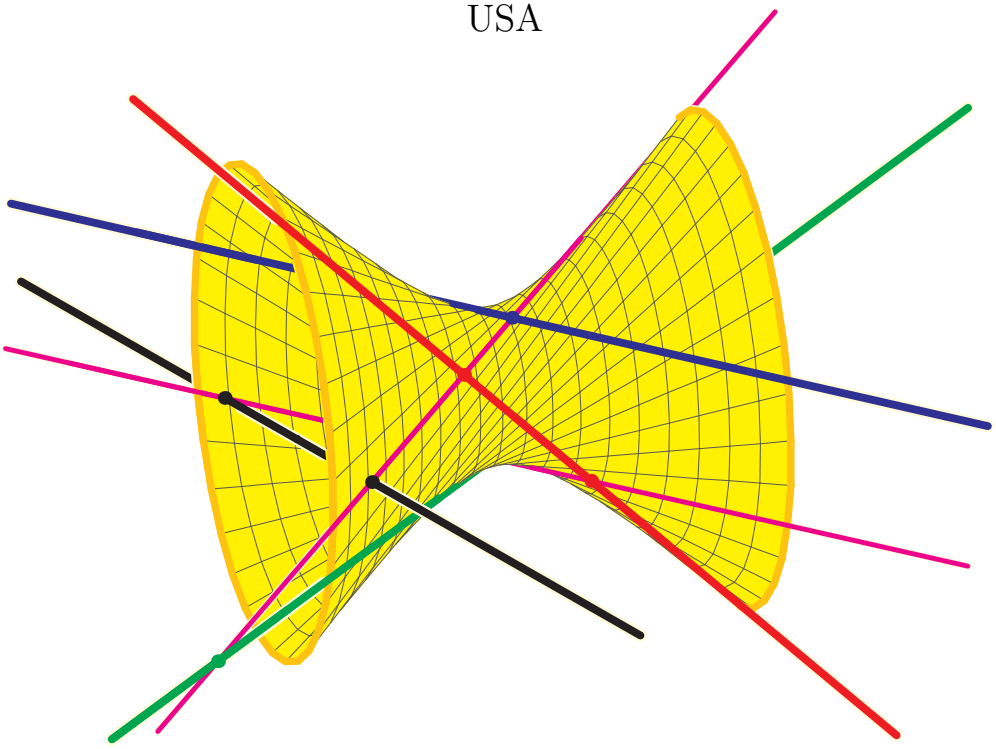
Discrete and Applicable Algebraic Geometry

Frank Sottile
Department of Mathematics
Texas A&M University
College Station
Texas 77843
USA



October 28, 2009

These notes were developed during a course that Sottile taught at Texas A&M University in January and February 2007, and were further refined for the IMA PI Summer Graduate Program on Applicable Algebraic Geometry. They are intended to accompany Sottile's Lecture series, " Introduction to algebraic geometry and Gröbner bases" and "Projective varieties and Toric Varieties" at the Summer School. Their genesis was in notes from courses taught at the University of Wisconsin at Madison, the University of Massachusetts, and Texas A&M University.

The lectures do not linearly follow the notes, as there was a need to treat material on Gröbner bases nearly at the beginning. These notes are an incomplete work in progress.

We thank Luis Garcia, who provided many of the exercises.

—Frank Sottile
College Station
20 August 2007

This material is based upon work supported by the National Science Foundation under Grant No. 0538734

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Chapter 1

Varieties

Outline:

1. Examples of affine varieties.
2. The algebra-geometry dictionary.
3. Zariski topology.
4. Irreducible decomposition and dimension.
5. Regular functions. The algebra-geometry dictionary II.
6. Rational functions.
7. Smooth and singular points.

1.1 Affine Varieties

Let $\mathbb{F}$ be a field, which for us will almost always be either the complex numbers $\mathbb{C}$, the real numbers $\mathbb{R}$, or the rational numbers $\mathbb{Q}$. These different fields have their individual strengths and weaknesses. The complex numbers are *algebraically closed*; every univariate polynomial has a complex root. Algebraic geometry works best when using an algebraically closed field, and most introductory texts restrict themselves to the complex numbers. However, quite often real number answers are needed, particularly for applications. Because of this, we will often consider real varieties and work over $\mathbb{R}$. Symbolic computation provides many useful tools for algebraic geometry, but it requires a field such as $\mathbb{Q}$, which can be represented on a computer. Much of what we do remains true for arbitrary fields, such as $\mathbb{C}(t)$, the field of rational functions in the variable t, or in finite fields. We will at times use this added generality.

The set of all n-tuples $(a_1, \ldots, a_n)$ of numbers in $\mathbb{F}$ is called *affine n-space* and written $\mathbb{A}^n$ or $\mathbb{A}^n_{\mathbb{F}}$ when we want to indicate our field. We write $\mathbb{A}^n$ rather than $\mathbb{F}^n$ to emphasize that we are not doing linear algebra. Let $x_1, \ldots, x_n$ be variables, which we regard as coordinate functions on $\mathbb{A}^n$ and write $\mathbb{F}[x_1, \ldots, x_n]$ for the ring of polynomials in the variables $x_1, \ldots, x_n$ with coefficients in the field $\mathbb{F}$. We may evaluate a polynomial $f \in$

$\mathbb{F}[x_1, \ldots, x_n]$ at a point $a \in \mathbb{A}^n$ to get a number $f(a) \in \mathbb{F}$, and so polynomials are also functions on $\mathbb{A}^n$. We make our main definition.

Definition 1.1.1 An *affine variety* is the set of common zeroes of a collection of polynomials. Given a set $S \subset \mathbb{F}[x_1, \ldots, x_n]$ of polynomials, the affine variety defined by S is the set

$$\mathcal{V}(S) \;:=\; \{a \in \mathbb{A}^n \mid f(a) = 0 \quad \text{for } f \in S\}\,.$$

This is a(n affine) *subvariety* of $\mathbb{A}^n$ or simply a *variety* or algebraic variety.

If X and Y are varieties with $Y \subset X$, then Y is a *subvariety* of X.

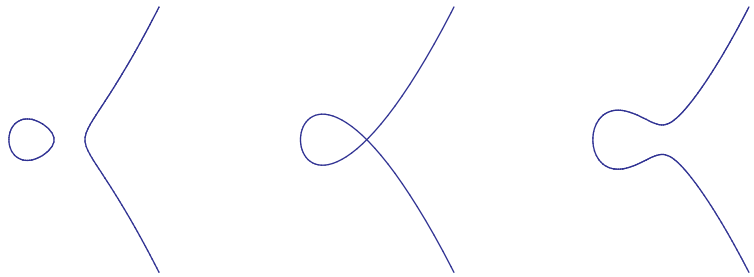
The empty set $\emptyset = \mathcal{V}(1)$ and affine space itself $\mathbb{A}^n = \mathcal{V}(0)$ are varieties. Any linear or affine subspace L of $\mathbb{A}^n$ is a variety. Indeed, an affine subspace L has an equation $Ax = b$, where A is a matrix and b is a vector, and so $L = \mathcal{V}(Ax - b)$ is defined by the linear polynomials which form the rows of the column vector $Ax - b$. An important special case is when $L = \{a\}$ is a point of $\mathbb{A}^n$. Writing $a = (a_1, \ldots, a_n)$, then L is defined by the equations $x_i - a_i = 0$ for $i = 1, \ldots, n$.

Any finite subset $Z \subset \mathbb{A}^1$ is a variety as $Z = \mathcal{V}(f)$, where

$$f \;:=\; \prod_{z \in Z} (x - z)$$

is the monic polynomial with simple zeroes in Z.

A non-constant polynomial $p(x, y)$ in the variables x and y defines a *plane curve* $\mathcal{V}(p) \subset \mathbb{A}^2$. Here are the plane cubic curves $\mathcal{V}(p + \frac{1}{20})$, $\mathcal{V}(p)$, and $\mathcal{V}(p - \frac{1}{20})$, where $p(x, y) := y^2 - x^3 - x^2$.



A *quadric* is a variety defined by a single quadratic polynomial. In $\mathbb{A}^2$, the shooth quadrics are the plane conics (circles, ellipses, parabolas, and hyperbolas in $\mathbb{R}^2$) and in $\mathbb{R}^3$, the smooth quadrics are the spheres, ellipsoids, paraboloids, and hyperboloids. Figure 1.1 shows a hyperbolic paraboloid $\mathcal{V}(xy+z)$ and a hyperboloid of one sheet $\mathcal{V}(x^2-x+y^2+yz)$.

These examples, finite subsets of $\mathbb{A}^1$, plane curves, and quadrics, are varieties defined by a single polynomial and are called *hypersurfaces*. Any variety is an intersection of hypersurfaces, one for each polynomial defining the variety.

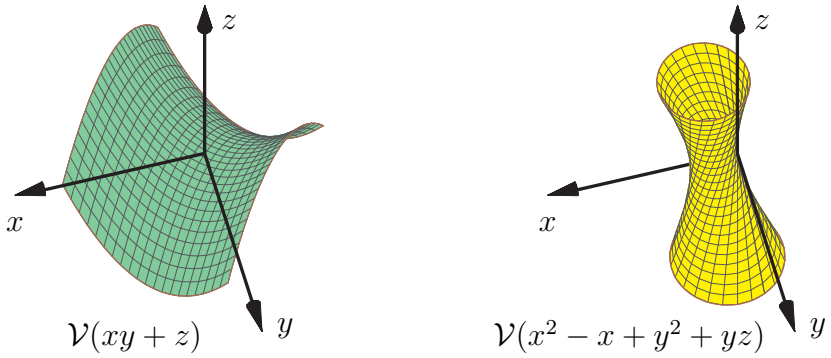
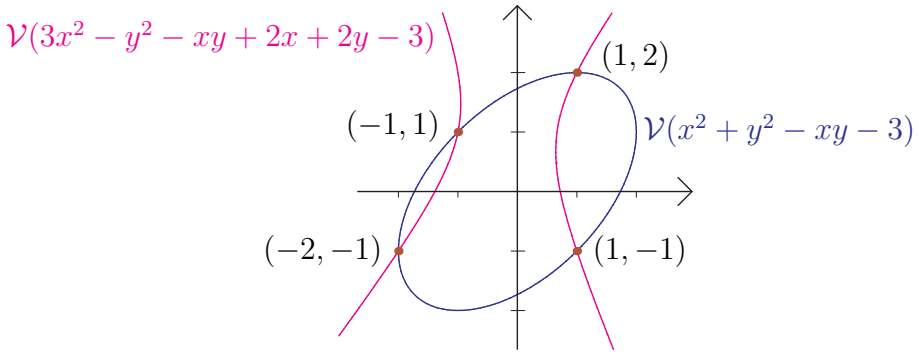


Figure 1.1: Two hyperboloids.

The set of four points $\{(-2,-1),(-1,1),(1,-1),(1,2)\}$ in $\mathbb{A}^2$ is a variety. It is the intersection of an ellipse $\mathcal{V}(x^2+y^2-xy-3)$ and a hyperbola $\mathcal{V}(3x^2-y^2-xy+2x+2y-3)$.



The quadrics of Figure 1.1 meet in the variety $\mathcal{V}(xy+z,\ x^2-x+y^2+yz)$, which is shown on the right in Figure 1.2. This intersection is the union of two space curves. One is the line $x=1, y+z=0$, while the other is the cubic space curve which has parametrization $(t^2,t,-t^3)$.

The intersection of the hyperboloid $x^2+(y-\frac{3}{2})^2-z^2=\frac{1}{4}$ with the sphere $x^2+y^2+z^2=4$ is a singular space curve (the figure ∞ on the left sphere in Figure 1.3). If we instead intersect the hyperboloid with the sphere centered at the origin having radius 1.9, then we obtain the smooth quartic space curve drawn on the right sphere in Figure 1.3.

The product $X\times Y$ of two varieties X and Y is again a variety. Suppose that $X\subset\mathbb{A}^n$ is defined by the polynomials $f_1,\ldots,f_s\in\mathbb{F}[x_1,\ldots,x_n]$ and the variety $Y\subset\mathbb{A}^m$ is defined by the polynomials $g_1,\ldots,g_t\in\mathbb{F}[y_1,\ldots,y_m]$. Then $X\times Y\subset\mathbb{A}^n\times\mathbb{A}^m=\mathbb{A}^{n+m}$ is defined by the polynomials $f_1,\ldots,f_s,g_1,\ldots,g_t\in\mathbb{F}[x_1,\ldots,x_n,\,y_1,\ldots,y_m]$. Given a point $x\in X$, the product $\{x\}\times Y$ is a subvariety of $X\times Y$ which may be identified with Y simply by forgetting the coordinate x.

The set $\mathrm{Mat}_{n\times n}$ or $\mathrm{Mat}_{n\times n}(\mathbb{F})$ of $n\times n$ matrices with entries in $\mathbb{F}$ is identified with

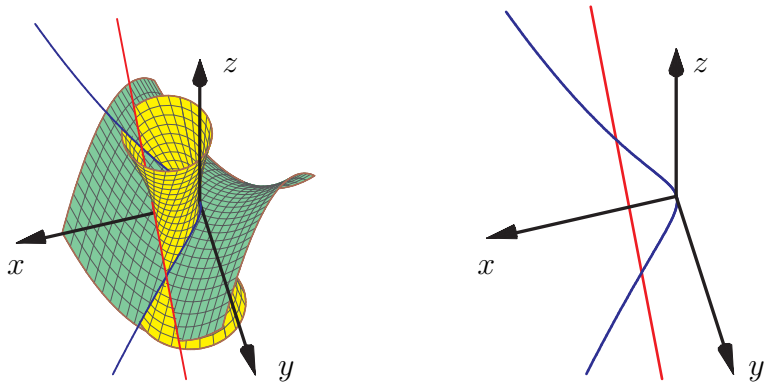


Figure 1.2: Intersection of two quadrics.

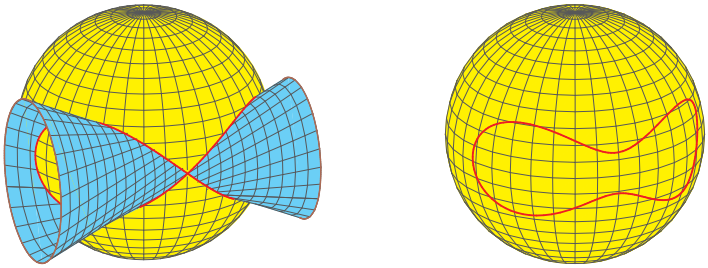


Figure 1.3: Quartics on spheres.

the affine space $\mathbb{A}^{n^2}$, which is sometimes written $\mathbb{A}^{n\times n}$. An interesting class of varieties are linear algebraic groups, which are algebraic subvarieties of $\mathrm{Mat}_{n\times n}$ which are closed under multiplication and taking inverse. The *special linear group* is the set of matrices with determinant 1,

$$SL_n \;:=\; \{M \in \mathrm{Mat}_{n\times n} \mid \det M = 1\}\,,$$

which is a linear algebraic group. Since the determinant of a matrix in $\mathrm{Mat}_{n\times n}$ is a polynomial in its entries, SL_n is the variety $\mathcal{V}(\det -1)$. We will later show that SL_n is smooth, irreducible, and has dimension $n^2 - 1$. (We must first, of course, define these notions.)

There is a general construction of other linear algebraic groups. Let g^T be the transpose of a matrix $g \in \mathrm{Mat}_{n\times n}$. For a fixed matrix $M \in \mathrm{Mat}_{n\times n}$, set

$$G_M \;:=\; \{g \in SL_n \mid gMg^T = M\}\,.$$

This a linear algebraic group, as the condition $gMg^T = M$ is n^2 polynomial equations in the entries of g, and G_M is closed under matrix multiplication and matrix inversion.

When M is skew-symmetric and invertible, G_M is a *symplectic group*. In this case, n is necessarily even. If we let J_n denote the $n \times n$ matrix with ones on its anti-diagonal,

then the matrix

$$\begin{bmatrix} 0 & J_n \\ -J_n & 0 \end{bmatrix}$$

is conjugate to every other invertible skew-symmetric matrix in $\mathrm{Mat}_{2n\times 2n}$. We assume M is this matrix and write Sp_{2n} for the symplectic group.

When M is symmetric and invertible, G_M is a *special orthogonal group*. When $\mathbb{F}$ is algebraically closed, all invertible symmetric matrices are conjugate, and we may assume $M = J_n$. For general fields, there may be many different forms of the special orthogonal group. For instance, when $\mathbb{F} = \mathbb{R}$, let k and l be, respectively, the number of positive and negative eigenvalues of M (these are conjugation invariants of M. Then we obtain the group $SO_{k,l}\mathbb{R}$. We have $SO_{k,l}\mathbb{R} \simeq SO_{l,k}\mathbb{R}$.

Consider the two extreme cases. When $l = 0$, we may take $M = I_n$, and when $|k-l| \leq 1$, we take $M = J_n$. We will write SO_n for the special orthogonal groups defined when $M = J_n$.

When $\mathbb{F} = \mathbb{R}$, this differs from the standard convention that the real special orthogonal group is $SO_{n,0}$, which is compact in the Euclidean topology. Our reason for this deviation is that we want $SO_n(\mathbb{R})$ to share more properties with $SO_n(\mathbb{C})$. Our group $SO_n(\mathbb{R})$ is often called the *split form* of the special linear group.

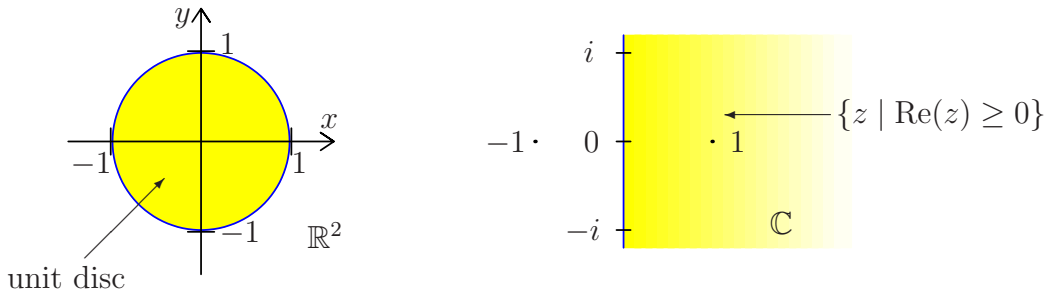
When $n = 2$, consider the two different real groups:

$$\begin{aligned} SO_{2,0}\mathbb{R} &:= \left\{ \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \mid \theta \in S^1 \right\} \\ SO_{1,1}\mathbb{R} &:= \left\{ \begin{bmatrix} a & 0 \\ 0 & a^{-1} \end{bmatrix} \mid a \in \mathbb{R}^\times \right\} \end{aligned}$$

Note that in the Euclidean topology $SO_{2,0}(\mathbb{R})$ is compact, while $SO_{1,1}(\mathbb{R})$ is not. The complex group $SO_2(\mathbb{C})$ is also not compact in the Euclidean topology.

We also point out some subsets of $\mathbb{A}^n$ which are not varieties. The set $\mathbb{Z}$ of integers is not a variety. The only polynomial vanishing at every integer is the zero polynomial, whose variety is all of $\mathbb{A}^1$. The same is true for any other infinite subset of $\mathbb{A}^1$, for example, the infinite sequence $\{\frac{1}{n} \mid n = 1, 2, \dots\}$ is not a subvariety of $\mathbb{A}^1$.

Other subsets which are not varieties (for the same reasons) include the unit disc in $\mathbb{R}^2$, $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$ or the complex numbers with positive real part.



Sets like these last two which are defined by inequalities involving real polynomials are called *semi-algebraic*. We will study them later.

Exercises for Section 1

1. Show that no proper nonempty open subset S of $\mathbb{R}^n$ or $\mathbb{C}^n$ is a variety. Here, we mean open in the usual (Euclidean) topology on $\mathbb{R}^n$ and $\mathbb{C}^n$. (Hint: Consider the Taylor expansion of any polynomial in $\mathcal{I}(S)$.)

2. Prove that in $\mathbb{A}^2$, we have $\mathcal{V}(y-x^2) = \mathcal{V}(y^3-y^2x^2,\ x^2y-x^4)$.

3. Express the cubic space curve C with parametrization (t, t^2, t^3) in each of the following ways.

   (a) The intersection of a quadric and a cubic hypersurface.

   (b) The intersection of two quadrics.

   (c) The intersection of three quadrics.

4. Let $\mathbb{A}^{n\times n}$ be the set of $n \times n$ matrices.

   (a) Show that the set $SL(n, \mathbb{F}) \subset \mathbb{A}^{n^2}_{\mathbb{F}}$ of matrices with determinant 1 is an algebraic variety.

   (b) Show that the set of singular matrices in $\mathbb{A}^{n^2}_{\mathbb{F}}$ is an algebraic variety.

   (c) Show that the set $GL(n, \mathbb{F})$ of invertible matrices is not an algebraic variety in $\mathbb{A}^{n\times n}$. Show that $GL_n(\mathbb{F})$ can be identified with an algebraic subset of $\mathbb{A}^{n^2+1} = \mathbb{A}^{n\times n} \times \mathbb{A}^1$ via a map $GL_n(\mathbb{F}) \to \mathbb{A}^{n^2+1}$.

5. An $n\times n$ matrix with complex entries is *unitary* if its columns are orthonormal under the complex inner product $\langle z, w\rangle = z \cdot \overline{w}^t = \sum_{i=1}^n z_i\overline{w_i}$. Show that the set $\mathbf{U}(n)$ of unitary matrices is not a complex algebraic variety. Show that it can be described as the zero locus of a collection of polynomials with real coefficients in $\mathbb{R}^{2n^2}$, and so it is a real algebraic variety.

6. Let $\mathbb{A}^{mn}_{\mathbb{F}}$ be the set of $m \times n$ matrices over $\mathbb{F}$.

   (a) Show that the set of matrices of rank $\leq r$ is an algebraic variety.

   (b) Show that the set of matrices of rank $= r$ is not an algebraic variety if $r > 0$.

7. (a) Show that the set $\{(t, t^2, t^3) \mid t \in \mathbb{F}\}$ is an algebraic variety in $\mathbb{A}^3_{\mathbb{F}}$.

   (b) Show that the following sets are not algebraic varieties

   (i) $\{(x, y) \in \mathbb{A}^2_{\mathbb{R}} | y = \sin x\}$

   (ii) $\{(\cos t, \sin t, t) \in \mathbb{A}^3_{\mathbb{R}} \mid t \in \mathbb{R}\}$

   (iii) $\{(x, e^x) \in \mathbb{A}^2_{\mathbb{R}} \mid x \in \mathbb{R}\}$

1.2 The algebra-geometry dictionary

The strength and richness of algebraic geometry as a subject and source of tools for applications comes from its dual nature. Intuitive geometric concepts are tamed via the precision of algebra while basic algebraic notions are enlivened by their geometric counterparts. The source of this dual nature is a correspondence between algebraic concepts and geometric concepts that we refer to as the algebraic-geometric dictionary.

We defined varieties $\mathcal{V}(S)$ associated to sets $S \subset \mathbb{F}[x_1, \ldots, x_n]$ of polynomials,

$$\mathcal{V}(S) \;=\; \{a \in \mathbb{A}^n \mid f(a) = 0 \text{ for all } f \in S\}\,.$$

We would like to invert this association. Given a subset Z of $\mathbb{A}^n$, consider the collection of polynomials that vanish on Z,

$$\mathcal{I}(Z) \;:=\; \{f \in \mathbb{F}[x_1, \ldots, x_n] \mid f(z) = 0 \text{ for all } z \in Z\}\,.$$

The map $\mathcal{I}$ reverses inclusions so that $Z \subset Y$ implies $\mathcal{I}(Z) \supset \mathcal{I}(Y)$.

These two inclusion-reversing maps

$$\{\text{Subsets } S \text{ of } \mathbb{F}[x_1, \ldots, x_n]\} \quad \underset{\mathcal{I}}{\overset{\mathcal{V}}{\rightleftarrows}} \quad \{\text{Subsets } Z \text{ of } \mathbb{A}^n\} \tag{1.1}$$

form the basis of the algebra-geometry dictionary of affine algebraic geometry. We will refine this correspondence to make it more precise.

An *ideal* is a subset $I \subset \mathbb{F}[x_1, \ldots, x_n]$ which is closed under addition and under multiplication by polynomials in $\mathbb{F}[x_1, \ldots, x_n]$. If $f, g \in I$ then $f + g \in I$ and if we also have $h \in \mathbb{F}[x_1, \ldots, x_n]$, then $hf \in I$. The ideal $\langle S \rangle$ generated by a subset S of $\mathbb{F}[x_1, \ldots, x_n]$ is the smallest ideal containing S. This is the set of all expressions of the form

$$h_1 f_1 + \cdots + h_m f_m$$

where $f_1, \ldots, f_m \in S$ and $h_1, \ldots, h_m \in \mathbb{F}[x_1, \ldots, x_n]$. We work with ideals because if f, g, and h are polynomials and $a \in \mathbb{A}^n$ with $f(a) = g(a) = 0$, then $(f + g)(a) = 0$ and $(hf)(a) = 0$. Thus $\mathcal{V}(S) = \mathcal{V}(\langle S \rangle)$, and so we may restrict $\mathcal{V}$ to the ideals of $\mathbb{F}[x_1, \ldots, x_n]$. In fact, we lose nothing if we restrict the left-hand-side of the correspondence (1.1) to the ideals of $\mathbb{F}[x_1, \ldots, x_n]$.

Lemma 1.2.1 *For any subset S of $\mathbb{A}^n$, $\mathcal{I}(S)$ is an ideal of $\mathbb{F}[x_1, \ldots, x_n]$.*

Proof. Let $f, g \in \mathcal{I}(S)$ be two polynomials which vanish at all points of S. Then $f + g$ vanishes on S, as does hf, where h is any polynomial in $\mathbb{F}[x_1, \ldots, x_n]$. This shows that $\mathcal{I}(S)$ is an ideal of $\mathbb{F}[x_1, \ldots, x_n]$. □

When S is infinite, the variety $\mathcal{V}(S)$ is defined by infinitely many polynomials. Hilbert's Basis Theorem tells us that only finitely many of these polynomials are needed.

Hilbert's Basis Theorem. *Every ideal I of $\mathbb{F}[x_1, \ldots, x_n]$ is finitely generated.*

Proof. We prove this by induction. When $n = 1$, it is not hard to show that an ideal $I \neq \{0\}$ of $\mathbb{F}[x]$ is generated by any non-zero element f of minimal degree.

Now suppose that ideals of $\mathbb{F}[x_1, \ldots, x_n]$ are finitely generated. Let $I \subset \mathbb{F}[x_1, \ldots, x_{n+1}]$ be an ideal which is not finitely generated. We may therefore inductively construct a sequence of polynomials $f_1, f_2, \ldots$ with the property that $f_i \in I$ is a polynomial of minimal degree in x_{n+1} which is not in the ideal $\langle f_1, \ldots, f_{i-1} \rangle$. Let d_i be the degree of x_{n+1} in f_i and observe that $d_1 \leq d_2 \leq \cdots$.

Let $g_i \in \mathbb{F}[x_1, \ldots, x_n]$ be the coefficient of the highest power of x_{n+1} in f_i. Then $\langle g_1, \ldots \rangle$ is an ideal of $\mathbb{F}[x_1, \ldots, x_n]$, and so it is finitely generated, say by $\langle g_1, g_2, \ldots, g_m \rangle$. In particular, there are polynomials $h_1, \ldots, h_m \in \mathbb{F}[x_1, \ldots, x_n]$ with $g_{i+1} = \sum_i h_i g_i$. Set

$$\overline{f} \;:=\; f_{m+1} \;-\; \sum_{i=1}^{m} h_i f_i x_{n+1}^{d_{m+1}-d_i} \;.$$

By construction of the h_i, the terms involving $x_{n+1}^{d_{m+1}}$ cancel on the right hand side, so that the degree of x_{n+1} in $\overline{f}$ is less than d_{m+1}. We also see that $\overline{f} \in \langle f_1, \ldots, f_m \rangle$. But this is a contradiction to our construction of $f_{m+1} \in I$ as a polynomial in with lowest degree in x_{n+1} that is not in the ideal $\langle f_1, \ldots, f_{i-1} \rangle$. □

Hilbert's Basis Theorem implies many important finiteness properties of algebraic varieties.

Corollary 1.2.2 *Any variety $Z \subset \mathbb{A}^n$ is the intersection of finitely many hypersurfaces.*

Proof. Let $Z = \mathcal{V}(I)$ be defined by the ideal I. By Hilbert's Basis Theorem, I is finitely generated, say by $f_1, \ldots, f_s$, and so $Z = \mathcal{V}(f_1, \ldots, f_s) = \mathcal{V}(f_1) \cap \cdots \cap \mathcal{V}(f_s)$. □

Example 1.2.3 The ideal of the cubic space curve C of Figure 1.2 with parametrization $(t^2, -t, t^3)$ not only contains the polynomials $xy{+}z$ and $x^2{-}x + y^2{+}yz$, but also $y^2{-}x$, $x^2{+}yz$, and $y^3{+}z$. Not all of these polynomials are needed to define C as $x^2 - x + y^2 + yz = (y^2 - x) + (x^2 + yz)$ and $y^3 + z = y(y^2 - x) + (xy + z)$. In fact three of the quadrics suffice,

$$\mathcal{I}(C) \;=\; \langle xy{+}z,\; y^2{-}x,\; x^2{+}yz \rangle \,.$$

Lemma 1.2.4 *For any subset Z of $\mathbb{A}^n$, if $X = \mathcal{V}(\mathcal{I}(Z))$ is the variety defined by the ideal $\mathcal{I}(Z)$, then $\mathcal{I}(X) = \mathcal{I}(Z)$ and X is the smallest variety containing Z.*

Proof. Set $X := \mathcal{V}(\mathcal{I}(Z))$. Then $\mathcal{I}(Z) \subset \mathcal{I}(X)$, since if f vanishes on Z, it will vanish on X. However, $Z \subset X$, and so $\mathcal{I}(Z) \supset \mathcal{I}(X)$, and thus $\mathcal{I}(Z) = \mathcal{I}(X)$.

If Y was a variety with $Z \subset Y \subset X$, then $\mathcal{I}(X) \subset \mathcal{I}(Y) \subset \mathcal{I}(Z) = \mathcal{I}(X)$, and so $\mathcal{I}(Y) = \mathcal{I}(X)$. But then we must have $Y = X$ for otherwise $\mathcal{I}(X) \subsetneq \mathcal{I}(Y)$, as is shown in Exercise 3. □

Thus we also lose nothing if we restrict the right-hand-side of the correspondence (1.1) to the subvarieties of $\mathbb{A}^n$. Our correspondence now becomes

$$\{\text{Ideals } I \text{ of } \mathbb{F}[x_1,\ldots,x_n]\} \quad \underset{\mathcal{I}}{\overset{\mathcal{V}}{\rightleftarrows}} \quad \{\text{Subvarieties } X \text{ of } \mathbb{A}^n\}\ . \tag{1.2}$$

This association is not a bijection. In particular, the map $\mathcal{V}$ is not one-to-one and the map $\mathcal{I}$ is not onto. There are several reasons for this.

For example, when $\mathbb{F}=\mathbb{Q}$ and $n=1$, we have $\emptyset=\mathcal{V}(1)=\mathcal{V}(x^2-2)$. The problem here is that the rational numbers are not algebraically closed and we need to work with a larger field (for example $\mathbb{Q}(\sqrt{2})$) to study $\mathcal{V}(x^2-2)$. When $\mathbb{F}=\mathbb{R}$ and $n=1$, $\emptyset\neq\mathcal{V}(x^2-2)$, but we have $\emptyset=\mathcal{V}(1)=\mathcal{V}(1+x^2)=\mathcal{V}(1+x^4)$. While the problem here is again that the real numbers are not algebraically closed, we view this as a manifestation of positivity. The two polynomials $1+x^2$ and $1+x^4$ only take positive values. When working over $\mathbb{R}$ (as our interest in applications leads us to do so) we will sometimes take positivity of polynomials into account.

The problem with the map $\mathcal{V}$ is more fundamental than these examples reveal and occurs even when $\mathbb{F}=\mathbb{C}$. When $n=1$ we have $\{0\}=\mathcal{V}(x)=\mathcal{V}(x^2)$, and when $n=2$, we invite the reader to check that $\mathcal{V}(y-x^2)=\mathcal{V}(y^2-yx^2,xy-x^3)$. Note that while $x\notin\langle x^2\rangle$, we have $x^2\in\langle x^2\rangle$. Similarly, $y-x^2\notin\mathcal{V}(y^2-yx^2,xy-x^3)$, but

$$(y-x^2)^2 \ = \ y^2-yx^2 \ - \ x(xy-x^3) \ \in \ \langle y^2-yx^2,xy-x^3\rangle\,.$$

In both cases, the lack of injectivity of the map $\mathcal{V}$ is because f and f^m have the same set of zeroes, for any positive integer m. For example, if $f_1,\ldots,f_s$ are polynomials, then the two ideals

$$\langle f_1,\,f_2,\,\ldots,\,f_s\rangle \qquad \text{and} \qquad \langle f_1,\,f_2^2,\,f_3^3,\,\ldots,\,f_s^s\rangle$$

both define the same variety, and if $f^m\in\mathcal{I}(Z)$, then $f\in\mathcal{I}(Z)$.

We clarify this point with a definition. An ideal $I\subset\mathbb{F}[x_1,\ldots,x_n]$ is *radical* if whenever $f^2\in I$, then $f\in I$. If I is radical and $f^m\in I$, then let s be an integer with $m\leq 2^s$. Then $f^{2^s}\in I$. As I is radical, this implies that $f^{2^{s-1}}\in I$, then that $f^{2^{s-2}}\in I$, and then by downward induction that $f\in I$. The radical $\sqrt{I}$ of an ideal I of $\mathbb{F}[x_1,\ldots,x_n]$ is

$$\sqrt{I} \ := \ \{f\in\mathbb{F}[x_1,\ldots,x_n]\mid f^2\in I\}\,.$$

This turns out to be an ideal which is the smallest radical ideal containing I. For example, we just showed that

$$\sqrt{\langle y^2-yx^2,xy-x^3\rangle} \ = \ \langle y-x^2\rangle\,.$$

The reason for this definition is twofold: first, $\mathcal{I}(Z)$ is radical, and second, an ideal I and its radical$\sqrt{I}$ both define the same variety. We record these facts.

Lemma 1.2.5 *For $Z\subset\mathbb{A}^n$, $\mathcal{I}(Z)$ is a radical ideal. If $I\subset\mathbb{F}[x_1,\ldots,x_n]$ is an ideal, then $\mathcal{V}(I)=\mathcal{V}(\sqrt{I})$.*

When $\mathbb{F}$ is algebraically closed, the precise nature of the correspondence (1.2) follows from Hilbert's Nullstellensatz (null=zeroes, stelle=places, satz=theorem), another of Hilbert's foundational results in the 1890's that helped to lay the foundations of algebraic geometry and usher in twentieth century mathematics. We first state a weak form of the Nullstellensatz, which describes the ideals defining the empty set.

Theorem 1.2.6 (Weak Nullstellensatz) *If I is an ideal of $\mathbb{C}[x_1, \ldots, x_n]$ with $\mathcal{V}(I) = \emptyset$, then $I = \mathbb{C}[x_1, \ldots, x_n]$.*

Let $a = (a_1, \ldots, a_n) \in \mathbb{A}^n$, which is defined by the linear polynomials $x_i - a_i$. A polynomial f is equal to the constant $f(a)$ modulo the ideal $\mathfrak{m}_a := \langle x_1 - a_1, \ldots, x_n - a_n \rangle$ generated by these polynomials, thus the quotient ring $\mathbb{F}[x_1, \ldots, x_n]/\mathfrak{m}_a$ is isomorphic to the field $\mathbb{F}$ and so $\mathfrak{m}_a$ is a maximal ideal. In the appendix we show that when $\mathbb{F} = \mathbb{C}$ (or any other algebraically closed field), these are the only maximal ideals.

Theorem 1.2.7 *Every maximal ideal of $\mathbb{C}[x_1, \ldots, x_n]$ has the form $\mathfrak{m}_a$ for some $a \in \mathbb{A}^n$.*

Proof of Weak Nullstellensatz. We prove the contrapositive, if $I \subsetneq \mathbb{C}[x_1, \ldots, x_n]$ is a proper ideal, then $\mathcal{V}(I) \neq \emptyset$. There is a maximal ideal $\mathfrak{m}_a$ with $a \in \mathbb{A}^n$ of $\mathbb{C}[x_1, \ldots, x_n]$ which contains I. But then

$$\{a\} \;=\; \mathcal{V}(\mathfrak{m}_a) \;\subset\; \mathcal{V}(I)\,,$$

and so $\mathcal{V}(I) \neq \emptyset$. Thus if $\mathcal{V}(I) = \emptyset$, we must have $I = \mathbb{C}[x_1, \ldots, x_n]$, which proves the weak Nullstellensatz. □

The Fundamental Theorem of Algebra states that any nonconstant polynomial $f \in \mathbb{C}[x]$ has a root (a solution to $f(x) = 0$). We recast the weak Nullstellensatz as the multivariate fundamental theorem of algebra.

Theorem 1.2.8 (Multivariate Fundamental Theorem of Algebra) *If the polynomials $f_1, \ldots, f_m \in \mathbb{C}[x_1, \ldots, x_n]$ generate a proper ideal of $\mathbb{C}[x_1, \ldots, x_n]$, then the system of polynomial equations*

$$f_1(x) \;=\; f_2(x) \;=\; \cdots \;=\; f_m(x) \;=\; 0$$

has a solution in $\mathbb{A}^n$.

We now deduce the full Nullstellensatz, which we will use to complete the characterization (1.2).

Theorem 1.2.9 (Nullstellensatz) *If $I \subset \mathbb{C}[x_1, \ldots, x_n]$ is an ideal, then $\mathcal{I}(\mathcal{V}(I)) = \sqrt{I}$.*

Proof. Since $\mathcal{V}(I) = \mathcal{V}(\sqrt{I})$, we have $\sqrt{I} \subset \mathcal{I}(\mathcal{V}(I))$. We show the other inclusion. Suppose that we have a polynomial $f \in \mathcal{I}(\mathcal{V}(I))$. Introduce a new variable t. Then the variety $\mathcal{V}(tf-1, I) \subset \mathbb{A}^{n+1}$ defined by I and $tf-1$ is empty. Indeed, if $(a_1, \ldots, a_n, b)$ were a point of this variety, then $(a_1, \ldots, a_n)$ would be a point of $\mathcal{V}(I)$. But then $f(a_1, \ldots, a_n) = 0$, and so the polynomial $tf - 1$ evaluates to 1 (and not 0) at the point $(a_1, \ldots, a_n, b)$.

By the weak Nullstellensatz, $\langle tf-1, I\rangle = \mathbb{C}[x_1, \ldots, x_n, t]$. In particular, $1 \in \langle tf-1, I\rangle$, and so there exist polynomials $f_1, \ldots, f_m \in I$ and $g, g_1, \ldots, g_m \in \mathbb{C}[x_1, \ldots, x_n, t]$ such that

$$1 \;=\; g(x,t)(tf(x)-1) \;+\; f_1(x)g_1(x,t) + f_2(x)g_2(x,t) + \cdots + f_m(x)g_m(x,t)\,.$$

If we apply the substitution $t = \frac{1}{f}$, then the first term with the factor $tf - 1$ vanishes and each polynomial $g_i(x,t)$ becomes a rational function in $x_1, \ldots, x_n$ whose denominator is a power of f. Clearing these denominators gives an expression of the form

$$f^N \;=\; f_1(x)G_1(x) + f_2(x)G_2(x) + \cdots + f_m(x)G_m(x)\,,$$

where $G_1, \ldots, G_m \in \mathbb{C}[x_1, \ldots, x_n]$. But this shows that $f \in \sqrt{I}$, and completes the proof of the Nullstellensatz. □

Corollary 1.2.10 (Algebraic-Geometric Dictionary I) *Over any field $\mathbb{F}$, the maps $\mathcal{V}$ and $\mathcal{I}$ give an inclusion reversing correspondence*

$$\{\textit{Radical ideals } I \textit{ of } \mathbb{F}[x_1, \ldots, x_n]\} \quad \underset{\mathcal{I}}{\overset{\mathcal{V}}{\rightleftarrows}} \quad \{\textit{Subvarieties } X \textit{ of } \mathbb{A}^n\} \tag{1.3}$$

with $\mathcal{V}(\mathcal{I}(X)) = X$. When $\mathbb{F}$ is algebraically closed, the maps $\mathcal{V}$ and $\mathcal{I}$ are inverses, and this correspondence is a bijection.

Proof. First, we already observed that $\mathcal{I}$ and $\mathcal{V}$ reverse inclusions and these maps have the domain and range indicated. Let X be a subvariety of $\mathbb{A}^n$. In Lemma 1.2.4 we showed that $X = \mathcal{V}(\mathcal{I}(X))$. Thus $\mathcal{V}$ is onto and $\mathcal{I}$ is one-to-one.

Now suppose that $\mathbb{F} = \mathbb{C}$. By the Nullstellensatz, if I is radical then $\mathcal{I}(\mathcal{V}(I)) = I$, and so $\mathcal{I}$ is onto and $\mathcal{V}$ is one-to-one. In particular, this shows that $\mathcal{I}$ and $\mathcal{V}$ are inverse bijections. □

Corollary 1.2.10 is only the beginning of the algebra-geometry dictionary. Many natural operations on varieties correspond to natural operations on their ideals. The *sum* $I + J$ and *product* $I \cdot J$ of ideals I and J are defined to be

$$\begin{aligned} I + J &:= \{f + g \mid f \in I \quad \text{and} \quad g \in J\} \\ I \cdot J &:= \langle f \cdot g \mid f \in I \quad \text{and} \quad g \in J\rangle\,. \end{aligned}$$

Lemma 1.2.11 *Let I, J be ideals in $\mathbb{F}[x_1, \ldots, x_n]$ and set $X := \mathcal{V}(I)$ and $Y = \mathcal{V}(J)$ to be their corresponding varieties. Then*

1. $\mathcal{V}(I+J) = X \cap Y$,
2. $\mathcal{V}(I \cdot J) = \mathcal{V}(I \cap J) = X \cup Y$,

If $\mathbb{F}$ is algebraically closed, then we also have

3. $\mathcal{I}(X \cap Y) = \sqrt{I+J}$, *and*
4. $\mathcal{I}(X \cup Y) = \sqrt{I \cap J} = \sqrt{I \cdot J}$.

Example 1.2.12 It can happen that $I \cdot J \neq I \cap J$. For example, if $I = \langle xy - x^3 \rangle$ and $J = \langle y^2 - x^2 y \rangle$, then $I \cdot J = \langle xy(y-x^2)^2 \rangle$, while $I \cap J = \langle xy(y-x^2) \rangle$.

This correspondence will be further refined in Section 1.5 to include maps between varieties. Because of this correspondence, each geometric concept has a corresponding algebraic concept, when $\mathbb{F}$ is algebraically closed. When $\mathbb{F}$ is not algebraically closed, this correspondence is not exact. In that case we will often use algebra to guide our geometric definitions.

Exercises for Section 2

1. Verify the claim in the text that the smallest ideal containing a set $S \subset \mathbb{F}[x_1, \ldots, x_n]$ of polynomials consists of all expressions of the form

$$h_1 f_1 + \cdots + h_m f_m$$

where $f_1, \ldots, f_m \in S$ and $h_1, \ldots, h_m \in \mathbb{F}[x_1, \ldots, x_n]$.

2. Let I be an ideal of $\mathbb{C}[x_1, \ldots, x_n]$. Show that

$$\sqrt{I} \;:=\; \{f \in \mathbb{F}[x_1, \ldots, x_n] \mid f^2 \in I\}$$

is an ideal, is radical, and is the smallest radical ideal containing I.

3. If $Y \subsetneq X$ are varieties, show that $\mathcal{I}(X) \subsetneq \mathcal{I}(Y)$.

4. Suppose that I and J are radical ideals. Show that $I \cap J$ is also a radical ideal.

5. Give radical ideals I and J for which $I + J$ is not radical.

6. Let I be an ideal in $\mathbb{F}[x_1, \ldots, x_n]$. Prove or find counterexamples to the following statements.

   (a) If $\mathcal{V}(I) = \mathbb{A}^n_{\mathbb{F}}$ then $I = \langle 0 \rangle$.

   (b) If $\mathcal{V}(I) = \emptyset$ then $I = \mathbb{F}[x_1, \ldots, x_n]$.

7. Give an example of two algebraic varieties Y and Z such that $\mathcal{I}(Y \cap Z) \neq \mathcal{I}(Y) + \mathcal{I}(Z)$.

8. (a) Let I be an ideal of $\mathbb{F}[x_1, x_2, \ldots, x_n]$. Show that if $\mathbb{F}[x_1, x_2, \ldots, x_n]/I$ is a finite dimensional $\mathbb{F}$-vector space then $\mathcal{V}(I)$ is a finite set.

   (b) Let $J = \langle xy, yz, xz \rangle$ be an ideal in $\mathbb{F}[x, y, z]$. Find the generators of $\mathcal{I}(\mathcal{V}(J))$. Show that J cannot be generated by two polynomials in $\mathbb{F}[x, y, z]$. Describe $V(I)$ where $I = \langle xy, xz - yz \rangle$. Show that $\sqrt{I} = J$.

9. Let $f, g \in \mathbb{F}[x, y]$ be coprime polynomials. Use Exercise 8(a) to show that $\mathcal{V}(f) \cap \mathcal{V}(g)$ is a finite set.

10. Prove that there are three points p, q, and r in $\mathbb{A}^2_{\mathbb{F}}$ such that

$$\sqrt{\langle x^2 - 2xy^4 + y^6, y^3 - y \rangle} \;=\; I(\{p\}) \cap I(\{q\}) \cap I(\{r\})\,.$$

Find a reason why you would know that the ideal $\langle x^2 - 2xy^4 + y^6, y^3 - y \rangle$ is not a radical ideal.

1.3 Generic properties of varieties

Many properties in algebraic geometry hold for almost all points of a variety or for almost all objects of a given type. For example, matrices are almost always invertible, univariate polynomials of degree d almost always have d distinct roots, and multivariate polynomials are almost always irreducible. We develop the terminology 'generic' and 'Zariski open' to describe this situation.

A starting point is that intersections and unions of affine varieties behave well.

Theorem 1.3.1 *The intersection of any collection of affine varieties is an affine variety. The union of any finite collection of affine varieties is an affine variety.*

Proof. For the first statement, let $\{I_t \mid t \in T\}$ be a collection of ideals in $\mathbb{F}[x_1, \ldots, x_n]$. Then we have

$$\bigcap_{t \in T} \mathcal{V}(I_t) \;=\; \mathcal{V}\Big(\bigcup_{t \in T} I_t\Big)\,.$$

Arguing by induction on the number of varieties, shows that it suffices to establish the second statement for the union of two varieties but that case is Lemma 1.2.11 (3). □

Theorem 1.3.1 shows that affine varieties have the same properties as the closed sets of a topology on $\mathbb{A}^n$. This was observed by Oscar Zariski.

Definition 1.3.2 We call an affine variety a *Zariski closed set*. The complement of a Zariski closed set is a *Zariski open set*. The *Zariski topology* on $\mathbb{A}^n$ is the topology whose closed sets are the affine varieties in $\mathbb{A}^n$. The *Zariski closure* of a subset $Z \subset \mathbb{A}^n$ is the smallest variety containing Z, which is $\overline{Z} := \mathcal{V}(\mathcal{I}(Z))$, by Lemma 1.2.4. Any subvariety X of $\mathbb{A}^n$ inherits its Zariski topology from $\mathbb{A}^n$, the closed subsets are simply the subvarieties of X. A subset $Z \subset X$ of a variety X is *Zariski dense* in X if its closure is X.

We emphasize that the purpose of this terminology is to aid our discussion of varieties, and not because we will use notions from topology in any essential way. This Zariski topology is behaves quite differently from the usual *Euclidean* topology on $\mathbb{R}^n$ or $\mathbb{C}^n$ with which we may be familiar. A topology on a space may be defined by giving a collection of basic open sets which generate the topology jn that any open set is a union or a finite intersection of basic open sets. In the Euclidean topology, the basic open sets are balls. Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. The *ball* with radius $\epsilon > 0$ centered at $z \in \mathbb{A}^n$ is

$$B(z, \epsilon) \;:=\; \{a \in \mathbb{A}^n \mid \sum |a_i - z_i|^2 \;<\; \epsilon\}\,.$$

In the Zariski topology, the basic open sets are complements of hypersurfaces, called principal open sets. Let $f \in \mathbb{F}[x_1, \ldots, x_n]$ and set

$$U_f \;:=\; \{a \in \mathbb{A}^n \mid f(a) \neq 0\}\,.$$

In both these topologies the open sets are unions of basic open sets—we do not need intersections to generate the given topology.

We give two examples to illustrate the Zariski topology.

Example 1.3.3 The Zariski closed subsets of $\mathbb{A}^1$ are the empty set, finite collections of points, and $\mathbb{A}^1$ itself. Thus when $\mathbb{F}$ is infinite the usual separation property of Hausdorff spaces (any two points are covered by two disjoint open sets) fails spectacularly as any two nonempty open sets meet.

Example 1.3.4 The Zariski topology on a product $X \times Y$ of affine varieties X and Y is in general not the product topology. In the product topology on $\mathbb{A}^2$, the closed sets are finite unions of sets of the following form: the empty set, points, vertical and horizontal lines of the form $\{a\} \times \mathbb{A}^1$ and $\mathbb{A}^1 \times \{a\}$, and the whole space $\mathbb{A}^2$. On the other hand, $\mathbb{A}^2$ contains a rich collection of 1-dimensional subvarieties (called *plane curves*), such as the cubic plane curves of Section 1.1.

We compare the Zariski topology with the Euclidean topology.

Theorem 1.3.5 *Suppose that $\mathbb{F}$ is one of $\mathbb{R}$ or $\mathbb{C}$. Then*

1. *A Zariski closed set is closed in the Euclidean topology on $\mathbb{A}^n$.*
2. *A Zariski open set is open in the Euclidean topology on $\mathbb{A}^n$.*
3. *A nonempty Euclidean open set is Zariski dense.*
4. *$\mathbb{R}^n$ is Zariski dense in $\mathbb{C}^n$.*
5. *A Zariski closed set is nowhere dense in the Euclidean topology on $\mathbb{A}^n$.*
6. *A nonempty Zariski open set is dense in the the Euclidean topology on $\mathbb{A}^n$.*

Proof. For statements 1 and 2, observe that a Zariski closed set $\mathcal{V}(I)$ is the intersection of the hypersurfaces $\mathcal{V}(f)$ for $f \in I$, so it suffices to show this for a hypersurface $\mathcal{V}(f)$. But then Statement 1 (and hence also 2) follows as the polynomial function $f\colon \mathbb{A}^n \to \mathbb{F}$ is continuous in the Euclidean topology, and $\mathcal{V}(f) = f^{-1}(0)$.

We show that any ball $B(z, \epsilon)$ is Zariski dense. If a polynomial f vanishes identically on $B(z, \epsilon)$, then all of its partial derivatives do as well. This implies that its Taylor series expansion at z is identically zero. But then f is the zero polynomial. This shows that $\mathcal{I}(B) = \{0\}$, and so $\mathcal{V}(\mathcal{I}(B)) = \mathbb{A}^n$, that is, B is dense in the Zariski topology on $\mathbb{A}^n$.

For statement 4, we use the same argument. If a polynomial vanishes on $\mathbb{R}^n$, then all of its partial derivatives vanish and so f must be the zero polynomial. Thus $\mathcal{I}(\mathbb{R}^n) = \{0\}$ and $\mathcal{V}(\mathcal{I}(\mathbb{R}^n)) = \mathbb{C}^n$.

For statements 5 and 6, observe that if f is nonconstant, then the interior of the (Euclidean) closed set $\mathcal{V}(f)$ is empty and so $\mathcal{V}(f)$ is nowhere dense. A subvariety is an intersection of nowhere dense hypersurfaces, so varieties are nowhere dense. The complement of a nowhere dense set is dense, so Zariski open sets are dense in $\mathbb{A}^n$. □

The last statement of Theorem 1.3.5 leads to the useful notions of genericity and generic sets and properties.

Definition 1.3.6 Let X be a variety. A subset $Y \subset X$ is called *generic* if it contains a Zariski dense open subset U of X. That is, we have $U \subset Y \subset X$ with U Zariski open and $\overline{U} = X$. A property is generic if the set of points on which it holds is a generic set. Points of a generic set are called general points.

This notion of general depends on the context, and so care must be exercised in its use. For example, we may identify $\mathbb{A}^3$ with the set of quadratic polynomials in x via

$$(a, b, c) \longmapsto ax^2 + bx + c\,.$$

Then, the general quadratic polynomial does not vanish when $x = 0$. (We just need to avoid quadratics with $c = 0$.) On the other hand, the general quadratic polynomial has two roots, as we need only avoid quadratics with $b^2 - 4ac = 0$. The quadratic $x^2 - 2x + 1$ is general in the first sense, but not in the second, while the quadratic $x^2 + x$ is general in the second sense, but not in the first. Despite this ambiguity, we will see that general is a very useful concept.

When $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$, generic sets are dense in the Euclidean topology, by Theorem 1.3.5(6). Thus generic properties hold almost everywhere, in the standard sense.

Example 1.3.7 The generic $n \times n$ matrix is invertible, as it is a nonempty principal open subset of $\mathrm{Mat}_{n\times n} = \mathbb{A}^{n\times n}$. It is the complement of the variety $\mathcal{V}(\det)$ of singular matrices. Define the *general linear group* GL_n to be the set of all invertible matrices,

$$GL_n \;:=\; \{M \in \mathrm{Mat}_{n\times n} \mid \det(M) \neq 0\} \;=\; U_{\det}\,.$$

Example 1.3.8 The general univariate polynomial of degree n has n distinct complex roots. Identify $\mathbb{A}^n$ with the set of univariate polynomials of degree n via

$$(a_1, \ldots, a_n) \;\in\; \mathbb{A}^n \;\longmapsto\; x^n + a_1 x^{n-1} + \cdots + a_{n-1}x + a_n \;\in\; \mathbb{F}[x]\,. \tag{1.4}$$

The classical discriminant $\Delta \in \mathbb{F}[a_1, \ldots, a_n]$ (See Example 2.3.6) is a polynomial of degree $2n - 2$ which vanishes precisely when the polynomial (1.4) has a repeated factor. This identifies the set of polynomials with n distinct complex roots as the set U_Δ. The discriminant of a quadratic $x^2 + bx + c$ is $b^2 - 4c$.

Example 1.3.9 The generic complex $n \times n$ matrix is semisimple (diagonalizable). Let $M \in \mathrm{Mat}_{n\times n}$ and consider the (monic) characteristic polynomial of M

$$\chi(x) \;:=\; \det(xI_n - M)\,.$$

We do not show this by providing an algebraic characterization of semisimplicity. Instead we observe that if a matrix $M \in \mathrm{Mat}_{n\times n}$ has n distinct eigenvalues, then it is semisimple. The coefficients of the characteristic polynomial $\chi(x)$ are polynomials in the entries of M. Evaluating the discriminant at these coefficients gives a polynomial ψ which vanishes when the characteristic polynomial $\chi(x)$ of M has a repeated root.

We see that the set of matrices with distinct eigenvalues equals the basic open set U_ψ, which is nonempty. Thus the set of semisimple matrices contains an open dense subset of $\mathrm{Mat}_{n\times n}$ and is therefore generic.

When $n = 2$,

$$\det\left(xI_2 \;-\; \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\right) \;=\; x^2 - x(a_{11}+a_{22}) + a_{11}a_{22} - a_{12}a_{21}\,,$$

and so the polynomial ψ is $(a_{11}+a_{22})^2 - 4(a_{11}a_{22} - a_{12}a_{21})$.

In each of these examples, we used the following easy fact.

Proposition 1.3.10 *A set $X \subset \mathbb{A}^n$ is generic if and only if there is a nonconstant polynomial that vanishes on its complement, if and only if it contains a basic open set U_f.*

More generally, if $X \subset \mathbb{A}^n$ is a variety and $f \in \mathbb{F}[x_1, \ldots, x_n]$ is a polynomial which is not identically zero on X ($f \notin \mathcal{I}(X)$), then we have the *principal open subset* of X,

$$X_f \;:=\; X - \mathcal{V}(F) \;=\; \{x \in X \mid f(x) \neq 0\}\,.$$

Lemma 1.3.11 *Any Zariski open subset U of a variety X is a finite union of principal open subsets.*

Proof. The complement $Y := X - U$ of a Zariski open subset U of X is a Zariski closed subset. The ideal $\mathcal{I}(Y)$ of Y in $\mathbb{A}^n$ contains the ideal $\mathcal{I}(X)$ of X. By the Hilbert Basis Theorem, there are polynomials $f_1, \ldots, f_m \in \mathcal{I}(Y)$ such that

$$\mathcal{I}(Y) \;=\; \langle \mathcal{I}(X),\; f_1, \ldots, f_m \rangle\,.$$

Then $X_{f_1} \cup \cdots \cup X_{f_m}$ is equal to

$$(X - \mathcal{V}(f_1)) \cup \cdots \cup (X - \mathcal{V}(f_m)) \;=\; X - (\mathcal{V}(f_1) \cap \cdots \cap \mathcal{V}(f_m)) \;=\; X - Y \;=\; U\,. \quad \square$$

Exercises

1. Look up the definition of a topology in a text book and verify the claim that the collection of affine subvarieties of $\mathbb{A}^n$ form the closed sets in a topology on $\mathbb{A}^n$.

2. Prove that a closed set in the Zariski topology on $\mathbb{A}^1$ is either the empty set, a finite collection of points, or $\mathbb{A}^1$ itself.

3. Let $n \leq m$. Prove that a generic $n \times m$ matrix has rank n.

4. Prove that the generic triple of points in $\mathbb{A}^2$ are the vertices of a triangle.

5. (a) Describe all the algebraic varieties in $\mathbb{A}^1$.
   (b) Show that any open set in $\mathbb{A}^1 \times \mathbb{A}^1$ is open in $\mathbb{A}^2$.
   (c) Find a Zariski open set in $\mathbb{A}^2$ which is not open in $\mathbb{A}1 \times \mathbb{A}^1$.
6. (a) Show that the Zariski topology in $\mathbb{A}^n$ is not Hausdorff if $\mathbb{F}$ is infinite.
   (b) Prove that any nonempty open subset of $\mathbb{A}^n$ is dense.
   (c) Prove that $\mathbb{A}^n$ is compact.

1.4 Unique factorization for varieties

Every polynomial factors uniquely as a product of irreducible polynomials. A basic structural result about algebraic varieties is an analog of unique factorization. Any algebraic variety is the finite union of irreducible varieties, and this decomposition is unique.

A polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ is *decomposable* if we may factor f notrivially, that is, if $f = gh$ with neither g nor h a constant polynomial. Otherwise f is *indecomposable*. Any polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ may be factored

$$f \;=\; g_1^{\alpha_1} g_2^{\alpha_2} \cdots g_m^{\alpha_m} \tag{1.5}$$

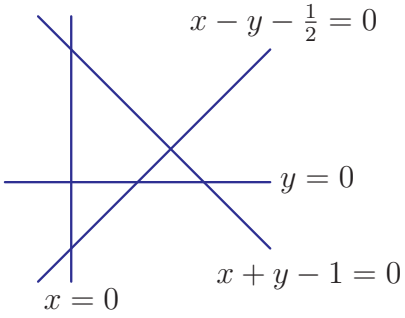
where the exponents α_i are positive integers, each polynomial g_i is irreducible and nonconstant, and when $i \neq j$ the polynomials g_i and g_j are not proportional. This factorization is essentially unique as any other such factorization is obtained from this by permuting the factors and possibly multiplying each polynomial g_i by a constant. The polynomials g_j are *irreducible factors* of f.

When $\mathbb{F}$ is algebraically closed, this algebraic property has a consequence for the geometry of hypersurfaces. Suppose that a polynomial f has a factorization (1.5) into irreducible polynomials. Then the hypersurface $X = \mathcal{V}(f)$ is the union of hypersurfaces $X_i := \mathcal{V}(g_i)$, and this decomposition

$$X \;=\; X_1 \cup X_2 \cup \cdots \cup X_m$$

of X into hypersurfaces X_i defined by irreducible polynomials is unique.

For example, $\mathcal{V}(xy(x+y-1)(x-y-\frac{1}{2}))$ is the union of four lines in $\mathbb{A}^2$.



This decomposition property is shared by general varieties.

Definition 1.4.1 A variety X is *reducible* if it is the union $X = Y \cup Z$ of proper closed subvarieties $Y, Z \subsetneq X$. Otherwise X is *irreducible*. In particular, if an irreducible variety is written as a union of subvarieties $X = Y \cup Z$, then either $X = Y$ or $X = Z$.

Example 1.4.2 Figure 1.2 in Section 1.2 shows that $\mathcal{V}(xy + z,\, x^2 - x + y^2 + yz)$ consists of two space curves, each of which is a variety in its own right. Thus it is reducible. To see this, we solve the two equations $xy + z = x^2 - x + y^2 + yz = 0$. First note that

$$x^2 - x + y^2 + yz \;-\; y(xy + z) \;=\; x^2 - x + y^2 - xy^2 \;=\; (x - 1)(x - y^2)\,.$$

Thus either $x = 1$ or else $x = y^2$. When $x = 1$, we see that $y + z = 0$ and these equations define the line in Figure 1.2. When $x = y^2$, we get $z = -y^3$, and these equations define the cubic curve parametrized by $(t^2, t, -t^3)$.

Figure 1.4 shows another reducible variety. It has six components, one is a surface,

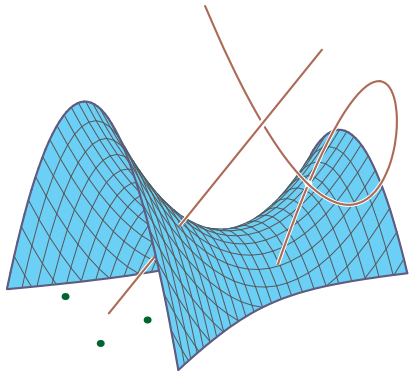


Figure 1.4: A reducible variety

two are space curves, and three are points.

Theorem 1.4.3 *A product $X \times Y$ of irreducible varieties is irreducible.*

Proof. Suppose that $Z_1, Z_2 \subset X \times Y$ are subvarieties with $Z_1 \cup Z_2 = X \times Y$. We assume that $Z_2 \neq X \times Y$ and use this to show that $Z_1 = X \times Y$. For each $x \in X$, identify the subvariety $\{x\} \times Y$ with Y. This irreducible variety is the union of two subvarieties,

$$\{x\} \times Y \;=\; \big((\{x\} \times Y) \cap Z_1\big) \;\cup\; \big((\{x\} \times Y) \cap Z_2\big)\,,$$

and so one of these must equal $\{x\} \times Y$. In particular, we must either have $\{x\} \times Y \subset Z_1$ or else $\{x\} \times Y \subset Z_2$. If we define

$$\begin{aligned} X_1 &= \{x \in X \mid \{x\} \times Y \subset Z_1\}\,, \quad \text{and} \\ X_2 &= \{x \in X \mid \{x\} \times Y \subset Z_2\}\,, \end{aligned}$$

then we have just shown that $X = X_1 \cup X_2$. Since $Z_2 \neq X \times Y$, we have $X_2 \neq X$. We claim that both X_1 and X_2 are subvarieties of X. Then the irreducibility of X implies that $X = X_1$ and thus $X \times Y = Z_1$.

We will show that X_1 is a subvariety of X. For $y \in Y$, set

$$X_y \;:=\; \{x \in X \mid (x, y) \in Z_1\}\,.$$

Since $X_y \times \{y\} = (X \times \{y\}) \cap Z_1$, we see that X_y is a subvariety of X. But we have

$$X_1 \;=\; \bigcap_{y \in Y} X_y\,,$$

which shows that X_1 is a subvariety of X. An identical argument for X_2 completes the proof. $\square$

The geometric notion of an irreducible variety corresponds to the algebraic notion of a prime ideal. An ideal $I \subset \mathbb{F}[x_1, \ldots, x_n]$ is *prime* if whenever $fg \in I$ with $f \notin I$, then we have $g \in I$. Equivalently, if whenever $f, g \notin I$ then $fg \notin I$.

Theorem 1.4.4 *A variety X is irreducible if and only if its ideal $\mathcal{I}(X)$ is prime.*

Proof. Let X be a variety. First suppose that X is irreducible. Let $f, g \notin \mathcal{I}(X)$. Then neither f nor g vanishes identically on X. Thus $Y := X \cap \mathcal{V}(f)$ and $Z := X \cap \mathcal{V}(z)$ are proper subvarieties of X. Since X is irreducible, $Y \cup Z = X \cap \mathcal{V}(fg)$ is also a proper subvariety of X, and thus $fg \notin \mathcal{I}(X)$.

Suppose now that X is reducible. Then $X = Y \cup Z$ is the union of proper subvarieties Y, Z of X. Since $Y \subsetneq X$ is a subvariety, we have $\mathcal{I}(X) \subsetneq \mathcal{I}(Y)$. Let $f \in \mathcal{I}(Y) - \mathcal{I}(X)$, a polynomial which vanishes on Y but not on X. Similarly, let $g \in \mathcal{I}(Z) - \mathcal{I}(X)$ be a polynomial which vanishes on Z but not on X. Since $X = Y \cup Z$, fg vanishes on X and therefore lies in $\mathcal{I}(X)$. This shows that I is not prime. □

We have seen examples of varieties with one, two, four, and six irreducible components. Taking products of distinct irreducible polynomials (or dually unions of distinct hypersurfaces), gives varieties having any finite number of irreducible components. This is all that can occur as Hilbert's Basis Theorem implies that a variety is a union of finitely many irreducible varieties.

Lemma 1.4.5 *Any affine variety is a finite union of irreducible subvarieties.*

Proof. An affine variety X either is irreducible or else we have $X = Y \cup Z$, with both Y and Z proper subvarieties of X. We may similarly decompose whichever of Y and Z are reducible, and continue this process, stopping only when all subvarieties obtained are irreducible. *A priori*, this process could continue indefinitely. We argue that it must stop after a finite number of steps.

If this process never stops, then X must contain an infinite chain of subvarieties, each properly contained in the previous,

$$X \supsetneq X_1 \supsetneq X_2 \supsetneq \cdots .$$

Their ideals form an infinite increasing chain of ideals in $\mathbb{F}[x_1, \ldots, x_n]$,

$$\mathcal{I}(X) \subsetneq \mathcal{I}(X_1) \subsetneq \mathcal{I}(X_2) \subsetneq \cdots .$$

The union I of these ideals is again an ideal. Note that no ideal $\mathcal{I}(X_m)$ is equal to I. By the Hilbert Basis Theorem, I is finitely generated, and thus there is some integer m for which $\mathcal{I}(X_m)$ contains these generators. But then $I = \mathcal{I}(X_m)$, a contradiction. □

A consequence of this proof is that any decreasing chain of subvarieties of a given variety must have finite length. When $\mathbb{F}$ is infinite, there are such decreasiung chains of arbitrary length. There is however a bound for the length of the longest decreasing chain of irreducible subvarieties.

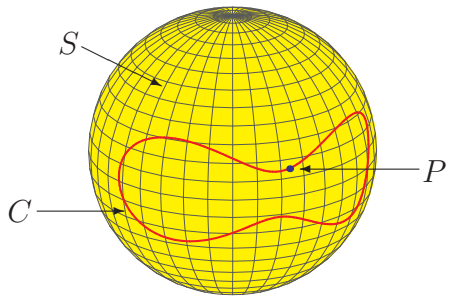
[Combinatorial Definition of Dimension] The *dimension* of a variety X is essentially the length of the longest decreasing chain of irreducible subvarieties of X. If

$$X \supset X_0 \supsetneq X_1 \supsetneq X_2 \supsetneq \cdots \supsetneq X_m \supsetneq \emptyset,$$

with each X_i irreducible is such a chain of maximal length, then X has dimension m.

Since maximal ideals of $\mathbb{C}[x_1, \ldots, x_n]$ necessarily have the form $\mathfrak{m}_a$, we see that X_m must be a point when $\mathbb{F} = \mathbb{C}$. The only problem with this definition is that we cannot yet show that it is well-founded, as we do not yet know that there is a bound on the length of such a chain. In Section 3.2 we shall prove that this definition is correct by relating it to other notions of dimension.

Example 1.4.6 The sphere S has dimension at least two, as we have the chain of subvarieties $S \supsetneq C \supsetneq P$ as shown below.



It is quite challenging to show that any maximal chain of irreducible subvarieties of the sphere has length 2 with what we know now.

By Lemma 1.4.5, an affine variety X may be written as a finite union

$$X = X_1 \cup X_2 \cup \cdots \cup X_m$$

of irreducible subvarieties. We may assume this is irredundant in that if $i \neq j$ then X_i is not a subvariety of X_j. If we did have $i \neq j$ with $X_i \subset X_j$, then we may remove X_i from the decomposition. We prove that this decomposition is unique, which is the main result of this section and a basic structural result about varieties.

Theorem 1.4.7 (Unique Decomposition of Varieites) *A variety X has a unique irredundant decomposition as a finite union of irreducible subvarieties*

$$X = X_1 \cup X_2 \cup \cdots \cup X_m.$$

We call these distinguished subvarieties X_i the *irreducible components* of X.

Proof. Suppose that we have another irredundant decomposition into irreducible subvarieties,

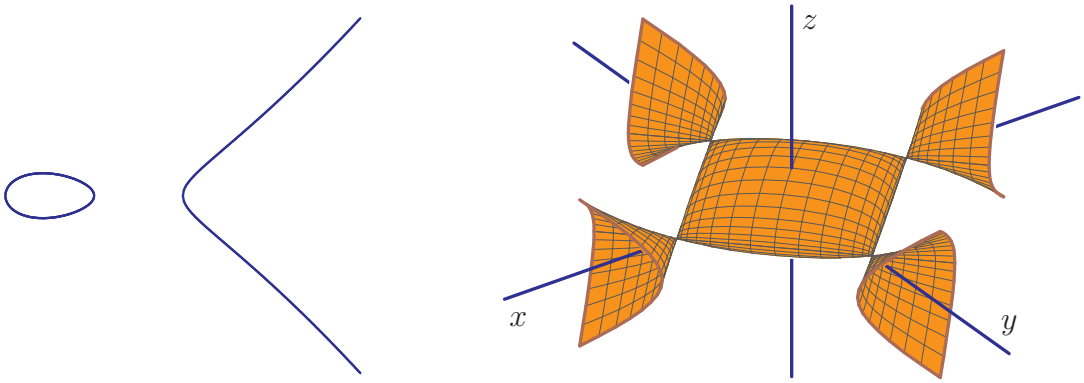
$$X = Y_1 \cup Y_2 \cup \cdots \cup Y_n ,$$

where each Y_i is irreducible. Then

$$X_i = (X_i \cap Y_1) \cup (X_i \cap Y_2) \cup \cdots \cup (X_i \cap Y_n) .$$

Since X_i is irreducible, one of these must equal X_i, which means that there is some index j with $X_i \subset Y_j$. Similarly, there is some index k with $Y_j \subset X_k$. Since this implies that $X_i \subset X_k$, we have $i = k$, and so $X_i = Y_j$. This implies that $n = m$ and that the second decomposition differs from the first solely by permuting the terms. □

When $\mathbb{F} = \mathbb{C}$, we will show that an irreducible variety is connected in the usual Euclidean topology. We will even show that the smooth points of an irreducible variety are connected. Neither of these facts are true over $\mathbb{R}$. Below, we display the irreducible cubic plane curve $\mathcal{V}(y^2 - x^3 + x)$ in $\mathbb{A}^2_{\mathbb{R}}$ and the surface $\mathcal{V}((x^2 - y^2)^2 - 2x^2 - 2y^2 - 16z^2 + 1)$ in $\mathbb{A}^3_{\mathbb{R}}$.



Both are irreducible hypersurfaces. The first has two connected components in the Euclidean topology, while in the second, the five components of smooth points meet at the four singular points.

Exercises

1. Show that the ideal of a hypersurface $\mathcal{V}(f)$ is generated by the *squarefree* part of f, which is the product of the irreducible factors of f, all with exponent 1.

2. For every positive integer n, give a decreasing chain of subvarieties of $\mathbb{A}^1$ of length $n+1$.

3. Prove that the dimension of a point is 0 and the dimension of $\mathbb{A}^1$ is 1.

4. Show that an irreducible affine variety is zero-dimensional if and only if it is a point.

5. Prove that the dimension of an irreducible plane curve is 1 and use this to show that the dimension of $\mathbb{A}^2$ is 2.

6. Write the ideal $\langle x^3 - x, x^2 - y \rangle$ as the intersection of two prime ideals. Describe the corresponding geometry.

7. Show that $f(x, y) = y^2 + x^2(x - 1)^2 \in \mathbb{R}[x, y]$ is an irreducible polynomial but that $V(f)$ is reducible.

8. Fix the hyperbola $H = V(xy - 5) \subset \mathbb{A}^2_{\mathbb{R}}$ and let C_t be the circle $x^2 + (y - t)^2 = 1$ for $t \in \mathbb{R}$.

   (a) Show that $H \cap C_t$ is zero-dimensional, for any choice of t.

   (b) Determine the number of points in $H \cap C_t$ (this number depends on t).

1.5 The algebra-geometry dictionary II

The algebra-geometry dictionary of Section 1.2 is strengthened when we include regular maps between varieties and the corresponding homomorphisms between rings of regular functions.

Let $X \subset \mathbb{A}^n$ be an affine variety and suppose that $\mathbb{F}$ is infinite. Any polynomial function $f \in \mathbb{F}[x_1, \ldots, x_n]$ restricts to give a *regular function* on X, $f\colon X \to \mathbb{F}$. We may add and multiply regular functions, and the set of all regular functions on X forms a ring, $\mathbb{F}[X]$, called the *coordinate ring* of the affine variety X or the ring of regular functions on X. The coordinate ring of an affine variety X is a basic invariant of X, which is, in fact equivalent to X itself.

The restriction of polynomial functions on $\mathbb{A}^n$ to regular functions on X defines a surjective ring homomorphism $\mathbb{F}[x_1, \ldots, x_n] \twoheadrightarrow \mathbb{F}[X]$. The kernel of this restriction homomorphism is the set of polynomials which vanish identically on X, that is, the ideal $\mathcal{I}(X)$ of X. Under the correspondence between ideals, quotient rings, and homomorphisms, this restriction map gives and isomorphism between $\mathbb{F}[X]$ and the quotient ring $\mathbb{F}[x_1, \ldots, x_n]/\mathcal{I}(X)$. When $\mathbb{F}$ is not infinite, we define the coordinate ring $\mathbb{F}[X]$ or ring of regular functions on X to be this quotient.

Example 1.5.1 The coordinate ring of the parabola $y = x^2$ is $\mathbb{F}[x, y]/\langle y - x^2\rangle$, which is isomorphic to $\mathbb{F}[x]$, the coordinate ring of $\mathbb{A}^1$. To see this, observe that substituting x^2 for y rewrites and polynomial $f(x, y)$ as a polynomial $g(x)$ in x alone, and $y - x^2$ divides the difference $f(x, y) - g(x)$.

Parabola Cuspidal Cubic

On the other hand, the coordinate ring of the cuspidal cubic $y^2 = x^3$ is $\mathbb{F}[x, y]/\langle y^2 - x^3\rangle$. This ring is not isomorphic to $\mathbb{F}[x, y]/\langle y - x^2\rangle$,. For example, the element $y^2 = x^3$ has two distinct factorizations into indecomposable elements, while polynomials $f(x)$ in one variable always factor uniquely.

Let X be a variety. Its quotient ring $\mathbb{F}[X] = \mathbb{F}[x_1, \ldots, x_n]/\mathcal{I}(X)$ is finitely generated by the images of the variables x_i. Since $\mathcal{I}(X)$ is radical, the quotient ring has no nilpotent elements (elements f such that $f^m = 0$ for some m). Such a ring with no nilpotents is called *reduced*. When $\mathbb{F}$ is algebraically closed, these two properties characterize coordinate rings of algebraic varieties.

Theorem 1.5.2 *Suppose that $\mathbb{F}$ is algebraically closed. Then an $\mathbb{F}$-algebra R is the coordinate ring of an affine variety if and only if R is finitely generated and reduced.*

Proof. We need only show that a finitely generated reduced $\mathbb{F}$-algebra R is the coordinate ring of an affine variety. Suppose that the reduced $\mathbb{F}$-algebra R has generators $r_1, \ldots, r_n$. Then there is a surjective ring homomorphism

$$\varphi : \mathbb{F}[x_1, \ldots, x_n] \longrightarrow\!\!\!\!\!\rightarrow R$$

given by $x_i \mapsto r_i$. Let $I \subset \mathbb{F}[x_1, \ldots, x_n]$ be the kernel of φ. This identifies R with $\mathbb{F}[x_1, \ldots, x_n]/I$. Since R is reduced, we see that I is radical.

When $\mathbb{F}$ is algebraically closed, the algebra-geometry dictionary of Corollary 1.2.10 shows that $I = \mathcal{I}(\mathcal{V}(I))$ and so $R \simeq \mathbb{F}[x_1, \ldots, x_n]/I \simeq \mathbb{F}[\mathcal{V}(I)]$. □

A different choice $s_1, \ldots, s_m$ of generators for R in this proof will give a different affine variety with coordinate ring R. One goal of this section is to understand this apparent ambiguity.

Example 1.5.3 Consider the finitely generated $\mathbb{F}$-algebra $R := \mathbb{F}[t]$. Choosing the generator t realizes R as $\mathbb{F}[\mathbb{A}^1]$. We could, however choose generators $x := t+1$ and $y := t^2+3t$. Since $y = x^2+x-2$, this also realizes R as $\mathbb{F}[x, y]/\langle y - x^2 - x + 2\rangle$, which is the coordinate ring of a parabola.

Among the coordinate rings $\mathbb{F}[X]$ of affine varieties are the polynomial algebras $\mathbb{F}[\mathbb{A}^n] = \mathbb{F}[x_1, \ldots, x_n]$. Many properties of polynomial algebras, including the algebra-geometry dictionary of Corollary 1.2.10 and the Hilbert Theorems hold for these coordinate rings $\mathbb{F}[X]$.

Given regular functions $f_1, \ldots, f_m \in \mathbb{F}[X]$ on an affine variety $X \subset \mathbb{A}^n$, their set of common zeroes

$$\mathcal{V}(f_1, \ldots, f_m) := \{x \in X \mid f_1(x) = \cdots = f_m(x) = 0\},$$

is a subvariety of X. To see this, let $F_1, \ldots, F_m \in \mathbb{F}[x_1, \ldots, x_n]$ be polynomials which restrict to $f_1, \ldots, f_m$. Then

$$\mathcal{V}(f_1, \ldots, f_m) = X \cap \mathcal{V}(F_1, \ldots, F_m).$$

As in Section 1.2, we may extend this notation and define $\mathcal{V}(I)$ for an ideal I of $\mathbb{F}[X]$. If $Y \subset X$ is a subvariety of X, then $\mathcal{I}(X) \subset \mathcal{I}(Y)$ and so $\mathcal{I}(Y)/\mathcal{I}(X)$ is an ideal in the coordinate ring $\mathbb{F}[X] = \mathbb{F}[\mathbb{A}^n]/\mathcal{I}(X)$ of X. Write $\mathcal{I}(Y) \subset \mathbb{F}[X]$ for the ideal of Y in $\mathbb{F}[X]$.

Both Hilbert's Basis Theorem and Hilbert's Nullstellensatz have analogs for affine varieties X and their coordinate rings $\mathbb{F}[X]$. These consequences of the original Hilbert Theorems follow from the surjection $\mathbb{F}[x_1, \ldots, x_n] \twoheadrightarrow \mathbb{F}[X]$ and corresponding inclusion $X \hookrightarrow \mathbb{A}^n$.

Theorem 1.5.4 (Hilbert Theorems for $\mathbb{F}[X]$) *Let X be an affine variety. Then*

1. *Any ideal of $\mathbb{F}[X]$ is finitely generated.*

2. *If Y is a subvariety of X then $\mathcal{I}(Y) \subset \mathbb{F}[X]$ is a radical ideal. The subvariety Y is irreducible if and only if $I(Y)$ is a prime ideal.*

3. *Suppose that $\mathbb{F}$ is algebraically closed. An ideal I of $\mathbb{F}[X]$ defines the empty set if and only if $I = \mathbb{F}[X]$.*

In the same way as in Section 1.2 we obtain a version of the algebra-geometry dictionary between subvarieties of an affine variety X and radical ideals of $\mathbb{F}[X]$. The proofs are nearly the same, so we leave them to the reader. For this, you will need to recall that ideals of a quotient ring R/I all have the form J/I, where J is an ideal of R which contains I.

Theorem 1.5.5 *Let X be an affine variety. Then the maps $\mathcal{V}$ and $\mathcal{I}$ give an inclusion reversing correspondence*

$$\{\textit{Radical ideals } I \textit{ of } \mathbb{F}[X]\} \quad \underset{\mathcal{I}}{\overset{\mathcal{V}}{\rightleftarrows}} \quad \{\textit{Subvarieties } Y \textit{ of } X\} \tag{1.6}$$

with $\mathcal{I}$ injective and $\mathcal{V}$ surjective. When $\mathbb{F} = \mathbb{C}$, the maps $\mathcal{V}$ and $\mathcal{I}$ are inverses and this correspondence is a bijection.

In algebraic geometry, we do not just study varieties, but also the maps between them.

Definition 1.5.6 A list $f_1, \ldots, f_m \in \mathbb{F}[X]$ of regular functions on an affine variety X defines a function

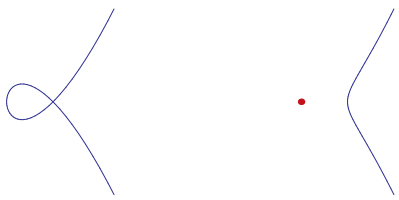
$$\begin{array}{rccl} \varphi : & X & \longrightarrow & \mathbb{A}^m \\ & x & \longmapsto & (f_1(x), f_2(x), \ldots, f_m(x)) \,, \end{array}$$

which we call a *regular map*.

Example 1.5.7 The elements $t^2, t, -t^3 \in \mathbb{F}[t] = \mathbb{F}[\mathbb{A}^1]$ define the map $\mathbb{A}^1 \to \mathbb{A}^3$ whose image is the cubic curve of Figure 1.2.

The elements t^2, t^3 of $\mathbb{F}[\mathbb{A}^1]$ define a map $\mathbb{A}^1 \to \mathbb{A}^2$ whose image is the cuspidal cubic that we saw earlier.

Let $x = t^2 - 1$ and $y = t^3 - t$, which are elements of $\mathbb{F}[t] = \mathbb{F}[\mathbb{A}^1]$. These define a map $\mathbb{A}^1 \to \mathbb{A}^2$ whose image is the nodal cubic curve $\mathcal{V}(y^2 - (x^3 + x^2))$ on the left below. If we instead take $x = t^2 + 1$ and $y = t^3 + t$, then we get a different map $\mathbb{A}^1 \to \mathbb{A}^2$ whose image is the curve on the right below.



In the curve on the right, the image of $\mathbb{A}^1_{\mathbb{R}}$ is the arc, while the isolated or *solitary point* is the image of the points $\pm\sqrt{-1}$.

Suppose that X is an affine variety and we have a regular map $\varphi\colon X \to \mathbb{A}^m$ given by regular functions $f_1, \ldots, f_m \in \mathbb{F}[X]$. A polynomial $g \in \mathbb{F}[x_1, \ldots, x_m] \in \mathbb{F}[A^m]$ *pulls back* along φ to give the regular function $\varphi^* g$, which is defined by

$$\varphi^* g \;:=\; g(f_1, \ldots, f_m)\,.$$

This element of the coordinate ring $\mathbb{F}[X]$ of X is the usual pull back of a function. For $x \in X$ we have

$$(\varphi^* g)(x) \;=\; g(\varphi(x)) \;=\; g(f_1(x), \ldots, f_m(x))\,.$$

The resulting map $\varphi^*\colon \mathbb{F}[\mathbb{A}^m] \to \mathbb{F}[X]$ is a homomorphism of $\mathbb{F}$-algebras. Conversely, given a homomorphism $\psi\colon \mathbb{F}[x_1, \ldots, x_m] \to \mathbb{F}[X]$ of $\mathbb{F}$-algebras, if we set $f_i := \psi(x_i)$, then $f_1, \ldots, f_m \in \mathbb{F}[X]$ define a regular map φ with $\varphi^* = \psi$.

We have just shown the following basic fact.

Lemma 1.5.8 *The association $\varphi \mapsto \varphi^*$ defines a bijection*

$$\left\{\begin{matrix}\textit{regular maps}\\ \varphi\colon X \to \mathbb{A}^m\end{matrix}\right\} \quad\longleftrightarrow\quad \left\{\begin{matrix}\textit{ring homomorphisms}\\ \psi\colon \mathbb{F}[\mathbb{A}^m] \to \mathbb{F}[X]\end{matrix}\right\}$$

In the examples that we gave, the image $\varphi(X)$ of X under φ was contained in a subvariety. This is always the case.

Lemma 1.5.9 *Let X be an affine variety, $\varphi\colon X \to \mathbb{A}^m$ a regular map, and $Y \subset \mathbb{A}^m$ a subvariety. Then $\varphi(X) \subset Y$ if and only if $\mathcal{I}(Y) \subset \ker \varphi^*$.*

Proof. Suppose that $\varphi(X) \subset Y$. If $f \in \mathcal{I}(Y)$ then f vanishes on Y and hence on $\varphi(X)$. But then $\varphi^* f$ is the zero function, and so $\mathcal{I}(Y) \subset \ker \varphi^*$.

For the other direction, suppose that $\mathcal{I}(Y) \subset \ker \varphi^*$ and let $x \in X$. If $f \in \mathcal{I}(Y)$, then $\varphi^* f = 0$ and so $0 = \varphi^* f(x) = f(\varphi(x))$. Since this is true for every $f \in \mathcal{I}(Y)$, we conclude that $\varphi(x) \in Y$. As this holds for every $x \in X$, we have $\varphi(X) \subset Y$. □

Corollary 1.5.10 *Let X be an affine variety, $\varphi\colon X \to \mathbb{A}^m$ a regular map, and $Y \subset \mathbb{A}^m$ a subvariety. Then*

(1) $\ker \varphi^*$ *is a radical ideal.*

(2) *If X is irreducible, then* $\ker \varphi^*$ *is a prime ideal.*

(3) $\mathcal{V}(\ker \varphi^*)$ *is the smallest affine variety containing $\varphi(X)$.*

(4) *If $\varphi\colon X \to Y$, then $\varphi^*\colon \mathbb{F}[\mathbb{A}^m] \to \mathbb{F}[X]$ factors through $\mathbb{F}[Y]$ inducing a homomorphism $\mathbb{F}[Y] \to \mathbb{F}[X]$.*

We write φ^* for the induced map $\mathbb{F}[Y] \to \mathbb{F}[X]$ of part (4).

Proof. For (1), suppose that $f^2 \in \ker \varphi^*$, so that $0 = \varphi^*(f^2) = (\varphi^*(f))^2$. Since $\mathbb{F}[X]$ has no nilpotent elements, we conclude that $\varphi^*(f) = 0$ and so $f \in \ker \varphi^*$.

For (2), suppose that $f \cdot g \in \ker \varphi^*$ with $g \notin \ker \varphi^*$. Then $0 = \varphi^*(f \cdot g) = \varphi^*(f) \cdot \varphi^*(g)$ in $\mathbb{F}[X]$, but $0 \neq \varphi^*(g)$. Since $\mathbb{F}[X]$ is a domain, we must have $0 = \varphi^*(f)$ and so $f \in \ker \varphi^*$, which shows that $\ker \varphi^*$ is a prime ideal.

Suppose that Y is an affine variety containing $\varphi(X)$. By Lemma 1.5.9, $\mathcal{I}(Y) \subset \ker \varphi^*$ and so $\mathcal{V}(\ker \varphi^*) \subset Y$. Statement (3) follows as we also have $X \subset \mathcal{V}(\ker \varphi^*)$.

For (4), we have $\mathcal{I}(Y) \subset \ker \varphi^*$ and so the map $\varphi^* \colon \mathbb{F}[\mathbb{A}^m] \to \mathbb{F}[X]$ factors through the quotient map $\mathbb{F}[\mathbb{A}^m] \twoheadrightarrow \mathbb{F}[\mathbb{A}^m]/\mathcal{I}(Y) = \mathbb{F}[Y]$. □

Thus we may refine the correspondence of Lemma 1.5.8. Let X and Y be affine varieties. Then the association $\varphi \mapsto \varphi^*$ gives a bijective correspondence

$$\left\{\begin{matrix}\text{regular maps} \\ \varphi \colon X \to Y\end{matrix}\right\} \quad \longleftrightarrow \quad \left\{\begin{matrix}\text{ring homomorphisms} \\ \psi \colon \mathbb{F}[Y] \to \mathbb{F}[X]\end{matrix}\right\}.$$

This map $X \mapsto \mathbb{F}[X]$ from affine varieties to finitely generated reduced $\mathbb{F}$-algebras not only maps objects to objects, but is an isomorphism on maps between objects (reversing their direction however). In mathematics, such an association is called a *contravariant equivalence of categories*. The point of this equivalence is that an affine variety and its coordinate ring are different packages for the same information. Each one determines and is determined by the other. Whether we study algebra or geometry, we are studying the same thing.

The prototypical example of a contravariant equivalence of categories comes from linear algebra. To a finite-dimensional vector space V, we may associate its dual space V^*. Given a linear transformation $L \colon V \to W$, its adjoint is a map $L^* \colon W^* \to V^*$. Since $(V^*)^* = V$ and $(L^*)^* = L$, this association is a bijection on the objects (finite-dimensional vector spaces) and a bijection on linear maps linear maps from V to W.

This equivalence of categories leads to the following question:

If affine varieties correspond to finitely generated reduced $\mathbb{F}$-algebras, which geometric objects correspond to finitely generated $\mathbb{F}$-algebras?

In modern algebraic geometry, these geometric objects are called *affine schemes*.

Exercises for Section 5

1. Give a proof of Theorem 1.5.4.
2. Show that a regular map $\varphi \colon X \to Y$ is continuous in the Zariski topology.
3. Show that if two varieties X and Y are isomorphic, then they are homeomorphic as topological spaces. Show that the converse does not hold.
4. Let $C = V(y^2 - x^3)$ show that the map $\phi : \mathbb{A}^1_{\mathbb{C}} \to C$, $\phi(t) = (t^2, t^3)$ is a homeomorphism in the Zariski topology but it is not an isomorphism of affine varieties.

5. Let $V = \mathcal{V}(y - x^2) \subset \mathbb{A}^2_{\mathbb{F}}$ and $W = \mathcal{V}(xy - 1) \subset \mathbb{A}^2_{\mathbb{F}}$. Show that

$$\begin{aligned} \mathbb{F}[V] &:= \mathbb{F}[x, y]/\mathcal{I}(V) \cong \mathbb{F}[t] \\ \mathbb{F}[W] &:= \mathbb{F}[x, y]/\mathcal{I}(W) \cong \mathbb{F}[t, t^{-1}] \end{aligned}$$

Conclude that the hyperbola $V(xy - 1)$ is not isomorphic to the affine line.

1.6 Rational functions

In algebraic geometry, we also use functions and maps between varieties which are not defined at all points of their domains. Workin with functions and maps not defined at all points is a special feature of algebraic geometry that sets it apart from other branches of geometry.

Suppose X is any irreducible affine variety. By Theorem 1.4.4, its ideal $\mathcal{I}(X)$ is prime, so its coordinate ring $\mathbb{F}[X]$ has no zero divisors ($0 \neq f, g \in \mathbb{F}[X]$ with $fg = 0$). A ring without zero divisors is called an *integral domain*. In exact analogy with the construction of the rational numbers $\mathbb{Q}$ as quotients of integers $\mathbb{Z}$, we may form the *function field* $\mathbb{F}(X)$ of X as the quotients of regular functions in $\mathbb{F}[X]$. Formally, $\mathbb{F}(X)$ is the collection of all quotients f/g with $f, g \in \mathbb{F}[X]$ and $g \neq 0$, where we identify

$$\frac{f_1}{g_1} = \frac{f_2}{g_2} \iff f_1 g_2 - f_2 g_1 = 0 \text{ in } \mathbb{F}[X]\,.$$

Example 1.6.1 The function field of affine space $\mathbb{A}^n$ is the collection of quotients of polynomials P/Q with $P, Q \in \mathbb{F}[x_1, \ldots, x_n]$. This field $\mathbb{F}(x_1, \ldots, x_n)$ is called the *field of rational functions* in the variables $x_1, \ldots, x_n$.

Given an irreducible affine variety $X \subset \mathbb{A}^n$, we may also express $\mathbb{F}(X)$ as the collection of quotients f/g of polynomials $f, g \in \mathbb{F}[\mathbb{A}^n]$ with $g \notin \mathcal{I}(X)$, where we identify

$$\frac{f_1}{g_1} = \frac{f_2}{g_2} \iff f_1 g_2 - f_2 g_1 \in \mathcal{I}(X)\,.$$

Rational functions on an affine variety X do not in general have unique representatives as quotients of polynomials or even quotients of regular functions.

Example 1.6.2 Let $X := \mathcal{V}(x^2 + y^2 + 2y) \subset \mathbb{A}^2$ be the circle of radius 1 and center at $(0, -1)$. In $\mathbb{F}(X)$ we have

$$-\frac{x}{y} = \frac{y^2 + 2y}{x}\,.$$

A point $x \in X$ is a *regular point* of a rational function $\varphi \in \mathbb{F}(X)$ if φ has a representative f/g with $f, g \in \mathbb{F}[X]$ and $g(x) \neq 0$. From this we see that all points of the

neighborhood X_g of x in X are regular points of φ. Thus the set of regular points of φ is a nonempty Zariski open subset of X. Call this the *domain of regularity of* φ.

When $x \in X$ is a regular point of a rational function $\varphi \in \mathbb{F}(X)$, we set $\varphi(x) := f(x)/g(x) \in \mathbb{F}$, where φ has representative f/g with $g(x) \neq 0$. The value of $\varphi(x)$ does not depend upon the choice of representative f/g of φ. In this way, φ gives a function from a dense subset of X (its domain of regularity) to $\mathbb{F}$. We write this as

$$\varphi : X \dashrightarrow \mathbb{F}$$

with the dashed arrow indicating that φ is not necessarily defined at all points of X.

The rational function φ of Example 1.6.2 has domain of regularity $X - \{(0,0)\}$. Here $\varphi : X \dashrightarrow \mathbb{F}$ is stereographic projection of the circle onto the line $y = -1$ from the point $(0,0)$.

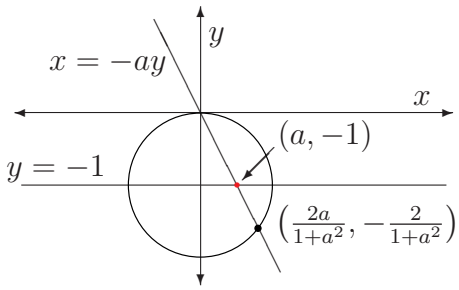


Figure 1.5: Projection of the circle $\mathcal{V}(x^2 + (y-1)^2 - 1)$ from the origin.

Example 1.6.3 Let $X = \mathbb{A}^1_{\mathbb{R}}$ and $\varphi = 1/(1+x^2) \in \mathbb{R}(X)$. Then every point of X is a regular point of φ. The existence of rational functions which are regular at every point, but are not elements of the coordinate ring, is a special feature of real algebraic geometry. Observe that φ is not regular at the points $\pm\sqrt{-1} \in \mathbb{A}^1_{\mathbb{C}}$.

Theorem 1.6.4 *When $\mathbb{F}$ is algebraically closed, a rational function that is regular at all points of an irreducible affine variety X is a regular function in $\mathbb{C}[X]$.*

Proof. For each point $x \in X$, there are regular functions $f_x, g_x \in \mathbb{F}[X]$ with $\varphi = f_x/g_x$ and $g_x(x) \neq 0$. Let $\mathcal{I}$ be the ideal generated by the regular functions g_x for $x \in X$. Then $\mathcal{V}(\mathcal{I}) = \emptyset$, as φ is regular at all points of X.

If we let $g_1, \ldots, g_s$ be generators of $\mathcal{I}$ and let $f_1, \ldots, f_s$ be regular functions such that $\varphi = f_i/g_i$ for each i. Then by the Weak Nullstellensatz for X (Theorem 1.5.4(3)), there are regular functions $h_1, \ldots, h_s \in \mathbb{C}[X]$ such that

$$1 \;=\; h_1 g_1 + \cdots + h_s g_s \,.$$

multiplying this equation by φ, we obtain

$$\varphi \;=\; h_1 f_1 + \cdots + h_s f_s \,,$$

which proves the theorem. □

A list $f_1, \ldots, f_m$ of rational functions gives a *rational map*

$$\begin{aligned} \varphi : X &\dashrightarrow \mathbb{A}^m, \\ x &\longmapsto (f_1(x), \ldots, f_m(x)). \end{aligned}$$

This rational map φ is only defined on the intersection U of the domains of regularity of each of the f_i. We call U the *domain of* φ and write $\varphi(X)$ for $\varphi(U)$.

Let X be an irreducible affine variety. Since $\mathbb{F}[X] \subset \mathbb{F}(X)$, any regular map is also a rational map. As with regular maps, a rational map $\varphi : X \dashrightarrow \mathbb{A}^m$ given by functions $f_1, \ldots, f_m \in \mathbb{F}(X)$ defines a homomorphism $\varphi^* : \mathbb{F}[\mathbb{A}^m] \to \mathbb{F}(X)$ by $\varphi^*(g) = g(f_1, \ldots, f_m)$. If Y is an affine subvariety of $\mathbb{A}^m$, then $\varphi(X) \subset Y$ if and only if $\varphi(\mathcal{I}(Y)) = 0$. In particular, the kernel J of the map $\varphi^* : \mathbb{F}[\mathbb{A}^m] \to \mathbb{F}(X)$ defines the smallest subvariety $Y = \mathcal{V}(J)$ containing $\varphi(X)$, that is, the Zariski closure of $\varphi(X)$. Since $\mathbb{F}(X)$ is a field, this kernel is a prime ideal, and so Y is irreducible.

When $\varphi : X \dashrightarrow Y$ is a rational map with $\varphi(X)$ dense in Y, then we say that φ is *dominant*. A dominant rational map $\varphi : X \dashrightarrow Y$ induces an embedding $\varphi^* : \mathbb{F}[Y] \hookrightarrow \mathbb{F}(X)$. Since Y is irreducible, this map extends to a map of function fields $\varphi^* : \mathbb{F}(Y) \to \mathbb{F}(X)$. Conversely, given a map $\psi : \mathbb{F}(Y) \to \mathbb{F}(X)$ of function fields, with $Y \subset \mathbb{A}^m$, we obtain a dominant rational map $\varphi : X \dashrightarrow Y$ given by the rational functions $\psi(x_1), \ldots, \psi(x_m) \in \mathbb{F}(X)$ where $x_1, \ldots, x_m$ are the coordinate functions on $Y \subset \mathbb{A}^m$.

Suppose we have two rational maps $\varphi\colon X \dashrightarrow Y$ and $\psi\colon Y \dashrightarrow Z$ with φ dominant. Then $\varphi(X)$ intersects the set of regular points of ψ, and so we may compose these maps $\psi \circ \varphi\colon X \dashrightarrow Z$. Two irreducible affine varieties X and Y are *birationally equivalent* if there is a rational map $\varphi : X \dashrightarrow Y$ with a rational inverse $\psi : Y \dashrightarrow X$. By this we mean that the compositions $\varphi \circ \psi$ and $\psi \circ \varphi$ are the identity maps on their respective domains. Equivalently, X and Y are birationally equivalent if and only if their function fields are isomorphic, if and only if they have isomorphic open subsets.

For example, the line $\mathbb{A}^1$ and the circle of Figure 1.5 are birationally equivalent. The inverse of stereographic projection from the circle to $\mathbb{A}^1$ is the map from $\mathbb{A}^1$ to the circle given by $a \mapsto (\frac{2a}{1+a^2}, -\frac{2}{1+a^2})$.

Exercises for Section 6

1. Show that irreducible affine varieties X and Y are birationally equivalent if and only if they have isomorphic open sets.

1.7 Smooth and singular points

Given a polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ and $a = (a_1, \ldots, a_n) \in \mathbb{A}^n$, we may write f as a polynomial in new variables $t = (x_1, \ldots, x_n)$, with $t_i = x_i - a_i$ and obtain

$$f \;=\; f(a) \;+\; \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a) \cdot t_i \;+\; \cdots\,, \tag{1.7}$$

where the remaining terms have degrees greater than 1 in the variables t. When $\mathbb{F}$ has characteristic zero, this is the usual Taylor expansion of f at the point a. The coefficient of the monomial t^α is the mixed partial derivative of f evaluated at a,

$$\frac{1}{\alpha_1!\alpha_2!\cdots\alpha_n!}\left(\frac{\partial}{\partial x_1}\right)^{\alpha_1}\left(\frac{\partial}{\partial x_2}\right)^{\alpha_2}\cdots\left(\frac{\partial}{\partial x_n}\right)^{\alpha_n} f(a)\,.$$

In the coordinates t for $\mathbb{F}^m$, the linear term in the expansion (1.7) is a linear map

$$d_a f \;:\; \mathbb{F}^n \;\longrightarrow\; \mathbb{F}$$

called the *differential* of f at the point a.

Definition 1.7.1 Let $X \subset \mathbb{A}^n$ be a subvariety. The *(Zariski) tangent space* T_aX to X at the point $a \in X$ is the joint kernel[†] of the linear maps $\{d_a f \mid f \in \mathcal{I}(X)\}$. Since

$$\begin{aligned} d_a(f+g) &= d_a f + d_a g \\ d_a(fg) &= f(a)d_a g + g(a)d_a f \end{aligned}$$

we do not need all the polynomials in $\mathcal{I}(X)$ to define T_aX, but may instead take any finite generating set.

Theorem 1.7.2 *Let X be an affine variety. Then the set of points of X whose tangent space has minimal dimension is a Zariski open subset of X.*

Proof. Let $f_1, \ldots, f_m$ be generators of $\mathcal{I}(X)$. Let $M \in \mathrm{Mat}_{m\times n}(\mathbb{F}[\mathbb{A}^n])$ be the matrix whose entry in row i and column j is $\partial f_i/\partial x_j$. For $a \in \mathbb{A}^n$, the components of the vector-valued function

$$\begin{aligned} M \;:\; \mathbb{F}^n &\longrightarrow \mathbb{F}^m \\ t &\longmapsto M(a)t \end{aligned}$$

are the differentials $d_a f_1, \ldots, d_a f_m$.

For each $\ell = 1, 2, \ldots, \max\{n, m\}$, the *degeneracy locus* $\Delta_\ell \subset \mathbb{A}^m$ is the variety defined by all $\ell \times \ell$ subdeterminants (*minors*) of the matrix M, and set $\Delta_{1+\max\{n,m\}} := \mathbb{A}^n$. Since

[†]Intersection of the nullspaces?

we may expand any $(i+1)\times(i+1)$ determinant along a row or column and express it in terms of $i\times i$ subdeterminants, these varieties are nested

$$\Delta_1 \subset \Delta_2 \subset \cdots \subset \Delta_{\max\{n,m\}} \subset \Delta_{1+\max\{n,m\}} = \mathbb{A}^n .$$

By definition, a point $a \in \mathbb{A}^n$ lies in $\Delta_{i+1} - \Delta_i$ if and only if the matrix $M(a)$ has rank exactly i. In particular, if $a \in \Delta_{i+1} - \Delta_i$, then the kernel of $M(a)$ has dimension $n-i$.

Let i be the minimal index with $X \subset \Delta_{i+1}$. Then

$$X - (X \cap \Delta_i) = \{a \in X \mid \dim T_a X = n-i\}$$

is a nonempty open subset of X and $n-i$ is the minimum dimension of a tangent space at a point of X. □

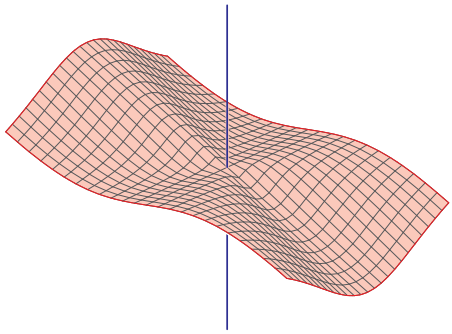
Suppose that X is irreducible and let m be the minimum dimension of a tangent space of X. Points $x \in X$ whose tangent space has this minimum dimension are call *smooth* and we write $X_{\rm sm}$ for the non-empty open subset of smooth points. The complement $X - X_{\rm sm}$ is the set $X_{\rm sing}$ of *singular* points of X. The set of smooth points is dense in X, for otherwise we may write the irreducible variety X as a union $\overline{X_{\rm sm}} \cup X_{\rm sing}$ of two proper closed subsets.

When X is irreducible, this minimum dimension of a tangent space is the dimension of X. This gives a second definition of dimension which is distinct from the combinatorial definition of Definition 1.4.

We have the following facts concerning the locus of smooth and singular points on a real or complex variety

Proposition 1.7.3 *The set of smooth points of an irreducible complex affine subvariety X of dimension d whose complex local dimension in the Euclidean topology is d is dense in the Euclidean topology.*

Example 1.7.4 Irreducible real algebraic varieties need not have this property. The Cartan umbrella $\mathcal{V}(z(x^2+y^2)-x^3)$



is a connected irreducible surface in $\mathbb{A}^3_{\mathbb{R}}$ where the local dimension of its smooth points is either 1 (along the z axis) or 2 (along the 'canopy' of the umbrella).

Chapter 2

Algorithms for Algebraic Geometry

Outline:

1. Gröbner basics.
2. Algorithmic applications of Gröbner bases.
3. Resultants and Bézout's Theorem.
4. Solving equations with Gröbner bases.
5. Numerical Homotopy continuation.

2.1 Gröbner basics

Gröbner bases are a foundation for many algorithms to represent and manipulate varieties on a computer. While these algorithms are important in applications, we shall see that Gröbner bases are also a useful theoretical tool.

A motivating problem is that of recognizing when a polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ lies in an ideal I. When the ideal I is radical and $\mathbb{F}$ is algebraically closed, this is equivalent to asking whether or not f vanishes on $\mathcal{V}(I)$. For example, we may ask which of the polynomials $x^3z - xz^3$, $x^2yz - y^2z^2 - x^2y^2$, and/or $x^2y - x^2z + y^2z$ lies in the ideal

$$\langle x^2y - xz^2 + y^2z,\ y^2 - xz + yz \rangle \ ?$$

This *ideal membership problem* is easy for univariate polynomials. Suppose that $I = \langle f(x), g(x), \ldots, h(x) \rangle$ is an ideal and $F(x)$ is a polynomial in $\mathbb{F}[x]$, the ring of polynomials in a single variable x. We determine if $F(x) \in I$ via a two-step process.

1. Use the Euclidean Algorithm to compute $\varphi(x) = \gcd(f(x), g(x), \ldots, h(x))$.

2. Use the Division Algorithm to determine if $\varphi(x)$ divides $F(x)$.

This is valid, as $I = \langle \varphi(x) \rangle$. The first step is a simplifications, where we find a simpler (lower-degree) polynomial which generates I, while the second step is a reduction, where we compute F modulo I. Both steps proceed systematically, operating on the terms of the polynomials involving the highest power of x. A good description for I is a prerequisite for solving our ideal membership problem.

We shall see how Gröbner bases give algorithms which extend this procedure to multivariate polynomials. In particular, a Gröbner basis of an ideal I gives a sufficiently good description of I to solve the ideal membership problem. Gröbner bases are also the foundation of algorithms that solve many other problems.

2.1.1 Monomial ideals

Monomial ideals are central to what follows. A *monomial* is a product of powers of the variables $x_1, \ldots, x_n$ with nonnegative integer exponents. The *exponent* α of a monomial $x^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$ is a vector $\alpha \in \mathbb{N}^n$. If we identify monomials with their exponent vectors, the multiplication of monomials corresponds to addition of vectors, and divisibility to the partial order on $\mathbb{N}^n$ of componentwise comparison.

Definition 2.1.1 A *monomial ideal* $I \subset \mathbb{F}[x_1, \ldots, x_n]$ is an ideal which satisfies the following two equivalent conditions.

(*i*) I is generated by monomials.

(*ii*) If $f \in I$, then every monomial of f lies in I.

One advantage of monomial ideals is that they are essentially combinatorial objects. By Condition (*ii*), a monomial ideal is determined by the set of monomials which it contains. Under the correspondence between monomials and their integer vector exponents, divisibility of monomials corresponds to componentwise comparison of vectors.

$$x^\alpha | x^\beta \iff \alpha_i \leq \beta_i \,,\ i = 1, \ldots, n \iff \alpha \leq \beta \,,$$

which defines a partial order on $\mathbb{N}^n$. Thus

$$(1,1,1) \leq (3,1,2) \qquad \text{but} \qquad (3,1,2) \not\leq (2,3,1) \,.$$

The set $O(I)$ of exponent vectors of monomials in a monomial ideal I has the property that if $\alpha \leq \beta$ with $\alpha \in O(I)$, then $\beta \in O(I)$. Thus $O(I)$ is an (upper) *order ideal* of the *poset* (*p*artially *o*rdered *set*) $\mathbb{N}^n$.

A set of monomials $G \subset I$ generates I if and only if every monomial in I is divisible by at least one monomial of G. A monomial ideal I has a unique minimal set of generators—these are the monomials x^α in I which are not divisible by any other monomial in I.

Let us look at some examples. When $n = 1$, monomials have the form x^d for some natural number $d \geq 0$. If d is the minimal exponent of a monomial in I, then $I = x^d$. Thus all monomial ideals have the form $\langle x^d \rangle$ for some $d \geq 0$.

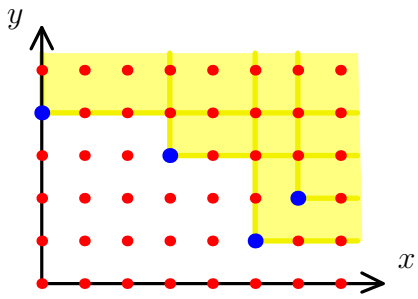
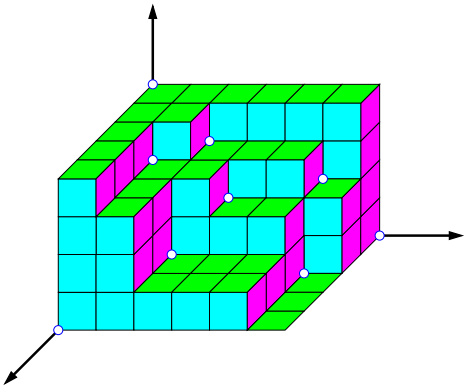


Figure 2.1: Exponents of monomials in the ideal $\langle y^4, x^3y^3, x^5y, x^6y^2\rangle$.

When $n = 2$, we may plot the exponents in the order ideal associated to a monomial ideal. For example, the lattice points in the shaded region of Figure 2.1 represent the monomials in the ideal $I := \langle y^4, x^3y^3, x^5y, x^6y^2\rangle$. From this picture we see that I is minimally generated by y^4, x^3y^3, and x^5y.

Since $x^ay^b \in I$ implies that $x^{a+\alpha}y^{b+\beta} \in I$ for any $(\alpha, \beta) \in \mathbb{N}^2$, a monomial ideal $I \subset \mathbb{F}[x, y]$ is the union of the shifted positive quadrants $(a, b) + \mathbb{N}^2$ for every monomial $x^ay^b \in I$. It follows that the monomials in I are those above the staircase shape that is the boundary of the shaded region. The monomials not in I lie under the staircase, and they form a vector space basis for the quotient ring $\mathbb{F}[x, y]/I$.

This notion of staircase for two variables makes sense when there are more variables. The *staircase* of an ideal consists of the monomials which are on the boundary of the ideal, in that they are visible from the origin of $\mathbb{N}^n$. For example, here is the staircase for the ideal $\langle x^5, x^2y^5, y^6, x^3y^2z, x^2y^3z^2, xy^5z^2, x^2yz^3, xy^2z^3, z^4\rangle$.



We offer a purely combinatorial proof that monomial ideals are finitely generated, which is independent of the Hilbert Basis Theorem.

Lemma 2.1.2 (Dickson's Lemma) *Monomial ideals are finitely generated.*

Proof. We prove this by induction on n. The case $n = 1$ was covered in the preceding examples.

Let $I \subset \mathbb{F}[x_1, \ldots, x_n, y]$ be a monomial ideal. For each $d \in \mathbb{N}$, observe that the monomials

$$\{x^\alpha \mid x^\alpha y^d \in I\}$$

form a monomial ideal I_d of $\mathbb{F}[x_1, \ldots, x_n]$, and the union of all such monomials

$$\{x^\alpha \mid x^\alpha y^d \in I \text{ for some } d \geq 0\}\,.$$

form a monomial ideal I_∞ of $\mathbb{F}[x_1, \ldots, x_n]$. By our inductive hypothesis, I_d has a finite generating set G_d, for each $d = 0, 1, \ldots, \infty$.

Note that $I_0 \subset I_1 \subset \cdots \subset I_\infty$. We must have $I_\infty = I_d$ for some $d < \infty$. Indeed, each generator $x^\alpha \in G_\infty$ of I_∞ comes from a monomial $x^\alpha y^b$ in I, and we may let d be the maximum of the numbers b which occur. Since $I_\infty = I_d$, we have $I_b = I_d$ for any $b > d$. Note that if $b > d$, then we may assume that $G_b = G_d$ as $I_b = I_d$.

We claim that the finite set

$$G \;=\; \bigcup_{b=0}^{d} \{x^\alpha y^b \mid x^\alpha \in G_b\}$$

generates I. Indeed, suppose that $x^\alpha y^b$ is a monomial in I. We find a monomial in G which divides $x^\alpha y^b$. Since $x^\alpha \in I_b$, there is a generator $x^\gamma \in G_b$ which divides x^α. If $b \leq d$, then $x^\gamma y^b \in G$ is a monomial dividing $x^\alpha y^b$. If $b > d$, then $x^\gamma y^d \in G$ as $G_b = G_d$ and $x^\gamma y^d$ divides $x^\alpha y^b$. □

A simple consequence of Dickson's Lemma is that any strictly increasing chain of monomial ideals is finite. Suppose that

$$I_1 \;\subset\; I_2 \;\subset\; I_3 \;\subset\; \cdots$$

is an increasing chain of monomial ideals. Let I_∞ be their union, which is another monomial ideal. Since I_∞ is finitely generated, there must be some ideal I_d which contains all generators of I_∞, and so $I_d = I_{d+1} = \cdots = I_\infty$. We used this fact crucially in our proof of Dickson's lemma.

2.1.2 Monomial orders and Gröbner bases

The key idea behind Gröbner bases is to determine what is meant by 'term of highest power' in a polynomial having two or more variables. There is no canonical way to do this, so we must make a choice, which is encoded in the notion of a term or monomial order. An order $\succ$ on monomials in $\mathbb{F}[x_1, \ldots, x_n]$ is *total* if for monomials x^α and x^β exactly one of the following holds

$$x^\alpha \;\succ\; x^\beta \qquad \text{or} \qquad x^\alpha \;=\; x^\beta \qquad \text{or} \qquad x^\alpha \;\prec\; x^\beta\,.$$

Definition 2.1.3 A *monomial order* on $\mathbb{F}[x_1, \ldots, x_n]$ is a total order $\succ$ on the monomials in $\mathbb{F}[x_1, \ldots, x_n]$ such that

(*i*) 1 is the minimal element under $\succ$.

(*ii*) $\succ$ respects multiplication by monomials: If $x^\alpha \succ x^\beta$ then $x^\alpha \cdot x^\gamma \succ x^\beta \cdot x^\gamma$, for any monomial x^γ.

Conditions (*i*) and (*ii*) in Definition 2.1.3 imply that if x^α is divisible by x^β, then $x^\alpha \succ x^\beta$. A *well-ordering* is a total order with no infinite descending chain, equivalently, one in which every subset has a minimal element.

Lemma 2.1.4 *Monomial orders are exactly the well-orderings $\succ$ on monomials that satisfy Condition (ii) of Definition* 2.1.3.

Proof. Let $\succ$ be a well-ordering on monomials which satisfies Condition (ii) of Definition 2.1.3. Suppose that $\succ$ is not a monomial order, then there is some monomial x^a with $1 \succ x^a$. By Condition (ii), we have $1 \succ x^a \succ x^{2a} \succ x^{3a} \succ \cdots$, which contradicts $\succ$ being a well-order.

Let $\succ$ be a monomial order and M be any set of monomials. Set I to be the ideal generated by M. By Dickson's Lemma, M has a finite subset G which generates I. Since G is finite, let x^γ be the minimal monomial in G under $\succ$. We claim that x^γ is the minimal monomial in M.

Let $x^\alpha \in M$. Since G generates I and $M \subset I$, there is some $x^\beta \in G$ which divides x^α and thus $x^\alpha \succ x^\beta$. But x^γ is the minimal monomial in G, so $x^\alpha \succ x^\beta \succ x^\gamma$. □

The well-ordering property of monomials orders is key to what follows, as many proofs use induction on $\succ$, which is only possible as $\succ$ is a well-ordering.

Example 2.1.5 The (*total*) *degree* $\deg(x^\alpha)$ of a monomial $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ is $\alpha_1 + \cdots + \alpha_n$. We describe four important monomial orders.

1. The *lexicographic order* $\succ_{\text{lex}}$ on $\mathbb{F}[x_1, \ldots, x_n]$ is defined by

$$x^\alpha \succ_{\text{lex}} x^\beta \iff \left\{ \begin{array}{l} \text{The first non-zero entry of the} \\ \text{vector } \alpha - \beta \text{ in } \mathbb{Z}^n \text{ is positive.} \end{array} \right\}$$

2. The *degree lexicographic order* $\succ_{\text{dlx}}$ on $\mathbb{F}[x_1, \ldots, x_n]$ is defined by

$$x^\alpha \succ_{\text{dlx}} x^\beta \iff \left\{ \begin{array}{ll} \deg x^\alpha > \deg x^\beta & \text{or}, \\ \deg x^\alpha = \deg x^\beta & \text{and } \ x^\alpha \succ_{\text{lex}} x^\beta . \end{array} \right.$$

3. The *degree reverse lexicographic order* $\succ_{\text{drl}}$ $\mathbb{F}[x_1, \ldots, x_n]$ is defined by

$$x^\alpha \succ_{\text{drl}} x^\beta \iff \left\{ \begin{array}{ll} \deg x^\alpha > x^\beta & \text{or}, \\ \deg x^\alpha = \deg x^\beta & \text{and the last non-zero entry of the} \\ & \text{vector } \alpha - \beta \text{ in } \mathbb{Z}^n \text{ is negative}. \end{array} \right.$$

4. More generally, we may have *weighted orders*. Let $c \in \mathbb{R}^n$ be a vector with non-negative components, called a weight. This defines a partial order $\succ_c$ on monomials

$$x^\alpha \succ_c x^\beta \iff c \cdot \alpha > c \cdot \beta \,.$$

If all components of c are positive, then $\succ_c$ satisfies the two conditions of Definition 2.1.3. Its only failure to be a monomial order is that it may not be a total order on monomials. (For example, consider $c = (1, 1, \ldots, 1)$.) This may be remedied by picking a monomial order to break ties. For example, if we use $\succ_{\text{lex}}$, then we get a monomial order

$$x^\alpha \succ_{c,\text{lex}} x^\beta \iff \begin{cases} \omega \cdot \alpha > \omega \cdot \beta & \text{or}\,, \\ \omega \cdot \alpha = \omega \cdot \beta & \text{and } x^\alpha \succ_{\text{lex}} x^\beta \end{cases}$$

Another way to do this is to break the ties with a different monomial order, or a different weight.

You are asked to prove these are monomial orders in Exercise 7.

Remark 2.1.6 We compare these three orders on monomials of degrees 1 and 2 in $\mathbb{F}[x, y, z]$ where the variables are ordered $x \succ y \succ z$.

$$\begin{aligned} &x^2 \succ_{\text{lex}} xy \succ_{\text{lex}} xz \succ_{\text{lex}} x \succ_{\text{lex}} y^2 \succ_{\text{lex}} yz \succ_{\text{lex}} y \succ_{\text{lex}} z^2 \succ_{\text{lex}} z \\ &x^2 \succ_{\text{dlx}} xy \succ_{\text{dlx}} xz \succ_{\text{dlx}} y^2 \succ_{\text{dlx}} yz \succ_{\text{dlx}} z^2 \succ_{\text{dlx}} x \succ_{\text{dlx}} y \succ_{\text{dlx}} z \\ &x^2 \succ_{\text{drl}} xy \succ_{\text{drl}} y^2 \succ_{\text{drl}} xz \succ_{\text{drl}} yz \succ_{\text{drl}} z^2 \succ_{\text{drl}} x \succ_{\text{drl}} y \succ_{\text{drl}} z \end{aligned}$$

For the remainder of this section, $\succ$ will denote a fixed, but arbitrary monomial order on $\mathbb{F}[x_1, \ldots, x_n]$. A *term* is a product ax^α of a scalar $a \in \mathbb{F}$ with a monomial x^α. We may extend any monomial order $\succ$ to an order on terms by setting $ax^\alpha \succ bx^\beta$ if $x^\alpha \succ x^\beta$ and $ab \neq 0$. Such a term order is no longer a partial order as different terms with the same monomial are incomparable. For example $3x^2y$ and $5x^2y$ are incomparable. Term orders are however well-founded in that they have no infinite strictly decreasing chains.

The *initial term* $\text{in}_\succ(f)$ of a polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ is the term of f that is maximal with respect to $\succ$ among all terms of f. For example, if $\succ$ is lexicographic order with $x \succ y$, then

$$\text{in}_\succ(3x^3y - 7xy^{10} + 13y^{30}) \;=\; 3x^3y\,.$$

When $\succ$ is understood, we may write $\text{in}(f)$. The *initial ideal* $\text{in}_\succ(I)$ (or $\text{in}(I)$) of an ideal $I \subset \mathbb{F}[x_1, \ldots, x_n]$ is the ideal generated by the initial terms of polynomials in I,

$$\text{in}_\succ(I) \;=\; \langle \text{in}_\succ(f) \mid f \in I \rangle\,.$$

We make the most important definition of this section.

Definition 2.1.7 Let $I \subset \mathbb{F}[x_1, \ldots, x_n]$ be an ideal and $\succ$ a monomial order. A set $G \subset I$ is a *Gröbner basis* for I with respect to the monomial order $\succ$ if the initial ideal $\mathrm{in}_\succ(I)$ is generated by the initial terms of polynomials in G, that is, if

$$\mathrm{in}_\succ(I) \;=\; \langle \mathrm{in}_\succ(g) \mid g \in G \rangle \,.$$

Notice that if G is a Gröbner basis and $G \subset G'$, then G' is also a Gröbner basis.

We justify our use of the term 'basis' in 'Gröbner basis'.

Lemma 2.1.8 *If G is a Gröbner basis for I with respect to a monomial order $\succ$, then G generates I.*

Proof. We will prove this by induction on $\mathrm{in}(f)$ for $f \in I$. Let $f \in I$. Since $\{\mathrm{in}(g) \mid g \in G\}$ generates $\mathrm{in}(I)$, there is a polynomial $g \in G$ whose initial term $\mathrm{in}(g)$ divides the initial term $\mathrm{in}(f)$ of f. Thus there is some term ax^α so that

$$\mathrm{in}(f) \;=\; ax^\alpha \mathrm{in}(g) \;=\; \mathrm{in}(ax^\alpha g)\,,$$

as $\succ$ respects multiplication. If we set $f_1 := f - cx^\alpha g$, then $\mathrm{in}(f) \succ \mathrm{in}(f_1)$.

It follows that if $\mathrm{in}(f)$ is the minimal monomial in $\mathrm{in}I$, then $f_1 = 0$ and so $f \in \langle G \rangle$. In fact, $f \in G$ up to a scalar multiple. Suppose now that whenever $\mathrm{in}(f) \succ \mathrm{in}(h)$ and $h \in I$, then $h \in \langle G \rangle$. But then $f_1 = f - cx^\alpha g \in \langle G \rangle$, and so $f \in \langle G \rangle$. □

An immediate consequence of Dickson's Lemma and Lemma 2.1.8 is the following Gröbner basis version of the Hilbert Basis Theorem.

Theorem 2.1.9 (Hilbert Basis Theorem) *Every ideal $I \subset \mathbb{F}[x_1, \ldots, x_n]$ has a finite Gröbner basis with respect to any given monomial order.*

Example 2.1.10 Different monomial orderings give different Gröbner bases, and the sizes of the Gröbner bases can vary. Consider the ideal generated by the three polynomials

$$xy^3 + xz^3 + x - 1, \quad yz^3 + yx^3 + y - 1, \quad zx^3 + zy^3 + z - 1$$

In the degree reverse lexicographic order, where $x \succ y \succ z$, this has a Gröbner basis

$x^3z + y^3z + z - 1$,
$xy^3 + xz^3 + x - 1$,
$x^3y + yz^3 + y - 1$,
$y^4z - yz^4 - y + z$,
$2xyz^4 + xyz + xy - xz - yz$,
$2y^3z^3 - x^3 + y^3 + z^3 + x^2 - y^2 - z^2$,
$y^6 - z^6 - y^5 + y^3z^2 - 2x^2z^3 - y^2z^3 + z^5 + y^3 - z^3 - x^2 - y^2 + z^2 + x$,
$x^6 - z^6 - x^5 - y^3z^2 - x^2z^3 - 2y^2z^3 + z^5 + x^3 - z^3 - x^2 - y^2 + y + z$,
$2z^7 + 4x^2z^4 + 4y^2z^4 - 2z^6 + 3z^4 - x^3 - y^3 + 3x^2z + 3y^2z - 2z^3 + x^2 + y^2 - 2xz - 2yz - z^2 + z - 1$,

$2yz^6 + y^4 + 2yz^3 + x^2y - y^3 + yz^2 - 2z^3 + y - 1,$
$2xz^6 + x^4 + 2xz^3 - x^3 + xy^2 + xz^2 - 2z^3 + x - 1,$

consisting of 11 polynomials with largest coefficient 4 and degree 7. If we consider instead the lexicographic monomial order, then this ideal has a Gröbner basis

$$64z^{34} - 64z^{33} + 384z^{31} - 192z^{30} - 192z^{29} + 1008z^{28} + 48z^{27} - 816z^{26} + 1408z^{25} + 976z^{24}$$
$$-1296z^{23} + 916z^{22} + 1964z^{21} - 792z^{20} - 36z^{19} + 1944z^{18} + 372z^{17} - 405z^{16} + 1003z^{15}$$
$$+879z^{14} - 183z^{13} + 192z^{12} + 498z^{11} + 7z^{10} - 94z^9 + 78z^8 + 27z^7 - 47z^6 - 31z^5 + 4z^3$$
$$-3z^2 - 4z - 1,$$

$$64yz^{21} + 288yz^{18} + 96yz^{17} + 528yz^{15} + 384yz^{14} + 48yz^{13} + 504yz^{12} + 600yz^{11} + 168yz^{10}$$
$$+200yz^9 + 456yz^8 + 216yz^7 + 120yz^5 + 120yz^4 - 8yz^2 + 16yz + 8y - 64z^{33} + 128z^{32}$$
$$-128z^{31} - 320z^{30} + 576z^{29} - 384z^{28} - 976z^{27} + 1120z^{26} - 144z^{25} - 2096z^{24} + 1152z^{23}$$
$$+784z^{22} - 2772z^{21} + 232z^{20} + 1520z^{19} - 2248z^{18} - 900z^{17} + 1128z^{16} - 1073z^{15} - 1274z^{14}$$
$$+229z^{13} - 294z^{12} - 966z^{11} - 88z^{10} - 81z^9 - 463z^8 - 69z^7 + 26z^6 - 141z^5 - 32z^4 + 24z^3$$
$$-12z^2 - 11z + 1$$

$$589311934509212912y^2 - 11786238690184258240yz^{20} - 9428990952147406592yz^{19}$$
$$-2357247738036851648yz^{18} - 483235786297554587 84yz^{17} - 4832357862975545878 4yz^{16}$$
$$-2003660577331323900 8yz^{15} - 8191435889678059476 8yz^{14} - 9782578112852934339 2yz^{13}$$
$$-5303807410582916208 0yz^{12} - 7867314325697992375 2yz^{11} - 9988837289931158858 4yz^{10}$$
$$-6364568892694994496yz^9 - 3712665187408041345 6yz^8 - 4390373912093636194 4yz^7$$
$$-3447474816878895535 2yz^6 - 9134334984892800136yz^5 - 589311934509212912 0yz^4$$
$$-4125183541564490384yz^3 - 1178623869018425824yz^2 - 2062591770782245192yz$$
$$-1178623869018425824y + 46665645155349846336z^{33} - 52561386330338650688z^{32}$$
$$+25195872352020329920z^{31} + 28156769162372952723 2z^{30} - 19392177430724378694 4z^{29}$$
$$-22383823960598695936z^{28} + 81706533724600969099 2z^{27} - 16308104685758723524 8z^{26}$$
$$-42770559036883403033 6z^{25} + 13905781683718208538 08z^{24} + 39000434368484674580 8z^{23}$$
$$-98032219788785598166 4z^{22} + 13454251172212979738 76z^{21} + 12879560659390367316 76z^{20}$$
$$-95338316228249822884 4z^{19} + 63120234731058122985 6z^{18} + 17043019678692273960 24z^{17}$$
$$-15520856778655514998 8z^{16} - 16764066862257396505z^{15} + 12574754032771507009 61z^{14}$$
$$+52668596890136716959 8z^{13} - 16475153000055626488 0z^{12} + 49124953163927565405 0z^{11}$$
$$+45712630887118688230 6z^{10} - 87008396189513562747z^9 + 15803768907185828750z^8$$
$$+13932068156394410127 3z^7 - 17355919586383317961z^6 - 50777365233910819054z^5$$
$$-4630862847055988750z^4 + 8085080238139562826z^3 + 1366850803924776890z^2$$
$$-3824545208919673161z - 2755936363893486164,$$

$$589311934509212912x + 589311934509212912y - 879663783965093185 92z^{33}$$
$$+13338340253167146649 6z^{32} - 59115312141727767552z^{31} - 50692680764859328012 8z^{30}$$
$$+52214177181017233427 2z^{29} + 48286434009450032640z^{28} - 14347259883387363887 52z^{27}$$
$$+62997181176686959171 2z^{26} + 91798600277439166526 4z^{25} - 23898711989748432051 36z^{24}$$
$$-24698231483106694188 8z^{23} + 20389689261052715195 36z^{22} - 21748963896433430866 20z^{21}$$
$$-17581387825462211569 76z^{20} + 20253901854065627985 52z^{19} - 77454264142036382836 4z^{18}$$
$$-23653906414512782784 84z^{17} + 62782483555936330499 2z^{16} + 39848463323285911590 7z^{15}$$
$$-15486831101309342203 22z^{14} - 50019266671009151041 9z^{13} + 55192142799847475851 0z^{12}$$

$-490368794345102286410z^{11} - 480504004841899057384z^{10} + 220514007454401175615z^9$
$+38515984901980047305z^8 - 136644301635686684609z^7 + 174107126941325207 94z^6$
$+58724552354094225803z^5 + 157023419718953073 56z^4 - 744005890769778933 2z^3$
$-13983410894686689 12z^2 + 391320563053161239 7z + 2689145244006168857,$

consisting of 4 polynomials with largest degree 34 and significantly larger coefficients.

Exercises for Section 1

1. Prove the equivalence of conditions (*i*) and (*ii*) in Definition 2.1.1.

2. Show that a monomial ideal is radical if and only if it is square-free. (Square-free means that it has generators in which no variable occurs to a power greater than 1.)

3. Show that the elements of a monomial ideal I which are minimal with respect to division form a minimal set of generators of I in that they generate I and are a subset of any generating set of I.

4. Which of the polynomials $x^3z - xz^3$, $x^2yz - y^2z^2 - x^2y^2$, and/or $x^2y - x^2z + y^2z$ lies in the ideal
$$\langle x^2y - xz^2 + y^2z,\ y^2 - xz + yz \rangle \ ?$$

5. Using Definition 2.1.1, show that a monomial order is a linear extension of the divisibility partial order on monomials.

6. Show that if an ideal I has a square-free initial ideal, then I is radical. Give an example to show that the converse of this statement is false.

7. Show that each of the order relations $\succ_{\text{lex}}$, $\succ_{\text{dlx}}$, and $\succ_{\text{drl}}$, are monomial orders. Show that if the coordinates of $\omega \in \mathbb{R}^n_>$ are linearly independent over $\mathbb{Q}$, then $\succ_\omega$ is a monomial order. Show that each of $\succ_{\text{lex}}$, $\succ_{\text{dlx}}$, and $\succ_{\text{drl}}$ are weighted orders.

8. Show that a term order is well-founded.

9. Show that for a monomial order $\succ$, $\text{in}(I)\text{in}(J) \subseteq \text{in}(IJ)$ for any two ideals I and J. Find I and J such that the inclusion is proper.

10. Let $I := \langle x^2 + y^2, x^3 + y^3 \rangle \subset \mathbb{Q}[x, y]$. Let our monomial order be $\succ_{\text{lex}}$, the lexicographic order with $x \succ_{\text{lex}} y$.

    (a) Prove that $y^4 \in I$.

    (b) Show that the reduced Gröbner basis for I is $\{y^4, xy^2 - y^3, x^2 + y^2\}$.

    (c) Show that $\{x^2 + y^2, x^3 + y^3\}$ cannot be a Gröbner basis for I for any monomial ordering.

2.2 Algorithmic applications of Gröbner bases

Many practical algorithms to study and manipulate ideals and varieties are based on Gröbner bases. The foundation of algorithms involving Gröbner bases is the multivariate division algorithm. The subject began with Buchberger's thesis which contained his algorithm to compute Gröbner bases [3, 4].

2.2.1 Ideal membership and standard monomials

Both steps in the algorithm for ideal membership in one variable relied on the the same elementary procedure: using a polynomial of low degree to simplify a polynomial of higher degree. This same procedure was also used in the proof of Lemma 2.1.8. Those ideas lead to the *multivariate division algorithm*, which is a cornerstone of the theory of Gröbner bases.

Algorithm 2.2.1 (Multivariate division algorithm)
INPUT: Polynomials $g_1, \ldots, g_m$ and f in $\mathbb{F}[x_1, \ldots, x_n]$ and a monomial order $\succ$.
OUTPUT: Polynomials $q_1, \ldots, q_m$ and r such that

$$f \;=\; q_1 g_1 + q_2 g_2 + \cdots + q_m g_m + r\,, \tag{2.1}$$

where no term of r is divisible by an initial term of any polynomial g_i and we also have $\mathrm{in}(f) \succeq \mathrm{in}(r)$, and $\mathrm{in}(f) \succeq \mathrm{in}(q_i g_i)$, for each $i = 1, \ldots, m$.
INITIALIZE: Set $r := f$ and $q_1 := 0$, ..., $q_m := 0$.

(1) If no term of r is divisible by an initial term of some g_i, then exit.

(2) Otherwise, let ax^α be the largest (with respect to $\succ$) term of r divisible by some $\mathrm{in}(g_i)$. Choose j minimal such that $\mathrm{in}(g_j)$ divides x^α and suppose that $ax^\alpha = bx^\beta \cdot \mathrm{in}(g_j)$. Replace r by $r - bx^\beta g_j$ and q_j by $q_j + bx^\beta$, and return to step (1).

Proof of correctness. Each iteration of (2) is a *reduction* of r by the polynomials $g_1, \ldots, g_m$. With each reduction, the largest term in r divisible by some $\mathrm{in}(g_i)$ decreases with respect to $\succ$. Since the monomial order $\succ$ is well-founded, this algorithm must terminate after a finite number of steps. Every time the algorithm executes step (1), condition (2.1) holds. We also always have $\mathrm{in}(f) \succeq \mathrm{in}(r)$ because it holds initially, and with every reduction the new terms of r are less than the term which was canceled. Lastly, $\mathrm{in}(f) \succeq \mathrm{in}(q_i g_i)$ always holds, because it held initially, and the initial terms of the $q_i g_i$ are always terms of r. □

Given a list $G = (g_1, \ldots, g_m)$ of polynomials and a polynomial f, let r be the remainder obtained by the multivariate division algorithm applied to G and f. Since $f - r$ lies in the ideal generated by G, we write $f \bmod G$ for this remainder r. While it is clear (and expected) that $f \bmod G$ depends on the monomial order $\succ$, in general it will also depend

upon the order of the polynomials $(g_1, \ldots, g_m)$. For example, in the degree lexicographic order

$$\begin{aligned} x^2y \mod (x^2,\, xy + y^2) &= 0\,, \quad \text{but} \\ x^2y \mod (xy + y^2,\, x^2) &= y^3\,. \end{aligned}$$

This example shows that we cannot reliably use the multivariate division algorithm to test when f is in the ideal generated by G. However, this does not occur when G is a Gröbner basis.

Lemma 2.2.2 (Ideal membership test) *Let G be a finite Gröbner basis for an ideal I with respect to a monomial order $\succ$. Then a polynomial $f \in I$ if and only if $f \mod G = 0$.*

Proof. Set $r := f \mod G$. If $r = 0$, then $f \in I$. Suppose $r \neq 0$. Since no term of r is divisible any initial term of a polynomial in G, its initial term $\mathrm{in}(r)$ is not in the initial ideal of I, as G is a Gröbner basis for I. But then $r \notin I$, and so $f \notin I$. □

When G is a Gröbner basis for an ideal I, the remainder $f \mod G$ is a linear combination of monomials that do not lie in the initial ideal of I. A monomial x^α is *standard* if $x^\alpha \notin \mathrm{in}(I)$. The images of standard monomials in the ring $\mathbb{F}[x_1, \ldots, x_n]/\mathrm{in}(I)$ form a vector space basis. Much more interesting is the following theorem of Macaulay [13].

Theorem 2.2.3 *Let $I \subset \mathbb{F}[x_1, \ldots, x_n]$ be an ideal and $\succ$ a monomial order. Then the images of standard monomials in $\mathbb{F}[x_1, \ldots, x_n]/I$ form a vector space basis.*

Proof. Let G be a finite Gröbner basis for I with respect to $\succ$. Given a polynomial f, both f and $f \mod G$ represent the same element in $\mathbb{F}[x_1, \ldots, x_n]/I$. Since $f \mod G$ is a linear combination of standard monomials, the standard monomials span $\mathbb{F}[x_1, \ldots, x_n]/I$.

A linear combination f of standard monomials is zero in $\mathbb{F}[x_1, \ldots, x_n]/I$ only if $f \in I$. But then $\mathrm{in}(f)$ is both standard and lies in $\mathrm{in}(I)$, and so we conclude that $f = 0$. Thus the standard monomials are linearly independent in $\mathbb{F}[x_1, \ldots, x_n]/I$. □

Because of this result, if we have a monomial order $\succ$ and an ideal I, then for every polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$, there is a unique polynomial $\overline{f}$ which involves only standard monomials such that f and $\overline{f}$ have the same image in the quotient ring $\mathbb{F}[x_1, \ldots, x_m]/I$. Moreover, this polynomial $\overline{f}$ may be computed from f with the division algorithm and any Gröbner basis G for I with respect to the monomial order $\succ$. This unique representative $\overline{f}$ of f is typically called the *normal form* of f modulo I and the division algorithm called with a Gröbner basis for I is often called normal form reduction.

Macaulay's Theorem shows that a Gröbner basis allows us to compute in the quotient ring $\mathbb{F}[x_1, \ldots, x_n]/I$ using the operations of the polynomial ring and ordinary linear algebra. Indeed, suppose that G is a finite Gröbner basis for an ideal I with respect to a given monomial order $\succ$ and that $f, g \in \mathbb{F}[x_1, \ldots, x_n]/I$ are in normal form, expressed as a linear combination of standard monomials. Then $f + g$ is a linear combination of standard monomials, and we can compute the product fg in the quotient ring as $fg \mod G$, where this product is taken in the polynomial ring.

2.2.2 Buchberger's algorithm

Theorem 2.1.9, which asserted the existence of a finite Gröbner basis, was purely existential. To use Gröbner bases, we need methods to detect and generate them. Such methods were given by Bruno Buchberger in his 1965 Ph.D. thesis [4]. Key ideas about Gröbner bases had appeared earlier in work of Gordan and of Macaulay, and in Hironaka's resolution of singularities [10]. Hironaka called Gröbner bases "standard bases", a term which persists. For example, in the computer algebra package `Singular` [8] the command `std(I);` computes the Gröbner basis of an ideal `I`. Despite these precedents, the theory of Gröbner bases rightly begins with these Buchberger's contributions.

A given set of generators for an ideal will fail to be a Gröbner basis if the initial terms of the generators fail to generate the initial ideal. That is, if there are polynomials in the ideal whose initial terms are not divisible by the initial terms of our generators. A necessary step towards generating a Gröbner basis is to generate polynomials in the ideal with 'new' initial terms. This is the *raison d'etre* for the following definition.

Definition 2.2.4 The *least common multiple*, $\mathrm{lcm}\{ax^\alpha, bx^\beta\}$ of two terms ax^α and bx^β is the minimal monomial x^γ divisible by both x^α and x^β. Here, γ is the exponent vector that is the componentwise maximum of α and β.

Let $0 \neq f, g \in \mathbb{F}[x_1, \ldots, x_n]$ and suppose $\succ$ is a monomial order. The *S-polynomial* of f and g, $\mathrm{Spol}(f, g)$, is the polynomial linear combination of f and g,

$$\mathrm{Spol}(f,g) \;:=\; \frac{\mathrm{lcm}\{\mathrm{in}(f),\mathrm{in}(g)\}}{\mathrm{in}(f)} f \;-\; \frac{\mathrm{lcm}\{\mathrm{in}(f),\mathrm{in}(g)\}}{\mathrm{in}(g)} g\,.$$

Note that both terms in this expression have initial term equal to $\mathrm{lcm}\{\mathrm{in}(f), \mathrm{in}(g)\}$.

Buchberger gave the following simple criterion to detect when a set G of polynomials is a Gröbner basis for the ideal it generates.

Theorem 2.2.5 (Buchberger's Criterion) *A set G of polynomials is a Gröbner basis for the ideal it generates if and only if for for all pairs $f, g \in G$,*

$$\mathrm{Spol}(f,g) \bmod G \;=\; 0\,.$$

Proof. Suppose first that G is a Gröbner basis for I with respect to $\succ$. Then, for $f, g \in G$, their S-polynomial $\mathrm{Spol}(f, g)$ lies in I and the ideal membership test implies that $\mathrm{Spol}(f, g) \bmod G = 0$.

Now suppose that $G = \{g_1, \ldots, g_m\}$ satisfies Buchberger's criterion. Let $f \in \langle G\rangle$ be a polynomial in the ideal generated by G. We will show that $\mathrm{in}(f)$ is divisible by $\mathrm{in}(g)$, for some $g \in G$. This implies that G is a Gröbner basis for $\langle G\rangle$.

Given a list $h = (h_1, \ldots, h_m)$ of polynomials in $\mathbb{F}[x_1, \ldots, x_n]$ let $m(h)$ be the largest monomial appearing in one of $h_1 g_1, \ldots, h_m g_m$. This will necessarily be the leading monomial in at least one of $\mathrm{in}(h_1 g_1), \ldots, \mathrm{in}(h_m g_m)$. Let $j(h)$ be the minimum index i for which $m(h)$ is the monomial of $\mathrm{in}(h_i g_i)$.

Consider lists $h = (h_1, \ldots, h_m)$ of polynomials with

$$f \;=\; h_1 g_1 + \cdots + h_m g_m \tag{2.2}$$

for which $m(h)$ minimal among all lists satisfying (2.2). Of these, let h be a list with $j := j(h)$ maximal. We claim that $m(h)$ is the monomial of $\mathrm{in}(f)$, which implies that $\mathrm{in}(g_j)$ divides $\mathrm{in}(f)$.

Otherwise, $m(h) \succ \mathrm{in}(f)$, and so the initial term $\mathrm{in}(h_j g_j)$ must be canceled in the sum (2.2). Thus there is some index k such that $m(h)$ is the monomial of $\mathrm{in}(h_k g_k)$. By our assumption on j, we have $k > j$. Let $x^\beta := \mathrm{lcm}\{\mathrm{in}(g_j), \mathrm{in}(g_k)\}$, the monomial which is canceled in $\mathrm{Spol}(g_j, g_k)$. Since $\mathrm{in}(g_j)$ and $\mathrm{in}(g_k)$ both divide $m(h)$ and thus both must divide $\mathrm{in}(h_j g_j)$, there is some term ax^α such that $ax^\alpha x^\beta = \mathrm{in}(h_j g_j)$. Let $cx^\gamma := \mathrm{in}(h_j g_j)/\mathrm{in}(g_k)$. Then

$$ax^\alpha \mathrm{Spol}(g_j, g_k) \;=\; ax^\alpha \frac{x^\beta}{\mathrm{in}(g_j)} g_j \;-\; ax^\alpha \frac{x^\beta}{\mathrm{in}(g_k)} g_k \;=\; \mathrm{in}(h_j) g_j \;-\; cx^\gamma g_k \,.$$

By our assumption that Buchberger's criterion is satisfied, there are polynomials $q_1, \ldots, q_m$ such that

$$\mathrm{Spol}(g_j, g_k) \;=\; q_1 g_1 + \cdots + q_m g_m \,.$$

Define a new list h' of polynomials,

$$h' \;=\; (h_1 + ax^\alpha q_1, \ldots, h_j - \mathrm{in}(h_j) + ax^\alpha q_j, \ldots, h_k + cx^\gamma + ax^\alpha q_k, \ldots, h_m + ax^\alpha q_m) \,,$$

and consider the sum $\sum h'_i g_i$, which is

$$\sum_i h_i g_i + ax^\alpha \sum_i q_i g_i - \mathrm{in}(h_j) g_j + cx^\gamma g_k$$
$$= \; f + ax^\alpha \mathrm{Spol}(g_j, g_k) - ax^\alpha \mathrm{Spol}(g_j, g_k) \;=\; f \,.$$

By the division algorithm, $\mathrm{in}(q_i g_i) \preceq \mathrm{in}(\mathrm{Spol}(g_j, g_k))$, so $\mathrm{in}(ax^\alpha q_i g_i) \prec x^\alpha x^\beta = m(h)$. But then $m(h') \preceq m(h)$. By our assumption, $m(h') = m(h)$. Since $\mathrm{in}(h_j - \mathrm{in}(h_j)) \prec \mathrm{in}(h_j)$, we have $j(h') > j = j(h)$, which contradicts our choice of h. □

Buchberger's algorithm to compute a Gröbner basis begins with a list of polynomials and augments that list by adding reductions of S-polynomials. It halts when the list of polynomials satisfy Buchberger's Criterion.

Algorithm 2.2.6 (Buchberger's Algorithm) Begin with a list of generators $G = (g_1, \ldots, g_m)$ for an ideal. For each $i < j$, let $h_{ij} := \mathrm{Spol}(g_i, g_j) \bmod G$. If each of these is zero, then we have a Gröbner basis, by Buchberger's Criterion. Otherwise append all the non-zero h_{ij} to the list G and repeat this process.

This algorithm terminates after finitely many steps, because the initial terms of polynomials in G after each step generate a strictly larger monomial ideal and Dickson's Lemma implies that any increasing chain of monomial ideals is finite. Since the manipulations in Buchberger's algorithm involve only algebraic operations using the coefficients of the input polynomials, we deduce the following corollary, which is important when studying real varieties. Let $\mathbb{K}$ be any field containing $\mathbb{F}$.

Corollary 2.2.7 *Let $f_1, \ldots, f_m \in \mathbb{F}[x_1, \ldots, x_n]$ be polynomials and $\succ$ a monomial order. Then there is a Gröbner basis $G \subset \mathbb{F}[x_1, \ldots, x_n]$ for the ideal $\langle f_1, \ldots, f_m \rangle$ in $\mathbb{K}[x_1, \ldots, x_n]$*

Example 2.2.8 Consider applying the Buchberger algorithm to $G = (x^2, xy + y^2)$ with any monomial order where $x \succ y$. First

$$\mathrm{Spol}(x^2, xy + y^2) \;=\; y \cdot x^2 - x(xy + y^2) \;=\; -xy^2 \,.$$

Then

$$-xy^2 \bmod (x^2, xy + y^2) \;=\; -xy^2 + y(xy + y^2) \;=\; y^3 \,.$$

Since all S-polynomials of $(x^2, xy + y^2, y^3)$ reduce to zero, this is a Gröbner basis.

Among the polynomials h_{ij} computed at each stage of the Buchberger algorithm are those where one of $\mathrm{in}(g_i)$ or $\mathrm{in}(g_j)$ divides the other. Suppose that $\mathrm{in}(g_i)$ divides $\mathrm{in}(g_j)$ with $i \neq j$. Then $\mathrm{Spol}(g_i, g_j) = g_j - mg_i$, where m is some term. This has strictly smaller initial term than does g_j and so we never use g_j to compute $h_{ij} := \mathrm{Spol}(g_i, g_j) \bmod G$. It follows that $g_j - h_{ij}$ lies in the ideal generated by $G \setminus \{g_j\}$, and so we may replace g_j by h_{ij} in G without changing the ideal generated by G, and only possibly increasing the ideal generated by the initial terms of polynomials in G.

This gives the following elementary improvement to the Buchberger algorithm:

$$\begin{array}{l}\text{In each step, initially compute } h_{ij} \text{ for those } i \neq j \\ \text{where } \mathrm{in}(g_i) \text{ divides } \mathrm{in}(g_j), \text{ and replace } g_j \text{ by } h_{ij}.\end{array} \tag{2.3}$$

In some important cases, this step computes the Gröbner basis. Another improvement, which identifies S-polynomials which reduce to zero and therefore do not need to be computed, is given in the exercises.

There are additional improvements in Buchberger's algorithm (see Ch. 2.9 in [5] for a discussion), and even a series of completely different algorithms due to Jean-Charles Faugère [6] based on linear algebra with vastly improved performance.

A Gröbner basis G is *reduced* if the initial terms of polynomials in G are monomials with coefficient 1 and if for each $g \in G$, no monomial of g is divisible by an initial term of another Gröbner basis element. A reduced Gröbner basis for an ideal is uniquely determined by the monomial order. Reduced Gröbner bases are the multivariate analog of unique monic polynomial generators of ideals of $\mathbb{F}[x]$. Elements f of a reduced Gröbner basis also have a special form,

$$x^\alpha - \sum_{\beta \in B} c_\beta x^\beta \,,$$

where $x^\alpha = \text{in}(f)$ is the initial term and B consists of exponent vectors of standard monomials. This rewrites the nonstandard initial monomial as a linear combination of standard monomials. The reduced Gröbner basis has one generator for every generator of the initial ideal.

Example 2.2.9 Let M be a $m \times n$ matrix, which we consider to be the matrix of coefficients of m linear forms $g_1, \ldots, g_m$ in $\mathbb{F}[x_1, \ldots, x_n]$, and suppose that $x_1 \succ x_2 \succ \cdots \succ x_n$. We can apply (2.3) to two forms g_i and g_j when their initial terms have the same variable. Then the S-polynomial and subsequent reductions are equivalent to the steps in the algorithm of Gaussian elimination applied to the matrix M. If we iterate our applications of (2.3) until the initial terms of the forms g_i have distinct variables, then the forms $g_1, \ldots, g_m$ are a Gröbner basis for the ideal they generate.

If the forms g_i are a reduced Gröbner basis and are sorted in decreasing order according to their initial terms, then the resulting matrix $\overline{M}$ of their coefficients is an *echelon matrix*: The initial non-zero entry in each row is 1 and is the only non-zero entry in its column and these columns increase with row number.

Gaussian elimination produces the same echelon matrix from M. In this way, we see that the Buchberger algorithm is a generalization of Gaussian elimination to non-linear polynomials.

Exercises for Section 2

1. Describe how Buchberger's algorithm behaves when it computes a Gröbner basis from a list of monomials. What if we use the elementary improvement (2.3)?

2. Use Buchberger's algorithm to compute the reduced Gröbner basis of $\langle y^2 - xz + yz,\ x^2y - xz^2 + y^2z \rangle$ in the degree reverse lexicographic order where $x \succ y \succ z$.

3. Let $f, g \in \mathbb{F}[x_1, \ldots, x_m]$ be such that $\text{in}(f)$ and $\text{in}(g)$ are relatively prime and the leading coefficients of f and g are 1. Show that

$$S(f,g) \;=\; -(g - \text{in}(g))f + (f - \text{in}(f))g\,.$$

   Deduce that the leading monomial of $S(f,g)$ is a multiple of either the leading monomial of f or of g in this case.

4. The following problem shows that every ideal has a finite generating set that is a Gröbner basis with respect to all term orderings. Such a generating set is called a *universal Gröbner basis*. Let $I \subset \mathbb{F}[x_1, \ldots, x_n]$ be an ideal.

   (a) Show that there are only finitely many initial ideals of I. More precisely show that

$$\{\text{in}_\succ(I) \mid \succ \text{ is a term order on } \mathbb{F}[x_1, \ldots, x_n]\}$$

   is a finite set.

(b) Show that every ideal $I \subset \mathbb{F}[x_1, \ldots, x_n]$ has a set of generators that is a Gröbner basis for every term order.

5. Let U be a universal Gröbner basis for an ideal I in $\mathbb{F}[x_1, \ldots, x_n]$. Show that for every subset $Y \subset \{x_1, \ldots, x_n\}$ the *elimination ideal* $I \cap \mathbb{F}[Y]$ is generated by $U \cap \mathbb{F}[Y]$.

6. Let I be a ideal generated by homogeneous linear polynomials. We call a nonzero linear form f in I a *circuit* of I if f has minimal support (with respect to inclusion) among all polynomials in I. Prove that the set of all circuits of I is a universal Gröbner basis of I.

7. (a) Prove that the ideal $\langle x, y \rangle \subset \mathbb{Q}[x, y]$ is not a principal ideal.
   (b) Is $\langle x^2 + y, x + y \rangle$ already a Gröbner basis with respect to some term ordering?
   (c) Use Buchberger's algorithm to compute a Gröbner basis of the ideal $I = \langle y - z^2, z - x^3 \rangle \in \mathbb{Q}[x, y, z]$ with lexicographic and the degree reverse lexicographic monomial orders.

8. This exercise illustrates an algorithm to compute the saturation of ideals. Let $I \subset \mathbb{F}[x_1, \ldots, x_n]$ be an ideal, and fix $f \in \mathbb{F}[x_1, \ldots, x_n]$. Then the *saturation* of I with respect to f is the set

$$(I : f^\infty) \;=\; \{g \in \mathbb{F}[x_1, \ldots, x_n] \mid f^m g \in I \text{ for some } m > 0\}\,.$$

   (a) Prove that $(I : f^\infty)$ is an ideal.
   (b) Prove that we have an ascending chain of ideals

$$(I : f) \;\subset\; (I : f^2) \;\subset\; (I : f^3) \;\subset \cdots$$

   (c) Prove that there exists a nonnegative integer N such that $(I : f^\infty) = (I : f^N)$.
   (d) Prove that $(I : f^\infty) = (I : f^m)$ if and only if $(I : f^m) = (I : f^{m+1})$.

When the ideal I is homogeneous and $f = x_n$ then one can use the following strategy to compute the saturation. Fix the degree reverse lexicographic order $\succ_{\text{drl}}$ where $x_1 \succ_{\text{drl}} x_2 \succ_{\text{drl}} \cdots \succ_{\text{drl}} x_n$ and let G be a reduced Gröbner basis of a homogeneous ideal $I \subset \mathbb{F}[x_1, \ldots, x_n]$.

   (e) Show that the set

$$G' \;=\; \{f \in G \mid x_n \text{ does not divide } f\} \bigcup \{f/x_n \mid f \in G \text{ and } x_n \text{ divides } f\}$$

   is a Gröbner basis of $(I : x_n)$.
   (f) Show that a Gröbner basis of $(I : x_n^\infty)$ is obtained by dividing each element $f \in G$ by the highest power of x_n that divides f.

9. Suppose that $\prec$ is the lexicographic order with $x \prec y \prec z$.

   (a) Apply Buchberger's algorithm to the ideal $\langle x + y, xy \rangle$.

   (b) Apply Buchberger's algorithm to the ideal $\langle x + y + z, xy + xz + yz, xyz \rangle$.

   (c) Define the *elementary symmetric polynomials* $e_i(x_1, \ldots, x_n)$ by

$$\sum_{i=0}^{n} t^{n-i} e_i(x_1, \ldots, x_n) \;=\; \prod_{i=1}^{n} (t + x_i)\,,$$

   that is, $e_0 = 1$ and if $i > 0$, then

$$e_i(x_1, \ldots, x_n) \;:=\; e_i(x_1, \ldots, x_{n-1}) \;+\; x_n e_{i-1}(x_1, \ldots, x_{n-1})\,.$$

   Alternatively, $e_i(x_1, \ldots, x_n)$ is also the sum of all square-free monomials of total degree i in $x_1, \ldots, x_n$.
   The *symmetric ideal* is $\langle e_i(x_1, \ldots, x_n) \mid 1 \leq i \leq n \rangle$. Describe its Gröbner basis, and the set of standard monomials with respect to lexicographic order when $x_1 \prec x_2 \prec \cdots \prec x_n$.
   What about degree reverse lexicographic order?

2.3 Resultants and Bézout's Theorem

Algorithms based on Gröbner bases are universal in that their input may be any list of polynomials. This comes at a price as Gröbner basis algorithms may have poor performance and the output is quite sensitive to the input. An alternative foundation for some algorithms is provided by resultants. These are are special polynomials having determinantal formulas which were introduced in the 19th century. A drawback is that they are not universal—different inputs require different algorithms, and for many inputs, there are no formulas for resultants.

The key algorithmic step in the Euclidean algorithm for the greatest common divisor (gcd) of two univariate polynomials f and g in $\mathbb{F}[x]$ with $n = \deg(g) \geq \deg(f) = m$,

$$\begin{aligned} f &= f_0 x^m + f_1 x^{m-1} + \cdots + f_{m-1} x + f_m \\ g &= g_0 x^n + g_1 x^{n-1} + \cdots + g_{n-1} x + g_n \,, \end{aligned} \tag{2.4}$$

is to replace g by

$$g - \frac{g_0}{f_0} x^{n-m} \cdot f \,,$$

which has degree at most $n-1$. In some applications (for example, when $\mathbb{F}$ is a function field), we will want to avoid division. Resultants give a way to detect common factors without using division. We will use them for much more than this.

2.3.1 Sylvester Resultant

Let S_ℓ or $S_\ell(x)$ be the set of polynomials in $\mathbb{F}[x]$ of degree at most ℓ. This is a vector space over $\mathbb{F}$ of dimension $\ell + 1$ with a canonical ordered basis of monomials $x^\ell, \ldots, x, 1$. Given f and g as above, consider the linear map

$$\begin{aligned} \varphi_{f,g} \,:\, S_{n-1} \times S_{m-1} &\longrightarrow S_{m+n-1} \\ (h(x), k(x)) &\longmapsto f \cdot h + g \cdot k \,. \end{aligned}$$

The domain and range of $\varphi_{f,g}$ each have dimension $m + n$.

Lemma 2.3.1 *The polynomials f and g have a nonconstant common divisor if and only if* $\ker \varphi_{f,g} \neq \{(0,0)\}$.

Proof. Suppose first that f and g have a nonconstant common divisor, p. Then there are polynomials h and k with $f = pk$ and $g = ph$. As p is nonconstant, $\deg k < \deg f = n$ and $\deg h < \deg g = m$ so that $(h, -k) \in S_{n-1} \times S_{m-1}$. Since

$$fh - gk \;=\; pkh - phk \;=\; 0 \,,$$

we see that $(h, -k)$ is a nonzero element of the kernel of $\varphi_{f,g}$.

Suppose that f and g are relatively prime and let $(h,k) \in \ker \varphi_{f,g}$. Since $\langle f, g\rangle = \mathbb{F}[x]$, there exist polynomials p and q with $1 = gp + fq$. Using $0 = fh + gk$ we obtain

$$k \;=\; k\cdot 1 \;=\; k(gp+fq) \;=\; gkp+fkq \;=\; -fhp+fkq \;=\; f(kq-hp)\,.$$

This implies that $k = 0$ for otherwise $m-1 \geq \deg k > \deg f = m$, which is a contradiction. We similarly see that $h = 0$, and so $\ker \varphi_{f,g} = \{(0,0)\}$. □

The matrix of the linear map $\varphi_{f,g}$ in the ordered bases of monomials for $S_{m-1}\times S_{n-1}$ and S_{m+n-1} is called the *Sylvester matrix*. When f and g have the form (2.4), it is

$$\mathrm{Syl}(f,g;x) \;=\; \mathrm{Syl}(f,g) \;:=\; \left(\begin{array}{ccccc|ccc} f_0 & & & & & g_0 & & 0 \\ f_1 & f_0 & & 0 & & g_1 & \ddots & \\ \vdots & \vdots & \ddots & & & \vdots & & g_0 \\ f_m & \vdots & & \ddots & & \vdots & & \vdots \\ & f_m & & & f_0 & g_{n-1} & & \vdots \\ & & \ddots & & \vdots & g_n & & \vdots \\ & 0 & & \ddots & \vdots & & \ddots & \vdots \\ & & & & f_m & 0 & & g_n \end{array}\right). \tag{2.5}$$

Note that the sequence $f_0,\ldots,f_0,g_n,\ldots,g_n$ lies along the main diagonal and the left side of the matrix has n columns while the right side has m columns.

The *(Sylvester) resultant* $\mathrm{Res}(f,g)$ is the determinant of the Sylvester matrix. To emphasize that the Sylvester matrix represents the map $\varphi_{f,g}$ in the basis of monomials in x, we also write $\mathrm{Res}(f,g;x)$ for $\mathrm{Res}(f,g)$. We summarize some properties of resultants, which follow from its formula as the determinant of the Sylvester matrix (2.5) and from Lemma 2.3.1.

Theorem 2.3.2 *The resultant of two nonconstant polynomials $f, g \in \mathbb{F}[x]$ is an irreducible integer polynomial in the coefficients of f and g. The resultant vanishes if and only if f and g have a nonconstant common factor.*

We only need to prove that the resultant is irreducible. The path we choose to this will give another expression for the resultant as well as a geometric interpretation of the resultant.

Lemma 2.3.3 *Suppose that $\mathbb{F}$ contains all the roots of the polynomials f and g so that we have*

$$f(x) \;=\; f_0\prod_{i=1}^{m}(x-\alpha_i) \qquad \text{and} \qquad g(x) \;=\; g_0\prod_{i=1}^{n}(x-\beta_i)\,,$$

where $\alpha_1,\ldots,\alpha_m$ are the roots of f and $\beta_1,\ldots,\beta_n$ are the roots of g. Then

$$\mathrm{Res}(f,g;x) \;=\; (-1)^{mn} f_0^n g_0^m \prod_{i=1}^{m}\prod_{j=1}^{n}(\alpha_i-\beta_j)\,. \tag{2.6}$$

A corollary of this lemma is the Poisson formula,

$$\mathrm{Res}(f,g;x) \;=\; (-1)^{mn} f_0^m \prod_{i=1}^{m} g(\alpha_i) \;=\; g_0^n \prod_{i=1}^{n} f(\beta_i)\,.$$

Proof. Consider these formulas as expressions in $\mathbb{Z}[f_0, g_0, \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_n]$. Recall that the coefficients of f and g are essentially the elementary symmetric polynomials in their roots,

$$f_i \;=\; (-1)^i f_0 e_i(\alpha_1, \ldots, \alpha_m) \qquad \text{and} \qquad g_i \;=\; (-1)^i g_0 e_i(\beta_1, \ldots, \beta_n)\,.$$

We claim that both sides of (2.6) are homogeneous polynomials of degree mn in the variables $\alpha_1, \ldots, \beta_n$. This is straightforward for the right hand side. To see this for the resultant, we extend our notation, setting $f_i := 0$ when $i < 0$ or $i > m$ and $g_i := 0$ when $i < 0$ or $i > n$. Then the entry in row i and column j of the Sylvester matrix is

$$\mathrm{Syl}(f,g;x)_{i,j} \;=\; \begin{cases} f_{i-j} & \text{if } j \le n\,, \\ g_{n+i-j} & \text{if } n < j \le m+n\,. \end{cases}$$

The determinant is a signed sum over permutations w of $\{1, \ldots, m+n\}$ of terms

$$\prod_{j=1}^{n} f_{w(j)-j} \;\cdot\; \prod_{j=n+1}^{m+n} g_{n+w(j)-j}\,.$$

Since f_i and g_i are each homogeneous of degree i in the variables $\alpha_1, \ldots, \beta_n$, this term is homogeneous of degree

$$\sum_{j=1}^{m} w(j)-j \quad + \quad \sum_{j=n+1}^{m+n} n+w(j)-j \;=\; mn + \sum_{j=1}^{m+n} w(j)-j \;=\; mn\,.$$

Both sides of (2.6) vanish exactly when some $\alpha_i = \beta_j$. Since they have the same degree, they are proportional. We compute this constant of proportionality. The term in $\mathrm{Res}(f,g)$ which is the product of diagonal entries of the Sylvester matrix is

$$f_0^n g_n^m \;=\; f_0^n g_0^m e_n(\beta_1, \ldots, \beta_n)^m \;=\; f_0^n g_0^m \beta_1^m \cdots \beta_n^m\,.$$

This is the only term of $\mathrm{Res}(f,g)$ involving the monomial $\beta_1^m \cdots \beta_n^m$. The corresponding term on the right hand side of (2.6) is

$$(-1)^{mn} f_0^n g_0^m (-\beta_1)^m \cdots (-\beta_n)^m \;=\; f_0^n g_0^m \beta_1^m \cdots \beta_n^m\,,$$

which completes the proof. □

Suppose that $\mathbb{F}$ is algebraically closed and consider the variety of all triples consisting of a pair of polynomials with a common root, together with a common root,

$$\Sigma \;:=\; \{(f,g,a) \in S_m \times S_n \times \mathbb{A}^1 \mid f(a) = g(a) = 0\}\,.$$

This has projections $p\colon \Sigma \to S_m \times S_n$ and $\pi\colon \Sigma \to \mathbb{A}^1$. The image $p(\Sigma)$ is the set of pairs of polynomials having a common root, which is the variety $\mathcal{V}(\mathrm{Res})$ of the resultant polynomial, $\mathrm{Res} \in \mathbb{Z}[f_0,\ldots,f_m,g_0,\ldots,g_n]$.

The fiber of π over a point $a \in \mathbb{A}^1$ consists all pairs of polynomials f,g with $f(a) = g(a) = 0$. Since each equation is linear in the coefficients of the polynomials f and g, this fiber is isomorphic to $\mathbb{A}^m \times \mathbb{A}^n$. Since $\pi\colon \Sigma \to \mathbb{A}^1$ has irreducible image ($\mathbb{A}^1$) and irreducible fibers, we see that Σ is irreducible, and has dimension $1+m+n$.[†]

This implies that $p(\Sigma)$ is irreducible. Furthermore, the fiber $p^{-1}(f,g)$ is the set of common roots of f and g. This is a finite set when $f,g \neq (0,0)$. Thus $p(\Sigma)$ has dimension $1+m+n$, and is thus an irreducible hypersurface in $S_m \times S_n$. Let F be a polynomial generating the ideal $\mathcal{I}(p(\Sigma))$, which is necessarily irreducible. As $\mathcal{V}(\mathrm{Res}) = p(\Sigma)$, we must have $\mathrm{Res} = F^N$ for some positive integer N. The formula (2.6) shows that $N = 1$ as the resultant polynomial is square-free.

Corollary 2.3.4 *The resultant polynomial is irreducible. It is the unique (up to sign) irreducible integer polynomial in the coefficients of f and g that vanishes on the set of pairs of polynomials (f,g) which have a common root.*

We only need to show that the greatest common divisor of the coefficients of the integer polynomial Res is 1. But this is clear as Res contains the term $f_0^n g_n^m$, as we showed in the proof of Lemma 2.3.3.

When both f and g have the same degree n, there is an alternative determinantal formula for their resultant. The *Bezoutian polynomial* of f and g is the bivariate polynomial

$$\Delta_{f,g}(y,z) \;:=\; \frac{f(y)g(z) - f(z)g(y)}{y-z} \;=\; \sum_{i,j=0}^{n-1} b_{i,j}\, y^i z^j\;.$$

The $n \times n$ matrix $\mathrm{Bez}(f,g)$ whose entries are the coefficients $(b_{i,j})$ of the Bezoutian polynomial is called the *Bezoutian matrix* of f and g. Note that each entry of the Bezoutian matrix $\mathrm{Bez}(f,g)$ is a linear combination of the brackets $[ij] := f_i g_j - f_j g_i$. For example, when $n = 2$ and $n = 3$, the Bezoutian matrices are

$$\begin{pmatrix} [02] & [12] \\ [01] & [02] \end{pmatrix} \qquad \begin{pmatrix} [03] & [13] & [23] \\ [02] & [03]+[12] & [13] \\ [01] & [02] & [03] \end{pmatrix}\;.$$

Theorem 2.3.5 *When f and g both have degree n, $\mathrm{Res}(f,g) = (-1)^{\binom{n}{2}} \det(\mathrm{Bez}(f,g))$.*

[†]This will need to cite results from Chapter 1 on dimension and irreducibility

Proof. Suppose that $\mathbb{F}$ is algebraically closed. Let B be the determinant of the Bezoutian matrix and Res the resultant of the polynomials f and g, both of which lie in the ring $\mathbb{F}[f_0, \ldots, f_n, g_0, \ldots, g_n]$. Then B is a homogeneous polynomial of degree $2n$, as is the resultant. Suppose that f and g are polynomials having a common root, $a \in \mathbb{F}$ with $f(a) = g(a) = 0$. Then the Bezoutian polynomial $\Delta_{f,g}(y, z)$ vanishes when $z = a$,

$$\Delta_{f,g}(y, a) \;=\; \frac{f(y)g(a) - f(a)g(y)}{y - a} \;=\; 0\,.$$

Thus

$$0 \;=\; \sum_{i,j=0}^{n-1} b_{i,j} y^i a^j \;=\; \sum_{i=0}^{n-1} \Big(\sum_{j=0}^{n-1} b_{i,j} a^j \Big)\, y^i\,.$$

Since every coefficient of this polynomial in y must vanish, the vector $(1, a, a^2, \ldots, a^{d-1})^T$ lies in the kernel of the Bezoutian matrix, and so the determinant $B(f, g)$ of the Bezoutian matrix vanishes.

Since the resultant generates the ideal of the pairs (f, g) of polynomial that are not relatively prime, Res divides B. As they have the same degree B is a constant multiple of Res. In Exercise 4 you are asked to show this constant is $(-1)^{\binom{n}{2}}$. □

Example 2.3.6 We give an application of resultants. A polynomial $f \in \mathbb{F}[x]$ of degree n has fewer than n distinct roots in the algebraic closure of $\mathbb{F}$ when it has a factor in $\mathbb{F}[x]$ of multiplicity greater than 1, and in that case f and its derivative f' have a factor in common. The *discriminant* of f is a polynomial in the coefficients of f which vanishes precisely when f has a repeated factor. It is defined to be

$$\mathrm{disc}(f) \;:=\; \frac{(-1)^{\binom{n}{2}}}{f_0} \mathrm{Res}(f, f')\,.$$

The discriminant is a polynomial of degree $2n - 2$ in the coefficients $f_1, \ldots, f_n$.

2.3.2 Resultants and Elimination

Resultants do much more than detect the existence of common factors in two polynomials. One of their most important uses is to eliminate variables from multivariate equations. The first step towards this is another interesting formula involving the Sylvester resultant. Not only is it a polynomial in the coefficients, but it has a canonical expression as a polynomial linear combination of f and g.

Lemma 2.3.7 *Given univariate polynomials $f, g \in \mathbb{F}[x]$, there are polynomials $h, k \in \mathbb{F}[x]$ whose coefficients are integer polynomials in the coefficients of f and g such that*

$$f(x)h(x) + g(x)k(x) \;=\; \mathrm{Res}(f, g)\,. \tag{2.7}$$

Proof. Set $\mathbb{F} := \mathbb{Q}(f_0, \ldots, f_m, g_0, \ldots, g_n)$, the field of rational functions (quotients of integer polynomials) in the indeterminates $f_0, \ldots, f_m, g_0, \ldots, g_n$ and let $f, g \in \mathbb{F}[x]$ be univariate polynomials as in (2.4). Then $\gcd(f, g) = 1$ and so the map $\varphi_{f,g}$ is invertible.

Set $(h, k) := \varphi_{f,g}^{-1}(\mathrm{Res}(f, g))$ so that

$$f(x)h(x) \;+\; g(x)k(x) \;=\; \mathrm{Res}(f, g)\,,$$

with $h, k \in \mathbb{F}[x]$ where $h \in S_{n-1}(x)$ and $k \in S_{m-1}(x)$.

Recall the adjoint formula for the inverse of a $n \times n$ matrix A,

$$\det(A) \cdot A^{-1} \;=\; \mathrm{ad}(A)\,. \tag{2.8}$$

Here $\mathrm{ad}(A)$ is the *adjoint* of the matrix A. Its (i, j)-entry is $(-1)^{i+j} \cdot \det A_{i,j}$, where $A_{i,j}$ is the $(n-1) \times (n-1)$ matrix obtained from A by deleting its ith column and jth row.

Since $\det \varphi_{f,g} = \mathrm{Res}(f, g) \in \mathbb{F}$, we have

$$\varphi_{f,g}^{-1}(\mathrm{Res}(f, g)) \;=\; \det \varphi_{f,g} \cdot \varphi_{f,g}^{-1}(1) \;=\; \mathrm{ad}(\mathrm{Syl}(f, g))(1)\,.$$

In the monomial basis for S_{m+n-1} the polynomial 1 corresponds to the vector $(0, \ldots, 0, 1)$. Thus, the coefficients of $\varphi_{f,g}^{-1}(\mathrm{Res}(f, g))$ are given by the entries of the last column of $\mathrm{ad}(\mathrm{Syl}(f, g))$, which are $\pm$ the minors of the Sylvester matrix $\mathrm{Syl}(f, g)$ with its last row removed. In particular, these are polynomials in the indeterminates $f_0, \ldots, g_n$ having integer coefficients. □

This proof shows that $h, k \in \mathbb{Z}[f_0, \ldots, f_m, g_0, \ldots, g_m][x]$ and that (2.7) holds as an expression in this polynomial ring with $m+n+3$ variables. It also shows that if $f, g \in \mathbb{F}[x_1, \ldots, x_n]$ are multivariate polynomials, and we hide all of the variables except x_1 in their coefficients, considering them as polynomials in x_n with coefficients in $\mathbb{F}(x_2, \ldots, x_n)$, then the resultant lies in both the ideal generated by f and g, and in the subring $\mathbb{F}[x_2, \ldots, x_n]$. We examine the geometry of this elimination of variables.

Suppose that $1 \leq i < n$ and let $\pi\colon \mathbb{A}^n \to \mathbb{A}^{n-i}$ be the coordinate projection

$$\pi \;:\; (a_1, \ldots, a_n) \;\longmapsto\; (a_{i+1}, \ldots, a_n)\,.$$

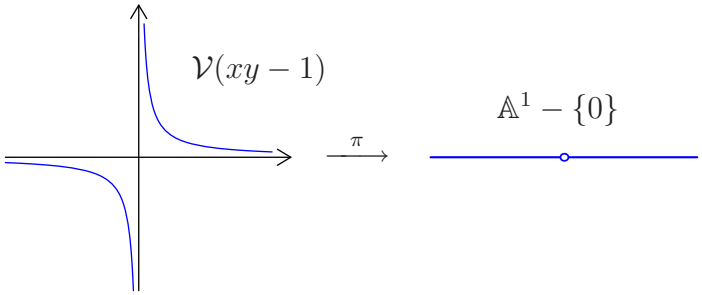
Lemma 2.3.8 *Let $I \subset \mathbb{F}[x_1, \ldots, x_n]$ be an ideal. Then $\pi(\mathcal{V}(I)) \subset \mathcal{V}(I \cap \mathbb{F}[x_{i+1}, \ldots, x_n])$.*

Proof. Suppose that $a = (a_1, \ldots, a_n) \in \mathcal{V}(I)$. If $f \in I \cap \mathbb{F}[x_{i+1}, \ldots, x_n]$, then

$$0 \;=\; f(a) \;=\; f(a_{i+1}, \ldots, a_n) \;=\; f(\pi(a))\,.$$

Note that we may view f as a polynomial in either $x_1, \ldots, x_n$ or in $x_{i+1}, \ldots, x_n$. The statement of the lemma follows. □

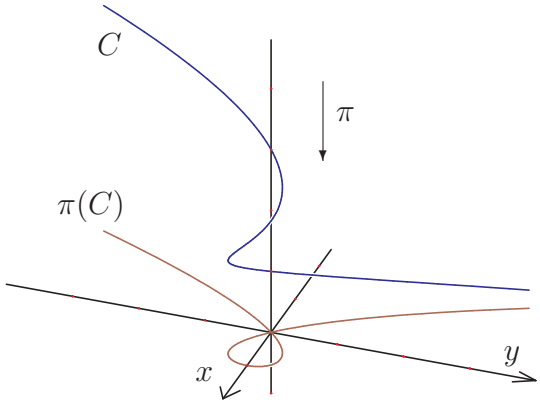
The ideal $I \cap \mathbb{F}[x_{i+1}, \ldots, x_n]$ is called an *elimination ideal* as the variables $x_1, \ldots, x_i$ have been eliminated from the ideal I. By Lemma 2.3.8, elimination is the algebraic counterpart to projection, but the correspondence is not exact. For example, the inclusion $\pi(\mathcal{V}(I)) \subset \mathcal{V}(I \cap \mathbb{F}[x_{i+1}, \ldots, x_n])$ may be strict. Let $\pi\colon \mathbb{A}^2 \to \mathbb{A}^1$ be the map which forgets the first coordinate. Then $\pi(\mathcal{V}(xy-1)) = \mathbb{A}^1 - \{0\} \subsetneq \mathbb{A}^1 = V(0)$ and $\{0\} = \langle xy-1 \rangle \cap F[y]$.



Elimination of variables enables us to solve the implicitization problem for plane curves. For example, consider the parametric plane curve

$$x \;=\; t^2 - 1, \qquad y \;=\; t^3 - t\,. \tag{2.9}$$

This is the image of the space curve $C := \mathcal{V}(t^2 - 1 - x, t^3 - t - y)$ under the projection $(t, x, y) \mapsto (x, y)$. We display this with the t-axis vertical and the xy-plane at $t = -2$.



By lemma 2.3.8, the equation for the plane curve is $\langle t^2 - x - x, t^3 - t - y \rangle \cap \mathbb{F}[x, y]$. If we set

$$f(t) \;:=\; t^2 - 1 - x \qquad \text{and} \qquad g(t) \;:=\; t^3 - t - y\,,$$

then the Sylvester resultant is

$$\det \left[\begin{array}{ccc|cc} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ -x-1 & 0 & 1 & -1 & 0 \\ 0 & -x-1 & 0 & -y & -1 \\ 0 & 0 & -x-1 & 0 & -y \end{array}\right] \;=\; y^2 + x^2 - x^3\,,$$

which is the implicit equation of the parameterized $\pi(C)$ cubic (2.9).

2.3.3 Resultants and Bézout's Theorem

We use resultants to study the variety $\mathcal{V}(f,g) \subset \mathbb{A}^2$ for $f, g \in \mathbb{F}[x,y]$. A by-product will be a form of Bézout's Theorem bounding the number of points in the variety $\mathcal{V}(f,g)$.

Suppose now that we have two variables, x and y. The ring $\mathbb{F}[x,y]$ of bivariate polynomials is a subring of the ring $\mathbb{F}(y)[x]$ of polynomials in x whose coefficients are rational functions in y. Suppose that $f, g \in \mathbb{F}[x,y]$. If we consider f and g as elements of $\mathbb{F}(y)[x]$, then the resultant $\mathrm{Res}(f,g;x)$ is the determinant of their Sylvester matrix expressed in the basis of monomials in x. Theorem 2.3.2 implies that $\mathrm{Res}(f,g;x)$ is a univariate polynomial in y which vanishes if and only if f and g have a common factor in $\mathbb{F}(y)[x]$. In fact it vanishes if and only if $f(x,y)$ and $g(x,y)$ have a common factor in $\mathbb{F}[x,y]$ with positive degree in y, by the following version of Gauss's lemma for $\mathbb{F}[x,y]$.

Lemma 2.3.9 *Polynomials f and g in $\mathbb{F}[x,y]$ have a common factor of positive degree in x if and only if they have a common factor in $\mathbb{F}(y)[x]$.*

Proof. The forward direction is clear. For the reverse, suppose that

$$f \;=\; h \cdot \overline{f} \qquad \text{and} \qquad g \;=\; h \cdot \overline{g} \tag{2.10}$$

is a factorization in $\mathbb{F}(y)[x]$ where h has positive degree in x.

There is a polynomial $d \in \mathbb{F}[y]$ which is divisible by every denominator of a coefficient of h, $\overline{f}$, and $\overline{g}$. Multiplying the expressions (2.10) by d^2 gives

$$d^2 f \;=\; (dh) \cdot (d\overline{f}) \qquad \text{and} \qquad d^2 g \;=\; (dh) \cdot (d\overline{g})\,,$$

where dh, $d\overline{f}$, and $d\overline{g}$ are polynomials in $\mathbb{F}[x,y]$. Let $k(x,y)$ be an irreducible factor of dh having positive degree in x. Then k divides both $d^2 f$ and $d^2 g$. However, k cannot divide d as $d \in \mathbb{F}[y]$ and k has positive degree in x. Therefore $k(x,y)$ is the desired common polynomial factor of f and g. □

Let $\pi\colon \mathbb{A}^2 \to \mathbb{A}^1$ be the projection which forgets the first coordinate, $\pi(x,y) = y$. Set $I := \langle f, g\rangle \cap \mathbb{F}[y]$. By Lemma 2.3.7, the resultant $\mathrm{Res}(f,g;x)$ lies in I. Combining this with Lemma 2.3.8 gives the chain of inclusions

$$\pi(\mathcal{V}(f,g)) \;\subset\; \mathcal{V}(I) \;\subset\; \mathcal{V}(\mathrm{Res}(f,g;x))\,.$$

We now suppose that $\mathbb{F}$ is algebraically closed. Let $f, g \in \mathbb{F}[x,y]$ and write

$$\begin{aligned} f &= f_0(y)x^m + f_1(y)x^{m-1} + \cdots + f_{m-1}(y)x + f_m(y) \\ g &= g_0(y)x^n + g_1(y)x^{n-1} + \cdots + g_{n-1}(y)x + g_n(y)\,, \end{aligned}$$

where neither $f_0(y)$ nor $g_0(y)$ is the zero polynomial.

Theorem 2.3.10 (Extension Theorem) *If $b \in \mathcal{V}(I) - \mathcal{V}(f_0(y), g_0(y))$, then there is some $a \in \mathbb{F}$ with $(a,b) \in \mathcal{V}(f,g)$.*

This establishes the chain of inclusions of subvarieties of $\mathbb{A}^1$

$$\mathcal{V}(I) - \mathcal{V}(f_0, g_0) \ \subset \ \pi(\mathcal{V}(f,g)) \ \subset \ \mathcal{V}(I) \subset \ \mathcal{V}(\mathrm{Res}(f,g;x)) \, .$$

Proof. Let $b \in \mathcal{V}(I) - \mathcal{V}(f_0, g_0)$. Suppose first that $f_0(b) \cdot g_0(b) \neq 0$. Then $f(x,b)$ and $g(x,b)$ are polynomials in x of degrees m and n, respectively. It follows that the Sylvester matrix $\mathrm{Syl}(f(x,b), g(x,b))$ has the same format (2.5) as the Sylvester matrix $\mathrm{Syl}(f,g;x)$, and it is in fact obtained from $\mathrm{Syl}(f,g;x)$ by substituting $y = b$.

This implies that $\mathrm{Res}(f(x,b), g(x,b))$ is the evaluation of the resultant $\mathrm{Res}(f,g;x)$ at $y = b$. Since $\mathrm{Res}(f,g;x) \in I$ and $b \in \mathcal{V}(I)$, this evaluation is 0. By Theorem 2.3.2, $f(x,b)$ and $g(x,b)$ have a nonconstant common factor. As $\mathbb{F}$ is algebraically closed, they have a common root, say a. But then $(a,b) \in \mathcal{V}(f,g)$, and so $b \in \pi(\mathcal{V}(f,g))$.

Now suppose that $f_0(b) \neq 0$ but $g_0(b) = 0$. Since $\langle f, g \rangle = \langle f, g + x^l f \rangle$, if we replace g by $g + x^l f$ where $l + m > n$, then we are in the previous case. □

Observe that if f_0 and g_0 are constants, then we showed that $\mathcal{V}(I) = \mathcal{V}(\mathrm{Res}(f,g;x)$. We record this fact.

Corollary 2.3.11 *If the coefficients of the highest powers of x in f and g do not involve y, then $\mathcal{V}(I) = \mathcal{V}(\mathrm{Res}(f,g;x))$.*

Lemma 2.3.12 *Suppose that $\mathbb{F}$ is algebraically closed. The system of bivariate polynomials*

$$f(x,y) \ = \ g(x,y) \ = \ 0$$

has finitely many solutions in $\mathbb{A}^2$ if and only if f and g have no common nonconstant factor.

Proof. We instead show that $\mathcal{V}(f,g)$ is infinite if and only if f and g do have a common nonconstant factor. If f and g have a common nonconstant factor $h(x,y)$ then their common zeroes $\mathcal{V}(f,g)$ include $\mathcal{V}(h)$ which is infinite as h is nonconstant and $\mathbb{F}$ is algebraically closed.

Now suppose that $\mathcal{V}(f,g)$ is infinite. Then the projection of $\mathcal{V}(f,g)$ to at least one of the two axes is infinite. Suppose that the projection π onto the y-axi is infinites. Set $I := \langle f, g \rangle \cap \mathbb{F}[y]$, the elimination ideal. By the Extension Theorem 2.3.10, we have $\pi(\mathcal{V}(f,g)) \subset \mathcal{V}(I) \subset \mathcal{V}(\mathrm{Res}(f,g;x))$. Since $\pi(\mathcal{V}(f,g))$ is infinite, $\mathcal{V}(\mathrm{Res}(f,g;x)) = \mathbb{A}^1$, which implies that $\mathrm{Res}(f,g;x)$ it is the zero polynomial. By Theorem 2.3.2 and Lemma 2.3.9, f and g have a common nonconstant factor. □

Let $f, g \in \mathbb{F}[x,y]$ and suppose that neither $\mathrm{Res}(f,g;x)$ nor $\mathrm{Res}(f,g;y)$ vanishes so that f and g have no nonconstant common factor. Then $\mathcal{V}(f,g)$ consists of finitely many points. The Extension Theorem gives the following algorithm to compute $\mathcal{V}(f,g)$.

Algorithm 2.3.13 (Elimination Algorithm)
INPUT: Polynomials $f, g \in \mathbb{F}[x, y]$.
OUTPUT: $\mathcal{V}(f, g)$.

First, compute the resultant $\mathrm{Res}(f, g; x)$, which is not the zero polynomial. Then, for every root b of $\mathrm{Res}(f, g; x)$, find all common roots a of $f(x, b)$ and $g(x, b)$. The finitely many pairs (a, b) computed are the points of $\mathcal{V}(f, g)$.

The Elimination Algorithm reduces the problem of solving a bivariate system

$$f(x, y) \;=\; g(x, y) \;=\; 0\,, \tag{2.11}$$

to that of finding the roots of univariate polynomials.

Often we only want to count the number of solutions to a system (2.11), or give a realistic bound for this number which is attained when f and g are generic polynomials. The most basic of such bounds was given by Etienne Bézout in his 1779 treatise *Théorie Générale des Équations Algébriques* [1, 2]. Our first step toward establishing Bézout's Theorem is an exercise in algebra and some book-keeping. The monomials in a polynomial of degree n in the variables x, y are indexed by the set

$$n\Delta \;:=\; \{(i, j) \in \mathbb{N}^2 \mid i + j \leq n\}\,.$$

Let $F := \{f_{i,j} \mid (i, j) \in m\Delta\}$ and $G := \{g_{i,j} \mid (i, j) \in n\Delta\}$ be indeterminates and consider generic polynomials f and g of respective degrees m and n in $\mathbb{F}[F, G][x, y]$,

$$f(x, y) \;:=\; \sum_{(i,j)\in m\Delta} f_{i,j} x^i y^j \qquad \text{and} \qquad g(x, y) \;:=\; \sum_{(i,j)\in n\Delta} g_{i,j} x^i y^j\,.$$

Lemma 2.3.14 *The generic resultant* $\mathrm{Res}(f, g; x)$ *is a polynomial in* y *of degree* mn.

Proof. Write

$$f \;:=\; \sum_{j=0}^{m} f_j(y) x^{m-j} \qquad \text{and} \qquad g \;:=\; \sum_{j=0}^{n} g_j(y) x^{n-j}\,,$$

where the coefficients are univariate polynomials in x

$$f_j(y) \;:=\; \sum_{i=0}^{j} f_{i,j} y^i \qquad \text{and} \qquad g_j(y) \;:=\; \sum_{i=0}^{j} g_{i,j} y^i\,.$$

Then the Sylvester matrix $\mathrm{Syl}(f, g; x)$ has the form

$$\mathrm{Syl}(f, g; x) \;:=\; \left(\begin{array}{cccc|cccc}
f_0(y) & & 0 & & g_0(y) & & & 0 \\
\vdots & \ddots & & & \vdots & \ddots & & \\
\vdots & & \ddots & & \vdots & & & g_0(y) \\
f_m(y) & & & f_0(y) & g_{n-1}(y) & & & \vdots \\
 & \ddots & & \vdots & g_n(y) & & & \vdots \\
 & & \ddots & \vdots & & \ddots & & \vdots \\
0 & & & f_m(y) & 0 & & & g_n(y)
\end{array}\right)\,,$$

and so the resultant $\mathrm{Res}(f, g; x) = \det(\mathrm{Syl}(f, g; x))$ is a univariate polynomial in y.

As in the proof of Lemma 2.3.3, if we set $f_j := 0$ when $j < 0$ or $j > m$ and $g_j := 0$ when $j < 0$ or $j > n$, then the entry in row i and column j of the Sylvester matrix is

$$\mathrm{Syl}(f, g; x)_{i,j} \;=\; \begin{cases} f_{i-j}(y) & \text{if } j \le n \\ g_{n+i-j}(y) & \text{if } n < j \le m+n \end{cases}$$

The determinant is a signed sum over permutations w of $\{1, \ldots, m+n\}$ of terms

$$\prod_{j=1}^{n} f_{w(j)-j}(y) \;\cdot\; \prod_{j=n+1}^{m+n} g_{n+w(j)-j}(y)\,.$$

This is a polynomial of degree at most

$$\sum_{j=1}^{m} w(j)-j \quad + \quad \sum_{j=n+1}^{m+n} n+w(j)-j \;=\; mn + \sum_{j=1}^{m+n} w(j)-j \;=\; mn\,.$$

Thus $\mathrm{Res}(f, g; x)$ is a polynomial of degree at most mn.

We complete the proof by showing that the resultant does indeed have degree mn. The product $f_0(y)^n \cdot g_n(y)^m$ of the entries along the main diagonal of the Sylvester matrix has constant term $f_{0,0}^n \cdot g_{0,n}^m$. The coefficient of y^{mn} in this product is $f_{0,0}^n \cdot g_{n,n}^m$, and these are the only terms in the expansion of the determinant of the Sylvester matrix involving either of these monomials in the coefficients $f_{i,j}, g_{k,l}$. □

We now state and prove Bézout's Theorem, which bounds the number of points in the variety $\mathcal{V}(f, g)$ in $\mathbb{A}^2$.

Theorem 2.3.15 (Bézout's Theorem) *Two polynomials $f, g \in \mathbb{F}[x, y]$ either have a common factor or else $|\mathcal{V}(f, g)| \le \deg f \cdot \deg g$.*

When $|\mathbb{F}|$ is at least $\max\{\deg f, \deg g\}$, this inequality is sharp in that the bound is attained. When $\mathbb{F}$ is algebraically closed, the bound is attained when f and g are general polynomials of the given degrees.

Proof. Suppose that $m := \deg f$ and $n = \deg g$. By Lemma 2.3.12, if f and g are relatively prime, then $\mathcal{V}(f, g)$ is finite. Let us extend our field to its algebraic closure $\overline{\mathbb{F}}$, which in infinite. If we change coordinates, replacing f by $f(A(x, y))$ and g by $g(A(x, y))$, where A is an invertible affine transformation,

$$A(x, y) \;=\; (ax + by + c,\; \alpha x + \beta y + \gamma)\,, \tag{2.12}$$

with $a, b, c, \alpha, \beta, \gamma \in \overline{\mathbb{F}}$ with $a\beta - \alpha b \neq 0$. We can choose these parameters so that both f and g have non-vanishing constant terms and nonzero coefficients of their highest powers (m and n, respectively) of y. By Lemma 2.3.14, this implies that the resultant $\mathrm{Res}(f, g; x)$

has degree at most mn and thus at most mn zeroes. If we set $I := \langle f, g\rangle \cap \overline{\mathbb{F}}[y]$, then this also implies that $\mathcal{V}(I) = \mathcal{V}(\mathrm{Res}(f, g; x))$, by Corollary 2.3.11.

We can furthermore choose the parameters in A so that the projection $\pi\colon (x, y) \mapsto y$ is 1-1 on $\mathcal{V}(f, g)$, as $\mathcal{V}(f, g)$ is a finite set. Thus

$$\pi(\mathcal{V}(f, g)) \;=\; \mathcal{V}(I) \;=\; \mathcal{V}(\mathrm{Res}(f, g; x))\,,$$

which implies the inequality of the theorem as $|\mathcal{V}(\mathrm{Res}(f, g; x))| \leq mn$.

To see that the bound is sharp when $|\mathbb{F}|$ is large enough, let $a_1, \ldots, a_m$ and $b_1, \ldots, b_n$ be distinct elements of $\mathbb{F}$. Note that the system

$$f \;:=\; \prod_{i=1}^{m}(x - a_i) \;=\; 0 \qquad \text{and} \qquad g \;:=\; \prod_{i=1}^{n}(y - b_i) \;=\; 0$$

has mn solutions $\{(a_i, b_j) \mid 1 \leq i \leq m, 1 \leq j \leq n\}$, so the inequality is sharp.

Suppose now that $\mathbb{F}$ is algebraically closed. If the resultant $\mathrm{Res}(f, g; x)$ has fewer than mn distinct roots, then either it has degree strictly less than mn or else it has a multiple root. In the first case, its leading coefficient vanishes and in the second case, its discriminant vanishes. But the leading coefficient and the discriminant of $\mathrm{Res}(f, g; x)$ are polynomials in the coefficients of f and g. Thus the set of pairs of polynomials (f, g) with $\mathcal{V}(f, g)$ consisting of mn points in $\mathbb{A}^2$ forms a nonempty open subset in the space $\mathbb{A}^{\binom{m+2}{2}+\binom{n+2}{2}}$ of pairs of polynomials (f, g) with $\deg f = m$ and $\deg g = n$. □

Exercises for Section 3

1. Using the formula (2.6) deduce the Poisson formula for the resultant of univariate polynomials f and g,

$$\mathrm{Res}(f, g; x) \;=\; (-1)^{mn} f_0^n \prod_{i=1}^{m} g(\alpha_i)\,,$$

   where $\alpha_1, \ldots, \alpha_m$ are the roots of f. where $\alpha_1, \ldots, \alpha_n$ are the roots of f.

2. Suppose that the polynomial $g = g_1 \cdot g_2$ factorizes. Show that the resultant also factorizes, $\mathrm{Res}(f, g; x) = \mathrm{Res}(f, g_1; x) \cdot \mathrm{Res}(f, g_2; x)$.

3. Compute the Bezoutian matrix when $n = 4$. Give a general formula for the entries of the Bezoutian matrix.

4. Compute the constant in the proof of Theorem 2.3.5, by computing the resultant and Bezoutian polynomials when $f(x) := x^m$ and $g(x) = x^n + 1$.

5. Write out the 5×5 matrix used to compute the discriminant of a general cubic $x^3 + ax^2 + bx + c$ and take its determinant to show that the discriminant is

$$27d^2 - 18b^2d - 18acd + 9bc^2 + 30a^2bd - 12ab^2c + 3b^4 - 8a^4d + 4a^3bc - a^2b^3 \,.$$

6. Show that the discriminant of a polynomial f of degree n may also be expressed as

$$\prod_{i \neq j} (\alpha_i - \alpha_j)^2 \,,$$

where $\alpha_1, \ldots, \alpha_n$ are the roots of f.

7. Prove the adjoint formula for the inverse of a matrix A, $\det(A) \cdot A^{-1} = \mathrm{ad}(A)$.

8. Let $f(x, y)$ be a polynomial of total degree n. Show that there is a non-empty Zariski open subset of parameters $(a, b, c, \alpha, \beta, \gamma) \in \mathbb{A}^6$ with $a\beta - \alpha b \neq 0$ such that if A is the affine transformation (2.12), then every monomial $x^i y^j$ with $0 \leq i, j$ and $i + j \leq n$ appears in the polynomial $f(A(x, y))$ with a non-zero coefficient.

9. Use Lemma 2.3.12 to show that $\mathbb{A}^2$ has dimension 2, in the sense of the combinatorial definition of dimension (1.4).

10. Use Lemma 2.3.12 and induction on the number of polynomials defining a proper subvariety X of $\mathbb{A}^2$ to show that X consists of finitely many irreducible curves and finitely many isolated points.

2.4 Solving equations with Gröbner bases

Algorithm 2.3.13 used resultants to reduce the problem of solving two equations in two variables to that of solving univariate polynomials. A modification of this algorithm can be used to find all roots of a zero-dimensional ideal $I \subset \mathbb{F}[x_1, \ldots, x_n]$, if we can compute elimination ideals $I \cap \mathbb{F}[x_i, x_{i+1}, \ldots, x_n]$. This may be accomplished using Gröbner bases and more generally, ideas from the theory of Gröbner bases can help us to understand solutions to systems of equations.

Suppose that we have N polynomial equations in n variables $(x_1, \ldots, x_n)$

$$f_1(x_1, \ldots, x_n) \;=\; \cdots \;=\; f_N(x_1, \ldots, x_n) \;=\; 0\,, \tag{2.13}$$

and we want to understand the solutions or *roots* to this system. By understand, we mean answering (any of) the following questions.

(i) Does (2.13) have finitely many solutions?

(ii) If not, can we understand the isolated solutions of (2.13)?

(iii) Can we count them, or give (good) upper bounds on their number?

(iv) Can we *solve* the system (2.13) and find all complex solutions?

(v) When the polynomials have real coefficients, can we count (or bound) the number of real solutions to (2.13)? Or simply find them?

We describe symbolic algorithms based upon Gröbner bases that begin to address these questions.

The solutions to (2.13) in $\mathbb{A}^n$ constitute the affine variety $\mathcal{V}(I)$, where I is the ideal generated by the polynomials $f_1, \ldots, f_N$. Symbolic algorithms to address Questions (i)-(v) involve studying the ideal I. An ideal I is *zero-dimensional* if, over the algebraic closure of $\mathbb{F}$, $\mathcal{V}(I)$ is finite, that is, if the dimension of $\mathcal{V}(I)$ is zero. Thus I is zero-dimensional if and only if the radical $\sqrt{I}$ of I is zero-dimensional.

Theorem 2.4.1 *Let $I \subset \mathbb{F}[x_1, \ldots, x_n]$ be an ideal. Then I is zero-dimensional if and only if $\mathbb{F}[x_1, \ldots, x_n]/I$ is a finite-dimensional $\mathbb{F}$-vector space.*

Proof. We may assume the $\mathbb{F}$ is algebraically closed, as this does not change the dimension of quotient rings.

Suppose first that I is radical. Then $\mathbb{F}[x_1, \ldots, x_n]/I$ is the coordinate ring $\mathbb{F}[X]$ of $X := \mathcal{V}(I)$, and therefore consists of all functions obtained by restricting polynomials to $\mathcal{V}(I)$. If X is finite, then $\mathbb{F}[X]$ is finite-dimensional as the space of functions on X has dimension equal to the number of points in X. Suppose that X is infinite. Then there is some coordinate, say x_1, such that the projection of X to the x_1-axis is infinite and

therefore dense. Restriction of polynomials in x_1 to X is an injective map from $\mathbb{F}[x_1]$ to $\mathbb{F}[X]$ which shows that $\mathbb{F}[X]$ is infinite-dimensional.

Now suppose that I is any ideal. If $\mathbb{F}[x_1,\ldots,x_n]/I$ is finite-dimensional, then so is $\mathbb{F}[x_1,\ldots,x_n]/\sqrt{I}$ as $I \subset \sqrt{I}$. For the other direction, we suppose that $\mathbb{F}[x_1,\ldots,x_n]/\sqrt{I}$ is finite-dimensional. For each variable x_i, there is some linear combination of $1, x_i, x_i^2, \ldots$ which is zero in $\mathbb{F}[x_1,\ldots,x_n]/\sqrt{I}$ and hence lies in $\sqrt{I}$. But this is a univariate polynomial $g_i(x_i) \in \sqrt{I}$, so there is some power $g_i(x_i)^{M_i}$ of g which lies in I. But then we have $\langle g_1(x_1)^{M_1}, \ldots, g_n(x_n)^{M_n}\rangle \subset I$, and so the map

$$\mathbb{F}[x_1,\ldots,x_n]/\langle g_1(x_1)^{M_1},\ldots,g_n(x_n)^{M_n}\rangle \quad\longrightarrow\quad \mathbb{F}[x_1,\ldots,x_n]/I$$

is a surjection. But $\mathbb{F}[x_1,\ldots,x_n]/\langle g_1(x_1)^{M_1},\ldots,g_n(x_n)^{M_n}\rangle$ has dimension $M_1M_2\cdots M_n$, which implies that $\mathbb{F}[x_1,\ldots,x_n]/I$ is finite-dimensional. □

A consequence of the proof is the following criterion for an ideal to be zero-dimensional.

Corollary 2.4.2 *An ideal $I \subset \mathbb{F}[x_1,\ldots,x_n]$ is zero-dimensional if and only if for every variable x_i, there is a univariate polynomial $g_i(x_i)$ which lies in I.*

Together with Macaulay's Theorem 2.2.3, Theorem 2.4.1 leads to a Gröbner basis algorithm to solve Question (i). Let $\succ$ be a monomial order on $\mathbb{F}[x_1,\ldots,x_n]$. Then $\mathbb{F}[x_1,\ldots,x_n]/I$ is finite-dimensional if and only if some power of every variable lies in the initial ideal of I. The lowest such power of a variable is necessarily a generator of the initial ideal. Thus we can determine if I is zero-dimensional and thereby answer Question (i) by computing a Gröbner basis for I and checking that the leading terms of elements of the Gröbner basis include pure powers of all variables.

When I is zero-dimensional, its *degree* is the dimension of $\mathbb{F}[x_1,\ldots,x_n]/I$ as a $\mathbb{F}$-vector space, which is the number of standard monomials, by Macaulay's Theorem 2.2.3. A Gröbner basis for I gives generators of the initial ideal which we can use to count the number of standard monomials to determine the degree of an ideal.

Suppose that the ideal I generated by the polynomials f_i of (2.13) is not zero-dimensional, and we still want to count the isolated solutions to (2.13). In this case, there are symbolic algorithms that compute a zero-dimensional ideal J with $J \supset I$ having the property that $\mathcal{V}(J)$ consists of all isolated points in $\mathcal{V}(I)$, that is all isolated solutions to (2.13). These algorithms successively compute the ideal of all components of $\mathcal{V}(I)$ of maximal dimension, and then strip them off. One such method would be to compute the primary decomposition of an ideal. Another method, when the non-isolated solutions are known to lie on a variety $\mathcal{V}(J)$, is to saturate I by J to remove the excess intersection.[†]

When I is a zero-dimensional radical ideal and $\mathbb{F}$ is algebraically closed, the degree of I equals the number of points in $\mathcal{V}(I) \subset \mathbb{A}^n$ (see Exercise 1) and thus we obtain an answer to Question (iii).

[†]Develop this further, either here or somewhere else, and then refer to that place.

Theorem 2.4.3 *Let I be the ideal generated by the polynomials f_i of* (2.13)*. If I is zero-dimensional, then the number of solutions to the system* (2.13) *is bounded by the degree of I. When $\mathbb{F}$ is algebraically closed, the number of solutions is equal to this degree if and only if I is radical.*

In many important cases, there are sharp upper bounds for the number of isolated solutions to the system (2.13) which do not involve computing a Gröbner basis. For example, Theorem 2.3.15 (Bézout's Theorem in the plane) gave such bounds when $N = n = 2$. Suppose that $N = n$ so that the number of equations equals the number of variables. This is called a *square system*. Bézout's Theorem in the plane has a natural extension in this case, which we will prove in Section 3.2. A common zero a to a square system of equations is *nondegenerate* if the differentials of the equations are linearly independent at a.

Theorem 2.4.4 (Bézout's Theorem) *Given polynomials $f_1, \ldots, f_n \in \mathbb{F}[x_1, \ldots, x_n]$ with $d_i = \deg(f_i)$, the number of nondegenerate solutions to the system*

$$f_1(x_1, \ldots, x_n) \;=\; \cdots \;=\; f_n(x_1, \ldots, x_n) \;=\; 0$$

in $\mathbb{A}^n$ is at most $d_1 \cdots d_n$. When $\mathbb{F}$ is algebraically closed, this is a bound for the number of isolated solutions, and it is attained for generic choices of the polynomials f_i.

This product of degrees $d_1 \cdots d_n$ is called the *Bézout bound* for such a system. While this bound is sharp for generic square systems, few practical problems involve generic systems and other bounds are often needed (see Exercise 6).

We discuss a symbolic method to solve systems of polynomial equations (2.13) based upon elimination theory and the Shape Lemma, which describes the form of a Gröbner basis of a zero-dimensional ideal I with respect to a lexicographic monomial order.

Let $I \subset \mathbb{F}[x_1, \ldots, x_n]$ be an ideal. A univariate polynomial $g(x_i)$ is an *eliminant for I* if g generates the elimination ideal $I \cap \mathbb{F}[x_i]$.

Theorem 2.4.5 *Suppose that $g(x_i)$ is an eliminant for an ideal $I \subset \mathbb{F}[x_1, \ldots, x_n]$. Then $g(a_i) = 0$ for every $a = (a_1, \ldots, a_n) \in \mathcal{V}(I) \in \mathbb{A}^n$. When $\mathbb{F}$ is algebraically closed, every root of g occurs in this way.*

Proof. We have $g(a_i) = 0$ as this is the value of g at the point a. Suppose that $\mathbb{F}$ is algebraically closed and that ξ is a root of $g(x_i)$ but there is no point $a \in \mathcal{V}(I)$ whose ith coordinate is ξ. Let $h(x_i)$ be a polynomial whose roots are the other roots of g. Then h vanishes on $\mathcal{V}(I)$ and so $h \in \sqrt{I}$ and so some power, h^N, of h lies in I. Thus $h^N \in I \cap \mathbb{F}[x_i] = \langle g \rangle$. But this is a contradiction as $h(\xi) \neq 0$ while $g(\xi) = 0$. □

Theorem 2.4.6 *If $g(x_i)$ is a monic eliminant for and ideal $I \subset \mathbb{F}[x_1, \ldots, x_n]$, then g is an element of any reduced Gröbner basis for I with respect to a lexicographic order in which x_i is the minimal variable.*

Proof. Suppose that $\succ$ is a lexicographic monomial order with x_i the minimal variable. Since g generates the elimination ideal $I \cap \mathbb{F}[x_i]$, it is the lowest degree monic polynomial in x_i lying in I. In particular $x_i^{\deg(g)}$ is a generator of the initial ideal of I. Therefore there is a polynomial f in the reduced Gröbner basis whose leading term is $x_i^{\deg(g)}$ and whose remaining terms involve smaller standard monomials. As x_i is the minimal variable and this is a lexicographic monomial order, the only smaller standard monomials are x^d with $d < \deg(g)$. Thus $f \in I$ is a monic polynomial in x_i with degree $\deg(g)$. By the uniqueness of g, $f = g$, which proves that g lies in the reduced Gröbner basis. □

Theorem 2.4.6 gives an algorithm to compute eliminants—simply compute a lexicographic Gröbner basis. This is typically not recommended, as lexicographic Gröbner bases appear to be the most expensive to compute in practice. We instead offer the following algorithm.

Algorithm 2.4.7
INPUT: Ideal $I \subset \mathbb{F}[x_1, \ldots, x_n]$ and a variable x_i.
OUTPUT: Either a univariate eliminant $g(x_i) \in I$ or else a certificate that one does not exist.

(1) Compute a Gröbner basis G for I with respect to any term order.

(2) If no initial term of any element of G is a pure power of x_i, then halt and declare that I does not contain a univariate eliminant for x_i.

(3) Otherwise, compute the sequence $1 \mod G$, $x_i \mod G$, $x_i^2 \mod G$, …, until a linear dependence is found among these,

$$\sum_{j=0}^{m} a_j (x_i^j \mod G) = 0\,, \tag{2.14}$$

where m is minimal. Then

$$g(x_i) = \sum_{j=0}^{m} a_j x_i^j$$

is the univariate eliminant.

Proof of correctness. If I does not have an eliminant in x_i, then $I \cap \mathbb{F}[x_i] = \{0\}$, so $1, x_i, x_i^2, \ldots$ are all standard, and no Gröbner basis contains a polynomial with initial monomial a pure power of x_i. This shows that the algorithm correctly identifies when no eliminant exists.

Suppose now that I does have an eliminant $g(x_i)$. Since $g \mod G = 0$, the Gröbner basis G must contain a polynomial whose initial monomial divides that of g and is hence a pure power of x_i. If $g = \sum b_j x_i^j$ and has degree N, then

$$0 = g \mod G = \Big(\sum_{j=0}^{N} b_j x_i^j\Big) \mod G = \sum_{j=0}^{N} b_j (x_i^j \mod G)\,,$$

which is a linear dependence among the elements of the sequence $1 \mod G$, $x_i \mod G$, $x_i^2 \mod G$, ... Thus the algorithm halts. The minimality of the degree of g implies that $N = m$ and the uniqueness of such minimal linear combinations implies that the coefficients b_j and a_j are proportional, which shows that the algorithm computes a scalar multiple of g, which is also an eliminant. □

Elimination using Gröbner bases leads to an algorithm for addressing Question (v). The first step is to understand the optimal form of a Gröbner basis of a zero-dimensional ideal.

Lemma 2.4.8 (Shape Lemma) *Suppose g is an eliminant of a zero-dimensional ideal I with $\deg(g) = \deg(I)$. Then I is radical if and only if g has no multiple factors.*

If we further have $g = g(x_n)$, then in the lexicographic term order with $x_1 \succ x_2 \succ \cdots \succ x_n$, the ideal I has a Gröbner basis of the form:

$$x_1 - g_1(x_n), \quad x_2 - g_2(x_n), \quad \ldots, \quad x_{n-1} - g_{n-1}(x_n), \quad g(x_n)\,, \tag{2.15}$$

where $\deg(g) > \deg(g_i)$ for $i = 1, \ldots, n-1$.

If I is generated by real polynomials, then the number of real roots of I equals the number of real roots of g.

Proof. We have

$$\#\text{roots of } g \;\leq\; \#\text{roots of } I \;\leq\; \deg(I) \;=\; \deg(g)\,,$$

the first inequality by Theorem 2.4.5 and the second by Theorem 2.4.3. If the roots of g are distinct, then their number is $\deg(g)$ and so these inequalities are equalities. This implies that I is radical, by Theorem 2.4.3. Conversely, if $g = g(x_i)$ has multiple roots, then there is a polynomial h with the same roots as g but with smaller degree. Since $\langle g \rangle = I \cap \mathbb{F}[x_i]$, we have that $h \notin I$, but since $h^{\deg(g)}$ is divisible by g, $h^{\deg(g)} \in I$, so I is not radical.

To prove the second statement, let d be the degree of the eliminant $g(x_n)$. Then each of $1, x_n, \ldots, x_n^{d-1}$ is a standard monomial, and since $\deg(g) = \deg(I)$, there are no others. Thus the initial ideal in the lexicographic monomial order is $\langle x_1, \ldots, x_{n-1}, x_n^d \rangle$. Each element of the reduced Gröbner basis for I expresses a generator of the initial ideal as a $\mathbb{F}$-linear combination of standard monomials. It follows that the reduced Gröbner basis has the form claimed.

For the last statement, we may reorder the variables if necessary and assume that $g = g(x_n)$. The common zeroes of the polynomials (2.15) are

$$\{(a_1, \ldots, a_n) \mid g(a_n) = 0 \text{ and } a_i = g_i(a_n)\,,\ i = 1, \ldots, n-1\}\,.$$

By Corollary 2.2.7, the polynomials g_i are all real, and so a component a_i is real if the root a_n of g is real. □

Not all ideals can have such a Gröbner basis. For example, the ideal

$$\langle x, y\rangle^2 \;=\; \langle x^2, xy, y^2\rangle$$

cannot. Nevertheless, the key condition on the eliminant g. that $\deg(g) = \deg(I)$, often holds after a generic change of coordinates, just as in the proof of Bézout's Theorem in the plane (Theorem 2.3.15). This gives the following symbolic algorithm to count the number of real solutions to a system of equations.

Algorithm 2.4.9 (Counting real roots) Given a system of polynomial equations (2.13), let I be the ideal generated by the polynomials. If I is zero-dimensional, then compute an eliminant $g(x_i)$ and check if $\deg(g) = \deg(I)$, if not, then perform a generic change of variables and compute another eliminant.[†]

Given such a eliminant g satisfying the hypotheses of the Shape Lemma, use Sturm sequences or any other method to determine its number of real solutions, which is the number of real solutions to the original system.

This simple algorithm is the idea behind the efficient algorithm REALSOLVING of Faugère and Roullier.

While the Shape Lemma describes an optimal form of a Gröbner basis for a zero-dimensional ideal, it is typically not optimal to compute such a Gröbner basis directly. An alternative to computation of a Lexicographic Gröbner basis is the *FGLM algorithm* of Faugère, Gianni, Lazard, and Mora [7], which is an algorithm for Gröbner basis conversion. That is, given a Gröbner basis for a zero dimensional ideal with respect to one monomial order and a different monomial order $\succ$, FGLM computes a Gröbner basis for the ideal with respect to $\succ$.

Algorithm 2.4.10 (FGLM)
INPUT: A Gröbner basis G for a zero-dimensional ideal $I \subset \mathbb{F}[x_1, \ldots, x_n]$ with respect to a monomial order $\geq$, and a different monomial order $\succ$.
OUTPUT: A Gröbner basis H for I with respect to $\succ$.
INITIALIZE: Set $H := \{\}$, $x^\alpha := 1$, and $S := \{\}$.

(1) Compute $\overline{x^\alpha} := x^\alpha \mod G$.

(2) If $\overline{x^\alpha}$ does not lie in the linear span of S, then set $S := S \cup \{\overline{x^\alpha}\}$.

Otherwise, there is a (unique) linear combination of elements of S such that

$$\overline{x^\alpha} \;=\; \sum_{\overline{x^\beta} \in S} c_\beta \overline{x^\beta}\,.$$

Set $H := H \cup \{x^\alpha - \sum_\beta c_\beta x^\beta\}$.

[†]This is flawed. Maybe we should instead present a form of the Rational univariate eliminant.

(3) If

$$\{x^\alpha \mid x^\alpha \succ x^\gamma\} \;\subset\; \mathrm{in}(H) := \{\mathrm{in}_\succ(h) \mid h \in H\}\,,$$

then halt and output H. Otherwise, set x^α to be the $\succ$-minimal monomial in the set $\{x^\gamma \notin \mathrm{in}(H) \mid x^\alpha \succ x^\gamma\}$ and return to (1).

Proof of correctness. By construction, H always consists of elements of I, and elements of S are linearly independent in the quotient ring $F[x_1,\ldots,x_n]/I$. Thus $\mathrm{in}_\succ(H) := \{\mathrm{in}_\succ(h) \mid h \in H\}$ is a subset of the initial ideal $\mathrm{in}_\succ I$, and we always have the inequalities

$$|S| \;\le\; \dim_{\mathbb{F}}(\mathbb{F}[x_1,\ldots,x_n]/I) \qquad \text{and} \qquad \mathrm{in}_\succ(H) \;\subset\; \mathrm{in}_\succ I\,.$$

Every time we return to (1) either the set S or the set H (and hence $\mathrm{in}_\succ(H)$) increases. Since the cardinality of S is bounded and the monomial ideals generated by the different $\mathrm{in}_\succ(H)$ form a strictly increasing chain, the algorithm must halt.

When it halts, every monomial is either in the set $\mathrm{SM} := \{\mathrm{x}^\beta \mid \overline{\mathrm{x}^\beta} \in \mathrm{S}\}$ or else in the monomial ideal generated by $\mathrm{in}_\succ(H)$. By our choice of x^α in (3), these two sets are disjoint, so that SM is the set of standard monomials for $\langle \mathrm{in}_\succ(H)\rangle$. Since $\mathrm{in}_\succ(H) \subset \mathrm{in}_\succ\langle H\rangle \subset \mathrm{in}_\succ I$, and elements of S are linearly independent modulo $\mathrm{in}_\succ I$, we have

$$|S| \;\le\; \dim_{\mathbb{F}}(\mathbb{F}[x]/\mathrm{in}_\succ I) \;\le\; \dim_{\mathbb{F}}(\mathbb{F}[x]/\mathrm{in}_\succ\langle H\rangle) \;\le\; \dim_{\mathbb{F}}(\mathbb{F}[x]/\langle \mathrm{in}_\succ(H)\rangle) \;=\; |S|\,.$$

Thus $\mathrm{in}_\succ I = \langle \mathrm{in}_\succ(H)\rangle$, which proves that H is a Gröbner basis for I with respect to the monomial order $\succ$. By the form of the elements of H, it is the reduced Gröbner basis. □

Exercises for Section 4

1. Suppose $I \subset \mathbb{F}[x_1,\ldots,x_n]$ is radical, $\mathbb{F}$ is algebraically closed, and $\mathcal{V}(I) \subset \mathbb{A}^n$ consists of finitely many points. Show that the coordinate ring $\mathbb{F}[x_1,\ldots,x_n]/I$ of restrictions of polynomial functions to $\mathcal{V}(I)$ has dimension as a $\mathbb{F}$-vector space equal to the number of points in $\mathcal{V}(I)$.

2. Verify Bézout on random examples.

3. Study multilinear equations on random examples.

4. Study sparse equations on random examples.

5. Study degrees os Schubert varieties on random examples.

6. Compute the number of solutions to the system of polynomials

$$1 + 2x + 3y + 5xy \;=\; 7 + 11xy + 13xy^2 + 17x^2y \;=\; 0\,.$$

Show that each is nondegenerate and compare this to the Bézout bound for this system. How many solutions are real?

2.5 Numerical Homotopy continuation

In the previous two sections, we discussed symbolic approaches to solving systems of polynomial equations. Resultants, for which we have formulas, do not always give precise information about the corresponding ideal, and are not universally applicable. On the other hand, Gröbner bases are universally applicable and may be readily computed. The drawback of Gröbner bases is that they contain too much information and therefore may be expensive or impossible to compute.

It is natural to ask for a method to find numerical solutions that does not require a Gröbner basis. *Homotopy algorithms* furnish one such class of methods. These find all isolated solutions to a system of polynomial equations [14]. For large classes of systems, there are optimal homotopy algorithms, and they have the additional advantage of being inherently parallelizable.

Example 2.5.1 Suppose we want to compute the (four) solutions to the equations

$$x(y-1) \;=\; 1 \qquad \text{and} \qquad 2x^2+y^2 \;=\; 9\,. \tag{2.16}$$

If we consider instead the system

$$x(y-1) \;=\; 0 \qquad \text{and} \qquad 2x^2+y^2 \;=\; 9\,, \tag{2.17}$$

then we obtain the following solutions by inspection

$$(0,\pm 3) \qquad \text{and} \qquad (\pm 2,1)\,.$$

Figure 2.2 shows the two systems, the first seeks the intersection of the hyperbola with the ellipse, while the second replaces the hyperbola by the two lines.

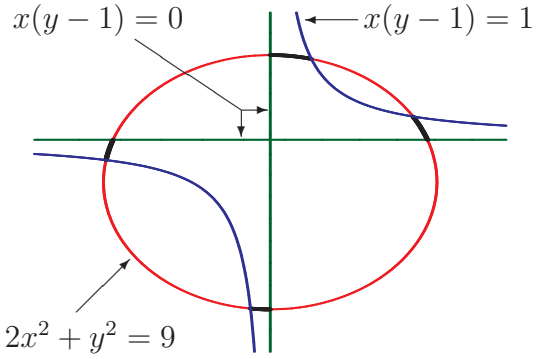


Figure 2.2: The intersection of a hyperbola with an ellipse.

The 1-parameter family of polynomial systems

$$x(y-1) \;=\; t \qquad \text{and} \qquad 2x^2+y^2 \;=\; 9\,,$$

interpolates between these two systems as t varies from 0 to 1. This defines four solution curves $z_i(t)$ (the thickened arcs in Figure 2.2) for $t \in [0,1]$, which connect each solution $(0, \pm 3)$, $(\pm 2, 1)$ of (2.17) to a solution of (2.16). We merely trace each solution curve $z_i(t)$ from the known solution $z_i(0)$ at $t = 0$ to the desired solution $z_i(1)$ at $t = 1$, obtaining all solutions of (2.16).

More generally, suppose that we want to find all solutions to a 0-dimensional *target system* of polynomial equations

$$f_1(x_1, \ldots, x_n) \;=\; f_2(x_1, \ldots, x_n) \;=\; \cdots \;=\; f_N(x_1, \ldots, x_n) \;=\; 0\,, \qquad (2.18)$$

written compactly as $F(x) = 0$. Numerical homotopy continuation finds these solutions if we have a *homotopy*, which is a system $H(x,t)$ of polynomials in $n+1$ variables such that

1. The systems $H(x,1) = 0$ and $F(x) = 0$ both have the same solutions;

2. We know all solutions to the *start system* $H(x,0) = 0$;

3. The components of the variety defined by $H(x,t) = 0$ include curves whose projection to $\mathbb{C}$ (via the second coordinate t) is dominant; and

4. The solutions to the system $H(x,t) = 0$, where $t \in [0,1)$, occur at smooth points of curves from (3) in the variety $H(x,t) = 0$.

We summarize these properties: The homotopy $H(x,t)$ interpolates between our original system (1) and a trivial system (2) such that all isolated solutions are attained (3) and we can do this avoiding singularities (4).

Given such a homotopy, we restrict the variety $H(x,t) = 0$ to $t \in [0,1]$ and obtain finitely many real arcs in $\mathbb{C}^n \times [0,1]$ which connect (possibly singular) solutions of the target system $H(x,1) = 0$ to solutions of the start system $H(x,0) = 0$. We then numerically trace each arc from $t = 0$ to $t = 1$, obtaining all isolated solutions to the target system.

The homotopy is *optimal* if every solution at $t = 0$ is connected to a unique solution at $t = 1$ along an arc. This is illustrated in Figure 2.3.

Remark 2.5.2 Homotopy continuation software often constructs a homotopy as follows. Let $F(x)$ be the target system (2.18) and suppose we have solutions to a *start system* $G(x)$. Then for a number $\gamma \in \mathbb{C}$ with $|\gamma| = 1$ define

$$H(x,t) \;:=\; \gamma t F(x) + (1-t)G(x)\,.$$

Then $H(x,t)$ satisfies the definition of a homotopy for all but finitely many γ. The software detects the probability 0 event that $H(x,t)$ does not satisfy the definition when it encounters a singularity, and then it recreates the homotopy with a different γ.

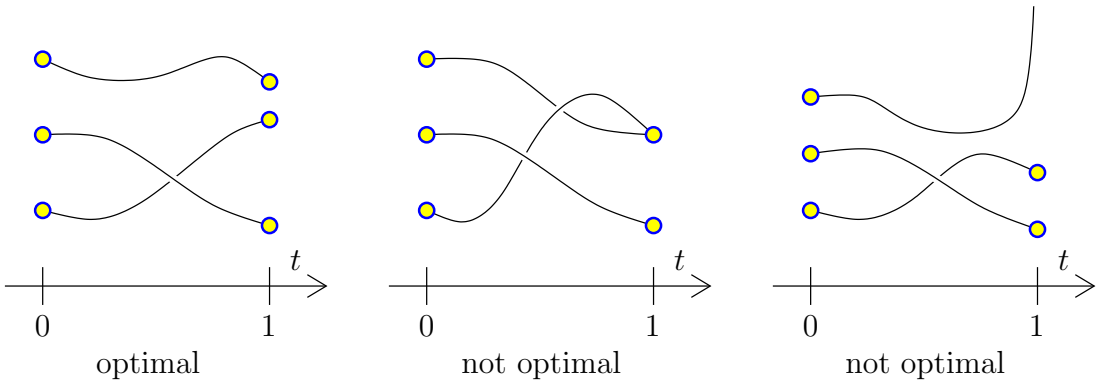


Figure 2.3: Optimal and non-optimal homotopies

Example 2.5.3 We illustrate these ideas. Suppose we have a square system

$$f_1(x_1,\ldots,x_n) \;=\; \cdots \;=\; f_n(x_1,\ldots,x_n) \;=\; 0\,, \tag{2.19}$$

where $\deg f_i = d_i$. If, for each $i = 1,\ldots,n$, we set

$$g_i \;:=\; \prod_{i=1}^{d_i}(x_i - j)\,, \tag{2.20}$$

then we immediately see that the start system

$$g_1(x_1,\ldots,x_n) \;=\; \cdots \;=\; g_n(x_1,\ldots,x_n) \;=\; 0 \tag{2.21}$$

has the $d_1 d_2 \cdots d_n$ solutions

$$\prod_{i=1}^{n}\{1,2,\ldots,d_i\} \;=\; \{(a_1,\ldots,a_n)\in\mathbb{Z}^n \mid 1\le a_i\le d_i,\ i=1,\ldots,n\}\,.$$

The *Bézout homotopy* $H(x,t)$ consists of the polynomials

$$h_i(x,t) \;:=\; \gamma t\cdot f_i(x) + (1-t)g_i(x) \text{ for } i = 1,\ldots,n\,, \tag{2.22}$$

where γ is an arbitrary complex number. If the isolated solutions to the original system (2.19) $F(x) = 0$ occur without multiplicities, then $H(x,t)$ is a homotopy, as Bézout's Theorem 2.4.4 implies that all intermediate systems $H(x,t_0)$ have at most $d_1\cdots d_n$ isolated solutions. Furthermore, if γ is chosen from a generic set, then there are no multiple solutions for $t\in[0,1]$. We may replace the polynomials g_i (2.20) by any other polynomials (of the same degree) so that the start system has $d_1\cdots d_n$ solutions. For example, we may take $g_i := x_i^{d_i} - 1$.

Path following algorithms use predictor-corrector methods, which are conceptually simple for *square systems*, where the number of equations equals the number of variables.

Given a point (x_0, t_0) on an arc such that $t_0 \in [0, 1)$, the $n \times n$ matrix

$$H_x := \left(\frac{\partial H_i}{\partial x_j}\right)_{i,j=1}^{n}$$

is regular at (x_0, t_0), which follows from the definition of the homotopy. Let $H_t = (\partial H_1/\partial t, \ldots, \partial H_n/\partial t)^T$, given δt we set

$$\delta x := -\delta t\, H_x(x_0, t_0)^{-1}\, H_t(x_0, t_0)\,.$$

Then the vector $(\delta x, \delta t)$ is tangent to the arc $H(x, t) = 0$ at the point (x_0, t_0). For $t_1 = t_0 + \delta t$, the point $(x', t_1) = (x_0 + \Delta x, t_1)$ is an approximation to the point (x_1, t_1) on the same arc. This constitutes a first order predictor step. A corrector step uses the multivariate Newton method for the system $H(x, t_1) = 0$, refining the approximate solution x' to a solution x_1. In practice, the points x_0 and x_1 are numerical (approximate) solutions, and both the prediction and correction steps require that $\det H_x \neq 0$ at every point where the computation of the Jacobian matrix H_x is done.

When the system is not square, additional strategies must be employed to enable the path following.

These numerical methods for following a solution curve $x(t)$ will break down if $x(t)$ approaches a singularity of the variety $H(x, t) = 0$, as the rank of the Jacobian matrix $J(x(t), x)$ will drop at such points. This will occur when t lies in an algebraic subvariety of $\mathbb{C}$, that is, for finitely many $t \in \mathbb{C}$. In practice, the choice of random complex number γ in (2.22) precludes this situation.

Another possibility is that the target system has singular solutions at which the homotopy must necessarily end. Numerical algorithms to detect and handle these and other end-game situations have been devised and implemented. This is illustrated in the middle picture in Figure 2.3.

Another, more serious possibility is when the solution curve $x(t)$ does not converge to an isolated solution of the original system. By assumptions (III) and (IV), this can only occur of there are fewer isolated solutions to our original system than to the start system. In this case, some solution curves $x(t)$ will diverge to ∞ as t approaches 1. A homotopy $H(x, t)$ is optimal if this does not occur; that is, if almost all intermediate systems $H(x, t) = 0$ (including $t = 0, 1$) have the same number of isolated solutions.

For example, suppose our original system (2.18) is a generic square system ($N = n$) with polynomials of degrees $d_1, \ldots, d_n$. (This is also called a generic dense system.) Then Bézout's Theorem implies that it has $d_1 d_2 \cdots d_n$ isolated solutions. Since the start system also has this number of solutions, almost all intermediate solutions will, too, and so we obtain the following fundamental result.

Theorem 2.5.4 *For generic dense systems, the Bézout homotopy is optimal.*

The only difficulty with Theorem 2.5.4 and the Bézout homotopy is that polynomial systems arising 'in nature' from applications are rarely generic dense systems. The Bézout bound for the number of isolated solutions is typically not achieved. A main challenge in this subject is to construct optimal homotopies.

For example, consider the system of cubic polynomials

$$\begin{array}{rcl} 1 + 2x + 3y + 4xy + 5x^2y + 6xy^2 & = & 0 \\ 5 + 7x + 11y + 13xy + 17x^2y + 19xy^2 & = & 0 \end{array} \tag{2.23}$$

we leave it an an exercise to check that this has 5 solutions and not 9, which is the Bézout bound.

Another source of optimal homotopies are *Cheater homotopies* [12], which are constructed from families of polynomial systems. For example, given a Schubert problem $(\lambda^1, \ldots, \lambda^m)$, let V be the space of all m-tuples $(F_\bullet^1, \ldots, F_\bullet^m)$ of flags. The total space of the Schubert problem

$$U \;:=\; \{(H, F_\bullet^1, \ldots, F_\bullet^m) \in G(k,n) \times V \mid H \in Y_{\lambda^i} F_\bullet^i \;\text{ for } i = 1, \ldots, m\}$$

is defined by equations (see Section **??**) depending upon the point $(F_\bullet^1, \ldots, F_\bullet^m) \in V$. If $\varphi\colon \mathbb{C} \to V$ is an embedding of $\mathbb{C}$ into V in which $\varphi(0)$ and $\varphi(1)$ are general m-tuples of flags and we write $\varphi(t) = (F_\bullet^1(t), \ldots, F_\bullet^m(t))$, then

$$\varphi^* U \;=\; \{(H, F_\bullet^1(t), \ldots, F_\bullet^m(t)) \mid H \in Y_{\lambda^i} F_\bullet^i(t) \;\text{ for } i = 1, \ldots, m\}\ .$$

This is defined by a system $H(x,t) = 0$, which gives an optimal homotopy.

The computational complexity of solving systems of polynomials using an optimal homotopy is roughly linear in the number of solutions, for a fixed number of variables. The basic idea is that the cost of following each solution curve is essentially constant. This happy situation is further enhanced as homotopy continuation algorithms are inherently massively parallelizable—once the initial precomputation of solving the start system and setting up the homotopies is completed, then each solution curve may be followed independently of all other solution curves.

Exercises for Section 5

1. Verify the claim in the text that the system (2.23) has five solutions. Show that only one is real.

Chapter 3

Projective and toric varieties

Outline:

1. Projective space and projective varieties.
2. Hilbert functions and degree.
3. Toric ideals.
4. Toric varieties.
5. Bernstein's Theorem.

3.1 Projective varieties

Projective space and projective varieties are undoubtedly the most important objects in algebraic geometry. We will first motivate projective space with an example.

Consider the intersection of the parabola $y = x^2$ in the affine plane $\mathbb{A}^2$ with a line, $\ell := \mathcal{V}(ay + bx + c)$. Solving these implied equations gives

$$ax^2 + bx + c \;=\; 0 \qquad \text{and} \qquad y \;=\; x^2\,.$$

There are several cases to consider.

(i) $a \neq 0$ and $b^2 - 4ac > 0$. Then ℓ meets the parabola in two distinct real points.

(i′) $a \neq 0$ and $b^2 - 4ac < 0$. While ℓ does not appear to meet the parabola, that is because we have drawn the real picture, and ℓ meets it in two complex conjugate points.

When $\mathbb{F}$ is algebraically closed, then cases (i) and (i′) coalesce, to the case of $a \neq 0$ and $b^2 - 4ac \neq 0$. These two points of intersection are predicted by Bézout's Theorem in the plane (Theorem 2.3.15).

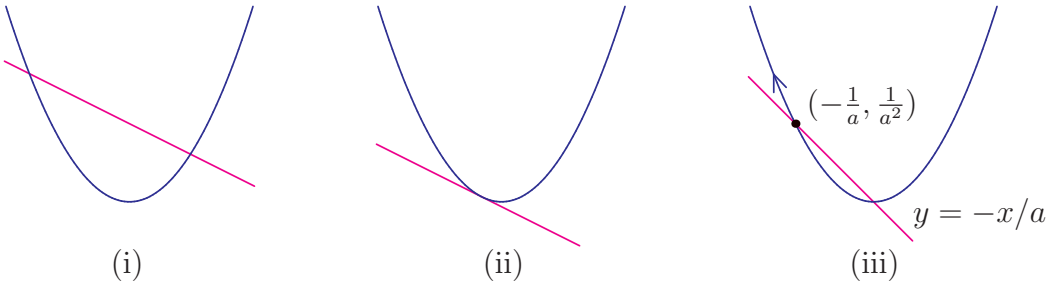
(ii) $a \neq 0$ but $b^2 - 4ac = 0$. Then ℓ is tangent to the parabola, and when we solve the equations, we get

$$a(x - \tfrac{b}{2a})^2 \;=\; 0 \qquad \text{and} \qquad y \;=\; x^2\,.$$

Thus there is one solution, $(\frac{b}{2a}, \frac{b^2}{4a^2})$. As $x = \frac{b}{2a}$ is a root of multiplicity 2 in the first equation, it is reasonable to say that this one solution to our geometric problem occurs with multiplicity 2.

(iii) $a = 0$. There is a single, unique solution, $x = -c/b$ and $y = c^2/b^2$.

Suppose now that $c = 0$ and let $b = 1$. For $a \neq 0$, there are two solutions $(0, 0)$ and $(-\frac{1}{a}, \frac{1}{a^2})$. In the limit as $a \to 0$, the second solution runs off to infinity.



One purpose of projective space is to prevent this last phenomenon from occurring.

Definition 3.1.1 The set of all 1-dimensional linear subspaces of $\mathbb{F}^{n+1}$ is called *n-dimensional projective space* and written $\mathbb{P}^n$ or $\mathbb{P}^n_{\mathbb{F}}$. If V is a finite-dimensional vector space, then $\mathbb{P}(V)$ is the set of all 1-dimensional linear subspaces of V. Note that $\mathbb{P}(V) \simeq \mathbb{P}^{\dim V - 1}$, but there are no preferred coordinates for $\mathbb{P}(V)$.

Example 3.1.2 $\mathbb{P}^1$ is the set of lines through the origin in $\mathbb{F}^2$. When $\mathbb{F} = \mathbb{R}$, we see that every line through the origin $x = ay$ intersects the circle $\mathcal{V}(x^2 + (y-1)^2 - 1)$ in the origin and in the point $(2a/(1+a^2),\, 2/(1+a^2))$, as shown in Figure 3.1. Identifying the x-axis with the origin and the lines $x = ay$ with this point of intersection gives a one-to-one map from $\mathbb{P}^1_{\mathbb{R}}$ to the circle, where the origin becomes the point at infinity.

Our definition of $\mathbb{P}^n$ leads to a system of global homogeneous coordinates for $\mathbb{P}^n$. We may represent a point, ℓ, of $\mathbb{P}^n$ by the coordinates $[a_0, a_1, \ldots, a_n]$ of any non-zero vector lying on the one-dimensional linear subspace $\ell \subset \mathbb{F}^{n+1}$. These coordinates are not unique. If $\lambda \neq 0$, then $[a_0, a_1, \ldots, a_n]$ and $[\lambda a_0, \lambda a_1, \ldots, \lambda a_n]$ both represent the same point. This non-uniqueness is the reason that we use rectangular brackets $[\ldots]$ in our notation for these *homogeneous coordinates*. Some authors prefer the notation $[a_0 : a_1 : \cdots : a_n]$.

Example 3.1.3 When $\mathbb{F} = \mathbb{R}$, note that a 1-dimensional subspace of $\mathbb{R}^{n+1}$ meets the sphere S^n in two antipodal points, v and $-v$. This identifies real projective space $\mathbb{P}^n_{\mathbb{R}}$ with the quotient $S^n/\{\pm 1\}$, showing that $\mathbb{P}^n_{\mathbb{R}}$ is a compact manifold.

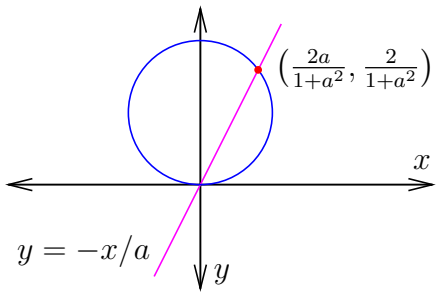


Figure 3.1: Lines through the origin meet the circle $\mathcal{V}(x^2+(y-1)^2-1)$ in a second point.

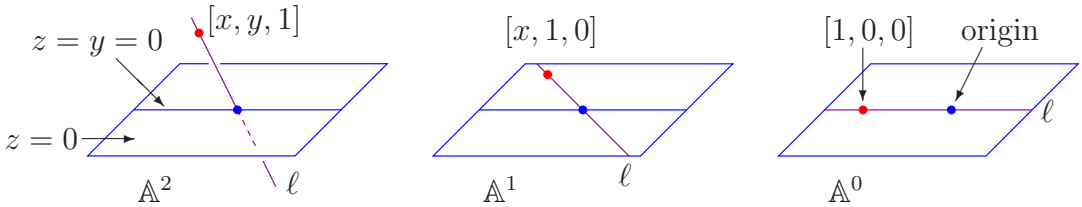
Suppose that $\mathbb{F} = \mathbb{C}$. Given a point $a \in \mathbb{P}^n_{\mathbb{C}}$, we can assume that $|a_0|^2 + |a_1|^2 + \cdots + |a_n|^2 = 1$. Identifying $\mathbb{C}$ with $\mathbb{R}^2$, this is the set of points a on the $2n+1$-sphere $S^{2n+1} \subset \mathbb{R}^{2n+2}$. If $[a_0, \ldots, a_n] = [b_0, \ldots, b_n]$ with $a, b \in S^{2n+1}$, then there is some $\zeta \in S^1$, the unit circle in $\mathbb{C}$, such that $a_i = \zeta b_i$. This identifies $\mathbb{P}^n_{\mathbb{C}}$ with the quotient of S^{2n+1}/S^1, showing that $\mathbb{P}^n_{\mathbb{C}}$ is a compact manifold. Since $\mathbb{P}^n_{\mathbb{R}} \subset \mathbb{P}^n_{\mathbb{C}}$, we again see that $\mathbb{P}^n_{\mathbb{R}}$ is a compact manifold.

Homogeneous coordinates of a point are not unique. Uniqueness may be restored, but at the price of non-uniformity. Let $A_i \subset \mathbb{P}^n$ be the set of points $[a_0, a_1, \ldots, a_n]$ in projective space $\mathbb{P}^n$ with $a_i \neq 0$, but $a_{i+1} = \cdots = a_n = 0$. Given a point $a \in A_i$, we may divide by its ith coordinate to get a representative of the form $[a_0, \ldots, a_{i-1}, 1, 0, \ldots, 0]$. These i numbers $(a_0, \ldots, a_{i-1})$ provide coordinates for A_i, identifying it with the affine space $\mathbb{A}^i$. This decomposes projective space $\mathbb{P}^n$ into a disjoint union of $n+1$ affine spaces

$$\mathbb{P}^n \;=\; \mathbb{A}^n \sqcup \cdots \sqcup \mathbb{A}^1 \sqcup \mathbb{A}^0 \,.$$

When a variety admits a decomposition as a disjoint union of affine spaces, we say that it is *paved by affine spaces*. Many important varieties admit such a decomposition.

It is instructive to look at this closely for $\mathbb{P}^2$. Below, we show the possible positions of a one-dimensional linear subspace $\ell \subset \mathbb{F}^3$ with respect to the x, y-plane $z = 0$, the x-axis $z = y = 0$, and the origin in $\mathbb{F}^3$.



There is also a scheme for local coordinates on projective space.

1. For $i = 0, \ldots, n$, let U_i be the set of points $a \in \mathbb{P}^n$ in projective space whose ith coordinate is non-zero. Dividing by this ith coordinate, we obtain a representative

of the point having the form

$$[a_0, \ldots, a_{i-1}, 1, a_{i+1}, \ldots, a_n]\,.$$

The n coordinates $(a_0, \ldots, a_{i-1},\ a_{i+1}, \ldots, a_n)$ determine this point, identifying U_i with affine n-space, $\mathbb{A}^n$. Every point of $\mathbb{P}^n$ lies in some U_i,

$$\mathbb{P}^n \ = \ U_0 \cup U_1 \cup \cdots \cup U_n\,.$$

When $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, these U_i are coordinate charts for $\mathbb{P}^n$ as a manifold.

For any field $\mathbb{F}$, these affine sets U_i provide coordinate charts for $\mathbb{P}^n$.

2. We give a coordinate-free description of these affine charts. Let $\Lambda\colon \mathbb{F}^{n+1} \to \mathbb{F}$ be a linear map, and let $H \subset \mathbb{F}^{n+1}$ be the set of points x where $\Lambda(x) = 1$. Then $H \simeq \mathbb{A}^n$, and the map

$$H \ni x \ \longmapsto \ [x] \in \mathbb{P}^n$$

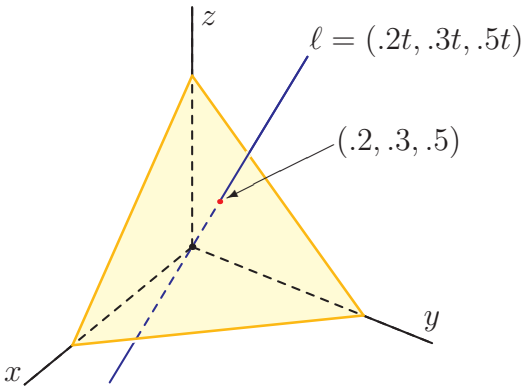
identifies H with the complement $U_\Lambda = \mathbb{P}^n - \mathcal{V}(\Lambda)$ of the points where Λ vanishes.

Example 3.1.4 (Probability simplex) This more general description of affine charts leads to the beginning of an important application of algebraic geometry to statistics. Here $\mathbb{F} = \mathbb{R}$, the real numbers and we set $\Lambda(x) := x_0 + \cdots + x_n$. If we consider those points x where $\Lambda(x) = 1$ which have positive coordinates, we obtain the *probability simplex*

$$\Delta \ := \ \{(p_0, p_1, \ldots, p_n) \in \mathbb{R}^{n+1}_+ \mid p_0 + p_1 + \cdots + p_n = 1\}\,,$$

where $\mathbb{R}^{n+1}_+$ is the *positive orthant*, the points of $\mathbb{R}^{n+1}$ with nonnegative coordinates. Here p_i represents the probability of an event i occurring, and the condition $p_0 + \cdots + p_n = 1$ reflects that every event does occur.

Here is a picture when $n = 2$.



3.1.1 Subvarieties of $\mathbb{P}^n$

We wish to extend the definitions and structures of affine algebraic varieties to projective space. One problem arises immediately: given a polynomial $f \in \mathbb{F}[x_0, \ldots, x_n]$ and a point $a \in \mathbb{P}^n$, we cannot in general define $f(a) \in \mathbb{F}$. To see why this is the case, for each non negative integer d, let f_d be the sum of the terms of f of degree d. We call f_d the dth *homogeneous component* of f. If $[a_0, \ldots, a_n]$ and $[\lambda a_0, \ldots, \lambda a_n]$ are two representatives of $a \in \mathbb{P}^n$, and f has degree m, then

$$f(\lambda a_0, \ldots, \lambda a_n) \;=\; f_0(a_0, \ldots, a_n) + \lambda f_1(a_0, \ldots, a_n) + \cdots + \lambda^m f_m(a_0, \ldots, a_n)\,, \qquad (3.1)$$

since we can factor λ^d from every monomial $(\lambda x)^\alpha$ of degree d. Thus $f(a)$ is a well-defined number only if the polynomial (3.1) in λ is constant. That is, if and only if

$$f_i(a_0, \ldots, a_n) \;=\; 0 \qquad i = 1, \ldots, \deg(f)\,.$$

In particular, a polynomial f vanishes at a point $a \in \mathbb{P}^n$ if and only if every homogeneous component f_d of f vanishes at a. A polynomial f is *homogeneous* of degree d when $f = f_d$. We also use the term *homogeneous form* for a homogeneous polynomial.

Definition 3.1.5 Let $f_1, \ldots, f_m \in \mathbb{F}[x_0, \ldots, x_n]$ be homogeneous polynomials. These define a *projective variety*

$$\mathcal{V}(f_1, \ldots, f_m) \;:=\; \{a \in \mathbb{P}^n \mid f_i(a) = 0,\ i = 1, \ldots, m\}\,.$$

An ideal $I \subset \mathbb{F}[x_0, \ldots, x_n]$ is *homogeneous* if whenever $f \in I$ then all homogeneous components of f lie in I. Thus projective varieties are defined by homogeneous ideals. Given a subset $Z \subset \mathbb{P}^n$ of projective space, its ideal is the collection of polynomials which vanish on Z,

$$\mathcal{I}(Z) \;:=\; \{f \in \mathbb{F}[x_0, x_1, \ldots, x_n] \mid f(z) = 0 \text{ for all } z \in Z\}\,.$$

In the exercises, we ask you to show that this ideal is homogeneous.

It is often convenient to work in an affine space when treating projective varieties. The *(affine) cone* $CZ \subset \mathbb{F}^{n+1}$ over a subset Z of projective space $\mathbb{P}^n$ is the union of the one-dimensional linear subspaces $\ell \subset \mathbb{F}^{n+1}$ corresponding to points of Z. Then the ideal $\mathcal{I}(X)$ of a projective variety X is equal to the ideal $\mathcal{I}(CX)$ of the affine cone over X.

Example 3.1.6 Let $\Lambda := a_0x_0 + a_1x_1 + \cdots + a_nx_n$ be a linear form. Then $\mathcal{V}(\Lambda)$ is a *hyperplane*. Let $V \subset \mathbb{F}^{n+1}$ be the kernel of Λ which is an n-dimensional linear subspace. It is also the affine variety defined by Λ. We have $\mathcal{V}(\Lambda) = \mathbb{P}(V)$.

The weak Nullstellensatz does not hold for projective space, as $\mathcal{V}(x_0, x_1, \ldots, x_n) = \emptyset$. We call this ideal, $\mathfrak{m}_0 := \langle x_0, x_1, \ldots, x_n \rangle$, the *irrelevant ideal*. It plays a special role in the projective algebraic geometric dictionary.

Theorem 3.1.7 (Projective Algebraic-Geometric Dictionary) *Over any field $\mathbb{F}$, the maps $\mathcal{V}$ and $\mathcal{I}$ give an inclusion reversing correspondence*

$$\left\{\begin{array}{c}\textit{Radical homogeneous ideals } I \textit{ of} \\ \mathbb{F}[x_0,\ldots,x_n] \textit{ properly contained in } \mathfrak{m}_0\end{array}\right\} \quad \begin{array}{c}\xrightarrow{\mathcal{V}} \\ \xleftarrow[\mathcal{I}]{}\end{array} \quad \{\textit{Subvarieties } X \textit{ of } \mathbb{P}^n\}$$

with $\mathcal{V}(\mathcal{I}(X)) = X$. When $\mathbb{F}$ is algebraically closed, the maps $\mathcal{V}$ and $\mathcal{I}$ are inverses, and this correspondence is a bijection.

We can deduce this from the algebraic-geometric dictionary for affine space (Corollary 1.2.10). if we replace a subvariety X of projective space by its affine cone CX.

Many of the basic notions from affine varieties extend to projective varieties for many of the the same reasons. In particular, we have the Zariski topology with open and closed sets, and the notion of generic sets. Projective varieties are finite unions of irreducible varieties, in an essentially unique way. The definitions, statements, and proofs are the same as in Sections 1.3 and 1.4.

If we relax the condition that an ideal be radical, then the corresponding geometric objects are *projective schemes*. This comes at a price, for many homogeneous ideals will define the same projective scheme. This non-uniqueness comes from the irrelevant ideal, $\mathfrak{m}_0$. Recall the construction of colon ideals. Let I and J be ideals. Then the *colon ideal* (or *ideal quotient* of I by J) is

$$(I : J) \;:=\; \{f \mid fJ \subset I\}\,.$$

An ideal $I \subset \mathbb{F}[x_0, x_1, \ldots, x_n]$ is *saturated* if

$$I \;=\; (I : \mathfrak{m}_0) \;:=\; \{f \mid x_i f \in I \text{ for } i = 0, 1, \ldots, n\}\,.$$

The reason for this definition is that I and $(I : \mathfrak{m}_0)$ define the same projective scheme.

3.1.2 Affine covers

Given a projective variety $X \subset \mathbb{P}^n$, we may consider its intersection with any affine open subset $U_i = \{x \in \mathbb{P}^n \mid x_i \neq 0\}$. For simplicity of notation, we will work with $U_0 = \{[1, x_1, \ldots, x_n] \mid (x_1, \ldots, x_n) \in \mathbb{A}^n\} \simeq \mathbb{A}^n$. Then

$$X \cap U_0 \;=\; \{a \in U_0 \mid f(a) = 0 \text{ for all } f \in \mathcal{I}(X)\}\,.$$

and

$$\mathcal{I}(X \cap U_0) \;=\; \{f(1, x_1, \ldots, x_n) \mid f \in \mathcal{I}(X)\}\,.$$

We call the polynomial $f(1, x_1, \ldots, x_n)$ the *dehomogenization* of the homogeneous polynomial f. This shows that the ideal of $X \cap U_0$ is obtained by dehomogenizing the polynomials in the ideal of X. Note that f and $x_0^m f$ both dehomogenize to the same polynomial.

Conversely, given an affine subvariety $Y \subset U_0$, we have its Zariski closure $\overline{Y} := \mathcal{V}(\mathcal{I}(Y)) \subset \mathbb{P}^n$. The relation between the ideal of the affine variety Y and homogeneous ideal of its closure $\overline{Y}$ is through homogenization.

$$\begin{aligned}\mathcal{I}(\overline{Y}) &= \{f \in \mathbb{F}[x_0,\dots,x_n] \mid f|_Y = 0\} \\ &= \{f \in \mathbb{F}[x_0,\dots,x_n] \mid f(1,x_1,\dots,x_n) \in \mathcal{I}(Y) \subset \mathbb{F}[x_1,\dots,x_n]\} \\ &= \{x_0^{\deg(g)+m} g(\tfrac{x_1}{x_0},\dots,\tfrac{x_n}{x_0}) \mid g \in \mathcal{I}(Y),\ m \geq 0\}\,.\end{aligned}$$

The point of this is that every projective variety X is naturally a union of affine varieties

$$X \;=\; \bigcup_{i=0}^{n} (X \cap U_i)\,.$$

Furthermore, a subset $Z \subset \mathbb{P}^n$ of projective space is Zariski closed if and only if its intersection with each U_i is closed. This gives a relationship between varieties and manifolds: Affine varieties are to varieties what open subsets of $\mathbb{R}^n$ are to manifolds.

3.1.3 Coordinate rings and maps

Given a projective variety $X \subset \mathbb{P}^n$, its *homogeneous coordinate ring* $\mathbb{F}[X]$ is the quotient

$$\mathbb{F}[X] \;:=\; \mathbb{F}[x_0,x_1,\dots,x_n]/\mathcal{I}(X)\,.$$

If we set $\mathbb{F}[X]_d$ to be the images of all degree d homogeneous polynomials, $\mathbb{F}[x_0,\dots,x_n]_d$, then this ring is graded,

$$\mathbb{F}[X] \;=\; \bigoplus_{d\geq 0} \mathbb{F}[X]_d\,,$$

where if $f \in \mathbb{F}[X]_d$ and $g \in \mathbb{F}[X]_e$, then $fg \in \mathbb{F}[X]_{d+e}$. More concretely, we have

$$\mathbb{F}[X]_d \;=\; \mathbb{F}[x_0,\dots,x_n]_d/\mathcal{I}(X)_d\,,$$

where $\mathcal{I}(X)_d = \mathcal{I}(X) \cap \mathbb{F}[x_0,\dots,x_n]_d$.

This differs from the coordinate ring of an affine variety as its elements are not functions on X, as we already observed that, apart from constant polynomials, elements of $\mathbb{F}[x_0,\dots,x_n]$ do not give functions on $\mathbb{P}^n$. However, given two homogeneous polynomials f and g which have the same degree, d, the quotient f/g does give a well-defined function, at least on $\mathbb{P}^n - \mathcal{V}(g)$. Indeed, if $[a_0,\dots,a_n]$ and $[\lambda a_0,\dots,\lambda a_n]$ are two representatives of the point $a \in \mathbb{P}^n$ and $g(a) \neq 0$, then

$$\frac{f(\lambda a_0,\dots,\lambda a_n)}{g(\lambda a_0,\dots,\lambda a_n)} \;=\; \frac{\lambda^d f(a_0,\dots,a_n)}{\lambda^d g(a_0,\dots,a_n)} \;=\; \frac{f(a_0,\dots,a_n)}{g(a_0,\dots,a_n)}\,.$$

It follows that if $f,g \in \mathbb{F}[X]$ with $g \neq 0$, then the quotient f/g gives a well-defined function on $X - \mathcal{V}(g)$.

More generally, let $f_0, f_1, \ldots, f_m \in \mathbb{F}[X]$ be elements of the same degree with at least one f_i non-zero on X. These define a *rational map*

$$\begin{array}{rccl} \varphi : & X & \dashrightarrow & \mathbb{P}^m \\ & x & \longmapsto & [f_0(x), f_1(x), \ldots, f_m(x)]\,. \end{array}$$

This is defined at least on the set $X - \mathcal{V}(f_0, \ldots, f_m)$. A second list $g_0, \ldots, g_m \in \mathbb{F}[X]$ of elements of the same degree (possible different from the degrees of the f_i) defines the same rational map if we have

$$\operatorname{rank} \begin{bmatrix} f_0 & f_1 & \cdots & f_m \\ g_0 & g_1 & \cdots & g_m \end{bmatrix} \;=\; 1 \qquad \text{i.e.} \quad f_i g_j - f_j g_i \;\in\; \mathcal{I}(X) \quad \text{for } i \neq j\,.$$

The map φ is regular at a point $x \in X$ if there is some system of representatives $f_0, \ldots, f_m$ for the map φ for which $x \notin \mathcal{V}(f_0, \ldots, f_m)$. The set of such points is an open subset of X called the *domain of regularity* of φ. The map φ is *regular* if it is regular at all points of X. The *base locus* of a rational map $\varphi\colon X \dashrightarrow Y$ is the set of points of X at which φ is not regular.

Example 3.1.8 An important example of a rational map is a linear projection. Let $\Lambda_0, \Lambda_1, \ldots, \Lambda_m$ be linear forms. These give a rational map φ which is defined at points of $\mathbb{P}^n - E$, where E is the common zero locus of the linear forms $\Lambda_0, \ldots, \Lambda_m$, that is $E = \mathbb{P}(\text{kernel}(L))$, where L is the matrix whose columns are the Λ_i.

The identification of $\mathbb{P}^1$ with the points on the circle $\mathcal{V}(x^2 + (y-1)^2 - 1) \subset \mathbb{A}^2$ from Example 3.1.2 is an example of a linear projection. Let $X := \mathcal{V}(x^2 + (y-z)^2 - z^2)$ be the plane conic which contains the point $[0, 0, 1]$. The identification of Example 3.1.2 was the map

$$\mathbb{P}^1 \;\ni\; [a, b] \;\longmapsto\; [2ab, 2a^2, a^2 + b^2] \;\in\; X\,.$$

Its inverse is the linear projection $[x, y, z] \mapsto [x, y]$.

Figure 3.2 shows another linear projection. Let C be the cubic space curve with parametrization $[1, t, t^2, 2t^3 - 2t]$ and $\pi\colon \mathbb{P}^3 \dashrightarrow L \simeq \mathbb{P}^1$ the linear projection defined by the last two coordinates, $\pi\colon [x_0, x_1, x_2, x_3] \mapsto [x_3, x_4]$. We have drawn the image $\mathbb{P}^1$ in the picture to illustrate that the inverse image of a linear projection is a linear section of the variety (after removing the base locus). The center of projection is a line, E, which meets the curve in a point, B.

Projective varieties $X \subset \mathbb{P}^n$ and $Y \subset \mathbb{P}^m$ are *isomorphic* if we have regular maps $\varphi\colon X \to Y$ and $\psi\colon Y \to X$ for which the compositions $\psi \circ \varphi$ and $\varphi \circ \psi$ are the identity maps on X and Y, respectively.

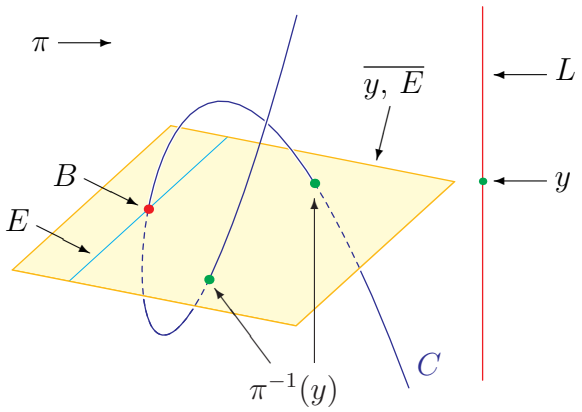


Figure 3.2: A linear projection π with center E.

Exercises

1. Write down the transition functions for $\mathbb{P}^n$ provided by the affine charts $U_0, \ldots, U_n$. Recall that a transition function $\varphi_{i,j}$ expresses how to change from the local coordinates from U_i of a point $p \in U_i \cap U_j$ to the local coordinates from U_j.

2. Show that an ideal I is homogeneous if and only if it is generated by homogeneous polynomials if and only if it has a finite homogeneous Gröbner basis.

3. Let $Z \subset \mathbb{P}^n$. Show that $\mathcal{I}(Z)$ is a homogeneous ideal.

4. Show that a radical homogeneous ideal is saturated.

5. Show that the homogeneous ideal $\mathcal{I}(Z)$ of a subset $Z \subset \mathbb{P}^n$ is equal to the ideal $\mathcal{I}(CZ)$ of the affine cone over Z.

6. Verify the claim in the text concerning the relation between the ideal of an affine subvariety $Y \subset U_0$ and of its Zariski closure $\overline{Y} \subset \mathbb{P}^n$:

$$\mathcal{I}(\overline{Y}) \;=\; \{x_0^{\deg(g)+m} g(\tfrac{x_1}{x_0}, \ldots, \tfrac{x_n}{x_0}) \mid g \in \mathcal{I}(Y) \subset \mathbb{F}[x_1, \ldots, x_n],\ m \geq 0\}\,.$$

7. Let $X \subset \mathbb{P}^n$ be a projective variety and suppose that $f, g \in \mathbb{F}[X]$ are homogeneous forms of the same degree with $g \neq 0$. Show that the quotient f/g gives a well-defined function on $X - \mathcal{V}(g)$.

8. Show that if I is a homogeneous ideal and J is its *saturation*,

$$J \;=\; \bigcup_{d \geq 0} (I : \mathfrak{m}_0^d)\,,$$

then there is some integer N such that

$$J_d \;=\; I_d \qquad \text{for} \quad d \geq N\,.$$

9. Verify the claim in the text that if $X \subset \mathbb{P}^n$ is a projective variety, then its homogeneous coordinate ring is graded with

$$\mathbb{F}[X]_d \;=\; \mathbb{F}[x_0, \ldots, x_n]_d / \mathcal{I}(X)_d \,.$$

3.2 Hilbert functions and degree

The homogeneous coordinate ring $\mathbb{F}[X]$ of a projective variety $X \subset \mathbb{P}^n$ is an invariant of the variety X which determines it up to a linear automorphism of $\mathbb{P}^n$. Basic numerical invariants, such as the dimension of X, are encoded in the combinatorics of $\mathbb{F}[X]$. Another invariant is the degree of X, which is the number of solutions on X to systems of linear equations.

The homogeneous coordinate ring $\mathbb{F}[X]$ of a projective variety $X \subset \mathbb{P}^n$ is a graded ring,

$$\mathbb{F}[X] \;=\; \bigoplus_{m=0}^{\infty} \mathbb{F}[X]_m \,,$$

with degree m piece $\mathbb{F}[X]_m$ equal to the quotient $\mathbb{F}[x_0, x_1, \ldots, x_n]_m / \mathcal{I}(X)_m$. The most basic numerical invariant of this ring is the *Hilbert function* of X, whose value at $m \in \mathbb{N}$ is the dimension of the m-th graded piece of $\mathbb{F}[X]$,

$$\mathrm{HF}_X(m) \;:=\; \dim_{\mathbb{F}}(\mathbb{F}[X]_m) \,.$$

This is also the number of linearly independent degree m homogeneous polynomials on X. We may also define the Hilbert function of a homogeneous ideal $I \subset \mathbb{F}[x_0, \ldots, x_n]$,

$$\mathrm{HF}_I(m) \;:=\; \dim_{\mathbb{F}}\big(\mathbb{F}[x_0, \ldots, x_n]_m / I_m\big) \,.$$

Note that $\mathrm{HF}_X = \mathrm{HF}_{\mathcal{I}(X)}$.

Example 3.2.1 Consider the space curve C of Figure 3.2. This is the image of $\mathbb{P}^1$ under the map

$$\varphi \;:\; \mathbb{P}^1 \ni [s,t] \;\longmapsto\; [s^3, s^2t, st^2, 2t^3 - 2s^2t] \in \mathbb{P}^3 \,.$$

If the coordinates of $\mathbb{P}^3$ are $[w,x,y,z]$, then $C = \mathcal{V}(2y^2{-}xz{-}2yw,\; 2xy{-}2xw{-}zw,\; x^2{-}yw)$. This map has the property that the pullback $\varphi^*(f)$ of a homogeneous form f of degree m is a homogenous polynomial of degree $3m$ in the variables s,t, and all homogeneous forms of degree $3m$ in s,t occur as pullbacks. Since there are $3m+1$ forms of degree $3m$ in s,t, we see that $\mathrm{HF}(m) = 3m+1$.

The Hilbert function of a homogeneous ideal I may be computed using Gröbner bases. First observe that any reduced Gröbner basis of I consists of homogeneous polynomials.

Theorem 3.2.2 *Any reduced Gröbner basis for a homogeneous ideal I consists of homogeneous polynomials.*

Proof. Buchberger's algorithm is friendly to homogeneous polynomials. If f and g are homogeneous, then so is $\mathrm{Spol}(f,g)$. Since Buchberger's algorithm consists of forming S-polynomials and of reductions (a form of S-polynomial), if given homogeneous generators of an ideal, it will compute a reduced Gröbner basis consisting of homogeneous polynomials.

A homogeneous ideal I has a finite generating set B consisting of homogeneous polynomials. Therefore, given a monomial order, Buchberger's algorithm will transform B into a reduced Gröbner basis G consisting of homogeneous polynomials. But reduced Gröbner bases are uniquely determined by the term order, so Buchberger's algorithm will transform any generating set into G. □

A consequence of Theorem 3.2.2 is that it is no loss of generality to use graded term orders when computing a Gröbner basis of a homogeneous ideal. Theorem 3.2.2 also implies that the linear isomorphism of Theorem 2.2.3 between $\mathbb{F}[x_0,\ldots,x_n]/I$ and $\mathbb{F}[x_0,\ldots,x_n]/\mathrm{in}(I)$ respects degree and so the Hilbert functions of I and of $\mathrm{in}(I)$ agree.

Corollary 3.2.3 *Let I be a homogeneous ideal. Then $\mathrm{HF}_I(m) = HF_{\mathrm{in}(I)}$, the number of standard monomials of degree m.*

Proof. The image in $\mathbb{F}[\mathbb{P}^n]/I$ of a standard monomial of degree m lies in the mth graded piece. Since the images of standard monomials are linearly independent, we only need to show that they span the degree m graded piece of this ring. Let $f \in \mathbb{F}[\mathbb{P}^n]$ be a homogeneous form of degree m and let G be a reduced Gröbner basis for I. Then the reduction f mod G is a linear combination of standard monomials. Each of these will have degree m as G consists of homogeneous polynomials and the division algorithm is homogeneous-friendly. □

Example 3.2.4 In the degree-reverse lexicographic monomial order where $x \succ y \succ z \succ w$, the polynomials

$$\underline{2y^2} - xz - 2yw\underline{2xy} - 2xw - zw,\ \underline{x^2} - yw,$$

form the reduced Gröbner basis for the ideal of the cubic space curve C of Example 3.2.1. As the underlined terms are the initial terms, the initial ideal is the monomial ideal $\langle y^2, xy, x^2\rangle$.

The standard monomials of degree m are exactly the set

$$\{z^a w^b,\ xz^c w^d,\ yz^c w^d \mid a+b=m,\ c+d=m-1\}$$

and so there are exactly $m+1+m+m = 3m$ standard monomials of degree m. This agrees with the Hilbert function of C, as computed in Example 3.2.1.

Thus we need only consider monomial ideals when studying Hilbert functions of arbitrary homogeneous ideals. Once again we see how some questions about arbitrary ideals may be reduced to the same questions about monomial ideals, where we may use combinatorics.

Because an ideal and its saturation both define the same projective scheme, and because Hilbert functions are difficult to compute, we introduce the Hilbert polynomial.

Definition 3.2.5 Two functions $f, g\colon \mathbb{N} \to \mathbb{N}$ are *stably equivalent*, $f \sim g$, if $f(d) = g(d)$ for d sufficiently large.

The following result may be proved purely combinatorially.[†]

[†]A proof is given in [5].

Proposition-Definition 3.2.6 *The Hilbert function of a monomial ideal I is stably equivalent to a polynomial, HP_I, called the Hilbert polynomial of I.*

The degree of HP_I is the dimension of the largest linear subspace contained in $\mathcal{V}(I)$.

Corollary 3.2.7 *If I is a monomial ideal, then the Hilbert polynomials HP_I and $\mathrm{HP}_{\sqrt{I}}$ have the same degree.*

Together with Corollary 3.2.3, we may define the Hilbert polynomial of any homogeneous ideal I and the Hilbert polynomial HP_X of any projective variety $X \subset \mathbb{P}^n$. The Hilbert polynomial of a projective variety encodes virtually all of its numerical invariants. We explore two such invariants.

Definition 3.2.8 Let $X \subset \mathbb{P}^n$ be a projective variety and suppose that the initial term of its Hilbert polynomial is

$$\mathrm{in}(\mathrm{HP}_X(t)) \;=\; a\frac{t^d}{d!}\,.$$

Then the *dimension of X* is the degree, d, of the Hilbert polynomial and the number a is the *degree of X*.

We computed the Hilbert function of the curve C of Example 3.2.1 to be $3m+1$. This is also its Hilbert polynomial, as we see that C has dimension 1 and degree 3, which justifies our calling it a cubic space curve.

We may similarly define the dimension and degree of a homogeneous ideal I.

Example 3.2.9 In Exercise 4 you are asked to show that if X consists of a distinct points, then the Hilbert polynomial of X is the constant, a. Thus X has dimension 0 and degree a.

Suppose that X is a linear space, $\mathbb{P}(V)$, where $V \subset \mathbb{F}^{n+1}$ has dimension $d+1$. We may choose coordinates $x_0, \ldots, x_n$ on $\mathbb{P}^n$ so that V is defined by $x_{d+1} = \cdots = x_n = 0$, and so $\mathbb{F}[X] \simeq \mathbb{F}[x_0, \ldots, x_d]$. Then $\mathrm{HF}_X(m) = \binom{m+d}{d}$, which has initial term $\frac{m^d}{d!}$ and so X has dimension d and degree 1.

Suppose that $I = \langle f \rangle$, where f is homogeneous of degree a. Then

$$\big(\mathbb{F}[x_0, \ldots, x_n]/I\big)_m \;=\; \frac{\mathbb{F}[x_0, \ldots, x_n]_m}{f \cdot \mathbb{F}[x_0, \ldots, x_n]_{m-a}}\,,$$

so that if $m > 0$ we have $\mathrm{HF}_I(m) = \binom{m+n}{n} - \binom{m-a+n}{n}$. Thus the leading term of the Hilbert polynomial of I is $a\frac{m^{n-1}}{(n-1)!}$, and so I has dimension $n-1$ and degree a. When f is square-free, so that $I = \mathcal{I}(\mathcal{V}(f))$, we see that the hypersurface defined by f has dimension $n-1$ and degree equal to the degree of f.

Theorem 3.2.10 *Let X be a subvariety of $\mathbb{P}^n$ and suppose that $f \in \mathbb{F}[X]_d$ has degree d and is not a zero divisor. Then the ideal $\langle \mathcal{I}(X), f \rangle$ has dimension $\dim(X) - 1$ and degree $d \cdot \deg(X)$.*

Proof. For $m \geq d$, the degree m piece of the quotient ring $\mathbb{F}[x_0, \ldots, x_n]/\langle \mathcal{I}(X), f \rangle$ is the quotient

$$\mathbb{F}[X]_m / f \cdot \mathbb{F}[X]_{m-d} \,,$$

and so it has dimension $\dim_{\mathbb{F}}(\mathbb{F}[X]_m) - \dim_{\mathbb{F}}(\Lambda \cdot \mathbb{F}[X]_{m-d})$.

Suppose that m is large enough so that the Hilbert function of X is equal to its Hilbert polynomial at $m-d$ and all larger integers. Since f is not a zero divisor, multiplication by f is injective. Thus this dimension is

$$\mathrm{HP}_X(m) - \mathrm{HP}_X(m-d) \,.$$

which is a polynomial of degree $\dim(X)-1$ and leading coefficient $d \cdot \deg(X)/(\dim(X)-1)!$, as you are asked to show in Exercise 3. □

Lemma 3.2.11 *Suppose that X is a projective variety of dimension d and degree a. All subvarieties of X have dimension at most d and at least one irreducible component of X has dimension d, and a is the sum of the degrees of the irreducible components of dimension d.*

Furthermore, if X is irreducible, then every proper subvariety has dimension at most $d-1$ and X has a subvariety of dimension $d-1$.

Proof. Let Y be a subvariety of X. Then the coordinate ring of Y is a quotient of the coordinate ring of X, so $HF_Y(m) \leq HF_X(m)$ for all m, which shows that the degree of the Hilbert polynomial of Y is bounded above by the degree of the Hilbert polynomial of X.

Suppose that $X = X_1 \cup \cdots \cup X_r$ is the decomposition of X into irreducible components. Consider the map of graded vector spaces which is induced by restriction

$$\mathbb{F}[X] \;\longrightarrow\; \mathbb{F}[X_1] \oplus \mathbb{F}[X_2] \oplus \cdots \oplus \mathbb{F}[X_r] \,.$$

This is injective, which gives the inequality

$$\mathrm{HF}_X(m) \;\leq\; \sum_{i=1}^{r} \mathrm{HF}_{X_i}(m) \,.$$

Thus at least one irreducible component must have dimension d. Fill in the argument about the sum.

Suppose now that X is irreducible, let Y be a proper subvariety of X and let $0 \neq f \in \mathcal{I}(Y) \subset \mathbb{F}[X]$. Since $\mathbb{F}[X]/\langle f \rangle \twoheadrightarrow \mathbb{F}[X]/\mathcal{I}(Y) = \mathbb{F}[Y]$, we see that the Hilbert polynomial of $\mathbb{F}[Y]$ has degree at most that of $\mathbb{F}[X]/\langle f \rangle$, which is $d-1$.

Let $I = \langle \mathcal{I}(X), f \rangle$, where we write f both for the element $f \in \mathcal{I}(Y)$ and a homogeneous polynomial which restricts to it. If I is radical, then we have just shown that $\mathcal{V}(I) \subset X$ is a subvariety of dimension $d-1$. Otherwise, let $\succ$ be a monomial order, and we have the chain of inclusions

$$\mathrm{in}(I) \;\subset\; \mathrm{in}(\sqrt{I}) \;\subset\; \sqrt{\mathrm{in}(I)} \,, \tag{3.2}$$

and thus,

$$\deg(\mathrm{HP}_I) \;=\; \deg(\mathrm{HP}_{\mathrm{in}(I)}) \;\geq\; \deg(\mathrm{HP}_{\sqrt{I}}) \;\geq\; \deg(\mathrm{HP}_{\sqrt{\mathrm{in}(I)}})\,.$$

Since $\mathrm{HP}_{\mathrm{in}(I)}$ and $\mathrm{HP}_{\sqrt{\mathrm{in}(I)}}$ have the same degree and $\sqrt{I}$ is the ideal of $\mathcal{V}(I)$, we conclude that $\mathcal{V}(I)$ is a subvariety of X having dimension $d-1$. □

We may now show that the combinatorial definition (Definition 1.4) of dimension is correct.

Corollary 3.2.12 (Combinatorial definition of dimension) *The dimension of a variety X is the length of the longest decreasing chain of irreducible subvarieties of X. If*

$$X \;\supset\; X_0 \;\supsetneq\; X_1 \;\supsetneq\; X_2 \;\supsetneq\; \cdots \;\supsetneq\; X_m \;\supsetneq\; \emptyset\,,$$

is such a chain of maximal length, then X has dimension m.

Proof. Suppose that

$$X \;\supset\; X_0 \;\supsetneq\; X_1 \;\supsetneq\; X_2 \;\supsetneq\; \cdots \;\supsetneq\; X_m \;\supsetneq\; \emptyset$$

is a chain of irreducible subvarieties of a variety X. By Lemma 3.2.11 $\dim(X_{i-1}) > \dim(X_i)$ for $i = 1, \ldots, m$, and so $\dim(X) \geq \dim(X_0) \geq m$.

For the other inequality, we may assume that X_0 is an irreducible component of X with $\dim(X) = \dim(X_0)$. Since X_0 has a subvariety X_1' with dimension $\dim(X_0) - 1$, we may let X_1 be an irreducible component of X' with the same dimension. In the same fashion, for each $i = 2, \ldots, \dim(X)$, we may construct an irreducible subvariety X_i of dimension $\dim(X)-i$. This gives a chain of irreducible subvarieties of X of length $\dim(X)+1$, which proves the combinatorial definition of dimension. □

A consequence of a Bertini's Theorem[†] is that if X is a projective variety, then for almost all homogeneous polynomials f of a fixed degree, $\langle \mathcal{I}(X), f\rangle$ is radical and f is not a zero-divisor in $\mathbb{F}[X]$.

Consequently, if Λ is a generic linear form and set $Y := \mathcal{V}(\Lambda) \cap X$, then $\mathcal{I}(Y) = \langle \mathcal{I}(X), \Lambda\rangle$, and so

$$\mathrm{HP}_Y \;=\; \mathrm{HP}_{\langle \mathcal{I}(X),\Lambda\rangle}\,,$$

and so by Theorem 3.2.10, $\deg(Y) = \deg(X)$. If $Y \subset \mathbb{P}^n$ has dimension d, then we say that Y has *codimension* $n - d$.

Corollary 3.2.13 (Geometric meaning of degree) *The degree of a projective variety $X \subset \mathbb{P}^n$ of dimension d is the number of points in an intersection*

$$X \cap L\,,$$

where $L \subset \mathbb{P}^n$ is a generic linear subspaces of codimension d.

For example, the cubic curve of Figure 3.2 has degree 3, and we see in that figure that it meets the plane $z = 0$ in 3 points.

[†]Not formulated here, yet!

Exercises

1. Show that the dimension of the space $\mathbb{F}[x_0, \ldots, x_n]_m$ of homogeneous polynomials of degree m is $\binom{m+n}{n} = \frac{m^n}{n!} +$ lower order terms in m.

2. Let I be a homogeneous ideal. Show that $HF_I \sim HF_{(I\colon \mathfrak{m}_0)} \sim HF_{I_{\geq d}}$.

3. Suppose that $f(t)$ is a polynomial of degree d with initial term $a_0 t^d$. Show that $f(t) - f(t-1)$ has initial term $ma_0 t^{m-1}$. Show that $f(t) - f(t-b)$ has initial term $mba_0 t^{m-1}$.

4. Show that if $X \subset \mathbb{P}^n$ consists of a points, then, for m sufficiently large, we have $\mathbb{F}[X]_m \simeq \mathbb{F}^a$, and so $HP_X(t) = a$.

5. Compute the Hilbert functions and polynomials the following projective varieties. What are their dimensions and degrees?

   (a) The union of three skew lines in P^3, say $\mathcal{V}(x-w, y-z) \cup \mathcal{V}(x+w, y+z) \cup \mathcal{V}(y-w, x+z)$, whose ideal has reduced Gröbner basis

   $$\langle \underline{x^2} + y^2 - z^2 - w^2,\ \underline{y^2 z} - xz^2 - z^3 + xyw + yzw - zw^2,\ \underline{xyz} - y^2 w - xzw + yw^2,$$
   $$\underline{y^3} - yz^2 - y^2 w + z^2 w,\ \underline{xy^2} - xyw - yzw + zw^2 \rangle$$

   (b) The union of two coplanar lines and a third line not meeting the first two, say the x- and y-axes and the line $x = y = 1$.

   (c) The union of three lines where the first meets the second but not the third and the second meets the third. For example $\mathcal{V}(wy, wz, xz)$.

   (d) The union of three coincident lines, say the x-, y-, and z- axes.

3.3 Toric ideals

Toric ideals are ideals of toric varieties, both of which play a special role in applications of algebraic geometry. We begin with some interesting geometry of polynomial equations. Here, we work with *Laurent polynomials*, which are polynomials whose monomials may have both positive and negative exponents.

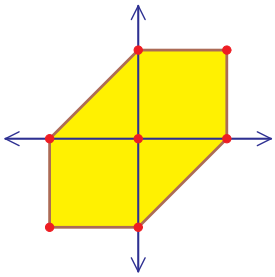
Let $\mathcal{A} \subset \mathbb{Z}^n$ be a finite collection of exponent vectors for Laurent monomials in $x_1, \ldots, x_n$. For example, when $n = 4$ the exponent vector $\alpha = (2, -3, -5, 1)$ corresponds to $x_1^2 x_2^{-3} x_3^{-5} x_4$. A *(Laurent) polynomial* f with *support* $\mathcal{A}$ is a linear combination of (Laurent) monomials,

$$f \;=\; \sum_{\alpha \in \mathcal{A}} c_\alpha x^\alpha \,. \tag{3.3}$$

For example, the polynomial

$$f \;:=\; 1 + 2x + 3xy + 5y + 7x^{-1} + 11x^{-1}y^{-1} + 13y^{-1}$$

has support $\mathcal{A}$ consisting of the integer points in the hexagon



$$\mathcal{A} \;=\; \begin{pmatrix} 0 & 1 & 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 & -1 & -1 \end{pmatrix} .$$

The polynomial f (3.3) lies in the ring of Laurent polynomials, $\mathbb{F}[x_1, x_1^{-1}, \ldots, x_n, x_n^{-1}]$, which is the coordinate ring of the *algebraic torus*

$$\begin{aligned} (\mathbb{F}^\times)^n \;&:=\; \{x \in \mathbb{F}^n \mid x_1 \cdots x_n \neq 1\} \\ &\simeq\; \mathcal{V}(x_1 \cdots x_n x_{n+1} - 1) \;\subset\; \mathbb{A}^{n+1} \end{aligned}$$

While it is often convenient to use the elements of $\mathcal{A}$ as indices, when the elements of $\mathcal{A}$ are in a list ($\mathcal{A} = \{\alpha_1, \ldots, \alpha_m\}$), then we will use the indices $1, \ldots, m$. For example, if we set $c_i := c_{\alpha_i}$, then the polynomial f in (3.3) becomes $\sum_{i=1}^m c_i x^{\alpha_i}$.

Given the set $\mathcal{A} \subset \mathbb{Z}^n$, we define a map,

$$\begin{aligned} \varphi_{\mathcal{A}} \,:\, (\mathbb{F}^\times)^n \;&\longrightarrow\; \mathbb{P}^{\mathcal{A}} := [y_\alpha \mid \alpha \in \mathcal{A}] \\ x \;&\longmapsto\; [x^\alpha \mid \alpha \in \mathcal{A}] \,. \end{aligned}$$

One reason that we consider this map is that it turns non-linear polynomials on $(\mathbb{F}^\times)^n$ into linear equations on $\varphi((\mathbb{F}^\times)^n)$. Let

$$\Lambda \;=\; \Lambda(y) \;:=\; \sum_{\alpha \in \mathcal{A}} c_\alpha y_\alpha$$

be a linear form on $\mathbb{P}^{\mathcal{A}}$. Then the pullback

$$\varphi_{\mathcal{A}}^*(\Lambda) \;=\; \sum_{\alpha\in\mathcal{A}} c_\alpha x^\alpha$$

is a polynomial with support $\mathcal{A}$. This construction gives a correspondence

$$\left\{\begin{array}{c}\text{Polynomials } f \text{ on}\\ (\mathbb{F}^\times)^n \text{ with support } \mathcal{A}\end{array}\right\} \iff \left\{\begin{array}{c}\text{Linear forms } \Lambda \text{ in } \mathbb{P}^{\mathcal{A}}\\ \text{on } \varphi_{\mathcal{A}}((\mathbb{F}^\times)^n)\end{array}\right\}$$
$$f \iff \varphi_{\mathcal{A}}^*(\Lambda)$$

In this way, a system of polynomials

$$f_1(x_1,\ldots,x_n) \;=\; f_2(x_1,\ldots,x_n) \;=\; \cdots \;=\; f_n(x_1,\ldots,x_n) \;=\; 0 \tag{3.4}$$

where each polynomial f_i has support $\mathcal{A}$ corresponds to a system of linear equations

$$\Lambda_1(y) \;=\; \Lambda_2(y) \;=\; \cdots \;=\; \Lambda_n(y) \;=\; 0$$

on $\varphi_{\mathcal{A}}((\mathbb{F}^\times)^n)$. Our approach to study these linear equations is to replace $\varphi_{\mathcal{A}}((\mathbb{F}^\times)^n)$ by its Zariski closure.

Definition 3.3.1 The *toric variety* $X_{\mathcal{A}} \subset \mathbb{P}^{\mathcal{A}}$ is the closure of the image $\varphi_{\mathcal{A}}((\mathbb{F}^\times)^n)$ of $\varphi_{\mathcal{A}}$. The *toric ideal* $I_{\mathcal{A}}$ is the ideal of the toric variety $X_{\mathcal{A}}$. It consists of all homogeneous polynomials which vanish on $\varphi_{\mathcal{A}}((\mathbb{F}^\times)^n)$.

The map $\varphi_{\mathcal{A}}\colon (\mathbb{F}^\times)^n \to X_{\mathcal{A}}$ parametrizes $X_{\mathcal{A}}$.

Corollary 3.3.2 *The number of solutions to a system of polynomials with support $\mathcal{A}$ (3.4) is at most the degree of the toric variety $X_{\mathcal{A}}$. When $\mathbb{F}$ is algebraically closed, it equals this degree when the polynomials are generic given their support $\mathcal{A}$.*

Proof. This follows from the geometric interpretation of degree of a projective variety (Corollary 3.2.13). The only additional argument that is needed is that the difference $\partial(X_{\mathcal{A}}) := X_{\mathcal{A}} - \varphi_{\mathcal{A}}((\mathbb{F}^\times)^n)$ has dimension less than n, and so a generic linear space of codimension n will not meet $\partial(X_{\mathcal{A}})$. □

It will greatly aid our discussion if we homogenize the map $\varphi_{\mathcal{A}}$.

$$\begin{array}{rcl}\mathbb{F}^\times \times (\mathbb{F}^\times)^n & \longrightarrow & \mathbb{P}^{\mathcal{A}}\\ (t,x) & \longmapsto & [tx^\alpha \mid \alpha\in\mathcal{A}]\end{array}$$

We can regard this homogenized version of $\varphi_{\mathcal{A}}$ as a map to $\mathbb{F}^{\mathcal{A}}$ whose image is equal to the cone over $\varphi_{\mathcal{A}}((\mathbb{F}^\times)^n)$ with the origin removed. Since $X_{\mathcal{A}}$ is projective, this does not change the image of $\varphi_{\mathcal{A}}$. The easiest way to do ensure that $\varphi_{\mathcal{A}}$ is homogeneous is to

assume that every vector in $\mathcal{A}$ has first coordinate 1, or more generally, to assume that the row space of $\mathcal{A}$ contains the vector all of whose components are 1.

For example, suppose that $\mathcal{A}$ consists of the 7 points in $\mathbb{Z}^2$ which are columns of the 2×7 matrix (also written $\mathcal{A}$),

$$\mathcal{A} = \begin{pmatrix} -1 & -1 & 0 & 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 & -1 & 0 \end{pmatrix}.$$

Then $\mathcal{A}$ is the integer points (in red) in the hexagon on the left of Figure 3.3. The

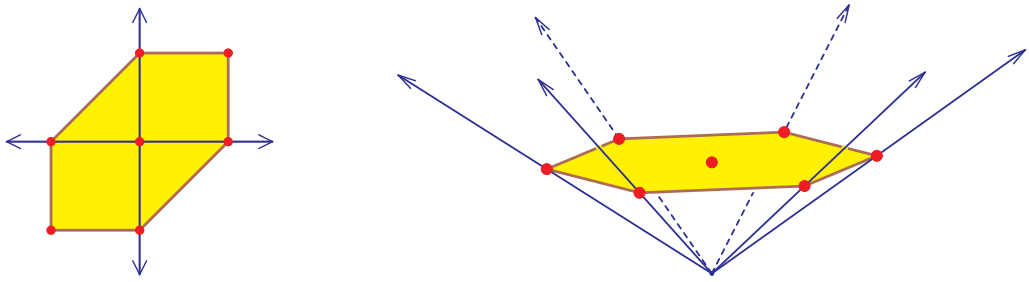


Figure 3.3: The hexagon and its lift

homogenized version of the hexagon is the column vectors of the 3×7 matrix,

$$\mathcal{A}^+ = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 0 & 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 & -1 & 0 \end{pmatrix}.$$

These are the red points in the picture on the right of Figure 3.3. (There, the first coordinate is vertical.)

Given such a homogeneous configuration of integer vectors $\mathcal{A}$ we have the associated map $\varphi_{\mathcal{A}}$ parametrizing the toric variety $X_{\mathcal{A}}$. The pull back of $\varphi_{\mathcal{A}}$ is the map

$$\begin{array}{rccc} \varphi_{\mathcal{A}}^* : & \mathbb{F}[y_\alpha \mid \alpha \in \mathcal{A}] & \longrightarrow & \mathbb{F}[x_1, x_1^{-1}, \ldots, x_n, x_n^{-1}] \\ & y_\alpha & \longmapsto & x^\alpha \end{array}$$

Its kernel is the toric ideal $I_{\mathcal{A}}$.

Theorem 3.3.3 *The toric ideal $I_{\mathcal{A}}$ is a prime ideal.*

Proof. The image of $\varphi_{\mathcal{A}}^*$ is a subring of $\mathbb{F}[x_1, x_1^{-1}, \ldots, x_n, x_n^{-1}]$, which is an integral domain. Since the image is isomorphic to the quotient $\mathbb{F}[y_\alpha \mid \alpha \in \mathcal{A}]/\ker \varphi^*$, this quotient is a domain and thus $\ker \varphi^* = I_{\mathcal{A}}$ is prime.

We may also see this as $(\mathbb{F}^\times)^n$ is irreducible, which implies that $X_{\mathcal{A}}$ is irreducible, and therefore its homogeneous ideal is prime. These two arguments are equivalent via the algebraic-geometric dictionary. □

We describe some generating sets for the toric ideal $I_{\mathcal{A}}$. Let $u = (u_\alpha \mid \alpha \in \mathcal{A}) \in \mathbb{N}^{\mathcal{A}}$ be an integer vector. Then

$$\varphi^*_{\mathcal{A}}(y^u) \;=\; \prod_{\alpha\in\mathcal{A}} (x^\alpha)^{u_\alpha} \;=\; x^{\sum u_\alpha \cdot \alpha}\,.$$

If we regard these exponents $u \in \mathbb{Z}^{\mathcal{A}}$ as column vectors, and the elements of $\mathcal{A}$ as the columns of a matrix $\mathcal{A}$ with $n+1$ rows, then

$$\varphi^*_{\mathcal{A}}(y^u) \;=\; x^{\mathcal{A}u}\,.$$

Notice that $\varphi^*_{\mathcal{A}}(y^u) = \varphi^*_{\mathcal{A}}(y^v)$ if and only if $Au = Av$, and thus $y^u - y^v \in I_{\mathcal{A}}$.

Theorem 3.3.4 *The toric ideal $I_{\mathcal{A}}$ is the linear span of binomials of the form*

$$\{y^u - y^v \mid Au = Av\}\,. \tag{3.5}$$

Proof. These binomials certainly lie in $I_{\mathcal{A}}$. Pick a monomial order $\prec$ on $\mathbb{F}[y_\alpha \mid \alpha \in \mathcal{A}]$. Let $f \in I_{\mathcal{A}}$,

$$f \;=\; c_u y^u \;+\; \sum_{v\prec u} c_v y^v \qquad c_u \neq 0\,,$$

so that $\mathrm{in}(f) = c_u y^u$. Then

$$0 \;=\; \varphi^*_{\mathcal{A}}(f) \;=\; c_u x^{\mathcal{A}u} \;+\; \sum_{v\prec u} c_v x^{\mathcal{A}v}\;.$$

There is some $v \prec u$ with $Av = Au$, for otherwise the initial term $c_u x^{\mathcal{A}u}$ is not canceled and $\varphi^*_{\mathcal{A}}(f) \neq 0$. Set $\overline{f} := f - c_u(y^u - y^v)$. Then $\varphi^*_{\mathcal{A}}(\overline{f}) = 0$ and $\mathrm{in}(\overline{f}) \prec \mathrm{in}(f)$.

If the leading term of f were $\prec$-minimal in $\mathrm{in}(I_{\mathcal{A}})$, then $\overline{f}$ must be zero, and so f is a linear combination of binomials of the form (3.5). Suppose by way of induction that every polynomial in $I_{\mathcal{A}}$ whose initial term is less that that of f is a linear combination of binomials of the form (3.5). Then $\overline{f}$ is a linear combination of binomials of the form (3.5), which implies that f is as well. □

Theorem 3.3.4 gives an infinite generating set for $I_{\mathcal{A}}$. We seek smaller generating sets. Suppose that $Au = Av$ with $u, v \in \mathbb{N}^{\mathcal{A}}$. Set

$$\begin{aligned} t_\alpha &:= \min(u_\alpha, v_\alpha) \\ w^+_\alpha &:= \max(u_\alpha - v_\alpha, 0) \\ w^-_\alpha &:= \max(v_\alpha - u_\alpha, 0) \end{aligned}$$

Then $u - v = w^+ - w^-$ and $u = t + w^+$ and $v = t + w^-$, and

$$y^t(y^{w^+} - y^{w^-}) \;=\; y^u - y^v \;\in\; I_{\mathcal{A}}\,.$$

For $w \in \mathbb{Z}^{\mathcal{A}}$, let w^+ be the coordinatewise maximum of w and the 0-vector, and let w^- be the coordinatewise maximum of $-w$ and the 0-vector.

Corollary 3.3.5 $I_{\mathcal{A}} = \langle y^{w^+} - y^{w^-} \mid \mathcal{A}w = 0 \rangle$.

Thus $I_{\mathcal{A}}$ is generated by binomials coming from the integer kernel of the matrix $\mathcal{A}$.

Theorem 3.3.6 *Any reduced Gröbner basis of $I_{\mathcal{A}}$ consists of binomials.*

The point is that Buchberger's algorithm is binomial-friendly, in the same sense as it was homogeneous-friendly in the proof of Theorem 3.2.2. We omit the nearly verbatim proof.

In practice, Buchberger's algorithm is not the best way to compute a Gröbner basis for a toric ideal. Here is an alternative method due to Hoşten and Sturmfels [11]. We remark that there are other, often superior algorithms available, for example the project-and-lift algorithm of Hemmecke and Malkin [9] which is implemented in the software `4ti2`.

Let A be an integer matrix whose row space contains the vector all of whose components are 1. This is a map from $\mathbb{Z}^{\mathcal{A}}$ to $\mathbb{Z}^{n+1}$ whose kernel is a free abelian subgroup of $\mathbb{Z}^{\mathcal{A}}$. Let $\mathcal{W} := \{w_1, \ldots, w_\ell\}$ be a $\mathbb{Z}$-basis for the kernel of $\mathcal{A}$, and set

$$I_{\mathcal{W}} \;:=\; \{y^{w_i^+} - y^{w_i^-} \mid i = 1, \ldots, \ell\}\,.$$

Then $I_{\mathcal{W}}$ defines the image $\varphi((\mathbb{F}^\times)^n)$ as a subvariety of the dense torus in $\mathbb{P}^{\mathcal{A}}$,

$$(\mathbb{F}^\times)^{|\mathcal{A}-1|} \;=\; \{y \in \mathbb{P}^{\mathcal{A}} \mid \prod_{\alpha} y_\alpha \neq 0\}\,.$$

Thus the ideal $I_{\mathcal{W}}$ agrees with $I_{\mathcal{A}}$ off the coordinate axes of $\mathbb{P}^{\mathcal{A}}$.

The idea behind this method of Hoşten and Sturmfels is to use saturation to pass from the ideal $I_{\mathcal{W}}$ to the full toric ideal $I_{\mathcal{A}}$. Let $\alpha \in \mathcal{A}$, and consider the saturation of an ideal I with respect to the variable y_α,

$$(I \;:\; y_\alpha^\infty) \;:=\; \{f \mid y_\alpha^m f \in I \text{ for some } m\}\,.$$

Note that in the affine set $U_\alpha = \mathbb{P}^{\mathcal{A}} - \mathcal{V}(y_\alpha)$ the two ideals agree, $\mathcal{V}(I) = \mathcal{V}(I : y_\alpha^\infty)$.

Lemma 3.3.7 *Let I be a homogeneous ideal. Then $\mathcal{V}(I : y_\alpha^\infty) = \overline{\mathcal{V}(I) - \mathcal{V}(y_\alpha)}$.*

Proof. We have that $\overline{\mathcal{V}(I) - \mathcal{V}(y_\alpha)} \subset \mathcal{V}(I : y_\alpha^\infty)$ as $\mathcal{V}(I) = \mathcal{V}(I : y_\alpha^\infty)$ in $U_\alpha = \mathbb{P}^{\mathcal{A}} - \mathcal{V}(y_\alpha)$. For the other inclusion, let x be a point of $\mathcal{V}(y_\alpha)$ which lies in $\mathcal{V}(I : y_\alpha^\infty)$ and let f be a homogeneous polynomial which vanishes on $\mathcal{V}(I) - \mathcal{V}(y_\alpha)$. We show that $f(x) - 0$. Then $y_\alpha f$ vanishes on $\mathcal{V}(I)$. By the Nullstellensatz, there is some m such that $y_\alpha^m f^m \in I$, and so $f^m \text{in}(I : y_\alpha^\infty)$. But then $f^m(x) = 0$ and so $f(x) = 0$. □

Lemma 3.3.7 gives an algorithm to compute $I_{\mathcal{A}}$, namely, first compute a $\mathbb{Z}$-basis $\mathcal{W}$ for the kernel of $\mathcal{A}$ to obtain the ideal $I_{\mathcal{W}}$. Next, saturate this with respect to a variable y_α, and then saturate the result with respect to a second variable, and repeat.

Exercises for Section 3.3

1. Let $\mathcal{A} \subset \mathbb{Z}^n$ be a finite collection of exponent vectors for Laurent monomials. Show that the map $\varphi_{\mathcal{A}} \colon (\mathbb{F}^\times)^n \to \mathbb{P}^{\mathcal{A}}$ is injective if and only if $\mathcal{A}$ affinely spans $\mathbb{Z}^n$. (That is, if the differences $\alpha - \beta$ for $\alpha, \beta \in \mathcal{A}$ linearly span $\mathbb{Z}^n$.)

2. Show that $\mathbb{F}[x_1, x_1^{-1}, x_2, x_2^{-1}, \ldots, x_n, x_n^{-1}] \simeq \mathbb{F}[X]$, where $X = \mathcal{V}(x_1 \cdots x_n x_{n+1} - 1) \subset \mathbb{F}^{n+1}$.

3. Complete the implied proof of Theorem 3.3.6.

3.4 Toric varieties

Intro

I'm going be talking about projective toric varieties because I have a specific goal in mind. However, some of the things (with extra work) can translate to toric varieties as well.

3.5 Projective Toric Varieties

Before, I took $\mathcal{A} \subset \mathbb{Z}^n$ and I looked at the affine hull of $\mathcal{A}$. Now, I'm going to take

$$P := \text{conv}(\mathcal{A}) \;=\; \{\sum \lambda_\alpha \cdot \alpha : \sum \lambda_\alpha = 1\}.$$

This forms a *polytope*. Draw some polytopes

Polytopes have faces of various dimensions: the codimension 1 faces are caled *facets*. The zero-dimensional ones are called *vertices*. Facets have the nice feature that there's a normal vector to them.

Sometimes, I'll write $X_\mathcal{A}$ or X_P, and what I say really just depends upon these structures here. I assume the affine span of $\mathcal{A}$ is $\mathbb{Z}^n$, and I might as well assume that $0 \in \mathcal{A}$. Now I consider a map

$$\varphi_\mathcal{A}(\mathbb{F}^\times)^n \to \mathbb{P}^\mathcal{A}$$

defined as

$$x \mapsto [x^{\beta+\alpha} : \alpha \in \mathcal{A}]$$

With β, I just shift the vector, and so I'm not really doing anything at all. If I multiply a whole polynomial by a given monomial, I haven't changed the zero-set of my polynomial by one bit. So this map doesn't depend on $\mathcal{A}$, it depends on $\mathcal{A}$ upto translation. With zero in there, it has the nice fact that $x^0 = 1$ and so we have the map to projective coordinates of the form

$$[1, x_1^{\alpha_1}, \cdots].$$

This is just a sketch of a theory.

Then $(\mathbb{F}^\times)^n$ acts on $\mathbb{P}^\mathcal{A}$. Here, $x \in (\mathbb{F}^\times)^n$ hits $y \in \mathbb{P}^\mathcal{A}$ with the *diagonal action*

$$xy = [x^\alpha y_\alpha : \alpha \in \mathcal{A}.$$

Lemma 3.5.1 *$\varphi_\mathcal{A}((\mathbb{F}^\times)^n)$ is the orbit $(\mathbb{F}^\times)^n$ on $[1, 1, \cdots, 1]$.*

This implies that the algebraic torus $(\mathbb{F}^\times)^n$ acts on the toric variety $X_\mathcal{A}$. Moreover, $\varphi_\mathcal{A}$ maps $(\mathbb{F}^\times)^n$ isomorphically to its image $\varphi_\mathcal{A}((\mathbb{F}^\times)^n)$, and this is a dense subset of X_A. There is an open subset of the image that is the torus. Here, we apply a compactification.

But rather than talk about this in complete generality, let's just look at some simple examples.

Example 3.5.2 Let's suppose we have $\mathcal{A} = \{(0,0),(1,0),(0,1)\}$. Then P is the unit simplex. Then, I have the map

$$(\mathbb{F}^\times)^2 \to \mathbb{P}^2$$
$$(t,u) \mapsto [1,t,u] \subset U_0 \sim \mathbb{A}^2$$

It turns out that $X_\mathcal{A}$ here is just $\mathbb{P}^2$.

I'd like to decompose $\mathbb{P}^2$ according to its orbits. What are the orbits? Well, there's $[1,x,y]$, sets of the form $xy \neq 0$. There's also $[1,x,0]$ and $[1,0,y]$. Another orbit of this action is $[0,1,y]$. Lastly, there's three zero-dimensional orbits, which correspond to the three basis points: $[1,0,0]$, $[0,1,0]$, and $[0,0,1]$.

Let's decompose my polytope P. The whole polytope is a face, but it also has a horizontal edge, a vertical edge, and a diagonal edge. It also has a lower-left corner point, the right vertex, and the top vertex. When I draw these faces, I really mean the (relative) interior of the face, because the boundary is composed of the lower-dimensional stuff. There's a map here from simplectic space that I won't discuss. There's a different map for non-projective varieties that are algebraic. In the projective case, the map is a *moment map*.

This always happens: you can always decompose a toric variety into pieces.

There's a very illustrative example where I take the unit square, but I won't do this. I might exemplify that by looking here. I'm going to show you how the facets glue together.

3.6 The Facets

(I'm going to talk about limits, which correspond with our usual notion in $\mathbb{R}$ or $\mathbb{C}$. In one-dimension, algebraic geometry gives us a notion. In complex analysis, this might be called the *Riemann extension theorem*.)

Now, let F be a facet of P. Let η be the inward-pointing normal (dual) vector. Now η a priori some vector, but I can choose $\eta \in (\mathbb{Z}^n)^*$, but I want it to be *primitive*: If I look at $\mathbb{Q}\eta \cap \mathbb{Z}^n$, then this should just be $\mathbb{Z} \cdot \eta$.

I need something to help me take the limit. Notice that

$$\mathcal{F}^\times \ni t \mapsto (t^{\eta_1}, t^{\eta_2}, \ldots, t^{\eta_n}) \in (\mathbb{F}^\times)^n$$

acts on $\mathbb{P}^\mathcal{A}$ by

$$t \cdot y = [t^{\eta\cdot\alpha} y_\alpha : \alpha \in \mathcal{A}],$$

which is exactly how η is evaluating.

Suppose that $a = \eta(F)$. Then η applied to anything on F should be constant. Notice that if $\alpha \in \mathcal{A}$, $\eta \cdot \alpha \geq a$ and equality holds iff $\alpha \in F$.

What happens if I apply t to $\varphi_{\mathcal{A}}(x)$, with $x \in (\mathbb{F}^\times)^n$? This looks like

$$\begin{aligned} t \cdot \varphi_{\mathcal{A}}(x) &= [t^{\eta_\alpha} \cdot y_\alpha : \alpha \in \mathcal{A}] \\ &= [t^{-a+\eta\cdot\alpha} \cdot x^\alpha : \alpha \in \mathcal{A}] \end{aligned}$$

So as we take the limit as $t \to 0$, $\alpha \notin F$, we get $t^{pos}x^\alpha$, $\alpha \in F$, x^α????????

If F is a facet of P, then I have that $\mathbb{P}^{\mathcal{A}\cap F}$ naturally embeds as a coordinate subspace of $\mathbb{P}^{\mathcal{A}}$. And if I consider the limit

$$\lim_{t\to 0} t\varphi_{\mathcal{A}}(x) = \varphi_{F\cap\mathcal{A}}(x).$$

In summary:

1. X_P is a disjoint union:

$$X_P = \bigcup_{\text{faces } F\subset P} \varphi_{F\cap\mathcal{A}}((\mathbb{F}^\times)^n).$$

   These factors of the coproduct are the orbits. If you take the closures, every face F of P gives rise to a sub toric variety $X_{F\cap\mathcal{A}}$ of $X_{\mathcal{A}}$.

Now, the question is, how do they fit together? I won't get into that, because toric varieties can be singular. It turns out that what really matters (in terms of the singularities) is the angles of how things meet.

3.7 Over $\mathbb{R}$

What about over the real numbers? Let's get even more extreme. Let's say our "field" is $\mathbb{R}_>$. Let

$$X^+ := X_{\mathcal{A}} \cap \Delta_{\mathcal{A}}$$

and this is just the positive part of $\mathbb{RP}^{\mathcal{A}}_2$. This turns out to be the closure of $\varphi_{\mathcal{A}}(\mathbb{R}^n_>)$.

Here's a great fact: This is homeomorphic to the polytope. A polytope is a manifold with boundary. The homeomorphism really respects the angles. You can either go to your symplectic geometry friends and ask them for a moment map, or you can go to friends in algebraic statistics and ask them for what ever they call it.

Luis called it a *tautolagical map*. I call it the *algebraic moment map*. This is the map

$$\tau : \Delta_{\mathcal{A}} \to P$$

Recall that you can identify $\Delta_{\mathcal{A}}$ with the standard unit vectors in $\mathcal{RP}^{\mathcal{A}}$. What do I do? The map is

$$e_\alpha \mapsto \alpha.$$

A typical point in the probability simplex is $[y_\alpha : \alpha \in \mathcal{A}]$ with $y_\alpha \geq 0$ and $\sum y_\alpha = 1$. What does the map do to this? The image is $\sum y_\alpha \cdot \alpha$, and I call this a tautological map for this reason. It's essentially a linear map, but it's really an algebraic moment map. The fact is that under τ, X_A is homeomorphic to P.

This is the positive part of toric variety.

Notice that inside the torus $(\mathbb{R}^\times)^n = \{\pm 1\}^n \times \mathbb{R}^n_>$, if I take $\varepsilon \in \{\pm 1\}^n$, note that ε acts on $\mathbb{RP}^\mathcal{A}$. In turn, I can act with ε on $X_A \cap \Delta_\mathcal{A}$, and this is homeomorphic to $X_A \cap \Delta_\mathcal{A}$. So for all 2^n possible ε, I get a copy of P. What happens to my edges? Some of them have zero coordinates. Each facets have 2^{n-1} elements under this sign group. The way they're glued together come from the signs of the primitive normal vector and looking at its coefficients modulo 2.

3.8 The Punchline

I've talked about the geometry and structure of toric varieties. I was also giving a dictionary between these and polytopes. Now, I'm going to go on to our main question:

How many solutions are there to a system of polynomial equations

$$f_1(x_1, \ldots, x_n) = \cdots = f_n(x_1, \ldots, x_n) = 0$$

where the suppose of $f_i = \mathcal{A}$?

This is called a *sparse system*.

Sometimes these people require that all of these vectors in $\mathcal{A}$ lie in the convex hull, but we don't do that here.

These f_i are polynomials on the algebraic torus $(\mathbb{F}^\times)^n$. Recall that these correspond to linear polynomials on the toric variety $X_\mathcal{A} = \overline{\varphi_\mathcal{A}((\mathbb{F}^\times)^n)}$. Recall that we have dimension-many linear polynomials, namely n of them.

Last time, we said that we can in this situation count the number of solutions. But I want to replace $\varphi_\mathcal{A}((\mathbb{F}^\times)^n)$ by $X_\mathcal{A}$. So, I'm just going to answer Question when they're generic.

We need to determine the Hilbert polynomial. Here's it's relatively easy to determine the degree of HP of $X_\mathcal{A}$. I'll sketch the argument of my bounds.

What I'm going to do is, as before, I'll homogenize my map, but I'll write it a little bit differently:

$$\begin{aligned} \psi : \mathbb{F}^\times \times (\mathbb{F}^\times)^n &\to \mathbb{P}^\mathcal{A} \\ (t, x) &\mapsto [tx^\alpha : \alpha \in \mathcal{A}] \end{aligned}$$

Then

$$\mathbb{F}[X_\mathcal{A}] = \mathbb{F}[y_\alpha]/\ker\psi = \text{image}\psi = \mathbb{F}[tx^\alpha : \alpha \in \mathcal{A}]$$

A typical element in here is $t^{\text{mess}}x^{\text{bigmess}}$. What is the degree of this? It's just "mess".

Let me give you a "geometry of numbers"-way of looking at this. My exponent vectors $\mathcal{A}$ lie inside some $\mathbb{Z}^n$. When I homogenize, they are in $1 \times \mathbb{Z}^n$. So essentially, I have

$$\mathbb{N}\mathcal{A}^+ \leftrightsquigarrow \text{ monomials}$$

Points in $\mathbb{N}A^+$ that are sums of d monomials in $\mathcal{A}^+$ $\mathbb{R}_{>0}\mathcal{A}^+ \cap (\text{first coordinate} = d) = d \cdot P$.

Thus, we have $HF_{\mathcal{A}}(d) \leq$ the Ehrhart polynomial (we have a lattice polytope). This has the form $\text{Vol}(P) \cdot d^n +$ lower order terms in d.

Thus

$$\text{EP}_P(d-m) \leq HP_{\mathcal{A}} \leq \text{EP}_P(d)$$

so my HP looks like

$$n! \cdot \text{Vol}(P) \cdot \frac{d^n}{n!} + \text{ lower order terms}$$

We have Kouchnerenko's Theorem (1976):

Theorem 3.8.1 *The number of non-degenerate solutions to a system of polynomials $f_1 = \cdots = f_n = 0$ with support of each f_i being $\mathcal{A}$ is at most*

$$n!\,\text{Vol}(\text{conv}(\mathcal{A})).$$

When $\mathbb{F}$ is closed and the f_i are generic, then this is equal to $n!\,\text{Vol}(\text{conv}(\mathcal{A}))$.

Then, the question over the real numbers is interesting.

3.9 Bernstein's theorem

You want to solve equations

$$f_1(x_1, \ldots, x_n) = \cdots = f_n(x_1, \ldots, x_n) = 0 \tag{3.6}$$

Note that numerics almost always place you in the generic situation. The polynomial f_i has support $A_i := P_i \cap \mathbb{Z}^n$, where $P_1, \ldots, P_n$ are integer polytopes.

Bernstein's Theorem says:

Theorem 3.9.1 *The number of non-degenerate solutions to (3.6) is less than or equal to the mixed volume* $\text{MixedVol}(P_1, \ldots, P_n)$.

When $\mathbb{F}$ is closed and the f_i are generic (given their support), the number is exactly the mixed volume.

Suppose each P_i is a segment (or rather just two points u_i and v_i). This is the same thing as the segment (0 and $v_i - u_i$). So, I'll assume that P_0 is of the form

$$P_0 = \text{conv}(0, v_i : v_i \in \mathbb{Z}^n)$$

So, we can say $x^{v_i} = c_i$ for $i = 1, \ldots, n$. This set of equations has the form

$$\varphi_V^{-1}(c_1, \ldots, c_n)$$

where

$$\begin{aligned} \varphi_V : (\mathbb{F}^\times)^n &\to (\mathbb{F}^\times)^n \\ x &\mapsto (x^{v_1}, \ldots, (x^{v_i}) \end{aligned}$$

Notice that φ_V is a group homomorphism. Thus, we compute the size:

$$|\varphi_V^{-1}(c)| = |\ker \varphi_V| = \{x : \varphi_V(x) = 1\}$$

It turns out that

$$\ker \varphi_V = \text{Hom}\left(\frac{\mathbb{Z}^n}{\langle v_1, \ldots, v_n \rangle}, \mathbb{F}^\times\right)$$

So, I just need to count this. There are a number of interpretations for this:

$$\left|\frac{\mathbb{Z}^n}{\langle v_1, \ldots, v_n \rangle}\right|$$

This is equal to the volume $\det(v_1|\cdots|v_n)$ of the fundamental parallelepiped Π determined by the vectors v_i:

$$\Pi = \{\sum \lambda_i v_i : 0 \leq \lambda_i \leq 1\} = P_1 + \cdots + P_n$$

the Minkowski sum of the polytopes P_i.

Then, the *mixed volume* is defined as the coefficient of $t_1 \cdots t_n$ in

$$\begin{aligned}
&\mathrm{Vol}(t_1 P_1 + \cdots t_n P_n) \\
&= \det(t_1 v_1 | \cdots | t_n v_n) \\
&= t_1 \cdots t_n \det(v_1 | \cdots | v_n) \\
&= \mathrm{MixedVol}(P_1, \ldots, P_n) \\
&= \text{Number of solutions}
\end{aligned}$$

The permutations $\sigma \in S_n$ index the simplices in a triangulation of the n-cube C_n. Given σ, take the convex hull of $0, 0 + e_{\sigma(1)}, 0 + e_{\sigma(1)} + e_{\sigma(2)}, \ldots, 0 + e_{\sigma(1)} + \cdots + e_{\sigma(n)}$.

3.10 Proof of Bernstein's Theorem

The polynomial f_i has the form

$$f_i = \sum_{\alpha \in A_i} c_{i,\alpha} x^\alpha$$

I will choose $v_{i,\alpha} \in \mathbb{Z}, \alpha \in A_i, i = 1, \ldots, n$. Given this choice, I will define:

$$f_i(x; t) := \sum_{\alpha \in A_i} c_{i,\alpha} t^{\nu_{i,\alpha}} x^\alpha \qquad t \in \mathbb{F}^\times \tag{3.7}$$

When $t = 1$, we get back the original system. This defines some algebraic curve C in $(\mathbb{F}^\times)^n \times \mathbb{F}^\times$. I'd like to look at C near $t = 0$. It will be a little strange because of how the exponents might behave.

3.10.1 Puiseaux series

To work on Puiseaux series, we really need our field to be algebraically closed. So, what I'll do here is prove the equality statement of Theorem 3.9.1.

Puiseaux series are the algebraic closure of the field of formal power series in t. Our formal power series are of the form

$$\sum_{n \geq N} c_n t^n$$

and thus our Puiseaux elements look like

$$\sum_{n \geq N} c_n t^{n/M}.$$

So, what are the Puiseaux solutions to (3.7)? We look for solutions of the form

$$X_i(t) = X_i(0) t^{u_i} + \text{ higher order terms in } t \qquad u_i \in \mathbb{Q}. \tag{3.8}$$

What is $X(t)$? Let's take (coordinate-wise) powers:

$$X(t)^\alpha = X(0)^\alpha t^{u\cdot\alpha} + \text{ higher order terms in } t$$

Now, when I plug this into (3.7), I obtain

$$f_i(X(t);t) = \sum_{\alpha\in A_i} c_{i,\alpha}\big(t^{\nu_{i,\alpha}+u\cdot\alpha}X(0)^\alpha + \text{ higher order terms in } t\big) = 0.$$

I have a system of equations like this. If a power series is zero, that means that all of the coefficients must cancel.

It's necessary[1] that

$$\min_{\alpha\in A_i}\{\nu_{i,\alpha} + U\cdot\alpha\} \tag{3.9}$$

occurs at **least** twice each $i = 1,\dots,n$.

Given (3.9), suppose the occurances are all exactly twice. Then the lowest order terms are binomials of the form (cancelling $t^{\min}$), you get the system

$$c_{i,\alpha}X(0)^\alpha + c_{i,\beta}X(0)^\beta = 0 \tag{3.10}$$

with $\alpha,\beta\in A_i$.

The number of solutions is exactly the volume of the parallelepiped Π, which is the sum of the volumes of line segments.

Now, we go back to the form for X in of our anzatz(sp) in (3.8) and we add a second term.

I need to thus count the number of solutions to the tropical problem (3.9), and for each of those, I need to solve the volume problem (3.10). So, that's what I'm going to do. From here on out, now it's going to be geometric combinatorics.

3.11 Counting the number of solutions in (3.10)

3.11.1 Some Geometric Combinatorics

Each $A_i\subset\mathbb{R}^n$, and I'm going to lift to $\mathbb{R}^{n+1}$ by considering $A_i\ni\alpha\mapsto(\alpha,\nu_{i,\alpha})$. I'll call this lifted set A_i^+.

Now, I'm going to take $P_i^+ := \operatorname{conv}(A_i^+)$, and I'll set $V = P_1^+ + \cdots + P_n^+$. The faces of V are places where a linear functional takes its minimum. These are sums of the corresponding faces of the P_i^+.

The facets whose inward-pointing normal vector has the form $(u,1)$ where $u\in\mathbb{Q}^n$ are called *lower facets*. These project to $\mathbb{R}^n$ to give some polytopes. The form a *polyhedral subdivision*, a union of polyhedra that meet certain intersection criteria. It is a subdivision of the projection of V, and this is of course our original Minkowski sum $P_1+\cdots+P_n$.

[1] No, that can't be right. Is that right?

We are going to try to identify parts of the subdivision with the parallelepipeds, and then compute the mixed volume. The proper term for this is called a *regular mixed subdivision*.

I have a lower facets F with normal vector of the form $(u, 1)$. Then $(u, 1)$ has a minimum face f_i on each P_i^+. Then

$$F = f_1 + \cdots + f_n. \tag{3.11}$$

If you take a point of coordinates $(\alpha, \nu_{i,\alpha})$ and dot this with $(u, 1)$, then you obtain

$$u \cdot \alpha + \nu_{i,\alpha}.$$

Notice that this is exactly the expression in (3.9). That means that on each of these, I need to have at least two of these f_i lifted.

The genericity assumption on these ν values is that if each is an edge, none of them contain two or points on them. It means that if you had a face, then one of them had to have been lifted higher. The point is $(u, 1)$ solves (3.9) exactly when each f_i is an edge with only two $(\alpha, \nu_{i,\alpha})$ on it.

Note (3.11) is a Minkowski sum of edges. So, the projection of the sum of the line segments $(\alpha, \nu_{i,\alpha}) - -(\beta, \nu_{i,\beta})$ are the Minkowski sums of the form $\alpha - -\beta$. Thus, the solutions to (3.9) are the parallelepipeds among the lower faces of $P_1^+, \cdots, P_n^+$ generated by edges $f_1, \ldots, f_n$, an edge in each P_i^+. Projecting one of these lower faces to $\mathbb{R}^n$ gives a parallelepiped generated by the segment $\alpha - -\beta$, where $\alpha, \beta \in A_i$.

I have an exact bijection between solutions to (3.9) and the Minkowski sums of edges. The term for this is *mixed cells*.

In this mixed subdivision, there are two classes of faces:

1. Mixed cells (These are special parallelepipeds. Each edge comes from a different P_i^+).
2. The rest (Each one excludes some P_i^+).

Recall the mixed volume is the coefficient $t_1 \cdots t_n$ in

$$t_1 P_1 + \cdots + t_n P_n.$$

What happens in each case?

1. Mixed cells: $\text{vol} = t_1 \cdots t_n \cdot \text{vol}(t = 1)$.
2. Here, each cell will lack some t_i.

The sum of the volumes of these mixed cells **IS** the mixed volume. (The second case doesn't contribute because some t_i is always missing).

Chapter 4

Unstructured material

Outline:

1. Case Study: Lines tangent to four spheres.

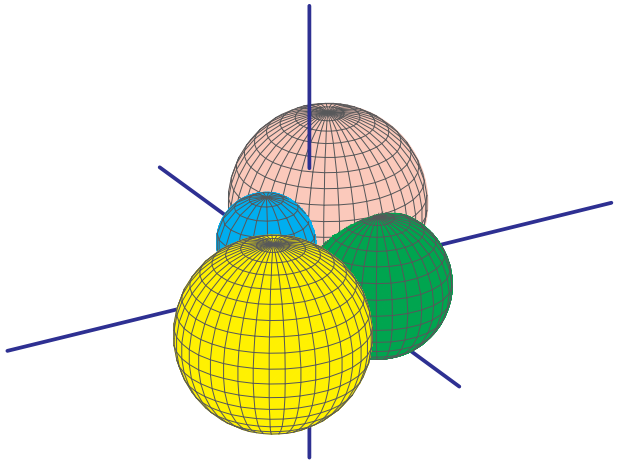
4.1 Lines tangent to four spheres

Methods and ideas from algebraic geometry can be fruitfully applied to solve problems in the ordinary geometry of three-dimensional space. To illustrate this, we consider the following geometrical problem:

How many lines are tangent to four general spheres?

We shall see that this simple problem contains quite a lot of interesting geometry.

We approach this problem computationally, formulating it as a system of equations, and then we shall solve one instance of the question, determining the lines that are tanget to the spheres with centers $(0, 1, 1)$, $(-1, -1, 0)$, $(1, -1, 1)$, $(-2, 2, 0)$, and respective radii 1, 3/2, 2, and 2, drawn below.



For this, we first give a system of coordinates for lines, then determine equations for a line to be tangent to a sphere, and finally, solve this instance.

4.1.1 Coordinates for lines

We will consider a line in $\mathbb{R}^3$ as a line in complex projective space, $\mathbb{P}^3$. Thus we study an *algebraic relaxation* of the original problem in $\mathbb{R}^3$, as our coordinates will include complex lines as well as lines at infinity. Recall that projective space $\mathbb{P}^3$ is the set of all 1-dimensional linear subspaces of $\mathbb{C}^4$. Under this correspondence, a line in $\mathbb{P}^3$ is the projective space of a 2-dimensional linear subspace of $\mathbb{C}^4$.

We will need some multilinear algebra[†]. Consider the inclusion $V \hookrightarrow \mathbb{C}^4$, where V is a 2-dimensional linear subspace of $\mathbb{C}^4$. Applying the second exterior power gives

$$\wedge^2 V \;\hookrightarrow\; \wedge^2 \mathbb{C}^4 \;\simeq\; \mathbb{C}^{\binom{4}{2}} \;\simeq\; \mathbb{C}^6 .$$

Since V is 2-dimensional, $\wedge^2 V \simeq \mathbb{C}$, and so $\wedge^2 V$ is a 1-dimensional linear subspace of $\wedge^2 \mathbb{C}^4$, which a point in the corresponding projective space. In this way, we associate each 2-plane in $\mathbb{C}^4$ to a point in the projective 5-space $\mathbb{P}(\wedge^2 \mathbb{C}^4) = \mathbb{P}^5$.

This map

$$\{\text{2-planes in } \mathbb{C}^4\} \;\longrightarrow\; \mathbb{P}^5 \;=\; \mathbb{P}(\wedge^2 \mathbb{C}^4)$$

is called the *Plücker embedding* and $\mathbb{P}(\wedge^2 \mathbb{C}^4)$ is called *Plücker space*. We identify its image. A tensor in $\wedge^2 \mathbb{C}^4$ is *decomposable* if it has the form $u \wedge v$. If $V = \langle u, v \rangle$ is the linear span of u and v, then $\wedge^2 V = \langle u \wedge v \rangle$ is spanned by decomposable tensors. Conversely, any non-zero decomposable tensor $u \wedge v \in \mathbb{P}(\wedge^2 \mathbb{C}^4)$ is the image of a unique 2-dimensional linear subspace $\langle u, v \rangle$ of $\mathbb{C}^4$ (Exercise 2).

Let us investigate decomposable tensors. A basis e_0, e_1, e_2, e_3 for $\mathbb{C}^4$, gives the basis

$$e_0 \wedge e_1 \,,\; e_0 \wedge e_2 \,,\; e_0 \wedge e_3 \,,\; e_1 \wedge e_2 \,,\; e_1 \wedge e_3 \,,\; e_2 \wedge e_3$$

for $\wedge^2 \mathbb{C}^4$, showing that it is six-dimensional. If

$$u \;=\; u_0 e_0 + u_1 e_1 + u_2 e_2 + u_3 e_3 \qquad \text{and} \qquad v \;=\; v_0 e_0 + v_1 e_1 + v_2 e_2 + v_3 e_3$$

are vectors in $\mathbb{C}^4$, their exterior product $u \wedge v$ is

$$\begin{aligned} &(u_0 v_1 - u_1 v_0) e_0 \wedge e_1 \;+\; (u_0 v_2 - u_2 v_0) e_0 \wedge e_2 \;+\; (u_0 v_3 - u_3 v_0) e_0 \wedge e_3 \\ +\;\; &(u_1 v_2 - u_2 v_1) e_1 \wedge e_2 \;+\; (u_1 v_3 - u_3 v_1) e_1 \wedge e_3 \;+\; (u_2 v_3 - u_3 v_2) e_2 \wedge e_3 \,. \end{aligned}$$

The components

$$p_{ij} \;:=\; u_i v_j - u_j v_i \;=\; \det \begin{pmatrix} u_i & u_j \\ v_i & v_j \end{pmatrix} \qquad 0 \le i < j \le 3$$

[†]Do this in the algebra appendix, Appendix A

of this decomposable tensor are the *Plücker coordinates* of the 2-plane $\langle u, v\rangle$.

Observe that $p_{03}p_{12}$ equals

$$(u_0v_3 - u_3v_0)(u_1v_2 - u_2v_1) \;=\; u_0u_1v_2v_3 - u_0u_2v_1v_3 - u_1u_3v_0v_2 + u_2u_3v_0v_1\,.$$

Another product which has the same leading term, $u_0u_1v_2v_3$, in the lexicographic monomial order where $u_0 > u_1 > \cdots > v_2 > v_3$ is $p_{02}p_{13}$, which is

$$(u_0v_2 - u_2v_0)(u_1v_3 - u_3v_1) \;=\; u_0u_1v_2v_3 - u_0u_3v_1v_2 - u_1u_2v_0v_3 + u_2u_3v_0v_1\,.$$

If we subtract these, $p_{03}p_{12} - p_{02}p_{13}$, we obtain

$$-(u_0u_2v_1v_3 - u_0u_3v_1v_2 - u_1u_2v_0v_3 + u_1u_3v_0v_2) \;=\; -(u_0v_1 - u_1v_0)(u_2v_3 - u_3v_2)\,,$$

which is $p_{01}p_{23}$. We have just applied the subduction algorithm to the polynomials p_{ij} to obtain the quadratic *Plücker relation*

$$p_{03}p_{12} - p_{02}p_{13} + p_{01}p_{23} \;=\; 0\,, \tag{4.1}$$

which holds on the Plücker coordinates of decomposable tensors. In the exercises, you are asked to show that any $p_{01}, \ldots, p_{23}$ satisfying this relation is a Plücker coordinate of a 2-plane in $\mathbb{C}^4$.

Definition 4.1.1 The *Grassmannian of 2-planes in* $\mathbb{C}^4$, $G(2,4)$, is the algebraic subvariety of Plücker space defined by the Plücker relation (4.1). We may also write $\mathbb{G}(1,3)$ for this Grassmannian, when we consider it as the space of lines in $\mathbb{P}^3$.

The Plücker coordinates of the Grassmannian as a subvariety of Plücker space give us a set of coordinates for lines.

4.1.2 Equation for a line to be tangent to a sphere

The points of a sphere S with center (a, b, c) and radius r are the homogeneous coordinates $X = [x_0, x_1, x_2, x_3]^T$ that satisfy the quadratic equation $X^TQX = 0$, where

$$Q \;=\; \begin{bmatrix} a^2 + b^2 + c^2 - r^2 & -a & -b & -c \\ -a & 1 & 0 & 0 \\ -b & 0 & 1 & 0 \\ -c & 0 & 0 & 1 \end{bmatrix}\,. \tag{4.2}$$

Indeed, X^TQX is $(x_1 - ax_0)^2 + (x_2 - bx_0)^2 + (x_3 - cx_0)^2 - r^2x_0^2$. This matrix Q defines an isomorphism $\mathbb{C}^4 \xrightarrow{Q} (\mathbb{C}^4)^\vee$, between $\mathbb{C}^4$ and its linear dual, $(\mathbb{C}^4)^\vee$. The quadratic form X^TQX is simply the pairing between $X \in \mathbb{C}^4$ and $QX \in (\mathbb{C}^4)^\vee$.

If V is a 2-plane in $\mathbb{C}^4$, then its intersection with the sphere S is the zero set of this quadratic form restricted to V. There are three possibilities for such a homogeneous

quadratic form on $V \simeq \mathbb{C}^2$. If it is non-zero, then it factors. Either it has two distinct factors, and thus the line corresponding to V meets the sphere in 2 distinct points, or else it is the square of a linear form, and thus the line is tangent to the sphere. The third possibility is that it is zero, in which case, the line lies on the sphere (such a line is necessarily imaginary) and is tangent to the sphere at every point of the line. The three cases are distinguished by the rank of the matrix representing the quadratic form (Exercise 3). In particular, the line is tangent to the sphere if and only if the determinant of this matrix vanishes.

We investigate its determinant. This restriction is defined by the composition of maps

$$V \;\hookrightarrow\; \mathbb{C}^4 \;\xrightarrow{Q}\; (\mathbb{C}^4)^\vee \;\twoheadrightarrow\; V^\vee\,, \tag{4.3}$$

where the last map is the restriction of a linear form on $\mathbb{C}^4$ to V. The line represented by V is tangent to S if and only if this quadratic form is degenerate, which means that the map does not have full rank. To take the determinant of this map between two-dimensional vector spaces, we apply the second exterior power $\wedge^2$ to the composition (4.3) and obtain

$$\wedge^2 V \;\hookrightarrow\; \wedge^2\mathbb{C}^4 \;\xrightarrow{\wedge^2 Q}\; \wedge^2(\mathbb{C}^4)^\vee \;\twoheadrightarrow\; \wedge^2 V^\vee\,.$$

Since the image of $\wedge^2 V$ in $\wedge^2\mathbb{C}^4$ is spanned by the Plücker vector p of $\wedge^2 V$ and we restrict a linear form on $\wedge^2\mathbb{C}^4$ to $\wedge^2 V^\vee$ by evaluating it at p, we obtain the equation

$$p^T \wedge^2 Q\; p \;=\; 0\,,$$

for the line with Plücker coordinate p to be tangent to the sphere defined by the quadratic form Q.

If we express $\wedge^2 Q$ as a matrix with respect to the basis $e_i \wedge e_j$ of $\wedge^2\mathbb{C}^4$, it will have rows and columns indexed by pairs ij with $0 \le i < j \le 3$, where

$$(\wedge^2 Q)_{ij,kl} \;:=\; Q_{ik}Q_{jl} - Q_{il}Q_{jk} \;=\; \det\begin{pmatrix} Q_{ik} & Q_{il} \\ Q_{jk} & Q_{jl} \end{pmatrix}\,.$$

For our sphere (4.2), this is

$$\wedge^2 Q \;=\; \begin{pmatrix} b^2+c^2-r^2 & -ab & -ac & b & c & 0 \\ -ab & a^2+c^2-r^2 & -bc & -a & 0 & c \\ -ac & -bc & a^2+b^2-r^2 & 0 & -a & -b \\ b & -a & 0 & 1 & 0 & 0 \\ c & 0 & -a & 0 & 1 & 0 \\ 0 & c & -b & 0 & 0 & 1 \end{pmatrix} \begin{matrix} 01 \\ 02 \\ 03 \\ 12 \\ 13 \\ 23 \end{matrix} \tag{4.4}$$

We remark that there is nothing special about spheres in this discussion.

Theorem 4.1.2 *If q is any smooth quadric in $\mathbb{P}^3$ defined by a quadratic form Q, then a line with Plücker coordinate p is tangent to q if and only if $p^T \wedge^2 Q\; p = 0$, if and only if p lies on the quadric in Plücker space defined by $\wedge^2 Q$.*

4.1.3 Solving the equations?

We may now formulate our problem of lines tangent to four spheres as the solutions to a system of equations. Namely, the set of lines tangent to four spheres have Plücker coordinates in $\mathbb{P}^5$ which satisfy

1. The Plücker equation (4.1), and
2. Four quadratic equations of the form $p^T \wedge^2 Q\, p = 0$, one for each sphere.

By Bézout's theorem, we expect that there will be 2^5 solutions to these five quadratic equations on $\mathbb{P}^5$.

Let us investigate these equations. We will use the symbolic computation package Singular [8] and display both annotated code and output in `typewriter font`. Output lines begin with `//`, which are comment-line characters in Singular. First, we define our ground ring R to be $\mathbb{Q}[u, v, w, x, y, z]$ with the degree reverse lexicographic monomial order where $u > v > \cdots > z$. This is the coordinate ring of Plücker space, where we identify p_{01} with u, p_{02} with v, p_{03} with w, and so on. We also declare the types of some variables.

```
ring R = 0, (u,v,w,x,y,z), dp;
matrix wQ[6][6];
matrix Pc[6][1] = u,v,w,x,y,z;
```

We give a procedure to compute $\wedge^2 Q$ (4.4),

```
proc Wedge_2_Sphere (poly r, a, b, c)
 {
  wQ = b^2+c^2-r^2 , -a*b , -a*c ,  b , c , 0,
       -a*b , a^2+c^2-r^2 , -b*c , -a , 0 , c,
       -a*c , -b*c , a^2+b^2-r^2 ,  0 ,-a ,-b,
                  b  , -a  ,  0  ,  1 , 0 , 0,
                  c  ,  0  , -a  ,  0 , 1 , 0,
                  0  ,  c  , -b  ,  0 , 0 , 1;
  return(wQ);
 }
```

and a procedure to compute the quadratic form $p^T \wedge^2 Q\, p$.

```
proc makeEquation (poly r, a, b, c)
 {
  return((transpose(Pc)*WedgeTwoSphere(r,a,b,c)*Pc)[1][1]);
 }
```

Now we create the ideal defining the lines tangent to four spheres with radii 1, 3/2, 2, and 2, and respective centers $(0, 1, 1)$, $(-1, -1, 0)$, $(1, -1, 1)$, and $(-2, 2, 0)$.

```
ideal I =
   w*x-v*y+u*z,
   makeEquation(1 , 0, 1, 1),
   makeEquation(3/2 ,-1,-1, 0),
   makeEquation(2 , 1,-1, 1),
   makeEquation(2 ,-2, 2,0);
```

Lastly, we compute a Gröbner basis for I and determine its dimension and degree.

```
 I=std(I);
 degree(I);
// dimension (proj.)  = 1
// degree (proj.)   = 4
```

This computation shows that the set of lines tangent to these four spheres has dimension 1, so that there are infinitely many common tangents! This is not what we expected from Bézout's Theorem and we must conclude that our equations are *not* sufficiently general. We will try to understand the special structure in our equations.

The key to this, as it turns out, is to use some classical facts about spheres. It is well-knkown that circles are exactly the conics in the plane $\mathbb{P}^2$ which contain the imaginary circular points at infinity $[0, 1, \pm i]$. This is clear if we set $x_0 = 0$ in the equation for a circle with center (a, b) and radius r,

$$(x_1 - ax_0)^2 + (x_2 - ax_0)^2 \;=\; r^2 x_0^2 \,.$$

For the same reason, spheres are the quadrics in $\mathbb{P}^3$ which contain the imaginary *circular conic at infinity*, which is defined by

$$x_0 \;=\; 0 \qquad \text{and} \qquad x_1^2 + x_2^2 + x_3^2 \;=\; 0 \,.$$

The lines at infinity have Plücker coordinates satisfying $p_{01} = p_{02} = p_{03} = 0$. For such a point p, the equation $p^T \wedge^2 Q\, p$, where Q is (4.4), becomes

$$p_{12}^2 + p_{13}^2 + p_{23}^2 \;=\; 0 \,. \tag{4.5}$$

This is the condition that the line at infinity is tangent to the spherical conic at infinity. Since the parameters r, a, b, c for the sphere do not appear in the equations $p_{01} = p_{02} = p_{03} = 0$ and (4.5), every line at infinity tangent to the spherical conic at infinity is tangent to every sphere.

In the language of enumerative geometry, this problem of lines tangent to four spheres has *excess intersection*. That is, our equations for a line to be tangent to four spheres not only define the lines we want (the tangent lines not at infinity), but also lines we did not intend, namely these lines tangent to the spherical conic at infinity.

If I is the ideal generated by our equations and J is the ideal of this excess component, then by Lemma 3.3.7, the saturation $(I : J^\infty)$ is the ideal of $\overline{\mathcal{V}(I) \setminus \mathcal{V}(J)}$, which should

be the tangents that we seek. (We could also saturate by the ideal $K = \langle p_{01}, p_{02}, p_{03} \rangle$ of lines at infinity.) We return to our Singular computation, defining the ideal J and computing the quotient ideal $(I\colon J)$. We do this instead of saturation, as saturation is typically computationally expensive.

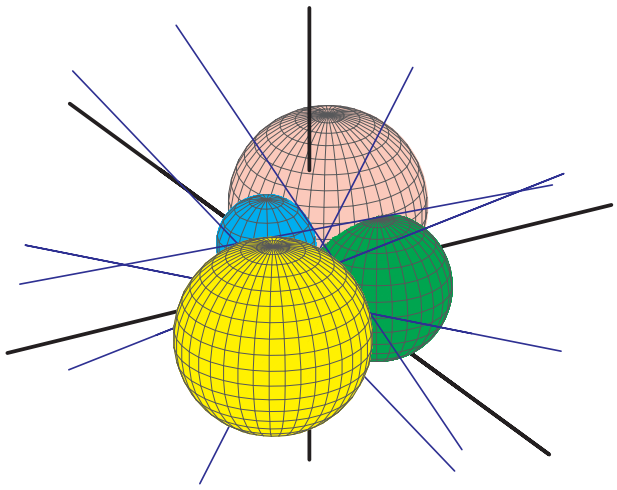
```
    ideal J = std(ideal(u,v,w,x^2+y^2+z^2));
    I = std(quotient(I,J));
    degree(I);
// dimension (proj.)  = 1
// degree (proj.)   = 2
```

While the degree of $(I\colon J)$ is less than that of I, it is still 1-dimensional, so we take the quotient ideal again.

```
    I = std(quotient(I,J));
    degree(I);
// dimension (proj.)  = 0
// degree (proj.)   = 12
```

The dimension is now zero and we have removed the excess component from $\mathcal{V}(I)$.

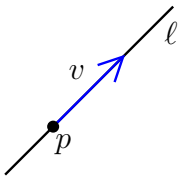
Since the dimension is 0 and the degree is 12, we expect that 12 is the answer to our original question. That is, we expect that there will be 12 *complex* lines tangent to amy four spheres in general position. In Exercise 6 we ask you to verify that there are indeed 12 complex common tangent lines to the four spheres. Of these 12, six are real, and we display them with the spheres below.



4.1.4 Twelve lines tanget to four general spheres

We remark that the computation of Section 4.1.3, while convincing, does not constitute a proof. We will give a rigorous proof here. Our basic idea to handle the excess component

is to simply define it away. Represent a line ℓ in $\mathbb{R}^3$ by a point $p \in \ell$ and a direction vector $v \in \mathbb{RP}^2$.



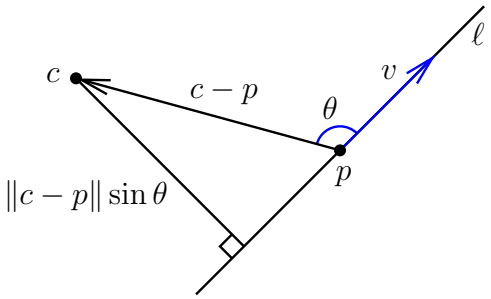
No such line can lie at infinity, so we are avoiding the excess component of lines at infinity tangent to the spherical conic at infinity.

Lemma 4.1.3 *The set of direction vectors $v \in \mathbb{RP}^2$ of lines tangent to four spheres with affinely independent centers consists of the common solutions to a cubic and a quartic equation on $\mathbb{RP}^2$. Each direction vector gives one common tangent.*

Proof. For vectors $x, y \in \mathbb{R}^3$, let $x \cdot y$ be their ordinary Euclidean dot product and write x^2 for $x \cdot x$, which is $\|x\|^2$. Fix p to be the point of ℓ closest to the origin, so that

$$p \cdot v \;=\; 0\,. \tag{4.6}$$

The distance from the line ℓ to a point c is $\|c-p\| \sin\theta$, where θ is the angle between v and the displacement vector from p to c.



This is the length of the cross product $(c-p) \times v$, multiplied by the length $\|v\|$ of the direction vector v. If ℓ is tangent to the sphere with radius r centered at $c \in \mathbb{R}^3$ if its distance to c is r, and so we have $\|(c-p) \times v\| = r\|v\|$. Squaring, we get

$$[(c-p) \times v]^2 \;=\; r^2 v^2\,. \tag{4.7}$$

This formulation requires that $v^2 \neq 0$. A line with $v^2 = 0$ has a complex direction vector and it meets the spherical conic at infinity.

We assume that one sphere is centered at the origin and has radius r, while the other three have centers and radii (c_i, r_i) for $i = 1, 2, 3$. The condition for the line to be tangent to the sphere centered at the origin is

$$p^2 \;=\; r^2\,. \tag{4.8}$$

For the other spheres, we expand (4.7), use vector product identities, and the equations (4.6) and (4.8) to obtain the vector equation

$$2v^2 \begin{pmatrix} c_1^T \\ c_2^T \\ c_3^T \end{pmatrix} \cdot p \;=\; -\begin{pmatrix} (c_1 \cdot v)^2 \\ (c_2 \cdot v)^2 \\ (c_3 \cdot v)^2 \end{pmatrix} \;+\; v^2 \begin{pmatrix} c_1^2 + r^2 - r_1^2 \\ c_2^2 + r^2 - r_2^2 \\ c_3^2 + r^2 - r_3^2 \end{pmatrix} \;. \tag{4.9}$$

Now suppose that the spheres have affinely independent centers. Then the matrix $(c_1, c_2, c_3)^T$ appearing in (4.9) is invertible. Assuming $v^2 \neq 0$, we may use (4.9) to write p as a quadratic function of v. Substituting this expression into equations (4.6) and (4.8), we obtain a cubic and a quartic equation for $v \in \mathbb{RP}^2$. The lemma now follows from Bézout's Theorem. □

Bézout's Theorem implies that there are at most $3 \cdot 4 = 12$ isolated solutions to these equations, and over $\mathbb{C}$ exactly 12 if they are generic. The equations are however far from generic as they involve only 13 parameters while the space of quartics has 14 parameters and the space of cubics has 9 parameters.

Example 4.1.4 Suppose that the spheres have equal radii, r, and have centers at the vertices of a regular tetrahedron with side length $2\sqrt{2}$,

$$(2,2,0)^T\,, \qquad (2,0,2)^T\,, \qquad (0,2,2)^T\,, \quad \text{and} \quad (0,0,0)^T\,.$$

In this symmetric case, the cubic factors into three linear factors. There are real common tangents only if $\sqrt{2} \leq r \leq 3/2$, and exactly 12 when the inequality is strict. If $r = \sqrt{2}$, then the spheres are pairwise tangent and there are three common tangents, one for each pair of non-intersecting edges of the tetrahedron. Each tangent has algebraic multiplicity 4. If $r = 3/2$, then there are six common tangents, each of multiplicity 2. The spheres meet pairwise in circles of radius $1/2$ lying in the plane equidistant from their centers. This plane also contains the centers of the other two spheres, as well as one common tangent which is parallel to the edge between those centers.

Figure 4.1 shows the cubic (which consists of three lines supporting the edges of an equilateral triangle) and the quartic, in an affine piece of the set $\mathbb{RP}^2$ of direction vectors. The vertices of the triangle are the standard coordinate directions $(1,0,0)^T$, $(0,1,0)^T$, and $(0,0,1)^T$. The singular cases, (i) when $r = \sqrt{2}$ and (ii) when $r = 3/2$, are shown first, and then (iii) when $r = 1.425$. The 12 points of intersection in this third case are visible in the expanded view in (iii'). Each point of intersection gives a real tangent, so there are 12 tangents to four spheres of equal radii 1.425 with centers at the vertices of the regular tetrahedron with edge length $2\sqrt{2}$.

One may also see this number 12 using group theory. The symmetry group of the tetrahedron, which is the group of permutations of the spheres, acts transitively on their common tangents and the isotropy group of any tangent has order 2. To see this, orient a common tangent and suppose that it meets the spheres a, b, c, d in order. Then the permutation $(a,d)(b,c)$ fixes that tangent but reverses its orientation, and the identity is the only other permutation fixing that tangent.

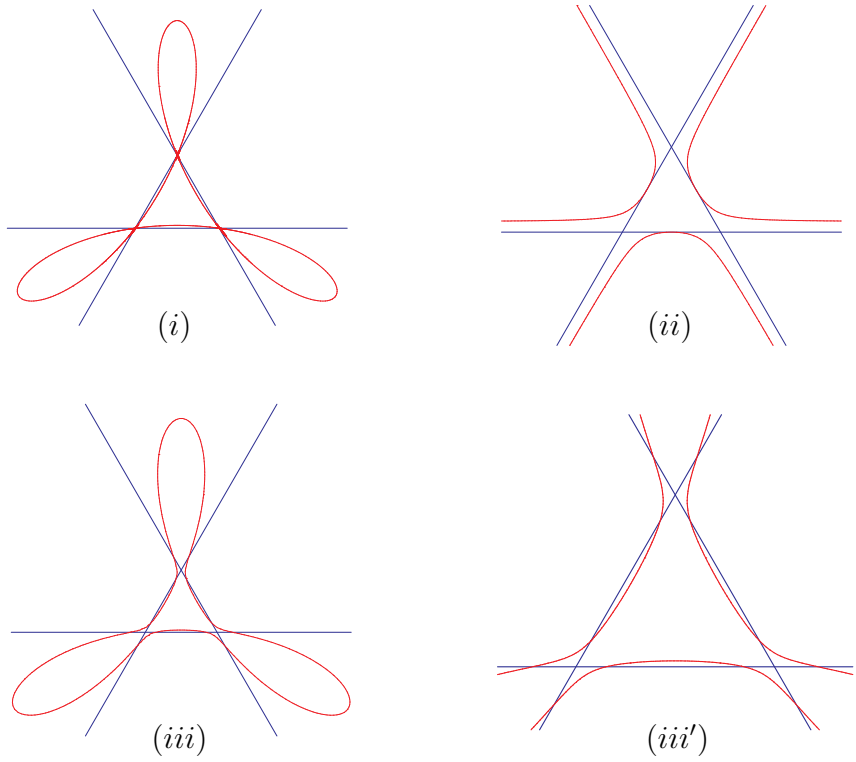


Figure 4.1: The cubic and quartic for symmetric configurations.

This example shows that the bound of 12 common tangents from Lemma 4.1.3 is in fact attained.

Theorem 4.1.5 *There are at most* 12 *common real tangent lines to four spheres whose centers are not coplanar, and there exist spheres with* 12 *common real tangents.*

Example 4.1.6 We give an example when the radii are distinct, namely 1.4, 1.42, 1.45, and 1.474. Figure 4.2 shows the quartic and cubic and the configuration of 4 spheres and their 12 common tangents.

Now suppose that the centers are coplanar. A continuity argument shows that four general such spheres will have 12 complex common tangents (or infinitely many, but this possibility is precluded by the following example). Three spheres of radius 4/5 centered at the vertices of an equilateral triangle with side length $\sqrt{3}$ and one of radius 1/3 at the triangle's center have 12 common real tangents. We display this configuration in Figure 4.3. This configuration of spheres has symmetry group $\mathbb{Z}_2 \times D_3$, which has order 12 and acts faithfully and transitively on the common tangents.

In the symmetric configuration of Example 4.1.4 having 12 common tangents, every pair of spheres meet. It is however not necessary for the spheres to meet pairwise when there are 12 common tangents. In fact, in both Figures 4.2 and 4.3 not all pairs of spheres

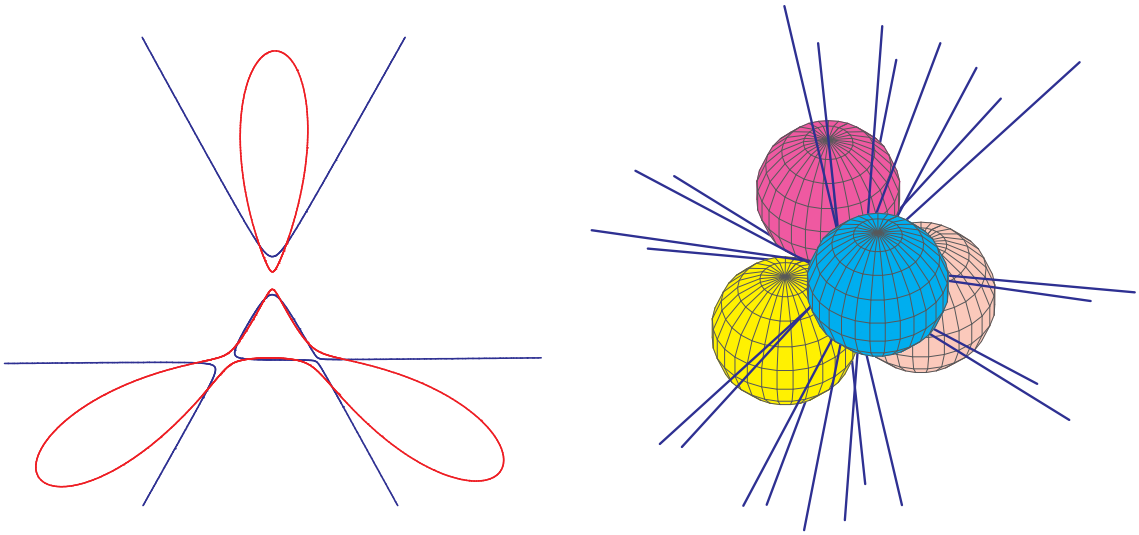


Figure 4.2: Spheres with 12 common tangents.

meet. However, the union of the spheres is connected. This turns out to be unnecessary. In the tetrahedral configuration, if one sphere has radius 1.38 and the other three have equal radii of 1.44, then the first sphere does not meet the others, but there are still 12 tangents.

More interestingly, it is possible to have 12 common real tangents to four *disjoint* spheres. Figure 4.4 displays such a configuration. The three large spheres have radius $4/5$ and are centered at the vertices of an equilateral triangle of side length $\sqrt{3}$, while the smaller sphere has radius $1/4$ and is centered on the axis of symmetry of the triangle, but at a distance of $35/100$ from the plane of the triangle.

Exercises

1. Show that if $u \wedge v$ and $x \wedge y$ are two decomposable tensors representing the same point in $\mathbb{P}(\wedge^2\mathbb{C}^4)$, then they come from the same 2-plane in $\mathbb{C}^4$. That is, show that $\langle u, v\rangle = \langle x, y\rangle$.

2. Show that if $[p_{01}, \ldots, p_{23}]$ are the homogeneous coordinates of a point of $\mathbb{P}(\wedge^2\mathbb{C}^4)$ that satisfies the Plücker relation, then this is the Plücker coordinate of a 2-plane in $\mathbb{C}^4$.

3. Let Q be a symmetric 2×2 matrix. Show that the quadratic form X^TQX has distinct factors if and only if Q is invertible. If Q has rank 1, then show that X^TQX is the square of a linear form, and that X^TQX is the zero polynomial only when Q has rank zero.

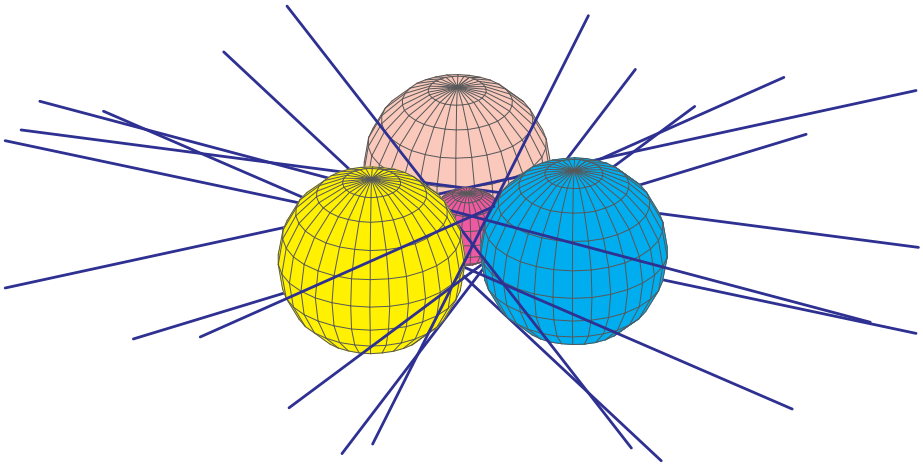


Figure 4.3: Spheres with coplanar centers and 12 common tangents.

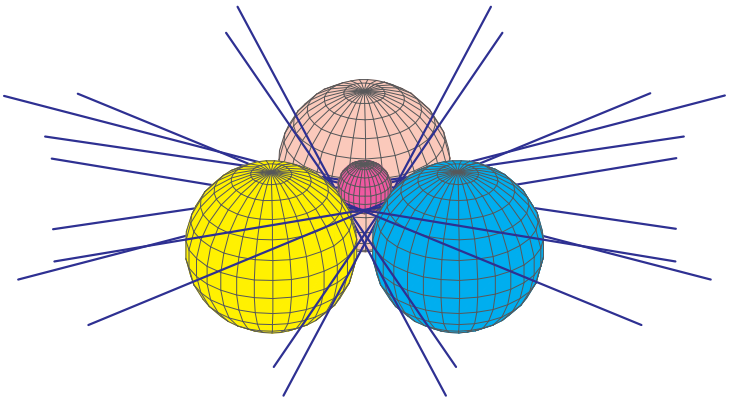


Figure 4.4: Four disjoint spheres with 12 common tangents.

4. Verify that there are indeed 12 lines tangent (four real and 8 complex) to the four spheres of the example in Section 4.1.3.

5. Show that four general quadrics in $\mathbb{P}^3$ will have 32 common tangents. By Bézout's theorem, it suffices to find a single instance of four quadrics with this number of tangents. You may do this by replacing the procedure `WedgeTwoSphere` with a procedure to compute $\wedge^2 Q$ for an arbitrary 4×4 symmetric matrix Q.

6. Verify the claim that there are 12 complex tangents to the four spheres discussed in Section 4.1.3. Show that six of these are real.

Appendix A

Appendix

A.1 Algebra

Algebra is the foundation of algebraic geometry; here we collect some of the basic algebra on which we rely. We develop some algebraic background that is needed in the text. This may not be an adequate substitute for a course in abstract algebra. Proofs can be found in give some useful texts.

A.1.1 Fields and Rings

We are all familiar with the real numbers, $\mathbb{R}$, with the rational numbers $\mathbb{Q}$, and with the complex numbers $\mathbb{C}$. These are the most common examples of *fields*, which are the basic building blocks of both the algebra and the geometry that we study. Formally and briefly, a field is a set $\mathbb{F}$ equipped with operations of addition and multiplication and distinguished elements 0 and 1 (the additive and multiplicative identities). Every number $a \in \mathbb{F}$ has an additive inverse $-a$ and every non-zero number $a \in \mathbb{F}^\times := \mathbb{F} - \{0\}$ has a multiplicative inverse $a^{-1} =: \frac{1}{a}$. Addition and multiplication are commutative and associative and multiplication distributes over addition, $a(b + c) = ab + ac$. To avoid triviality, we require that $0 \neq 1$.

The set of integers $\mathbb{Z}$ is not a field as $\frac{1}{2}$ is not an integer. While we will mostly be working over $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, at times we will need to discuss other fields. Most of what we do in algebraic geometry makes sense over any field, including the finite fields. In particular, linear algebra (except numerical linear algebra) works over any field.

Linear algebra concerns itself with *vector spaces*. A vector space V over a field $\mathbb{F}$ comes equipped with an operation of addition—we may add vectors and an operation of multiplication—we may multiply a vector by an element of the field. A linear combination of vectors $v_1, \ldots, v_n \in V$ is any vector of the form

$$a_1 v_1 + a_2 v_2 + \cdots + a_n v_n \,,$$

where $a_1, \ldots, a_n \in \mathbb{F}$. A collection S of vectors *spans* V if every vector in V is a linear

combination of vectors from S. A collection S of vectors is *linearly independent* if zero is not nontrivial linear combination of vectors from S. A *basis* S of V is a linearly independent spanning set. When a vector space V has a finite basis, every other basis has the same number of elements, and this common number is called the *dimension* of V.

A *ring* is the next most complicated object we encounter. A ring R comes equipped with an addition and a multiplication which satisfy almost all the properties of a field, except that we do not necessarily have multiplicative inverses. While the integers $\mathbb{Z}$ do not form a field, they do form a ring. An *ideal* I of a ring R is a subset which is closed under addition and under multiplication by elements of R. Every ring has two trivial ideals, the zero ideal $\{0\}$ and the unit ideal consisting of R itself. Given a set $S \subset R$ of elements, the smallest ideal containing S, also called the ideal *generated by* S, is

$$\langle S \rangle \;:=\; \{r_1 s_1 + r_2 s_2 + \cdots + r_m s_m \mid r_1, \ldots, r_m \in R \text{ and } s_1, \ldots, s_m \in S\}\,.$$

A primary use of ideals in algebra is through the construction of quotient rings. Let $I \subset R$ be an ideal. Formally, the *quotient ring* R/I is the collection of all sets of the form

$$[r] \;:=\; r + I \;=\; \{r + s \mid s \in I\}\,,$$

as r ranges over R. Addition and multiplication of these sets are defined in the usual way

$$\begin{aligned} [r] + [s] &= \{r' + s' \mid r' \in [r] \text{ and } s' \in [s]\} \overset{!}{=} [r+s], \quad \text{and} \\ [r] \cdot [s] &= \{r' \cdot s' \mid r' \in [r] \text{ and } s' \in [s]\} \overset{!}{=} [rs]\,. \end{aligned}$$

The last equality ($\overset{!}{=}$) in each line is meant to be surprising, it is a theorem and due to I being an ideal. Thus addition and multiplication on R/I are inherited from R. With these definitions (and also $-[r] = [-r]$, $0 := [0]$, and $1 := [1]$), the set R/I becomes a ring.

We say 'R-mod-I' for R/I because the arithmetic in R/I is just the arithmetic in R, but considered modulo the ideal I, as $[r] = [s]$ in R/I if and only if $r - s \in I$.

Ideals also arise naturally as kernels of homomorphisms. A *homomorphism* $\varphi\colon R \to S$ from the ring R to the ring S is a function that preserves the ring structure. Thus for $r, s \in R$, $\varphi(r+s) = \varphi(r) + \varphi(s)$ and $\varphi(rs) = \varphi(r)\varphi(s)$. We also require that $\varphi(1) = 1$. The *kernel* of a homomorphism $\varphi\colon R \to S$,

$$\ker \varphi \;:=\; \{r \in R \mid \varphi(r) = 0\}$$

is an ideal: If $r, s \in \ker\varphi$ and $t \in R$, then

$$\varphi(r+s) \;=\; \varphi(r) + \varphi(s) \;=\; 0 \;=\; t\varphi(r) \;=\; \varphi(tr)\,.$$

Homomorphisms are deeply intertwined with ideals. If I is an ideal of a ring R, then the association $r \mapsto [r]$ defines a homomorphism $\varphi\colon R \to R/I$ whose kernel is I. Dually, given a homomorphism $\varphi\colon R \to S$, the image of R in S is identified with $R/\ker\varphi$. More

generally, if $\varphi\colon R \to S$ is a homomorphism and $I \subset R$ is an ideal with $I \subset \ker \varphi$ (that is, $\varphi(I) = 0$), then φ induces a homomorphism $\varphi\colon R/I \to S$.

Properties of ideals induce natural properties in the associated quotient rings. An element r of a ring R is *nilpotent* if $r \neq 0$, but some power of r vanishes. A ring R is *reduced* if it has no nilpotent elements, that is, whenever $r \in R$ and n is a natural number with $r^n = 0$, then we must have $r = 0$. An ideal *radical* if whenever $r \in R$ and n is a natural number with $r^n \in I$, then we must have $r \in I$. It follows that a quotient ring R/I is reduced if and only if I is radical.

A ring R is a *domain* if whenever we have $r \cdot s = 0$ with $r \neq 0$, then we must have $s = 0$. An ideal is *prime* if whenever $r \cdot s \in I$ with $r \notin I$, then we must have $s \in I$. It follows that a quotient ring R/I is a domain if and only if I is prime.

A ring R with no nontrivial ideals must be a field. Indeed, if $0 \neq r \in R$, then the ideal rR of R generated by r is not the zero ideal, and so it must equal R. But then $1 = rs$ for some $s \in R$, and so r is invertible. Conversely, if R is a field and $0 \neq r \in R$, then $1 = r \cdot r^{-1} \in rR$, so the only ideals of R are $\{0\}$ and R. An ideal $\mathfrak{m}$ of R is *maximal* if $\mathfrak{m} \subsetneq R$, but there is no ideal I strictly contained between $\mathfrak{m}$ and R; if $\mathfrak{m} \subset I \subset R$ and $I \neq R$, then $I = \mathfrak{m}$. It follows that a quotient ring R/I is a field if and only if I is maximal.

Lastly, we remark that any ideal I of R with $I \neq R$ is contained in some maximal ideal. Suppose not. Then we may find an infinite chain of ideals

$$I \;=:\; I_0 \subsetneq I_1 \subsetneq I_2 \subsetneq \cdots$$

where each is proper so that 1 lies in none of them. Set $J := \bigcup_n I_n$. Then we claim that the union $I := \bigcup_n I_n$ of these ideals is an ideal. Indeed, if $r, s \in I$ then there are indices i, j with $r \in I_i$ and $s \in I_j$. Since $I_i, I_j \subset I_{\max(i,j)}$, we have $r + s \in I_{\max(i,j)} \subset J$. If $t \in R$, then $tr \in I_i \subset J$

A.1.2 Fields and polynomials

Our basic algebraic objects are polynomials. A *univariate polynomial* p is an expression of the form

$$p \;=\; p(x) \;:=\; a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m\,, \tag{A.1}$$

where m is a nonnegative integer and the coefficients $a_0, a_1, \ldots, a_m$ lie in $\mathbb{F}$. Write $\mathbb{F}[x]$ for the set of all polynomials in the variable x with coefficients in $\mathbb{F}$. We may add, subtract, and multiply polynomials and $\mathbb{F}[x]$ is a ring.

While a polynomial p may be regarded as a formal expression (A.1), evaluation of a polynomial defines a function $p\colon \mathbb{F} \to \mathbb{F}$: The value of the function p at a point $a \in \mathbb{F}$ is simply $p(a)$. When $\mathbb{F}$ is infinite, the polynomial and the function determine each other, but this is not the case when $\mathbb{F}$ is finite.

Our study requires polynomials with more than one variable. We first define a monomial.

Definition A.1.1 A *monomial* in the variables $x_1, \ldots, x_n$ is a product of the form

$$x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} ,$$

where the exponents $\alpha_1, \ldots, \alpha_n$ are nonnegative integers. For notational convenience, set $\alpha := (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^{n\dagger}$ and write x^α for the expression $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. The *(total) degree* of the monomial x^α is $|\alpha| := \alpha_1 + \cdots + \alpha_n$.

A *polynomial* $f = f(x_1, \ldots, x_n)$ in the variables $x_1, \ldots, x_n$ is a linear combination of monomials, that is, a finite sum of the form

$$f \;=\; \sum_{\alpha \in \mathbb{N}^n} a_\alpha x^\alpha ,$$

where each *coefficient* a_α lies in $\mathbb{F}$ and all but finitely many of the coefficients vanish. The product $a_\alpha x^\alpha$ of an element a_α of $\mathbb{F}$ and a monomial x^α is called a *term*. The *support* $\mathcal{A} \subset \mathbb{N}^n$ of a polynomial f is the set of all exponent vectors that appear in f with a nonzero coefficient. We will say that f has support $\mathcal{A}$ when we mean that the support of f is a subset of $\mathcal{A}$.

After 0 and 1 (the additive and multiplicative identities), the most distinguished integers are the prime numbers, those $p > 1$ whose only divisors are 1 and themselves. These are the numbers $2, 3, 5, 7, 11, 13, 17, 19, 23, \ldots$ Every integer $n > 1$ has a unique factorization into prime numbers

$$n \;=\; p_1^{\alpha_1} p_1^{\alpha_2} \cdots p_n^{\alpha_n} ,$$

where $p_1 < \cdots < p_n$ are distinct primes, and $\alpha_1, \ldots, \alpha_n$ are (strictly) positive integers. For example, $999 = 3^3 \cdot 37$. Polynomials also have unique factorization.

Definition A.1.2 A nonconstant polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ is *irreducible* if whenever we have $f = gh$ with g, h polynomials, then either g or h is a constant. That is, f has no nontrivial factors.

Theorem A.1.3 *Every polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ is a product of irreducible polynomials*

$$f \;=\; p_1 \cdot p_2 \cdots p_m ,$$

where the polynomials $p_1, \ldots, p_m$ are irreducible and nonconstant. Moreover, this factorization is essentially unique. That is, if

$$f \;=\; q_1 \cdot q_2 \cdots q_s ,$$

is another such factorization, then $m = s$, and after permuting the order of the factors, each polynomial q_i is a scalar multiple of the corresponding polynomial p_i.

---

†Where have we defined $\mathbb{N}$?

A.1.3 Polynomials in one variable

While rings of polynomials have many properties in common with the integers, the relation is the closest for univariate polynomials. The *degree*, $\deg(f)$ of a univariate polynomial f is the largest degree of a monomial appearing in f. If this monomial has coefficient 1, then the polynomial is *monic*. This allows us to remove the ambiguity in the uniqueness of factorizations in Theorem A.1.3. A polynomial $f(x) \in \mathbb{F}[x]$ has a unique factorization of the form

$$f \;=\; f_m \cdot p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots p_s^{\alpha_s}\,,$$

where $f_m \in \mathbb{F}^\times$ is the leading coefficient of f, the polynomials $p_1, \ldots, p_s$ are monic and irreducible, and the exponents α_i are positive integers.

Definition A.1.4 A *greatest common divisor* of two polynomials $f, g \in \mathbb{F}[x]$ (or $\gcd(f, g)$) is a polynomial h such that h divides each of f and g, and if there is another polynomial k which divides both f and g, then k divides h.

Any two polynomials f and g have a monic greatest common divisor which is the product of the common monic irreducible factors of f and g, each raised to the highest power that divides both f and g. Finding greatest common divisor would seem challenging as factoring polynomials is not an easy task. There is, however, a very fast and efficient algorithm for computing the greatest common divisor of two polynomials.

Suppose that we have polynomials f and g in $\mathbb{F}[x]$ with $\deg(g) \geq \deg(f)$,

$$\begin{aligned} f &= f_0 + f_1 x + f_2 x^2 + \cdots + f_m x^m \\ g &= g_0 + g_1 x + g_2 x^2 + \cdots + g_m x^m + \cdots + g_n x^n\,, \end{aligned}$$

where f_m and g_n are nonzero. Then the polynomial

$$S(f, g) \;:=\; g - \frac{g_n}{f_m} x^{n-m} \cdot f$$

has degree strictly less than $n = \deg(g)$. This simple operation of *reducing* f by the polynomial g forms the basis of the Division Algorithm and the Euclidean Algorithm for computing the greatest common divisor of two polynomials.

We describe the *Division Algorithm* in *pseudocode*, which is a common way to explain algorithms without reference to a specific programming language.

Algorithm A.1.5 [Division Algorithm]
INPUT: Polynomials $f, g \in \mathbb{F}[x]$.
OUTPUT: Polynomials $q, r \in \mathbb{F}[x]$ with $g = qf + r$ and $\deg(r) < \deg(f)$.
Set $r := g$ and $q := 0$.
(1) If $\deg(r) < \deg(f)$, then exit.
(2) Otherwise, reduce r by f to get the expression

$$r \;=\; \frac{r_n}{f_m} x^{n-m} \cdot f \;+\; S(f, r)\,,$$

where $n = \deg(r)$ and $m = \deg(f)$. Set $q := q + \frac{r_n}{f_m} x^{n-m}$ and $r := S(f, r)$, and return to step (1).

To see that this algorithm does produce the desired expression $g = qf + r$ with the degree of r less than the degree of f, note first that whenever we are at step (1), we will always have $g = qf + r$. Also, every time step (2) is executed, the degree of r must drop, and so after at most $\deg(g) - \deg(f) + 1$ steps, the algorithm will halt with the correct answer.

The *Euclidean Algorithm* computes the greatest common divisor of two polynomials f and g.

Algorithm A.1.6 [Euclidean Algorithm]
INPUT: Polynomials $f, g \in \mathbb{F}[x]$.
OUTPUT: The greatest common divisor h of f and g.
(1) Call the Division Algorithm to write $g = qf + r$ where $\deg(r) < \deg(f)$.
(2) If $r = 0$ then set $h := f$ and exit.
Otherwise, set $g := f$ and $f := r$ and return to step (1).

To see that the Euclidean algorithm performs as claimed, first note that if $g = qf + r$ with $r = 0$, then $f = \gcd(f, g)$. If $r \neq 0$, then $\gcd(f, g) = \gcd(f, r)$. Thus the greatest common divisor h of f and g is always the same whenever step (1) is executed. Since the degree of r must drop upon each iteration, r will eventually become 0, which shows that the algorithm will halt and return h.[†]

An ideal is *principal* if it has the form

$$\langle f \rangle \;=\; \{h \cdot f \mid h \in \mathbb{F}[x]\}\,,$$

for some $f \in \mathbb{F}[x]$. We say that f *generates* $\langle f \rangle$. Since $\langle f \rangle = \langle \alpha f \rangle$ for any $\alpha \in \mathbb{F}$, the principal ideal has a unique monic generator.

Theorem A.1.7 *Every ideal I of $\mathbb{F}[x]$ is principal.*

Proof. Suppose that I is a nonzero ideal of $\mathbb{F}[x]$, and let f be a nonzero polynomial of minimal degree in I. If $g \in I$, then we may apply the Division Algorithm and obtain polynomials $q, r \in \mathbb{F}[x]$ with

$$g \;=\; qf + r \qquad \text{with} \qquad \deg(r) < \deg(f)\,.$$

Since $r = g - qf$, we have $r \in I$, and since $\deg(r) < \deg(f)$, but f had minimal degree in I, we conclude that f divides g, and thus $I = \langle f \rangle$. □

[†]This is poorly written!

The ideal generated by univariate polynomials $f_1, \ldots, f_s$ is the principal ideal $\langle p \rangle$, where p is the greatest common divisor of $f_1, \ldots, f_s$.

For univariate polynomials p the quotient ring $\mathbb{F}[x]/\langle p \rangle$ has a concrete interpretation. Given $f \in \mathbb{F}[x]$, we may call the Division Algorithm to obtain polynomials q, r with

$$f \;=\; q \cdot p + r, \text{ where } \deg(r) < \deg(q)\,.$$

Then $[f] = f + \langle p \rangle = r + \langle p \rangle = [r]$ and in fact r is the unique polynomial of minimal degree in the coset $f + \langle p \rangle$. We call this the *normal form* of f in $\mathbb{F}[x]/\langle p \rangle$.

Since, if $\deg(r), \deg(s) < \deg(p)$, we cannot have $r - s \in \langle p \rangle$ unless $r = s$, we see that the monomials $1, x, x^2, \ldots, x^{\deg(p)-1}$ form a basis for the $\mathbb{F}$-vector space $\mathbb{F}[x]/\langle p \rangle$. This describes the additive structure on $\mathbb{F}[x]/\langle p \rangle$.

To describe its multiplicative structure, we only need to show how to write a product of monomials $x^a \cdot x^b$ with $a, b < \deg(p)$ in this basis. Suppose that p is monic with $\deg(p) = n$ and write $p(x) = x^n - q(x)$, where q has degree strictly less than p. Since $x^a \cdot x^b = (x^a \cdot x) \cdot x^{b-1}$, we may assume that $b = 1$. When $a < n$, we have $x^a \cdot x^1 = x^{a+1}$. When $a = n - 1$, then $x^{n-1} \cdot x^1 = x^n = q(x)$,

- Relate algebraic properties of $p(x)$ to properties of R, for example, zero divisors and domain.

- Prove that a field is a ring with only trivial ideals.

  Prove $I \subset J \subset R$ are ideals, then J/I is an ideal of R/I, and deduce that $R = \mathbb{F}[x]/p(x)$ is a field only if $p(x)$ is irreducible.

  Example $\mathbb{Q}[x]/(x^2 - 2)$ and explore $\mathbb{Q}(\sqrt{2})$.

  Example $\mathbb{R}[x]/(x^2 + 1)$ and show how it is isomorphic to $\mathbb{C}$.

  Work up to algebraically closed fields, the fundamental theorem of algebra (both over $\mathbb{C}$ and over $\mathbb{R}$).

  Explain that an algebraically closed field has no algebraic extensions (hence the name).

- Define the maximal ideal $\mathfrak{m}_a$ for $a \in \mathbb{A}^n$.

  Theorem A.1.8 *The maximal ideals of* $\mathbb{C}[x_1, \ldots, x_n]$ *all have the form* $\mathfrak{m}_a$ *for some* $a \in \mathbb{A}^n$.

A.2 Topology

Collect some topological statements here. Definition of topology, Closed/open duality, dense, nowhere dense... Describe the usual topology.

Recall that a function $f\colon X \to Y$ is continuous if and only if whenever $Z \subset Y$ is a closed set $f^{-1}(Z) \subset X$ is also closed.

Bibliography

[1] Etienne Bézout, *Théorie générale des équations algébriques*, Ph.-D. Pierres, 1779.

[2] ———, *General theory of algebraic equations*, Princeton University Press, 2006, Translated from French original by Eric Feron.

[3] B. Buchberger, *Ein algorithmisches Kriterium für die Lösbarkeit eines algebraischen Gleichungssystems*, Aequationes Math. **4** (1970), 374–383.

[4] Bruno Buchberger, *An algorithm for finding the basis elements of the residue class ring of a zero dimensional polynomial ideal*, J. Symbolic Comput. **41** (2006), no. 3-4, 475–511, Translated from the 1965 German original by Michael P. Abramson.

[5] David Cox, John Little, and Donal O'Shea, *Ideals, varieties, and algorithms*, third ed., Undergraduate Texts in Mathematics, Springer, New York, 2007, An introduction to computational algebraic geometry and commutative algebra.

[6] J.-C. Faugère, *FGb*, See `http://fgbrs.lip6.fr/jcf/Software/FGb/index.html`.

[7] J. C. Faugère, P. Gianni, D. Lazard, and T. Mora, *Efficient computation of zero-dimensional Gröbner bases by change of ordering*, J. Symbolic Comput. **16** (1993), no. 4, 329–344.

[8] G.-M. Greuel, G. Pfister, and H. Schönemann, SINGULAR *3.0*, A Computer Algebra System for Polynomial Computations, Centre for Computer Algebra, University of Kaiserslautern, 2005, `http://www.singular.uni-kl.de`.

[9] Raymond Hemmecke and Peter Malkin, *Computing generating sets of lattice ideals*, `math.CO/0508359`.

[10] H. Hironaka, *Resolution of singularities of an algebraic variety over a field of characteristic zero*, Ann. Math. **79** (1964), 109–326.

[11] Serkan Hoşten and Bernd Sturmfels, *GRIN: an implementation of Gröbner bases for integer programming*, Integer programming and combinatorial optimization (Copenhagen, 1995), Lecture Notes in Comput. Sci., vol. 920, Springer, Berlin, 1995, pp. 267–276.

[12] T. Y. Li, Tim Sauer, and J. A. Yorke, *The cheater's homotopy: an efficient procedure for solving systems of polynomial equations*, SIAM J. Numer. Anal. **26** (1989), no. 5, 1241–1251.

[13] F.S. Macaulay, *Some properties of enumeration in the theory of modular systems*, Proc. London Math. Soc. **26** (1927), 531–555.

[14] Andrew J. Sommese and Charles W. Wampler, II, *The numerical solution of systems of polynomials*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005, Arising in engineering and science.