Introduction to Algebraic Computation

Tim Duff

March 10, 2023

Notation and conventions: All rings are commutative with 1, ring homomorphisms are assumed to preserve 1, and we allow $1 = 0$ in the zero ring. Given a ring R and elements $r_1, \ldots, r_k \in R$, we write $\langle r_1, \ldots, r_k \rangle$ for the ideal that they generate. If n is a positive integer, we define $[n] = \{1, \ldots, n\}$.

Contents

1 Fundamental Theorem of Algebra

Let $\mathbb{K}$ be a field. Recall that $\mathbb{K}$ is said to be *algebraically closed* if every nonconstant univariate polynomial with coefficients in $\mathbb{K}$ has a root. For example, neither the field of real numbers $\mathbb{R}$ nor its subfield consisting of rational numbers $\mathbb{Q}$ is algebraically closed, since the polynomial $x^2 + 1$ has no real roots.

Nevertheless, starting from the reals it is easy to construct the field of *complex numbers* $\mathbb{C} \supset \mathbb{R}$ in which the polynomial $x^2 + 1$ *does* have a root: we may take the quotient of the univariate polynomial ring $\mathbb{R}[x]$ by its ideal $\langle x^2 + 1 \rangle$:

$$\mathbb{C} = \mathbb{R}[x]/\langle x^2 + 1 \rangle \tag{1}$$

If we write 1 and i for the cosets $1 + \langle x^2 + 1 \rangle$ and $x + \langle x^2 + 1 \rangle$, respectively, you can check that 1 and i span $\mathbb{C}$ as a vector space over $\mathbb{R}$. Moreover, arithmetic with this definition of $\mathbb{C}$ works as you'd expect:

$$\begin{aligned} (a + bi) + (c + di) &= (a + c) + (b + d)i, \\ (a + bi) \cdot (c + di) &= (ac - bd) + (ad + bc)i. \end{aligned} \tag{2}$$

Here is one reason why mathematicians love the complex numbers.

Theorem 1 (Fundamental Theorem of Algebra)**.** $\mathbb{C}$ is algebraically closed.

It's likely you've seen this theorem stated in a high school algebra course. It's also possible you've seen a proof—there are several "standard" arguments based on undergraduate-level mathematics.

Exercise 1. Prove that Theorem 1 is equivalent to the assertion that every $n \times n$ matrix with complex entries has a complex eigenvalue.

I will explain a proof of Theorem 1 that hopefully helps you understand *why* it is true. The proof is essentially constructive, and introduces the main ideas behind *homotopy continuation*, which can be used to numerically solve polynomial equations in one or more variables. It seems that a proof involving similar ideas was known to Weierstraß near the end of the 19th century. More recently, the "Bézout series" of papers by Smale and collaborators have led to a number of results concerning this type of proof.

The key idea, which we will use again and again, is rather simple. For fixed $n \geq 1$, a root r of the monic polynomial $x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in \mathbb{C}[x]$ corresponds to a point $(a_0, \ldots, a_{n-1}, r)$ in a *space of problem-solution pairs*,

$$P_n := \{(a_0, \ldots, a_{n-1}, r) \in \mathbb{C}^{n+1} \mid r^n + a_{n-1}r^{n-1} + \ldots + a_0 = 0\}. \tag{3}$$

Exactly what kind of "space" is P_n? Here are two valid answers:

Answer 1. P_n is a smooth manifold, diffeomorphic to $\mathbb{R}^{2n}$.

Answer 2. P_n is a complex algebraic variety. More precisely, it is a Zariski-closed subset of $(n+1)$-dimensional affine space over $\mathbb{C}$.

In this course, we will cover aspects of algebraic geometry which may allow you to better parse the words appearing in Answer 2. Answer 1 is also useful, but better suited for those who already know something about differential topology.

Our space of problem-solution pairs is naturally equipped with two projections: one onto $\mathbb{C}^n$ where the coefficients live, the other onto $\mathbb{C}$ where the roots live. Let's make the first of these projections explicit:

$$\begin{aligned} \pi_n : P_n &\to \mathbb{C}^n \\ (a_0, \ldots, a_{n-1}, r) &\mapsto (a_0, \ldots, a_{n-1}). \end{aligned}$$

The map π_n allows us to formulate Theorem 1 in more geometric terms as saying, *the map π_n is surjective*, or equivalently, *the fiber $\pi^{-1}(\mathbf{a})$ is nonempty for every $\mathbf{a} \in \mathbb{C}^n$*. Indeed, the map π_n is the proptotypical example of a *branched cover*—for *almost every* $\mathbf{a} \in \mathbb{C}^n$, some small neighborhood $N_{\mathbf{a}} \in \mathbb{C}^n$ containing $\mathbf{a}$ pulls back under π_n to a union of n disjoint neighborhoods in P_n:

$$\pi_n^{-1}(N_{\mathbf{a}}) = \sqcup_{i=1}^n N_{(\mathbf{a}, r_i)}. \tag{4}$$

However, there is a "small" subset of points $\mathbf{a} \in \mathbb{C}^n$ where this property fails. This is called the *branch locus* of the map π_n, also known as the *discriminant locus*. The discriminant locus turns out to be a *hypersurface* in $\mathbb{C}^n$: that is, the set of all points where a single polynomial equation in n variables vanishes. To compute that equation, let

$$\begin{aligned} f(x) &= x^n + a_{n-1}x^{n-1} + \ldots + a_0, \\ f'(x) &= nx^{n-1} + (n-1)a_{n-1}x^{n-2} + \ldots + a_1. \end{aligned}$$

If f and f' have a common root, then they have a common linear factor, which implies there exists a polynomial identity of the form

$$(b_{n-2}x^{n-2} + \cdots + b_0) \cdot f + (c_{n-1}x^{n-1} + \cdots + c_0) \cdot g = 0. \tag{5}$$

When the coefficients of f are given, we can think of eq. (5) as a system of $2n-1$ linear equations in $2n-1$ unknowns—set the coefficients of $x^{2n-2}, x^{2n-2}, \ldots, 1$ all equal to zero. This linear system can be represented in matrix form as

$$\begin{pmatrix} 1 & 0 & \cdots & 0 & n & 0 & \cdots & 0 \\ a_{n-1} & 1 & \cdots & 0 & (n-1)a_{n-1} & n & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ a_0 & a_1 & \ddots & 1 & a_1 & 2a_2 & \ddots & n \\ 0 & a_0 & \ddots & a_{n-1} & 0 & a_1 & \ddots & (n-1)a_{n-1} \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & a_0 & 0 & 0 & \cdots & a_1 \end{pmatrix} \begin{pmatrix} b_{n-2} \\ b_{n-3} \\ \vdots \\ b_0 \\ c_{n-1} \\ c_{2n-2} \\ \vdots \\ c_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{6}$$

The $(2n-1) \times (2n-1)$ coefficient matrix in eq. (6) is the *Sylvester matrix* $S_x(f, f')$. We may define the discriminant of f in x to be $\Delta_{x,f} = \det S_x(f, f')$.

Example 1. When $n = 2$, the space of ps-pairs P_n is a quadric cone in $\mathbb{C}^3$. The discriminant locus in $\mathbb{C}^2$ is the parabola defined by $\Delta_{x,f} = a_1^2 - 4a_0 = 0$.

Example 2. For $n = 3$, the Sylvester matrix and discriminant are as follows:

$$S_x(f, f') = \begin{pmatrix} 1 & 0 & 3 & 0 & 0 \\ a_2 & 1 & 2\,a_2 & 3 & 0 \\ a_1 & a_2 & a_1 & 2\,a_2 & 3 \\ a_0 & a_1 & 0 & a_1 & 2\,a_2 \\ 0 & a_0 & 0 & 0 & a_1 \end{pmatrix}, \quad \Delta_{x,f} = -a_1^2 a_2^2 + 4\,a_0 a_2^3 + 4\,a_1^3 - 18\,a_0 a_1 a_2 + 27\,a_0^2.$$

This can be verified using the code below. We note that Macaulay2 uses a different convention than ours, in which the Sylvester matrix is transposed.

```
d = 3
R = QQ[x,a_0..a_(d-1)]
f = x^d + sum(0..d-1, i -> a_i * x^i)
fx = diff(x,f)
S = sylvesterMatrix(f, fx, x)
det S
discriminant(f, x)
```

Exercise 2. For particular values of its coefficients $a_0, \ldots, a_{n-1} \in \mathbb{C}$, the degree-$n$ polynomial f has a repeated root $r \in \mathbb{C}$ if and only if f and its derivative f' both vanish at r. If either condition holds, then the discriminant $\Delta_{x,f}$ vanishes at the point $(a_0, \ldots, a_{n-1})$.

Exercise 3. The discriminant is a polynomial of degree $2n - 2$ in the coefficients $a_0, \ldots, a_{n-1}$.

Exercise 4. (Assumes Galois theory background) The discriminant of the "factored polynomial" $\prod_{1 \le i \le n}(x - r_i)$ in x is given, up to sign, by the formula $\prod_{1 \le i < j \le n}(r_i - r_j)^2$.

For given n, we define the *discriminant locus* Δ_n to be the set of all points $(a_0, \ldots, a_{n-1}) \in \mathbb{C}^n$ where $\Delta_{x,f}$ vanishes. With our interpretation of P_n is a space of problem-solution pairs branched along Δ_n, let us return to the fundamental theorem of algebra. Computing the roots of a polynomial may sometimes be a trivial problem to solve. For example, we may take the coefficient vector $\mathbf{a}_0 = (0, \ldots, 0, -1)$, which corresponds to a polynomial $x^n - 1$ whose roots are precisely the n-th roots of unity: thus

$$\pi_n^{-1}(\mathbf{a}_0) = \mathbf{a}_0 \times \{\exp(2\pi i/n)\}.$$

Homotopy continuation is the art of deforming the solutions of some trivially solvable problem like $\mathbf{a}_0$ into the solutions of some other problem $\mathbf{a}_1$ that you'd actually like to solve. The key player in making this work is a *homotopy function* whose *solution curves* connect points in the fiber $\pi_n^{-1}(\mathbf{a}_0)$ to those in the fiber $\pi^{-1}(\mathbf{a}_1)$. The simplest example is called the *straight-line homotopy.*

Example 3. Suppose we want to solve the equation

$$g(x) = x^2 - x - 1 = 0.$$

We refer to this as the *target system.* It is given by the parameter point $\mathbf{a}_1 = (-1, -1) \in \mathbb{C}^2$. Our goal is to recover the fiber

$$\pi_n^{-1}(\mathbf{a}_1) = \left\{ \left(\mathbf{a}_1, (1 + \sqrt{5})/2\right), \left(\mathbf{a}_1, (1 - \sqrt{5})/2\right) \right\}$$

We realize this system as the deformation of the *total degree start system*

$$f(x) = x^2 - 1 = 0,$$

whose parameter values $\mathbf{a}_0 = (0, -1)$ correspond to the fiber

$$\pi_n^{-1}(\mathbf{a}_0) = \{(\mathbf{a}_0, 1), (\mathbf{a}_0, -1)\}$$

The straight-line homotopy allows us to interpolate between roots of f and g:

$$H(x, t) = (1 - t)f(x) + tg(x). \tag{7}$$

The key insight is as follows: if we are interested in some *solution function* $x(t)$ satisfying

$$H\left(x(t),t)\right) = 0 \quad \forall t \in [0,1],$$

then basic calculus also implies that

$$\frac{d}{dt} H\Big|_{(x(t),t)} = \left(\frac{dH}{dx} \cdot \frac{dx}{dt} + \frac{dH}{dt}\right)\Big|_{(x(t),t)} = 0 \quad \forall t \in [0,1].$$

Provided that $\frac{dH}{dx}(x(t),t) \neq 0$ for all $t \in [0,1]$, we obtain the differential equation

$$\frac{dx}{dt} = -\left(\frac{dH}{dx}\right)^{-1} \cdot \frac{dH}{dt}. \tag{8}$$

Using the initial values $x(0) = \pm 1$, we can numerically integrate this ODE using predictor/corrector methods. Roughly speaking, if we have a good approximation of the solution function $x(t)$ at some $t \in [0,1]$, we use a predictor subroutine to estimate $x(t+\Delta t)$ for some small timestep $\Delta t > 0$, and then refine our estimate with one or more iterations of a corrector subroutine. A particularly simple predictor routine is based on Euler's method: using 8,

$$\begin{aligned} x(t+\Delta t) &\approx x(t) + \Delta t \cdot \frac{dx}{dt}\Big|_{(x(t),t+\Delta t)} \\ &= x(t) - \Delta t \cdot \left(\left(\frac{dH}{dx}\right)^{-1} \cdot \frac{dH}{dt}\right)\Big|_{(x(t),t+\Delta t)}. \end{aligned} \tag{9}$$

The corrector subroutine is usually *Newton's method*: a single iteration updates the predicted value from 9 using the formula

$$x(t+\Delta t) \leftarrow x(t+\Delta t) - \left(\left(\frac{dH}{dx}\right)^{-1} \cdot H\right)\Big|_{(x(t+\Delta t),t+\Delta t)}.$$

Although homotopy continuation is implemented "under the hood" in Macaulay2 methods like `solveSystem`, it may be useful to see what it looks like to implement a predictor/corrector method "from scratch." Here is a very simple version applied to our example.

```
R = CC[t,x]
f = x^2 - 1
g = x^2 - x - 1
H = (1-t) * f + t * g
Ht = diff(t, H)
Hx = diff(x, H)
(t0, x0) = (0.0_CC, 1.0_CC)
sub(H, matrix{{t0,x0}}) -- the residual is small, so x0 is a valid starting solution
(ti, xi) = (t0, x0)
dt = 1e-2
while ti < 1.0 do (
    ti = min(ti + dt, 1.0);
    -- predict w/ Euler's method
    Hxi = sub(Hx, matrix{{ti, xi}});
    Hti = sub(Ht, matrix{{ti, xi}});
    xti = -Hti / Hxi;
    xi = xi + dt * xti;
    -- correct with Newton's method
    Hxi = sub(Hx, matrix{{ti, xi}});
    Hi = sub(H, matrix{{ti, xi}});
    xi = xi - Hi / Hxi;
```

```
    << "estimate x(" << ti << ") = " << xi << " w/ residual " << sub(H, matrix{{ti, xi}})
    << endl;
)
```

You should try playing around with this code. Here are a few respects in which this numerical routine is naive:

1. In practice, the number of starting solutions is potentially much larger than 2.
2. Euler's method is a simplistic predictor: in practice, it its common to use the fourth-order Runge Kutta method, or an even more sophisticated method.
3. Only one step of Newton's method is used.
4. A fixed stepsize of 10^{-2} is used. In practice, we want to choose this *adaptively* since the solution function could be quite "curvy."
5. We do not check whether or not the printed *solution residuals* remain small for all $t \in [0,1]$, or that the derivatives "remain reasonable."
6. If we want to solve some other target system, then $\left(\frac{dH}{dx}\right)^{-1}$ *need not exist* for all solutions of all systems corresponding to points along the straight-line segment $(1-t)\mathbf{a}_0 + t\mathbf{a}_1$. This occurs precisely when this segment $(1-t)\mathbf{a}_0 + t\mathbf{a}_1$ intersects Δ_n. Notice that this may occur even if $\mathbf{a}_1 \notin \Delta_n$.

To get around the last of these difficulties, we could try to design a path connecting $\mathbf{a}_0$ to $\mathbf{a}_1$ that would avoid Δ_n. If we are ok with a little bit of randomness, then an extremely elegant solution may be found in the so-called *gamma trick*. With this trick, we modify the straight-line homotopy of equation eq. (7) by multiplying the start system by a random complex γ number of modulus one:

$$H(x,t) = \gamma \cdot (1-t)f(x) + tg(x) \tag{10}$$

The importance of this maneuver is that can multiplication by γ puts the starting parameters $\gamma \cdot \mathbf{a}_0$ in "general position" with respect to the target parameters $\mathbf{a}_1$. Along these lines, we have the following proposition.

Proposition 1.1. Suppose, for some fixed target parameters $\mathbf{a}_1 \in \mathbb{C}^n$, that $\Delta_{x,f}(\mathbf{a}_1) \neq 0$. Then, for almost all starting parameters $\mathbf{a}_0 \in \mathbb{C}^n$ with $\Delta_{x,f}(\mathbf{a}_0) \neq 0$, the straight-line segment connecting $\mathbf{a}_0$ and $\mathbf{a}_1$ is disjoint from the discriminant locus for all real t.

Proof. The discriminant locus $\Delta_n \subset \mathbb{C}^n \cong \mathbb{R}^{2n}$ is a hypersurface of complex dimension $\dim_{\mathbb{C}}(\Delta_n) = (n-1)$, and so it has $\dim_{\mathbb{R}}(\Delta_n) = 2n-2$. The union of all real lines connecting $\mathbf{a}_0$ to some point in Δ has dimension $(2n-2)+1 = 2n-1$ in $\mathbb{R}^{2n}$, so it has Lebesgue measure zero. The result holds for all $\mathbf{a}_1$ in the complement of this set. □

The gamma-trick mentioned above is analagous to Proposition 1.1—note however, that $\gamma \cdot \mathbf{a}_1$ does not encode a *monic* polynomial. We will return to this trick later on in a more general setting. In the meantime, Proposition 1.1 already gives us enough for us to prove Theorem 1.

(Proof of Theorem 1): Let $\mathbf{a}_1$ be the parameter values corresponding to some univariate polynomial. For some nearby parameter values $\mathbf{a}_0$, the corresponding polynomial has exactly n complex roots. This can be seen by numerically continuing the roots of the total-degree start system, given by

$$f(x) = x^n - 1 \quad \text{w/ roots} \quad x = e^{2\pi i k/n}, \quad k = 0, \ldots, n-1,$$

to some randomly chosen $\mathbf{a}_0$ in a small Euclidean ball around $\mathbf{a}_1$ using a straight-line homotopy and appealing to Proposition 1.1.

Now consider the segment $(1-t)\mathbf{a}_0 + t\mathbf{a}_1$. Reparametrizing if necessary, we may assume that this segment is disjoint from Δ_n for all $t \in [0,1)$—indeed, this segment intersects Δ_n in at most $n(n-1)$ points. If $\mathbf{a}_1 \notin \Delta_n$, this holds for all $t \in [0,1]$, and the roots corresponding to points in the fiber $\pi_n^{-1}(\mathbf{a}_0)$ can be numerically continued to the roots corresponding to $\pi_n^{-1}(\mathbf{a}_1)$. Otherwise, $\mathbf{a}_1 \in \Delta_n$, in which case we have a root with some multiplicity greather than 1. In the space of problem-solution pairs we may take a sequence $(\mathbf{a}^{(m)}, x^{(m)}) \in P_n$ such that $\mathbf{a}^{(m)} \to \mathbf{a}_1$ lies along this segment. We get a limiting value $x_* = \lim_{m\to\infty} x^{(m)}$ on the Riemann sphere $\mathbb{P}^1_{\mathbb{C}}$, which cannot be a pole of the limiting function $x \mapsto f(x;\mathbf{a}_1)$ since it is a polynomial. Thus x_* is the desired root.

2 Symbolic vs numeric

It is certainly nice to know that polynomials have complex roots. However, we would like to actually to *compute* these roots. There are several basic difficulties: consider, for instance, the roots of $f(x) = x^2 - 2 = 0$. As any good algebraist would do, we could simply define them into existence, extending $\mathbb{Q}$ by the field $\mathbb{Q}[x]/\langle f(x)\rangle$. But this does not help us with basic questions like what the *size* of a root is. Can we actually make sense sense of

$$\sqrt{2} \approx 1.41421356237309504880168872420969807 8, \tag{11}$$

and various other well-known "identities"? In, say, a real analysis course, you might have proved that every real number can be represented in *base-b representation*: for any fixed integer $b \geq 2$, any $r \in \mathbb{R}$ may be written as

$$r = \sum_{i=k}^{\infty} c_i b^{-i} \tag{12}$$

for some $k \in \mathbb{Z}$, with each $c_i \in \{0, 1, \ldots, b-1\}$. One may show that $r \in \mathbb{Q}$ if and only if there exists such a representation with only finitely many $c_i = 0$. When $r = \sqrt{2}$, such a representation would require storing infinitely many c_i, which cannot be done on a computer with finite memory.

Turing had the idea that we could represent $r = \sqrt{2}$ as a computer program that can compute a representation of the form eq. (12) to any desired accuracy. Concretely, we could use Newton's method in rational arithmetic. The same idea allows us to represent any *algebraic number*, provided we know its *minimal polynomial* over $\mathbb{Q}$. Here is a Macaulay2 function giving such a representation:

```
sqrtAppx = eps -> (
    R := QQ[x];
    f := x^2-2;
    df := diff(x,f);
    xHat := 3/2;
    while (xHat^2 - 2) > eps^2 do (
        fxHat := sub(f, x => xHat);
        dfxHat := sub(df, x => xHat);
        xHat = xHat - (1/dfxHat) * fxHat;
        );
    xHat
    )
```

For example, with $\epsilon = 1/1000$ we obtain the rational approximation $\frac{665857}{470832}$. In fact, this agrees with the approximation of eq. (11) to 11 digits.

```
i1 : r = sqrtAppx(1/1000)

     665857
o1 = ------
     470832

o1 : QQ

i2 : sub(r, RR)

o2 = 1.41421356237469

o2 : RR (of precision 53)
```

Exercise 5. Show that for any given input $\epsilon \in \mathbb{Q}_{>0}$, the output of `sqrtAppx` is a rational number $\hat{x} \in \mathbb{Q}$ with $|\hat{x} - \sqrt{2}| < \epsilon$.

Notice that the output of `sqrtAppx` was a rational number, but to obtain an approximate decimal expansion (namely eq. (12) with $b = 10$), it was subsequently converted to a real number data type. In practice, real numbers

are usually represented using *fixed precision floating point arithmetic*. To be more precise, most often what is used is the `IEEE binary64` standard, also called *double-precision floating-point*. This uses numbers of the form

$$\pm 2^e \cdot (1 + \sum_{i=1}^{52} c_i 2^{-i}) \tag{13}$$

where e is an 11-bit integer in the range $[-1022, 1023]$ and $c_1, \ldots, c_k \in \{0, 1\}$. The "64" refers to the $52 + 1 + 11$ bits of information appearing in this number: note that the "leading 1" in eq. (13) does not need to be explicitly stored.

Since the set of such numbers is *not* closed under addition or multiplication, floating-point operations involve rounding. Rounding of individual addition and multiplication operations are close to the "true values." However, the result of multiple floating point operations may be inaccurate, particularly when quantities of large or varying magnitudes are involved. For example,

```
i1 : eps = 10.0^(-16)
o1 = 1e-16
o1 : RR (of precision 53)
i2 : epsPlus = 1.0 + eps
o2 = 1
o2 : RR (of precision 53)
i3 : notEps = epsPlus - 1.0
o3 = 0
o3 : RR (of precision 53)
```

In Macaulay2, it is possible to work with extended-precision real numbers, which are implemented in an external library called MPFR.

```
i1 : defaultPrecision = 54
i2 : eps = 10.0^(-16)
o2 = 1e-16
o2 : RR (of precision 53)
i3 : epsPlus = 1.0 + eps
o3 = 1
o3 : RR (of precision 53)
i4 : epsPlus - 1.0
o4 = 1e-16
o4 : RR (of precision 53)
```

A more dramatic example of cancellation comes from *Wilkinson's polynomial*,

$$p(x) = (x-1)(x-2)(x-3)\cdots(x-20). \tag{14}$$

Although innocent-looking when in factored form, the coefficients of p in the monomial basis are huge. In fact, the coefficients of $x^2, \ldots, x^{10}$ cannot be exactly represented as numbers in the form of eq. (13). Thus, if we were to expand eq. (13) using floating-point arithmetic, the coefficients computed would differ from their true values, and the roots of the resulting polynomial would differ noticeably from the original values.

```
i2 : ringQ = QQ[x]

o2 = ringQ

o2 : PolynomialRing

i3 : wQ = product for i from 1 to 20 list (x-i)

      20       19          18           17             16
o3 = x   - 210x   + 20615x   - 1256850x   + 53327946x

             15                14                 13
-1672280820x   + 40171771630x   - 756111184500x
```

```
                 12                     11                     10
+11310276995381x   - 135585182899530x   + 1307535010540395x

                   9                     8
-10142299865511450x  + 63030812099294896x

                    7                      6
-311333643161390640x  + 1206647803780373360x

                     5                      4
-3599979517947607200x  + 8037811822645051776x

                      3                      2
-1287093124515098880 0x  + 13803759753640704000x

                     1
-8752948036761600000x + 2432902008176640000

o3 : ringQ
i4 : sort roots wQ
o4 = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20}
o4 : List
i5 : ringR = RR[x]
o5 = ringR
o5 : PolynomialRing
i6 : wR = product for i from 1 to 20 list (x-i);
i7 : sort roots wR
o7 = {1, 2, 3, 4, 5, 6, 7, 8.00002, 8.99992, 10.0002, 10.9996, 12.0005, 12.9994, 14.0005,
    14.9996, 16.0002, 16.9999, 18, 19, 20}
o7 : List
```

Floating-point arithmetic is undeniably useful in many domains of computational mathematics, and computational algebraic geometry is no exception. However, the toy example of Wilkinson's polynomial already suggests we might be able to gain something by postponing inexact computation until we really need it. Moreover, a mathematician trying to *prove* something would wisely be skeptical of relying on such inexact calculations. Surprisingly, one can sometimes *certify* the correctness of particular types of inexact computations. Notwithstanding, methods of *symbolic computation*, most prominently *Gröbner bases*, are an extremely effective tool in computational algebraic geometry, which can also be used to prove many foundational results (such as Hilbert's basis theorem (Theorem 2 and Nullstellensatz (Theorem 5).) Thus, we will spend a lot of time studying them.

3 Gröbner bases and normal forms

In general, we are interested in solving problems of the following form

$$f_1(\mathbf{x};\mathbf{p}) = f_2(\mathbf{x};\mathbf{p}) = \cdots = f_s(\mathbf{x};\mathbf{p}) = 0,$$

where $\mathbf{x} \in \mathbb{C}^n$ are *unknowns*, or *variables*, $\mathbf{p} \in \mathbb{C}^m$ are *given data* or *parameters*, and $f_1, \ldots, f_s$ are equations are polynomial functions[1] of $\mathbf{x}$ and $\mathbf{p}$. Here is a simple example with $n > 1$.

Example 4. For rectangle with length x_1 and width x_2, we can easily compute the area p_1 and the perimeter p_2. The *inverse problem* asks: given $\mathbf{p} = (p_1, p_2) \in \mathbb{R}^2$, can we recover $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ satisfying

$$\begin{aligned} f_1(\mathbf{x};\mathbf{p}) &= 2(x_1 + x_2) - p_1 = 0, \\ f_2(\mathbf{x};\mathbf{p}) &= x_1 x_2 - p_2 = 0. \end{aligned} \tag{15}$$

Here is some Macaulay2 code exploring symbolic and numerical solutions to this problem:

[1] Or perhaps even rational/algebraic functions

```
(p1, p2) = (10, 6)
R = QQ[x_1,x_2]
f1 = 2*(x_1+x_2) - p1
f2 = x_1*x_2 - p2
needsPackage "NumericalAlgebraicGeometry"
solveSystem {f1,f2}
I = ideal(f1,f2)
Ielim1 = eliminate(I, x_1)
fx2 = first Ielim1_*
roots fx2
```

The blackbox solver `solveSystem` uses a homotopy continuation procedure analagous to that of the previous section. The command `eliminate` relies on Gröbner bases. To eliminate variables "by hand", we can try to work with various "polynomial consequences of eq. (15). For instance, if $g_1, g_2 \in \mathbb{C}[\mathbf{x}]$ are *arbitrary* polynomials in the unknowns, these equations also imply that

$$g_1(\mathbf{x}) \cdot f_1(\mathbf{x}; \mathbf{p}) + g_2(\mathbf{x}) \cdot f_2(\mathbf{x}; \mathbf{p}) = 0.$$

A fortuitous choice is given by $(g_1, g_2) = (x_2, -2)$, from which we obtain

$$2x_2^2 - p_1 x_2 + 2p_2 = 0.$$

The roots of this univariate polynomial can be computed in radicals:

$$x_2 = \frac{p_1 \pm \sqrt{p_1^2 - 16p_2}}{4} = 2 \text{ or } 3,$$

from which we also easily obtain, using the equation for perimeter,

$$x_1 = p_1/2 - x_1 = \frac{p_1 \mp \sqrt{p_1^2 - 16p_2}}{4} = 3 \text{ or } 2.$$

Our search for "polynomial consequences" in the previous example motivates the following definition.

Definition 3.1. Let $\mathbb{K}$ be a field, and $\mathbb{K}[\mathbf{x}] = \mathbb{K}[x_1, \ldots, x_n]$ be the polynomial ring over $\mathbb{K}$ in n indeterminates. For fixed $f_1, \ldots, f_s \in \mathbb{K}[\mathbf{x}]$, the *ideal generated* by these polynomials is the set

$$\langle f_1, \ldots, f_s \rangle = \left\{ \sum_{i=1}^{s} g_i f_i \mid g_1, \ldots, g_s \in \mathbb{K}[\mathbf{x}] \right\}.$$

Hilbert's basis theorem states that every polynomial ideal has the form given in Definition 3.1. In hopes of reducing multivariate polynomial system solving to univariate polynomial system solving, we pose the *elimination problem*: for a given ideal $\langle f_1, \ldots, f_s \rangle \subset \mathbb{C}[\mathbf{x}]$, how can we compute generators for the ideal $\langle f_1, \ldots, f_s \rangle \cap \mathbb{K}[x_2, \ldots, x_n]$? We will show how a complete solution to this problem can be obtained by computing a Gröbner basis with respect to a *lexicographic order* (Definition 3.3.)

In addition to the elimination problem, we also consider the *ideal membership problem*: for given $f, f_1, \ldots, f_s \in \mathbb{K}[\mathbf{x}]$, can we decide whether or not $f \in \langle f_1, \ldots, f_s \rangle$? In the *univariate case* $n = 1$, it is easy to solve this problem using the *division algorithm.*

Example 5. Let $g = \underline{x^2} - 1$, and consider the ideal $I = \langle g \rangle$. We use the division algorithm to show that $f = \underline{x^4} + x^3 - x - 1 \in I$. Note that we have underlined the terms of highest degree. Anticipating the multivariate case, we write $\mathrm{LT}(g) = x^2$ and $\mathrm{LT}(f) = x^4$ to denote the *leading term* of f and g. The condition $f \in I$ is the same as saying $f \equiv 0 \mod I$, or, since I principal, that I is a polynomial multiple of g. The division algorithm proceeds as

follows:

$$\begin{aligned}
\underline{x^4} + x^3 - x - 1 &= x^2 \cdot \mathrm{LT}(g) + x^3 - x - 1 \\
&= x^2 \cdot \left(g + \underbrace{(\mathrm{LT}(g) - g)}_{1} \right) + x^3 - x - 1 \\
&\equiv x^2 + x^3 - x - 1 \mod I \\
&= \underline{x^3} + x^2 - x - 1 \\
&= x \cdot (x^2 - 1) + x + x^2 - x - 1 \\
&\equiv \underline{x^2} - 1 \mod I \\
&\equiv 0 \mod I.
\end{aligned}$$

There are several obstacles to adapting polynomial division to the *multivariate case* $n > 1$. One obstacle is that most ideals are not principal. Another obstacle is that the concept of a leading term does not extend uniquely. Indeed, the usual ordering of monomials when $n = 1$,

$$1 < x < x^2 < x^3 < \cdots$$

has a number of properties that are easy to take for granted. These properties are crystalized in the following definition. Recall that a *monomial* in $\mathbb{K}[x_1, \ldots, x_n]$ is a polynomial with exactly one term whose coefficient equsls $1_{\mathbb{K}}$. A monomial $x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdots x_n^{\alpha_n}$ may be written more compactly in *multi-index notation* as $\mathbf{x}^{\boldsymbol{\alpha}}$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n_{\geq 0}$ is a lattice point in the positive orthant of $\mathbb{R}^n$. For visualization purposes, it is standard to identity monomials and lattice points (especially when $n = 2$ or 3.)

Definition 3.2. A *monomial order* is any total, multiplicative order $<$ on the set of monomials in $\mathbb{K}[x_1, \ldots, x_n]$ such that 1 is the minimum element.

Exercise 6. There is a unique monomial order on the univariate polynomial ring $\mathbb{K}[x]$.

Though they may seem unmotivated at first, it is worthwhile to build up a repertoire of several different monomial orders. For now, we define two classes of monomials orders that are easy to understand, though not always the most useful.

Definition 3.3. The *lexicographical order* with $x_1 > x_2 > \cdots > x_{n-1} > x_n$, denoted in Macaulay2 by `Lex`, is defined as follows:

$$\mathbf{x}^{\boldsymbol{\alpha}} > \mathbf{x}^{\boldsymbol{\beta}} \quad \Leftrightarrow \quad \boldsymbol{\alpha} - \boldsymbol{\beta} = (0, 0, \ldots, 0, \underbrace{\alpha_i - \beta_i}_{>0}, \ldots).$$

Definition 3.4. The *graded lexicographical order* with $x_1 > x_2 > \cdots > x_{n-1} > x_n$, denoted in Macaulay2 by `GLex`, is defined as follows:

$$\begin{aligned}
\mathbf{x}^{\boldsymbol{\alpha}} > \mathbf{x}^{\boldsymbol{\beta}} \quad \Leftrightarrow \quad & \sum_{i=1}^{n} (\alpha_i - \beta_i) > 0, \quad \textbf{OR} \\
& \sum_{i=1}^{n} (\alpha_i - \beta_i) = 0, \quad \mathbf{x}^{\boldsymbol{\alpha}} >_{\texttt{Lex}} \mathbf{x}^{\boldsymbol{\beta}}.
\end{aligned}$$

In more plain language, `Lex` compares monomials as though they were words in a dictionary, whereas `GLex` compares monomials based on their total degree, breaking any ties with `Lex` as needed. Note that both orders depend on the chosen ordering of the variables: in other words, Definitions 3.3 and 3.4 describe a total of $2n!$ monomial orders on $\mathbb{K}[x_1, \ldots, x_n]$.

An important property of monomial orders is that they are all *well orders*; that is, given $<$ as in Definition 3.2, any nonempty set of monomials has a smallest element with respect to $<$. This can proved using the following special case of Hilbert's basis theorem.

Lemma 3.5 (Gordan's Lemma)**.** Every monomial ideal is finitely generated by monomials. That is, if $I \subset \mathbb{K}[\mathbf{x}] = \mathbb{K}[x_1, \ldots, x_n]$ is an ideal such that every element of I has the form

$$g_1 \mathbf{x}^{\boldsymbol{\alpha}_1} + \cdots + g_s \mathbf{x}^{\boldsymbol{\alpha}_s} \quad \text{w/} \quad \mathbf{x}^{\boldsymbol{\alpha}_1}, \ldots, \mathbf{x}^{\boldsymbol{\alpha}_s} \in I, \tag{16}$$

then $I = \langle \mathbf{x}^{\boldsymbol{\beta}_1}, \ldots, \mathbf{x}^{\boldsymbol{\beta}_k} \rangle$ for some finite subset $\{\mathbf{x}^{\boldsymbol{\beta}_1}, \ldots, \mathbf{x}^{\boldsymbol{\beta}_k}\} \subset I$.

Remark: Lemma 3.5 is sometimes called "Dickson's Lemma", despite the fact that Gordan proved it well before Dickson did.

Proof. Induction on n. If $n = 1$, then I is a principal ideal. Writing $I = \langle p \rangle$, then writing p in the form 16 shows that p is divisible by some $x^k \in I$, so the chain of inclusions

$$\langle x^k \rangle \subset I = \langle p \rangle \subset \langle x^k \rangle$$

shows that $I = \langle x^k \rangle$ is finitely generated by a single monomial.

For $n > 1$, assume the result for all smaller n. Define for each $j \in \mathbb{Z}_{\geq 0}$ the monomial ideal

$$I_j = \langle x_1^{\alpha_1} \cdots x_{n-1}^{\alpha_{n-1}} \mid x_1^{\alpha_1} \cdots x_{n-1}^{\alpha_{n-1}} x_n^j \in I \rangle.$$

By inductive hypothesis, each I_j is finitely generated by monomials. Moreover, since we have an ascending chain $I_0 \subset I_1 \subset I_2 \subset \cdots$, the union $\cup_{k \geq 0} I_k$ is also an ideal that is finitely generated by monomials. This implies that the ascending chain eventually stabilizes: that is, there exists some r such that $I_r = I_{r+k}$ for all $k \geq 0$. It follows that $\mathbf{x}^{\boldsymbol{\alpha}} \in I$ iff $x_1^{\alpha_1} \cdots x_{n-1}^{\alpha_{n-1}} \in I_r$. If $B_0, \ldots, B_r$ are monomial generating sets for $I_0, \ldots, I_r$, then it follows that

$$I = \langle B_0 \sqcup x_n B_1 \sqcup x_n^2 B_2 \sqcup \cdots \sqcup x_n^r B_r \rangle,$$

since monomial in I (and consequently, also every polynomial in I) belongs to the ideal on the right. □

Corollary 3.6. Every monomial order is a well-order.

Proof. Let $S = \{\mathbf{x}^{\boldsymbol{\alpha}_i} \mid i \in I\}$ be a nonempty set of monomials. Gordan's lemma implies there exist monomials $\mathbf{x}^{\boldsymbol{\beta}_1}, \ldots, \mathbf{x}^{\boldsymbol{\beta}_s}$ such that

$$\langle S \rangle = \langle \mathbf{x}^{\boldsymbol{\beta}_1}, \ldots, \mathbf{x}^{\boldsymbol{\beta}_s} \rangle.$$

Moreover, since each $\mathbf{x}^{\boldsymbol{\beta}_i}$ is divisible by some element of S, we may assume WLOG that $\mathbf{x}^{\boldsymbol{\beta}_i} \in S$ for all i. Since any element of S is divisible by one of these monomial generators, the smallest generator with respect to $<$ is the minimum element of S. □

For any fixed monomial order $<$ on $\mathbb{K}[\mathbf{x}]$ and any nonzero polynomial $f \in \mathbb{K}[x]$, we may write

$$f = c_1 \mathbf{x}^{\boldsymbol{\alpha}_1} + c_2 \mathbf{x}^{\boldsymbol{\alpha}_2} + \cdots + c_k \mathbf{x}^{\boldsymbol{\alpha}_k}$$

with its coefficients in sorted order, i.e.

$$\mathbf{x}^{\boldsymbol{\alpha}_k} < \cdots < \mathbf{x}^{\boldsymbol{\alpha}_2} < \mathbf{x}^{\boldsymbol{\alpha}_1}.$$

The leading term/coefficient/monomial of f with respect to $<$ are then defined as follows:

$$\begin{aligned} \mathrm{LT}_<(f) &= c_1 x^{\boldsymbol{\alpha}_1}, \\ \mathrm{LC}_<(f) &= c_1, \\ \mathrm{LM}_<(f) &= x^{\boldsymbol{\alpha}_1}. \end{aligned}$$

When $f = 0$, those notions are left undefined. When $\mathrm{LC}_<(f) = 1$, we say f is *monic* with respect to $<$.

Emulating the pattern of Example 5, let us try to solve the ideal membership problem on an example, using some of the monomial orders introduced so far.

Example 6. Consider the `Lex` order on $\mathbb{Q}[x, y, z]$ with $x < y < z$, and let

$$\begin{aligned} f &= \underline{z^2} - y \\ f_1 &= \underline{y} - x, \\ f_2 &= \underline{z^2} - x. \end{aligned}$$

We would like to decide the ideal membership query

$$f \in I = \langle f_1, f_2 \rangle?$$

We begin by trying to divide $\mathrm{LT}_<(f)$ by $\mathrm{LT}_<(f_1)$ or $\mathrm{LT}_<(f_2)$—if that succeeds, then we can write

$$f = \mathbf{x}^{\boldsymbol{\alpha}} \cdot f_i + \tilde{f}$$

for some $\tilde{f}$ with strictly smaller leading monomial: $\mathrm{LM}_<(\tilde{f}) < \mathrm{LM}_<(f)$. Applying the same procedure with $\tilde{f}$ in place of f, we obtain the following sequence of operations:

$$\begin{aligned} f &= \underline{z^2} - y \\ &= \underbrace{(z^2 - x)}_{f_2} + x - y \\ &\equiv \underline{-y} + x \mod I \\ &= -f_1 \\ &\equiv 0 \mod I. \end{aligned}$$

This calculation produces a certificate of ideal membership in the form of the multipliers $(g_1, g_2) = (1, -1)$ appearing in Definition 3.1:

$$f = 1 \cdot f_1 + (-1) \cdot f_2 \in I.$$

Now suppose instead that we chose the Lex order with the order of variables reversed: $x > y > z$. Our ideal-membership query is the same (up to sign) as before), but we have different leading terms:

$$\underline{y} - z^2 \in I = \langle \underline{x} - y, \underline{x} - z^2 \rangle?$$

Applying the same algorithm as before, we see that $\mathrm{LT}_<(f)$ is not divisible by $\mathrm{LT}_<(f_1)$ or $\mathrm{LT}_<(f_2)$, so we do not succeed in our strategy of rewriting f as an element of I. Notice how, in the previous case, the leading monomials $\mathrm{LM}(f_1), \mathrm{LM}(f_2)$ function like "pivots" in the familiar algorithm of Gaussian elimination. When we change the monomial order in this example, the number of these "pivots" drops from 2 to 1! Fortunately, this is not a deficiency of the monomial order, but rather of the *generating set* used to represent I. Indeed, if we were to discover independently that $f \in I$, we could add it to our set of generators for I, thus obtaining a new leading term $\underline{y^2}$. The definition of a Gröbner basis captures in precise terms when a generating set of an ideal has "enough" leading terms to make the division algorithm work.

Definition 3.7. Fix a monomial order $<$ on $\mathbb{K}[\mathbf{x}]$, and let $I \subset \mathbb{K}[\mathbf{x}]$ be an ideal. The *initial ideal* of I with respect to $<$ is defined as follows:

$$\mathrm{in}_<(I) = \langle \mathrm{LM}_<(f) \mid f \in I \rangle. \tag{17}$$

A *Gröbner basis* $G = \{g_1, \ldots, g_s\} \subset I$ with respect to $<$ is a finite subset of I whose leading monomials generate the initial ideal: $\mathrm{in}_<(I) = \langle \mathrm{LM}_<(g_1), \ldots, \mathrm{LM}_<(g_s) \rangle$.

Example 7. Continuing with Example 6, consider the following Macaulay2 session:

```
i1 : R = QQ[z,y,x];
i2 : f = z^2 - y;
i3 : f1 = y-x;
i4 : f2 = z^2 - x;
i5 : I = ideal(f1, f2);
o5 : Ideal of R
i6 : G = gb I
o6 = GroebnerBasis[status: done; S-pairs encountered up to degree 1]
o6 : GroebnerBasis
i7 : gens G
o7 = | y-x z2-x |
```

It seems that $\{f_1, f_2\}$ is a Gröbner basis for I, but with respect to which monomial order? The following comparisons rule out the possibility of Lex or GLex.

```
i8 : y < z
o8 = true
i9 : x < y -- so z > y > x
o9 = true
```

```
i10 : x^2 < y -- not Lex!
o10 = false
i11 : y^2 < z*x -- not GLex!
o11 = false
```

As it turns out, any object of class `PolynomialRing` such as `R` in this example represents not just a polynomial ring, but a polynomial ring together with several pieces of satellite data, including a monomial order. The mystery monomial order, used by default in Macaulay2, is revealed to be `GRevLex`.

```
i12 : describe R
o12 = QQ[z, y, x, Degrees => {3:1}, Heft => {1},
MonomialOrder => {MonomialSize => 32}, DegreeRank => 1]
                 {GRevLex => {3:1}  }
                 {Position => Up    }
```

Definition 3.8. The *graded reverse lexicographical order* with $x_1 > x_2 > \cdots > x_{n-1} > x_n$, denoted in Macaulay2 by `GRevLex`, is defined as follows:

$$\mathbf{x}^{\boldsymbol{\alpha}} > \mathbf{x}^{\boldsymbol{\beta}} \quad \Leftrightarrow \quad \sum_{i=1}^{n}(\alpha_i - \beta_i) > 0, \quad \textbf{OR}$$
$$\sum_{i=1}^{n}(\alpha_i - \beta_i) = 0, \quad \boldsymbol{\alpha} - \boldsymbol{\beta} = (\ldots, \underbrace{\alpha_i - \beta_i}_{<0}, \ldots, 0)$$

Thus `GRevLex` first compares monomials by total degree, then breaks ties by picking the greater monomial to be the one with the *smaller* power of x_n, then breaking further ties using x_{n-1}, and so on.

To compute Gröbner bases using the `Lex` orders considered originally in Example 6, we must specify these orders manually:

```
i8 : S = newRing(R, MonomialOrder => Lex);
i9 : gens gb sub(I, S)
o9 = | y-x z2-x |

             1       2
o9 : Matrix S  <--- S
i10 : T = QQ[reverse gens R, MonomialOrder => Lex];
i11 : gens gb sub(I, T)
o11 = | y-z2 x-z2 |

              1       2
o11 : Matrix T  <--- T
```

Before explaining how to compute Gröbner bases in the next section, we will show that they lead to a simple, constructive proof of Hilbert's basis theorem, and that they enable us to solve both the ideal membership problem and the elimination problem. To begin, we observe that a Gröbner basis, *a priori* only a subset of some ideal, is in fact a generating set for that ideal.

Proposition 3.9. Let G be a Gröbner basis for I. Then G generates I.

Proof. Suppose not—then, since $\langle G \rangle \subsetneq I$, there exists a polynomial $f \in I \setminus \langle G \rangle$. Appealing to Corollary 3.6, we may choose such an f with $\mathrm{LM}_<(f)$ minimal. Then, since G is a Gröbner basis, we have $\mathrm{LT}_<(f) = cm\,\mathrm{LT}_<(g)$ for some $g \in G$, $c \in \mathbb{K}$ and monomial m. If we set $\tilde{f} = f - cmg$, then we have $\tilde{f} \in I$ and $\mathrm{LM}_<(\tilde{f}) < \mathrm{LM}_<(f)$, contradicting the minimality of f. □

Proposition 3.9 leads directly to a cornerstone result in commutative algebra.

Theorem 2 (Hilbert's Basis Theorem)**.** Every ideal in $\mathbb{K}[\mathbf{x}]$ is finitely generated.

Proof. Let $I \subset \mathbb{K}[\mathbf{x}]$ be any ideal. Gordan's lemma (Lemma 3.5) implies that I has a Gröbner basis $G = \{g_1, \ldots, g_s\}$. Thus $I = \langle g_1, \ldots, g_s \rangle$ by Proposition 3.9. □

Exercise 7. Show that every ascending chain of ideals in $\mathbb{K}[\mathbf{x}]$ stabilizes. That is, if we have ideals $I_1, I_2, \ldots$ in this ring with

$$I_1 \subset I_2 \subset \cdots,$$

then there exists some $n \in \mathbb{Z}_{\geq 0}$ such that for all $m \in \mathbb{Z}_{\geq 0}$ we have $I_n = I_{n+m}$.

An important property of the univariate division algorithm as that the remainder and quotient representation are *unique*. The next example illustrates some subtleties in the multivariate case.

Example 8. As in Example 6, let $I = \langle \underline{x} - y, \underline{x} - z^2 \rangle$, with the `Lex` $x > y > z$ order. Division of $f = x$ by the given generators depends on how they are ordered: we could get a "remainder" of y or z^2, depending on the order in which we test the divisibility of $\mathrm{LM}_<(f)$ by the leading monomials of the generators.

Thus, in general, the quotient and remainder when we try to divide a polynomial by a generating set of an ideal are not unique. However, no such ambiguity can arise when the generators form a Gröbner basis.

Proposition 3.10. Fix a monomial order $<$ and an ideal $I \subset \mathbb{K}[\mathbf{x}]$. Then any $f \in \mathbb{K}[\mathbf{x}]$ has a unique *normal form* $\mathrm{NF}_{I,<}(f) \in \mathbb{K}[\mathbf{x}]$ such that $f - \mathrm{NF}_{I,<}(f) \in I$ and no monomial of $\mathrm{NF}_{I,<}(f)$ is contained in $\mathrm{in}_<(I)$.

The monomials not contained in $\mathrm{in}_<(I)$ are called the *standard monomials* for I with respect to $<$.

Proof. For the existence statement, let G be a Gröbner basis for I. For any polynomial f, we can run the naive division algorithm, first rewriting any term of f that is divisible by $\mathrm{in}_<(g)$ for some $g \in G$. This terminates in finitely many steps by Gordan's lemma, and we are left with a remainder which is either 0 or whose leading monomial is standard. Continuing in this way for any non-leading terms in f, we obtain a remainder r which is either 0 or such that *all* monomials in r are standard.

For uniqueness, suppose r, r' are both such that $f - r, f - r' \in I$ and r, r' are in the span of standard monomials. This implies $r - r' \in I$ is also in the span of the standard monomials. We cannot have $r \neq r'$, since this would imply that $\mathrm{LM}_<(r - r') \in \mathrm{in}_<(I)$ is standard. □

Example 9. Let $I \subset S = \mathbb{Q}[r_{11} \ldots r_{33}]$ be the ideal defining al 3×3 orthogonal matrices: that is,

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \quad , \quad I = \langle f_1, \ldots, f_9 \rangle = \langle \text{entries of } RR^T - I_{3\times 3} \rangle.$$

For any monomial order, the normal form of $f = (\det R)^2$ is 1. This can be computed using the operator `%`.

```
i1 : S = QQ[r_(1,1)..r_(3,3)];
i2 : R = genericMatrix(S,3,3);

             3       3
o2 : Matrix S  <--- S
i3 : I = ideal(R * transpose R - id_(S^3));
o3 : Ideal of S
i4 : f = (det R)^2;
i5 : f % I
o5 = 1
o5 : S
```

Similarly, you can use the operator `//` find coefficients $h_1, \ldots, h_9 \in S$ expressing $f = \sum_{i=1}^{9} h_i f_i + 1$.

Exercise 8. Show that the mapping from $\mathbb{K}[\mathbf{x}]$ into itself that associates a polynomial with its normal form is $\mathbb{K}$-linear. Can you describe its image and kernel?

MEMBERSHIP (f, I)

1. Compute a Gröbner basis G for I wrt. some monomial order $<$,
2. Let $r = \mathrm{NF}_{I,<}(f)$, computed using the division algorithm and G from the first step.
3. Output YES if $r = 0$ and NO otherwise.

Figure 1: An algorithm for deciding ideal membership $f \in I$.

The normal form furnishes a simple algorithm that solves the ideal membership problem. This algorithm is described in Figure 1. Its correctness follows from Proposition 3.10. To make it effective, all that we need is a procedure for computing the Gröbner basis G in step 1. This can be done using *Buchberger's algorithm*, given in Figure 2.

It is important to note that Gröbner bases are not unique: indeed, if G is a Gröbner basis for I, we can add in more polynomials in I and still have a Gröbner basis. However, if and when we need uniqueness, we may appeal to the notion of a *reduced* Gröbner basis.

Definition 3.11. We say a Gröbner basis G is *reduced* if every element of $g \in G$ is monic, all non-leading monomials of g are standard, and the set $\{\mathrm{LM}_<(g) \mid g \in G\}$ minimally generates $\mathrm{in}_<(I)$: that is, no proper subset of the leading monomials generates $\mathrm{in}_<(I)$.

Proposition 3.12. For any ideal $I \subset \mathbb{K}[\mathbf{x}]$ and monomial order $<$, there exists a unique reduced Gröbner basis for I with respect to $<$.

Proof. To get a reduced Gröbner basis from an arbitrary Gröbner basis G, replace every polynomial in G with its normal form and remove any normal forms that equal zero. For uniqueness, suppose G and G' are two reduced Gröbner bases for I. Then for any $g \in G$ there exists a $g' \in G'$ such that $\mathrm{LT}_<(g) = \mathrm{LT}_<(g')$, and reducedness implies that $g - g'$ is its own normal form. On the other hand, $g - g' \in I$, so we must have $g = g'$. □

The commands `gb` and `groebnerBasis` produce "almost-reduced" Gröbner bases in the sense that the generators might not be monic, but the other conditions of Definition 3.11 are satisfied.

Finally, we address the elimination problem. As it turns out, there is a wide class of *elimination orders* that can be useful for this task. In what follows, we consider polynomial rings in which the variables form two "groups." Generalizing to the case of more than two groups is straightforward.

Definition 3.13. Consider a polynomial ring $\mathbb{K}[\mathbf{x}, \mathbf{y}] = \mathbb{K}[x_1, \ldots, x_n, y_1, \ldots, y_m]$. We say that $<$ is an *elimination order* with $\mathbf{x} > \mathbf{y}$ if any monomial involving a single variable from $\mathbf{x}$ alone is greater than all monomials in $\mathbf{y}$ alone.

It may help to think as variables in the group $\mathbf{x}$ as being "expensive" and variables in $\mathbf{y}$ as being "cheap". The normal form maps defined by an elimination order try to rewrite "expensive" monomials in terms of "cheap" ones. For example, the `Lex` order on $\mathbb{K}[x_1, \ldots, x_n]$ with $x_1 > \cdots > x_n$ is an elimination order with respect to the grouping $\mathbf{x} = \{x_1\}$, $\mathbf{y} = \{x_2, \ldots, x_n\}$. For the singleton grouping $\mathbf{x}_1 = \{x_1\}, \ldots, \mathbf{x}_n = \{x_n\}$, this `Lex` is also an elimination order with $\mathbf{x}_1 > \cdots > \mathbf{x}_n$.

Theorem 3. Let $I \subset \mathbb{K}[\mathbf{x}, \mathbf{y}]$ be an ideal and $<$ an elimination order with $\mathbf{x} > \mathbf{y}$. Suppose G is a Gröbner basis for I with respect to $<$. Then $G_{\mathbf{y}} = G \cap \mathbb{K}[\mathbf{y}]$ is a Gröbner basis for the elimination ideal $I_{\mathbf{y}} = I \cap \mathbb{K}[\mathbf{y}]$. Moreover, if G is reduced, then $G_{\mathbf{y}}$ is also reduced.

Proof. If $f \in I_{\mathbf{y}}$, then $\mathrm{LM}_<(f)$ must be by divisible $\mathrm{LM}_<(g)$ for some $g \in G$. Since $\mathrm{LM}_<(f) \in \mathbb{C}[\mathbf{y}]$, we must have $\mathrm{LM}_<(g) \in \mathbb{C}[\mathbf{y}]$ as well. The fact that $<$ is an elimination order then implies that $g \in \mathbb{C}[\mathbf{y}]$. Thus, for the order on $\mathbb{C}[\mathbf{y}]$ induced by $<$, we see that $G_{\mathbf{y}}$ is a Gröbner basis. When G is reduced, reducedness of $G_{\mathbf{y}}$ follows straightforwardly from Definition 3.11. □

Example 10. If we want to know all polynomial relations on the set of 2×2 minors of the $2 \times n$ matrix

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ x_{21} & \cdots & x_{2n} \end{pmatrix},$$

we should first form an ideal with $\binom{n}{2}$ generators in a ring with $2n + \binom{n}{2}$ variables, namely

$$I = \langle y_S - \det(X_S) \mid S \subset [n],\ \#S = 2 \rangle \subset \mathbb{Q}[\mathbf{x}, \mathbf{y}],$$

BUCHBERGER $(I, <)$:

1. Initialize:
   1. A set of unprocessed S-pairs, S-pairs $= \{(f_1, f_2), \ldots, (f_{s-1}, f_s)\}$
   2. A partial Gröbner basis, $G = \{f_1, \ldots, f_s\}$
2. while $\exists$ an unprocessed S-pair, $(f, p) \in$ S-pairs:
   i. $h \leftarrow S_{f,p}$
   ii. while $\exists g \in G$, terms t, t_h w/ t_h a term of h and $t_h = t \cdot \mathrm{LT}_<(g)$:
   update $h \leftarrow h - t \cdot g$
   iii if $h \neq 0$
   update $G \leftarrow G \cup \{h\}$
   update unprocessed S-pairs, S-pairs $= (\text{S-pairs} \setminus \{(f, p)\}) \cup \{(g, h) \mid g \in G\}$
3. Output G

Figure 2: Buchberger's algorithm for computing a Gröbner basis of an ideal $I = \langle f_1, \ldots, f_s \rangle$ in a polynomial ring $\mathbb{K}[\mathbf{x}]$ with respect to a monomial order $<$.

and then compute the elimination ideal for an appropriate elimination order. The code below does exactly this for $n = 9$ using one of the so-called *block* or *product* orders. This is a monomial order that compares monomials using `GRevLex` in the variables $\mathbf{x}$ first and then breaks ties using `GRevLex` in the variables in $\mathbf{y}$. We see that the reduced Gröbner basis G for I has 330 elements. For the elimination ideal $I_{\mathbf{y}}$, we have a reduced Gröbner basis $G_{\mathbf{y}}$ of cardinality 126. What happens if you use `Lex` instead?

```
n = 9
R = QQ[x_(1,1)..x_(2,n), apply(subsets(n,2), S -> y_S), MonomialOrder => Eliminate(2*n)]
X = transpose genericMatrix(R,n,2)
I = ideal apply(subsets(n,2), S -> y_S - det X_S)
elapsedTime G = gens gb I;
```

4 Buchberger's algorithm

Suppose we are given a polynomial ideal specified by a finite set of generators: $I = \langle f_1, \ldots, f_s \rangle$. We would like to compute a Gröbner basis for I with respect to a particular monomial order $<$. In particular, this will allow us to determine whether or not the original generators form a Gröbner basis. To make progress towards computing a Gröbner basis, we need to generate leading terms that aren't already in the ideal $< \mathrm{in}_<(f_1), \ldots, \mathrm{in}_<(f_s) \rangle$. One way to do this is to take a pair (f_i, f_j) and cancel leading terms by producing the following element of I:

$$S_{f_i,f_j} = \frac{\mathrm{lcm}(\mathrm{LM}_<(f_i), \mathrm{LM}_<(f_j))}{\mathrm{LT}_<(f_i)} \cdot f_i - \frac{\mathrm{lcm}(\mathrm{LM}_<(f_i), \mathrm{LM}_<(f_j))}{\mathrm{LT}_<(f_j)} \cdot f_j. \tag{18}$$

Equation (18) is called the *S-polynomial* associated to the *S-pair* (f_i, f_j). If you look back at examples 4 and 6, you will see that these calculations were really computing S-pairs in disguise. A more systematic procedure generalizing these examples can be found in Figure 2. This is called *Buchberger's algorithm.*

Buchberger's algorithm may be summarized as follows. For each of the possibile S-pairs, we apply a division procedure analogous to that described in Proposition 3.10. If h is the polynomial obtained from $S_{f_i f_j}$ in step 2.ii., we say $S_{f_i f_j}$ *reduces* to h. In fact, many authors would define the normal form $\mathrm{NF}_{G,<}(h)$ with respect to an *ordered* set G as the ouput of this procedure. With that definition, we would then have $\mathrm{NF}_{I,<} = \mathrm{NF}_{G,<}$ precisely when G is a Gröbner basis (regardless of how we order the elements of G.) If some S-pair reduces to a nonzero polynomial h, we add h to our partial Gröbner basis, and we now need to reduce further S-pairs involving h. Once all S-pairs are processed, our partial Gröbner basis is, in fact, a Gröbner basis. The following theorem establishes this fact, and much more.

Theorem 4. Fix $G = \{g_1, \ldots, g_s\} \subset \mathbb{K}[\mathbf{x}]$ and $<$ a monomial order. The following are equivalent:

1. G is a Gröbner basis with respect to $<$
2. Buchberger's algorithm run on $(<, \langle G \rangle)$ outputs G.
3. Every S-polynomial formed from G has a *standard representation*: that is, whenever $1 \leq i < j \leq s$ we can write
$$S_{g_i,g_j} = \sum_{k=1}^{s} h_k g_k \tag{19}$$
where $\mathrm{LM}_<(h_k g_k) \leq \mathrm{LM}_<(S_{g_i g_j})$ for all k with $h_k g_k \neq 0$.
4. Every S-polynomial formed from G has a *lcm representation*: that is, whenever $1 \leq i < j \leq s$ we can write
$$S_{g_i,g_j} = \sum_{k=1}^{s} h_k g_k \tag{20}$$
where $\mathrm{LM}_<(h_k g_k) < \mathrm{lcm}(\mathrm{LM}_<(g_i), \mathrm{LM}_<(g_j))$ for all k with $h_k g_k \neq 0$.

Here is a (unrealistically simple) example of Buchberger's algorithm in action:

Example 11. Let $f_1 = \underline{x^2}, f_2 = \underline{xy} + y^2$. We use the `Lex` order with $x > y$. We compute

$$\begin{aligned} S_{f_1 f_2} &= y f_1 - x f_2 \\ &= -\underline{xy^2} && \text{(divisible by } \mathrm{LM}_<(f_2)\text{)} \\ &= -f_2 + \underline{y^3}. \end{aligned}$$

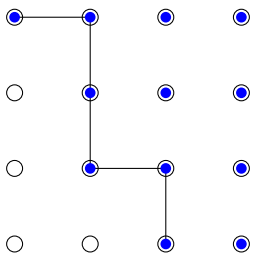
Since y^2 is not divisible by $\mathrm{LM}_<(f_1)$ or $\mathrm{LM}_<(f_2)$, we set $f_3 = y^2$, and set $G = \{f_1, f_2, f_3\}$. Now we have two more S-polynomials to check:

$$S_{f_1 f_3} = y^3 x^2 - x^2 y^3 = 0,$$

and

$$\begin{aligned} S_{f_2 f_3} &= y^2 f_2 - x f_3 \\ &= \underline{y^4} && \text{(divisible by } \mathrm{LM}_<(f_3)\text{)} \\ &= y f_2 + 0. \end{aligned}$$

Theorem 4 implies G is a Gröbner basis, and $\mathrm{in}_<(I) = \langle x^2, xy, y^3 \rangle$. The standard monomials $1, x, y, y^2$ can be visualized as the lattice points in $\mathbb{Z}^2_{\geq 0}$ below a "staircase" formed by the generators of the initial ideal.



Proposition 4.1 establishes that Buchberger's algorithm *terminates* in finite time. Combined with Theorem 4, it's straightforward to see that the output G forms a Gröbner basis, since the S-pairs formed from G are among the (potentially very large) set of S-pairs that are processed.

Proposition 4.1. For any input $(I, <)$, Buchberger's algorithm (Figure 2) terminates after finitely-many steps.

Proof. To see that Buchberger's algorithm terminates, let

$$I_1 = \langle \mathrm{LM}_<(f_1), \ldots, \mathrm{LM}_<(f_s) \rangle.$$

Note that $I_1 \subset \mathrm{in}_<(I)$, and that the inclusion is strict iff $\{f_1, \ldots, f_s\}$ is not a Gröbner basis with respect to $<$. If h is the result of reducing some S-pair when running the algorithm, set $I_2 = I_1 + \langle \mathrm{LM}_<(h) \rangle$. Constructing $I_3, I_4, \ldots$ in a similar way, we obtain an ascending chain of monomial ideals which must stabilize by Exercise 7. Suppose the chain stabilizes after processing n S-pairs, and consider the reduction of any subsequent S-pair. This will be some polynomial h with $\mathrm{LM}_<(h) \in I_{n+1}$. We claim $h = 0$; if not, then we would have $\mathrm{LM}_<(h) \notin I_n$, however $I_n = I_{n+1}$. Thus, after n steps, all remaining S-polynomials reduce to zero. □

Proof of Theorem 4. (1) ⇒ (2): If h is the result of reducing any S-pair formed from G, we must show that h is zero. If that were not the case, then we would have, just as in the proof of termination, that $\mathrm{LM}_<(h)$ was not divisible by any $\mathrm{LM}_<(g_i)$, contradicting the fact that G is a Gröbner basis.

(2) ⇒ (3): Suppose we were to trace the "quotients" produced in each reduction step (step 2.ii) of Buchberger's algorithm. Since we assume each S-pair reduces to zero, this would give us a representation

$$S_{g_i,g_j} = \sum_{k=1}^{s} h_k g_k.$$

If $h_k \neq 0$, then h_k is a sum of polynomials whose leading terms have the form $t_{i,j}/\mathrm{LM}_<(g_k)$ for some term $t_{ij} < \mathrm{LM}_<(S_{g_i g_j})$, thus showing that this representation is standard.

(3) ⇒ (4): Every standard representation is also an lcm representation.

(4) ⇒ (1): Proof by contradiction. Let $f \in \langle G \rangle$, and suppose that $\mathrm{LM}_<(f)$ is not divisible by $\mathrm{LM}_<(g)$ for any $g \in G$. Consider the following representation of f as an element of $\langle G \rangle$:

$$f = \sum_{j=1}^{s} h_s g_s. \tag{21}$$

Without loss of generality, we may assume the $h_s g_s$ are sorted by leading monomial[2],

$$\mathrm{LM}_<(h_s g_s) \leq \mathrm{LM}_<(h_{s-1} g_{s-1}) \leq \cdots \leq \mathrm{LM}_<(h_{\mu+1} g_{\mu+1}) < \mathrm{LM}_<(h_\mu g_\mu) = \mathrm{LM}_<(h_{\mu-1} g_{\mu-1}) = \cdots = \mathrm{LM}_<(h_1 g_1).$$

We choose a representation 21 such that $\mathrm{LM}_<(h_1 g_1)$ is minimal, and further such that the number μ of leading monomials is also minimal.

If $\mu = 1$, then $\mathrm{LM}_<(h_1 g_1)$ occurs as a monomial of some $h_s g_s$ iff $s = 1$. Thus $\mathrm{LM}_<(f) = \mathrm{LM}_<(h_1 g_1)$, which implies $\mathrm{LM}_<(g_1)$ divides $\mathrm{LM}_<(f)$, a contradiction.

Since $\mu > 1$, we may consider the monomial

$$m = \frac{\mathrm{LM}_<(h_1 g_1)}{\mathrm{lcm}(\mathrm{LM}_<(g_1), \mathrm{LM}_<(g_2))} = \frac{\mathrm{LM}_<(h_2 g_2)}{\mathrm{lcm}(\mathrm{LM}_<(g_1), \mathrm{LM}_<(g_2))}. \tag{22}$$

In particular, for some $c \in \mathbb{K}$ we may write

$$\mathrm{LT}_<(h_1)\,\mathrm{LT}_<(g_1) = cm\,\mathrm{lcm}(\mathrm{LM}_<(g_1), \mathrm{LM}_<(g_s)). \tag{23}$$

Now consider an lcm representation of $S_{g_1 g_2}$,

$$S_{g_1 g_2} = \sum_{k=1}^{s} \hat{h}_k g_k, \quad \text{where } \hat{h}_k g_k \neq 0 \;\Rightarrow\; \mathrm{LM}_<(\hat{h}_k g_k) < \mathrm{lcm}(\mathrm{LM}_<(g_1), \mathrm{LM}_<(g_2)). \tag{24}$$

Multiplying this equation by cm and then subtracting $cm\, S_{g_1 g_2}$ from both sides, we obtain (using 22) a representation of 0 as an element of $\langle G \rangle$,

$$0 = \Big(cm\hat{h}_1 - \mathrm{LT}_<(h_1)\Big)\, g_1 + \Big(cm\hat{h}_2 + c'\,\mathrm{LT}_<(h_2)\Big)\, g_2 + \sum_{k=3}^{s} (cm\hat{h}_s)\, g_s, \tag{25}$$

[2]In this case, if $h_i g_s = 0$, we should take $\mathrm{LM}_<(h_i g_i) = 1$.

where $c' \in \mathbb{K}$ may depend on c and $\mathrm{LC}_<(g_2)$. Adding 25 to 21, we obtain a new representation of f as an element of $\langle G \rangle$. For this representation, observe that

$$\begin{aligned}
\mathrm{LM}_< \left(\left(h_1 - \mathrm{LT}_<(h_1) + cm\hat{h}_1 \right) g_1 \right) &\leq \max \left((h_1 - \mathrm{LT}_<(h_1)) \, \mathrm{LM}_<(g_1), \, m \, \mathrm{LM}_<(\hat{h}_1 g_1) \right) \\
&< \max \left(\mathrm{LM}_<(h_1 g_1), \, m \, \mathrm{lcm}(\mathrm{LM}_<(g_1), \mathrm{LM}_<(g_2)) \right) \qquad \text{(using 24)} \\
&= \mathrm{LM}_<(h_1 g_1) \qquad \text{(using 23.)}
\end{aligned}$$

Similarly, one may show

$$\begin{aligned}
\mathrm{LM}_< \left(\left(h_2 - c' \, \mathrm{LT}_<(h_2) + cm\hat{h}_2 \right) g_2 \right) \leq \mathrm{LM}_<(g_2 h_2), \quad \text{strict iff } c = c', \\
\mathrm{LM}_< \left(\left(h_s + cm\hat{h}_s \right) g_s \right) \leq \mathrm{LM}_<(h_s g_s) \quad \forall s \geq 3.
\end{aligned}$$

Thus, for this new representation, we have either fewer leading monomials, or if $\mu = 2$ and $c' = c$, a smaller leading monomial. In either case, this contradicts the minimality of 21. □

A weakness of Buchberger's algorithm is that it spends a huge amount of time reducing *superfluous* S-pairs which can ultimately be reduced to 0. Thus, it is a huge advantage to be able to predict in advance when this will occur. This leads naturally to *Buchberger's criteria*. The first of these criteria is the simplest to use, and its proof follows easily from the lcm representation appearing in Theorem 4.

Proposition 4.2. [Buchberger's first criterion] Suppose $f, g \in G$ are such that $\mathrm{LM}_<(f)$ and $\mathrm{LM}_<(g)$ are relatively prime. Then S_{fg} has a lcm representation with respect to G and $<$.

Proof. We define the "tails" of f and g wrt $<$ to be

$$\mathrm{tail}_<(f) = f - \mathrm{LT}_<(f), \quad \mathrm{tail}_<(g) = g - \mathrm{LT}_<(g).$$

WLOG assume $\mathrm{LC}_<(f) = \mathrm{LC}_<(g) = 1$. We then calculate

$$\begin{aligned}
S_{fg} &= \mathrm{LM}_<(g) f - \mathrm{LM}_<(f) g \\
&= (g - \mathrm{tail}_<(g)) f - (f - \mathrm{tail}_<(f)) g \\
&= \mathrm{tail}_<(f) g - \mathrm{tail}_<(g) f.
\end{aligned}$$

The last of these formulae is a lcm representation, since

$$\begin{aligned}
\mathrm{LM}_< \left(\mathrm{tail}_<(f) g \right) \; &< \; \mathrm{LM}_<(fg) = \mathrm{lcm}(\mathrm{LM}_<(f), \mathrm{LM}_<(g)), \\
\mathrm{LM}_< \left(\mathrm{tail}_<(g) f \right) \; &< \; \mathrm{LM}_<(fg) = \mathrm{lcm}(\mathrm{LM}_<(f), \mathrm{LM}_<(g)).
\end{aligned}$$

□

There are many examples which show that Gröbner bases are not preserved under specialization of variables. For instance, if we take $G = \{\underline{ax} + y + b, \underline{by} + z\}$, then Proposition 4.2 implies this is a Gröbner basis for the `Lex` order with $a > y > b > z > x$. However, if we set $a = 1$, and work with the induced `Lex` order on the remaining variables, our polynomials become $G = \{x + \underline{y} + b, \underline{by} + z\}$, and we get the S-polynomial

$$b(x + \underline{y} + b) - (\underline{by} + z) = \underline{b^2} + bx - z \quad \text{w/} \quad b^2 \notin \langle y, by \rangle.$$

Nevertheless, we *can* prove a specialization property for elimination orders with $\mathbf{x} > \mathbf{y}$, provided that we specialize the cheap variables $\mathbf{y}$ to *sufficiently generic* values.

Proposition 4.3. Let $G = \{g_1, \ldots, g_s\} \subset \mathbb{K}[\mathbf{x}, \mathbf{y}] = \mathbb{K}[x_1, \ldots, x_n, y_1, \ldots, y_m]$ be a Gröbner basis with respect to an elimination order with $\mathbf{x} > \mathbf{y}$. Let us partition the set G as

$$G = \{g_1, \ldots, g_{s'}\} \cup \{g_{s'+1}, \ldots, g_s\}$$

where $g_1, \ldots, s_{s'} \in \mathbb{K}[\mathbf{y}]$ and $g_{s'+1}, \ldots, g_s \notin \mathbb{K}[\mathbf{y}]$. We may write for $i = s' + 1, \ldots, s$,

$$g_i(\mathbf{x}, \mathbf{y}) = c_i(\mathbf{y}) \mathbf{x}^{\boldsymbol{\alpha}_i} + \text{l.o.t.},$$

where $\mathrm{LT}_<(g_i) = \mathrm{LT}_<(c_i) \cdot \mathbf{x}^{\boldsymbol{\alpha}_i}$ where $x^{\boldsymbol{\alpha}_i} > 1$. Then, if $\bar{y} \in \mathbb{K}^m$ is a point such that $c_i(\bar{y}) \neq 0$ for all $s' + 1 \leq i \leq s'$, and $g_i(\bar{y}) = 0$ for $1 \leq i \leq s'$, the set of specialized polynomials $\{g_1(\mathbf{x}, \bar{y}), \ldots, g_s(\mathbf{x}, \bar{y})\}$ is a Gröbner basis.

To prove Proposition 4.3, we develop further the notion of a standard representation appearing in Theorem 4.

Proposition 4.4. Let $G = \{g_1, \ldots, g_s\} \subset \mathbb{K}[\mathbf{x}]$ and fix a monomial order $<$.

1. If we have
$$\mathrm{LM}_<(g_1 + g_2 + \cdots g_s) < \mathrm{LM}_<(g_1) = \mathrm{LM}_<(g_2) = \cdots = \mathrm{LM}_<(g_s), \tag{26}$$
then $g_1 + \cdots + g_s$ is a $\mathbb{K}$-linear combination of S-polynomials formed from G.
2. If G is a Gröbner basis, then every $f \in \langle G \rangle$ has a standard representation,
$$f = \sum_{i=1}^{s} h_i g_i \quad \text{w/} \quad h_i g_i \neq 0 \Rightarrow \mathrm{LM}_<(h_i g_i) \leq \mathrm{LM}_<(f).$$

Proof. For part 1, our assumption 26 implies that
$$\mathrm{LC}_<(g_1) + \mathrm{LC}_<(g_2) + \cdots + \mathrm{LC}_<(g_s) = 0. \tag{27}$$
It follows that a suitable $\mathbb{K}$-linear combination is given by
$$\begin{aligned}
\sum_{i=1}^{s-1} \mathrm{LC}_<(g_i) S_{g_i g_s} &= \sum_{i=1}^{s-1} \mathrm{LC}_<(g_i) \left(\frac{g_i}{\mathrm{LC}_<(g_i)} - \frac{g_s}{\mathrm{LC}_<(g_s)} \right) \\
&= \sum_{i=1}^{s-1} g_i - \left(\sum_{i=1}^{s-1} \frac{\mathrm{LC}_<(g_i)}{\mathrm{LC}_<(g_s)} \right) g_s \\
&= \sum_{i=1}^{s} g_i \qquad \text{(by 27.)}
\end{aligned}$$

For part 2, take any $f = \sum_{i=1}^{s} h_i g_i \in I$. Let $\mathbf{x}^{\boldsymbol{\alpha}}$ be the maximum element of $\{\mathrm{LM}_<(h_i g_i) \mid 1 \leq i \leq s\}$, and write
$$f = \sum_{\substack{i \text{ w/} \\ \mathrm{LM}_<(h_i g_i) = \mathbf{x}^{\boldsymbol{\alpha}}}} \mathrm{LT}_<(h_i) g_i \; + \sum_{\substack{i \text{ w/} \\ \mathrm{LM}_<(h_i g_i) = \mathbf{x}^{\boldsymbol{\alpha}}}} \mathrm{tail}_<(h_i) g_i \; + \text{ l.o.t.}$$
By Part 1, the first summand as a linear combination of S-polynomials formed from G, and thus Theorem 4 implies it has a standard representation. Since the second and third summands contribute smaller leading terms than the first, we conclude that f has a standard representation. □

Proof of Proposition 4.3. Consider the "partial S-polynomials" defined by
$$S_{ij}(\mathbf{x}, \mathbf{y}) = \frac{\mathrm{lcm}(\mathbf{x}_i^{\boldsymbol{\alpha}}, \mathbf{x}_j^{\boldsymbol{\alpha}})}{c_i(\bar{y})\, \mathbf{x}_i^{\boldsymbol{\alpha}}} g_i(\mathbf{x}, \mathbf{y}) - \frac{lcm(\mathbf{x}_i^{\boldsymbol{\alpha}}, \mathbf{x}_j^{\boldsymbol{\alpha}})}{c_j(\bar{y})\, \mathbf{x}_j^{\boldsymbol{\alpha}}} g_j(\mathbf{x}, \mathbf{y}).$$
If we specialize, we get an honest S-polynomial with respect to the induced order on $\mathbb{K}[\mathbf{x}]$,
$$S_{ij}(\mathbf{x}, \bar{y}) = S_{g_i(\mathbf{x}, \bar{y}) g_j(\mathbf{x}, \bar{y})}.$$
By Proposition 4.4, $S_{ij}(\mathbf{x}, \mathbf{y})$ has a standard representation in $\mathbb{K}[\mathbf{x}, \mathbf{y}]$,
$$S_{ij}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{s} h_i(\mathbf{x}, \mathbf{y}) g_i(\mathbf{x}, \mathbf{y}). \tag{28}$$
For each summand whose specialization doesn't vanish, $h_i(\mathbf{x}\bar{y}) g_i(\mathbf{x}, \bar{y}) \neq 0$, we have
$$\begin{aligned}
\mathrm{LM}_< \left(h_i(\mathbf{x}, \bar{y}) g_i(\mathbf{x}, \bar{y}) \right) &\leq \mathrm{LM}_< \left(h_i(\mathbf{x}, \mathbf{y}) g_i(\mathbf{x}, \mathbf{y}) \right) \\
&\leq \mathrm{LM}_<(S_{ij}) \\
&< \mathrm{lcm}(\mathbf{x}^{\boldsymbol{\alpha}_i}, \mathbf{x}^{\boldsymbol{\alpha}_j})
\end{aligned}$$
where the first and third inequalities use the fact that $<$ is an elimination order. Thus, specializing the standard representation 28 in $\mathbb{K}[\mathbf{x}, \mathbf{y}]$, we obtain a lcm representation for the corresponding S-polynomial in $\mathbb{K}[\mathbf{x}]$,
$$S_{g_i(\mathbf{x}, \bar{y}) g_j(\mathbf{x}, \bar{y})} = \sum_{i=1}^{s} h_i(\mathbf{x}, \bar{y}) g_i(\mathbf{x}, \bar{y}).$$
Thus, Theorem 4 implies that $\{g_1(\mathbf{x}, \bar{y}), \ldots, g_s(\mathbf{x}, \bar{y})\}$ forms a Gröbner basis. □

5 Algebra/geometry dictionary

5.1 Nullstellensaätz

For any field $\mathbb{K}$, the n-dimensional affine space $\mathbb{A}^n_{\mathbb{K}}$ is defined to be the set $\mathbb{K}^n$ consisting of all ordered n-tuples of elements of $\mathbb{K}$. The most basic objects studied in algebraic geometry are the Zariski-closed subsets of $\mathbb{A}^n_{\mathbb{K}}$, also known as affine algebraic varieties.

Definition 5.1. For any set of polynomials $S \subset \mathbb{K}[x_1, \ldots, x_n]$, the vanishing locus $\mathbf{V}(S)$ is defined to be the set of all points in affine n-space over $\mathbb{K}$ where every polynomial in S vanishes:

$$\mathbf{V}_{\mathbb{K}}(S) = \{a \in \mathbb{A}^n_{\mathbb{K}} \mid f(a) = 0 \quad \forall f \in S\}.$$

When the field is clear we write $\mathbf{V}(S)$. A set of the form $\mathbf{V}_{\mathbb{K}}(S)$ is called a *Zariski-closed* subset of $\mathbb{A}^n_{\mathbb{K}}$.

It would not be much of a restriction to assume the set S appearing in Definition 5.1 is an ideal.

Exercise 9. Show that $\mathbf{V}(S) = \mathbf{V}(\langle S \rangle)$.

As long as it's clear that we are talking about a subset of $\mathbb{A}^n_{\mathbb{K}}$, it will be ok to simply call $\mathbf{V}(S)$ a Zariski-closed set. However, after we develop parallel notions for projective algebraic varieties, this shorthand will become ambiguous. When in doubt, say where things live!

Exercise 10. (For readers familiar with topology) Show that the Zariski-closed sets are the closed sets with respect to a topology on $\mathbb{A}^n_{\mathbb{K}}$. This is called the *Zariski topology.*

Definition 5.2. If $X \subset \mathbb{A}^n_{\mathbb{K}}$ is any set, we define its Zariski closure in $\mathbb{A}^n_{\mathbb{K}}$ to be the smallest (inclusion-wise) Zariski-closed set containing X. We denote the Zariski closure of X in $\mathbb{A}^n_{\mathbb{K}}$ by $\overline{X}$.

A basic result in algebraic geometry is Hilbert's Nullstellensatz, which establishes a 1-1 correspondence between Zariski-closed and radical ideals. We recall the definition of a radical ideal.

Definition 5.3. An ideal $I \subset \mathbb{K}[\mathbf{x}]$ is said to be *radical* if whenever $f^m \in I$ for some $f \in \mathbb{K}[\mathbf{x}]$, $m \in \mathbb{Z}_{\geq 0}$, we have $f \in I$.

The operation of taking the vanishing locus of an ideal has a "sort of" inverse operation.

Definition 5.4. For any subset $X \subset \mathbb{A}^n_{\mathbb{K}}$, we define the vanishing ideal $\mathbf{I}(X)$ to be the set of all polynomials that vanish on all points of X:

$$\mathbf{I}(X) = \{f \in \mathbb{K}[\mathbf{x}] \mid f(a) = 0 \quad \forall a \in X\}.$$

One can easily check the following properties of the operations $\mathbf{V}$ and $\mathbf{I}$:

1. $\mathbf{I}(X) = \mathbf{I}(\overline{X})$
2. $\mathbf{V}(\mathbf{I}(\mathbf{V}(X))) = \mathbf{V}(X)$
3. $\mathbf{I}(\mathbf{V}(\mathbf{I}(X))) = \mathbf{I}(X)$
4. For Zariski-closed $X, Y \subset \mathbb{A}^n_{\mathbb{K}}$, if $\mathbf{I}(X) = \mathbf{I}(Y)$, then $X = Y$.

The natural analogue of property 4 for ideals does not hold: for example, the ideals $\langle x \rangle, \langle x^2 \rangle \subset \mathbb{K}[x]$ have the same vanishing locus, but they are not equal.

Another obstruction to getting a bijection between ideals and varieties occurs when the field $\mathbb{K}$ is not algebraically closed. Indeed, we have two distinct ideals $\langle 1 \rangle, \langle x^2 + 1 \rangle \subset \mathbb{R}[\mathbf{x}]$, whose vanishing loci in $\mathbb{A}^1_{\mathbb{R}}$ are both the empty set.

Fortunately, as long as we assume $\mathbb{K}$ is algebraically closed, we can establish a bijection between Zariski-closed sets and radical ideals.

Theorem 5. [Hilbert's Nullstellensatz] If $\mathbb{K}$ is algebraically closed and $I \subset \mathbb{K}[\mathbf{x}]$ is a radical ideal, then $\mathbf{I}(\mathbf{V}(I)) = I$.

Thus, if $I, J \subset \mathbb{K}[\mathbf{x}]$ are two radical ideals with $\mathbf{V}(I) = \mathbf{V}(J)$, for $\mathbb{K}$ algebraically closed it follows that

$$I = \mathbf{I}(\mathbf{V}(I)) = \mathbf{I}(\mathbf{V}(J)) = J.$$

Remarkably, Theorem 5 can be reduced to the following "weak" form.

Theorem 6. [Weak Nullstellensatz] Let $I \subset \mathbb{K}[\mathbf{x}]$ be an ideal, $\mathbb{K}$ algebraically closed. Then $1 \in I$ iff $\mathbf{V}(I) = \emptyset$.

Notice that the weak Nullstellensatz reduces to Theorem 1 when $\mathbb{K} = \mathbb{C}$ and $n = 1$. Also, the "only if" direction of Theorem 6 still holds when the field $\mathbb{K}$ is not algebraically closed. Since Buchberger's algorithm uses the same operations whether we work over the field $\mathbb{K}$ or any of its extensions, Gröbner bases can provide a simple *infeasibility certificate* that $\mathbf{V}_{\mathbb{K}}(I)$ is empty.

Corollary 5.5. Let $I \subset \mathbb{K}[\mathbf{x}]$ be an ideal, $\overline{\mathbb{K}}$ be an algebraically closed field containing $\mathbb{K}$. Then $\mathbf{V}_{\overline{\mathbb{K}}}(I) = \emptyset$ iff the reduced Gröbner basis of I is the set $\{1\}$.

Example 12. Let $G = ([n], E)$ be a simple, undirected graph on n vertices. Recall that a proper k-coloring of G is a function $\chi : [n] \to [k]$ such that $\chi(i) \neq \chi(j)$ whenever there is an edge $(i, j) \in E$. If we replace $[k]$ with the set of k-th roots of unity, it is easy to write down a system of polynomials whose set of solutions in $\mathbb{C}$ are precisely the k-colorings of G. Indeed, we may define the k-th *coloring ideal* of G to be

$$I_{\text{col}}(k, G) = \langle x_i^k - 1 \mid i \in [n] \rangle + \langle x_i^{k-1} + x_i^{k-2}x_j + \cdots + x_i x_j^{k-2} + x_j^{k-1} \mid (i, j) \in E \rangle.$$

By Corollary 5.5, if we take $\mathbb{K} = \mathbb{Q}$ to be our field of definition, then it follows that I has a proper k-coloring if and only if the reduced Gröbner basis of $I_{\text{col}}(k, G)$ equals $\{1\}$. The following code shows that the Petersen graph (an example of a Kneser graph) is 3-colorable, but not 2-colorable.

```
needsPackage "Graphs"
coloringIdeal = (k, KK, G) -> (
    (V, E) := (vertices G, toList \ edges G);
    n := length V;
    R := KK[(symbol x)_0..(symbol x)_(n-1)];
    eq1 := apply(n, i -> x_i^k - 1);
    eq2 := apply(E, e -> sum(0..k-1, i -> x_(first e)^i * x_(last e)^(k-1-i)));
    ideal(eq1 | eq2)
    )
G = kneserGraph(5,2)
gens gb coloringIdeal(2, QQ, G) -- no 2-colorings by NSZ
gens gb coloringIdeal(3, QQ, G) -- chromatic number is 3
```

Let's now prove both Nullstellensätze.

Proof of Weak Nullstellensatz. Clearly $\mathbf{V}(I) \neq \emptyset$ implies $1 \notin I$. We prove the converse by induction on n. When $n = 1$, the ideal I is generated by a single polynomial p which has a root $r \in \mathbf{V}(I)$. For $n > 1$, we write $\mathbf{x} = \{x_1, \ldots, x_{n-1}\}$. Let $I \subset \mathbb{K}[\mathbf{x}, x_n]$ with $1 \notin I$. For any $a \in \mathbb{K}$ we consider the ideal $I|_{x_n=a} \subset \mathbb{K}[\mathbf{x}]$ obtained by specializing every element of I to $x_n = a$. If we can show $1 \notin I|_{x_n=a_n}$ for some $a_n \in \mathbb{K}$, then by induction we would get a point $(a_1, \ldots, a_{n-1}) \in \mathbf{V}(I|_{x_n=a})$, which would then imply $(a_1, \ldots, a_n) \in \mathbf{V}(I)$.

Consider the elimination ideal $J = I \cap \mathbb{K}[x_n]$. We treat separately the cases when J contains a nonzero polynomial and when it does not.

In the first case, we have $J = \langle \prod_{j=1}^{d} (x_n - b_j) \rangle$ for some $b_1, \ldots, b_d \in \mathbb{K}$. We claim that $1 \notin I|_{x_n=b_i}$ for some b_i. If this were not the case, then by induction we would have for each i a polynomial $B_i \in \mathbb{K}[\mathbf{x}]$ with $B_i(\mathbf{x}, b_i) = 1$. Note, however, that there exist polynomials $A_i \in \mathbb{K}[\mathbf{x}, x_n]$ such that

$$B_i(\mathbf{x}, x_n) = B_i(\mathbf{x}, b_i + (x_n - b_i)) = B_i(\mathbf{x}, b_i) + (x_n - b_i)A(\mathbf{x}, x_n) = 1 + (x_n - b_i)A(\mathbf{x}, x_n).$$

Re-arranging these identities and taking their product, we obtain

$$1 = \prod_{i=1}^{d} \big(B_i(\mathbf{x}, x_n) + (b_i - x_n)A(\mathbf{x}, x_n)\big),$$

and hence, for some $A \in \mathbb{K}[\mathbf{x}, x_n]$, we have

$$1 = \prod_{j=1}^{d} B_i(\mathbf{x}, x_n) + A(\mathbf{x}, x_n) \prod_{j=1}^{d} (x_n - b_j) \in I$$

a contradiction.

In the second case, we apply Proposition 4.3 for the Lex order with cheap variables $\mathbf{y} = \{x_n\}$ and expensive variables $\mathbf{x} = \{x_1, \ldots, x_{n-1}\}$. If G is a Gröbner basis for I with respect to this order, choosing a such that all leading coefficients in x_n do not vanish implies that $G|_{x_n=a}$ is a Gröbner basis for $I_{x_n=a}$. Moreover, none of the leading monomials of $G|_{x_n=a}$ is 1, as this would imply $G \cap \mathbb{K}[x_n] \neq \emptyset$. Thus $1 \notin I|_{x_n=a}$, as desired. □

Proof of Hilbert's Nullstellensatz. The inclusion $I \subset \mathbf{I}(\mathbf{V}(I))$ is trivial. For the reverse inclusion, suppose $I = \langle f_1, \ldots, f_r \rangle$, and suppose $f \in \mathbf{I}(\mathbf{V}(I))$. Let y be a new variable and consider the ideal

$$J = \langle f_1, \ldots, f_s, yf - 1 \rangle \subset \mathbb{K}[\mathbf{x}, y].$$

Then we have $\mathbf{V}_{\mathbb{K}}(J) = \emptyset$, so Theorem 6 implies that $1 \in J$. More explicitly, we can write

$$1 = h_1(\mathbf{x}, y)f_1(\mathbf{x}) + \cdots + h_s(\mathbf{x}, y)f_s(\mathbf{x}) + h_{s+1}(\mathbf{x}, y)(yf(\mathbf{x}) - 1).$$

If we set $y = 1/f(\mathbf{x})$ and clear denominators, we obtain

$$f(\mathbf{x})^m = H_1(\mathbf{x})f_1(\mathbf{x}) + \cdots + H_s(\mathbf{x})f_s(\mathbf{x})$$

for some $H_1, \ldots, H_s \in \mathbb{K}[\mathbf{x}]$. Since I is radical, we conclude that $f \in I$. □

The last proof illustrates a general trick for turning a polynomial *inequation* $f(\mathbf{x}) \neq 0$ into a polynomial equation with one new variable, $f(\mathbf{x})y = 1$. This is sometimes called the *Rabinowitsch trick.*

5.2 Irreducibility and decomposition

Next, we discuss the decomposition of varieties into simpler building blocks, the *irreducible varieties.*

Definition 5.6. Let $X \subset \mathbb{A}^n_{\mathbb{C}}$ be an affine variety. We say that X is *irreducible* if we cannot write X is the union of two affine varieties properly contained in X. Otherwise we say X is *reducible.*

In the next proposition, we collect some useful facts about irreducible varieties and their vanishing ideals.

Proposition 5.7.

1. Every affine variety X can be written as uniquely as the union of irreducible subvarieties,
$$X = X_1 \cup X_2 \cup \cdots \cup X_k, \tag{29}$$
with $X_i \subset X$ irreducible such that $X_i \not\subset X_j$ whenever $1 \leq i < j \leq k$.
2. The variety X is irreducible iff its vanishing ideal $\mathcal{I}_X$ is prime.
3. The vanishing ideal of any variety can be written uniquely as the intersection of *prime* polynomial ideals. This is the *prime decomposition* of a radical ideal in $\mathbb{K}[\mathbf{x}]$.

The $X_1, \ldots, X_k$ appearing in are called the *irreducible components* of X, and 29 is called the (irredundant or minimal) *irreducible decomposition* of X.

Proof.

1. To see that a decomposition of the form 29 exists, let X be any variety. If X is irreducible, we are done. Otherwise, we can write $X = X_1 \cup X_1'$ for proper subvarieties $X_1, X_1' \subsetneq X$. We are again done if both subvarieties are irreducible, so suppose X_1 is not. We can continue to play this game, but not forever: for otherwise, we would generate a strictly descending chain of subvarieties
$$X_1 \supsetneq X_2 \supsetneq \cdots$$
and a corresponding strictly ascending chain of vanishing ideals
$$\mathcal{I}_{X_1} \subsetneq \mathcal{I}_{X_2} \subsetneq \cdots,$$
which would contradict Hilbert's Basis Theorem 2. Thus we obtain a decomposition of X as the union of a finite number of irreducible subvarieties. After deleting some of these subvarieties, we may also assume the irredundancy condition $X_i \not\subset X_j$. To see uniqueness, suppose we have another irredundant irreducible decomposition $X = X_1' \cup \cdots X_j'$. Then, for each X_j', we have that
$$X_j' = X_j' \cap X = \left(X_j' \cap X_1\right) \cup \cdots \cup \left(X_j' \cap X_k\right),$$
and since X_j' is irreducible we must have $X_j' \subset X_{i_j}$ for some i_j between 1 and k. Applying this argument in reverse, we may deduce that $X_j' = X_{i_j}$.

2. Suppose X is irreducible and let $f_1, f_2 \in \mathbb{K}[\mathbf{x}]$ such that their product $f_1 f_2$ vanishes on X. We may write

$$X = (X \cap \mathbf{V}(f_1)) \cup (X \cap \mathbf{V}(f_2)) . \tag{30}$$

If neither f_1 nor f_2 vanish on X, then we have written X in eq. (30) as the union of two proper subvarieties, contradicting irreducibility. Thus either f_1 or f_2 is in the vanishing ideal $\mathcal{I}_X$, proving thus ideal is prime. Conversely, if X is reducible, then we can write $X = X_1 \cup X_2$ with $X_i \subsetneq X$. The strict inclusions reverse for the vanishing ideals: $\mathcal{I}_X \subsetneq \mathcal{I}_{X_i}$. So take $f_i \in \mathcal{I}_{X_i} \setminus \mathcal{I}_X$ for $i = 1, 2$. Then $f_1 f_2 \in \mathcal{I}_X$ shows that $\mathcal{I}_X$ is not prime.

3. Any variety X has a decomposition 29, and we may consider the prime ideals $P_1 = \mathcal{I}_{X_1}, \ldots, P_k = \mathcal{I}_{X_k}$. Define $I = \bigcap_{j=1}^{k} P_j$. It is easy to check that I is a radical ideal vanishing on X. So, when $\mathbb{K}$ is algebraically closed, Hilbert's Nullstellensatz (Theorem 5) implies $I = \mathcal{I}_X$. The general case follows by pulling back a prime decomposition after extending scalars to an algebraic closure, $\varphi : \mathbb{K}[\mathbf{x}] \to \overline{\mathbb{K}}[\mathbf{x}]$, and eliminating any redundant primes that appear after pulling back.

□

When doing symbolic computation, we are a bit of a disadvantage, since a general element favorite algebraically closed field $\mathbb{C}$ is not computable. So we instead work over $\mathbb{Q}$, or even a finite field, and hope for the best. Sometimes this causes no issues.

Example 13. In the example below, we work with an ideal $I \subset \mathbb{Q}[x, y, z]$. We verify that the ideal I, although not prime, is radical. The command `decompose` writes I as the intersection of two prime ideals,

$$I = \langle x^2 - y \rangle \cap \langle z + 1, y - 2, x - 1 \rangle.$$

This corresponds to the fact that $\mathbf{V}_{\mathbb{Q}}(I)$ defines the union of a quadric surface and the point in $(-1, 2, 1) \in \mathbb{Q}^3$. In this case, the same is true if we work over the complex numbers.

```
R = QQ[x,y,z]
I = ideal(x^2*z+x^2-y*z-y, x^2*y-2*x^2-y^2+2*y, x^3-x^2-x*y+y)
isPrime I, I == radical I -- false, true
decompose I
```

The next two example illustrates that we should be somewhat cautious when decomposing ideals.

Example 14. Consider $I = \langle x^2 + xy + y^2 \rangle \subset \mathbb{Q}[x, y]$. You can check that this is a prime ideal. Unfortunately, this does not remain true when we *extend scalars* to $\mathbb{C}$. What this means is that for the ring homomorphism $\varphi : \mathbb{Q}[x, y] \to \mathbb{C}[x, y]$, the complex polynomial ideal generated by $\varphi(f)$ for all $f \in I$ is *not* prime. Ineeded, we have

$$x^2 + xy + y^2 = (x + \omega y)(x + \omega^2 y)$$

where $\omega \in \mathbb{C}$ is a primitive cube root of unity. Thus, the vanishing locus of this polynomial in $\mathbb{A}^2_{\mathbb{C}}$ is the union of two lines.

There is a generalization of the prime decomposition in Proposition 5.7 for ideals that are not radical, called *primary decomposition*. We say an ideal I is *primary* if its radical is prime, and that I is irreducible if whenever $I = I_1 \cap I_2$ then either $I = I_1$ or $I = I_2$.

Theorem 7 (Lasker-Noether Theorem)**.** Any ideal $I \subset \mathbb{K}[\mathbf{x}]$ can be written an irredundant intersection of primary ideals,

$$I = Q_1 \cap \cdots \cap Q_s,$$

such that $P_1 = \sqrt{Q_1}, \ldots, P_s = \sqrt{Q_s}$ are distinct prime ideals, and $\cap_{j \neq i} P_j \not\subset P_i$ for all i. Moreover, the set of *associated primes*

$$\mathrm{Ass}(I) = \{P_1, \ldots, P_s\}$$

appearing in such a decomposition is unique.

Example 15. Consider the ideal $I = \langle x^2y, x^3 \rangle \subset \mathbb{C}[x, y]$. The vanishing locus $\mathbf{V}_{\mathbb{C}}(I)$ is the line $x = 0$ in $\mathbb{C}^2$. This is an irreducible variety, whose prime vanishing ideal is $\sqrt{I} = \langle x \rangle \supsetneq I$. Running `decompose` tells us about the prime decomposition of $\sqrt{I}$. To get information about I instead, we use `primaryDecomposition`, whose output informs us that

$$I = \langle x^2 \rangle \cap \langle x^3, y \rangle.$$

For the first of these ideals, we have the *associated prime* $\sqrt{\langle x^2 \rangle} = \sqrt{I} = \langle x \rangle$. On the other hand, we have another associated prime, $\sqrt{\langle x^3, y \rangle} = \langle x, y \rangle \supset \langle x \rangle$. The first prime is said to be *minimal* and the second is *embedded*. This primary decomposition, despite being irredundant, is not unique: for instance

$$I = \langle x^2 \rangle \cap \langle x^3, x^2y, xy^2, y^3 \rangle.$$

We have already seen through Gröbner bases how monomial ideals play a special role in understanding general polynomial ideals. An extremely special class of monomial ideals of all are the ones that are also radical.

Exercise 11. Show that a monomial ideal is radical iff its minimal generators are all squarefree.

Squarefree monomial ideals are very combinatorial objects. This is because any subset $S \subset [n]$ gives rise to the squarefree monomial

$$\mathbf{x}^S = \prod_{i \in S} x_i \in \mathbb{K}[x_1, \ldots, x_n].$$

Given a squarefree monomial ideal $I = \langle \mathbf{x}^{S_1}, \ldots, \mathbf{x}^{S_k} \rangle$, we may construct an associated simplicial complex called the *Stanley-Reisner complex*:

$$\Delta_I = \{S \subset [n] \mid \mathbf{x}^S \notin I\}.$$

This description can be applied in reverse: the *Stanley-Reisner ideal* of an arbitrary simplicial complex is the monomial ideal generated by its minimal non-faces. These ideals played a surprising role in Stanley's proof of the upper bound theorem.

Example 16. The following code shows an example of how to visualize a 2D simplicial complexes and play with the associated Stanley-Reisner ideals. How do the associated primes of I relate to the faces of Δ_I?

```
R = QQ[a..f]
help simplicialComplex
I = monomialIdeal(c*d,c*d,e*f,e*f,a*c,a*f,b*d,b*e)
netList decompose I
B2 = basis(2, R/I) -- contains minimal nonfaces on 2 vertices
needsPackage "SimplicialDecomposability"
S = simplicialComplex I
faces S
isShellable S -- false
netList primaryDecomposition I
needsPackage "Visualize";
openPort "222";
visualize S
```

5.3 Functions and mappings

A common philosophy in math is that the properties of a "space" should be reflected by the properties of the functions one can define on it. On affine varieties, there are two important types of functions.

Definition 5.8. If $X \subset \mathbb{A}^n_{\mathbb{K}}$ is an affine variety, its *coordinate ring* is defined to be the quotient ring $\mathbb{K}[X] = \mathbb{K}[\mathbf{x}]/\mathcal{I}_X$. When X is irreducible, Proposition 5.7 implies that $\mathbb{K}[X]$ is an integral domain, in which case its field of fractions is called the *function field* of X and denoted $\mathbb{K}(X)$.

To make sense of Definition 5.8, note that if we have, say, two polynomial functions (also called regular functions) $f, g : \mathbb{A}^n_{\mathbb{K}} \to \mathbb{K}$, then their *restrictions* to X are equal iff $f - g \in \mathcal{I}_X$. Thus, the coordinate ring of X can be thought of as the ring of all polynomial functions on X. Similarly, for irreducible X, the function field of X can be thought of as the field of rational functions on X.

Regular and rational functions can be throught of as mappings from the variety X to the variety $\mathbb{A}^1_{\mathbb{K}}$. The following definitions include these as special cases. For an affine variety $X \subset \mathbb{A}^n_{\mathbb{K}}$, we say that $U \subset X$ is *Zariski-open* in X if $U = X \cap (\mathbb{A}^n_{\mathbb{K}} \setminus Y)$ for some other affine variety $Y \subset \mathbb{A}^n_{\mathbb{K}}$. This is precisely what it means for a set to be open in the topology on X induced by the Zariski topology on $\mathbb{A}^n_{\mathbb{K}}$.

Definition 5.9. For affine varieties $X \subset \mathbb{A}^n_{\mathbb{K}}$, $Y \subset \mathbb{A}^m_{\mathbb{K}}$, and polynomials $f_1, \ldots, f_m \in \mathbb{K}[\mathbf{x}] = \mathbb{K}[x_1, \ldots, x_n]$, the map

$$\begin{aligned} X &\to Y \\ \mathbf{x} &\mapsto (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})) \end{aligned}$$

is said to be *regular*. For X irreducible, suppose the $f_1, \ldots, f_m \in \mathbb{K}(X)$ are such that $f_i(x)$ exists for all i and x contained in some *nonempty* Zariski-open $U \subset X$. Then then the map

$$\begin{aligned} f : U &\to Y \\ \mathbf{x} = (x_1, \ldots, x_n) &\mapsto (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})) \end{aligned}$$

is said to be a *rational map*, also denoted

$$f : X \dashrightarrow Y.$$

A rational map is really better thought of as an equivalence class of the pairs (f, U) appearing in Definition 5.9, such that the functions f agree on the common overlap of any two open sets U. The union over all such open sets is also open, and is the maximal *domain of definition* for f.

6 Dimension of an affine variety

Throughout this section, $\mathbb{K}$ denotes an algebraically closed field. The dimension of an affine variety $X \subset \mathbb{A}^n_{\mathbb{K}}$ can be thought of as the number of "free variables" in the vanishing ideal of X. To make this precise, it is technically convenient, although not strictly necessary, to assume that X is irreducible, and then define the dimension of an arbitrary variety to be the maximum dimension of its irreducible components.

For a general field extension $\mathbb{F}/\mathbb{K}$, we recall the notion of a set $B \subset \mathbb{F}$ that is *algebraically independent* over $\mathbb{K}$. For our purposes, $B = \{b_1, \ldots, b_k\}$ will always be finite, in which case this means that there exists no polynomial $p \in \mathbb{K}[x_1, \ldots, x_k]$ such that $p(b_1, \ldots, b_k) = 0$. We say $\mathbb{F}/\mathbb{K}$ is algebraic over $\mathbb{F}$ if there exists no subset of $\mathbb{F}$ that is algebraically independent over $\mathbb{K}$. A *transcendence basis* for $\mathbb{F}$ over $\mathbb{K}$ is a set $B \subset \mathbb{F}$ algebraically independent over $\mathbb{K}$ such that $\mathbb{F}$ is algebraic over the subfield generated by B, denoted $\mathbb{K}(B)$. The cardinality of B in this case is called the *transcendence degree* of $\mathbb{K}$ over $\mathbb{F}$, which makes sense because of the following result.

Proposition 6.1. The transcendence degree of a field extension is well-defined.

To prove Proposition 6.1 , we make a detour into the realm of *matroids.*

Definition 6.2. A *matroid* is a pair $(S, \mathcal{B})$, where S is a set and $\mathcal{B}$ is a collection of finite subsets of S called *bases* that satisfy the following *exchange property*: for any $B, B' \in \mathcal{B}$ and $b \in B$, there exists $b' \in \mathcal{B}'$ such that $B' - b' + b$ (this is an abbreviation for $(B' \setminus \{b'\}) \cup \{b\}$) is also in $\mathcal{B}$.

Exercise 12. Show that two bases of a matroid have the same cardinality.

Matroids appear everywhere. For one motivating example, we may take S to be a vector space, and $\mathcal{B}$ to be the set of bases in the sense that is familiar from linear algebra. Similarly, we can show that the transcendence bases of a field extension are also the bases of a matroid.

Proof of Proposition 6.1. Suppose $B = \{b_1, \ldots, b_s\}, B' = \{b'_1, \ldots, b'_r\}$ are two transcendence bases for $\mathbb{F}/\mathbb{K}$. If we delete some element from B', say b'_1, then we need to find some $b_i \in B$ such that $B - b'_1 + b_i$ is also a transcendence basis. We know for all $i = 1, \ldots, s$ that there exists a univariate polynomial p_i with coefficients in $\mathbb{K}(B')$ such that $p_i(b_i) = 0$. At least one p_i must involve b'_1 nontrivially; otherwise, all elements of B would be algebraic over of $\mathbb{K}(b'_2, \ldots, b'_r)$, and since b_1 is algebraic over $\mathbb{K}(B)$ this would imply that B' is not algebraically independent. Now, for any p_i involving b'_i nontrivially, a similar argument shows that $B' - b'_1 + b_i$ is also a transcendence basis. This proves that the transcendence bases of $\mathbb{F}/\mathbb{K}$ give a matroid. Since any two bases in a matroid have the same size, we are done. $\square$

Definition 6.3. The dimension of an irreducible affine variety $X \subset \mathbb{A}^n_{\mathbb{K}}$ is defined to be the transcendence degree of $\mathbb{K}(X)$ over $\mathbb{K}$. In general, $\dim X$ is defined to be the maximum dimension of an irreducible component of X.

It is worth observing that the coordinate functions of affine n-space restricted to an irreducible variety will always contain a transcendence basis. If $X \subset \mathbb{A}^n_{\mathbb{K}}$ is an affine variety (not necessarily irreducible), we say a subset of variables $\{x_i\}_{i\in B}$ for $B \subset [n]$ is *free on* X if

$$\mathcal{I}_X \cap \mathbb{K}[\{x_i\}_{i\in B}] = \langle 0 \rangle.$$

Geometrically, this means the set-theoretic image of X under the projection onto the affine space with coordiantes indexed by B is Zariski-dense: such a map is said to be *dominant*.

Proposition 6.4. The dimension of an affine variety $X \subset \mathbb{A}^n_{\mathbb{K}}$ is the cardinality of a set of free variables on X.

Proof. Decompose X into irreducible components using Proposition 5.7, $X = X_1 \cup \cdots \cup X_k$. Consider any $\{x_i\}_{i\in B}$ with $B \subset [n]$. Observe that this set is free on X if and only if it is free on some X_i. In particular, when $\dim(X_i) = \dim(X)$, we have $|B| \leq \dim(X)$, with equality iff the set is a transcendence basis for $\mathbb{K}(X_i)$. □

When X is irreducible, the transcendence bases formed by free sets of variables form a matroid on the ground set $[n]$, usually called the *algebraic matroid* of X.

Example 17. Suppose $X = \mathbf{V}(f)$ for some nonzero $f \in \mathbb{K}[x_1, \ldots, x_n]$. Such a variety is called a *hypersurface* in $\mathbb{A}^n_{\mathbb{K}}$. If x_i is any variable appearing in f, then Theorem 3 implies that $B = \{x_1, \ldots, x_n\} - x_i$ is free on X. In particular, when X is irreducible, this set is a transcendence basis for $\mathbb{K}(X)$. It follows that the dimension of any (possibly reducible) hypersurface in $\mathbb{A}^n_{\mathbb{K}}$ is $n - 1$.

Exercise 13. Use the *primitive element theorem* from field theory to show that any affine variety is birationally equivalent to a hypersurface.

Let us revisit the problem of solving polynomial systems of equations from the geometric point of view. In order for this task to be feasible, we need to assume that the set of solutions to these equations is finite. Thus, we would like to characterize those ideals $I \subset \mathbb{K}[\mathbf{x}]$ such that $\mathbf{V}_{\mathbb{K}}(I)$ is a finite set. Our assumption that $\mathbb{K}$ is algebraically closed gives the following clean characterization.

Theorem 8. [Finiteness Theorem] Let $I \subset \mathbb{K}[x_1, \ldots, x_n]$ be an ideal, $<$ a monomial order. The following are equivalent:

1. For each x_i with $1 \leq i \leq n$, we have $x_i^{m_i} \in \mathrm{in}_<(I)$ for some $m_i \in \mathbb{Z}_{>0}$.
2. The set of standard monomials with respect to $<$ is finite.
3. The quotient ring $\mathbb{K}[\mathbf{x}]/I$ is a finite-dimensional vector space over $\mathbb{K}$.
4. The vanishing locus $\mathbf{V}_{\mathbb{K}}(I)$ is a finite set.
5. $\dim(\mathbf{V}_{\mathbb{K}}(I)) = 0$.

An ideal satisfying any one of the equivalent conditions of this theorem is said to be *zero dimensional*.

Remark The implications (1) ⇔ (2) ⇔ (3) ⇒ (4), (5) hold even when $\mathbb{K}$ is not algebraically closed.

Proof. (1) ⇒ (2): The hypothesis implies that all standard monomials have bounded degree in each variable, and hence there are finitely-many.
(2) ⇒ (3) Using Proposition 3.10, this follows from the isomorphism of $\mathbb{K}[\mathbf{x}]/I$ and $\mathrm{NF}_{I,<}(\mathbb{K}[\mathbf{x}])$ as vector spaces.
(3) ⇒ (4): For each x_i, there is a nontrivial $\mathbb{K}$-linear dependence on $1, x_i, \ldots, x_i^{m_i}$ in $\mathbb{K}[\mathbf{x}]/I$, which encodes a nonzero univariate polynomial $p_i(x_i) \in I$. Thus $\mathbf{V}(I) \subset \mathbf{V}(p_1, \ldots, p_n)$, a finite set.
(4) ⇒ (5): Suppose $\mathbf{V}(I) = \{p_1, \ldots, p_d\}$. Decomposing into irreducibles, it is enough to show that each component $\{p_i\}$ has dimension 0, which follows since no set of variables is free on $\{p_i\}$.
(5) ⇒ (1): Since $\{x_i\}$ is not free on $\mathbf{V}(I)$, we may take m_i to be the degree of some nonzero $p_i(x_i) \in I$. □

Proposition 6.5 (Shape Lemma)**.** If I is a radical zero-dimensional ideal, then $\dim_{\mathbb{K}}(\mathbb{K}[\mathbf{x}]/I)$ (ie. the number of standard monomials for any monomial order) is equal to cardinality of the set $\mathbf{V}_{\mathbb{K}}(I)$.

Proof. Consider a change of coordinates $\mathbf{y} = A\mathbf{x}$ given by $A \in \mathrm{GL}_n(\mathbb{K})$. Then $\dim_{\mathbb{K}}(\mathbb{K}[\mathbf{x}]/I) = \dim_{\mathbb{K}}(\mathbb{K}[\mathbf{y}]/A^*(I))$, where

$$A^*(I) = \langle f(A^{-1}\mathbf{y}) \mid f \in I \rangle.$$

For a *sufficiently generic coordinate change* A, we may assume that

$$\mathbf{V}(A^*(I)) = A \cdot \mathbf{V}(I) = \{p_1, \ldots, p_d\}$$

where the n-th coordinates of the points p_i, which we denote by $p_1^{(n)}, \ldots, p_d^{(n)}$, are *pairwise-distinct*. Now, if $<$ is a `Lex` order with y_n last, consider the polynomial $p(y_n) = \prod_{i=1}^{n} \left(y_n - p_i^{(n)}\right)$. Using Theorem 3, we can see that

$$A^*(I) \cap \mathbb{K}[y_n] = \langle p(y_n) \rangle.$$

The assumption that I is radical is needed so that $p(y_n) \in I$. Thus the monomials $1, y_n, \ldots, y_n^{d-1}$ are standard. It remains to show these are the *only* standard monomials. To see this, we construct for each $i = 1, \ldots, n-1$ a univariate polynomial $h_i(y_n)$ such that

$$y_i - h_i(y_n) \in A^*(I),$$

thus proving $y_i \in \mathrm{in}_<(A^*(I))$. This can be achieved using *Lagrange interpolation*—set

$$h_i(y_n) = \sum_{j=1}^{d} p_j^{(i)} \prod_{k \neq j} \frac{y_n - p_k^{(n)}}{p_j^{(n)} - p_k^{(n)}}.$$

□

7 Four ways to solve

Consider the hypersurface $\mathbf{V}(f) \subset \mathbb{A}^2_{\mathbb{C}}$ given by a bivariate polynomial of degree-4, $f \in \mathbb{C}[x, y]_{\leq 4}$. For concreteness, we consider the Trott curve

$$f = 144(x^4 + y^4) + 350x^2y^2 - 225(x^2 + y^2) + 81.$$

From algebraic geometry, we know that f has 28 real bitangent lines. How can we find them?

The solution we gave in class (see `lecture-2-15.m2`) developed the notions of *ideal quotients* and *saturations*. Let $I, J \subset \mathbb{K}[\mathbf{x}]$ be ideals. Define the ideal quotient

$$I : J = \{f \in \mathbb{K}[\mathbf{x}] \mid fJ \subset I\}$$

and saturation

$$I : J^\infty = \bigcup_{d=0}^{\infty} I : J^d.$$

An ascending chain argument shows that there exists an integer $d_{I,J}$ such that

$$I : J^\infty = I : J^{d_{I,J}}. \tag{31}$$

Proposition 7.1. The following relation holds

$$\sqrt{I} : J = \sqrt{I : J^\infty}. \tag{32}$$

Additionally, for $\mathbb{K}$ algebraically closed we have

1. $\mathbf{V}(I : J^\infty) = \overline{\mathbf{V}(I) \setminus \mathbf{V}(J)}$.
2. If $I = \mathbf{I}(X)$ is the vanishing ideal of a variety X, then $I : J$ is the vanishing ideal of $X \setminus \mathbf{V}(J)$.

Proof. First we prove 32. Suppose $f \in \sqrt{I : J^\infty}$, so that $f^n \in I : J^\infty$ for some n. Choosing $d_{I,J}$ as in 31, this implies for any $g \in J$ that $(fg)^{\max(n, d_{I,J})} \in I$. Hence $fg \in \sqrt{I}$. Since $g \in J$ was arbitrary, we have $f \in \sqrt{I} : J$ and conclude that $\sqrt{I : J^\infty} \subset \sqrt{I} : J$. For the reverse containment, write $J = \langle g_1, \ldots, g_s \rangle$. If $f \in \sqrt{I} : J$, then for some M we have $(fg_i)^M \in I$ for all $i = 1, \ldots, s$. Now, for any $g \in J$, it follows that $f^M g^{sM} \in I$, since expanding

g^{sM} in terms of $g_1, \ldots, g_s$ shows that $g^{sM} \in \langle g_1^M, \ldots, g_s^M \rangle$. So $f^M \in I : J^{sM} \subset I : J^\infty$ gives $f \in \sqrt{I : J^\infty}$, as needed.

Now, to prove 1, it suffices to show that $\mathbf{V}(\sqrt{I} : J) = \overline{\mathbf{V}(I) \setminus \mathbf{V}(J)}$. To see this, we show that $\sqrt{I} : J$ is the vanishing ideal of $\overline{\mathbf{V}(I) \setminus \mathbf{V}(J)}$, which by the Nullstellensatz also implies the second part. Suppose f vanishes on all points of $\mathbf{V}(I) \setminus \mathbf{V}(J)$. Then for any $g \in J$ we have that fg vanishes on all points of $\mathbf{V}(I)$, and hence $f \in I : J \subset \sqrt{I} : J$. Conversely, if $f \in \sqrt{I} : J$, then fg vanishes on $\mathbf{V}(I)$ for all $g \in J$. In particular, this holds whenever g does not vanish on some point of $\mathbf{V}(I) \setminus \mathbf{V}(J)$, implying f must vanish on all of these points. $\square$