POLYNOMIAL INTERPOLATION: AN INTRODUCTION TO ALGEBRAIC GEOMETRY

JACK HUIZENGA

Abstract. These are notes for the Fall 2018 Honors MASS Algebra course at Penn State. They cover a one-semester advanced undergraduate course in linear algebra and algebraic geometry.

Contents

1. Lagrangian Interpolation

First let's fix some notation. Let $x_1, \dots, x_n \in \mathbb{R}$ be distinct *points*, and let $y_1, \dots, y_n \in \mathbb{R}$ be arbitrary *values*. The *interpolation problem* for this data is to find a function $f : \mathbb{R} \to \mathbb{R}$ which takes the prescribed values at the points:

$$f(x_i) = y_i \qquad (i = 1, \dots, n).$$

Already in high school algebra, students want to solve the interpolation problem. How can we find a function which takes the prescribed values at the prescribed points? On the face of it this question is not really very interesting: a function $f : \mathbb{R} \to \mathbb{R}$ is an arbitrary assignment of a value $f(x) \in \mathbb{R}$ for each $x \in \mathbb{R}$, so we can for instance construct a

Date: January 24, 2019.

function which has $f(x_i) = y_i$ for all i by defining

$$f(x) = \begin{cases} y_i & \text{if } x = x_i \\ 0 & \text{otherwise.} \end{cases}$$

There is nothing particular about the value 0 here; we can instead (independently) assign whatever value we like at any points x different from $x_1, \ldots, x_n$. The set of all functions $f : \mathbb{R} \to \mathbb{R}$ is enormous, and contains all kinds of nasty functions that no high schooler would be happy with. Their graphs look like the static fuzz of an old TV screen.

Of course what the high schooler really wants to do is come up with a "nice," "named" function $f : \mathbb{R} \to \mathbb{R}$ that satisfies $f(x_i) = y_i$. They don't want an f defined piecewise, and hopefully f is differentiable. For the interpolation problem to be interesting, we need to restrict the class of functions that we are considering. The simplest, most natural family of functions to look at are the polynomial functions.

1.1. **Basics on polynomials, functions, and polynomial functions.**

Definition 1.1 (Polynomials)**.** A *polynomial* p (with real coefficients) is an expression of the form

$$p = a_n x^n + \cdots + a_1 x + a_0$$

where the $a_i \in \mathbb{R}$.

The set of all such polynomials is denoted by $\mathbb{R}[x]$. We can *add* polynomials coefficient-by-coefficient: if $p = \sum_{i=0}^{n} a_i x^i$ and $q = \sum_{i=0}^{m} b_i x^i$, then

$$p + q = \sum_{i=0}^{\max\{m,n\}} (a_i + b_i) x^i.$$

The *product* of p and q is the polynomial

$$pq = \sum_{k=0}^{m+n} c_k x^k \qquad \text{with} \qquad c_k = \sum_{i+j=k} a_i b_j.$$

With these operations, the set $\mathbb{R}[x]$ is what is called a *ring* in algebra. (Basically, this just means that the usual rules you are accustomed to for addition and multiplication hold. For example, $p(q+r) = pq + pr$.)

Two polynomials p and q are equal iff they are given by the same expression. In other words, their coefficients are equal.

Definition 1.2 (Functions)**.** Let X and Y be sets. A function $f : X \to Y$ assigns to each $x \in X$ an element $f(x) \in Y$. Two functions

$f, g : X \to Y$ are equal iff $f(x) = g(x)$ for all $x \in X$. (This is just set theory.)

Definition 1.3 (Real-valued functions)**.** Let X be a set and consider functions $f : X \to \mathbb{R}$. (A particularly interesting case is where also $X = \mathbb{R}$.) The set of all such functions is denoted by $\mathbb{R}^X$. We can use the addition and multiplication operations on $\mathbb{R}$ to define addition and multiplication operations on $\mathbb{R}^X$. Specifically, if $f, g : X \to \mathbb{R}$ then we define functions $f + g$ and fg *pointwise* by the rules

$$(f + g)(x) := f(x) + g(x)$$
$$(fg)(x) := f(x)g(x).$$

With these operations, the set $\mathbb{R}^X$ again becomes an example of a ring.

Definition 1.4 (Polynomial functions)**.** Given a polynomial $p \in \mathbb{R}[x]$, we can substitute a real value $t \in \mathbb{R}$ for the variable x. Thus if $p = \sum a_i x^i$, we can define a function $f : \mathbb{R} \to \mathbb{R}$ by the rule

$$f(t) = \sum a_i t^i.$$

The function f is called a *polynomial function.* It can be *represented by the polynomial* p. The set of all polynomial functions is a subset of $\mathbb{R}^{\mathbb{R}}$, the set of all functions from $\mathbb{R}$ to $\mathbb{R}$.

Warning 1.5. The polynomial p representing a polynomial function $f : \mathbb{R} \to \mathbb{R}$ is unique, but this requires proof. (See below.) In particular, two polynomial functions $f, g : \mathbb{R} \to \mathbb{R}$ are equal if and only if the polynomials p, q representing them are equal. Polynomial functions $f : \mathbb{R} \to \mathbb{R}$ and polynomials $p \in \mathbb{R}[x]$ are "basically" the same thing, but in more general contexts this requires some care.

Definition 1.6 (Degree)**.** The *degree* of the polynomial $p = a_n x^n + \cdots + a_0$ is $\deg p = n$ if $a_n \neq 0$. We define $\deg 0 = -\infty$ by convention. The degree of a polynomial function is the degree of a polynomial which represents it.

Lemma 1.7 (Basic properties of degree)**.** *Let* $p, q \in \mathbb{R}[x]$.

(1) We have $\deg(pq) = \deg p + \deg q$.
(2) We have $\deg(p+q) \leq \max\{\deg p, \deg q\}$. *If* $\deg p \neq \deg q$, *then equality holds.*

1.2. **Lagrangian interpolation: existence.** Back to our motivating question. Let $x_1, \ldots, x_n \in \mathbb{R}$ be distinct points, and let $y_1, \ldots, y_n \in \mathbb{R}$ be arbitrary values. How can we construct a polynomial function $f : \mathbb{R} \to \mathbb{R}$ such that $f(x_i) = y_i$ for all i?

First, let's solve a special case of this problem. What if $y_1 = 1$ and all other $y_i = 0$? We can easily arrange that a polynomial function $f(x)$ has $f(x_i) = 0$ by giving it a factor of $(x - x_i)$. Thus, the polynomial function

$$f(x) = (x - x_2) \cdots (x - x_n)$$

has $f(x_i) = 0$ for $i = 2, \ldots, n$, and $f(x_1) \neq 0$. To get a value of 1 at x_1, we can just divide by the constant $f(x_1)$. Therefore

$$F_1(x) = \frac{(x - x_2) \cdots (x - x_n)}{(x_1 - x_2) \cdots (x_1 - x_n)}$$

is a polynomial function such that $F_1(x_1) = 1$ and $F_1(x_i) = 0$ for $i \neq 1$.

Similarly, we can define for each i a polynomial function $F_i(x)$ by the rule

$$F_i(x) = \prod_{\substack{j=1 \\ j \neq i}}^{n} \frac{(x - x_j)}{(x_i - x_j)}.$$

Then

$$F_i(x_j) = \delta_{ij} := \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

We can piece these special solutions together to prove the following theorem.

Theorem 1.8 (Lagrangian interpolation: existence)**.** *Let $x_1, \ldots, x_n \in \mathbb{R}$ be distinct points, and let $y_1, \ldots, y_n \in \mathbb{R}$ be arbitrary values. Then there exists a polynomial function $f : \mathbb{R} \to \mathbb{R}$ of degree at most $n - 1$ such that $f(x_i) = y_i$.*

Proof. In the notation preceding the statement, the function

$$\sum_{i=1}^{n} y_i F_i$$

has the required properties. □

The basic method of proof used here was to first find a polynomial that vanishes at all but one of the points, to construct a "characteristic function" supported at just one of the points. Then it is a simple matter to combine these various functions to get whatever value we want at each of the points. This technique continues to be useful in more general problems, as we will see.

Could there have been other solutions to the interpolation problem? If we don't bound the degree of the function, then yes. Notice that the

polynomial $(x - x_1) \cdots (x - x_n)$ vanishes at all the points $x_1, \ldots, x_n$. Therefore, if $f(x)$ is any solution to the interpolation problem, then

$$f(x) + (x - x_1) \cdots (x - x_n)$$

is also a solution. More generally, if g is any polynomial, then

$$f(x) + (x - x_1) \cdots (x - x_n)g(x)$$

is also a solution. However, if $\deg f \leq n - 1$, then all these other solutions have degree n or greater, so we have only constructed one solution of degree at most $n - 1$.

1.3. **The division algorithm.** It can't be overstated how important the division algorithm is in the study of 1-variable polynomials. In algebra terminology, $\mathbb{R}[x]$ is what is called a *Euclidean domain*, and it is one of the simplest possible kinds of rings out there. This fact is directly responsible for the simple story of Lagrangian interpolation.

Theorem 1.9 (The division algorithm)**.** *Let $p, m \in \mathbb{R}[x]$ with $m \neq 0$. Then there are unique polynomials $q, r \in \mathbb{R}[x]$ (the* quotient *and* remainder*) with $\deg r < \deg m$ such that*

$$p = qm + r.$$

The proof of the theorem is essentially an analysis of the high school algorithm for polynomial long division.

Proof. (Existence) First we show that we can find such polynomials q and r. We go by (strong) induction on the degree of p. If $\deg p < \deg m$ we take $q = 0$ and $r = p$. Suppose $\deg p \geq \deg m$ and that the result is known for polynomials p' of degree $\deg p' < \deg p$.

Since $\deg p \geq \deg m$ and $m \neq 0$, there is a monomial cx^a such that

$$p' := p - cx^a m$$

has degree $< \deg p$. (Choose c and a to cancel the leading term of p.) Then by induction we can write

$$p' = q'm + r' \qquad (\deg r' < \deg m).$$

But then

$$p = p' + cx^a m = (q' + cx^a)m + r'.$$

That is, if we put $q = q' + cx^a$ and $r' = r$, then

$$p = qm + r \qquad (\deg r < \deg m).$$

Therefore, it is possible to write p in the required way.

(Uniqueness) Suppose we can write

$$p = qm + r = q'm + r'$$

where both r, r' have degree less than $\deg m$. Then

$$r - r' = (q' - q)m.$$

The LHS has degree $< \deg m$. If $q' \neq q$, the RHS has degree $\geq \deg m$. Therefore, $q' = q$, and so $r = r'$. Therefore the expression is unique. □

The division algorithm allows us to conclude that polynomials with roots have a very special form.

Proposition 1.10. *Suppose $p \in \mathbb{R}[x]$ is a polynomial, $a \in \mathbb{R}$, and $p(a) = 0$. Then $x - a$ divides p: we can write $p = (x - a)q$ for some $q \in \mathbb{R}[x]$.*

Proof. Use the division algorithm to divide p by $x - a$. We find that there are polynomials q and r with $p = (x - a)q + r$, and $\deg r \leq 0$. Therefore r is a constant. Plugging a in to both sides forces $r = 0$. □

Corollary 1.11. *If a polynomial $p \in \mathbb{R}[x]$ of degree at most $n - 1$ has n zeroes, then $p = 0$.*

Proof. Suppose $p(a_i) = 0$ for $i = 1, \ldots, n$. Using the proposition repeatedly, we can write

$$p = (x - a_1) \cdots (x - a_n)q$$

for some $q \in \mathbb{R}[x]$. This forces either $p = 0$ or $\deg p \geq n$, but since $\deg p \leq n - 1$ we conclude $p = 0$. □

Corollary 1.12. *If the polynomial functions given by two polynomials p, q are equal, then $p = q$.*

Proof. The difference $p - q$ has infinitely many zeroes, so $p - q = 0$. □

The consequences of the division algorithm allow us to immediately settle the uniqueness question for Lagrangian interpolation.

Theorem 1.13 (Lagrangian interpolation: existence and uniqueness)**.** *Let $x_1, \ldots, x_n \in \mathbb{R}$ be distinct points, and let $y_1, \ldots, y_n \in \mathbb{R}$ be arbitrary values. There is a unique polynomial function $f : \mathbb{R} \to \mathbb{R}$ of degree at most $n - 1$ such that $f(x_i) = y_i$ for all i.*

Proof. We already showed that such a function exists. Suppose f and g are two such functions. Then $f - g$ has n zeroes and degree at most $n - 1$. Therefore $f - g = 0$ and $f = g$. □

We can also address the question of arbitrary solutions of the interpolation problem, without bounding the degree.

Theorem 1.14 (Lagrangian interpolation: all solutions)**.** *Let $x_1, \dots, x_n \in \mathbb{R}$ be distinct points, and let $y_1, \dots, y_n \in \mathbb{R}$ be arbitrary values. Let f be a polynomial function such that $f(x_i) = y_i$ for all i. If g is any other polynomial function such that $g(x_i) = y_i$ for all i, then there is a unique polynomial q such that*

$$g = f + (x - x_1) \cdots (x - x_n)q.$$

Thus, any solution to Lagrangian interpolation can be obtained from a "particular solution" by adding on a multiple of $(x - x_1) \cdots (x - x_n)$.

Proof. Use the division algorithm to divide $g - f$ by $(x - x_1) \cdots (x - x_n)$. Then there are unique polynomials q, r with $\deg r < n$ such that

$$g - f = (x - x_1) \cdots (x - x_n)q + r.$$

Plugging in x_i shows $r(x_i) = 0$ for all i, and therefore r has n roots. Therefore $r = 0$. □

1.4. **Repeated points.** In the statement of Lagrangian interpolation we required all the points to be distinct. What happens in the limiting scenario where two of the points "collide" with one another? Suppose we have two points $x_1, x_2 \in \mathbb{R}$ and

$$x_2 = x_2(t) = x_1 + t,$$

so that x_2 depends on the parameter t in a linear way. As $t \to 0$, we have $x_2(t) \to x_1$. Also suppose we have two values y_1 and $y_2(t)$, where y_2 also depends on the parameter t. Suppose $y_2(t) \in \mathbb{R}[t]$ is a polynomial in t.

Now suppose we have for each time t a polynomial function

$$f_t(x) = a_n(t)x^n + a_{n-1}(t)x^{n-1} + \cdots + a_0(t)$$

such that

$$f_t(x_1) = y_1$$
$$f_t(x_2(t)) = y_2(t),$$

and suppose the $a_i(t) \in \mathbb{R}[t]$ are polynomials in t.

Question 1.15. *In this setup, what does the polynomial $f_0(x)$ "look like?"*

Example 1.16. Take $x_1 = 0$, so $x_2(t) = t$. Let $y_1 = 0$ and $y_2(t) = t + t^2$. There are many possible choices of the family $f_t(x)$ of polynomials; for example, both

$$f_t(x) = x + x^2$$
$$f_t(x) = (1 + t)x$$

work. The time $t = 0$ polynomials are $f_0(x) = x + x^2$ and $f_0(x) = x$, respectively.

First observe that this setup forces that $y_2(0) = y_1$. Indeed, we have

$$y_2(0) = f_0(x_2(0)) = f_0(x_1) = y_1.$$

That is, as the points x_1 and x_2 collide, the values y_1 and y_2 also have to collide in order for there to be a *function* which takes on value y_1 at x_1 and value y_2 at x_2.

What is the *derivative* of f_0 at x_1? Intuitively, for any small time t the polynomial f_t takes value y_1 at x_1 and value $y_2(t)$ at $x_2 = x_1 + t$. Therefore the graph of f_t has a secant line passing between the points (x_1, y_1) and $(x_1 + t, y_2(t))$. The slopes of these secant lines are

$$\frac{y_2(t) - y_1}{(x_1 + t) - x_1} = \frac{y_2(t) - y_2(0)}{t}.$$

As $t \to 0$ these slopes converge to the derivative $y_2'(0)$, and thus we should expect that the polynomial $f_0(x)$ has derivative $y_2'(0)$ at x_1! The remarkable thing here is that the particular family of polynomials $f_t(x)$ did not matter at all—all that matters is the way in which the value $y_2(t)$ approaches y_1.

Proposition 1.17. *Let $x_1 \in \mathbb{R}$, write $x_2(t) = x_1 + t$, let $y_2(t) \in \mathbb{R}[t]$, and put $y_1 = y_2(0)$. Suppose $f_t(x) = \sum_{i=0}^{n} a_i(t)x^i$ for some $a_i(t) \in \mathbb{R}[t]$. If $f_t(x_1) = y_1$ and $f_t(x_2(t)) = y_2(t)$, then*

$$f_0'(x_1) = y_2'(0).$$

Proof. Without loss of generality, we may shift and translate the picture so that $x_1 = 0$ and $y_1 = 0$. Then $a_0(t) = 0$, so we can write $f_t(x) = a_1(t)x + a_2(t)x^2 + \cdots$, and

$$f_0'(0) = a_1(0).$$

Since $y_1 = 0$ we can write $y_2(t) = b_1 t + b_2 t^2 + \cdots$, so that $y_2'(0) = b_1$. We are given that $f_t(t) = y_2(t)$. Plugging in t into $f_t(x)$, all the terms are of degree ≥ 2 except for $a_1(0)t$. Therefore $a_1(0) = b_1$, and $f_0'(0) = y_2'(0)$. □

In other words, as two points collide together the limiting interpolation problem is to specify the derivative of the polynomials at the limiting point.

1.5. **Derivative data.** Motivated by the previous computation, we introduce a generalized version of Lagrangian interpolation that allows for "repeated points." Suppose we are given (still distinct) points $x_1, \ldots, x_n$, multiplicities $m_1, \ldots, m_n$, and for each point x_i a list $y_{i,0}, \ldots, y_{i,m_i-1}$ of m_i values. We seek to find a polynomial $f(x)$ such that for each i and each $0 \leq j \leq m_i - 1$,

$$f^{(j)}(x_i) = y_{i,j}.$$

Here $f^{(j)}$ denotes the jth derivative, as usual.

Thus, we are prescribing the value of f and its first $m_i - 1$ derivatives at the point x_i. The original Lagrangian interpolation problem is recovered when all the multiplicities are 1. The number of equalities that f is required to satisfy is $m_1 + \cdots + m_n$.

To solve this version of the interpolation problem, you'll first prove a basic fact about polynomials.

Lemma 1.18. *Let $f \in \mathbb{R}[x]$, $a \in \mathbb{R}$, and let $m \geq 1$. Then $f^{(k)}(a) = 0$ for all $0 \leq k \leq m-1$ if and only if $(x-a)^m$ divides f.*

Proof. Homework. □

Theorem 1.19 (Lagrangian interpolation with derivatives)**.** *With the above setup, there exists a unique polynomial $f(x)$ of degree at most $\sum m_i - 1$ such that $f^{(j)}(x_i) = y_{i,j}$ for all i, j.*

Proof. (Existence) As with ordinary Lagrangian interpolation, it is helpful to focus on one point at a time. It will be enough to construct a polynomial $f(x)$ such that the values of the derivatives at the first point are all correct, but the values of the derivatives at all of the other points are 0. Then, to get the desired values at all the points we can add together the solutions for each of these sub-problems.

If all the required derivatives of f at the points $x_2, \ldots, x_n$ vanish, this means exactly that f is divisible by $g := (x - x_2)^{m_2} \cdots (x - x_n)^{m_n}$. Let us consider the m_1 different polynomials

$$f_k(x) := (x - x_1)^k g,$$

where $0 \leq k \leq m_1 - 1$. We claim that some linear combination

$$F = a_0 f_0 + \cdots + a_{m_1-1} f_{m_1-1}$$

satisfies $F^{(k)}(x_1) = y_{1,k}$ for each $0 \leq k \leq m_1 - 1$. Note that F has degree at most $\sum m_i - 1$, as required.

The list of equations $F^{(k)}(x_1) = y_{1,k}$ is a system of m_1 linear equations in the m_1 variables $a_0, \dots, a_{m_1-1}$. More precisely, it is the system

$$\begin{aligned} f_0(x_1)a_0 + \cdots + f_{m_1-1}(x_1)a_{m_1-1} &= y_{1,0} \\ f_0'(x_1)a_0 + \cdots + f_{m_1-1}'(x_1)a_{m_1-1} &= y_{1,1} \\ &\vdots \\ f_0^{(m_1-1)}(x_1)a_0 + \cdots + f_{m_1-1}^{(m_1-1)}(x_1)a_{m_1-1} &= y_{1,m_1-1}. \end{aligned}$$

Or, in matrix form,

$$\begin{pmatrix} f_0(x_1) & \cdots & f_{m_1-1}(x_1) \\ f_0'(x_1) & \cdots & f_{m_1-1}'(x_1) \\ \vdots & \ddots & \vdots \\ f_0^{(m_1-1)}(x_1) & \cdots & f_{m_1-1}^{(m_1-1)}(x_1) \end{pmatrix} \begin{pmatrix} a_0 \\ \vdots \\ a_{m_1-1} \end{pmatrix} = \begin{pmatrix} y_{1,0} \\ y_{1,1} \\ \vdots \\ y_{1,m_1-1} \end{pmatrix}.$$

But, our choice of the polynomials $f_k(x)$ makes this matrix have a very simple form: since $f_k^{(i)}(x_1) = 0$ for $i < k$, the matrix is lower triangular. Furthermore, since f_k is divisible by $(x - x_1)$ exactly k times, the diagonal entries are nonzero. Then it is a simple matter to solve for the coefficients a_i by back-substitution: the first equation $f_0(x_1)a_0 = y_{1,0}$ lets us read off a_0, then the second equation $f_0'(x_1)a_0 + f_1'(x_1)a_1 = y_{1,1}$ gives us a_1, and so in. Thus it is possible to find numbers $a_0, \dots, a_{m_1-1}$ such that $F^{(k)}(x_1) = y_{1,k}$ for each $0 \le k \le m_1 - 1$. And of course, since g divides F all the required derivatives of F at the other points vanish.

As discussed at the beginning of the proof, we can carry about the above procedure for each point x_i and add the resulting solutions to get the required polynomial.

(Uniqueness) Given two solutions f and g to the problem, the difference $f - g$ has $(f-g)^{(k)}(x_i) = 0$ for all i and $0 \le k \le m_i - 1$. Therefore $\prod (x - x_i)^{m_i}$ divides $f - g$. But since the degree of $f - g$ is at most $\sum m_i - 1$, this is only possible if $f - g = 0$. Therefore $f = g$. $\square$

2. Linear Algebra: Vector Spaces

While we were able to solve the Lagrangian interpolation problem with purely elementary techniques, linear algebra will give a unifying framework of techniques that greatly simplifies the picture. In both the original interpolation problem and the interpolation problem with derivatives, the existence proof was relatively painful and the uniqueness proof was relatively easy. In fact, with more knowledge and a better point of view the existence proof can be seen to be essentially unnecessary!

2.1. **A note on fields.** In section 1 we stated all our results for polynomials over the real numbers. This is unnecessarily restrictive, and there are good reasons to allow other fields. Algebraically closed fields, (e.g. the field $\mathbb{C}$ of complex numbers) make the study of polynomial factorization much easier. Finite fields (e.g. the field of integers mod a prime p, $\mathbb{Z}/p\mathbb{Z}$) are frequently useful in number theory and cryptography. In what follows we let K be an arbitrary field, but you should feel free to think of $K = \mathbb{R}$ or $\mathbb{C}$ if it makes you happier. (The choice of letter K is traditional: the German word for field is *Körper*).

2.2. **Systems of linear equations.** On a first approach, linear algebra is fundamentally about the solution of systems of linear equations. This is hopefully a review, but it is fundamental to everything that follows so we have to be solid on it.

Consider a system of m linear equations in n unknowns

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

Here the numbers $a_{ij}, b_i \in K$ are constants, and we seek to find all tuples $(x_1, \ldots, x_n)$ which satisfy all of the equations. It is customary to encode this system in a vector equation as

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix},$$

or more compactly as

$$A\mathbf{x} = \mathbf{b},$$

where A is the $m \times n$ matrix (a_{ij}) and $\mathbf{x} \in K^n$ and $\mathbf{b} \in K^m$ are the evident vectors of length n and m, respectively. Thus, we seek to

describe the solution set

$$\{\mathbf{x} \in K^n : A\mathbf{x} = \mathbf{b}\}.$$

Definition 2.1. The system $A\mathbf{x} = \mathbf{b}$ is called *consistent* if it has a solution, and *inconsistent* otherwise.

2.2.1. *Row echelon form and back-substitution.* First let us study what the solutions of the system $A\mathbf{x} = \mathbf{b}$ are in a case where the coefficient matrix A has a particularly nice form. We say that the (i, j) entry (row i, column j) is a *pivot* of A if $a_{ij} = 1$ and all the entries in row i left of the (i, j) entry are zero.

Definition 2.2. Let $A = (a_{ij})$ be an $m \times n$ matrix, and suppose it has exactly k nonzero rows. We say A is in *row echelon form* if every nonzero row has a pivot, and they are in positions $(1, \ell_1), \dots, (k, \ell_k)$, where the sequence $\ell_1 < \ell_2 < \cdots < \ell_k$ is strictly increasing.

We say that the variables $x_{\ell_1}, \dots, x_{\ell_k}$ corresponding the columns where there are pivots are *bound variables*. The variables among $x_1, \dots, x_n$ which are not bound are *free*. The number k is the *rank* of the matrix.

Note that if row i in the matrix A is all zeroes, then the system $A\mathbf{x} = \mathbf{b}$ is inconsistent unless $b_i = 0$. For row echelon matrices, the consistency of the system can be read off using this observation. The next result also explains the terminology of bound and free variables: if the system is consistent, then the free variables can have their values assigned arbitrarily and we can solve for the bound variables by back-substitution.

Proposition 2.3. *Suppose A is an $m \times n$ matrix in row echelon form, and that A has rank k. Let $\mathbf{b} = (b_1, \dots, b_m) \in K^m$.*

1. *The system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $b_{k+1} = \cdots = b_m = 0$.*
2. *Suppose the system $A\mathbf{x} = \mathbf{b}$ is consistent. For each free variable x_i, let $c_i \in K$ be any constant. Then for each bound variable x_j there is a uniquely determined constant $c_j \in K$ such that the vector $\mathbf{c} = (c_1, \dots, c_n)$ has $A\mathbf{c} = \mathbf{b}$.*

Proof. We saw before the proof that if $A\mathbf{x} = \mathbf{b}$ is consistent then $b_{k+1} = \cdots = b_m = 0$.

Conversely, suppose $b_{k+1} = \cdots = b_m = 0$, and assign arbitrary fixed values to each of the free variables. Read off the system of equations from the bottom up. The equations in rows $k + 1$ through m all say $0 = 0$, so are satisfied.

Suppose the pivots of A are in positions $(1, \ell_1), \ldots, (1, \ell_k)$. In row k, the system reads

$$x_{\ell_k} + a_{k,\ell_k+1} x_{\ell_k+1} + \cdots + a_n x_n = b_k.$$

But, all the variables $x_{\ell_k+1}, \ldots, x_n$ are free, so we are seeking a solution where those variables take the corresponding assigned values. Thus, we know x_{ℓ_k} has to satisfy

$$x_{\ell_k} + a_{k,\ell_k+1} c_{\ell_k+1} + \cdots + a_n c_n = b_k.$$

Thus the only possible value for x_{ℓ_k} is

$$x_{\ell_k} = b_k - (a_{k,\ell_k+1} c_{\ell_k+1} + \cdots + a_n c_n) =: c_{\ell_k},$$

so the value of x_{ℓ_k} has been forced on us.

Next, in row $k-1$, we similarly read the equation

$$x_{\ell_{k-1}} + a_{k-1,\ell_{k-1}+1} x_{\ell_{k-1}+1} + \cdots + a_n x_n = b_{k-1}.$$

The values of each of the variables other than $x_{\ell_{k-1}}$ are known (either because they are free or because they were determined in the previous step), and so the value of $x_{\ell_{k-1}}$ is readily determined.

Continuing in this fashion, we can uniquely determine the values the bound variables which solve the system. □

Thus systems where the coefficient matrix are in row echelon form are readily solved by back-substitution.

2.2.2. *Gaussian elimination.* The most important algorithm in linear algebra allows us to convert a linear system $A\mathbf{x} = \mathbf{b}$ to a different linear system $A'\mathbf{x} = \mathbf{b}'$ where A' is in row echelon form and the solution set is the same. Then since we know how to read off the solutions of a row echelon system, we can solve a general system $A\mathbf{x} = \mathbf{b}$.

Lemma 2.4 (Row operations)**.** *The solution set of a system $A\mathbf{x} = \mathbf{b}$ is unchanged by the following transformations.*

(1) Scale an equation by a nonzero constant.
(2) Add one equation to another.
(2') Add a scalar multiple of one equation to another.
(3) Swap the position of two equations.

Proof. Let $A\mathbf{x} = \mathbf{b}$ denote the original system, and let $A'\mathbf{x} = \mathbf{b}'$ denote the system obtained by performing one of the above operations. Let $S = \{\mathbf{x} \in K^n : A\mathbf{x} = \mathbf{b}\}$, and let $T = \{\mathbf{x} \in K^n : A'\mathbf{x} = \mathbf{b}'\}$. We need to show that $S = T$ in each case.

It is clear that operations (1) and (3) do not affect the solution set.

To see that operation (2) does not affect the solution set, notice that if $\mathbf{x} \in S$ then $\mathbf{x} \in T$. Indeed, if $\mathbf{x}$ satisfies two equations it also

satisfies the sum of those two equations. Therefore $S \subset T$. Conversely, if $\mathbf{x} \in T$, we can see that $\mathbf{x} \in S$ since $\mathbf{x}$ will satisfy the difference of any two equations in T, and we can transform our system back to the original system by subtracting one equation from another. Therefore, $S = T$.

Operation (2') can be seen to be a combination of operations of type (1) and (2), so it preserves the solution set. □

When the corresponding operations in the lemma are performed on the coefficient matrix A, they are called *elementary row operations.*

Theorem 2.5 (Gaussian elimination)**.** *By using the operations of Lemma 2.4, any system $A\mathbf{x} = \mathbf{b}$ can be replaced by a system $A'\mathbf{x} = \mathbf{b}'$ such that*

(1) the solution sets are the same, and
(2) A' is in row echelon form.

Proof. Call two matrices A and A' *row equivalent* if you can obtain A' from A by a sequence of elementary row operations.

Say that the *reduction number* of a matrix A is k if $0 \leq k \leq m$ is the largest integer such that

(1) the first k rows of A are a row echelon matrix,
(2) any rows of zeroes in A are the last rows in the matrix, and
(3) if there are any nonzero entries of A in rows $k+1, \ldots, m$, then they lie to the right of the leading nonzero entry in row k.

Note that A is in row echelon form if and only if its reduction number is m. Intuitively, the larger the reduction number, the closer A is to being in row echelon form.

Now let A' be any matrix which is row equivalent to A and such that the reduction number of A' is *as large as possible.* We claim that A' is actually in reduced row echelon form, i.e. its reduction number is m.

Suppose the reduction number of A' is $k < m$. The rows $k+1, \ldots, m$ can't all be zero, since then A' is actually in row echelon form and its reduction number is m. One of these rows has a leading nonzero entry which is as far left as possible. Swap that row with row $k+1$. Scalar multiply row $k+1$ so the leading entry is 1. This entry will be our $(k+1)$st pivot. Now for each row from $k+2$ to m, subtract the appropriate multiple of row $k+1$ to clear the entries below the pivot. The resulting matrix is row equivalent to A' (and hence row equivalent to A) and has reduction number at least $k+1$. This contradicts our choice of A'.

Thus by a sequence of elementary row operations we can convert A to a row echelon form matrix A'. Performing the corresponding sequence

of operations on the system of equations $A\mathbf{x} = \mathbf{b}$, we get the required system $A'\mathbf{x} = \mathbf{b}'$ with the same solution set. □

It is really hard to overstate how important the following fact is; we will eventually have more impressive sounding versions of this statement (see the "rank-nullity theorem") but the core foundations of linear algebra depend crucially on the next simple fact.

Theorem 2.6. *Consider the homogeneous system of equations*

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= 0 \end{aligned}$$

where $a_{ij} \in K$. If $n > m$, then this system has a nonzero solution.

Proof. Use Gaussian elimination to replace the system with an equivalent system $A'\mathbf{x} = \mathbf{0}$ where the coefficient matrix is in row echelon form. (Notice that the RHS of the equation remains $\mathbf{0}$ after each row operation.) The number of bound variables is then at most m, the number of rows, since there is at most one pivot in each row. Then there are at least $n - m$ free variables. The system is consistent, so if we assign values to the free variables (at least some of them nonzero) then we can determine values of the bound variables that solve the system. □

2.3. **Polynomial interpolation and systems of equations.** Familiar polynomial interpolation questions can be rephrased in terms of solutions of linear systems of equations. For example, consider the original Lagrangian interpolation question. Let $x_1, \ldots, x_n \in K$ be distinct and let $y_1, \ldots, y_n \in K$. Consider the problem of finding a polynomial $f \in K[x]$ of degree at most $n - 1$ such that $f(x_i) = y_i$ for all i. Write down a general polynomial of degree at most $n - 1$ as

$$f = a_{n-1}x^{n-1} + \cdots + a_0.$$

Then we want f to satisfy the system of n equations $f(x_i) = y_i$ ($i = 1, \ldots, n$). Written out more explicitly, we need to find $a_{n-1}, \ldots, a_0 \in K$ such that

$$\begin{aligned} a_{n-1}x_1^{n-1} + \cdots + a_0 &= 0 \\ a_{n-1}x_2^{n-1} + \cdots + a_0 &= 0 \\ &\vdots \\ a_{n-1}x_{n-1}^{n-1} + \cdots + a_0 &= 0. \end{aligned}$$

Alternately, in matrix form, we want to solve

$$\begin{pmatrix} x_1^{n-1} & \cdots & x_1 & 1 \\ x_2^{n-1} & \cdots & x_2 & 1 \\ \vdots & \ddots & \vdots & \vdots \\ x_n^{n-1} & \cdots & x_n & 1 \end{pmatrix} \begin{pmatrix} a_{n-1} \\ \vdots \\ a_2 \\ a_0 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

(The matrix on the left is called a *Vandermonde matrix*). We showed that for any RHS vector $\mathbf{y} = (y_1, \ldots, y_n) \in K^n$ there exists a unique solution vector $\mathbf{a} = (a_{n-1}, \ldots, a_1) \in K^n$. In other words, after a row reduction the matrix on the left can be brought into a row echelon form with n pivots (there can't be any free variables). Notice that this last fact depends on the assumption that $x_1, \ldots, x_n \in K$ are distinct: if two of them are the same, then the matrix will have two rows that are the same, and we can get a row of 0's after a row reduction.

In this way we see that a polynomial can in many ways be regarded as its list or vector of coefficients. We add two polynomials coefficient-by-coefficient, so that the addition of polynomials behaves in the same way as the addition of vectors. Since the conversion process between polynomials and coefficient vectors is a bit annoying and not worth repeating every time we need to do it, it is preferably to work with the polynomials directly. The concept of a vector space will allow us to do this.

2.4. **Vector spaces.** A vector space is an abstract algebraic structure that shares many of the familiar properties of the n-dimensional Euclidean space K^n.

Definition 2.7. Let K be a field, and let V be a set. Suppose there are *addition* and *scalar multiplication* operations on V: for all $\mathbf{v}, \mathbf{w} \in V$ and $\lambda \in K$, there are elements $\mathbf{v} + \mathbf{w} \in V$ and $\lambda \cdot \mathbf{v} \in V$. Then V is a *vector space over* K if the following axioms are satisfied.

(1) $(V, +)$ is an abelian group:
   (a) $+$ is commutative: $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$ for all $\mathbf{v}, \mathbf{w} \in V$.
   (b) $+$ is associative: $(\mathbf{v}+\mathbf{w})+\mathbf{u} = \mathbf{v}+(\mathbf{w}+\mathbf{u})$ for all $\mathbf{v}, \mathbf{w}, \mathbf{u} \in V$.
   (c) There is an element $\mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$.
   (d) For any $\mathbf{v} \in V$, there is some $\mathbf{w} \in V$ such that $\mathbf{v} + \mathbf{w} = \mathbf{0}$.
(2) The operation $\cdot$ is compatible with the addition, in the sense that
   (a) $1 \cdot \mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in V$, and

(b) $\lambda(\mathbf{v}+\mathbf{w}) = \lambda\mathbf{v}+\lambda\mathbf{w}$ and $(\lambda+\mu)\mathbf{v} = \lambda\mathbf{v}+\mu\mathbf{v}$ for all $\lambda, \mu \in K$ and $\mathbf{v}, \mathbf{w} \in V$.

The #1 most important example is Euclidean space.

Theorem 2.8. *Give K^n the usual operations of vector addition and scalar multiplication:*

$$(v_1, \ldots, v_n) + (w_1, \ldots, w_n) = (v_1 + w_1, \ldots, v_n + w_n)$$
$$\lambda \cdot (v_1, \ldots, v_n) = (\lambda v_1, \ldots, \lambda v_n)$$

Then K^n is a vector space over K.

Proof. Since K is a field, its addition and multiplication satisfy the field axioms. All the axioms for a vector space are inherited from the field axioms for K. For example, let's check $\lambda(\mathbf{v}+\mathbf{w}) = \lambda\mathbf{v}+\lambda\mathbf{w}$.

Let $\lambda \in K$ and let $\mathbf{v}, \mathbf{w} \in K^n$. Then we can write

$$\mathbf{v} = (v_1, \ldots, v_n)$$
$$\mathbf{w} = (w_1, \ldots, w_n).$$

Now

$$\begin{aligned}
\lambda(\mathbf{v}+\mathbf{w}) &= \lambda((v_1, \ldots, v_n) + (w_1, \ldots, w_n)) \\
&= \lambda(v_1 + w_1, \ldots, v_n + w_n) \\
&= (\lambda(v_1 + w_1), \ldots, \lambda(v_n + w_n)) \\
&= (\lambda v_1 + \lambda w_1, \ldots, \lambda v_n + \lambda w_n) \\
&= (\lambda v_1, \ldots, \lambda v_n) + (\lambda w_1, \ldots, \lambda w_n) \\
&= \lambda(v_1, \ldots, v_n) + \lambda(w_1, \ldots, w_n) \\
&= \lambda\mathbf{v} + \lambda\mathbf{w}.
\end{aligned}$$

The rest of the verifications are left as an exercise. □

For our purposes, spaces of polynomials give the next most important examples.

Theorem 2.9. *Let $K[x]$ be the ring of polynomials in x with coefficients in K. Then $K[x]$ is a vector space over K, with the evident operations.*

Let $d \geq 0$, and let $V \subset K[x]$ be the subset of polynomials of degree at most d. Then V is a vector space over K, with the evident operations.

Proof. For the first part, you have to check the axioms directly. The second part is better proved after we have the concept of a subspace. □

Example 2.10. If $d \geq 0$, the set of polynomials in x of degree d is not a vector space. Indeed, there is no additive identity for addition.

Example 2.11. The previous theorem has obvious generalizations, for example to multivariable polynomials.

Remark 2.12. The definition of a vector space has several immediate consequences, such as the following:

(1) The additive identity $\mathbf{0} \in V$ is unique.
(2) For $\mathbf{v} \in V$, the additive inverse $\mathbf{w} \in V$ of $\mathbf{v}$ is unique. Call it $-\mathbf{v}$. (Parts (1) and (2) don't discuss the scalar multiplication, and are more generally true for any abelian group.)
(3) We have $-\mathbf{v} = (-1) \cdot \mathbf{v}$ for all $\mathbf{v} \in V$.
(4) We have $0 \cdot \mathbf{v} = 0$ for all $\mathbf{v} \in V$.

If these facts are not familiar you should prove them.

Theorem 2.13. *Let X be a set, and let $V = K^X$ be the set of all functions $f : X \to K$. Give V the operations of pointwise addition and scalar multiplication. Then X is a vector space.*

Proof. Check the axioms directly. □

2.5. **Subspaces.** As we saw in the previous section, there are a few main examples of vector spaces—Euclidean spaces, polynomials, and spaces of functions. Most other vector spaces that come up (e.g. the space of polynomials of degree $\leq d$) arise as subsets of these big known examples, and this makes the verification of the vector space axioms much easier in these cases.

Definition 2.14. Let $(V, +, \cdot)$ be a vector space, and let $W \subset V$ be a subset. Then W is a *subspace* of V if the following properties are satisfied:

(1) The zero vector $\mathbf{0} \in V$ is in W.
(2) (Closed under $+$) For all $\mathbf{w}_1, \mathbf{w}_2 \in W$, we have $\mathbf{w}_1 + \mathbf{w}_2 \in W$.
(3) (Closed under $\cdot$) For all $\mathbf{w} \in W$ and $\lambda \in K$, we have $\lambda \cdot \mathbf{w} \in W$.

The axioms in the definition of a subspace ensure that the addition and scalar multiplication operations on V make sense as operations on W: if we add two vectors in W we get a vector in W, and if we scalar multiply a vector in W by a scalar, we get a vector in W.

Remark 2.15. The empty subset $\emptyset \subset V$ is *not* a subspace! (Why?)

Theorem 2.16. *Let V be a vector space, and let $W \subset V$ be a subspace. Then W is a vector space when given the operations from V.*

Proof. Again you have to check that W satisfies the axioms of a vector space. Each of the axioms for W follows from the same axiom for V.

To do one, let us show that the addition on W is commutative. To be pedantic, let's write the operations on V as $(V, +_V, \cdot_V)$ and the operations on W as $(W, +_W, \cdot_W)$. The definition of the operations on W is that if $\mathbf{w}, \mathbf{w}_1, \mathbf{w}_2 \in W$ and $\lambda \in K$ then

$$\begin{aligned} \mathbf{w}_1 +_W \mathbf{w}_2 &:= \mathbf{w}_1 +_V \mathbf{w}_2 \\ \lambda \cdot_W \mathbf{w} &:= \lambda \cdot_V \mathbf{w}. \end{aligned}$$

(We add and scalar multiply the vectors in W as if they are in V.) Now we can check

$$\begin{aligned} \mathbf{v} +_W \mathbf{w} &= \mathbf{v} +_V \mathbf{w} \\ &= \mathbf{w} +_V \mathbf{v} \\ &= \mathbf{w} +_W \mathbf{v}. \end{aligned}$$

The other axioms are proved similarly. □

Example 2.17. For $d \geq 0$, let $V \subset K[x]$ be the subset of polynomials of degree d. Then V is a subspace of $K[x]$, so it is a vector space in its own right. Indeed, a sum of two polynomials of degree $\leq d$ has degree $\leq d$, a scalar multiple of a polynomial of degree $\leq d$ has degree $\leq d$, and the zero polynomial has degree $\leq d$.

Example 2.18. The main source of subspaces of K^n comes from solutions of homogeneous systems of linear equations. Consider a system

$$A\mathbf{x} = \mathbf{0}$$

of m equations in n unknowns $\mathbf{x} = (x_1, \ldots, x_n)$. If $\mathbf{x}, \mathbf{x}' \in K^n$ are two solutions of the system, then $\mathbf{x} + \mathbf{x}'$ and $\lambda\mathbf{x}$ are also solutions. Also, the zero vector is a solution. Thus

$$\{\mathbf{x} \in K^n : A\mathbf{x} = \mathbf{0}\} \subset K^n$$

is a subspace of K^n.

(The converse: every subspace $V \subset K^n$ can be expressed as the set of solutions of a homogeneous system of linear equations, is also true, but best approached later.)

Example 2.19. Subspaces of polynomials can be constructed by considering evaluations, derivatives, and other constructions. For example, the subset

$$V = \{f(x) \in K[x] : f(1) = f(2) = f'(3) = 0\} \subset K[x]$$

is a subspace, as you should verify.

2.6. **Span and spanning sets.** It is convenient to specify vector spaces or subspaces by specifying certain vectors that can be combined to form an arbitrary vector.

Example 2.20. In K^n, every vector can be written in the form

$$(x_1, \dots, x_n) = x_1(1, 0, \dots, 0) + \cdots + x_n(0, \dots, 0, 1) = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n,$$

where $\mathbf{e}_i$ is the ith unit vector.

Example 2.21. In $K[x]$, every vector can be written in the form

$$f(x) = a_0 + a_1x + \cdots + a_nx^n$$

for some n.

Example 2.22. Let $U \subset K[x]$ be the subspace of polynomials

$$U = \{f \in K[x] : f(1) = 0\} \subset K[x].$$

Every polynomial in U can be expressed in the form $(x-1)g(x)$ for some polynomial g. Writing g out explicitly, we can write f in the form

$$a_0(x-1) + a_1(x-1)x + a_2(x-1)x^2 + \cdots + a_n(x-1)x^n.$$

Thus the polynomials $(x-1)x^k$ $(k \geq 0)$ suffice to describe all the "degrees of freedom" in U.

Example 2.23. Some of the vectors can be redundant. For example, there are lots of different ways of writing a linear polynomial $ax + b$ as a sum of multiples of x, 1, and $1 + x$. E.g.

$$2x + 1 = 2 \cdot x + 1 \cdot 1 + 0 \cdot (x + 1) = 1 \cdot x + 0 \cdot 1 + 1 \cdot (x + 1).$$

These examples motivate the definitions of linear combinations and spans.

Definition 2.24. Let V be a vector space. A *linear combination* of vectors $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ is any expression of the form

$$\lambda_1\mathbf{v}_1 + \cdots + \lambda_n\mathbf{v}_n$$

where $\lambda_i \in K$.

Definition 2.25. Let $S \subset V$ be any subset. The *span* of S is the set of all linear combinations of vectors in S. More formally,

$$\operatorname{span}(S) = \{\lambda_1\mathbf{v}_1 + \cdots + \lambda_n\mathbf{v}_n : \mathbf{v}_i \in S \text{ and } \lambda_i \in K\}.$$

Remark 2.26. By convention the empty sum is always in $\operatorname{span}(S)$, so for example $\operatorname{span}(\emptyset) = \{0\}$.

Proposition 2.27. *Let V be a vector space and let $S \subset V$ be a subset. The span* $\operatorname{span}(S)$ *is a subspace of V. Furthermore, it is the* smallest *subspace of V that contains S: if $U \subset V$ is any subspace that contains S, then* $\operatorname{span}(S) \subset U$.

Proof. Exercise. □

Closely related to the idea of a span is the concept of a spanning set.

Definition 2.28. A subset $S \subset V$ is said to *span* V if $\operatorname{span}(S) = V$. (Or, it is a *spanning set*.)

Loosely speaking, all the "directions" in V have to be describable by vectors in S in order for S to span V.

Example 2.29. A set $S \subset V$ is always a spanning set of $\operatorname{span}(S)$.

Example 2.30. The standard unit vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are a spanning set of K^n. If any of these vectors are removed, they will fail to span K^n.

The space K^2 can be spanned by $(1, -1)$ and $(1, 1)$.

Example 2.31. The polynomials $1, x, x^2, \ldots$ are a spanning set of $K[x]$. So are the polynomials $1, (x-1), (x-1)^2, \ldots$. (Why?)

2.7. **Operations on subspaces.** Let V be a vector space. There are various ways of constructing new subspaces from old ones.

Proposition 2.32. *The intersection of any collection of subspaces of V is a subspace of V.*

Proof. Exercise. □

Example 2.33. The union of subspaces is usually not a subspace; it is rare that such a union is closed under addition. (Consider $\mathbb{R}^2$ with the two subspaces given by the coordinate axes.)

If $U, W \subset V$ are two subspaces of V then the *sum* $U + W$ is the subset

$$U + W = \{\mathbf{u} + \mathbf{w} : \mathbf{u} \in U, \mathbf{w} \in W\} \subset V.$$

More generally, given a collection $\{W_\alpha\}_{\alpha \in A}$ of subspaces of V, the *sum* of the W_α is the set

$$\sum_{\alpha \in A} W_\alpha := \left\{ \sum_{\alpha \in A} \lambda_\alpha \mathbf{w}_\alpha : \mathbf{w}_\alpha \in W_\alpha \text{ and } \lambda_\alpha \in K \right\} \subset V.$$

(If the index set A is infinite, we require that all but finitely many of the coefficients λ_α are 0. Thus, there is no issue with convergence of the sum.)

Proposition 2.34. *We have*

$$\sum_{\alpha\in A} W_\alpha = \operatorname{span}\left(\bigcup_{\alpha\in A} W_\alpha\right).$$

Therefore, $\sum_{\alpha\in A} W_\alpha$ *is the smallest subspace of* V *that contains each of the subspaces* W_α*.*

Proof. Exercise. □

A sum $U + W \subset V$ of two subspaces is called an *(internal) direct sum* if every vector $\mathbf{v} \in U + W$ is uniquely expressible in the form $\mathbf{v} = \mathbf{u} + \mathbf{w}$ for some $\mathbf{u} \in U$ and $\mathbf{w} \in W$. In this case, it is customary to write $U + W = U \oplus W$. More generally, a finite sum $U_1 + \cdots + U_n \subset V$ of subspaces $U_1, \ldots, U_n \subset V$ is direct if every vector $\mathbf{v} \in U_1 + \cdots + U_n$ is uniquely expressible as $\mathbf{u}_1 + \cdots + \mathbf{u}_n$ for some $\mathbf{u}_i \in U_i$.

Example 2.35. In K^n, let $U_i = \operatorname{span}(\mathbf{e}_i)$. Then $K^n = U_1 \oplus \cdots \oplus U_n$.

Example 2.36. In $K[x]$, call a polynomial $f \in K[x]$ *even* if $f(-x) = x$ and *odd* if $f(-x) = -x$. There are subspaces $U \subset K[x]$ and $V \subset K[x]$ consisting of even and odd polynomials, respectively. Then $U = \operatorname{span}(1, x^2, x^4, \ldots)$ and $V = \operatorname{span}(x, x^3, x^5, \ldots)$. We have $K[x] = U \oplus V$.

When we are considering two subspaces there is a simple criterion to tell if the sum is direct.

Lemma 2.37. *Let* $U, W \subset V$ *be subspaces of the vector space* V*. Then the sum* $U + W$ *is direct if and only if* $U \cap W = \{0\}$*.*

Proof. Exercise. □

Example 2.38. The analogous statement for 3 or more subspaces is false. For example, in $\mathbb{R}^2$ consider the subspaces given by the two coordinate axes and the line $y = x$.

2.8. **Linear independence.** The notion of linear independence captures what it means for a collection of vectors to all point "in different directions."

Definition 2.39. Let V be a vector space. A list $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ of vectors is *linearly independent* if the only constants $\lambda_1, \ldots, \lambda_n \in K$ such that

$$\lambda_1 \mathbf{v}_1 + \ldots + \lambda_n \mathbf{v}_n = 0$$

are $\lambda_1 = \cdots = \lambda_n = 0$. Otherwise, the list is *linearly dependent.*

Example 2.40. In K^n, the list of vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is linearly independent. If some vectors are removed from this list, the resulting list is still linearly independent.

In K^2, the list of vectors $(1,1)$, $(1,-1)$ is independent. The list of vectors $(1,0), (1,1), (1,-1)$ is not linearly independent.

Example 2.41. In $K[x]$, the list of vectors $1, x, \ldots, x^n$ is independent for any $n \geq 0$.

The next result establishes an alternative criterion for linear dependence.

Lemma 2.42. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$. Then this list is linearly dependent if and only if one of the vectors can be written as a linear combination of the others.*

Proof. ($\Rightarrow$) Since the vectors are dependent, we can find a linear combination

$$\lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n = 0$$

where at least one of the coefficients is nonzero. Without loss of generality, suppose $\lambda_1 \neq 0$. Then we can rearrange the equation and find

$$\mathbf{v}_1 = -\frac{1}{\lambda_1}(\lambda_2 \mathbf{v}_2 + \cdots + \lambda_n \mathbf{v}_n) = -\frac{\lambda_2}{\lambda_1}\mathbf{v}_2 - \cdots - \frac{\lambda_n}{\lambda_1}\mathbf{v}_n.$$

($\Leftarrow$) An equation

$$\mathbf{v}_1 = \lambda_2 \mathbf{v}_2 + \cdots + \lambda_n \mathbf{v}_n$$

which expresses one of the vectors as a linear combination of the others can be rearranged to

$$(-1) \cdot \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_n \mathbf{v}_n = 0.$$

Since the coefficient of $\mathbf{v}_1$ is nonzero, we conclude the list is linearly dependent. □

The most important consequence of linear independence is that a vector in the span of a linearly independent list is uniquely expressible as a linear combination of the vectors in the list.

Proposition 2.43. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ be linearly independent. If $\mathbf{w} \in \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$, then $\mathbf{w}$ can be written as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in exactly one way.*

Proof. Since $\mathbf{w} \in \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$, we can write it as a linear combination

$$\mathbf{w} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n.$$

Suppose it can also be written as

$$\mathbf{w} = \mu_1 \mathbf{v}_1 + \cdots + \mu_n \mathbf{v}_n.$$

Subtracting these two equations from each other,

$$\mathbf{0} = (\lambda_1 - \mu_1)\mathbf{v}_1 + \cdots + (\lambda_n - \mu_n)\mathbf{v}_n.$$

Since $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are independent, this requires $\lambda_i - \mu_i = 0$, and so $\lambda_i = \mu_i$. □

We can give an inductive description of what it means for a list $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of vectors to be independent:

(1) A list $\mathbf{v}_1$ of length 1 is independent iff $\mathbf{v}_1 \neq \mathbf{0}$.
(2) Suppose the list $\mathbf{v}_1$ is independent. The list $\mathbf{v}_1, \mathbf{v}_2$ is independent iff $\mathbf{v}_2$ is not a scalar multiple of $\mathbf{v}_1$.
(3) Suppose the list $\mathbf{v}_1, \mathbf{v}_2$ is independent. The list $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ is independent iff $\mathbf{v}_3$ is not a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$.
(4) Continue...

The justification for the inductive step is the following.

Proposition 2.44 (Building independent sets)**.** *Suppose* $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$*. Then* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *is linearly independent if and only if* $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$ *is independent and*

$$\mathbf{v}_n \notin \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}).$$

Proof. ($\Rightarrow$) Suppose $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ are independent. It is clear from the definition that a sublist of an independent list is independent; therefore $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$ are independent. We know $\mathbf{v_n}$ is not a linear combination of the other vectors, so also $\mathbf{v}_n \notin \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1})$.

($\Leftarrow$) Suppose $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1} \in V$ are independent and $\mathbf{v}_n \notin \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1})$. Let $\lambda_i \in K$ be such that

$$\lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n = 0.$$

Then

$$-\lambda_n \mathbf{v}_n = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_{n-1} \mathbf{v}_{n-1}.$$

Since $\mathbf{v}_n \notin \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1})$, the only way this is possible is if $\lambda_n = 0$. But then since $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$ are independent we must also have $\lambda_1 = \cdots = \lambda_{n-1} = 0$. Therefore all the λ_i are zero, and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are independent. □

2.9. **Bases.** A list of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ spans V if every vector in V can be written as a linear combination of vectors in the list. If the list is also independent, then every vector in V can be written as a linear combination of the vectors in one and only one way. Thus we are able to record an arbitrary vector in V by the corresponding linear combination.

Definition 2.45. A list $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ is a *basis* of V if it is a linearly independent spanning list.

Corollary 2.46. *If $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ is a basis of V, then every vector $\mathbf{w} \in V$ can be written uniquely as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$.*

Example 2.47. In K^n, the standard unit vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are a basis.

In K^2, the vectors $(1,3)$,$(2,4)$ are a basis. Every vector (x,y) in K^2 can be uniquely described as a linear combination

$$(x,y) = a(1,3) + b(2,4).$$

Example 2.48. Let $V \subset K[x]$ be the subspace of polynomials of degree at most d. Then the vectors $1, x, x^2, \ldots, x^d$ are a basis of V. Can you give another basis of V?

Example 2.49. The vector space $K[x]$ does not have a basis of any given finite length n. Indeed, it does not even have a finite spanning set. Suppose $f_1, \ldots, f_n \in K[x]$; then any linear combination $\lambda_1 f_1 + \cdots + \lambda_n f_n$ has degree at most $\max_i \deg f_i$, so no polynomials of degree higher than this are in $\operatorname{span}(f_1, \ldots, f_n)$.

It is possible to generalize the concept of basis to lists of infinite length, and then $1, x, x^2, \ldots$ is a basis of $K[x]$. But, we will mostly be concerned with finite bases and so won't develop this theory.

When does a vector space have a (finite) basis? A necessary condition is that the space has a finite spanning set.

Definition 2.50. A vector space V is *finite-dimensional* if it can be spanned by a finite list of vectors.

Given a spanning set, any linear dependencies can be "pruned away" to yield a basis.

Proposition 2.51. *If V is a vector space and $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ span V, then we can construct a basis of V by deleting some (possibly 0) of the vectors in the list.*

Proof. If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are independent then they are a basis and we are done. So, suppose they are depdendent. Without loss of generality, the vector $\mathbf{v}_n$ is a linear combination of the other vectors. Then the

list $\mathbf{v}_1, \dots, \mathbf{v}_{n-1}$ still spans V. Indeed, given $\mathbf{w} \in V$, we can write it as a linear combination

$$\mathbf{w} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n,$$

and we can also write $\mathbf{v}_n$ as a linear combination

$$\mathbf{v}_n = \mu_1 \mathbf{v}_1 + \cdots + \mu_{n-1} \mathbf{v}_{n-1}.$$

Substituting this expression for $\mathbf{v}_n$ in the expression for $\mathbf{w}$ and rearranging, we find that $\mathbf{w}$ is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_{n-1}$.

Now since $\mathbf{v}_1, \dots, \mathbf{v}_{n-1}$ still spans V, either it is a basis or we can repeat the above process. Eventually this process stops (there are only finitely many vectors in the list), and that means we have arrived at a basis. □

Corollary 2.52. *Every finite-dimensional vector space has a basis.*

Proof. Shrink any spanning set to a basis. □

Conversely, for a finite-dimensional vector space, any independent set can be extended to a basis.

Proposition 2.53. *Let V be a finite-dimensional vector space and let $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ be independent. Then this list can be extended to a basis $\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{u}_1, \dots, \mathbf{u}_m \in V$ of V.*

Proof. Exercise. □

2.10. **Dimension.** We have arrived at one of the fundamental concepts and results in linear algebra. Roughly speaking, the *dimension* of a vector space should be the number of independent parameters it takes to specify a vector. More precisely, it should be the number of vectors in a basis.

Unfortunately, while we now know that every finite dimensional vector space has a basis, it is not clear that any two bases have the same number of vectors in them. This is a major theorem which has several approaches; we will appeal to our earlier study of Gaussian elimination.

Lemma 2.54. *Suppose $\mathbf{v}_1, \dots, \mathbf{v}_m \in K^n$ and $m > n$. Then the list $\mathbf{v}_1, \dots, \mathbf{v}_m$ is linearly dependent.*

Proof. Write the vectors $\mathbf{v}_1, \dots, \mathbf{v}_m$ as column vectors in the matrix

$$A = (\mathbf{v}_1 | \cdots | \mathbf{v}_m).$$

The vectors are dependent if and only if there are constants $x_1, \dots, x_m \in K$, not all zero, such that $x_1 \mathbf{v}_1 + \cdots + x_n \mathbf{v}_n = \mathbf{0}$. Equivalently, the vector $(x_1, \dots, x_m)$ satisfies the system $A\mathbf{x} = \mathbf{0}$. But this is

a linear system of n equations in $m > n$ unknowns, so it has a nonzero solution. □

Proposition 2.55. *If a vector space V has an independent list $\mathbf{v}_1, \ldots, \mathbf{v}_m$ and a spanning list $\mathbf{w}_1, \ldots, \mathbf{w}_n$, then $m \leq n$.*

Proof. By way of contradiction, assume $m > n$. Since $\mathbf{w}_1, \ldots, \mathbf{w}_n$ span V, we can write each of the vectors $\mathbf{v}_j$ as a linear combination of $\mathbf{w}_1, \ldots, \mathbf{w}_n$. Explicitly, write

$$\mathbf{v}_j = a_{1j}\mathbf{w}_1 + \cdots + a_{nj}\mathbf{w}_n.$$

Consider the vectors in K^n made from the coefficients of each of these expressions:

$$\begin{aligned} \mathbf{u}_1 &= (a_{11}, \ldots, a_{n1}) \\ \mathbf{u}_2 &= (a_{12}, \ldots, a_{n2}) \\ &\vdots \\ \mathbf{u}_m &= (a_{1m}, \ldots, a_{nm}). \end{aligned}$$

These are m vectors in K^n, so there is an equation of linear dependence

$$\lambda_1\mathbf{u}_1 + \cdots + \lambda_m\mathbf{u}_m = \mathbf{0}$$

with the λ_i not all 0.

I claim that also

$$\lambda_1\mathbf{v}_1 + \cdots + \lambda_m\mathbf{v}_m = \mathbf{0}.$$

This will imply that $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are dependent, a contradiction. Expand out the left hand side and gather the coefficients of the $\mathbf{w}_i$:

$$\begin{aligned} \lambda_1\mathbf{v}_1 + \cdots + \lambda_m\mathbf{v}_m &= \lambda_1(a_{11}\mathbf{w}_1 + \cdots + a_{n1}\mathbf{w}_n) \\ &\quad + \lambda_2(a_{12}\mathbf{w}_1 + \cdots + a_{n2}\mathbf{w}_n) \\ &\quad \vdots \\ &\quad + \lambda_m(a_{1m}\mathbf{w}_1 + \cdots + a_{nm}\mathbf{w}_n) \\ &= (\lambda_1 a_{11} + \cdots + \lambda_m a_{1m})\mathbf{w}_1 \\ &\quad \vdots \\ &\quad + (\lambda_1 a_{n1} + \cdots + \lambda_m a_{nm})\mathbf{w}_m. \end{aligned}$$

Here we notice that the coefficient of $\mathbf{w}_i$ is exactly the ith entry of the vector $\lambda_1\mathbf{u}_1 + \cdots + \lambda_m\mathbf{u}_m = \mathbf{0}$. Therefore all the coefficients are 0, and $\lambda_1\mathbf{v}_1 + \cdots + \lambda_m\mathbf{v}_m = \mathbf{0}$ follows. □

Corollary 2.56. *In a finite-dimensional vector space V, any two bases have the same length.*

Proof. Supose $\mathbf{v}_1, \dots, \mathbf{v}_m$ and $\mathbf{w}_1, \dots, \mathbf{w}_n$ are bases of V. Then $\mathbf{v}_1, \dots, \mathbf{v}_m$ is independent and $\mathbf{w}_1, \dots, \mathbf{w}_n$ spans V, so $m \leq n$. By symmetry, $n \leq m$ and $m = n$. □

Thus, the next definition is justified.

Definition 2.57. The *dimension* $\dim V$ of a finite-dimensional vector space V is the length of any basis.

Example 2.58. We have $\dim K^n = n$.

2.11. **More examples of dimension.** Various spaces of polynomials give several additional examples of vector spaces where the dimension can be computed.

Example 2.59. Let $V \subset K[x]$ be the subspace of polynomials of degree at most d. Then $\dim V = d + 1$. Indeed, this vector space has a basis $1, x, \dots, x^d$.

Example 2.60 (Spaces of homogeneous polynomials)**.** Let $V = K[x_0, \dots, x_n]$ be the vector space of polynomials in $x_0, \dots, x_n$. A polynomial $f \in V$ in $n + 1$ variables is *homogeneous of degree d* if every monomial which appears in f with a nonzero coefficient has degree d. For example,

$$x_0^2 x_1 + x_1^3 + 2x_1^2 x_2 + x_2 x_3^2$$

is homogeneous of degree 3 in the four variables x_0, x_1, x_2, x_3. Observe that a sum of two polynomials which are homogeneous of the same degree is again homogeneous of that degree, unless the sum is 0. Therefore, for a fixed integer $d \geq 0$ there is a subspace

$$V_d = \{f \in V : f \text{ is homogeneous of degree } d\} \cup \{0\} \subset V.$$

This subspace has a basis given by the collection of monomials $x_0^{d_0} \cdots x_n^{d_n}$ of degree d (so $\sum d_i = d$).

Therefore, the dimension of this subspace is the number of monomials of degree d. How many are there? This number can be computed by a standard combinatorial argument called *stars and bars*. Suppose we are given a string that consists of d *stars* and n *bars*, written in some order. Here are some examples for $d = 7$ and $n = 3$:

$$||\ast\ast\ast|\ast\ast\ast\ast \qquad |\ast\ast\ast|\ast\ast|\ast\ast \qquad \ast\ast\ast|\ast|\ast\ast|\ast\,.$$

Such an expression determines and is determined by a monomial of degree d: The exponent on variable x_i is the number of stars between bar i and bar $i + 1$. (We label the variables from 0 to n and label the

bars from 1 to n. The exponent on x_0 is the number of stars left of bar 1, and the exponent on x_n is the number of stars right of bar n.) Thus, the above expressions correspond to the monomials

$$x_2^3x_3^4 \qquad x_1^3x_2^2x_3^2 \qquad x_0^3x_1x_2^2x_3,$$

respectively.

Conversely, given a monomial it is easy to write down the stars-and-bars diagram that corresponds to it, so we have a bijection between the set of monomials of degree d in $n+1$ variables and the number of stars-and-bars diagrams with d stars and n bars.

How many stars-and-bars diagrams are there with d stars and n bars? Imagine we start with a string of $d+n$ stars. Then we just have to change n of those stars to bars; the number of ways of doing this is equal to the number of n-element subsets of a set of cardinality $d+n$. So, the number of diagrams is

$$\binom{d+n}{n} = \frac{(d+n)!}{d! \cdot n!}.$$

We conclude

$$\dim(K[x_0, \ldots, x_n])_d = \binom{d+n}{n}.$$

So far we've mostly talked about vector spaces of polynomials of degree $\leq d$. These spaces have a simple relationship to spaces of homogeneous polynomials.

Example 2.61 (Multivariable polynomials of bounded degree)**.** For an integer $d \geq 0$, consider the space $V \subset K[x_1, \ldots, x_n]$ of polynomials of degree $\leq d$ in n variables $x_1, \ldots, x_n$. Given a polynomial $f \in V$, its *homogenization* is the polynomial obtained by taking each monomial in f and multiplying it by an appropriate power of a new variable x_0 to get a monomial of degree d. More formally, the homogenization of f is the polynomial

$$F(x_0, \ldots, x_n) := x_0^d f\left(\frac{x_1}{x_0}, \ldots, \frac{x_n}{x_0}\right)$$

(Why is this actually a polynomial?) Conversely, given a homogeneous polynomial of degree d in $x_0, \ldots, x_n$, we can set $x_0 = 1$ to get a polynomial in $x_1, \ldots, x_n$ of degree $\leq d$.

For example, the homogenization of

$$f(x_1, \ldots, x_3) = x_1^2 + 2x_2 + 3x_1x_3 + x_3^4$$

to a homogeneous polynomial of degree 4 is

$$F(x_0, \ldots, x_4) = x_0^2x_1^2 + 2x_0^3x_2 + 3x_0^2x_1x_3 + x_3^4.$$

Observe that we can recover f from F by setting $x_0 = 1$.

Now the space V has a basis given by the monomials $x_1, \dots, x_n$ of degree $\leq d$. Each such monomial uniquely determines a monomial in $x_0, \dots, x_n$ of degree d. Therefore the number of basis vectors is the number of monomials in $x_1, \dots, x_n$ of degree d, and so

$$\dim V = \binom{d+n}{n}$$

by our earlier computation.

2.12. **Solving linear systems, take 2.** Our earlier discussion of the solution of systems of linear equations can now be discussed more conceptually. First consider a homogeneous linear system $A\mathbf{x} = \mathbf{0}$, where $A = (a_{ij})$ is an $m \times n$ matrix and $\mathbf{x} = (x_1, \dots, x_n) \in K^n$.

Remark 2.62. In an earlier example and your homework, you saw that the set of solutions

$$U = \{\mathbf{x} \in K^n : A\mathbf{x} = \mathbf{0}\}$$

is a subspace of K^n.

Problem 2.63. Compute a basis (and thus the dimension) of U.

By our study of Gaussian elimination, we may as well assume A is in row echelon form, since row operations do not change the solution set U of the system. Once A is in row echelon form, we identify the sets of bound and free variables. We know that if we assign values to the free variables, then there are uniquely determined values for the bound variables which solve the system.

To give a basis of U, it is useful to assign the values of the free variables in an intelligent way. Suppose the free variables are $x_{\ell_1}, \dots, x_{\ell_k}$. For the free variable x_{ℓ_i}, construct a solution $\mathbf{v}_i$ of the system by setting the value of x_{ℓ_i} to be 1, setting the value of all the other free variables to be 0, and determining the values of the bound variables by requiring that the vector solve the system.

Example 2.64. For the row echelon matrix

$$A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

the variables x_2 and x_4 are free, while x_1 and x_3 are bound. (In the notation before the example, we could put $\ell_1 = 2$ and $\ell_2 = 4$, so that

the free variables are x_{ℓ_1}, x_{ℓ_2}.) Then we construct two solution vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ by starting from

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ \\ 0 \end{pmatrix} \qquad \mathbf{v}_2 = \begin{pmatrix} 0 \\ \\ 1 \end{pmatrix}$$

(leaving blank spaces for the bound variables) and using back substitution on the system $A\mathbf{x} = \mathbf{0}$ to determine the remaining entries:

$$\mathbf{v}_1 = \begin{pmatrix} -3 \\ 1 \\ 0 \\ 0 \end{pmatrix} \qquad \mathbf{v}_2 = \begin{pmatrix} -2 \\ 0 \\ -1 \\ 1 \end{pmatrix}.$$

Observe that these vectors are linearly independent.

Theorem 2.65. *The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are a basis of U. In particular, every solution of the system $A\mathbf{x} = \mathbf{0}$ can be uniquely written in the form*

$$\mathbf{x} = a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k$$

for some $a_i \in K$. The dimension of U is k, the number of free variables.

Proof. We show that $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent and span U.

(Independent) Suppose $a_1, \ldots, a_k \in K$ are such that

$$a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k = 0.$$

Then the entry in slot ℓ_i of the LHS vector is just a_i, by the definition of the vectors $\mathbf{v}_j$. Therefore all the a_i must be zero, and the vectors are independent.

(Spanning) Suppose $\mathbf{x} \in U$ is an arbitrary solution of the system. Let a_i be the ℓ_i-th entry of $\mathbf{x}_i$, and consider the vector

$$\mathbf{w} = a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k.$$

Notice that $\mathbf{x}$ and $\mathbf{w}$ are the same in all the entries corresponding to free variables. But the solution of $A\mathbf{x} = \mathbf{0}$ that has particular values assigned to the free variables is unique. Therefore $\mathbf{x} = \mathbf{w}$, and $\mathbf{x}$ is in the span of $\mathbf{v}_1, \ldots, \mathbf{v}_n$. □

In the general case of a potentially inhomogeneous system $A\mathbf{x} = \mathbf{b}$, recall that the system is either consistent or inconsistent, and that we can use Gaussian elimination to determine which possibility is true. If there is one solution $\mathbf{p}$, then any other solution $\mathbf{p}'$ satisfies

$$A(\mathbf{p}' - \mathbf{p}) = 0,$$

so that $\mathbf{h} := \mathbf{p}' - \mathbf{p}$ is a solution of the homogeneous system $A\mathbf{h} = 0$. We know how to express solutions $\mathbf{h}$ of the homogeneous system, so we arrive at the following description of the solutions of the inhomogeneous system.

Theorem 2.66. *The system $A\mathbf{x} = \mathbf{b}$ is either consistent or inconsistent. If it is consistent, let $\mathbf{p}$ be one solution to the system. Then any solution $\mathbf{p}'$ of $A\mathbf{x} = \mathbf{b}$ can be written uniquely in the form*

$$\mathbf{p}' = \mathbf{p} + a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k,$$

where $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are a basis of the space U of solutions to the homogeneous system

$$A\mathbf{x} = \mathbf{0}.$$

3. Linear Algebra: Linear Transformations

The central concept in linear algebra is the notion of a linear transformation. Roughly speaking, a linear transformation between two vector spaces is a function that preserves the vector space operations. Linear transformations allow us to study the relationships between different vector spaces. Many important questions in mathematics that sound like they have nothing to do with linear algebra can actually be phrased in terms of vector spaces and linear transformations.

3.1. **Linear transformations.**

Let V and W be vector spaces over the same field K.

Definition 3.1. A *linear transformation* from V to W is a function $T : V \to W$ which preserves the vector space operations. That is, for all $\mathbf{v}, \mathbf{v}' \in V$ and $\lambda \in K$, we have

$$T(\mathbf{v} + \mathbf{v}') = T(\mathbf{v}) + T(\mathbf{v}')$$
$$T(\lambda \mathbf{v}) = \lambda T(\mathbf{v}).$$

Remark 3.2. It is customary to write $T\mathbf{v}$ instead of $T(\mathbf{v})$ so long as no confusion will occur.

There are many important examples.

Example 3.3. The function $T : K^2 \to K^3$ defined by

$$T(x, y) = (3x + y, 2y, x + y)$$

is a linear transformation. Indeed, if $(x, y), (x', y') \in K^2$, then

$$\begin{aligned} T((x, y) + (x', y')) &= T(x + x', y + y') \\ &= (3(x + x') + (y + y'), 2(y + y'), (x + x') + (y + y')) \\ &= ((3x + y) + (3x' + y'), 2y + 2y', (x + y) + (x' + y')) \\ &= (3x + y, 2y, x + y) + (3x' + y', 2y', x' + y') \\ &= T(x, y) + T(x', y'). \end{aligned}$$

A similar argument shows that T preserves scalar multiplication.

Example 3.4. More generally, if $A = (a_{ij})$ is an $m \times n$ matrix with entries in K, then we can define a linear transformation $T : K^n \to K^m$ by the rule

$$T\mathbf{x} = A\mathbf{x} := \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix},$$

where $\mathbf{x} = (x_1, \ldots, x_n)$. The proof is essentially the same as in the previous example. Notice that $T\mathbf{e}_i$ is the ith column of A.

Example 3.5. Let $\theta \in \mathbb{R}$. The function $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ which rotates a vector by an angle of θ is a linear transformation. Indeed, if we add two vectors and then rotate the result by θ, we get the same answer as if we rotate the two vectors by θ and add the results. (Draw a picture, recalling that vectors are added tip-to-tail.)

Example 3.6. For a vector space of polynomials, we can use facts about polynomials to define interesting linear transformations. Let $\mathcal{P}_d \subset K[x]$ denote the vector space of polynomials of degree $\leq d$.

There is a *derivative* transformation

$$\begin{aligned} D : \mathcal{P}_d &\to \mathcal{P}_{d-1} \\ p(x) &\mapsto p'(x). \end{aligned}$$

There are *evaluation* transformations, for example evaluation at 1:

$$\begin{aligned} T : \mathcal{P}_d &\to K \\ p(x) &\mapsto p(1). \end{aligned}$$

We can simultaneously evaluate at several points:

$$\begin{aligned} T : \mathcal{P}_d &\to K^3 \\ p(x) &\mapsto (p(1), p(3), p(7)). \end{aligned}$$

We can evaluate p and some of its derivatives at some points:

$$\begin{aligned} T : \mathcal{P}_d &\to K^4 \\ p(x) &\mapsto (p(2), p'(2), p(5), p'''(6)). \end{aligned}$$

We can send a polynomial to its list of coefficients:

$$\begin{aligned} T : \mathcal{P}_d &\to K^{d+1} \\ \sum_i a_i x^i &\mapsto (a_0, a_1, \ldots, a_d) \end{aligned}$$

We can take the integral from 0 to x:

$$\begin{aligned} T : \mathcal{P}_d &\to \mathcal{P}_{d+1} \\ p(x) &\mapsto \int_0^x p(t)\, dt \end{aligned}$$

(Note that a function $p(x) \mapsto \int p(x)\, dx$ is ambiguous, since there is a constant of integration...).

We can make new linear transformations from old ones by several common operations.

Proposition 3.7. *Let $T, T' : V \to W$ and $S : W \to U$ be linear transformations. Then $T+T'$ (defined pointwise) is a linear transformation, as is λT (defined pointwise) and the composition $S \circ T : V \to U$.*

In fact, linear transformations give us new examples of vector spaces as well.

Proposition 3.8. *Let V and W be vector spaces over K. The set $\mathrm{Hom}(V, W)$ of all linear transformations from V to W is a vector space over K when given the operations of pointwise sum and scalar multiplication.*

3.2. **Constructing linear transformations.** The mantra of linear transformations is that "a linear transformation is uniquely determined by its action on a basis." More precisely, we have the following theorem.

Theorem 3.9. *Let V and W be vector spaces over K, and suppose V is finite-dimensional. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ be a basis of V, and let $\mathbf{w}_1, \ldots, \mathbf{w}_n \in W$ be arbitrary vectors. Then there is a unique linear transformation $T : V \to W$ such that*

$$T\mathbf{v}_i = \mathbf{w}_i.$$

If we are willing to dispense with uniqueness, the hypotheses can be weakened:

Proposition 3.10. *Let V and W be vector spaces over K, and suppose V is finite-dimensional. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ be a linearly independent list, and let $\mathbf{w}_1, \ldots, \mathbf{w}_n \in W$ be arbitrary vectors. Then there is a linear transformation $T : V \to W$ such that*

$$T\mathbf{v}_i = \mathbf{w}_i \qquad (i = 1, \ldots, n).$$

Both of these results look like interpolation problems! We are finding some "nice function" which takes the specified points (vectors in V) to the specified values (vectors in W).

Remark 3.11. In the proposition, if you are comfortable with Zorn's Lemma then the finite-dimensional hypothesis is not necessary. The theorem is also true without the finite-dimensional hypothesis if you define what a basis of an infinite-dimensional space is.

We prove the theorem first.

Proof of Theorem 3.9. First suppose $T : V \to W$ is a linear transformation that satisfies $T\mathbf{v}_i = \mathbf{w}_i$ for $i = 1, \ldots, n$. Then by linearity we must have

$$T(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) = a_1\mathbf{w}_1 + \cdots + a_n\mathbf{w}_n$$

for any $a_i \in K$.

Since any vector $\mathbf{v} \in V$ can be uniquely expressed as a linear combination

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n,$$

this suggests that we should define a function $T : V \to W$ by the rule

$$T(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) = a_1\mathbf{w}_1 + \cdots + a_n\mathbf{w}_n.$$

Clearly $T\mathbf{v}_i = \mathbf{w}_i$. Since any linear transformation satisfying $T\mathbf{v}_i = \mathbf{w}_i$ must satisfy this rule, there is at most one linear transformation satisfying $T\mathbf{v}_i = \mathbf{w}_i$ for $i = 1, \ldots, n$.

We have to show that this function $T : V \to W$ is actually linear. So, let $\mathbf{v}, \mathbf{v}' \in V$. Then we can write

$$\begin{aligned}\mathbf{v} &= a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n \\ \mathbf{v}' &= b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n,\end{aligned}$$

so

$$\begin{aligned}T(\mathbf{v} + \mathbf{v}') &= T((a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) + (b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n)) \\ &= T((a_1 + b_1)\mathbf{v}_1 + \cdots + (a_n + b_n)\mathbf{v}_n) \\ &= (a_1 + b_1)\mathbf{w}_1 + \cdots + (a_n + b_n)\mathbf{w}_n \\ &= (a_1\mathbf{w}_1 + \cdots + a_n\mathbf{w}_n) + (b_1\mathbf{w}_1 + \cdots + b_n\mathbf{w}_n) \\ &= T(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) + T(b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n) \\ &= T(\mathbf{v} + \mathbf{v}').\end{aligned}$$

The proof that $T(\lambda\mathbf{v}) = \lambda T\mathbf{v}$ for $\lambda \in K$ is similar. Therefore T is linear. □

The proposition is an easy application of the theorem.

Proof of Proposition 3.10. Extend the independent list to a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{u}_1, \ldots, \mathbf{u}_k$ of V. Then by the theorem there is a linear transformation $T : V \to W$ such that $T\mathbf{v}_i = \mathbf{w}_i$ and $T\mathbf{u}_i = 0$ (actually we can send the $\mathbf{u}_i$'s to whatever vectors we want). □

3.3. **The matrix of a linear transformation.** Since a linear transformation is uniquely specified by its action on a basis, we can record a linear transformation by just remembering what it does to a basis. Matrices can be used to conveniently record this information.

Suppose V and W are finite dimensional vector spaces with bases $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ and $\mathbf{w}_1, \ldots, \mathbf{w}_m \in W$, respectively. Let $T : V \to W$ be a linear transformation. For each basis vector $\mathbf{v}_j$, the vector $T\mathbf{v}_j$ is some

vector in W, and therefore it can be written as a linear combination of the basis vectors $\mathbf{w}_i$. Explicitly, write

$$T\mathbf{v}_j = a_{1j}\mathbf{w}_1 + \cdots + a_{mj}\mathbf{w}_m.$$

Then the coefficients a_{ij} $(1 \leq i \leq m,\ 1 \leq j \leq n)$ contain all the information needed to specify the linear transformation. We put them into a matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix}.$$

Remark 3.12. The information needed to determine $T\mathbf{v}_j$ is contained in the jth column of the matrix A. More specifically, the entry a_{ij} is the coefficient of $\mathbf{w}_i$ when $T\mathbf{v}_j$ is expressed as a linear combination of the basis vectors $\mathbf{w}_1, \ldots, \mathbf{w}_m$.

Note that the matrix A representing T very much depends on the bases used to record vectors in V and W! The matrix of T can be a useful tool for computing T. If $\mathbf{v} \in V$, express $\mathbf{v}$ as a linear combination

$$\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$$

of the basis vectors on V. Let $\mathbf{x} = (x_1, \ldots, x_n)$ be the vector of coefficients and let $\mathbf{y} = A\mathbf{x}$. Write $\mathbf{y} = (y_1, \ldots, y_m)$. Then we have

$$T\mathbf{v} = y_1\mathbf{w}_1 + \cdots + y_m\mathbf{w}_m.$$

Thus, multiplying the matrix A by the vector $\mathbf{x}$ computes the coefficients when the vector $\mathbf{w}$ is expressed as a linear combination of the basis vectors $\mathbf{w}_!, \ldots, \mathbf{w}_m$ on W.

Exercise 3.13. Justify the assertions in the previous paragraph.

3.4. **Injectivity and surjectivity; kernel and image.** Let $T : V \to W$ be a linear transformation. A common problem is to determine if for $\mathbf{w} \in W$ the equation $T\mathbf{v} = \mathbf{w}$ has a solution. When it has a solution, we often want to know if that solution is unique.

Recall that a function of sets $f : X \to Y$ is *injective* (or *1-1*) if whenever $f(x_1) = f(x_2)$ for some $x_1, x_2 \in X$, we must have $x_1 = x_2$. The *image* of f is the subset

$$\operatorname{Im} f = \{f(x) : x \in X\} \subset Y,$$

and f is *surjective* (or *onto*) if $\operatorname{Im} f = Y$. Equivalently, for every $y \in Y$, there must exist some $x \in X$ such that $f(x) = y$. The function f is *bijective* if it is injective and surjective.

A linear transformation $T : V \to W$ is in particular a function from V to W, so the concepts of injective/surjective/bijective and images make sense for linear transformations. The next definition makes use of the vector space structure.

Definition 3.14. Let $T : V \to W$ be a linear transformation. The *kernel* of T is the subset

$$\ker T = \{\mathbf{v} \in V : T\mathbf{v} = \mathbf{0}\} \subset V.$$

Proposition 3.15. *Let $T : V \to W$ be a linear transformation. Then $\ker T$ is a subspace of V, and $\operatorname{Im} T$ is a subspace of W.*

Proof. (Important) Exercise. Check the three subspace axioms. □

Clearly T is surjective if and only if $\operatorname{Im} T = W$. In case W is finite-dimensional, this is equivalent to the statement that $\dim \operatorname{Im} T = \dim W$, by the next result.

Lemma 3.16. *Suppose V is a finite-dimensional vector space and $W \subset V$ is a subspace. Then $\dim W \leq \dim V$, with equality if and only if $W = V$.*

Proof. From your homework, W is finite-dimensional. Let $\mathbf{w}_1, \ldots, \mathbf{w}_m \in W$ be a basis of W. This is a linearly independent list in V, so it can be extended to a basis of V. Therefore $\dim W \leq \dim V$, with equality if and only if $\mathbf{w}_1, \ldots, \mathbf{w}_m$ is already a basis of V; in that case, $V = W$. □

The dimension of the image occurs very frequently, and deserves its own name.

Definition 3.17. The *rank* of a linear transformation is $\dim \operatorname{Im} T$.

Thus if W is finite-dimensional, then $T : V \to W$ is surjective if and only if its rank is $\dim W$. On the other hand, kernels give us a simple criterion for injectivity.

Proposition 3.18. *Let $T : V \to W$ be a linear transformation. Then T is injective if and only if $\ker T = \{0\}$.*

Proof. First suppose T is injective; we show $\ker T = \{0\}$. Suppose $\mathbf{v} \in \ker T$. Then $T\mathbf{v} = 0$ and $T0 = 0$, so $T\mathbf{v} = T0$. Since T is injective, this gives $\mathbf{v} = 0$. Therefore 0 is the only vector in $\ker T$, and $\ker T = \{0\}$.

Conversely, suppose $\ker T = \{0\}$, and suppose $\mathbf{v}, \mathbf{v}' \in V$ have $T\mathbf{v} = T\mathbf{v}'$. Then

$$0 = T\mathbf{v} - T\mathbf{v}' = T(\mathbf{v} - \mathbf{v}'),$$

so $\mathbf{v} - \mathbf{v}' \in \ker T$. Therefore $\mathbf{v} - \mathbf{v}' = 0$, and $\mathbf{v} = \mathbf{v}'$. Therefore T is injective. □

Since the only vector space of dimension 0 is the zero vector space, this can be rephrased as saying T is injective iff $\dim \ker T = \{0\}$.

Definition 3.19. The *nullity* of a linear transformation $T : V \to W$ is $\dim \ker T$.

The notions of injectivity and surjectivity are also related to the concepts of spanning and independent sets, as the next exercise shows.

Exercise 3.20. Let $T : V \to W$ be a linear transformation.

(1) Show that T is injective if and only if it carries independent lists to independent lists. (That is, for every independent list $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$, we have that $T\mathbf{v}_1, \ldots, T\mathbf{v}_n$ are independent.)
(2) Suppose that V is finite dimensional. Show that T is surjective if and only if it carries spanning lists to spanning lists.
(3) Show by example that (2) is not true for infinite dimensional vector spaces.

3.5. **The rank-nullity theorem.** The rank-nullity theorem is one of the most fundemantal theorems in linear algebra. It connects the study of the image of a linear transformation to the study of its kernel. Taken to its extreme, this connects the study of surjectivity of linear transformations to the the study of injectivity of linear transformations. When a linear transformation is thought of as a matrix and the corresponding systems of linear equations, this connects the existence and uniqueness problems for solutions of linear equations.

Theorem 3.21 (Rank-nullity theorem)**.** *Let $T : V \to W$ be a linear transformation, and suppose V is finite-dimensional. Then*

$$\dim V = \dim \ker T + \dim \operatorname{Im} T.$$

Proof. Write $U = \ker T \subset V$. Since U is a subspace of a finite-dimensional vector space, it is finite-dimensional and it has a basis $\mathbf{u}_1, \ldots, \mathbf{u}_k$. This basis can be extended to a basis $\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{v}_1, \ldots, \mathbf{v}_\ell \in V$ of V. To prove the theorem, we must show $\operatorname{Im} T$ has a basis of length ℓ. To this end, we claim that $T\mathbf{v}_1, \ldots, T\mathbf{v}_\ell$ are a basis of $\operatorname{Im} T$. We show they are independent and span $\operatorname{Im} T$.

(Independent.) Suppose

$$a_1 T\mathbf{v}_1 + \cdots + a_\ell T\mathbf{v}_\ell = 0.$$

Then

$$T(a_1\mathbf{v}_1 + \cdots + a_\ell\mathbf{v}_\ell) = 0,$$

so $\mathbf{v} := a_1\mathbf{v}_1 + \cdots + a_\ell\mathbf{v}_\ell$ is in $\ker T$. Therefore we can write $\mathbf{v}$ as a linear combination

$$a_1\mathbf{v}_1 + \cdots + a_\ell\mathbf{v}_\ell = b_1\mathbf{u}_1 + \cdots + b_k\mathbf{u}_k.$$

This is only possible if all a_i, b_j are zero, so $T\mathbf{v}_1, \ldots, T\mathbf{v}_\ell$ are independent.

(Spanning.) Suppose $\mathbf{w} \in \operatorname{Im} T$. Then there is some $\mathbf{v} \in V$ such that $T\mathbf{v} = \mathbf{w}$. Express $\mathbf{v}$ as a linear combination of basis vectors

$$\mathbf{v} = a_1\mathbf{u}_1 + \cdots + a_k\mathbf{u}_k + b_1\mathbf{v}_1 + \cdots + b_\ell\mathbf{v}_\ell.$$

Applying T to both sides and recalling $T\mathbf{u}_i = 0$, we get

$$\mathbf{w} = T\mathbf{v} = T(a_1\mathbf{u}_1 + \cdots a_k\mathbf{u}_k + b_1\mathbf{v}_1 + \cdots + b_\ell\mathbf{v}_\ell) = b_1T\mathbf{v}_1 + \cdots + b_\ell T\mathbf{v}_\ell.$$

Therefore $\mathbf{w}$ is a linear combination of $T\mathbf{v}_1, \ldots T\mathbf{v}_\ell$. □

The rank-nullity theorem has many immediate consequences.

Corollary 3.22. *Suppose $T : V \to W$ is a linear transformation and V is finite-dimensional. Then*

$$\dim \ker T \geq \dim V - \dim W.$$

In particular, if $\dim V > \dim W$, then T is not injective.

Proof. Since $\operatorname{Im} T$ is a subspace of W, we have $\dim \operatorname{Im} T \leq \dim W$. Then we have

$$\dim \ker T = \dim V - \dim \operatorname{Im} T \geq \dim V - \dim W.$$

If $\dim V > \dim W$, we conclude $\ker T$ is positive-dimensional, and so is nonzero. □

A similar argument proves the next result.

Corollary 3.23. *Suppose $T : V \to W$ is a linear transformation and V is finite-dimensional. Then $\dim V \geq \dim \operatorname{Im} T$. In particular, if $\dim V < \dim W$, then T is not surjective.*

Proof. Exercise. □

Corollary 3.24. *Suppose $T : V \to W$ is a linear transformation and $\dim V = \dim W$ (and both are finite). Then T is injective iff it is surjective iff it is bijective.*

Proof. Write down the rank-nullity theorem:

$$\dim W = \dim V = \dim \ker T + \dim \operatorname{Im} T.$$

Then T is injective iff $\ker T = 0$ iff $\dim \ker T = 0$ iff $\dim \operatorname{Im} T = \dim W$ iff $\operatorname{Im} T = W$ iff T is surjective. □

3.6. **Applications to interpolation.** The rank-nullity theorem and ideas from linear algebra enable us to give slick new "non-constructive" proofs of many of the results from Section 1. A general theme is that for theoretical reasons it can be easier to compute the kernel of a linear transformation than it is to compute the image. But, by the rank-nullity theorem, if we know the dimension of the kernel then we know the dimension of the image.

Here we give new proofs of the main theorems from Section 1.

Theorem 3.25 (Division algorithm)**.** *Let $p, m \in \mathbb{R}[x]$ with $m \neq 0$. Then there are unique polynomials $q, r \in \mathbb{R}[x]$ (the* quotient *and* remainder*) with $\deg r < \deg m$ such that*

$$p = qm + r.$$

Proof. Let V_d be the vector space of polynomials of degree $\leq d$. Suppose $\deg p = d$ and $\deg m = e$. Define a linear transformation

$$\begin{aligned} T : V_{d-e} \times V_{e-1} &\to V_d \\ T(q, r) &= qm + r. \end{aligned}$$

(Check T is linear.) Note that $V_{d-e} \times V_{e-1}$ and V_d have the same dimension; the theorem claims that T is a bijection. By a corollary of the rank-nullity theorem, it is enough to show that T is injective.

Suppose $(q, r) \in \ker T$. Then $qm + r = 0$, so

$$-r = qm.$$

The LHS has degree $\leq e - 1$. If $q \neq 0$, then the RHS has degree $\geq e$. Therefore $q = 0$ and $r = 0$. So, $\ker T = \{(0, 0)\}$ and T is injective. □

Theorem 3.26 (Lagrangian interpolation: all solutions)**.** *Let $x_1, \ldots, x_n \in K$ be distinct, and let $y_1, \ldots, y_n \in K$. Suppose $d \geq n - 1$. Then there is a polynomial $f(x) \in K[x]$ of degree at most d such that $f(x_i) = y_i$. Furthermore, if g is any other such polynomial, then there is a polynomial q such that*

$$g = f + (x - x_1) \cdots (x - x_n)q.$$

Proof. Again let V_d be the vector space of polynomials of degree $\leq d$. Define a linear transformation

$$\begin{aligned} T : V_d &\to K^n \\ Tf &= (f(x_1), \ldots, f(x_n)) \end{aligned}$$

(Check T is linear.) Let us compute the kernel of T.

If $f \in \ker T$, then $f(x_i) = 0$ for all i. Therefore, by the division algorithm (and its corollaries in Section 1) it follows that $(x-x_1)\cdots(x-x_n)$ divides f. Thus

$$\ker T = \{(x-x_1)\cdots(x-x_n)q : q \in V_{d-n}\},$$

and $\ker T$ has a basis given by the polynomials $(x-x_1)\cdots(x-x_n)x^i$ $(i = 0, \ldots, d-n)$. (If $d = n-1$ then the basis is empty and $\ker T = \{0\}$.) Therefore,

$$\dim \ker T = d - n + 1.$$

By the rank-nullity theorem,

$$d + 1 = \dim V_d = \dim \ker T + \dim \operatorname{Im} T = d - n + 1 + \dim \operatorname{Im} T,$$

and therefore $\dim \operatorname{Im} T = n$. Therefore T is surjective, and the equation $Tf = \mathbf{y}$ has a solution f. Any two solutions f and g of the equation differ by an element of $\ker T$, since $T(f-g) = 0$. The description of an arbitrary solution follows. □

Theorem 3.27 (Lagrangian interpolation with derivatives: all solutions)**.** *Let $x_1, \ldots, x_n \in K$ be distinct, let $m_1, \ldots, m_n$ be positive integers, and for each i let $y_{i,0}, \ldots, y_{i,m_i-1} \in K$. Let $N = \sum m_i$, and suppose $d \geq N - 1$. Then there is a polynomial $f(x) \in K[x]$ of degree at most d such that $f^{(j)}(x_i) = y_{i,j}$ for all $i = 1, \ldots, n$ and $j = 0, \ldots, m_i-1$. Furthermore, if g is any other such polynomial, then there is a unique polynomial q such that*

$$g = f + (x-x_1)^{m_1}\cdots(x-x_n)^{m_n}q.$$

Proof. Carry out the same procedure as in the proof of the previous result, this time using the linear transformation $T : V_d \to K^N$ defined by

$$\begin{aligned} Tf = (&f^{(0)}(x_1), \ldots, f^{(m_1-1)}(x_1), \\ &f^{(0)}(x_2), \ldots, f^{(m_2-1)}(x_2), \\ &\vdots \\ &f^{(0)}(x_n), \ldots, f^{(m_n-1)}(x_n)). \end{aligned}$$

□

4. Multivariate interpolation: the Hilbert function

The general theme of single variable interpolation is that results can be proven for arbitrary points $x_1, \ldots, x_n \in K$ in the domain: if the points are somehow "special" then it is not really any easier or harder to solve the interpolation problem. This changes dramatically in several variables.

4.1. **Interpolation in the plane and the Hilbert function.** In this subsection, let $R = K[x, y]$ be the polynomial ring in two variables and let $V_d \subset K[x, y]$ be the subspace of polynomials of degree at most d. Consider a collection of distinct *points* $p_1, \ldots, p_n \in K^2$, so each $p_i = (x_i, y_i)$ is an ordered pair, and let $z_1, \ldots, z_n \in K$ be values.

Problem 4.1. Can we find a polynomial $f \in V_d$ such that $f(p_i) = z_i$ for $i = 1, \ldots, n$?

It is easy to rephrase this question in terms of linear algebra. Define an *evaluation transformation*

$$\begin{aligned} T : V_d &\to K^n \\ Tf &= (f(p_1), \ldots, f(p_n)). \end{aligned}$$

The dimension of V_d is $\binom{d+2}{2}$, since V_d has a basis given by the monomials $x^i y^j$ of degree $\leq d$, and there are $\binom{d+2}{2}$ of them. Then by the results in Chapter 3, we can make several immediate observations.

- If $\binom{d+2}{2} < n$, then for some choice of values $z_1, \ldots, z_n \in K$ there is no solution f to the problem. Indeed, T is not surjective, and any vector $\mathbf{z} = (z_1, \ldots, z_n) \in K^n$ which is not in the image of T is a list of values that cannot be achieved by any polynomial.
- If $\binom{d+2}{2} = n$, then if there is a solution to the problem for any list $z_1, \ldots, z_n \in K$, then that solution is unique. Conversely, if the only $f \in V_d$ with $f(p_i) = 0$ $(i = 1, \ldots, n)$ is $f = 0$, then the problem has a unique solution for any list $z_1, \ldots, z_n$. In other words, T is surjective if and only if it is injective.
- If $\binom{d+2}{2} > n$, then for a list of values $z_1, \ldots, z_n \in K$ there may or may not be a solution to the problem. If there is a solution f, then it is not unique: we have $\ker T \neq 0$, and adding an element of $\ker T$ to f provides additional solutions. (In fact, every solution to $Tf = \mathbf{z}$ can be obtained in this way.

In general, the problem of describing all the possible lists $\mathbf{z} \in K^n$ of values such that the interpolation problem $Tf = \mathbf{z}$ has a solution is not terribly interesting; this is the problem of computing the image of

a linear transformation, and it can be done by standard linear algebra techniques.

On the other hand, more qualitative analysis (e.g., is it injective or surjective? What is its rank?) of the map T is very interesting and deeply connected with the geometric configuration of the collection of points $p_1, \dots, p_n$. We thus switch our attention to this program.

Problem 4.2 (Interpolation problem: enlightened version)**.** Let $Z = \{p_1, \dots, p_n\} \subset K^2$ be a collection of n distinct points in the plane. For an integer $d \geq 0$, what is the rank of the evaluation map

$$\begin{aligned} T_{Z,d} &: V_d \to K^n \\ T_{Z,d}(f) &= (f(p_1), \dots, f(p_n))? \end{aligned}$$

In particular, is the map $T_{Z,d}$ either injective or surjective?

The answer to the problem is encoded in the *Hilbert function.*

Definition 4.3. Let $Z = \{p_1, \dots, p_n\} \subset K^2$ be a collection of n distinct points in the plane. The *Hilbert function of* Z is the function

$$\begin{aligned} h_Z &: \mathbb{N} \to \mathbb{N} \\ h_Z(d) &= \operatorname{rk} T_{Z,d}. \end{aligned}$$

Some simple properties are worth discussing now, before we study a bunch of examples:

Proposition 4.4. *Let $Z = \{p_1, \dots, p_n\} \subset K^2$ be a collection of $n \geq 1$ distinct points in the plane. The Hilbert function $h_Z(d)$ satisfies the following properties.*

(1) $h_Z(0) = 1$.

(2) For each $d \geq 0$,

$$h_Z(d) \leq \min\left\{ \binom{d+2}{2}, n \right\}$$

(3) The Hilbert function is increasing: if $d \leq e$, *then* $h_Z(d) \leq h_Z(e)$.

(4) If d *is sufficiently large, then* $h_Z(d) = n$. *In fact, if* $d \geq n-1$ *then* $h_Z(d) = n$.

Proof. (1) Since V_0 is just the constant polynomials, the image of $T_{Z,0} : V_0 \to K^n$ has basis given by the vector $(1, \dots, 1)$.

(2) The rank of the linear transformation $T_{Z,d} : V_d \to K^n$. is bounded by the rank of the domain and the rank of the codomain.

(3) Notice that if $d \leq e$, then $V_d \subset V_e$ is a subspace. Consider the linear transformations $T_{Z,e} : V_e \to K^n$ and $T_{Z,d} : V_d \to K^n$. Then if

$f \in V_d$, we have $T_{Z,e}(f) = T_{Z,d}(f)$. Therefore, we have a containment $\operatorname{Im} T_{Z,d} \subset \operatorname{Im} T_{Z,e}$, and it follows that $h_Z(d) \leq h_Z(e)$.

(4) By (3), it is enough to show that $h_Z(d) = n$ for some $d \geq 0$. The statement that $h_Z(d) = n$ means that the linear transformation $T_{Z,d} : V_d \to K^n$ is surjective. One way to show that $T_{Z,d}$ is surjective is to show that the image of $T_{Z,d}$ contains the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of K^n. In other words, for each $i = 1, \ldots, n$ we need to construct a polynomial $f \in V_d$ such that $f(p_i) = 1$ but $f(p_j) = 0$ for $j \neq i$. Since we can scale a polynomial, it is actually good enough to construct a polynomial $f \in V_d$ such that $f(p_i) \neq 0$ but $f(p_j) = 0$ for $j \neq i$. The next lemma shows that so long as $d \geq n - 1$, we can do this. □

Lemma 4.5. *Let $p \in K^2$ and let $q_1, \ldots, q_{n-1} \in K^2$ be points different from p. Then there is a polynomial $f \in V_{n-1}$ such that $f(p) \neq 0$ and $f(q_i) = 0$.*

Proof. Suppose $p = (a, b)$ and $q_1 = (x_1, y_1)$. Since $p \neq q_1$, either $a \neq x_1$ or $b \neq y_1$. Then either the vertical line $x = x_1$ through q_1 or the horizontal line $y = y_1$ through q_1 does not pass through p. Let $L_1 \in V_1$ be the equation of one of these lines that does not pass through p_1 (so that L_1 is either $x - x_1$ or $y - y_1$).

Repeat this procedure for each of the points $q_1, \ldots, q_{n-1}$ to find lines $L_1, \ldots, L_{n-1}$ such that L_i passes through q_i but does not pass through p. Then the product $L_1 \cdots L_{n-1}$ is the desired polynomial. □

Remark 4.6. Linear transformations $T : V \to W$ such that

$$\dim \operatorname{Im} T = \min\{\dim V, \dim W\}$$

are said to have *maximal rank*, since the inequality

$$\dim \operatorname{Im} T \leq \min\{\dim V, \dim W\}$$

is always true. If T has maximal rank, then it is either injective (if $\dim \operatorname{Im} T = \dim V$) or surjective (if $\dim \operatorname{Im} T = \dim W$) or both (if $\dim \operatorname{Im} T = \dim V = \dim W$). Conversely, if T has maximal rank, then it is either injective or surjective (or both).

4.2. **The Hilbert function: examples.** In this section we discuss the Hilbert function of some collections of points in the plane.

Example 4.7 (2 points)**.** Let $Z = \{p, q\} \subset K^2$ be two points in the plane. The Hilbert function of Z is

$$h_Z(d) = \begin{cases} 1 & d = 0 \\ 2 & d \geq 1. \end{cases}$$

This follows directly from our general results in the previous section.

Example 4.8 (3 noncollinear points)**.** Let $Z = \{p, q, r\} \subset K^2$ be three points in the plane which are not collinear. Then $h_Z(0) = 1$. Consider $T_{Z,1} : V_1 \to K^3$. By definition, there is no degree 1 polynomial that vanishes on Z, so $\ker T_{Z,1} = \{0\}$. Since $\dim V_1 = 3$, we find that $T_{Z,1}$ is a bijection. Therefore,

$$h_Z(d) = \begin{cases} 1 & d = 0 \\ 3 & d \geq 1. \end{cases}$$

Example 4.9 (3 collinear points)**.** Let $Z = \{p, q, r\} \subset K^2$ be three points in the plane which are collinear. Then $h_Z(0) = 1$. Consider $T_{Z,1} : V_1 \to K^3$. Since Z lies on a unique line, we have $\dim \ker T_{Z,1} = 1$, and therefore $T_{Z,1}$ has rank 2 by the rank-nullity theorem. We also know $h_Z(2) = 3$ by general principles. So,

$$h_Z(d) = \begin{cases} 1 & d = 0 \\ 2 & d = 1 \\ 3 & d \geq 2. \end{cases}$$

Example 4.10 (4 points, not all on a line)**.** Let $Z = \{p, q, r, s\} \subset K^2$ be four points in the plane which do not all lie on the same line. As in the example of 3 noncollinear points, we find that $h_Z(0) = 1$ and $h_Z(1) = 3$.

We claim that $h_Z(2) = 4$. To do this, we can show that $T_{Z,2}$ is surjective by constructing polynomials that map to the standard basis vectors of K^4. Without loss of generality, we construct a polynomial of degree ≤ 2 that is nonzero at p and zero at q, r, s.

If the line $L = \overline{qr}$ does not pass through p, then let M be a line that passes through s and not p. Then LM is the desired polynomial of degree ≤ 2.

Next suppose the line $L = \overline{qr}$ does pass through p. By assumption, it does not pass through s. Then the line $M = \overline{qs}$ does not pass through p. Let M' be any line through r that does not pass through p. Then MM' is the desired polynomial of degree ≤ 2.

We conclude that

$$h_Z(d) = \begin{cases} 1 & d = 0 \\ 3 & d = 1 \\ 4 & d \geq 2. \end{cases}$$

Example 4.11 (4 collinear points)**.** Let $Z = \{p, q, r, s\} \subset K^2$ be four collinear points. As in the case of three collinear points, we have $h_Z(0) = 1$ and $h_Z(1) = 2$.

Consider the map $T_{Z,2} : V_2 \to K^4$. The kernel consists of the degree 2 polynomials that vanish on the 4 points. Actually, a degree 2 polynomial which vanishes on 4 collinear points must be divisible by the equation $L = 0$ of the line (see the next proposition). Thus any element of $\ker T_{Z,2}$ can be written in the form $L \cdot q$ for some $q \in V_1$. It follows that $\ker T_{Z,2}$ has a basis given by Lx, Ly, L, and so $\dim \ker T_{Z,2} = 3$. Therefore $\operatorname{rk} T_{Z,2} = 3$, and the Hilbert function is

$$h_Z(d) = \begin{cases} 1 & d = 0 \\ 2 & d = 1 \\ 3 & d = 2 \\ 4 & d \geq 3. \end{cases}$$

Proposition 4.12. *Let $p_1, \ldots, p_{d+1} \in K^2$ be $d+1$ collinear points in the plane, and let $L = 0$ be the defining equation of the line. If a polynomial $f \in V_d$ of degree at most d vanishes at $p_1, \ldots, p_{d+1}$, then f is divisible by L.*

Proof. By making a linear change of coordinates, we may as well assume the points $p_1, \ldots, p_{d+1}$ all lie on the x-axis $y = 0$. Say $p_i = (x_i, 0)$. The polynomial $f(x, 0)$ of x has $d+1$ roots since $f(x_i, 0) = 0$. Since the degree of $f(x, 0)$ is at most d, it must be the zero polynomial. This means that every monomial of $f(x, y)$ has a y in it, and therefore y divides f. □

4.3. **The Hilbert function of "random" points.** As we saw in the previous section, if $Z \subset K^2$ is a collection of n points which have some special relationship with one another (e.g. they are collinear), then the Hilbert function $h_Z(d)$ can be complicated. At the other extreme, if there are no interesting relationships between the points, then we should be able to easily compute the Hilbert function.

Definition 4.13. A collection $Z \subset K^2$ of n points is in *boring position* if

$$h_Z(d) = \min \left\{ \binom{d+2}{2}, n \right\}$$

for all $d \geq 0$. In other words, the map $T_{Z,d}$ has maximal rank for all $d \geq 0$.

In this section, we will show that for any $n \geq 1$ there is a collection $Z \subset K^2$ of n points in boring position.

Example 4.14. In the previous section we showed several collections of points are in boring position. Any two points in K^2 are in boring position. Three points in K^2 are in boring position iff they are not

collinear. Four points in K^2 are in boring position iff they are not collinear.

Note that points in boring position aren't necessarily as boring as possible: if we have four points in K^2 with exactly three on a line, then the Hilbert function is the same as for a more random collection of points.

Theorem 4.15. *For any $n \geq 1$, there is a collection $Z \subset K^2$ of n points in boring position.*

Proof. The proof is by induction on n. The Hilbert function of a single point $Z = \{p\}$ is constant $h_Z(d) = 1$.

Suppose there is a collection $Z' \subset K^2$ of $n-1$ points in boring position, and let $p \in K^2$ be some point of K^2 not in Z'. We determine conditions on p that ensure that $Z = Z' \cup \{p\}$ is in boring position.

Let $D \geq 0$ be the first integer such that $h_{Z'}(D) = n - 1 < \binom{D+2}{2}$. Then for $d < D$, we have $h_{Z'}(d) = \binom{d+2}{2}$. Then the map $T_{Z',d} : V_d \to K^{n-1}$ is injective. Clearly $T_{Z,d} : V_d \to K^n$ is also injective, since if $T_{Z,d}(f) = 0$ then also $T_{Z',d}(f) = 0$, so that $\ker T_{Z,d} \subset \ker T_{Z',d}$. Therefore

$$h_Z(d) = \binom{d+2}{2} = \min\left\{\binom{d+2}{2}, n\right\}$$

for $d < D$.

Degree D is more interesting, and requires that we be careful about how the point p is chosen. Let $U = \ker T_{Z',D} \subset V_D$ be the subspace of polynomials of degree at most D which vanish on Z'. Observe that $U \neq \{0\}$, since

$$\binom{D+2}{2} = \dim V_D = \dim U + n - 1$$

so that

$$\dim U = \binom{D+2}{2} - n + 1 > 0.$$

Pick some polynomial $0 \neq f \in U$. Then by an old homework problem, we can find some point $p \in K^2$ such that $f(p) \neq 0$. Then the subspace

$$U(-p) = \{f \in U : f(p) = 0\} \subset U$$

is a *proper* subspace of U. Its dimension can be computed by observing that there is a surjective linear transformation

$$S : U \to K^1$$
$$Sf = f(p)$$

with kernel $U(-p)$. Therefore $\dim U(-p) = \dim U - 1$.

Finally, consider the map $T_{Z,D} : V_D \to K^n$. Clearly $\ker T_{Z,D} = U(-p)$. Therefore

$$\operatorname{rk} T_{Z,D} = \dim V_D - \dim U(-p) = \binom{D+2}{2} - \dim U + 1 = n,$$

and $h_Z(D) = n$. Since $h_Z(d)$ is increasing and bounded by n, we conclude $h_Z(d) = n$ for all $d \geq D$. □

In a sense, *almost every* collection of points is in boring position: if a collection of points does something interesting, it is an "accident." To phrase this phenomenon more rigorously, we have to introduce some more terminology from algebraic geometry.

5. Algebraic geometry: varieties in affine space

We now turn to studying algebraic geometry in earnest. Algebraic geometry is roughly about the study of solutions of systems of polynomial equations. On the one hand, this includes the question of *finding* or *describing* all the solutions of a system of polynomial equations. This can be viewed as an essentially algebraic question: find the common roots of some polynomials. On the other hand, in practice this question can be very difficult and not particularly enlightening; in that case, we instead aim to describe the *shape* or *geometric properties* of the solutions.

5.1. **First examples.** The purpose of this section is to introduce you to the huge sea of examples that are available in algebraic geometry. Most of the assertions here are made without justification; some are easy, some are hard.

Example 5.1. In $\mathbb{R}^2$, consider the common solutions of $y = x$ and $x^2+y^2 = 2$ (sketch the curves!). The intersection is two points, $(x, y) = (\pm\sqrt{2}, \pm\sqrt{2})$.

Example 5.2. In $\mathbb{R}^3$, consider the common solutions of $x^2+y^2-z^2 = 1$ and $z = 0$. The horizontal cross-sections $z = c$ of the surface $x^2 + y^2 - z^2 = 1$ are circles in the plane $z = c$; the common solutions of $x^2 + y^2 - z^2 = 1$ and $z = 0$ are the unit circle $x^2 + y^2 = 1$. We can parameterize the solutions by a trigonometric function $f(t) = (\cos(t), \sin(t), 0)$. (Can you find *rational* functions that parameterize the solution set?)

Example 5.3 (Lines and conics)**.** A single equation $f(x, y) \in K[x, y]$ describes a *curve* $f(x, y) = 0$ in the plane K^2. Several examples are familiar from high school analytic geometry:

(1) A degree 1 equation $ax + by + c = 0$ describes a line in K^2.
(2) A degree 2 equation $F = ax^2 + bxy + cy^2 + dx + ey + f = 0$ describes a conic C in K^2. (Warning: it is possible that C is empty. For example, if $K = \mathbb{R}$, consider $x^2+y^2 = -1$. There are no solutions. This is a very good reason for studying algebraic geometry over algebraically closed fields, such as $\mathbb{C}$.)
(3) If $K = \mathbb{R}$, conics can be further separated into cases by the sign of the *discriminant* $b^2 - 4ac$. Suppose that F is not a product of two linear forms, and that the zero locus $F = 0$ is nonempty.
    - If $b^2 - 4ac > 0$, then C is a hyperbola.
    - If $b^2 - 4ac = 0$, then C is a parabola.
    - If $b^2 - 4ac < 0$, then C is an ellipse.

Any two hyperbolas/parabolas/ellipses can be brought to one another by a change of variables of the form

$$x' = ax + by + e$$
$$y' = cx + dy + f$$

where the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible. Conversely, such changes of coordinates preserve the above trichotomy. Such a change of coordinates is called an *invertible affine linear transformation*.

(4) If $K = \mathbb{C}$, the above classification of conics becomes simpler: the discriminant $b^2 - 4ac$ is complex, so its sign is not a meaningful concept.
   - If $b^2 - 4ac \neq 0$, then by an invertible affine linear transformation we can carry C to a circle $x^2 + y^2 = 1$.
   - If $b^2 - 4ac = 0$, then by an invertible affine linear transformation we can carry C to a parabola $y = x^2$.

Example 5.4 (Cubic curves)**.** Curves of higher degree are both tremendously complicated and tremendously interesting. A degree 3 curve $f(x, y) = 0$ is called an *elliptic curve*, and these play a huge role in modern number theory and cryptography. If $K = \mathbb{C}$, then by linear changes of coordinates we can essentially bring such curves into one of three normal forms:

$$y^2 = x(x-1)(x-\lambda) \qquad (\lambda \neq 0, 1)$$
$$y^2 = x^3 + x^2$$
$$y^2 = x^3.$$

These respectively describe a *smooth cubic*, a *nodal cubic*, and a *cuspidal cubic*. In the case of the smooth cubic, the parameter λ controls the "geometry" or "shape" of the curve, and different values of λ give curves that are legitimately different from one another. (Compare this situation with conics, where huge classes of conics that perhaps look different from one another are actually the same in a different coordinate system!)

Example 5.5. The study of curves of degree ≥ 4 in K^2 is very much a current subject of active research.

Example 5.6 (Surfaces)**.** In three variables, the solutions $f(x, y, z) = 0$ of an equation $f \in K[x, y, z]$ define a *surface* in K^3. As with curves, low degree situations are easy to understand, and things get progressively more difficult.

(1) A degree 1 equation $ax + by + cz + d = 0$ describes a plane.
(2) In degree 2, the generalization of conic curves is quadric surfaces. The most important examples are:
   - $x^2 + y^2 + z^2 = 1$, a unit sphere.
   - $z^2 = xy$, a quadric cone.
   - $z = xy$, a smooth quadric surface.

   As with conics there is a notion of equivalence between different conics given by invertible affine linear transformations, but we won't go into this here.

   Over $\mathbb{C}$, any quadric surface is swept out by lines that lie on the surface. For example, for each $\lambda \in \mathbb{C}$, the quadric surface $z = xy$ contains all the lines of the form

$$L_\lambda = \{(t, \lambda, \lambda t) : t \in \mathbb{C}\}.$$

   It also contains all the lines of the form

$$M_\lambda = \{(\lambda, t, \lambda t) : t \in \mathbb{C}\}.$$

   (Can you find lines on the sphere $x^2 + y^2 + z^2 = 1$? Clearly the real sphere doesn't contain any lines, so any lines you find will have to involve complex numbers.)
(3) In degree 3, we have cubic surfaces such as the *Fermat cubic*

$$x^3 + y^3 + z^3 = 1.$$

   These have a lot of beautiful geometry; for example, (up to small lies) a cubic contains 27 lines arranged in an interesting combinatorial pattern.
(4) Past degree 3, we quickly get into open research questions.

So far our examples have focused on *hypersurfaces*, which are the zero loci of a single polynomial. A lot of algebraic geometry is considerably easier in this case–this is the difference between solving a single polynomial equation and solving a system of polynomial equations. On the other hand, even if your goal is to study hypersurfaces then it often becomes necessary to study systems of several equations.

Example 5.7. A *line* in K^3 is defined by a system of equations

$$ax + by + cz + d = 0$$
$$a'x + b'y + c'z + d' = 0$$

where the vectors (a, b, c) and (a', b', c') are not parallel. Equivalently, a line is the intersection of two planes. (These vectors are the normal vectors to the planes, so the planes are not parallel iff these vectors are not parallel.)

Example 5.8. We saved our most important example for last. Let $\{p_1, \ldots, p_n\} = Z \subset K^2$ be a collection of points in the plane. We claim that Z is the common zero locus of a system of polynomial equations.

First let us consider the simple case of two points. Let $p_1 = (x_1, y_1)$ and let $p_2 = (x_2, y_2)$. Then p_i is the common zero locus of the system of equations

$$x - x_i = 0$$
$$y - y_i = 0$$

Consider the system of equations

$$(x - x_1)(x - x_2) = 0$$
$$(x - x_1)(y - y_2) = 0$$
$$(y - y_1)(x - x_2) = 0$$
$$(y - y_1)(y - y_2) = 0.$$

If a point $p = (x_0, y_0)$ satisfies this system, then we see that x_0 is x_1 or x_2 (from the first equation) and y_0 is y_1 or y_2 (from the fourth equation). This leaves the four possibilities $p \in \{(x_1, y_1), (x_1, y_2), (x_2, y_1), (x_2, y_2)\}$. If (x_1, y_2) satisfies the system, then the third equation gives $y_1 = y_2$ (so $(x_1, y_2) = (x_1, y_1) = p_1$) or $x_1 = x_2$ (so $(x_1, y_2) = (x_2, y_2) = p_2$)). Similarly, if (x_2, y_1) satisfies the system then the second equation shows that either $(x_2, y_1) = p_1$ or $(x_2, y_1) = p_2$. We conclude that the only solutions of the system are p_1 and p_2.

In the general case of n points, write each point as $p_i = (x_i, y_i)$. Then p_i is the common zero locus of the equations

$$F_i^1 := x - x_i = 0$$
$$F_i^2 := y - y_i = 0.$$

Consider the set of equations

$$\{F_1^{a_1} F_2^{a_2} \cdots F_n^{a_n} : a_i \in \{1, 2\}\}.$$

This directly generalizes the system of equations from the two point case above, and it can be shown that the common zero locus of these equations is $Z = \{p_1, \ldots, p_n\}$. (This will become more clear after we make a more general argument about unions of zero loci in the next section.)

5.2. **Affine space.** Until now, when we've looked at collections of points or zero loci of polynomials, we've considered them as subsets of K^n. However, since K^n is a vector space, the natural "transformations" of K^n are the linear mappings $T : K^n \to K^n$. In particular, the origin $\mathbf{0}$ is a distinguished point in K^n.

When we compare configurations of points or shapes in algebraic geometry, the origin in K^n is not really a special point—it has nice coordinates, sure, but for example we don't care about the relative position between our shapes and the origin. The vector space structure on K^n is not helpful.

Definition 5.9. As a set, *affine n-space over K* is the set $\mathbb{A}^n_K = K^n$. We don't give $\mathbb{A}^n_K$ any other operations—in particular, it is not a vector space. If the field K is clear or unimportant, we will drop it from the notation and write $\mathbb{A}^n$.

Since $\mathbf{0} \in \mathbb{A}^n$ is no longer a distinguished point, the natural transformations of $\mathbb{A}^n$ should allow us to move $\mathbf{0}$ to any other point $\mathbf{x}_0 \in \mathbb{A}^n$. This can be accomplished by allowing translations.

Definition 5.10. An *affine linear transformation* $T : \mathbb{A}^n \to \mathbb{A}^n$ is any function of the form

$$T\mathbf{x} = A\mathbf{x} + \mathbf{x}_0,$$

where A is an $n \times n$ matrix with entries in K and $\mathbf{x}_0 \in K^n$.

Lemma 5.11. *The composition of two affine linear transformations is an affine linear transformation.*

Proof. Exercise. □

An affine linear transformation is *invertible* if there is an affine linear transformation $S : \mathbb{A}^n \to \mathbb{A}^n$ which is a two-sided inverse: $TS = ST = I$.

Proposition 5.12. *Let $T : \mathbb{A}^n \to \mathbb{A}^n$ be an affine linear transformation, given by a formula*

$$T\mathbf{x} = A\mathbf{x} + \mathbf{x}_0.$$

Then T is invertible if and only if the matrix A is invertible.

Proof. Exercise. □

When two subsets of A^n can be carried to one another by an invertible affine linear transformation, it is natural to think of them as being equivalent in some sense.

Definition 5.13. Two subsets $X, Y \subset \mathbb{A}^n$ are *affinely equivalent* if there is an invertible affine linear transformation $T : \mathbb{A}^n \to \mathbb{A}^n$ such that $T(X) = Y$.

Small number of points in $\mathbb{A}^n$ can be brought into a "normal form" by an affine linear transformation if they are not in some kind of special position:

Theorem 5.14. *Let $p_1, \dots, p_{n+1} \in \mathbb{A}^n$, and suppose they do not lie on a hyperplane: there is no nontrivial linear equation*

$$a_1 x_1 + \cdots + a_n x_n + c = 0$$

that all of the points satisfy. Then there is an invertible affine linear transformation $T : \mathbb{A}^n \to \mathbb{A}^n$ such that $T(p_{n+1}) = \mathbf{0}$ and $T(p_i) = \mathbf{e}_i$, where as usual $\mathbf{e}_i$ denotes the coordinate point with 1 in the ith position and 0 in the other positions.

Proof. Since the composition of invertible affine linear transformations is invertible affine linear, we can construct our transformation $\mathbb{A}^n \to \mathbb{A}^n$ in several stages. First, there is an invertible affine linear transformation that sends p_{n+1} to $\mathbf{0}$: just take a translation by $-p_{n+1}$, which has an inverse given by translation by p_{n+1}. For $i = 1, \dots, n$, let $\mathbf{q}_i = p_i - p_{n+1}$ be the image of p_i under translation by $-p_{n+1}$. We are going to think of the $\mathbf{q}$'s as vectors in K^n instead of just as points in $\mathbb{A}^n$, which is why we wrote them bold.

We claim that the vectors $\mathbf{q}_1, \dots, \mathbf{q}_n$ are linearly independent in K^n. The alternative is that they span a proper subspace of K^n. Let $U \subset K^n$ be their span, and arbitrarily pick an $(n-1)$-dimensional subspace $V \subset K^n$ such that $U \subseteq V \subset K^n$. Then V is a hyperplane in K^n that contains $\mathbf{q}_1, \dots, \mathbf{q}_n$ and $\mathbf{0}$. Translating back to our original coordinate system, we obtain a hyperplane in $\mathbb{A}^n$ that contains $p_1, \dots, p_{n+1}$, contradicting our assumption.

Since $\mathbf{q}_1, \dots, \mathbf{q}_n$ are linearly independent in K^n, they are a basis of K^n. Therefore, we can find a linear transformation $S : K^n \to K^n$ such that $S\mathbf{q}_i = \mathbf{e}_i$ for $i = 1, \dots, n$. It has an invertible matrix A with respect to the standard bases, since it takes a basis to a basis.

Finally, the desired invertible affine linear tranfsormation is the composition of these two transformations:

$$T\mathbf{x} = A(\mathbf{x} - p_{n+1}) = A\mathbf{x} - Ap_{n+1}. \quad \square$$

By composing several transformations we can get more impressive sounding/memorable results:

Corollary 5.15. *Let $p_1, \dots, p_{n+1} \in \mathbb{A}^n$ and $q_1, \dots, q_{n+1} \in \mathbb{A}^n$ be two collections of points, with neither collection lying on a hyperplane. Then there is an invertible affine linear transformation $T : \mathbb{A}^n \to \mathbb{A}^n$ such that $Tp_i = q_i$ for all i. In other words, any two collections of $n+1$ points which do not lie on a hyperplane are affinely equivalent.*

Example 5.16. Any triangle in $\mathbb{A}^2$ can be carried to any other triangle in $\mathbb{A}^2$ by an invertible affine linear transformation. Indeed, if a transformation carries points p_1 and p_2 to q_1 and q_2, then it must carry

the line segment $\overline{p_1p_2}$ to the line segment $\overline{q_1q_2}$. (Check!) So all we have to do is carry the vertices to the vertices, and the triangle will come along for the ride. Any two triangles in $\mathbb{A}^2$ are affinely equivalent.

It is particularly important to understand how affine linear transformations act on the zero loci of polynomials. Unfortunately, the answer is maybe not the one you would guess unless you really think about it:

Proposition 5.17. *Let $F(x_1, \dots, x_n) \in K[x_1, \dots, x_n]$, and let $T : \mathbb{A}^n \to \mathbb{A}^n$ be an invertible affine linear transformation. Let $S : \mathbb{A}^n \to \mathbb{A}^n$ be the inverse of T. Then if $X : F(x_1, \dots, x_n) = 0$ is the zero locus of F, its image $T(X)$ is the zero locus of $F(S(x_1, \dots, x_n))$.*

Proof. Exercise. □

Warning 5.18. In particular, it is really easy to forget that you have to take the inverse of T!

Example 5.19. Let $X : F = x^2 + y^2 - 1 = 0$ be the unit circle, and consider a translation $T : \mathbb{A}^2 \to \mathbb{A}^2$ by $(1, 1)$, so

$$T(x, y) = (x + 1, y + 1).$$

Its inverse is the translation by $(-1, -1)$

$$S(x, y) = (x - 1, y - 1).$$

Then the equation of $T(X)$ is

$$0 = F(S(x, y)) = F(x - 1, y - 1) = (x - 1)^2 + (y - 1)^2 - 1,$$

which is indeed the equation of a unit circle centered at $(1, 1)$.

5.3. **Affine varieties and the Zariski topology.** We finally come to the central definition of algebraic geometry. For simplicity, we now assume that our field K is algebraically closed. Think $K = \mathbb{C}$ if you prefer.

Definition 5.20. Let $\{f_\alpha\}_{\alpha \in A}$ be a (possibly infinite) set of polynomials in $K[x_1, \dots, x_n]$. The *affine variety* cut out by these polynomials is

$$V(\{f_\alpha\}_{\alpha \in A}) := \{p \in \mathbb{A}^n : f_\alpha(p) = 0 \text{ for all } \alpha \in A\}.$$

In case the set of polynomials is a finite list $f_1, \dots, f_k$, we will write $V(f_1, \dots, f_k)$ for their common zero locus.

Remark 5.21. For technical reasons it is useful to allow infinite lists of equations in the definition of an affine variety. However, this is not truly necessary if you do some more work. A consequence of the *Hilbert basis theorem* is that any affine variety can be cut out by finitely many equations. The proof of this requires more commutative algebra than we have developed, but this would be a good project topic.

In the previous sections we have seen that many interesting geometric objects are affine varieties. In particular, curves, surfaces, finite collections of points, etc., can all be realized as the common zero loci of collections of polynomials. We now start to learn some of the general properties of affine varieties.

Example 5.22. The empty set $\emptyset \subset \mathbb{A}^n$ is an affine variety, since $V(1) = \emptyset$.

Example 5.23. The full affine space $\mathbb{A}^n$ is an affine variety, since either $V(0) = \mathbb{A}^n$ or $V(\emptyset) = \mathbb{A}^n$.

Proposition 5.24. *An arbitrary intersection of affine varieties in $\mathbb{A}^n$ is an affine variety.*

Proof. For ease of notation, let's discuss the case of two varieties $X, Y \subset \mathbb{A}^n$. Then $X = V(\{f_\alpha\})$ and $Y = V(\{g_\beta\})$ for some sets of polynomials $\{f_\alpha\}$ and $\{g_\beta\}$. But then

$$X \cap Y = V(\{f_\alpha\} \cup \{g_\beta\}).$$

Indeed a point $p \in \mathbb{A}^n$ satisfies all the equations of X and all the equations of Y if and only if p is in $X \cap Y$. The generalization to an arbitrary intersection is immediate. □

Proposition 5.25. *A union of two affine varieties in $\mathbb{A}^n$ is an affine variety. Therefore, a finite union of affine varieties in $\mathbb{A}^n$ is an affine variety.*

Proof. We foreshadowed this proof when discussing finite collections of points. Let $X, Y \subset \mathbb{A}^n$ be affine varieties, defined by collections of equations $\{f_\alpha\}_{\alpha \in A}$ and $\{g_\beta\}_{\beta \in B}$. Consider the set of equations

$$S = \{f_\alpha g_\beta : \alpha \in A \text{ and } \beta \in B\}.$$

We claim that $V(S) = X \cup Y$.

Suppose $p \in X \cup Y$. Then $p \in X$ or $p \in Y$; without loss of generality, say $p \in X$. Then $f_\alpha(p) = 0$ for all $\alpha \in A$, and consequentially $(f_\alpha g_\beta)(p) = 0$ for all $f_\alpha g_\beta \in S$. Therefore, $p \in V(S)$.

Conversely, suppose $p \notin X \cup Y$. Then $p \notin X$ and $p \notin Y$. Since X is exactly the zero locus of the f_α, this means that we can find some $\alpha_0 \in A$ such that $f_{\alpha_0}(p) \neq 0$. Likewise, we can find some $\beta_0 \in B$ such that $g_{\beta_0}(p) \neq 0$. But then $(f_{\alpha_0} g_{\beta_0})(p) \neq 0$, and so $p \notin V(S)$.

Therefore, $X \cup Y = V(S)$ is an affine variety. The result for finite unions follows by induction. □

Proposition 5.26. *Let $X \subset \mathbb{A}^n$ and $Y \subset \mathbb{A}^m$ be affine varieties. Then the Cartesian product*

$$X \times Y = \{(x, y) : x \in X \text{ and } y \in Y\} \subset \mathbb{A}^n \times \mathbb{A}^m$$

is an affine variety.

Note that by abuse of notation we think of $\mathbb{A}^n \times \mathbb{A}^m$ as being $\mathbb{A}^{n+m}$. If we think of $\mathbb{A}^n$ as having coordinates $x_1, \ldots, x_n$ and $\mathbb{A}^m$ as having coordinates $y_1, \ldots, y_m$, we can think of $\mathbb{A}^{n+m}$ as having coordinates $x_1, \ldots, x_n, y_1, \ldots, y_m$. Thus affine varieties in $\mathbb{A}^{n+m}$ can be cut out by polynomials from the ring $K[x_1, \ldots, x_n, y_1, \ldots, y_m]$.

Proof. Suppose that X is cut out by polynomials $f_\alpha(x_1, \ldots, x_n)$ and Y is cut out by polynomials $g_\beta(y_1, \ldots, y_m)$. Consider the union S of these two sets of equations in the polynomial ring $K[x_1, \ldots, x_n, y_1, \ldots, y_m]$. Write a point $p \in \mathbb{A}^n \times \mathbb{A}^m$ as a pair (p', p''), where $p' \in \mathbb{A}^n$ and $p'' \in \mathbb{A}^m$. Then

$$\begin{aligned} p \in X \times Y &\Leftrightarrow p' \in X \text{ and } p'' \in Y \\ &\Leftrightarrow p' \in V(\{f_\alpha\}) \text{ and } p'' \in V(\{g_\beta\}) \\ &\Leftrightarrow p \in V(\{f_\alpha\}) \text{ and } p \in V(\{g_\beta\}) \\ &\Leftrightarrow p \in V(S). \end{aligned}$$

Therefore, $X \times Y = V(S)$ is an affine variety. □

The facts we have proved about unions and intersections show that the affine varieties in $\mathbb{A}^n$ form the closed sets of a *topology* on $\mathbb{A}^n$. We quickly recall this basic concept.

Definition 5.27. Let X be a set, and let $\tau \subset \mathcal{P}(X)$ be a collection of subsets of X. (Here $\mathcal{P}(X)$ is the power set of X, the set of all subsets of X.) The set τ is called a *topology on* X if it satisfies the following properties:

(1) We have $\emptyset \in \tau$ and $X \in \tau$.
(2) An arbitrary union of elements of τ is in τ.
(3) Finite intersections of elements of τ are in τ.

The elements $U \in \tau$ are called *open sets*. A subset $Z \subset X$ is called *closed* if its complement $X \setminus Z$ is open.

Remark 5.28. The standard definition of a topology defines the open sets first, and the definition of closedness is a consequence. In algebraic geometry the closed sets are more fundamental, so we define them first and call a subset U open if its complement is closed.

Definition 5.29. A subset $X \subset \mathbb{A}^n$ is *closed* if it is an affine variety. That is, X is cut out by a collection of polynomial equations. A subset $U \subset \mathbb{A}^n$ is *open* if its complement is closed.

Our earlier results imply similar sounding facts for open sets:

Example 5.30. In $\mathbb{A}^n$, the empty set and $\mathbb{A}^n$ are open.

Corollary 5.31. *In $\mathbb{A}^n$, an arbitrary union of open sets is open.*

Proof. Let $\{U_\alpha\}_{\alpha \in A}$ be open sets. Then by De Morgan's law,

$$X \setminus \bigcup_\alpha U_\alpha = \bigcap_\alpha (X \setminus U_\alpha)$$

is an intersection of closed sets, so is closed. Therefore the complement of $\bigcup U_\alpha$ is closed, and $\bigcup U_\alpha$ is open. □

Corollary 5.32. *In $\mathbb{A}^n$, an intersection of two open sets is open. Therefore, finite intersections of open sets are open.*

Proof. Let $U, V \subset \mathbb{A}^n$ be open. Then

$$X \setminus (U \cap V) = (X \setminus U) \cup (X \setminus V)$$

is a finite union of closed sets, so is closed. Therefore $U \cap V$ is open. □

Thus we have proved the following result.

Theorem and Definition 5.33. *Let $\tau \subset \mathcal{P}(\mathbb{A}^n)$ be the collection of open subsets of $\mathbb{A}^n$. Then τ is a topology on $\mathbb{A}^n$, called the* Zariski topology.

5.4. **Irreducibility.** Geometry really begins once topology is involved—topology allows to speak about all kinds of geometric properties, such as continuity, convergence of sequences, "closeness," as well as things like boundedness or compactness. But, the Zariski topology is incredibly weird! You can be easily mislead if you try to make arguments about the Zariski topology while thinking about more usual topologies such as the standard topology on $\mathbb{R}^n$.

Example 5.34 (Zariski topology on $\mathbb{A}^1$)**.** The only closed subsets of $\mathbb{A}^1$ are the whole space $\mathbb{A}^1$, the empty set, and finite collections of points. Indeed, if $f \in K[x]$ is a nonzero polynomial, then $V(f) \subset \mathbb{A}^1$ is the set of roots of f, and there are only finitely many of them. Given a list $\{f_\alpha\}$ of polynomials, the common zero locus of them all is contained in the zero locus of any one of them, so any closed set except $\mathbb{A}^1$ is finite.

Conversely, we know that every finite subset of $\mathbb{A}^1$ is closed: if $Z = \{x_1, \ldots, x_k\} \subset \mathbb{A}^1$, then $Z = V((x - x_1) \cdots (x - x_k))$.

Consequently, the open sets in $\mathbb{A}^1$ are the empty set and complements of finite sets. Open sets are huge! This topology on $\mathbb{A}^1$ is a standard example in topology with many pathological properties.

Example 5.35. A topological space X is called *Hausdorff* if for any two distinct points $x, y \in X$, you can find open sets U, V such that $x \in U$, $y \in V$, and $U \cap V = \emptyset$. In other words, distinct points can be "bubbled off" in their own little neighborhoods. Reasonable topological spaces (e.g. the usual topology on $\mathbb{R}^n$) are Hausdorff. The Zariski topology on $\mathbb{A}^n$ ($n \geq 1$) is *not Hausdorff*! The next result proves something even worse.

Proposition 5.36. *Let $U, V \subset \mathbb{A}^n$ be nonempty open sets. Then $U \cap V \neq \emptyset$.*

Proof. Let $X = \mathbb{A}^n \setminus U$ and $Y = \mathbb{A}^n \setminus V$ be the complementary closed sets. Then $U \cap V \neq \emptyset$ means that $X \cup Y \neq \mathbb{A}^n$. The closed sets X and Y are each cut out by some equations; pick one nonzero equation f from the list defining X, and one nonzero equation g from the list defining Y. Then $X \subset V(f)$ and $Y \subset V(g)$, and it will be enough to show that $V(f) \cup V(g) \neq \mathbb{A}^n$. But $V(f) \cup V(g)$ is $V(fg)$, so it is enough to show that $V(fg) \neq \mathbb{A}^n$. Here fg is a nonzero polynomial, so we need to see that there is some point p such that $(fg)(p) \neq 0$. The next lemma shows this and more. □

Lemma 5.37. *Let K be an algebraically closed field, and let $f(x_1, \ldots, x_n) \in K[x_1, \ldots, x_n]$ be a polynomial.*

1. *If $f \neq 0$, then there are infinitely many points p such that $f(p) \neq 0$.*
2. *If $n \geq 2$ and f is nonconstant, then there are infinitely many points p such that $f(p) = 0$.*

Proof. (1) The result is clear if f is a nonzero constant, so suppose $\deg f \geq 1$. Without loss of generality, assume x_1 appears in f. Then we can write

$$f = g_d(x_2, \ldots, x_n)x_1^d + \cdots + g_0(x_2, \ldots, x_n),$$

where g_d is nonzero. By induction on n, we can find some point $(a_2, \ldots, a_n) \in \mathbb{A}^{n-1}$ such that $g_d(a_2, \ldots, a_n) \neq 0$. Now consider the one-variable polynomial

$$f(x_1, a_2, \ldots, a_n) = g_d(a_2, \ldots, a_n)x_1^d + \cdots .$$

By our choice of $a_2, \ldots, a_n$, it has degree d. Therefore, it has at most d roots. Then if a_1 is not a root of this polynomial, we have $f(a_1, \ldots, a_n) \neq 0$. Since K is algebraically closed it is infinite (this is a

theorem whose proof requires some more algebra), and therefore there are infinitely many values of a_1 that work.

(2) In the previous proof, for each choice of point $(a_2, \ldots, a_n) \in \mathbb{A}^{n-1}$ such that $g_d(a_2, \ldots, a_n) \neq 0$, the polynomial $f(x_1, a_2, \ldots, a_n)$ is non-constant and therefore has a root since K is algebraically closed. By (1) there are infinitely many choices of $(a_2, \ldots, a_n)$ such that $g_d(a_2, \ldots, a_n) \neq 0$, so there are infinitely many points $p \in \mathbb{A}^n$ such that $f(p) = 0$. □

Spaces with the property of the proposition have a special name:

Definition 5.38. A topological space (X, τ) is *irreducible* if whenever there are closed sets Z_1, Z_2 such that $X = Z_1 \cup Z_2$, then either $X = Z_1$ or $X = Z_2$.

Corollary 5.39. *$\mathbb{A}^n_K$ with the Zariski topology is irreducible.*

Proof. Suppose $\mathbb{A}^n_K = Z_1 \cup Z_2$ for some closed sets Z_1, Z_2. Taking complements,

$$\emptyset = \mathbb{A}^n_K \setminus (Z_1 \cup Z_2) = (\mathbb{A}^n_K \setminus Z_1) \cap (\mathbb{A}^n_K \setminus Z_2).$$

Thus $\mathbb{A}^n_K \setminus Z_1$ and $\mathbb{A}^n_K \setminus Z_2$ are disjoint open sets. By the proposition, one of them has to be empty, and therefore one of Z_1 or Z_2 is all of $\mathbb{A}^n_K$. □

Irreducible topological spaces, despite having a nice name, are horribly pathological. Things like Euclidean space $\mathbb{R}^n$ are very far from being irreducible, since for example $\mathbb{R}^n$ is the union of two closed half-spaces overlapping along a line. (There are also a billion other examples to show $\mathbb{R}^n$ is not irreducible.) Intuitively, an irreducible space is a space where the (proper) closed subsets are "very small": the space is not a finite union of proper closed subsets. Correspondingly, the nonempty open sets are "very large": any two open sets are forced to overlap.

Another way to quantify the size of open sets in an irreducible space is to think about closure.

Definition 5.40. Let (X, τ) be a topological space, and let $S \subset X$ be a subset. The *closure* of S is the smallest closed set $\overline{S}$ of X that contains S. It can be shown that $\overline{S}$ is the intersection of all closed sets containing S, and therefore that the closure actually exists.

A subset $S \subset X$ is called *dense* if its closure is all of X.

Example 5.41. A closed set is its own closure.

Lemma 5.42. *Let (X, τ) be a topological space. Then $S \subset X$ is dense if and only if every nonempty open set $U \subset X$ meets S.*

Proof. ($\Rightarrow$) Suppose $S \subset X$ is dense. Let $U \subset X$ be an open set, and suppose it doesn't meet S. Then its complement $Z = X \setminus U$ is closed and contains S. Since the smallest closed set that contains S is all of X, we must have $Z = X$. Therefore U is empty.

($\Leftarrow$) Suppose every nonempty open set $U \subset X$ meets S. Let $Z \subset X$ be a closed set that contains S. Then its complement $U = X \setminus Z$ is an open set that doesn't meet S, and therefore it must be empty. Therefore $Z = X$, and the closure of S is X. □

Example 5.43. In $\mathbb{R}^2$ with the usual topology, the closure of an open ball $\{(x, y) : x^2 + y^2 < 1\}$ is the closed ball $\{(x, y) : x^2 + y^2 \leq 1\}$. The closure of the subset $\mathbb{Q}^2 \subset \mathbb{R}^2$ of points with rational coordinates is all of $\mathbb{R}^2$ (use the lemma: any open set contains an open ball, and any open ball contains points with rational coordinates). The set $\mathbb{Z}^2 \subset \mathbb{R}^2$ is already closed, since its complement is open: any point with a non-integer coordinate can be surrounded by a little ball that doesn't meet any points with integer coordinates.

Example 5.44. Many examples of closure in $\mathbb{A}^n_K$ still have the intuitive "expected" answer. For example, Let $S \subset \mathbb{A}^2$ be the set

$$V(x^2 + y^2 - 1) \setminus \{(0, 1)\},$$

a circle with a point deleted. Then the closure is the full circle $V(x^2 + y^2 - 1)$.

But, there are also pathological examples. For example, the integer lattice $\mathbb{Z}^2$ is dense in $\mathbb{A}^2_{\mathbb{R}}$: a polynomial that vanishes at every point with integer coordinates has to be the zero polynomial, so there is no nonzero f such that $V(f)$ contains $\mathbb{Z}^2$. The only closed set that contains $\mathbb{Z}^2$ is all of $\mathbb{A}^2_{\mathbb{R}}$.

Lemma 5.45. *A topological space (X, τ) is irreducible if and only if every nonempty open set is dense.*

Proof. ($\Rightarrow$) Suppose X is irreducible, and let $U \subset X$ be a nonempty open set. Let $V \subset X$ be an open set, and suppose it doesn't meet U, so $U \cap V = \emptyset$. Taking complements,

$$X = (X \setminus U) \cup (X \setminus V)$$

is a union of closed sets. Since X is irreducible, we must have $X \setminus V = X$, and therefore $V = \emptyset$. Therefore U is dense in X.

($\Leftarrow$) Suppose every nonempty open set is dense, and suppose $X = Z_1 \cup Z_2$ is a union of two proper closed sets. Then taking complements,

$$\emptyset = (X \setminus Z_1) \cap (X \setminus Z_2),$$

so $X \setminus Z_1$ and $X \setminus Z_2$ are nonempty open sets which don't intersect. This contradicts that $X \setminus Z_1$ is dense. □

Corollary 5.46. *In $\mathbb{A}^n_K$, any nonempty open set is dense.*

5.5. **The Zariski topology on an affine variety.** In the previous two sections we discussed the Zariski topology on an affine space: a closed set in $\mathbb{A}^n$ is the zero locus $V(\{f_\alpha\}_\alpha)$ of some collection of polynomials. Here we extend this notion to define a topology on any closed set $X \subset \mathbb{A}^n$.

Definition 5.47. Let $X \subset \mathbb{A}^n$ be closed. A set $Z \subset X$ is closed if it is a closed set in $\mathbb{A}^n$. Equivalently, Z is the intersection of a closed set in $\mathbb{A}^n$ with X. The Zariski topology on X is the topology with these closed sets.

If $X = V(f_1, \ldots, f_k)$ is cut out by some equations, then any closed subset $Z \subset X$ can be cut out by the defining equations of X together with some additional equations. So for example, we can write

$$Z = V(f_1, \ldots, f_k, g_1, \ldots, g_\ell).$$

Remark 5.48. This construction of a topology on a closed set is a special case of the *subspace topology* from general topology.

Example 5.49 (The Zariski topology on a set of points)**.** Let $Z \subset \mathbb{A}^n$ be a finite set of points. Then it is a closed subset of $\mathbb{A}^n$, since individual points are closed and finite unions of closed sets are closed. For the same reason, every subset $Z' \subset Z$ is closed. Equivalently, every subset of Z is open (in the topology on Z). This topology on a finite set is called the *discrete topology*.

Example 5.50 (The Zariski topology on a plane curve)**.** Let $f(x, y) \in K[x, y]$ be a nonconstant polynomial. Recall that a polynomial f is called *irreducible* if whenever we factor it as $f = gh$ then either g or h is constant. Assume K is algebraically closed and f is irreducible. Let $X = V(f)$. We saw above that there are infinitely many points p such that $f(p) = 0$, and therefore X is infinite. Here we sketch an argument that the closed sets of X are:

- All of X.
- Finite subsets of X.

(Compare with the Zariski topology on $\mathbb{A}^1$.) It is clear that these sets are all closed, so we have to show there are not any other closed subsets of X.

Let $Z \subset X$ be a proper subset. Then Z is cut out from X by at least one more equation, so we can find some $g \in K[x, y]$ such that

$$Z \subset V(f, g) \subsetneq X.$$

This gives that g is not divisible by f, since in that case $V(f, g) = V(f) = X$. The next result asserts that the system $f = g = 0$ has only finitely many solutions.

Proposition 5.51. *Let $f \in K[x, y]$ be nonconstant and irreducible, and let $g \in K[x, y]$ be a polynomial that is not divisible by g. Then the set $V(f, g)$ is finite.*

Proof. The proof needs somewhat more algebra than we have been using. See for example Shafarevich "Basic Algebraic Geometry I," the Lemma in the first section. In particular, the fact that $K[x, y]$ is a UFD is crucial, so if you are pursuing the project on UFD's this could be a payoff for your project. □

Example 5.52 (The Zariski topology on $\mathbb{A}^2$)**.** We can now discuss the structure of an arbitrary closed set $Z \subset \mathbb{A}^2$. Suppose $\emptyset \neq Z \neq \mathbb{A}^2$, and pick one equation $f \in K[x, y]$ such that $Z \subset V(f)$. Factor f into irreducible factors $f = f_1 \cdots f_k$, so that $Z \subset V(f_1) \cup \cdots \cup V(f_k)$. Then for each i, $Z \cap V(f_i)$ is a closed subset of the plane curve $V(f_i)$, so it either equals $V(f_i)$ or it is a finite (possibly empty) subset of $V(f_i)$, by the previous example. It follows that Z is one of

- a point;
- a plane curve $V(f)$, where f is irreducible;
- or a finite union of closed sets of the previous two types.

The irreducible closed sets in $\mathbb{A}^2$ are points and irreducible plane curves.

Example 5.53. In $\mathbb{A}^2$, the set

$$V(xy) \cup \{(1, 1)\} = V(x) \cup V(y) \cup \{(1, 1)\}$$

is closed.

5.6. **Polynomial mappings.** In order to compare and relate several varieties to one another, it is important to have the concept of a map between two varieties. The natural functions in algebra come from polynomials, so the natural mappings between varieties come from collections of polynomials.

Definition 5.54. A *polynomial mapping* (or *regular map*) $F : \mathbb{A}^n \to \mathbb{A}^m$ is a function of the form

$$F(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)),$$

where the components $f_i \in K[x_1, \ldots, x_n]$ are polynomials.

More generally, if $X \subset \mathbb{A}^n$ is closed, a polynomial mapping $F : X \to \mathbb{A}^m$ is a function which is the restriction of a polynomial mapping $\mathbb{A}^n \to \mathbb{A}^m$.

Even more generally, if $X \subset \mathbb{A}^n$ and $Y \subset \mathbb{A}^m$ are closed, a polynomial mapping $F : X \to Y$ is a polynomial mapping $F : X \to \mathbb{A}^m$ such that the image of F is contained in Y.

Example 5.55. Affine linear transformations $T : \mathbb{A}^n \to \mathbb{A}^m$ are polynomial mappings where all the component functions $f_1, \ldots, f_m$ have degree at most 1.

Example 5.56. There is a polynomial mapping $F : \mathbb{A}^1 \to \mathbb{A}^3$ defined by the formula

$$F(t) = (t, t^2, t^3).$$

Its image is called the *twisted cubic curve*. The image X is in fact a closed set in $\mathbb{A}^3$: on the one hand, every point in the image satisfies the equations $y - x^2 = 0$ and $z - x^3 = 0$, so

$$X \subset V(y - x^2, z - x^3).$$

On the other hand, if $(x, y, z) \in V(y - x^2, z - x^3)$, then $y = x^2$ and $z = x^3$, so

$$(x, y, z) = (x, x^2, x^3) = F(x).$$

Therefore,

$$X = V(y - x^2, z - x^3).$$

The twisted cubic gives an example of an embedding:

Definition 5.57. A polynomial mapping $F : X \to Y$ is an *isomorphism* between X and Y if there is a polynomial mapping $G : Y \to X$ such that $F \circ G = \mathrm{id}_Y$ and $G \circ F = \mathrm{id}_X$.

A polynomial mapping $F : X \to Y$ is an *embedding* of X into Y if it is an isomorphism onto its image.

Remark 5.58. Set theory shows that an isomorphism is a bijection, and an embedding is an injection. An embedding is a bijection onto its image.

Example 5.59. Continuing the previous example, to show that the twisted cubic map $F : \mathbb{A}^1 \to \mathbb{A}^3$ is an embedding of $\mathbb{A}^1$ into $\mathbb{A}^3$, we need to give a polymomial mapping $G : X \to \mathbb{A}^1$ such that $F \circ G = \mathrm{id}_X$ and $G \circ F = \mathrm{id}_{\mathbb{A}^1}$. Recall that a polynomial map $X \to \mathbb{A}^1$ has to be the restriction of a polynomial map $\mathbb{A}^3 \to \mathbb{A}^1$; therefore, we can define G

by defining its extension $\widetilde{G}$ on all of $\mathbb{A}^3$. Consider $\widetilde{G} : \mathbb{A}^3 \to \mathbb{A}^1$ defined by

$$\widetilde{G}(x, y, z) = x.$$

Clearly $\widetilde{G}$ is a polynomial mapping. For $t \in \mathbb{A}^1$, we have

$$G(F(t)) = G(t, t^2, t^3) = t,$$

and for $(t, t^2, t^3) \in \mathbb{A}^3$, we have

$$F(G(t, t^2, t^3)) = F(t) = (t, t^2, t^3).$$

Therefore F is an isomorphism between $\mathbb{A}^1$ and X, and it is an embedding of $\mathbb{A}^1$ in $\mathbb{A}^3$.

A first property of polynomial mappings is that they are continuous in the Zariski topology. We recall the basic definitions from topology.

Definition 5.60. Let X and Y be topological spaces. A function $f : X \to Y$ is *continuous* if for every open set $U \subset Y$, the preimage

$$f^{-1}(U) := \{x \in X : f(x) \in U\}$$

is open.

Continuity can also be characterized in terms of closed sets, which is preferable for working with the Zariski topology.

Lemma 5.61. *Let X and Y be topological spaces. A function $f : X \to Y$ is continuous if and only if for every closed set $Z \subset Y$, the preimage $f^{-1}(Z)$ is closed.*

Proof. Exercise. □

Proposition 5.62. *Let $F : X \to Y$ be a polynomial mapping, where $X \subset \mathbb{A}^n$ and $Y \subset \mathbb{A}^m$ are closed sets. Then F is continuous.*

Proof. In case $X = \mathbb{A}^n$ and $Y = \mathbb{A}^m$, you did this on your homework.

In the general case, let $F : X \to Y$ be a polynomial mapping, and let $G : \mathbb{A}^n \to \mathbb{A}^m$ be a polynomial mapping such that $G|_X = F$. Let $Z \subset Y$ be a closed set. Then Z is closed in $\mathbb{A}^m$, so $G^{-1}(Z)$ is closed. Then $G^{-1}(Z) \cap X$ is closed in X, and $G^{-1}(Z) \cap X = F^{-1}(Z)$. Therefore $F^{-1}(Z)$ is closed in X, and F is continuous. □

The continuous image of an irreducible space is irreducible, so the same is true for polynomial mappings. First we need to slightly extend the notion of an irreducible space to the subsets of a topological space:

Definition 5.63. Let X be a topological space and let $S \subset X$ be a subset. Then S is *irreducible* if whenever $Z_1, Z_2 \subset X$ are closed and

$$S \subset Z_1 \cup Z_2,$$

then either $S \subset Z_1$ or $S \subset Z_2$.

Remark 5.64. Alternately, we could introduce the correct notion of the *subspace topology* on $S \subset X$: a subset of S is *closed* if it is the intersection $Z \cap S$ of a closed set in X with S. Then S is irreducible in the sense of the previous definition if and only if S (with the subspace topology) is an irreducible topological space.

Lemma 5.65. *Let $f : X \to Y$ be a continuous map of topological spaces, and let $S \subset Y$ be the image of f. Suppose X is irreducible. Then S is irreducible.*

Proof. Suppose that $S \subset Z_1 \cup Z_2$, where Z_1, Z_2 are closed subsets of Y. Then

$$X = f^{-1}(S) = f^{-1}(Z_1) \cup f^{-1}(Z_2).$$

Since f is continuous, $f^{-1}(Z_1)$ and $f^{-1}(Z_2)$ are closed subsets of X. Since X is irreducible, we find that either $X = f^{-1}(Z_1)$ or $X = f^{-1}(Z_2)$. But since f maps X *onto* S, this implies that either Z_1 or Z_2 contains S. Therefore S is irreducible. □

Corollary 5.66. *Let X, Y be affine varieties and let $F : X \to Y$ be a polynomial mapping. If X is irreducible, then the image of F is irreducible.*

Some examples are called for.

Example 5.67. We know that the preimage of a closed set is closed under a polynomial mapping. We also saw the example of the twisted cubic, where the image of $\mathbb{A}^1$ in $\mathbb{A}^3$ was closed. Unfortunately, in general the image of a closed set under a polynomial mapping is not closed.

For example, consider $X = V(xy - 1) \subset \mathbb{A}^2$, and let $F : \mathbb{A}^2 \to \mathbb{A}^1$ be the projection $F(x, y) = x$. The image of F is $\mathbb{A}^1 \setminus \{0\}$. (Check!) This is clearly not closed in $\mathbb{A}^1$.

Remark 5.68. The previous example shows that even if $F : A^n \to \mathbb{A}^m$ is an affine linear transformation, then the image of a closed set is not necessarily closed. On the other hand, if $F : \mathbb{A}^n \to \mathbb{A}^n$ is an *invertible* affine transformation, then the image of a closed set is closed. Indeed, if G is the inverse of F and $Z \subset \mathbb{A}^n$ is closed, then $F(Z) = G^{-1}(Z)$ is the preimage of a closed set under the map G, so it is closed. (Compare with an earlier HW exercise.)

Example 5.69. The *graph* of a polynomial $f(x_1, \ldots, x_{n-1}) \in K[x_1, \ldots, x_{n-1}]$ is the subset X of $\mathbb{A}^n$ defined by the equation

$$x_n = f(x_1, \ldots, x_{n-1}).$$

Then X is irreducible, and in fact X is isomorphic to $\mathbb{A}^{n-1}$. (What are the inverse polynomial maps $\mathbb{A}^{n-1} \to X$ and $X \to \mathbb{A}^{n-1}$?)

6. Parameter spaces

6.1. **Parameter spaces and general properties.** One of the main features of algebraic geometry that distinguishes it from other areas of mathematics is that collections of objects in algebraic geometry can often be described within the language of algebraic geometry.

Example 6.1. In analysis, the set of all continuous (or differentiable, etc.) functions $f : \mathbb{R} \to \mathbb{R}$ is enormous, and not really a suitable space to do analysis on. The subject of *functional analysis* is an entirely additional subject introduced to study these basic objects from analysis.

Example 6.2. In algebraic geometry, a polynomial function $f : \mathbb{A}^1 \to \mathbb{A}^1$ of degree d can be recorded as a single polynomial

$$f(x) = a_d x^d + \cdots + a_0.$$

Then we can regard f as its list of coefficients $(a_d, \ldots, a_0)$, and therefore we can think of it as a point in affine space $\mathbb{A}^{d+1}$.

The general theme is that objects in algebraic geometry are specified by only a finite amount of data. If we record that data in the right way, we can capture the object as a point in some algebraic variety.

Example 6.3 (n points)**.** Consider an (ordered) list $p_1, \ldots, p_n \in \mathbb{A}^2$ of points in the plane. Each p_i has coordinates (x_i, y_i), so we can record the $2n$ coordinates of these n points to get a point in $\mathbb{A}^{2n}$. Maybe we record the point as an $n \times 2$ matrix

$$\begin{pmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix}$$

and think of the coordinates on $\mathbb{A}^{2n}$ as being the x_i and y_i.

Inside the $\mathbb{A}^{2n}$ of collections of n points, we can describe various geometric loci.

Example 6.4 (Distinct points)**.** In the $\mathbb{A}^{2n}$ of collections of n points in $\mathbb{A}^2$, consider the subset $X \subset \mathbb{A}^{2n}$ where two of the points are the same. Let us see that X is closed. First, we have $p_1 = p_2$ iff $x_1 = x_2$ and $y_1 = y_2$. So, the subvariety $V(x_1 - x_2, y_1 - y_2)$ of $\mathbb{A}^{2n}$ describes the locus where $p_1 = p_2$. Similarly, for $i \neq j$ we have a subvariety $V(x_i - x_j, y_i - y_j)$ where $p_i = p_j$. Then X is the union of these subvarieties over all $i \neq j$, and therefore X is closed.

The complement $\mathbb{A}^{2n} \setminus X$ is the locus of collections $p_1, \ldots, p_n$ of distinct points. It is open in $\mathbb{A}^{2n}$, and it is nonempty since there are obviously collections of n distinct points in $\mathbb{A}^2$. Since $\mathbb{A}^{2n}$ is irreducible,

it is also dense in $\mathbb{A}^{2n}$. We say that *in a general collection of n points in $\mathbb{A}^2$, all the points are distinct.*

The word *general* has a technical meaning that we are now ready to introduce.

Definition 6.5. Suppose that a collection of geometric objects are parameterized by an irreducible variety S. We say that the general object has some property P if there is an open dense subset U of S such that all the objects corresponding to points in U have property P.

Example 6.6. In the previous example, $S = \mathbb{A}^{2n}$ parameterizes collections of n points in $\mathbb{A}^2$. Property P is "the points are distinct." The open set U is $\mathbb{A}^{2n} \setminus X$.

Example 6.7 (Collinear triples of points)**.** Now consider $\mathbb{A}^6 = \mathbb{A}^{2\cdot 3}$, the parameter space for collections of 3 points in $\mathbb{A}^2$. We wish to describe the locus of collinear triple of points. Three points $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ lie on a line if and only if there are some $a, b, c \in K$, not all zero, such that

$$\begin{aligned} ax_1 + by_1 + c &= 0 \\ ax_2 + by_2 + c &= 0 \\ ax_3 + by_3 + c &= 0 \end{aligned}$$

In matrix form, this reads

$$\begin{pmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0.$$

Thus, the three points are collinear iff the matrix on the left has a nonzero vector in its kernel. Since it is a 3×3 matrix, this happens if and only if its determinant is nonzero. The determinant of the matrix is

$$x_1y_2 + x_2y_3 + x_3y_1 - x_1y_3 - x_2y_1 - x_3y_2,$$

so the locus in $\mathbb{A}^6$ of collinear triples of points is the subvariety

$$X = V(x_1y_2 + x_2y_3 + x_3y_1 - x_1y_3 - x_2y_1 - x_3y_2) \subset \mathbb{A}^6.$$

Its complement $U = \mathbb{A}^6 \setminus X$ is an open set in $\mathbb{A}^6$. Since there are triples of points that aren't collinear, U is dense in $\mathbb{A}^6$. Therefore, the general triple of points in $\mathbb{A}^6$ is not collinear.

Example 6.8 (m collinear points in $\mathbb{A}^2$)**.** Let $\mathbb{A}^{2m}$ be the parameter space for m points in $\mathbb{A}^2$, with $m \geq 3$. Let $Y \subset \mathbb{A}^{2m}$ be the locus of m

points that all lie on the same line. The m points all lie on the same line if and only if any 3 of the points are collinear. For $1 \leq i, j, k \leq m$, let $X_{ijk} \subset \mathbb{A}^{2m}$ be the subset where points p_i, p_j, p_k are collinear. Then by a very similar computation to the previous example, we can see that X_{ijk} is cut out by the determinant of the matrix

$$\begin{pmatrix} x_i & y_i & 1 \\ x_j & y_j & 1 \\ x_k & y_k & 1 \end{pmatrix}.$$

Therefore, X_{ijk} is closed in $\mathbb{A}^{2m}$. But now

$$Y = \bigcap_{i,j,k} X_{ijk},$$

so also Y is closed in $\mathbb{A}^{2m}$.

Example 6.9 (m points in $\mathbb{A}^2$ with at least k on a line.)**.** Let $\mathbb{A}^{2m}$ be the parameter space for m points in $\mathbb{A}^2$, with $m \geq 3$. Let $3 \leq k \leq m$, and let $Z \subset \mathbb{A}^{2m}$ be the subset of collections of m points with at least k lying on some line. For a size k subset I of $\{1, \ldots, m\}$, we can look at the locus $X_I \subset \mathbb{A}^{2m}$ of collections of m points such that the k points $\{p_i : i \in I\}$ are all collinear. Then equations for X_I can be determined by the method in the previous example. Therefore, X_I is closed. We have

$$Z = \bigcup_I X_I,$$

so Z is also closed.

Example 6.10. A line in $\mathbb{A}^2$ is given by an equation

$$ax + by + c = 0$$

with a, b not both zero. We can record it as a point in $\mathbb{A}^3$ (with coordinates a, b, c) by remembering the coefficients (a, b, c). Then for any point in the open subset $\mathbb{A}^3 \backslash V(a, b)$, we can form the line $ax+by+c = 0$.

This correspondence between lines in $\mathbb{A}^2$ and points in $\mathbb{A}^3 \setminus V(a, b)$ isn't totally perfect: if we scale the equation of a line it still describes the same line. So, each line $ax + by + c = 0$ in $\mathbb{A}^2$ is represented by all the points of the form $(\lambda a, \lambda b, \lambda c)$ in $\mathbb{A}^3$ (here $0 \neq \lambda \in K$).

The best solution to this problem is to introduce the concept of projective space. At the level of set theory, we can define

$$\mathbb{P}^2 = (\mathbb{A}^3 \setminus \{0\})/\sim,$$

where two points $(a, b, c), (a', b', c')$ in $\mathbb{A}^3 \backslash \{0\}$ are regarded as *equivalent* if there is some $0 \neq \lambda \in K$ such that $\lambda(a, b, c) = (a', b', c')$. Then $\mathbb{P}^2$ is the set of equivalence classes of points in $\mathbb{A}^3 \backslash \{0\}$ under this equivalence

relation. We then have a bijective correspondence between lines in $\mathbb{A}^2$ and the points of $\mathbb{P}^2 \setminus \{(0,0,1)\}$.

The set $\mathbb{P}^2$ is not an algebraic variety in the sense we have been talking about; there is no way to think of it as a closed set in an affine space. But, there is a mild generalization of the concept of variety where $\mathbb{P}^2$ becomes an algebraic variety.

Even though the correspondence between lines in $\mathbb{A}^2$ and points in $\mathbb{A}^2 \setminus V(a,b)$ is not perfect, it is still good enough to work with and doesn't really cause any problems. So, rather than develop the theory of projective spaces we will stick with this imperfect correspondence.

Example 6.11 (Affine linear transformations)**.** Consider the set of affine linear transformations

$$S = \{T : \mathbb{A}^n \to \mathbb{A}^n | T \text{ is affine linear}\}.$$

Such a transformation can be written uniquely in the form

$$T\mathbf{x} = A\mathbf{x} + \mathbf{b},$$

where $A \in \mathrm{Mat}_{n\times n}(K)$ and $\mathbf{b} \in K^n$, so we can record it by a point in $\mathbb{A}^{n^2+n}$ that consists of all these numbers.

Example 6.12 (Invertible affine linear transformations)**.** An affine linear transformation is invertible if and only if the matrix A is invertible if and only if $\det A \neq 0$. The determinant of A is a polynomial in the entries of A, so the open set $\mathbb{A}^{n^2+n} \setminus V(\det(A))$ parameterizes the invertible affine linear transformations. It is nonempty, so it is dense, and the general affine linear transformation is invertible.

Often times when we show that a general object has some property P, it is actually the case that the locus where property P does not hold is closed. This has been the case in all the previous examples. In such a case, we can show that the general object has property P in two steps:

(1) Show that the locus where property P holds is open.
(2) Show that there actually is some object that satisfies property P.

Both steps are critical, and often times the second step is the more interesting one! (Although it hasn't been in the previous examples.)

The above outlined strategy works very often, but there are some examples where you have to be more careful.

Exercise 6.13. A *fixed point* of an affine linear transformation $T : \mathbb{A}^n \to \mathbb{A}^n$ is a point $\mathbf{x} \in \mathbb{A}^n$ such that $T\mathbf{x} = \mathbf{x}$. The general affine linear transformation has a unique fixed point. The complement consists of

affine linear transformations that either have no fixed points or have more than one fixed point, and this is **not** closed.

Example 6.14. Let $\mathbb{A}^2$ be the parameter space for a single point in $\mathbb{A}^2$ (yes this is as dumb as it sounds). Let $S \subset \mathbb{A}^2$ be the subset

$$S = \{(x, y) : x \neq 0 \text{ or } x = y = 0\}.$$

The set S is neither open nor closed. Let property P be the property that $(x, y) \in S$. Then the locus where property P holds is not open; however, there is an open dense subset of $\mathbb{A}^2$ where property P holds (namely, property P holds on $\mathbb{A}^2 \setminus V(x)$). Therefore, the general point in $\mathbb{A}^2$ has property P, even though the set of points in $\mathbb{A}^2$ with property P is not open.

6.2. **Spaces of matrices.** Spaces of matrices are some of the most important examples of parameter spaces. We clearly have a bijection between $\mathrm{Mat}_{m\times n}(K)$ and $\mathbb{A}^{mn}_K$, given by writing down all the entries of an $m \times n$ matrix A as a vector of length mn. Instead of constantly making this identification, we can just think of $\mathrm{Mat}_{m\times n}(K)$ as actually being an affine space. We can then discuss the Zariski topology on $\mathrm{Mat}_{m\times n}(K)$, closed sets, open sets, general properties of matrices, and so on.

When $m = n$ and we are considering square matrices, there is an open subset

$$\mathrm{GL}_n(K) \subset \mathrm{Mat}_{n\times n}(K)$$

that consists of the invertible matrices. The fact that this is actually open comes from the fact that non-invertibility is characterized by the equation $\det(A) = 0$, so that

$$\mathrm{GL}_n(K) = \mathrm{Mat}_{n\times n}(K) \setminus V(\det).$$

Let us recall some of the basic properties of the determinant. The $n \times n$ determinant is a function

$$\det : \mathrm{Mat}_{n\times n}(A) \to K$$

such that

(1) If two rows/columns of A are swapped, the sign of the determinant changes.
(2) If a row/column of A is scaled, the determinant is scaled by the same factor.
(3) If one row/column of A is added to another, the determinant is unchanged.
(4) The determinant of the identity matrix I is 1.

The first main theorem on determinants is that they actually exist:

Theorem 6.15. *There is a unique function* $\det : \mathrm{Mat}_{n\times n}(A) \to K$ *satisfying the above four properties. It is given by*

$$\det A = \sum_{\sigma \in S_n} \mathrm{sign}(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}.$$

Proof. First check that the formula satisfies the properties. Next, show that the properties determine the value of $\det A$ for any A, and therefore there is at most one such determinant function. □

Another key consequence of the properties of the determinant is the following.

Theorem 6.16. *We have* $\det A = 0$ *if and only if* A *is not invertible.*

Proof. From the properties of the determinant, if A and B are related by row operations then $\det A = 0$ iff $\det B = 0$ and A is invertible iff B is invertible. So, we can replace A by a reduced row echelon matrix B. If B is not invertible then it has a row of zeroes, and in this case $\det B = 0$ since scaling the row of zeroes is supposed to scale the determinant, but the matrix is unchanged by scaling that row. On the other hand if B is invertible then $B = I$ and $\det B = 1 \neq 0$. □

In particular, notice that the determinant is in fact a polynomial in the entries of the matrix A, and therefore $V(\det)$ is a closed subset of $\mathrm{Mat}_{n\times n}(K)$.

More generally, we want to consider the loci in $\mathrm{Mat}_{m\times n}(K)$ given by various conditions on the rank of the matrix. Our main goal in this section is to prove the following theorem:

Theorem 6.17. *Let* $M_k \subset \mathrm{Mat}_{m\times n}(K)$ *be the locus of matrices of rank at most* k*:*

$$M_k = \{A \in \mathrm{Mat}_{m\times n}(K) : \mathrm{rk}(A) \leq k\}.$$

Then M_k *is closed.*

The subvariety M_k is often called a *determinantal variety*.

Remark 6.18. You might wonder why we don't instead consider a locus like

$$\{A \in \mathrm{Mat}_{m\times n}(K) : \mathrm{rk}(A) = k\}.$$

In general these sets are neither open nor closed. If $k \leq \min\{m, n\}$ (the maximal rank of an $m \times n$ matrix) then the closure of this set is M_k from the theorem.

To prove the theorem, we may as well assume $k \leq \min\{m, n\}$, since otherwise $M_k = \mathrm{Mat}_{m\times n}(K)$ and there is nothing to prove. The key to proving the theorem is to introduce the concept of the *minors* of a matrix.

Definition 6.19. Given an $m \times n$ matrix A and an integer $k \leq \min\{m, n\}$, we can form a $k \times k$ submatrix of A by picking any k rows and any k columns. The determinant of such a $k \times k$ submatrix is a $k \times k$ *minor* of A.

Clearly each $k \times k$ minor of A is a polynomial of degree k in the entries of A.

Example 6.20. The 2×2 minors of the matrix

$$\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$$

are $ae - bd$, $af - cd$, and $bf - ce$. The first minor is zero if and only if the first two columns are multiples of each other. The second minor is zero if and only if the first and third columns are multiples of each other. The third minor is zero if and only if the second and third columns are multiples of each other. All three minors are zero if and only if all three columns are multiples of each other, if and only if the columns span an at most 1-dimensional space, if and only if the rank of the matrix is at most 1.

Theorem 6.21. *An $m \times n$ matrix A has rank less than k if and only if every $k \times k$ minor of A is zero.*

Proof. ($\Rightarrow$) First suppose matrix A has rank less than k. This means that the image of the linear transformation $T : K^n \to K^m$ associated to A has dimension less than k. In terms of the columns of A, this says that any k columns of A are linearly dependent. If we form any $k \times k$ submatrix B of A, then the columns of that matrix are also linearly dependent. Therefore, B is not invertible and $\det B = 0$. Therefore, every $k \times k$ minor of A is zero.

($\Leftarrow$) Let us suppose matrix A has rank at least k, and show that there is some $k \times k$ minor of A that is not zero. Since A has rank at least k, the columns of A span a subspace of K^m of dimension at least k. Throw out columns that depend on other columns to get a basis of the image. If it has dimension greater than k, throw out some more columns until we have exactly k independent columns. In this way we get an $m \times k$ submatrix B of A, and it has rank k. Notice that every $k \times k$ minor of B is also a $k \times k$ minor of A. Therefore, it is enough to show that B has a nonzero $k \times k$ minor.

Since B has rank k, we can perform elementary column operations to it to reduce it to column echelon form B'. Since B has k columns, to specify a $k \times k$ submatrix of B we just have to specify k rows. Consider the $k \times k$ submatrix C of B given by the rows where B' has pivots,

and let C' be the corresponding $k \times k$ submatrix of B'. Performing column operations on B changed the determinant of C in a predictable way: its sign changed if we swapped columns, it was scaled if we scaled columns, and it was unaffected if we added one column to another. (It is crucial here that C is as "wide" as B: if C had fewer columns than B, then adding a column outside of C to a column inside C would change the determinant unpredictably. This is why we reduced the $m \times k$ case in the previous paragraph.) It follows that $\det C$ is nonzero if and only if $\det C'$ is nonzero. But, the matrix C' is in column echelon form and has a pivot in every column. Therefore C' is invertible, $\det C' \neq 0$, and $\det C \neq 0$. We conclude that B has a nonzero $k \times k$ minor, and therefore A also has a nonzero $k \times k$ minor. □

Corollary 6.22. *If $k < \min\{m, n\}$, then the set $M_k \subset \mathrm{Mat}_{m\times n}(K)$ of matrices of rank at most k is closed. If the coordinates on $\mathrm{Mat}_{m\times n}(K)$ are a_{ij}, then M_k is cut out by the $(k+1)\times(k+1)$ minors of the matrix (a_{ij}).*

Note that we have a chain of containments

$$\{0\} = M_0 \subset M_1 \subset M_2 \subset \cdots \subset M_{\min\{m,n\}} = \mathrm{Mat}_{m\times n}(K).$$

Each M_i is closed inside of M_{i+1}.

Remark 6.23. In fact, each subvariety $M_k \subset \mathrm{Mat}_{m\times n}(K)$ is irreducible, althgouh we can't prove that here. Each containment $M_k \subsetneq M_{k+1}$ is actually proper if k is in the right range (exhibit a matrix of rank $k+1$ that isn't of rank k), so the complement $M_{k+1} \setminus M_k$ is an open dense subset of M_{k+1}. It consists of the matrices of rank exactly $k+1$.

(You can also show that the closure of $M_{k+1} \setminus M_k$ is M_{k+1} by more elementary means; you'll (try to) do this on your homework.)

Theorem 6.24. *If $k \leq \min\{m, n\}$, then the closure of $M_k \setminus M_{k-1}$ is M_{k-1}.*

Proof. Let $Y \subset \mathrm{Mat}_{m\times n}(K)$ be the closure of $M_k \setminus M_{k-1}$. Since M_k is closed, it is clear that $M_k \setminus M_{k-1} \subset Y \subset M_k$. Conversely, let $A \in M_{k-1}$; we have to show that $A \in Y$. It will be enough to produce a polynomial map $F : \mathbb{A}^1 \to \mathrm{Mat}_{m\times n}(K)$ such that $F(0) = A$ and $F(t) \in M_k \setminus M_{k-1}$ for all $t \neq 0$. Indeed, then $F^{-1}(M_k \setminus M_{k-1})$ is a nonempty (hence dense) open set in $\mathbb{A}^1$. Then $F^{-1}(Y)$ is a closed set that contains $F^{-1}(M_k \setminus M_{k-1})$, so it must be all of $\mathbb{A}^1$. Therefore, $F(0) = A$ is in Y.

To produce such a map F, let $r = \mathrm{rk}\, A$. Without loss of generality, assume the first r columns of A are independent, and write the column

vectors of A as

$$\mathbf{v}_1, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n$$

Let $V = \operatorname{Im} A = \operatorname{span}(\mathbf{v}_1, \dots, \mathbf{v}_r) \subset K^m$, so V has dimension r. We can find a subspace $U \subset K^m$ of dimension $k-r$ such that $U \cap V = \{0\}$, and then the subspace $U + V \subset K^m$ has dimension k. Let $\mathbf{u}_{r+1}, \dots, \mathbf{u}_k$ be a basis of U. Finally, let B be the $m \times n$ matrix with columns

$$0, \dots, 0, \mathbf{u}_{r+1}, \dots, \mathbf{u}_k, 0, \dots, 0$$

Now consider the map $F(t) = A + tB$. Then $F(0) = A$. For time $t \neq 0$, we compute the rank of $F(t)$ by using column operations. The matrix has columns

$$\mathbf{v}_1, \dots, \mathbf{v}_r, \mathbf{v}_{r+1} + t\mathbf{u}_{r+1}, \dots, \mathbf{v}_k + t\mathbf{u}_r, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n.$$

Since the matrix A has rank r, we can use column operations on A to reduce it to the matrix with columns

$$\mathbf{v}_1, \dots, \mathbf{v}_r, 0, \dots, 0, 0, \dots, 0;$$

we only have to use the first r columns to clear out all the other columns, since all the other columns are linear combinations of the first r columns. Performing the same operations on $A + tB$, we get a matrix with columns

$$\mathbf{v}_1, \dots, \mathbf{v}_r, t\mathbf{u}_{r+1}, \dots, t\mathbf{u}_k, 0, \dots, 0.$$

Scaling the $\mathbf{u}$ columns we get a matrix with columns

$$\mathbf{v}_1, \dots, \mathbf{v}_r, \mathbf{u}_{r+1}, \dots, \mathbf{u}_k, 0, \dots, 0.$$

The first k columns here are a basis of the the space $U+V$, and therefore the rank of the matrix is k. Therefore $F(t) \in M_k \setminus M_{k-1}$ for all times $t \neq 0$. We conclude that $A \in Y$. □

6.3. **The semicontinuity principle.** A common situation in algebraic geometry is that to each point in a parameter space we can associate some matrix. We typically care about the rank of these matrices. Our study of the determinantal varieties $M_k \subset \operatorname{Mat}_{m\times n}(K)$ shows that the rank has to vary in a very particular way.

Definition 6.25. Let $X \subset \mathbb{A}^N$ be closed. A *matrix-valued polynomial map*

$$F : X \to \operatorname{Mat}_{m\times n}(K)$$

is just a polynomial map $F : X \to \mathbb{A}^{mn}$, where we think of $\mathbb{A}^{mn}$ as being the space $\operatorname{Mat}_{m\times n}(K)$.

Equivalently, in the case of $X = \mathbb{A}^N$, a matrix-valued polynomial map $F : \mathbb{A}^N \to \mathrm{Mat}_{m\times n}(K)$ is a map of the form

$$F(x_1,\dots,x_N) = \begin{pmatrix} f_{11}(x_1,\dots,x_N) & \cdots & f_{1n}(x_1,\dots,x_N) \\ \vdots & \ddots & \vdots \\ f_{m1}(x_1,\dots,x_N) & \cdots & f_{mn}(x_1,\dots,x_N) \end{pmatrix}$$

In general, a matrix-valued polynomial map $F : X \to \mathrm{Mat}_{m\times n}(K)$ is the restriction of a matrix-valued polynomial map $\mathbb{A}^N \to \mathrm{Mat}_{m\times n}(K)$.

Example 6.26. Let $\mathbb{A}^n = \mathbb{A}^{1\cdot n}$ be the parameter space for n points in $\mathbb{A}^1$. To a point $(x_1,\dots,x_n) \in \mathbb{A}^n$, we associate the $(n+1)\times(n+1)$ Vandermonde matrix

$$F(x_1,\dots,x_n) = \begin{pmatrix} x_1^n & x_1^{n-1} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ x_n^n & x_n^{n-1} & \cdots & 1 \end{pmatrix}.$$

Then $F(x_1,\dots,x_n)$ has rank $n+1$ if and only if Lagrangian interpolation $f(x_i) = y_i$ has a unique solution for any choice of $y_i \in K$; this is the case if and only if the points $x_1,\dots,x_n$ are all distinct.

In terms of the determinantal varieties M_k, we find that $F^{-1}(M_n)$ is the locus of n points $x_1,\dots,x_n$ with at least one repetition. Then $F^{-1}(M_{n+1} \setminus M_n)$ is the locus of distinct n-uples of points.

Example 6.27. Let $\mathbb{A}^{12} = \mathbb{A}^{2\cdot 6}$ be the parameter space for 6 points in $\mathbb{A}^2$. To a point $((x_1,y_1),\dots,(x_6,y_6)) \in \mathbb{A}^{12}$, we associate the matrix

$$F((x_1,y_1),\dots,(x_6,y_6)) = \begin{pmatrix} x_1^2 & x_1y_1 & y_1^2 & x_1 & y_1 & 1 \\ x_2^2 & x_2y_2 & y_2^2 & x_2 & y_2 & 1 \\ x_3^2 & x_3y_3 & y_3^2 & x_3 & y_3 & 1 \\ x_4^2 & x_4y_4 & y_4^2 & x_4 & y_4 & 1 \\ x_5^2 & x_5y_5 & y_5^2 & x_5 & y_5 & 1 \\ x_6^2 & x_6y_6 & y_6^2 & x_6 & y_6 & 1 \end{pmatrix}.$$

This assignment defines a matrix-valued polynomial map $F : \mathbb{A}^{12} \to \mathrm{Mat}_{6\times 6}(K)$. The rank of this matrix is 6 if and only if there is no nonzero conic that passes through all 6 points: if the rank is less than 6 then there is a vector (a,b,c,d,e,f) in the kernel, and then all 6 points satisfy the equation

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

The rank of the matrix is 5 if the six points lie on a unique conic (up to scale).

The rank of the matrix is 4 if the six points lie on two independent conics. The only way for this to happen is if five of the six points lie on a line.

The rank of the matrix is 3 if the six points lie on three independent conics. The only way for this to happen is if all six points lie on a line.

In each example, we observe that the rank of the matrix is lower when the points are more and more "special;" the rank is as high as possible when the points are general. This is a general fact that follows from our study of determinantal varieties.

Theorem 6.28. *Let $F : X \to \mathrm{Mat}_{m\times n}(K)$ be a matrix-valued polynomial map, and suppose that X is irreducible. Let*

$$r_0 = \max_{x\in X} \mathrm{rk}(F(x)).$$

(1) The set

$$F^{-1}(M_{r_0} \setminus M_{r_0-1}) = \{x \in X : \mathrm{rk}(F(x)) = r_0\} \subset X$$

is an open dense subset of X.

(2) For each k, the set

$$F^{-1}(M_k) = \{x \in X : \mathrm{rk}(F(x)) \leq k\} \subset X$$

is closed in X.

(3) In particular, we have a chain of closed subvarieties of X:

$$F^{-1}(M_0) \subset F^{-1}(M_1) \subset \cdots \subset F^{-1}(M_{r_0}) = X.$$

Proof. (2) The map F is continuous and $M_k \subset \mathrm{Mat}_{m\times n}(K)$ is closed, so $F^{-1}(M_k)$ is closed.

(1) The complement of $F^{-1}(M_{r_0} \setminus M_{r_0-1})$ is $F^{-1}(M_{r_0-1})$, which is closed. Therefore $F^{-1}(M_{r_0} \setminus M_{r_0-1})$ is open. It is nonempty by our choice of r_0, and therefore it is dense since X is irreducible.

(3) Take preimages of the chain of subvarieties

$$M_0 \subset M_1 \subset \cdots \subset M_{r_0}$$

of $\mathrm{Mat}_{m\times n}(K)$. □

Remark 6.29. Note that in (3) the inclusions do not all have to be strict. For example, it is possible that $F(x)$ has rank 2 for all $x \in X$, and $F^{-1}(M_0)$ and $F^{-1}(M_1)$ are both empty.

The most important part of the theorem is part (2). This can be rephrased as saying that the function $x \mapsto \mathrm{rk}\, F(x)$ on X is *lower-semicontinuous* on X.

Definition 6.30. Let X be a topological space. A function $\lambda : X \to \mathbb{Z}$ is called *lower-semicontinuous* if for every $k \in \mathbb{Z}$, the set $\lambda^{-1}((-\infty, t])$ is closed in X.

A function $\lambda : X \to \mathbb{Z}$ is called *upper-semicontinuous* if for every $k \in \mathbb{Z}$, the set $\lambda^{-1}([k, \infty))$ is closed in X.

Remark 6.31. Observe that a function $\lambda : X \to \mathbb{Z}$ is continuous if and only if it is lower- and upper-semicontinuous. In this case, for every integer k the set $\lambda^{-1}(\{k\})$ is both open and closed; this means that it is a *connected component* of X (if it is nonempty).

Intuitively, a lower-semicontinuous function $\lambda : X \to \mathbb{Z}$ can decrease at special $x \in X$. An upper-semicontinuous function can increase at special $x \in X$. A lower-semicontinuous function takes its largest value at a general point $x \in X$. An upper-semicontinuous function takes its smallest value at a general point $x \in X$.

Exercise 6.32. Let $F : X \to \mathrm{Mat}_{m \times n}(K)$ be a matrix-valued polynomial map. Show that the function $\lambda : X \to \mathbb{Z}$ defined by $\lambda(x) = \dim \ker F(x)$ is upper-semicontinuous.

6.4. **Specialization.** When we try to show that a general geometric object has some property, recall that there are often two steps:

(1) Show that the locus where the property holds is open. (This isn't always true, but it is true often enough to be useful.)
(2) Show that the locus where the property holds is nonempty.

The second step is often the harder one. The technique of *specialization* often helps us show that some such object exists. The idea is that it can be hard to show that a completely random object satisfies the property (even though this must be true for a property that holds generally). The property can be much easier to check for an object that is special in some way.

Example 6.33. Let us show that the general collection of 10 points in $\mathbb{A}^2$ does not lie on a cubic curve.

Notice that the space of cubics on $\mathbb{A}^2$ has basis $x^3, x^2y, \ldots, 1$ of length 10. Then for any collection $(x_1, y_1), \ldots, (x_{10}, y_{10})$, we can set up a 10×10 matrix such that the kernel of the matrix consists of the vectors of coefficients of cubics that pass through the 10 points. Therefore, the determinant of this matrix (a polynomial in the 20 variables of the parameter space $\mathbb{A}^{2 \cdot 10}$) cuts out the locus of 10 points that lie on at least one cubic. Its complement, the locus of 10 points that don't lie on a cubic, is therefore open. We need to show that this locus is nonempty; equivalently we need to give 10 points that don't lie on a cubic.

Suppose we vary (at least some of) our points p_i with a parameter t. You can think of this variation as being given by a map $P : \mathbb{A}^1 \to \mathbb{A}^{20}$, so $P(t) = (p_1(t), \ldots, p_{10}(t))$. If for some time t our collection of points $P(t)$ does not lie on a cubic, then by semicontinuity we know that for a general time t our collection of points does not lie on a cubic.

Now imagine that we vary our points in the following way: starting from a random configuration of points, move 4 of them onto a line. We can think of this as giving a map $P : \mathbb{A}^1 \to \mathbb{A}^{20}$ such that $P(0)$ has four collinear points and $P(t)$ for $t \neq 0$ is somewhat arbitrary. Any cubic that contains the points $P(0)$ must contain the line; therefore, $P(0)$ is not on a cubic if and only if the 6 points off the line are not on a conic.

The remaining 6 points are still unspecified, so let us further specialize them by moving three of them onto a line. Any conic that contains the 6 points must then contain that line. If the 10 points lie on a cubic, then residually the remaining 3 points must lie on a line. But those 3 points are still not fixed, so we may as well take them to not lie on a line. Thus no cubic exists which contains these points.

Remark 6.34. The previous argument actually shows the following: if we take three different lines L_1, L_2, L_3 and put 4 points on L_1 (avoid the points of intersection of the lines), 3 points on L_2, 2 points on L_1, and one point off all three lines, then there is no cubic that contains all 10 points. We could have just shown that directly instead of thinking of the points as "specializing" onto the lines, but the specialization procedure is a good method for "guessing" what a convenient configuration of points might look like.

There is a delicate balance between trying to take a special enough geometric object that it becomes easy to prove a property holds, and not taking too special an object so that the property actually becomes false.

Example 6.35. Given 10 points in $\mathbb{A}^2$ such that some 5 of the points lie on the line, there *is* a cubic that passes through them: take a line through the 5 points and multiply it by the equation of a conic through the 5 points. Thus the previous example could not have been proven by placing more than 4 points on a line.

Let us revisit the result that for any integer $n \geq 1$ there exists a collection $Z \subset \mathbb{A}^2$ of k distinct points in boring position. Recall that this means that the evaluation map

$$T_{Z,d} : V_d \to K^n$$

has maximal rank for every d. Let us let D be the smallest number such that

$$\binom{D+2}{2} = \dim V_D > n.$$

Then in particular we are supposed to show that $T_{Z,d}$ is injective for $d < D$ and surjective for $d \geq D$.

Proposition 6.36. *Suppose $d < D$. Then for a general collection $Z \subset \mathbb{A}^2$ of n distinct points, the evaluation map $T_{Z,d}$ is injective.*

Proof. First, in the parameter space $\mathbb{A}^{2n}$ of n points in $\mathbb{A}^2$, there is an open subset where the evaluation map $T_{Z,d}$ is injective. Indeed, we can write down a map

$$F : \mathbb{A}^{2n} \to \mathrm{Mat}_{n \times \binom{d+2}{2}}(K)$$

where row i consists of all the degree $\leq d$ monomials in x, y evaluated at the ith point $p_i = (x_i, y_i)$. Then $F(Z)$ is just the matrix of the evaluation map $T_{Z,d}$, written down with respect to the standard bases of V_d and K^n. Therefore $T_{Z,d}$ is not injective if and only if $F(Z)$ has rank $\leq \binom{d+2}{2} - 1$. The locus of Z such that $T_{Z,d}$ is not injective is $F^{-1}(M_{\binom{d+2}{2}-1})$, which is closed.

Next we have to show that the locus of Z such that $T_{Z,d}$ is injective is actually nonempty. We did this in the Hilbert scheme chapter, but here is another argument motivated by specialization.

Pick d lines $L_1, \ldots, L_d$. Since

$$n > \binom{d+2}{2} = (d+1) + d + (d-1) + \cdots + 1,$$

we can specialize $i+1$ points onto line L_i, and we will still have some additional points left over that don't lie on any of the lines. Then consider the kernel of $T_{Z,d}$, which consists of degree d polynomials vanishing on all the points. Any such polynomial f has to vanish on L_d, since L_d contains $d+1$ points. Residually, the degree $d-1$ polynomial that is left over has to vanish on L_{d-1} since it contains d points, and so on. It follows that f is a scalar multiple of $L_1 \cdots L_d$. But then f can't still vanish at the additional points that aren't on the lines. Therefore $T_{Z,d}$ is injective. □

Proposition 6.37. *Suppose $d \geq D$. Then for a general collection $Z \subset \mathbb{A}^2$ of n distinct points, the evaluation map $T_{Z,d}$ is surjective.*

Proof. The fact that the locus where $T_{Z,d}$ is surjective is open is proved in the same way as in the first paragraph of the previous result.

Let's show $T_{Z,d}$ is surjective for general Z by induction on n. If $n < d+1$ then we always know $h_Z(d) = n$, so we may as well assume $n \geq d+1$. Specialize $d+1$ of the points onto a line $L = 0$, and let $Z' \subset Z$ be the $n' = n - d - 1$ remaining points. Then everything in the kernel of $T_{Z,d}$ looks like $L \cdot f$ where $f \in \ker T_{Z',d-1}$. Therefore,

$$\dim \ker T_{Z,d} = \dim \ker T_{Z',d-1}.$$

Furthermore, from the facts that

$$\binom{d+1}{2} = \binom{d+2}{2} - (d+1)$$

and

$$n \leq \binom{d+2}{2}$$

and $n' = n - d - 1$ we deduce that

$$n' \leq \binom{d+1}{2}.$$

Therefore by induction, $T_{Z',d-1}$ is surjective (the remaining points Z' are general), and

$$\dim \ker T_{Z',d-1} = \binom{d+1}{2} - n'.$$

But then

$$\dim \ker T_{Z,d} = \binom{d+1}{2} - n' = \binom{d+1}{2} - n + (d+1) = \binom{d+2}{2} - n,$$

and by rank-nullity the rank of $T_{Z,d}$ is n. Therefore $T_{Z,d}$ is surjective. □

Combining the previous two results proves the following statement.

Corollary 6.38. *Let $Z \subset \mathbb{A}^2$ be a general collection of n distinct points. Then the evaluation map $T_{Z,d}$ has maximal rank. Therefore,*

$$h_Z(d) = \min\left\{\binom{d+2}{2}, n\right\}$$

Corollary 6.39. *Let $Z \subset \mathbb{A}^2$ be a general collection of n distinct points. Then Z is in boring position: the evaluation map $T_{Z,d}$ has maximal rank for every d.*

Proof. Let $\mathbb{A}^{2n}$ be the parameter space for n points, and let $V \subset \mathbb{A}^{2n}$ be the open subset parameterizing distinct collections of points. For each d, there is a dense open set $U_d \subset V$ parameterizing distinct collections of points Z such that $T_{Z,d}$ has maximal rank. Let $U = \bigcap_{d \geq 0} U_d$; then U parameterizes distinct collections of points in boring position. A priori this is an infinite intersection of open sets, but actually we know $U_d = V$ for $d \geq n+1$. Therefore only finitely many of the terms in the intersection contribute, and it is actually a finite intersection of open sets. Therefore U is open and dense. $\square$

7. The Alexander-Hirschowitz theorem

7.1. **Double points.** We end the course by generalizing our discussion of interpolation for points in the plane to the case of "fat points." This is the natural generalization of the single variable Lagrangian interpolation problem with derivative data to the several variable case.

Remark 7.1. You showed on your homework that if $F_t(x, y) \in K[x, y, t]$ is a polynomial such that

$$F_t(0, 0) = F_t(t, 0) = F_t(0, t) = 0$$

for all t, then the "limit" polynomial $F_0(x, y)$ satisfies

$$F_0(0, 0) = \frac{\partial F_t}{\partial x}(0, 0) = \frac{\partial F_t}{\partial y}(0, 0) = 0.$$

We say that a polynomial $F(x, y) \in K[x, y]$ (or the curve $X = V(F)$) has a *double point* at $p = (x_0, y_0) \in \mathbb{A}^2$ if

$$F(p) = \frac{\partial F}{\partial x}(p) = \frac{\partial F}{\partial y}(p) = 0.$$

Thus in the remark, the limiting polynomial $F_0(x, y)$ has a double point at the origin.

Remark 7.2. Let $F \in K[x, y]$ and let $X = V(F) \subset \mathbb{A}^2$ be the plane curve defined by F. If $p \in X$ and

$$\nabla F(p) := \left(\frac{\partial F}{\partial x}(p), \frac{\partial F}{\partial y}(p) \right) \neq 0,$$

then the gradient vector $\nabla F(p)$ is a normal vector to the curve at p. A vector orthogonal to it is a tangent vector to the curve at p. In this case the curve X is *smooth* or *nonsingular* at p. So, the polynomial F has a double point at p if and only if the curve X is singular at p.

Exercise 7.3. The line $X = V(y)$ is smooth at each of its points.

The simplest source of polynomials double at a point is to take products.

Lemma 7.4. *Suppose $F, G \in K[x, y]$, and let $p \in \mathbb{A}^2$. Then the product FG has a double point at p if and only if one of the following occurs:*

(1) We have $F(p) = G(p) = 0$.
(2) Either F or G has a double point at p.

Proof. ($\Leftarrow$) In either case it is clear that $(FG)(p) = 0$. Compute the derivatives with the product rule:

$$\frac{\partial (FG)}{\partial x}(p) = \frac{\partial F}{\partial x}(p) G(p) + F(p) \frac{\partial G}{\partial x}(p).$$

Clearly this vanishes in either case (1) or (2).

($\Rightarrow$) Suppose FG has a double point at p. Since $(FG)(p) = 0$, clearly either $F(p) = 0$ or $G(p) = 0$. If both hold, we are done, so without loss of generality suppose $F(p) = 0$ and $G(p) \neq 0$. Then we compute the partial derivatives

$$0 = \frac{\partial(FG)}{\partial x}(p) = \frac{\partial F}{\partial x}(p)G(p) + F(p)\frac{\partial G}{\partial x}(p) = \frac{\partial F}{\partial x}(p)G(p)$$
$$0 = \frac{\partial(FG)}{\partial x}(p) = \frac{\partial F}{\partial y}(p)G(p) + F(p)\frac{\partial G}{\partial y}(p) = \frac{\partial F}{\partial y}(p)G(p).$$

Since $G(p) \neq 0$, this is only possible if $\nabla F(p) = 0$. □

This result is often used in the following form.

Corollary 7.5. *Let $X = V(F)$ and suppose $p \in X$ is a smooth point. If FG has a double point at p, then $G(p) = 0$.*

7.2. **Double point interpolation.** Given a collection of points $p_1, \dots, p_k \in \mathbb{A}^2$, we can study the vector space W of polynomials of degree at most d that are double at $p_1, \dots, p_k$. If we define an evaluation map

$$T_{2Z,d} : V_d \to K^{3k}$$

by the rule

$$T_{2Z,d}(F) = \left(F(p_1), \frac{\partial F}{\partial x}(p_1), \frac{\partial F}{\partial y}(p_1), \dots, F(p_k), \frac{\partial F}{\partial x}(p_k), \frac{\partial F}{\partial y}(p_k)\right),$$

then this space is exactly the kernel

$$W = \ker T_{2Z,d}.$$

As usual, we are particularly interested in the dimension of W. From that information, we can deduce if $T_{2Z,d}$ is injective or surjective, and more generally we can compute $\operatorname{rk} T_{2Z,d}$.

In contrast with the case of "simple" points, even if the points $p_1, \dots, p_k$ are general it may happen that $T_{2Z,d}$ does not have maximal rank!

Example 7.6. Let $Z = \{p, q\} \in \mathbb{A}^2$ be two general points. Then there is a line $L = 0$ that passes through them. The square L^2 of this equation is double at both p and q. Therefore

$$T_{2Z,2} : V_2 \to K^6$$

is not injective, even though both V_2 and K^6 have dimension 6.

Example 7.7. Let $Z = \{p_1, \dots, p_5\} \in \mathbb{A}^2$ be five general points. Then there is a conic $F = 0$ that passes through them. The square F^2 of this equation is double at all five points. Therefore

$$T_{2z,4} : V_4 \to K^{15}$$

is not injective, even though both V_4 and K^{15} have dimension 15.

Slightly more generally, given two sets of points $p_1, \dots, p_k$ and $q_1, \dots, q_\ell$ in $\mathbb{A}^2$, we can study the vector space W of polynomials of degree at most d that are double at $p_1, \dots, p_k$ and pass through $q_1, \dots, q_\ell$. This is the kernel $W = \ker T$ of the evaluation map

$$T : V_d \to K^{3k+\ell}$$

defined by

$$T(F) = (F(p_1), \frac{\partial F}{\partial x}(p_1), \frac{\partial F}{\partial y}(p_1), \dots, F(p_k), \frac{\partial F}{\partial x}(p_k), \frac{\partial F}{\partial y}(p_k), \\ F(q_1), \dots, F(q_\ell)).$$

Our expectation is that such maps usually have maximal rank, which motivates the following definition.

Definition 7.8. Let $W \subset V_d$ be the subspace of polynomials which have double points at $p_1, \dots, p_k$ and vanish at $q_1, \dots, q_\ell$. Then the *expected dimension* of W is

$$\operatorname{edim} W = \max\left\{ \binom{d+2}{2} - 3k - \ell, 0 \right\}.$$

Equivalently, this number would be the dimension of W if the linear transformation T has the maximal rank $\min\{\binom{d+2}{2}, 3k+\ell\}$. Therefore, we always have

$$\dim W \geq \operatorname{edim} W.$$

In the parameter space $\mathbb{A}^{2(k+\ell)}$ of lists of $k+\ell$ points in $\mathbb{A}^2$, there is an open subset where the rank of T is as large as possible; by rank-nullity there is an open subset where the dimension of W is as small as possible. In particular, if there is one list $p_1, \dots, p_k, q_1, \dots, q_\ell$ such that the dimension of W is the expected dimension, then this holds for a general such list.

The Alexander-Hirschowitz theorem shows that the only exceptions to our expectation for the dimension of W were already discussed above:

Theorem 7.9 (Alexander-Hirschowitz theorem for the plane)**.** *Let $p_1, \dots, p_k, q_1, \dots, q_\ell \in \mathbb{A}^2$ be $k+\ell$ general points, and let $W \subset V_d$ be the*

subspace of polynomials which are double at $p_1, \dots, p_k$ and pass through $q_1, \dots, q_\ell$. Then $\dim W = \operatorname{edim} W$ except in the following cases:

(1) the double line: $(k, \ell, d) = (2, 0, 2)$;
(2) the double conic: $(k, \ell, d) = (5, 0, 4)$.

The proof of the theorem will occupy the rest of the section. First let us reduce to a special case.

Lemma 7.10. *If the Alexander-Hirschowitz theorem is true whenever (k, ℓ, d) are such that the expected dimension is 0, it is true.*

Proof. Suppose (k, ℓ, d) are such that the expected dimension $r = \binom{d+2}{2} - 3k - \ell$ is positive, and let W be the space of polynomials of degree d which are double at general points $p_1, \dots, p_k$ and pass through general points $q_1, \dots, q_\ell$. Consider r additional general points $q_{\ell+1}, \dots, q_{\ell+r}$, and let $W' \subset W$ be the subspace of polynomials which additionally pass through $q_{\ell+1}, \dots, q_{\ell+r}$. Then $(k, \ell+r, d)$ has expected dimension 0 and is not one of the exceptions in the Alexander-Hirschowitz theorem (all the exceptions have no simple points). Therefore $\dim W' = 0$. But W' is the kernel of an evaluation map

$$T : W \to K^r$$
$$T(F) = (F(q_{r+1}), \dots, F(q_\ell)),$$

so T is injective and $\dim W \leq r$. On the other hand $\operatorname{edim} W = r$ and $\dim W \geq \operatorname{edim} W$ is always true, so $\dim W = \operatorname{edim} W = r$. □

7.3. **Specialization.** Many cases of the Alexander-Hirschowitz theorem can be proved by a simple specialization technique.

Observation 7.11. Suppose $p_1, \dots, p_k$ and $q_1, \dots, q_\ell$ are distinct points that all lie on the x-axis $y = 0$; write $p_i = (x_i, 0)$ and $q_i = (x_i', 0)$. If $F(x, y)$ is double at $p_1, \dots, p_k$ and vanishes at $q_1, \dots, q_\ell$, then the polynomial $f(x) = F(x, 0)$ factors as

$$(x - x_1)^2 \cdots (x - x_k)^2 (x - x_1') \cdots (x - x_\ell') g(x)$$

for some polynomial $g(x)$. Indeed, $f(x)$ vanishes at all the points, and

$$\frac{df}{dx}(x) = \left.\frac{\partial F}{\partial x}(x, y)\right|_{y=0},$$

so $f'(x)$ must vanish at $x_1, \dots, x_k$.

In particular, if $p_1, \dots, p_k, q_1, \dots, q_\ell$ all lie on a line $L = 0$ and $2k + \ell \geq d + 1$, then any polynomial F of degree d that is double at $p_1, \dots, p_k$ and passes through $q_1, \dots, q_\ell$ must be divisible by the line: $F = GL$. Furthermore, since the line is smooth at each of its points, the residual polynomial G must still pass through $p_1, \dots, p_k$.

Example 7.12. Consider the space W of polynomials of degree ≤ 5 which are double at 7 general points; in our earlier notation, $(k, \ell, d) = (7, 0, 5)$. We have $\dim V_5 = 21$, so $\operatorname{edim} W = 21 - 3 \cdot 7 = 0$. We show that $W = \{0\}$.

Specialize 3 of the double points to be collinear. Then any $F \in W$ has 6 zeroes on the line, so is divisible by it. Any $F \in W$ can then be written in the form GL, where $G \in W'$ is an element of the space of degree ≤ 4 polynomials with 4 general double points and 3 collinear simple points. To show $W = \{0\}$ it is enough to show $W' = \{0\}$.

Further specialize one of the double points onto the line with the 3 collinear simple points. Then any $G \in W'$ has 5 zeroes on the line, is divisible by it. Any $G \in W'$ can then be written in the form HL, where $H \in W''$ is an element of the space of degree ≤ 3 polynomials with 3 general double points and one simple point (at the point where the double point specialized onto the line).

At this point we don't have to do any further specialization: W'' is visibly zero. Suppose $H \in W''$. Then it is a cubic double at two (collinear) points, so it is divisble by a line. Residually, we have a conic which is double at one point and passes through 3 points (all general). It has to be divisible by the line through the double point and a simple point. Residually, we have a degree 1 polynomial that vanishes at 3 general points, and this is absurd. Therefore $H = 0$.

On the other hand, it is possible to become "stuck" with the specialization. Intuitively, we can't place $d+2$ zeroes on a line without losing independence of conditions: once a polynomial F vanishes at $d + 1$ places on a line (say $y = 0$) it will be divisible by a line, and then the vanishing of the partial derivative $(\partial F/\partial x)(p)$ at any point of $y = 0$ is automatic. In terms of the evaluation map $T : V_d \to K^{3k+\ell}$, this would say that one of the rows of the corresponding matrix depends on the others, and so it won't necessarily have maximal rank. (If the number of columns $\binom{d+2}{2}$ is at least as big as the number of rows $3k + \ell$ then it definitely doesn't have maximal rank; in the other case things are more delicate.)

Example 7.13. Consider the space W of polynomials of degree ≤ 4 which are double at 5 general points; we know $\dim W > \operatorname{edim} W = 0$, so the specialization method should fail in this case. Put three of the double points on a line. Then any $F \in W$ factors as $F = GL$, where $G \in W'$ is of degree ≤ 3, has two general double points and passes through three collinear points. But already W' has $\operatorname{edim} W' = 10 - 3 \cdot 2 - 3 = 1$, so $W' \neq \{0\}$. Thus it is not possible to conclude $\dim W = 0$ by this specialization. (Which is a good thing!)

Example 7.14. Unfortunately, there are cases that are not exceptions to the Alexander-Hirschowitz theorem where this specialization is not sufficient. Consider $(k, \ell, d) = (15, 0, 8)$, so there are 15 double points and polynomials of degree at most 8. Then $\operatorname{edim} W = 0$, and the Alexander-Hirschowitz theorem asserts $W = \{0\}$. Our only option is to try and specialize 5 double points onto the line, to get 10 zeroes along the line. Residually we get polynomials $G \in W'$ of degree ≤ 7 with 10 general double points and 5 simple points on the line. The expected dimension is $36 - 3 \cdot 10 - 5 = 1$, so $W' \neq \{0\}$. The specialization was too drastic, and can't show $W = \{0\}$.

7.4. **Tangency conditions.** We have the following problem. Suppose $x_1, \ldots, x_d \in K$ are nonzero numbers (possibly repeated) and let $W \subset V_d$ be the subspace of polynomials such that $F(x, 0)$ is of the form

$$F(x, 0) = c \prod_{i=1}^{d} (x - x_i).$$

If we further impose that $F(0, 0) = 0$, then it follows that $F(x, 0) = 0$, so F is divisible by y. Then $(\partial F/\partial x)(0, 0) = 0$, so it is only one more linear condition for F to be double at the origin $(0, 0)$. Therefore we shouldn't expect a double point at $(0, 0)$ to impose three conditions on the coefficients of F.

The solution to this problem is to "dynamically" specialize a double point to the origin: instead of asking for F's that are double at the origin, we ask for F's that are limits of polynomials with a double point tending to the origin. It turns out that such F's are still divisible by the line, but furthermore if $F = yG$ then the residual polynomial G must additionally pass through the origin and be tangent to the line! This means there are actually two conditions that the coefficients of G have to satisfy.

Proposition 7.15. *Let $F_t(x, y) \in K[x, y, t]$ be a family of polynomials of degree at most d in x and y. Suppose there are nonzero numbers $x_1, \ldots, x_d \in K$ and a polynomial $H(t)$ such that*

$$F_t(x, 0) = H(t) \prod_{i=1}^{d} (x - x_i).$$

Furthermore suppose that at time t the polynomial $F_t(x, y)$ has a double point at $(0, t)$, so

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

Then $F_0(x,y)$ factors as

$$F_0(x,y) = yG(x,y),$$

where G is a polynomial such that

$$G(0,0) = \frac{\partial G}{\partial x}(0,0) = 0.$$

In other words, G passes through $(0,0)$ and (if it is smooth) it is tangent to the x-axis there.

Proof. We basically have to chase coefficients. Let us expand $F_t(x,y)$ as a polynomial in t with coefficients in $K[x,y]$:

$$F_t(x,y) = (a+bx+cy+dx^2+exy+fy^2+\cdots)+(g+hx+iy+\cdots)t+\cdots.$$

The constant term of the polynomial $0 = F_t(0,t)$ is a, so $a = 0$:

$$F_t(x,y) = (bx+cy+dx^2+exy+fy^2+\cdots)+(g+hx+iy+\cdots)t+\cdots.$$

Next consider $F_t(x,0)$. Written as a polynomial in t with coefficients in $K[x]$, this looks like

$$F_t(x,0) = (bx+dx^2+\cdots)+(g+hx+\cdots)t+\cdots.$$

Since $F_t(x,0)$ takes the form $c(t)\prod(x-x_i)$, the coefficient of each power of t is a constant multiple of $\prod(x-x_i)$. Since all the x_i are nonzero and the coefficient of t^0, namely $(bx+\cdots)$, is divisible by x, this forces the coefficient of t^0 to be 0. Therefore $(bx+cy+\cdots)$ is divisible by y, and we can write

$$F_t(x,y) = y(c+ex+fy+\cdots)+(g+hx+iy+\cdots)t+\cdots.$$

Consider the derivative $(\partial F_t/\partial y)(0,t)$. First compute

$$\frac{\partial F_t}{\partial y}(x,y) = (c+ex+2fy+\cdots)+(i+\cdots)t.$$

Then the constant term of $(\partial F_t/\partial y)(0,t)$ is c, so $c = 0$:

$$F_t(x,y) = y(ex+fy+\cdots)+(g+hx+iy+\cdots)t+\cdots.$$

Look again at $F_t(0,t) = 0$, this time at the linear term: we get $g = 0$, so

$$F_t(x,y) = y(ex+fy+\cdots)+(hx+iy+\cdots)t+\cdots.$$

Consider $F_t(x,0)$ again:

$$F_t(x,0) = (hx+\cdots)t+\cdots.$$

As before, the coefficient $(hx+\cdots)$ must be a constant multiple of $\prod(x-x_i)$, and therefore y divides $(hx+iy+\cdots)$. So,

$$F_t(x,y) = y(ex+fy+\cdots)+y(i+\cdots)t+\cdots.$$

Finally, consider the derivative $(\partial F_t/\partial x)(0,t)$. We have

$$\frac{\partial F_t}{\partial x}(x,y) = (ey + \cdots) + y(\cdots)t + \cdots,$$

and looking at the coefficient of t in $(\partial F_t/\partial x)(0,t)$ we find that $e = 0$. Finally, we conclude that

$$F_t(x,y) = y(fy + \cdots) + y(i + \cdots)t + \cdots,$$

so

$$F_0(x,y) = y(fy + h.o.t.).$$

Therefore y divides F_0 and the residual polynomial G satisfies $G(0,0) = 0$ and $(\partial G/\partial x)(0,0) = 0$. □

7.5. **Collections of points with some on a line.** We will prove the Alexander-Hirschowitz theorem by induction on d. Since our specialization produces points and tangency conditions which lie along a line, the theorem doesn't directly apply to the systems of residual curves that arise after we split off a line. In order to get a clean induction, we need to generalize the theorem to allow such conditions.

Fix a line $L : y = 0$, and nonnegative numbers $(k, \ell, \ell', \epsilon, d)$ with $\epsilon = 0$ or 1 and

$$0 \leq \ell' \leq (d + 2 - \epsilon)/2.$$

Let $W \subset V_d$ be the subspace of polynomials with

- k general double points,
- ℓ general simple points,
- ℓ' prescribed simple points along L, and
- ϵ tangency conditions on L (i.e. the polynomials pass through the point and the derivative in the direction of the line vanishes.)

The expected dimension of W is defined to be

$$\operatorname{edim} W = \max\left\{ \binom{d+2}{2} - 3k - (\ell + \ell') - 2\epsilon, 0 \right\};$$

this is what the dimension would be if the obvious evaluation map has maximal rank.

Theorem 7.16 (Generalized Alexander-Hirschowitz theorem)**.** *With the above assumptions, we have* $\dim W = \operatorname{edim} W$ *except in the following cases:*

- *a double line:* $(k, \ell, \ell', \epsilon, d) = (2, 0, 0, 0, 2)$*;*
- *a double conic:* $(k, \ell, \ell', \epsilon, d) = (5, 0, 0, 0, 4)$*;*
- *three collinear zeroes:* $(k, \ell, \ell', \epsilon, d) = (0, 0, 1, 1, 1)$*;*

- *two double points and 4 collinear zeroes:* $(k, \ell, \ell', \epsilon, d) = (2, 0, 2, 1, 3)$.

As with the original Alexander-Hirschowitz theorem, it is enough to prove the theorem in the case edim $= 0$, by adding additional general simple points.

Lemma 7.17. *With the above assumptions, we have* $\ell' + 2\epsilon \leq d + 1$. *Thus, the number of prescribed zeroes along* L *is at most* $d + 1$.

Proof. We have

$$\ell' + 2\epsilon \leq (d + 2 + 3\epsilon)/2 \leq (d + 5)/2.$$

If $d \geq 3$ then $\ell' + 2\epsilon \leq d + 1$ follows. The cases $d = 0, 1, 2$ are easily checked independently. □

7.6. **Specialization, take 2.** Suppose numbers $(k, \ell, \ell', \epsilon, d)$ have been given, that edim $= 0$, and we are trying to exhibit positions for the points such that the space W is zero. Let the line be $y = 0$, place the points on the line at points other than $(0, 0)$, and at time t let $W_t \subset V_d$ be the subspace where all but one of the points is fixed, and the last point is positioned at $(0, t)$. If W_t is ever zero we win, so suppose W_t is nonzero for every time t.

Lemma 7.18. *If* W_t *is nonzero for every time* t, *then there is a polynomial* $F_t(x, y) \in K[x, y, t]$ *such that* $F_t \in W_t$ *for all* t. *Furthermore, we can assume* $F_t \neq 0$ *for all* t.

Proof. The space W_t is the kernel of an evaluation map

$$T_t : V_d \to K^N.$$

If we write down the matrix of this map with respect to standard bases, then the entries of the matrix will be polynomials in t. For every time t, the matrix doesn't have maximal rank (since edim $= 0$ but $\dim W_t \neq 0$), so the maximal minors of this matrix are all zero. Think of this family of matrices as a matrix $A(t)$ whose entries come from the field $K(t)$ of rational functions in t. Then the minors of $A(t)$ are also zero, since they are polynomials in t which are zero when evaluated at every time t. Therefore $A(t)$ doesn't have maximal rank, and there is a vector in its kernel. The entries of this vector are rational functions of t. By clearing denominators and cancelling common factors, we can make them be polynomials in t with no common zero. These polynomials in t are the coefficients of the monomials (in x and y) of a polynomial $F_t(x, y)$ such that $F_t \in W_t$ for all t. Furthermore, since the polynomials in t had no common zero, every F_t has at least one monomial that appears in it with a nonzero coefficient. □

Now we pursue the following specialization strategy. Suppose we have numbers $(k, \ell, \ell', \epsilon, d)$ with $\ell' + 2\epsilon \leq d + 1$ and suppose edim $= 0$ but $W \neq \{0\}$. If $\ell' + 2\epsilon = d + 1$ then every $F \in W$ is divisible by L, $F = GL$, and we can instead examine the residual polynomials $G \in W'$. Here W' corresponds to the data $(k, \ell, 0, 0, d - 1)$. The expected dimension of W' is the same as W (the binomial coefficient went down by $d + 1$, and so did the number of conditions on the line). Thus, so long as this data is a case where the AH-theorem is true, we win.

Suppose instead that $\ell' + 2\epsilon \leq d - 1$. Then we can specialize some number m of double and/or n of simple points onto the line until the total number of zeroes on the line is d or $d + 1$. If it is $d + 1$, then $F = GL$ and the residual polynomials $G \in W'$ correspond to the data $(k - m, \ell - n, m, 0, d - 1)$. Observe that $m \leq (d + 1)/2$ since the total number of zeroes on the line was $d + 1$. Again the expected dimension hasn't changed (check!), so if the generalized AH-theorem holds for this data, we win again.

(The hard case.) Finally it is possible that we have arrived in a case where there are d zeroes along the line after specializing the double and/or simple points (this is analogous to Example 7.14 where we got "stuck"). In this case, if we want to specialize another double point onto the line then we need to consider tangency conditions. Specialize a double point onto the x-axis $y = 0$ by approaching along the y-axis. If there is a curve with the required properties for all times t, then the space W' corresponding to the data

$$(k - m - 1, \ell - n, m, 1, d - 1)$$

must be nonzero. We have $m \leq d/2$, so the ℓ' here satisfies the required inequality

$$\ell' \leq ((d - 1) + 2 - 1)/2.$$

Additionally, the expected dimension is still the same. So, we apply the contrapositive: if W' is zero, then $W = \{0\}$ and we win.

7.7. **The combinatorial game.** We're done with the algebraic geometry of the problem; now it's time to analyze the specialization and make sure that we can always carry it out except in the cases where the theorem is false.

Example 7.19. One good example is probably more illustrative than the full proof of the theorem. Again consider degree $d = 8$ with $k = 15$ double points. Specialize 4 double points onto a line, and then send in a 5th. Residually we have the data $(10, 0, 4, 1, 7)$; degree 7 polynomials

with 4 simple points and a tangency condition along a line, as well as 10 general double points.

We already have 6 zeroes along the line, so specialize one more double point onto the line to get 8. The line splits and leaves us with the residual system $(9, 0, 1, 0, 6)$. It takes 7 zeroes on the line to split it off again, so specialize 3 more double points. Residually we get $(6, 0, 3, 0, 5)$. It takes 6 zeroes on the line to make it split again and we already have 3, so send in 2 more double points. We get $(4, 0, 1, 1, 4)$, since we had an excesss zero along the line. We need 5 zeroes on the line and we have 3, so specialize another double point and get $(3, 0, 1, 0, 3)$. We have 1 zero on the line and need 4, so send in two double points to get $(1, 0, 1, 1, 2)$. We have 3 zeroes on the line, which makes it already split, and we get $(1, 0, 0, 0, 1)$. This is zero, since there are no lines with a double point.

Since the proof of the main theorem is by induction, we just have to analyze one step of this procedure and make sure we don't end up with any of the systems where the theorem is false.

Proof of the Generalized Alexander-Hirschowitz Theorem 7.16. We prove the theorem by induction on d, starting from high degree cases and assuming the result holds for lower degree cases. This mimics the computation in Example 7.19, but for a general case.

Notice that all the supposed counterexamples to the generalized theorem have degree at most 4. There is no obstruction to repeatedly specializing points onto lines and computing the residual systems before we get down to systems with $d = 5$.

Consider a system $(k, \ell, \ell', \epsilon, 5)$. We are given $\ell' \leq (d + 2 - \epsilon)/2 \leq \frac{7}{2}$ and $\epsilon \leq 1$. So $\ell' + 2\epsilon \leq 5$. If there are any double points, specialize at least one of them onto the line, and then continue to specialize points onto the line until either the number of zeroes on the line will be 6, or there are already 5 zeroes and a double point will cause there to be 7. Then in either case we can perform the specialization (keep track of tangency data as needed) and the resulting system of degree 4 has edim $= 0$ and it isn't just 5 double points (there are some simple points or a tangency residual to the line). Thus the resulting system W' is $\{0\}$, and also $W = \{0\}$.

Consider a system $(k, \ell, \ell', \epsilon, 4)$ other than $(5, 0, 0, 0, 4)$. We are given $\ell' \leq 3$ and $\epsilon \leq 1$. If $\ell' + 2\epsilon \neq 0$, we can specialize points onto the line arbitrarily until there are 5 or 6 zeroes on the line and the residual system will have at most 3 zeroes on the line; we win in this case. If $\ell' = \epsilon = 0$, then either there are at least 3 simple points or 6 double points. If there are 6 double points then emptiness of $(5, 1, 0, 0, 4)$ implies emptiness with 6 double points. In any case we can specialize

exactly 5 zeroes onto the line, and the resulting W' is $\{0\}$ (it has no tangency condition).

Consider a system $(k, \ell, \ell', \epsilon, 3)$ other than $(2, 0, 2, 1, 3)$. We have $\ell' \leq \frac{5}{2}$ and $\epsilon \leq 1$. If $\ell' + 2\epsilon = 4$ we must have $k \geq 3$ or $\ell \geq 3$, the line splits off without specialization, and the resulting W' is empty. If $\ell' + 2\epsilon < 4$, specialize points onto L until we get 4 or 5 zeroes on L; start with the double points if there are any. If the resulting W' has any double points then it also has a simple point or tangency along the line, and thus W' is $\{0\}$.

Consider a system $(k, \ell, \ell', \epsilon, 2)$ other than $(2, 0, 0, 0, 2)$. We have $\ell' \leq 1$ and $\epsilon \leq 1$. If $\ell' + 2\epsilon = 3$, the line splits off and W' is empty. If there are two double points and any other point, then clearly $W' = \{0\}$ since two copies of the line between the two double points split off. If there is one or fewer double points then we can specialize at will and there can be at most two residual zeroes on the line.

The degree 1 case is clear.

8. Exercises

Problem 8.1. Find, with proof, a polynomial $f \in \mathbb{R}[x]$ of *smallest degree* such that

$$\begin{aligned} f(0) &= 2 \\ f(1) &= 4 \\ f(2) &= 12 \\ f(3) &= 32 \\ f(4) &= 70 \\ f(5) &= 132. \end{aligned}$$

What is the second smallest degree of a polynomial satisfying the above equalities?

Problem 8.2. Let $f(x) \in \mathbb{R}[x]$, $a \in \mathbb{R}$, and let $m \geq 1$. Show that $f^{(k)}(a) = 0$ for all $0 \leq k \leq m-1$ if and only if $(x-a)^m$ divides f.

Problem 8.3 (Multivariate interpolation—a preview)**.** Consider three distinct points

$$(x_1, y_1), (x_2, y_2), (x_3, y_3) \in \mathbb{R}^2$$

and three arbitrary values $z_1, z_2, z_3 \in \mathbb{R}$. We answer the question: when can we find a linear polynomial $f(x, y)$ of the form $f(x, y) = ax + by + c$ such that $f(x_i, y_i) = z_i$ for all i?

(1) Show that if the three points (x_i, y_i) are not collinear, then we can find an f such that $f(x_i, y_i) = z_i$ for all i. (*Hint:* think about the existence proof for single-variable Lagrangian interpolation.)
(2) Show that if the three points (x_i, y_i) are collinear, then for some choice of the values z_i it is not possible to find such an f.

Problem 8.4. Let $f_t(x) \in \mathbb{R}[x, t]$ be a polynomial in two variables, viewed as a family of polynomials in x parameterized by t. Suppose that

$$\begin{aligned} f_t(0) &= 0 \\ f_t(t) &= t^2 \\ f_t(t^2) &= t^4 \end{aligned}$$

(That is, the polynomial at time t passes through the points $(0, 0)$, (t, t^2), and (t^2, t^4) on the parabola $y = x^2$.)

(1) Give at least two families $f_t(x)$ satisfying the assumptions. The more examples the better!

(2) What can you say is always true about the the polynomial $f_0(x)$, regardless of what the family $f_t(x)$ is? (*Hint:* as $t \to 0$, three points are "colliding." Can you say three interesting things?)

Problem 8.5. Let R and S be commutative rings with 1. A *ring homomorphism* $\phi : R \to S$ is a function which preserves the addition, multiplication, and unit element. That is,

(1) $\phi(f+g) = \phi(f) + \phi(g)$,
(2) $\phi(fg) = \phi(f)\phi(g)$, and
(3) $\phi(1_R) = 1_S$.

Recall that $\mathbb{R}[x]$ is the polynomial ring and $\mathbb{R}^{\mathbb{R}}$ is the ring of functions from $\mathbb{R}$ to $\mathbb{R}$. There is a function $\phi : \mathbb{R}[x] \to \mathbb{R}^{\mathbb{R}}$ which takes a polynomial to its corresponding polynomial function. Formally, if $p \in \mathbb{R}[x]$ then $\phi(p)$ is the function $x \mapsto p(x)$.

(1) Show that ϕ is a ring homomorphism.
(2) Show that ϕ is injective.
(3) (For those of you with more algebra background.) Let p be a prime number, and let $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ be the finite field with p elements. We analogously define a function $\phi : \mathbb{F}_p[x] \to \mathbb{F}_p^{\mathbb{F}_p}$. Describe the kernel of ϕ.

Problem 8.6. Let R be a commutative ring. An *ideal* I in R is an additive subgroup which is closed under multiplication by elements in R. If $f \in R$, the *principal ideal generated by* f is the subset

$$(f) := \{af : a \in R\}.$$

An ideal I is *principal* if $I = (f)$ for some $f \in R$.

(1) Let $f \in R$. Show that the principal ideal (f) is actually an ideal.
(2) Show that in the polynomial ring $\mathbb{R}[x]$ every ideal is principal. (*Hint:* division algorithm.)

Problem 8.7. Find, with proof, a polynomial $f \in \mathbb{R}[x]$ of *smallest degree* such that

$$\begin{aligned} f(0) &= 2 \\ f'(0) &= -3 \\ f''(0) &= 2 \\ f(1) &= 0 \\ f'(1) &= -1 \\ f''(1) &= 2. \end{aligned}$$

Problem 8.8 (Collisions of points in the plane)**.** Consider the following three points in the plane $\mathbb{R}^2$, which depend on a time parameter t:

$$(x_1(t), y_1(t)) = (0,0) \qquad (x_2(t), y_2(t)) = (t,0) \qquad (x_3(t), y_3(t)) = (0,t).$$

Let $z_1 = 0$, and let $z_2(t)$ and $z_3(t)$ be polynomials of t. Suppose that $f_t(x,y) \in \mathbb{R}[x,y,t]$ is a polynomial of three variables x, y, t such that

$$f_t(x_i(t), y_i(t)) = z_i(t).$$

(1) Show that $z_2(0) = z_3(0) = 0$.
(2) Consider the polynomial $f_0(x,y)$. Show that

$$\begin{aligned} f_0(0,0) &= 0 \\ \frac{\partial f_0}{\partial x}(0,0) &= z_2'(0) \\ \frac{\partial f_0}{\partial y}(0,0) &= z_3'(0). \end{aligned}$$

Thus, the limiting interpolation problem is to specify the value and the first-order partial derivatives of f_0 at the point $(0,0)$.

Problem 8.9. Let $(x_0, y_0) \in \mathbb{R}^2$, and let $a, b, c \in \mathbb{R}$. Describe all polynomials $f(x,y)$ such that

$$\begin{aligned} f(x_0, y_0) &= a \\ \frac{\partial f}{\partial x}(x_0, y_0) &= b \\ \frac{\partial f}{\partial y}(x_0, y_0) &= c. \end{aligned}$$

(*Hint:* one approach is to first solve the problem when $(x_0, y_0) = (0,0)$. Then make a change of coordinates.)

Problem 8.10 (Interpolation with non-consecutive derivatives)**.** In the interpolation problem with derivatives, we specified the 0th through $(m_i - 1)$st derivatives of a polynomial at point x_i. We can instead try to specify more arbitrary derivatives. This problem shows that this story is more delicate.

(1) Let $x_1, x_2, x_3 \in \mathbb{R}$ be distinct, and let $y_1, y_2, y_3 \in \mathbb{R}$. Show that there is a unique polynomial $f \in \mathbb{R}[x]$ of degree at most 2 such that

$$f'(x_1) = y_1 \qquad f'(x_2) = y_2 \qquad f(x_3) = y_3.$$

(2) Let $x_1, x_2, x_3 \in \mathbb{R}$ be distinct, and let $y_1, y_2 \in \mathbb{R}$. Show that if $f \in \mathbb{R}[x]$ is a polynomial of degree at most 2 and

$$f'(x_1) = y_1 \qquad \text{and} \qquad f'(x_2) = y_2,$$

then $f'(x_3)$ can be determined in terms of x_1, x_2, x_3, y_1, y_2. Thus, for most choices of $y_3 \in \mathbb{R}$, there is no polynomial $f \in \mathbb{R}[x]$ of degree at most 2 such that

$$f'(x_1) = y_1 \qquad f'(x_2) = y_2 \qquad f'(x_3) = y_3.$$

Problem 8.11. Formulate and prove the analogue of Theorem 1.14 in the notes for Lagrangian interpolation with multiplicities.

Problem 8.12. Consider a linear system

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

For short, write it in matrix form as $A\mathbf{x} = \mathbf{b}$.

(1) The *associated homogeneous system* is the system $A\mathbf{x} = \mathbf{0}$. Show that the set of solutions

$$U = \{\mathbf{h} \in K^n : A\mathbf{h} = \mathbf{0}\} \subset K^n$$

is a subspace of K^n.

(2) Suppose $\mathbf{p} \in K^n$ solves the original system: $A\mathbf{p} = \mathbf{b}$. (We say $\mathbf{p}$ is a *particular solution*.) Show that any solution $\mathbf{p}'$ of $A\mathbf{x} = \mathbf{b}$ can be written uniquely in the form

$$\mathbf{p}' = \mathbf{p} + \mathbf{h}$$

for some solution $\mathbf{h} \in U$ of the associated homogeneous system.

Problem 8.13. Let V be a vector space. Show that the intersection of any collection of subspaces of V is a subspace of V.

Problem 8.14. Let V be a vector space, and let $S \subset V$ be a subset.

(1) Show that $\operatorname{span}(S)$ is a subspace of V.
(2) Show that any subspace of V which contains S contains $\operatorname{span}(S)$. Therefore, $\operatorname{span}(S)$ is the smallest subspace of V which contains S.

Problem 8.15. Let $\{W_\alpha\}_{\alpha \in A}$ be any collection of subspaces. Show that

$$\sum_{\alpha \in A} W_\alpha = \operatorname{span}\left(\bigcup_{\alpha \in A} W_\alpha\right).$$

Problem 8.16. Let V be a finite-dimensional vector space, and let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ be a linearly independent list. Show that this list can be extended to a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{u}_1, \ldots, \mathbf{u}_m \in V$ of V.

Problem 8.17. Let $(x_1, y_1), \ldots, (x_5, y_5) \in K^2$ be five points in the plane. Prove that there is a nonzero polynomial $f(x, y) \in K[x, y]$ of degree at most 2 such that $f(x_i, y_i) = 0$ for $i = 1, \ldots, 5$.

(The degree of a monomial $cx^a y^b$ is $a + b$. The degree of a multivariable polynomial is the highest degree of a monomial with a nonzero coefficient.)

Problem 8.18. Let $V \subset K[x]$ be the subset of polynomials of degree at most 5 which satisfy $f(1) = f(2) = f'(2) = 0$.

(1) Show V is a subspace of $K[x]$.
(2) Give a basis of V.

Problem 8.19. Let X be a finite set, and let K^X be the vector space of all functions from X to K. What is the dimension of K^X? Compute a basis of K^X.

Problem 8.20. Show that a subspace of a finite-dimensional vector space is finite-dimensional.

Problem 8.21. Show that each of the following vector spaces is infinite-dimensional.

(1) V is the vector space over $\mathbb{R}$ of polynomial functions $f : \mathbb{R} \to \mathbb{R}$.
(2) V is the vector space over $\mathbb{R}$ of continuous functions $f : \mathbb{R} \to \mathbb{R}$.
(3) V is the vector space $K[x, y]$ of polynomials in two variables x, y.
(4) V is the vector space

$$\bigoplus_{i=1}^{\infty} K := \{(a_1, a_2, \ldots) : a_i \in K \text{ and all but finitely many } a_i \text{ are } 0\},$$

where the addition is defined componentwise and the scalar multiplication is evident. (This vector space is called an *infinite direct sum*.)
(5) V is the vector space

$$\prod_{i=1}^{\infty} K := \{(a_1, a_2, \ldots) : a_i \in K\},$$

with the evident operations. (This vector space is called an *infinite direct product*.)

Problem 8.22. Let $f(x, y) \in K[x, y]$ be a polynomial in two variables.

(1) Suppose that K is infinite, and suppose that $f(a, b) = 0$ for all $a, b \in K$. Show that f is the zero polynomial.

(2) Suppose that K is a finite field. Show that there is a *nonzero* polynomial $f(x, y) \in K[x, y]$ such that $f(a, b) = 0$ for all $a, b \in K$.

Problem 8.23. Let $V \subset K[x, y]$ be a subspace of $K[x, y]$. The *common zero locus of* V is the set

$$Z(V) = \{(a, b) \in K^2 : f(a, b) = 0 \text{ for all } f \in V\} \subset K^2.$$

(1) For the rest of the problem, suppose K is infinite. Show that if $V \neq \{0\}$, then $Z(V)$ is a *proper* subset of K^2.
(2) For a point $p = (a, b) \in K^2$, define a subset

$$V(-p) = \{f \in V : f(p) = 0\} \subset V.$$

Show that $V(-p)$ is a subspace of V. Furthermore, it is a *proper* subspace of V if and only if $p \notin Z(V)$.
(3) Suppose V is finite-dimensional and $p \in K^2 \setminus Z(V)$. Show that

$$\dim V(-p) = \dim V - 1.$$

Problem 8.24. Let V and W be vector spaces, and let $T, S : V \to W$ be linear transformations. Show that $T + S : V \to W$, defined pointwise, is a linear transformation.

Problem 8.25. Let $U \subset K^n$ be a subspace of dimension $n - m$.

(1) Show that there is a linear transformation $T : K^n \to K^m$ such that $\ker T = U$.
(2) Show that there is an $m \times n$ matrix A such that

$$U = \{\mathbf{x} \in K^n : A\mathbf{x} = \mathbf{0}\}.$$

Problem 8.26. Let $T : V \to W$ be a linear transformation.

(1) Suppose $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ are linearly independent and T is injective. Show that $T\mathbf{v}_1, \dots, T\mathbf{v}_n$ are linearly independent.
(2) Suppose $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ span V and T is surjective. Show that $T\mathbf{v}_1, \dots, T\mathbf{v}_n$ span W.
(3) Suppose $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ are a basis of V. Show that T is bijective if and only if $T\mathbf{v}_1, \dots, T\mathbf{v}_n$ are a basis of W.

Problem 8.27. Let $T : V \to W$ be a linear transformation, and suppose V and W are finite-dimensional. (Do not use the rank-nullity theorem for the next parts!)

(1) Suppose that V and W are finite-dimensional and $\dim V > \dim W$. Show that T is not injective.
(2) Suppose that V and W are finite-dimensional and $\dim V < \dim W$. Show that T is not surjective.

(3) Suppose that V and W are finite-dimensional and $\dim V = \dim W$. Show that T is injective iff it is surjective iff it is bijective.

Problem 8.28. Let $V_{n-1} \subset K[x]$ be the vector space of polynomials of degree at most $n-1$. Let $x_1, \ldots, x_n \in K$ be distinct, and define a linear transformation

$$\begin{aligned} T : V_{n-1} &\to K^n \\ Tf &= (f(x_1), \ldots, f(x_n)). \end{aligned}$$

Show that T is injective, and use this to give a simpler "non-constructive" proof of our first statement on Lagrangian interpolation. (Theorem 1.13 in the notes.)

Problem 8.29. Carry out a procedure similar to the one outlined in Problem 5 to give a simple proof of our theorem on Lagrangian interpolation with derivatives. (Theorem 1.19 in the notes.)

Problem 8.30. Let V, W, and W' be vector spaces, and let $T : V \to W$ and $S : V \to W'$ be linear transformations. Recall that the direct product $W \times W'$ is the vector space

$$\{(\mathbf{w}, \mathbf{w}') : \mathbf{w} \in W, \mathbf{w}' \in W'\}$$

with the evident operations.

(1) Show that the function

$$T \times S : V \to W \times W'$$

defined by

$$(T \times S)(\mathbf{v}) = (T\mathbf{v}, S\mathbf{v})$$

is a linear transformation.

(2) Show that the function

$$\Psi : \mathrm{Hom}(V, W) \times \mathrm{Hom}(V, W') \to \mathrm{Hom}(V, W \times W')$$

defined by

$$\Psi(T, S) = T \times S$$

is a bijective linear transformation. In other words, giving the data of a pair of linear transformations $V \to W$ and $V \to W'$ is "the same" as giving a linear transformation $V \to W \times W'$.

Problem 8.31. Determine the possible Hilbert functions $h_Z(d)$ of a collection of 5 points $Z = \{p_1, \ldots, p_5\} \in K^2$. For each possible Hilbert function h, geometrically describe necessary and sufficient conditions on Z such that $h_Z = h$.

Problem 8.32. Let $Z = \{p_1, \dots, p_n\} \in K^2$ be a collection of n collinear points. Compute the Hilbert function of Z.

Problem 8.33. Classify all collections $Z = \{p_1, \dots, p_n\} \in K^2$ of n points such that $h_Z(n-2) = n$. (Remember that we always know $h_Z(n-1) = n$.)

Problem 8.34. Inscribe a regular hexagon in the unit circle in $\mathbb{R}^2$, and let $Z = \{p_1, \dots, p_6\} \in \mathbb{R}^2$ be the six vertices. Compute the Hilbert function of Z.

Problem 8.35. For an integer $e \geq 0$, consider the set of $n = \binom{e+2}{2}$ points

$$Z = \{(i,j) : i,j \in \mathbb{N} \text{ and } i+j \leq e\} \subset \mathbb{R}^2.$$

Show that

$$h_Z(d) = \min\left\{\binom{d+2}{2}, n\right\}$$

for every $d \geq 0$. Therefore, the evaluation map $T_{Z,d} : V_d \to K^n$ has maximal rank for every d.

Problem 8.36. Let $T, S : \mathbb{A}^n \to \mathbb{A}^n$ be affine linear transformations. Show that $T \circ S$ is affine linear.

Problem 8.37. Show that an affine linear transformation $T : \mathbb{A}^n \to \mathbb{A}^n$ given by

$$T\mathbf{x} = A\mathbf{x} + \mathbf{x}_0$$

(where A is an $n \times n$ matrix and $\mathbf{x}_0 \in K^n$) is invertible if and only if the matrix A is invertible.

Problem 8.38. Let $Z = \{p_1, \dots, p_n\} \in \mathbb{A}^2$ be n distinct points in the plane. Show that Z is the common zero locus of two polynomials $f, g \in K[x,y]$ such that $\deg f \leq n-1$ and $\deg g \leq n$. (*Hint:* first handle the case where all the points p_i have different x-coordinates.)

Problem 8.39. Let $T : \mathbb{A}^2 \to \mathbb{A}^2$ be an invertible affine linear transformation and suppose its inverse $S : \mathbb{A}^2 \to \mathbb{A}^2$ is given by

$$S\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}.$$

Let $F \in K[x,y]$ be any polynomial, and let $X \subset \mathbb{A}^2$ be the zero locus $F(x,y) = 0$. Show that $T(X)$ is the zero locus of $F(ax+by+e, cx+dy+f)$.

Problem 8.40. With the same notation as in Problem 5, let $K = \mathbb{R}$ and let

$$F = Ax^2 + Bxy + Cy^2 + Dx + Ey + F.$$

Show that the discriminants of $F(x, y)$ and $F(ax + by + e, cx + dy + f)$ have the same sign. Therefore, invertible affine linear transformations carry conics to conics with discriminants of the same sign.

Problem 8.41. Let $T \subset \mathbb{A}^2_{\mathbb{R}}$ be a triangle in the plane. Show that there is an ellipse inscribed in T which is tangent to T at each of the midpoints of T. (*Hint:* this is easy if T is equilateral.)

Problem 8.42. The following construction gives one of the most general ways that topologies arise in geometry. Let X be a set and let $d : X \times X \to \mathbb{R}_{\geq 0}$ be a function that takes nonnegative real values. We think of $d(x, y)$ as the *distance* between x and y. The pair (X, d) is called a *metric space* if the distance function satisfies the following properties for all $x, y, z \in X$.

(1) (Definiteness) $d(x, y) = 0$ iff $x = y$.
(2) (Symmetry) $d(x, y) = d(y, x)$.
(3) (Triangle inequality) $d(x, z) \leq d(x, y) + d(y, z)$.

Let (X, d) be a metric space. The *open ball of radius ϵ centered at $x \in X$* is the set

$$B_\epsilon(x) = \{y \in X : d(x, y) < \epsilon\}.$$

A subset $U \subset X$ is called *open* if for every $x \in U$ there is some $\epsilon > 0$ such that $B_\epsilon(x) \subset U$.

(1) Show that any open ball $B_\epsilon(x)$ is an open set.
(2) Show that the collection τ of open sets defines a topology on X.
(3) Show that the topological space (X, τ) is Hausdorff: given two distinct points $x, y \in X$, there are open sets $U, V \subset X$ such that $x \in U$, $y \in V$, and $U \cap V = \emptyset$.

On the other hand, the Zariski topology on affine space is not Hausdorff. Therefore, there is not a distance function $d : \mathbb{A}^n \times \mathbb{A}^n \to \mathbb{R}$ such that the corresponding topology is the Zariski topology.

Problem 8.43. A polynomial map $F : \mathbb{A}^n \to \mathbb{A}^m$ is a function of the form

$$F(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)),$$

where $f_1, \ldots, f_m \in K[x_1, \ldots, x_n]$. Let $Y \subset \mathbb{A}^m$ be a closed set. Show that the preimage of Y,

$$F^{-1}(Y) := \{p \in \mathbb{A}^n : F(p) \in Y\},$$

is a closed set in $\mathbb{A}^n$. (In topology language, this says F is *continuous*. *Hint:* handle the case where $Y = V(g)$ is defined by one equation first.)

Problem 8.44. Let (X, τ) be a topological space, and let $Z \subset X$ be a subset. The *closure* of Z is the smallest closed set that contains Z. Show that the closure of Z actually exists, and that it is the intersection of all closed sets that contain Z.

Problem 8.45. Let $\mathbb{Z}^2 \subset \mathbb{A}^2_{\mathbb{R}}$ be the lattice of points with integer coordinates. Show that the closure of $\mathbb{Z}^2$ in $\mathbb{A}^2_{\mathbb{R}}$ is $\mathbb{A}^2_{\mathbb{R}}$. (We say $\mathbb{Z}^2$ is *dense* in $\mathbb{A}^2_{\mathbb{R}}$.)

Problem 8.46 (Challenge)**.** Let $X \subset \mathbb{A}^2_{\mathbb{R}}$ be the subset defined by the equation $y = e^x$. Show that the closure of X in $\mathbb{A}^2_{\mathbb{R}}$ is $\mathbb{A}^2_{\mathbb{R}}$.

Problem 8.47. Think of the space $X = \mathrm{Mat}_{2\times 2}(K)$ of 2×2 matrices with entries in K as an affine space $\mathbb{A}^4$. For each of the following subsets of X, determine (with proof) if the set is either open, closed or neither.

(1) The set of matrices A such that
$$A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$
(2) The set of invertible matrices.
(3) The set of matrices of rank 1.
(4) The set of matrices of rank at least 1.
(5) The set of matrices of rank at most 1.

Problem 8.48. Let $X \subset \mathbb{A}^3$ be the twisted cubic curve, which is the image of the map $F : \mathbb{A}^1 \to \mathbb{A}^3$ defined by
$$F(t) = (t, t^2, t^3).$$
Consider the projections of X onto the xy-, yz-, and xz-planes. For each projection, determine if it is closed, and if it is, give equations for the image.

Problem 8.49. Show that the parabola $y = x^2$ in $\mathbb{A}^2$ is isomorphic to $\mathbb{A}^1$.

Problem 8.50. Show that if $f(t) \in \mathbb{C}[t]$, then there are (at most) finitely many constants $c \in K$ such that the polynomial $f(t) + c$ has a repeated root.

Problem 8.51. Let $f(t), g(t) \in \mathbb{C}[t]$ be nonzero, and consider the polynomial mapping

$$\begin{aligned} F : \mathbb{A}^1_{\mathbb{C}} &\to \mathbb{A}^2_{\mathbb{C}} \\ F(t) &= (f(t), g(t)). \end{aligned}$$

Let $a, b \in \mathbb{C}$ be chosen so that $\deg(af(t) + bg(t)) = \max\{\deg f(t), \deg g(t)\}$. (Why can we do this?) For a constant $c \in \mathbb{C}$, consider the line

$$L = V(ax + by + c).$$

Show that for all but finitely many choices of c, the preimage $F^{-1}(L)$ consists of exactly

$$\max\{\deg f(t), \deg g(t)\}$$

points.

Problem 8.52. Show that the circle $X : x^2 + y^2 = 1$ in $\mathbb{A}^2_{\mathbb{C}}$ is **not** isomorphic to $\mathbb{A}^1_{\mathbb{C}}$. (*Hint:* suppose $F : \mathbb{A}^1_{\mathbb{C}} \to X$ given by $F(t) = (f(t), g(t))$ is an isomorphism. Use the previous problem to bound the degrees of f and g, and analyze the remaining cases to get a contradiction.)

Problem 8.53 (Challenge)**.** Let K be algebraically closed, and let $f \in K[x, y]$ be nonconstant. Show that the cardinality of $X = V(f) \subset \mathbb{A}^2$ is the same as the cardinality of K. Conclude that any two irreducible plane curves in $\mathbb{A}^2$ are *homeomorphic* (there is a continuous bijection between them, with continuous inverse).

Problem 8.54. Show that if $p_1, \ldots, p_m$ are m general points in $\mathbb{A}^n$, then no two points have a coordinate that is equal.

Problem 8.55. Let $\mathbb{A}^6$ be the parameter space for homogeneous degree 2 polynomials

$$F = ax^2 + by^2 + cz^2 + dxy + eyz + fzx.$$

Show that the general such polynomial is not a product of two homogeneous degree 1 polynomials.

Hint: the hard part is to find a nontrivial equation in the six coefficients that is satisfied by any product of two degree 1 polynomials. Consider the *associated symmetric matrix*

$$\begin{pmatrix} 2a & d & f \\ d & 2b & e \\ f & e & 2c \end{pmatrix}.$$

Show that if F is a product of degree 1 polynomials, then this matrix is not invertible. Conceptually, this matrix arises in the computation of the gradient of F: we have

$$\nabla F = \begin{pmatrix} \frac{\partial F}{\partial x} \\ \frac{\partial F}{\partial y} \\ \frac{\partial F}{\partial z} \end{pmatrix} = \begin{pmatrix} 2a & d & f \\ d & 2b & e \\ f & e & 2c \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Problem 8.56. (1) Show that a general affine linear transformation $T : \mathbb{A}^n \to \mathbb{A}^n$ has a **unique** fixed point.

(2) Explicitly describe the subset of $\mathbb{A}^2$ corresponding to affine linear transformations $T : \mathbb{A}^1 \to \mathbb{A}^1$ that do not have a unique fixed point. Show that this set is not closed.

Thus, this is a natural example of a general property that holds on an open dense subset, even though the set of points where the property holds is not open.

Problem 8.57. This problem is often useful for computing closures. Let $S \subset \mathbb{A}^n$ be a subset, and let $F : \mathbb{A}^1 \to \mathbb{A}^n$ be a polynomial map. Suppose that $F(t) \in S$ for infinitely many $t \in \mathbb{A}^1$. Show that $F(t) \in \overline{S}$ for all $t \in \mathbb{A}^1$.

Problem 8.58. Let $k \leq \min\{m, n\}$ and let

$$M_k = \{A : \operatorname{rk} A \leq k\} \subset \operatorname{Mat}_{m\times n}(K).$$

Let

$$S_k = \{A : \operatorname{rk} A = k\} \subset \operatorname{Mat}_{m\times n}(K).$$

Show that the closure of S_k is M_k.

Problem 8.59. Let $p_1, \ldots, p_n \in K$ be distinct, and let $q_1, \ldots, q_n \in K$ be arbitrary. Show that if $d \geq n - 1$, there is a polynomial $f \in K[x]$ of degree at most d such that $f(p_i) = q_i$ for all i.

Problem 8.60. Let $p_1, \ldots, p_n \in \mathbb{A}^n$. Show that there is a hyperplane in $\mathbb{A}^n$ which contains $p_1, \ldots, p_n$.

Problem 8.61. Let V and W be vector spaces of dimension n. Show that there is an isomorphism $T : V \to W$.

Problem 8.62. Let $V = \operatorname{Mat}_{n\times n}(K)$, and let $M_1 \subset V$ be the subset of matrices of rank at most 1. Show that

$$\operatorname{span}(M_1) = V.$$

Problem 8.63. Consider the following collection of 6 points in $\mathbb{A}^2$:

$$Z = \{(1, 0), (2, 0), (3, 0), (4, 0), (0, 1), (0, 2)\}.$$

Compute the Hilbert function $h_Z(d)$.

Problem 8.64. Show that if $Z \subset \mathbb{A}^n$ is a general collection of $n+1$ points, then they do not lie on a hyperplane.

Problem 8.65. Let $X \subset \mathbb{A}^N$ be closed, and let $F : X \to \mathrm{Mat}_{m\times n}(K)$ be a matrix-valued polynomial map. Show that the function $\lambda : X \to \mathbb{Z}$ defined by $\lambda(x) = \dim \ker F(x)$ is upper-semicontinuous.

Problem 8.66. Show that if $Z \subset \mathbb{A}^2$ is a general collection of 5 points, then they lie on a unique conic.

9. Topics for further study

A component of the final exam for the original MASS course was to complete a self-guided project on a topic related to the course. Here is a discussion of topics that were suggested for the project.

Difficulty. The proposed projects below are given approximate difficulty rankings from * to ****. Since you all have widely ranging levels of preparation for the course, the most important thing is to pick a project that will be a beneficial learning experience, and I hope this can guide you.

Topics in Linear Algebra

Polynomial approximation. (*-**) We have concerned ourselves with finding polynomials which interpolate points *exactly*. This is an interesting problem in algebra, but in applied areas we often want to fit a small degree polynomial to a very large data set. The data in such applications is often "fuzzy," so the data won't actually lie on a small degree polynomial, even though it might be very close to one. Therefore from an applied standpoint it doesn't make sense to try and solve the interpolation problem exactly; instead we should look for approximate solutions.

Least squares minimization for prescribed points. If $x_i \in \mathbb{R}$ are points and $y_i \in \mathbb{R}$ are values, we might try to find a polynomial f of degree $\leq d$ that minimizes a sum

$$\sum_i (f(x_i) - y_i)^2.$$

By developing the theory of inner product spaces in linear algebra, this problem has a very clean solution.

Least squares minimization for a prescribed function. If $g : [0, 1] \to \mathbb{R}$ is an integrable function, we can try to find a polynomial f of degree $\leq d$ that minimizes the integral

$$\int_0^1 (f(x) - g(x))^2 \, dx.$$

Again, inner product spaces in linear algebra make this easy.

Stone-Weierstrass theorem. If $g : [0, 1] \to \mathbb{R}$ is a continuous function, we might try to find a polynomial f such that

$$\max_{x \in [0,1]} |f(x) - g(x)|$$

is small. The Stone-Weierstrass theorem says we can make the above maximum as small as we like, so long as we let the degree of f get arbitrarily large. This theorem has more of an analysis flavor.

Possible references. Axler, "Linear algebra done right," Chapter 6. Rudin, "Principles of Mathematical Analysis."

Eigenvalues and Google; the Billion Dollar Eigenvector. (*-**) The natural next topic in linear algebra is eigenvalues and eigenvectors and diagonalization. The original PageRank algorithm that Google used to achieve global domination is based on a simple eigenvector computation. The internet is encoded in a graph, with vertices representing pages and edges representing links. We let a viewer randomly click on links and record what fraction of time they spend at each page, with the hypothesis being that heavily linked pages are more important. This fraction of time spent at each page can be phrased in terms of an eigenvector computation. The result helps you find better cat pictures.

Possible references. Axler, "Linear algebra done right," Chapter 5. Heffron, "Linear Algebra," has an appendix on PageRank.

Topics in Algebraic geometry

Beginning projective geometry. (*-**) We have already seen in class that by considering "points at infinity" we can streamline the exposition of various topics. For example, the nature of a plane conic is determined by its points at infinity.

Projective geometry can be viewed as fixing the (false) theorem that "any two distinct lines in the plane meet at a point." For each equivalence class of parallel lines, we add an "ideal" infinite point at infinity

where those parallel lines all meet. These infinite points form an entire line of points at infinity.

The introduction of homogeneous coordinates makes projective geometry fairly straightforward from an algebraic standpoint. For example, it is easy to algebraically compute the points at infinity on a plane curve.

One goal of this project could be to classify conics in the projective plane up to projective equivalence.

Possible references. Bix, "Conics and cubics," or Shafarevich, "Basic Algebraic Geometry I" (substantially harder).

Intersection multiplicity and Bézout's theorem. (**-***) We've seen that it would be useful to know how many times curves in $\mathbb{A}^2$ of degree d and e should intersect. As a baby theorem, we showed that a curve of degree d and a line intersect in at most d points, unless the curve of degree d contains the line. You also had a homework problem where you saw that if there are five points with no three collinear then there is only one conic that vanishes at all five; in other words, this says two irreducible conics meet in at most 4 points. The important theorem of Bézout says that in $\mathbb{A}^2$ a curve C of degree d and a curve D of degree e intersect at most de times counting multiplicity, unless they have a component curve in common. (If you also learn about projective space, this theorem can be improved: in $\mathbb{P}^2_{\mathbb{C}}$ there are exactly de intersections.)

Possible references. Bix, "Conics and cubics."

Unique factorization domains. (**-***) When studying hypersurfaces $F = 0$ in $\mathbb{A}^n$, it is very useful to be able to factor F into irreducible factors. A simple induction shows that this can always be done, but it is not so clear that the factorization is unique. A classical theorem says that the polynomial ring $\mathbb{R}[x_1, \ldots, x_n]$ in any number of variables is a unique factorization domain, implying that this is the case. The main argument goes by induction on the number of variables. One shows using *Gauss' Lemma* that if a ring R is a UFD then so is $R[x]$. The goal of this project would be to learn a proof of this theorem. Along the way, you will learn about commutative rings, prime and irreducible elements, integral domains, the *content* of a polynomial, and the aforementioned Gauss' Lemma.

Possible references. Dummit & Foote "Abstract Algebra," but we can probably find an easier text as well.

Elliptic curves. (**-***) Elliptic curves are of tremendous importance in both algebraic geometry and number theory, as well as in cryptography. The geometric side of the story can be understood in terms of projective cubic plane curves. If a flex point is placed at infinity, then a group law on C is defined by saying that $P + Q + R = 0$ whenever P, Q, R are collinear. (Some care has to be taken in case there are tangent lines...) A good goal for this project would be to see that this rule (when all the technicalities are resolved) actually defines a group law. In particular, the proof of associativity takes some real work.

Possible references. Bix, "Conics and cubics," and Silverman and Tate, "Rational points on elliptic curves."

Beginning higher-dimensional algebraic geometry. (***-****) As we will see soon in class, an *algebraic variety* is a locus cut out by a *system* (or *ideal*) of defining polynomials.

Pretty quickly there are some algebraic formalities needed to make sense of things. The first major theorem in this area is the *Hilbert basis theorem* which roughly says that although we could consider varieties cut out by *infinitely* many polynomial equations, it is actually the case that only finitely many such equations are needed to describe the same variety. The next major theorem in this area is Hilbert's *Nullstellensatz*, which makes precise the idea that a locus cut out by a system of polynomial equations is essentially the same thing as that system of polynomial equations—just as a curve in the plane is essentially the same thing as its defining equation.

Possible references. Kunz, "Introduction to Commutative Algebra and Algebraic Geometry" and Cox, Little, O'Shea "Ideals, Varieties, and Algorithms," along with many, many others.

Intersection theory and Schubert calculus. (***-****) Suppose I am given 4 "random" lines in 3-dimensional projective space $\mathbb{P}^3$ (or $\mathbb{C}^3$ would be just fine too, to keep things simple.) How many lines in space meet all 4? Answer: 2.

Bézout's theorem (coming up shortly in the course) is the first major theorem in intersection theory: over the complex numbers, two "random" plane curves C and D of degrees d and e, respectively, meet in de points.

More generally, given two geometric loci in some space we can ask what their intersection looks like. If we expect that the intersection is a finite number of points, we can ask how many points there are. Intersection theory makes these notions precise.

The Grassmannian $\mathbb{G}(1,3)$ is a space whose points are the lines in $\mathbb{P}^3$ (i.e. the two-dimensional subspaces of $\mathbb{C}^4$, much like projective space $\mathbb{P}^3$ itself is a space whose points are the 1-dimensional subspaces of $\mathbb{C}^4$. It is a 4-dimensional space, which can be naturally thought of as a subvariety of $\mathbb{P}^5$ defined by one equation. For a line $L \subset \mathbb{P}^3$, there is a locus $\Sigma_L \subset \mathbb{G}(1,3)$ consisting of all the lines that meet L. It is a 3-dimensional subvariety of the Grassmannian. Now given 4 lines $L_1, \ldots, L_4$, what is the intersection

$$\Sigma_{L_1} \cap \cdots \cap \Sigma_{L_4} \subset \mathbb{G}(1,3)?$$

It should consist of all lines that meet $L_1, \ldots, L_4$. Schubert calculus is a remarkable tool which allows us to predict the number of intersection points in such an intersection in an essentially computational way.

Possible references. Harris, "3264 and all that," some old course notes, and some pep talks from me.

Algebraic geometry and manifolds. (**-***) A plane curve C : $F(x,y) = 0$ is *nonsingular* if the gradient vector

$$(\nabla F)(p) = \left(\frac{F}{\partial x}(p), \frac{\partial F}{\partial y}(p) \right)$$

is nonzero at every point $p \in C$. More generally, there is a similar (but more complicated) definition of a nonsingular variety defined by a system of polynomial equations.

Considerable progress in algebraic geometry has been made by realizing that nonsingular algebraic varieties over $\mathbb{R}$ or $\mathbb{C}$ are also manifolds; that is, they locally look like a Euclidean space. A reasonable goal for this project is to learn what a manifold is and learn enough of the topic to show that a nonsingular algebraic variety is actually a manifold.

Guillemin and Pollack, "Differential Topology" for the differential geometry side of things. Shafarevich, "Basic Algebraic Geometry I," for the algebraic geometry side of things.

Applications of algebraic geometry

Elliptic curves and Fermat's Last Theorem. (*-****) Fermat's Last Theorem is the statement that the equation

$$x^n + y^n = z^n$$

has no solutions in positive integers x, y, z if $n \geq 3$. This theorem, originally conjectured by Fermat, was proved 358 years later by Andrew Wiles in 1994. A key part of Wiles' proof makes use of an elliptic (think cubic) curve closely related to this equation, and its group law.

This project could go in many directions of wildly varying difficulty. A good goal from a mathematical standpoint would be to understand how the theorem is related to elliptic curves. For this, the survey article Cox, "Introduction to Fermat's Last Theorem" might be appropriate.

A lighter goal would be to read some "popular" math on the topic. The book "Fermat's Enigma" by Simon Singh is a wonderful read on the history of the problem, and a great discussion of what being a mathematician is all about; in my case it was influential in my decision to become a mathematician. Since this book is a fairly light read, ideally the project writeup would go a bit beyond the Singh book and incorporate some more mathematical detail, such as from the Cox survey. There is additionally a very interesting NOVA documentary on the subject, loosely along the lines of the Singh book.

Possible references. Cox, "Introduction to Fermat's Last Theorem" and Singh, "Fermat's Enigma," as well as many others.

Elliptic curves and factorization of numbers. (**-***) One of the main methods used for factoring really large integers is the so-called Lenstra elliptic curve factorization method. This algorithm uses the group law on an elliptic curve defined over a finite field (think: the integers mod p) to efficiently factor large numbers.

Possible references. Silverman and Tate, "Rational Points on Elliptic Curves."

Elliptic curve cryptography. (**-***) Elliptic curves have also been recently used to make more efficient public-key cryptography possible. The background for this project overlaps considerably with the previous one.

Possible references. Silverman and Tate, "Rational Points on Elliptic Curves."

Algebraic statistics and computational biology. (***-****) The field of algebraic statistics—which is roughly a mashup of statistics and algebraic geometry—has recently had great applications to computational biology and, in particular, phylogenetics. A goal for this project would be to make first contact with some of these connections. Possible references include Part I of Pachter and Sturmfels,

Possible references. Part I of Pachter and Sturmfels, "Algebraic statistics for computational biology" (probably challenging) or various sets of lecture notes, including Hosten and Ruffa, "Introductory Notes to Algebraic Statsitics," among others.