, a_0 + x_1)$ is the zero polynomial in $K[x_1]$.

(c) Conclude, from here, that the polynomial $f(x_0, x_1, x_0 + x_1)$, viewed as a polynomial in the variable x_0 with coefficients in $K[x_1]$, is the zero polynomial.

(d) Argue that, since $f(x_0, x_1, x_0 + x_1) = 0$, we have

$$f \in \langle x_0 + x_1 - x_2 \rangle.$$

(e) You have now shown that every homogeneous element of $\mathcal{I}(X)$ is in $\langle x_0 + x_1 - x_2 \rangle$. Explain why it follows that, in fact, every element of $\mathcal{I}(X)$ is in $\langle x_0 + x_1 - x_2 \rangle$.

Section 9.4 Affine restrictions

In order to relate a projective variety to a more easily visualizable affine variety, we recall the second perspective on projective space from Section 9.1, wherein we view $\mathbb{P}^n$ as the result of adding points at infinity to $\mathbb{A}^n$. By ignoring these points, each projective variety has an affine restriction, and this restriction gives a helpful—though incomplete—picture of what the projective variety looks like. Before explaining the concept in general, we look at a specific example.

9.35 **EXAMPLE** Affine restriction of a quadric curve in $\mathbb{P}^2$

Consider the projective variety

$$X = \mathcal{V}_{\mathbb{P}}(x_0^2 + x_1^2 - x_2^2) \subseteq \mathbb{P}^2.$$

Any point in X whose first coordinate is nonzero can be expressed in homogenous coordinates as $[1 : a_1 : a_2]$, where $1 + a_1^2 - a_2^2 = 0$. Thus, setting

In settings involving both affine and projective varieties, we often use the notation $\mathcal{V}_{\mathbb{A}}$ and $\mathcal{I}_{\mathbb{A}}$ for the affine $\mathcal{V}$- and $\mathcal{I}$-operators to distinguish them from their projective counterparts.

$$X_0 = \mathcal{V}_{\mathbb{A}}(1 + x_1^2 - x_2^2) \subseteq \mathbb{A}^2,$$

there is a natural bijection

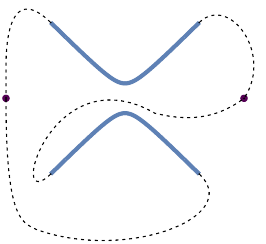
$$X = X_0 \sqcup \{[0 : a_1 : a_2] \mid a_1^2 - a_2^2 = 0\}.$$

There are two points of the second type: a point $[0 : a_1 : a_2]$ satisfying

$$a_1^2 - a_2^2 = (a_1 - a_2)(a_1 + a_2) = 0$$

must be of the form $[0 : a : a]$ or $[0 : a : -a]$, and under the equivalence relation on $\mathbb{P}^2$, this means that it is either $[0 : 1 : 1]$ or $[0 : 1 : -1]$. Thus,

$$X = X_0 \sqcup \{[0 : 1 : 1], [0 : 1 : -1]\}.$$



Working over the real numbers, we can draw the affine variety X_0: it is the hyperbola shown at right, which captures almost all of X. The two additional points in X are the points at infinity that one reaches by walking along the two asymptotes of X_0. Namely, the two asymptotes of X_0 are the lines through the origin of slope 1 and -1, which—as we saw in the discussion following Proposition 9.11—tend toward the points $[0 : 1 : 1]$ and $[0 : 1 : -1]$ in $\mathbb{P}^2$. As you can see, following the path that X makes with X_0 and the two points at infinity, we can see that X forms a single loop, just like $\mathbb{P}^1_{\mathbb{R}}$. In fact, once we have discussed isomorphisms of projective varieties, we will see that there is an isomorphism $X \cong \mathbb{P}^1$.

Generalizing Example 9.35, if $X = \mathcal{V}_{\mathbb{P}}(\mathcal{S}) \subseteq \mathbb{P}^n$ is a projective variety, then

$$X = X_0 \sqcup \{\text{points at infinity in } X\},$$

where

$$X_0 = \{[a_0 : \cdots : a_n] \in X \mid a_0 \neq 0\},$$

and the points at infinity in X are those with $a_0 = 0$. That is, $X_0 = X \cap \mathbb{A}^n$ under the natural bijection between $\mathbb{A}^n$ and points in $\mathbb{P}^n$ with nonzero first coordinate. Given that points of X_0 can be expressed in the form $[1 : b_1 : \cdots : b_n]$, we see that

$$X_0 = \mathcal{V}_{\mathbb{A}}(\mathcal{S}_0) \subseteq \mathbb{A}^n,$$

where

$$\mathcal{S}_0 = \{f(1, x_1, \ldots, x_n) \mid f \in \mathcal{S}\} \subseteq K[x_1, \ldots, x_n].$$

In particular, X_0 is an affine variety, called the *affine restriction* of X. Before giving a general definition of affine restrictions, we discuss one more class of examples.

9.36 **EXAMPLE** Restricting lines in $\mathbb{P}^2$

The affine restriction of the projective variety $X = \mathcal{V}_{\mathbb{P}}(x_0 + x_1 - x_2) \subseteq \mathbb{P}^2$ from Example 9.22 is the line $X_0 = \mathcal{V}_{\mathbb{A}}(1 + x_1 - x_2) \subseteq \mathbb{A}^2$, and the only point at infinity of X is $[0 : 1 : 1]$. More generally, the affine restriction of the projective variety

$$L = \mathcal{V}_{\mathbb{P}}(bx_0 + mx_1 - x_2) \subseteq \mathbb{P}^2.$$

is

$$L_0 = \mathcal{V}_{\mathbb{A}}(b + mx_1 - x_2) \subseteq \mathbb{A}^2,$$

which is a line with vertical intercept b and slope m. Some moments reflecting should convince the reader that L again contains just one point at infinity: $[0 : 1 : m]$. This is a more precise manifestation of what we saw informally in Section 9.1: the point $[0 : 1 : m]$ is the point at infinity that one reaches by "walking along L_0."

In particular, we see again that the point at infinity reached by walking along L_0 only depends on the slope of L_0. Parallel lines, then, such as

$$\mathcal{V}_{\mathbb{A}}(1 + 3x_1 - x_2),\ \mathcal{V}_{\mathbb{A}}(2 + 3x_1 - x_2) \subseteq \mathbb{A}^2,$$

do not meet in $\mathbb{A}^2$, yet when viewed as the affine restrictions of the projective lines

$$\mathcal{V}_{\mathbb{P}}(x_0 + 3x_1 - x_2),\ \mathcal{V}_{\mathbb{P}}(2x_0 + 3x_1 - x_2) \subseteq \mathbb{P}^2,$$

they meet at the point at infinity $[0 : 1 : 3]$, dictated by their common slope. For this reason, $\mathbb{P}^2$ is sometimes referred to as the setting in which "parallel lines meet." See Exercise 9.4.1 for a more complete exploration of this phenomenon.

The role played by x_0 in the above discussion, as opposed to any other variable, is arbitrary. More generally, restricting a projective variety to the points of $\mathbb{P}^n$ with nonzero ith coordinate yields an affine variety, described in the following definition.

9.37 **DEFINITION** *Affine patches and affine restrictions*

For each $i \in \{0, 1, \ldots, n\}$, the *ith affine patch* of $\mathbb{P}^n$ is the set

$$\mathbb{A}^n_i = \{[a_0 : a_1 : \cdots : a_n] \in \mathbb{P}^n \mid a_i \neq 0\},$$

and for any set $X \subseteq \mathbb{P}^n$, the intersection $X \cap \mathbb{A}^n_i$ is called the *ith affine restriction* of X.

The reader should pause to convince themselves (Exercise 9.4.2) that for any i, there is a natural bijection $\mathbb{A}^n_i = \mathbb{A}^n$, and under this bijection, the ith affine restriction of $X = \mathcal{V}_{\mathbb{P}}(\mathcal{S})$ is $\mathcal{V}_{\mathbb{A}}(\mathcal{S}_i) \subseteq \mathbb{A}^n$, where

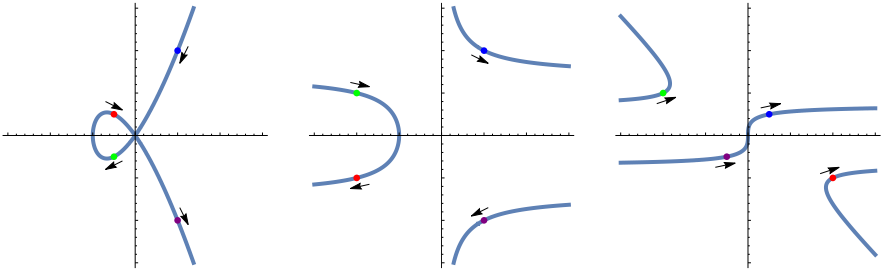
$$\mathcal{S}_i = \{f(x_0, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n) \mid f \in \mathcal{S}\}.$$

9.38 **EXAMPLE** Affine restrictions of a cubic curve in $\mathbb{P}^2$

Consider the three affine restrictions of $X = \mathcal{V}_{\mathbb{P}}(x_0x_2^2 - 2x_1^3 - 2x_0x_1^2) \subseteq \mathbb{P}^2$:

$$X_0 = \mathcal{V}_{\mathbb{A}}(x_2^2 - 2x_1^3 - 2x_1^2) \subseteq \mathbb{A}^2,$$
$$X_1 = \mathcal{V}_{\mathbb{A}}(x_0x_2^2 - 2 - 2x_0) \subseteq \mathbb{A}^2,$$
$$X_2 = \mathcal{V}_{\mathbb{A}}(x_0 - 2x_1^3 - 2x_1^2x_0) \subseteq \mathbb{A}^2.$$

The full projective variety X is the union of these three subsets, which intersect in points with more than one nonzero coordinate. Thus, one can construct X by "gluing together"—with substantial overlap—three affine varieties. For example, the point $[1 : 1 : 2] \in X$ can be found in each of the three affine patches: in X_0, it has affine coordinates $(x_1, x_2) = (1, 2)$; in X_1, it has affine coordinates $(x_0, x_2) = (1, 2)$; and in X_2, it has affine coordinates $(x_0, x_1) = (\frac{1}{2}, \frac{1}{2})$. In the images below, we have depicted the three affine patches over the real numbers, marking four color-coded points on each that are identified within X. The arrows suggest the orientation in which these affine patches are glued together.



By carefully tracing the curve, moving between affine patches when necessary, one sees that, when viewed over the real numbers, X forms a "figure-eight." While the first affine patch is missing only a single point (the point at infinity on a vertical line), the second and third are each missing two points, one of which (the point at infinity on a horizontal line) is the point at which the figure-eight crosses itself.

The passage from a projective variety X to its affine restrictions involves simply setting one of the coordinates equal to 1, but can we reverse this procedure? Namely, starting from an affine variety X, can one find a projective variety $\overline{X}$ such that X is one of the affine restrictions of $\overline{X}$? The answer is yes: the associated projective variety is called the *projective closure* of X—the topic of the next section.

Exercises for Section 9.4

9.4.1 A *line* in $\mathbb{P}^2$ is defined as a projective variety of the form

$$\mathcal{V}_{\mathbb{P}}(ax_0 + bx_1 + cx_2) \subseteq \mathbb{P}^2$$

where $a, b, c \in K$ are not all zero. For this exercise, let X and Y be a pair of *distinct* lines in $\mathbb{P}^2$.

(a) Prove that $X \cap Y$ consists of a single point.

(b) Suppose that the first affine restrictions X_0 and Y_0 are nonempty. Prove that X_0 and Y_0 are lines in $\mathbb{A}^2$.

(c) Prove that $X \cap Y = X_0 \cap Y_0$ whenever X_0 and Y_0 are not parallel.

(d) Now suppose that X_0 and Y_0 are parallel lines. What can we say about the defining equations of X and Y? In what point do X and Y intersect?

This clarifies the fact that parallel lines in $\mathbb{A}^2$ meet "at infinity" in $\mathbb{P}^2$.

9.4.2 (a) Prove that there is a natural bijection between the affine patch $\mathbb{A}^n_i \subseteq \mathbb{P}^n$ and $\mathbb{A}^n$.

(b) If $X = \mathcal{V}_{\mathbb{P}}(\mathcal{S}) \subseteq \mathbb{P}^n$, prove that the bijection in (a) identifies $X \cap \mathbb{A}^n_i$ with $\mathcal{V}_{\mathbb{A}}(\mathcal{S}_i) \subseteq \mathbb{A}^n$, where

$$\mathcal{S}_i = \{f(x_0, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n) \mid f \in \mathcal{S}\}.$$

9.4.3 Let $X = \mathcal{V}_{\mathbb{P}}(x_0x_2 - x_1^2) \subseteq \mathbb{P}^2$.

(a) Calculate the three affine restrictions X_0, X_1, and X_2, and draw a picture of each over the real numbers.

(b) Consider the point $(2, 4) \in X_0 \subseteq \mathbb{A}^2$. As an element of X, this is the point $[1 : 2 : 4]$, which also lies in X_1. What are the coordinates of this point in $X_1 \subseteq \mathbb{A}^2$?

(c) Repeat the reasoning of part (b) for several other points in $X_0 \cap X_1$ to illustrate, visually, how X_0 and X_1 fit together inside X. Then, do the same for X_1 and X_2 and for X_0 and X_2.

Section 9.5 Projective closures

In the previous section, we learned how to view a projective variety as the disjoint union of an affine variety and a collection of points at infinity, and in this section, we reverse that process, describing a method for producing a projective variety by adding points at infinity to an affine variety. Let $j_0 : \mathbb{A}^n \to \mathbb{P}^n$ be the function

$$j_0(a_1, \ldots, a_n) = [1 : a_1 : \cdots : a_n],$$

which is a bijection onto the affine patch $\mathbb{A}^n_0 \subseteq \mathbb{P}^n$. If $X \subseteq \mathbb{A}^n$ is an affine variety, then $j_0(X) \subseteq \mathbb{P}^n$ is a subset of $\mathbb{P}^n$ whose first affine restriction is X by construction, but $j_0(X)$ is not, in general, a projective variety. In order to extend $j_0(X)$ to a projective variety, one must add some additional points; the minimal projective variety obtained in this way is called the *projective closure* of X.

9.39 **DEFINITION** *Projective closure*

Let $X \subseteq \mathbb{A}^n$ be an affine variety, and let $j_0(X)$ be the image of X in the first affine patch $\mathbb{A}^n_0 \subseteq \mathbb{P}^n$. The *projective closure* of X, denoted $\overline{X} \subseteq \mathbb{P}^n$, is the intersection of all projective varieties that contain $j_0(X)$.

That $\overline{X}$ is, itself, a projective variety follows from the fact that (even infinite) intersections of projective varieties are projective varieties (Exercise 9.2.4). By definition, we see that $\overline{X}$ is the "smallest" projective variety containing $j_0(X)$; in other words, it is contained within every other projective variety containing X.

9.40 **EXAMPLE** Projective closure of a line

If $X = \mathcal{V}_{\mathbb{A}}(1 + x_1 - x_2) \subseteq \mathbb{A}^2$, then

$$j_0(X) = \{[1 : a : 1 + a] \mid a \in K\} \subseteq \mathbb{P}^2.$$

In particular, since

$$[1 : a : 1 + a] = [\tfrac{1}{a} : 1 : \tfrac{1}{a} + 1]$$

when $a \neq 0$, we see that $j_0(X)$ contains $[b : 1 : b + 1]$ for any $b \in K \setminus \{0\}$. But a projective variety that contains all of these points must also contain the corresponding point with $b = 0$; see Exercise 9.5.1. Thus, given that $[0 : 1 : 1] \notin j_0(X)$, we see that $j_0(X)$ cannot be a projective variety.

Adding this one missing point yields a projective variety:

$$\mathcal{V}_{\mathbb{P}}(x_0 + x_1 - x_2) = j_0(X) \sqcup \{[0 : 1 : 1]\},$$

as one verifies by splitting into cases depending on whether the first coordinate is zero or nonzero. Thus, $\mathcal{V}_{\mathbb{P}}(x_0 + x_1 - x_2)$ is the smallest projective variety containing $j_0(X)$, implying that $\overline{X} = \mathcal{V}_{\mathbb{P}}(x_0 + x_1 - x_2)$.

9.41 **EXAMPLE** Projective closure of a parabola

Consider the parabola

$$X = \mathcal{V}_{\mathbb{A}}(x_2 - x_1^2) \subseteq \mathbb{A}^2.$$

Then

$$j_0(X) = \{[1 : a : a^2] \mid a \in K\} \subseteq \mathbb{P}^2.$$

As in the previous example, from the fact that

$$[1 : a : a^2] = [\tfrac{1}{a^2} : \tfrac{1}{a} : 1]$$

for $a \neq 0$, we see that $j_0(X)$ contains $[b^2 : b : 1]$ for any $b \in K \setminus \{0\}$. But again, a projective variety containing all of these points must contain the point $[0 : 0 : 1]$; since $j_0(X)$ does not contain this point, it cannot be a projective variety.

By adding the missing point, we find a projective variety:

$$\mathcal{V}_{\mathbb{P}}(x_0x_2 - x_1^2) = j_0(X) \sqcup \{[0 : 0 : 1]\},$$

and we conclude that $\overline{X} = \mathcal{V}_{\mathbb{P}}(x_0x_2 - x_1^2)$.

In each of the previous two examples, the affine variety X was defined by an inhomogeneous polynomial f, and the defining polynomial of $\overline{X}$ could be obtained from f by "homogenizing": multiplying each term of f by a power of x_0 to produce a homogeneous polynomial. To illustrate the idea in another example, let

$$f(x_1, x_2, x_3) = x_1^2 + x_2 + x_1x_3^4.$$

Then the highest-degree term is the last one, which has degree 5, and we homogenize f by multiplying each term by the necessary power of x_0 to give it degree 5. The result is the homogeneous polynomial

$$\overline{f}(x_0, x_1, x_2, x_3) = x_0^3x_2^2 + x_0^4x_2 + x_1x_3^4.$$

The following definition describes this procedure in general.

9.42 **DEFINITION** *Homogenization of a polynomial*

Let $f \in K[x_1, \ldots, x_n]$ be a polynomial of degree d. The *homogenization* of f is defined by

$$\overline{f} = x_0^d \cdot f\Big(\frac{x_1}{x_0}, \ldots, \frac{x_n}{x_0}\Big) \in K[x_0, x_1, \ldots, x_n].$$

The reader should verify (Exercise 9.5.3) that $\overline{f}$ is a homogeneous polynomial of degree d satisfying

$$\overline{f}(1, x_1, \ldots, x_n) = f(x_1, \ldots, x_n), \tag{9.43}$$

and that this definition agrees with the term-by-term procedure described above.

In Examples 9.40 and 9.41, we had $X = \mathcal{V}_{\mathbb{A}}(f)$ and $\overline{X} = \mathcal{V}_{\mathbb{P}}(\overline{f})$, which might lead one to postulate that if $X = \mathcal{V}_{\mathbb{A}}(f_1, \ldots, f_r)$, then $\overline{X} = \mathcal{V}_{\mathbb{P}}(\overline{f}_1, \ldots, \overline{f}_r)$. This would be convenient if it were the case; not only would it make computing projective closures a straightforward process, but it would imply, via equation (9.43), that the first affine restriction of $\overline{X}$ is X. Unfortunately, the passage from X to $\overline{X}$ is not always quite so simple, as the next example illustrates.

9.44 **EXAMPLE** The twisted cubic curve

Let $X = \mathcal{V}_{\mathbb{A}}(x_2 - x_1^2, x_3 - x_1^3)$ so that

$$j_0(X) = \{[1 : a : a^2 : a^3] \mid a \in K\} \subseteq \mathbb{P}^3.$$

Similarly to Examples 9.40 and 9.41, any projective variety containing $j_0(X)$ must also contain the point $[0 : 0 : 0 : 1]$. It follows that $j_0(X)$ is not a projective variety, but direct computation (Exercise 9.5.6) shows that $j_0(X) \sqcup \{[0 : 0 : 0 : 1]\}$ is:

$$\overline{X} = \mathcal{V}_{\mathbb{P}}(x_0x_2 - x_1^2, x_0x_3 - x_1^3, x_1x_3 - x_2^2) = j_0(X) \cup \{[0 : 0 : 0 : 1]\}.$$

This projective variety is called the *twisted cubic curve*.

The first two defining polynomials of $\overline{X}$ are obtained by homogenizing the defining polynomials of X, but the third is also necessary. Without it,

$$\mathcal{V}_{\mathbb{P}}(x_0x_2 - x_1^2,\ x_0x_3 - x_1^3) \tag{9.45}$$

contains $j_0(X)$, but also contains *all* of the points $[0 : 0 : b : c] \in \mathbb{P}^3$. Thus, while (9.45) is a projective variety containing $j_0(X)$, it is quite a bit larger than $\overline{X}$.

While Example 9.44 shows that the projective closure of $X = \mathcal{V}_{\mathbb{A}}(f_1, \ldots, f_r)$ is not, in general, obtained by homogenizing $f_1, \ldots, f_r$, there is a fix: instead of homogenizing only an arbitrarily chosen set of defining polynomials, we should homogenize *every* polynomial in the vanishing ideal of X.

9.46 **DEFINITION** *Homogenization of an ideal*

Let $I \subseteq K[x_1, \ldots, x_n]$ be an ideal. The *homogenization* of I is the set

$$\overline{I} = \{\overline{f} \mid f \in I\} \subseteq K[x_0, \ldots, x_n].$$

Careful: $\overline{I}$ is not generally an ideal; can you see why?

By homogenizing every polynomial in the vanishing ideal of an affine variety, we then obtain enough polynomials to describe its projective closure.

9.47 **PROPOSITION** *Projective closures via homogenization*

If X is any affine variety, then $\overline{X} = \mathcal{V}_{\mathbb{P}}\big(\overline{\mathcal{I}_{\mathbb{A}}(X)}\big)$.

PROOF We prove both inclusions.

($\subseteq$) Note that

$$j_0(X) \subseteq \mathcal{V}_{\mathbb{P}}\big(\overline{\mathcal{I}_{\mathbb{A}}(X)}\big),$$

since if $[1 : a_1 : \cdots : a_n] \in j_0(X)$ and $g \in \overline{\mathcal{I}_{\mathbb{A}}(X)}$, then $g = \overline{f}$ for some $f \in \mathcal{I}_{\mathbb{A}}(X)$ and hence $g(1, a_1, \ldots, a_n) = f(a_1, \ldots, a_n) = 0$. Thus, $\mathcal{V}_{\mathbb{P}}\big(\overline{\mathcal{I}_{\mathbb{A}}(X)}\big)$

is a projective variety containing $j_0(X)$, and since $\overline{X}$ is the smallest projective variety containing $j_0(X)$, it follows that $\overline{X} \subseteq \mathcal{V}_{\mathbb{P}}(\overline{\mathcal{I}_{\mathbb{A}}(X)})$.

$(\supseteq)$ We prove that

$$\mathcal{I}_{\mathbb{P}}(\overline{X}) \subseteq \mathcal{I}_{\mathbb{P}}\left(\mathcal{V}_{\mathbb{P}}(\overline{\mathcal{I}_{\mathbb{A}}(X)})\right), \tag{9.48}$$

then applying $\mathcal{V}_{\mathbb{P}}$ to both sides of this containment implies, by Proposition 9.31, that

$$\overline{X} \supseteq \mathcal{V}_{\mathbb{P}}\left(\overline{\mathcal{I}_{\mathbb{A}}(X)}\right),$$

as required. To prove (9.48), let $g \in \mathcal{I}_{\mathbb{P}}(\overline{X})$. Given that $\mathcal{I}_{\mathbb{P}}(\overline{X})$ is a homogeneous ideal and thus admits a set of homogeneous generators, it suffices to assume for the proof of (9.48) that g is homogeneous.

Even though g is homogeneous, it may not be the homogenization of an element of $K[x_1,\ldots,x_n]$, since it could be the case that every term of g contains x_0. However, if $k \in \mathbb{Z}_{\geq 0}$ is the maximum power of x_0 such that $g = x_0^k h$ for some $h \in K[x_0,x_1,\ldots,x_n]$, then Exercise 9.5.8 implies that h is the homogenization of the polynomial $h_0 = h(1,x_1,\ldots,x_n) \in K[x_1,\ldots,x_n]$. Using that g vanishes on $\overline{X} \supseteq j_0(X)$, we see that h_0 vanishes on X: for any $(a_1,\ldots,a_n) \in X$, we have

$$h_0(a_1,\ldots,a_n) = h(1,a_1,\ldots,a_n) = g(1,a_1,\ldots,a_n) = 0.$$

Thus, $h(1,x_1,\ldots,x_n) \in \mathcal{I}_{\mathbb{A}}(X)$, so $h \in \overline{\mathcal{I}_{\mathbb{A}}(X)}$, and it follows that

$$g \in \langle \overline{\mathcal{I}_{\mathbb{A}}(X)} \rangle \subseteq \mathcal{I}_{\mathbb{P}}\left(\mathcal{V}_{\mathbb{P}}\left(\overline{\mathcal{I}_{\mathbb{A}}(X)}\right)\right),$$

where the containment is another application of Proposition 9.31. This completes the proof of (9.48) and hence the proof of the proposition. □

Proposition 9.47 is a statement about vanishing ideals, but combining it with the Nullstellensatz yields a result on defining ideals, which can be useful in practice.

9.49 COROLLARY *Projective closures via defining ideals*

If $I \subseteq K[x_1,\ldots,x_n]$ is an ideal, then $\overline{\mathcal{V}_{\mathbb{A}}(I)} = \mathcal{V}_{\mathbb{P}}(\overline{I})$.

PROOF Let $I \subseteq K[x_1,\ldots,x_n]$ be any ideal. By Proposition 9.47 and the Nullstellensatz, we have

$$\overline{\mathcal{V}_{\mathbb{A}}(I)} = \mathcal{V}_{\mathbb{P}}\left(\overline{\mathcal{I}_{\mathbb{A}}(\mathcal{V}_{\mathbb{A}}(I))}\right) = \mathcal{V}_{\mathbb{P}}\left(\overline{\sqrt{I}}\right). \tag{9.50}$$

From here, Exercise 9.5.9 shows that

$$\left\langle \overline{\sqrt{I}} \right\rangle = \sqrt{\langle \overline{I} \rangle},$$

so the equalities in (9.50) can be continued with

$$V_{\mathbb{P}}\left(\sqrt{\langle \overline{I} \rangle}\right) = \mathcal{V}_{\mathbb{P}}(\langle \overline{I} \rangle) = \mathcal{V}_{\mathbb{P}}(\overline{I}),$$

in which the first equality is an application of Exercise 9.5.10. □

9.51 EXAMPLE The twisted cubic revisited

In light of Corollary 9.49, we can make further sense of the phenomenon observed in Example 9.44. In that case, $X = \mathcal{V}_{\mathbb{A}}(I)$ for the ideal

$$I = \mathcal{V}(x_2 - x_1^2,\ x_3 - x_1^3).$$

The homogenization of I contains the homogenizations of $x_2 - x_1^2$ and $x_3 - x_1^3$, meaning

$$\overline{I} \supseteq \langle x_0x_2 - x_1^2,\ x_0x_3 - x_1^3\rangle.$$

But the containment is strict: to see why, notice that there is a homogeneous polynomial in I itself, namely

$$x_1x_3 - x_2^2 = (x_2 - x_1^2)\cdot(-x_2 - x_1^2) + (x_3 - x_1^3)\cdot x_1.$$

This polynomial is therefore in $\overline{I}$, but it cannot be in $\langle x_0x_2 - x_1^2,\ x_0x_3 - x_1^3\rangle$ because any element of this ideal vanishes at $[0:0:1:0]$ and $x_1x_3 - x_2^2$ does not. In fact, one can show that

$$\overline{I} = \langle x_0x_2 - x_1^2, x_0x_3 - x_1^3, x_1x_3 - x_2^2\rangle,$$

at which point Proposition 9.47 again gives the result

$$\overline{X} = \mathcal{V}_{\mathbb{P}}(x_0x_2 - x_1^2,\ x_0x_3 - x_1^3,\ x_1x_3 - x_2^2).$$

This example highlights the fact that, in general,

$$\overline{\langle f_1,\dots,f_k\rangle} \neq \langle \overline{f_1},\dots,\overline{f_k}\rangle.$$

Fortunately, however, this issue does not arise when the ideal is principal: one has

$$\overline{\langle f\rangle} = \langle\, \overline{f}\,\rangle \tag{9.52}$$

for any $f \in K[x_1,\dots,x_n]$; see Exercise 9.5.11. Combining (9.52) with Corollary 9.49 gives the following useful result.

9.53 COROLLARY *Projective closures of hypersurfaces*

If $f \in K[x_1,\dots,x_n]$, then $\overline{\mathcal{V}_{\mathbb{A}}(f)} = \mathcal{V}_{\mathbb{P}}(\overline{f})$.

A final corollary of Proposition 9.47 is that taking affine restrictions is, in some sense, the inverse of taking projective closures, as stated in the next result.

9.54 COROLLARY $(\overline{X})_0 = X$

If X is an affine variety, then the affine restriction of the projective closure of X is X. More succinctly, $(\overline{X})_0 = X$.

PROOF Let $\mathcal{S} = \overline{\mathcal{I}_{\mathbb{A}}(X)}$. Combining Proposition 9.47 with Definition 9.37, we see that the affine restriction of $\overline{X}$ is $\mathcal{V}_{\mathbb{A}}(\mathcal{S}_0)$, where

$$\begin{aligned}\mathcal{S}_0 &= \{g(1, x_1, \ldots, x_n) \mid g \in \overline{\mathcal{I}_{\mathbb{A}}(X)}\} \\ &= \{\overline{f}(1, x_1, \ldots, x_n) \mid f \in \mathcal{I}_{\mathbb{A}}(X)\} \\ &= \mathcal{I}_{\mathbb{A}}(X).\end{aligned}$$

Thus, the affine restriction of $\overline{X}$ is $\mathcal{V}_{\mathbb{A}}(\mathcal{I}_{\mathbb{A}}(X)) = X$, as claimed. □

Affine restrictions and projective closures set up a close connection between affine varieties and projective varieties, and as we have seen, the theory in the two settings is essentially parallel. The reader may expect, then, that there is a projective version of the Nullstellensatz. This is indeed the case, and the last section of this chapter is devoted to developing it.

Exercises for Section 9.5

9.5.1 Let $X \subseteq \mathbb{P}^2$ be a projective variety containing the points $[b : 1 : b+1]$ for all $b \in K \setminus \{0\}$. Prove that $[0 : 1 : 1] \in X$.

9.5.2 Generalizing the previous exercise, let $f_0, \ldots, f_n \in K[x]$ be single-variable polynomials, not all zero, and let $X \subseteq \mathbb{P}^n$ be a projective variety such that

$$[f_0(b) : \cdots : f_n(b)] \in X$$

for all $b \in K \setminus \{0\}$. Prove that

$$[f_0(0) : \cdots : f_n(0)] \in X.$$

9.5.3 Prove that, for any polynomial $f \in [x_1, \ldots, x_n]$ of degree d, the homogenization $\overline{f}$ is a homogeneous polynomial of degree d satisfying

$$\overline{f}(1, x_1, \ldots, x_n) = f(x_1, \ldots, x_n).$$

9.5.4 Let $f, g \in K[x_1, \ldots, x_n]$.

(a) Prove that $\overline{fg} = \overline{f} \cdot \overline{g}$.

(b) Is it the case that $\overline{f+g} = \overline{f} + \overline{g}$? Prove or give a counterexample.

9.5.5 Mimic the arguments of Examples 9.40 and 9.41 to prove that the projective closure of the affine variety

$$X = \mathcal{V}_{\mathbb{A}}(x_1 x_2 - 1) \subseteq \mathbb{A}^2$$

must contain the points $[0 : 0 : 1]$ and $[0 : 1 : 0]$, and conclude that

$$\overline{X} = \mathcal{V}_{\mathbb{P}}(x_1 x_2 - x_0^2).$$

What are the asymptotes of X, and how does your answer relate to the points at infinity of $\overline{X}$?

9.5.6 Let

$$X = \mathcal{V}_{\mathbb{A}}(x_2 - x_1^2, x_3 - x_1^3) \subseteq \mathbb{A}^3$$

be the affine twisted cubic. Prove that

$$\overline{X} = \mathcal{V}_{\mathbb{P}}(x_0x_2 - x_1^2, x_0x_3 - x_1^3, x_1x_3 - x_2^2).$$

9.5.7 Calculate the projective closure of the affine variety

$$X = \mathcal{V}(x - 1, y) \subseteq \mathbb{A}^2.$$

Prove your answer.

9.5.8 Let $h \in K[x_0, x_1, \ldots, x_n]$. Prove that if $x_0 \nmid h$, then h is the homogenization of

$$h(1, x_1, \ldots, x_n) \in K[x_1, \ldots, x_n].$$

9.5.9 Let $I \subseteq K[x_1, \ldots, x_n]$ be any ideal. Prove that

$$\langle \overline{\sqrt{I}} \rangle = \sqrt{\langle \overline{I} \rangle}\,.$$

9.5.10 Let $I \subseteq K[x_1, \ldots, x_n]$ be a homogeneous ideal. Prove that $\sqrt{I}$ is also a homogeneous ideal, and that

$$\mathcal{V}_{\mathbb{P}}(I) = \mathcal{V}_{\mathbb{P}}\left(\sqrt{I}\right).$$

9.5.11 Prove that

$$\overline{\langle f \rangle} = \langle \overline{f} \rangle$$

for any $f \in K[x_1, \ldots, x_n]$. (**Hint**: Use Exercise 9.5.4.)

Section 9.6 The projective Nullstellensatz

How might one generalize the statement of the Nullstellensatz to the projective setting? The most straightforward generalization one might hope for is that

$$\mathcal{I}_{\mathbb{P}}(\mathcal{V}_{\mathbb{P}}(I)) = \sqrt{I} \tag{9.55}$$

for any ideal $I \subseteq K[x_0, x_1, \ldots, x_n]$. Unfortunately, this is not quite correct. First, since vanishing ideals are homogeneous, (9.55) can only be true if I is a homogeneous ideal. But even among homogeneous ideals, there is a case when (9.55) fails:

$$I = \langle x_0, x_1, \ldots, x_n \rangle.$$

More explicitly, if $I = \langle x_0, x_1, \ldots, x_n \rangle$, notice that $\mathcal{V}_{\mathbb{P}}(I)$ consists of all points $[a_0 : a_1 : \cdots : a_n] \in \mathbb{P}^n$ at which each of the polynomials $x_0, x_1, \ldots, x_n$ vanishes, which can only be the case if $a_i = 0$ for each i. Since no such points exist in $\mathbb{P}^n$, it follows that $\mathcal{V}_{\mathbb{P}}(I) = \emptyset$. Every polynomial vacuously vanishes at every point of $\emptyset$, implying that $\mathcal{I}_{\mathbb{P}}(\mathcal{V}_{\mathbb{P}}(I)) = K[x_0, x_1, \ldots, x_n]$. On the other hand, $I = \langle x_0, x_1, \ldots, x_n \rangle$ is a radical (in fact, maximal) ideal, so $\sqrt{I} = \langle x_0, x_1, \ldots, x_n \rangle$. Tying together these observations, we conclude that

$$\mathcal{I}_{\mathbb{P}}(\mathcal{V}_{\mathbb{P}}(I)) = K[x_0, \ldots, x_n] \neq \langle x_0, \ldots, x_n \rangle = \sqrt{I}.$$

To avoid this pesky exception to (9.55), we give the ideal $I = \langle x_0, x_1, \ldots, x_n \rangle$ a name that emphasizes our unwillingness to consider it.

9.56 **DEFINITION** *Irrelevant ideal*

The ideal $\langle x_0, x_1, \ldots, x_n \rangle \subseteq K[x_0, x_1, \ldots, x_n]$ is called the *irrelevant ideal*. An ideal that is not the irrelevant ideal is called *relevant*.

The projective Nullstellensatz asserts that, aside from the irrelevant ideal, the affine Nullstellensatz directly generalizes.

9.57 **THEOREM** *Projective Nullstellensatz*

Let $I \subseteq K[x_0, x_1, \ldots, x_n]$ be a homogeneous ideal. If $\sqrt{I}$ is relevant, then

$$\mathcal{I}_{\mathbb{P}}(\mathcal{V}_{\mathbb{P}}(I)) = \sqrt{I}.$$

If $\sqrt{I}$ is the irrelevant ideal, then $\mathcal{I}_{\mathbb{P}}(\mathcal{V}_{\mathbb{P}}(I)) = K[x_0, x_1 \ldots, x_n]$.

Fortunately, the projective Nullstellensatz can be derived as a consequence of its affine cousin without re-developing the algebraic machinery. The key idea is to leverage a different relationship between affine and projective geometry than the ones we considered in the previous two sections.

Notice that a homogeneous ideal $I \subseteq K[x_0, x_1, \ldots, x_n]$ can be used in two different ways. On one hand, it defines a projective variety $\mathcal{V}_{\mathbb{P}}(I) \subseteq \mathbb{P}^n$. But on the other hand, before learning about projective varieties, we would have simply viewed I as an ideal in $n+1$ variables, defining an affine variety $\mathcal{V}_{\mathbb{A}}(I) \subseteq \mathbb{A}^{n+1}$. What is the relationship between $\mathcal{V}_{\mathbb{P}}(I)$ and $\mathcal{V}_{\mathbb{A}}(I)$? Let us consider an example.

9.58 EXAMPLE $\mathcal{V}_{\mathbb{P}}(I)$ versus $\mathcal{V}_{\mathbb{A}}(I)$

Consider $K = \mathbb{R}$ and let $I = \langle -x_0^2 + x_1^2 + x_2^2 \rangle \subseteq \mathbb{R}[x_0, x_1, x_2]$. Then

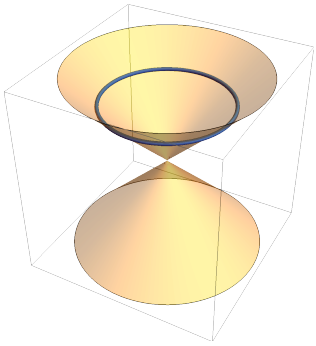
$$\mathcal{V}_{\mathbb{P}}(I) = \{[a_0 : a_1 : a_2] \in \mathbb{P}^2 \mid -a_0^2 + a_1^2 + a_2^2 = 0\}.$$

If $a_0 \neq 0$, then by rescaling all three coordinates we can assume that $a_0 = 1$. If, on the other hand, $a_0 = 0$, then the defining equation becomes $a_1^2 + a_2^2 = 0$, implying that $a_1 = a_2 = 0$. Since no such point with $a_0 = a_1 = a_2 = 0$ exists in $\mathbb{P}^2$, we conclude that

$$\mathcal{V}_{\mathbb{P}}(I) = \{[1 : a_1 : a_2] \in \mathbb{P}^2 \mid -1 + a_1^2 + a_2^2 = 0\},$$

which is in bijection with the unit circle in $\mathbb{A}^2$—that is, with $\mathcal{V}_{\mathbb{A}}(-1 + x_1^2 + x_2^2)$.

Alternatively viewing I as defining an affine variety in $\mathbb{A}^3$, we see that $\mathcal{V}_{\mathbb{A}}(I)$ is the circular cone with vertex at the origin depicted at right. (In the image, x_0 is the vertical coordinate). The projective variety $\mathcal{V}_{\mathbb{P}}(I)$ is visible in this image as the intersection of $\mathcal{V}_{\mathbb{A}}(I)$ with the plane $x_0 = 1$. Further, we note that $\mathcal{V}_{\mathbb{A}}(I)$ is the union of all lines through the origin that correspond, under the bijection of Corollary 9.11, to points of $\mathcal{V}_{\mathbb{P}}(I)$.



The fact that $\mathcal{V}(I)$ is a cone in the previous example is no accident: a cone $C \subseteq \mathbb{A}^3$ has the property that whenever $(a_0, a_1, a_2) \in C$, the entire line

$$\{(\lambda a_0, \lambda a_1, \lambda a_2) \mid \lambda \in K\}$$

is contained in C; from the projective perspective, this is the statement that membership of $[a_0 : a_1 : a_2]$ in C is well-defined. To make this observation precise in general, we introduce the language of *affine cones*.

9.59 DEFINITION *Affine cone*

Let $X \subseteq \mathbb{P}^n$ be any subset. The *affine cone* over X is the set

$$C(X) = \{(0, \ldots, 0)\} \cup \{(a_0, \ldots, a_n) \mid [a_0 : \cdots : a_n] \in X\} \subseteq \mathbb{A}^{n+1}.$$

For instance, the affine cone over $\mathcal{V}_{\mathbb{P}}(I) \subseteq \mathbb{P}^2$ in Example 9.58 is equal to $\mathcal{V}_{\mathbb{A}}(I) \subseteq \mathbb{A}^3$. This is a special case of the following lemma.

9.60 LEMMA $\mathcal{V}_{\mathbb{A}}(I) = C(\mathcal{V}_{\mathbb{P}}(I))$

If $I \subsetneq K[x_0, x_1, \ldots, x_n]$ is a proper homogeneous ideal, then

$$\mathcal{V}_{\mathbb{A}}(I) = C\big(\mathcal{V}_{\mathbb{P}}(I)\big).$$

PROOF First, note that $(0,\dots,0) \in \mathcal{V}_{\mathbb{A}}(I)$ for every proper homogeneous ideal $I \subsetneq K[x_0,\dots,x_n]$. To see this, recall from Proposition 9.29 that I has a set of homogeneous generators. If $(0,\dots,0) \notin \mathcal{V}_{\mathbb{A}}(I)$, then at least one of these generators must be a homogeneous polynomial f such that $f(0,\dots,0) \neq 0$. But the only homogeneous polynomials that do not vanish at $(0,\dots,0)$ are the nonzero constant polynomials, and if such a polynomial is among the generators of I, then $I = K[x_0,x_1,\dots,x_n]$, contradicting the assumption that I is a proper ideal.

Thus, $(0,\dots,0) \in \mathcal{V}_{\mathbb{A}}(I)$. Moreover, if at least one coordinate of $(a_0,\dots,a_n)$ is nonzero, then it follows from the definitions of the $\mathcal{V}$-operators that

$$(a_0,\dots,a_n) \in \mathcal{V}_{\mathbb{A}}(I) \Leftrightarrow [a_0 : \cdots : a_n] \in \mathcal{V}_{\mathbb{P}}(I) \Leftrightarrow (a_0,\dots,a_n) \in C(\mathcal{V}_{\mathbb{P}}(I)).$$

This implies that $\mathcal{V}_{\mathbb{A}}(I) = C(\mathcal{V}_{\mathbb{P}}(I))$, as claimed. □

We have now used affine cones to relate the $\mathcal{V}_{\mathbb{A}}$- and $\mathcal{V}_{\mathbb{P}}$-operators, and they can analogously be used to relate the $\mathcal{I}_{\mathbb{A}}$- and $\mathcal{I}_{\mathbb{P}}$-operators, as the next lemma shows.

9.61 LEMMA $\mathcal{I}_{\mathbb{P}}(X) = \mathcal{I}_{\mathbb{A}}(C(X))$

If $X \subseteq \mathbb{P}^n$ is a nonempty subset, then

$$\mathcal{I}_{\mathbb{P}}(X) = \mathcal{I}_{\mathbb{A}}\big(C(X)\big).$$

The exceptions to Lemma 9.60 and 9.61, when $I = K[x_0,\dots,x_n]$ in the first case or $X = \emptyset$ in the second, are necessary; see Exercise 9.6.2.

PROOF By Propositions 9.30 and 9.29, there exists a set of homogeneous generators of $\mathcal{I}_{\mathbb{P}}(X)$. None of these can be a nonzero constant polynomial, since $X \neq \emptyset$, and it follows that they are all homogeneous polynomials of positive degree and hence vanish at $(0,\dots,0) \in \mathbb{A}^{n+1}$. Thus, we have $f(0,\dots,0) = 0$ for all $f \in \mathcal{I}_{\mathbb{P}}(X)$, and with this, the definition of $\mathcal{I}_{\mathbb{P}}(X)$ can be re-expressed as follows:

$$\begin{aligned}\mathcal{I}_{\mathbb{P}}(X) &= \{f \in K[x_0,\dots,x_n] \mid f(0,\dots,0) = 0 \text{ and } \\ &\qquad f(a_0,\dots,a_n) = 0 \ \forall [a_0 : \cdots : a_n] \in X\} \\ &= \mathcal{I}_{\mathbb{A}}\big(\{(0,\dots,0)\} \cup \{(a_0,\dots,a_n) \mid [a_0 : \cdots : a_n] \in X\}\big) \\ &= \mathcal{I}_{\mathbb{A}}\big(C(X)\big),\end{aligned}$$

as claimed □

Equipped with Lemmas 9.60 and 9.61, the proof of the projective Nullstellensatz is an application of the affine Nullstellensatz, as we now discuss.

PROOF OF THEOREM 9.57 Let $I \subseteq K[x_0,x_1,\dots,x_n]$ be a homogeneous ideal. If $I = K[x_0,x_1,\dots,x_n]$, then

$$\mathcal{I}_{\mathbb{P}}(\mathcal{V}_{\mathbb{P}}(I)) = \mathcal{I}_{\mathbb{P}}(\emptyset) = K[x_0,x_1,\dots,x_n] = \sqrt{I},$$

so the theorem holds in this case.

Assume, now, that I is a proper ideal. We claim, in this case, that $\sqrt{I}$ is the irrelevant ideal if and only if $\mathcal{V}_{\mathbb{P}}(I) = \emptyset$. To see this, note first that if $\sqrt{I}$ is the irrelevant ideal, then

$$\mathcal{V}_{\mathbb{P}}(I) = \mathcal{V}_{\mathbb{P}}(\sqrt{I}) = \mathcal{V}_{\mathbb{P}}(\langle x_0, \ldots, x_n \rangle) = \emptyset,$$

where the first equality is Exercise 9.5.10. Conversely, if $\mathcal{V}_{\mathbb{P}}(I) = \emptyset$, then Lemma 9.60 implies

$$\mathcal{V}_{\mathbb{A}}(I) = C\Big(\mathcal{V}_{\mathbb{P}}(I)\Big) = C(\emptyset) = \{(0, \ldots, 0)\},$$

from which the affine Nullstellensatz shows that

$$\sqrt{I} = \mathcal{I}_{\mathbb{A}}(\mathcal{V}_{\mathbb{A}}(I)) = \mathcal{I}_{\mathbb{A}}(\{(0, \ldots, 0)\}) = \langle x_0, \ldots, x_n \rangle.$$

Now, to prove the projective Nullstellensatz, suppose first that $\sqrt{I}$ is not the irrelevant ideal, so that $\mathcal{V}_{\mathbb{P}}(I) \neq \emptyset$. Then Lemma 9.61 shows that

$$\mathcal{I}_{\mathbb{P}}(\mathcal{V}_{\mathbb{P}}(I)) = \mathcal{I}_{\mathbb{A}}\big(C(\mathcal{V}_{\mathbb{P}}(I))\big),$$

and then Lemma 9.60 (which applies because I is a proper ideal), along with the affine Nullstellensatz, show that

$$\mathcal{I}_{\mathbb{A}}\big(C(\mathcal{V}_{\mathbb{P}}(I))\big) = \mathcal{I}_{\mathbb{A}}(\mathcal{V}_{\mathbb{A}}(I)) = \sqrt{I}.$$

Finally, if $\sqrt{I}$ is the irrelevant ideal, then $\mathcal{V}_{\mathbb{P}}(I) = \emptyset$ and the statement

$$\mathcal{I}_{\mathbb{P}}(\mathcal{V}_{\mathbb{P}}(I)) = K[x_0, x_1, \ldots, x_n]$$

is immediate from the definitions. □

As a first application of the projective Nullstellensatz, we see that the computation of projective vanishing ideals is now greatly simplified.

9.62 **EXAMPLE** Vanishing ideal of a line in $\mathbb{P}^2$, revisited

Let $X = \mathcal{V}(x_0 + x_1 - x_2) \subseteq \mathbb{P}^2$, as in Example 9.27. We outlined a direct computation of $\mathcal{I}_{\mathbb{P}}(X)$ in Exercise 9.3.8, but with the projective Nullstellensatz, we now need only observe that $\langle x_0 + x_1 - x_2 \rangle$ is a radical homogeneous ideal that is not the irrelevant ideal. Thus,

$$\mathcal{I}_{\mathbb{P}}(X) = \mathcal{I}_{\mathbb{P}}(\mathcal{V}_{\mathbb{P}}(\langle x_0 + x_1 - x_2 \rangle)) = \langle x_0 + x_1 - x_2 \rangle.$$

Furthermore, just as in the affine case, the projective Nullstellensatz clarifies the domains on which the projective $\mathcal{V}$- and $\mathcal{I}$-operators are inverses, allowing us to describe the following dictionary between projective varieties and associated ideals.

9.63 PROPOSITION *Projective varieties and ideals*

The $\mathcal{V}_{\mathbb{P}}$- and $\mathcal{I}_{\mathbb{P}}$-operators are inverse, inclusion-reversing bijections that translate between the following hierarchies of ideals and varieties:

$$\begin{array}{ccc}
\left\{\begin{array}{c}\text{relevant homogeneous}\\ \text{radical ideals in } K[x_0,\ldots,x_n]\end{array}\right\} & \longleftrightarrow & \{\text{projective varieties in } \mathbb{P}^n\} \\
\cup\mathrm{I} & & \cup\mathrm{I} \\
\left\{\begin{array}{c}\text{relevant homogeneous}\\ \text{prime ideals in } K[x_0,\ldots,x_n]\end{array}\right\} & \longleftrightarrow & \{\text{irreducible varieties in } \mathbb{P}^n\} \\
\cup\mathrm{I} & & \cup\mathrm{I} \\
\left\{\begin{array}{c}\text{relevant homogeneous}\\ \text{maximal ideals in } K[x_0,\ldots,x_n]\end{array}\right\} & \longleftrightarrow & \{\text{points in } \mathbb{P}^n\}.
\end{array}$$

PROOF Let $X \subseteq \mathbb{P}^n$ be a projective variety. First, note that $\mathcal{I}_{\mathbb{P}}(X)$ is indeed a homogeneous radical ideal in $K[x_0, x_1, \ldots, x_n]$ by Proposition 9.30. Furthermore, $\mathcal{I}_{\mathbb{P}}(X)$ is not the irrelevant ideal. To see why not, suppose to the contrary that

$$\mathcal{I}_{\mathbb{P}}(X) = \langle x_0, \ldots, x_n \rangle.$$

Then Proposition 9.31 implies

$$X = \mathcal{V}_{\mathbb{P}}(\mathcal{I}_{\mathbb{P}}(X)) = \mathcal{V}(\langle x_0, \ldots, x_n \rangle) = \emptyset.$$

But then

$$\mathcal{I}_{\mathbb{P}}(X) = \mathcal{I}_{\mathbb{P}}(\emptyset) = K[x_0, \ldots, x_n],$$

contradicting our assumption that $\mathcal{I}_{\mathbb{P}}(X)$ is the irrelevant ideal. Thus, $\mathcal{I}_{\mathbb{P}}$ maps any projective variety in $\mathbb{P}^n$ to a relevant homogeneous radical ideal in $K[x_0, \ldots, x_n]$, and conversely, $\mathcal{V}_{\mathbb{P}}$ maps any such ideal to a projective variety, by definition. The fact that

$$\mathcal{V}_{\mathbb{P}}(\mathcal{I}_{\mathbb{P}}(X)) = X \quad \text{and} \quad \mathcal{I}_{\mathbb{P}}(\mathcal{V}_{\mathbb{P}}(I)) = I$$

on these domains follows from Proposition 9.31 and the projective Nullstellensatz, respectively, justifying the first bijection in the proposition. The other two bijections then follow from Exercises 9.3.3 and 9.3.4 □

Exercises for Section 9.6

9.6.1 Draw a picture, over the real numbers, of the affine cone over the projective variety

$$X = \mathcal{V}_{\mathbb{P}}(x^2 - yz) \subseteq \mathbb{P}^2.$$

(Computer graphing software might help.) Where, in your picture, do you see the affine restriction

$$X_0 = \mathcal{V}(x^2 - y) \subseteq \mathbb{A}^2?$$

9.6.2 (a) Show that Lemma 9.60 fails if $I = K[x_0, x_1, \ldots, x_n]$.

(b) Show that Lemma 9.61 fails if $X = \emptyset$.

9.6.3 Use the projective Nullstellensatz to calculate (with proof) $\mathcal{I}_{\mathbb{P}}(X)$ for the projective variety

$$X = \mathcal{V}(x^2 - yz) \subseteq \mathbb{P}^2.$$

9.6.4 Use the projective Nullstellensatz to calculate (with proof) $\mathcal{I}_{\mathbb{P}}(X)$ for the twisted cubic

$$X = \{[a^3 : a^2b : ab^2 : b^3] \mid [a : b] \in \mathbb{P}^1\} \subseteq \mathbb{P}^3.$$

Chapter 10

Maps of Projective Varieties

LEARNING OBJECTIVES FOR CHAPTER 10

- Become acquainted with regular maps between projective varieties.
- Study isomorphisms in the setting of projective varieties.
- Build a toolkit of examples of regular maps, including polynomial maps, isomorphisms, projective equivalences, and Veronese embeddings.
- Become familiar with Segre maps and use them to define products of projective varieties.

In the introduction to Chapter 4, the reader was encouraged to ask a key question whenever a new type of mathematical object is introduced: which maps between these objects preserve their relevant structure? For affine varieties, we landed upon polynomial maps as the appropriate notion of structure-preserving maps, and the goal of this chapter is to define a corresponding notion of maps between projective varieties.

This goal is complicated by a number of crucial differences between affine and projective varieties. For starters, due to the equivalence relation in the definition of $\mathbb{P}^m$, a polynomial $f \in K[x_0, \ldots, x_m]$ does not give a well-defined function on a projective variety $X \subseteq \mathbb{P}^m$, simply because the value of the polynomial is sensitive to scaling the homogeneous coordinates in $\mathbb{P}^m$. However, if f happens to be homogeneous of degree d, then we have seen that scaling homogeneous coordinates has a fairly simple effect on the value of f:

$$f(\lambda a_0, \ldots, \lambda a_m) = \lambda^d f(a_0, \ldots, a_m) \quad \text{for all} \quad \lambda \in K \setminus \{0\}.$$

It then follows that, given polynomials $f_0, \ldots, f_n \in K[x_0, \ldots, x_m]$ that are all homogeneous *of the same degree*—at least one of which does not vanish at a—we obtain a well-defined value

$$[f_0(a) : \cdots : f_n(a)] \in \mathbb{P}^n,$$

which is independent of the choice of homogeneous coordinates for a. Motivated by this observation, such tuples of polynomials will form the foundation of our study of maps between projective varieties.

In general, if $X \subseteq \mathbb{P}^m$ and $Y \subseteq \mathbb{P}^n$ are projective varieties, we will say that a function $F : X \to Y$ is a *regular map* if it can be realized (at least locally) by a tuple of polynomials $f_0, \ldots, f_n$ that are homogeneous of the same degree. Regular maps serve as the structure-preserving maps between projective varieties, and our primary aim in this section is to initiate their study.

Section 10.1 Regular maps of projective varieties

Our goal in this section is to familiarize ourselves with the precise notion of a regular map between two projective varieties. As we alluded to in the introduction to this chapter, regular maps between projective varieties are locally modeled by collections of polynomials that are homogeneous of the same degree. Before introducing the key definitions, a few observations are in order.

Suppose that $f \in K[x_0, \ldots, x_m]$ and $a \in \mathbb{P}^m$. Upon choosing homogeneous coordinates and writing $a = [a_0 : \cdots : a_n]$, we can evaluate f to obtain a value:

$$f(a_0, \ldots, a_n) \in K.$$

However, as we have seen, different choices of homogeneous coordinates for a lead to different values when evaluating f, so the value of f is not well-defined at points of $\mathbb{P}^m$. As a workaround, let us suppose that $f_0, \ldots, f_n \in K[x_0, \ldots, x_m]$ are homogeneous of the same degree d and that at least one of them does not vanish at a. Then we can collect the values together and view them as a point of projective space:

$$[f_0(a_0, \ldots, a_n) : \cdots : f_n(a_0, \ldots, a_n)] \in \mathbb{P}^n.$$

Something quite nice has occurred in doing this: the corresponding point in $\mathbb{P}^n$ is actually independent of the choice of homogeneous coordinates for a. Indeed, if we choose any other homogeneous coordinates $a = [\lambda a_0 : \cdots : \lambda a_n]$, then

$$\begin{aligned}[f_0(\lambda a_0, \ldots, \lambda a_n) : \cdots : & f_n(\lambda a_0, \ldots, \lambda a_n)] \\ &= [\lambda^d f_0(a_0, \ldots, a_n) : \cdots : \lambda^d f_n(a_0, \ldots, a_n)] \\ &= [f_0(a_0, \ldots, a_n) : \cdots : f_n(a_0, \ldots, a_n)],\end{aligned}$$

where the first equality uses homogeneity of the f_i and the second uses the equivalence relation in $\mathbb{P}^n$. In this situation, we henceforth adopt the shorthand notation

$$[f_0(a) : \cdots : f_n(a)] = [f_0(a_0, \ldots, a_n) : \cdots : f_n(a_0, \ldots, a_n)],$$

which is a slight abuse of notation, given that the individual values $f_i(a)$ are not well-defined. By evaluating at all points of $\mathbb{P}^m$ where at least one of the f_i does not vanish, we thus obtain a function

$$[f_0 : \cdots : f_n] : \mathbb{P}^m \setminus \mathcal{V}(f_0, \ldots, f_n) \to \mathbb{P}^n.$$

Functions arising in this way lead to a natural notion of *polynomial maps* between projective varieties, which serve as a first approximation to the more general notion of *regular maps*, which will be introduced later in the section.

10.1 **DEFINITION** *Polynomial map between projective varieties*

Let $X \subseteq \mathbb{P}^m$ and $Y \subseteq \mathbb{P}^n$ be projective varieties. A map $F : X \to Y$ is called a *polynomial map* if there exist polynomials $f_0, \ldots, f_n \in K[x_0, \ldots, x_m]$, all homogeneous of the same degree, such that $X \cap \mathcal{V}(f_0, \ldots, f_n) = \emptyset$ and

$$F(a) = [f_0(a) : \cdots : f_n(a)] \quad \text{for all} \quad a \in X.$$

If $f_0, \ldots, f_n$ can be taken to be linear, we say that F is a *linear map*.

Note that the conditions that $f_0, \ldots, f_n$ are homogeneous of the same degree and that $X \cap \mathcal{V}(f_0, \ldots, f_n) = \emptyset$ together ensure that $[f_0(a) : \cdots : f_n(a)]$ is a well-defined point of $\mathbb{P}^n$ for all $a \in X$. Let us consider a few concrete examples.

10.2 **EXAMPLE** A polynomial map from $\mathbb{P}^1$ to a conic

Let $X = \mathcal{V}(y^2 - xz) \subseteq \mathbb{P}^2$ and consider the function

$$\begin{aligned} G : \mathbb{P}^1 &\to X \\ [a : b] &\mapsto [a^2 : ab : b^2]. \end{aligned}$$

Observe that G is a polynomial map that can be realized by the three polynomials $x^2, xy, y^2 \in K[x, y]$. That G is defined at every point of $\mathbb{P}^1$ follows from the observation that

$$\mathcal{V}(x^2, xy, y^2) = \emptyset \subseteq \mathbb{P}^1,$$

and to verify that the image of G lies in X, note that $y^2 - xz$ vanishes when evaluated at $[a^2 : ab : b^2]$, for any $[a : b] \in \mathbb{P}^1$.

10.3 **EXAMPLE** A linear map from a quadric surface to $\mathbb{P}^2$

Let $X = \mathcal{V}(w^2 + x^2 + y^2 - z^2) \subseteq \mathbb{P}^3$, and note that $[0 : 0 : 0 : 1] \notin X$. Thus, we obtain a linear map

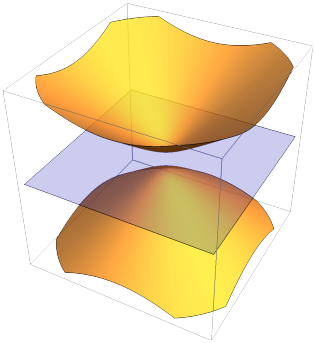
$$\begin{aligned} H : X &\to \mathbb{P}^2 \\ [a : b : c : d] &\mapsto [a : b : c], \end{aligned}$$

given by the homogeneous linear polynomials $w, x, y \in K[w, x, y, z]$.

To visualize H, let us consider the affine patch X_0 where $w \neq 0$:

$$X_0 = \mathcal{V}(1 + x^2 + y^2 - z^2) \subseteq \mathbb{A}^3.$$

Restricting to this patch, the map H sends $(b, c, d) \in X_0$ to $(b, c) \in \mathbb{A}^2$, which can be visualized over the real numbers as the two-to-one map that vertically projects the two-sheeted hyperboloid depicted at right onto the horizontal coordinate plane.



10.4 **EXAMPLE** Linear projections

Generalizing the previous example, suppose that $X \subseteq \mathbb{P}^n$ is a projective variety that does not contain the point $[0 : \cdots : 0 : 1]$. The *linear projection of X onto* $\mathbb{P}^{n-1}$ is the linear map

$$\begin{aligned} H : X &\to \mathbb{P}^{n-1} \\ [a_0 : \cdots : a_{n-1} : a_n] &\mapsto [a_0 : \cdots : a_{n-1}]. \end{aligned}$$

One way to visualize the map H is as follows: given a point $a = [a_0 : \cdots : a_n] \in X$, there is a unique line $L_a \subseteq \mathbb{P}^n$ that passes through both $[0 : \cdots : 0 : 1]$ and a:

$$L_a = \{[ca_0 : \cdots : ca_{n-1} : da_n] \mid [c : d] \in \mathbb{P}^1\}.$$

This line intersects the hyperplane $\mathcal{V}(x_n) \subseteq \mathbb{P}^n$ at the point $[a_0 : \cdots : a_{n-1} : 0]$, and making the natural identification of $\mathcal{V}(x_n)$ with $\mathbb{P}^{n-1}$, this intersection point is $H(a)$. Thus, intuitively, we view $H(X)$ as the shadow that X casts on $\mathcal{V}(x_n)$ when a light shines from a. In the affine patch where $x_0 \neq 0$, the lines L_a are "vertical" lines for which only the last coordinate varies and the hyperplane $V(x_n)$ is the "horizontal" hyperplane for which the last coordinate is zero; thus, as in Example 10.3, the linear projection in this patch can be viewed as the vertical projection.

To motivate the more general definition of *regular maps*, we note that the constraint in Definition 10.1 that $X \cap \mathcal{V}(f_0, \ldots, f_n) = \emptyset$ is often too restrictive, and it is useful to allow ourselves the flexibility to work with maps that are described by polynomials only "locally." To better understand this, let us consider an example.

10.5 **EXAMPLE** A piecewise polynomial function from a conic to $\mathbb{P}^1$

Let $X = \mathcal{V}(y^2 - xz) \subseteq \mathbb{P}^2$ and consider the pair of homogeneous linear polynomials $x, y \in K[x, y, z]$. Note that these polynomials do *not* give rise to a polynomial map from X to $\mathbb{P}^1$, simply because $\mathcal{V}(x, y) \cap X = \{[0 : 0 : 1]\} \neq \emptyset$. However, we still obtain a function from a subset of X to $\mathbb{P}^1$:

$$\begin{aligned} [x : y] : X \setminus \mathcal{V}(x, y) &\to \mathbb{P}^1 \\ [a : b : c] &\mapsto [a : b]. \end{aligned}$$

Importantly, it's not the pair (x, y) that interests us, but the function $[x : y]$ that the pair defines. For example, the function $[5x : 5y]$ is the same as $[x : y]$, simply because points of $\mathbb{P}^1$ are invariant under scaling coordinates. More generally, recalling that two points of $\mathbb{P}^1$ are equal when their cross-ratio vanishes (Exercise 9.1.2), we see that two functions $[f : g]$ and $[f' : g']$ are equal—at all points where they are both defined—exactly when their cross-ratio vanishes on X:

$$fg' - gf' \in \mathcal{I}(X).$$

For example, the maps $[x : y]$ and $[y : z]$ agree at all points of their common domain within X because their cross-ratio is $xz - y^2$, which vanishes on X. Let us consider, then, the second of these two functions:

$$\begin{aligned} [y : z] : X \setminus \mathcal{V}(y, z) &\to \mathbb{P}^1 \\ [a : b : c] &\mapsto [b : c]. \end{aligned}$$

Observe that the domains of $[x : y]$ and $[y : z]$ collectively cover all of X, since the first only omits $[0 : 0 : 1]$ while the second only omits $[1 : 0 : 0]$. Since the maps agree at all points where both are defined, we can then combine them to obtain a piecewise-defined function on all of X:

$$\begin{aligned} F : X &\to \mathbb{P}^1 \\ [a : b : c] &\mapsto \begin{cases} [a : b] & [a : b : c] \notin \mathcal{V}(x, y) \\ [b : c] & [a : b : c] \notin \mathcal{V}(y, z). \end{cases} \end{aligned}$$

The function $F : X \to \mathbb{P}^1$ in the previous example is not described by a single pair of polynomials, but it is "locally" polynomial in the following sense: for every point $p \in X$, there exists a pair $f, g \in K[x,y,z]$—possibly different pairs for different points—such that $p \notin \mathcal{V}(f,g)$ and

$$F(a) = [f(a) : g(a)] \quad \text{for all} \quad a \in X \setminus \mathcal{V}(f,g).$$

This characterization of F motivates the definition of regular maps.

10.6 **DEFINITION** *Regular map between projective varieties*

Let $X \subseteq \mathbb{P}^m$ and $Y \subseteq \mathbb{P}^n$ be projective varieties. A map $F : X \to Y$ is said to be a *regular map* if, for every $p \in X$, there exist polynomials $f_0, \ldots, f_n \in K[x_0, \ldots, x_m]$, all homogeneous of the same degree, such that $p \notin \mathcal{V}(f_0, \ldots, f_n)$ and

$$F(a) = [f_0(a) : \cdots : f_n(a)] \quad \text{for all} \quad a \in X \setminus \mathcal{V}(f_0, \ldots, f_n).$$

In other words, a map $F : X \to Y$ is regular if, for every $p \in X$, we can find a polynomial expression for F that is well-defined at p, even though it may not be well-defined on all of X. It follows from the definitions that every polynomial map is regular, but the converse is not true: not every regular map can be described globally by a single tuple of polynomials. We verify this in the next example.

10.7 **EXAMPLE** A regular map that is not a polynomial map

Let $X = \mathcal{V}(y^2 - xz) \subseteq \mathbb{P}^2$ and consider again the regular map of Example 10.5:

$$F : X \to \mathbb{P}^1$$
$$[a : b : c] \mapsto \begin{cases} [a : b] & [a : b : c] \notin \mathcal{V}(x,y) \\ [b : c] & [a : b : c] \notin \mathcal{V}(y,z). \end{cases}$$

We claim that F is not a polynomial map. To justify this, suppose to the contrary that there exists $f, g \in K[x,y,z]$, homogeneous of the same degree d, such that $\mathcal{V}(f,g) \cap X = \emptyset$ and $F = [f : g]$. Using that $y^2 = xz$ on X, we may replace every instance of y^2 in both f and g with xz without affecting the map $[f : g]$. This allows us to reduce to the case that both f and g are linear in y:

$$f = f_0 + yf_1 \quad \text{and} \quad g = g_0 + yg_1 \quad \text{for some} \quad f_0, f_1, g_0, g_1 \in K[x,z].$$

Since $F = [f : g]$, the formula for F then implies that f vanishes at $[0 : 0 : 1]$ but not at $[1 : 0 : 0]$, and it follows that $f_0 = ax^d$ for some nonzero $a \in K$. Similarly, $g_0 = bz^d$ for some nonzero $b \in K$. Since the function $[f : g]$ must be equal to $[x : y]$ at all points in their common domain, we have

$$yf - xg \in \mathcal{I}(X) = \langle y^2 - xz \rangle,$$

implying that

$$ax^d y + y^2 f_1 - bxz^d - xyg_1 \in \langle y^2 - xz \rangle.$$

Subtracting $(y^2 - xz)f_1 \in \langle y^2 - xz \rangle$, we then see that

$$ax^d y + xzf_1 - bxz^d - xyg_1 \in \langle y^2 - xz \rangle,$$

and since every nonzero element of $\langle y^2 - xz \rangle$ is at least quadratic in y, we have

$$ax^d y + xzf_1 - bxz^d - xyg_1 = 0.$$

Identifying the constant and linear coefficients in y, we see that

$$f_1 = bz^{d-1} \quad \text{and} \quad g_1 = ax^{d-1},$$

from which it follows that

$$f = ax^d + byz^{d-1} \quad \text{and} \quad g = bz^d + ayx^{d-1}.$$

However, letting $c \in K$ be any value such that $c^{2d-1} = -b/a$, one readily checks that $[c : 1 : c^{-1}]$ is a point of X at which both f and g vanish, a contradiction of the assumption that $\mathcal{V}(f, g) \cap X = \emptyset$. Thus, no such f and g can exist.

We close this section by mentioning that, in general, a method for describing a regular map is to define it piecewise, much like we did in Example 10.5. More precisely, a regular map $F : X \to Y$ between projective varieties $X \subseteq \mathbb{P}^m$ and $Y \subseteq \mathbb{P}^n$ can always be described by a collection of functions of the form

$$[f_0 : \cdots : f_n] : X \setminus \mathcal{V}(f_0, \ldots, f_n) \to Y.$$

This collection of functions must satisfy the following two properties:

1. If $[f_0 : \cdots : f_n]$ and $[g_0 : \cdots : g_n]$ are both in the collection, then
$$f_i g_j - f_j g_i \in \mathcal{I}(X) \quad \text{for all} \quad i, j.$$
2. Every $p \in X$ must be in the domain of at least one function in this collection.

The first condition ensures that two functions in this collection agree on their common domain, which allows us to combine them to obtain a well-defined function on their union, and the second condition ensures that the union of all the domains covers X. In fact, while it is not obvious from the definition, the reader is encouraged to verify that every regular map can be described by a *finite* collection of such functions (Exercise 10.1.7).

Exercises for Section 10.1

10.1.1 Let $X \subseteq \mathbb{P}^m$ and $Y \subseteq \mathbb{P}^n$ be projective varieties, and let $F : X \to Y$ be a regular map. If $Z \subseteq \mathbb{P}^m$ is any projective variety such that $Z \subseteq X$, explain why the restriction $F|_Z : Z \to Y$ is also a regular map.

10.1.2 Let $X \subseteq \mathbb{P}^\ell$, $Y \subseteq \mathbb{P}^m$, and $Z \subseteq \mathbb{P}^n$ be projective varieties. If $F : X \to Y$ and $G : Y \to Z$ are regular maps, prove that $G \circ F : X \to Z$ is also regular.

10.1.3 Let $X = \mathcal{V}(wz - xy) \subseteq \mathbb{P}^3$.

(a) Construct a regular map $F : X \to \mathbb{P}^1$ that extends the function

$$[w : x] : X \setminus \mathcal{V}(w, x) \to \mathbb{P}^1.$$

(b) Prove that F is surjective.

(c) Describe the preimage of F over any point in $\mathbb{P}^1$.

10.1.4 Prove that every regular map $F : \mathbb{P}^1 \to \mathbb{P}^n$ is a polynomial map.

10.1.5 Prove that the map

$$[x_0 : \cdots : x_{n-1}] : \mathbb{P}^n \setminus \mathcal{V}(x_0, \ldots, x_{n-1}) \to \mathbb{P}^{n-1}$$

cannot be extended to a regular map on all of $\mathbb{P}^n$.

10.1.6 Let $X \subseteq \mathbb{P}^n$ be a projective variety that does not contain $[0 : \cdots : 0 : 1]$ and consider the linear projection

$$H : X \to \mathbb{P}^{n-1}$$
$$[a_0 : \cdots : a_{n-1} : a_n] \mapsto [a_0 : \cdots : a_{n-1}]$$

Prove that H is finite-to-one. In other words, prove that, for any $a \in \mathbb{P}^{n-1}$, the preimage $H^{-1}(\{a\})$ is a finite set.

10.1.7 Let $X \subseteq \mathbb{P}^m$ and $Y \subseteq \mathbb{P}^n$ be projective varieties. Prove that every regular map $F : X \to Y$ can be described by a *finite* collection of functions of the form

$$[f_0 : \cdots : f_n] : X \setminus \mathcal{V}(f_0, \ldots, f_n) \to Y.$$

Section 10.2 Isomorphisms of projective varieties

Having developed an appropriate notion of maps between projective varieties, we now turn to a discussion of isomorphisms. We begin with the natural definition.

10.8 **DEFINITION** *Isomorphism of projective varieties*

An *isomorphism* of projective varieties $X \subseteq \mathbb{P}^m$ and $Y \subseteq \mathbb{P}^n$ is a regular map $F : X \to Y$ for which there exists a regular map $G : Y \to X$ that is inverse to F. If there exists an isomorphism between X and Y, we say that X and Y are *isomorphic* and write $X \cong Y$.

As in the case of affine varieties, we view the intrinsic nature of a projective variety to be the structure that is preserved under isomorphism.

A few moments reflecting should convince the reader that isomorphisms are an equivalence relation on the set of projective varieties—verifying this assertion carefully requires knowing that compositions of regular maps are regular (see Exercise 10.1.2). As a first example, building on the maps discussed in the previous section, we observe that the conic $\mathcal{V}(y^2 - xz) \subseteq \mathbb{P}^2$ is isomorphic to $\mathbb{P}^1$.

10.9 **EXAMPLE** $\mathcal{V}(y^2 - xz) \cong \mathbb{P}^1$

Let $X = \mathcal{V}(y^2 - xz) \subseteq \mathbb{P}^2$ and let $F : X \to \mathbb{P}^1$ and $G : \mathbb{P}^1 \to X$ be the regular maps introduced in Examples 10.7 and 10.2, respectively:

$$F([a:b:c]) = \begin{cases} [a:b] & [a:b:c] \notin \mathcal{V}(x,y) \\ [b:c] & [a:b:c] \notin \mathcal{V}(y,z). \end{cases}$$

and

$$G([a:b]) = [a^2 : ab : b^2].$$

We claim that these two regular maps are inverse to one another. Given a point $[a:b] \in \mathbb{P}^1$, we compute

$$F(G([a:b])) = F([a^2:ab:b^2]) = \left\{ \begin{array}{ll} [a^2:ab] & \text{if } a \neq 0 \\ [ab:b^2] & \text{if } b \neq 0 \end{array} \right\} = [a:b].$$

Conversely, given a point $[a:b:c] \in X$, we compute

$$\begin{aligned} G(F([a:b:c])) &= G\left(\left\{ \begin{array}{ll} [a:b] & \text{if } (a,b) \neq (0,0) \\ [b:c] & \text{if } (b,c) \neq (0,0) \end{array} \right\} \right) \\ &= \left\{ \begin{array}{ll} [a^2:ab:b^2] & \text{if } (a,b) \neq (0,0) \\ [b^2:bc:c^2] & \text{if } (b,c) \neq (0,0) \end{array} \right\} \\ &= [a:b:c], \end{aligned}$$

where the final equality uses the fact that $b^2 = ac$ for every $[a:b:c] \in X$. Since F and G are inverse regular maps, we conclude that $X \cong \mathbb{P}^1$.

A special class of isomorphisms arises from linear transformations on projective space. To set up notation, suppose that A is an invertible $(n+1)\times(n+1)$ matrix with entries in K. For each point $p \in \mathbb{P}^n$, we define a new point $Ap \in \mathbb{P}^n$ by choosing homogeneous coordinates for p, viewing them as a column vector, and multiplying the column vector on the left by the matrix A. For example, if

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 1 & -1 \end{pmatrix}$$

and $p = [a_0 : a_1 : a_2]$, then $Ap = [a_0 + a_1 : a_0 + 2a_2 : a_1 - a_2]$.

More generally, if we write $A = (a_{ij})$, the procedure described above gives rise to a linear map $F_A : \mathbb{P}^n \to \mathbb{P}^n$ defined by linear polynomials whose coefficients are the rows of A:

$$F_A = \Big[\sum_{j=0}^{n} a_{0j}x_j : \cdots : \sum_{j=0}^{n} a_{nj}x_j \Big].$$

The invertibility of A implies that F_A is defined at every point of $\mathbb{P}^n$, and, moreover, that the inverse of F_A is the regular map

$$F_{A^{-1}} : \mathbb{P}^n \to \mathbb{P}^n$$

(see Exercise 10.2.3). Thus, F_A is an isomorphism from $\mathbb{P}^n$ to itself, and we give isomorphisms arising from invertible matrices in this way a special name.

10.10 **DEFINITION** *Projective equivalence*

A *projective equivalence of* $\mathbb{P}^n$ is an isomorphism of the form

$$\begin{aligned} F_A : \mathbb{P}^n &\to \mathbb{P}^n \\ p &\mapsto Ap, \end{aligned}$$

where A is an invertible $(n+1)\times(n+1)$ matrix with entries in K. We say that two projective varieties $X, Y \subseteq \mathbb{P}^n$ are *projectively equivalent* if there exists a projective equivalence $F_A : \mathbb{P}^n \to \mathbb{P}^n$ such that $F_A(X) = Y$.

Intuitively, we view a projective equivalence simply as a change of coordinates in $\mathbb{P}^n$, similar to a change of basis in a vector space. Let us consider an example.

10.11 **EXAMPLE** Projectively equivalent conics

Consider the two projective varieties

$$X = \mathcal{V}(x^2 + y^2 + z^2) \subseteq \mathbb{P}^2_{\mathbb{C}} \quad \text{and} \quad Y = \mathcal{V}(y^2 - xz) \subseteq \mathbb{P}^2_{\mathbb{C}}.$$

We claim that X and Y are projectively equivalent, and we exhibit this through a change of coordinates in the defining equations. More precisely, notice that

$$x^2 + y^2 + z^2 = y^2 - (z\mathrm{i} + x)(z\mathrm{i} - x).$$

This implies that $[a : b : c] \in X$ if and only if $[ci + a : b : ci - a] \in Y$. In other words, letting A be the invertible matrix

$$A = \begin{pmatrix} 1 & 0 & i \\ 0 & 1 & 0 \\ -1 & 0 & i \end{pmatrix},$$

we see that $F_A(X) = Y$, so X and Y are projectively equivalent.

In fact, the previous example can be generalized in a somewhat surprising way.

10.12 EXAMPLE All irreducible conics are projectively equivalent

Let $X = \mathcal{V}(f) \subseteq \mathbb{P}^2_{\mathbb{C}}$ where $f \in \mathbb{C}[x,y,z]$ is irreducible and homogeneous of degree 2. We claim that X is projectively equivalent to $\mathcal{V}(x^2 - yz)$.

First, suppose that f does not have any nonzero coefficients of x^2, y^2, or z^2:

$$f = axy + bxz + cyz \quad \text{for some} \quad a,b,c \in \mathbb{C}.$$

By irreducibility, we must have $a \neq 0$, and the invertible change of variables

$$(x,y,z) \mapsto (x+y, x-y, z)$$

transforms f into $ax^2 - ay^2 + bxz + cxz + (b-c)yz$, which has a nonzero coefficient of x^2. Therefore, working up to projective equivalence, we may assume that $X = \mathcal{V}(f)$ where f has at least one nonzero square term. Up to reordering variables and scaling the polynomial, we may assume that the coefficient of x^2 is 1:

$$f = x^2 + axy + by^2 + cxz + dz^2 + exz \quad \text{for some} \quad a,b,c,d,e \in \mathbb{C}.$$

By "completing the square," we can write

$$x^2 + axy + by^2 = \left(x + \frac{a}{2}y\right)^2 + \left(b - \frac{a^2}{4}\right)y^2,$$

and it follows that the invertible change of coordinates

$$(x,y,z) \mapsto \left(x - \frac{a}{2}y, y, z\right)$$

transforms f into a polynomial without an xy-term. Similarly, we can transform f into a polynomial without an xz-term. Thus, working up to projective equivalence, we may assume that f has the form

$$f = x^2 + ay^2 + byz + cz^2 \quad \text{for some} \quad a,b,c \in \mathbb{C}.$$

Since $\mathbb{C}$ is algebraically closed, $ay^2 + byz + cz^2$ factors as a product of linear polynomials $\ell_1, \ell_2 \in \mathbb{C}[y,z]$, and by irreducibility of f, it must be the case that ℓ_1 and ℓ_2 are linearly independent. Thus, $f = x^2 - (-\ell_1)\ell_2$, and the invertible change of coordinates

$$(x,y,z) \mapsto (x, -\ell_1, \ell_2)$$

transforms $x^2 - yz$ into f, so $\mathcal{V}(f)$ is projectively equivalent to $\mathcal{V}(x^2 - yz)$.

Combining Examples 10.9 and 10.12, we have essentially proved the following.

10.13 **PROPOSITION** *All irreducible conics are isomorphic to* $\mathbb{P}^1$

Let $f \in K[x, y, z]$ be irreducible and homogeneous of degree 2. Then

$$\mathcal{V}(f) \cong \mathbb{P}^1.$$

PROOF The combination of Examples 10.9 and 10.12 proves the case $K = \mathbb{C}$. For the more general setting, we note that the computations in the examples readily extend to any (algebraically closed) field of characteristic not equal to 2 (when competing the square, we divided by 2, which is not valid if the characteristic is 2). With a slightly different argument, however, the result of Example 10.12 continues to hold in characteristic 2 (Exercise 10.2.4). □

While Proposition 10.13 tells us that the intrinsic nature of irreducible conics in $\mathbb{P}^2$ is rather simplistic—they are all isomorphic to $\mathbb{P}^1$—we caution the reader against being misled into thinking that the intrinsic nature of curves of higher degree in $\mathbb{P}^2$ is just as simple. In fact, upon increasing the degree by one, it can be shown that there are infinitely many distinct isomorphism classes of irreducible cubic curves in $\mathbb{P}^2$, none of which are isomorphic to $\mathbb{P}^1$. Cubic curves in $\mathbb{P}^2$ form the basis of the study of *elliptic curves*, which is a fascinating branch of mathematics to which a great many researchers have devoted their entire careers.

The discussion of isomorphisms naturally leads to the question: how might we study the intrinsic nature of projective varieties? In the affine setting, we introduced the coordinate ring of an affine variety as a key tool in this regard: two affine varieties are isomorphic if and only their coordinate rings are isomorphic, allowing us study the intrinsic nature of affine varieties by studying the algebraic structure of their coordinate rings. Motivated by the affine setting, we might wonder, then, if there is an algebraic object that captures the intrinsic nature of projective varieties. Following our developments in the affine setting, it is completely natural to introduce the *homogeneous coordinate ring* of a projective variety $X \subseteq \mathbb{P}^n$:

$$\frac{K[x_0, \ldots, x_n]}{\mathcal{I}(X)}.$$

Unfortunately, unlike in the affine setting, the homogeneous coordinate ring is not preserved by isomorphisms, as the next example illustrates.

10.14 **EXAMPLE** Isomorphisms and homogeneous coordinate rings

As we have seen in Example 10.9, $X = \mathcal{V}(y^2 - xz) \subseteq \mathbb{P}^2$ is isomorphic to $\mathbb{P}^1$. However, the homogeneous coordinate rings of these two projective varieties are

$$\frac{K[x, y, z]}{\langle y^2 - xz \rangle} \quad \text{and} \quad K[s, t],$$

respectively, and these are not isomorphic rings. In particular, the equation $y^2 = xz$ can be used to show that the former is not a unique factorization domain, while the

latter is. Another way to see that these rings are not isomorphic is that, if they were, then the equivalence of algebra and geometry in the setting of affine varieties would imply an isomorphism of affine varieties

$$\mathcal{V}_{\mathbb{A}}(y^2 - xz) \cong \mathbb{A}^2,$$

but the former has a singularity at the origin while the latter is nonsingular.

Homogeneous coordinate rings play an important role in projective geometry, but they will not be central to the developments in this text.

While the previous example implies that homogeneous coordinate rings are not complete algebraic invariants of projective varieties, it still leaves open the possibility that maybe there exists a different algebraic object associated to projective varieties that encodes all of their intrinsic structure. As we will see in the next chapter, however, there is not a single algebraic object that encodes the intrinsic nature of a projective variety, but an entire family of algebras, leading us to the notion of a *sheaf* of algebras. Loosely speaking each algebra in the sheaf records local geometric information about the projective variety, while the way in which the algebras fit together records the global geometry of the variety. Once we have developed these ideas, we will have a much better understanding of the intrinsic nature of projective varieties.

Before getting too far ahead of ourselves, however, we will continue to hone our understanding of maps between projective varieties in the next two sections, each of which is devoted to classical family of maps: the Veronese and Segre maps.

Exercises for Section 10.2

10.2.1 Let $X = \mathcal{V}(wy - xz, x^2 - wy, y^2 - xz) \subseteq \mathbb{P}^3$, and consider the map

$$\begin{aligned} F : \mathbb{P}^1 &\to X \\ [a : b] &\mapsto [a^3 : a^2b : ab^2 : b^3]. \end{aligned}$$

Prove that F is an isomorphism by constructing a regular inverse.

10.2.2 Let $\ell_1, \ldots, \ell_k$ be homogeneous linear polynomials that are linearly independent. Prove that $\mathcal{V}(\ell_1, \ldots, \ell_k) \subseteq \mathbb{P}^n$ is isomorphic to $\mathbb{P}^{n-k}$.

10.2.3 Let A be an invertible $(n+1) \times (n+1)$ matrix A with entries in K.

(a) Prove that $F_A : \mathbb{P}^n \to \mathbb{P}^n$ is a regular map.

(b) Prove that $F_{A^{-1}}$ is the inverse of F_A.

10.2.4 Let K be an algebraically closed field of characteristic 2 and let $f \in K[x, y, z]$ be an irreducible homogeneous polynomial of degree 2. Prove that $\mathcal{V}(f)$ is projectively equivalent to $\mathcal{V}(x^2 - yz)$.

10.2.5 Let $X, Y \subseteq \mathbb{P}^n$ be projective varieties that are projectively equivalent. Prove that

$$\frac{K[x_0, \ldots, x_n]}{\mathcal{I}(X)} \cong \frac{K[x_0, \ldots, x_n]}{\mathcal{I}(Y)}.$$

10.2.6 Let $X \subseteq \mathbb{P}^n$ be a projective variety and let $\ell_0, \ldots, \ell_k \in K[x_0, \ldots, x_n]$ be homogeneous linear polynomials such that $\mathcal{V}(\ell_0, \ldots, \ell_k) \cap X = \emptyset$. Consider the linear map

$$\begin{aligned} L : X &\to \mathbb{P}^k \\ a &\mapsto [\ell_0(a) : \cdots : \ell_n(a)]. \end{aligned}$$

(a) If $\ell_0, \ldots, \ell_k$ are linearly independent, prove that there exists a projective equivalence $F_A : \mathbb{P}^n \to \mathbb{P}^n$ such that, for all $a = [a_0 : \cdots : a_n] \in X$,

$$L(F_A(a)) = [a_0 : \cdots : a_k].$$

(b) If $\ell_0, \ldots, \ell_k$ are linearly independent, prove that L is finite-to-one. (Hint: Exercise 10.1.6).

(c) Prove that L is finite-to-one even if $\ell_0, \ldots, \ell_k$ are not linearly independent.

10.2.7 Let X and Y be isomorphic projective varieties. Prove that X is irreducible if and only if Y is irreducible.

Section 10.3 Veronese maps

In this section, we introduce the Veronese maps, named in honor of the Italian mathematician Giuseppe Veronese (1854–1917). These are a family of regular maps between projective spaces that can be described by the collection of all monomials of a fixed degree. The utility of Veronese maps is that they allow us to reduce the degree of projective varieties and maps between them. More precisely, as we will see in this section, Veronese maps can be leveraged to prove the following two somewhat surprising properties.

1. Up to isomorphism, every projective variety can be realized as the vanishing of a collection of polynomials of degree at most two (Proposition 10.22).
2. Up to isomorphism, every polynomial map of projective varieties is a linear map (Proposition 10.23).

Let us dive in and begin our discussion of Veronese maps with the definition.

10.15 **DEFINITION** *Veronese maps*

For any $d, n \geq 1$, the *Veronese map* $F_{d,n} : \mathbb{P}^n \to \mathbb{P}^{\binom{d+n}{d}-1}$ is the polynomial map

$$F_{d,n} = [x_0^d : x_0^{d-1}x_1 : x_0^{d-2}x_1^2 : x_0^{d-2}x_1x_2 : \cdots : x_n^d],$$

where the monomials appearing in the definition are all possible monomials of degree d in the variables $x_0, \ldots, x_n$.

Exercise 10.3.1 outlines a strategy to prove that there are precisely $\binom{d+n}{d}$ monomials of degree d in the variables $x_0, \ldots, x_n$ (which explains the dimension of the projective space that serves as the codomain of $F_{d,n}$). Furthemore, Exercise 10.3.2 asks the reader to verify that these monomials do not simultaneously vanish at any point of $\mathbb{P}^n$, explaining why the domain of $F_{d,n}$ is all of $\mathbb{P}^n$.

As a first example, setting $d = 2$ and $n = 1$, the Veronese map is

$$\begin{aligned} F_{2,1} : \mathbb{P}^1 &\to \mathbb{P}^2 \\ [a : b] &\mapsto [a^2 : ab : b^2], \end{aligned}$$

which is the map F that we considered in Examples 10.2 and 10.9. In that particular case, we described the image $F_{2,1}(\mathbb{P}^1)$ as the projective variety $\mathcal{V}(y^2 - xz) \subseteq \mathbb{P}^2$, and our next aim is to generalize this description to $F_{d,n}(\mathbb{P}^n)$ for any d and n. In order to describe the image of any Veronese map as a projective variety, we first establish convenient notation for the coordinates of the codomain of $F_{d,n}$.

Note that degree-d monomials in the variables $x_0, \ldots, x_n$ can be indexed by tuples $D = (d_0, \ldots, d_n)$ of non-negative integers with $d_0 + \cdots + d_n = d$; more specifically, the monomial associated to $D = (d_0, \ldots, d_n)$ is

$$x^D = x_0^{d_0} x_1^{d_1} \cdots x_n^{d_n}.$$

In light of this, we denote the coordinates of $\mathbb{P}^{\binom{d+n}{d}-1}$ by y_D for each such tuple D, so that $F_{d,n}$ is the map given by the sequence of polynomials whose y_D-coordinate

is x^D. For instance, in the case $d = 2$ and $n = 1$ considered above, we denote the coordinates in $\mathbb{P}^2$ by $y_{(2,0)}$, $y_{(1,1)}$, and $y_{(0,2)}$, in which case $F_{2,1}$ sends a point $a = [a_0 : a_1]$ to the point whose $y_{(2,0)}$-coordinate is a_0^2, whose $y_{(1,1)}$-coordinate is a_0a_1, and whose $y_{(0,2)}$-coordinate is a_1^2. From this, one sees that $F_{2,1}$ maps to the variety

$$\mathcal{V}(y_{(2,0)}y_{(0,2)} - y_{(1,1)}^2) \subseteq \mathbb{P}^2.$$

More generally, images of Veronese maps are described by the following result.

10.16 **PROPOSITION** *Images of Veronese maps*

For any $d, n \geq 1$, the image of the Veronese map $F_{d,n} : \mathbb{P}^n \to \mathbb{P}^{\binom{d+n}{d}-1}$ is

$$X_{d,n} = \mathcal{V}(\{y_D y_{D'} - y_E y_{E'} \mid D + D' = E + E'\}) \subseteq \mathbb{P}^{\binom{d+n}{d}-1}.$$

PROOF That $F_{d,n}(\mathbb{P}^n) \subseteq X_{d,n}$ is a result of the observation that

$$a^D a^{D'} - a^E a^{E'} = a_0^{d_0+d_0'} \cdots a_n^{d_n+d_n'} - a_0^{e_0+e_0'} \cdots a_n^{e_n+e_n'} = 0$$

for any $a \in \mathbb{P}^n$ and for any tuples D, D', E, E' such that $D + D' = E + E'$.

Conversely, to show that $F_{d,n}(\mathbb{P}^n) \supseteq X_{d,n}$, let $b \in X_{d,n}$, so that the coordinates of b satisfy $b_D b_{D'} = b_E b_{E'}$ for all tuples D, D', E, E' with $D + D' = E + E'$. From the relations $b_D b_{D'} = b_E b_{E'}$, one can show (Exercise 10.3.3) that

$$b_{D_1} b_{D_2} \cdots b_{D_k} = b_{E_1} b_{E_2} \cdots b_{E_k} \tag{10.17}$$

for all tuples $D_1, \ldots, D_k, E_1, \ldots, E_k$ such that $D_1 + \cdots + D_k = E_1 + \cdots + E_k$. In particular, for $D = (d_0, \ldots, d_n)$, we have

$$b_D^d = b_{(d,0,\ldots,0)}^{d_0} b_{(0,d,0,\ldots,0)}^{d_1} \cdots b_{(0,\ldots,0,d)}^{d_n}. \tag{10.18}$$

This implies that at least one of the coordinates

$$b_{(d,0,\ldots,0)}, b_{(0,d,0,\ldots,0)}, \ldots, b_{(0,\ldots,0,d)}$$

is nonzero, for otherwise (10.18) would yield $b_D = 0$ for all D, which is impossible in projective space.

Suppose, then, without loss of generality, that $b_{(d,0,\ldots,0)}$ is nonzero. Define an element $a = [a_0 : \cdots : a_n] \in \mathbb{P}^n$ by choosing a_0 to be any element of K satisfying $a_0^d = b_{(d,0,\ldots,0)}$, and for $1 \leq i \leq n$, set

$$a_i = a_0 \cdot \frac{b_{(d-1,0,\ldots,0,1,0,\ldots,0)}}{b_{(d,0,\ldots,0)}},$$

in which the 1 in the index of the numerator is in the ith coordinate. We claim that

$$F_{d,n}(a) = b. \tag{10.19}$$

To prove (10.19), note that the definition of $F_{d,n}$ and of a_i implies that, for any $D = (d_0, \ldots, d_n)$, the y_D-coordinate of $F_{d,n}(a)$ is

$$a_0^{d_0} a_1^{d_1} \cdots a_n^{d_n} = a_0^{d_0+d_1+\cdots+d_n} \frac{b_{(d-1,1,0,\ldots,0)}^{d_1}}{b_{(d,0,\ldots,0)}^{d_1}} \frac{b_{(d-1,0,1,0,\ldots,0)}^{d_2}}{b_{(d,0,\ldots,0)}^{d_2}} \cdots \frac{b_{(d-1,0,\ldots,0,1)}^{d_n}}{b_{(d,0,\ldots,0)}^{d_n}}$$
$$= b_{(d,0,\ldots,0)} \frac{b_{(d-1,1,0,\ldots,0)}^{d_1}}{b_{(d,0,\ldots,0)}^{d_1}} \frac{b_{(d-1,0,1,0,\ldots,0)}^{d_2}}{b_{(d,0,\ldots,0)}^{d_2}} \cdots \frac{b_{(d-1,0,\ldots,0,1)}^{d_n}}{b_{(d,0,\ldots,0)}^{d_n}}.$$

From here, an application of equation (10.17) shows that the above equals b_D. This proves (10.19) and thus completes the proof that $X_{d,n} = F_{d,n}(\mathbb{P}^n)$. □

Proposition 10.16 generalizes one aspect of Example 10.9—it gives explicit defining equations for the variety into which $F_{d,n}$ maps—but in the case of Example 10.9, we actually proved more: we showed that the Veronese map was an isomorphism onto its image. This fact is also true of all Veronese maps.

10.20 PROPOSITION *Veronese maps are embeddings*

For any $d, n \geq 1$, the Veronese map $F_{d,n} : \mathbb{P}^n \to X_{d,n}$ is an isomorphism.

PROOF To prove that $F_{d,n}$ is an isomorphism onto $X_{d,n}$, we construct a regular inverse $G_{d,n} : X_{d,n} \to \mathbb{P}^n$. Given any $b \in X_{d,n}$, define $G_{d,n}(b)$ as follows:

In many mathematical contexts, an "embedding" refers to a map that is an isomorphism onto its image.

$$G_{d,n}(b) = \begin{cases} [b_{(d,0,\ldots,0)} : b_{(d-1,1,\ldots,0)} : \cdots : b_{(d-1,0,\ldots,1)}] & b_{(d,0,\ldots,0)} \neq 0 \\ [b_{(1,d-1,\ldots,0)} : b_{(0,d,\ldots,0)} : \cdots : b_{(0,d-1,\ldots,1)}] & b_{(0,d,\ldots,0)} \neq 0 \\ \vdots & \vdots \\ [b_{(1,0,\ldots,d-1)} : b_{(0,1,\ldots,d-1)} : \cdots : b_{(0,0,\ldots,d)}] & b_{(0,0,\ldots,d)} \neq 0. \end{cases}$$

From the defining equations of $X_{d,n}$, it follows that the cross-ratios of any two of the above expressions for $G_{d,n}$ vanish on $X_{d,n}$, showing that $G_{d,n}$ is well-defined at all points where at least one of the coordinates $b_{(d,0,\ldots,0)}$, $b_{(0,d,\ldots,0)}$, $\ldots b_{(0,0,\ldots,d)}$ is nonzero. Furthermore, as we saw in the proof of Proposition 10.16, at least one of these coordinates is nonzero for every $b \in X_{d,n}$. Thus, $G_{d,n} : X \to \mathbb{P}^n$ is a regular map. It remains to check that $F_{d,n}$ and $G_{d,n}$ are inverse functions, which is left as an exercise to the reader (Exercise 10.3.4). □

The varieties $X_{d,n} \subseteq \mathbb{P}^{\binom{n+d}{d}}$ are called *Veronese varieties*. Up to isomorphism, Proposition 10.20 shows that the Veronese variety $X_{d,n}$ is intrinsically nothing more than a different perspective on projective space $\mathbb{P}^n$. However, by studying the various Veronese models of $\mathbb{P}^n$, it is possible to reduce the maximum degree of the defining polynomials of a projective variety $X \subseteq \mathbb{P}^n$. Before describing the result in general, let us see how this process works in an example.

10.21 **EXAMPLE** Veronese image of a cubic curve

Let $X = \mathcal{V}(x_0^3 + x_1^3 + x_2^3) \subseteq \mathbb{P}^2$, and consider the Veronese map

$$F_{3,2} : X \to \mathbb{P}^{\binom{5}{3}-1} = \mathbb{P}^9.$$

Recall that the coordinates y_D of $\mathbb{P}^3$ are indexed by triples $D = (d_0, d_1, d_2)$ such that $d_0 + d_1 + d_2$, and observe that, for any $a = [a_0 : a_1 : a_2] \in \mathbb{P}^2$, we have

$$a \in X \iff a_0^3 + a_1^3 + a_2^3 = 0 \iff F_{3,2}(a) \in \mathcal{V}(y_{(3,0,0)} + y_{(0,3,0)} + y_{(0,0,3)}).$$

It then follows that

$$F_{3,2}(X) = X_{3,2} \cap \mathcal{V}(y_{(3,0,0)} + y_{(0,3,0)} + y_{(0,0,3)}).$$

Moreover, by Proposition 10.20, $F_{3,2}$ gives an isomorphism $\mathbb{P}^2 \cong X_{3,2}$, and it follows that the restriction of $F_{3,2}$ gives an isomorphism

$$X \cong X_{3,2} \cap \mathcal{V}(y_{(3,0,0)} + y_{(0,3,0)} + y_{(0,0,3)}).$$

Since $X_{3,2}$ is defined by quadratic polynomials (Proposition 10.16), this shows that X is isomorphic to a projective variety that can be defined by polynomials of degree at most two, even though our original expression for X described it as the vanishing of a cubic polynomial.

Building upon the previous example, we now describe the general result.

10.22 **PROPOSITION** *Projective varieties are defined by quadratics*

Up to isomorphism, every projective variety can be written as the vanishing of a finite set of homogeneous polynomials of degree at most two.

PROOF Let $X \subseteq \mathbb{P}^n$ be a projective variety. We first observe (Exercise 10.3.5) that we can write $X = \mathcal{V}(f_1, \ldots, f_k)$ where each $f_i \in K[x_0, \ldots, x_n]$ is homogeneous of the same degree d. In other words, we can write each f_i as

$$f_i = \sum_D a_{i,D} x_0^{d_0} \cdots x_n^{d_n}$$

where the sum is over all tuples $D = (d_0, \ldots, d_n)$ of nonnegative integers that sum to d and $a_{i,D} \in K$. Consider the Veronese map

$$F_{d,n} : \mathbb{P}^n \to \mathbb{P}^{\binom{n+d}{d}-1},$$

and define linear polynomials $\ell_1, \ldots, \ell_k$ in the variables of the codomain by

$$\ell_i = \sum_D a_{i,D} y_D.$$

As in Example 10.21, for any $a \in \mathbb{P}^n$, we have

$$a \in X \iff F_{d,n}(a) \in X_{d,n} \cap \mathcal{V}(\ell_1, \ldots, \ell_k),$$

implying that $F_{d,n}(X) = X_{d,n} \cap \mathcal{V}(\ell_1, \ldots, \ell_k)$. Since $F_{d,n} : \mathbb{P}^n \to \mathbb{P}^{\binom{n+d}{d}-1}$ is a regular map, it restricts to a regular map

$$F_{d,n} : X \to X_{d,n} \cap \mathcal{V}(\ell_1, \ldots, \ell_k).$$

Moreover, since $F_{d,n}$ is an isomorphism onto $X_{d,n}$ (Proposition 10.20), it then follows that $F_{d,n}^{-1} : X_{d,n} \to \mathbb{P}^n$ restricts to an inverse regular map

$$F_{d,n}^{-1} : X_{d,n} \cap \mathcal{V}(\ell_1, \ldots, \ell_k) \to X.$$

Thus, $X \cong X_{d,n} \cap \mathcal{V}(\ell_1, \ldots, \ell_k)$, and the result now follows from the fact that $X_{d,n}$ can be defined by quadratics (Proposition 10.16) while each ℓ_i is linear. □

It is worth noting that, while Proposition 10.22 reduces the maximum degree of the defining polynomials of a projective variety, it generally increases the number of defining polynomials quite drastically. For instance, in Example 10.21, dimension arguments (which will be made precise in the next chapter) imply that we require at least 8 linear and quadratic polynomials to describe the Veronese image of the cubic curve, even though the cubic curve, itself, only required one defining polynomial.

As a final application of Veronese maps, the next result says that all polynomial maps are, up to isomorphism on the domain, linear maps.

10.23 **PROPOSITION** *Up to isomorphism, polynomial maps are linear*

If $F : X \to Y$ is a polynomial map of projective varieties, then there exists an isomorphism $G : X \to Z$ and a linear map $L : Z \to Y$ such that $F = L \circ G$.

PROOF See Exercise 10.3.6. □

An interesting consequence of Proposition 10.23 is the following result.

10.24 **COROLLARY** *Polynomial maps are finite-to-one*

If $F : X \to Y$ is a polynomial map of projective varieties, then for any $b \in Y$, there are finitely many $a \in X$ such that $F(a) = b$.

PROOF With notation as in the statement of Proposition 10.23, any polynomial map F can be written as a composition $L \circ G$ where G is an isomorphism and L is linear. The result then follows from the observation that isomorphisms are one-to-one, while linear maps are finite-to-one (Exercise 10.2.6). □

Corollary 10.24 is another indication of just how restrictive polynomial maps are, providing additional justification for why it is important to work with the more flexible notion of regular maps, where polynomiality is only assumed locally.

Exercises for Section 10.3

10.3.1 Let $d, n \geq 1$ be integers and let $SB_{d,n}$ denote the sequences of d "stars" and n "bars. For example, if $(d, n) = (4, 3)$, a few elements of $SB_{d,n}$ are

$$(\star \mid \star\ \star \mid \mid \star), \qquad (\mid \mid \mid \star\ \star\ \star\ \star), \qquad \text{and} \qquad (\star \mid \star \mid \star \mid \star).$$

(a) Explain why $SB_{d,n}$ has $\binom{d+n}{d}$ elements.

(b) Describe a bijection between $SB_{d,n}$ and the set of monomials of degree d in the variables $x_0, x_1, \ldots, x_n$.

(c) Conclude from (a) and (b) that there are exactly $\binom{d+n}{d}$ monomials of degree d in the variables $x_0, x_1, \ldots, x_n$.

10.3.2 Prove that the dth Veronese map $F_{d,n}$ is regular at every point of $\mathbb{P}^n$.

10.3.3 Let $X_{d,n} \subseteq \mathbb{P}^{\binom{d+n}{d}-1}$ be the Veronese variety and consider a point $b \in X_{d,n}$. The exercise proves that the product of coordinates

$$b_{D_1} \cdots b_{D_k} \in K$$

only depends on $D_1 + \cdots + D_k \in \mathbb{N}^{n+1}$. Fix a tuple $(e_1, \ldots, e_n) \in \mathbb{N}^{n+1}$ such that $e_0 + \cdots + e_n = dk$ and let $\mathcal{S}$ denote the sequences of length dk that contain e_0 entries equal to 0, e_1 entries equal to 1, and so on. For each $\sigma \in \mathcal{S}$, define $b_\sigma \in K$ by setting

$$b_\sigma = b_{D_1} \cdots b_{D_k}$$

where the ith entry of D_j is the number of is in the jth subsequence of σ of length d.

(a) To parse notation, consider the tuple $(e_0, e_1, e_2) = (3, 2, 4)$ where we take $n = 2$ and $d = k = 3$. Write down three examples of sequences $\sigma \in \mathcal{S}$ and the corresponding values $b_\sigma \in K$.

(b) Prove that the function $\varphi : \mathcal{S} \to K$ sending σ to b_σ is a surjection onto

$$\{b_{D_1} \cdots b_{D_k} \mid D_1 + \cdots + D_k = (e_1, \ldots, e_n)\}.$$

(c) Suppose that σ and σ' are two sequences in $\mathcal{S}$ that differ by a transposition of adjacent terms. Use the defining equations of $X_{d,n}$ to prove that $\varphi(\sigma) = \varphi(\sigma')$.

(d) Using the fact that adjacent transpositions generate all permutations, conclude that $b_{D_1} \cdots b_{D_k}$ only depends on $D_1 + \cdots + D_k$.

10.3.4 With notation as in the proof of Proposition 10.20, prove that $F_{d,n}$ and $G_{d,n}$ are inverse functions.

10.3.5 (a) Let $f \in K[x_0, \ldots, x_n]$. Prove that

$$\mathcal{V}_{\mathbb{P}}(f) = \mathcal{V}_{\mathbb{P}}(x_0^d f, \ldots, x_n^d f).$$

(b) Prove that every projective variety can be defined by a finite set of homogeneous polynomials that all have the same degree.

10.3.6 Prove Proposition 10.23. (Hint: Take G to be the restriction of a Veronese map.)

Section 10.4 Segre maps and products

In the final section of this chapter, our aim is to study products of projective varieties. The notion of products is quite a bit more involved in the projective case than it was in the affine case. The complications stem from the fact that, while there was a natural identification $\mathbb{A}^m \times \mathbb{A}^n = \mathbb{A}^{m+n}$ in the affine setting, no such identification exists between $\mathbb{P}^m \times \mathbb{P}^n$ and $\mathbb{P}^{m+n}$, as the reader is encouraged to ponder. In fact, it's not even obvious at the onset how to think of $\mathbb{P}^m \times \mathbb{P}^n$, itself, as a projective variety. Our first task is to interpret $\mathbb{P}^m \times \mathbb{P}^n$ as a projective variety, and we accomplish this using Segre maps, named in honor of the Italian mathematician Corrado Segre (1863–1924).

10.25 **DEFINITION** *Segre maps*

For any $m, n \geq 0$, the *Segre map* $S_{m,n} : \mathbb{P}^m \times \mathbb{P}^n \to \mathbb{P}^{(m+1)(n+1)-1}$ is the function

$$S_{m,n} = [x_0y_0 : x_0y_1 : \cdots : x_my_{n-1} : x_my_n],$$

where the monomials in the definition are all possible products x_iy_j.

To parse the definition, let us take a look at the first interesting example.

10.26 **EXAMPLE** The Segre map $S_{1,1}$

Consider the Segre map in the case $m = n = 1$:

$$\begin{aligned} S_{1,1} : \mathbb{P}^1 \times \mathbb{P}^1 &\to \mathbb{P}^3 \\ ([a_0 : a_1], [b_0 : b_1]) &\mapsto [a_0b_0 : a_0b_1 : a_1b_0 : a_1b_1]. \end{aligned}$$

Note that scaling the homogeneous coordinates within either $\mathbb{P}^1$ simply results in a uniform scaling of all of the coordinates in the image. Moreover, one checks that

$$a_0b_0 = a_0b_1 = a_1b_0 = a_1b_1 = 0 \quad \Longrightarrow \quad a_0 = a_1 = 0 \quad \text{or} \quad b_0 = b_1 = 0.$$

These two observations, together, imply that $S_{1,1}$ is well-defined at every point of $\mathbb{P}^1 \times \mathbb{P}^1$. Generalizing this, the reader is encouraged to verity that $S_{m,n}$ is well-defined at every point of $P^m \times \mathbb{P}^n$ (Exercise 10.4.1).

Even though Segre maps are well-defined at every point of their domain, it does not make sense to ask whether they are regular maps, because we have not yet given the domain the structure of a projective variety. In fact, we will use the Segre map to do just that, by showing that the Segre map is a bijection of $\mathbb{P}^m \times \mathbb{P}^n$ with a projective variety in $\mathbb{P}^{(m+1)(n+1)-1}$. To accomplish this task, it will be useful to introduce notation that allows us to conveniently organize the coordinates of $\mathbb{P}^{(m+1)(n+1)-1}$.

Observe that $\mathbb{P}^{(m+1)(n+1)-1}$ has $(m+1)(n+1)$ homogeneous coordinates; we denote these coordinates by z_{ij} with $0 \leq i \leq m$ and $0 \leq j \leq n$. Conveniently, these coordinates naturally organize into an $m \times n$ matrix $A(z)$, whose ij-entry is z_{ij}. With this labeling, we take the convention that the z_{ij}-coordinate of the Segre map $S_{m,n}$ is x_iy_j. Using this notation, we now describe the image of the Segre maps.

10.27 **PROPOSITION** *The Segre map injects onto a projective variety*

For any $m, n \geq 0$, the Segre map $S_{m,n}$ is an injection of $\mathbb{P}^m \times \mathbb{P}^n$ onto

$$Z_{m,n} = \mathcal{V}(\{z_{ij}z_{k\ell} - z_{i\ell}z_{kj} \mid 0 \leq i,k \leq m,\ 0 \leq j,\ell \leq n\}) \subseteq \mathbb{P}^{(m+1)(n+1)-1}.$$

$Z_{m,n}$ is defined by all 2×2 minors of the matrix $A(z)$, which is an example of a "determinantal variety."

PROOF We prove that the image of $S_{m,n}$ is equal to $Z_{m,n}$, and we leave the verification that $S_{m,n}$ is injective as an exercise (Exercise 10.4.2).

To prove that the image of $S_{m,n}$ is contained in $Z_{m,n}$, it suffices to show that $z_{ij}z_{k\ell} - z_{i\ell}z_{kj}$ vanishes when evaluated at any point in the image of $S_{m,n}$, which follows from the observation that

$$(a_ib_j)(a_kb_\ell) - (a_ib_\ell)(a_kb_j) = 0 \quad \text{for any} \quad (a,b) \in \mathbb{P}^m \times \mathbb{P}^n.$$

Conversely, to prove that $Z_{m,n}$ is contained in the image of $S_{m,n}$, let $c \in Z_{m,n}$. This means that all of the 2×2 minors of $A(c)$ vanish, implying that the rank of $A(c)$ is at most one, or in other words, that every row of $A(c)$ is a multiple of some row. Since there must be at least one nonzero row, assume without loss of generality that the 0th row is nonzero, and denote it by $b = (b_0, \ldots, b_n)$. Let $a_i \in K$ be the value such that the ith row of $A(c)$ is equal to a_ib. One readily checks that

$$c = S_{m,n}([1 : a_1 : \cdots : a_m], [b_0 : b_1 : \cdots : b_n]),$$

showing that c is in the image of $S_{m,n}$. □

10.28 **EXAMPLE** The image of $S_{1,1}$

Considering again the case $m = n = 1$, denote the coordinates on $\mathbb{P}^3$ by z_{00}, z_{01}, z_{10}, and z_{11}. Proposition 10.27 shows that $S_{1,1}$ is an injection onto the projective variety in $\mathbb{P}^3$ defined by vanishing of the 2×2 determinant

$$\det \begin{bmatrix} z_{00} & z_{01} \\ z_{10} & z_{11} \end{bmatrix} = z_{00}z_{11} - z_{01}z_{10}.$$

The importance of Proposition 10.27 is that it identifies the product $\mathbb{P}^m \times \mathbb{P}^n$ with a projective variety in $\mathbb{P}^{(m+1)(n+1)-1}$, thereby giving us a natural way to view the product, itself, as a projective variety. The next result shows that we can use Segre maps to naturally view the product of any two projective varieties—not just projective spaces—as a projective variety.

10.29 **PROPOSITION** *Products of projective varieties*

If $X \subseteq \mathbb{P}^m$ and $Y \subseteq \mathbb{P}^n$ are projective varieties, then the Segre map $S_{m,n}$ restricts to an injection of $X \times Y$ onto a projective variety in $\mathbb{P}^{(m+1)(n+1)-1}$.

PROOF Suppose that $X = \mathcal{V}(\mathcal{S})$ and $Y = \mathcal{V}(\mathcal{T})$, where $\mathcal{S} \subseteq K[x_0, \ldots, x_m]$ and $\mathcal{T} \subseteq K[y_0, \ldots, y_n]$ are finite sets of homogeneous polynomials. For every $f \in \mathcal{S}$ and $g \in \mathcal{T}$ and for each $0 \leq i \leq n$ and $0 \leq j \leq m$, define

$$f_i = f(z_{0i}, z_{1i}, \ldots, z_{mi}) \quad \text{and} \quad g_j = g(z_{j0}, z_{j1}, \ldots, z_{jn}).$$

Let $\mathcal{S}'$ be the collection of all such f_i and let $\mathcal{T}'$ be the collection of all such g_j. The reader is encouraged to verify (Exercise **??**) that

$$S_{m,n}(X \times Y) = S_{m,n}(\mathbb{P}^m \times \mathbb{P}^n) \cap \mathcal{V}(\mathcal{S}' \cup \mathcal{T}').$$

By Proposition 10.27, we know that $S_{m,n}$ is an injection and that $S_{m,n}(\mathbb{P}^m \times \mathbb{P}^n)$ is a projective variety. Since intersections of projective varieties are, themselves, projective varieties, we conclude that $S_{m,n}(X \times Y)$ is a projective variety. □

Since the Segre map identifies $X \times Y$ with $S_{m,n}(X \times Y)$, we now use this identification to define products within the realm of projective varieties.

10.30 **DEFINITION** *Product of projective varieties*

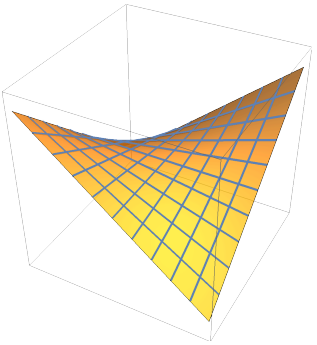
If $X \subseteq \mathbb{P}^m$ and $Y \subseteq \mathbb{P}^n$ are projective varieties, then their *product* $X \times Y$ (as a projective variety) is the projective variety

$$S_{m,n}(X \times Y) \subseteq \mathbb{P}^{(m+1)(n+1)-1}.$$

10.31 **EXAMPLE** A doubly-ruled surface

Let us pause to visualize the projective variety $\mathbb{P}^1 \times \mathbb{P}^1$, which, by Definition 10.30 and Example 10.28, is equal to the projective variety $\mathcal{V}(z_{00}z_{11} - z_{01}z_{10}) \subseteq \mathbb{P}^3$. Consider the affine chart where $z_{00} \neq 0$, which is the affine surface defined by

$$\mathcal{V}_{\mathbb{A}}(z_{11} - z_{01}z_{10}) \subseteq \mathbb{A}^3.$$



The image to the right is a depiction of this affine chart over $\mathbb{R}$, from which we can see that the surface is doubly-ruled: it can be viewed as a disjoint union of lines in two different ways. The double-ruling in this affine chart is reflecting the more general fact that the projective surface $\mathbb{P}^1 \times \mathbb{P}^1 \subseteq \mathbb{P}^3$ is also doubly-ruled: it can be realized as a disjoint union of (projective) lines in $\mathbb{P}^3$ in two different ways. The two different rulings of $\mathbb{P}^1 \times \mathbb{P}^1$ are given by the subsets in $\mathbb{P}^3$ of the form

$$\{a\} \times \mathbb{P}^1 \quad \text{and} \quad \mathbb{P}^1 \times \{b\},$$

as the reader is encouraged to explore in Exercise 10.4.4.

Having realized the product $X \times Y$ as a projective variety, it is now possible discuss regular maps to and from $X \times Y$. The following result addresses the two most important examples of regular maps from a product: the projection maps.

10.32 **PROPOSITION** *Projection maps are regular*

If $X \subseteq \mathbb{P}^m$ and $Y \subseteq \mathbb{P}^n$ are projective varieties, then the projection maps $\pi_1 : X \times Y \to X$ and $\pi_2 : X \times Y \to Y$ are regular.

PROOF Without loss of generality, we focus on π_1. It suffices to note that, for any $c = (a, b) \in X \times Y \subseteq \mathbb{P}^{(m+1)(n+1)-1}$, we can write $\pi_1(c)$ using the following piecewise polynomial expressions:

$$\pi_1(c) = \begin{cases} [c_{00} : \cdots : c_{m0}] & c \notin \mathcal{V}(z_{00}, \ldots, z_{m0}) \\ \quad \vdots & \quad \vdots \\ [c_{0n} : \cdots : c_{mn}] & c \notin \mathcal{V}(z_{0n}, \ldots, z_{mn}). \end{cases}$$

To verify the above expression for π_1, suppose that $c = (a, b) \in X \times Y$. Then the z_{ij}-coordinate of c is $a_i b_j$. Note that at least one of the coordinates of a and at least one of the coordinates of b is nonzero. If $b_j \neq 0$, it then follows that $c \notin \mathcal{V}(z_{0j}, \ldots, z_{m,j})$, and we compute

$$[c_{0j} : \cdots : c_{mj}] = [a_0 b_j : \cdots : a_m b_j] = [a_0 : \cdots : a_m] = \pi_1(c),$$

verifying the above piecewise polynomial expressions for $\pi_1(c)$. □

Just as we can take products of projective varieties, we can also take products of maps to each factor. The next result addresses products of regular maps.

10.33 **PROPOSITION** *Products of regular maps are regular*

Let $W \subseteq \mathbb{P}^\ell$, $X \subseteq \mathbb{P}^m$, and $Y \subseteq \mathbb{P}^n$ be projective varieties. If $F : W \to X$ and $G : W \to Y$ are regular maps, then the function

$$\begin{aligned} F \times G : W &\to X \times Y \\ a &\mapsto (F(a), G(a)) \end{aligned}$$

is a regular map.

PROOF Let $p \in W$. By the definition of regular maps, there exist homogeneous polynomials

$$f_0, \ldots, f_m, g_0, \ldots, g_n \in K[w_0, \ldots, x_\ell]$$

such that $p \notin \mathcal{V}(f_0, \ldots, f_m) \cup \mathcal{V}(g_0, \ldots, g_n)$, and

$$F(a) = [f_0(a) : \cdots : f_m(a)] \quad \text{for all} \quad a \in W \setminus \mathcal{V}(f_0, \ldots, f_m)$$

and

$$G(a) = [g_0(a) : \cdots : g_n(a)] \quad \text{for all} \quad a \in W \setminus \mathcal{V}(g_0, \ldots, g_n).$$

It then follows that $p \notin \mathcal{V}(\{f_i g_j \mid 0 \le i \le m,\ 0 \le j \le n\})$ and, by the definition of products of projective varieties, we have

$$(F \times G)(a) = [f_0(a)g_0(a) : f_0(a)g_1(a) : \cdots : f_m(a)g_{n-1}(a) : f_m(a)g_n(a)]$$

for all $a \in W \setminus \mathcal{V}(\{f_i g_j \mid 0 \le i \le m,\ 0 \le j \le n\})$. Thus, $F \times G$ can be locally described by the polynomials $f_i g_j$, verifying that $F \times G$ is regular. □

An important consequence of the previous two results is the following, which shows that the intrinsic nature of products depends only on the intrinsic nature of each factor.

10.34 **COROLLARY** *Products preserve isomorphisms*

Let X, X', Y, Y' be projective varieties. If $X \cong X'$ and $Y \cong Y'$, then

$$X \times Y \cong X' \times Y'.$$

PROOF Let $F : X \to X'$ and $G : Y \to Y'$ be isomorphisms. Since the projection maps are regular, precomposing F and G with the projection maps yield regular maps

$$F \circ \pi_1 : X \times Y \to X' \quad \text{and} \quad G \circ \pi_2 : X \times Y \to Y'.$$

The product of these two regular maps is then the regular map

$$\begin{aligned} X \times Y &\to X' \times Y' \\ (a, b) &\mapsto (F(a), G(b)). \end{aligned}$$

Since F and G are isomorphisms, they have regular inverses, and repeating the above procedure with F^{-1} and G^{-1} yields a regular map

$$\begin{aligned} X' \times Y' &\to X \times Y \\ (a, b) &\mapsto (F^{-1}(a), G^{-1}(b)). \end{aligned}$$

The above pair of regular maps between $X \times Y$ and $X' \times Y'$ are inverse to each other, from which we conclude that $X \times Y \cong X' \times Y'$. □

Exercises for Section 10.4

10.4.1 For any $a = [a_0 : \cdots : a_m] \in \mathbb{P}^m$ and $b = [b_0 : \cdots : b_n] \in \mathbb{P}^n$, let

$$S_{m,n}(a, b) = [a_0 b_0 : a_0 b_1 : \cdots : a_m b_{n-1} : a_m a_n] \in \mathbb{P}^{(m+1)(n+1)-1}.$$

(a) Prove that $S_{m,n}(a, b)$ does not depend on the choice of homogeneous coordinates for a or b.

(b) Prove that at least one coordinate of $S_{m,n}(a, b)$ must be nonzero.

10.4.2 Prove that the Segre map $S_{m,n} : \mathbb{P}^m \times \mathbb{P}^n \to \mathbb{P}^{(m+1)(n+1)-1}$ is an injection.

10.4.3 Let $X = \mathcal{V}(\mathcal{S}) \subseteq \mathbb{P}^m$ and $Y = \mathcal{V}(\mathcal{T}) \subseteq \mathbb{P}^n$ where $\mathcal{S} \subseteq K[x_0, \ldots, x_m]$ and $\mathcal{T} \subseteq K[y_0, \ldots, y_n]$ are finite sets of homogeneous polynomials. For every $f \in \mathcal{S}$ and $g \in \mathcal{T}$ and for each $0 \leq i \leq n$ and $0 \leq j \leq m$, define

$$f_i = f(z_{0i}, z_{1i}, \ldots, z_{mi}) \quad \text{and} \quad g_j = g(z_{j0}, z_{j1}, \ldots, z_{jn}).$$

Let S' be the collection of all such f_i and let $\mathcal{T}'$ be the collection of all such g_j. Prove that

$$S_{m,n}(X \times Y) = S_{m,n}(\mathbb{P}^m \times \mathbb{P}^n) \cap \mathcal{V}(\mathcal{S}' \cup \mathcal{T}').$$

10.4.4 Prove that the subsets of $\mathbb{P}^1 \times \mathbb{P}^1$ of the form $\{a\} \times \mathbb{P}^1$ and $\mathbb{P}^1 \times \{b\}$ are lines in $\mathbb{P}^3$. In other words, describe each of these subsets as the image of a linear map $\mathbb{P}^1 \to \mathbb{P}^3$.

10.4.5 Let $X \subseteq \mathbb{P}^m$ and $Y \subseteq \mathbb{P}^n$ be projective varieties. Prove that $X \times Y$ and $Y \times X$ are projectively equivalent in $\mathbb{P}^{(m+1)(n+1)-1}$. Conclude that $X \times Y \cong Y \times X$.

Chapter 11

Quasiprojective Varieties

Coming soon!