Chapter 2

Solving Polynomial Equations

In this chapter we will discuss several approaches to solving systems of polynomial equations. First, we will discuss a straightforward attack based on the elimination properties of lexicographic Gröbner bases. Combining elimination with numerical root-finding for one-variable polynomials we get a conceptually simple method that generalizes the usual techniques used to solve systems of linear equations. However, there are potentially severe difficulties when this approach is implemented on a computer using finite-precision arithmetic. To circumvent these problems, we will develop some additional algebraic tools for root-finding based on the algebraic structure of the quotient rings $k[x_1, \ldots, x_n]/I$. Using these tools, we will present alternative numerical methods for approximating solutions of polynomial systems and consider methods for real root-counting and root-isolation. In Chapters 3, 4 and 7, we will also discuss polynomial equation solving. Specifically, Chapter 3 will use resultants to solve polynomial equations, and Chapter 4 will show how to assign a well-behaved multiplicity to each solution of a system. Chapter 7 will consider other numerical techniques (homotopy continuation methods) based on bounds for the total number of solutions of a system, counting multiplicities.

§1 Solving Polynomial Systems by Elimination

The main tools we need are the Elimination and Extension Theorems. For the convenience of the reader, we recall the key ideas:

- (Elimination Ideals) If I is an ideal in $k[x_1, \ldots, x_n]$, then the *ℓth elimination ideal* is

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

  Intuitively, if $I = \langle f_1, \ldots, f_s \rangle$, then the elements of I_ℓ are the linear combinations of the $f_1, \ldots, f_s$, with polynomial coefficients, that eliminate $x_1, \ldots, x_\ell$ from the equations $f_1 = \cdots = f_s = 0$.

- (The Elimination Theorem) If G is a Gröbner basis for I with respect to the *lex* order ($x_1 > x_2 > \cdots > x_n$) (or any order where monomials involving at least one of $x_1, \ldots, x_\ell$ are greater than all monomials involving only the remaining variables), then
$$G_\ell = G \cap k[x_{\ell+1}, \ldots, x_n]$$
is a Gröbner basis of the ℓth elimination ideal I_ℓ.
- (Partial Solutions) A point $(a_{\ell+1}, \ldots, a_n) \in \mathbf{V}(I_\ell) \subset k^{n-\ell}$ is called a *partial solution*. Any solution $(a_1, \ldots, a_n) \in \mathbf{V}(I) \subset k^n$ truncates to a partial solution, but the converse may fail—not all partial solutions extend to solutions. This is where the Extension Theorem comes in. To prepare for the statement, note that each f in $I_{\ell-1}$ can be written as a polynomial in x_ℓ, whose coefficients are polynomials in $x_{\ell+1}, \ldots, x_n$:
$$f = c_q(x_{\ell+1}, \ldots, x_n)x_\ell^q + \cdots + c_0(x_{\ell+1}, \ldots, x_n).$$
We call c_q the leading coefficient polynomial of f if x_ℓ^q is the highest power of x_ℓ appearing in f.
- (The Extension Theorem) If k is algebraically closed (e.g., $k = \mathbb{C}$), then a partial solution $(a_{\ell+1}, \ldots, a_n)$ in $\mathbf{V}(I_\ell)$ extends to $(a_\ell, a_{\ell+1}, \ldots, a_n)$ in $\mathbf{V}(I_{\ell-1})$ provided that the leading coefficient polynomials of the elements of a *lex* Gröbner basis for $I_{\ell-1}$ do not all vanish at $(a_{\ell+1}, \ldots, a_n)$.

For the proofs of these results and a discussion of their geometric meaning, see Chapter 3 of [CLO]. Also, the Elimination Theorem is discussed in §6.2 of [BW] and §2.3 of [AL], and [AL] discusses the geometry of elimination in §2.5.

The Elimination Theorem shows that a *lex* Gröbner basis G successively eliminates more and more variables. This gives the following strategy for finding all solutions of the system: start with the polynomials in G with the fewest variables, solve them, and then try to extend these partial solutions to solutions of the whole system, applying the Extension Theorem one variable at a time.

As the following example shows, this works especially nicely when $\mathbf{V}(I)$ is finite. Consider the system of equations

$$\begin{aligned} x^2 + y^2 + z^2 &= 4 \\ x^2 + 2y^2 &= 5 \\ xz &= 1 \end{aligned} \tag{1.1}$$

from Exercise 4 of Chapter 3, §1 of [CLO]. To solve these equations, we first compute a *lex* Gröbner basis for the ideal they generate using Maple:

```
with(Groebner):
PList := [x^2+y^2+z^2-4, x^2+2*y^2-5, x*z-1];
G := gbasis(PList,plex(x,y,z));
```

This gives output

$$G := [1 + 2z^4 - 3z^2, y^2 - z^2 - 1, x + 2z^3 - 3z].$$

From the Gröbner basis it follows that the set of solutions of this system in $\mathbb{C}^3$ is finite (why?). To find all the solutions, note that the last polynomial depends only on z (it is a generator of the second elimination ideal $I_2 = I \cap \mathbb{C}[z]$) and factors nicely in $\mathbb{Q}[z]$. To see this, we may use

```
factor(2*z^4 - 3*z^2 + 1);
```

which generates the output

$$(z - 1)(z + 1)(2z^2 - 1).$$

Thus we have four possible z values to consider:

$$z = \pm 1, \pm 1/\sqrt{2}.$$

By the Elimination Theorem, the first elimination ideal $I_1 = I \cap \mathbb{C}[y, z]$ is generated by

$$\begin{aligned} &y^2 - z^2 - 1 \\ &2z^4 - 3z^2 + 1. \end{aligned}$$

Since the coefficient of y^2 in the first polynomial is a nonzero constant, *every* partial solution in $\mathbf{V}(I_2)$ extends to a solution in $\mathbf{V}(I_1)$. There are eight such points in all. To find them, we substitute a root of the last equation for z and solve the resulting equation for y. For instance,

```
subs(z=1,G);
```

will produce:

$$[-1 + x, y^2 - 2, 0],$$

so in particular, $y = \pm\sqrt{2}$. In addition, since the coefficient of x in the first polynomial in the Gröbner basis is a nonzero constant, we can extend each partial solution in $\mathbf{V}(I_1)$ (uniquely) to a point of $\mathbf{V}(I)$. For this value of z, we have $x = 1$.

Exercise 1. Carry out the same process for the other values of z as well. You should find that the eight points

$$(1, \pm\sqrt{2}, 1),\ (-1, \pm\sqrt{2}, -1),\ (\sqrt{2}, \pm\sqrt{6}/2, 1/\sqrt{2}),\ (-\sqrt{2}, \pm\sqrt{6}/2, -1/\sqrt{2})$$

form the set of solutions.

The system in (1.1) is relatively simple because the coordinates of the solutions can all be expressed in terms of square roots of rational numbers. Unfortunately, general systems of polynomial equations are rarely this nice. For instance it is known that there are *no general formulas* involving only the field operations in k and extraction of roots (i.e., radicals) for solving single variable polynomial equations of degree 5 and higher. This is a famous result of Ruffini, Abel, and Galois (see [Her]). Thus, if elimination leads to a one-variable equation of degree 5 or higher, then we may not be able to give radical formulas for the roots of that polynomial.

We take the system of equations given in (1.1) and change the first term in the first polynomial from x^2 to x^5. Then executing

```
PList2 := [x^5+y^2+z^2-4, x^2+2*y^2-5, x*z-1];
G2 := gbasis(PList2,plex(x,y,z));
```

produces the following *lex* Gröbner basis:

$$(1.2)\qquad [2+2z^7-3z^5-z^3, 4y^2-2z^5+3z^3+z-10, 2x+2z^6-3z^4-z^2].$$

In this case, the command

```
factor(2*z^7 - 3*z^5 - z^3 + 2);
```

gives the factorization

$$2z^7-3z^5-z^3+2=(z-1)(2z^6+2z^5-z^4-z^3-2z^2-2z-2),$$

and the second factor is irreducible in $\mathbb{Q}[z]$. In a situation like this, to go farther in equation solving, we need to decide what kind of answer is required.

If we want a purely algebraic, "structural" description of the solutions, then Maple can represent solutions of systems like this via the `solve` command. Let's see what this looks like. Entering

```
solve(convert(G2,set),{x,y,z});
```

you should generate the following output:

$$\begin{aligned}
&\{\{y=\mathrm{RootOf}(\_Z^2-2, label=\_L4), x=1, z=1\},\\
&\quad\{y=1/2\mathrm{RootOf}(\_Z^2\\
&\qquad -2\mathrm{RootOf}(2\_Z^6+2\_Z^5-\_Z^4-\_Z^3-2\_Z^2-2\_Z-2)^5\\
&\qquad +3\mathrm{RootOf}(2\_Z^6+2\_Z^5-\_Z^4-\_Z^3-2\_Z^2-2\_Z-2)^3\\
&\qquad +\mathrm{RootOf}(2\_Z^6+2\_Z^5-\_Z^4-\_Z^3-2\_Z^2-2\_Z-2)\\
&\qquad -10, label=\_L1),\\
&\quad x=\mathrm{RootOf}(2\_Z^6+2\_Z^5-\_Z^4-\_Z^3-2\_Z^2-2\_Z-2)^4\\
&\qquad -1/2\mathrm{RootOf}(2\_Z^6+2\_Z^5-\_Z^4-\_Z^3-2\_Z^2-2\_Z-2)^2-1\\
&\qquad +\mathrm{RootOf}(2\_Z^6+2\_Z^5-\_Z^4-\_Z^3-2\_Z^2-2\_Z-2)^5\\
&\qquad -1/2\mathrm{RootOf}(2\_Z^6+2\_Z^5-\_Z^4-\_Z^3-2\_Z^2-2\_Z-2)^3\\
&\qquad -\mathrm{RootOf}(2\_Z^6+2\_Z^5-\_Z^4-\_Z^3-2\_Z^2-2\_Z-2),\\
&\quad z=\mathrm{RootOf}(2\_Z^6+2\_Z^5-\_Z^4-\_Z^3-2\_Z^2-2\_Z-2)\}\}
\end{aligned}$$

Here $\mathrm{RootOf}(2\_Z^6+2\_Z^5-\_Z^4-\_Z^3-2\_Z^2-2\_Z-2)$ stands for any one root of the polynomial equation $2\_Z^6+2\_Z^5-\_Z^4-\_Z^3-2\_Z^2-2\_Z-2=0$. Similarly, the other RootOf expressions stand for any solution of the corresponding equation in the dummy variable $\_Z$.

Exercise 2. Verify that the expressions above are obtained if we solve for z from the Gröbner basis G_2 and then use the Extension Theorem. How many solutions are there of this system in $\mathbb{C}^3$?

On the other hand, in many practical situations where equations must be solved, knowing a *numerical approximation* to a real or complex solution is often more useful, and perfectly acceptable provided the results are sufficiently accurate. In our particular case, one possible approach would be to use a numerical root-finding method to find approximate solutions of the one-variable equation

$$2z^6+2z^5-z^4-z^3-2z^2-2z-2=0, \tag{1.3}$$

and then proceed as before using the Extension Theorem, except that we now use floating point arithmetic in all calculations. In some examples, numerical methods will also be needed to solve for the other variables as we extend.

One well-known numerical method for solving one-variable polynomial equations in $\mathbb{R}$ or $\mathbb{C}$ is the *Newton-Raphson method* or, more simply but less accurately, *Newton's method.* This method may also be used for equations involving functions other than polynomials, although we will not discuss those here. For motivation and a discussion of the theory behind the method, see [BuF] or [Act].

The Newton-Raphson method works as follows. Choosing some initial approximation z_0 to a root of $p(z)=0$, we construct a sequence of numbers by the rule

$$z_{k+1}=z_k-\frac{p(z_k)}{p'(z_k)} \qquad \text{for } k=0,1,2,\ldots,$$

where $p'(z)$ is the usual derivative of p from calculus. In *most* situations, the sequence z_k will converge rapidly to a solution $\overline{z}$ of $p(z)=0$, that is, $\overline{z}=\lim_{k\to\infty} z_k$ will be a root. Stopping this procedure after a finite number of steps (as we must!), we obtain an approximation to $\overline{z}$. For example, we might stop when z_{k+1} and z_k agree to some desired accuracy, or when a maximum allowed number of terms of the sequence have been computed. See [BuF], [Act], or the comments at the end of this section for additional information on the performance of this technique. When trying to find *all* roots of a polynomial, the trickiest part of the Newton-Raphson method is making appropriate choices of z_0. It is easy to find the same root repeatedly and to miss other ones if you don't know where to look!

Fortunately, there are elementary bounds on the absolute values of the roots (real or complex) of a polynomial $p(z)$. Here is one of the simpler bounds.

Exercise 3. Show that if $p(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_0$ is a monic polynomial with complex coefficients, then all roots $\overline{z}$ of p satisfy $|\overline{z}| \leq B$, where

$$B = \max\{1, |a_{n-1}| + \cdots + |a_1| + |a_0|\}.$$

Hint: The triangle inequality implies that $|a + b| \geq |a| - |b|$.

See Exercise 10 below for another better bound on the roots. Given any bound of this sort, we can limit our attention to z_0 in this region of the complex plane to search for roots of the polynomial.

Instead of discussing searching strategies for finding roots, we will use a built-in Maple function to approximate the roots of the system from (1.2). The Maple function `fsolve` finds numerical approximations to all real (or complex) roots of a polynomial by a combination of root location and numerical techniques like Newton-Raphson. For instance, the command

```
fsolve(2*z^6+2*z^5-z^4-z^3-2*z^2-2*z-2);
```

will compute approximate values for the *real* roots of our polynomial (1.3). The output should be:

$$-1.395052015, \qquad 1.204042437.$$

(Note: In Maple, 10 digits are carried by default in decimal calculations; more digits can be used by changing the value of the Maple system variable `Digits`. Also, the actual digits in your output may vary slightly if you carry out this computation using another computer algebra system.) To get approximate values for the complex roots as well, try:

```
fsolve(2*z^6+2*z^5-z^4-z^3-2*z^2-2*z-2,complex);
```

We illustrate the Extension Step in this case using the approximate value

$$z = 1.204042437.$$

We substitute this value into the Gröbner basis polynomials using

```
subs(z=1.204042437,G2);
```

and obtain

$$[2x - 1.661071025, -8.620421528 + 4y^2, -.2 * 10^{-8}].$$

Note that the value of the last polynomial was *not exactly zero* at our approximate value of z. Nevertheless, as in Exercise 1, we can extend this approximate partial solution to two approximate solutions of the system:

$$(x, y, z) = (.8305355125, \pm 1.468027718, 1.204042437).$$

Checking one of these by substituting into the equations from (1.2), using

```
subs(z=1.204042437,y=1.468027718,x=.8305355125, G2);
```

we find

$$[0, -.4 * 10^{-8}, -.2 * 10^{-8}],$$

so we have a reasonably good approximate solution, in the sense that our computed solution gives values very close to zero in the polynomials of the system.

Exercise 4. Find approximate values for all other real solutions of this system by the same method.

In considering what we did here, one potential pitfall of this approach should be apparent. Namely, since our solutions of the one-variable equation are only approximate, when we substitute and try to extend, the remaining polynomials to be solved for x and y are themselves only approximate. Once we substitute approximate values for one of the variables, we are in effect solving a system of equations that is *different from the one we started with*, and there is little guarantee that the solutions of this new system are close to the solutions of the original one. Accumulated errors after several approximation and extension steps can build up quite rapidly in systems in larger numbers of variables, and the effect can be particularly severe if equations of high degree are present.

To illustrate how bad things can get, we consider a famous cautionary example due to Wilkinson, which shows how much the roots of a polynomial can be changed by very small changes in the coefficients.

Wilkinson's example involves the following polynomial of degree 20:

$$p(x) = (x + 1)(x + 2) \cdots (x + 20) = x^{20} + 210x^{19} + \cdots + 20!.$$

The roots are the 20 integers $x = -1, -2, \ldots, -20$. Suppose now that we "perturb" just the coefficient of x^{19}, adding a very small number. We carry 20 decimal digits in all calculations. First we construct $p(x)$ itself:

```
Digits := 20:
p := 1:
for k to 20 do p := p*(x+k) end do:
```

Printing `expand(p)` out at this point will show a polynomial with some large coefficients indeed! But the polynomial we want is actually this:

```
q := expand(p + .000000001*x^19):
fsolve(q,x,complex);
```

The approximate roots of $q = p + .000000001\, x^{19}$ (truncated for simplicity) are:

$$\begin{aligned}
&-20.03899, -18.66983 - .35064\ I,\ \ -18.66983 + .35064\ I,\\
&-16.57173 - .88331\ I,\ \ -16.57173 + .88331\ I,\\
&-14.37367 - .77316\ I,\ \ -14.37367 + .77316\ I,\\
&-12.38349 - .10866\ I,\ \ -12.38349 + .10866\ I,\\
&-10.95660,\ \ -10.00771,\ \ -8.99916,\ \ -8.00005,\\
&-6.999997,\ \ -6.000000,\ \ -4.99999,\ \ -4.00000,\\
&-2.999999,\ \ -2.000000,\ \ -1.00000.
\end{aligned}$$

Instead of 20 real roots, the new polynomial has 12 real roots and 4 complex conjugate pairs of roots. Note that the imaginary parts are not even especially small!

While this example is admittedly pathological, it indicates that we should use care in finding roots of polynomials whose coefficients are only approximately determined. (The reason for the surprisingly bad behavior of this p is essentially the equal spacing of the roots! We refer the interested reader to Wilkinson's paper [Wil] for a full discussion.)

Along the same lines, even if nothing this spectacularly bad happens, when we take the approximate roots of a one-variable polynomial and try to extend to solutions of a system, the results of a numerical calculation can still be unreliable. Here is a simple example illustrating another situation that causes special problems.

Exercise 5. Verify that if $x > y$, then

$$G = [x^2 + 2x + 3 + y^5 - y, y^6 - y^2 + 2y]$$

is a *lex* Gröbner basis for the ideal that G generates in $\mathbb{R}[x, y]$.

We want to find all *real* points $(x, y) \in \mathbf{V}(G)$. Begin with the equation

$$y^6 - y^2 + 2y = 0,$$

which has exactly two real roots. One is $y = 0$, and the second is in the interval $[-2, -1]$ because the polynomial changes sign on that interval. Hence there must be a root there by the Intermediate Value Theorem from calculus. Using `fsolve` to find an approximate value, we find the nonzero root is

$$-1.267168305 \tag{1.4}$$

to 10 decimal digits. Substituting this approximate value for y into G yields

$$[x^2 + 2x + .999999995, .7 * 10^{-8}].$$

Then use

```
fsolve(x^2 + 2*x + .999999995);
```

to obtain

$$-1.000070711, \ -.9999292893.$$

Clearly these are both close to $x = -1$, but they are different. Taken uncritically, this would seem to indicate two distinct real values of x when y is given by (1.4).

Now, suppose we used an approximate value for y with *fewer* decimal digits, say $y \doteq -1.2671683$. Substituting this value for y gives us the quadratic

$$x^2 + 2x + 1.000000054.$$

This polynomial has no real roots at all. Indeed, using the `complex` option in `fsolve`, we obtain two complex values for x:

$$-1. - .0002323790008\, I, \qquad -1. + .0002323790008\, I.$$

To see what is really happening, note that the nonzero real root of $y^6 - y^2 + 2y = 0$ satisfies $y^5 - y + 2 = 0$. When the exact root is substituted into G, we get

$$[x^2 + 2x + 1, 0]$$

and the resulting equation has a double root $x = -1$.

The conclusion to be drawn from this example is that equations with double roots, such as the *exact* equation

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

we got above, are *especially* vulnerable to the errors introduced by numerical root-finding. It can be very difficult to tell the difference between a pair of real roots that are close, a real double root, and a pair of complex conjugate roots.

From these examples, it should be clear that finding solutions of polynomial systems is a delicate task in general, especially if we ask for information about how many real solutions there are. For this reason, numerical methods, for all their undeniable usefulness, are not the whole story. And they should never be applied blindly. The more information we have about the structure of the set of solutions of a polynomial system, the better a chance we have to determine those solutions accurately. For this reason, in §2 and §3 we will go to the algebraic setting of the quotient ring $k[x_1, \ldots, x_n]/I$ to obtain some additional tools for this problem. We will apply those tools in §4 and §5 to give better methods for finding solutions.

For completeness, we conclude with a few additional words about the numerical methods for equation solving that we have used. First, if $\overline{z}$ is a multiple root of $p(z) = 0$, then the convergence of the Newton-Raphson sequence z_k can be quite slow, and a large number of steps and high precision may be required to get really close to a root (though we give a method for avoiding this difficulty in Exercise 8). Second, there are some choices of z_0 where the sequence z_k will *fail to converge* to a root of $p(z)$. See Exercise 9 below for some simple examples. Finally, the location of $\overline{z}$ in relation to z_0 can be somewhat unpredictable. There could be other roots lying closer to z_0. These last two problems are related to the fractal pictures associated to the Newton-Raphson method over $\mathbb{C}$—see, for example, [PR]. We should also mention that there are multivariable versions of Newton-Raphson for systems of equations and other iterative methods that do not depend on elimination. These have been much studied in numerical analysis. For more details on these and other numerical root-finding methods, see [BuF] and [Act]. Also, we will discuss homotopy continuation methods in Chapter 7, §5 of this book.

Additional Exercises for §1

Exercise 6. Use elimination to solve the system

$$\begin{aligned}0 &= x^2 + 2y^2 - y - 2z\\ 0 &= x^2 - 8y^2 + 10z - 1\\ 0 &= x^2 - 7yz.\end{aligned}$$

How many solutions are there in $\mathbb{R}^3$; how many are there in $\mathbb{C}^3$?

Exercise 7. Use elimination to solve the system

$$\begin{aligned}0 &= x^2 + y^2 + z^2 - 2x\\ 0 &= x^3 - yz - x\\ 0 &= x - y + 2z.\end{aligned}$$

How many solutions are there in $\mathbb{R}^3$; how many are there in $\mathbb{C}^3$?

Exercise 8. In this exercise we will study exactly why the performance of the Newton-Raphson method is poor for multiple roots, and suggest a remedy. Newton-Raphson iteration for any equation $p(z) = 0$ is an example of *fixed point iteration*, in which a starting value z_0 is chosen and a sequence

$$z_{k+1} = g(z_k) \qquad \text{for } k = 0, 1, 2, \ldots \tag{1.5}$$

is constructed by iteration of a fixed function $g(z)$. For Newton-Raphson iteration, the function $g(z)$ is $g(z) = N_p(z) = z - p(z)/p'(z)$. If the sequence produced by (1.5) converges to some limit $\overline{z}$, then $\overline{z}$ is a *fixed point* of g (that is, a solution of $g(z) = z$). It is a standard result from analysis (a special case of the Contraction Mapping Theorem) that iteration as in

(1.5) will converge to a fixed point $\overline{z}$ of g provided that $|g'(\overline{z})| < 1$, and z_0 is chosen sufficiently close to $\overline{z}$. Moreover, the smaller $|g'(\overline{z})|$ is, the faster convergence will be. The case $g'(\overline{z}) = 0$ is especially favorable.

a. Show that each simple root of the polynomial equation $p(z) = 0$ is a *fixed point* of the rational function $N_p(z) = z - p(z)/p'(z)$.
b. Show that multiple roots of $p(z) = 0$ are *removable singularities* of $N_p(z)$ (that is, $|N_p(z)|$ is bounded in a neighborhood of each multiple root). How should N_p be defined at a multiple root of $p(z) = 0$ to make N_p continuous at those points?
c. Show that $N_p'(\overline{z}) = 0$ if $\overline{z}$ is a *simple* root of $p(z) = 0$ (that is, if $p(\overline{z}) = 0$, but $p'(\overline{z}) \neq 0$).
d. On the other hand, show that if $\overline{z}$ is a root of multiplicity k of $p(z)$ (that is, if $p(\overline{z}) = p'(\overline{z}) = \cdots = p^{(k-1)}(\overline{z}) = 0$ but $p^{(k)}(\overline{z}) \neq 0$), then
$$\lim_{z\to\overline{z}} N_p'(z) = 1 - \frac{1}{k}.$$
Thus Newton-Raphson iteration converges much faster to a simple root of $p(z) = 0$ than it does to a multiple root, and the larger the multiplicity, the slower the convergence.
e. Show that replacing $p(z)$ by
$$p_{red}(z) = \frac{p(z)}{\mathrm{GCD}(p(z), p'(z))}$$
(see [CLO], Chapter 1, §5, Exercises 14 and 15) eliminates this difficulty, in the sense that the roots of $p_{red}(z) = 0$ are all simple roots.

Exercise 9. There are cases when the Newton-Raphson method fails to find a root of a polynomial for *lots* of starting points z_0.

a. What happens if the Newton-Raphson method is applied to solve the equation $z^2 + 1 = 0$ starting from a *real* z_0? What happens if you take z_0 with nonzero imaginary parts? Note: It can be shown that Newton-Raphson iteration for the equation $p(z) = 0$ is *chaotic* if z_0 is chosen in the *Julia set* of the rational function $N_p(z) = z - p(z)/p'(z)$ (see [PR]), and exact arithmetic is employed.
b. Let $p(z) = z^4 - z^2 - 11/36$ and, as above, let $N_p(z) = z - p(z)/p'(z)$. Show that $\pm 1/\sqrt{6}$ satisfies $N_p(1/\sqrt{6}) = -1/\sqrt{6}$, $N_p(-1/\sqrt{6}) = 1/\sqrt{6}$, and $N_p'(1/\sqrt{6}) = 0$. In the language of dynamical systems, $\pm 1/\sqrt{6}$ is a *superattracting* 2-*cycle* for $N_p(z)$. One consequence is that for *any* z_0 close to $\pm 1/\sqrt{6}$, the Newton-Raphson method will *not* locate a root of p. This example is taken from Chapter 13 of [Dev].

Exercise 10. This exercise improves the bound on roots of a polynomial given in Exercise 3. Let $p(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$ be a monic polynomial in $\mathbb{C}[z]$. Show that all roots $\overline{z}$ of p satisfy $|\overline{z}| \leq B$, where
$$B = 1 + \max\{|a_{n-1}|, \ldots, |a_1|, |a_0|\}.$$

This upper bound can be much smaller than the one given in Exercise 3. Hint: Use the Hint from Exercise 3, and consider the evaluation of $p(z)$ by nested multiplication:

$$p(z) = (\cdots((z + a_{n-1})z + a_{n-2})z + \cdots + a_1)z + a_0.$$

§2 Finite-Dimensional Algebras

This section will explore the "remainder arithmetic" associated to a Gröbner basis $G = \{g_1, \ldots, g_t\}$ of an ideal $I \subset k[x_1, \ldots, x_n]$. Recall from Chapter 1 that if we divide $f \in k[x_1, \ldots, x_n]$ by G, the division algorithm yields an expression

$$f = h_1 g_1 + \cdots + h_t g_t + \overline{f}^G, \tag{2.1}$$

where the remainder $\overline{f}^G$ is a linear combination of the monomials $x^\alpha \notin \langle \mathrm{LT}(I)\rangle$. Furthermore, since G is a Gröbner basis, we know that $f \in I$ if and only if $\overline{f}^G = 0$, and the remainder is uniquely determined for all f. This implies

$$\overline{f}^G = \overline{g}^G \iff f - g \in I. \tag{2.2}$$

Since polynomials can be added and multiplied, given $f, g \in k[x_1, \ldots, x_n]$ it is natural to ask how the remainders of $f + g$ and fg can be computed if we know the remainders of f, g themselves. The following observations show how this can be done.

- The sum of two remainders is again a remainder, and in fact one can easily show that $\overline{f}^G + \overline{g}^G = \overline{f+g}^G$.
- On the other hand, the product of remainders need not be a remainder. But it is also easy to see that $\overline{\overline{f}^G \cdot \overline{g}^G}^G = \overline{fg}^G$, and $\overline{\overline{f}^G \cdot \overline{g}^G}^G$ is a remainder.

We can also interpret these observations as saying that the set of remainders on division by G has naturally defined addition and multiplication operations which produce remainders as their results.

This "remainder arithmetic" is closely related to the quotient ring $k[x_1, \ldots, x_n]/I$. We will assume the reader is familiar with quotient rings, as described in Chapter 5 of [CLO] or in a course on abstract algebra. Recall how this works: given $f \in k[x_1, \ldots, x_n]$, we have the *coset*

$$[f] = f + I = \{f + h : h \in I\},$$

and the crucial property of cosets is

$$[f] = [g] \iff f - g \in I. \tag{2.3}$$

The quotient ring $k[x_1, \ldots, x_n]/I$ consists of all cosets $[f]$ for $f \in k[x_1, \ldots, x_n]$.

From (2.1), we see that $\overline{f}^G \in [f]$, and then (2.2) and (2.3) show that we have a one-to-one correspondence

$$\begin{aligned} \text{remainders} &\longleftrightarrow \text{cosets} \\ \overline{f}^G &\longleftrightarrow [f]. \end{aligned}$$

Thus we can think of the remainder $\overline{f}^G$ as a standard representative of its coset $[f] \in k[x_1, \ldots, x_n]/I$. Furthermore, it follows easily that remainder arithmetic is *exactly* the arithmetic in $k[x_1, \ldots, x_n]/I$. That is, under the above correspondence we have

$$\begin{aligned} \overline{f}^G + \overline{g}^G &\longleftrightarrow [f] + [g] \\ \overline{\overline{f}^G \cdot \overline{g}^G}^G &\longleftrightarrow [f] \cdot [g]. \end{aligned}$$

Since we can add elements of $k[x_1, \ldots, x_n]/I$ and multiply by constants (the cosets $[c]$ for $c \in k$), $k[x_1, \ldots, x_n]/I$ also has the structure of a vector space over the field k. A ring that is also a vector space in this fashion is called an *algebra*. The algebra $k[x_1, \ldots, x_n]/I$ will be denoted by A throughout the rest of this section, which will focus on its vector space structure.

An important observation is that remainders are the linear combinations of the monomials $x^\alpha \notin \langle \textsc{lt}(I) \rangle$ in this vector space structure. (Strictly speaking, we should use cosets, but in much of this section we will identify a remainder with its coset in A.) Since this set of monomials is linearly independent in A (why?), it can be regarded as a basis of A. In other words, the monomials

$$B = \{x^\alpha : x^\alpha \notin \langle \textsc{lt}(I) \rangle\}$$

form a basis of A (more precisely, their cosets are a basis). We will refer to elements of B as *basis monomials*. In the literature, basis monomials are often called *standard monomials*.

The following example illustrates how to compute in A using basis monomials. Let

$$(2.4) \qquad G = \{x^2 + 3xy/2 + y^2/2 - 3x/2 - 3y/2, xy^2 - x, y^3 - y\}.$$

Using the *grevlex* order with $x > y$, it is easy to verify that G is a Gröbner basis for the ideal $I = \langle G \rangle \subset \mathbb{C}[x, y]$ generated by G. By examining the leading monomials of G, we see that $\langle \textsc{lt}(I) \rangle = \langle x^2, xy^2, y^3 \rangle$. The only monomials not lying in this ideal are those in

$$B = \{1, x, y, xy, y^2\}$$

so that by the above observation, these five monomials form a vector space basis for $A = \mathbb{C}[x, y]/I$ over $\mathbb{C}$.

We now turn to the structure of the quotient ring A. The addition operation in A can be viewed as an ordinary vector sum operation once we express elements of A in terms of the basis B in (2.4). Hence we will consider the addition operation to be completely understood.

Perhaps the most natural way to describe the multiplication operation in A is to give a table of the remainders of all products of pairs of elements from the basis B. Since multiplication in A distributes over addition, this information will suffice to determine the products of all pairs of elements of A.

For example, the remainder of the product $x \cdot xy$ may be computed as follows using Maple. Using the Gröbner basis G, we compute

```
normalf(x^2*y,G,tdeg(x,y));
```

and obtain

$$\frac{3}{2}xy - \frac{3}{2}x + \frac{3}{2}y^2 - \frac{1}{2}y.$$

Exercise 1. By computing all such products, verify that the multiplication table for the elements of the basis B is:

(2.5)

$\cdot$	1	x	y	xy	y^2
1	1	x	y	xy	y^2
x	x	α	xy	β	x
y	y	xy	y^2	x	y
xy	xy	β	x	α	xy
y^2	y^2	x	y	xy	y^2

where

$$\begin{aligned}\alpha &= -3xy/2 - y^2/2 + 3x/2 + 3y/2\\ \beta &= 3xy/2 + 3y^2/2 - 3x/2 - y/2.\end{aligned}$$

This example was especially nice because A was finite-dimensional as a vector space over $\mathbb{C}$. In general, for any field $k \subset \mathbb{C}$, we have the following basic theorem which describes when $k[x_1, \ldots, x_n]/I$ is finite-dimensional.

- (Finiteness Theorem) Let $k \subset \mathbb{C}$ be a field, and let $I \subset k[x_1, \ldots, x_n]$ be an ideal. Then the following conditions are equivalent:
  a. The algebra $A = k[x_1, \ldots, x_n]/I$ is finite-dimensional over k.
  b. The variety $\mathbf{V}(I) \subset \mathbb{C}^n$ is a finite set.
  c. If G is a Gröbner basis for I, then for each i, $1 \leq i \leq n$, there is an $m_i \geq 0$ such that $x_i^{m_i} = \textsc{lt}(g)$ for some $g \in G$.

For a proof of this result, see Theorem 6 of Chapter 5, §3 of [CLO], Theorem 2.2.7 of [AL], or Theorem 6.54 of [BW]. An ideal satisfying any of the above conditions is said to be *zero-dimensional*. Thus

$$A \text{ is a finite-dimensional algebra} \Longleftrightarrow I \text{ is a zero-dimensional ideal.}$$

A nice consequence of this theorem is that I is zero-dimensional if and only if there is a nonzero polynomial in $I \cap k[x_i]$ for each $i = 1, \ldots, n$. To see why this is true, first suppose that I is zero-dimensional, and let G be a reduced Gröbner basis for any *lex* order with x_i as the "last" variable (i.e., $x_j > x_i$ for $j \neq i$). By item c above, there is some $g \in G$ with $\text{LT}(g) = x_i^{m_i}$. Since we're using a *lex* order with x_i last, this implies $g \in k[x_i]$ and hence g is the desired nonzero polynomial. Note that g generates $I \cap k[x_i]$ by the Elimination Theorem.

Going the other way, suppose $I \cap k[x_i]$ is nonzero for each i, and let m_i be the degree of the unique monic generator of $I \cap k[x_i]$ (remember that $k[x_i]$ is a principal ideal domain—see Corollary 4 of Chapter 1, §5 of [CLO]). Then $x_i^{m_i} \in \langle \text{LT}(I) \rangle$ for any monomial order, so that all monomials not in $\langle \text{LT}(I) \rangle$ will contain x_i to a power strictly less than m_i. In other words, the exponents α of the monomials $x^\alpha \notin \langle \text{LT}(I) \rangle$ will all lie in the "rectangular box"

$$R = \{\alpha \in \mathbb{Z}^n_{\geq 0} : \text{ for each } i,\ 0 \leq \alpha_i \leq m_i - 1\}.$$

This is a finite set of monomials, which proves that A is finite-dimensional over k.

Given a zero-dimensional ideal I, it is now easy to describe an algorithm for finding the set B of all monomials not in $\langle \text{LT}(I) \rangle$. Namely, no matter what monomial order we are using, the exponents of the monomials in B will lie in the box R described above. For each $\alpha \in R$, we know that $x^\alpha \notin \langle \text{LT}(I) \rangle$ if and only if $\overline{x^\alpha}^G = x^\alpha$. Thus we can list the $\alpha \in R$ in some systematic way and compute $\overline{x^\alpha}^G$ for each one. A vector space basis of A is given by the set of monomials

$$B = \{x^\alpha : \alpha \in R \text{ and } \overline{x^\alpha}^G = x^\alpha\}.$$

See Exercise 13 below for a Maple procedure implementing this method.

The vector space structure on $A = k[x_1, \ldots, x_n]/I$ for a zero-dimensional ideal I can be used in several important ways. To begin, let us consider the problem of finding the monic generators of the elimination ideals $I \cap k[x_i]$. As indicated above, we could find these polynomials by computing several different *lex* Gröbner bases, reordering the variables each time to place x_i last. This is an extremely inefficient method, however. Instead, let us consider the set of non-negative powers of $[x_i]$ in A:

$$S = \{1, [x_i], [x_i]^2, \ldots\}.$$

Since A is finite-dimensional as a vector space over the field k, S must be *linearly dependent* in A. Let m_i be the *smallest* positive integer for which $\{1, [x_i], [x_i]^2, \ldots, [x_i]^{m_i}\}$ is linearly dependent. Then there is a linear combination

$$\sum_{j=0}^{m_i} c_j [x_i]^j = [0]$$

in A in which the $c_j \in k$ are not all zero. In particular, $c_{m_i} \neq 0$ since m_i is minimal. By the definition of the quotient ring, this is equivalent to saying that

$$p_i(x_i) = \sum_{j=0}^{m_i} c_j x_i^j \in I. \tag{2.6}$$

Exercise 2. Verify that $p_i(x_i)$ as in (2.6) is a generator of the ideal $I \cap k[x_i]$, and develop an algorithm based on this fact to find the monic generator of $I \cap k[x_i]$, given any Gröbner basis G for a zero-dimensional ideal I as input.

The algorithm suggested in Exercise 2 often requires far less computational effort than a *lex* Gröbner basis calculation. Any ordering (e.g. *grevlex*) can be used to determine G, then only standard linear algebra (matrix operations) are needed to determine whether the set $\{1, [x_i], [x_i]^2, \ldots, [x_i]^m\}$ is linearly dependent. We note that the `univpoly` function from Maple's `Groebner` package is an implementation of this method.

We will next discuss how to find the *radical* of a zero-dimensional ideal (see Chapter 1 for the definition of radical). To motivate what we will do, recall from §1 how multiple roots of a polynomial can cause problems when trying to find roots numerically. When dealing with a one-variable polynomial p with coefficients lying in a subfield of $\mathbb{C}$, it is easy to see that the polynomial

$$p_{red} = \frac{p}{\mathrm{GCD}(p, p')}$$

has the same roots as p, but all with multiplicity one (for a proof of this, see Exercises 14 and 15 of Chapter 1, §5 of [CLO]). We call p_{red} the *square-free part* of p.

The radical $\sqrt{I}$ of an ideal I generalizes the idea of the square-free part of a polynomial. In fact, we have the following elementary exercise.

Exercise 3. If $p \in k[x]$ is a nonzero polynomial, show that $\sqrt{\langle p \rangle} = \langle p_{red} \rangle$.

Since $k[x]$ is a PID, this solves the problem of finding radicals for *all* ideals in $k[x]$. For a general ideal $I \subset k[x_1, \ldots, x_n]$, it is more difficult to find $\sqrt{I}$, though algorithms are known and have been implemented in *Macaulay 2*, REDUCE, and `Singular`. Fortunately, when I is zero-dimensional, computing the radical is much easier, as shown by the following proposition.

(2.7) Proposition. *Let $I \subset \mathbb{C}[x_1, \ldots, x_n]$ be a zero-dimensional ideal. For each $i = 1, \ldots, n$, let p_i be the unique monic generator of $I \cap \mathbb{C}[x_i]$, and let $p_{i,red}$ be the square-free part of p_i. Then*

$$\sqrt{I} = I + \langle p_{1,red}, \ldots, p_{n,red} \rangle.$$

Proof. Write $J = I + \langle p_{1,red}, \ldots, p_{n,red}\rangle$. We first prove that J is a radical ideal, i.e., that $J = \sqrt{J}$. For each i, using the fact that $\mathbb{C}$ is algebraically closed, we can factor each $p_{i,red}$ to obtain $p_{i,red} = (x_i - a_{i1})(x_i - a_{i2}) \cdots (x_i - a_{id_i})$, where the a_{ij} are distinct. Then

$$J = J + \langle p_{1,red}\rangle = \bigcap_j (J + \langle x_1 - a_{1j}\rangle),$$

where the first equality holds since $p_{1,red} \in J$ and the second follows from Exercise 9 below since $p_{1,red}$ has distinct roots. Now use $p_{2,red}$ to decompose each $J + \langle x_1 - a_{1j}\rangle$ in the same way. This gives

$$J = \bigcap_{j,k} (J + \langle x_1 - a_{1j}, x_2 - a_{2k}\rangle).$$

If we do this for all $i = 1, 2, \ldots, n$, we get the expression

$$J = \bigcap_{j_1,\ldots,j_n} (J + \langle x_1 - a_{1j_1}, \ldots, x_n - a_{nj_n}\rangle).$$

Since $\langle x_1 - a_{1j_1}, \ldots, x_n - a_{nj_n}\rangle$ is a maximal ideal, the ideal $J + \langle x_1 - a_{1j_1}, \ldots, x_n - a_{nj_n}\rangle$ is either $\langle x_1 - a_{1j_1}, \ldots, x_n - a_{nj_n}\rangle$ or the whole ring $\mathbb{C}[x_1, \ldots, x_n]$. It follows that J is a finite intersection of maximal ideals. Since a maximal ideal is radical and an intersection of radical ideals is radical, we conclude that J is a radical ideal.

Now we can prove that $J = \sqrt{I}$. The inclusion $I \subset J$ is built into the definition of J, and the inclusion $J \subset \sqrt{I}$ follows from the Strong Nullstellensatz, since the square-free parts of the p_i vanish at all the points of $\mathbf{V}(I)$. Hence we have

$$I \subset J \subset \sqrt{I}.$$

Taking radicals in this chain of inclusions shows that $\sqrt{J} = \sqrt{I}$. But J is radical, so $\sqrt{J} = J$ and we are done. □

A Maple procedure that implements an algorithm for the radical of a zero-dimensional ideal based on Proposition (2.7) is discussed in Exercise 16 below. It is perhaps worth noting that even though we have proved Proposition (2.7) using the properties of $\mathbb{C}$, the actual computation of the polynomials $p_{i,red}$ will involve only rational arithmetic when the input polynomials are in $\mathbb{Q}[x_1, \ldots, x_n]$.

For example, consider the ideal

$$(2.8) \qquad I = \langle y^4x + 3x^3 - y^4 - 3x^2, x^2y - 2x^2, 2y^4x - x^3 - 2y^4 + x^2\rangle$$

Exercise 4. Using Exercise 2 above, show that

$$I \cap \mathbb{Q}[x] = \langle x^3 - x^2\rangle$$

and

$$I \cap \mathbb{Q}[y] = \langle y^5 - 2y^4\rangle.$$

Writing $p_1(x) = x^3 - x^2$ and $p_2(y) = y^5 - 2y^4$, we can compute the square-free parts in Maple as follows. The command

```
p1red := simplify(p1/gcd(p1,diff(p1,x)));
```

will produce

$$p_{1,red}(x) = x(x-1).$$

Similarly,

$$p_{2,red}(y) = y(y-2).$$

Hence by Proposition (2.7), $\sqrt{I}$ is the ideal

$$\langle y^4x + 3x^3 - y^4 - 3x^2, x^2y - 2x^2, 2y^4x - x^3 - 2y^4 + x^2, x(x-1), y(y-2)\rangle.$$

We note that Proposition (2.7) yields a basis, but usually *not a Gröbner basis,* for $\sqrt{I}$.

Exercise 5. How do the dimensions of the vector spaces $\mathbb{C}[x,y]/I$ and $\mathbb{C}[x,y]/\sqrt{I}$ compare in this example? How could you determine the number of distinct points in $\mathbf{V}(I)$? (There are *two.*)

We will conclude this section with a very important result relating the dimension of A and the number of points in the variety $\mathbf{V}(I)$, or what is the same, the number of solutions of the equations $f_1 = \cdots = f_s = 0$ in $\mathbb{C}^n$. To prepare for this we will need the following lemma.

(2.9) Lemma. *Let $S = \{p_1, \ldots, p_m\}$ be a finite subset of $\mathbb{C}^n$. There exist polynomials $g_i \in \mathbb{C}[x_1, \ldots, x_n]$, $i = 1, \ldots, m$, such that*

$$g_i(p_j) = \begin{cases} 0 & \text{if } i \neq j, \text{ and} \\ 1 & \text{if } i = j. \end{cases}$$

For instance, if $p_i = (a_{i1}, \ldots, a_{in})$ and the first coordinates a_{i1} are *distinct*, then we can take

$$g_i = g_i(x_1) = \frac{\prod_{j\neq i}(x_1 - a_{j1})}{\prod_{j\neq i}(a_{i1} - a_{j1})}$$

as in the *Lagrange interpolation formula.* In any case, a collection of polynomials g_i with the desired properties can be found in a similar fashion. We leave the proof to the reader as Exercise 11 below. The following theorem ties all of the results of this section together, showing how the dimension of the algebra A for a zero-dimensional ideal gives a bound on the number of points in $\mathbf{V}(I)$, and also how radical ideals are special in this regard.

(2.10) Theorem. *Let I be a zero-dimensional ideal in $\mathbb{C}[x_1, \ldots, x_n]$, and let $A = \mathbb{C}[x_1, \ldots, x_n]/I$. Then $\dim_{\mathbb{C}}(A)$ is greater than or equal to the* *number of points in* $\mathbf{V}(I)$*. Moreover, equality occurs if and only if* I *is a radical ideal.*

Proof. Let I be a zero-dimensional ideal. By the Finiteness Theorem, $\mathbf{V}(I)$ is a finite set in $\mathbb{C}^n$, say $\mathbf{V}(I) = \{p_1, \ldots, p_m\}$. Consider the mapping

$$\begin{aligned} \varphi : \mathbb{C}[x_1, \ldots, x_n]/I &\longrightarrow \mathbb{C}^m \\ [f] &\mapsto (f(p_1), \ldots, f(p_m)) \end{aligned}$$

given by evaluating a coset at the points of $\mathbf{V}(I)$. In Exercise 12 below, you will show that φ is a well-defined linear map.

To prove the first statement in the theorem, it suffices to show that φ is onto. Let $g_1, \ldots, g_m$ be a collection of polynomials as in Lemma (2.9). Given an arbitrary $(\lambda_1, \ldots, \lambda_m) \in \mathbb{C}^m$, let $f = \sum_{i=1}^m \lambda_i g_i$. An easy computation shows that $\varphi([f]) = (\lambda_1, \ldots, \lambda_m)$. Thus φ is onto, and hence $\dim(A) \geq m$.

Next, suppose that I is radical. If $[f] \in \ker(\varphi)$, then $f(p_i) = 0$ for all i, so that by the Strong Nullstellensatz, $f \in \mathbf{I}(\mathbf{V}(I)) = \sqrt{I} = I$. Thus $[f] = [0]$, which shows that φ is one-to-one as well as onto. Then φ is an isomorphism, which proves that $\dim(A) = m$ if I is radical.

Conversely, if $\dim(A) = m$, then φ is an isomorphism since it is an onto linear map between vector spaces of the same dimension. Hence φ is one-to-one. We can use this to prove that I is radical as follows. Since the inclusion $I \subset \sqrt{I}$ always holds, it suffices to consider $f \in \sqrt{I} = \mathbf{I}(\mathbf{V}(I))$ and show that $f \in I$. If $f \in \sqrt{I}$, then $f(p_i) = 0$ for all i, which implies $\varphi([f]) = (0, \ldots, 0)$. Since φ is one-to-one, we conclude that $[f] = [0]$, or in other words that $f \in I$, as desired. □

In Chapter 4, we will see that in the case I is *not* radical, there are well-defined multiplicities at each point in $\mathbf{V}(I)$ so that the sum of the multiplicities equals $\dim(A)$.

Additional Exercises for §2

Exercise 6. Using the *grevlex* order, construct the monomial basis B for the quotient algebra $A = \mathbb{C}[x, y]/I$, where I is the ideal from (2.8) and construct the multiplication table for B in A.

Exercise 7. In this exercise, we will explain how the ideal $I = \langle x^2 + 3xy/2 + y^2/2 - 3x/2 - 3y/2, xy^2 - x, y^3 - y\rangle$ from (2.4) was constructed. The basic idea was to start from a finite set of points and construct a system of equations, rather than the reverse.

To begin, consider the maximal ideals

$$\begin{gathered} I_1 = \langle x, y\rangle, \qquad I_2 = \langle x-1, y-1\rangle, \\ I_3 = \langle x+1, y-1\rangle, \qquad I_4 = \langle x-1, y+1\rangle, \\ I_5 = \langle x-2, y+1\rangle \end{gathered}$$

in $\mathbb{C}[x,y]$. Each variety $\mathbf{V}(I_j)$ is a single point in $\mathbb{C}^2$, indeed in $\mathbb{Q}^2 \subset \mathbb{C}^2$. The union of the five points forms an affine variety V, and by the algebra-geometry dictionary from Chapter 1, $V = \mathbf{V}(I_1 \cap I_2 \cap \cdots \cap I_5)$.

An algorithm for intersecting ideals is described in Chapter 1. Use it to compute the intersection $I = I_1 \cap I_2 \cap \cdots \cap I_5$ and find the reduced Gröbner basis for I with respect to the *grevlex* order $(x > y)$. Your result should be the Gröbner basis given in (2.4).

Exercise 8.

a. Use the method of Proposition (2.7) to show that the ideal I from (2.4) is a radical ideal.
b. Give a non-computational proof of the statement from part a using the following observation. By the form of the generators of each of the ideals I_j in Exercise 7, $\mathbf{V}(I_j)$ is a single point and I_j is the ideal $\mathbf{I}(\mathbf{V}(I_j))$. As a result, $I_j = \sqrt{I_j}$ by the Strong Nullstellensatz. Then use the general fact about intersections of radical ideals from part a Exercise 9 from §4 of Chapter 1.

Exercise 9. This exercise is used in the proof of Proposition (2.7). Suppose we have an ideal $I \subset k[x_1,\ldots,x_n]$, and let $p = (x_1 - a_1)\cdots(x_1 - a_d)$, where $a_1,\ldots,a_d$ are distinct. The goal of this exercise is to prove that

$$I + \langle p\rangle = \bigcap_j (I + \langle x_1 - a_j\rangle).$$

a. Prove that $I + \langle p\rangle \subset \bigcap_j (I + \langle x_1 - a_j\rangle)$.
b. Let $p_j = \prod_{i\neq j}(x_1 - a_i)$. Prove that $p_j \cdot (I + \langle x_1 - a_j\rangle) \subset I + \langle p\rangle$.
c. Show that $p_1,\ldots,p_n$ are relatively prime, and conclude that there are polynomials $h_1,\ldots,h_n$ such that $1 = \sum_j h_j p_j$.
d. Prove that $\bigcap_j (I + \langle x_1 - a_j\rangle) \subset I + \langle p\rangle$. Hint: Given h in the intersection, write $h = \sum_j h_j p_j h$ and use part b.

Exercise 10. (The Dual Space of $k[x_1,\ldots,x_n]/I$) Recall that if V is a vector space over a field k, then the *dual space* of V, denoted V^*, is the k-vector space of linear mappings $L : V \to k$. If V is finite-dimensional, then so is V^*, and $\dim V = \dim V^*$. Let I be a zero-dimensional ideal in $k[x_1,\ldots,x_n]$, and consider $A = k[x_1,\ldots,x_n]/I$ with its k-vector space structure. Let G be a Gröbner basis for I with respect to some monomial ordering, and let $B = \{x^{\alpha(1)},\ldots,x^{\alpha(d)}\}$ be the corresponding monomial basis for A, so that for each $f \in k[x_1, \ldots, x_n]$,

$$\overline{f}^G = \sum_{j=1}^{d} c_j(f) x^{\alpha(j)}$$

for some $c_j(f) \in k$.

a. Show that each of the functions $c_j(f)$ is a linear function of $f \in k[x_1, \ldots, x_n]$. Moreover, show that $c_j(f) = 0$ for all j if and only if $f \in I$, or equivalently $[f] = 0$ in A.

b. Deduce that the collection B^* of mappings c_j given by $f \mapsto c_j(f)$, $j = 1, \ldots, d$ gives a *basis* of the dual space A^*.

c. Show that B^* is the *dual basis* corresponding to the basis B of A. That is, show that

$$c_j(x^{\alpha(i)}) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Exercise 11. Let $S = \{p_1, \ldots, p_m\}$ be a finite subset of $\mathbb{C}^n$.

a. Show that there exists a linear polynomial $\ell(x_1, \ldots, x_n)$ whose values at the points of S are *distinct*.

b. Using the linear polynomial ℓ from part a, show that there exist polynomials $g_i \in \mathbb{C}[x_1, \ldots, x_n]$, $i = 1, \ldots, m$, such that

$$g_i(p_j) = \begin{cases} 0 & \text{if } i \neq j, \text{ and} \\ 1 & \text{if } i = j. \end{cases}$$

Hint: Mimic the construction of the Lagrange interpolation polynomials in the discussion after the statement of Lemma (2.9).

Exercise 12. As in Theorem (2.10), suppose that $\mathbf{V}(I) = \{p_1, \ldots, p_m\}$.

a. Prove that the map $\varphi : \mathbb{C}[x_1, \ldots, x_n]/I \to \mathbb{C}^m$ given by evaluation at $p_1, \ldots, p_m$ is a well-defined linear map. Hint: $[f] = [g]$ implies $f - g \in I$.

b. We can regard $\mathbb{C}^m$ as a ring with coordinate-wise multiplication. Thus

$$(a_1, \ldots, a_m) \cdot (b_1, \ldots, b_m) = (a_1 b_1, \ldots, a_m b_m).$$

With this ring structure, $\mathbb{C}^m$ is a direct product of m copies of $\mathbb{C}$. Prove that the map φ of part a is a ring homomorphism.

c. Prove that φ is a ring isomorphism if and only if I is radical. This means that in the radical case, we can express A as a direct product of the simpler rings (namely, m copies of $\mathbb{C}$). In Chapter 4, we will generalize this result to the nonradical case.

Exercise 13. In Maple, the `SetBasis` command finds a monomial basis B for the quotient algebra $A = k[x_1, \ldots, x_n]/I$ for a zero-dimensional ideal I. However, it is instructive to have the following "home-grown" version called `kbasis` which makes it easier to see what is happening.

```
kbasis := proc(GB,VList,torder)

 # returns a list of monomials forming a basis of the quotient
 # ring, where GB is a Groebner basis for a zero-dimensional
 # ideal, and generates an error message if the ideal is not
 # 0-dimensional.

local B,C,v,t,l,m,leadmons,i;

if is_finite(GB,VList) then
  leadmons:={seq(leadterm(GB[i],torder),i=1..nops(GB))};
  B:=[1];
  for v in VList do
    m:=degree(univpoly(v,GB),v);
    C:=B;
    for t in C do
      for l to m-1 do
        t:=t*v;
        if evalb(not(1 in map(u->denom(t/u),leadmons))) then
          B:=[op(B),t];
        end if;
      end do;
    end do;
  end do;
  return B;
else
  print(‘ideal is not zero-dimensional‘);
end if
end proc:
```

a. Show that `kbasis` correctly computes $\{x^\alpha : x^\alpha \notin \langle \textsc{lt}(I)\rangle\}$ if A is finite-dimensional over k and terminates for all inputs.
b. Use either `kbasis` or `SetBasis` to check the results for the ideal from (2.4).
c. Use either `kbasis` or `SetBasis` to check your work from Exercise 6 above.

Exercise 14. The algorithm used in the procedure from Exercise 13 can be improved considerably. The "box" R that `kbasis` searches for elements of the complement of $\langle \textsc{lt}(I)\rangle$ is often much larger than necessary. This is because the call to `univpoly`, which finds a monic generator for $I \cap k[x_i]$ for each i, gives an m_i such that $x_i^{m_i} \in \langle \textsc{lt}(I)\rangle$, but m_i might not be as small as possible. For instance, consider the ideal I from (2.4). The monic generator of $I \cap \mathbb{C}[x]$ has degree 4 (check this). Hence `kbasis` computes

$\overline{x^2}^G$, $\overline{x^3}^G$ and rejects these monomials since they are not remainders. But the Gröbner basis G given in (2.4) shows that $x^2 \in \langle \textsc{lt}(I) \rangle$. Thus a smaller set of α containing the exponents of the monomial basis B can be determined directly by examining the leading terms of the Gröbner basis G, without using `univpoly` to get the monic generator for $I \cap k[x_i]$. Develop and implement an improved `kbasis` that takes this observation into account.

Exercise 15. Using either `Setbasis` or `kbasis`, develop and implement a procedure that computes the multiplication table for a finite-dimensional algebra A.

Exercise 16. Implement the following Maple procedure for finding the radical of a zero-dimensional ideal given by Proposition (2.7) and test it on the examples from this section.

```
zdimradical := proc(PList,VList)

 # constructs a set of generators for the radical of a
 # zero-dimensional ideal.

local p,pred,v,RList;

if is_finite(PList,VList) then
  RList := PList;
  for v in VList do
    p := univpoly(v,PList);
    pred := simplify(p/gcd(p,diff(p,v)));
    RList:=[op(RList),pred]
  end do;
  return RList
else
  print(‘Ideal not zero-dimensional; method does not apply‘)
end if
end proc:
```

Exercise 17. Let $I \subset \mathbb{C}[x_1, \ldots, x_n]$ be an ideal such that for every $1 \leq i \leq n$, there is a square-free polynomial p_i such that $p_i(x_i) \in I$. Use Proposition (2.7) to show that I is radical.

Exercise 18. For $1 \leq i \leq n$, let p_i be a square-free polynomial. Also let $d_i = \deg(p_i)$. The goal of this exercise is to prove that $\langle p_1(x_1), \ldots, p_n(x_n) \rangle$ is radical using only the division algorithm.

a. Let r be the remainder of $f \in \mathbb{C}[x_1, \ldots, x_n]$ on division by the $p_i(x_i)$. Prove that r has degree at most $d_i - 1$ in x_i.

b. Prove that r vanishes on $\mathbf{V}(p_1(x_1), \ldots, p_n(x_n))$ if and only if r is identically 0.
c. Conclude that $\langle p_1(x_1), \ldots, p_n(x_n)\rangle$ is radical without using Proposition (2.7).

Exercise 19. In this exercise, you will use Exercise 18 to give an elementary proof of the result of Exercise 17. Thus we assume that $I \subset \mathbb{C}[x_1, \ldots, x_n]$ is an ideal such that for every $1 \leq i \leq n$, there is a square-free polynomial p_i such that $p_i(x_i) \in I$. Take $f \in \mathbb{C}[x_1, \ldots, x_n]$ such that $f^N \in I$ for some $N > 0$. Let z be a new variable and set $J = \langle p_1(x_1), \ldots, p_n(x_n), z - f\rangle \subset \mathbb{C}[x_1, \ldots, x_n, z]$.
a. Prove that there is a ring isomorphism

$$\mathbb{C}[x_1, \ldots, x_n, z]/J \cong \mathbb{C}[x_1, \ldots, x_n]/\langle p_1(x_1), \ldots, p_n(x_n)\rangle$$

and conclude via Exercise 18 that J is zero-dimensional and radical.
b. Without using Proposition (2.7), show that there is a square-free polynomial g such that $g(z) \in J$.
c. Explain why $\mathrm{GCD}(g, z^N)$ is 1 or z, and conclude that $z = p(z)g(z) + q(z)z^N$ for some polynomials p, q.
d. Under the isomorphism of part a, show that $z = p(z)g(z) + q(z)z^N$ maps to $f = q(f)f^N + h$, where $h \in \langle p_1(x_1), \ldots, p_n(x_n)\rangle$. Conclude that $f \in I$.

This argument is due to M. Mereb.

§3 Gröbner Basis Conversion

In this section, we will use linear algebra in $A = k[x_1, \ldots, x_n]/I$ to show that a Gröbner basis G for a zero-dimensional ideal I with respect to one monomial order can be converted to a Gröbner basis G' for the same ideal with respect to *any* other monomial order. The process is sometimes called *Gröbner basis conversion*, and the idea comes from a paper of Faugère, Gianni, Lazard, and Mora [FGLM]. We will illustrate the method by converting from an arbitrary Gröbner basis G to a *lex* Gröbner basis G_{lex} (using any ordering on the variables). The Gröbner basis conversion method is often used in precisely this situation, so that a more favorable monomial order (such as *grevlex*) can be used in the application of Buchberger's algorithm, and the result can then be converted into a form more suited for equation solving via elimination. For another discussion of this topic, see [BW], §1 of Chapter 9.

The basic idea of the Faugère-Gianni-Lazard-Mora algorithm is quite simple. We start with a Gröbner basis G for a zero-dimensional ideal I, and we want to convert G to a *lex* Gröbner basis G_{lex} for some *lex* order. The algorithm steps through monomials in $k[x_1, \ldots, x_n]$ in increasing *lex* order. At each step of the algorithm, we have a list $G_{lex} = \{g_1, \ldots, g_k\}$ of elements in I (initially empty, and at each stage a subset of the eventual *lex* Gröbner basis), and a list B_{lex} of monomials (also initially empty, and at each stage a subset of the eventual *lex* monomial basis for A). For each input monomial x^α (initially 1), the algorithm consists of three steps:

(3.1) Main Loop. Given the input x^α, compute $\overline{x^\alpha}^G$. Then:

a. If $\overline{x^\alpha}^G$ is *linearly dependent* on the remainders (on division by G) of the monomials in B_{lex}, then we have a linear combination
$$\overline{x^\alpha}^G - \textstyle\sum_j c_j \overline{x^{\alpha(j)}}^G = 0,$$
where $x^{\alpha(j)} \in B_{lex}$ and $c_j \in k$. This implies that
$$g = x^\alpha - \textstyle\sum_j c_j x^{\alpha(j)} \in I.$$
We add g to the list G_{lex} as the last element. Because the x^α are considered in increasing *lex* order (see (3.3) below), whenever a polynomial g is added to G_{lex}, its leading term is $\text{LT}(g) = x^\alpha$ with coefficient 1.

b. If $\overline{x^\alpha}^G$ is *linearly independent* from the remainders (on division by G) of the monomials in B_{lex}, then we add x^α to B_{lex} as the last element.

After the Main Loop acts on the monomial x^α, we test G_{lex} to see if we have the desired Gröbner basis. This test needs to be done only if we added a polynomial g to G_{lex} in part a of the Main Loop.

(3.2) Termination Test. If the Main Loop added a polynomial g to G_{lex}, then compute $\text{LT}(g)$. If $\text{LT}(g)$ is a power of x_1, where x_1 is the greatest variable in our *lex* order, then the algorithm terminates.

The proof of Theorem (3.4) below will explain why this is the correct way to terminate the algorithm. If the algorithm does not stop at this stage, we use the following procedure to find the next input monomial for the Main Loop:

(3.3) Next Monomial. Replace x^α with the next monomial in *lex* order which is not divisible by any of the monomials $\text{LT}(g_i)$ for $g_i \in G_{lex}$.

Exercise 3 below will explain how the Next Monomial procedure works. Now repeat the above process by using the new x^α as input to the Main Loop, and continue until the Termination Test tells us to stop.

Before we prove the correctness of this algorithm, let's see how it works in an example.

Exercise 1. Consider the ideal
$$I = \langle xy + z - xz, x^2 - z, 2x^3 - x^2yz - 1\rangle$$

in $\mathbb{Q}[x,y,z]$. For *grevlex* order with $x > y > z$, I has a Gröbner basis $G = \{f_1, f_2, f_3, f_4\}$, where

$$\begin{aligned} f_1 &= z^4 - 3z^3 - 4yz + 2z^2 - y + 2z - 2 \\ f_2 &= yz^2 + 2yz - 2z^2 + 1 \\ f_3 &= y^2 - 2yz + z^2 - z \\ f_4 &= x + y - z. \end{aligned}$$

Thus $\langle \mathrm{LT}(I) \rangle = \langle z^4, yz^2, y^2, x \rangle$, $B = \{1, y, z, z^2, z^3, yz\}$, and a remainder $\overline{f}^G$ is a linear combination of elements of B. We will use basis conversion to find a *lex* Gröbner basis for I, with $z > y > x$.

a. Carry out the Main Loop for $x^\alpha = 1, x, x^2, x^3, x^4, x^5, x^6$. At the end of doing this, you should have

$$\begin{aligned} G_{lex} &= \{x^6 - x^5 - 2x^3 + 1\} \\ B_{lex} &= \{1, x, x^2, x^3, x^4, x^5\}. \end{aligned}$$

Hint: The following computations will be useful:

$$\begin{aligned} \overline{1}^G &= 1 \\ \overline{x}^G &= -y + z \\ \overline{x^2}^G &= z \\ \overline{x^3}^G &= -yz + z^2 \\ \overline{x^4}^G &= z^2 \\ \overline{x^5}^G &= z^3 + 2yz - 2z^2 + 1 \\ \overline{x^6}^G &= z^3. \end{aligned}$$

Note that $\overline{1}^G, \ldots, \overline{x^5}^G$ are linearly independent while $\overline{x^6}^G$ is a linear combination of $\overline{x^5}^G$, $\overline{x^3}^G$ and $\overline{1}^G$. This is similar to Exercise 2 of §2.

b. After we apply the Main Loop to x^6, show that the monomial provided by the Next Monomial procedure is y, and after y passes through the Main Loop, show that

$$\begin{aligned} G_{lex} &= \{x^6 - x^5 - 2x^3 + 1, y - x^2 + x\} \\ B_{lex} &= \{1, x, x^2, x^3, x^4, x^5\}. \end{aligned}$$

c. Show that after y, Next Monomial produces z, and after z passes through the Main Loop, show that

$$\begin{aligned} G_{lex} &= \{x^6 - x^5 - 2x^3 + 1, y - x^2 + x, z - x^2\} \\ B_{lex} &= \{1, x, x^2, x^3, x^4, x^5\}. \end{aligned}$$

d. Check that the Termination Test (3.2) terminates the algorithm when G_{lex} is as in part c. Hint: We're using *lex* order with $z > y > x$.

e. Verify that G_{lex} from part c is a *lex* Gröbner basis for I.

We will now show that the algorithm given by (3.1), (3.2) and (3.3) terminates and correctly computes a *lex* Gröbner basis for the ideal I.

(3.4) Theorem. *The algorithm described above terminates on every input Gröbner basis G generating a zero-dimensional ideal I, and correctly computes a lex Gröbner basis G_{lex} for I and the lex monomial basis B_{lex} for the quotient ring A.*

Proof. We begin with the key observation that monomials are added to the list B_{lex} in strictly increasing *lex* order. Similarly, if $G_{lex} = \{g_1, \ldots, g_k\}$, then

$$\mathrm{LT}(g_1) <_{lex} \cdots <_{lex} \mathrm{LT}(g_k),$$

where $>_{lex}$ is the *lex* order we are using. We also note that when the Main Loop adds a new polynomial g_{k+1} to $G_{lex} = \{g_1, \ldots, g_k\}$, the leading term $\mathrm{LT}(g_{k+1})$ is the input monomial in the Main Loop. Since the input monomials are provided by the Next Monomial procedure, it follows that for all k,

(3.5) $\mathrm{LT}(g_{k+1})$ is divisible by none of $\mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_k)$.

We can now prove that the algorithm terminates for all inputs G generating zero-dimensional ideals. If the algorithm did not terminate for some input G, then the Main Loop would be executed infinitely many times, so one of the two alternatives in (3.1) would be chosen infinitely often. If the first alternative were chosen infinitely often, G_{lex} would give an infinite list $\mathrm{LT}(g_1), \mathrm{LT}(g_2), \ldots$ of monomials. However, we have:

- (Dickson's Lemma) Given an infinite list $x^{\alpha(1)}, x^{\alpha(2)}, \ldots$ of monomials in $k[x_1, \ldots, x_n]$, there is an integer N such that every $x^{\alpha(i)}$ is divisible by one of $x^{\alpha(1)}, \ldots, x^{\alpha(N)}$.

(See, for example, Exercise 7 of [CLO], Chapter 2, §4). When applied to $\mathrm{LT}(g_1), \mathrm{LT}(g_2), \ldots$, Dickson's Lemma would contradict (3.5). On the other hand, if the second alternative were chosen infinitely often, then B_{lex} would give infinitely many monomials $x^{\alpha(j)}$ whose remainders on division by G were linearly independent in A. This would contradict the assumption that I is zero-dimensional. As a result, the algorithm always terminates for G generating a zero-dimensional ideal I.

Next, suppose that the algorithm terminates with $G_{lex} = \{g_1, \ldots, g_k\}$. By the Termination Test (3.2), $\mathrm{LT}(g_k) = x_1^{a_1}$, where $x_1 >_{lex} \cdots >_{lex} x_n$. We will prove that G_{lex} is a *lex* Gröbner basis for I by contradiction. Suppose there were some $g \in I$ such that $\mathrm{LT}(g)$ is not a multiple of any of the $\mathrm{LT}(g_i)$, $i = 1, \ldots, k$. Without loss of generality, we may assume that g is *reduced* with respect to G_{lex} (replace g by $\overline{g}^{G_{lex}}$).

If $\text{LT}(g)$ is greater than $\text{LT}(g_k) = x_1^{a_1}$, then one easily sees that $\text{LT}(g)$ is a multiple of $\text{LT}(g_k)$ (see Exercise 2 below). Hence this case can't occur, which means that

$$\text{LT}(g_i) < \text{LT}(g) \le \text{LT}(g_{i+1})$$

for some $i < k$. But recall that the algorithm places monomials into B_{lex} in strictly increasing order, and the same is true for the $\text{LT}(g_i)$. All the non-leading monomials in g must be less than $\text{LT}(g)$ in the *lex* order. They are not divisible by any of $\text{LT}(g_j)$ for $j \le i$, since g is reduced. So, the non-leading monomials that appear in g would have been included in B_{lex} by the time $\text{LT}(g)$ was reached by the Next Monomial procedure, and g would have been the next polynomial after g_i included in G_{lex} by the algorithm (i.e., g would equal g_{i+1}). This contradicts our assumption on g, which proves that G_{lex} is a *lex* Gröbner basis for I.

The final step in the proof is to show that when the algorithm terminates, B_{lex} consists of *all* basis monomials determined by the Gröbner basis G_{lex}. We leave this as an exercise for the reader. □

In the literature, the basis conversion algorithm discussed here is called the *FGLM algorithm* after the authors Faugère, Gianni, Lazard, and Mora of the paper [FGLM] in which the algorithm first appeared. We should also mention that while the FGLM algorithm assumes that I is zero-dimensional, there are methods which apply to the positive-dimensional case. For instance, if degree bounds on the elements of the Gröbner basis with respect to the desired order are known, then the approach described above can also be adapted to treat ideals that are not zero-dimensional. An interesting related "Hilbert function-driven" basis conversion method for homogeneous ideals has been proposed by Traverso (see [Trav]). However, general basis conversion methods that apply even when information such as degree bounds is not available are also desirable. Such a method is the *Gröbner Walk* to be described in Chapter 8.

The ideas used in Gröbner basis conversion can be applied in other contexts. In order to explain this, we need to recast the above discussion using linear maps. Recall that we began with a Gröbner basis G of a zero-dimensional ideal I and our goal was to find a *lex* Gröbner basis G_{lex} of I. However, for G, the main thing we used was the normal form $\overline{f}^G$ of a polynomial $f \in k[x_1, \ldots, x_n]$.

Let's write this out carefully. Let B be the monomial basis of $A = k[x_1, \ldots, x_n]/I$ determined by G. Denote $\overline{f}^G$ by $L(f)$ and $\text{Span}(B)$ by V, so that $L(f) = \overline{f}^G \in V = \text{Span}(B)$. Thus we have a map

$$L : k[x_1, \ldots, x_n] \longrightarrow V. \tag{3.6}$$

In Exercise 10 of §2, you showed that L is linear with kernel equal to I. Using this, the Main Loop (3.1) can be written as follows.

(3.7) Main Loop, Restated. Given the input x^α, compute $L(x^\alpha)$. Then:

a. If $L(x^\alpha)$ is *linearly dependent* on the images under L of the monomials in B_{lex}, then we have a linear combination

$$L(x^\alpha) - \textstyle\sum_j c_j L(x^{\alpha(j)}) = 0,$$

where $x^{\alpha(j)} \in B_{lex}$ and $c_j \in k$. This implies that $L\bigl(x - \sum_j c_j x^{\alpha(j)}\bigr) = 0$. Since I is the kernel of L, we have

$$g = x^\alpha - \textstyle\sum_j c_j x^{\alpha(j)} \in I.$$

We add g to G_{lex} as the last element.

b. If $L(x^\alpha)$ is *linearly independent* from the images under L of the monomials in B_{lex}, then we add x^α to B_{lex} as the last element.

If we combine (3.7) with the Termination Test (3.2) and Next Monomial (3.3), then we get the same algorithm as before. But even more is true, for this algorithm computes a *lex* Gröbner basis of the kernel for *any* linear map (3.6), provided that V has finite dimension and the kernel is an ideal of $k[x_1, \ldots, x_n]$. You will prove this in Exercise 9 below.

As an example of how this works, pick distinct points $p_1, \ldots, p_m \in k^n$ and consider the evaluation map

$$L : k[x_1, \ldots, x_n] \longrightarrow k^m, \quad L(f) = (f(p_1), \ldots, f(p_m)).$$

The kernel is the ideal $\mathbf{I}(p_1, \ldots, p_m)$ of polynomials vanishing at the given points. It follows that we now have an algorithm for computing a *lex* Gröbner basis of this ideal! This is closely related to the Buchberger-Möller algorithm described in [BuM]. You will work out an explicit example in Exercise 10.

For another example, consider

$$(3.8) \quad I = \{f \in \mathbb{C}[x,y] : f(0,0) = f_x(0,0) = f_y(0,0) - f_{xx}(0,0) = 0\}.$$

In Exercise 11, you will show that I is an ideal of $\mathbb{C}[x,y]$. Since I is the kernel of the linear map

$$L : \mathbb{C}[x,y] \longrightarrow \mathbb{C}^3, \quad L(f) = (f(0,0), f_x(0,0), f_y(0,0) - f_{xx}(0,0)),$$

the above algorithm can be used to show that $\{y^2, xy, x^2 + 2y\}$ is a *lex* Gröbner basis with $x > y$ for the ideal I. See Exercise 11 for the details.

There are some very interesting ideas related to these examples. Differential conditions like those in (3.8), when combined with primary decomposition, can be used to describe any zero-dimensional ideal in $k[x_1, \ldots, x_n]$. This is explained in [MMM1] and [MöS] (and is where we got (3.8)). The paper [MMM1] also describes other situations where these ideas are useful, and [MMM2] makes a systematic study of the different representations of a zero-dimensional ideal and how one can pass from one representation to another.

Additional Exercises for §3

Exercise 2. Consider the *lex* order with $x_1 > \cdots > x_n$ and fix a power x_1^a of x_1. Then, for any monomial x^α in $k[x_1, \ldots, x_n]$, prove that $x^\alpha > x_1^a$ if and only if x^α is divisible by x_1^a.

Exercise 3. Suppose $G_{lex} = \{g_1, \ldots, g_k\}$, where $\mathrm{LT}(g_1) < \cdots < \mathrm{LT}(g_k)$, and let x^α be a monomial. This exercise will show how the Next Monomial (3.3) procedure works, assuming that our *lex* order satisfies $x_1 > \cdots > x_n$. Since this procedure is only used when the Termination Test fails, we can assume that $\mathrm{LT}(g_k)$ is *not* a power of x_1.

a. Use Exercise 2 to show that none of the $\mathrm{LT}(g_i)$ divide $x_1^{a_1+1}$.

b. Now consider the *largest* $1 \le k \le n$ such that none of the $\mathrm{LT}(g_i)$ divide the monomial
$$x_1^{a_1} \cdots x_{k-1}^{a_{k-1}} x_k^{a_k+1}.$$
By part a, $k = 1$ has this property, so there must be a largest such k. If x^β is the monomial corresponding to the largest k, prove that $x^\beta > x^\alpha$ is the smallest monomial (relative to our *lex* order) greater than x^α which is not divisible by any of the $\mathrm{LT}(g_i)$.

Exercise 4. Complete the proof of Theorem (3.4) by showing that when the basis conversion algorithm terminates, the set B_{lex} gives a monomial basis for the quotient ring A.

Exercise 5. Use Gröbner basis conversion to find *lex* Gröbner bases for the ideals in Exercises 6 and 7 from §1. Compare with your previous results.

Exercise 6. What happens if you try to apply the basis conversion algorithm to an ideal that is *not* zero-dimensional? Can this method be used for general Gröbner basis conversion? What if you have more information about the *lex* basis elements, such as their total degrees, or bounds on those degrees?

Exercise 7. Show that the output of the basis conversion algorithm is actually a monic *reduced lex* Gröbner basis for $I = \langle G \rangle$.

Exercise 8. Implement the basis conversion algorithm outlined in (3.1), (3.2) and (3.3) in a computer algebra system. Hint: Exercise 3 will be useful. For a more complete description of the algorithm, see pages 428–433 of [BW].

Exercise 9. Consider a linear map $L : k[x_1, \ldots, x_n] \to V$, where V has finite dimension and the kernel of L is an ideal. State and prove a version of Theorem (3.4) which uses (3.7), (3.2), and (3.3).

Exercise 10. Use the method described at the end of the section to find a *lex* Gröbner basis with $x > y$ for the ideal of all polynomials vanishing at $(0,0), (1,0), (0,1) \in k^2$.

Exercise 11. Prove that (3.8) is an ideal of $\mathbb{C}[x,y]$ and use the method described at the end of the section to find a *lex* Gröbner basis with $x > y$ for this ideal.

§4 Solving Equations via Eigenvalues and Eigenvectors

The central problem of this chapter, finding the solutions of a system of polynomial equations $f_1 = f_2 = \cdots = f_s = 0$ over $\mathbb{C}$, rephrases in fancier language to finding the points of the variety $\mathbf{V}(I)$, where I is the ideal generated by $f_1, \ldots, f_s$. When the system has only finitely many solutions, i.e., when $\mathbf{V}(I)$ is a finite set, the Finiteness Theorem from §2 says that I is a zero-dimensional ideal and the algebra $A = \mathbb{C}[x_1, \ldots, x_n]/I$ is a finite-dimensional vector space over $\mathbb{C}$. The first half of this section exploits the structure of A in this case to evaluate an arbitrary polynomial f at the points of $\mathbf{V}(I)$; in particular, evaluating the polynomials $f = x_i$ gives the coordinates of the points (Corollary (4.6) below). The values of f on $\mathbf{V}(I)$ turn out to be *eigenvalues* of certain linear mappings on A. We will discuss techniques for computing these eigenvalues and show that the corresponding *eigenvectors* contain useful information about the solutions.

We begin with the easy observation that given a polynomial $f \in \mathbb{C}[x_1, \ldots, x_n]$, we can use multiplication to define a linear map m_f from $A = \mathbb{C}[x_1, \ldots, x_n]/I$ to itself. More precisely, f gives the coset $[f] \in A$, and we define $m_f : A \to A$ by the rule: if $[g] \in A$, then

$$m_f([g]) = [f] \cdot [g] = [fg] \in A.$$

Then m_f has the following basic properties.

(4.1) Proposition. *Let $f \in \mathbb{C}[x_1, \ldots, x_n]$. Then*

a. *The map m_f is a linear mapping from A to A.*

b. *We have $m_f = m_g$ exactly when $f - g \in I$. Thus two polynomials give the same linear map if and only if they differ by an element of I. In particular, m_f is the zero map exactly when $f \in I$.*

Proof. The proof of part a is just the distributive law for multiplication over addition in the ring A. If $[g], [h] \in A$ and $c \in k$, then

$$m_f(c[g] + [h]) = [f] \cdot (c[g] + [h]) = c[f] \cdot [g] + [f] \cdot [h] = cm_f([g]) + m_f([h]).$$

Part b is equally easy. Since $[1] \in A$ is a multiplicative identity, if $m_f = m_g$, then

$$[f] = [f] \cdot [1] = m_f([1]) = m_g([1]) = [g] \cdot [1] = [g],$$

so $f - g \in I$. Conversely, if $f - g \in I$, then $[f] = [g]$ in A, so $m_f = m_g$. □

Since A is a finite-dimensional vector space over $\mathbb{C}$, we can represent m_f by its matrix with respect to a basis. For our purposes, a monomial basis B such as the ones we considered in §2 will be the most useful, because once we have the multiplication table for the elements in B, the matrices of the multiplication operators m_f can be read off immediately from the table. We will denote this matrix also by m_f, and whether m_f refers to the matrix or the linear operator will be clear from the context. Proposition (4.1) implies that $m_f = m_{\overline{f}^G}$, so that we may assume that f is a remainder.

For example, for the ideal I from (2.4) of this chapter, the matrix for the multiplication operator by f may be obtained from the table (2.5) in the usual way. Ordering the basis monomials as before,

$$B = \{1, x, y, xy, y^2\},$$

we make a 5×5 matrix whose jth column is the vector of coefficients in the expansion in terms of B of the image under m_f of the jth basis monomial. With $f = x$, for instance, we obtain

$$m_x = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 3/2 & 0 & -3/2 & 1 \\ 0 & 3/2 & 0 & -1/2 & 0 \\ 0 & -3/2 & 1 & 3/2 & 0 \\ 0 & -1/2 & 0 & 3/2 & 0 \end{pmatrix}.$$

Exercise 1. Find the matrices m_1, m_y, m_{xy-y^2} with respect to B in this example. How do m_{y^2} and $(m_y)^2$ compare? Why?

We note the following useful general properties of the matrices m_f (the proof is left as an exercise).

(4.2) Proposition. *Let f, g be elements of the algebra A. Then*
a. $m_{f+g} = m_f + m_g$.
b. $m_{f \cdot g} = m_f \cdot m_g$ *(where the product on the right means composition of linear operators or matrix multiplication).*

This proposition says that the map sending $f \in \mathbb{C}[x_1, \ldots, x_n]$ to the matrix m_f defines a *ring homomorphism* from $\mathbb{C}[x_1, \ldots, x_n]$ to the ring $M_{d\times d}(\mathbb{C})$ of $d \times d$ matrices, where d is the dimension of A as a $\mathbb{C}$-vector space. Furthermore, part b of Proposition (4.1) and the Fundamental Theorem of Homomorphisms show that $[f] \mapsto m_f$ induces a one-to-one homomorphism $A \to M_{d\times d}(\mathbb{C})$. A discussion of ring homomorphisms and the

Fundamental Theorem of Homomorphisms may be found in Chapter 5, §2 of [CLO], especially Exercise 16. But the reader should note that $M_{d\times d}(\mathbb{C})$ is not a commutative ring, so we have here a slightly more general situation than the one discussed there.

For use later, we also point out a corollary of Proposition (4.2). Let $h(t) = \sum_{i=0}^m c_i t^i \in \mathbb{C}[t]$ be a polynomial. The expression $h(f) = \sum_{i=0}^m c_i f^i$ makes sense as an element of $\mathbb{C}[x_1, \ldots, x_n]$. Similarly $h(m_f) = \sum_{i=0}^m c_i (m_f)^i$ is a well-defined matrix (the term c_0 should be interpreted as $c_0 I$, where I is the $d \times d$ identity matrix).

(4.3) Corollary. *In the situation of Proposition (4.2), let $h \in \mathbb{C}[t]$ and $f \in \mathbb{C}[x_1, \ldots, x_n]$. Then*

$$m_{h(f)} = h(m_f).$$

Recall that a polynomial $f \in \mathbb{C}[x_1, \ldots, x_n]$ gives the coset $[f] \in A$. Since A is finite-dimensional, as we noted in §2 for $f = x_i$, the set $\{1, [f], [f]^2, \ldots\}$ must be *linearly dependent* in the vector space structure of A. In other words, there is a linear combination

$$\sum_{i=0}^m c_i [f]^i = [0]$$

in A, where $c_i \in \mathbb{C}$ are not all zero. By the definition of the quotient ring, this is equivalent to saying that

$$\sum_{i=0}^m c_i f^i \in I. \tag{4.4}$$

Hence $\sum_{i=0}^m c_i f^i$ vanishes at every point of $\mathbf{V}(I)$.

Now we come to the most important part of this discussion, culminating in Theorem (4.5) and Corollary (4.6) below. We are looking for the points in $\mathbf{V}(I)$, I a zero-dimensional ideal. Let $h(t) \in \mathbb{C}[t]$, and let $f \in \mathbb{C}[x_1, \ldots, x_n]$. By Corollary (4.3),

$$h(m_f) = 0 \qquad \Longleftrightarrow \qquad h([f]) = [0] \text{ in } A.$$

The polynomials h such that $h(m_f) = 0$ form an ideal in $\mathbb{C}[t]$ by the following exercise.

Exercise 2. Given a $d \times d$ matrix M with entries in a field k, consider the collection I_M of polynomials $h(t)$ in $k[t]$ such that $h(M) = 0$, the $d \times d$ zero matrix. Show that I_M is an ideal in $k[t]$.

The nonzero monic generator h_M of the ideal I_M is called the *minimal polynomial* of M. By the basic properties of ideals in $k[t]$, if h is any polynomial with $h(M) = 0$, then the minimal polynomial h_M divides h. In particular, the Cayley-Hamilton Theorem from linear algebra tells us that

h_M divides the characteristic polynomial of M. As a consequence, if $k = \mathbb{C}$, the roots of h_M are eigenvalues of M. Furthermore, all eigenvalues of M occur as roots of the minimal polynomial. See [Her] for a more complete discussion of the Cayley-Hamilton Theorem and the minimal polynomial of a matrix.

Let h_f denote the minimal polynomial of the multiplication operator m_f on A. We then have three interesting sets of numbers:

- the *roots* of the equation $h_f(t) = 0$,
- the *eigenvalues* of the matrix m_f, and
- the *values* of the function f on $\mathbf{V}(I)$, the set of points we are looking for.

The amazing fact is that all three sets are equal.

(4.5) Theorem. *Let $I \subset \mathbb{C}[x_1, \ldots, x_n]$ be zero-dimensional, let $f \in \mathbb{C}[x_1, \ldots, x_n]$, and let h_f be the minimal polynomial of m_f on $A = \mathbb{C}[x_1, \ldots, x_n]/I$. Then, for $\lambda \in \mathbb{C}$, the following are equivalent:*
a. *λ is a root of the equation $h_f(t) = 0$,*
b. *λ is an eigenvalue of the matrix m_f, and*
c. *λ is a value of the function f on $\mathbf{V}(I)$.*

Proof. a $\Leftrightarrow$ b follows from standard results in linear algebra.

b $\Rightarrow$ c: Let λ be an eigenvalue of m_f. Then there is a corresponding eigenvector $[z] \neq [0] \in A$ such that $[f - \lambda][z] = [0]$. Aiming for a contradiction, suppose that λ is not a value of f on $\mathbf{V}(I)$. That is, letting $\mathbf{V}(I) = \{p_1, \ldots, p_m\}$, suppose that $f(p_i) \neq \lambda$ for all $i = 1, \ldots, m$.

Let $g = f - \lambda$, so that $g(p_i) \neq 0$ for all i. By Lemma (2.9) of this chapter, there exist polynomials g_i such that $g_i(p_j) = 0$ if $i \neq j$, and $g_i(p_i) = 1$. Consider the polynomial $g' = \sum_{i=1}^m 1/g(p_i) g_i$. It follows that $g'(p_i)g(p_i) = 1$ for all i, and hence $1 - g'g \in \mathbf{I}(\mathbf{V}(I))$. By the Nullstellensatz, $(1 - g'g)^\ell \in I$ for some $\ell \geq 1$. Expanding by the binomial theorem and collecting the terms that contain g as a factor, we get $1 - \tilde{g}g \in I$ for some $\tilde{g} \in \mathbb{C}[x_1, \ldots, x_n]$. In A, this last inclusion implies that $[1] = [\tilde{g}][g]$, hence g has a multiplicative inverse $[\tilde{g}]$ in A.

But from the above we have $[g][z] = [f - \lambda][z] = [0]$ in A. Multiplying both sides by $[\tilde{g}]$, we obtain $[z] = [0]$, which is a contradiction. Therefore λ must be a value of f on $\mathbf{V}(I)$.

c $\Rightarrow$ a: Let $\lambda = f(p)$ for $p \in \mathbf{V}(I)$. Since $h_f(m_f) = 0$, Corollary (4.3) shows $h_f([f]) = [0]$, and then (4.4) implies $h_f(f) \in I$. This means $h_f(f)$ vanishes at every point of $\mathbf{V}(I)$, so that $h_f(\lambda) = h_f(f(p)) = 0$. □

Exercise 3. We saw earlier that the matrix of multiplication by x in the 5-dimensional algebra $A = \mathbb{C}[x, y]/I$ from (2.4) of this chapter is given by the matrix displayed before Exercise 1 in this section.

a. Using the `minpoly` command in Maple (part of the `linalg` package) or otherwise, show that the minimal polynomial of this matrix is

$$h_x(t) = t^4 - 2t^3 - t^2 + 2t.$$

The roots of $h_x(t) = 0$ are thus $t = 0, -1, 1, 2$.

b. Now find all points of $\mathbf{V}(I)$ using the methods of §1 and show that the roots of h_x are exactly the distinct values of the function $f(x, y) = x$ at the points of $\mathbf{V}(I)$. (Two of the points have the same x-coordinate, which explains why the degree and the number of roots are 4 instead of 5!) Also see Exercise 7 from §2 to see how the ideal I was constructed.

c. Finally, find the minimal polynomial of the matrix m_y, determine its roots, and explain the degree you get.

When we apply Theorem (4.5) with $f = x_i$, we get a general result exactly parallel to this example.

(4.6) Corollary. *Let $I \subset \mathbb{C}[x_1, \ldots, x_n]$ be zero-dimensional. Then the eigenvalues of the multiplication operator m_{x_i} on A coincide with the x_i-coordinates of the points of $\mathbf{V}(I)$. Moreover, substituting $t = x_i$ in the minimal polynomial h_{x_i} yields the unique monic generator of the elimination ideal $I \cap \mathbb{C}[x_i]$.*

Corollary (4.6) indicates that it is possible to solve equations by computing eigenvalues of the multiplication operators m_{x_i}. This has been studied in papers such as [Laz], [Möl], and [MöS], among others. As a result a whole array of numerical methods for approximating eigenvalues can be brought to bear on the root-finding problem, at least in favorable cases. We include a brief discussion of some of these methods for the convenience of some readers; the following two paragraphs may be safely ignored if you are familiar with numerical eigenvalue techniques. For more details, we suggest [BuF] or [Act].

In elementary linear algebra, eigenvalues of a matrix M are usually determined by solving the *characteristic polynomial equation*:

$$\det(M - tI) = 0.$$

The degree of the polynomial on the left hand side is the size of the matrix M. But computing $\det(M - tI)$ for large matrices is a large job itself, and as we have seen in §1, exact solutions (and even accurate approximations to solutions) of polynomial equations of high degree over $\mathbb{R}$ or $\mathbb{C}$ can be hard to come by, so the characteristic polynomial is almost never used in practice. So other methods are needed.

The most basic numerical eigenvalue method is known as the *power method*. It is based on the fact that if a matrix M has a unique *dominant eigenvalue* (i.e., an eigenvalue λ satisfying $|\lambda| > |\mu|$ for all other eigenvalues μ of M), then starting from a randomly chosen vector x_0, and forming the sequence

$$x_{k+1} = \text{ unit vector in direction of } Mx_k,$$

we almost always approach an eigenvector for λ as $k \to \infty$. An approximate value for the dominant eigenvalue λ may be obtained by computing the norm $\|Mx_k\|$ at each step. If there is no unique dominant eigenvalue, then the iteration may not converge, but the power method can also be modified to eliminate that problem and to find other eigenvalues of M. In particular, we can find the eigenvalue of M *closest to* some fixed s by applying the power method to the matrix $M' = (M - sI)^{-1}$. For almost all choices of s, there will be a unique dominant eigenvalue of M'. Moreover, if λ' is that dominant eigenvalue of M', then $1/\lambda' + s$ is the eigenvalue of M closest to s. This observation makes it possible to search for *all* the eigenvalues of a matrix as we would do in using the Newton-Raphson method to find all the roots of a polynomial. Some of the same difficulties arise, too. There are also much more sophisticated iterative methods, such as the LR and QR algorithms, that can be used to determine *all* the (real or complex) eigenvalues of a matrix except in some very uncommon degenerate situations. It is known that the QR algorithm, for instance, converges for all matrices having no more than two eigenvalues of any given magnitude in $\mathbb{C}$. Some computer algebra systems (e.g., Maple and Mathematica) provide built-in procedures that implement these methods.

A legitimate question at this point is this: *Why* might one consider applying these eigenvalue techniques for root finding instead of using elimination? There are two reasons.

The first concerns the amount of calculation necessary to carry out this approach. The direct attack—solving systems via elimination as in §1—imposes a *choice of monomial order* in the Gröbner basis we use. Pure *lex* Gröbner bases frequently require a large amount of computation. As we saw in §3, it is possible to compute a *grevlex* Gröbner basis first, then convert it to a *lex* basis using the FGLM basis conversion algorithm, with some savings in total effort. But basis conversion is unnecessary if we use Corollary (4.6), because the algebraic structure of $\mathbb{C}[x_1, \ldots, x_n]/I$ is *independent of the monomial order* used for the Gröbner basis and remainder calculations. Hence any monomial order can be used to determine the matrices of the multiplication operators m_{x_i}.

The second reason concerns the amount of numerical versus symbolic computation involved, and the potential for numerical instability. In the frequently-encountered case that the generators for I have rational coefficients, the entries of the matrices m_{x_i} will also be rational, and hence can be determined *exactly* by symbolic computation. Thus the numerical component of the calculation is restricted to the eigenvalue calculations.

There is also a significant difference even between a naive first idea for implementing this approach and the elimination method discussed in §1. Namely, we could begin by computing all the m_{x_i} and their eigenvalues separately. Then with some additional computation we could determine exactly which vectors $(x_1, \ldots, x_n)$ formed using values of the coordinate functions actually give approximate solutions. The difference here is that the computed values of x_i *are not used* in the determination of the x_j, $j \neq i$. In §1, we saw that a major source of error in approximate solutions was the fact that small errors in one variable could produce larger errors in the other variables when we substitute them and use the Extension Theorem. Separating the computations of the values x_i from one another, we can avoid those *accumulated error* phenomena (and also the numerical stability problems encountered in other non-elimination methods).

We will see shortly that it is possible to reduce the computational effort involved even further. Indeed, it suffices to consider the eigenvalues of only one suitably-chosen multiplication operator $m_{c_1x_1+\cdots+c_nx_n}$. Before developing this result, however, we present an example using the more naive approach.

Exercise 4. We will apply the ideas sketched above to find approximations to the complex solutions of the system:

$$\begin{aligned} 0 &= x^2 - 2xz + 5 \\ 0 &= xy^2 + yz + 1 \\ 0 &= 3y^2 - 8xz. \end{aligned}$$

a. First, compute a Gröbner basis to determine the monomial basis for the quotient algebra. We can use the *grevlex* (Maple `tdeg`) monomial order:

```
PList := [x^2 - 2*x*z + 5, x*y^2 + y*z + 1, 3*y^2 - 8*x*z];
G := gbasis(PList,tdeg(x,y,z));
B := SetBasis(G,tdeg(x,y,z))[1];
```

(this can also be done using the `kbasis` procedure from Exercise 13 in §2) and obtain the eight monomials:

$$[1, x, y, xy, z, z^2, xz, yz].$$

(You should compare this with the output of `SetBasis` or `kbasis` for *lex* order. Also print out the *lex* Gröbner basis for this ideal if you have a taste for complicated polynomials.)

b. Using the monomial basis B, check that the matrix of the full multiplication operator m_x is

$$\begin{pmatrix} 0 & -5 & 0 & 0 & 0 & -3/16 & -3/8 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3/20 & 0 & 0 & 0 & 3/40 \\ 0 & 0 & 0 & 0 & 0 & 5/2 & 0 & 0 \\ 0 & 0 & 0 & -2 & 0 & 0 & 0 & -1 \\ 0 & 2 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -3/10 & 0 & -3/16 & -3/8 & -3/20 \end{pmatrix}.$$

This matrix can also be computed using the `MulMatrix` command in Maple.

c. Now, applying the numerical eigenvalue routine `eigenvals` from Maple, check that there are two approximate real eigenvalues:

$$-1.100987715, \qquad .9657124563,$$

and 3 complex conjugate pairs. (This computation can be done in several different ways and, due to roundoff effects, the results can be slightly different depending on the method used. The values above were found by expressing the entries of the matrix of m_x as floating point numbers, and applying Maple's `eigenvals` routine to that matrix.)

d. Complete the calculation by finding the multiplication operators m_y, m_z, computing their real eigenvalues, and determining which triples (x, y, z) give solutions. (There are exactly two real points.) Also see Exercises 9 and 10 below for a second way to compute the eigenvalues of m_x, m_y, and m_z.

In addition to eigenvalues, there are also eigenvectors to consider. In fact, every matrix M has two sorts of eigenvectors. The *right eigenvectors* of M are the usual ones, which are column vectors $v \neq 0$ such that

$$M\, v = \lambda v$$

for some $\lambda \in \mathbb{C}$. Since the transpose M^T has the same eigenvalues λ as M, we can find a column vector $v' \neq 0$ such that

$$M^T\, v' = \lambda v'.$$

Taking transposes, we can write this equation as

$$w\, M = \lambda w,$$

where $w = v'^T$ is a row vector. We call w a *left eigenvector* of M.

The right and left eigenvectors for a matrix are connected in the following way. For simplicity, suppose that M is a *diagonalizable* $n \times n$ matrix, so that there is a basis for $\mathbb{C}^n$ consisting of right eigenvectors for M. In Exercise 7 below, you will show that there is a matrix equation $MQ = QD$, where Q is the matrix whose columns are the right eigenvectors in a basis for $\mathbb{C}^n$, and D is a diagonal matrix whose diagonal entries are the eigenvalues of M. Rearranging the last equation, we have $Q^{-1}M = DQ^{-1}$. By the second part of Exercise 7 below, the rows of Q^{-1} are a collection of left eigenvectors of M that also form a basis for $\mathbb{C}^n$.

For a zero-dimensional ideal I, there is also a strong connection between the points of $\mathbf{V}(I)$ and the left eigenvectors of the matrix m_f relative to the monomial basis B coming from a Gröbner basis. We will assume that I is radical. In this case, Theorem (2.10) implies that A has dimension m, where m is the number of points in $\mathbf{V}(I)$. Hence, we can write the monomial basis B as the cosets

$$B = \{[x^{\alpha(1)}], \ldots, [x^{\alpha(m)}]\}.$$

Using this basis, let m_f be the matrix of multiplication by f. We can relate the left eigenvectors of m_f to points of $\mathbf{V}(I)$ as follows.

(4.7) Proposition. *Suppose $f \in \mathbb{C}[x_1, \ldots, x_n]$ is chosen such that the values $f(p)$ are distinct for $p \in \mathbf{V}(I)$, where I is a radical ideal not containing 1. Then the left eigenspaces of the matrix m_f are 1-dimensional and are spanned by the row vectors $(p^{\alpha(1)}, \ldots, p^{\alpha(m)})$ for $p \in \mathbf{V}(I)$.*

Proof. If we write $m_f = (m_{ij})$, then for each j between 1 and m,

$$[x^{\alpha(j)} f] = m_f([x^{\alpha(j)}]) = m_{1j}[x^{\alpha(1)}] + \cdots + m_{mj}[x^{\alpha(m)}].$$

Now fix $p \in \mathbf{V}(f_1, \ldots, f_n)$ and evaluate this equation at p to obtain

$$p^{\alpha(j)} f(p) = m_{1j} p^{\alpha(1)} + \cdots + m_{mj} p^{\alpha(m)}$$

(this makes sense by Exercise 12 of §2). Doing this for $j = 1, \ldots, m$ gives

$$f(p)(p^{\alpha(1)}, \ldots, p^{\alpha(m)}) = (p^{\alpha(1)}, \ldots, p^{\alpha(m)})\, m_f.$$

Exercise 14 at the end of the section asks you to check this computation carefully. Note that one of the basis monomials in B is the coset [1] (do you see why this follows from $1 \notin I$?), which shows that $(p^{\alpha(1)}, \ldots, p^{\alpha(m)})$ is nonzero and hence is a left eigenvector for m_f, with $f(p)$ as the corresponding eigenvalue.

By hypothesis, the $f(p)$ are distinct for $p \in \mathbf{V}(I)$, which means that the $m \times m$ matrix m_f has m distinct eigenvalues. Linear algebra then implies that the corresponding eigenspaces (right and left) are 1-dimensional. □

This proposition can be used to find the points in $\mathbf{V}(I)$ for *any* zero-dimensional ideal I. The basic idea is as follows. First, we can assume that I is radical by replacing I with $\sqrt{I}$ as computed by Proposition (2.7). Then compute a Gröbner basis G and monomial basis B as usual. Now consider the function

$$f = c_1 x_1 + \cdots + c_n x_n,$$

where $c_1, \ldots, c_n$ are randomly chosen integers. This will ensure (with small probability of failure) that the values $f(p)$ are distinct for $p \in \mathbf{V}(I)$. Relative to the monomial basis B, we get the matrix m_f, so that we can use standard numerical methods to find an eigenvalue λ and corresponding left eigenvector v of m_f. This eigenvector, when combined with the Gröbner basis G, makes it *trivial* to find a solution $p \in \mathbf{V}(I)$.

To see how this is done, first note that Proposition (4.7) implies

$$v = c(p^{\alpha(1)}, \ldots, p^{\alpha(m)}) \tag{4.8}$$

for some nonzero constant c and some $p \in \mathbf{V}(I)$. Write $p = (a_1, \ldots, a_n)$. Our goal is to compute the coordinates a_i of p in terms of the coordinates of v. Equation (4.8) implies that each coordinate of v is of the form $cp^{\alpha(j)}$.

The Finiteness Theorem implies that for each i between 1 and n, there is $m_i \geq 1$ such that $x_i^{m_i}$ is the leading term of some element of G. If $m_i > 1$, it follows that $[x_i] \in B$ (do you see why?), so that ca_i is a coordinate of v. As noted above, we have $[1] \in B$, so that c is also a coordinate of v. Consequently,

$$a_i = \frac{ca_i}{c}$$

is a ratio of coordinates of v. This way, we get the x_i-coordinate of p for all i satisfying $m_i > 1$.

It remains to study the coordinates with $m_i = 1$. These variables appear in *none* of the basis monomials in B (do you see why?), so that we turn instead to the Gröbner basis G for guidance. Suppose the variables with $m_i = 1$ are $x_{i_1}, \ldots, x_{i_\ell}$. We will assume that the variables are labeled so that $x_1 > \cdots > x_n$ and $i_1 > \cdots > i_\ell$. In Exercise 15 below, you will show that for $j = 1, \ldots, \ell$, there are elements $g_j \in G$ such that

$$g_j = x_{i_j} + \text{terms involving } x_i \text{ for } i > i_j.$$

If we evaluate this at $p = (a_1, \ldots, a_n)$, we obtain

$$0 = a_{i_j} + \text{terms involving } a_i \text{ for } i > i_j. \tag{4.9}$$

Since we already know a_i for $i \notin \{i_1, \ldots, i_\ell\}$, these equations make it a simple matter to find $a_{i_1}, \ldots, a_{i_\ell}$. We start with a_{i_ℓ}. For $j = \ell$, (4.9) implies that a_{i_ℓ} is a polynomial in the coordinates of p we already know. Hence we get a_{i_ℓ}. But once we know a_{i_ℓ}, (4.9) shows that $a_{i_{\ell-1}}$ is also a polynomial in known coordinates. Continuing in this way, we get *all* of the coordinates of p.

Exercise 5. Apply this method to find the solutions of the equations given in Exercise 4. The x-coordinates of the solutions are distinct, so you can assume $f = x$. Thus it suffices to compute the left eigenvectors of the matrix m_x of Exercise 4.

The idea of using eigenvectors to find solutions first appears in the pioneering work of Auzinger and Stetter [AS] in 1988 and was further developed in [MöS], [MT], and [Ste]. Our treatment focused on the radical case since our first step was to replace I with $\sqrt{I}$. In general, whenever a multiplication map m_f is *nonderogatory* (meaning that all eigenspaces have dimension one), one can use Proposition (4.7) to find the solutions. Unfortunately, when I is not radical, it can happen that m_f is derogatory for *all* $f \in k[x_1, \ldots, x_n]$. Rather than replacing I with $\sqrt{I}$ as we did above, another approach is to realize that the *family* of operators $\{m_f : f \in k[x_1, \ldots, x_n]\}$ is nonderogatory, meaning that its joint left eigenspaces are one-dimensional and hence are spanned by the eigenvectors described in Proposition (4.7). This result and its consequences are discussed in [MT] and [Mou1]. We will say more about multiplication maps in §2 of Chapter 4.

Since the left eigenvectors of m_f help us find solutions in $\mathbf{V}(I)$, it is natural to ask about the right eigenvectors. In Exercise 17 below, you will show that these eigenvectors solve the *interpolation problem*, which asks for a polynomial that takes preassigned values at the points of $\mathbf{V}(I)$.

This section has discussed several ideas for solving polynomial equations using linear algebra. We certainly do not claim that these ideas are a computational panacea for all polynomial systems, but they do give interesting alternatives to other, more traditional methods in numerical analysis, and they are currently an object of study in connection with the implementation of the next generation of computer algebra systems. We will continue this discussion in §5 (where we study real solutions) and Chapter 3 (where we use resultants to solve polynomial systems).

Additional Exercises for §4

Exercise 6. Prove Proposition (4.2).

Exercise 7. Let M, Q, P, D be $n \times n$ complex matrices, and assume D is a diagonal matrix.

a. Show that the equation $MQ = QD$ holds if and only if each nonzero column of Q is a right eigenvector of M and the corresponding diagonal entry of D is the corresponding eigenvalue.
b. Show that the equation $PM = DP$ holds if and only if each nonzero row of P is a left eigenvector of M and the corresponding diagonal entry of D is the corresponding eigenvalue.
c. If $MQ = QD$ and Q is invertible, deduce that the rows of Q^{-1} are left eigenvectors of M.

Exercise 8.

a. Apply the eigenvalue method from Corollary (4.6) to solve the system from Exercise 6 of §1. Compare your results.

b. Apply the eigenvalue method from Corollary (4.6) to solve the system from Exercise 7 from §1. Compare your results.

Exercise 9. Let V_i be the subspace of A spanned by the non-negative powers of $[x_i]$, and consider the *restriction* of the multiplication operator $m_{x_i} : A \to A$ to V_i. Assume $\{1, [x_i], \ldots, [x_i]^{m_i - 1}\}$ is a basis for V_i.

a. What is the matrix of the restriction $m_{x_i}|_{V_i}$ with respect to this basis? Show that it can be computed by the same calculations used in Exercise 4 of §2 to find the monic generator of $I \cap \mathbb{C}[x_i]$, without computing a *lex* Gröbner basis. Hint: See also Exercise 11 of §1 of Chapter 3.
b. What is the characteristic polynomial of $m_{x_i}|_{V_i}$ and what are its roots?

Exercise 10. Use part b of Exercise 9 and Corollary (4.6) to give another determination of the roots of the system from Exercise 4.

Exercise 11. Let I be a zero-dimensional ideal in $\mathbb{C}[x_1, \ldots, x_n]$, and let $f \in \mathbb{C}[x_1, \ldots, x_n]$. Show that $[f]$ has a multiplicative inverse in $\mathbb{C}[x_1, \ldots, x_n]/I$ if and only if $f(p) \neq 0$ for all $p \in \mathbf{V}(I)$. Hint: See the proof of Theorem (4.5).

Exercise 12. Prove that a zero-dimensional ideal is radical if and only if the matrices m_{x_i} are diagonalizable for each i. Hint: Linear algebra tells us that a matrix is diagonalizable if and only if its minimal polynomial is square-free. Proposition (2.7) and Corollary (4.6) of this chapter will be useful.

Exercise 13. Let $A = \mathbb{C}[x_1, \ldots, x_n]/I$ for a zero-dimensional ideal I, and let $f \in \mathbb{C}[x_1, \ldots, x_n]$. If $p \in \mathbf{V}(I)$, we can find $g \in \mathbb{C}[x_1, \ldots, x_n]$ with $g(p) = 1$, and $g(p') = 0$ for all $p' \in \mathbf{V}(I)$, $p' \neq p$ (see Lemma (2.9)). Prove that there is an $\ell \geq 1$ such that the coset $[g^\ell] \in A$ is a *generalized eigenvector* for m_f with eigenvalue $f(p)$. (A generalized eigenvector of a matrix M is a nonzero vector v such that $(M - \lambda I)^m v = 0$ for some $m \geq 1$.) Hint: Apply the Nullstellensatz to $(f - f(p))g$. In Chapter 4, we will study the generalized eigenvectors of m_f in more detail.

Exercise 14. Verify carefully the formula $f(p)(p^{\alpha(1)}, \ldots, p^{\alpha(m)}) = (p^{\alpha(1)}, \ldots, p^{\alpha(m)})\, m_f$ used in the proof of Proposition (4.7).

Exercise 15. Let $>$ be some monomial order, and assume $x_1 > \cdots > x_n$. If $g \in k[x_1, \ldots, x_n]$ satisfies $\mathrm{LT}(g) = x_j$, then prove that

$$g = x_j + \text{terms involving } x_i \text{ for } i > j.$$

Exercise 16. (The Shape Lemma) Let I be a zero-dimensional radical ideal such that the x_n-coordinates of the points in $\mathbf{V}(I)$ are distinct. Let

G be a reduced Gröbner basis for I relative to a *lex* monomial order with x_n as the *last* variable.

a. If $\mathbf{V}(I)$ has m points, prove that the cosets $1, [x_n], \ldots, [x_n^{m-1}]$ are linearly independent and hence are a basis of $A = k[x_1, \ldots, x_n]/I$.

b. Prove that G consists of n polynomials

$$\begin{aligned} g_1 &= x_1 + h_1(x_n) \\ &\vdots \\ g_{n-1} &= x_{n-1} + h_{n-1}(x_n) \\ g_n &= x_n^m + h_n(x_n), \end{aligned}$$

where $h_1, \ldots, h_n$ are polynomials in x_n of degree at most $m-1$. Hint: Start by expressing $[x_1], \ldots, [x_{n-1}], [x_n^m]$ in terms of the basis of part a.

c. Explain how you can find *all* points of $\mathbf{V}(I)$ once you know their x_n-coordinates. Hint: Adapt the discussion following (4.9).

Exercise 17. This exercise will study the right eigenvectors of the matrix m_f and their relation to interpolation. Assume that I is a zero-dimensional radical ideal and that the values $f(p)$ are distinct for $p \in \mathbf{V}(I)$. We write the monomial basis B as $\{[x^{\alpha(1)}], \ldots, [x^{\alpha(m)}]\}$.

a. If $p \in \mathbf{V}(I)$, Lemma (2.9) of this chapter gives us g such that $g(p) = 1$ and $g(p') = 0$ for all $p' \neq p$ in $\mathbf{V}(I)$. Prove that the coset $[g] \in A$ is a right eigenvector of m_f and that the corresponding eigenspace has dimension 1. Conclude that *all* eigenspaces of m_f are of this form.

b. If $v = (v_1, \ldots, v_m)^t$ is a right eigenvector of m_f corresponding to the eigenvalue $f(p)$ for p as in part a, then prove that the polynomial

$$\tilde{g} = v_1 x^{\alpha(1)} + \cdots + v_m x^{\alpha(m)}$$

satisfies $\tilde{g}(p) \neq 0$ and $\tilde{g}(p') = 0$ for $p' \neq p$ in $\mathbf{V}(I)$.

c. Show that we can take the polynomial g of part a to be

$$g = \frac{1}{\tilde{g}(p)}\,\tilde{g}.$$

Thus, once we know the solution p and the corresponding right eigenvector of m_f, we get an *explicit formula* for the polynomial g.

d. Given $\mathbf{V}(I) = \{p_1, \ldots, p_m\}$ and the corresponding right eigenvectors of m_f, we get polynomials $g_1, \ldots, g_m$ such that $g_i(p_j) = 1$ if $i = j$ and 0 otherwise. Each g_i is given explicitly by the formula in part c. The *interpolation problem* asks to find a polynomial h which takes preassigned values $\lambda_1, \ldots, \lambda_m$ at the points $p_1, \ldots, p_m$. This means $h(p_i) = \lambda_i$ for all i. Prove that one choice for h is given by

$$h = \lambda_1 g_1 + \cdots + \lambda_m g_m.$$

Exercise 18. Let $A = k[x_1, \ldots, x_n]/I$, where I is zero-dimensional. In Maple, `MulMatrix` computes the matrix of the multiplication map m_{x_i} relative to a monomial basis computed by `SetBasis`. However, in §5, we will need to compute the matrix of m_f, where $f \in k[x_1, \ldots, x_n]$ is an arbitrary polynomial. Develop and code a Maple procedure `getmatrix` which, given a polynomial f, a monomial basis B, a Gröbner basis G, and a term order, produces the matrix of m_f relative to B. You will use `getmatrix` in Exercise 6 of §5.

§5 Real Root Location and Isolation

The eigenvalue techniques for solving equations from §4 are only a first way that we can use the results of §2 for finding roots of systems of polynomial equations. In this section we will discuss a second application that is more sophisticated. We follow a recent paper of Pedersen, Roy, and Szpirglas [PRS] and consider the problem of determining the *real* roots of a system of polynomial equations with coefficients in a field $k \subset \mathbb{R}$ (usually $k = \mathbb{Q}$ or a finite extension field of $\mathbb{Q}$). The underlying principle here is that for many purposes, explicitly determined, bounded regions $R \subset \mathbb{R}^n$, each guaranteed to contain *exactly one* solution of the system can be just as useful as a collection of numerical approximations. Note also that if we wanted numerical approximations, once we had such an R, the job of finding that one root would generally be much simpler than a search for *all* of the roots! (Think of the choice of the initial approximation for an iterative method such as Newton-Raphson.) For one-variable equations, this is also the key idea of the *interval arithmetic* approach to computation with real algebraic numbers (see [Mis]). We note that there are also other methods known for locating and isolating the real roots of a polynomial system (see §8.8 of [BW] for a different type of algorithm).

To define our regions R in $\mathbb{R}^n$, we will use polynomial functions in the following way. Let $h \in k[x_1, \ldots, x_n]$ be a nonzero polynomial. The real points where h takes the value 0 form the variety $\mathbf{V}(h) \cap \mathbb{R}^n$. We will denote this by $\mathbf{V}_\mathbb{R}(h)$ in the discussion that follows. In typical cases, $\mathbf{V}_\mathbb{R}(h)$ will be a *hypersurface*—an $(n-1)$-dimensional variety in $\mathbb{R}^n$. The complement of $\mathbf{V}_\mathbb{R}(h)$ in $\mathbb{R}^n$ is the union of connected open subsets on which h takes either all positive values or all negative values. We obtain in this way a decomposition of $\mathbb{R}^n$ as a disjoint union

$$\mathbb{R}^n = H^+ \cup H^- \cup \mathbf{V}_\mathbb{R}(h), \tag{5.1}$$

where $H^+ = \{a \in \mathbb{R}^n : h(a) > 0\}$, and similarly for H^-. Here are some concrete examples.

Exercise 1.

a. Let $h = (x^2+y^2-1)(x^2+y^2-2)$ in $\mathbb{R}[x, y]$. Identify the regions H^+ and H^- for this polynomial. How many connected components does each of them have?

b. In this part of the exercise, we will see how regions like rectangular "boxes" in $\mathbb{R}^n$ may be obtained by intersecting several regions H^+ or H^-. For instance, consider the box

$$R = \{(x, y) \in \mathbb{R}^2 : a < x < b,\ c < y < d\}.$$

If $h_1(x, y) = (x - a)(x - b)$ and $h_2(x, y) = (y - c)(y - d)$, show that

$$R = H_1^- \cap H_2^- = \{(x, y) \in \mathbb{R}^2 : h_i(x, y) < 0,\ i = 1, 2\}.$$

What do H_1^+, H_2^+ and $H_1^+ \cap H_2^+$ look like in this example?

Given a region R like the box from part b of the above exercise, and a system of equations, we can ask whether there are roots of the system in R. The results of [PRS] give a way to answer questions like this, using an extension of the results of §2 and §4. Let I be a zero-dimensional ideal and let B be the monomial basis of $A = k[x_1, \ldots, x_n]/I$ for any monomial order. Recall that the *trace* of a square matrix is just the sum of its diagonal entries. This gives a mapping Tr from $d \times d$ matrices to k. Using the trace, we define a *symmetric bilinear form* S by the rule:

$$S(f, g) = \mathrm{Tr}(m_f \cdot m_g) = \mathrm{Tr}(m_{fg})$$

(the last equality follows from part b of Proposition (4.2)).

Exercise 2.

a. Prove that S defined as above is a symmetric bilinear form on A, as claimed. That is, show that S is symmetric, meaning $S(f, g) = S(g, f)$ for all $f, g \in A$, and linear in the first variable, meaning

$$S(cf_1 + f_2, g) = cS(f_1, g) + S(f_2, g)$$

for all $f_1, f_2, g \in A$ and all $c \in k$. It follows that S is linear in the second variable as well.

b. Given a symmetric bilinear form S on a vector space V with basis $\{v_1, \ldots, v_d\}$, the matrix of S is the $d \times d$ matrix $M = (S(v_i, v_j))$. Show that the matrix of S with respect to the monomial basis $B = \{x^{\alpha(i)}\}$ for A is given by:

$$M = (\mathrm{Tr}(m_{x^{\alpha(i)}x^{\alpha(j)}})) = (\mathrm{Tr}(m_{x^{\alpha(i)+\alpha(j)}})).$$

Similarly, given the polynomial $h \in k[x_1, \ldots, x_n]$ used in the decomposition (5.1), we can construct a bilinear form

$$S_h(f, g) = \mathrm{Tr}(m_{hf} \cdot m_g) = \mathrm{Tr}(m_{hfg}).$$

Let M_h be the matrix of S_h with respect to B.

Exercise 3. Show that S_h is also a symmetric bilinear form on A. What is the i, j entry of M_h?

Since we assume $k \subset \mathbb{R}$, the matrices M and M_h are symmetric matrices with *real* entries. It follows from the real spectral theorem (or principal axis theorem) of linear algebra that all of the eigenvalues of M and M_h will be *real*. For our purposes the exact values of these eigenvalues are much less important than their *signs*.

Under a change of basis defined by an invertible matrix Q, the matrix M of a symmetric bilinear form S is taken to Q^tMQ. There are two fundamental invariants of S under such changes of basis—the *signature* $\sigma(S)$, which equals the difference between the number of positive eigenvalues and the number of negative eigenvalues of M, and the *rank* $\rho(S)$, which equals the rank of the matrix M. (See, for instance, Chapter 6 of [Her] for more information on the signature and rank of bilinear forms.)

We are now ready to state the main result of this section.

(5.2) Theorem. *Let I be a zero-dimensional ideal generated by polynomials in $k[x_1, \ldots, x_n]$ ($k \subset \mathbb{R}$), so that $\mathbf{V}(I) \subset \mathbb{C}^n$ is finite. Then, for $h \in k[x_1, \ldots, x_n]$, the signature and rank of the bilinear form S_h satisfy:*

$$\begin{aligned}\sigma(S_h) &= \#\{a \in \mathbf{V}(I) \cap \mathbb{R}^n : h(a) > 0\} - \#\{a \in \mathbf{V}(I) \cap \mathbb{R}^n : h(a) < 0\}\\ \rho(S_h) &= \#\{a \in \mathbf{V}(I) : h(a) \neq 0\}.\end{aligned}$$

Proof. This result is essentially a direct consequence of the reasoning leading up to Theorem (4.5) of this chapter. However, to give a full proof it is necessary to take into account the *multiplicities* of the points in $\mathbf{V}(I)$ as defined in Chapter 4. Hence we will only sketch the proof in the special case when I is radical. By Theorem (2.10), this means that $\mathbf{V}(I) = \{p_1, \ldots, p_m\}$, where m is the dimension of the algebra A. Given the basis $B = \{[x^{\alpha(i)}]\}$ of A, Proposition (4.7) implies that $(p_j^{\alpha(i)})$ is an invertible matrix.

By Theorem (4.5), for any f, we know that the set of eigenvalues of m_f coincides with the set of values of the f at the points in $\mathbf{V}(I)$. The key new fact we will need is that using the structure of the algebra A, for each point p in $\mathbf{V}(I)$ it is possible to define a positive integer $m(p)$ (the multiplicity) so that $\sum_p m(p) = d = \dim(A)$, and so that $(t - f(p))^{m(p)}$ is a factor of the characteristic polynomial of m_f. (See §2 of Chapter 4 for the details.)

By definition, the i, j entry of the matrix M_h is equal to

$$\mathrm{Tr}(m_{h \cdot x^{\alpha(i)} \cdot x^{\alpha(j)}}).$$

The trace of the multiplication operator equals the sum of its eigenvalues. By the previous paragraph, the sum of these eigenvalues is

$$\sum_{p \in \mathbf{V}(I)} m(p)h(p)p^{\alpha(i)}p^{\alpha(j)}, \tag{5.3}$$

where $p^{\alpha(i)}$ denotes the value of the monomial $x^{\alpha(i)}$ at the point p. List the points in $\mathbf{V}(I)$ as $p_1, \ldots, p_d$, where each point p in $\mathbf{V}(I)$ is repeated $m(p)$ times consecutively. Let U be the $d \times d$ matrix whose jth column consists of the values $p_j^{\alpha(i)}$ for $i = 1, \ldots, d$. From (5.3), we obtain a matrix factorization $M_h = UDU^t$, where D is the diagonal matrix with entries $h(p_1), \ldots, h(p_d)$. The equation for the rank follows since U is invertible. Both U and D may have nonreal entries. However, the equation for the signature follows from this factorization as well, using the facts that M_h has real entries and that the nonreal points in $\mathbf{V}(I)$ occur in complex conjugate pairs. We refer the reader to Theorem 2.1 of [PRS] for the details. □

The theorem may be used to determine how the real points in $\mathbf{V}(I)$ are distributed among the sets H^+, H^- and $\mathbf{V}_{\mathbb{R}}(h)$ determined by h in (5.1). Theorem (5.2) implies that we can count the number of real points of $\mathbf{V}(I)$ in H^+ and in H^- as follows. The signature of S_h gives the *difference* between the number of solutions in H^+ and the number in H^-. By the same reasoning, computing the signature of S_{h^2} we get the number of solutions in $H^+ \cup H^-$, since $h^2 > 0$ at every point of $H^+ \cup H^-$. From this we can recover $\#\mathbf{V}(I) \cap H^+$ and $\#\mathbf{V}(I) \cap H^-$ by simple arithmetic. Finally, we need to find $\#\mathbf{V}(I) \cap \mathbf{V}_{\mathbb{R}}(h)$, which is done in the following exercise.

Exercise 4. Using the form S_1 in addition to S_h and S_{h^2}, show that the three signatures $\sigma(S), \sigma(S_h), \sigma(S_{h^2})$ give all the information needed to determine $\#\mathbf{V}(I) \cap H^+$, $\#\mathbf{V}(I) \cap H^-$ and $\#\mathbf{V}(I) \cap \mathbf{V}_{\mathbb{R}}(h)$.

From the discussion above, it might appear that we need to compute the eigenvalues of the forms S_h to count the numbers of solutions of the equations in H^+ and H^-, but the situation is actually *much better than that.* Namely, the entire calculation can be done symbolically, so no recourse to numerical methods is needed. The reason is the following consequence of the classical Descartes Rule of Signs.

(5.4) Proposition. *Let M_h be the matrix of S_h, and let*

$$p_h(t) = \det(M_h - tI)$$

be its characteristic polynomial. Then the number of positive eigenvalues of S_h is equal to the number of sign changes in the sequence of coefficients of $p_h(t)$. (In counting sign changes, any zero coefficients are ignored.)

Proof. See Proposition 2.8 of [PRS], or Exercise 5 below for a proof. □

For instance, consider the real symmetric matrix

$$M = \begin{pmatrix} 3 & 1 & 5 & 4 \\ 1 & 2 & 6 & 9 \\ 5 & 6 & 7 & -1 \\ 4 & 9 & -1 & 0 \end{pmatrix}.$$

The characteristic polynomial of M is $t^4 - 12t^3 - 119t^2 + 1098t - 1251$, giving *three* sign changes in the sequence of coefficients. Thus M has three positive eigenvalues, as one can check.

Exercise 5. The usual version of Descartes' Rule of Signs asserts that the number of positive roots of a polynomial $p(t)$ in $\mathbb{R}[t]$ equals the number of sign changes in its coefficient sequence minus a non-negative even integer.

a. Using this, show that the number of negative roots equals the number of sign changes in the coefficient sequence of $p(-t)$ minus another non-negative even integer.
b. Deduce (5.4) from Descartes' Rule of Signs, part a, and the fact that all eigenvalues of M_h are real.

Using these ideas to find and isolate roots requires a good searching strategy. We will not consider such questions here. For an example showing how to certify the presence of exactly one root of a system in a given region, see Exercise 6 below.

The problem of counting real solutions of polynomial systems in regions $R \subset \mathbb{R}^n$ defined by several polynomial inequalities and/or equalities has been considered in general by Ben-Or, Kozen, and Reif (see, for instance, [BKR]). Using the signature calculations as above gives an approach which is very well suited to *parallel* computation, and whose complexity is relatively manageable. We refer the interested reader to [PRS] once again for a discussion of these issues.

For a recent exposition of the material in this section, we refer the reader to Chapter 6 of [GRRT]. One topic not mentioned in our treatment is *semidefinite programming*. As explained in Chapter 7 of [Stu5], this has interesting relations to real solutions and sums of squares.

Additional Exercises for §5

Exercise 6. In this exercise, you will verify that the equations

$$\begin{aligned} 0 &= x^2 - 2xz + 5 \\ 0 &= xy^2 + yz + 1 \\ 0 &= 3y^2 - 8xz \end{aligned}$$

have exactly one real solution in the rectangular box

$$R = \{(x, y, z) \in \mathbb{R}^3 : 0 < x < 1,\ -3 < y < -2,\ 3 < z < 4\}.$$

a. Using *grevlex* monomial order with $x > y > z$, compute a Gröbner basis G for the ideal I generated by the above equations. Also find the corresponding monomial basis B for $\mathbb{C}[x, y, z]/I$.

b. Implement the following Maple procedure getform which computes the matrix of the symmetric bilinear form S_h.

```
getform := proc(h,B,G,torder)

 # computes the matrix of the symmetric bilinear form S_h,
 # with respect to the monomial basis B for the quotient
 # ring. G should be a Groebner basis with respect to
 # torder.

local d,M,i,j,p,q;

d:=nops(B);
M := array(symmetric,1..d,1..d);
for i to d do
  for j from i to d do
    p := normalf(h*B[i]*B[j],G,torder);
    M[i,j]:=trace(getmatrix(p,B,G,torder));
    end do;
  end do;
return eval(M)
end proc:
```

The call to getmatrix computes the matrix $m_{hx^{\alpha(i)}x^{\alpha(j)}}$ with respect to the monomial basis $B = \{x^{\alpha(i)}\}$ for A. Coding getmatrix was Exercise 18 in §4 of this chapter.

c. Then, using

```
h := x*(x-1);
S := getform(h,B,G,tdeg(x,y,z));
```

compute the matrix of the bilinear form S_h for $h = x(x - 1)$.

d. The actual entries of this 8 × 8 rational matrix are rather complicated and not very informative; we will omit reproducing them. Instead, use

```
charpoly(S,t);
```

to compute the characteristic polynomial of the matrix. Your result should be a polynomial of the form:

$$t^8 - a_1 t^7 + a_2 t^6 + a_3 t^5 - a_4 t^4 - a_5 t^3 - a_6 t^2 + a_7 t + a_8,$$

where each a_i is a positive rational number.

e. Use Proposition (5.4) to show that S_h has 4 positive eigenvalues. Since $a_8 \neq 0$, $t = 0$ is not an eigenvalue. Explain why the other 4 eigenvalues are strictly negative, and conclude that S_h has signature

$$\sigma(S_h) = 4 - 4 = 0.$$

f. Use the second equation in Theorem (5.2) to show that h is nonvanishing on the real or complex points of $\mathbf{V}(I)$. Hint: Show that S_h has rank 8.

g. Repeat the computation for h^2:

```
T := getform(h*h,B,G,tdeg(x,y,z));
```

and show that in this case, we get a second symmetric matrix with exactly 5 positive and 3 negative eigenvalues. Conclude that the signature of S_{h^2} (which counts the total number of real solutions in this case) is

$$\sigma(S_{h^2}) = 5 - 3 = 2.$$

h. Using Theorem (5.2) and combining these two calculations, show that

$$\#\mathbf{V}(I) \cap H^+ = \#\mathbf{V}(I) \cap H^- = 1,$$

and conclude that there is exactly one real root between the two planes $x = 0$ and $x = 1$ in $\mathbb{R}^3$. Our desired region R is contained in this infinite slab in $\mathbb{R}^3$. What can you say about the other real solution?

i. Complete the exercise by applying Theorem (5.2) to polynomials in y and z chosen according to the definition of R.

Exercise 7. Use the techniques of this section to determine the number of real solutions of

$$\begin{aligned} 0 &= x^2 + 2y^2 - y - 2z \\ 0 &= x^2 - 8y^2 + 10z - 1 \\ 0 &= x^2 - 7yz \end{aligned}$$

in the box $R = \{(x, y, z) \in \mathbb{R}^3 : 0 < x < 1, 0 < y < 1, 0 < z < 1\}$. (This is the same system as in Exercise 6 of §1. Check your results using your previous work.)

Exercise 8. The alternative real root isolation methods discussed in §8.8 of [BW] are based on a result for real one-variable polynomials known as Sturm's Theorem. Suppose $p(t) \in \mathbb{Q}[t]$ is a polynomial with no multiple roots in $\mathbb{C}$. Then $\mathrm{GCD}(p(t), p'(t)) = 1$, and the sequence of polynomials produced by

$$\begin{aligned} p_0(t) &= p(t) \\ p_1(t) &= p'(t) \\ p_i(t) &= -\mathrm{rem}(p_{i-1}(t), p_{i-2}(t), t), i \geq 2 \end{aligned}$$

(so $p_i(t)$ is the *negative* of the remainder on division of $p_{i-1}(t)$ by $p_{i-2}(t)$ in $\mathbb{Q}[t]$) will eventually reach a nonzero constant, and all subsequent terms will be zero. Let $p_m(t)$ be the last nonzero term in the sequence. This sequence of polynomials is called the *Sturm sequence* associated to $p(t)$.

a. (Sturm's Theorem) If $a < b$ in $\mathbb{R}$, and neither is a root of $p(t) = 0$, then show that the number of real roots of $p(t) = 0$ in the interval $[a, b]$ is the difference between the number of sign changes in the sequence of real numbers $p_0(a), p_1(a), \ldots, p_m(a)$ and the number of sign changes in the sequence $p_0(b), p_1(b), \ldots, p_m(b)$. (Sign changes are counted in the same way as for Descartes' Rule of Signs.)

b. Give an algorithm based on part a that takes as input a polynomial $p(t) \in \mathbb{Q}[t]$ with no multiple roots in $\mathbb{C}$, and produces as output a collection of intervals $[a_i, b_i]$ in $\mathbb{R}$, each of which contains exactly one root of p. Hint: Start with an interval guaranteed to contain all the real roots of $p(t) = 0$ (see Exercise 3 of §1, for instance) and bisect repeatedly, using Sturm's Theorem on each subinterval.