3

Elimination Theory

This chapter will study systematic methods for eliminating variables from systems of polynomial equations. The basic strategy of elimination theory will be given in two main theorems: the Elimination Theorem and the Extension Theorem. We will prove these results using Groebner bases and the classic theory of resultants. The geometric interpretation of elimination will also be explored when we discuss the Closure Theorem. Of the many applications of elimination theory, we will treat two in detail: the implicitization problem and the envelope of a family of curves.

§1 The Elimination and Extension Theorems

To get a sense of how elimination works, let us look at an example similar to those discussed at the end of Chapter 2. We will solve the system of equations

$$\begin{aligned} x^2 + y + z &= 1, \\ x + y^2 + z &= 1, \\ x + y + z^2 &= 1. \end{aligned} \tag{1}$$

If we let I be the ideal

$$I = \langle x^2 + y + z - 1, x + y^2 + z - 1, x + y + z^2 - 1 \rangle, \tag{2}$$

then a Groebner basis for I with respect to lex order is given by the four polynomials

$$\begin{aligned} g_1 &= x + y + z^2 - 1, \\ g_2 &= y^2 - y - z^2 + z, \\ g_3 &= 2yz^2 + z^4 - z^2, \\ g_4 &= z^6 - 4z^4 + 4z^3 - z^2. \end{aligned} \tag{3}$$

It follows that equations (1) and (3) have the same solutions. However, since

$$g_4 = z^6 - 4z^4 + 4z^3 - z^2 = z^2(z-1)^2(z^2 + 2z - 1)$$

involves only z, we see that the possible z's are 0,1 and $-1 \pm \sqrt{2}$. Substituting these values into $g_2 = y^2 - y - z^2 + z = 0$ and $g_3 = 2yz^2 + z^4 - z^2 = 0$, we can determine the possible y's, and then finally $g_1 = x + y + z^2 - 1 = 0$ gives the corresponding x's. In this way, one can check that equations (1) have exactly five solutions:

$$\begin{aligned}&(1,0,0),\ (0,1,0),\ (0,0,1),\\&(-1+\sqrt{2},-1+\sqrt{2},-1+\sqrt{2}),\\&(-1-\sqrt{2},-1-\sqrt{2},-1-\sqrt{2}).\end{aligned}$$

What enabled us to find these solutions? There were two things that made our success possible:

- (Elimination Step) We could find a consequence $g_4 = z^6 - 4z^4 + 4z^3 - z^2 = 0$ of our original equations which involved only z, i.e., we eliminated x and y from the system of equations.
- (Extension Step) Once we solved the simpler equation $g_4 = 0$ to determine the values of z, we could extend these solutions to solutions of the original equations.

The basic idea of *elimination theory* is that both the Elimination Step and the Extension Step can be done in great generality.

To see how the Elimination Step works, notice that our observation concerning g_4 can be written

$$g_4 \in I \cap \mathbb{C}[z],$$

where I is the ideal given in equation (2). In fact, $I \cap \mathbb{C}[z]$ consists of *all* consequences of our equations which eliminate x and y. Generalizing this idea leads to the following definition.

Definition 1. *Given $I = \langle f_1, \ldots, f_s\rangle \subset k[x_1, \ldots, x_n]$ the l-th* **elimination ideal** *I_l is the ideal of $k[x_{l+1}, \ldots, x_n]$ defined by*

$$I_l = I \cap k[x_{l+1}, \ldots, x_n].$$

Thus, I_l consists of all consequences of $f_1 = \cdots = f_s = 0$ which eliminate the variables $x_1, \ldots, x_l$. In the exercises, you will verify that I_l is an ideal of $k[x_{l+1}, \ldots, x_n]$. Note that $I = I_0$ is the 0th elimination ideal. Also observe that different orderings of the variables lead to different elimination ideals.

Using this language, we see that eliminating $x_1, \ldots, x_l$ means finding nonzero polynomials in the l-th elimination ideal I_l. Thus *a solution of the Elimination Step means giving a systematic procedure for finding elements of I_l*. With the proper term ordering, Groebner bases allow us to do this instantly.

Theorem 2 (The Elimination Theorem). *Let $I \subset k[x_1, \ldots, x_n]$ be an ideal and let G be a Groebner basis of I with respect to lex order where $x_1 > x_2 > \cdots > x_n$. Then, for every $0 \le l \le n$, the set*

$$G_l = G \cap k[x_{l+1}, \ldots, x_n]$$

is a Groebner basis of the l-th elimination ideal I_l.

Proof. Fix l between 0 and n. Since $G_l \subset I_l$ by construction, it suffices to show that

$$\langle \text{LT}(I_l)\rangle = \langle \text{LT}(G_l)\rangle$$

by the definition of Groebner basis. One inclusion is obvious, and to prove the other inclusion $\langle \text{LT}(I_l)\rangle \subset \langle \text{LT}(G_l)\rangle$, we need only show that the leading term $\text{LT}(f)$, for an arbitrary $f \in I_l$, is divisible by $\text{LT}(g)$ for some $g \in G_l$.

To prove this, note that f also lies in I, which tells us that $\text{LT}(f)$ is divisible by $\text{LT}(g)$ for some $g \in G$ since G is a Groebner basis of I. Since $f \in I_l$, this means that $\text{LT}(g)$ involves only the variables $x_{l+1}, \ldots, x_n$. Now comes the crucial observation: since we are using lex order with $x_1 > \cdots > x_n$, any monomial involving $x_1, \ldots, x_l$ is greater than all monomials in $k[x_{l+1}, \ldots, x_n]$, so that $\text{LT}(g) \in [x_{l+1}, \ldots, x_n]$ implies $g \in k[x_{l+1}, \ldots, x_n]$. This shows that $g \in G_l$, and the theorem is proved. □

For an example of how this theorem works, let us return to example (1) from the beginning of the section. Here, $I = \langle x^2 + y + z - 1, x + y^2 + z - 1, x + y + z^2 - 1\rangle$, and a Groebner basis with respect to lex order is given in (3). It follows from the Elimination Theorem that

$$\begin{aligned} I_1 &= I \cap \mathbb{C}[y, z] = \langle y^2 - y - z^2 + z, 2yz^2 + z^4 - z^2, z^6 - 4z^4 + 4z^3 - z^2\rangle,\\ I_2 &= I \cap \mathbb{C}[z] = \langle z^6 - 4z^4 + 4z^3 - z^2\rangle. \end{aligned}$$

Thus, $g_4 = z^6 - 4z^4 + 4z^3 - z^2$ is not just some random way of eliminating x and y from our equations—it is the best possible way to do so since any other polynomial that eliminates x and y is a multiple of g_4.

The Elimination Theorem shows that a Groebner basis for lex order eliminates not only the first variable, but also the first two variables, the first three variables, and so on. In some cases (such as the implicitization problem to be studied in §3), we only want to eliminate certain variables, and we do not care about the others. In such a situation, it is a bit of overkill to compute a Groebner basis using lex order. This is especially true since lex order can lead to some very unpleasant Groebner bases (see Exercise 13 of Chapter 2, §9 for an example). In the exercises, you will study versions of the Elimination Theorem that use more efficient monomial orderings than lex.

We next discuss the Extension Step. Suppose that we have an ideal $I \subset k[x_1, \ldots, x_n]$. As in Chapter 2, we have the affine variety

$$\mathbf{V}(I) = \{(a_1, \ldots, a_n) \in k^n : f(a_1, \ldots, a_n) = 0 \text{ for all } f \in I\}.$$

To describe points of $\mathbf{V}(I)$, the basic idea is to build up solutions one coordinate at a time. Fix some l between 1 and n. This gives us the elimination ideal I_l, and we will call a solution $(a_{l+1}, \ldots, a_n) \in \mathbf{V}(I_l)$ a *partial solution* of the original system of equations. To extend $(a_{l+1}, \ldots, a_n)$ to a complete solution in $\mathbf{V}(I)$, we first need to add one more coordinate to the solution. This means finding a_l so that $(a_l, a_{l+1}, \ldots, a_n)$ lies in the variety $\mathbf{V}(I_{l-1})$ of the next elimination ideal. More concretely, suppose that $I_{l-1} = \langle g_1, \ldots, g_r\rangle$ in $k[x_l, x_{l+1}, \ldots, x_n]$. Then we want to find solutions $x_l = a_l$ of the equations

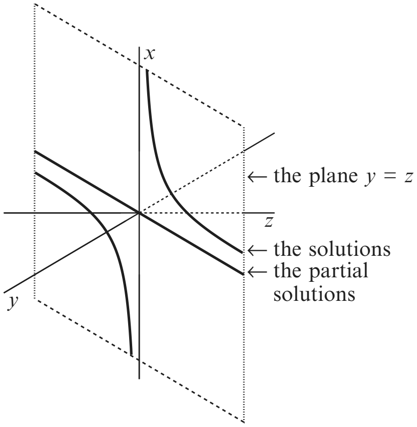
$$g_1(x_l, a_{l+1}, \ldots, a_n) = \cdots = g_r(x_l, a_{l+1}, \ldots, a_n) = 0.$$

Here we are dealing with polynomials of one variable x_l, and it follows that the possible a_l's are just the roots of the GCD of the above r polynomials.

The basic problem is that the above polynomials may not have a common root, i.e., there may be some partial solutions which do not extend to complete solutions. For a simple example, consider the equations

$$\begin{aligned} xy &= 1, \\ xz &= 1. \end{aligned} \tag{4}$$

Here, $I = \langle xy - 1, xz - 1\rangle$, and an easy application of the Elimination Theorem shows that $y - z$ generates the first elimination ideal I_1. Thus, the partial solutions are given by (a, a), and these all extend to complete solutions $(1/a, a, a)$ *except* for the partial solution (0, 0). To see what is going on geometrically, note that $y = z$ defines a plane in 3-dimensional space. Then the variety (4) is a hyperbola lying in this plane:



It is clear that the variety defined by (4) has no points lying over the partial solution (0,0). Pictures such as the one here will be studied in more detail in §2 when we study the geometric interpretation of eliminating variables. For now, our goal is to see if we can determine *in advance* which partial solutions extend to complete solutions.

Let us restrict our attention to the case where we eliminate just the first variable x_1. Thus, we want to know if a partial solution $(a_2, \ldots, a_n) \in \mathbf{V}(I_1)$ can be extended to a solution $(a_1, a_2, \ldots, a_n) \in \mathbf{V}(I)$. The following theorem tells us when this can be done.

Theorem 3 (The Extension Theorem). *Let $I = \langle f_1, \ldots, f_s\rangle \subset \mathbb{C}[x_1, \ldots, x_n]$ and let I_1 be the first elimination ideal of I. For each $1 \leq i \leq s$, write f_i in the form*

$$f_i = g_i(x_2, \ldots, x_n)x_1^{N_i} + \text{terms in which } x_1 \text{ has degree } < N_i,$$

where $N_i \geq 0$ and $g_i \in \mathbb{C}[x_2, \ldots, x_n]$ is nonzero. Suppose that we have a partial solution $(a_2, \ldots, a_n) \in \mathbf{V}(I_1)$. If $(a_2, \ldots, a_n) \notin \mathbf{V}(g_1, \ldots, g_s)$, then there exists $a_1 \in \mathbb{C}$ such that $(a_1, a_2, \ldots, a_n) \in \mathbf{V}(I)$.

The proof of this theorem uses resultants and will be given in §6. For the rest of the section, we will explain the Extension Theorem and discuss its consequences. A geometric interpretation will be given in §2.

A first observation is that the theorem is stated only for the field $k = \mathbb{C}$. To see why $\mathbb{C}$ is important, assume that $k = \mathbb{R}$ and consider the equations

$$\begin{aligned} x^2 &= y, \\ x^2 &= z. \end{aligned} \tag{5}$$

Eliminating x gives $y = z$, so that we get the partial solutions (a, a) for all $a \in \mathbb{R}$. Since the leading coefficients of x in $x^2 - y$ and $x^2 - z$ never vanish, the Extension Theorem guarantees that (a, a) extends, *provided* we work over $\mathbb{C}$. Over $\mathbb{R}$, the situation is different. Here, $x^2 = a$ has no real solutions when a is negative, so that only those partial solutions with $a \geq 0$ extend to real solutions of (5). This shows that the Extension Theorem is false over $\mathbb{R}$.

Turning to the hypothesis $(a_2, \ldots, a_n) \notin \mathbf{V}(g_1, \ldots, g_s)$, note that the g_i's are the leading coefficients with respect to x_l of the f_i's. Thus, $(a_2, \ldots, a_n) \notin \mathbf{V}(g_1, \ldots, g_s)$ says that the leading coefficients do not vanish simultaneously at the partial solution. To see why this condition is necessary, let us look at example (4). Here, the equations

$$\begin{aligned} xy &= 1, \\ xz &= 1 \end{aligned}$$

have the partial solutions $(y, z) = (a, a)$. The only one that does not extend is (0,0), which is the partial solution where the leading coefficients y and z of x vanish. The Extension Theorem tells us that *the Extension Step can fail only when the leading coefficients vanish simultaneously*.

Finally, we should mention that the variety $\mathbf{V}(g_1, \ldots, g_s)$ where the leading coefficients vanish depends on the basis $\{f_1, \ldots, f_s\}$ of I: changing to a different basis may cause $\mathbf{V}(g_1, \ldots, g_s)$ to change. In Chapter 8, we will learn how to choose $(f_1, \ldots, f_s)$ so that $\mathbf{V}(g_1, \ldots, g_s)$ is as small as possible. We should also point out that if one works in projective space (to be defined in Chapter 8), then one can show that *all* partial solutions extend.

Although the Extension Theorem is stated only for the case of eliminating the first variable x_1, it can be used when eliminating any number of variables. For example, consider the equations

$$\begin{aligned} x^2 + y^2 + z^2 &= 1, \\ xyz &= 1. \end{aligned} \tag{6}$$

A Groebner basis for $I = \langle x^2 + y^2 + z^2 - 1, xyz - 1\rangle$ with respect to lex order is

$$\begin{aligned} g_1 &= y^4z^2 + y^2z^4 - y^2z^2 + 1, \\ g_2 &= x + y^3z + yz^3 - yz. \end{aligned}$$

By the Elimination Theorem, we obtain

$$\begin{aligned} I_1 &= I \cap \mathbb{C}[y,z] = \langle g_1 \rangle, \\ I_2 &= I \cap \mathbb{C}[z] = \{0\}. \end{aligned}$$

Since $I_2 = \{0\}$, we have $\mathbf{V}(I_2) = \mathbb{C}$, and, thus, *every* $c \in \mathbb{C}$ is a partial solution. So we ask:

Which partial solutions $c \in \mathbb{C} = \mathbf{V}(I_2)$ extend to $(a,b,c) \in \mathbf{V}(I)$?

The idea is to extend c one coordinate at a time: first to (b,c), then to (a,b,c). To control which solutions extend, we will use the Extension Theorem at each step. The crucial observation is that I_2 is the first elimination ideal of I_1. This is easy to see here, and the general case is covered in the exercises. Thus, we will use the Extension Theorem once to go from $c \in \mathbf{V}(I_2)$ to $(b,c) \in \mathbf{V}(I_1)$, and a second time to go to $(a,b,c) \in \mathbf{V}(I)$. This will tell us exactly which c's extend.

To start, we apply the Extension Theorem to go from I_2 to $I_1 = \langle g_1 \rangle$. The coefficient of y^4 in g_1 is z^2, so that $c \in \mathbb{C} = \mathbf{V}(I_2)$ extends to (b,c) whenever $c \neq 0$. Note also that $g_1 = 0$ has *no* solution when $c = 0$. The next step is to go from I_1 to I; that is, to find a so that $(a,b,c) \in \mathbf{V}(I)$. If we substitute $(y,z) = (b,c)$ into (6), we get two equations in x, and it is not obvious that there is a common solution $x = a$. This is where the Extension Theorem shows its power. The leading coefficients of x in $x^2 + y^2 + z^2 - 1$ and $xyz - 1$ are 1 and yz, respectively. Since 1 never vanishes, the Extension Theorem *guarantees* that a always exists. We have thus proved that *all* partial solutions $c \neq 0$ extend to $\mathbf{V}(I)$.

The Extension Theorem is especially easy to use when one of the leading coefficients is constant. This case is sufficiently useful that we will state it as a separate corollary.

Corollary 4. *Let $I = \langle f_1, \dots, f_s \rangle \subset \mathbb{C}[x_1, \dots, x_n]$, and assume that for some i, f_i is of the form*

$$f_i = cx_1^N + \text{terms in which } x_1 \text{ has degree } < N,$$

where $c \in \mathbb{C}$ is nonzero and $N > 0$. If I_1 is the first elimination ideal of I and $(a_2, \dots, a_n) \in \mathbf{V}(I_1)$, then there is $a_1 \in \mathbb{C}$ so that $(a_1, a_2, \dots, a_n) \in \mathbf{V}(I)$.

Proof. This follows immediately from the Extension Theorem: since $g_i = c \neq 0$ implies $\mathbf{V}(g_1, \dots, g_s) = \emptyset$, we have $(a_2, \dots, a_n) \notin \mathbf{V}(g_1, \dots, g_s)$ for all partial solutions. □

We will end this section with an example of a system of equations that does not have nice solutions. Consider the equations

$$\begin{aligned} xy &= 4, \\ y^2 &= x^3 - 1. \end{aligned}$$

Using lex order, the Groebner basis is given by

$$\begin{aligned} g_1 &= 16x - y^2 - y^4, \\ g_2 &= y^5 + y^3 - 64, \end{aligned}$$

but if we proceed as usual, we discover that $y^5 + y^3 - 64$ has *no* rational roots (in fact, it is *irreducible* over $\mathbb{Q}$, a concept we will discuss in §5). One option is to compute the roots numerically. A variety of methods (such as the Newton-Raphson method) are available, and for $y^5 + y^3 - 64 = 0$, one obtains

$$y = 2.21363, -1.78719 \pm 1.3984i, \text{ or } 0.680372 \pm 2.26969i.$$

These solutions can then be substituted into $g_1 = 16x - y^2 - y^4 = 0$ to determine the values of x. Thus, unlike the previous examples, we can only find numerical approximations to the solutions.

There are many interesting problems that arise when one tries to find numerical solutions of polynomial equations. For further reading on this topic, we recommend LAZARD (1993) and MANOCHA (1994). The reader may also wish to consult COX, LITTLE and O'SHEA (1998), MIGNOTTE (1992) and MISHRA (1993).

EXERCISES FOR §1

1. Let $I \subset k[x_1, \ldots, x_n]$ be an ideal.
   a. Prove that $I_l = I \cap k[x_{l+1}, \ldots, x_n]$ is an ideal of $k[x_{l+1}, \ldots, x_n]$.
   b. Prove that the ideal $I_{l+1} \subset k[x_{l+2}, \ldots, x_n]$ is the first elimination ideal of $I_l \subset k[x_{l+1}, \ldots, x_n]$. This observation allows us to use the Extension Theorem multiple times when eliminating more than one variable.
2. Consider the system of equations

$$\begin{aligned} x^2 + 2y^2 &= 3, \\ x^2 + xy + y^2 &= 3. \end{aligned}$$

   a. If I is the ideal generated by these equations, find bases of $I \cap k[x]$ and $I \cap k[y]$.
   b. Find all solutions of the equations.
   c. Which of the solutions are *rational*, i.e., lie in $\mathbb{Q}^2$?
   d. What is the smallest field k containing $\mathbb{Q}$ such that all solutions lie in k^2?
3. Determine all solutions $(x, y) \in \mathbb{Q}^2$ of the system of equations

$$\begin{aligned} x^2 + 2y^2 &= 2 \\ x^2 + xy + y^2 &= 2. \end{aligned}$$

   Also determine all solutions in $\mathbb{C}^2$.
4. Find bases for the elimination ideals I_1 and I_2 for the ideal I determined by the equations:

$$\begin{aligned} x^2 + y^2 + z^2 &= 4, \\ x^2 + 2y^2 &= 5, \\ xz &= 1. \end{aligned}$$

   How many rational (i.e., in $\mathbb{Q}^3$) solutions are there?
5. In this exercise, we will prove a more general version of the Elimination Theorem. Fix an integer $1 \leq l \leq n$. We say that a monomial order $>$ on $k[x_1, \ldots, x_n]$ is of *l-elimination*

type provided that any monomial involving one of $x_1, \ldots, x_l$ is greater than all monomials in $k[x_{l+1}, \ldots, x_n]$. Prove the following generalized Elimination Theorem. If I is an ideal in $k[x_1, \ldots, x_n]$ and G is a Groebner basis of I with respect to a monomial order of l-elimination type, then $G \cap k[x_{l+1}, \ldots, x_n]$ is a basis of the lth elimination ideal $I \cap k[x_{l+1}, \ldots, x_n]$.

6. To exploit the generalized Elimination Theorem of Exercise 5, we need some interesting examples of monomial orders of l-elimination type. We will consider two such orders.
   a. Fix an integer $1 \leq l \leq n$, and define the order $>_l$ as follows: if $\alpha, \beta \in \mathbb{Z}^n_{\geq 0}$, then $\alpha >_l \beta$ if
   $$\alpha_1 + \cdots + a_l > \beta_i + \cdots + \beta_l, \text{ or } \alpha_1 + \cdots + \alpha_l = \beta_1 + \cdots + \beta_l \text{ and } \alpha >_{grevlex} \beta.$$
   This is the l-th *elimination order* of BAYER and STILLMAN (1987b). Prove that $>_l$ is a monomial order and is of l-elimination type. Hint: If you did Exercise 12 of Chapter 2, §4, then you have already done this problem.
   b. In Exercise 10 of Chapter 2, §4, we considered an example of a product order that mixed lex and grlex orders on different sets of variables. Explain how to create a product order that induces grevlex on both $k[x_1, \ldots, x_l]$ and $k[x_{l+1}, \ldots, x_n]$ and show that this order is of l-elimination type.
   c. If G is a Groebner basis for $I \subset k[x_1, \ldots, x_n]$ for either of the monomial orders of parts a or b, explain why $G \cap k[x_{l+1}, \ldots, x_n]$ is a Groebner basis with respect to grevlex.
7. Consider the equations
   $$\begin{aligned} t^2 + x^2 + y^2 + z^2 &= 0, \\ t^2 + 2x^2 - xy - z^2 &= 0, \\ t + y^3 - z^3 &= 0. \end{aligned}$$
   We want to eliminate t. Let $I = \langle t^2 + x^2 + y^2 + z^2, t^2 + 2x^2 - xy - z^2, t + y^3 - z^3 \rangle$ be the corresponding ideal.
   a. Using lex order with $t > x > y > z$, compute a Groebner basis for I, and then find a basis for $I \cap k[x, y, z]$. You should get four generators, one of which has total degree 12.
   b. Use grevlex to compute a Groebner basis for $I \cap k[x, y, z]$. You will get a simpler set of two generators.
   c. Combine the answer to part b with the polynomial $t + y^3 - z^3$ and show that this gives a Groebner basis for I with respect to the elimination order $>_1$ (this is $>_l$ with $l = 1$) of Exercise 6. Notice that this Groebner basis is much simpler than the one found in part a. If you have access to a computer algebra system that knows elimination orders, then check your answer.
8. In equation (6), we showed that $z \neq 0$ could be specified arbitrarily. Hence, z can be regarded as a "parameter." To emphasize this point, show that there are formulas for x and y in terms of z. Hint: Use g_1 and the quadratic formula to get y in terms of z. Then use $xyz = 1$ to get x. The formulas you obtain give a "parametrization" of $\mathbf{V}(I)$ which is different from those studied in §3 of Chapter 1. Namely, in Chapter 1, we used parametrizations by *rational* functions, whereas here, we have what is called a parametrization by *algebraic* functions. Note that x and y are *not* uniquely determined by z.
9. Consider the system of equations given by
   $$\begin{aligned} x^5 + \frac{1}{x^5} &= y, \\ x + \frac{1}{x} &= z. \end{aligned}$$
   Let I be the ideal in $\mathbb{C}[x, y, z]$ determined by these equations.

a. Find a basis of $I_1 \subset \mathbb{C}[y, z]$ and show that $I_2 = \{0\}$.
b. Use the Extension Theorem to prove that each partial solution $c \in \mathbf{V}(I_2) = \mathbb{C}$ extends to a solution in $\mathbf{V}(I) \subset \mathbb{C}^3$.
c. Which partial solutions $(y, z) \in \mathbf{V}(I_1) \subset \mathbb{R}^2$ extend to solutions in $\mathbf{V}(I) \subset \mathbb{R}^3$. Explain why your answer does not contradict the Extension Theorem.
d. If we regard z as a "parameter" (see the previous problem), then solve for x and y as algebraic functions of z to obtain a "parametrization" of $\mathbf{V}(I)$.

§2 The Geometry of Elimination

In this section, we will give a geometric interpretation of the theorems proved in §1. The main idea is that elimination corresponds to projecting a variety onto a lower dimensional subspace. We will also discuss the Closure Theorem, which describes the relation between partial solutions and elimination ideals. For simplicity, we will work over the field $k = \mathbb{C}$.

Let us start by defining the projection of an affine variety. Suppose that we are given $V = \mathbf{V}(f_1, \ldots, f_s) \subset \mathbb{C}^n$. To eliminate the first l variables $x_1, \ldots, x_l$, we will consider the *projection map*

$$\pi_l : \mathbb{C}^n \to \mathbb{C}^{n-l}$$

which sends $(a_1, \ldots, a_n)$ to $(a_{l+1}, \ldots, a_n)$. If we apply π_l to $V \subset \mathbb{C}^n$, then we get $\pi_l(V) \subset \mathbb{C}^{n-l}$. We can relate $\pi_l(V)$ to the l-th elimination ideal as follows.

Lemma 1. *With the above notation, let $I_l = \langle f_1, \ldots, f_s\rangle \cap \mathbb{C}[x_{l+1}, \ldots, x_n]$ be the l-th elimination ideal. Then, in $\mathbb{C}^{n-l}$, we have*

$$\pi_l(V) \subset \mathbf{V}(I_l).$$

Proof. Fix a polynomial $f \in I_l$. If $(a_1, \ldots, a_n) \in V$, then f vanishes at $(a_1, \ldots, a_n)$ since $f \in \langle f_1, \ldots, f_s\rangle$. But f involves only $x_{l+1}, \ldots, x_n$, so that we can write

$$f(a_{l+1}, \ldots, a_n) = f(\pi_l(a_1, \ldots, a_n)) = 0.$$

This shows that f vanishes at all points of $\pi_l(V)$. □

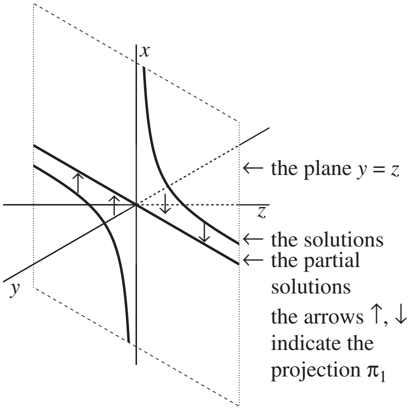
As in §1, points of $\mathbf{V}(I_l)$ will be called *partial solutions*. Using the lemma, we can write $\pi_l(V)$ as follows:

$$\begin{aligned}\pi_l(V) = \{(a_{l+1}, \ldots, a_n) \in \mathbf{V}(I_l) : \exists a_1, \ldots, a_l \in \mathbb{C} \\ \text{with } (a_1, \ldots, a_l,\ a_{l+1}, \ldots, a_n) \in V\}.\end{aligned}$$

Thus, $\pi_l(V)$ consists *exactly* of the partial solutions that extend to complete solutions. For an example of this, consider the variety V defined by equations (4) from §1:

$$\begin{aligned} xy &= 1, \\ xz &= 1. \end{aligned} \tag{1}$$

Here, we have the following picture that simultaneously shows the solutions and the partial solutions:



Note that $\mathbf{V}(I_1)$ is the line $y = z$ in the yz-plane, and that

$$\pi_1(V) = \{(a, a) \in \mathbb{C}^2 : a \neq 0\}.$$

In particular, $\pi_1(V)$ is *not an affine variety*—it is missing the point (0, 0).

The basic tool to understand the missing points is the Extension Theorem from §1. It only deals with π_1 (i.e., eliminating x_1), but gives us a good picture of what happens in this case. Stated geometrically, here is what the Extension Theorem says.

Theorem 2. *Given $V = \mathbf{V}(f_1, \ldots, f_s) \subset \mathbb{C}^n$, let g_i be as in the Extension Theorem from §1. If I_1 is the first elimination ideal of $\langle f_l, \ldots f_s \rangle$, then we have the equality in $\mathbb{C}^{n-1}$*

$$\mathbf{V}(I_1) = \pi_1(V) \cup (\mathbf{V}(g_1, \ldots, g_s) \cap \mathbf{V}(I_1)),$$

where $\pi_1 : \mathbb{C}^n \to \mathbb{C}^{n-1}$ is projection onto the last $n - 1$ components.

Proof. The proof follows from Lemma 1 and the Extension Theorem. The details will be left as an exercise. □

This theorem tells us that $\pi_1(V)$ fills up the affine variety $\mathbf{V}(I_1)$, except possibly for a part that lies in $\mathbf{V}(g_1, \ldots, g_s)$. Unfortunately, it is not clear how big this part is, and sometimes $\mathbf{V}(g_1, \ldots, g_s)$ is unnaturally large. For example, one can show that the equations

$$(2) \qquad \begin{aligned} (y - z)x^2 + xy &= 1, \\ (y - z)x^2 + xz &= 1 \end{aligned}$$

generate the same ideal as equations (1). Since $g_1 = g_2 = y - z$ generate the elimination ideal I_1, the Geometric Extension Theorem tells us *nothing* about the size of $\pi_1(V)$ in this case.

Nevertheless, we can still make the following strong statements about the relation between $\pi_l(V)$ and $\mathbf{V}(I_l)$.

Theorem 3 (The Closure Theorem). *Let $V = \mathbf{V}(f_1, \ldots, f_s) \subset \mathbb{C}^n$ and let I_l be the l-th elimination ideal of $\langle f_1, \ldots, f_s \rangle$. Then:*
(i) *$\mathbf{V}(I_l)$ is the smallest affine variety containing $\pi_l(V) \subset \mathbb{C}^{n-l}$.*
(ii) *When $V \neq \emptyset$, there is an affine variety $W \subsetneqq \mathbf{V}(I_l)$ such that $\mathbf{V}(I_l) - W \subset \pi_l(V)$.*

Proof. When we say "smallest variety" in part (i), we mean "smallest with respect to set-theoretic inclusion." Thus, $\mathbf{V}(I_l)$ being smallest means two things:

- $\pi_l(V) \subset \mathbf{V}(I_l)$
- If Z is *any other* affine variety in $\mathbb{C}^{n-l}$ containing $\pi_l(V)$, then $\mathbf{V}(I_l) \subset Z$.

In Chapter 4, we will express this by saying that $\mathbf{V}(I_l)$ is the *Zariski closure* of $\pi_l(V)$. This is where the theorem gets its name. We cannot yet prove part (i) of the theorem, for it requires the Nullstellensatz. The proof will be given in Chapter 4.

The second part of the theorem says that although $\pi_l(V)$ may not equal $\mathbf{V}(I_l)$, it fills up "most" of $\mathbf{V}(I_l)$ in the sense that what is missing lies in a strictly smaller affine variety. We will only prove this part of the theorem in the special case when $l = 1$. The proof when $l > 1$ will be given in §6 of Chapter 5.

The main tool we will use is the decomposition

$$\mathbf{V}(I_1) = \pi_1(V) \cup (\mathbf{V}(g_1, \ldots, g_s) \cap \mathbf{V}(I_1))$$

from the Geometric Extension Theorem. Let $W = \mathbf{V}(g_1, \ldots, g_s) \cap \mathbf{V}(I_1)$ and note that W is an affine variety by Lemma 2 of Chapter 1, §2. The above decomposition implies that $\mathbf{V}(I_1) - W \subset \pi_1(V)$, and thus we are done if $W \neq \mathbf{V}(I_1)$. However, as example (2) indicates, it can happen that $W = \mathbf{V}(I_1)$.

In this case, we need to change the equations defining V so that W becomes smaller. The key observation is that

$$(3) \qquad \text{if } W = \mathbf{V}(I_1), \text{ then } V = \mathbf{V}(f_1, \ldots, f_s, g_1, \ldots, g_s).$$

This is proved as follows. First, since we are adding more equations, it is obvious that $\mathbf{V}(f_1, \ldots, f_s, g_1, \ldots, g_s) \subset \mathbf{V}(f_1, \ldots, f_s) = V$. For the opposite inclusion, let $(a_1, \ldots, a_n) \in V$. Certainly each f_i vanishes at this point, and since $(a_2, \ldots, a_n) \in \pi_1(V) \subset \mathbf{V}(I_1) = W$, it follows that the g_i's vanish here. Thus, $(a_1, \ldots, a_n) \in \mathbf{V}(f_1, \ldots, f_s, g_1, \ldots, g_s)$, which completes the proof of (3).

Let $I = \langle f_1, \ldots, f_s \rangle$ be our original ideal and let $\tilde{I}$ be the ideal $\langle f_1, \ldots, f_s, g_1, \ldots, g_s \rangle$. Notice that I and $\tilde{I}$ may be different, even though they have the same variety V [proved in (3) above]. Thus, the corresponding elimination ideals I_1 and $\tilde{I}_1$ may differ. However, since $\mathbf{V}(I_1)$ and $\mathbf{V}(\tilde{I}_1)$ are both the smallest variety containing $\pi_1(V)$ [by part (i) of the theorem], it follows that $\mathbf{V}(I_1) = \mathbf{V}(\tilde{I}_1)$.

The next step is to find a better basis of $\tilde{I}$. First, recall that the g_i's are defined by writing

$$f_i = g_i(x_2, \ldots, x_n)x_1^{N_i} + \text{terms in which } x_1 \text{ has degree} < N_i,$$

where $N_i \geq 0$ and $g_i \in \mathbb{C}[x_2, \ldots, x_n]$ is nonzero. Now set

$$\tilde{f}_i = f_i - g_i x_1^{N_i}.$$

For each i, note that $\tilde{f}_i$ is either zero or has strictly smaller degree in x_1 than f_i. We leave it as an exercise to show that

$$\tilde{I} = \langle \tilde{f}_1, \ldots, \tilde{f}_s, g_1, \ldots, g_s \rangle.$$

Now apply the Geometric Extension Theorem to $V = \mathbf{V}(\tilde{f}_1, \ldots, \tilde{f}_s, g_1, \ldots, g_s)$. Note that the leading coefficients of the generators are different, so that we get a different decomposition

$$\mathbf{V}(I_1) = \mathbf{V}(\tilde{I}_1) = \pi_1(V) \cup \widetilde{W},$$

where $\widetilde{W}$ consists of those partial solutions where the leading coefficients of $\tilde{f}_1, \ldots, \tilde{f}_s, g_1, \ldots, g_s$ vanish.

Before going further with the proof, let us give an example to illustrate how $\widetilde{W}$ can be smaller than W. As in example (2), let $I = \langle (y-z)x^2 + xy - 1, (y-z)x^2 + xz - 1 \rangle$. We know that $I_1 = \langle y - z \rangle$ and $g_1 = g_2 = y - z$, so that $W = \mathbf{V}(I_1)$ in this case. Then it is easy to check that the process described earlier yields the ideal

$$\tilde{I} = \langle (y-z)x^2 + xy - 1, (y-z)x^2 + xz - 1, y - z \rangle = \langle xy - 1, xz - 1, y - z \rangle.$$

Applying the Geometric Extension Theorem to $\tilde{I}$, one finds that $\widetilde{W}$ consists of the partial solutions where y and z vanish, i.e., $\widetilde{W} = \{(0, 0)\}$, which is strictly smaller than $W = \mathbf{V}(I_1)$.

Unfortunately, in the general case, there is nothing to guarantee that $\widetilde{W}$ will be strictly smaller. So it still could happen that $\widetilde{W} = \mathbf{V}(I_1)$. If this is the case, we simply repeat the above process. If at any subsequent stage we get something strictly smaller than $\mathbf{V}(I_1)$, then we are done.

It remains to consider what happens when we *always* get $\mathbf{V}(I_1)$. Each time we do the above process, the degrees in x_1 of the generators drop (or remain at zero), so that eventually all of the generators will have degree 0 in x_1. This means that V can be defined by the vanishing of polynomials in $\mathbb{C}[x_2, \ldots, x_n]$. Thus, if $(a_2, \ldots, a_n)$ is a partial solution, it follows that $(a_1, a_2, \ldots, a_n) \in V$ for *any* $a_1 \in \mathbb{C}$ since x_1 does not appear in the defining equations. Hence *every* partial solution extends, which proves that $\pi_1(V) = \mathbf{V}(I_1)$. In this case, we see that part (ii) of the theorem is satisfied when $W = \emptyset$ (this is where we use the assumption $V \neq \emptyset$). The theorem is now proved. □

The Closure Theorem gives us a partial description of $\pi_l(V)$ since it fills up $\mathbf{V}(I_l)$, except for some missing points that lie in a variety strictly smaller than $\mathbf{V}(I_l)$. Unfortunately, the missing points might not fill up all of the smaller variety. The precise structure of $\pi_l(V)$ can be described as follows: there are affine varieties $Z_i \subset W_i \subset \mathbb{C}^{n-l}$ for $1 \leq i \leq m$ such that

$$\pi_l(V) = \bigcup_{i=1}^{m} (W_i - Z_i).$$

In general, a set of this form is called *constructible*. We will prove this in §6 of Chapter 5.

In §1, we saw that the nicest case of the Extension Theorem was when one of the leading coefficients g_i was a nonzero constant. Then the g_i's can *never* simultaneously vanish at a point $(a_2, \ldots, a_n)$, and, consequently, partial solutions *always* extend in this case. Thus, we have the following geometric version of Corollary 4 of §1.

Corollary 4. *Let $V = \mathbf{V}(f_1, \ldots, f_s) \subset \mathbb{C}^n$, and assume that for some i, f_i is of the form*

$$f_i = cx_1^N + \textit{terms in which } x_1 \textit{ has degree} < N,$$

where $c \in \mathbb{C}$ is nonzero and $N > 0$. If I_1 is the first elimination ideal, then in $\mathbb{C}^{n-1}$

$$\pi_1(V) = \mathbf{V}(I_1),$$

where π_1 is the projection on the last $n - 1$ components.

A final point we need to make concerns fields. The Extension Theorem and the Closure Theorem (and their corollaries) are stated for the field of complex numbers $\mathbb{C}$. In §6, we will see that the Extension Theorem actually holds for any *algebraically closed* field k, and in Chapter 4, we will show that the same is true for the Closure Theorem.

EXERCISES FOR §2

1. Prove the Geometric Extension Theorem (Theorem 2) using the Extension Theorem and Lemma 1.
2. In example (2), verify carefully that $\langle (y-z)x^2+xy-1, (y-z)x^2+xz-1 \rangle = \langle xy-1, xz-1 \rangle$. Also check that $y - z$ vanishes at all partial solutions in $\mathbf{V}(I_1)$.
3. In this problem, we will work through the proof of Theorem 3 in the special case when $I = \langle f_1, f_2, f_3 \rangle$, where

$$\begin{aligned} f_1 &= yx^3 + x^2, \\ f_2 &= y^3x^2 + y^2, \\ f_3 &= yx^4 + x^2 + y^2. \end{aligned}$$

   a. Find a Groebner basis for I and show that $I_1 = \langle y^2 \rangle$.
   b. Show that $\mathbf{V}(I_1) = \mathbf{V}(I_1) \cap \mathbf{V}(g_1, g_2, g_3)$, where g_i is the coefficient of the highest power of x in f_i. In the notation of Theorem 3, this is a case when $W = \mathbf{V}(I_1)$.
   c. Let $\tilde{I} = \langle f_1, f_2, f_3, g_1, g_2, g_3 \rangle$. Show that $I \neq \tilde{I}$ and that $\mathbf{V}(I) = \mathbf{V}(\tilde{I})$. Also check that $\mathbf{V}(I_1) = \mathbf{V}(\tilde{I}_1)$.
   d. Follow the procedure described in the text for producing a new basis for $\tilde{I}$. Using this new basis, show that $\widetilde{W} \neq \mathbf{V}(I_1)$.

4. Let $f_i, g_i, h_i \in k[x_1, \ldots, x_n]$ for $1 \leq i \leq s$. If we set $\tilde{f_i} = f_i + h_i g_i$, then prove that

$$\langle f_1, \ldots, f_s, g_1, \ldots, g_s \rangle = \langle \tilde{f_1}, \ldots, \tilde{f_s}, g_1, \ldots, g_s \rangle.$$

   Then explain how the proof of Theorem 3 used a special case of this result.

5. To see how the Closure Theorem can fail over $\mathbb{R}$, consider the ideal

$$I = \langle x^2 + y^2 + z^2 + 2, 3x^2 + 4y^2 + 4z^2 + 5 \rangle.$$

   Let $V = \mathbf{V}(I)$, and let π_1 be the projection taking (x, y, z) to (y, z).
   a. Working over $\mathbb{C}$, prove that $V(I_1) = \pi_1(V)$.
   b. Working over $\mathbb{R}$, prove that $V = \emptyset$ and that $\mathbf{V}(I_1)$ is infinite. Thus, $\mathbf{V}(I_1)$ may be much larger than the smallest variety containing $\pi_1(V)$ when the field is not algebraically closed.

§3 Implicitization

In Chapter 1, we saw that a variety V can sometimes be described using parametric equations. The basic idea of the *implicitization problem* is to convert the parametrization into defining equations for V. The name "implicitization" comes from Chapter 1, where the equations defining V were called an "implicit representation" of V. However, some care is required in giving a precise formulation of implicitization. The problem is that the parametrization need not fill up all of the variety V—an example is given by equation (4) from Chapter 1, §3. So the implicitization problem really asks for the equations defining the *smallest* variety V containing the parametrization. In this section, we will use the elimination theory developed in §§1 and 2 to give a complete solution of the implicitization problem.

Furthermore, once the smallest variety V has been found, two other interesting questions arise. First, does the parametrization fill up all of V? Second, if there are missing points, how do we find them? As we will see, Groebner bases and the Extension Theorem give us powerful tools for studying this situation.

To illustrate these issues in a specific case, let us look at the tangent surface to the twisted cubic in $\mathbb{R}^3$, first studied in Chapter 1, §3. Recall that this surface is given parametrically by

$$\begin{aligned} x &= t + u, \\ y &= t^2 + 2tu, \\ z &= t^3 + 3t^2u. \end{aligned} \tag{1}$$

In §8 of Chapter 2, we used these equations to show that the tangent surface lies on the variety V in $\mathbb{R}^3$ defined by

$$x^3z - (3/4)x^2y^2 - (3/2)xyz + y^3 + (1/4)z^2 = 0.$$

However, we do not know if V is the smallest variety containing the tangent surface and, thus, we cannot claim to have solved the implicitization problem. Furthermore, even if V is the smallest variety, we still do not know if the tangent surface fills it up completely. So there is a lot of work to do.

We begin our solution of the implicitization problem with the case of a *polynomial parametrization*, which is specified by the data

$$\begin{aligned} x_1 &= f_1(t_1,\ldots,t_m),\\ &\ \vdots\\ x_n &= f_n(t_1,\ldots,t_m). \end{aligned} \tag{2}$$

Here, $f_1,\ldots,f_n$ are polynomials in $k[t_1,\ldots,t_m]$. We can think of this geometrically as the function

$$F : k^m \longrightarrow k^n$$

defined by

$$F(t_1,\ldots,t_m) = (f_1(t_1,\ldots,t_m),\ldots,f_n(t_1,\ldots,t_m)).$$

Then $F(k^m) \subset k^n$ is the subset of k^n parametrized by equations (2). Since $F(k^m)$ may not be an affine variety (examples will be given in the exercises), a solution of the implicitization problem means finding the smallest affine variety that contains $F(k^m)$.

We can relate implicitization to elimination as follows. Equations (2) define a variety $V = \mathbf{V}(x_1 - f_1,\ldots,x_n - f_n) \subset k^{n+m}$. Points of V can be written in the form

$$(t_1,\ldots,t_m, f_1(t_1,\ldots,t_m),\ldots,f_n(t_1,\ldots,t_m)),$$

which shows that V can be regarded as the *graph* of the function F. We also have two other functions

$$\begin{aligned} i &: k^m \longrightarrow k^{n+m},\\ \pi_m &: k^{n+m} \longrightarrow k^n \end{aligned}$$

defined by

$$\begin{aligned} i(t_1,\ldots,t_m) &= (t_1,\ldots,t_m, f_1(t_1,\ldots,t_m),\ldots,f_n(t_1,\ldots,t_m))\\ \pi_m(t_1,\ldots,t_m,x_1,\ldots,x_n) &= (x_1,\ldots,x_n). \end{aligned}$$

This gives us the following diagram of sets and maps:

$$\begin{tikzcd} & k^{n+m} \arrow[dr, "\pi_m"] & \\ k^m \arrow[ur, "i"] \arrow[rr, "F"] & & k^n \end{tikzcd} \tag{3}$$

Note that F is then the composition $F = \pi_m \circ i$. It is also straightforward to show that $i(k^m) = V$. Thus, we obtain

$$F(k^m) = \pi_m(i(k^m)) = \pi_m(V). \tag{4}$$

In more concrete terms, this says that the image of the parametrization is the projection of its graph. We can now use elimination theory to find the smallest variety containing $F(k^m)$.

Theorem 1 (Polynomial Implicitization). *If k is an infinite field, let $F : k^m \to k^n$ be the function determined by the polynomial parametrization* (2). *Let I be the ideal $I = \langle x_1 - f_1, \dots, x_n - f_n\rangle \subset k[t_1, \dots, t_m, x_1, \dots, x_n]$ and let $I_m = I \cap k[x_1, \dots, x_n]$ be the m-th elimination ideal. Then $\mathbf{V}(I_m)$ is the smallest variety in k^n containing $F(k^m)$.*

Proof. Let $V = \mathbf{V}(I) \subset k^{n+m}$. The above discussion shows that V is the graph of $F : k^m \to k^n$. Now assume that $k = \mathbb{C}$. By (4), we have $F(\mathbb{C}^m) = \pi_m(V)$, and by the Closure Theorem from §2, we know that $\mathbf{V}(I_m)$ is the smallest variety containing $\pi_m(V)$. This proves the theorem when $k = \mathbb{C}$.

Next, suppose that k is a subfield of $\mathbb{C}$. This means that $k \subset \mathbb{C}$ and that k inherits its field operations from $\mathbb{C}$. Such a field always contains the integers $\mathbb{Z}$ (in fact, it contains $\mathbb{Q}$—do you see why?) and, thus, is infinite. Since k may be strictly smaller than $\mathbb{C}$, we cannot use the Closure Theorem directly. Our strategy will be to switch back and forth between k and $\mathbb{C}$, and we will use the subscript k or $\mathbb{C}$ to keep track of which field we are working with. Thus, $\mathbf{V}_k(I_m)$ is the variety we get in k^n, whereas $\mathbf{V}_{\mathbb{C}}(I_m)$ is the larger set of solutions in $\mathbb{C}^n$. (Note that going to the larger field does not change the elimination ideal I_m. This is because the algorithm used to compute the elimination ideal is unaffected by changing from k to $\mathbb{C}$.) We need to prove that $\mathbf{V}_k(I_m)$ is the smallest variety in k^n containing $F(k^m)$.

By equation (4) of this section and Lemma 1 of §2, we know that $F(k^m) = \pi_m(V_k) \subset \mathbf{V}_k(I_m)$. Now let $Z_k = \mathbf{V}_k(g_1, \dots, g_s) \subset k^n$ be any variety of k^n such that $F(k^m) \subset Z_k$. We must show $\mathbf{V}_k(I_m) \subset Z_k$. We begin by noting that $g_i = 0$ on Z_k and, hence, $g_i = 0$ on the smaller set $F(k^m)$. This shows that each $g_i \circ F$ vanishes on all of k^m. But g_i is a polynomial in $k[x_1, \dots, x_n]$, and $F = (f_1, \dots, f_n)$ is made up of polynomials in $k[t_1, \dots, t_m]$. It follows that $g_i \circ F \in k[t_1, \dots, t_m]$.

Thus, the $g_i \circ F$'s are polynomials that vanish on k^m. Since k is infinite, Proposition 5 of Chapter 1, §1 implies that each $g_i \circ F$ is the zero polynomial. In particular, this means that $g_i \circ F$ also vanishes on $\mathbb{C}^m$, and thus the g_i's vanish on $F(\mathbb{C}^m)$. Hence, $Z_{\mathbb{C}} = \mathbf{V}_{\mathbb{C}}(g_1, \dots, g_s)$ is a variety of $\mathbb{C}^n$ containing $F(\mathbb{C}^m)$. Since the theorem is true for $\mathbb{C}$, it follows that $\mathbf{V}_{\mathbb{C}}(I_m) \subset Z_{\mathbb{C}}$ in $\mathbb{C}^n$. If we then look at the solutions that lie in k^n, it follows immediately that $\mathbf{V}_k(I_m) \subset Z_k$. This proves that $\mathbf{V}_k(I_m)$ is the smallest variety of k^n containing $F(k^m)$.

Finally, if k is a field not contained in $\mathbb{C}$, one can prove that there is an algebraically closed field K such that $k \subset K$ [see Chapter VII, §2 of LANG (1965)]. As we remarked at the end of §2, the Closure Theorem holds over any algebraically closed field. Then the theorem follows using the above argument with $\mathbb{C}$ replaced by K. □

Theorem 1 gives the following **implicitization algorithm for polynomial parametrizations:** if we have $x_i = f_i(t_1, \dots, t_m)$ for polynomials $f_1, \dots, f_n \in k[t_1, \dots, t_m]$, consider the ideal $I = \langle x_1 - f_1, \dots, x_n - f_n\rangle$ and compute a Groebner basis with respect to a lexicographic ordering where every t_i is greater than every x_i. By the Elimination Theorem, the elements of the Groebner basis not involving $t_1, \dots, t_m$ form a basis of I_m, and by Theorem 1, they define the smallest variety in k^n containing the parametrization.

For an example of how this algorithm works, let us look at the tangent surface to the twisted cubic in $\mathbb{R}^3$, which is given by the polynomial parametrization (1). Thus, we need to consider the ideal

$$I = \langle x - t - u, y - t^2 - 2tu, z - t^3 - 3t^2u\rangle \subset \mathbb{R}[t, u, x, y, z].$$

Using lex order with $t > u > x > y > z$, a Groebner basis for I is given by

$$\begin{aligned}
g_1 &= t + u - x,\\
g_2 &= u^2 - x^2 + y,\\
g_3 &= ux^2 - uy - x^3 + (3/2)xy - (1/2)z,\\
g_4 &= uxy - uz - x^2y - xz + 2y^2,\\
g_5 &= uxz - uy^2 + x^2z - (1/2)xy^2 - (1/2)yz,\\
g_6 &= uy^3 - uz^2 - 2x^2yz + (1/2)xy^3 - xz^2 + (5/2)y^2z,\\
g_7 &= x^3z - (3/4)x^2y^2 - (3/2)xyz + y^3 + (1/4)z^2.
\end{aligned}$$

The Elimination Theorem tells us that $I_2 = I \cap \mathbb{R}[x, y, z] = \langle g_7\rangle$, and thus by Theorem 1, $\mathbf{V}(g_7)$ solves the implicitization problem for the tangent surface of the twisted cubic. The equation $g_7 = 0$ is exactly the one given at the start of this section, but now we know it defines the smallest variety in $\mathbb{R}^3$ containing the tangent surface.

But we still do not know if the tangent surface fills up *all* of $\mathbf{V}(g_7) \subset \mathbb{R}^3$. To answer this question, we must see whether all partial solutions $(x, y, z) \in \mathbf{V}(g_7) = \mathbf{V}(I_2)$ extend to $(t, u, x, y, z) \in \mathbf{V}(I)$. We will first work over $\mathbb{C}$ so that we can use the Extension Theorem. As usual, our strategy will be to add one coordinate at a time.

Let us start with $(x, y, z) \in \mathbf{V}(I_2) = \mathbf{V}(g_7)$. In §1, we observed that I_2 is the first elimination ideal of I_1. Further, the Elimination Theorem tells us that $I_1 = \langle g_2, \ldots, g_7\rangle$. Then the Extension Theorem, in the form of Corollary 4 of §1, implies that (x, y, z) always extends to $(u, x, y, z) \in \mathbf{V}(I_1)$ since I_1 has a generator with a constant leading coefficient of u (we leave it to you to find which of $g_2, \ldots, g_7$ has this property). Going from $(u, x, y, z) \in \mathbf{V}(I_1)$ to $(t, u, x, y, z) \in \mathbf{V}(I)$ is just as easy: using Corollary 4 of §1 again, we can always extend since $g_1 = t + u - x$ has a constant leading coefficient of t. We have thus proved that the tangent surface to the twisted cubic equals $\mathbf{V}(g_7)$ in $\mathbb{C}^3$.

It remains to see what happens over $\mathbb{R}$. If we start with a real solution $(x, y, z) \in \mathbb{R}^3$ of $g_7 = 0$, we know that it extends to $(t, u, x, y, z) \in \mathbf{V}(I) \subset \mathbb{C}^5$. But are the parameters t and u real? This is not immediately obvious. However, if you look at the above Groebner basis, you can check that t and u are real when $(x, y, z) \in \mathbb{R}^3$ (see Exercise 4). It follows that the tangent surface to the twisted cubic in $\mathbb{R}^3$ equals the variety defined by

$$x^3z - (3/4)x^2y^2 - (3/2)xyz + y^3 + (1/4)z^2 = 0.$$

In general, the question of whether a parametrization fills up all of its variety can be difficult to answer. Each case has to be analyzed separately. But as indicated by the example just completed, the combination of Groebner bases and the Extension Theorem can shed considerable light on what is going on.

In our discussion of implicitization, we have thus far only considered polynomial parametrizations. The next step is to see what happens when we have a parametrization by rational functions. To illustrate the difficulties that can arise, consider the following rational parametrization:

$$\begin{aligned} x &= \frac{u^2}{v}, \\ y &= \frac{v^2}{u}, \\ z &= u. \end{aligned} \tag{5}$$

It is easy to check that the point (x, y, z) always lies on the surface $x^2y = z^3$. Let us see what happens if we clear denominators in the above equations and apply the polynomial implicitization algorithm. We get the ideal

$$I = \langle vx - u^2, uy - v^2, z - u\rangle \subset k[u, v, x, y, z],$$

and we leave it as an exercise to show that $I_2 = I \cap k[x, y, z]$ is given by $I_2 = \langle z(x^2y - z^3)\rangle$. This implies that

$$\mathbf{V}(I_2) = \mathbf{V}(x^2y - z^3) \cup \mathbf{V}(z),$$

and, in particular, $\mathbf{V}(I_2)$ is *not* the smallest variety containing the parametrization. So the above ideal I is not what we want—simply "clearing denominators" is too naive. To find an ideal that works better, we will need to be more clever.

In the general situation of a rational parametrization, we have

$$\begin{aligned} x_1 &= \frac{f_1(t_1, \ldots, t_m)}{g_1(t_1, \ldots, t_m)}, \\ &\ \vdots \\ x_n &= \frac{f_n(t_1, \ldots, t_m)}{g_n(t_1, \ldots, t_m)}, \end{aligned} \tag{6}$$

where $f_1, g_1, \ldots, f_n, g_n$ are polynomials in $k[t_1, \ldots, t_m]$. The map F from k^m to k^n given by (6) may not be defined on all of k^m because of the denominators. But if we let $W = \mathbf{V}(g_1g_2\cdots g_n) \subset k^m$, then it is clear that

$$F(t_1, \ldots, t_m) = \left(\frac{f_1(t_1, \ldots, t_m)}{g_1(t_1, \ldots, t_m)}, \ldots, \frac{f_n(t_1, \ldots, t_m)}{g_n(t_1, \ldots, t_m)}\right)$$

defines a map

$$F : k^m - W \longrightarrow k^n.$$

To solve the implicitization problem, we need to find the smallest variety of k^n containing $F(k^m - W)$.

We can adapt diagram (3) to this case by writing

(7)
$$\begin{tikzcd} & k^{n+m} \arrow[dr, "\pi_m"] & \\ k^m - W \arrow[ur, "i"] \arrow[rr, "F"] & & k^n \end{tikzcd}$$

It is easy to check that $i(k^m - W) \subset \mathbf{V}(I)$, where $I = \langle g_1x_1 - f_1, \ldots, g_nx_n - f_n\rangle$ is the ideal obtained by "clearing denominators." The problem is that $\mathbf{V}(I)$ may *not* be the smallest variety containing $i(k^m - W)$. In the exercises, you will see that (5) is such an example.

To avoid this difficulty, we will alter the ideal I slightly by using an extra dimension to control the denominators. Consider the polynomial ring $k[y, t_1, \ldots, t_m, x_1, \ldots, x_n]$ which gives us the affine space k^{n+m+1}. Let g be the product $g = g_1 \cdot g_2 \cdots g_n$, so that $W = \mathbf{V}(g)$. Then consider the ideal

$$J = \langle g_1x_1 - f_1, \ldots, g_nx_n - f_n, 1 - gy\rangle \subset k[y, t_1, \ldots, t_m, x_1, \ldots, x_n].$$

Note that the equation $1-gy = 0$ means that the denominators $g_1, \ldots, g_n$ never vanish on $\mathbf{V}(J)$. To adapt diagram (7) to this new situation, consider the maps

$$j : k^m - W \longrightarrow k^{n+m+1},$$
$$\pi_{m+1} : k^{n+m+1} \longrightarrow k^n$$

defined by

$$j(t_1, \ldots, t_m) = \left(\frac{1}{g(t_1, \ldots, t_m)}, t_1, \ldots, t_m, \frac{f_1(t_1, \ldots, t_m)}{g_1(t_1, \ldots, t_m)}, \ldots, \frac{f_n(t_1, \ldots, t_m)}{g_n(t_1, \ldots, t_m)}\right),$$
$$\pi_{m+1}(y, t_1, \ldots, t_m, x_1, \ldots, x_n) = (x_1, \ldots, x_n),$$

respectively. Then we get the diagram

$$\begin{tikzcd} & k^{n+m+1} \arrow[dr, "\pi_{m+1}"] & \\ k^m - W \arrow[ur, "j"] \arrow[rr, "F"] & & k^n \end{tikzcd}$$

As before, we have $F = \pi_{m+1} \circ j$. The surprise is that $j(k^m - W) = \mathbf{V}(J)$ in k^{n+m+1}. To see this, note that $j(k^m - W) \subset \mathbf{V}(J)$ follows easily from the definitions of j and J. Going the other way, if $(y, t_1, \ldots, t_m, x_1, \ldots, x_n) \in \mathbf{V}(J)$, then $g(t_1, \ldots, t_m)y = 1$ implies that *none* of the g_i's vanish at $(t_1, \ldots, t_m)$ and, thus, $g_i(t_i, \ldots, t_m)x_i = f_i(t_1, \ldots, t_m)$ can be solved for $x_i = f_i(t_1, \ldots, t_m)/g_i(t_1, \ldots, t_m)$. Since $y = 1/g(t_1, \ldots, t_m)$, it follows that our point lies in $j(k^m - W)$. This proves $\mathbf{V}(J) \subset j(k^m - W)$.

From $F = \pi_{m+1} \circ j$ and $j(k^m - W) = \mathbf{V}(J)$, we obtain

(8)
$$F(k^m - W) = \pi_{m+1}(j(k^m - W)) = \pi_{m+1}(\mathbf{V}(J)).$$

Thus, the image of the parametrization is the projection of the variety $\mathbf{V}(J)$. As with the polynomial case, we can now use elimination theory to solve the implicitization problem.

Theorem 2 (Rational Implicitization). *If k is an infinite field, let $F : k^m - W \to k^n$ be the function determined by the rational parametrization* (6). *Let J be the ideal $J = \langle g_1x_1 - f_1, \ldots, g_nx_n - f_n, 1 - gy\rangle \subset k[y, t_1, \ldots, t_m, x_1, \ldots, x_n]$, where $g = g_1 \cdot g_2 \cdots g_n$, and let $J_{m+1} = J \cap k[x_1, \ldots, x_n]$ be the* $(m+1)$*-th elimination ideal. Then $\mathbf{V}(J_{m+1})$ is the smallest variety in k^n containing $F(k^m - W)$.*

Proof. The proof is similar to the proof of Theorem 1. One uses equation (8) rather than equation (4). The only tricky point is showing that a polynomial vanishing on $k^m - W$ must be the zero polynomial. The exercises at the end of the section will take you through the details. □

The interpretation of Theorem 2 is very nice: given the rational parametrization (6), consider the equations

$$\begin{aligned} g_1x_1 &= f_1,\\ &\vdots\\ g_nx_n &= f_n,\\ g_1g_2\cdots g_ny &= 1. \end{aligned}$$

These equations are obtained from (6) by "clearing denominators" and adding a final equation (in the new variable y) to prevent the denominators from vanishing. Then eliminating $y, t_1, \ldots, t_m$ gives us the equations we want.

More formally, Theorem 2 implies the following **implicitization algorithm for rational parametrizations**: if we have $x_i = f_i/g_i$ for polynomials $f_1, g_1, \ldots, f_n, g_n \in k[t_1, \ldots, t_m]$, consider the new variable y and $J = \langle g_1x_1 - f_1, \ldots, g_nx_n - f_n, 1 - gy\rangle$, where $g = g_1 \cdots g_n$. Compute a Groebner basis with respect to a lexicographic ordering where y and every t_i are greater than every x_i. Then the elements of the Groebner basis not involving $y, t_1, \ldots, t_m$ define the smallest variety in k^n containing the parametrization.

Let us see how this algorithm works for example (5). Let w be the new variable, so that

$$J = \langle vx - u^2, uy - v^2, z - u, 1 - uvw\rangle \subset k[w, u, v, x, y, z].$$

One easily calculates that $J_3 = J \cap k[x, y, z] = \langle x^2y - z^3\rangle$, so that $\mathbf{V}(x^2y - z^3)$ is the variety determined by the parametrization (5). In the exercises, you will study how much of $\mathbf{V}(x^2y - z^3)$ is filled up by the parametrization.

We should also mention that in practice, resultants are often used to solve the implicitization problem. Implicitization for curves and surfaces is discussed in ANDERSON, GOLDMAN and SEDERBERG (1984a) and (1984b). Another reference is CANNY and MANOCHA (1992), which shows how implicitization of parametric surfaces can be done using multipolynomial resultants.

EXERCISES FOR §3

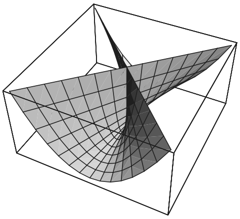
1. In diagram (3) in the text, prove carefully that $F = \pi_m \circ i$ and $i(k^m) = V$.
2. When $k = \mathbb{C}$, the conclusion of Theorem 1 can be strengthened. Namely, one can show that there is a variety $W \subsetneqq \mathbf{V}(I_m)$ such that $\mathbf{V}(I_m) - W \subset F(\mathbb{C}^m)$. Prove this using the Closure Theorem.
3. Give an example to show that Exercise 2 is false over $\mathbb{R}$. Hint: t^2 is always positive.
4. In the text, we proved that over $\mathbb{C}$, the tangent surface to the twisted cubic is defined by the equation
$$g_7 = x^3z - (3/4)x^2y^2 - (3/2)xyz + y^3 + (1/4)z^2 = 0.$$
We want to show that the same is true over $\mathbb{R}$. If (x, y, z) is a real solution of the above equation, then we proved (using the Extension Theorem) that there are $t, u \in \mathbb{C}$ such that
$$\begin{aligned} x &= t + u, \\ y &= t^2 + 2tu, \\ z &= t^3 + 3t^2u. \end{aligned}$$
Use the Groebner basis given in the text to show that t and u are real. This will prove that (x, y, z) is on the tangent surface in $\mathbb{R}^3$. Hint: First show that u is real.
5. In the parametrization of the tangent surface of the twisted cubic, show that the parameters t and u are uniquely determined by x, y, and z. Hint: The argument is similar to what you did in Exercise 4.
6. Let S be the parametric surface defined by
$$\begin{aligned} x &= uv, \\ y &= u^2, \\ z &= v^2. \end{aligned}$$
   a. Find the equation of the smallest variety V that contains S.
   b. Over $\mathbb{C}$, use the Extension Theorem to prove that $S = V$. Hint: The argument is similar to what we did for the tangent surface of the twisted cubic.
   c. Over $\mathbb{R}$, show that S only covers the "half" of V. What parametrization would cover the other "half"?
7. Let S be the parametric surface
$$\begin{aligned} x &= uv, \\ y &= uv^2, \\ z &= u^2. \end{aligned}$$
   a. Find the equation of the smallest variety V that contains S.
   b. Over $\mathbb{C}$, show that V contains points which are not on S. Determine exactly which points of V are not on S. Hint: Use lexicographic order with $u > v > x > y > z$.
8. The *Enneper surface* is defined parametrically by
$$\begin{aligned} x &= 3u + 3uv^2 - u^3, \\ y &= 3v + 3u^2v - v^3, \\ z &= 3u^2 - 3v^2. \end{aligned}$$

   a. Find the equation of the smallest variety V containing the Enneper surface. It will be a very complicated equation!
   b. Over $\mathbb{C}$, use the Extension Theorem to prove that the above equations parametrize the entire surface V. Hint: There are a lot of polynomials in the Groebner basis. Keep looking—you will find what you need.

9. The *Whitney umbrella surface* is given parametrically by

$$\begin{aligned} x &= uv, \\ y &= v, \\ z &= u^2. \end{aligned}$$

   A picture of this surface is:



   a. Find the equation of the smallest variety containing the Whitney umbrella.
   b. Show that the parametrization fills up the variety over $\mathbb{C}$ but not over $\mathbb{R}$. Over $\mathbb{R}$, exactly what points are omitted?
   c. Show that the parameters u and v are not always uniquely determined by x, y, and z. Find the points where uniqueness fails and explain how your answer relates to the picture.

10. Consider the curve in $\mathbb{C}^n$ parametrized by $x_i = f_i(t)$, where $f_1, \ldots f_n$ are polynomials in $\mathbb{C}[t]$. This gives the ideal

$$I = \langle x_1 - f_1(t), \ldots, x_n - f_n(t) \rangle \subset \mathbb{C}[t, x_1, \ldots, x_n].$$

   a. Prove that the parametric equations fill up all of the variety $\mathbf{V}(I_1) \subset \mathbb{C}^n$.
   b. Show that the conclusion of part a may fail if we let $f_1 \ldots, f_n$ be rational functions. Hint: See §3 of Chapter 1.
   c. Even if all of the f_i's are polynomials, show that the conclusion of part a may fail if we work over $\mathbb{R}$.

11. This problem is concerned with the proof of Theorem 2.
   a. Let k be an infinite field and let $f, g \in k[t_1, \ldots, t_m]$. Assume that $g \neq 0$ and that f vanishes on $k^m - V(g)$. Prove that f is the zero polynomial. Hint: Consider the product fg.
   b. Prove Theorem 2 using the hints given in the text.

12. Consider the parametrization (5) given in the text. For simplicity, let $k = \mathbb{C}$. Also let $I = \langle vx - u^2, uy - v^2, z - u \rangle$ be the ideal obtained by "clearing denominators."
   a. Show that $I_2 = \langle z(x^2y - z^3) \rangle$.
   b. Show that the smallest variety in $\mathbb{C}^5$ containing $i(\mathbb{C}^2 - W)$ [see diagram (7)] is $\mathbf{V}(vx - u^2, uy - v^2, z - u, x^2y - z^3, vz - xy)$. Hint: Show that $i(\mathbb{C}^2 - W) = \pi_1(\mathbf{V}(J))$, and then use the Closure Theorem.

c. Show that $\{(0, 0, x, y, 0) : x, y \text{ arbitrary}\} \subset \mathbf{V}(I)$ and conclude that $\mathbf{V}(I)$ is *not* the smallest variety containing $i(\mathbb{C}^2 - W)$.

d. Determine exactly which portion of $x^2y = z^3$ is parametrized by (5).

13. Given a rational parametrization as in (6), there is one case where the naive ideal $I = \langle g_1x_1 - f_1, \ldots, g_nx_n - f_n\rangle$ obtained by "clearing denominators" gives the right answer. Suppose that $x_i = f_i(t)/g_i(t)$ where there is only one parameter t. We can assume that for each i, $f_i(t)$ and $g_i(t)$ are relatively prime in $k[t]$ (so in particular, they have no common roots). If $I \subset k[t, x_1, \ldots, x_n]$ is as above, then prove that $\mathbf{V}(I_1)$ is the smallest variety containing $F(k - W)$, where as usual $g = g_1 \cdots g_n \in k[t]$ and $W = V(g) \subset k$. Hint: In diagram (7), show that $i(k^m - W) = \mathbf{V}(I)$, and adapt the proof of Theorem 1.

14. The *folium of Descartes* can be parametrized by

$$x = \frac{3t}{1 + t^3},$$
$$y = \frac{3t^2}{1 + t^3}.$$

a. Find the equation of the folium. Hint: Use Exercise 13.

b. Over $\mathbb{C}$ or $\mathbb{R}$, show that the above parametrization covers the entire curve.

15. In Exercise 16 to §3 of Chapter 1, we studied the parametric equations over $\mathbb{R}$

$$x = \frac{(1-t)^2x_1 + 2t(1-t)wx_2 + t^2x_3}{(1-t)^2 + 2t(1-t)w + t^2},$$
$$y = \frac{(1-t)^2y_1 + 2t(1-t)wy_2 + t^2y_3}{(1-t)^2 + 2t(1-t)w + t^2},$$

where $w, x_1, y_1, x_2, y_2, x_3, y_3$ are constants and $w > 0$. By eliminating t, show that these equations describe a portion of a conic section. Recall that a conic section is described by an equation of the form

$$ax^2 + bxy + cy^2 + dx + ey + f = 0.$$

Hint: In most computer algebra systems, the Groebner basis command allows polynomials to have coefficients involving symbolic constants like $w, x_1, y_1, x_2, y_2, x_3, y_3$.

§4 Singular Points and Envelopes

In this section, we will discuss two topics from geometry:

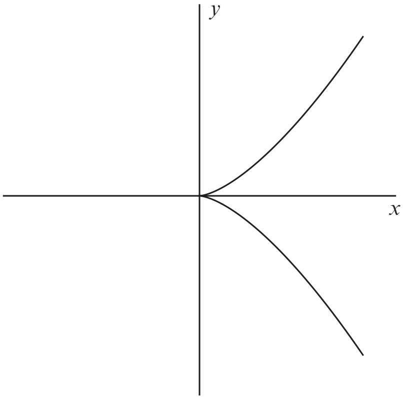
- the *singular points* on a curve,
- the *envelope* of a family of curves.

Our goal is to show how geometry provides interesting equations that can be solved by the techniques studied in §§1 and 2.

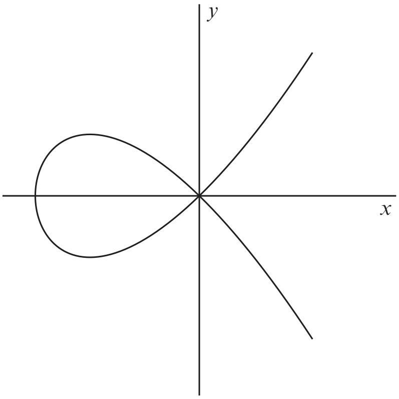
We will introduce some of the basic ideas concerning singular points and envelopes, but our treatment will be far from complete. One could write an entire book on these topics [see, for example, BRUCE and GIBLIN (1992)]. Also, our discussion of envelopes will not be fully rigorous. We will rely on some ideas from calculus to justify what is going on.

Singular Points

Suppose that we have a curve in the plane k^2 defined by $f(x, y) = 0$, where $f \in k[x, y]$. We expect that $\mathbf{V}(f)$ will have a well-defined tangent line at most points, although this may fail where the curve crosses itself or has a kink. Here are two examples:



$y^2 = x^3$



$y^2 = x^2(1 + x)$

If we demand that a tangent line be unique and follow the curve on both sides of the point, then each of these curves has a point where there is no tangent. Intuitively, a *singular point* of $\mathbf{V}(f)$ is a point such as above where the tangent line fails to exist.

To make this notion more precise, we first must give an algebraic definition of tangent line. We will use the following approach. Given a point $(a, b) \in \mathbf{V}(f)$, a line L through (a, b) is given parametrically by

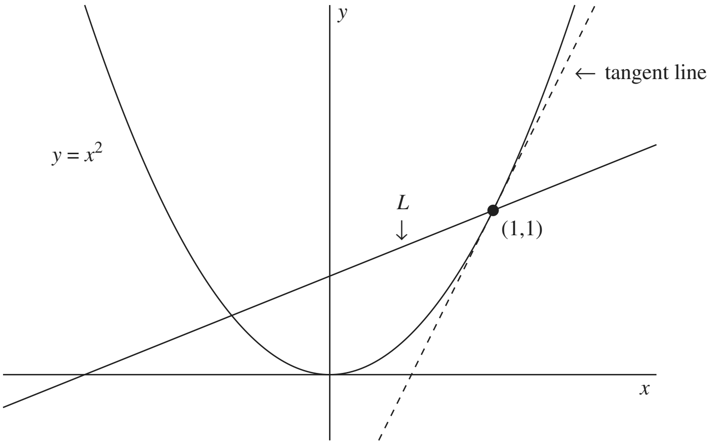
$$\begin{aligned} x &= a + ct, \\ y &= b + dt. \end{aligned} \tag{1}$$

This line goes through (a, b) when $t = 0$. Notice also that $(c, d) \neq (0, 0)$ is a vector parallel to the line. Thus, by varying (c, d), we get all lines through (a, b). But how do we find the one that is *tangent* to $\mathbf{V}(f)$? Can we find it without using calculus?

Let us look at an example. Consider the line L

$$\begin{aligned} x &= 1 + ct, \\ y &= 1 + dt, \end{aligned} \tag{2}$$

through the point (1, 1) on the parabola $y = x^2$:



From calculus, we know that the tangent line has slope 2, which corresponds to the line with $d = 2c$ in the above parametrization. To find this line by algebraic means, we will study the *polynomial* that describes how the line meets the parabola. If we substitute (2) into the left-hand side of $y - x^2 = 0$, we get the polynomial

$$(3) \qquad g(t) = 1 + dt - (1 + ct)^2 = -c^2t^2 + (d - 2c)t = t(-c^2t + d - 2c).$$

The roots of g determine where the line intersects the parabola (be sure you understand this). If $d \neq 2c$, then g has two distinct roots when $c \neq 0$ and one root when $c = 0$. But if $d = 2c$, then g has a root of multiplicity 2. Thus, we can detect when the line (2) is tangent to the parabola by looking for a *multiple root*.

Based on this example, let us make the following definition.

Definition 1. *Let k be a positive integer. Suppose that we have a point* $(a, b) \in \mathbf{V}(f)$ *and let L be line through (a, b). Then L* **meets** $\mathbf{V}(f)$ **with multiplicity** k **at** (a, b) *if L can be parametrized as in* (1) *so that* $t = 0$ *is a root of multiplicity k of the polynomial* $g(t) = f(a + ct, b + dt)$.

In this definition, note that $g(0) = f(a, b) = 0$, so that $t = 0$ is a root of g. Also, recall that $t = 0$ is a root of multiplicity k when $g = t^k h$, where $h(0) \neq 0$. One ambiguity with this definition is that a given line has many different parametrizations. So we need to check that the notion of multiplicity is independent of the parametrization. This will be covered in the exercises.

For an example of how this definition works, consider the line given by (2) above. It should be clear from (3) that the line meets the parabola $y = x^2$ with multiplicity 1 at (1, 1) when $d \neq 2c$ and with multiplicity 2 when $d = 2c$. Other examples will be given in the exercises.

We will use the notion of multiplicity to pick out the tangent line. To make this work, we will need the *gradient vector* of *f*, which is defined to be

$$\nabla f = \left(\frac{\partial}{\partial x} f, \frac{\partial}{\partial y} f\right).$$

We can now state our result.

Proposition 2. *Let* $f \in k[x, y]$, *and let* $(a, b) \in \mathbf{V}(f)$.

(i) *If* $\nabla f(a, b) \neq (0, 0)$, *then there is a unique line through* (a, b) *which meets* $\mathbf{V}(f)$ *with multiplicity* ≥ 2.

(ii) *If* $\nabla f(a, b) = (0, 0)$, *then every line through* (a, b) *meets* $\mathbf{V}(f)$ *with multiplicity* ≥ 2.

Proof. Let a line L through (a, b) be parametrized as in equation (1) and let $g(t) = f(a + ct, b + dt)$. Since $(a, b) \in \mathbf{V}(f)$, it follows that $t = 0$ is a root of g. The following observation will be proved in the exercises:

$$(4) \qquad t = 0 \text{ is a root of } g \text{ of multiplicity } \geq 2 \Leftrightarrow g'(0) = 0.$$

Using the chain rule, one sees that

$$g'(t) = \frac{\partial}{\partial x} f(a + ct, b + dt) \cdot c + \frac{\partial}{\partial y} f(a + ct, b + dt) \cdot d.$$

and thus

$$g'(0) = \frac{\partial}{\partial x} f(a, b) \cdot c + \frac{\partial}{\partial y} f(a, b) \cdot d.$$

If $\nabla f(a, b) = (0, 0)$, then the above equation shows that $g'(0)$ always equals 0. By (4), it follows that L always meets $\mathbf{V}(f)$ at (a, b) with multiplicity ≥ 2. This proves the second part of the proposition. Turning to the first part, suppose that $\nabla f(a, b) \neq (0, 0)$. We know that $g'(0) = 0$ if and only if

$$(5) \qquad \frac{\partial}{\partial x} f(a, b) \cdot c + \frac{\partial}{\partial y} f(a, b) \cdot d = 0.$$

This is a linear equation in the unknowns c and d. Since the coefficients $\frac{\partial}{\partial x} f(a, b)$ and $\frac{\partial}{\partial y} f(a, b)$ are not both zero, the solution space is 1-dimensional. Thus, there is $(c_0, d_0) \neq (0, 0)$ such that (c, d) satisfies the above equation if and only if $(c, d) = \lambda(c_0, d_0)$ for some $\lambda \in k$. It follows that the (c, d)'s giving $g'(0) = 0$ all parametrize the same line L. This shows that there is a unique line which meets $\mathbf{V}(f)$ at (a, b) with multiplicity ≥ 2. Proposition 2 is proved. □

Using Proposition 2, it is now obvious how to define the tangent line. From the second part of the proposition, it is also clear what a singular point should be.

Definition 3. *Let $f \in k[x, y]$ and let $(a, b) \in \mathbf{V}(f)$.*

(i) *If $\nabla f(a, b) \neq (0, 0)$, then the* **tangent line** *of $\mathbf{V}(f)$ at (a, b) is the unique line through (a, b) which meets $\mathbf{V}(f)$ with multiplicity ≥ 2. We say that (a, b) is a* **nonsingular point** *of $\mathbf{V}(f)$.*

(ii) *If $\nabla f(a, b) = (0, 0)$, then we say that (a, b) is a* **singular point** *of $\mathbf{V}(f)$.*

Over $\mathbb{R}$, the tangent line and the gradient have the following geometric interpretation. If the tangent to $\mathbf{V}(f)$ at (a, b) is parametrized by (1), then the vector (c, d) is parallel to the tangent line. But we also know from equation (5) that the dot product $\nabla f(a, b) \cdot (c, d)$ is zero, which means that $\nabla f(a, b)$ is perpendicular to (c, d). Thus, we have an algebraic proof of the theorem from calculus that *the gradient $\nabla f(a, b)$ is perpendicular to the tangent line of $\mathbf{V}(f)$* at (a, b).

For any given curve $\mathbf{V}(f)$, we can compute the singular points as follows. The gradient ∇f is zero when $\frac{\partial}{\partial x} f$ and $\frac{\partial}{\partial y} f$ vanish simultaneously. Since we also have to be on $\mathbf{V}(f)$, we need $f = 0$. It follows that the singular points of $\mathbf{V}(f)$ are determined by the equations

$$f = \frac{\partial}{\partial x} f = \frac{\partial}{\partial y} f = 0.$$

As an example, consider the curve $y^2 = x^2(1 + x)$ shown earlier. To find the singular points, we must solve

$$\begin{aligned} f &= y^2 - x^2 - x^3 = 0, \\ \frac{\partial}{\partial x} f &= -2x - 3x^2 = 0, \\ \frac{\partial}{\partial y} f &= 2y = 0. \end{aligned}$$

From these equations, it is easy to see that $(0, 0)$ is the only singular point of $\mathbf{V}(f)$. This agrees with the earlier picture.

Using the methods learned in §§1 and 2, we can tackle much more complicated problems. For example, later in this section we will determine the singular points of the curve defined by the sixth degree equation

$$\begin{aligned} 0 = &-1156 + 688x^2 - 191x^4 + 16x^6 + 544y + 30x^2y - 40x^4y \\ &+ 225y^2 - 96x^2y^2 + 16x^4y^2 - 136y^3 - 32x^2y^3 + 16y^4. \end{aligned}$$

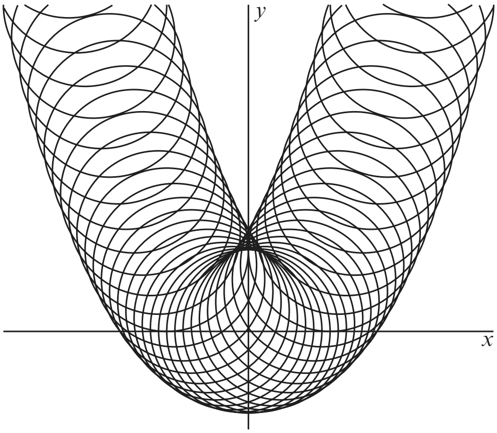
The exercises will explore some other aspects of singular points. In Chapter 9, we will study singular and nonsingular points on an arbitrary affine variety.

Envelopes

In our discussion of envelopes, we will work over $\mathbb{R}$ to make the geometry easier to see. The best way to explain what we mean by envelope is to compute an example. Let $t \in \mathbb{R}$, and consider the circle in $\mathbb{R}^2$ defined by the equation

$$(x - t)^2 + (y - t^2)^2 = 4. \tag{6}$$

Since (t, t^2) parametrizes a parabola, we can think of equation (6) as describing the *family* of circles of radius 2 in $\mathbb{R}^2$ whose centers lie on the parabola $y = x^2$. The picture is as follows:



A Family of Circles in the Plane

Note that the "boundary" curve is *simultaneously tangent* to all the circles in the family. This is a special case of the *envelope* of a family of curves. The basic idea is that the envelope of a family of curves is a single curve that is tangent to all of the curves in the family. Our goal is to study envelopes and learn how to compute them. In particular, we want to find the equation of the envelope in the above example.

Before we can give a more careful definition of envelope, we must first understand the concept of a *family* of curves in $\mathbb{R}^2$.

Definition 4. *Given a polynomial $F \in \mathbb{R}[x, y, t]$, fix a real number $t \in \mathbb{R}$. Then the variety in $\mathbb{R}^2$ defined by $F(x, y, t) = 0$ is denoted $\mathbf{V}(F_t)$, and the* **family of curves** *determined by F consists of the varieties $\mathbf{V}(F_t)$ as t varies over $\mathbb{R}$.*

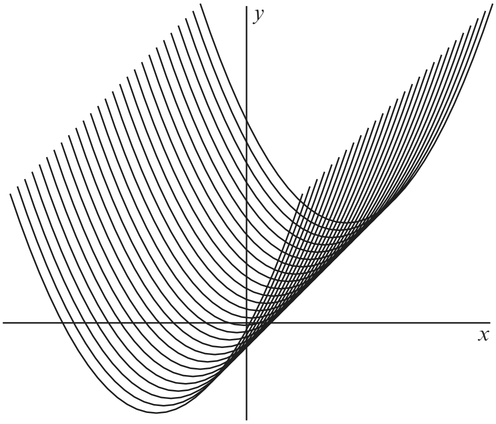
In this definition, we think of t as a *parameter* that tells us which curve in the family we are looking at. Strictly speaking, we should say "family of varieties" rather than "family of curves," but we will use the latter to emphasize the geometry of the situation.

For another example of a family and its envelope, consider the curves defined by

$$F(x, y, t) = (x - t)^2 - y + t = 0. \tag{7}$$

Writing this as $y - t = (x - t)^2$, we see in the picture at the top of the next page that (7) describes the family $\mathbf{V}(F_t)$ of parabolas obtained by translating the standard parabola $y = x^2$ along the straight line $y = x$. In this case, the envelope is clearly the straight line that just touches each parabola in the family. This line has slope 1, and from here, it is easy to check that the envelope is given by $y = x - 1/4$ (the details are left as an exercise).

Not all envelopes are so easy to describe. The remarkable fact is that we can characterize the envelope in the following completely algebraic way.



A Family of Parabolas in the Plane

Definition 5. *Given a family* $\mathbf{V}(F_t)$ *of curves in* $\mathbb{R}^2$*, its* **envelope** *consists of all points* $(x, y) \in \mathbb{R}^2$ *with the property that*

$$F(x, y, t) = 0,$$
$$\frac{\partial}{\partial t} F(x, y, t) = 0$$

for some $t \in \mathbb{R}$.

We need to explain how this definition corresponds to the intuitive idea of envelope. The argument given below is not rigorous, but it does explain where the condition on $\frac{\partial}{\partial t} F$ comes from. A complete treatment of envelopes requires a fair amount of theoretical machinery. We refer the reader to Chapter 5 of Bruce and Giblin (1992) for more details.

Given a family $\mathbf{V}(F_t)$, we think of the envelope as a curve C with the property that at each point on the curve, C is tangent to one of the curves $\mathbf{V}(F_t)$ in the family. Suppose that C is parametrized by

$$x = f(t),$$
$$y = g(t).$$

We will assume that at time t, the point $(f(t), g(t))$ is on the curve $\mathbf{V}(F_t)$. This ensures that C meets all the members of the family. Algebraically, this means that

$$(8) \qquad F(f(t), g(t), t) = 0 \quad \text{for all } t \in \mathbb{R}.$$

But when is C *tangent* to $\mathbf{V}(F_t)$ at $(f(t), g(t))$? This is what is needed for C to be the envelope of the family. We know from calculus that the tangent vector to C is

$(f'(t), g'(t))$. As for $\mathbf{V}(F_t)$, we have the gradient $\nabla F = \left(\frac{\partial}{\partial x}F, \frac{\partial}{\partial y}F\right)$, and from the first part of this section, we know that ∇F is perpendicular to the tangent line to $\mathbf{V}(F_t)$. Thus, for C to be tangent to $\mathbf{V}(F_t)$, the tangent $(f'(t), g'(t))$ must be perpendicular to the gradient ∇F. In terms of dot products, this means that $\nabla F \cdot (f'(t), g'(t)) = 0$ or, equivalently,

$$\frac{\partial}{\partial x}F(f(t), g(t), t) \cdot f'(t) + \frac{\partial}{\partial y}F(f(t), g(t), t) \cdot g'(t) = 0. \tag{9}$$

We have thus shown that the envelope is determined by conditions (8) and (9). To relate this to Definition 5, differentiate (8) with respect to t. Using the chain rule, we get

$$\frac{\partial}{\partial x}F(f(t), g(t), t) \cdot f'(t) + \frac{\partial}{\partial y}F(f(t), g(t), t) \cdot g'(t) + \frac{\partial}{\partial t}F(f(t), g(t), t) = 0.$$

If we subtract equation (9) from this, we obtain

$$\frac{\partial}{\partial t}F(f(t), g(t), t) = 0. \tag{10}$$

From (8) and (10), it follows that $(x, y) = (f(t), g(t))$ has exactly the property described in Definition 5.

As we will see later in the section, the above discussion of envelopes is rather naive. For us, the main consequence of Definition 5 is that the envelope of $\mathbf{V}(F_t)$ is determined by the equations

$$\begin{aligned} F(x, y, t) &= 0, \\ \frac{\partial}{\partial t}F(x, y, t) &= 0. \end{aligned}$$

Note that x and y tell us where we are on the envelope and t tells us which curve in the family we are tangent to. Since these equations involve x, y, and t, we need to eliminate t to find the equation of the envelope. Thus, we can apply the theory from §§1 and 2 to study the envelope of a family of curves.

Let us see how this works in example (6). Here, $F = (x - t)^2 + (y - t^2)^2 - 4$, so that the envelope is described by the equations

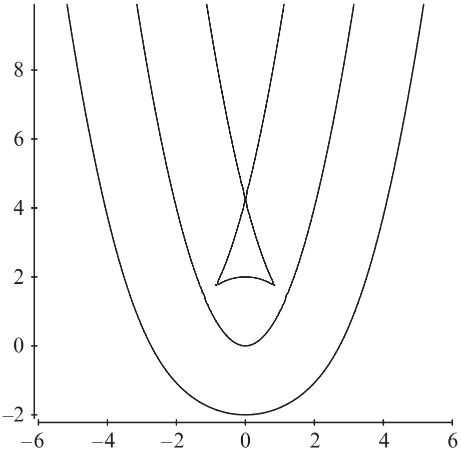
$$\begin{aligned} F &= (x - t)^2 + (y - t^2)^2 - 4 = 0, \\ \frac{\partial}{\partial t}F &= -2(x - t) - 4t(y - t^2) = 0. \end{aligned} \tag{11}$$

Using lexicographic order with $t > x > y$, a Groebner basis is given by

$$\begin{aligned} g_1 =& -1156 + 1688x^2 - 191x^4 + 16x^6 + 544y + 30x^2y - 40x^4y \\ &+ 225y^2 - 96x^2y^2 + 16x^4y^2 - 136y^3 - 32x^2y^3 + 16y^4, \\ g_2 =& (7327 - 1928y - 768y^2 - 896y^3 + 256y^4)t + 6929x - 2946x^3 \\ &+ 224x^5 + 2922xy - 1480x^3y + 128x^5y - 792xy^2 - 224x^3y^2 \\ &- 544xy^3 + 128x^3y^3 - 384xy^4, \end{aligned}$$

$$\begin{aligned}
g_3 =&(431x - 12xy - 48xy^2 - 64xy^3)t + 952 - 159x^2 - 16x^4 + 320y\\
&- 214x^2y + 32x^4y - 366y^2 - 32x^2y^2 - 80y^3 + 32x^2y^3 + 32y^4,\\
g_4 =&(697 - 288x^2 + 108y - 336y^2 + 64y^3)t + 23x - 174x^3\\
&+ 32x^5 + 322xy - 112x^3y + 32xy^2 + 32x^3y^2 - 96xy^3,\\
g_5 =&135t^2 + (26x + 40xy + 32xy^2)t - 128 + 111x^2\\
&- 16x^4 + 64y + 8x^2y + 32y^2 - 16x^2y^2 - 16y^3.
\end{aligned}$$

We have written the Groebner basis as polynomials in t with coefficients in $\mathbb{R}[x, y]$, The Elimination Theorem tells us that g_1 generates the first elimination ideal. Thus, the envelope lies on the curve $g_1 = 0$. Here is a picture of $\mathbf{V}(g_1)$ together with the parabola $y = x^2$:



The surprise is the "triangular" portion of the graph that was somewhat unclear in the earlier picture of the family. By drawing some circles centered on the parabola, you can see how the triangle is still part of the envelope.

We have proved that the envelope lies on $\mathbf{V}(g_1)$, but the two may not be equal. In fact, there are two interesting questions to ask at this point:

- Is every point of $\mathbf{V}(g_1)$, on the envelope? This is the same as asking if every partial solution (x, y) of (11) extends to a complete solution (x, y, t).
- Given a point on the envelope, how many curves in the family are tangent to the envelope at the point? This asks how many t's are there for which (x, y) extends to (x, y, t).

Since the leading coefficient of t in g_5 is the constant 135, the Extension Theorem (in the form of Corollary 4 of §1) guarantees that *every* partial solution extends, *provided* we work over the complex numbers. Thus, t exists, but it might be complex. This illustrates the power and limitation of the Extension Theorem: it can guarantee that there is a solution, but it might lie in the wrong field.

In spite of this difficulty, the equation $g_5 = 0$ does have something useful to tell us: it is quadratic in t, so that a given (x, y) extends in at most *two* ways to a complete solution. Thus, *a point on the envelope of (6) is tangent to at most two circles in the family*. Can you see any points where there are two tangent circles?

To get more information on what is happening, let us look at the other polynomials in the Groebner basis. Note that g_2, g_3, and g_4 involve t only to the first power. Thus, we can write them in the form

$$g_i = A_i(x, y)t + B_i(x, y), \quad i = 2, 3, 4.$$

If A_i does not vanish at (x, y) for one of $i = 2, 3, 4$, then we can solve $A_i t + B_i = 0$ to obtain

$$t = -\frac{B_i(x, y)}{A_i(x, y)}.$$

Thus, we see that t is real whenever x and y are. More importantly, this formula shows that t is uniquely determined when $A_i(x, y) \neq 0$. Thus, *a point on the envelope of (6) not in* $\mathbf{V}(A_2, A_3, A_4)$ *is tangent to exactly one circle in the family.*

It remains to understand where A_2, A_3, and A_4 vanish simultaneously. These polynomials might look complicated, but, using the techniques of §1, one can show that the real solutions of $A_2 = A_3 = A_4 = 0$ are

$$(x, y) = (0, 17/4) \text{ and } (\pm 0.936845,\ 1.63988). \tag{12}$$

Looking back at the picture of $\mathbf{V}(g_1)$, it appears that these are the singular points of $\mathbf{V}(g_1)$. Can you see the two circles tangent at these points? From the first part of this section, we know that the singular points of $\mathbf{V}(g_1)$ are determined by the equations $g_1 = \frac{\partial}{\partial x} g_1 = \frac{\partial}{\partial y} g_1 = 0$. Thus, to say that the singular points coincide with (12) means that

$$\mathbf{V}(A_2, A_3, A_4) = \mathbf{V}\left(g_1, \frac{\partial}{\partial x} g_1, \frac{\partial}{\partial y} g_1\right). \tag{13}$$

To prove this, we will show that

$$\begin{aligned} &g_1, \frac{\partial}{\partial x} g_1, \frac{\partial}{\partial y} g_1 \in \langle A_2, A_3, A_4 \rangle, \\ &A_2^2, A_3^2, A_4^2 \in \left\langle g_1, \frac{\partial}{\partial x} g_1, \frac{\partial}{\partial y} g_1 \right\rangle. \end{aligned} \tag{14}$$

The proof of (14) is a straightforward application of the ideal membership algorithm discussed in Chapter 2. For the first line, one computes a Groebner basis of $\langle A_2, A_3, A_4 \rangle$ and then applies the algorithm for the ideal membership problem to each of $g_1, \frac{\partial}{\partial x} g_1, \frac{\partial}{\partial y} g_1$ (see §8 of Chapter 2). The second line of (14) is treated similarly—the details will be left as an exercise.

Since (13) follows immediately from (14), we have proved that *a nonsingular point on* $\mathbf{V}(g_1)$, *is in the envelope of (6) and, at such a point, the envelope is tangent to exactly one circle in the family*. Also note that the singular points of $\mathbf{V}(g_1)$ are the most interesting points in the envelope, for they are the ones where there are two tangent circles. This last observation shows that singular points are not always bad—they can be a useful indication that something unusual is happening. An important part of algebraic geometry is devoted to the study of singular points.

In this example, equations (11) for the envelope were easy to write down. But to understand the equations, we had to use a Groebner basis and the Elimination and Extension Theorems. Even though the Groebner basis looked complicated, it told us exactly which points on the envelope were tangent to more than one circle. This illustrates nicely the power of the theory we have developed so far.

As we said earlier, our treatment of envelopes has been a bit naive. Evidence of this comes from the above example, which shows that the envelope can have singularities. How can the envelope be "tangent" to a curve in the family at a singular point? In the exercises, we will indicate another reason why our discussion has been too simple. We have also omitted the fascinating relation between the family of curves $\mathbf{V}(F_t) \subset \mathbb{R}^2$ and the surface $\mathbf{V}(F) \subset \mathbb{R}^3$ defined by $F(x, y, t) = 0$. We refer the reader to Chapter 5 of Bruce and Giblin (1992) for a more complete treatment of these aspects of envelopes.

EXERCISES FOR §4

1. Let C be the curve in k^2 defined by $x^3 - xy + y^2 = 1$ and note that $(1, 1) \in C$. Now consider the straight line parametrized by
$$\begin{aligned} x &= 1 + ct, \\ y &= 1 + dt. \end{aligned}$$
Compute the multiplicity of this line when it meets C at $(1, 1)$. What does this tell you about the tangent line? Hint: There will be two cases to consider.
2. In Definition 1, we need to show that the notion of multiplicity is independent of how the line is parametrized.
   a. Show that two parametrizations
$$\begin{aligned} x &= a + ct, & x &= a + c't, \\ y &= b + dt, & y &= b + d't, \end{aligned}$$
correspond to the same line if and only if there is a nonzero number $\lambda \in k$ such that $(c, d) = \lambda(c', d')$. Hint: In the parametrization $x = a + ct$, $y = b + dt$ of a line L, recall that L is parallel to the vector (c, d).
   b. Suppose that the two parametrizations of part a correspond to the same line L that meets $\mathbf{V}(f)$ at (a, b). Prove that the polynomials $g(t) = f(a + ct, b + dt)$ and $g'(t) = f(a + c't, b + d't)$ have the same multiplicity at $t = 0$. Hint: Use part a to relate g and g'. This will prove that the multiplicity of how L meets $\mathbf{V}(f)$ at (a, b) is well defined.
3. Consider the straight lines
$$\begin{aligned} x &= t, \\ y &= b + t. \end{aligned}$$
These lines have slope 1 and y-intercept b. For which values of b is the line tangent to the circle $x^2 + y^2 = 2$? Draw a picture to illustrate your answer. Hint: Consider $g(t) = t^2 + (b + t)^2 - 2$. The roots of this quadratic determine the values of t where the line meets the circle.

4. If $(a, b) \in \mathbf{V}(f)$ and $\nabla f(a, b) \neq (0, 0)$, prove that the tangent line of $\mathbf{V}(f)$ at (a, b) is defined by the equation

$$\frac{\partial}{\partial x} f(a, b) \cdot (x - a) + \frac{\partial}{\partial y} f(a, b) \cdot (y - b) = 0.$$

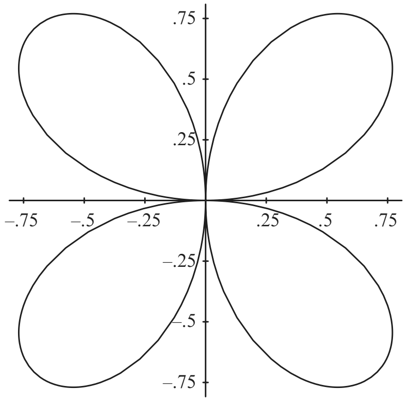
5. Let $g \in k[t]$ be a polynomial such that $g(0) = 0$. Assume that $\mathbb{Q} \subset k$.
   a. Prove that $t = 0$ is a root of multiplicity ≥ 2 of g if and only if $g'(0) = 0$. Hint: Write $g(t) = th(t)$, and use the product rule.
   b. More generally, prove that $t = 0$ is a root of multiplicity $\geq k$ if and only if $g'(0) = g''(0) = \cdots = g^{(k-1)}(0) = 0$.
6. As in the Definition 1, let a line L be parametrized by (1), where $(a, b) \in \mathbf{V}(f)$. Also let $g(t) = f(a + ct, b + dt)$. Prove that L meets $\mathbf{V}(f)$ with multiplicity k if and only if $g'(0) = g''(0) = \cdots = g^{(k-1)}(0) = 0$ but $g^{(k)}(0) \neq 0$. Hint: Use the previous exercise.
7. In this exercise, we will study how a tangent line can meet a curve with multiplicity *greater* than 2. Let C be the curve defined by $y = f(x)$, where $f \in k[x]$. Thus, C is just the graph of f.
   a. Give an algebraic proof that the tangent line to C at $(a, f(a))$ is parametrized by

   $$\begin{aligned} x &= a + t, \\ y &= f(a) + f'(a)t. \end{aligned}$$

   Hint: Consider $g(t) = f(a) + f'(a)t - f(a + t)$.
   b. Show that the tangent line at $(a, f(a))$ meets the curve with multiplicity ≥ 3 if and only if $f''(a) = 0$. Hint: Use the previous exercise.
   c. Show that the multiplicity is exactly 3 if and only if $f''(a) = 0$ but $f'''(a) \neq 0$.
   d. Over $\mathbb{R}$, a *point of inflection* is defined to be a point where $f''(x)$ changes sign. Prove that if the multiplicity is 3, then $(a, f(a))$ is a point of inflection.
8. In this problem, we will compute some singular points.
   a. Show that (0, 0) is the *only* singular point of $y^2 = x^3$.
   b. In Exercise 8 of §3 of Chapter 1, we studied the curve $y^2 = cx^2 - x^3$, where c is some constant. Find all singular points of this curve and explain how your answer relates to the picture of the curve given in Chapter 1.
   c. Show that the circle $x^2 + y^2 = a^2$ has no singular points.
9. One use of multiplicities is to show that one singularity is "worse" than another.
   a. For the curve $y^2 = x^3$, show that most lines through the origin meet the curve with multiplicity exactly 2.
   b. For $x^4 + 2xy^2 + y^3 = 0$, show that all lines through the origin meet the curve with multiplicity ≥ 3.

   This suggests that the singularity at the origin is "worse" on the second curve. Using the ideas behind this exercise, one can define the notion of the *multiplicity* of a singular point.
10. We proved in the text that (0, 0) is a singular point of the curve C defined by $y^2 = x^2(1+x)$. But in the picture of C, it looks like there are two "tangent" lines through the origin. Can we use multiplicities to pick these out?
    a. Show that with two exceptions, all lines through the origin meet C with multiplicity 2. What are the lines that have multiplicity 3?
    b. Explain how your answer to part (a) relates to the picture of C in the text. Why should the "tangent" lines have higher multiplicity?

11. The four-leaved rose is defined in polar coordinates by the equation $r = \sin(2\theta)$:



In Cartesian coordinates, this curve is defined by the equation $(x^2 + y^2)^3 = 4x^2y^2$.

a. Show that most lines through the origin meet the rose with multiplicity 4 at the origin. Can you give a geometric explanation for this number?

b. Find the lines through the origin that meet the rose with multiplicity > 4. Give a geometric explanation for the numbers you get.

12. Consider a surface $\mathbf{V}(f) \subset k^3$ defined by $f \in k[x, y, z]$.

a. Define what it means for $(a, b, c) \in \mathbf{V}(f)$ to be a singular point.

b. Determine all singular points of the sphere $x^2 + y^2 + z^2 = 1$. Does your answer make sense?

c. Determine all singular points on the surface $\mathbf{V}(x^2 - y^2z^2 + z^3)$. How does your answer relate to the picture of the surface drawn in §2 of Chapter 1?

13. Consider the family of curves given by $F = xy - t \in \mathbb{R}[x, y, t]$. Draw various of the curves $\mathbf{V}(F_t)$ in the family. Be sure to include a picture of $\mathbf{V}(F_0)$.

14. This problem will study the envelope of the family $F = (x - t)^2 - y + t$ considered in example (7).

a. It is obvious that the envelope is a straight line of slope 1. Use elementary calculus to show that the line is $y = x - 1/4$.

b. Use Definition 5 to compute the envelope.

c. Find a parametrization of the envelope so that at time t, the point $(f(t), g(t))$ is on the parabola $\mathbf{V}(F_t)$. Note that this is the kind of parametrization used in our discussion of Definition 5.

15. This problem is concerned with the envelope of example (6).

a. Copy the picture in the text onto a sheet of paper and draw in the two tangent circles for each of the points in (12).

b. For the point $(0, 4.25) = (0, 17.4)$, find the exact values of the t's that give the two tangent circles.

c. Show that the exact values of the points (12) are given by

$$(0, \tfrac{17}{4}) \quad \text{and} \quad (\pm\sqrt{15 + 6\sqrt[3]{2} - 12\sqrt[3]{4}}, \tfrac{1}{4}(-1 + 6\sqrt[3]{2})).$$

Hint: Most computer algebra systems have commands to factor polynomials and solve cubic equations.

16. Consider the family determined by $F = (x - t)^2 + y^2 - (1/2)t^2$.
    a. Compute the envelope of this family.
    b. Draw a picture to illustrate your answer.
17. Consider the family of circles defined by $(x - t)^2 + (y - t^2)^2 = t^2$ in the plane $\mathbb{R}^2$.
    a. Compute the equation of the envelope of this family and show that the envelope is the union of two varieties.
    b. Use the Extension Theorem and a Groebner basis to determine, for each point in the envelope, how many curves in the family are tangent to it. Draw a picture to illustrate your answer. Hint: You will use a different argument for each of the two curves making up the envelope.
18. Prove (14) using the hints given in the text. Also show that $A_2 \notin \langle g_1, \frac{\partial}{\partial x} g_1, \frac{\partial}{\partial y} g_1 \rangle$. This shows that the ideals $\langle g_1, \frac{\partial}{\partial x} g_1, \frac{\partial}{\partial y} g_1 \rangle$ and $\langle A_2, A_3, A_4 \rangle$ are not equal, even though they define the same variety.
19. In this exercise, we will show that our definition of envelope is too naive.
    a. Given a family of circles of radius 1 with centers lying on the x-axis, draw a picture to show that the envelope consists of the lines $y = \pm 1$.
    b. Use Definition 5 to compute the envelope of the family given by $F = (x - t)^2 + y^2 - 1$. Your answer should not be surprising.
    c. Use Definition 5 to find the envelope when the family is $F = (x - t^3)^2 + y^2 - 1$. Note that one of the curves in the family is part of the answer. This is because using t^3 allows the curves to "bunch up" near $t = 0$, which forces $\mathbf{V}(F_0)$ to appear in the envelope.

    In our intuitive discussion of envelope, recall that we assumed we could parametrize the envelope so that $(f(t), g(t))$ was in $\mathbf{V}(F_t)$ at time t. This presumes that the envelope is tangent to *different* curves in the family. Yet in the example given in part (c), part of the envelope lies in the *same* curve in the family. Thus, our treatment of envelope was too simple.
20. Suppose we have a family of curves in $\mathbb{R}^2$ determined by $F \in \mathbb{R}[x, y, t]$. Some of the curves $\mathbf{V}(F_t)$ may have singular points, whereas others may not. Can we find the ones that have a singularity?
    a. By considering the equations $F = \frac{\partial}{\partial x} F = \frac{\partial}{\partial y} F = 0$ in $\mathbb{R}^3$ and using elimination theory, describe a procedure for determining those t's corresponding to curves in the family which have a singular point.
    b. Apply the method of part (a) find the curves in the family of Exercise 13 that have singular points.

§5 Unique Factorization and Resultants

The main task remaining in Chapter 3 is to prove the Extension Theorem. This will require that we learn some new algebraic tools concerning *unique factorization* and *resultants*. Both of these will be used in §6 when we prove the Extension Theorem. We will also make frequent use of unique factorization in later chapters of the book.

Irreducible Polynomials and Unique Factorization

We begin with a basic definition.

Definition 1. *Let k be a field. A polynomial $f \in k[x_1, \ldots, x_n]$ is* **irreducible over** *k iff f is nonconstant and is not the product of two nonconstant polynomials in $k[x_1, \ldots, x_n]$.*

This definition says that if a nonconstant polynomial f is irreducible over k, then up to a constant multiple, its only nonconstant factor is f itself. Also note that the concept of irreducibility depends on the field. For example, $x^2 + 1$ is irreducible over $\mathbb{Q}$ and $\mathbb{R}$, but, over $\mathbb{C}$, we have $x^2 + 1 = (x - i)(x + i)$.

Every polynomial is a product of irreducible polynomials as follows.

Proposition 2. *Every nonconstant polynomial $f \in k[x_1, \dots, x_n]$ can be written as a product of polynomials which are irreducible over k.*

Proof. If f is irreducible over k, then we are done. Otherwise, we can write $f = gh$, where $g, h \in k[x_1, \dots, x_n]$ are nonconstant. Note that the total degrees of g and h are less than the total degree of f. Now apply this process to g and h: if either fails to be irreducible over k, we factor it into nonconstant factors. Since the total degree drops each time we factor, this process can be repeated at most finitely many times. Thus, f must be a product of irreducibles. □

In Theorem 5 we will show that the factorization of Proposition 2 is essentially unique. But first, we have to prove the following crucial property of irreducible polynomials.

Theorem 3. *Let $f \in k[x_1, \dots, x_n]$ be irreducible over k and suppose that f divides the product gh, where $g, h \in k[x_1, \dots, x_n]$. Then f divides g or h.*

Proof. We will use induction on the number of variables. When $n = 1$, we can use the GCD theory developed in §5 of Chapter 1. If f divides gh, then consider $p = \mathrm{GCD}(f, g)$. If p is nonconstant, then f must be a constant multiple of p since f is irreducible, and it follows that f divides g. On the other hand, if p is constant, we can assume $p = 1$, and then we can find $A, B \in k[x_1]$ such that $Af + Bg = 1$ (see Proposition 6 of Chapter 1, §5). If we multiply this by h, we get

$$h = h(Af + Bg) = Ahf + Bgh.$$

Since f divides gh, f is a factor of $Ahf + Bgh$, and, thus, f divides h. This proves the case $n = 1$.

Now assume that the theorem is true for $n - 1$. We first discuss the special case where the irreducible polynomial does not involve x_1:

$$(1) \qquad u \in k[x_2, \dots, x_n] \text{ irreducible, } u \text{ divides } gh \in k[x_1, \dots x_n] \Rightarrow u \text{ divides } g \text{ or } h.$$

To prove this, write $g = \Sigma_{i=0}^{l} a_i x_1^i$ and $h = \Sigma_{i=0}^{m} b_i x_1^i$, where $a_i, b_i \in k[x_2, \dots, x_n]$. If u divides every a_i, then u divides g, and similarly for h. Hence, if u divides neither, we can find $i, j \geq 0$ such that u divides neither a_i nor b_j. We will assume that i and j are the smallest subscripts with this property. Then consider

$$c_{i+j} = (a_0 b_{i+j} + a_1 b_{i+j-1} + \cdots + a_{i-1} b_{j+1}) + a_i b_j + (a_{i+1} b_{j-1} + \cdots + a_{i+j} b_0).$$

By the way we chose i, u divides every term inside the first set of parentheses and, by the choice of j, the same is true for the second set of parentheses. But u divides neither

a_i nor b_j, and since u is irreducible, our inductive assumption implies that u does not divide a_ib_j. Since u divides all other terms of c_{i+j}, it cannot divide c_{i+j}. We leave it as an exercise to show that c_{i+j} is the coefficient of x_1^{i+j} in gh, and, hence, u cannot divide gh. This contradiction completes the proof of (1).

Now, given (1), we can treat the general case. Suppose that f divides gh. If f doesn't involve x_1, then we are done by (1). So assume that f is nonconstant in x_1. We will use the ring $k(x_2, \ldots, x_n)[x_1]$, which is a polynomial ring in one variable over the *field* $k(x_2, \ldots, x_n)$. Remember that elements of $k(x_2, \ldots, x_n)$ are quotients of polynomials in $k[x_2, \ldots, x_n]$. We can regard $k[x_1, \ldots, x_n]$ as lying inside $k(x_2, \ldots, x_n)[x_1]$. The strategy will be to work in the larger ring, where we know the theorem to be true, and then pass back to the smaller ring $k[x_1, \ldots, x_n]$.

We claim that f is still irreducible when regarded as an element of $k(x_2, \ldots, x_n)[x_1]$. To see why, suppose we had a factorization of f in the larger ring, say $f = AB$. Here, A and B are polynomials in x_1 with coefficients in $k(x_2, \ldots, x_n)$. To prove that f is irreducible here, we must show that A or B has degree 0 in x_1. Let $d \in k[x_2, \ldots, x_n]$ be the product of all denominators in A and B. Then $\tilde{A} = dA$ and $\tilde{B} = dB$ are in $k[x_1, \ldots, x_n]$, and

$$d^2 f = \tilde{A}\tilde{B} \tag{2}$$

in $k[x_1, \ldots, x_n]$. By Proposition 2, we can write d^2 as a product of irreducible factors in $k[x_2, \ldots, x_n]$, and, by (1), each of these divides $\tilde{A}$ or $\tilde{B}$. We can cancel such a factor from both sides of (2), and after we have cancelled all of the factors, we are left with

$$f = \tilde{A}_1\tilde{B}_1$$

in $k[x_1, \ldots, x_n]$. Since f is irreducible in $k[x_1, \ldots, x_n]$, this implies that $\tilde{A}_1$ or $\tilde{B}_1$ is constant. Now these polynomials were obtained from the original A, B by multiplying and dividing by various elements of $k[x_2, \ldots, x_n]$. This shows that either A or B does not involve x_1, and our claim follows.

Now that f is irreducible in $k(x_2, \ldots, x_n)[x_1]$, we know by the $n = 1$ case of the theorem that f divides g or h in $k(x_2, \ldots, x_n)[x_1]$. Say $g = Af$ for some $A \in k(x_2, \ldots, x_n)[x_1]$. If we clear denominators, we can write

$$dg = \tilde{A}f \tag{3}$$

in $k[x_1, \ldots, x_n]$, where $d \in k[x_1, \ldots, x_n]$. By (1), every irreducible factor of d divides $\tilde{A}$ or f. The latter is impossible since f is irreducible and has positive degree in x_1. But each time an irreducible factor divides $\tilde{A}$, we can cancel it from both sides of (3). When all the cancellation is done, we see that f divides g in $k[x_1, \ldots, x_n]$. This completes the proof of the theorem. □

In §6, we will need the following consequence of Theorem 3.

Corollary 4. *Suppose that $f, g \in k[x_1, \ldots, x_n]$ have positive degree in x_1. Then f and g have a common factor in $k[x_1, \ldots, x_n]$ of positive degree in x_1 if and only if they have a common factor in $k(x_2, \ldots, x_n)[x_1]$.*

Proof. If f and g have a common factor h in $k[x_1, \ldots, x_n]$ of positive degree in x_1, then they certainly have a common factor in the larger ring $k(x_2, \ldots, x_n)[x_1]$. Going the other way, suppose that f and g have a common factor $\tilde{h} \in k(x_2, \ldots, x_n)[x_1]$. Then

$$f = \tilde{h}\tilde{f}_1, \ \tilde{f}_1 \in k(x_2, \ldots, x_n)[x_1].$$
$$g = \tilde{h}\tilde{g}_1, \ \tilde{g}_1 \in k(x_2, \ldots, x_n)[x_1].$$

Now $\tilde{h}$, $\tilde{f}_1$ and $\tilde{g}_1$ may have denominators that are polynomials in $k[x_2, \ldots, x_n]$. Letting $d \in k[x_2, \ldots, x_n]$ be a common denominator of these polynomials, we get $h = d\tilde{h}$, $f_1 = d\tilde{f}_1$ and $g_1 = d\tilde{g}_1$ in $k[x_1, \ldots, x_n]$. If we multiply each side of the above equations by d^2, we obtain

$$d^2 f = hf_1,$$
$$d^2 g = hg_1$$

in $k[x_1, \ldots, x_n]$. Now let h_1 be an irreducible factor of h of positive degree in x_1. Since $\tilde{h} = h/d$ has positive degree in x_1, such an h_1 must exist. Then h_1 divides $d^2 f$, so that it divides d^2 or f by Theorem 3. The former is impossible because $d^2 \in k[x_2, \ldots, x_n]$ and, hence, h_1 must divide f in $k[x_1, \ldots, x_n]$. A similar argument shows that h_1 divides g, and thus h_1 is the required common factor. This completes the proof of the corollary. □

Theorem 3 says that irreducible polynomials behave like prime numbers, in that if a prime divides a product of two integers, it must divide one or the other. This property of primes is the key to unique factorization of integers, and the same is true for irreducible polynomials.

Theorem 5. *Every nonconstant $f \in k[x_1, \ldots, x_n]$ can be written as a product $f = f_1 \cdot f_2 \cdots f_r$ of irreducibles over k. Further, if $f = g_1 \cdot g_2 \cdots g_s$ is another factorization into irreducibles over k, then $r = s$ and the g_i's can be permuted so that each f_i is a constant multiple of g_i.*

Proof. The proof will be covered in the exercises. □

For polynomials in $\mathbb{Q}[x_1, \ldots, x_n]$, there are algorithms for factoring into irreducibles over $\mathbb{Q}$, A classical algorithm due to Kronecker is discussed in Theorem 4.8 of MINES, RICHMAN, and RUITENBERG (1988), and a more efficient method is given in DAVENPORT, SIRET and TOURNIER (1993) or MIGNOTTE (1992). Most computer algebra systems have a command for factoring polynomials in $\mathbb{Q}[x_1, \ldots, x_n]$. Factoring polynomials in $\mathbb{R}[x_1, \ldots, x_n]$ or $\mathbb{C}[x_1, \ldots, x_n]$ is much more difficult.

Resultants

Although resultants have a different flavor from what we have done so far, they play an important role in elimination theory. We will introduce the concept of resultant by asking when two polynomials in $k[x]$ have a common factor. This might seem far removed from elimination, but we will see the connection by the end of the section. In §6, we will study the resultant of two polynomials in $k[x_1, \ldots, x_n]$, and we will then use resultants to prove the Extension Theorem. Suppose that we want to know whether two polynomials $f, g \in k[x]$ have a common factor (which means a polynomial $h \in k[x]$ of degree > 0 which divides f and g). One way would be to factor f and g into irreducibles. Unfortunately, factoring can be a time-consuming process. A more efficient method would be to compute the GCD of f and g using the Euclidean Algorithm from Chapter 1. A drawback is that the Euclidean Algorithm requires divisions in the field k. As we will see later, this is something we want to avoid when doing elimination. So is there a way of determining whether a common factor exists without doing any divisions in k? Here is a first answer.

Lemma 6. *Let $f, g \in k[x]$ be polynomials of degrees $l > 0$ and $m > 0$, respectively. Then f and g have a common factor if and only if there are polynomials $A, B \in k[x]$ such that:*

(i) *A and B are not both zero.*

(ii) *A has degree at most $m - 1$ and B has degree at most $l - 1$.*

(iii) *$Af + Bg = 0$.*

Proof. First, assume that f and g have a common factor $h \in k[x]$. Then $f = hf_1$ and $g = hg_1$, where $f_1, g_1 \in k[x]$. Note that f_1 has degree at most $l - 1$, and similarly $\deg(g_1) \leq m - 1$. Then

$$g_1 \cdot f + (-f_1) \cdot g = g_1 \cdot hf_1 - f_1 \cdot hg_1 = 0.$$

and, thus, $A = g_1$ and $B = -f_1$ have the required properties.

Conversely, suppose that A and B have the above three properties. By (i), we may assume $B \neq 0$. If f and g have no common factor, then their GCD is 1, so we can find polynomials $\tilde{A}, \tilde{B} \in k[x]$ such that $\tilde{A}f + \tilde{B}g = 1$ (see Proposition 6 of Chapter 1, §5). Now multiply by B and use $Bg = -Af$:

$$B = (\tilde{A}f + \tilde{B}g)B = \tilde{A}Bf + \tilde{B}Bg = \tilde{A}Bf - \tilde{B}Af = (\tilde{A}B - \tilde{B}A)f.$$

Since B is nonzero, this equation shows that B has degree at least l, which contradicts (ii). Hence, there must be a common factor of positive degree. □

The answer given by Lemma 6 may not seem very satisfactory, for we still need to decide whether the required A and B exist. Remarkably, we can use *linear algebra* to answer this last question. The idea is to turn $Af + Bg = 0$ into a system of linear equations. Write:

$$\begin{aligned} A &= c_0x^{m-1} + \cdots + c_{m-1}, \\ B &= d_0x^{l-1} + \cdots + d_{l-1}, \end{aligned}$$

where for now we will regard the $l + m$ coefficients $c_0, \ldots, c_{m-1}, d_0, \ldots, d_{l-1}$ as unknowns. Our goal is to find $c_i, d_i \in k$, not all zero, so that the equation

$$Af + Bg = 0 \tag{4}$$

holds. Note that this will automatically give us A and B as required in Lemma 6.

To get a system of linear equations, let us also write out f and g:

$$\begin{aligned} f &= a_0x^l + \cdots + a_l, \quad a_0 \neq 0,\\ g &= b_0x^m + \cdots + b_m, \quad b_0 \neq 0, \end{aligned}$$

where $a_i, b_i \in k$. If we substitute these formulas for f, g, A, and B into equation (4) and compare the coefficients of powers of x, then we get the following system of linear equations with unknowns c_i, d_i and coefficients a_i, b_i, in k:

$$\begin{array}{lclcl} a_0c_0 & + & b_0d_0 & = 0 & \text{coefficient of } x^{l+m-1}\\ a_1c_0 + a_0c_1 & + & b_1d_0 + b_0d_1 & = 0 & \text{coefficient of } x^{l+m-2}\\ \quad\ddots & & \quad\ddots & \vdots & \\ \qquad a_lc_{m-1} & + & \qquad b_md_{l-1} & = 0 & \text{coefficient of } x^0. \end{array} \tag{5}$$

Since there are $l + m$ linear equations and $l + m$ unknowns, we know from linear algebra that there is a nonzero solution if and only if the coefficient matrix has zero determinant. This leads to the following definition.

Definition 7. *Given polynomials $f, g \in k[x]$ of positive degree, write them in the form*

$$\begin{aligned} f &= a_0x^l + \cdots + a_l, \quad a_0 \neq 0,\\ g &= b_0x^m + \cdots + b_m, \quad b_0 \neq 0. \end{aligned}$$

Then the **Sylvester matrix** *of f and g with respect to x, denoted* $\mathrm{Syl}(f, g, x)$ *is the coefficient matrix of the system of equations given in (5). Thus,* $\mathrm{Syl}(f, g, x)$ *is the following $(l + m) \times (l + m)$ matrix:*

$$\mathrm{Syl}(f,g,x) = \left(\underbrace{\begin{matrix} a_0 & & & \\ a_1 & a_0 & & \\ a_2 & a_1 & \ddots & \\ \vdots & & \ddots & a_0 \\ & \vdots & & a_1 \\ a_l & & & \\ & a_l & & \vdots \\ & & \ddots & \\ & & & a_l \end{matrix}}_{m \text{ columns}} \underbrace{\begin{matrix} b_0 & & & \\ b_1 & b_0 & & \\ b_2 & b_1 & \ddots & \\ \vdots & & \ddots & b_0 \\ & \vdots & & b_1 \\ b_m & & & \\ & b_m & & \vdots \\ & & \ddots & \\ & & & b_m \end{matrix}}_{m \text{ columns}}\right),$$

where the empty spaces are filled by zeros. The **resultant** *of f and g with respect to x, denoted* $\mathrm{Res}(f, g, x)$*, is the determinant of the Sylvester matrix. Thus,*

$$\mathrm{Res}(f, g, x) = \det(\mathrm{Syl}(f, g, x)).$$

From this definition, we get the following properties of the resultant. A polynomial is called an *integer polynomial* provided that all of its coefficients are integers.

Proposition 8. *Given $f, g \in k[x]$ of positive degree, the resultant* $\text{Res}(f, g, x) \in k$ *is an integer polynomial in the coefficients of f and g. Furthermore, f and g have a common factor in $k[x]$ if and only if* $\text{Res}(f, g, x) = 0$.

Proof. The standard formula for the determinant of an $s \times s$ matrix $A = (a_{ij})_{l \le i, j \le s}$ is

$$\det(A) = \sum_{\substack{\sigma \text{ a permutation} \\ \text{of}\{1,\ldots,s\}}} \text{sgn}(\sigma) a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdots a_{s\sigma(s)},$$

where $\text{sgn}(\sigma)$ is + 1 if σ interchanges an even number of pairs of elements of $\{1, \ldots, s\}$ and -1 if σ interchanges an odd number of pairs (see Appendix A for more details). This shows that the determinant is an integer polynomial (in fact, the coefficients are ± 1) in its entries, and the first statement of the proposition then follows immediately from the definition of resultant.

The second statement is just as easy to prove: the resultant is zero $\Leftrightarrow$ the coefficient matrix of equations (5) has zero determinant $\Leftrightarrow$ equations (5) have a nonzero solution. We observed earlier that this is equivalent to the existence of A and B as in Lemma 6, and then Lemma 6 completes the proof of the proposition. □

As an example, let us see if $f = 2x^2 + 3x + 1$ and $g = 7x^2 + x + 3$ have a common factor in $\mathbb{Q}[x]$. One computes that

$$\text{Res}(f, g, x) = \det \begin{pmatrix} 2 & 0 & 7 & 0 \\ 3 & 2 & 1 & 7 \\ 1 & 3 & 3 & 1 \\ 0 & 1 & 0 & 3 \end{pmatrix} = 153 \neq 0,$$

so that there is no common factor.

One disadvantage to using resultants is that large determinants are hard to compute. In the exercises, we will explain an alternate method for computing resultants that is similar to the Euclidean Algorithm. Most computer algebra systems have a resultant command that implements this algorithm.

To link resultants to elimination, let us compute the resultant of the polynomials $f = xy - 1$ and $g = x^2 + y^2 - 4$. Regarding f and g as polynomials in x whose coefficients are polynomials in y, we get

$$\text{Res}(f, g, x) = \det \begin{pmatrix} y & 0 & 1 \\ -1 & y & 0 \\ 0 & -1 & y^2 - 4 \end{pmatrix} = y^4 - 4y^2 + 1.$$

More generally, if f and g are *any* polynomials in $k[x, y]$ in which x appears to a positive power, then we can compute $\text{Res}(f, g, x)$ in the same way. Since the coefficients are polynomials in y, Proposition 8 guarantees that $\text{Res}(f, g, x)$ is a polynomial in y. Thus, given $f, g \in k[x, y]$, we can use the resultant to eliminate x. But is this the same kind of elimination that we did in §§1 and 2? In particular, is $\operatorname{Res}(f, g, x)$ in the first elimination ideal $\langle f, g\rangle \cap k[y]$? To answer these questions, we will need the following result.

Proposition 9. *Given $f, g \in k[x]$ of positive degree, there are polynomials $A, B \in k[x]$ such that*

$$Af + Bg = \operatorname{Res}(f, g, x).$$

Furthermore, the coefficients of A and B are integer polynomials in the coefficients of f and g.

Proof. The definition of resultant was based on the equation $Af + Bg = 0$. In this proof, we will apply the same methods to the equation

$$\tilde{A}f + \tilde{B}g = 1. \tag{6}$$

The reason for using $\tilde{A}$ rather than A will soon be apparent.

The proposition is trivially true if $\operatorname{Res}(f, g, x) = 0$ (simply choose $A = B = 0$), so we may assume $\operatorname{Res}(f, g, x) \neq 0$. Now let

$$\begin{aligned} f &= a_0x^l + \cdots + a_l, \quad a_0 \neq 0,\\ g &= b_0x^m + \cdots + b_m, \quad b_0 \neq 0,\\ \tilde{A} &= c_0x^{m-1} + \cdots + c_{m-1},\\ \tilde{B} &= d_0x^{l-1} + \cdots + d_{l-1}, \end{aligned}$$

where the coefficients $c_0, \ldots, c_{m-1}, d_0, \ldots, d_{l-1}$ are unknowns in k. If we substitute these formulas into (6) and compare coefficients of powers of x, then we get the following system of linear equations with unknowns c_i, d_i and coefficients a_i, b_i in k:

$$\begin{array}{lllll} & a_0c_0 & + \quad b_0d_0 & = 0 & \text{coefficient of } x^{l+m-1}\\ & a_1c_0 + a_0c_1 & + \quad b_1d_0 + b_0d_1 & = 0 & \text{coefficient of } x^{l+m-2}\\ (7) & \quad\ddots & \qquad\ddots & \vdots & \\ & \qquad a_lc_{m-1} & + \qquad b_md_{l-1} & = 1 & \text{coefficient of } x^0. \end{array}$$

These equations are the same as (5) except for the 1 on the right-hand side of the last equation. Thus, the coefficient matrix is the Sylvester matrix of f and g, and then $\operatorname{Res}(f, g, x) \neq 0$ guarantees that (7) has a unique solution in k.

In this situation, we can use *Cramer's Rule* to give a *formula* for the unique solution. Cramer's Rule states that the i-th unknown is a ratio of two determinants, where the denominator is the determinant of the coefficient matrix and the numerator is the determinant of the matrix where the i-th column of the coefficient matrix has been replaced by the right-hand side of the equation. For a more precise statement of Cramer's rule, the reader should consult Appendix A. In our case, Cramer's rule gives formulas for the c_i's and d_i's. For example, the first unknown c_0 is given by

$$c_0 = \frac{1}{\operatorname{Res}(f, g, x)} \det \begin{pmatrix} 0 & & & & b_0 & & \\ 0 & a_0 & & & \vdots & \ddots & \\ \vdots & \vdots & \ddots & & \vdots & & b_0 \\ 0 & a_l & & a_0 & b_m & & \vdots \\ \vdots & & \ddots & \vdots & & \ddots & \vdots \\ 1 & & & a_l & & & b_m \end{pmatrix}.$$

Since a determinant is an integer polynomial in its entries, it follows that

$$c_0 = \frac{\text{an integer polynomial in } a_i, b_i}{\operatorname{Res}(f, g, x)}.$$

There are similar formulas for the other c_i's and the other d_i's. Since $\tilde{A} = c_0 x^{m-1} + \cdots + c_{m-1}$, we can pull out the common denominator $\operatorname{Res}(f, g, x)$ and write $\tilde{A}$ in the form

$$\tilde{A} = \frac{1}{\operatorname{Res}(f, g, x)} A,$$

where $A \in k[x]$ and the coefficients of A are integer polynomials in a_i, b_i. Similarly, we can write

$$\tilde{B} = \frac{1}{\operatorname{Res}(f, g, x)} B,$$

where $B \in k[x]$ has the same properties as A. Since $\tilde{A}$ and $\tilde{B}$ satisfy $\tilde{A}f + \tilde{B}g = 1$, we can multiply through by $\operatorname{Res}(f, g, x)$ to obtain

$$Af + Bg = \operatorname{Res}(f, g, x).$$

Since A and B have the required kind of coefficients, the proposition is proved. □

Most courses in linear algebra place little emphasis on Cramer's rule, mainly because Gaussian elimination is much more efficient (from a computational point of view) than Cramer's rule. But for theoretical uses, where one needs to worry about the *form* of the solution, Cramer's rule is very important (as shown by the above proposition).

We can now explain the relation between the resultant and the GCD. Given $f, g \in k[x]$, $\operatorname{Res}(f, g, x) \neq 0$ tells us that f and g have no common factor, and hence their GCD is 1. Then Proposition 6 of Chapter 1, §5 says that there are $\tilde{A}$ and $\tilde{B}$ such that $\tilde{A}f + \tilde{B}g = 1$. As the above formulas for $\tilde{A}$ and $\tilde{B}$ make clear, the coefficients of $\tilde{A}$ and $\tilde{B}$ have a denominator given by the resultant (though the resultant need not be the smallest denominator). Then clearing these denominators leads to $Af + Bg = \operatorname{Res}(f, g, x)$.

To see this more explicitly, let us return to the case of $f = xy - 1$ and $g = x^2 + y^2 - 4$. If we regard these as polynomials in x, then we computed that $\operatorname{Res}(f, g, x) = y^4 - 4y^2 + 1 \neq 0$. Thus, their GCD is 1, and we leave it as an exercise to check that

$$-\left(\frac{y}{y^4-4y^2+1}x+\frac{1}{y^4-4y^2+1}\right)f+\frac{y^2}{y^4-4y^2+1}g=1.$$

Note that this equation takes place in $k(y)[x]$, i.e., the coefficients are *rational functions* in y. This is because the GCD theory from §5 of Chapter 1 requires *field* coefficients. If we want to work in $k[x, y]$, we must clear denominators, which leads to

$$-(yx+1)f+y^2g=y^4-4y^2+1. \tag{8}$$

This, of course, is just a special case of Proposition 9. Hence, we can regard the resultant as a "denominator-free" version of the GCD.

We have now answered the question posed before Proposition 9, for (8) shows that the resultant $y^4 - 4y^2 + 1$ is in the elimination ideal. More generally, it is clear that if $f, g \in k[x, y]$ are any polynomials of positive degree in x, then $\mathrm{Res}(f, g, x)$ always lies in the first elimination ideal of $\langle f, g\rangle$. In §6, we see how these ideas apply to the case of $f, g \in k[x_1, \ldots, x_n]$.

In addition to the resultant of two polynomials discussed here, the resultant of three or more polynomials can be defined. Readers interested in *multipolynomial resultants* should consult MACAULAY (1902) or VAN DER WAERDEN (1931). Modern introductions to this theory can be found in BAJAJ, GARRITY and WARREN (1988), CANNY and MANOCHA (1993), or COX, LITTLE and O'SHEA (1998). A more sophisticated treatment of resultants is presented in JOUANOLOU (1991), and a vast generalization of the concept of resultant is discussed in GELFAND, KAPRANOV and ZELEVINSKY (1994).

EXERCISES FOR §5

1. Here are some examples of irreducible polynomials.
   a. Show that every $f \in k[x]$ of degree 1 is irreducible over k.
   b. Let $f \in k[x]$ have degree 2 or 3. Show that f is irreducible over k if and only if f has no roots in k.
   c. Use part (b) to show that $x^2 - 2$ and $x^3 - 2$ are irreducible over $\mathbb{Q}$ but not over $\mathbb{R}$.
   d. Prove that $x^4 + 1$ is irreducible over $\mathbb{Q}$ but not over $\mathbb{R}$. This one is a bit harder.
   e. Use part (d) to show that part (b) can fail for polynomials of degree ≥ 4.
2. Prove that a field k is algebraically closed if and only if every irreducible polynomial in $k[x]$ has degree 1.
3. This exercise is concerned with the proof of Theorem 3. Suppose that $g = \Sigma_i a_i x_1^i$ and $h = \Sigma_i b_i x_1^i$, where $a_i, b_i \in k[x_2, \ldots, x_n]$.
   a. Let $u \in k[x_2, \ldots, x_n]$. Show that u divides g in $k[x_1, \ldots, x_n]$ if and only if, in $k[x_2, \ldots, x_n]$, u divides every a_i.
   b. If we write $gh = \Sigma_i c_i x_1^i$, verify that c_{i+j} is given by the formula that appears in the proof of Theorem 3.
4. In this exercise, we will prove Theorem 5.
   a. If f is irreducible and divides a product $h_1 \cdots h_s$, then prove that f divides h_i for some i.
   b. The existence of a factorization follows from Proposition 2. Now prove the uniqueness part of Theorem 5. Hint: If $f = f_i \cdots f_r = g_1 \cdots g_s$, where the f_i's and g_j's are

irreducible, apply part a to show that f_1 divides some g_j. Then argue g_j is a constant multiple of f_1, and hence we can cancel f_1 from each side. Use induction on the total, degree of f.

5. Compute the resultant of $x^5 - 3x^4 - 2x^3 + 3x^2 + 7x + 6$ and $x^4 + x^2 + 1$. Do these polynomials have a common factor in $\mathbb{Q}[x]$? Explain your reasoning.
6. In Exercise 14 of Chapter 1, §5, we proved that if $f = c(x - a_1)^{r_1} \cdots (x - a_l)^{r_1} \in \mathbb{C}[x]$, then
$$\mathrm{GCD}(f, f') = (x - a_1)^{r_1 - 1} \cdots (x - a_l)^{r_l - 1}.$$

   Over an arbitrary field k, a given polynomial $f \in k[x]$ of positive degree may not be a product of linear factors. But by unique factorization, we know that f can be written in the form
$$f = cf_1^{r_1} \cdots f_l^{r_l}, \quad c \in k,$$

   where $f_1, \ldots, f_l \in k[x]$ are irreducible and no f_i is a constant multiple of f_j for $j \neq i$. Prove that if k contains the rational numbers $\mathbb{Q}$, then
$$\mathrm{GCD}(f, f') = f_1^{r_1 - 1} \cdots f_l^{r_l - 1}.$$

   Hint: Follow the outline of Exercise 14 of Chapter 1, §5. Your proof will use unique factorization. The hypothesis $\mathbb{Q} \subset k$ is needed to ensure that $f' \neq 0$.
7. If $f, g \in \mathbb{C}[x]$ are polynomials of positive degree, prove that f and g have a common root in $\mathbb{C}$ if and only if $\mathrm{Res}(f, g, x) = 0$. Hint: Use Proposition 8 and the fact that $\mathbb{C}$ is algebraically closed.
8. If $f = a_0x^l + \cdots + a_l \in k[x]$, where $a_0 \neq 0$ and $l > 0$, then the *discriminant* of f is defined to be
$$\mathrm{disc}(f) = \frac{(-1)^{l(l-1)/2}}{a_0}\mathrm{Res}(f, f', x).$$

   Prove that f has a multiple factor (i.e., f is divisible by h^2 for some $h \in k[x]$ of positive degree) if and only if disc $(f) = 0$. Hint: Use Exercise 6 (you may assume $\mathbb{Q} \subset k$). Over the complex numbers, Exercise 7 implies that a polynomial has a multiple root if and only if its discriminant is zero.
9. Use the previous exercise to determine whether or not $6x^4 - 23x^3 + 32x^2 - 19x + 4$ has a multiple root in $\mathbb{C}$. What is the multiple root?
10. Compute the discriminant of the quadratic polynomial $f = ax^2 + bx + c$. Explain how your answer relates to the quadratic formula, and, without using Exercise 8, prove that f has a multiple root if and only if its discriminant vanishes.
11. Consider the polynomials $f = 2x^2 + 3x + 1$ and $g = 7x^2 + x + 3$.
    a. Use the Euclidean Algorithm (by hand, not computer) to find the GCD of these polynomials.
    b. Find polynomials $A, B \in \mathbb{Q}[x]$ such that $Af + Bg = 1$. Hint: Use the calculations you made in part a.
    c. In the equation you found in part b, clear the denominators. How does this answer relate to the resultant?
12. If $f, g \in \mathbb{Z}[x]$, explain why $\mathrm{Res}(f, g, x) \in \mathbb{Z}$.
13. Let $f = xy - 1$ and $g = x^2 + y^2 - 4$. We will regard f and g as polynomials in x with coefficients in $k(y)$.

a. With f and g as above, set up the system of equations (7) that describes $\tilde{A}f + \tilde{B}g = 1$. Hint: $\tilde{A}$ is linear and $\tilde{B}$ is constant. Thus, you should have three equations in three unknowns.

b. Use Cramer's rule to solve the system of equations obtained in part (a). Hint: The denominator is the resultant.

c. What equation do you get when you clear denominators in part (b)? Hint: See equation (8) in the text.

14. In the text, we defined $\mathrm{Res}(f, g, x)$ when $f, g \in k[x]$ have positive degree. In this problem, we will explore what happens when one (or both) of f and g are constant.

a. First, assume that f has degree $l > 0$ and $g = b_0$ is constant (possibly 0). Show that the Sylvester matrix of f and g is the $l \times l$ matrix with b_0 on the main diagonal and 0's elsewhere. Conclude that $\mathrm{Res}(f, b_0, x) = b_0^l$.

b. When f and g are as in part a, show that Propositions 8 and 9 are still true.

c. What is $\mathrm{Res}(a_0, g, x)$ when $f = a_0$ is constant (possibly zero) and g has degree $m > 0$? Explain your reasoning.

d. The one case not covered so far is when both $f = a_0$ and $g = b_0$ are constants. In this case, one defines.

$$\mathrm{Res}(a_0, b_0) = \begin{cases} 0 & \text{if either } a_0 = 0 \text{ or } b_0 = 0 \\ 1 & \text{if } a_0 \neq 0 \text{ and } b_0 \neq 0. \end{cases}$$

By considering $f = g = 2$ in $\mathbb{Q}[x]$, show that Propositions 8 and 9 can fail when f and g are constant. Hint: Look at the statements requiring that certain things be integer polynomials in the coefficients of f and g.

15. Prove that if f has degree l and g has degree m, then the resultant has the following symmetry property:

$$\mathrm{Res}(f, g, x) = (-1)^{lm}\mathrm{Res}(g, f, x).$$

Hint: A determinant changes sign if you switch two columns.

16. Let $f = a_0x^l + \cdots + a_l$ and $g = b_0x^m + \cdots + b_m$ be polynomials in $k[x]$, and assume that $l \geq m$.

a. Let $\tilde{f} = f - (a_0/b_0)x^{l-m}g$, so that $\deg(\tilde{f}) \leq l - 1$. If $\deg(\tilde{f}) = l - 1$, then prove

$$\mathrm{Res}(f, g, x) = (-1)^m b_0 \mathrm{Res}(\tilde{f}, g, x).$$

Hint: Use column operations on the Sylvester matrix. You will subtract a_0/b_0 times the first m columns in the g part from the columns in the f part. Then expand by minors along the first row. [See Theorem 5.7 of FINKBEINER (1978) for a description of expansion by minors.]

b. Let $\tilde{f}$ be as in part (a), but this time we allow the possibility that the degree of $\tilde{f}$ could be strictly smaller than $l - 1$. Prove that

$$\mathrm{Res}(f, g, x) = (-1)^{m(l-\deg(\tilde{f}))} b_0^{l-\deg(\tilde{f})} \mathrm{Res}(\tilde{f}, g, x).$$

Hint: The exponent $l - \deg(\tilde{f})$ tells you how many times to expand by minors.

c. Now use the division algorithm to write $f = qg + r$ in $k[x]$, where $\deg(r) < \deg(g)$. Then use part (b) to prove that

$$\mathrm{Res}(f, g, x) = (-1)^{m(l-\deg(r))} b_0^{l-\deg(r)} \mathrm{Res}(r, g, x).$$

17. In this exercise, we will modify the Euclidean Algorithm to give an algorithm for computing resultants. The basic idea is the following: to find the GCD of f and g, we used the division

algorithm to write $f = qg + r$, $g = q'r + r'$, etc. In equation (5) of Chapter 1, §5, the equalities

$$\mathrm{GCD}(f, g) = \mathrm{GCD}(g, r) = \mathrm{GCD}(r, r') = \cdots$$

enabled us to compute the GCD since the degrees were decreasing. Using Exercises 15 and 16, we get the following "resultant" version of the first two equalities above:

$$\begin{aligned}
\mathrm{Res}(f, g, x) &= (-1)^{\deg(g)(\deg(f)-\deg(r))} b_0^{\deg(f)-\deg(r)} \mathrm{Res}(r, g, x)\\
&= (-1)^{\deg(f)\deg(g)} b_0^{\deg(f)-\deg(r)} \mathrm{Res}(g, r, x)\\
&= (-1)^{\deg(f)\deg(g)+\deg(r)(\deg(g)-\deg(r'))} b_0^{\deg(f)-\deg(r)} b_0'^{\deg(g)-\deg(r')} \mathrm{Res}(r', r, x)\\
&= (-1)^{\deg(f)\deg(g)+\deg(g)\deg(r)} b_0^{\deg(f)-\deg(r)} b_0'^{\deg(g)-\deg(r')} \mathrm{Res}(r, r', x)
\end{aligned}$$

where b_0 (resp. b_0') is the leading coefficient of g (resp. r). Continuing in this way, we can reduce to the case where the second polynomial is constant, and then we can use Exercise 14 to compute the resultant.

To set this up as pseudocode, we need to introduce two functions: let $r = \mathrm{remainder}(f, g)$ be the remainder on division of f by g and let $\mathrm{lead}(f)$ be the leading coefficient of f. We can now state the algorithm for finding $\mathrm{Res}(f, g, x)$

```
Input: $f, g$
Output: $res$
$h := f$
$s := g$
$res := 1$
WHILE $\deg(s) > 0$ DO
        $r := \mathrm{remainder}(h, s)$
        $res := (-1)^{\deg(h)\deg(s)}\ \mathrm{lead}\ (s)^{\deg(h)-\deg(r)} res$
        $h := s$
        $s := r$
IF $h = 0$ or $s = 0$ THEN $res := 0$ ELSE
IF $\deg(h) > 0$ THEN $res := s^{\deg(h)} res$
```

Prove that this algorithm computes the resultant of f and g. Hint: Use Exercises 14, 15, and 16, and follow the proof of Proposition 6 of Chapter 1, §5.

§6 Resultants and the Extension Theorem

In this section we will prove the Extension Theorem using the results of §5. Our first task will be to adapt the theory of resultants to the case of polynomials in n variables. Thus, suppose we are given $f, g \in k[x_1, \ldots, x_n]$ of positive degree in x_1. As in §5, we write

$$\begin{aligned}
f &= a_0 x_1^l + \cdots + a_l, \quad a_0 \neq 0,\\
g &= b_0 x_1^m + \cdots + b_m, \quad b_0 \neq 0,
\end{aligned} \tag{1}$$

where $a_i, b_i \in k[x_2, \ldots, x_n]$, and we define the *resultant* of f and g with respect to x_1

to be the determinant

$$\operatorname{Res}(f, g, x_1) = \det \begin{array}{c} \begin{pmatrix} a_0 & & & & b_0 & & & \\ a_1 & a_0 & & & b_1 & b_0 & & \\ & a_1 & \ddots & & & b_1 & \ddots & \\ \vdots & & \ddots & a_0 & \vdots & & \ddots & b_0 \\ & \vdots & & a_1 & & \vdots & & b_1 \\ a_l & & & & b_m & & & \\ & a_l & & \vdots & & b_m & & \vdots \\ & & \ddots & & & & \ddots & \\ & & & a_l & & & & b_m \end{pmatrix} \\ \underbrace{\hphantom{a_0 \quad a_0 \quad \ddots \quad a_0}}_{m \text{ columns}} \quad \underbrace{\hphantom{b_0 \quad b_0 \quad \ddots \quad b_0}}_{l \text{ columns}} \end{array},$$

where the empty spaces are filled by zeros.

For resultants of polynomials in several variables, the results of §5 can be stated as follows.

Proposition 1. *Let $f, g \in k[x_1, \ldots, x_n]$ have positive degree in x_1. Then:*

(i) *$Res(f, g, x_1)$ is in the first elimination ideal $\langle f, g\rangle \cap k[x_2, \ldots, x_n]$.*

(ii) *$Res(f, g, x_1) = 0$ if and only if f and g have a common factor in $k[x_1, \ldots, x_n]$ which has positive degree in x_1.*

Proof. When we write f, g in terms of x_1, the coefficients a_i, b_i, lie in $k[x_2, \ldots, x_n]$. Since the resultant is an integer polynomial in a_i, b_i, (Proposition 9 of §5), it follows that $\operatorname{Res}(f, g, x_1) \in k[x_2, \ldots, x_n]$. We also know that

$$Af + Bg = \operatorname{Res}(f, g, x_1),$$

where A and B are polynomials in x_1 whose coefficients are again integer polynomials in a_i, b_i (Proposition 9 of §5). Thus, $A, B \in k[x_2, \ldots, x_n][x_1] = k[x_1, \ldots, x_n]$, and then the above equation implies $\operatorname{Res}(f, g, x_1) \in \langle f, g\rangle$. This proves part (i) of the proposition.

To prove the second part, we will use Proposition 8 of §5 to interpret the vanishing of the resultant in terms of common factors. In §5, we worked with polynomials in one variable with coefficients in a field. Since f and g are polynomials in x_1 with coefficients in $k[x_2, \ldots, x_n]$ the *field* the coefficients lie in is $k(x_2, \ldots, x_n)$. Then Proposition 8 of §5, applied to $f, g \in k(x_2, \ldots, x_n)[x_1]$, tells us that $\operatorname{Res}(f, g, x_1) = 0$ if and only if f and g have a common factor in $k(x_2, \ldots, x_n)[x_1]$ which has positive degree in x_1. But then we can apply Corollary 4 of §5, which says that this is equivalent to having a common factor in $k[x_1, \ldots, x_n]$ of positive degree in x_1. The proposition is proved. □

Over the complex numbers, two polynomials in $\mathbb{C}[x]$ have a common factor if and only if they have a common root (this is easy to prove). Thus, we get the following corollary of Proposition 1.

Corollary 2. *If $f, g \in \mathbb{C}[x]$, then* $\operatorname{Res}(f, g, x) = 0$ *if and only if f and g have a common root in $\mathbb{C}$.*

We will prove the Extension Theorem by studying the interaction between resultants and partial solutions. Given $f, g \in \mathbb{C}[x_1, \ldots, x_n]$, we get the resultant

$$h = \operatorname{Res}(f, g, x_1) \in \mathbb{C}[x_2, \ldots, x_n]$$

as in Proposition 1. If we substitute $\mathbf{c} = (c_2, \ldots, c_n) \in \mathbb{C}^{n-1}$ into h, we get a *specialization* of the resultant. However, this need not equal the resultant of the specialized polynomials $f(x_1, \mathbf{c})$ and $g(x_1, \mathbf{c})$. See Exercises 8 and 9 for some examples. Fortunately, there is one situation where the exact relation between these resultants is easy to state.

Proposition 3. *Let $f, g \in \mathbb{C}[x_1, \ldots, x_n]$ have degree l, m respectively, and let $\mathbf{c} = (c_2, \ldots, c_n) \in \mathbb{C}^{n-1}$ satisfy the following:*
(i) *$f(x_1, \mathbf{c}) \in \mathbb{C}[x_1]$ has degree l.*
(ii) *$g(x_1, \mathbf{c}) \in \mathbb{C}[x_1]$ has degree $p \leq m$.*
Then the polynomial $h = \operatorname{Res}(f, g, x_1) \in \mathbb{C}[x_2, \ldots, x_n]$ *satisfies*

$$h(\mathbf{c}) = a_0(\mathbf{c})^{m-p} \operatorname{Res}(f(x_1, \mathbf{c}), g(x_1, \mathbf{c}), x_1),$$

where a_0 is as in (1).

Proof. If we substitute $\mathbf{c} = (c_2, \ldots, c_n)$ for $x_2, \ldots, x_n$ in the determinantal formula for $h = \operatorname{Res}(f, g, x_1)$, we obtain

$$h(\mathbf{c}) = \det \begin{array}{c} \left(\begin{array}{cccccc} a_0(\mathbf{c}) & & & b_0(\mathbf{c}) & & \\ \vdots & \ddots & & \vdots & \ddots & \\ \vdots & & a_0(\mathbf{c}) & \vdots & & b_0(\mathbf{c}) \\ a_l(\mathbf{c}) & & \vdots & b_m(\mathbf{c}) & & \vdots \\ & \ddots & \vdots & & \ddots & \vdots \\ & & a_l(\mathbf{c}) & & & b_m(\mathbf{c}) \end{array}\right) \\ \underbrace{\hphantom{a_0(\mathbf{c}) \quad \ddots \quad a_0(\mathbf{c})}}_{m \text{ columns}} \quad \underbrace{\hphantom{b_0(\mathbf{c}) \quad \ddots \quad b_0(\mathbf{c})}}_{l \text{ columns}} \end{array}.$$

First suppose that $g(x_1, \mathbf{c})$ has degree $p = m$. Then our assumptions imply that

$$\begin{aligned} f(x_1, \mathbf{c}) &= a_0(\mathbf{c})x_1^l + \cdots + a_l(\mathbf{c}), \qquad a_0(\mathbf{c}) \neq 0, \\ g(x_1, \mathbf{c}) &= b_0(\mathbf{c})x_1^m + \cdots + b_m(\mathbf{c}), \quad b_0(\mathbf{c}) \neq 0. \end{aligned}$$

Hence the above determinant is the resultant of $f(x_1, \mathbf{c})$ and $g(x_1, \mathbf{c})$, so that

$$h(\mathbf{c}) = \operatorname{Res}(f(x_1, \mathbf{c}), g(x_1, \mathbf{c}), x_1).$$

This proves the proposition when $p = m$. When $p < m$, the above determinant is no longer the resultant of $f(x_1, \mathbf{c})$ and $g(x_1, \mathbf{c})$ (it has the wrong size). Here, we get the desired resultant by repeatedly expanding by minors along the first row. We leave the details to the reader (see Exercise 10). □

We can now prove the Extension Theorem. Let us recall the statement of the theorem.

Theorem 4 (The Extension Theorem). *Let $I = \langle f_1, \ldots, f_s\rangle \subset \mathbb{C}[x_1, \ldots, x_n]$, and let I_1 be the first elimination ideal of I. For each $1 \le i \le s$, write f_i in the form*

$$f_i = g_i(x_2, \ldots, x_n)x_1^{N_i} + \textit{terms in which } x_1 \textit{ has degree} < N_i,$$

where $N_i \ge 0$ and $g_i \in \mathbb{C}[x_2, \ldots, x_n]$ is nonzero. Suppose that we have a partial solution $(c_2, \ldots, c_n) \in \mathbf{V}(I_1)$. If $(c_2, \ldots, c_n) \notin \mathbf{V}(g_1, \ldots, g_s)$, then there exists $c_1 \in \mathbb{C}$ such that $(c_1, c_2, \ldots, c_n) \in \mathbf{V}(I)$.

Proof. As above, we set $\mathbf{c} = (c_2, \ldots, c_n)$. Then consider the ring homomorphism

$$\mathbb{C}[x_1, \ldots, x_n] \longrightarrow \mathbb{C}[x_1]$$

defined by $f(x_1, x_2, \ldots, x_n) \mapsto f(x_1, \mathbf{c})$. In Exercise 11, you will show that the image of I under this homomorphism is an ideal of $\mathbb{C}[x_1]$. Since $\mathbb{C}[x_1]$ is a PID, the image of I is generated by a single polynomial $u(x_1)$. In other words,

$$\{f(x_1, \mathbf{c}) : f \in I\} = \langle u(x_1)\rangle.$$

If $u(x_1)$ is nonconstant, then there is $c_1 \in \mathbb{C}$ such that $u(c_1) = 0$ by the Fundamental Theorem of Algebra. It follows that $f(c_1, \mathbf{c}) = 0$ for all $f \in I$, so that $(c_1, \mathbf{c}) = (c_1, c_2, \ldots, c_n) \in \mathbf{V}(I)$. Note that this argument also works if $u(x_1)$ is the zero polynomial.

It remains to consider what happens when $u(x_1)$ is a nonzero constant u_0. By the above equality, there is $f \in I$ such that $f(x_1, \mathbf{c}) = u_0$. We will show that this case cannot occur. By hypothesis, our partial solution satisfies $\mathbf{c} \notin \mathbf{V}(g_1, \ldots, g_s)$. Hence $g_i(\mathbf{c}) \neq 0$ for some i. Then consider

$$h = \mathrm{Res}(f_i, f, x_1) \in \mathbb{C}[x_2, \ldots, x_n].$$

Applying Proposition 3 to f_i and f, we obtain

$$h(\mathbf{c}) = g_i(\mathbf{c})^{\deg(f)}\mathrm{Res}(f_i(x_1, \mathbf{c}), u_0, x_1)$$

since $f(x_1, \mathbf{c}) = u_0$. We also have $\mathrm{Res}(f_i(x_1, \mathbf{c}), u_0, x_1) = u_0^{N_i}$ by part (a) of Exercise 14 of §5. Hence

$$h(\mathbf{c}) = g_i(\mathbf{c})^{\deg(f)}u_0^{N_i} \neq 0.$$

However, $f_i, f \in I$ and Proposition 1 imply that $h \in I_1$, so that $h(\mathbf{c}) = 0$ since $\mathbf{c} \in \mathbf{V}(I_1)$. This contradiction completes the proof of the Extension Theorem. □

The proof of the Extension Theorem just given is elegant but nonconstructive, since it does not explain how to construct the polynomial $u(x_1)$. Exercise 12 will describe a constructive method for getting a polynomial whose root gives the desired c_1.

A final observation to make is that the Extension Theorem is true over any algebraically closed field. For concreteness, we stated the theorem only for the complex numbers $\mathbb{C}$, but if you examine the proof carefully, you will see that the required c_1 exists because a nonconstant polynomial in $\mathbb{C}[x_1]$ has a root in $\mathbb{C}$. Since this property is true for any algebraically closed field, it follows that the Extension Theorem holds over such fields (see Exercise 13 for more details).

EXERCISES FOR §6

1. In $k[x, y]$, consider the two polynomials
$$\begin{aligned} f &= x^2y - 3xy^2 + x^2 - 3xy, \\ g &= x^3y + x^3 - 4y^2 - 3y + 1. \end{aligned}$$
   a. Compute $\operatorname{Res}(f, g, x)$.
   b. Compute $\operatorname{Res}(f, g, y)$.
   c. What does the result of part b imply about f and g?
2. Let $f, g \in \mathbb{C}[x, y]$. In this exercise, you will prove that
$$\mathbf{V}(f, g) \text{ is infinite } \iff f \text{ and } g \text{ have a nonconstant common factor in } \mathbb{C}[x, y].$$
   a. Prove that $\mathbf{V}(f)$ is infinite when f is nonconstant. Hint: Suppose f has positive degree in x. Then the leading coefficient of x in f can vanish for at most finitely many values of y. Now use the fact that $\mathbb{C}$ is algebraically closed.
   b. If f and g have a nonconstant common factor $h \in [x, y]$, then use part a to show that $\mathbf{V}(f, g)$ is infinite.
   c. If f and g have no nonconstant common factor, show that $\operatorname{Res}(f, g, x)$ and $\operatorname{Res}(f, g, y)$ are nonzero and conclude that $\mathbf{V}(f, g)$ is finite.
3. If $f, g \in k[x, y]$, Proposition 1 shows that $\operatorname{Res}(f, g, x) \in I_1 = \langle f, g \rangle \cap k[y]$. The interesting fact is that the resultant *need not* generate I_1.
   a. Show that $\operatorname{Res}(f, g, x)$ generates I_1 when $f = xy - 1$ and $g = x^2 + y^2 - 4$.
   b. Show that $\operatorname{Res}(f, g, x)$ does not generate I_1 when $f = xy - 1$ and $g = yx^2 + y^2 - 4$. Do you see any connection between part b and the Extension Theorem?
4. Suppose that $f, g \in \mathbb{C}[x]$ are polynomials of positive degree. The goal of this problem is to construct a polynomial whose roots are all sums of a root of f plus a root of g.
   a. Show that a complex number $\gamma \in \mathbb{C}$ can be written $\gamma = \alpha + \beta$, where $f(\alpha) = g(\beta) = 0$, if and only if the equations $f(x) = g(y - x) = 0$ have a solution with $y = \gamma$.
   b. Using Proposition 3, show that γ is a root of $\operatorname{Res}(f(x), g(y - x), x)$ if and only if $\gamma = \alpha + \beta$, where $f(\alpha) = g(\beta) = 0$.
   c. Construct a polynomial with coefficients in $\mathbb{Q}$ which has $\sqrt{2} + \sqrt{3}$ as a root. Hint: What are f and g in this case?
   d. Modify your construction to create a polynomial whose roots are all differences of a root of f minus a root of g.
5. Suppose that $f, g \in \mathbb{C}[x]$ are polynomials of positive degree. If all of the roots of f are nonzero, adapt the argument of Exercise 4 to construct a polynomial whose roots are all products of a root of f times a root of g.
6. Suppose that $f, g \in \mathbb{Q}[x]$ are polynomials of positive degree.
   a. Most computer algebra systems have a command for factoring polynomials over $\mathbb{Q}$ into irreducibles over $\mathbb{Q}$. In particular, one can determine if a given polynomial has any integer roots. Combine this with part (d) of Exercise 4 to describe an algorithm for determining when f and g have roots α and β, respectively, which differ by an integer.
   b. Show that the polynomials $f = x^5 - 2x^3 - 2x^2 + 4$ and $g = x^5 + 5x^4 + 8x^3 + 2x^2 - 5x + 1$ have roots which differ by an integer. What is the integer?
7. In §3, we mentioned that resultants are sometimes used to solve implicitization problems. For a simple example of how this works, consider the curve parametrized by
$$u = \frac{t^2}{1 + t^2}, \quad v = \frac{t^3}{1 + t^2}.$$

To get an implicit equation, form the equations

$$u(1+t^2) - t^2 = 0, \qquad v(1+t^2) - t^3 = 0$$

and use an appropriate resultant to eliminate t. Then compare your result to the answer obtained by the methods of §3. (Note that Exercise 13 of §3 is relevant.)

8. In the discussion leading up to Proposition 3, we claimed that the specialization of a resultant need not be the resultant of the specialized polynomials. Let us work out some examples.
   a. Let $f = x^2y + 3x - 1$ and $g = 6x^2 + y^2 - 4$. Compute $h = \mathrm{Res}(f, g, x)$ and show that $h(0) = -180$. But if we set $y = 0$ in f and g, we get the polynomials $3x - 1$ and $6x^2 - 4$. Check that $\mathrm{Res}(3x - 1, 6x^2 - 4) = -30$. Thus, $h(0)$ is *not* a resultant—it is off by a factor of 6. Note why equality fails: $h(0)$ is a 4×4 determinant, whereas $\mathrm{Res}(3x - 1, 6x^2 - 4)$ is a 3×3 determinant.
   b. Now let $f = x^2y + 3xy - 1$ and $g = 6x^2 + y^2 - 4$. Compute $h = \mathrm{Res}(f, g, x)$ and verify that $h(0) = 36$. Setting $y = 0$ in f and g gives polynomials -1 and $6x^2 - 4$. Use Exercise 14 of §5 to show that the resultant of these polynomials is 1. Thus, $h(0)$ is off by a factor of 36.

   When the degree of f drops by 1 (in part a), we get an extra factor of 6, and when it drops by 2 (in part b), we get an extra factor of $36 = 6^2$. And the leading coefficient of x in g is 6. In Exercise 11, we will see that this is no accident.
9. Let $f = x^2y + x - 1$ and $g = x^2y + x + y^2 - 4$. If $h = \mathrm{Res}(f, g, x) \in \mathbb{C}[y]$, show that $h(0) = 0$. But if we substitute $y = 0$ into f and g, we get $x - 1$ and $x - 4$. Show that these polynomials have a nonzero resultant. Thus, $h(0)$ is *not* a resultant.
10. In this problem you will complete the proof of Theorem 4 by determining what happens to a resultant when specializing causes the degree of one of the polynomials to drop. Let $f, g \in \mathbb{C}[x_1, \ldots, x_n]$ and set $h = \mathrm{Res}(f, g, x_1)$. If $\mathbf{c} = (c_2, \ldots, c_n) \in \mathbb{C}^{n-1}$, let $f(x_1, \mathbf{c})$ be the polynomial in $k[x_1]$ obtained by substituting in $\mathbf{c}$. As in (1), let $a_0, b_0 \in \mathbb{C}[x_2, \ldots, x_n]$ be the leading coefficients of x_1 in f, g, respectively. We will assume that $a_0(\mathbf{c}) \neq 0$ and $b_0(\mathbf{c}) = 0$, and our goal is to see how $h(\mathbf{c})$ relates to $\mathrm{Res}(f(x_1, \mathbf{c}), g(x_1, \mathbf{c}), x_1)$.
    a. First suppose that the degree of g drops by exactly 1, which means that $b_1(\mathbf{c}) \neq 0$. In this case, prove that

    $$h(\mathbf{c}) = a_0(\mathbf{c}) \cdot \mathrm{Res}(f(x_1, \mathbf{c}), g(x_1, \mathbf{c}, x_1).$$

    Hint: $h(\mathbf{c})$ is given by the following determinant:

$$h(\mathbf{c}) = \det \begin{pmatrix}
a_0(\mathbf{c}) & & & & 0 & & & \\
a_1(\mathbf{c}) & a_0(\mathbf{c}) & & & b_1(\mathbf{c}) & 0 & & \\
 & a_1(\mathbf{c}) & \ddots & & & b_1(\mathbf{c}) & \ddots & \\
\vdots & & \ddots & a_0(\mathbf{c}) & \vdots & & \ddots & 0 \\
 & \ddots & & a_1(\mathbf{c}) & & \vdots & & b_1(\mathbf{c}) \\
a_l(\mathbf{c}) & & & & b_m(\mathbf{c}) & & & \\
 & a_l(\mathbf{c}) & & \vdots & & b_m(\mathbf{c}) & & \vdots \\
 & & \ddots & & & & \ddots & \\
 & & & a_l(\mathbf{c}) & & & & b_m(\mathbf{c})
\end{pmatrix}.$$

$$\underbrace{\qquad\qquad\qquad\qquad}_{m \text{ columns}} \qquad \underbrace{\qquad\qquad\qquad\qquad}_{l \text{ columns}}$$

The determinant is the wrong size to be the resultant of $f(x_1, \mathbf{c})$ and $g(x_1, \mathbf{c})$. If you expand by minors along the first row [see Theorem 5.7 of FINKBEINER (1978)], the desired result will follow.

b. Now let us do the general case. Suppose that the degree of $g(x_1, \mathbf{c})$ is $m - p$, where $p \geq 1$. Then prove that

$$h(\mathbf{c}) = a_0(\mathbf{c})^p \cdot \mathrm{Res}(f(x_1, \mathbf{c}), g(x_1, \mathbf{c}), x_1).$$

Hint: Expand by minors p times. Note how this formula explains the results of Exercise 8.

11. Suppose that k is a field and $\varphi : k[x_1, \ldots, x_n] \to k[x_1]$ is a ring homomorphism that is the identity on k and maps x_1 to x_1. Given an ideal $I \subset k[x_1, \ldots, x_n]$, prove that $\varphi(I) \subset k[x_1]$ is an ideal. (In the proof of Theorem 4, we use this result when φ is the map that evaluates x_i at c_i for $2 \leq i \leq n$.)

12. Suppose that $I = \langle f_1, \ldots, f_s \rangle \subset \mathbb{C}[x_1, \ldots, x_n]$ and $\mathbf{c} = (c_2, \ldots, c_s) \in \mathbf{V}(I_1)$ satisfy the hypotheses of Theorem 4. To get the desired $c_1 \in \mathbb{C}$, the proof of Theorem 4 given in the text uses a polynomial $u(x_1)$ found by nonconstructive means. But now that we know the theorem is true, we can give a constructive method for finding c_1 by considering the polynomial

$$v(x_1) = \mathrm{GCD}(f_1(x_1, \mathbf{c}), \ldots, f_s(x_1, \mathbf{c})).$$

(a) Show that $v(x_1)$ is nonconstant and that every root c_1 of $v(x_1)$ satisfies the conclusion of the Theorem 4. Hint: Show that $u(x_1)$ divides $v(x_1)$.

(b) Show that $v(x_1)$ and $u(x_1)$ have the same roots. Hint: Express $u(x_1)$ in terms of the $f_i(x, \mathbf{c})$.

13. Show that the Extension Theorem holds over any algebraically closed field. Hint: You will need to see exactly where the proof of Theorem 4 uses the complex numbers $\mathbb{C}$.