Commutative Algebra in Context
Course Notes

Fall 2020
Drew Armstrong

Contents

Introduction

August 17: Examples of Curves

To begin this course we will consider equations of the form

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

where $f(x, y) \in \mathbb{Z}[x, y]$ is a polynomial in two variables with integer coefficients. What questions about this equation might be interesting?

- Find all positive integer solutions.
- Find all integer solutions.
- Find all rational solutions.

These are called *Diophantine problems* and they are extremely hard. For example, Fermat's Last Theorem (proved by Wiles, 1994) says that for $n \geq 3$ the equation $x^n + y^n - 1 = 0$ has no rational solution $(x, y) \in \mathbb{Q}^2$ except when $xy = 0$. [Equivalently, the homogeneous equation $x^n + y^n - z^n = 0$ has no integer solution $(x, y, z) \in \mathbb{Z}^3$ except when $xyz = 0$.] Hilbert's 10th problem asked for an algorithm to determine whether an equation such as $f(x, y) = 0$ has an integer solution. Matiyasevich (1970) proved that no general algorithm exists.

One reason that Diophantine problems are hard is because we lack tools to study them. If we work over the real numbers then we can use tools from geometry and physics. The real solutions to $f(x, y) = 0$ form a curve in the real 2D plane $\mathbb{R}^2$. What questions about this curve might be interesting?

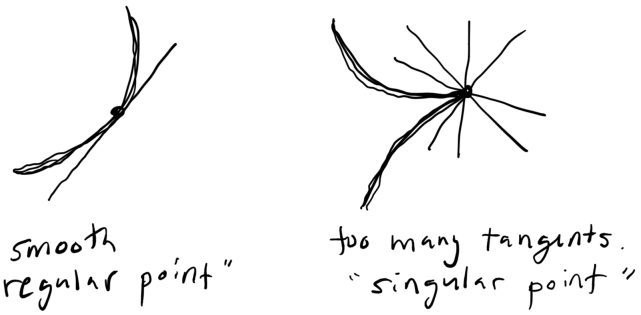
- Find and classify the singular points.
- Find and classify inflection points, double tangents, etc.
- How many connected components?

If $f(a, b) = 0$ for some $(a, b) \in \mathbb{R}^2$, then the *tangent line* at this point is given by

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

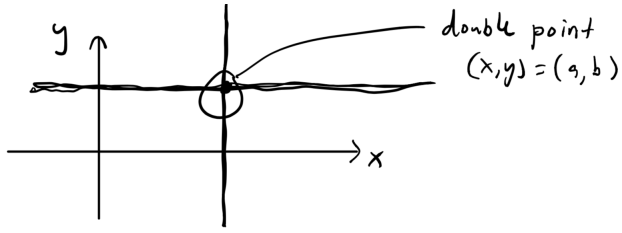
If $(\partial f/\partial x)(a, b) = (\partial f/\partial y)(a, b) = 0$ then the tangent line is not well-defined.[1] In this case we say that (a, b) is a *singular point* of the curve. Picture:

[1] You could also say that the *tangent space* is two dimensional.

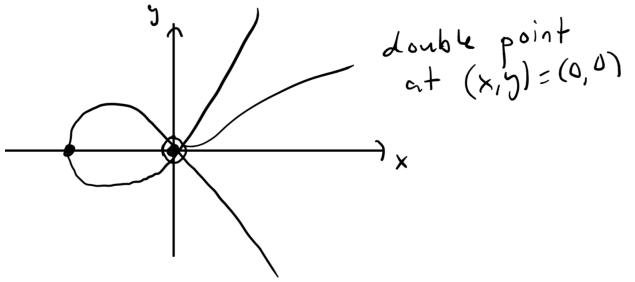


Examples:

(a) The curve $(x-a)(y-b)=0$ has a singular point at $(x,y)=(a,b)$. Picture:

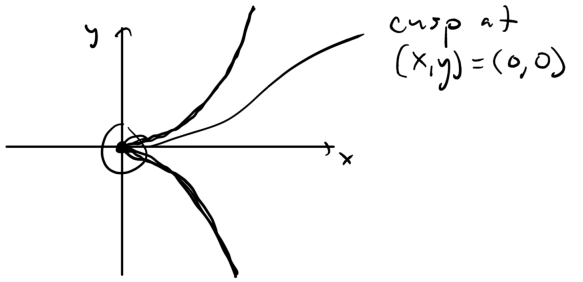


(b) The curve $y^2-x^2(x+1)=0$ has a singular point at $(x,y)=(0,0)$. Picture:



The singularities (a) and (b) are called *double points*.

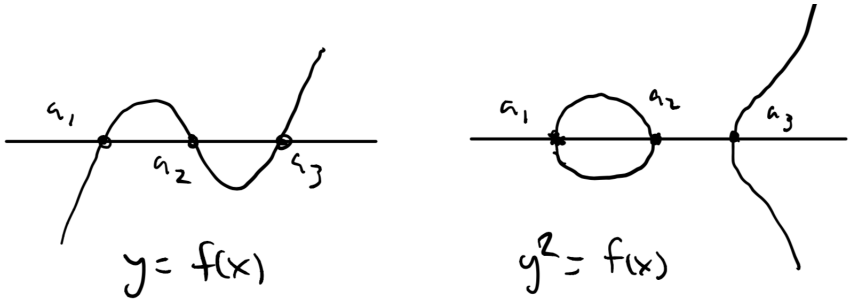
(c) The curve $y^2-x^3=0$ has a singularity at $(x,y)=(0,0)$. This singularity is called a *cusp*. Picture:



Equivalence of singularities is difficult to define. For now let me just tell you that double points and cusps are the two simplest equivalence classes.

A curve with no singular points is called *smooth*. Example:

(d) If $f(x) \in \mathbb{R}[x]$ is a polynomial of degree 3 with no multiple roots then the curve $y^2 = f(x)$ has no singularities. Such a curve is called *elliptic*. Proof: If $(a, b) \in \mathbb{R}^2$ is a singular point of the curve then we must have $b^2 = f(a)$ and $2b = f'(a) = 0$. This implies that $f(a) = f'(a) = 0$ and hence a is a multiple root of $f(x)$. /// Picture of $y^2 = (x - a_1)(x - a_2)(x - a_3)$:



August 19: Some Famous Theorems

We begin with some formalities about polynomials.

Let $f(x, y) \in \mathbb{Z}[x, y]$ be a polynomial in two variables with integer coefficients. Explicitly, this means that

$$f(x, y) = \sum_{i,j \geq 0} a_{ij} x^i y^j, \quad a_{i,j} \in \mathbb{Z} \text{ and } a_{i,j} = 0 \text{ for all but finitely many } i, j.$$

The *degree* of a monomial $a_{i,j} x^i y^j$ with $a_{ij} \neq 0$ is defined to be $i + j$. Then the *degree* of a nonzero polynomial $f(x, y)$ is defined as the maximal degree of monomials that occur with nonzero coefficient:

$$\deg(f) = \max\{i + j : a_{ij} \neq 0\}.$$

Note that nonzero constants have degree 0 and the degree of the zero polynomial is undefined. For any two monomials $m = ax^i y^j$ and $n = bx^k x^\ell$ with $a, b \neq 0$ we note that $ab \neq 0$ and hence

$$\deg(mn) = \deg(abx^{i+k} y^{j+\ell}) = (i + k) + (j + \ell) = (i + j) + (k + \ell) = \deg(m) + \deg(n).$$

It follows from this that for any two nonzero polynomials $f, g \in \mathbb{Z}[x, y]$ we have

$$\deg(fg) = \deg(f) + \deg(g).$$

For any ring $(R, 0_R, 1_R, +, \times)$ one can check that there exists a unique ring homomorphism $\iota_R : \mathbb{Z} \to R$. Namely,

$$\iota(n) := \begin{cases} 1_R + \cdots + 1_R & n > 0, \\ 0_R & n = 0, \\ -(1_R + \cdots + 1_R) & n < 0. \end{cases}$$

Then for any elements $\alpha, \beta \in R$ one can check that there exists a unique ring homomorphism $\iota_{a,b} : \mathbb{Z}[x,y] \to R$ sending $x \mapsto \alpha$ and $y \mapsto \beta$. Namely,

$$\iota_{\alpha,\beta}\left(\sum_{i,j\geq 0} a_{i,j}x^i y^j\right) = \sum_{i,j\geq 0} a_{i,j}\alpha^i\beta^j \in R.$$

This homomorphism is called *evaluation at the point* $(\alpha,\beta) \in R^2$. Less formally, it is convenient to denote the function $\iota_{\alpha,\beta}$ by

$$\mathbb{Z}[x,y] \ni f(x,y) \mapsto f(\alpha,\beta) \in R.$$

This formality allows us to be more precise about the concepts discussed last time. For any polynomial $f(x,y) \in \mathbb{Z}[x,y]$ we can define an "abstract curve" C_f. What kind of a thing is an abstract curve? You can think of it as a recipe that assigns to each ring R the set of "R-points of the curve":

$$C_f(R) := \{(\alpha,\beta) \in R^2 : f(\alpha,\beta) = 0\} \subseteq R^2.$$

What kind of questions about C_f might be interesting? Diophantine problems ask us to describe the $\mathbb{Z}$-points or the $\mathbb{Q}$-points of the curve. Analytic geometry asks us to describe the $\mathbb{R}$-points of the curve. Modern number theory asks us to describe the $\mathbb{F}_q$-points of the curve, where $\mathbb{F}_q$ is the finite field of size q (some power of a prime). Note that these problems all look very different. However, I claim that there is a deep structure underlying them all.

Miracle. For any $f(x,y) \in \mathbb{Z}[x,y]$, there is an integer $g \geq 0$, called the *genus of the abstract curve* C_f, which influences the structure of the R-points over any ring.

The nature of this influence is quite subtle. Let me sketch out some of the major theorems in this direction. To make the following statements as clean as possible we will assume that $f(x,y)$ is irreducible in $\mathbb{C}[x,y]$ and has no singular points in $\mathbb{C}^2$. We should also include "points at infinity." Don't worry, I'll define these notions soon.

- $\mathbb{Z},\mathbb{Q}$-points. Assume that $f(x,y)$ is irreducible over $\mathbb{Z}$.
  - $g = 0$: For any point of $C_f(\mathbb{Q})$ we obtain a bijection $C_f(\mathbb{Q}) \leftrightarrow \mathbb{Q}$ sending the given point to the point at infinity. The set $C_f(\mathbb{Z})$ is harder but still completely understood. It is related to continued fractions and Pell's equation ($x^2 - ny^2 = \pm 1$).
  - $g = 1$: (Mordell, 1922): $C_f(\mathbb{Q})$ is a finitely generated abelian group.
  - $g \geq 1$: (Siegel, 1929): $C_f(\mathbb{Z})$ is a finite set.
  - $g \geq 2$: (Faltings, 1983): $C_f(\mathbb{Q})$ is a finite set.
- $\mathbb{R}$-points.
  - $g = 0$: $C_f(\mathbb{R})$ is a conic section.

- $g = 1$: $C_f(\mathbb{R})$ is equivalent to $y^2 = (x - a_1)(x - a_2)(x - a_3)$ with a_1, a_2, a_3 distinct.
- (Harnack, 1876): # connected components of $C_f(\mathbb{R})$ is $\leq g + 1$.

- $\mathbb{F}_q$-points. If q is a power of p then we should also assume that the coefficients of $f(x, y)$ are not divisible by p.
  - $g = 0$: $\#C(\mathbb{F}_q) = q + 1$
  - $g = 1$: (Hasse, 1933): $|\#C(\mathbb{F}_q) - (q + 1)| \leq 2\sqrt{q}$
  - $g \geq 2$: (Weil, 1949): $|\#C(\mathbb{F}_q) - (q + 1)| \leq 2g\sqrt{q}$

This last theorem is a consequence of the Riemann hypothesis for curves, which was proved by Andre Weil while he was a prisoner of war. It is the easiest example of the famous Weil conjectures, which inspired the development of much of modern commutative algebra.

Aug 21: More Famous Theorems

So what **is** the genus g? And how can we compute it? The key is to consider the **complex points** of the curve.

- $\mathbb{C}$-points. Let $f(x, y) \in \mathbb{C}[x, y]$ have degree d.
  - (Newton): A line meets the curve C_f in at most d distinct points.
  - (Maclaurin, Bézout): A curve of degree n meets C_f in at most nd distinct points.
  - (Plücker, 1830s): There exists an integer $g \geq 0$ with the following property. Let n be large. Then for $nd - g$ distinct points on C_f there exists a curve of degree n intersecting C_f at exactly these points. However, for $nd - g + 1$ distinct points on C_f there is no such curve of degree n. This g was called the "deficiency" of C_f. If the curve C_f has only double points and cusps then the deficiency satisfies
$$g = \frac{(d-1)(d-2)}{2} - \#\text{ double points} - \#\text{ cusps}.$$
  - Modern version of Plücker (Serre, 1950s): If $\alpha = (a, b) \in \mathbb{C}^2$ is a singular point of C_f then $f^\alpha(x, y) := f(x + a, y + b)$ has a singular point at $(0, 0)$. Then the ring $\mathbb{C}[[x, y]]/(\partial f^\alpha/\partial x, \partial f^\alpha/\partial y)$ has even finite dimension $2\delta_\alpha$ as a complex vector space, and the genus is given by
$$g = \frac{(d-1)(d-2)}{2} - \sum_\alpha \delta_\alpha,$$
    where the sum is over singular points α. This generalizes Plücker's formula because $\delta_\alpha = 1$ when α is a double point or cusp.
  - (Abel, 1820s): An "abelian integral" has the form $A(a) = \int_{a_0}^{a} r(x, y)\, dx$ where $r(x)$ is a rational function and y is defined implicitly by $f(x, y) = 0$.[2] There

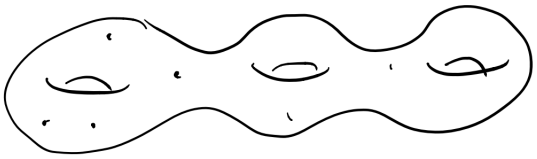
[2]Example: For $f(x, y) = y^2 + x^2 - 1$ and $r(y) = 1/y$ we have $A(a) = \arcsin(a)$.

exists an integer $g \geq 0$ depending on $f(x,y)$ with the following property. For any $a_1, \ldots, a_k \in \mathbb{C}$ there exist $b_1, \ldots, b_g \in \mathbb{C}$ such that[3]

$$A(a_1) + \cdots + A(a_k) = A(b_1) + \cdots + A(b_g) + e(a_1, \ldots, a_k)$$

where e is some elementary function.

– (Riemann, 1850s, Möbius, 1863): We can think of $C_f(\mathbb{C})$ as a compact orientable surface (a Riemann surface) with finitely many points deleted (the singular points and the points at infinity). The genus g is the number of handles:



g = # handles.

– Riemann's version of Abel: The genus g is equal to the dimension of the vector space of holomorphic 1-forms on the Riemann surface.

It is amazing that all of these definitions of genus coincide, and that they have anything do with the R-points of the curve for $R \in \{\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{F}_q\}$.

Points at Infinity

Aug 24: Equivalence of Curves

Our next job is to define "points at infinity." We can motivate this by first considering equivalence of curves.

Given two polynomials $f(x,y)$ and $g(x,y)$ in $\mathbb{Z}[x,y]$ we consider the real algebraic curves $C_f(\mathbb{R})$ and $C_g(\mathbb{R})$, which are subsets of $\mathbb{R}^2$. I propose that we should consider $C_f(\mathbb{R})$ and $C_g(\mathbb{R})$ to be "the same curve" if they differ by a translation; that is, if $f(x,y) = g(x+r, y+t)$ for some real numbers $(r,t) \in \mathbb{R}^2$. Furthermore, it seems clear that we should allow rotations and reflections:

$$f(x,y) = g(x\cos\theta \mp y\sin\theta + r, x\sin\theta \pm y\cos\theta + t).$$

In this case we say that $C_f(\mathbb{R})$ and $C_g(\mathbb{R})$ are equivalent up to Euclidean symmetries of the plane $\mathbb{R}^2$.

Example: The curves $f(x,y) = x^2 - y^2 = 0$ and $g(u,v) = uv = 0$ are equivalent under a 45° rotation $(u,v) = (x-y, x+y)/\sqrt{2}$. More generally, a classical theorem says that any degree 2 real curve

$$f(x,y) = \alpha x^2 + \beta xy + \gamma y^2 + \delta x + \varepsilon y + \lambda = 0$$

[3]Example: $\arcsin(a) + \arcsin(b) = \arcsin(c)$ where $c = a\sqrt{1-b^2} + b\sqrt{1-a^2}$.

that is nondegenerate (i.e., $\beta^2 - 4\alpha\gamma \neq 0$) is equivalent under translation and rotation to a curve of the form

$$g(u, v) = au^2 + bv^2 + c = 0,$$

i.e., an ellipse or a hyperbola (or a single point or the empty set). ///

However, since rotation by 45° does not preserve rational points $\mathbb{Q}^2 \subseteq \mathbb{R}^2$, maybe we should also allow transformations such as $(u, v) = (x - y, x + y)$. In which case, we will also obtain a bijection between rational points of $x^2 - y^2 = 0$ and $uv = 0$.

We define an *affine transformation* $F := \mathbb{R}^2 \to \mathbb{R}^2$ as

$$F\begin{pmatrix} u \\ v \end{pmatrix} := \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} r \\ t \end{pmatrix} = \begin{pmatrix} ax + by + r \\ cx + dy + t \end{pmatrix}$$

The transformation $F(x, y) = A(x, y) + (r, t)$ is invertible if and only if the matrix A is invertible ($ad - bc \neq 0$), in which case the inverse is the affine transformation $F^{-1}(x, y) = A^{-1}(x, y) - A^{-1}(r, t)$. We say that real curves $C_f(\mathbb{R})$ and $C_g(\mathbb{R})$ are *affinely equivalent* if there exists an invertible affine transformation $F : \mathbb{R}^2 \to \mathbb{R}^2$ such that $F(C_f(\mathbb{R})) = C_g(\mathbb{R})$.

Should we allow more general kinds of transformations? In other words, in which category do the algebraic curves live? Here are some possible answers: We can say that $C_f(\mathbb{R}) \cong C_g(\mathbb{R})$ if $F(C_f(\mathbb{R})) = C_g(\mathbb{R})$ for some function $F : \mathbb{R}^2 \to \mathbb{R}^2$ that is a

- homeomorphism,
- diffeomorphism of class C^k (all kth partials are continuous),
- smooth isomorphism (all partials are continuous),
- real analytic isomorphism (power series converge).

Each of these is more restrictive (more "rigid") than the last. For example, the function $\mathbb{R} \to \mathbb{R}$ defined by $x \mapsto e^{-1/x^2}$ is smooth (all derivatives exist and are continuous), but it is not real analytic because the power series at $x = 0$ is $0 + 0x + 0x^2 + 0x^3 + \cdots$ which has zero radius of convergence. Picture:

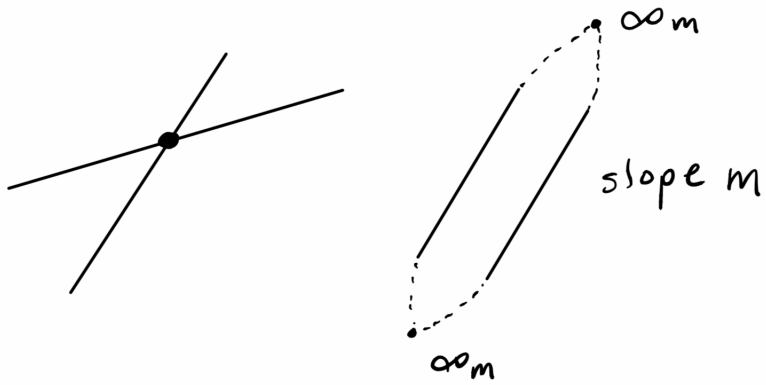
However, even real analytic geometry is still too general for us, because it does not preserve the property of "being defined by polynomials." But I claim that affine equivalence is not quite general enough to build an interesting theory of curves.

Next time I will define a notion of equivalence for algebraic curves that turns out to be just right. The key idea is to define points at infinity.

Aug 26: Points at Infinity

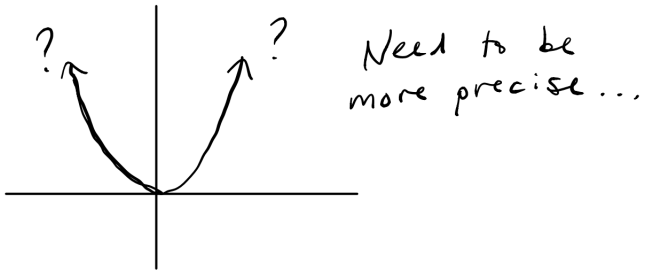
Projective geometry emerged from the theory of perspective drawing, often attributed to Brunelleschi in 1410. The systematic study of projective geometry began with Poncelet's *Treatise on the projective properties of figures* (1822), based on work that he did in a Russian prison camp. Möbius (1827) introduced coordinates into projective geometry and then Plücker (1830s) applied these to the study of algebraic curves. Today I will present the modern system of "homogeneous coordinates" for projective geometry of the plane.

Idea: There is a line of points at infinity corresponding to slopes. Any two lines of the same slope $m \in \mathbb{R} \cup \{\infty\}$ intersect at the point "∞_m".Thus the projective plane has the property that any two non-equal lines meet at a unique point:



It follows from this that a hyperbola is really a connected loop with two points at infinity:

What about a parabola?



To understand the behavior of a parabola at infinity we need to be more precise. Define the *real projective plane* as the set of nonzero points in $\mathbb{R}^3$ modulo scalar multiplication:

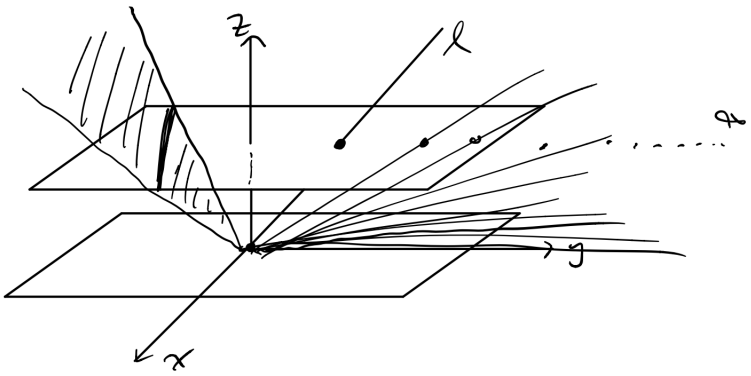
$$\mathbb{RP}^2 := \left(\mathbb{R}^3 \setminus \{(0,0,0)\}\right) / (\text{nonzero scalars}).$$

In other words, we have $(x, y, z) \sim (x', y', z')$ if and only if $(\lambda x, \lambda y, \lambda z) = (x', y', z')$ for some $\lambda \neq 0$. Let $(x : y : z)$ denote the equivalence class of the point (x, y, z).

If $z \neq 0$ then we have $(x : y : z) = (x : y : 1)$ for unique values of $x, y \in \mathbb{R}$. In this case we say that $(x : y : 1) \in \mathbb{RP}^2$ is a *finite point*. Thus we have a bijection $(x : y : 1) \leftrightarrow (x, y)$ between finite points of $\mathbb{RP}^2$ and all points of $\mathbb{R}^2$.

On the other hand, points of the form $(x : y : 0)$ are called *points at infinity*. We observe that $(x : y : 0) = (x' : y' : 0)$ if and only if $xy' = x'y$, so that points at infinity are in bijection with slopes $y/x \in \mathbb{R} \cup \{\infty\}$. The points at infinity corresponding to horizontal and vertical slopes are $(1 : 0 : 0)$ and $(0 : 1 : 0)$, respectively.

This is often visualized as follows:



We can view the finite points of $\mathbb{RP}^2$ as the points $(x, y, 1)$ in $\mathbb{R}^3$. Each such point determines a unique non-horizontal line in $\mathbb{R}^3$ through the origin. Points at infinity correspond to horizontal lines through the origin in $\mathbb{R}^3$. Furthermore, we can view *lines* in $\mathbb{RP}^2$ as the intersection of the plane $(x, y, 1)$ with a plane of the form $ax + by + cz = 0$. The unique such plane $z = 0$ that does not intersect $z = 1$ is called the *line at infinity*. [This picture is not fully satisfying, but it should give you enough hope to continue reading for a few paragraphs.]

So much for lines. Let's consider how to define higher degree curves in the projective plane. We say that a polynomial $F(x, y, z) \in \mathbb{R}[x, y, z]$ is *homogeneous of degree* d if we have

$$F(\lambda x, \lambda y, \lambda y) = \lambda^d F(x, y, z) \quad \text{for all } \lambda \in \mathbb{R}.$$

Warning: Such a polynomial **does not define a function** $F : \mathbb{RP}^2 \to \mathbb{R}$. However, it is true that the equation

$$F(x, y, z) = 0$$

is preserved by scaling, hence it defines a subset $C_F \subseteq \mathbb{RP}^2$, called the *projective curve* corresponding to $F(x, y, z)$.

Setting $z = 1$ gives the equation

$$f(x, y) := F(x, y, 1) = 0,$$

which defines a curve $C_f \subseteq \mathbb{R}^2 \subseteq \mathbb{RP}^2$ in the affine part of the plane. Observe that we have $C_f \subseteq C_F$. We say that the set $C_F \setminus C_f$ consists of the *infinite points of the curve* C_f.

To see that this notion is well-defined, let $f(x, y) \in \mathbb{R}[x, y]$ be any polynomial of degree d. Then we define its *homogenization* as

$$F(x, y, z) := z^d f(x/z, y/z).$$

We observe that $F(x, y, z)$ is a homogeneous polynomial of degree d with $F(x, y, 1) = f(x, y)$. Thus the curve $C_f \subseteq \mathbb{R}^2$ extends to $C_f \subseteq C_F \subseteq \mathbb{RP}^2$ in a unique way. This is called the *projective completion* of the curve.

Examples:

- **Hyperbola.** Let $f(x, y) = x^2 - y^2 - 1$. The homogenization is

  $$F(x, y, z) = z^2 f(x/z, y/z) = x^2 - y^2 - z^2.$$

  The projective curve C_F consists of the finite points C_f together with the points at infinity $(x : y : 0)$ satisfying the following equation:

  $$\begin{aligned} F(x, y, 0) &= 0 \\ x^2 - y^2 &= 0 \\ (x - y)(x + y) &= 0. \end{aligned}$$

  In other words, we have two points at infinity corresponding to slopes $y/x = \pm 1$.

- **Parabola.** Let $f(x, y) = x^2 - y$. The homogenization is

  $$F(x, y, z) = z^2 f(x/z, y/z) = x^2 - yz.$$

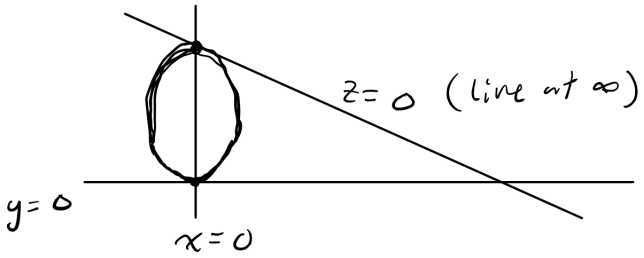
  The projective curve C_F consists of the finite points C_f together with the points at infinity $(x : y : 0)$ satisfying the following equation:

  $$\begin{aligned} F(x, y, 0) &= 0 \\ x^2 &= 0. \end{aligned}$$

This means that the point $(0:1:0)$ (vertical slope) is a *double point* at infinity. In fact, I claim that the projective curve C_F is **tangent to the point at infinity**. To see this, we note that there is nothing special about the line $z=0$ in $\mathbb{RP}^2$. We could equally well allow any line in $\mathbb{RP}^2$ to serve the role of the "line at infinity." For example, if we say that $y=0$ is the line at infinity then we obtain the de-homogenization

$$g(x,z) := F(x,1,z) = x^2 - z,$$

and we observe that the curve C_g is tangent to the line $z=0$ in the finite x,z-plane. Sometimes you will see this depicted as follows:



We will interpret this picture next time.

Aug 28: Projective Equivalence

Last time we defined the real projective plane:

$$\mathbb{RP}^2 := \left(\mathbb{R}^3 \setminus \{(0,0,0)\}\right) / (\text{nonzero scalars}).$$

We visualized this as the plane $(x,y,1)$ in $\mathbb{R}^3$ (consisting of "finite points") together with the "line at infinity" consisting of slopes $(x:y:0)$. Today we will use a different visualization.

Note that $(x,y,z) \sim (x',y',z')$ if and only if these points are on the same line through the origin in $\mathbb{R}^3$. Thus we have a bijection

$$\mathbb{RP}^2 \leftrightarrow (\text{lines through the origin in } \mathbb{R}^3).$$

This emphasizes that all points of $\mathbb{RP}^2$ "look the same."[4] By intersecting each line through $\mathbf{0} \in \mathbb{R}^3$ with the unit sphere $S^2 \subseteq \mathbb{R}^3$ we obtain a bijection

$$\mathbb{RP}^2 \leftrightarrow S^2/(\text{antipodal map}),$$

and in this language we have

$$(\text{points in } \mathbb{RP}^2) \leftrightarrow (\text{pairs of antipodal points in } S^2),$$
$$(\text{lines in } \mathbb{RP}^2) \leftrightarrow (\text{great circles in } S^2).$$

[4]Technically: $\mathbb{RP}^2$ is called a *homogeneous space*.

Furthermore, we observe that this sets up a bijection between points and lines in $\mathbb{RP}^2$ because each pair of antipodal points (north and south pole) in S^2 corresponds to a unique great circle in S^2 (the equator):

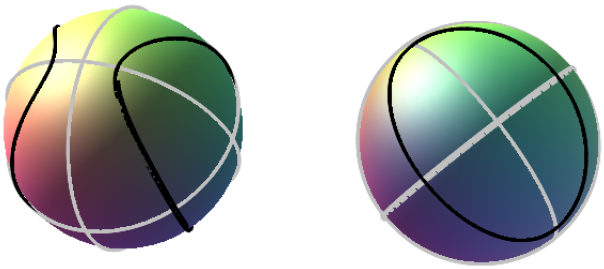
$$(\text{points in } \mathbb{RP}^2) \leftrightarrow (\text{lines in } \mathbb{RP}^2).$$

Algebraically: To each point $(a : b : c)$ we associate the line $ax + by + cz = 0$. Let us denote this line by $[a : b : c]$. Then we observe that point $(a : b : c)$ is contained in line $[a' : b' : c']$ if and only if point $(a' : b' : c')$ is contained in line $[a : b : c]$. This is called *point-line duality*. [In higher dimensions there is a duality between points and hyperplanes.]

This model of $\mathbb{RP}^2$ gives us a clearer picture of projective curves. For any homogeneous polynomial $F(x, y, z) \in \mathbb{R}[x, y, z]$, the curve $C_F \subseteq \mathbb{RP}^2$ can be viewed as the intersection of the surface $F(x, y, z) = 0$ in $\mathbb{R}^3$ with the unit sphere $S^2 \in \mathbb{R}^3$. From this point of view the lines $x = 0$ and $y = 0$ and $z = 0$ are mutually perpendicular and serve as the projective axes.

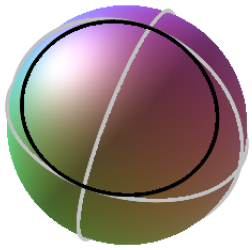
Examples:

- **Hyperbola.** Let $f(x, y) = x^2 - y^2 - 0.1$ so that $F(x, y, z) = x^2 - y^2 - 0.1z^2$. Here are two views of the projective curve:



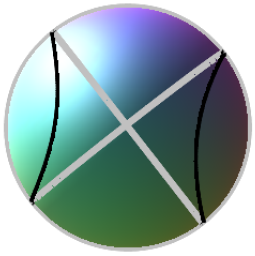
  We observe that the hyperbola in the $z = 0$ chart becomes an ellipse in the $x = 0$ chart. Algebraically, by setting $x = 1$ we obtain $F(1, y, z) = 1 - y^2 - 0.1z^2 = 0$, or $y^2 + z^2/10 = 1$, which is an ellipse in the finite y, z-plane.

- **Parabola.** Let $f(x, y) = x^2 - 5y$ so that $F(x, y, z) = x^2 - 5yz$. Here is a view of the projective curve:



  We observe that the parabola $y = x^2/5$ in the finite x, y-plane is tangent to the line $z = 0$ at infinity. Indeed, sending $y = 0$ to infinity gives $F(x, 1, z) = x^2 - 5z = 0$, which

is a parabola $z = x^2/5$ in the finite x, z-plane. Out of curiosity, what happens if we send $x = 0$ to infinity? We get $F(1, y, z) = 1 - 5zy = 0$, which is a hyperbola $yz = 1/5$ in the finite y, z-plane. Projective view:



From these pictures it seems clear that ellipses, hyperbolae and parabolae are all the same when viewed in the projective plane. Let's make this precise.

A function $\mathbb{RP}^2 \to \mathbb{RP}^2$ is called a *projective transformation* if it is induced by an invertible linear function $\mathbb{R}^3 \to \mathbb{R}^3$. Naturally, two linear functions $\mathbb{R}^3 \to \mathbb{R}^3$ determine the same projective transformation $\mathbb{RP}^2 \to \mathbb{RP}^2$ if and only if they differ by a nonzero constant multiple. More precisely, let PGL_3 denote the group of projective transformations. Then we have a surjective group homomorphism $\mathrm{PGL}_3 \to \mathrm{GL}_3$ with kernel given by scalar matrices $\{\lambda I : \lambda \neq 0\}$, and we can identify view this as a quotient group:

$$\mathrm{PGL}_3 = \mathrm{GL}_3/(\text{nonzero scalar matrices}).$$

Exercise: Let $(u, v, w) = \Phi(x, y, z)$ where $\Phi \in \mathrm{PGL}_3$ is any projective transformation. For any homogeneous polynomial $F(x, y, z) \in \mathbb{R}[x, y, z]$ of degree d, prove that the polynomial $G(x, y, z) := F(u, v, w) \in \mathbb{R}[x, y, z]$ is also homogeneous of degree d.

In this case we say that the projective curves $C_F, C_G \subseteq \mathbb{RP}^2$ are *projectively equivalent*. Furthermore, for any affine curves $C_f, C_g \subseteq \mathbb{R}^2$ we say that C_f, C_g are *projectively equivalent* if and only if their (unique) projective completions are projectively equivalent.

This is easiest to visualize when $\Phi : \mathbb{R}^3 \to \mathbb{R}^3$ is a **rotation**. Then equivalence corresponds to rotation of the sphere model of $\mathbb{RP}^2$ and we declare that the affine views of the curve in any hemisphere are equivalent. For example, if $F(x, y, z)$ is homogeneous then the following three affine curves in the x, y-plane are projectively equivalent:

$$F(x, y, 1) = 0, \quad F(1, x, y) = 0, \quad F(y, 1, x) = 0.$$

Exercise: Find the rotation matrix that permutes these curves. More generally, we can send any affine line $ax + by + c = 0$ to infinity by rotating the plane $ax + by + cz = 0$ onto the plane $z = 0$ in $\mathbb{R}^3$. The matrix of this rotation is hard to describe.

Aug 31, Sept 2: The Fundamental Theorem

We defined affine and projective equivalence of curves in terms of matrices. What is the geometric motivation for these definitions? The following results are special cases of the

"fundamental theorems of affine and projective geometry." They relate synthetic geometry (based on incidence axioms) to analytic geometry (based on coordinates). See Rey Casse, Theorem 4.27, page 64.

Fundamental Theorems of the Projective and Affine Plane.

(1) If $\Phi : \mathbb{RP}^2 \to \mathbb{RP}^2$ is bijective and sends lines to lines, then we have

$$\Phi \begin{pmatrix} x \\ y \\ z \end{pmatrix} = A \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

for some invertible matrix $A \in \mathrm{GL}_3(\mathbb{R})$, unique up to scalar multiplication.

(2) If $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ is bijective and sends lines to lines, then we have

$$\varphi \begin{pmatrix} x \\ y \end{pmatrix} = A' \begin{pmatrix} x \\ y \end{pmatrix} + \mathbf{t}$$

for some unique invertible matrix $A' \in \mathrm{GL}_2(\mathbb{R})$ and vector $\mathbf{t} = (s, t) \in \mathbb{R}^2$.

Proof that (1)⇒(2): Suppose that $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ is bijective and sends lines to lines. Since φ is bijective it sends parallel lines to parallel lines, hence it defines a permutation of the points at infinity. In other words, φ extends to a unique function $\Phi : \mathbb{RP}^2 \to \mathbb{RP}^2$ which is bijective and sends lines to lines. It follows from (1) that

$$\Phi \begin{pmatrix} x \\ y \\ z \end{pmatrix} = A \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

for some invertible matrix $A \in \mathrm{GL}_3(\mathbb{R})$. We know that this matrix stabilizes the plane $z = 0$, hence we can write

$$A = \left(\begin{array}{cc|c} & A' & \mathbf{t} \\ \hline 0 & 0 & c \end{array}\right)$$

for some matrix $A' \in \mathrm{GL}_2(\mathbb{R})$, vector $\mathbf{t} \in \mathbb{R}^2$ and nonzero constant $c \in \mathbb{R}^2$. After scaling A by $1/c$, we may assume that $c = 1$. Therefore A acts on the finite x, y-plane as follows:

$$A \begin{pmatrix} x \\ y \\ \hline 1 \end{pmatrix} = \left(\begin{array}{cc|c} & A' & \mathbf{t} \\ \hline 0 & 0 & 1 \end{array}\right) \begin{pmatrix} x \\ y \\ \hline 1 \end{pmatrix} = \begin{pmatrix} A' \begin{pmatrix} x \\ y \end{pmatrix} + \mathbf{t} \\ \hline 1 \end{pmatrix}.$$

□

Remark: This shows that affine transformations of the x, y-plane $\mathbb{R}^2$ are the same as projective transformations of $\mathbb{RP}^2$ that stabilize the line $z = 0$.

The proof of (1) would take us too far afield, but here is a sketch. For details see Rey Casse, *Projective Geometry: An Introduction*, Theorem 4.27.

Jargon: Any set of four points $\mathbb{RP}^2$, no three collinear, is called a *quadrangle*. The *fundamental quadrangle* is $(0:0:1), (0:1:0), (1:0:0), (1:1:1)$.

Sketch proof of (1):

- Let $\Phi : \mathbb{RP}^2 \to \mathbb{RP}^2$ be a collineation, i.e., bijective and sending lines to lines.
- Suppose Φ sends the fundamental quadrangle to points $PQRS$. It is straightforward to check that there exists a unique matrix $A \in \mathrm{PGL}_3(\mathbb{R})$ sending $PQRS$ back to the fundamental quadrangle.
- Now $A \circ \Phi : \mathbb{RP}^2 \to \mathbb{RP}^2$ is a collineation fixing the fundamental quadrangle.
- **The Hard Part.** If a collineation fixes the fundamental quadrangle then it must act on homogeneous coordinates as an automorphism of the field $\mathbb{R}$. This involves von Staudt's construction of coordinates for a synthetic projective plane and the fact that Pappus' Theorem holds when the coordinates come from a field. See Rey Casse for details.
- Finally, since any field automorphism $\mathbb{R} \to \mathbb{R}$ is trivial, we conclude that $A \circ \Phi : \mathbb{RP}^2 \to \mathbb{RP}^2$ is the identity map, hence $\Phi = A^{-1} \in \mathrm{PGL}_3(\mathbb{R})$.

□

[Remark: I was planning to give the full proof but it is way too complicated. This resulted in two topics being spread over three lectures.]

Sept 2,4: Projective Equivalence of Conics

In this lecture $\mathbb{F}$ will denote a field in which $2 \neq 0$ (i.e., the characteristic of the field is not 2).

Homogeneous polynomials of degree 1 are called *linear forms*. If $\mathbf{x} = (x_1, \ldots, x_n)$ is a vector of independent variables then we can express a linear form $F(\mathbf{x}) \in \mathbb{F}[\mathbf{x}] = \mathbb{F}[x_1, \ldots, x_n]$ in terms of matrix multiplication:

$$F(\mathbf{x}) = a_1 x_1 + \cdots + a_n x_n = \mathbf{a}^T \mathbf{x},$$

where $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{F}^n$ is some vector of constants. The zero set $V_F \subseteq \mathbb{RP}^n$ is called a *projective hyperplane*.[5] If the vector $\mathbf{a}$ is nonzero (which it must be because F has degree 1) then we can always find a matrix $A \in \mathrm{GL}_n(\mathbb{R})$ such that $A^T \mathbf{a} = (1, 0, 0, \ldots, 0)$ and hence

$$F(A\mathbf{x}) = \mathbf{a}^T A \mathbf{x} = (A^T \mathbf{a})^T \mathbf{x} = x_1.$$

In other words, every hyperplane is projectively equivalent to $x_1 = 0$. Pretty boring.

[5] The letter V is for *variety*, a generalization of *curve*.

The degree 2 case is more interesting. Homogeneous polynomials of degree 2 are called *quadratic forms*. These can also be expressed in terms of matrix multiplication. Let $\mathbb{F}(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ be homogeneous of degree 2, so that $F(\lambda\mathbf{x}) = \lambda^2 F(\mathbf{x})$ for all $0 \neq \lambda \in \mathbb{F}$. We can use the *quadratic form* F to define a *symmetric bilinear form*:

$$\langle \mathbf{x}, \mathbf{y} \rangle_F := \frac{1}{4}\left[F(\mathbf{x}+\mathbf{y}) - F(\mathbf{x}-\mathbf{y})\right] \in \mathbb{F}[\mathbf{x}, \mathbf{y}].$$

It is easy to see that $\langle \mathbf{x}, \mathbf{y} \rangle_F = \langle \mathbf{y}, \mathbf{x} \rangle_F$ because $F(\mathbf{x}-\mathbf{y}) = (-1)^2 F(\mathbf{y}-\mathbf{x}) = F(\mathbf{y}-\mathbf{x})$. To show that the form is bilinear we need to be more explicit. By assumption we have $F(\mathbf{x}) = \sum_{i \leq j} c_{ij} x_i x_j$ for some coefficients $c_{ij} \in \mathbb{F}$ with $i \leq j$. Let us also define the symbols $c_{ji} := c_{ij}$ when $j \geq i$. Then we compute

$$\begin{aligned} F(\mathbf{x}+\mathbf{y}) &= \sum_{i \leq j} c_{ij}(x_i + y_i)(x_j + y_j) \\ &= \sum_{i \leq j} c_{ij} x_i x_j + \sum_{i \leq j} c_{ij} y_i y_j + \sum_{i \leq j} c_{ij} x_i y_j + \sum_{i \leq j} c_{ij} y_i x_j \\ &= F(\mathbf{x}) + F(\mathbf{y}) + 2\sum_i c_{ii} x_i y_i + \sum_{i \neq j} c_{ij} x_i y_j, \end{aligned}$$

and, similarly,

$$F(\mathbf{x}-\mathbf{y}) = F(\mathbf{x}) + F(\mathbf{y}) - 2\sum_i c_{ii} x_i y_i - \sum_{i \neq j} c_{ij} x_i y_j,$$

so that

$$\langle \mathbf{x}, \mathbf{y} \rangle_F = \sum_i c_{ii} x_i y_i + \sum_{i \neq j} (c_{ij}/2) x_i y_j$$

Finally, we define the **symmetric matrix** C with entries c_{ii} and $c_{ij}/2$ when $i \neq j$, so that

$$\langle \mathbf{x}, \mathbf{y} \rangle_F = \mathbf{x}^T C \mathbf{y}.$$

This function is clearly bilinear.

Example: Given $F(x, y) = ax^2 + bxy + cy^2 \in \mathbb{F}[x, y]$ we have

$$F(\mathbf{x}) = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Now we can apply linear algebra to prove the following important theorem.

Diagonalization of Quadratic Forms. Again, let $\mathbb{F}$ be any field where $2 \neq 0$. Then for any quadratic form $F(\mathbf{x}) = \mathbf{x}^T C \mathbf{x} \in \mathbb{F}[\mathbf{x}]$ we can find $A \in \mathrm{GL}_n(\mathbb{F})$ such that

$$F(A\mathbf{x}) = d_1 x_1^2 + d_2 x_2^2 + \cdots + d_n x_n^2$$

for some $d_1, \ldots, d_n \in \mathbb{F}$ where $d_1 \ldots, d_k \neq 0$ and $d_{k+1}, \ldots, d_n = 0$ for some $k \geq 1$. In matrix language: For any symmetric matrix $C^T = C$ (invertible or not) we can find an invertible matrix $A \in \mathrm{GL}_n(\mathbb{F})$ such that

$$A^T C A = D = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix}$$

and hence $F(A\mathbf{x}) = (A\mathbf{x})^T C(A\mathbf{x}) = \mathbf{x}^T(A^T C A)\mathbf{x} = \mathbf{x}^T D\mathbf{x}$, as desired.

Proof. Given a matrix $A \in \mathrm{GL}_n(\mathbb{F})$ we will consider the quadratic form

$$G(\mathbf{x}) := F(A\mathbf{x}) = \mathbf{x}^T A^T C A\mathbf{x},$$

with associated bilinar form

$$\langle \mathbf{x}, \mathbf{y} \rangle_G = \langle A\mathbf{x}, A\mathbf{y} \rangle_F = \mathbf{x}^T A^T C A\mathbf{y}.$$

If $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbb{F}^n$ is the standard basis then $\mathbf{a}_i := A\mathbf{e}_i$ is the ith column of A and

$$\langle \mathbf{a}_i, \mathbf{a}_j \rangle_F = \langle A\mathbf{e}_i, A\mathbf{e}_j \rangle_F = \langle \mathbf{e}_i, \mathbf{e}_j \rangle_G = \mathbf{e}_i^T (A^T C A)\mathbf{e}_j,$$

which is the ij entry of the matrix $A^T C A$. Thus it is enough to find a basis of vectors $\mathbf{a}_i \in \mathbb{F}^n$ (the columns of A) with the property that $\langle \mathbf{a}_i, \mathbf{a}_j \rangle_F = 0$ for all $i \neq j$.

To begin we observe that $\langle \mathbf{x}, \mathbf{y} \rangle_F$ for some $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$, otherwise $F(\mathbf{x})$ is the zero polynomial. If $\langle \mathbf{x}, \mathbf{x} \rangle_F = \langle \mathbf{y}, \mathbf{y} \rangle_F = 0$ then we have[6]

$$\langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle_F = \langle \mathbf{x}, \mathbf{x} \rangle_F + \langle \mathbf{y}, \mathbf{y} \rangle_F + 2\langle \mathbf{x}, \mathbf{y} \rangle_F = 2\langle \mathbf{x}, \mathbf{y} \rangle_F \neq 0.$$

Thus we can choose $\mathbf{a}_1 \in \{\mathbf{x}, \mathbf{y}, \mathbf{x}+\mathbf{y}\}$ so that $d_1 := \langle \mathbf{a}_1, \mathbf{a}_1 \rangle_F \neq 0$. Now consider the subspace

$$V_1 := \{\mathbf{x} \in \mathbb{F}^n : \langle \mathbf{a}_1, \mathbf{x} \rangle_F \neq 0\} \subseteq \mathbb{F}^n.$$

Since $\langle \mathbf{a}_i, \mathbf{a}_i \rangle_F = \mathbf{a}_i^T C \mathbf{a}_i \neq 0$ we see that $\mathbf{a}_i^T C$ is a nonzero vector. Then since $\langle \mathbf{a}_1, \mathbf{x} \rangle = \mathbf{a}_1^T C\mathbf{x}$ we see that V_1 is $(n-1)$-dimensional. Furthermore, we observe that $\mathbf{a}_1 \notin V_1$.

If $\langle \mathbf{x}, \mathbf{y} \rangle_F = 0$ for all $\mathbf{x}, \mathbf{y} \in V_1$ then we can choose any basis $\mathbf{a}_2, \ldots, \mathbf{a}_n \in V$. Otherwise, we repeat the argument to choose some $\mathbf{a}_2 \in V_1$ with $\langle \mathbf{a}_2, \mathbf{a}_2 \rangle_F \neq 0$. Now consider the subspace

$$V_2 := \{\mathbf{x} \in \mathbb{F}^n : \langle \mathbf{a}_1, \mathbf{x} \rangle_F \neq 0 \text{ and } \langle \mathbf{a}_2, \mathbf{x} \rangle_F \neq 0\} \subseteq \mathbb{F}^n.$$

Since $\mathbf{a}_1^T C \mathbf{a}_1 \neq 0$ and $\mathbf{a}_2^T C \mathbf{a}_2 \neq 0$ we see that $\mathbf{a}_1^T C$ and $\mathbf{a}_2^T C$ are nonzero vectors. Furthermore, if $\mathbf{a}_1^T C = t(\mathbf{a}_2^T C)$ for some scalar t then we obtain the contradiction

$$0 = \mathbf{a}_1^T C \mathbf{a}_2 = t(\mathbf{a}_2^T C \mathbf{a}_2) \neq 0.$$

[6] Recall that $2 \neq 0$.

Thus V_2 is the intersection of two non-parallel hyperplanes $\mathbf{a}_1^T C\mathbf{x} = 0$ and $\mathbf{a}_2^T C\mathbf{x} = 0$, hence is $(n-2)$-dimensional. Now the result follows by induction. □

Projective equivalence vastly simplifies the classification of conic sections.

Corollary. Any curve $f(x, y) = 0$ in $\mathbb{R}^2$ of degree 2 is projectively equivalent to one of:

(1) $x^2 = 0$ (a double line)

(2) $x^2 \pm y^2 = 0$ (intersecting lines or a single point)

(3) $x^2 + y^2 \pm 1 = 0$ (a circle or the empty set)

In particular, any non-singular quadric curve is projectively equivalent to a circle.

Proof. Let $\mathbf{x} = (x, y, z)$ and let $F(\mathbf{x}) = \mathbf{x}^T C\mathbf{x}$ be the homogenization of $f(x, y)$. From the theorem we can find a matrix $A \in \mathrm{GL}_3(\mathbb{R})$ such that

$$F(A\mathbf{x}) = d_1x^2 + d_2y^2 + d_3z^2$$

for some $d_1, d_2, d_3 \in \mathbb{R}$. Now let $S = (s_{ij}) \in \mathrm{GL}_3(\mathbb{R})$ be defined by $s_{ij} = 0$ when $i \neq j$ and

$$s_{ii} = \begin{cases} 1/\sqrt{d_i} & d_i > 0, \\ 1/\sqrt{-d_i} & d_i < 0, \\ 1 & d_i = 0. \end{cases}$$

Thus

$$F(SA\mathbf{x}) = \delta_1x^2 + \delta_2y^2 + \delta_3z^2,$$

where $\delta_1, \delta_2, \delta_3 \in \{-1, 0, 1\}$, not all zero. Finally, we can choose a permutation matrix $P \in \mathrm{GL}_3(\mathbb{R})$ so that $F(\pm PSA\mathbf{x})$ has one of the forms

(1) x^2,

(2) $x^2 \pm y^2$,

(3) $x^2 + y^2 \pm z^2$,

and de-homogenizing at $z = 1$ gives the desired result. □

Remark: If we allow a "complex projective change of variables" $A \in \mathrm{PGL}_3(\mathbb{C})$ then we can define the elements of S by $s_{ii} = 1/\sqrt{d_i}$ when $d_i \neq 0$ and $s_{ii} = 1$ when $d_i = 0$. Then the standard forms have no negative signs:

(1) x^2,

(2) $x^2 + y^2$,

(3) $x^2 + y^2 + z^2$.

The geometric meaning of this is not so clear.

Principal Ideal Domains

The first homework will take you through the basic properties of homogeneous polynomials over integral domains. This section of the notes will provide background discussion. The solutions will be posted elsewhere.

Sept 9: Fields and Domains

A commutative ring $\mathbb{F}$ is called a *field* if every nonzero element $a \in \mathbb{F}$ has a multiplicative inverse $a^{-1} \in \mathbb{F}$. There is a deep analogy between the ring of polynomials $\mathbb{F}[x]$ in one variable and the ring of integers $\mathbb{Z}$, which is based on the fact that they both have a *division algorithm.*

The Division Algorithm in $\mathbb{Z}$ and $\mathbb{F}[x]$.

- For all $a, b \in \mathbb{Z}$ with $b \neq 0$ there exist $q, r \in \mathbb{Z}$ such that

$$\begin{cases} a = qb + r, \\ |r| < |b|. \end{cases}$$

- For all $f(x), g(x) \in \mathbb{F}[x]$ with $g(x) \neq 0$ there exist $q(x), r(x) \in \mathbb{F}[x]$ such that

$$\begin{cases} f(x) = q(x)g(x) + r(x), \\ \deg(r) < \deg(g) \text{ or } r(x) = 0. \end{cases}$$

The division algorithm for polynomials allows us to prove the following result.

Problem 1.1. A nonzero polynomial $f(x) \in \mathbb{F}[x]$ of degree d has at most d roots in $\mathbb{F}$.

A commutative ring R is called an *integral domain* (or just a *domain*) if for all $a, b \neq 0$ in R we have $ab \neq 0$. Domains have the important property of *multiplicative cancellation*: For all $a, b, c \in R$ with $a \neq 0$ we have

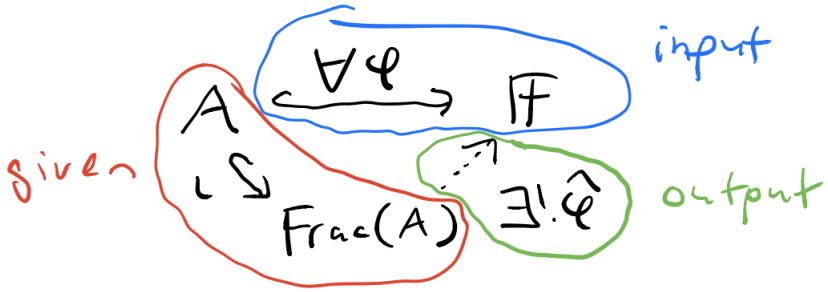
$$\begin{aligned} ac &= bc \\ a(b-c) &= 0 \\ b - c &= 0 \\ b &= c. \end{aligned}$$

Every subring of a field $R \subseteq \mathbb{F}$ (including $\mathbb{F}$ itself) is a domain. Indeed, suppose we have $ab = 0$ for some $a, b \in R$ with $a \neq 0$. Then in $\mathbb{F}$ we have $b = a^{-1}ab = a^{-1}0 = 0$, which also holds in R. Conversely, I claim that every domain R can be realized as a subring of a field, and because of this domains inherit many nice properties from fields.

Problem 1.2. (a): Every domain R is a subring of a field. (b): It follows that a nonzero polynomial $f(x) \in R[x]$ with coefficients in a domain R has at most finitely many roots in R.

Problem 1.3. If R is an infinite domain and if $f(x), g(x) \in R[x]$ satisfy $f(\alpha) = g(\alpha)$ for infinitely many $\alpha \in R$, use Problem 2 to show that $f(x) = g(x)$ as polynomials, i.e., have the same coefficients. In this case there is no distinction between "polynomial expressions" and "polynomial functions."

In fact, for a given domain R, there is a unique "smallest" field containing R as a subfield, called the *field of fractions* $\iota : R \hookrightarrow \mathrm{Frac}(R)$. The field of fractions satisfies the following universal property: If $\varphi : R \hookrightarrow \mathbb{F}$ is an injective ring homomorphism from a domain R to a field $\mathbb{F}$ then there exists a (unique, injective) homomorphism $\hat{\varphi} : \mathrm{Frac}(R) \hookrightarrow \mathbb{F}$ such that $\varphi = \hat{\varphi} \circ \iota$. Picture:



In fact, any field $\mathrm{Frac}(R)$ satisfying this property is unique up to a unique isomorphism. The only issue is to prove that it exists. You will do this on the homework.

Example: $\mathbb{Q} = \mathrm{Frac}(\mathbb{Z})$. In this case the universal property says that if $\mathbb{F} \supseteq \mathbb{Z}$ is a field containing the integers, then $\mathbb{F} \supseteq \mathbb{Q} \supseteq \mathbb{Z}$ also contains the rational numbers. There is really nothing more to it than that.

Another way to describe fields and domains is in terms of maximal and prime ideals in a general ring. Recall that an ideal $I \subseteq R$ in a ring is an additive subgroup satisfying one additional property:

$$a \in R, b \in I \quad \Rightarrow \quad ab \in I.$$

It follows from this definition that the additive quotient group R/I has a natural ring structure defined by

$$(a + I)(b + I) := ab + I.$$

We need only check that this equation is well-defined.

Proof. Suppose that $a + I = a' + I$ and $b + I = b' + I$, so that $a - a', b - b' \in I$. It follows that

$$ab - a'b' = ab - a'b + a'b - a'b' = (a - a')b + a'(b - b') \in I,$$

and hence $ab + I = a'b' + I$. □

Next time we will show that R/I is a field if and only if $I \subseteq R$ is *maximal* (i.e., any ideal $I \subseteq J \subseteq R$ satisfies $I = J$ or $J = R$). We will also prove that R/I is a domain if and only if $I \subseteq R$ is *prime* (i.e., the set $R \setminus I$ is closed under multiplication).

Sept 11: Maximal and Prime Ideals

A *commutative ring* $(R, +, \cdot, 0, 1)$ consists of a commutative group $(R, +, 0)$ and a commutative semigroup $(R, \cdot, 1)$ subject to the *distributive law*:

$$a(b + c) = ab + ac.$$

A ring homomorphism $\varphi : R \to S$ is a homomorphism for addition and multiplication. This implies that $\varphi(a) = \varphi(a + 0) = \varphi(a) + \varphi(0)$ and hence $\varphi(0) = 0$. We also insist that a ring homomorphism satisfies $\varphi(1) = 1$, since this is does not automatically follow from $\varphi(a) = \varphi(a)\varphi(1)$. The kernel and image of $\varphi : R \to S$ are defined as follows:

$$\ker \varphi = \{a \in R : \varphi(a) = 0\},$$
$$\operatorname{im} \varphi = \{b \in S : \exists a \in S, \varphi(a) = b\}.$$

One easily checks that $\operatorname{im} \varphi \subseteq S$ is a subring. The set $\ker \varphi \subseteq R$ is not generally a subring, however it is an ideal because $a \in R$ and $b \in \ker \varphi$ imply $\varphi(ab) = \varphi(a)\varphi(b) = \varphi(a)0 = 0$. Conversely, any ideal $I \subseteq R$ is the kernel of the *canonical surjection* $R \to R/I$ defined by $a \mapsto a + I$. Here is the fundamental theorem of ring homomorphisms.

The Correspondence Theorem. For any ideal $I \subseteq R$ we have an order-preserving bijection:

$$\{\text{ideals of } R \text{ containing } I\} \longleftrightarrow \{\text{ideals of } R/I\}$$

Explicitly, let $\varphi : R \to R/I$ be the canonical surjection. Then the bijection above is defined by sending each ideal $I \subseteq J \subseteq R$ to the image set $\varphi[J] = \{a + I : a \in J\} \subseteq R/I$, and sending each ideal $J' \subseteq R/I$ to the pre-image set $\varphi^{-1}[J'] = \{a \in R : \varphi(a) \in J'\}$.

Proof. There are many small things to check. The key steps are to prove the following for all ideals $J \subseteq A$ and $J' \subseteq R/I$:

- $\varphi[J] \subseteq R/I$ and $\varphi^{-1}[J'] \subseteq R$ are ideals,
- $\varphi[J] \subseteq J' \Leftrightarrow J \subseteq \varphi^{-1}[J']$,
- $\varphi^{-1}[\varphi(J)] = I + J$,
- $\varphi[\varphi^{-1}(J')]] = J'$.

The first property implies that (φ, φ^{-1}) is an order-preserving bijection between ideals $J \subseteq R$ satisfying $\varphi^{-1}[\varphi(J)] = J$ and ideals $J' \subseteq R/I$ satisfying $\varphi[\varphi^{-1}(J')]] = J'$. The second and third properties tell us that this bijection includes all ideals $I \subseteq J \subseteq R$ and all ideals $J' \subseteq A/I$. □

Now we will translate the concepts of field and domain into the language of ideals.

Maximal and Prime Ideals. Let $I \subseteq R$ be an ideal. Then

(1) R/I is a field if and only if $I \subseteq R$ is maximal,

(2) R/I is a domain if and only if $I \subseteq R$ is prime.

Proof. (1): Every ring has the trivial ideals $\{0\}$ and R. A field is a ring that has no other ideals. Indeed, any ideal that contains a unit u must contain $1 = uu^{-1}$ and hence for all $a \in R$ we have $a = a1 \in I$. This implies that a field has no nontrivial ideal. Conversely, if R has no nontrivial ideal then for any nonzero $u \in R$, the principal ideal $\{0\} \subsetneq uR$ must be the whole ring: $uR = R$. It follows that $1 \in R = uR$ and hence $1 = uu^{-1}$ for some $u^{-1} \in R$. /// Combining this with the correspondence theorem gives

$$\begin{aligned}(R/I \text{ is a field}) &\Leftrightarrow (R/I \text{ has no nontrivial ideal}) \\ &\Leftrightarrow (\text{there are no ideals between } I \text{ and } R) \\ &\Leftrightarrow (I \subseteq R \text{ is maximal}).\end{aligned}$$

(2): The correspondence theorem preserves primeness of ideals. Indeed, if $P' \subseteq R/I$ is prime and $a, b \notin \varphi^{-1}[P']$ then $\varphi(a), \varphi(b) \notin P'$, which implies that $\varphi(ab) = \varphi(a)\varphi(b) \notin P'$ and hence $ab \notin \varphi^{-1}[P]$. Conversely, if $I \subseteq P \subseteq R$ is prime and $a + I, b + I \notin \varphi[P]$ then $a, b \notin P$, which implies that $ab \notin P$. If $ab + I = p + I$ for some $p \in P$ then we would have $ab - p \in I \subseteq P$, which would imply $ab \in P$, contradiction. It follows that $ab + I \notin \varphi[P]$ as desired. /// Furthermore, we observe that R is a domain if and only if the zero ideal is prime. Finally, since $I = \ker \varphi = \varphi^{-1}[0]$ we have

$$\begin{aligned}(R/I \text{ is a domain}) &\Leftrightarrow (0 \subseteq R/I \text{ is prime}) \\ &\Leftrightarrow (\varphi^{-1}[0] \subseteq R \text{ is prime}) \\ &\Leftrightarrow (I \subseteq R \text{ is prime}).\end{aligned}$$

□

Sept 14: Principal Ideal Domains

As mentioned above, the rings $\mathbb{Z}$ and $\mathbb{F}[x]$ (where $\mathbb{F}$ is a field) each have a division algorithm. Many properties follow from this algorithm, so we make the following definition.

Definition. A ring R is called *Euclidean* if there exists a size function $\sigma : A \setminus 0 \to \mathbb{N}$ satisfying the following property:

- For all $a, b \in R$ with $b \neq 0$ there exist $q, r \in R$ such that

$$\begin{cases} a = qb + r, \\ r = 0 \text{ or } \sigma(r) < \sigma(b). \end{cases}$$

  The elements q, r need not be unique.

Euclidean ⇒ PIR. Any Euclidean ring is a *principal ideal ring*, in the sense that every ideal $I \subseteq R$ has the form $I = mR$ for some (non-unique) element $m \in R$.

Proof. If $I = 0 = 0R$ then we are done, so let $I \neq 0$ and let $m \in I$ be a nonzero element of minimal size. By definition we have $mR \subseteq I$. On the other hand, I claim that $I \subseteq mR$ and hence $I = mR$. To see this, we can divide any $a \in R$ by the nonzero m to obtain

$$\begin{cases} a = qm + r, \\ r = 0 \text{ or } \sigma(r) < \sigma(m). \end{cases}$$

But now we must have $r = 0$, otherwise we obtain nonzero element $r = a - qm \in I$ with size strictly smaller than m. Thus we conclude that $a = qm \in mR$ as desired. □

Corollary: Any two elements $a, b \in R$ in a Euclidean ring have a (non-unique) *greatest common divisor*. Indeed, we observe that $aR + bR = \{ar + bs : r, s \in S\} \subseteq R$ is an ideal, hence $aR + bR = dR$ for some $d \in R$. It follows that

- $d|a$ and $d|b$,
- if $e|a$ and $e|b$ then $e|d$.

Indeed, the symbol "$m|n$" is defined to mean "$n \in mR$." Since $a \in aR \subseteq dR$ and $b \in bR \subseteq dR$ we have $d|a$ and $d|b$. And if $a, b \in eR$ then we $aR + bR \subseteq eR$ and hence

$$d \in dR = aR + bR \subseteq eR.$$

Let us translate prime and maximal ideals into the language of PIRs.

Prime Ideals in a PIR. Let R be a PIR. Then pR is prime if and only if

$$(p \nmid a \text{ and } p \nmid b) \quad \Rightarrow \quad p \nmid ab$$

for all $a, b \in R$. Indeed, the notation $p \nmid a$ means that $a \notin pR$, or $a \in R \setminus pR$. In this case we say that p is a *prime element of the ring* R. In other words:

$$(pR \subseteq R \text{ is a prime ideal}) \quad \Leftrightarrow \quad (p \in R \text{ is a prime element}).$$

The interpretation of maximal ideals in a PIR is more complicated,[7] so we now restrict our attention to domains.

Definition. A ring R that is PIR and a domain is called PID (*principal ideal domain*). The major new property of PIDs is the following:

$$aR = bR \quad \Leftrightarrow \quad a = ub \text{ for some unit } u \in R.$$

Proof. If $a \sim b$ then we have $a = ub$ so that $a \in bR$, but we also have $b = u^{-1}a$ so that $b \in aR$. It follows that $aR \subseteq bR$ and $bR \subseteq aR$, hence $aR = bR$. Conversely, suppose that $aR = bR$,

[7] See Hungerford (1968), *On the structure of principal ideal rings.*

so that $b = ak$ and $a = b\ell$ for some $k, \ell \in R$. If $a = 0$ then we have $a = b = 0$ so we are done. Otherwise, we use the fact that R is a domain to obtain

$$\begin{aligned} a &= b\ell \\ a &= ak\ell \\ a(1 - k\ell) &= 0 \\ 1 - k\ell &= 0 \\ 1 &= k\ell, \end{aligned}$$

hence $k, \ell \in R$ are units. Now we can characterize maixmal ideals in a PID.

Maximal Ideals in a PID. Let R be a PID and let $mR \subseteq R$ be maximal. Then I claim that

$$a|m \quad \Rightarrow \quad (a \sim m \text{ or } a \sim 1),$$

in which case we say that m is an *irreducible element of* R. Conversely, if $m \in R$ is an irreducible element then $mR \subseteq R$ is a maximal ideal. In other words:

$$(mR \subseteq R \text{ is a maximal ideal}) \quad \Leftrightarrow \quad (m \in R \text{ is an irreducible element}).$$

Proof. Let $mR \subseteq R$ be maximal and let $a|m$ so that $mR \subseteq aR$. Then we have $mR = aR$ (hence $a \sim m$) or $aR = R$ (hence $a \sim 1$). Conversely, let $m \in R$ be irreducible and consider any ideal $mR \subseteq aR \subseteq R$, so that $a|m$. It follows that $a \sim m$ (hence $mR = aR$) or $a \sim 1$ (hence $aR = R$). □

Remark: The ideal $R = 1R$ consists of all units of R. By convention we do not call this a maximal ideal; equivalently, we do not say that units are irreducible. The reason for this convention will show up when we discuss unique factorization.

Sept 16: PID Implies UFD

Review: In any ring R, we say that an element $p \in R$ is *prime* if and only if the ideal $pR \subseteq R$ is prime, i.e., if and only if the following holds for all $a, b \in R$:

$$p \nmid a \text{ and } p \nmid b \quad \Rightarrow \quad p \nmid ab.$$

[Equivalently, $p|ab$ implies $p|a$ or $p|b$.] By definition we observe that $0|a$ (i.e., $a \in 0R$) if and only if $a = 0$, so that $0 \in R$ is prime if and only if R is a domain.

If R is a domain then we have $aR = bR$ if and only if $a = ub$ for some unit $u \in R$, in which case we write $a \sim b$. Then we say that an element $m \in R$ is *irreducible* if and only if the following holds for all $a \in R$:

$$a|m \quad \Rightarrow \quad a \sim m \text{ or } a \sim 1.$$

Equivalently, the ideal $mR \subseteq R$ is maximal among principal ideals. If R is a PID then $m \in R$ is irreducible if and only if $mR \subseteq R$ is maximal among all ideals. [Conventions: The unit ideal

$1R = R$ is not called maximal, hence units are not called irreducible. The zero ideal $0R \subseteq R$ is maximal if and only if R is a field, in which case you could say that $0 \in R$ is irreducible. I don't have strong feelings about it.]

You might observe that the definitions of prime and irreducible elements both express well known properties of prime integers. This equivalence is due to the fact that $\mathbb{Z}$ is a PID.

Prime $\Rightarrow$ Irreducible in a Domain. Let R be a domain. If $p \in R$ is prime we will show that $a|p$ implies $a \sim p$ or $a \sim 1$. Indeed, since $a|p$ we have $p = ab$ for some $b \in R$. But we also have $p|ab$ which since p is prime implies that $p|a$ or $p|b$. In the first case we have $a|p$ and $p|a$, hence $a \sim p$. In the second case we have $b = pu$ for some $u \in R$ and hence $p = ab = aup$ iomplies $1 = au$, hence $a \sim 1$. □

Euclid's Lemma (Irreducible $\Rightarrow$ Prime in a PID).[8] Let R be a PID. Then

$$\begin{aligned}(m \in R \text{ is irreducible}) &\Rightarrow (mR \subseteq R \text{ is maximal}) \\ &\Rightarrow (R/mR \text{ is a field}) \\ &\Rightarrow (R/mR \text{ is a domain}) \\ &\Rightarrow (mR \subseteq R \text{ is prime}) \\ &\Rightarrow (m \in R \text{ is prime}).\end{aligned}$$

The PID hypothesis was used in the first implication. □

Notice that I did not mention "Euclidean rings" today. It is more elegant to begin with a PID, since this avoids mention of an awkward "size function" $\sigma : R \setminus 0 \to \mathbb{N}$. The purpose of the size function is to allow proofs by induction, but it is not really necessary because a PIR has its own intrinsic version of induction.

PIR $\Rightarrow$ Noetherian. Any strictly ascending chain of ideals in a PIR is finite.

Proof. Assume for contradiction that we have an infinite strictly ascending chain of ideals:

$$a_1R \subsetneq a_2R \subsetneq a_3R \subsetneq \cdots \subsetneq R.$$

Note that the union $I = \cup_i a_iR$ is an ideal, hence $I = bR$ for some $b \in R$. But then $b \in a_jR$ for some j which gives the contradiction $bR \subseteq a_jR \subsetneq a_{j+1}R \subseteq I = bR$. □

Finally, by combining Euclid's Lemma and "generalized induction" we obtain the important result that every element of a PID has a unique factorization into prime elements.

PID $\Rightarrow$ UFD. Let R be a PID. Since R is a PIR, we will show that any nonzero element $a \in R \setminus 0$ is similar to a product of irreducible elements:

$$a \sim p_1p_2 \cdots p_k.$$

[8] The original version comes from Euclid's Elements Prop VII.30, which shows that an irreducible integer is prime.

(The empty product corresponds to $a \sim 1$.) Then since R is a PID we know that each irreducible factor is also prime. It follows that the factorization is unique in the sense that if $p_1 p_2 \cdots p_k \sim q_1 q_2 \cdots q_\ell$ for prime elements $p_i, q_j \in R$ then we must have $k = \ell$ and by relabeling the factors we can assume that $p_i \sim q_i$ for all i. In other words, every PID is a UFD (*unique factorization domain*).

Proof. **Existence.** Let R be a PIR. For any $a \in R \setminus 0$ I claim that a can be expressed as a finite product of irreducible elements. Indeed, if this were not the case then by successively factoring a we would obtain an infinite sequence $a = a_0, a_1, \ldots \in R$ with $a_{i+1} | a_i$ and $a_{i+1} \not\sim a_i$. In other words, we would obtain an infinite ascending chain $a_0 R \subsetneq a_1 R \subsetneq \cdots \subsetneq R$.

Uniqueness. We have shown that $a \sim p_1 \cdots p_k$ where $p_1, \ldots, p_k \in R$ are irreducible, and hence prime because R is a PID. Suppose that we also have $a \sim q_1 \cdots q_\ell$ for some primes $q_1, \ldots, q_\ell$. Then $p_1 | q_1 \cdots q_\ell$ implies that $p_1 | q_i$ for some i because p_1 is prime. And since q_i is irreducible we must have $p_1 \sim 1$ or $p_1 \sim q_i$. Since p_1 is not a unit we must have $p_1 \sim q_i$ and after relabeling we may assume that $p_1 \sim q_1$. Then canceling this factor from both sides gives

$$p_2 \cdots p_k \sim q_2 \cdots q_\ell,$$

and the result follows by induction. □

Corollary. If R is a PID then for any nonzero element $a \in R \setminus 0$ I claim that:

- The quotient ring R/aR has finitely many ideals.
- Any two maximal chains of ideals have the same length.

Proof. By the correspondence theorem, there is an order-preserving bijection between ideals of R/aR and ideals of R containing aR. Let $a \sim p_1^{e_1} \cdots p_k^{e_k}$ be the unique prime factorization. Then for any ideal $bR \supseteq aR$, the element b has unique factorization

$$b \sim p_1^{d_1} \cdots p_k^{d_k} \quad \text{with } d_i \leq e_i \text{ for all } i.$$

There are only finitely many such expressions. Furthermore, any maximal chain of ideals corresponds to adding prime factors one at a time, hence has length $e_1 + e_2 + \cdots + e_k$. □

Sept 18: Applications to $\mathbb{Z}$ and $\mathbb{F}[x]$.

The previous lectures covered quite a bit of theory. Here are some applications to the protoypical PIDs; namely, $\mathbb{Z}$ and $\mathbb{F}[x]$.

Applications to $\mathbb{Z}$.

- In $\mathbb{Z}$ the units are just $\{\pm 1\}$, so that $a \sim b$ if and only if $a = \pm b$.
- Every ideal has the form $n\mathbb{Z} \subseteq \mathbb{Z}$ for some unique $n \in \{0, 1, 2, \ldots\}$.
- Every maximal ideal has the form $p\mathbb{Z}$ where $p > 0$ is prime, and there is one non-maximal prime ideal corresponding to $p = 0$.

- Every ring R has a *characteristic* $\mathrm{char}(R) \in \mathbb{N}$ which is defined by considering the unique ring homomorphism $\iota_R : \mathbb{Z} \to R$. If $\ker \iota_R = n\mathbb{Z}$ $(n \geq 0)$ then we say that $\mathrm{char}(R) := n$.
- If R is a domain then $\mathbb{Z}/\ker \iota_R \approx \mathrm{im}\, \iota_R \subseteq R$, being a subring of a domain, is also a domain, hence $\ker \iota_R$ is a prime ideal. If follows that $\mathrm{char}(R) = 0$ or $\mathrm{char}(R) = p > 0$ for some prime.
- We observe that $\mathrm{char}(R) = 0$ if and only if R contains $\mathbb{Z}$ as a subring. Indeed, if $\mathbb{Z} \subseteq R$ then the canonical homomorphism is the identity $\iota_R : \mathbb{Z} \to \mathbb{Z} \subseteq R$. Conversely, if $\ker \iota_R = 0\mathbb{Z}$ then $R \supseteq \mathrm{im}\, \iota_R \approx \mathbb{Z}/\ker \iota_R = \mathbb{Z}/0\mathbb{Z} = \mathbb{Z}$. If $\mathbb{F}$ is a **field** of characteristic zero then it follows that $\mathbb{Q} = \mathrm{Frac}(\mathbb{Z}) \subseteq \mathbb{F}$ is a subfield. In fact, it is the **smallest subfield**, called the *prime subfield* of $\mathbb{F}$.
- If $\mathbb{F}$ is a field of prime characteristic $p > 0$ then the prime subfield is $\mathbb{F} \supseteq \mathrm{im}\, \iota_{\mathbb{F}} \approx \mathbb{Z}/\ker \iota_{\mathbb{F}} = \mathbb{Z}/p\mathbb{Z}$. If $\mathbb{F}$ is finite then it is a finite-dimensional vector space over $\mathbb{Z}/p\mathbb{Z}$, hence $\#\mathbb{F} = p^d$ for some $d \in \mathbb{N}$.
- For every prime power p^d $(p, d > 0)$, it is true (but much harder to show) that there exists a field of size p^d, and this field is unique up to isomorphism. I will say more about this in the applications to $\mathbb{F}[x]$.

Applications to $\mathbb{F}[x]$.

- In $\mathbb{F}[x]$ the units are the nonzero constants, i,.e., the polynomials of degree 0. Thus we have $f(x) \sim g(x)$ if and only if $f(x) = cg(x)$ for some $c \in \mathbb{F} \setminus 0$.
- Hence every nonzero ideal has the form $m(x) \subseteq \mathbb{F}[x]$ for some unique *monic polynomial* $m(x) \in \mathbb{F}[x]$, i.e., with leading coefficient 1.
- The prime ideals are $0\mathbb{F}[x]$ (the non-maximal prime) and $m(x)\mathbb{F}[x]$, where $m(x) \in \mathbb{F}[x]$ is an irreducible monic polynomial (though we could also call it a "prime" polynomial).
- For any (commutative) ring $R \supseteq \mathbb{F}$ and any element $\alpha \in R$ we have an evaluation homomorphism $\iota_\alpha : \mathbb{F}[x] \to R$ sending $f(x) \in \mathbb{F}[x]$ to $f(\alpha) \in R$. The image $\mathbb{F}[\alpha] := \mathrm{im}\, \iota_\alpha \subseteq R$ is the smallest subring of R containing the set $\mathbb{F} \cup \{\alpha\}$.
- If $R \supseteq \mathbb{F}$ is a domain, it follows that $\mathbb{F}[x]/\ker \iota_\alpha \approx \mathbb{F}[\alpha] \subseteq R$ is a domain, hence $\ker \iota_\alpha = m_\alpha(x)\mathbb{F}[x]$, where $m_\alpha(x) = 0$ or $m_\alpha(x)$ is irreducible.
- If $m_\alpha(x) = 0$ then we say that α is *transcendental over* $\mathbb{F}$, in which case $\mathbb{F}[\alpha] \approx \mathbb{F}[x]$, and we can treat α as a "variable."
- If $m_\alpha(x) \neq 0$ then we say that α is *algebraic over* $\mathbb{F}$, in which case $m_\alpha(x)$ is called the *minimal polynomial of* $\alpha/\mathbb{F}$. Since a nonzero prime ideal in a PID is maximal, this implies that $\mathbb{F}[\alpha] \approx \mathbb{F}[x]/m_\alpha(x)\mathbb{F}[x]$ is actually a **field**, in which case we observe that $\mathbb{F}[\alpha] = \mathbb{F}(\alpha) \subseteq R$ is the smallest sub**field** of R containing the set $\mathbb{F} \cup \{\alpha\}$.
- Furthermore, if $\deg(m_\alpha) = d$ then I claim that $\mathbb{F}(\alpha)$ is a d-dimensional vector space over $\mathbb{F}$ with basis $1, \alpha, \ldots, \alpha^{d-1}$. Proof: Every element of $\mathbb{F}(\alpha) = \mathbb{F}[\alpha]$ has the form $f(\alpha)$ for

some $f(x) \in \mathbb{F}[x]$. Divide by $m_\alpha(x)$ to obtain

$$f(x) = q(x)m_\alpha(x) + r(x),$$

where $r(x) = 0$ or $\deg(r) < \deg(m_\alpha) = d$. In either case, we have $r(x) = a_0 + a_1\alpha + \cdots + a_{d-1}\alpha^{d-1}$ for some $a_0, \ldots, a_{d-1} \in \mathbb{F}$. Evaluate at $x = \alpha$ to get

$$f(\alpha) = q(\alpha)m_\alpha(\alpha) + r(\alpha) = 0 + r(\alpha) = a_0 + a_1\alpha + \cdots + a_{d-1}\alpha^{d-1}.$$

Hence $1, \alpha, \ldots, \alpha^{d-1}$ is a spanning set. To prove independence, suppose that we have two such representations:

$$a_0 + a_1\alpha + \cdots + a_{d-1}\alpha^{d-1} = b_0 + b_1\alpha + \cdots + b_{d-1}\alpha^{d-1}.$$

We can rephrase this as $r_1(\alpha) = r_2(\beta)$ for two polynomials $r_1(x), r_2(x) \in \mathbb{F}[x]$ of degree $< d$ and our goal is to show that these are equal as polynomials, i.e., have the same coefficients. To see this, we note that $r_1(x) - r_2(x) \in \ker \iota_\alpha = m_\alpha(x)\mathbb{F}[x]$ and hence $r_1(x) - r_2(x) = m_\alpha(x)g(x)$ for some $g(x) \in \mathbb{F}[x]$. If $r_1(x) - r_2(x) \neq 0$ then this implies that $d = \deg(m_\alpha) \leq \deg(r_1 - r_2)$, which contradicts the fact that $\deg(r_1), \deg(r_2) < d$.

- Conversely, let $f(x) \in \mathbb{F}[x]$ be any nonzero polynomial of degree d. The same proof as above shows that $R := \mathbb{F}[x]/f(x)\mathbb{F}[x]$ is a d-dimensional vector space over $\mathbb{F}$ with basis (the images of) $1, x, \ldots, x^{d-1}$. If $f(x)$ is irreducible then R is a field. If $\mathbb{F} = \mathbb{Z}/p\mathbb{Z}$ then R is a finite field of size p^d. Thus the existence of finite fields can be proved by showing that there exist irreducible polynomials in $\mathbb{Z}/p\mathbb{Z}[x]$ of all degrees. (It's still not easy.)

Remark: The analogy between $\mathbb{Z}$ and $\mathbb{F}[x]$ is the heart of commutative algebra, and the general theory of PIDs is the most beautiful way to capture this analogy.

Tangent Spaces

Sept 21: Homogeneous Polynomials

Let R be a ring and let $\mathbf{x} = \{x_1, \ldots, x_n\}$ be *independent variables* over R. Technically: We assume that there exists some ring $E \supseteq R$ such that $\mathbf{x} \subseteq E$ and such that each $x_i \in E$ is *transcendental* over the subring $R[x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n]$.[9] In this case $R[\mathbf{x}]$ is called the *polynomial ring* generated by $\mathbf{x}$. If $\mathbf{x}$ and $\mathbf{y}$ are mutually independent sets of variables then we have $R[\mathbf{x}, \mathbf{y}] = R[\mathbf{x}][\mathbf{y}] = R[\mathbf{y}][\mathbf{x}]$.

For any vector $I = (i_1, \ldots, i_n) \in \mathbb{N}^n$ of natural numbers, we define the *monomial*

$$\mathbf{x}^I := x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n} \in R[\mathbf{x}],$$

[9] In André Weil's *Foundations of Algebraic Geometry* (1946), he always assumed the existence of a *universal domain* $\Omega \supseteq R$ containing an infinite supply of independent variables. The modern (post-Grothendieck) approach thinks of polynomials as elements of a *free R-algebra*, defined by mapping properties. Both approaches have their advantages.

with the property that $\mathbf{x}^I\mathbf{x}^J = \mathbf{x}^{I+J}$. By definition, these monomials are an R-linear basis of $R[\mathbf{x}]$. That is, every polynomial $f(\mathbf{x}) \in R[\mathbf{x}]$ has a unique expression of the form

$$f(\mathbf{x}) = \sum_{I\in\mathbb{N}^n} a_I\mathbf{x}^I,$$

where all but finitely many of the coefficients $a_I \in R$ are zero. By collecting terms with a common total degree $\sum I = i_1 + i_2 + \cdots + i_n$ we obtain a filtration into *homogeneous components*:

$$f(\mathbf{x}) = f^{(0)}(\mathbf{x}) + f^{(1)}(\mathbf{x}) + f^{(2)}(\mathbf{x}) + \cdots,$$

where

$$f^{(d)} = \sum_{\sum I=d} a_I\mathbf{x}^I.$$

If $f(\mathbf{x}) \neq 0$ then the *degree* $\deg(f) \in \mathbb{N}$ is defined as the maximum d such that $f^{(d)}(\mathbf{x}) \neq 0$, in which case this $f^{(d)}(\mathbf{x})$ is called the *leading form*. If $f(\mathbf{x}) = f^{(d)}(\mathbf{x})$ then we say that $f(\mathbf{x})$ is *homogeneous of degree d*. The polynomials of degree 0 are the nonzero constants.

Problem 1.4. If R is a domain, show that the degree is additive:

$$\deg(fg) = \deg(f) + \deg(g).$$

Use this to conclude that $R[\mathbf{x}]$ is also a domain.

Sometimes it is easier to use the following scaling property to prove that a polynomial is homogeneous.

Problem 1.5. Let $f(\mathbf{x}) \in R[\mathbf{x}]$ and consider the following two conditions:

(H1) $f(\mathbf{x}) = f^{(d)}(\mathbf{x})$,

(H2) $f(\mathbf{x}) \neq 0$ and $f(\lambda\mathbf{x}) = \lambda^d f(\mathbf{x})$ for all $\lambda \in R \setminus 0$.

Prove that (H1)⇒(H2) for any ring R. Prove that (H2)⇒(H1) when R is an infinite domain.

Proof. For any $\mathbf{x}^I$ we have $(\lambda\mathbf{x})^I = \lambda^{\sum I}\mathbf{x}^I$, hence for any $k \geq 0$ we have $f^{(k)}(\lambda\mathbf{x}) = \lambda^k f(\mathbf{x})$. This proves the first direction. Conversely, suppose that R is an infinite domain and define the polynomials $g(\mathbf{x}, y) = y^d f(\mathbf{x}) \in R[\mathbf{x}][y]$ and $h(\mathbf{x}, y) = f(y\mathbf{x}) = \sum y^k f^{(k)}(\mathbf{x}) \in R[\mathbf{x}][y]$, where y is a variable independent from $\mathbf{x}$. By assumption, the polynomial $g(\mathbf{x}, y) - h(\mathbf{x}, y) \in R[\mathbf{x}][y]$ has infinitely many roots $y = \lambda \in R \setminus 0$. Since $R[\mathbf{x}]$ is an infinite domain, it follows from Problem 1.3 that the polynomials $g(\mathbf{x}, y), h(\mathbf{x}, y) \in R[\mathbf{x}][y]$ have the same y-coefficients. In other words, we have $f^{(d)}(\mathbf{x}) = f(\mathbf{x})$ and $f^{(k)}(\mathbf{x}) = 0$ for all $k \neq d$. □

The following application shows that the degree of a curve (more generally, the degree of a hypersurface) is a projective invariant.

Application (Problem 1.6). Let R be an infinite domain and let $A \in \mathrm{GL}_n(R)$ be an invertible $n \times n$ matrix. If $F(\mathbf{x}) \in R[x_1, \ldots, x_n]$ is homogeneous of degree d, then $G(\mathbf{x}) := F(A\mathbf{x}) \in R[\mathbf{x}]$ is also homogeneous of degree d.[10]

Proof. First we observe that $G(\mathbf{x}) \neq 0(\mathbf{x})$, otherwise we obtain the contradiction $0(\mathbf{x}) \neq F(\mathbf{x}) = 0(A^{-1}\mathbf{x}) = 0(\mathbf{x})$. Then since matrix multiplication is linear, we have

$$G(\lambda\mathbf{x}) = F(A\lambda\mathbf{x}) = F(\lambda A\mathbf{x}) = \lambda^d F(A\mathbf{x}) = \lambda^d G(\mathbf{x})$$

for all $\lambda \in R \setminus 0$. □

Sept 23: Formal Derivatives and the Chain Rule

For a variable x over a ring R, let $D_x : R[x] \to R[x]$ be the unique R-linear map defined by

$$D_x(x^k) := \begin{cases} kx^{k-1} & k > 0, \\ 0 & k = 0. \end{cases}$$

Problem 1.7. Prove that the following properties are satisfied for all $f(x), g(x) \in R[x]$.

(a) $D_x(fg) = D_x(f)g + fD_x(g)$,

(b) $D_x(g^k) = kg^{k-1}D_x(g)$,

(c) $D_x(f \circ g) = (D_x f \circ g)D_x(g)$.

Remark: If $f(x) = \sum a_k x^k$ then the *formal composition* $(f \circ g)(x) \in R[x]$ is defined as $\sum a_k (g(x))^k$. If R is an infinite domain then this agrees with the composition of the corresponding "polynomial functions" $R \to R$.

Proof. (a): One can check that $\Phi(f, g) := D_x(fg)$ and $\Psi(f, g) := D_x(f)g + fD_x(g)$ are both R-bilinear functions $R[x]^2 \to R[x]$. Thus it suffices to check this identity on the basis of monomials. If $f(x) = x^m$ and $g(x) = x^n$ then we observe that

$$D_x(f)g + fD_x(g) = mx^{m-1}x^n + x^m nx^{n-1} = (m+n)x^{m+n-1} = D_x(x^{m+n}) = D_x(fg),$$

as desired.

(b): We observe that the statement $D_x(g^k) = kg^{k-1}D_x(g)$ is true for $k = 0, 1$. Now let us assume for induction that $D_x(g^k) = kg^{k-1}D_x(g)$ for some $k \geq 1$. It follows from part (a) that

$$D_x(g^{k+1}) = D_x(gg^k) = D_x(g)g^k + gkg^{k-1}D_x(g) = (k+1)g^k D_x(g).$$

(c): Let $f(x) = \sum a_k x^k$, so that $f \circ g = \sum a_k g^k$. Then it follows from (b) that

$$D_x(f \circ g) = \sum a_k D_x(g^k) = \left(\sum a_k k g^{k-1}\right) D_x(g) = [D_x(f) \circ g]D_x(g).$$

[10] This should also hold for any **nonzero** matrix, but in this case it is harder to show that $G(\mathbf{x}) \neq 0(\mathbf{x})$.

□

Now we extend the notion of differential operators to multivariable polynomials. If $\mathbf{x} = \{x_1, \ldots, x_n\}$ are independent variables over R then we define $D_{x_i} : R[\mathbf{x}] \to R[\mathbf{x}]$ by thinking of $R[\mathbf{x}] = R'[x_i] = R[x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n][x_i]$. Since D_{x_i} is R'-linear and since $R \subseteq R'$ is a subring, we observe that D_{x_i} is also R-linear. Alternatively, we can define $D_{x_i} : R[\mathbf{x}] \to R[\mathbf{x}]$ as the unique R-linear map defined on monomials as follows:

$$D_{x_i}(x^{k_1} \cdots x^{k_n}) := \begin{cases} k_i \cdot x_1^{k_1} \cdots x_{i_1}^{k_i} x_i^{k_i - 1} x_{i+1}^{k_{i+1}} \cdots x_n^{k_n} & k_i > 0, \\ 0 & k_i = 0. \end{cases}$$

We observe that "mixed partials commute," since for any monomial $m(\mathbf{x}) = x_1^{k_1} \cdots x_n^{k_n} \in R[\mathbf{x}]$ and for any indices $i \neq j$ we have

$$D_{x_i} D_{x_j}(m) = D_{x_j} D_{x_i}(m) = k_i k_j \cdot x_1^{k_1} \cdots x_i^{k_i - 1} \cdots x_j^{k_j - 1} \cdots x_n^{k_n}.$$

The multivariable chain rule explains how differentiation is related to change of coordinates, including projective equivalence as a special case.

Definition of Total Derivative. Let $\mathbf{x} = \{x_1, \ldots, x_n\}$ be independent variables over a ring R and let $f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})) \in R[\mathbf{x}]^m$ be a vector of polynomials. Then the *total derivative* $Df \in R[\mathbf{x}]^{m \times n}$ is the $m \times n$ matrix of polynomials, whose i, j entry is $D_{x_j} f_i$:

$$Df = \begin{pmatrix} D_{x_1} f_1 & D_{x_2} f_1 & \cdots & D_{x_n} f_1 \\ D_{x_1} f_2 & D_{x_2} f_2 & \cdots & D_{x_n} f_2 \\ \vdots & \vdots & \ddots & \vdots \\ D_{x_1} f_m & D_{x_2} f_m & \cdots & D_{x_n} f_m \end{pmatrix}.$$

This is also called the *Jacobian matrix*. If $m = 1$ then Df is a row vector with m entries, which we sometimes call the *gradient vector* $\nabla f := Df$.

Example: If R is an infinite domain then we can identify $f(\mathbf{x}) \in R[\mathbf{x}]^m$ with a function $f : R^n \to R^m$ by evaluation. If this function is linear with matrix A then I claim that $Df = A$ (i.e., the linearization of a linear function is the function itself). Proof: If $A = (a_{ij}) \in R^{m \times n}$ then by definition of matrix multiplication we have $f_i(\mathbf{x}) = a_{i1} x_1 + \cdots + a_{in} x_n \in R[\mathbf{x}]$, so that the i, j entry of Df is $D_{x_j} f_i = a_{ij}$. ///

The Chain Rule. Let $\mathbf{x} = \{x_1, \ldots, x_n\}$ be independent variables over a ring R, let $f(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})) \in R[\mathbf{x}]^m$ be a vector of m polynomials and let $\Phi(\mathbf{x}) = (\Phi_1(\mathbf{x}), \ldots, \Phi_n(\mathbf{x})) \in R[\mathbf{x}]^n$ be a vector of n polynomials in the same variables. [If R is an infinite domain then we can think of these as "polynomial functions" $f : R^n \to R^m$ and $\Phi : R^n \to R^n$.] We define the *formal composition* by evaluation:

$$(f \circ \Phi)(\mathbf{x}) := ((f_1 \circ \Phi)(\mathbf{x}), \ldots, (f_m \circ \Phi)(\mathbf{x})) \in R[\mathbf{x}]^m,$$

where $(f_i \circ \Phi)(\mathbf{x}) := f_i(\Phi_1(\mathbf{x}), \ldots, \Phi_n(\mathbf{x})) \in R[\mathbf{x}]$. [If R is an infinite domain then we can think of this as the composite function $f \circ \Phi : R^n \to R^m$.] Then I claim that

$$D(f \circ \Phi) = (Df \circ \Phi) \cdot D\Phi,$$

where the operation on the right hand side is matrix multiplication.

Proof. Sufficient...

Note that the kth row of $D(f \circ \Phi)$ is equal to the kth row of $Df \circ \Phi$ times the matrix $D\Phi$. Thus it suffices to prove the case when this matrix has only one row, i.e., when $m = 1$ and $f(\mathbf{x}) \in R[\mathbf{x}]$ is just a single polynomial.

Now our goal is to show that the row vector $\nabla(f \circ \Phi)$ is equal to the row vector $(\nabla f \circ \Phi)$ times the matrix $D\Phi$. That is, we want to show that the jth entry of the vector $\nabla(f \circ \Phi)$ is equal to the dot product of the row $(\nabla f \circ \Phi)$ with the jth column of the matrix $D\Phi$.

To prove this, suppose that $f(\mathbf{x}) = \sum_K b_K \mathbf{x}^K$. Then the ith entry of ∇f is

$$D_{x_i} f = \sum_K b_K k_i \cdot x_1^{k_1} \cdots x_i^{k_i - 1} \cdots x_n^{k_n},$$

and hence the ith entry of $(\nabla f \circ \Phi)$ is by definition

$$(D_{x_i} f \circ \Phi)(\mathbf{x}) = \sum_K b_K k_i \cdot \Phi_1(\mathbf{x})^{k_1} \cdots \Phi_i(\mathbf{x})^{k_i - 1} \cdots \Phi_n(\mathbf{x})^{k_n}.$$

On the other hand, by applying (a),(b),(c) from Problem 1.7, the jth entry of $\nabla(f \circ \Phi)$ is

$$\begin{aligned}
D_{x_j}(f \circ \Phi) &= D_{x_j} \left(\sum_K b_K \cdot \Phi_1(\mathbf{x})^{k_1} \cdots \Phi_n(\mathbf{x})^{k_n} \right) \\
&= \sum_K b_K D_{x_j} \left(\Phi_1(\mathbf{x})^{k_1} \cdots \Phi_n(\mathbf{x})^{k_n} \right) \\
&= \sum_K b_K \sum_i k_i \cdot \Phi_1(\mathbf{x})^{k_1} \cdots \Phi_i(\mathbf{x})^{k_i - 1} \cdots \Phi_n(\mathbf{x})^{k_n} \cdot D_{x_j} \Phi_i(\mathbf{x}) \\
&= \sum_i \left(\sum_K b_K k_i \cdot \Phi_1(\mathbf{x})^{k_1} \cdots \Phi_i(\mathbf{x})^{k_i - 1} \cdots \Phi_n(\mathbf{x})^{k_n} \right) D_{x_j} \Phi_i(\mathbf{x}) \\
&= \sum_i \left(i\text{th entry of } \nabla f \circ \Phi \right) \left(i, j \text{ entry of } D\Phi \right) \\
&= \left(\nabla f \circ \Phi \right) \left(j\text{th column of } D\Phi \right)
\end{aligned}$$

□

The point of this lecture is to show that the machinery of differential geometry works perfectly well over any ring, as long as we restrict our attention to polynomial functions. Later we will translate this machinery into the language of maximal ideals and local rings.

Sept 25: Taylor Expansion

Taylor series are used to investigate the behavior of a function in the neighborhood of a point. In the case of a polynomial $f(\mathbf{x}) \in R[\mathbf{x}] = R[x_1, \ldots, x_n]$ and a point $\mathbf{a} \in R^n$, we will see that the Taylor expansion of f near the point $\mathbf{x} = \mathbf{a}$ has the following form:

$$f(\mathbf{x}) = f(\mathbf{a}) + (\nabla f)_{\mathbf{a}}(\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T (Hf)_{\mathbf{a}}(\mathbf{x} - \mathbf{a}) + \text{higher terms}.$$

To interpret this formula, we recall the definition of the *gradient row vector*:

$$\nabla f = (D_{x_1} f, D_{x_2} f, \ldots, D_{x_n} f) \in R[\mathbf{x}]^n.$$

And we introduce the *Hessian matrix* of second derivatives:

$$Hf := \begin{pmatrix} D_{x_1}D_{x_1}f & D_{x_1}D_{x_2}f & \cdots & D_{x_1}D_{x_n}f \\ D_{x_2}D_{x_1}f & D_{x_2}D_{x_2}f & \cdots & D_{x_2}D_{x_n}f \\ \vdots & \vdots & \ddots & \vdots \\ D_{x_n}D_{x_1}f & D_{x_n}D_{x_2}f & \cdots & D_{x_n}D_{x_n}f \end{pmatrix} \in R[\mathbf{x}]^{n \times n}$$

Note that this matrix is symmetric because mixed partials commute. The notations $(\nabla f)_{\mathbf{a}}$ and $(Hf)_{\mathbf{a}}$ indicate that we should evaluate all of the entries at $\mathbf{x} = \mathbf{a}$, to obtain a vector and matrix of elements of R. Finally, we interpret $(\mathbf{x} - \mathbf{a}) = (x_1 - a_1, x_2 - a_2, \ldots, x_n - a_n)$ as a column vector.

Today we will prove this formula and next time we will begin to explore its geometric meaning.

Taylor Expansion. Let $\mathbf{x} = \{x_1, \ldots, x_n\}$ be independent variables over a ring R. For any index vector $I = (i_1, \ldots, i_k) \in \mathbb{N}^n$ we define the differential operator $D^I_{\mathbf{x}} : R[\mathbf{x}] \to R[\mathbf{x}]$ by

$$D^I_{\mathbf{x}} = D^{i_1}_{x_1} D^{i_2}_{x_2} \cdots D^{i_n}_{x_n} = \frac{\partial^{i_1 + \cdots + i_n}}{\partial x_1^{i_1} \cdots \partial x_n^{i_n}}.$$

which is well-defined because the mixed partials commute. Then for any polynomial $f(\mathbf{x}) = R[\mathbf{x}]$ and for any point $\mathbf{a} \in R^n$ I claim that we have

$$f(\mathbf{x}) = \sum_{I \in \mathbb{N}^n} \frac{(D^I_{\mathbf{x}} f)(\mathbf{a})}{I!} (\mathbf{x} - \mathbf{a})^I,$$

where $I! = i_1! i_2! \cdots i_n!$ and $(\mathbf{x} - \mathbf{a})^I = (x_1 - a_1)^{i_1} \cdots (x_1 - a_n)^{i_n}$. The sum is finite because $D^I_{\mathbf{x}} f$ is the zero polynomial whenever $i_k > \deg(f)$ for some k.

Proof. Given a point $\mathbf{a} \in R^n$ and a polynomial $f(\mathbf{x}) \in R[\mathbf{x}]$, we consider the polynomial $g(\mathbf{x}) := f(\mathbf{x} + \mathbf{a}) \in R[\mathbf{x}]$. Since $g(\mathbf{x}) \in R[\mathbf{x}]$ we can write $g(\mathbf{x}) = \sum_I c_I \mathbf{x}^I$ for some coefficients $c_I \in R$. I claim that we will be done if we can show that

$$c_I = \frac{(D^I_{\mathbf{x}} g)(\mathbf{0})}{I!}. \qquad (*)$$

Indeed, we observe from the basic chain rule that $(D_{x_k}g)(\mathbf{x}) = (D_{x_k}f)(\mathbf{x}+\mathbf{a})$ for all k, and hence$(D^I_{\mathbf{x}}g)(\mathbf{x}) = (D^I_{\mathbf{x}}f)(\mathbf{x}+\mathbf{a})$ for all $I \in \mathbb{N}^n$. By combining these facts we will conclude that

$$f(\mathbf{x}) = g(\mathbf{x}-\mathbf{a}) = \sum_{I\in\mathbb{N}^n} \frac{(D^I_{\mathbf{x}}g)(\mathbf{0})}{I!}(\mathbf{x}-\mathbf{a})^I = \sum_{I\in\mathbb{N}^n} \frac{(D^I_{\mathbf{x}}f)(\mathbf{a})}{I!}(\mathbf{x}-\mathbf{a})^I.$$

In order to prove $(*)$ we define a partial ordering on $\mathbb{N}^n$ by saying that "$I \le J$" when $i_k \le j_k$ for all k and "$I < J$" when $I \le J$ and $I \ne J$, i.e., when $i_k \le j_k$ for all k and $i_k < j_k$ for some k. Then I claim that the differential operator $D^I_{\mathbf{x}}$ acts on the monomial $\mathbf{x}^J$ as follows:

$$D^I_{\mathbf{x}}(\mathbf{x}^J) = \begin{cases} (J!_I)\mathbf{x}^{J-I} & I < J, \\ I! & I = J, \\ 0 & \text{otherwise,} \end{cases}$$

where we use the notations $(r)_s = r(r-1)\cdots(r-s+1)$ and $J!_I = (j_1)_{i_1}\cdots(j_n)_{i_n}$. Indeed, if $I \not\le J$ then we have $i_k > j_k$ for some k. Then it follows that $D^{i_k}_{x_k}\mathbf{x}^J = 0$ and hence $D^I_{\mathbf{x}}(\mathbf{x}^J) = 0$. On the other hand, if $I \le J$ then we have $i_k \le j_k$ and hence $D^{i_k}_{x_k}\mathbf{x}^J = (j_k)_{i_k}\mathbf{x}^J/x_k^{i_k}$ for all k, which implies that $D^I_{\mathbf{x}}(\mathbf{x}^J) = (J!_I)\mathbf{x}^J/\mathbf{x}^I = (J!_I)\mathbf{x}^{J-I}$. Finally, if $I = J$ then we observe that $\mathbf{x}^{J-I} = \mathbf{x}^{\mathbf{0}} = 1$ and $J!_I = I!_I = (i_1)_{i_1}\cdots(i_n)_{i_n} = i_1!\cdots i_n! = I!$.

By applying this rule to the polynomial $g(\mathbf{x}) = \sum_J c_J\mathbf{x}^J$ we obtain

$$D^I_{\mathbf{x}}g = c_I I! + \sum_{J>I} c_J(J!_I)\mathbf{x}^{J-I},$$

and then evaluating at $\mathbf{x} = \mathbf{0}$ gives $(D^I_{\mathbf{x}}g)(\mathbf{0}) = c_I I!$ as desired. □

To complete the discussion from above, we investigate the first few terms of the Taylor series. The vectors $I \in \mathbb{N}^n$ satisfying $\sum I = 0$ are just $I = \mathbf{0}$. The corresponding term is:

$$\frac{(D^{\mathbf{0}}_{\mathbf{x}}f)(\mathbf{a})}{\mathbf{0}!}(\mathbf{x}-\mathbf{a})^{\mathbf{0}} = \frac{f(\mathbf{a})}{0!0!\cdots 0!}1 = f(\mathbf{a}).$$

The vectors satisfying $\sum I = 1$ are $I_k := (0,\ldots,0,1,0,\ldots,0)$, with a 1 in the kth position. We observe that $I_k! = 0!\cdots 0!1!0!\cdots 0! = 1$, so the sum of the corresponding terms is

$$\sum_k \frac{(D^{I_k}_{\mathbf{x}}f)(\mathbf{a})}{I_k!}(\mathbf{x}-\mathbf{a})^{I_k} = \sum_k (D_{x_k}f)(\mathbf{a})(x_k-a_k) = (\nabla f)_{\mathbf{a}}(\mathbf{x}-\mathbf{a}).$$

The vectors satisfying $\sum I = 2$ are $2I_k$ and $I_{k\ell} := (0,\ldots,0,1,0,\ldots,0,1,0,\ldots,0)$, with 1s in the kth and ℓth position, where $k < \ell$. Let's use the notation $I_{kk} = 2I_k$. Then we observe

that $I_{kk}! = 2$ and $I_{k\ell}! = 0$ for $k < \ell$, so the sum of the corresponding terms is

$$\begin{aligned}
&\sum_{k\le\ell} \frac{(D_{\mathbf{x}}^{I_{k\ell}} f)(\mathbf{a})}{I_{k\ell}!}(\mathbf{x}-\mathbf{a})^{I_{k\ell}} \\
&= \frac{1}{2}\sum_k (D_{x_k}D_{x_k}f)(\mathbf{a})(x_k-a_k)^2 + \sum_{k<\ell}(D_{x_k}D_{x_\ell}f)(\mathbf{a})(x_k-a_k)(x_\ell-a_\ell) \\
&= \frac{1}{2}\sum_k (D_{x_k}D_{x_k}f)(\mathbf{a})(x_k-a_k)^2 + \frac{1}{2}\sum_{k\neq\ell}(D_{x_k}D_{x_\ell}f)(\mathbf{a})(x_k-a_k)(x_\ell-a_\ell) \\
&= \frac{1}{2}\sum_k\sum_\ell (D_{x_k}D_{x_\ell}f)(\mathbf{a})(x_k-a_k)(x_\ell-a_\ell) \\
&= \frac{1}{2}(\mathbf{x}-\mathbf{a})^T (Hf)_{\mathbf{a}}(\mathbf{x}-\mathbf{a}).
\end{aligned}$$

In the second equality we used the fact that $D_{x_k}D_{x_\ell} = D_{x_\ell}D_{x_k}$ for all k, ℓ.

Sept 28,30: Tangent Spaces

What does it mean for a line to be tangent to a curve, more generally to a hypersurface?

Let $\mathbb{F}$ be a field. A line $L \subseteq \mathbb{F}^n$ containing a point $\mathbf{p} \in \mathbb{F}^n$ can be parametrized as $L : \mathbf{p} + t\mathbf{v}$ where $\mathbf{v} \in \mathbb{F}^n$ is a direction vector and $t \in \mathbb{F}$ is a parameter. We can think of the parametrization as an injective function $\mathbb{F} \to \mathbb{F}^n$ sending $t \mapsto \mathbf{p} + t\mathbf{v}$. Now let $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}] = \mathbb{F}[x_1, \ldots, x_n]$ and consider the hypersurface $V : f(\mathbf{x}) = 0$. By composing $f : \mathbb{F}^n \to \mathbb{F}$ with the parametrized line $\mathbb{F} \to \mathbb{F}^2$ we obtain a polynomial

$$\varphi(t) := f(\mathbf{p} + t\mathbf{v}) \in \mathbb{F}[t],$$

whose roots correspond to the points of intersection $L \cap V$.

Exercise. Show that $\deg(f) = d$ implies $\deg(\varphi) = d$, hence there are at most d distinct points of intersection. In other words: $\#(L \cap V) \le \deg(f)$.

In particular, we observe that $\mathbf{p} \in L \cap V$ if and only if $t = 0$ is a root of $\varphi(t)$. To examine this point of intersection more closely, we consider the Taylor expansion of $f(\mathbf{x})$ near $\mathbf{x} = \mathbf{p}$:

$$f(\mathbf{x}) = f(\mathbf{p}) + (\nabla f)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}) + \frac{1}{2}(\mathbf{x}-\mathbf{p})^T (Hf)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}) + \text{higher terms}.$$

Then we substitute $\mathbf{x} = \mathbf{p} + t\mathbf{v}$ to obtain

$$\varphi(t) = f(\mathbf{p}) + t\cdot(\nabla f)_{\mathbf{p}}\mathbf{v} + \frac{t^2}{2}\cdot \mathbf{v}^T (Hf)_{\mathbf{p}}\mathbf{v} + \text{higher terms}.$$

We see again that $\varphi(0) = f(\mathbf{p})$ so that $\mathbf{p} \in V$ if and only if $\varphi(0) = 0$. More generally, suppose that $t = 0$ is a root of $\varphi(t)$ with multiplicity m, i.e., that $t^m | \varphi(t)$ and $t^{m+1} \nmid \varphi(t)$. In this case we define the *intersection multiplicity*:

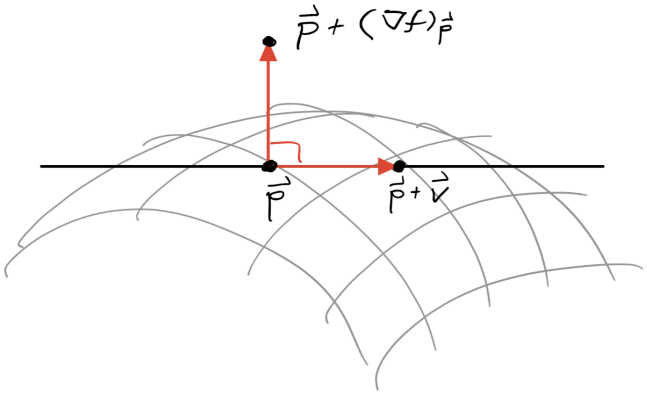
$$m = I_{\mathbf{p}}(L, V) = \text{ “intersection multiplicity of } L \text{ and } V \text{ at the point } \mathbf{p}.\text{”}$$

We say that the line L is *tangent* to the hypersurface V when the intersection multiplicity is at least 2.

Tangent Spaces and Singular Points of a Hypersurface. Given a parametrized line $L : \mathbf{p}+t\mathbf{v}$ and a hypersurface $V : f(\mathbf{x}) = 0$, we consider the polynomial $\varphi(t) := f(\mathbf{p}+t\mathbf{v}) \in \mathbb{F}[t]$ and we say that L is tangent to V at the point $\mathbf{p}$ when $t = 0$ is a root of $\varphi(t)$ of multiplicity at least two. From the above equation we see that this is equivalent to the following two conditions:

- $f(\mathbf{p}) = 0$,
- $(\nabla f)_{\mathbf{p}}\mathbf{v} = 0$.

In geometric terms, the second condition says that the vector $\mathbf{v}$ is perpendicular to the gradient vector $(\nabla f)_{\mathbf{p}}$ at the point $\mathbf{p}$. Picture:



If $(\nabla f)_{\mathbf{p}} \neq \mathbf{0}$ then the collection of tangent vectors at $\mathbf{p}$ forms an $(n-1)$-dimensional hyperplane called the *tangent space* $T_{\mathbf{p}}V \subseteq R^n$:

$$T_{\mathbf{p}}V : (\nabla f)_{\mathbf{p}}(\mathbf{x} - \mathbf{p}) = 0.$$

If we use the notation $f_{x_i} := D_{x_i} f$ then this becomes

$$f_{x_1}(\mathbf{p})(x_1 - p_1) + f_{x_2}(\mathbf{p})(x_2 - p_2) + \cdots + f_{x_n}(\mathbf{p})(x_n - p_n) = 0.$$

In this case we will say that $\mathbf{p}$ is a *smooth point* of the hypersurface V. On the other hand, if $(\nabla f)_{\mathbf{p}} = \mathbf{0}$ then **every** vector $\mathbf{v}$ satisfies $(\nabla f)_{\mathbf{p}}\mathbf{v} = 0$ and we could say that the tangent space is all of $\mathbb{F}^n$. In this case we say that $\mathbf{p}$ is a *singular point* of V.

To emphasize, we say that $\mathbf{p} \in V$ is a smooth point when $\dim T_{\mathbf{p}}V = n - 1$ and a singular point when $\dim T_{\mathbf{p}}V = n$. ///

You might wonder how the tangent space transforms under change of coordinates. Let $\Phi : R^n \to R^n$ be a polynomial change of coordinates, i.e., a vector of polynomials $\Phi(\mathbf{x}) \in R[\mathbf{x}]^n$,

and consider polynomials $f, g \in \mathbb{F}[\mathbf{x}]$ satisfying $f(\mathbf{x}) = (g \circ \Phi)(\mathbf{x}) = g(\Phi(\mathbf{x}))$. We recall from the chain rule that

$$\begin{aligned} \nabla(g \circ \Phi) &= (\nabla g \circ \Phi) D\Phi \\ \nabla f &= (\nabla g \circ \Phi) D\Phi, \end{aligned}$$

and evaluating this at a point $\mathbf{p}$ gives

$$(\nabla f)_{\mathbf{p}} = (\nabla g)_{\Phi(\mathbf{p})} (D\Phi)_{\mathbf{p}},$$

If the Jacobian matrix $(D\Phi)_{\mathbf{p}}$ is invertible then we observe that $(\nabla f)_{\mathbf{p}} = \mathbf{0}$ if and only if $(\nabla g)_{\Phi(\mathbf{p})} = \mathbf{0}$. In other words, the hypersurface $f(\mathbf{x}) = 0$ is singular at $\mathbf{p}$ if and only if the hypersurface $g(\mathbf{x}) = 0$ is singular at $\Phi(\mathbf{p})$. If $\Phi(\mathbf{x}) = A\mathbf{x}$ is an invertible linear transformation then we have $(D\Phi)_{\mathbf{p}} = A$ for all $\mathbf{p}$ and this remark applies at every point.

Oct 2: Projective Space in General

Let $\mathbb{F}$ be a field. We define the *n-dimensional projective space* $\mathbb{FP}^n$ by analogy with the real projective plane. That is, we set

$$\mathbb{FP}^n := (\mathbb{F}^{n+1} \setminus \mathbf{0})/(\text{nonzero scalars}).$$

In other words, we have $\mathbf{x} = (x_1, \ldots, x_{n+1}) \sim \mathbf{x}' = (x'_1, \ldots, x'_{n+1})$ if and only if there exists a nonzero scalar $\lambda \in \mathbb{F}$ such that $x'_i = \lambda x_i$ for all i. Let $(x_1 : \cdots : x_{n+1})$ denote the equivalence class of $(x_1, \ldots, x_{n+1})$.

We observe that projective space $\mathbb{FP}^n$ is covered by $n+1$ overlapping copies of the affine space $\mathbb{F}^n$, which are called *affine charts*. To see this, let $U_i \subseteq \mathbb{FP}^n$ be the set of points with nonzero ith coordinate. After scaling by this coordinate we obtain a unique expression with ith coordinate 1, hence we obtain the following bijection $U_i \leftrightarrow \mathbb{F}^n$:

$$(x_1 : \cdots : x_{i-1} : 1 : x_{i+1} : \cdots : x_{n+1}) \leftrightarrow (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{n+1}).$$

Polynomials do **not** define functions on projective space. However, if $F[\mathbf{x}] \in \mathbb{F}[x_1, \ldots, x_{n+1}]$ is a **homogeneous** polynomial, then for all $\lambda \in \mathbb{F} \setminus 0$ we have

$$F(\mathbf{x}) = 0 \quad \Leftrightarrow \quad F(\lambda \mathbf{x}) = 0,$$

and hence we obtain a *projective hypersurface* $V_F \subseteq \mathbb{FP}^n$ defined by $F(\mathbf{x}) = 0$.

If $F(\mathbf{x}) = \mathbf{a} \bullet \mathbf{x} = a_1 x_2 + \cdots + a_{n+1} x_{n+1}$ is a *linear form* then the corresponding hypersurface is called a *projective hyperplane*:

$$H_F = H_{\mathbf{a}} : a_1 x_1 + \cdots + a_{n+1} x_{n+1} = 0.$$

In particular, let $H_i = H_{x_i} : x_i = 0$, which is the complement of the ith affine chart, $H_i = \mathbb{FP}^n \setminus U_i$. We call this the *ith coordinate hyperplane* and we note that there is a bijection

$$\begin{aligned} H_i &\leftrightarrow \mathbb{RP}^{n-1} \\ (x_1 : \cdots : x_{i-1} : 0 : x_{i+1} : \cdots : x_{n+1}) &\leftrightarrow (x_1 : \cdots : x_{i-1} : x_{i+1} : \ldots, x_{n+1}). \end{aligned}$$

More generally, I claim that any projective hyperplane $H \subseteq \mathbb{FP}^n$ is "projectively equivalent" to $\mathbb{FP}^{n-1}$. The *projective linear group* $\mathrm{PGL}_{n+1}(\mathbb{F})$ is defined as $\mathrm{GL}_{n+1}(\mathbb{F})$ modulo nonzero scalar multiplication.[11] If a polynomial $F(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ is homogeneous of degree d and if $A \in \mathrm{PGL}_{n+1}(\mathbb{F})$ then we recall from Homework Problem 1.6 that $G(\mathbf{x}) := F(A\mathbf{x})$ is also homogeneous of degree d. In this case we say that the hypersurfaces V_F and V_G are *projectively equivalent*. This shows that the degree of a hypersurface is a *projective invariant*.[12]

Theorem. Let us identify $\mathbb{FP}^{n-1} := H_{n+1} \subseteq \mathbb{FP}^n$ with the coordinate hyperplane $x_{n+1} = 0$. Then every projective hyperplane (i.e., every projective hypersurface of degree 1) is projectively equivalent to $\mathbb{FP}^{n-1}$.

Proof. Every projective hyperplane has the form $H_{\mathbf{a}} : a_1x_1 + \cdots + a_{n+1}x_n = 0$ for some nonzero vector $\mathbf{a} = (a_1, \ldots, a_{n+1})$. Let $A \in \mathrm{PGL}_{n+1}(\mathbb{F})$ be an invertible matrix with $\mathbf{a}$ as its $(n+1)$st column vector. Then we have $A^{-1}\mathbf{a} = (0, \ldots, 0, 1)$ and hence $H_{A^{-1}\mathbf{a}} \approx H_{n+1} = \mathbb{FP}^{n-1}$. In fact, we may choose A to be an orthogonal matrix, so the equivalence is achieved by generalized rotations in $\mathbb{F}^{n+1}$. □

In summary, we view n-dimensional projective space as

$$\mathbb{FP}^n = \mathbb{F}^n \cup (\text{a hyperplane at infinity}),$$

where **any** hyperplane can play the role of the hyperplane at infinity. ///

As with curves, there is a tight relationship between **affine** hypersurfaces and **projective** hypersurfaces. Let $F(\mathbf{x}) \in \mathbb{F}[x_1, \ldots, x_{n+1}]$ be a homogeneous polynomial of degree d and define the *ith de-homogenization* as the (possibly non-homogeneous) polynomial

$$F_i := F(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_{n+1}).$$

If $x_i^m | F$ and $x_i^{m+1} \nmid F$ then F_i has degree $d-m$. Conversely, if $f \in \mathbb{F}[x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{n+1}]$ is any polynomial of degree d then we define the *ith homogenization*:

$$f^i(x_1, \ldots, x_{n+1}) := x_i^d \cdot f\left(\frac{x_1}{x_i}, \ldots, \frac{x_{i-1}}{x_i}, \frac{x_{i+1}}{x_i}, \ldots, \frac{x_{n+1}}{x_i}\right).$$

We oberve that $f^i(\mathbf{x})$ is homogeneous of degree d, with $(f^i)_i = f$. Furthermore, if $F(\mathbf{x})$ is homogeneous with $x_i^m | F$ and $x_i^{m+1} \nmid F$ then we observe that $(F_i)^i = F/x_i^m$.

Exercise. Check this.

Geometric Meaning: Any affine hypersurface $V_f \in U_i$ of degree d in the ith affine chart has a unique projective completion $V_f \subseteq V_F \subseteq \mathbb{FP}^n$, where V_F is a projective hypersurface of degree d that does not contain the ith hyperplane at infinity. (If a projective hypersurface in $\mathbb{FP}^n$ does contain the ith hyperplane at infinity then the de-homogenization in the affine chart U_i

[11] There is some conflict between the notations PGL_n and PGL_{n+1}.

[12] Remark: The fundamental theorem of projective geometry, which is tricky to prove, says that every bijective map $\Phi : \mathbb{RP}^n \to \mathbb{RP}^n$ of real projective space that "preserves incidence relations" among projective subspaces has the form $\Phi = A$ for some matrix.

will have degree less than d.) Note that we can define these concepts independently of any topology on the field $\mathbb{F}$.

Remark: If $F(\mathbf{x})$ is a multiple of x_i then V_F contains H_i. What about the other direction? Suppose $a_i = 0$ implies $F(\mathbf{a}) = 0$. Expand $F = \sum x_i^k F^{(d-k)}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{n+1})$ where $F^{(d-k)}$ is homogeneous of degree $d-k$. Then we have $F^{(d)}(\mathbf{a}) = 0$ for all $\mathbf{a}$. If $\mathbb{F}$ is infinite then this implies that $F^{(d)} = 0$, hence $x_i | F$. This is an easy case of the Nullstellensatz.

Oct 5,7: Intersection Multiplicity of Lines and Hypersurfaces

Next we want to investigate projective tangent spaces of projective hypersurfaces. In order to do this we first need to discuss the concept of projective subspaces.

Projective Subspaces. If $V \subseteq \mathbb{F}^{n+1}$ is a linear subspace of dimension $d+1$ (with $d \geq 0$) then we say that $\mathbb{P}(V) := (V \setminus \{\mathbf{0}\})/(\text{scalars}) \subseteq \mathbb{FP}^n$ is a *projective subspace of dimension d.* Since linear subspaces are closed under scalar multiplication, this defines a bijection:

$$(\text{projective } d\text{-dim subspaces of } \mathbb{FP}^n) \leftrightarrow (\text{linear } d+1\text{-dim subspaces of } \mathbb{F}^{n+1})$$

[Convention: The unique 0-dimensional subspace $\{\mathbf{0}\} \subseteq \mathbb{F}^{n+1}$ corresponds to the empty set $\mathbb{P}(\mathbf{0}) = \emptyset \subseteq \mathbb{FP}^n$, which is sometimes called "the (-1)-dimensional projective subspace."] Given a linear subspace $V \subseteq \mathbb{F}^{n+1}$ we define the *orthogonal complement*:

$$V^\perp := \{\mathbf{x} \in \mathbb{F}^{n+1} : \mathbf{v} \bullet \mathbf{x} = 0 \text{ for all } \mathbf{v} \in V\} \subseteq \mathbb{F}^{n+1}.$$

A theorem of linear algebra[13] implies that $\dim V + \dim V^\perp = n+1$, and hence $V^{\perp\perp} = V$. Then we define the *projective dual* of a projective subspace $\mathbb{P}(V) \subseteq \mathbb{FP}^n$ by

$$\mathbb{P}(V)^\vee := \mathbb{P}(V^\perp),$$

and it follows that $\mathbb{P}(V)^{\vee\vee} = \mathbb{P}(V^{\perp\perp}) = \mathbb{P}(V)$. Thus projective duality gives a bijection between d-dimensional and $(n-1-d)$-dimensional projective subspaces of $\mathbb{FP}^n$.

Example: A 0-dimensional projective subspace is just a point $\mathbf{a} \in \mathbb{FP}^n$. The projective dual is the projective hyperplane $\{\mathbf{a}\}^\vee = H_{\mathbf{a}} : a_1 x_1 + \cdots + a_{n+1} x_n = 0$. The the fact that $H_{\mathbf{a}}^\vee = \{\mathbf{a}\}$ follows from the rank-nullity theorem, as above. ///

I claim that any d-dimensional projective subspace $\mathbb{P}(V) \subseteq \mathbb{FP}^n$ is projectively equivalent to the standard embedded copy of $\mathbb{FP}^d$ in $\mathbb{FP}^n$, which we can define as an intersection of coordinate hyperplanes at infinity ($H_i : x_i = 0$):

$$\mathbb{FP}^d := H_{d+2} \cap \cdots \cap H_{n+1}.$$

[13] The fundamental theorem of linear algebra says that $\dim(\text{im } A) = \dim(\text{im } A^T)$ for any linear function. Let $A : \mathbb{F}^{n+1} \to \mathbb{F}^{d+1}$ be given by a $(d+1) \times (n+1)$ matrix whose rows are a basis for V, so that $V = \text{im } A^T$ and $V^\perp = \ker A$. Then the isomorphism $\mathbb{F}^{n+1}/\ker A \approx \text{im } A$ implies that $\dim(\text{im } A) + \dim(\ker A) = n+1$, and hence $\dim V + \dim V^\perp = n+1$.

Or we can define it in terms of the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_{n+1} \in \mathbb{F}^{n+1}$:

$$\mathbb{FP}^d := \mathbb{P}(t_1\mathbf{e}_1 + \cdots + t_{d+1}\mathbf{e}_{d+1}).$$

Proof. Let $\mathbb{P}(V) \subseteq \mathbb{FP}^n$ be a d-dimensional projective subspace, where $V \subseteq \mathbb{F}^{n+1}$ is a linear subspace with basis $\mathbf{a}_1, \ldots, \mathbf{a}_{d+1} \in \mathbb{F}^{n+1}$. Let $A \in \mathrm{GL}_{n+1}(\mathbb{F})$ be any invertible matrix with first $d+1$ columns equal to $\mathbf{a}_1, \ldots, \mathbf{a}_{d+1}$. Then A^{-1} maps $\mathbb{P}(V)$ onto $\mathbb{FP}^d \subseteq \mathbb{FP}^n$. □

Example: Any projective line $L \subseteq \mathbb{FP}^n$ has the form $L = t_1\mathbf{p}_1 + t_2\mathbf{p}_2$, where $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{FP}^n$ are two distinct points, and we obtain an equivalence with $\mathbb{FP}^1$:

$$\begin{array}{ccccc} L & \leftrightarrow & \mathbb{FP}^1 & \hookrightarrow & \mathbb{FP}^n \\ t_1\mathbf{p}_1 + t_2\mathbf{p}_2 & \leftrightarrow & (t_1 : t_2) & \to & (t_1 : t_2 : 0 : \cdots : 0). \end{array}$$

By scaling, we can identify "finite points" of the line with $\mathbf{p}_1 + t\mathbf{p}_2$ ($t \in \mathbb{F}$) and the unique "point at infinity" with $\mathbf{p}_2 = 0\mathbf{p}_1 + 1\mathbf{p}_2$. If an invertible matrix $A \in \mathrm{GL}_2(\mathbb{F})$ acts on the coordinates $(t_1 : t_2) \in \mathbb{FP}^2$ then we call this a "re-parametrization" of the line:

$$\begin{array}{ccc} L & \leftrightarrow & L \\ t_1\mathbf{p}_1 + t_2\mathbf{p}_2 & \mapsto & (at_1 + bt_2)\mathbf{p}_1 + (ct_1 + dt_2)\mathbf{p}_2. \end{array}$$

///

Now let us return to the topic of projective tangent spaces. Let $L, V \subseteq \mathbb{FP}^n$ be a line and a hypersurface of degree d in projective n-dimensional space. By definition this means that $L = \{t_1\mathbf{p}_1 + t_2\mathbf{p}_2 : t_1, t_2 \in \mathbb{F}\}$ for distinct points $\mathbf{u} \neq \mathbf{v} \in \mathbb{FP}^n$ and $V = V_F : F(\mathbf{x}) = 0$ for some homogeneous polynomial $F(\mathbf{x}) \in \mathbb{F}[x_1, \ldots, x_{n+1}]$ of degree d. Using matrix notation we can write $L : P\mathbf{t}$ where P is the matrix with columns $\mathbf{p}_1, \mathbf{p}_2$ and $\mathbf{t} = (t_1, t_2)$. By substitution we obtain a homogeneous polynomial in t_1, t_2 of degree d:

$$\Phi(t_1, t_2) := F(P\mathbf{t}) \in \mathbb{F}[t_1, t_2].$$

Thus $t_1\mathbf{p}_2 + t_2\mathbf{p}_2 \in \mathbb{FP}^n$ is a point of intersection of L and V if and only if $(t_1 : t_2) \in \mathbb{FP}^1$ is a root of $\Phi(t_1, t_2)$. We have the following projective version of Descartes' Theorem.

Projective Descartes' Theorem. A projective point $(a_1 : a_2) \in \mathbb{FP}^1$ is a root of the homogeneous polynomial $\Phi(t_1, t_2) \in \mathbb{F}[t_1, t_2]$ if and only if $(a_2t_1 - a_1t_2)$ divides $\Phi(t_1, t_2)$.[14]

Proof. If $(a_2t_1 - a_1t_2) | \Phi(t_1, t_2)$ then $\Phi(a_1, a_2) = 0$. Conversely, suppose that $\Phi(a_1, a_2) = 0$ where a_1, a_2 are not both zero. There are two cases:

- $a_2 \neq 0$: Suppose that $\deg \Phi = n$ and let $\Phi(t_1, t_2) = t_2^m \Phi'(t_1, t_2)$ where $\Phi'(t_1, t_2)$ is homogeneous of degree $n - m$. Note that we still have $\Phi'(a_1, a_2) = 0$. Now consider the de-homogenization $\varphi(t_1) := \Phi(t_1, 1) = \Phi'(t_1, 1)$. Since $(a_1 : a_2) \sim (a_1/a_2 : 1)$ we have

[14] Later we will show that every factor of a homogeneous polynomial is necessarily homogeneous.

$\varphi(a_1/a_2) = \Phi(a_1/a_2, 1) = \Phi(a_1, a_2) = 0$ and the usual Descartes' Theorem implies that $(t_1 - a_1/a_2)|\varphi(t_1)$. Let's say $\varphi(t_1) = (t_1 - a_1/a_2)\psi(t_1)$ where $\psi(t_1)$ has degree $n-m-1$. Then since $t_2 \nmid \Phi'(t_1, t_2)$ we may re-homogenize to obtain

$$\Phi'(t_1, t_2) = t_2^{n-m}\varphi(t_1/t_2) = (t_1 - t_2 a_1/a_2) \cdot t_2^{n-m-1}\psi(t_1/t_2) = \frac{1}{a_2}(a_2 t_1 - a_2 t_2)\Psi(t_1, t_2),$$

where $\Psi(t_1, t_2) \in \mathbb{F}[t_1, t_2]$ is homogeneous of degree $n - m - 1$.

- $a_1 \neq 0$: The proof is symmetric.

□

We define the *multiplicity* of the root $(a_1 : a_2)$ as the highest power of $(a_2 t_1 - a_1 t_2)$ that divides the homogeneous polynomial $\Phi(t_1, t_2)$. For any point $\mathbf{p} \in \mathbb{FP}^n$, this concept allows us to define the "intersection multiplicity" of a line and hypersurface meeting at $\mathbf{p}$.[15]

Intersection Multiplicity of a Line and a Hypersurface. Consider a line and a hypersurface of degree d in projective space, $L, V \subseteq \mathbb{FP}^n$. For any point $\mathbf{p} \in L$ we want to define a number $[L \cdot V]_{\mathbf{p}} \in \mathbb{N}$ representing the *intersection multiplicity* of L and V at the point $\mathbf{p}$.

To do this, we parametrize L as $P\mathbf{t} = t_1\mathbf{p}_1 + t_2\mathbf{p}_2$ and we let $V = V_F$ for some homogeneous polynomial $F(\mathbf{x}) \in \mathbb{F}[x_1, \ldots, x_{n+1}]$ of degree d. Thus we obtain a homogeneous polynomial of degree d in the variables $\mathbf{t} = (t_1, t_2)$:

$$\Phi(t_1, t_2) := F(P\mathbf{t}) \in \mathbb{F}[t_1, t_2].$$

If $\mathbf{p} = P\mathbf{a} = a_1\mathbf{p}_1 + a_2\mathbf{p}_2 \in L$ then we define

$$[L \cdot V]_{\mathbf{p}} := \text{the multiplicity of } (a_1 : a_2) \text{ as a root of } \Phi(t_1, t_2).$$

I claim that this number is well-defined up to projective automorphisms of the ambient space and projective automorphisms of the line.

[Remark: If $\mathbb{F}$ is algebraically closed then it will follow from the Nullstellensatz that a hypersurface has the form V_F for some unique homogeneous polynomial F each of whose (necessarily homogeneous) irreducible factors occurs with multiplicity 1. (You will prove this on HW2 for the case of curves in the plane.) Let's assume that we have chosen such a polynomial, so the intersection multiplicity does not depend on F.]

Proof. First we show that the intersection multiplicity is well-defined up to projective equivalence. Consider any invertible matrix $A \in \mathrm{GL}_{n+1}(\mathbb{F})$. Then the line $L = P\mathbf{t}$ gets sent to $AL = AP\mathbf{t}$ and the hypersurface V_F gets sent to $AV = V_G$ where $G(\mathbf{x}) = F(A^{-1}\mathbf{x})$. This leaves the polynomial defining the intersection unchanged:

$$G(AP\mathbf{t}) = F(A^{-1}AP\mathbf{t}) = F(P\mathbf{t}).$$

[15] The uniqueness of multiplicity also follows from the fact that $\mathbb{F}[t_1, t_2]$ is a UFD, which you will prove on Homework 2 Problem 2.

In other words, we have $[L \cdot V]_{\mathbf{p}} = [AL \cdot AV]_{A\mathbf{p}}$.

Now we show that intersection multiplicity is unchanged by re-parametrizing the line. Consider the polynomial $\Phi(\mathbf{t}) = F(P\mathbf{t})$ and let $\Sigma \in \mathrm{GL}_2(\mathbb{F})$ be a re-parametrization of the line, so that $L = P\mathbf{t} = P\Sigma\mathbf{t}$. Then we observe that $\Sigma\mathbf{t}$ is a root of $\varphi(\Sigma\mathbf{t})$ with multiplicity m if and only if $\mathbf{t}$ is a root of $\varphi(\mathbf{t})$ with multiplicity m. Reason: Linear substitution preserves the degrees of the homogeneous factors. □

Remark: The idea is that the intersection multiplicity $[L \cdot V]_{\mathbf{p}}$ is an "intrinsic" property of the geometry, independent of any algebraic mode of expression. Historically it has been **very difficult** to make this precise. The intersection of lines and hypersurfaces is the model for further generalization.

Jargon:

$[L \cdot V]_{\mathbf{p}}$	we say that L and V
0	do not intersect at $\mathbf{p}$
1	intersect transversely at $\mathbf{p}$
≥ 2	are tangent at $\mathbf{p}$

Oct 12: Projective Tangent Spaces

Now we will use the concept of intersection multiplicity to define and compute projective tangent spaces to projective hypersurfaces.

Let $F(\mathbf{x}) \in \mathbb{F}[x_1, \ldots, x_{n+1}]$ be a homogeneous polynomial and let $L \subseteq \mathbb{FP}^n$ be a line in projective space containing a point $\mathbf{p} \in \mathbb{FP}^n$. We can parametrize this as $L : \mathbf{p} + t\mathbf{q}$ where $\mathbf{q} \neq \mathbf{p}$ is any another point on the line. For a specific representation $\mathbf{p} = (p_1, \ldots, p_{n+1})$ we we will compute the Taylor series expansion of F near $\mathbf{x} = \mathbf{p}$:

$$F(\mathbf{x}) = F(\mathbf{p}) + (\nabla F)_{\mathbf{p}}(\mathbf{x} - \mathbf{p}) + \frac{1}{2}(\mathbf{x} - \mathbf{p})^T (HF)_{\mathbf{p}}(\mathbf{x} - \mathbf{p}) + \text{higher terms.}$$

Note that the Taylor series is an **affine** concept, but we can still use it to get **projective** information. Choose a specific representation $\mathbf{q} = (q_1, \ldots, q_{n+1})$ and substitute this into the polynomial F to obtain

$$\Phi(t) := F(\mathbf{p} + t\mathbf{q}) = F(\mathbf{p}) + t \cdot (\nabla F)_{\mathbf{p}}\mathbf{q} + \frac{t^2}{2} \cdot \mathbf{q}^T (HF)_{\mathbf{p}}\mathbf{q} + \text{higher terms.}$$

We showed in the previous lecture that the intersection multiplicity $[L \cdot V_F]_{\mathbf{p}}$ of the projective line L and the projective hypersurface V_F is independent of the parametrization of the line. Therefore $[L \cdot V_F]_{\mathbf{p}}$ equals the multiplicity of $t = 0$ as a root of $\Phi(t)$.

Projective Tangent Spaces and Singular Points of Hypersurfaces. By definition we say that $L : \mathbf{p} + t\mathbf{q}$ and $V_F : F(\mathbf{x}) = 0$ are *tangent* when $[L \cdot V_F]_{\mathbf{p}} \geq 2$, which happens if and only if

- $F(\mathbf{p}) = 0$,
- $(\nabla F)_{\mathbf{p}}\mathbf{q} = \mathbf{0}$.

If $(\nabla F)_{\mathbf{p}} \neq \mathbf{0}$ then this second equation defines a projective hyperplane $T_{\mathbf{p}}V_F = H_{(\nabla F)_{\mathbf{p}}}$, called the *projective tangent space* at $\mathbf{p} \in V_F$. In this case we say that $\mathbf{p}$ is a *smooth (regular) point* of V_F. If $(\nabla F)_{\mathbf{p}} = \mathbf{0}$ then every line through $\mathbf{p}$ is tangent to V_F and we declare that $T_{\mathbf{p}}V_F = \mathbb{FP}^n$ is the whole projective space. In this case we say that $\mathbf{p}$ is a *singular point* of V_F. By the theorem in the previous lecture, these notions are well-defined up to projective automorphisms of the ambient space and projective automorphisms of the line.

Summary: We say that $\mathbf{p} \in V_F$ is a smooth point when $\dim T_{\mathbf{p}}V_F = n-1$ and a singular point when $\dim T_{\mathbf{p}}V_F = n$, as projective subspaces. ///

To understand this concept completely we should investigate how it behaves with respect to affine charts. The following Lemma is the key fact.

Problem 1.8. Euler's Homogeneous Function Theorem.[16] Let $F(\mathbf{x}) \in \mathbb{F}[x_1, \ldots, x_{n+1}]$ be a polynomial over a ring $\mathbb{F}$ and consider the following three conditions:

(H1) $F(\mathbf{x}) = F^{(d)}(\mathbf{x})$,

(H2) $F(\mathbf{x}) \neq 0$ and $F(\lambda\mathbf{x}) = \lambda^d F(\mathbf{x})$ for all $\lambda \in \mathbb{F} \setminus 0$,

(H3) $d \cdot F(\mathbf{x}) = (\nabla F)_{\mathbf{x}}\mathbf{x} = x_1 F_{x_1}(\mathbf{x}) + \cdots + x_{n+1}F_{x_n}(\mathbf{x})$.

We saw in Problem 1.5 that (H1)⇒(H2) for any field $\mathbb{F}$ and that (H2)⇒(H1) when $\mathbb{F}$ is infinite. Now I claim that (H1)⇒(H3) for any field $\mathbb{F}$ and that (H3)⇒(H1) when $\mathbb{F}$ has characteristic zero (hence also is infinite).

Remark: In particular, if $F(\mathbf{p}) = 0$ then (H3) implies that the vectors $(\nabla F)_{\mathbf{p}}$ and $\mathbf{p}$ are perpendicular. Geometrically, the zero set of a homogeneous polynomial is a (generalized) cone over the origin. Thus the normal vector is always perpendicular to the radial vector. ///

Proof. (H1)⇒(H3): Assume that $F(\mathbf{x}) = \sum_I a_I \mathbf{x}^I$ is homogeneous of degree d, i.e., with each monomial degree $I = (i_1, \ldots, i_{n+1})$ satisfying $\sum I = i_1 + \cdots + i_{n+1} = d$. Note that for each

[16] Actually, Euler proved this for differentiable functions over $\mathbb{R}$; not just polynomials.

variable x_k and each monomial $\mathbf{x}^I$ we have $x_k D_{x_k}\mathbf{x}^I = i_k\mathbf{x}^I$ and hence

$$\begin{aligned}(\nabla F)_{\mathbf{x}}\mathbf{x} &= \sum_k x_k D_{x_k} F \\ &= \sum_I a_I \sum_{x_k} D_{x_k}\mathbf{x}^I \\ &= \sum_I a_I \sum_k i_k \mathbf{x}^I \\ &= \sum_I a_I (i_1 + \cdots + i_{n+1})\mathbf{x}^I \\ &= \sum_I a_I d\mathbf{x}^I \\ &= d \cdot F(\mathbf{x}).\end{aligned}$$

(H3)⇒(H1): Now let us assume that $d \cdot F(\mathbf{x}) = (\nabla F)_{\mathbf{x}}\mathbf{x}$ and that $\mathbb{F}$ has characteristic zero. Let $F(\mathbf{x}) = \sum_k F^{(k)}(\mathbf{x})$ be the filtration of F into homogeneous parts. Since $(\nabla F)_{\mathbf{x}}\mathbf{x}$ is a linear function of F and since (H1)⇒(H3) we see that

$$\begin{aligned}d \cdot F &= (\nabla F)_{\mathbf{x}}\mathbf{x} \\ &= \sum_k (\nabla F^{(k)})_{\mathbf{x}}\mathbf{x} \\ &= \sum_k k \cdot F^{(k)}.\end{aligned}$$

Now let y be another variable and substitute $\mathbf{x} \mapsto y\mathbf{x}$. Then since (H1)⇒(H2) we have

$$\begin{aligned}d \cdot F(y\mathbf{x}) &= \sum_k k \cdot F^{(k)}(y\mathbf{x}) \\ d\sum_k F^{(k)}(y\mathbf{x}) &= \sum_k k \cdot F^{(k)}(y\mathbf{x}) \\ \sum_k dy^k F^{(k)}(\mathbf{x}) &= \sum_k k y^k F^{(k)}(\mathbf{x}).\end{aligned}$$

We can regard this as an identity of polynomials in $\mathbb{F}[\mathbf{x}][y]$ and hence the coefficient of y^k on each side is the same:

$$\begin{aligned}d \cdot F^{(k)}(\mathbf{x}) &= k \cdot F^{(k)}(\mathbf{x}) \\ (d-k)F^{(k)}(\mathbf{x}) &= 0.\end{aligned}$$

Finally, since $\mathbb{F}$ has characteristic zero we see that $d \neq k$ in $\mathbb{Z}$ implies that $d - k \neq 0$ in $\mathbb{F}$ and hence $F^{(k)}(\mathbf{x}) = 0$. It follows that $F(\mathbf{x}) = F^{(d)}(\mathbf{x})$ as desired. □

This lemma allows us to describe the behavior of projective tangent spaces in affine charts.

Affine vs Projective Tangent Spaces. Let $\mathbf{p} \in V_F \subseteq \mathbb{FP}^n$ be a point on a projective hypersurface and let $T_{\mathbf{p}}V_F$ be the projective tangent space at this point. Let f be the ith de-homogenization of F (which is obtained by substituting $x_i = 1$). If $\mathbf{p} \notin H_i$ is not on the ith hyperplane at infinity then I claim that the affine tangent space $T_{\mathbf{p}}V_f \subseteq U_i \subseteq \mathbb{FP}^n$ is the ith de-homogenization of the projective tangent space. Conversely, if the projective tangent space V_F does not contain H_i then V_F is the ith homogenization of V_f.

Proof. Let F be homogeneous of degree d. The projective tangent space $T_{\mathbf{p}}V_F$ is defined by the folllowing equation:

$$\begin{aligned} (\nabla F)_{\mathbf{p}}\mathbf{x} &= 0 \\ x_1F_{x_1}(\mathbf{p}) + \cdots + x_{n+1}F_{x_{n+1}}(\mathbf{p}) &= 0. \end{aligned} \tag{$*$}$$

Since $\mathbf{p} \in V_F$ we have $F(\mathbf{p}) = 0$ and then Euler's formula tells us that

$$p_1F_{x_1}(\mathbf{p}) + \cdots + p_{n+1}F_{x_{n+1}}(\mathbf{p}) = d \cdot F(\mathbf{p}) = 0.$$

Thus we can also express the projective tangent space as

$$(x_1 - p_1)F_{x_1}(\mathbf{p}) + \cdots + (x_{n+1} - p_{n+1})F_{x_{n+1}}(\mathbf{p}) = 0.$$

Now suppose that $\mathbf{p} \in U_i$ (i.e., $p_i \neq 0$) and let $\mathbf{p} = (p_1, \ldots, p_{i-1}, 1, p_{i+1}, \ldots, p_{n+1}) \in U_i$ be the standard representation in this chart. Let f be the ith de-homogenization of F (i.e., set $x_i = 1$). If $k \neq i$ then we observe that f_{x_k} is the de-homogenization F_{x_k} and if $x_i \nmid F$ then $x_i \nmid F_{x_k}$, so that F_{x_k} is also the homogenization of f_{x_k}.

Now the affine tangent space $T_{\mathbf{p}}V_f \subseteq U_i$ is defined by

$$\begin{aligned} f_{x_1}(\mathbf{p})(x_1 - p_1) + \cdots + f_{x_{i-1}}(\mathbf{p})(x_{i-1} - p_{i-1}) \\ + f_{x_{i+1}}(\mathbf{p})(x_{i+1} - p_{i+1}) + \cdots + f_{x_{n+1}}(\mathbf{p})(x_{n+1} - p_{n+1}) = 0. \end{aligned}$$

And since homogenization distributes over sums and products, the ith homogenization of this affine space is the projective space defined by

$$\begin{aligned} F_{x_1}(\mathbf{p})(x_1 - p_1x_i) + \cdots + F_{x_{i-1}}(\mathbf{p})(x_{i-1} - p_{i-1}x_i) \\ + F_{x_{i+1}}(\mathbf{p})(x_{i+1} - p_{i+1}x_i) + \cdots + F_{x_{n+1}}(\mathbf{p})(x_{n+1} - p_{n+1}x_i) = 0. \end{aligned}$$

To see that this is the same as $(*)$ we again use Euler's formula:

$$\begin{aligned} p_1F_{x_1}(\mathbf{p}) + \cdots + 1F_{x_i}(\mathbf{p}) + \cdots + p_{n+1}F_{x_{n+1}}(\mathbf{p}) &= 0 \\ p_1F_{x_i}(\mathbf{p}) + \cdots + p_{i-1}F_{x_{i-1}}(\mathbf{p}) + p_{i+1}F_{x_{i+1}}(\mathbf{p}) + \cdots + \mathbf{p}_{n+1}F_{x_{n+1}}(\mathbf{p}) &= -F_{x_i}(\mathbf{p}). \end{aligned}$$

□

Geometric meaning: The tangent space at a point is determined by any local neighborhood. Of course this is intuitively obvious. The difficulty of algebraic geometry is always to show that the algebra matches the geometric intuition.

Sept 30 and Oct 26: Zariski Tangent Spaces

To end this chapter, we begin to translate the notion of singular points and tangent spaces into the language of rings and ideals. (The reason for doing this will become more clear in the next chapter.) The following result is an abstract version of Taylor expansion.

The Ideal of a Point. Let $\mathbb{F}$ be a field and consider the ring $\mathbb{F}[\mathbf{x}] = \mathbb{F}[x_1, \ldots, x_n]$. Then for any point $\mathbf{p} = (p_1, \ldots, p_n) \in \mathbb{F}^n$ we consider the ideal

$$M_{\mathbf{p}} := \sum_{i=1}^{n} (x_i - p_i)\mathbb{F}[\mathbf{x}].$$

(a) The ideal $M_{\mathbf{p}}$ is the kernel of the evaluation homomorphism $\mathbb{F}[\mathbf{x}] \to \mathbb{F}$ defined by $f(\mathbf{x}) \mapsto f(\mathbf{p})$. Since this homomorphism is surjective onto a field, the kernel is a maximal ideal.

(b) Given ideals $A, B \subseteq \mathbb{F}[\mathbf{x}]$ we define the *product ideal* $AB \subseteq \mathbb{F}[\mathbf{x}]$ as the smallest ideal containing the set $\{fg : f \in A, g \in B\}$. Then for all $k \geq 1$ we have

$$M_{\mathbf{p}}^k = \sum_{i_1, \ldots, i_k = 1}^{n} (x_{i_1} - p_{i_1}) \cdots (x_{i_k} - p_{i_k})\mathbb{F}[\mathbf{x}].$$

(c) The Taylor expansion at $\mathbf{x} = \mathbf{p}$ defines an isomorphism of $\mathbb{F}$-vector spaces:

$$\mathbb{F}[\mathbf{x}] \approx (\mathbb{F}[\mathbf{x}]/M_{\mathbf{p}}) \oplus (M_{\mathbf{p}}/M_{\mathbf{p}}^2) \oplus (M_{\mathbf{p}}^2/M_{\mathbf{p}}^3) \oplus \cdots,$$

where the vector space $M_{\mathbf{p}}^k/M_{\mathbf{p}}^{k+1}$ has dimension $\binom{n+k-1}{k}$. (Say $M_{\mathbf{p}}^0 := \mathbb{F}[\mathbf{x}]$.)

(d) Any generating set for the ideal $M_{\mathbf{p}}$ maps to a spanning set for the vector space $M_{\mathbf{p}}/M_{\mathbf{p}}^2$, hence the ideal $M_{\mathbf{p}}$ cannot be generated by fewer than n elements. [In particular, if $n \geq 2$ then the ring $\mathbb{F}[\mathbf{x}]$ is not a PID.]

Remarks:

- From linear algebra know that the point $\mathbf{p} = (p_1, \ldots, p_n) \in \mathbb{F}^n$ can be expressed as the intersection of the n hyperplanes $x_i - p_i = 0$, but no fewer than n hyperplanes. Part (c) of this theorem says that (morally) a point in $\mathbb{F}^n$ cannot be expressed as an intersection of fewer than n **hypersurfaces**. However, for this to be strictly true we need to work over an algebraically closed field.

- To be precise, if $\mathbb{F}$ is algebraically closed then *Hilbert's Nullstellensatz* (zero places theorem) says that **every** maximal ideal of $\mathbb{F}[\mathbf{x}]$ has the form $M_{\mathbf{p}}$ for some point. In other words, we have a bijection:

$$\begin{aligned} \text{points of } \mathbb{F}^n &\leftrightarrow \text{maximal ideals of } \mathbb{F}[x_1, \ldots, x_n] \\ \mathbf{p} &\leftrightarrow M_{\mathbf{p}}. \end{aligned}$$

  The proof is quite involved so we defer it to the next chapter.

- To see that algebraic closure is necessary, one can check that $M = x\mathbb{R}[x,y] + (y^2+1)\mathbb{R}[x,y]$ is a maximal ideal of $\mathbb{R}[x,y]$ and does not have the form $M_{\mathbf{p}}$. Hint: It is maximal because $\mathbb{R}[x,y]/M \approx \mathbb{C}$ is a field, and it is not of the form $M_{\mathbf{p}}$ because y^2+1 is irreducible in $\mathbb{R}[x,y]$.

Proof. (a): We want to show that $M_{\mathbf{p}} = \{f \in \mathbb{F}[\mathbf{x}] : f(\mathbf{p}) = 0\}$. For one direction, let $f \in M_{\mathbf{p}}$ so that $f = (x_1 - p_1)f_1 + \cdots (x_n - p_n)f_n$ for some $f_1, \ldots, f_n \in \mathbb{F}[\mathbf{x}]$. Then we have

$$f(\mathbf{p}) = (p_1 - p_1)f_1(\mathbf{p}) + \cdots (p_n - p_n)f_n(\mathbf{p}) = 0 + \cdots + 0 = 0.$$

For the other direction, let $f \in \mathbb{F}[\mathbf{x}]$ and consider the Taylor expansion at $\mathbf{x} = \mathbf{p}$:

$$f(\mathbf{x}) = f(\mathbf{p}) + \sum_{i_1+\cdots+i_n \geq 1} \frac{1}{i_1! \cdots i_n!} \left(\frac{\partial^{i_1+\cdots+i_n} f}{\partial x_1^{i_1} \cdots \partial x_n^{i_n}} \right)_{\mathbf{p}} (x_1 - p_1)^{i_1} \cdots (x_n - p_n)^{i_n}.$$

If $f(\mathbf{p}) = 0$ then since each term of the sum has $i_k \geq 1$ for some k, we conclude that $f \in M_{\mathbf{p}}$.

(b): Suppose that $A, B \subseteq \mathbb{F}[\mathbf{x}]$ are finitely generated ideals:

$$A = f_1\mathbb{F}[\mathbf{x}] + \cdots + f_\ell\mathbb{F}[\mathbf{x}],$$
$$B = g_1\mathbb{F}[\mathbf{x}] + \cdots + g_m\mathbb{F}[\mathbf{x}].$$

Then I claim that

$$AB = \sum_{i=1}^{\ell} \sum_{j=1}^{m} f_i g_j \mathbb{F}[\mathbf{x}].$$

Indeed, since $f_i g_j$ is in AB for all i, j we see that any $\mathbb{F}[\mathbf{x}]$-linear combination of these polynomials is in AB. Conversely, any element of AB is a sum of terms of the form fgh with $f \in A$, $g \in B$, $h \in \mathbb{F}[\mathbf{x}]$. Then writing $f = f_1\varphi_1 + \cdots + f_\ell\varphi_\ell$ and $g = g_1\gamma_1 + \cdots + g_m\gamma_m$ gives

$$fgh = \sum_{i=1}^{\ell} \sum_{j=1}^{m} f_i f_j (\varphi_i \gamma_j h) \in \sum_{i=1}^{\ell} \sum_{j=1}^{m} f_i g_j \mathbb{F}[\mathbf{x}].$$

Since $M_{\mathbf{p}}$ is generated by the polynomials $x_1 - p_1, \ldots, x_n - p_n$, we apply the above result and induction to conclude that $M_{\mathbf{p}}^k$ is generated by the polynomials $(x_{i_1} - p_{i_1}) \cdots (x_{i_k} - p_{i_k})$.

(c): The Taylor expansion tells us that each polynomial $f \in \mathbb{F}[\mathbf{x}]$ has a unique expression of the form $f(x) = \sum_{I \in \mathbb{N}^n} a_I (\mathbf{x} - \mathbf{p})^I$. Namely, we must have $a_I = D_{\mathbf{x}}^I(f)_{\mathbf{p}}/I!$, where $D_{\mathbf{x}}^I = (\partial^{i_1+\cdots+i_n})/(\partial x_1^{i_1} \cdots \partial x_n^{i_n})$ and $I! = i_1! \cdots i_n!$. Now for any f we define the k-th homogeneous piece at $\mathbf{x} = \mathbf{p}$:

$$f_{\mathbf{p}}^{(k)}(\mathbf{x}) := \sum_{\sum I = k} a_I (\mathbf{x} - \mathbf{p})^I \in M_{\mathbf{p}}^k.$$

Then we send f to a formal sequence sequence of cosets:

$$f \mapsto \left(f_{\mathbf{p}}^{(0)} + M_{\mathbf{p}}\,,\, f_{\mathbf{p}}^{(1)} + M_{\mathbf{p}}^2\,,\, f_{\mathbf{p}}^{(2)} + M_{\mathbf{p}}^3\,,\, \ldots \right)$$

To see that this is an isomorphism of vector spaces, we will show that the set $\{(\mathbf{x}-\mathbf{p})^I + M_{\mathbf{p}}^{k+1} : \sum I = k\}$ is a basis for the vector space $M_{\mathbf{p}}^k/M_{\mathbf{p}}^{k+1}$. Part (b) shows that is is a spanning set. Now assume for contradiction that we have a non-trivial linear relation. That is, suppose that for some $I \in \mathbb{N}^n$ with $\sum I = k$ we have

$$(\mathbf{x}-\mathbf{p})^I + \sum_{\sum J=k, J\neq I} a_J(\mathbf{x}-\mathbf{p})^J \in M_{\mathbf{p}}^{k+1}. \qquad (*)$$

For any $I, J \in \mathbb{N}^n$, we recall that the differential operator $D_{\mathbf{x}}^I$ satisfies[17]

$$D_{\mathbf{x}}^I(\mathbf{x}-\mathbf{p})^J = \begin{cases} \text{non-constant} & I < J \\ \text{non-zero constant} & I = J, \\ 0 & I \not\leq J. \end{cases}$$

Since $\sum I = \sum J$ and $I \neq J$ imply that $I \not\leq J$, we see that $D_{\mathbf{x}}^I$ applied to the polynomial in $(*)$ gives a non-zero constant. On the other hand, since $\sum I < \sum J$ implies that $I < J$ or $I \not\leq J$, we observe that $D_{\mathbf{x}}^I$ applied to any element of $M_{\mathbf{p}}^{k+1}$ gives either zero or a non-constant polynomial. Contradiction.

Now we compute the dimension of the vector space $M_{\mathbf{p}}^k/M_{\mathbf{p}}^{k+1}$. By the previous result this is just the number of n-tuples $I = (i_1, \ldots, i_n) \in \mathbb{N}^n$ satisfying $\sum I = k$. We count these with the "stars and bars" trick: For each such n-tuple we write down the $(n+k-1)$-tuple consisting of i_1 copies of 0 followed by a single 1, followed by i_2 copies of 0 followed by a single 1, etc., and ending with i_n copies of 0. The result is a binary word of length $(n+k-1)$ containing k copies of 0 and $n-1$ copies of 1. The number of these is the binomial coefficient $\binom{n+k-1}{k} = \binom{n+k-1}{n-1}$.

Remarks: The generating function for the dimensions is particularly nice:

$$\sum_{k\geq 0} \dim_{\mathbb{F}}(M_{\mathbf{p}}^k/M_{\mathbf{p}}^{k+1})\lambda^k = \sum_{k\geq 0} \binom{n+k-1}{k}\lambda^k = 1/(1-\lambda)^k.$$

This is called the *Hilbert series* of the graded ring $\mathbb{F}[\mathbf{x}]$.

(d): From part (c) we have a surjective linear map $M_{\mathbf{p}} \to M_{\mathbf{p}}/M_{\mathbf{p}}^2$ defined by sending the polynomial $f \in M_{\mathbf{p}}$ to the coset $(\nabla f)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}) + M_{\mathbf{p}}^2$. Now suppose that the ideal $M_{\mathbf{p}}$ can be generated by m elements:

$$M_{\mathbf{p}} = f_1\mathbb{F}[\mathbf{x}] + \cdots + f_m\mathbb{F}[\mathbf{x}].$$

Our goal is to show that every element of $M_{\mathbf{p}}/M_{\mathbf{p}}^2$ is an $\mathbb{F}$-linear combination of the cosets $(\nabla f_i)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}) + M_{\mathbf{p}}^2$. To see this, we consider an arbitrary coset $(\nabla f)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}) + M_{\mathbf{p}}^2$ with $f \in M_{\mathbf{p}}$. By hypothesis we can write $f = f_1g_1 + \cdots + f_mg_m$ for some $g_1, \ldots, g_m \in \mathbb{F}[\mathbf{x}]$. Then by applying the chain rule for ∇ and the fact that $f_i(\mathbf{p}) = 0$ for all i, we obtain

$$(\nabla f)_{\mathbf{p}} = \sum_i \left[(\nabla f_i)_{\mathbf{p}} g_i(\mathbf{p}) + \cancel{f_i(\mathbf{p})}(\nabla g_i)_{\mathbf{p}}\right] = \sum_i (\nabla f_i)_{\mathbf{p}} g_i(\mathbf{p}).$$

[17] See the Sept 24 lecture for more details.

This implies that $(\nabla f)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}) + M_{\mathbf{p}}^{k+1} = \sum_i g_i(\mathbf{p})(\nabla f_i)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}) + M_{\mathbf{p}}^{k+1}$ is an $\mathbb{F}$-linear combination of the cosets $(\nabla f_i)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}) + M_{\mathbf{p}}^{k+1}$, as desired. □

The fact that the vector space $M_{\mathbf{p}}/M_{\mathbf{p}}^2$ is n-dimensional captures the fact that the affine space $\mathbb{F}^n$ is n-dimensional in a neighborhood of the point $\mathbf{p}$. That's not very interesting, but now we will extend this idea to points on a hypersurface.

Zariski Tangent Spaces to Hypersurfaces. Let $f(\mathbf{x}) \in \mathbb{F}[x_1, \ldots, x_n]$ and consider the affine hypersurface $V = V_f \subseteq \mathbb{F}^n$. Recall from the lectures on Sept 28,30 that the affine tangent space $T_{\mathbf{p}}V_f \subseteq \mathbb{F}^n$ at a point $\mathbf{p} \in V_f$ has the equation

$$\begin{aligned}(\nabla f)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}) &= 0\\ f_{x_1}(\mathbf{p})(x_1-p_n) + \cdots + f_{x_n}(\mathbf{p})(x_n-p_n) &= 0.\end{aligned}$$

Our goal is to express this tangent space (and hence its dimension) in terms of the maximal ideal $M_{\mathbf{p}} \subseteq \mathbb{F}[\mathbf{x}]$. For this purpose, it turns out that we should really consider the *cotangent space* $(T_{\mathbf{p}}V_f)^*$, which is the vector space of linear functions $T_{\mathbf{p}}V_f \to \mathbb{F}$. The idea is to regard the gradient $(\nabla f)_{\mathbf{p}}$ not as a vector, but as a *covector*, sending each vector $\mathbf{v} \in \mathbb{F}^n$ to the dot product $(\nabla f)_{\mathbf{p}}\mathbf{v} \in \mathbb{F}$. Thus we have a linear function from $\mathbb{F}[\mathbf{x}]$ to the dual space $(\mathbb{F}^n)^*$:

$$\begin{aligned}\mathbb{F}[\mathbf{x}] &\to (\mathbb{F}^n)^*\\ g &\mapsto [\mathbf{v} \mapsto (\nabla g)_{\mathbf{p}}\mathbf{v}]\end{aligned}$$

This map is surjective because it sends the polynomial x_i to the ith coordinate function $\mathbf{v} \mapsto v_i$. Indeed, we observe that ∇x_i, and also $(\nabla x_i)_{\mathbf{p}}$, is the constant vector $(0, \ldots, 0, 1, 0, \ldots, 0)$ with a 1 in the ith position, hence $(\nabla x_i)_{\mathbf{p}}\mathbf{v} = v_i$. In fact, the polynomial $x_i - p_i \in M_{\mathbf{p}}$ also gets sent to the ith coordinate function, so the restriction $M_{\mathbf{p}} \to (\mathbb{F}^n)^*$ is still surjective. By composing this with the (surjective) restriction $(\mathbb{F}^n)^* \to (T_{\mathbf{p}}V_f)^*$ we obtain a surjective map:

$$\begin{aligned}M_{\mathbf{p}} &\to (T_{\mathbf{p}}V_f)^*\\ g &\mapsto [\mathbf{v} \mapsto (\nabla g)_{\mathbf{p}}\mathbf{v}]\end{aligned}$$

If $I \subseteq M_{\mathbf{p}}$ is the kernel of this map then I claim that $I = M_{\mathbf{p}}^2 + f(\mathbf{x})\mathbb{F}[\mathbf{x}]$ (which is an ideal of $\mathbb{F}[\mathbf{x}]$). Thus we obtain an isomorphism of vector spaces:

$$(T_{\mathbf{p}}V_f)^* \approx \frac{M_{\mathbf{p}}}{M_{\mathbf{p}}^2 + f(\mathbf{x})\mathbb{F}[\mathbf{x}]}.$$

Proof. Let $I \subseteq M_{\mathbf{p}}$ be the kernel surjective linear map $M_{\mathbf{p}} \to (T_{\mathbf{p}}V_f)^*$. By definition we have

$$\begin{aligned}I &= \{g \in M_{\mathbf{p}} : (\nabla g)_{\mathbf{p}}\mathbf{v} = 0 \text{ for all } \mathbf{v} \in T_{\mathbf{p}}V_f\}\\ &= \{g \in M_{\mathbf{p}} : (\nabla g)_{\mathbf{p}}\mathbf{v} = 0 \text{ for all } \mathbf{v} \in \mathbb{F}^n \text{ such that } (\nabla f)_{\mathbf{p}}\mathbf{v} = 0\}\\ &= \{g \in M_{\mathbf{p}} : \text{ for all } \mathbf{v} \in \mathbb{F}^n \text{ we have } (\nabla f)_{\mathbf{p}}\mathbf{v} = 0 \Rightarrow (\nabla g)_{\mathbf{p}}\mathbf{v} = 0\}.\end{aligned}$$

First we will show that $M_{\mathbf{p}}^2 + f(\mathbf{x})\mathbb{F}[\mathbf{x}] \subseteq I$. So let $g \in M_{\mathbf{p}}^2$ and $\varphi \in f(\mathbf{x})\mathbb{F}[\mathbf{x}]$. By definition this means that $g(\mathbf{p}) = 0$, $(\nabla g)_{\mathbf{p}} = \mathbf{0}$ and $\varphi = fh$ for some $h \in \mathbb{F}[\mathbf{x}]$. Since $(\nabla g)_{\mathbf{p}} = \mathbf{0}$ we have $(\nabla g)_{\mathbf{p}}\mathbf{v} = 0$ for all $\mathbf{v} \in \mathbb{F}^n$ and hence I. On the other hand, since $\mathbf{p} \in V_f$ we have $f(\mathbf{p}) = 0$. Then using the product rule for ∇ gives

$$(\nabla\varphi)_{\mathbf{p}} = (\nabla h)_{\mathbf{p}} + h(\mathbf{p})(\nabla f)_{\mathbf{p}} = 0 + h(\mathbf{p})(\nabla f)_{\mathbf{p}}.$$

It follows that $(\nabla f)_{\mathbf{p}}\mathbf{v} = 0$ implies $(\nabla\varphi)_{\mathbf{p}}\mathbf{v} = 0$ and hence $\varphi \in I$. Finally, since I is closed under addition we have $g+\varphi \in I$. Next we will show that $I \subseteq M_{\mathbf{p}}^2+f(\mathbf{x})\mathbb{F}[\mathbf{x}]$. Let $g \in I$ so that for all $\mathbf{v} \in \mathbb{F}^n$ we have $(\nabla f)_{\mathbf{p}}\mathbf{v} = 0 \Rightarrow (\nabla g)_{\mathbf{p}}\mathbf{v} = 0$. We want to show that $g \in M_{\mathbf{p}}^2 + f(\mathbf{x})\mathbb{F}[\mathbf{x}]$. On the one hand, if $(\nabla g)_{\mathbf{p}} = \mathbf{0}$ then since $g(\mathbf{p}) = 0$ we have $g \in M_{\mathbf{p}}^2 \subseteq M_{\mathbf{p}}^2 + f(\mathbf{x})\mathbb{F}[\mathbf{x}]$. Otherwise, we observe that the hyperplane $\{\mathbf{v} : (\nabla f)_{\mathbf{p}}\mathbf{v} = 0\}$ is contained in the hyperplane $\{\mathbf{v} : (\nabla g)_{\mathbf{p}}\mathbf{v} = 0\}$. Since both hyperplanes have dimension $n-1$ they must be equal,[18] hence their one-dimensional orthogonal complements are also equal. It follows that $(\nabla g)_{\mathbf{p}} = \lambda(\nabla f)_{\mathbf{p}}$ for some $\lambda \in \mathbb{F} \setminus 0$. Finally, let $h := g - \lambda f \in \mathbb{F}[\mathbf{x}]$ and observe that

$$\begin{aligned} h(\mathbf{p}) &= g(\mathbf{p}) - \lambda f(\mathbf{p}) = 0, \\ (\nabla h)_{\mathbf{p}} &= (\nabla g)_{\mathbf{p}} - \lambda(\nabla f)_{\mathbf{p}} = \mathbf{0}. \end{aligned}$$

In other words, we have $h \in M_{\mathbf{p}}^2$ and hence $g = h + \lambda f \in M_{\mathbf{p}}^2 + f(\mathbf{x})\mathbb{F}[\mathbf{x}]$. We conclude that

$$(T_{\mathbf{p}}V_f)^* \approx \frac{M_{\mathbf{p}}}{I} = \frac{M_{\mathbf{p}}}{M_{\mathbf{p}}^2 + f(\mathbf{x})\mathbb{F}[\mathbf{x}]}.$$

□

Remarks:

- It follows from the proof that $I = M_{\mathbf{p}}^2 + \{\lambda f(\mathbf{x}) : \lambda \in \mathbb{F}\}$. But I prefer to write $I = M_{\mathbf{p}}^2 + f(\mathbf{x})\mathbb{F}[\mathbf{x}]$ to emphasize the fact that I is an ideal of $\mathbb{F}[\mathbf{x}]$.
- This result looks fancy, but it is merely a translation of calculus into the language of ring theory. Why would anyone **want** to do this? The original motivation comes from number theory, where derivatives sometimes give the "wrong answer."
- The field in this theorem is completely arbitrary. However, if $\mathbb{F}$ is algebraically closed then the ideal $f(\mathbf{x})\mathbb{F}[\mathbf{x}]$ takes on more meaning. (See Study's Lemma in the next Chapter.) In this case the Zariski tangent space is key to proving that the dimension of a variety is "intrinsic," i.e., independent of how the variety is embedded.

The Nullstellensatz

The second homework will take you through the details of the Nullstellensatz for curves in the plane. This section will provide background discussion.

[18] If $U_1 \subseteq U_2$ are vector spaces of the same dimension then any basis for U_1 is also be a basis for U_2.

Oct 19: Gauss' Lemma

Greatest common divisors exist in UFDs. To be precise, we have the following result.

GCD in a UFD. Let R be a UFD. Then for any finitely generated ideal $I = a_1R + \cdots + a_nR$ there exists a **unique minimal principal ideal** dR containing I. In this case we say that d is a *greatest common divisor* of $a_1, \ldots, a_n$ and we write

$$\gcd(a_1, \ldots, a_n) \sim d.$$

The greatest common divisor is unique up to multiplication by units.

Proof. Suppose that a_i has unique prime factorization $\prod_k p_k^{\alpha_{ik}}$ and define

$$d := \prod_k p_k^{\delta_k} \quad \text{where} \quad \delta_k = \min_i\{\alpha_{ik}\}.$$

Since d is a common divisor of the a_i we have $a_1R + \cdots + a_nR \subseteq dR$. On other other hand, suppose that $a_1R + \cdots + a_nR \subseteq d'R$ for some $d' \in R$. In particular, since $a_iR \subseteq d'R$ we see that $d'|a_i$ and hence $d' = \prod_k p_k^{\delta'_k}$ for some $\delta'_k \leq \alpha_{ik}$. Since this holds for all i we have $\delta'_k \leq \min_i\{\alpha_{ik}\} = \delta_k$. It follows that $d'|d$ and hence $dR \subseteq d'R$. □

GCD in a PID (Bézout's Identity). If R is also PID and if $\gcd(a_1, \ldots, a_n) \sim d$ then we can find elements $b_1, \ldots, b_n \in R$ such that

$$a_1b_1 + \cdots + a_nb_n = d.$$

Proof. In this case $a_1R + \cdots + a_nR$ is itself principal, so that $a_1R + \cdots + a_nR = dR$. □

These concepts allow us to extend many properties of the ring R to the rings $R[x]$ and $\mathbb{F}[x]$, where $\mathbb{F} = \mathrm{Frac}(R)$. Our goal for today's lecture is to prove the following theorem.

Theorem. *R* **UFD implies** *$R[x]$* **UFD.**

The key step in the proof of this theorem is called *Gauss' Lemma*, which Gauss proved in the case $R = \mathbb{Z}$. Gauss' purpose was to show that the real number $\cos(2\pi/n)$ is expressible in terms of integers and square roots if and only if $\phi(n)$ is a power of 2. This result was an important precursor to Galois theory.

Gauss' Lemma. Let R be a UFD and let $\mathbb{F} = \mathrm{Frac}(R)$. For any polynomial $f(x) \in R[x]$ we let $c(f)$ denote the greatest common divisor of the coefficients, which we call the *content* of f.

(a) For all $f, g \in R[x]$, if $c(f) = c(g) = 1$ then $c(fg) = 1$. More generally, the content is multiplicative: $c(fg) = c(f)c(g)$.

(b) For any $f(x) \in \mathbb{F}[x]$ we have a unique factorization $f(x) = \alpha f'(x)$ where $\alpha \in \mathbb{F} \setminus 0$ and $f'(x) \in R[x]$ with $c(f') = 1$. This f' is called the *primitive part* of f.

(c) If $f(x) = \prod g_i(x)$ for some $f, g_i \in \mathbb{F}[x]$ then we have $f'(x) = \prod_i g_i'(x)$ in $R[x]$.

(d) If $f(x), g(x)$ are coprime in $R[x]$, then they are still coprime in $\mathbb{F}[x]$.

Proof. (a): Recall that prime and irreducible elements coincide in a UFD. For any prime $p \in R$ we have a ring homomorphism $R[x] \to (R/pR)[x]$ denoted by $f(x) \mapsto f_p(x)$, and we observe that $c(f) = 1$ if and only if $f_p(x) \neq 0$ for all primes p. Suppose that $c(f) = 1$ and $c(g) = 1$ so that $f_p(x) \neq 0$ and $g_p(x) \neq 0$ for all p. Since R/pR is a domain we see that $(R/pR)[x]$ is a domain, and it follows that $(fg)_p(x) = f_p(x)g_p(x) \neq 0$ for all p.

To prove that content is multiplicative, consider any $f, g \in R[x]$ and factor out the gcd of the coefficients to obtain $f(x) = c(f)f'(x)$ and $g(x) = c(g)g'(x)$, where $c(f') = c(g') = 1$. Then

$$f(x)g(x) = c(f)c(g)f'(x)g'(x).$$

Since $c(f'g') = 1$ we compare the content on each side to obtain $c(fg) = c(f)c(g)$.

(b): Let $f(x) \in \mathbb{F}[x]$. To prove existence, let $d \in R$ be any common multiple of the denominators of the coefficients of f, so that $df(x) \in R[x]$. Then we have $df(x) = c(df)f'(x)$ for some primitive polynomial $f'(x) \in R[x]$. Finally, we let $\alpha = c(df)/d$ so that $f(x) = \alpha f'(x)$. To prove uniqueness, suppose that we have $f(x) = \alpha f'(x) = \beta f''(x)$ with $\alpha, \beta \in \mathbb{F} \setminus 0$ and $f'(x), f''(x) \in R[x]$ primitive. Let $d \in R$ be such that $d\alpha, d\beta \in R$. Then from uniqueness[19] of the gcd we have $d\alpha = c(df) = d\beta$ and canceling d gives $\alpha = \beta$.

(c): Suppose that $f(x) = \prod_i g_i(x)$ for some $f, g_i \in \mathbb{F}[x]$. Using part (b), let $f(x) = \alpha f'(x)$ and $g_i(x) = \beta_i g_i'(x)$ where $\alpha, \beta_i \in \mathbb{F} \setminus 0$ and $f'(x), g_i'(x) \in R[x]$ are primitive. Then we have

$$\alpha f'(x) = \left(\prod \beta_i\right) \prod g_i'(x).$$

Choose any $d \in R$ such that $d\alpha \in R$ and $d(\prod \beta_i) \in R$ and multiply to obtain

$$d\alpha f'(x) = d\left(\prod \beta_i\right) \prod g_i'(x).$$

From (a) we know that $\prod g_i'(x)$ is primitive. Then comparing the content on each side gives $d\alpha = d(\prod \beta_i)$ and cancelling this quantity gives $f'(x) = \prod g_i'(x)$.

(d): Suppose that $f, g \in R[x]$ are coprime, meaning that they have no non-constant common divisor in $R[x]$. We will show that they still have no non-constant common divisor in $\mathbb{F}[x]$. To see this, we assume for contradiction that $f = qa$ and $g = qb$ for some $q, a, b \in \mathbb{F}[x]$ with $q \in \mathbb{F}[x]$ non-constant. From part (c) we have $f' = q'a'$ and $g' = q'b'$ in $R[x]$. And since $q = \alpha q'$ for some $\alpha \in \mathbb{F} \setminus 0$ we observe that $q' \in R[x]$ is non-constant. Finally, since $f = c(f)f'$ and $g = c(g)g'$ with $c(f), c(g) \in R$, we conclude that $q' | f$ and $q' | g$ in $R[x]$. □

[19] As always, the uniqueness is up to multiplication by units.

Now we can prove the theorem.

Proof that R UFD implies $R[x]$ UFD. We will show that any $f(x) \in R[x]$ has a unique factorization into irreducible elements of $R[x]$.

Existence. Consider $f(x)$ as an element of $\mathbb{F}[x]$. Since $\mathbb{F}[x]$ is a PID, and hence Noetherian, we can write

$$f(x) = q_1(x)q_2(x)\cdots q_k(x),$$

where $q_i(x) \in \mathbb{F}[x]$ are irreducible. It follows from the previous lemma that

$$f'(x) = q_1'(x)q_2'(x)\cdots q_k'(x),$$

and hence $f(x) = c(f)f'(x) = c(f)q_1'(x)q_2'(x)\cdots q_k'(x)$, where $q_i'(x) \in R[x]$ are irreducible and primitive. (Indeed, if $q_i'(x)$ were reducible in $R[x]$ then $q_i(x)$ would be reducible in $\mathbb{F}[x]$.) Then since R is a UFD we can factor $c(f) \in R$ into primes:

$$f(x) = up_1\cdots p_\ell q_1'(x)\cdots q_k'(x).$$

These primes $p_i \in R$ are clearly also irreducible in $R[x]$.

Uniqueness. First we will show that every primitive irreducible $q(x) \in R[x]$ is prime in $R[x]$. So suppose that $q(x) \in R[x]$ is primitive and irreducible, with $q(x)|f(x)g(x)$ for some $f, g \in R[x]$. Since $\mathbb{F}[x]$ is a PID we know that $q(x)$ is a prime element of $\mathbb{F}[x]$ and hence we have $q|f$ or $q|g$ in $\mathbb{F}[x]$. Without loss of generality, suppose that $f(x) = q(x)h(x)$ for some $h(x) \in \mathbb{F}[x]$. Then from Gauss' Lemma we have $f'(x) = q'(x)h'(x)$ in $R[x]$. Since $q(x) = q'(x)$ is primitive we conclude that $q|f'$ and hence $q|f$ in $R[x]$.

Finally, suppose that $p_1\cdots p_k q_1(x)\cdots q_\ell(x) \sim p_1'\cdots p_m' q_1'(x)\cdots q_n'(x)$ where $p_i, p_i' \in R$ are prime and $q_i(x), q_i'(x) \in R[x]$ are irreducible and primitive.[20] By comparing content we obtain $p_1\cdots p_k \sim p_1'\cdots p_m'$ and since R is a UFD we must have $k = m$ and by reordering the factors we may assume that $p_i \sim p_i'$ for all i. Now cancel this quantity to obtain $q_1(x)\cdots q_\ell(x) \sim q_1'(x)\cdots q_n'(x)$. Since $q_1(x)$ is prime in $R[x]$ we have $q_1(x)|q_i'(x)$ for some i and without loss of generality we may assume that $q_1(x)|q_1'(x)$. Since both are irreducible this implies $q_1(x) \sim q_1'(x)$. After canceling this factor from both sides we conclude by induction that $\ell = n$ and we may reindex the factors to obtain $q_i(x) \sim q_i'(x)$ for all i. □

Remark: By induction we conclude that the polynomial rings $\mathbb{Z}[x_1, \ldots, x_n]$ and $\mathbb{F}[x_1, \ldots, x_n]$ are unique factorization domains. In the next section we will use this result to study curves and hypersurfaces.

Oct 21: Study's Lemma for Curves

In the next two sections we will prove *Study's Lemma*, which says that any hypersurface V (either affine or projective) over an algebraically closed field is determined by a unique *minimal polynomial* f with the following properties:

[20] For a general ring R, the units of $R[x]$ are just the units of R. Thus the notation $f(x) \sim g(x)$ is unambiguous.

- f is square-free (i.e., has no repeated irreducible factor),
- $V = V_f$,
- $V_f \subseteq V_g$ implies $f|g$.

We say that a hypersurface V is *irreducible* if it cannot be expressed as a union of proper sub-hypersurfaces, i.e., $V = V_1 \cup V_2$ with $V_1, V_2 \subsetneq V$. We will see that irreducible hypersurfaces correspond to irreducible polynomials, hence from the unique factorization of polynomials we will obtain a unique decomposition of hypersurfaces into irreducible parts.

To see that algebraic closure is necessary, we observe that Study's Lemma is false over the real numbers. Indeed, let $f(x,y) = x^2 + y^2$ and $g(x,y) = x$ and consider the real curves $C_f, C_g \subseteq \mathbb{R}^2$. Note that C_f is just the single point $(0,0) \in \mathbb{R}^2$ and C_g is the y-axis, so that $C_f \subseteq C_g$. Furthermore, note that f is square-free (in fact, irreducible) in the ring $\mathbb{R}[x,y]$. However, x is clearly not divisible by $x^2 + y^2$. What went wrong? The problem is that C_f should really be thought of as $x^2+y^2 = (x+iy)(x-iy) = 0$, which defines a pair of intersecting "lines" in the "plane" $\mathbb{C}^2$. Thus we really have $C_f \not\subseteq C_g$, which is why $f \nmid g$.

The name "Study's Lemma" is standard in modern textbooks. The result is stated (for complex numbers) on page 202 of Eduard Study's *Methods for the theory of ternary forms* (1889). (A ternary form is just a homogeneous polynomial in three variables, i.e., a projective curve.) But Study actually attributes the result to Otto Hölder, *On the concept of invariants* (1884). So it should probably be "Hölder's Lemma."[21]

For pedagogical reasons I will first prove the theorem just for affine curves, even though this is not logically necessary. (This is the very first theorem in Shafarevich's Algebraic Geometry.) Then will follow the proof for affine hypersurfaces, and finally for projective hypersurfaces.

Study's Lemma for Affine Curves. Let $\mathbb{F}$ be an algebraically closed field and consider two polynomials $f, g \in \mathbb{F}[x,y]$ with corresponding affine curves $C_f, C_g \subseteq \mathbb{F}^2$.

(a) If f is irreducible and $f \nmid g$ then $C_f \cap C_g$ consists of finitely many points. [This result does not use the fact that $\mathbb{F}$ is algebraically closed.]

(b) If f is non-constant then the curve $C_f \subseteq \mathbb{F}^2$ has infinitely many points.

(c) If f is irreducible and $C_f \subseteq C_g$ then $f|g$.

(d) More generally, if f is square-free (no repeated irreducible factor) and $C_f \subseteq C_g$ then $f|g$. It follows that there is a bijection between affine curves and square-free polynomials.

(e) Finally, we say that a curve $C \subseteq \mathbb{F}^2$ is *irreducible* if it cannot be expressed as a union of curves $C = C_1 \cup C_2$ with $C_1, C_2 \subsetneq C$. It follows that there is a bijection between irreducible curves in $\mathbb{F}^2$ and irreducible polynomials in $\mathbb{F}[x,y]$.

[21] See Carlo Beenakker's answer to my question on mathoverflow: `https://mathoverflow.net/questions/374566/history-of-studys-lemma`

Proof. (a): Since f is irreducible and $f \nmid g$ we see that f and g are coprime. That is, the smallest principal ideal containing $f(x,y)\mathbb{F}[x,y] + g(x,y)\mathbb{F}[x,y]$ is the whole ring $\mathbb{F}[x,y]$. Now let $\mathbb{F}(x)$ be the field of fractions of $\mathbb{F}[x]$ and consider $f(x,y), g(x,y) \subseteq \mathbb{F}[x][y]$ as elements of the larger ring $\mathbb{F}(x)[y]$. I claim that f and g are still coprime in this larger ring. Indeed, suppose that $q|f$ and $q|g$ for some irreducible element $q \in \mathbb{F}(x)[y]$. Then from part (c) of Gauss' Lemma we have $q'|f'$ and $q'|g'$ in $\mathbb{F}[x,y]$, where f', g', q' are the primitive parts of f, g, q, and $q' \in \mathbb{F}[x,y]$ is irreducible.[22] Finally, since $f'|f$ and $g'|g$ in $\mathbb{F}[x,y]$ we conclude that q' is a common irreducible of f, g in $\mathbb{F}[x,y]$. Contradiction.

Now since $\mathbb{F}(x)$ is a field we know that $\mathbb{F}(x)[y]$ is a PID. And since f, g are coprime in $\mathbb{F}(x)[y]$ it follows from Bézout's identity (previous lecture) that there exist $F, G \in \mathbb{F}(x)[y]$ such that

$$f(x,y)F + g(x,y)G = 1.$$

The y-coefficients of F and G are rational functions of x. If $h(x) \in \mathbb{F}[x]$ is any common multiple of the denominators of these coefficients then multiplying both sides by $h(x)$ gives

$$f(x,y)\tilde{f}(x,y) + g(x,y)\tilde{g}(x,y) = h(x)$$

for some **polynomials** $\tilde{f}(x,y), \tilde{g}(x,y) \in \mathbb{F}[x,y]$. If $(a,b) \in \mathbb{F}^2$ is any point in the intersection $C_f \cap C_g$ then evaluating the previous equation at $(x,y) = (a,b)$ gives $h(a) = 0$, which has finitely many possible solutions $x = a$. We would like to finish the proof by claiming that each polynomial $f(a,y) \in \mathbb{F}[y]$ has finitely many roots $y = b$. However, it might be the case that one of the polynomials $f(a,y)$ is the zero polynomial, which has infinitely many roots $y = b$. To get around this problem we can run the whole proof again using the ring $\mathbb{F}(y)[x]$ to see that there are finitely many $y = b$ with $f(a,b) = 0$. (Unfortunately this trick does not work in higher dimensions. In that case we will have to precede the proof with a tiny change of variables called a "normalization." See below.)

(b): First we will show that an algebraically closed field $\mathbb{F}$ is infinite. Indeed, if $\mathbb{F}$ were finite then the polynomial $1 + \prod_a (x - a) \in \mathbb{F}[x]$ would have no roots in $\mathbb{F}$, which contradicts the fact that $\mathbb{F}$ is algebraically closed. Now let $f \in \mathbb{F}[\mathbf{x}, y]$ be non-constant. This means that we can write $f(x,y) = \sum c_k(x)y^k \in \mathbb{F}[x,y]$, where $c_k(x) \in \mathbb{F}[x]$ is non-zero for some $k \geq 1$. Since $\mathbb{F}$ is infinite there exist infinitely many $a \in \mathbb{F}$ such that $c_k(a) \neq 0$, i.e., such that $f(a,y) \in \mathbb{F}[y]$ is non-constant. Then since $\mathbb{F}$ is algebraically closed, each such polynomial $f(a,y) \in \mathbb{F}[y]$ has a root $y = b$, giving infinitely many points $(a,b) \in C_f$.

(c) Now suppose that $f, g \in \mathbb{F}[x,y]$ with f irreducible (in particular, non-constant) and $C_f \subseteq C_g$. If $f \nmid g$ then from part (a) the set $C_f = C_f \cap C_g$ would be finite, contradicting part (b). Hence we must have $f|g$.

(d): Let $f(x,y) \in \mathbb{F}[x,y]$ be square-free, with prime factorization $f = q_1 \cdots q_k$ and observe that $C_{q_i} \subset C_f$ for all i. If $C_f \subseteq C_g$ then it follows from (c) that $q_i|g$ for all i. Since $\mathbb{F}[x,y]$ is a UFD this implies that $f|g$.

[22] Gauss' Lemma is not needed here. By definition we have $q = \alpha q'$ for some $\alpha \in \mathbb{F}(x)$. If $q' = \varphi\psi$ for some non-constant $\varphi, \psi \in \mathbb{F}[x,y]$ then we have $q = \alpha\varphi\psi$ in $\mathbb{F}(x)[y]$, contradicting the fact that q is irreducible.

Now let $C = C_f \subseteq \mathbb{F}^2$ be any affine curve and let $f = q_1^{e_1} \cdots q_k^{e_k}$ be the prime factorization, with unique square-free divisor $\sqrt{f} = q_1 \cdots q_k$ (called the *radical* of f). We observe that $C_{\sqrt{f}} = C_f$, so that every curve is defined by some square-free polynomial. To see that this polynomial is unique, suppose that $C_f = C_g$ with f, g square-free. Then from the above remark we have $f|g$ and $g|f$, which implies that $f \sim g$.

(e): Let $C = C_f$ with f square-free. If f is reducible then we can write $f = g_1 g_2$ where g_1, g_2 are square-free and $g_1, g_2 \not\sim f$. Since $f = g_1 g_2$ we have $C_f = C_{g_1} \cup C_{g_2}$ and since $g_1, g_2 \not\sim f$, part (d) implies that $C_{g_1}, C_{g_2} \subsetneq C_f$, hence C_f is reducible. Conversely, suppose that C_f is reducible, so that $C_f = C_1 \cup C_2$ with $C_1, C_2 \subsetneq C_f$. From part (d) we can write $C_1 = C_{g_1}$ and $C = C_{g_2}$ for some square-free polynomials g_1, g_2. Then since $C_{g_1}, C_{g_2} \subsetneq C_f$ we have $g_1, g_2 | f$ and $g_1, g_2 \not\sim f$. In other words, f is reducible. □

Oct 23: Study's Lemma for Hypersurfaces

Now let's beef up the proof to include hypersurfaces, both affine and projective.

Study's Lemma for Affine Hypersurfaces. Lef $\mathbb{F}$ be algebraically closed (hence infinite) and consider the ring of polynomials in n variables: $\mathbb{F}[\mathbf{x}] = \mathbb{F}[x_1, \ldots, x_n]$.

(a) For any non-zero polynomial $f \in \mathbb{F}[\mathbf{x}]$ there exists a point $\mathbf{p} \in \mathbb{F}^n$ such that $f(\mathbf{p}) \neq 0$. [Here we only use the fact that $\mathbb{F}$ infinite.]

(b) Given $f, g \in \mathbb{F}[\mathbf{x}]$ with f irreducible and $V_f \subseteq V_g$, we must have $f|g$.

(c) We have bijections:

$$\text{hypersurfaces} \leftrightarrow \text{square-free polynomials},$$
$$\text{irreducible hypersurfaces} \leftrightarrow \text{irreducible polynomials}.$$

Proof. (a): We will use induction on n. *Base Case*: A polynomial $f(x)$ in one variable has finitely many roots by Descartes' Theorem. If $\mathbb{F}$ is infinite then there exists some $a \in \mathbb{F}$ with $f(a) \neq 0$. *Induction Step*: So let us suppose that $n \geq 2$ and let $f \in \mathbb{F}[\mathbf{x}]$ be nonzero. If $\mathbf{x}' = (x_1, \ldots, x_{n-1})$ then we can write

$$f(\mathbf{x}) = \sum f_k(\mathbf{x}') x_n^k,$$

where the coefficients $f_k(\mathbf{x}') \in \mathbb{F}[\mathbf{x}']$ are not all zero. Let $g \in \mathbb{F}[\mathbf{x}']$ denote the product of the non-zero coefficients, which is nonzero. Then by induction there exists a point $\mathbf{p}' \in \mathbb{F}^{n-1}$ such that $g(\mathbf{p}') \neq 0$. This implies that at least one of the coefficients $f_k(\mathbf{p}') \in \mathbb{F}$ of $f(\mathbf{p}', x_n) \in \mathbb{F}[x_n]$ is non-zero. Hence from the base case there exists some $p_n \in \mathbb{F}$ such that $f(\mathbf{p}) = f(\mathbf{p}', p_n) \neq 0$.

(b): *Normalization Step*: If f is non-constant of degree $d \geq 1$ then I claim that we can make an invertible linear substitution Φ such that

$$f(\Phi \mathbf{x}) = c x_n^d + \text{lower terms in } x_n,$$

where $c \in \mathbb{F} \setminus 0$. To see this, let $f = f^{(d)} + \cdots + f^{(1)} + f^{(0)}$ be the homogeneous filtration. Since $f^{(d)}$ is nonzero we know from (a) that there exists $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{F}^n$ such that $f^{(d)}(\mathbf{a}) \neq 0$. Furthermore, since $f^{(d)}$ is homogeneous, we know that $a_i \neq 0$ for some i. Now let A be the linear substitution sending $x_i \mapsto a_i x_i$ and $x_j \mapsto x_j + a_j x_i$ for all $j \neq i$, which is invertible because $a_i \neq 0$. Since linear transformations preserve homogeneous degree, we have

$$\begin{aligned} f(A\mathbf{x}) &= f^{(d)}(A\mathbf{x}) + \text{lower terms} \\ &= f^{(d)}(\mathbf{a})\mathbf{x}_i^d + \text{lower terms in } x_i. \end{aligned}$$

Finally, we can take $\Phi = PA$, where P switches the variables x_i and x_n.

Elimination Step: If $f, g \in \mathbb{F}[\mathbf{x}]$ with f irreducible and $V_f \subseteq V_g$ then we will show that $f|g$. Note that divisibility of polynomials and containment of hypersurfaces are preserved under invertible linear substitutions. Thus since f is non-constant we may assume from the previous step that $f = cx_n^d + \text{lower terms in } x_n$. Now suppose for contradiction that $f \nmid g$, so that f, g are coprime in the ring $\mathbb{F}[\mathbf{x}]$. Let $\mathbf{x}' = (x_1, \ldots, x_{n-1})$ as in part (a) and let $\mathbb{F}(\mathbf{x}')$ be the field of fractions of $\mathbb{F}[\mathbf{x}']$. Then Gauss' Lemma implies that f, g are still coprime in the larger ring $\mathbb{F}(\mathbf{x}')[x_n] \supseteq \mathbb{F}[\mathbf{x}'][x_n] = \mathbb{F}[\mathbf{x}]$. Since $\mathbb{F}(\mathbf{x}')$ is a field, this larger ring is a PID, hence from Bézout's identity there exist some $F, G \in \mathbb{F}(\mathbf{x}')[x_n]$ with $fF + gG = 1$. Let $h(\mathbf{x}') \in \mathbb{F}[\mathbf{x}']$ be any common denominator of the coefficients of F, G and multiply both sides to obtain

$$f(\mathbf{x})\tilde{f}(\mathbf{x}) + g(\mathbf{x})\tilde{g}(\mathbf{x}) = h(\mathbf{x}') \in \mathbb{F}[\mathbf{x}']$$

for some **polynomials** $\tilde{f}, \tilde{g} \in \mathbb{F}[\mathbf{x}]$. Since h is nonzero, we know from part (a) that there exists $\mathbf{p}' \in \mathbb{F}^{n-1}$ such that $h(\mathbf{p}') \neq 0$. Since $f(\mathbf{x})$ contains the term cx_n^d, we observe that the polynomial $f(\mathbf{p}', x_n) \in \mathbb{F}[x_n]$ is nonconstant. And **since $\mathbb{F}$ is algebraically closed** this polynomial has a root $x_n = p_n$, so that $f(\mathbf{p}) = f(\mathbf{p}', p_n) = 0$. Finally, since $V_f \subseteq V_g$ we also have $g(\mathbf{p}) = 0$, which gives the desired contradiction:

$$0 = f(\mathbf{p})\tilde{f}(\mathbf{p}) + g(\mathbf{p})\tilde{g}(\mathbf{p}) = h(\mathbf{p}') \neq 0.$$

(c): Using part (b), the proof follows verbatim from Study's Lemma for curves. □

Remark:

- One should compare Study's Lemma to Descartes' Factor Theorem, which says that a polynomial $f(x) \in \mathbb{F}[x]$ vanishes at $x = a$ if and only if it is divisible by $x - a$. Furthermore, a polynomial vanishes on a finite set $\{a_1, \ldots, a_k\} \subseteq \mathbb{F}$ if and only if it is divisible by $(x - a_1) \cdots (x - a_k)$, which we may consider the *minimal polynomial* of this set. If $\mathbb{F}$ is algebraically closed then each irreducible polynomial has the form $x - a$, which corresponds to a single point $a \in \mathbb{F}$. (A finite set is a hypersurface in $\mathbb{F}^1$ and a single point is an irreducible hypersurface in $\mathbb{F}^1$.)

To end this section we extend Study's Lemma to the projective case.

Study's Lemma for Projective Hypersurfaces. Let $\mathbb{F}$ be algebraically closed.

(a) Factors of homogeneous polynomials are homogeneous. [This holds over any field.]

(b) Let $F, G \in \mathbb{F}[\mathbf{x}] = \mathbb{F}[x_1, \ldots, x_{n+1}]$ be homogeneous with corresponding projective hypersurfaces $V_F, V_G \subseteq \mathbb{FP}^n$. If F is irreducible and $V_F \subseteq V_G$ then $F|G$.

(c) We have bijections:

$$\text{projective hypersurfaces} \leftrightarrow \text{square-free homogeneous polynomials},$$
$$\text{irreducible projective hypersurfaces} \leftrightarrow \text{irreducible homogeneous polynomials}.$$

Proof. (a) Let $F(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ be homogeneous of degree $d \geq 1$ and let $F = q_1 \cdots q_m$ be the unique irreducible factorization in $\mathbb{F}[\mathbf{x}]$. Let $q_i = \sum q_i^{(k)}$ be the filtrations into homogeneous pieces and suppose that each q_i has degree d_i so that $q_i^{(d_i)}$ is its leading form. Since the leading form of a product is the product of the leading forms we conclude that

$$q_1 \cdots q_m = F = F^{(d)} = q_1^{(d_1)} \cdots q_m^{(d_m)}.$$

Then since each $q_i^{(d_i)}$ is non-constant, the uniqueness of prime factorization in $\mathbb{F}[\mathbf{x}]$ tells us that each $q_i^{(d_i)}$ must be irreducible, with $q_i^{(d_i)} \sim q_j$ for some j. Since $q_i^{(d_i)}$ is homogeneous this implies that q_j is homogeneous, and in fact $q_j = q_j^{(d_j)}$. Finally, since the unique prime factors are homogeneous, and since the product of homogeneous polynomials is homogeneous, it follows that any factor of F is homogeneous.

(b): Let $F, G \in \mathbb{F}[\mathbf{x}]$ be homogeneous with corresponding projective hypersurfaces $V_F, V_G \subseteq \mathbb{FP}^n$. If F is irreducible and $V_F \subseteq V_G$ then we will show that $F|G$. First suppose that $F = x_1$, so that V_F is the coordinate hyperplane $H_1 : x_1 = 0$ and V_G contains this hyperplane. Suppose that G has degree d and write $G(\mathbf{x}) = G_d(\mathbf{x}') + x_1 G_{d-1}(\mathbf{x}') + \cdots + x_1^d G_0(\mathbf{x}')$, where $\mathbf{x}' = (x_2, \ldots, x_{n+1})$ and $G_k \in \mathbb{F}[\mathbf{x}']$ is homogeneous of degree k. Note that $G(0, \mathbf{x}') = G_d(\mathbf{x}')$. Then since V_G contains the hyperplane H_1 we have $G_d(\mathbf{a}') = G(0, \mathbf{a}') = 0$ for all $\mathbf{a}' \in \mathbb{F}^n$. Since $\mathbb{F}$ is infinite implies that $G_d(\mathbf{x}') \in \mathbb{F}[\mathbf{x}']$ is the zero polynomial, hence $x_1|G$.

Next suppose that $F \neq x_1$ (which since F is irreducible implies that $x_1 \nmid F$) and assume that $V_F \subseteq V_G$. Consider the de-homogenizations $f(\mathbf{x}') = F(1, \mathbf{x}')$ and $g(\mathbf{x}') = G(1, \mathbf{x}')$, with corresponding affine hypersurfaces $V_f, V_g \subseteq \mathbb{F}^n = \mathbb{FP}^n \setminus H_1$. Since $V_F \subseteq V_G$ we have

$$f(\mathbf{a}') = 0 \quad \Rightarrow \quad F(1, \mathbf{a}') = 0 \quad \Rightarrow \quad G(1, \mathbf{a}') = 0 \quad \Rightarrow \quad g(\mathbf{a}') = 0,$$

and hence $V_f \subseteq V_g$. Furthermore, since F is irreducible I claim that $f(x, y)$ is irreducible. Indeed, if $f = f_1 f_2$ with $\deg(f_1) = d_1 \geq 1$ and $\deg(f_2) = d_2 \geq 1$, then since $x_1 \nmid F$ we have

$$F(\mathbf{x}) = x_1^d f(\mathbf{x}'/x_1) = [x_1^{d_1} f_1(\mathbf{x}'/x_1)][x_1^{d_2} f_2(\mathbf{x}'/x_i)] = F_1(\mathbf{x}) F_2(\mathbf{x}),$$

for some (homogeneous) polynomials F_1, F_2 of degrees $d_1, d_2 \geq 1$. Since f is irreducible and $V_f \subseteq V_g$ it follows from the affine Study's Lemma that $f|g$, say $g(\mathbf{x}') = f(\mathbf{x}')h(\mathbf{x}')$. Finally, we re-homogenize to obtain $x_1^k G(\mathbf{x}) = x_1^\ell F(\mathbf{x}) H(\mathbf{x})$ for some (homogeneous) $H(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ and $k, \ell \geq 0$. Since F is irreducible and $x_1 \nmid F$, we conclude by unique factorization that $F|G$.

(c): Using parts (a) and (b), the proof follows verbatim from Study's Lemma for curves. $\square$

Remark: Study's Lemma for **hyperplanes** holds over any infinite field $\mathbb{F}$, not necessarily algebraically closed. To see this, let $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} \in \mathbb{F}[\mathbf{x}] = \mathbb{F}[x_1, \ldots, x_{n+1}]$ be a homogeneous linear polynomial and let $H_{\mathbf{a}} : f(\mathbf{x}) = 0$ be the corresponding projective hyperplane in $\mathbb{FP}^n$. Recall that any such hyperplane is projectively equivalent to the coordinate hyperplane $H_1 : x_1 = 0$. Indeed, if $A \in \mathrm{GL}_{n+1}(\mathbb{F})$ is any invertible matrix satisfying $\mathbf{e}_1^T A = \mathbf{a}^T$ (i.e., if A has first row $\mathbf{a}^T$) then we have

$$(Af)(\mathbf{x}) = f(A^{-1}\mathbf{x}) = \mathbf{a}^T A^{-1}\mathbf{x} = \mathbf{e}_1^T\mathbf{x} = x_1,$$

so that $AH_{\mathbf{a}} = H_1$. Now let $g \in \mathbb{F}[\mathbf{x}]$ be any homogeneous polynomial that vanishes on $H_{\mathbf{a}}$, so that $h(\mathbf{x}) := (Ag)(\mathbf{x}) = g(A^{-1}\mathbf{x})$ vanishes on H_1. We will prove that $f(\mathbf{x})$ divides $g(\mathbf{x})$.

Let $\mathbf{x}' = (x_2, \ldots, x_{n+1})$ and suppose that g has degree d. Then we can write

$$h(\mathbf{x}) = \sum_{k=0}^{d} h_k(\mathbf{x}')x_1^k,$$

where $h_k(\mathbf{x}') \in \mathbb{F}[x_2, \ldots, x_{n+1}]$ is homogeneous of degree $d-k$. Let $\mathbf{p} = (0, \mathbf{p}') \in \mathbb{F}^{n+1}$ be an arbitrary point of the hyperplane H_1. Then since h vanishes on H_1 we have

$$0 = h(\mathbf{p}) = h_0(\mathbf{p}'),$$

and since $\mathbb{F}$ is infinite this implies that $h_0(\mathbf{x}')$ is the zero polynomial. In other words, $h(\mathbf{x})$ is divisible by $x_1 = f(A^{-1}\mathbf{x})$. Finally, since linear substitution is a ring homomorphism, we conclude that $h(A\mathbf{x}) = g(\mathbf{x})$ is divisible by $f(A^{-1}A\mathbf{x}) = f(\mathbf{x})$, as desired.

Oct 28: Sylvester's Resultant

Recall the following key step from the proof of Study's Lemma. Let $f, g \in \mathbb{F}[\mathbf{x}]$ be coprime and let $\mathbf{x}_i'$ be the subset of the variables $\mathbf{x} = (x_1, \ldots, x_n)$ with x_i deleted. Then we can find polynomials $\tilde{f}, \tilde{g} \in \mathbb{F}[\mathbf{x}]$ and $h \in \mathbb{F}[\mathbf{x}_i']$ such that

$$f\tilde{f} + g\tilde{g} = h.$$

This is some sort of weak version of Bézout's identity that holds in a UFD (i.e., not only in a PID). In this section we will pursue this idea to its natural conclusion.

Sylvester's Resultant Theorem. Let R be a UFD (hence $R[x]$ is also a UFD by Gauss' Lemma) and consider two polynomials $f, g \in R[x]$ of degrees d and e (i.e., $a_0, b_0 \neq 0$):

$$f(x) = a_0x^d + \cdots + a_{d-1}x + a_d,$$
$$g(x) = b_0x^e + \cdots + b_{e-1}x + b_e.$$

Then the following are equivalent:

(1) The polynomials f and g are **not** coprime in $R[x]$.

(2) There exist polynomials $\varphi, \gamma \in R[x]$ satisfying

- $\deg \varphi < \deg g$ and $\deg \gamma < \deg g$,
- $f\gamma = g\varphi$.

(3) Sylvester's *resultant* $\operatorname{Res}(f,g) \in R$ is nonzero:

$$\operatorname{Res}(f,g) := \det \begin{pmatrix} a_0 & a_1 & \cdots & a_d & & \\ & \ddots & \ddots & & \ddots & \\ & & a_0 & a_1 & \cdots & a_d \\ b_0 & b_1 & \cdots & b_e & & \\ & \ddots & \ddots & & \ddots & \\ & & b_0 & b_1 & \cdots & b_e \end{pmatrix} \neq 0$$

[Note that this matrix has shape $(d+e) \times (d+e)$.]

Proof. (1)⇒(2): Suppose that $h|f$ and $h|g$ for some non-constant h. Then we can write $f = h\varphi$ and $g = h\gamma$ with $\deg \varphi < \deg f$ and $\deg \gamma < \deg g$, so that $f\gamma = h\varphi\gamma = g\varphi$.

(2)⇒(1): Let $f\gamma = g\varphi$ with $\deg \varphi < \deg f$ and $\deg \gamma < \deg g$, and suppose for contradiction that f and g are coprime. Then since $f|g\varphi$ it follows from unique factorization that $f|\varphi$, which contradicts the fact that $\deg \varphi < \deg f$.

(2)⇔(3): The desired polynomials $\varphi, \gamma \in R[x]$ have the form

$$\begin{aligned} \varphi(x) &= u_0 x^{d-1} + \cdots + u_{d-2}x + u_{d-1}, \\ \gamma(x) &= v_0 x^{e-1} + \cdots + v_{e-2}x + v_{e-1}, \end{aligned}$$

and satisfy the equation

$$\begin{aligned} &(a_0x^d + \cdots + a_{d-1}x + a_d)(v_0x^{e-1} + \cdots + v_{e-2}x + v_{e-1}) \\ &\qquad = (b_0x^e + \cdots + b_{e-1}x + b_e)(u_0x^{d-1} + \cdots + u_{d-2}x + u_{d-1}). \end{aligned}$$

By expanding and equating x-coefficients, this is equivalent to the following system of $m+n$ linear equations in the $m+n$ unknown coefficients $u_0, \ldots, u_{d-1}, v_0, \ldots, v_{e-1}$:

$$\begin{array}{rcccccccc} a_0v_0 & & & & & = & b_0u_0 & & \\ a_1v_0 & + & a_0v_1 & & & = & b_1u_0 & + & b_0u_1 \\ & & \vdots & & & = & & & \vdots \\ & & & a_dv_{e-1} & & = & & & b_eu_{d-1}. \end{array}$$

Then by moving all variables to the left we see that this linear system has a solution if and

only if the determinant of the matrix of coefficients is nonzero:

$$\det \begin{pmatrix} a_0 & & & -b_0 & & \\ a_0 & \ddots & & -b_1 & \ddots & \\ \vdots & \ddots & a_0 & \vdots & \ddots & -b_0 \\ a_d & & a_1 & -b_e & & -b_1 \\ & \ddots & \vdots & & \ddots & \vdots \\ & & a_d & & & -b_e \end{pmatrix} \neq 0.$$

Scaling each of the final d columns by -1 and then transposing the matrix just scales the determinant by $(-1)^d$. □

Examples:

- A polynomial $f(x) \in \mathbb{C}[x]$ has a multiple root $\alpha \in \mathbb{C}$ if and only if $f(x), f'(x) \in \mathbb{C}[x]$ have the common irreducible factor $(x-\alpha) \in \mathbb{C}[x]$. If $f(x) = ax^2 + bx + c$ (with $a \neq 0$), and hence $f'(x) = 2ax + b$, then f has a multiple root in $\mathbb{C}$ if and only if

$$0 \neq \begin{pmatrix} a & b & c \\ 2a & b & \\ & 2a & b \end{pmatrix} = a(b^2 - 0) - 2a(b^2 - 2ac) = -a(b^2 - 4ac).$$

  In general we call $\mathrm{Res}(f, f') \in R$ is the *discriminant* of the polynomial $f(x) \in R[x]$.

- From Descartes we know that a polynomial $f(x) = a_0x^d + \cdots + a_{d-1}x + a_d \in R[x]$ has a root $\alpha \in R$ if and only if $f(x)$ and $g(x) = x - \alpha$ are not coprime in the ring $R[x]$. And, indeed, one can show by induction that

$$\det \begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_d \\ 1 & -\alpha & & & \\ & 1 & -\alpha & & \\ & & \ddots & \ddots & \\ & & & 1 & -\alpha \end{pmatrix} = (-1)^d(a_0\alpha^d + \cdots + a_{d-1}\alpha + a_d) = (-1)^d f(\alpha).$$

  In other words, if $\deg(f) = d$ then for all $\alpha \in R$ we have $\mathrm{Res}(f, x-\alpha) = (-1)^d f(\alpha)$. If $f(x)$ can be split as $\prod_{i=1}^{d}(x - \lambda_i)$ then it follows that

$$\mathrm{Res}(f, x-\alpha) = (-1)^d \prod_{i=1}^{d}(\alpha - \lambda_i) = \prod_{i=1}^{d}(\lambda_i - \alpha).$$

- More generally, If $f, g \in R[x]$ can be split as $f(x) = \prod_{i=1}^{d}(x-\lambda_i)$ and $g(x) = \prod_{i=1}^{e}(x-\mu_i)$ then we will prove below that

$$\mathrm{Res}(f, g) = \prod_{i,j}(\lambda_i - \mu_j).$$

If R is a UFD then by induction the polynomial ring $R[x_1, \ldots, x_n]$ is also a UFD, so we can make the following definition.

Resultants of Multivariable Polynomials. Let R be a UFD. For any index i we let $\mathbf{x}_i'$ denote the subset of the variables $\mathbf{x} = (x_1, \ldots, x_n)$ with x_i deleted. Let $f, g \in R[\mathbf{x}]$ be polynomials of degrees m, n and and expand each in powers of x_i:

$$f(\mathbf{x}) = a_0(\mathbf{x}_i')x_i^d + \cdots + a_{d-1}(\mathbf{x}_i')x_i + a_d(\mathbf{x}_i'),$$
$$g(x) = b_0(\mathbf{x}_i')x_i^e + \cdots + b_{e-1}(\mathbf{x}_i')x + b_e(\mathbf{x}_i'),$$

where $a_0(\mathbf{x}_i'), \ldots, a_d(\mathbf{x}_i'), b_0(\mathbf{x}_i'), \ldots, b_e(\mathbf{x}_i) \in R[\mathbf{x}_i']$. Then we define the *resultant* of f and g with respect to the variable x_i:

$$\mathrm{Res}_{x_i}(f,g) := \det \begin{pmatrix} a_0 & a_1 & \cdots & a_d & & \\ & \ddots & \ddots & & \ddots & \\ & & a_0 & a_1 & \cdots & a_d \\ b_0 & b_1 & \cdots & b_e & & \\ & \ddots & \ddots & & \ddots & \\ & & b_0 & b_1 & \cdots & b_e \end{pmatrix} \in R[\mathbf{x}_i'].$$

Here are some useful and interesting properties of resultants:

(a) We have $\mathrm{Res}_{x_i}(f,g) = \pm \mathrm{Res}_{x_i}(g,f)$.

(b) There exist polynomials $\tilde{f}, \tilde{g} \in R[\mathbf{x}]$ such that $\mathrm{Res}_{x_i}(f,g) = f\tilde{f} + g\tilde{g}$. [This is a sharper version of the elimination step used in the proof of Study's Lemma.]

(c) Suppose that f and g are **homogeneous** of degrees d and e, so that the coefficients $a_k(\mathbf{x}_i')$ and $b_k(\mathbf{x}_i')$ are homogeneous of degree k. Then one of the following holds:

- $\mathrm{Res}_{x_i}(f,g) = 0$, or
- $\mathrm{Res}_{x_i}(f,g)$ is homogeneous of degree $de = \deg(f)\deg(g)$.

(d) Suppose that $f, g \in R[x]$ polynomials in one variable that split as $f(x) = \prod_{i=1}^d (x - \lambda_i)$ and $g(x) = \prod_{i=1}^e (x - \mu_i)$. Then we have

$$\mathrm{Res}(f,g) = \mathrm{Res}_x(f,g) = \prod_{i,j} (\lambda_i - \mu_j) = \prod_{i=1}^d g(\lambda_i).$$

Furthermore, the identity $\mathrm{Res}(f,g) = \prod_i g(\lambda_i)$ holds even when g does not split in $R[x]$, since we can pass to a larger ring in which it does split.

(e) Finally, if $f'(x)$ is the derivative of $f(x)$ then it follows from (d) that

$$\mathrm{Res}(f,f') = \prod_{(i,j):i\neq j} (\lambda_i - \lambda_j) = \pm \prod_{(i,j):i<j} (\lambda_i - \lambda_j)^2.$$

Proof. (a): Note we can permute the rows cyclically by a sequence of row swaps. Each row swap multiplies the determinant by -1. We can move the b-rows to the top by a sequence of cyclic permutations.

(b): The determinant is unchanged if we replace the first column of Sylvester's matrix by

$$\sum_{k=0}^{d+e} x_i^k (\text{the } k\text{th column}).$$

Then the new first column has the form

$$\Big(f(\mathbf{x}), x_i f(\mathbf{x}), \dots, x_i^{d-1} f(\mathbf{x}), g(\mathbf{x}), x_i g(\mathbf{x}), \dots, x_i^{e-1} g(\mathbf{x})\Big)^T,$$

and expanding the determinant along this new column shows that

$$\mathrm{Res}_{x_i}(f,g) = f(\mathbf{x})\tilde{f}(\mathbf{x}) + g(\mathbf{x})\tilde{g}(\mathbf{x})$$

for some polynomials $\tilde{f}, \tilde{g} \in R[\mathbf{x}]$.

(c): Since $a_k, b_k \in R[\mathbf{x}_i']$ are homogeneous of degree k we have $a_k(\lambda \mathbf{x}_i') = \lambda^k a_k(\mathbf{x}_i')$ and $b_k(\lambda \mathbf{x}_i') = \lambda^k b_k(\mathbf{x}_i')$ for any $\lambda \in R \setminus 0$, and hence

$$\mathrm{Res}_{x_i}(f,g)(\lambda \mathbf{x}_i') = \det \begin{pmatrix} a_0 & \lambda a_1 & \cdots & \lambda^d a_d & & \\ & \ddots & \ddots & & \ddots & \\ & & a_0 & \lambda a_1 & \cdots & \lambda^d a_d \\ b_0 & \lambda b_1 & \cdots & \lambda^e b_e & & \\ & \ddots & \ddots & & \ddots & \\ & & b_0 & \lambda b_1 & \cdots & \lambda^e b_e \end{pmatrix} \in R[\mathbf{x}_i'].$$

Now multiply the first e rows by $\lambda, \lambda^2, \dots, \lambda^e$, respectively, and multiply the last d rows by $\lambda, \lambda^2, \dots, \lambda^d$, respectively, to obtain

$$\lambda^1\lambda^2\cdots\lambda^e\lambda^1\lambda^2\cdots\lambda^d \mathrm{Res}_{x_i}(f,g)(\lambda \mathbf{x}_i') = \det \begin{pmatrix} \lambda a_0 & \lambda^2 a_1 & \cdots & \lambda^{d+1} a_d & & \\ & \ddots & \ddots & & \ddots & \\ & & \lambda^e a_0 & \lambda^{e+1} a_1 & \cdots & \lambda^{d+e} a_d \\ \lambda b_0 & \lambda^2 b_1 & \cdots & \lambda^{e+1} b_e & & \\ & \ddots & \ddots & & \ddots & \\ & & \lambda^d b_0 & \lambda^{d+1} b_1 & \cdots & \lambda^{d+e} b_e \end{pmatrix}.$$

On the other hand, we can obtain the same matrix by multiplying the kth **column** of the original Sylvester matrix by λ^k,[23] so that

$$\begin{aligned} \lambda^1\lambda^2\cdots\lambda^d\lambda^1\lambda^2\cdots\lambda^e \mathrm{Res}_{x_i}(f,g)(\lambda \mathbf{x}_i') &= \lambda^1\lambda^2\cdots\lambda^{d+e} \mathrm{Res}_{x_i}(f,g)(\mathbf{x}_i) \\ \lambda^{d(d+1)/2}\lambda^{e(e+1)/2} \mathrm{Res}_{x_i}(f,g)(\lambda \mathbf{x}_i') &= \lambda^{(d+e)(d+e+1)/2} \mathrm{Res}_{x_i}(f,g)(\mathbf{x}_i) \\ \lambda^{(d^2+d+e^2+e)/2} \mathrm{Res}_{x_i}(f,g)(\lambda \mathbf{x}_i') &= \lambda^{(d^2+2de+e^2+d+e)/2} \mathrm{Res}_{x_i}(f,g) \\ \mathrm{Res}_{x_i}(f,g)(\lambda \mathbf{x}_i') &= \lambda^{de} \mathrm{Res}_{x_i}(f,g)(\mathbf{x}_i'). \end{aligned}$$

[23] Typesetting difficulties make this a bit hard to see. I recommend writing down an explicit example.

Since R is a domain this implies that $\mathrm{Res}_{x_i}(f,g) \in R[\mathbf{x}_i']$ is homogeneous of degree de.

(d): Consider the polynomials $f(x) = (x-y_1)\cdots(x-y_d)$ and $g(x) = (x-z_1)\cdots(x-z_e)$ as elements of the ring $R[x, y_1, \ldots, y_d, z_1, \ldots, z_e]$. Since f and g are homogeneous of degrees d and e, then we conclude from part (c) that $R(\mathbf{y},\mathbf{z}) := \mathrm{Res}_x(f,g) \in R[\mathbf{y},\mathbf{z}]$ is homogeneous of degree de. On the other hand, we observe that the polynomial

$$\Phi(\mathbf{y},\mathbf{z}) := \prod_{i=1}^{d}\prod_{j=1}^{e}(y_i - z_j) \in R[\mathbf{y},\mathbf{z}]$$

is also homogeneous of degree de. To show that these two polynomials are equal, consider what happens if we substitute z_j for y_i in the polynomial $R(\mathbf{y},\mathbf{z})$. Since substitution is a ring homomorphism, this is the same as first substituting z_j for y_i in f and then computing the resultant. In this case we observe that f and g have the common factor $(x - z_j)$, so the resultant is zero by Sylvester's theorem. In other words, if we think of $R(\mathbf{y},\mathbf{z})$ as a polynomial in the variable y_i then z_j is a root, hence it follows from Descartes' Factor Theorem that $y_i - z_j$ divides $R(\mathbf{y},\mathbf{z})$. Since the same argument holds for any i, j we conclude that Φ divides R, hence we must have $\Phi = \lambda R$ for some $\lambda \in R \setminus 0$. Finally, we observe that the monomial $(z_1\cdots z_e)^d$ has the coefficient $(-1)^{de}$ in R. Indeed, this term can only come from the diagonal term of the determinant. But this is also the coefficient in Φ, hence $\lambda = 1$.

(e): If $f(x) = \prod_{i=1}^{d}(x-\lambda_i)$ then we have $f'(x) = \sum_k \prod_{j:j\neq k}(x-\lambda_j)$. If we evaluate this at λ_i then the only non-zero summand corresponds to $k = i$ and hence $f'(\lambda_i) = \prod_{j:j\neq i}(\lambda_i - \lambda_j)$. It follows from (d) that

$$\mathrm{Res}(f,f') = \prod_{i=1}^{d} f'(\lambda_i) = \prod_{(i,j):i\neq j}(\lambda_i - \lambda_j) = \pm \prod_{(i,j):i<j}(\lambda_i - \lambda_j)^2.$$

□

As a corollary, we obtain a classification of the prime ideals of the ring $\mathbb{F}[x,y]$ when $\mathbb{F}$ is algebraically closed. We already know about three kinds of prime ideals:

- The zero ideal $0 \subseteq \mathbb{F}[x,y]$, which is prime because $\mathbb{F}[x,y]$ is a domain.
- Principal prime ideals $f\mathbb{F}[x,y]$ where $f(x,y)$ is irreducible.
- Maximal (prime) ideals $M_{a,b} = (x-a)\mathbb{F}[x,y] + (y-b)\mathbb{F}[x,y]$.

Resultants allow us to prove that these are the **only** prime ideals of $\mathbb{F}[x,y]$.

Proof. Let $0 \subsetneq P \subseteq \mathbb{F}[x,y]$ be a non-zero prime ideal. I claim that P contains an irreducible element f. Indeed, take any non-zero element $g \in P$. Since P is prime at least one irreducible factor of g must lie in P. If $P = f\mathbb{F}[x,y]$ then we are done. Otherwise, take $g \in \mathbb{F}[x,y] \setminus f\mathbb{F}[x,y]$. Since f is irreducible and $f \nmid g$ we see that f and g have no common irreducible factor. It

follows from the above results that $\mathrm{Res}_y(f,g) \in \mathbb{F}[x]$ is a non-zero element of P.[24] Since $\mathbb{F}$ is algebraically closed we may write

$$\mathrm{Res}_y(f,g) = \prod_i (x - a_i).$$

Then since P is prime we must have $x - a \in P$ for some $a = a_i \in \mathbb{F}$. A similar argument shows that $y - b \in P$ for some $b \in \mathbb{F}$, so that $M_{a,b} \subseteq P$. Finally, since $M_{a,b}$ is a maximal ideal we conclude that $P = M_{a,b}$. □

Remark: On the homework you will show that this result can be generalized to any ring $R[x]$ where R is a PID. In this case the prime ideals are:

- 0,
- $fR[x]$ for irreducible f,
- $M_{f,p} = fR[x] + pR[x]$ where $p \in R$ is a prime element of R and $f(x)$ is a polynomial whose reduction mod p is irreducible in $(R/pR)[x]$.

Oct 31 and Nov 2: Hilbert's Nullstellensatz

Study's Lemma describes the polynomials that vanish on a hypersurface V_f over an algebraically closed field: They are just the polynomials that are divisible by the square-free part of f. Equivalently, if g vanishes on V_f then some power of g is divisible by f. Hilbert's Nullstellensatz generalizes this result to **intersections** of hypersurfaces. The result is usually divided into a weak version and strong version, though the two results are logically equivalent. The weak version was originally much easier to prove. However, after the discovery of the "trick of Rabinowitsch" this is no longer the case.

The Nullstellensatz is preceded by a necessary lemma, which is also due to Hilbert. First I will state the classical, geometric version of these results. Afterwards I will prove them in their modern, ring-theoretic forms.

Hilbert's Basis Theorem (Classical Form). Any intersection of hypersurfaces (affine or projective) can be expressed as an intersection of **finitely many** hypersurfaces.

Hilbert's Nullstellensatz (Classical Form). Let $f_1, \ldots, f_m \in \mathbb{F}[x_1, \ldots, x_n]$ with $\mathbb{F}$ algebraically closed, and consider the intersection of the corresponding hypersurfaces:

$$V = V_{f_1} \cap V_{f_2} \cap \cdots \cap V_{f_m} \subseteq \mathbb{F}^n.$$

- **(Weak):** If $V = \emptyset$ then there exist $\tilde{f}_1, \ldots, \tilde{f}_m \in \mathbb{F}[\mathbf{x}]$ such that

$$1 = f_1\tilde{f}_1 + f_2\tilde{f}_2 + \cdots + f_m\tilde{f}_m.$$

[24] Non-zero because f and g have no common factor, and an element of P because it is an $\mathbb{F}[x,y]$-linear combination of f and g.

- **(Strong):** If g vanishes on V then there exist $r \geq 1$ and $\tilde{f}_1, \ldots, \tilde{f}_m \in \mathbb{F}[\mathbf{x}]$ such that

$$g^r = f_1\tilde{f}_1 + f_2\tilde{f}_2 + \cdots + f_m\tilde{f}_m.$$

The strong form implies the weak form since every function vanishes on the empty set. We will see below that the weak form also implies the strong form.

Hilbert's Basis Theorem was revolutionary because its proof was non-constructive and ideal-theoretic. The modern form is expressed in terms of *Noetherian rings*. These are named for Emmy Noether, who is responsible for the abstract, ring-theoretic approach to polynomials. As we saw in the case of PIDs, the Noetherian condition is an abstract substitute for well-ordering.

Noetherian Rings. Let R be a ring. Then the following are equivalent:

(1) Ideals of R are *finitely generated*. That is, for any ideal $I \subseteq R$ there exist some (non-unique) elements $a_1, \ldots, a_n \in R$ such that $I = a_1R + \cdots + a_nR$.

(2) Ideals of R satisfy the *ascending chain condition*. That is, if $I_1 \subseteq I_2 \subseteq \cdots$ is any ascending chain of ideals, then there exists some n such that $I_n = I_{n+1} = \cdots$.

(3) Every non-empty set of ideals contains a maximal element.

Proof. (1)⇒(2): Assume for contradiction that we have an infinite ascending chain of ideals $I_1 \subsetneq I_2 \subsetneq \cdots$. Since the union $J = \cup_k I_k$ is an ideal,[25] we have that $J = a_1R + \cdots + a_nR$ for some $a_1, \ldots, a_n \in R$. By definition, each a_i occurs in some I_{k_i}. If $k = \max_i\{k_i\}$ then we have $\{a_1, \ldots, a_n\} \subseteq I_k$, and then since I_k is an ideal we obtain the following contradiction:

$$J = a_1R + \cdots + a_nR \subseteq I_k \subsetneq I_{k+1} \subsetneq J.$$

(2)⇒(1): Assume for contradiction that there exists an ideal $I \subseteq R$ that is not finitely generated. For any $a_1 \in I$ this implies that $I = a_1R$, so we may choose some $a_2 \in I \setminus a_1R$. Then by induction we may choose $a_k \in I \setminus (a_1R + \cdots + a_{k_1}R)$. Thus we obtain an infinite ascending chain of ideals $a_1R \subsetneq (a_1R + a_2R) \subsetneq \cdots$.

(2)⇒(3): Assume for contradiction that there exists a non-empty set of ideals S with no maximal element and let $I_1 \in S$. Since I_1 is not maximal, there exists some $I_2 \in S$ such that $I_1 \subsetneq I_2$. By induction we obtain an infinite ascending chain of ideals: $I_1 \subsetneq I_2 \subsetneq \cdots$.

(3)⇒(2): Consider any infinite chain of ideals $I_1 \subseteq I_2 \subseteq \cdots$ and let $S = \{I_k\}$. By assumption this set has a maximal element $I_n \in S$ so that $I_n = I_{n+1} = \cdots$. □

Compare the following statement to the modern form of Gauss' Lemma, which says that if R is a UFD then $R[x]$ is a UFD. Many properties of polynomial rings are inherited in this way.

[25] We proved this earlier in the chapter on PIDs.

Hilbert's Basis Theorem (Modern Form). If R is Noetherian then $R[x]$ is Noetherian. In particular, if $\mathbb{F}$ is a field then $\mathbb{F}[x_1, \ldots, x_n]$ is Noetherian for any n.

Proof. Assume for contradiction that $R[x]$ is not Noetherian and let $I \subseteq R[x]$ be a non-zero ideal that is not finitely generated. Choose any non-zero $f_1(x) \in I$ of minimal degree. Since $I \neq f_1 R[x]$ we may choose $f_2(x) \in I \setminus f_1 R[x]$ of minimal degree, and then by induction we may choose some $f_k(x) \in I \setminus (f_1 R[x] + \cdots + f_{k-1} R[x])$ of minimal degree, so that $\deg(f_1) \leq \deg(f_2) \leq \cdots$.[26] To obtain a contradiction we let $a_k \in R$ be the leading coefficient of $f_k(x)$. Then I claim for all k that

$$a_1 R + \cdots a_{k-1} R \subsetneq a_1 R + \cdots + a_k R,$$

which contradicts the fact that R is Noetherian.

To prove this, suppose for contradiction that we have $a_k \in a_1 R + \cdots a_{k-1} R$, so that $a_k = a_1 b_1 + \cdots a_{k-1} b_{k-1}$ for some $b_1, \ldots, b_{k-1} \in R$. It follows that

$$g_k(x) := f_k(x) - b_1 x^{\deg(f_k) - \deg(f_1)} f_1(x) - \cdots - b_{k-1} x^{\deg(f_k) - \deg(f_{k-1})} f_{k-1}(x)$$

has leading coefficient $a_k - a_1 b_1 - \cdots - a_{k-1} b_{k-1} = 0$, and hence $\deg(g_k) < \deg(f_k)$. We must also have $g_k \in I \setminus (f_1 R[x] + \cdots + f_{k-1} R[x])$ since otherwise we obtain the contradiction $f_k \in f_1 R[x] + \cdots f_{k-1} R[x]$:

$$f_k(x) := g_k(x) + b_1 x^{\deg(f_k) - \deg(f_1)} f_1(x) + \cdots + b_{k-1} x^{\deg(f_k) - \deg(f_{k-1})} f_{k-1}(x).$$

In summary, we have found an element of $I \setminus (f_1 R[x] + \cdots + f_{k-1} R[x])$ with degree strictly less than $f_k(x)$. Contradiction. □

Hilbert's Nullstellensatz (Modern Form). Let $\mathbb{F}$ be algebraically closed and consider the ring $\mathbb{F}[\mathbf{x}] = \mathbb{F}[x_1, \ldots, x_n]$. For any ideal $I \subseteq \mathbb{F}[\mathbf{x}]$ we define the set of points

$$\mathbf{V}(I) = \{\mathbf{p} \in \mathbb{F}^n : f(\mathbf{p}) = 0 \text{ for all } f \in I\}.$$

To connect this with the classical case, we observe from the basis theorem that $I = f_1 \mathbb{F}[\mathbf{x}] + \cdots + f_m \mathbb{F}[\mathbf{x}]$ for some finite set of polynomials $f_1, \ldots, f_m \in \mathbb{F}[\mathbf{x}]$, and hence $\mathbf{V}(I)$ is an intersection of finitely many hypersurfaces:

$$\mathbf{V}(I) = V_{f_1} \cap V_{f_2} \cap \cdots \cap V_{f_m}.$$

Indeed, if $\mathbf{p} \in \mathbf{V}(I)$ then since $f_1, \ldots, f_m \in I$ we have $f_i(\mathbf{p}) = 0$ for all i. On the other hand, if $f_i(\mathbf{p}) = 0$ for all i, then for any $f(x) = f_1(x) g_1(x) + \cdots f_m(x) g_m(x) \in I$ we have

$$f(\mathbf{p}) = f_1(\mathbf{p}) g_1(\mathbf{p}) + \cdots f_m(\mathbf{p}) g_m(\mathbf{p}) = 0 g_1(\mathbf{p}) + \cdots 0 g_m(\mathbf{p}) = 0.$$

Here are the weak and strong versions of the theorem.

[26] Note that we make use of the well-ordering principle for integers. Polynomial rings always maintain some relationship to the integers through their degree. They are not completely abstract.

(Weak). If $\mathbf{V}(I) = \emptyset$ then $I = \mathbb{F}[\mathbf{x}]$ (equivalently, $1 \in I$). Geometrically: If an intersection of hypersurfaces is empty, $V_{f_1} \cap \cdots \cap V_{f_m} = \emptyset$, then we can find some $\tilde{f}_1, \ldots, \tilde{f}_m \in \mathbb{F}[\mathbf{x}]$ such that $1 = f_1\tilde{f}_1 + \cdots + f_m\tilde{f}_m$.

(Strong). If $g \in \mathbb{F}[\mathbf{x}]$ vanishes on $\mathbf{V}(I)$ then we must have $g^r \in I$ for some $r \geq 0$. Geometrically: If g vanishes on an intersection of hypersurfaces $V_{f_1} \cap \cdots \cap V_{f_m}$ then we must have $g^r = f_1\tilde{f}_1 + \cdots + f_m\tilde{f}_m$ for some $r \geq 0$ and $\tilde{f}_1, \ldots, \tilde{f}_m \in \mathbb{F}[\mathbf{x}]$.

Proof. **(Weak):** We will use induction on the number of variables n. *Base Case* ($n = 1$): Given $I \neq \mathbb{F}[x_1]$ we will show that $\mathbf{V}(I) \neq \emptyset$. Since $\mathbb{F}[x_1]$ is a PID we must have $I = f\mathbb{F}[x_1]$ for some $f(x_1) \in \mathbb{F}[x_1]$. And since $I \neq \mathbb{F}[x_1]$ we see that f is non-constant (or zero). Finally, since $\mathbb{F}$ is algebraically closed we must have $f(p) = 0$ for some $p \in \mathbb{F}$, giving the contradiction $p \in V_f = \mathbf{V}(I)$. *Induction Step* ($n \geq 2$): Assuming that $I \neq \mathbb{F}[\mathbf{x}]$ (equivalently, $1 \notin I$) we will show that $\mathbf{V}(I) \neq \emptyset$. This time the ring $\mathbb{F}[\mathbf{x}]$ is not a PID we we will have to be more clever.

Normalization Step: For any invertible linear transformation $A \in \mathrm{GL}_n(\mathbb{F})$ we consider the set of polynomials $AI = \{f(A^{-1}\mathbf{x}) : f(\mathbf{x}) \in I\}$ and the set of points $A\mathbf{V}(I) := \{A\mathbf{p} : \mathbf{p} \in \mathbf{V}(I)\}$, and we observe that

$$A\mathbf{V}(I) = \mathbf{V}(AI).$$

Since $A : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{x}]$ is a ring homomorphism we see that $AI \subseteq \mathbb{F}[\mathbf{x}]$ is an ideal. And since A is invertible we see that $A\mathbf{V}(I) = \emptyset \Leftrightarrow \mathbf{V}(I) = \emptyset$ and $AI = \mathbb{F}[\mathbf{x}] \Leftrightarrow I = \mathbb{F}[\mathbf{x}]$. Finally, from the Normalization Lemma we can choose A so that some polynomial $f \in AI$ has the form $f = cx_n^d + \text{lower terms in } x_n$. Thus, for the purpose of the proof we might as well assume that some element of I has this form.

Elimination Step: Let $\mathbf{x}' = (x_1, \ldots, x_{n-1})$ and consider the ideal $I' := I \cap \mathbb{F}[\mathbf{x}']$ of $\mathbb{F}[\mathbf{x}']$. Since $1 \notin I$ we observe that $1 \notin I'$. Hence by induction there exists a point $\mathbf{p}' = (p_1, \ldots, p_{n-1}) \in \mathbb{F}^{n-1}$ such that $f(\mathbf{p}') = 0$ for all $f \in I'$.[27] Now consider the set

$$J = \{f(\mathbf{p}', x_n) : f \in I\} \subseteq \mathbb{F}[x_n].$$

Since evaluation is a ring homomorphism, this set is an ideal. If $1 \notin J$ then from the base case there exists $p_n \in \mathbb{F}$ such that $f(\mathbf{p}) = f(\mathbf{p}', p_n) = 0$ for all $f \in I$. This implies that $\mathbf{p} \in \mathbf{V}(I)$, which completes the proof. So let us assume for contradiction that $1 \in J$, so that $1 = g(\mathbf{p}', x_n)$ for some $g(\mathbf{x}) \in I$. Assuming that $\deg(g) = e$ we expand in terms of x_n to obtain

$$\begin{aligned} g(\mathbf{x}) &= b_0(\mathbf{x}')x_n^e + b_1(\mathbf{x})x_n^{e-1} + \cdots b_{e-1}(\mathbf{x})x_n + b_e(\mathbf{x}) \\ 1 = g(\mathbf{p}', x_n) &= b_0(\mathbf{p}')x_n^e + b_1(\mathbf{p}')x_n^{e-1} + \cdots b_{e-1}(\mathbf{p}')x_n + b_e(\mathbf{p}'), \end{aligned}$$

which implies that $b_0(\mathbf{p}') = \cdots = b_{e-1}(\mathbf{p}') = 0$ and $b_e(\mathbf{p}') = 1$. Furthermore, we may assume **from the normalization step** that there exists some $f(\mathbf{x}) \in I$ of the form

$$\begin{aligned} f(\mathbf{x}) &= cx_n^d + a_1(\mathbf{x}')a_n^{d-1} + \cdots a_{d-1}(\mathbf{x}')x_n + a_d(\mathbf{x}') \\ f(\mathbf{p}', x_n) &= cx_n^d + a_1(\mathbf{p}')x_n^{d-1} + \cdots a_{d-1}(\mathbf{p}')x_n + a_d(\mathbf{p}'), \end{aligned}$$

[27] There is a subtle issue here because the ideal I' is not obviously finitely generated. It seems impossible to prove the classical Nullstellensatz (regarding finite intersections of hypersurfaces) without invoking the Hilbert Basis Theorem.

with $c \in \mathbb{F} \setminus 0$. Finally, we will obtain a contradiction by looking at the resultant polynomial $\mathrm{Res}_{x_n}(f,g) \in \mathbb{F}[\mathbf{x}']$. Since $f, g \in I$ and since the resultant is an $\mathbb{F}[\mathbf{x}]$-linear combination of f and g, we observe that $\mathrm{Res}_{x_n}(f,g) \in I \cap \mathbb{F}[\mathbf{x}'] = I'$. Thus it follows from the original definition of $\mathbf{p}'$ that $\mathrm{Res}_{x_n}(f,g)(\mathbf{p}') = 0$. On the other hand, Sylvester's determinant tells us that

$$\mathrm{Res}_{x_n}(f,g) = \pm \det \begin{pmatrix} c & a_{d-1}(\mathbf{p}') & \cdots & a_0(\mathbf{p}') & & \\ & \ddots & \ddots & & \ddots & \\ & & c & a_{d-1}(\mathbf{p}') & \cdots & a_0(\mathbf{p}') \\ 0 & \cdots & 0 & 1 & & \\ & \ddots & \ddots & & \ddots & \\ & & 0 & \cdots & 0 & 1 \end{pmatrix} = c^e \neq 0.$$

This completes the proof of the Weak Nullstellensatz.

(Strong): We will use the "trick of Rabinowitsch"[28] to derive the strong version from the weak version. From the Hilbert Basis Theorem we can write $I = f_1\mathbb{F}[\mathbf{x}] + \cdots + f_m\mathbb{F}[\mathbf{x}]$, and hence $\mathbf{V}(I) = V_{f_1} \cap \cdots \cap V_{f_m}$ is an intersection of finitely many hyperplanes. Suppose that $g(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ vanishes on the set $\mathbf{V}(I)$, so that $f_1(\mathbf{p}) = \cdots = f_m(\mathbf{p}) = 0$ implies $g(\mathbf{p}) = 0$. Now we introduce another variable y and consider the consider the polynomials $f_1, \ldots, f_m, 1 - yg \in \mathbb{F}[\mathbf{x}, y]$. By construction, the following intersection of hyperplanes in one higher dimension is empty:

$$V^+ = V_{f_1} \cap \cdots \cap V_{f_m} \cap V_{1-yg} \subseteq \mathbb{F}^{n+1}.$$

Indeed, if $\mathbf{p}^+ = (p_1, \ldots, p_n, q)$ satisfies $f_i(\mathbf{p}) = f_i(\mathbf{p}^+) = 0$ for all i then we must have $(1 - yg)(\mathbf{p}^+) = 1 - qg(\mathbf{p}) = 1 - q0 = 1 \neq 0$. Thus from the Weak Nullstellensatz there exist polynomials $\tilde{h}_1, \ldots, \tilde{h}_m, \tilde{g} \in \mathbb{F}[\mathbf{x}, y]$ such that

$$1 = f_1(\mathbf{x})\tilde{h}_1(\mathbf{x}, y) + \cdots + f_m(\mathbf{x})\tilde{h}_m(\mathbf{x}, y) + (1 - yg(\mathbf{x}))\tilde{g}(\mathbf{x}, y).$$

Now we substitute $y = 1/g(\mathbf{x})$ to obtain an identity in the field of fractions $\mathbb{F}(\mathbf{x})$:

$$1 = f_1(\mathbf{x})\tilde{h}_1(\mathbf{x}, 1/g) + \cdots + f_m(\mathbf{x})\tilde{h}_m(\mathbf{x}, 1/g) + 0.$$

We observe that the least common denominator of the right side has the form $g(\mathbf{x})^r$ for some $r \geq 0$. Hence there exist some polynomials $\tilde{f}_1, \ldots, \tilde{f}_m \in \mathbb{F}[\mathbf{x}]$ satisfying

$$1 = \left(f_1\tilde{f}_1 + f_2\tilde{f}_2 + \cdots + f_m\tilde{f}_m\right)/g^r$$
$$g^r = f_1\tilde{f}_1 + f_2\tilde{f}_2 + \cdots + f_m\tilde{f}_m.$$

In other words, $g^r \in I$. □

Remark: Observe that the classical and modern versions of the theorem cannot be separated. The classical proof of the Weak Nullstellensatz must invoke the modern statement, and the modern proof of the Strong Nullstellensatz must invoke the classical statement. The Hilbert Basis Theorem provides the connection between the classical and modern forms. This was one of the first motivations for abstract ring theory in the foundations of algebraic geometry.

[28] Rabinowitsch, *Zum Hilberten Nullstellensatz* (1929)

The Zariski Topology

Nov 2: Minimal and Maximal Prime ideals

For any field $\mathbb{F}$ we have seen that a point $\mathbf{p} \in \mathbb{F}^n$ determines a maximal (prime) ideal $M_{\mathbf{p}} \subseteq \mathbb{F}[x_1, \ldots, x_n]$. And if $\mathbb{F}$ is algebraically closed then Study's Lemma provides a bijection between irreducible hypersurfaces in $\mathbb{F}^n$ and irreducible polynomials in $\mathbb{F}[x_1, \ldots, x_n]$. The following theorems translate these facts into modern language.

[Conventions: The unit ideal is not prime. The zero and unit ideals are not minimal or maximal. A *minimal prime ideal* is minimal among prime ideals, not necessarily among all ideals.]

Minimal Prime Ideals (Extension of Study's Lemma). Let R be a UFD. Then minimal prime ideals are the same as principal prime ideals. If $\mathbb{F}$ is algebraically closed then applying this result to $R = \mathbb{F}[x_1, \ldots, x_n]$ gives a bijection between minimal prime ideals in $\mathbb{F}[x_1, \ldots, x_n]$ and irreducible hypersurfaces in $\mathbb{F}^n$.

Proof. Let $0 \subsetneq pR \subsetneq R$ be a principal prime ideal and let Q be any prime ideal satisfying $0 \subsetneq Q \subseteq pR$. (At least one such Q exists because pR is prime.) Choose any non-zero, non-unit element $f \in Q$ with irreducible factorization $f = p_1 \cdots p_k$. Since Q is prime there exists some i such that $p_i \in Q$ and hence $p_i R \subseteq Q \subseteq pR$. Since p_i is irreducible with $p | p_i$ and $p \not\sim 1$ we must have $p_i \sim p$ and hence $p_i R = Q = pR$. Thus pR is a minimal prime. Conversely, let $0 \subsetneq P \subsetneq R$ be a minimal prime ideal. Choose any non-zero, non-unit $f \in P$ with irreducible factorization $f = p_1 \cdots p_k$. Since P is prime there exists some i such that $p_i \in P$ and hence $p_i R \subseteq P$. Since irreducible elements of a UFD are prime we see that $p_i R$ is a prime ideal and hence $p_i R = P$ by the minimality of P. Thus P is principal. □

Maximal Prime Ideals (Weak Nullstellensatz). Let $\mathbb{F}$ be algebraically closed. Then every maximal ideal of $\mathbb{F}[x_1, \ldots, x_n]$ has the form $M_p = \{f \in \mathbb{F}[\mathbf{x}] : f(\mathbf{p}) = 0\}$ for some point $\mathbf{p} \in \mathbb{F}^n$. This result is logically equivalent to the Weak Nullstellensatz (hence also to the Strong Nullstellensatz).

Proof. Let $M \subsetneq \mathbb{F}[\mathbf{x}]$ be a maximal (prime) ideal. Since $M \neq \mathbb{F}[\mathbf{x}]$ we know from the Weak Nullstellensatz that $\mathbf{V}(M) \neq \emptyset$, say $\mathbf{p} \in \mathbf{V}(M)$. But then $M_{\mathbf{p}} \subseteq \mathbf{V}(M)$, which implies that $M_{\mathbf{p}} = V(M)$ because $M_{\mathbf{p}}$ is maximal. Conversely, we will show that this result implies the Weak Nullstellensatz. So suppose that every maximal ideal has the form $M_{\mathbf{p}}$ and let $I \subseteq \mathbb{F}[\mathbf{x}]$ be an ideal satisfying $I \neq \mathbb{F}[\mathbf{x}]$. I claim that I is contained in a maximal ideal. Indeed, if this is not the case then we obtain an infinite ascending chain $I = I_1 \subsetneq I_2 \subsetneq \cdots$, contradicting the fact that $\mathbb{F}[\mathbf{x}]$ is Noetherian (Hilbert's Basis Theorem).[29] Thus we have $I \subseteq M_{\mathbf{p}}$ for some point $\mathbf{p}$ and it follows that $\mathbf{p} \in \mathbf{V}(I)$, hence $\mathbf{V}(I) \neq \emptyset$. □

[29] More generally, one can invoke Zorn's Lemma to show that any non-unit ideal in any ring is contained in some maximal ideal. But I prefer not to invoke Zorn's Lemma.

Nov 4: Galois Connections

For any ideal $I \subseteq \mathbb{F}[\mathbf{x}]$ we have defined the zero set $\mathbf{V}(I) \subseteq \mathbb{F}^n$, which is called an *algebraic variety*. More generally, we define the zero set $\mathbf{V}(S) \subseteq \mathbb{F}^n$ for any **set** of polynomials $S \subseteq \mathbb{F}[\mathbf{x}]$. On the other hand, for any set of points $S \subseteq \mathbb{F}^n$ we define the set of functions that vanish on this set, $\mathbf{I}(S) := \{f \in \mathbb{F}[\mathbf{x}] : f(\mathbf{p}) = 0 \text{ for all } f \in S\}$, which one can check is an ideal, hence the notation. Thus we have a pair of maps $\mathbf{I}, \mathbf{V}$ sending sets of polynomials to sets of points, and vice versa. The Zariski topology provides a dictionary between varieties and ideals. The framework of the dictionary is purely "formal," having nothing whatsoever to do with polynomials. We we will take care of this first.

Abstract Galois Connections. let $(\mathcal{P}, \leq)$ and $(\mathcal{Q}, \leq)$ be partially ordered sets, and suppose we have maps $* : \mathcal{P} \to \mathcal{Q}$ and $* : \mathcal{Q} \to \mathcal{P}$ satisfying the following property:

$$\forall p \in \mathcal{P}, \forall q \in \mathcal{Q}, \quad p \leq q^* \Leftrightarrow q \leq p^*.$$

We use the same name for both maps because the definition is symmetric. Then for all $p, p_1, p_2 \in \mathcal{P}$ and $q, q_1, q_2 \in \mathcal{Q}$ we have the following properties:

(a) We have $p \leq p^{**}$ and $q \leq q^{**}$.

(b) We have $p_1 \leq p_2 \Rightarrow p_2^* \leq p_1^*$ and $q_1 \leq q_2 \Rightarrow q_2^* \leq q_1^*$.

(c) We have $p^* = p^{***}$ and $q^* = q^{***}$.

(d) We let $\mathrm{cl}(\mathcal{P}) := \{p : p = p^{**}\} \subseteq \mathcal{P}$ be the set of *closed elements* of $\mathcal{P}$ and we let $\mathcal{P}^* := \{p^* : p \in \mathcal{P}\} \subseteq \mathcal{Q}$ denote the image of $* : \mathcal{P} \to \mathcal{Q}$. Similarly, we have $\mathrm{cl}(\mathcal{Q}) \subseteq \mathcal{Q}$ and $Q^* \subseteq \mathcal{P}$. Then I claim that

$$\mathrm{cl}(\mathcal{P}) = \mathcal{Q}^* \quad \text{and} \quad \mathrm{cl}(\mathcal{Q}) = \mathcal{P}^*.$$

(e) The maps $* : \mathcal{P} \rightleftarrows \mathcal{Q} : *$ restrict to an *anti-isomorphism* (a bijection that reverses order) between the sub-posets of closed elements:

$$* : \mathrm{cl}(\mathcal{P}) \xleftrightarrow{\sim} \mathrm{cl}(\mathcal{Q}) : *.$$

(f) Suppose further that $(\mathcal{P}, \leq, \vee, \wedge)$ and $(\mathcal{Q}, \leq, \vee, \wedge)$ are lattices. This means that for any elements $\{p_i\} \subseteq \mathcal{P}$ there exists a least upper bound $\vee_i p_i \in \mathcal{P}$ and a greatest lower bound $\wedge_i p_i \in \mathcal{P}$, and similarly for any elements $\{q_i\} \subseteq \mathcal{Q}$. In this case we have

$$\begin{array}{lll} \vee_i p_i^* \leq (\wedge_i p_i)^* & \text{and} & \vee_i q_i^* \leq (\wedge_i q_i)^*, \\ \wedge_i p_i^* = (\vee_i p_i)^* & \text{and} & \wedge_i q_i^* = (\vee_i q_i)^*. \end{array}$$

It follows from the second identities that the greatest lower bound of closed elements is closed. However, the least upper bound of closed elements is **not** generally closed.

Proof. Because of symmetry we only need to prove one half of each statement.

(a): Let $q = p^*$ so that $p^* \leq p^* \Rightarrow q \leq p^* \Rightarrow p \leq q^* \Rightarrow p \leq p^{**}$.

(b): Let $p_1 \leq p_2$. Then from (a) we have $p_1 \leq p_2 \leq p_2^{**}$, and hence $p_1 \leq p_2^{**} = (p_2^*)^*$. Then by definition we have $p_2^* \leq p_1^*$.

(c): On the one hand, since $(p^{**}) \leq p^{**} = (p^*)^*$ we have by definition that $(p^*) \leq (p^{**})^* = p^{***}$. On the other hand, from (a) we have $(p) \leq (p^{**})$ then from (b) we have $(p^{**})^* \leq (p)^*$; or, in other words, $p^{***} \leq p^*$.

(d): If $p = p^{**}$ is closed then we have $p = q^* \in \mathcal{Q}^*$ where $q = p^*$. Conversely, if $p = q^* \in \mathcal{Q}^*$ for some q then from part (c) we have $p^{**} = q^{***} = q^* = p$.

(e): First we show that the map $* : \mathrm{cl}(\mathcal{P}) \to \mathrm{cl}(\mathcal{Q})$ is is bijective. *Surjective.* From (d) we know that every element of $\mathrm{cl}(\mathcal{Q}) = \mathcal{P}^*$ looks like p^*. Then from (c) we have $(p^{**})^* = p^*$, so that $*$ sends $p^{**} = (p^*)^* \in \mathcal{Q}^* = \mathrm{cl}(\mathcal{P})$ to p^* as desired. /// *Injective.* Suppose that there exist $q_1^*, q_2^* \in \mathcal{Q}^* = \mathrm{cl}(\mathcal{P})$ that get send to the same element $q_1^{**} = q_2^{**}$. Then from (c) we have $q_1^* = (q_1^{**})^* = (q_2^{**})^* = q_2^*$. /// Finally, we observe from part (b) that this bijection preserves order in both directions.

(f): For the first identities we observe from the definition of lower bound that $\wedge_i p_i \leq p_j$ for all j, and then from part (b) we have $p_j^* \leq (\wedge_i p_i)^*$ for all j. In other words, $(\wedge_i p_i)^*$ is an upper bound of the set $\{p_j\}$. Thus from the definition of **least** upper bound we have $\vee_j p_j^* \leq (\wedge_i p_i)^*$.

Now we prove the second identities. On the one hand, a symmetric version of the previous proof shows that $(\vee_i p_i)^* \leq \wedge_i p_i^*$. Indeed, since $\vee_i p_i$ is an upper bound of the set $\{p_i\}$ it follows from (b) that $(\vee_i p_i)^*$ is a lower bound of the set $\{p_i^*\}$, and hence is less than the **greatest** lower bound $\wedge_i p_i^*$. On the other hand, by the definition of lower bound we have $\wedge_j p_j^* \leq p_i^*$ for all i, hence from the definition of Galois connection we have $p_i \leq (\wedge_j p_j^*)^*$. In other words, $(\wedge_j p_j^*)^*$ is an upper bound of the set $\{p_i\}$. Thus the **least** upper bound satisfies $\vee_i p_i \leq (\wedge_j p_j^*)^*$ and we apply the definition of Galois connection one more to conclude that $\wedge_j p_j^* \leq (\vee_i p_i)^*$. In summary, we have shown that $\wedge_j p_j^* = (\vee_i p_i)^*$. □

[Remark: This proof waas not "hard," but it was . . . delicate. The notion of a Galois connection between posets is the easiest example of an adjunction between categories.]

Example:

- We already saw an example of a Galois connection when we discussed the correspondence theorem for ideals. Let $\varphi : R \to S$ a ring homomorphism and let $\mathscr{L}(R), \mathscr{L}(S)$ be the sets of ideals of R, S. Then the image/pre-image are a Galois connection of posets:[30]

  $$\varphi : (\mathscr{L}(R), \subseteq) \rightleftarrows (\mathscr{L}(S), \supseteq) : \varphi^{-1}.$$

  Indeed, we have already seen that image and pre-image send ideals to ideals. Then we observe for any ideals $I \in \mathscr{L}(R)$ and $J \in \mathscr{L}(S)$ that

  $$I \subseteq \varphi^{-1}[J] \quad \Leftrightarrow \quad \forall a \in I, \varphi(a) \in J \quad \Leftrightarrow \quad J \supseteq \varphi[I].$$

[30]Note that we must use **reverse** the inclusion order on $\mathscr{L}(S)$.

Thus from part (e) above we obtain an **order-preserving** bijection between the subposets of "closed ideals" on each side:

$$\varphi : \varphi^{-1}[\mathscr{L}(S)] \xleftrightarrow{\sim} \varphi[\mathscr{L}(R)] : \varphi^{-1}.$$

And it only remains to determine exactly which ideals are "closed." One can show that

$$\varphi^{-1}[\varphi[I]] = I + \ker\varphi,$$
$$\varphi[\varphi^{-1}[J]] = J.$$

The second identity says that **every** ideal of S is closed. The first identity says that $I \subseteq R$ is closed if and only if $I = I + \ker\varphi$, i.e., if and only if $\ker\varphi \subseteq I$. This completes the proof of the correspondence theorem from Sept 11. ///

As the name suggests, the notion of "Galois closure" is related to the notion of "topological closure." To capture this, we define the following notion of a *Galois closure space*, which is nearly identical with Kuratowski's axiomatization of topology using closure operators.[31]

Galois Closure Spaces. Let X be a set and let $\mathrm{cl} : 2^X \to 2^X$ be a function taking subsets to subsets. We say that cl is a *Galois closure* if for all $S, T \subseteq X$ the following hold:

(T1) We have $S \subseteq \mathrm{cl}(S)$.

(T2) If $S \subseteq T$ then $\mathrm{cl}(S) \subseteq \mathrm{cl}(T)$. In particular, the full set X is closed.

(T3) We have $\mathrm{cl}(S) = \mathrm{cl}(\mathrm{cl}(S))$.

(T4) An arbitrary intersection of closed sets is closed.

If X, Y are sets and if $* : 2^X \rightleftarrows 2^Y : *$ is a Galois connection, then we observe that the functions $** : 2^X \to 2^X$ and $** : 2^Y \to 2^Y$ are Galois closures.

Proof. (T1): This is just (a). (T2): This follows from two applications of (b): $S \subseteq T \Rightarrow T^* \subseteq S^* \Rightarrow S^{**} \subseteq T^{**}$. (T3): This follows from (c): $S^* = S^{***} \Rightarrow S^{**} = S^{****} = (S^{**})^{**}$. (T4): This is the conclusion of (f). ///

Nov 6: The Affine Zariski Topology

A Galois closure does not necessarily determine a topology in the modern sense. However, it will determine a topology if it satisfies the following additional properties:[32]

(T5) The empty set is closed.

(T6) A union of finitely many closed sets is closed.

[31] Kasimierz Kuratowski, doctoral thesis (1921).

[32] These become one property if we say that the empty set is the union of the empty collection of sets.

These extra properties hold, for example, in the case of the Zariski topology.

The Affine Zariski Topology (Strong Nullstellensatz). Let $\mathbb{F}$ be a field and consider the ring of polynomials $\mathbb{F}[\mathbf{x}] = \mathbb{F}[x_1, \ldots, x_n]$. Let the functions

$$\mathbf{I} : (\text{subsets of } \mathbb{F}^n) \rightleftarrows (\text{ideals of } \mathbb{F}[\mathbf{x}]) : \mathbf{V}$$

be defined as follows:

$$\begin{aligned} \mathbf{I}(S) &:= \{f \in \mathbb{F}[\mathbf{x}] : f(\mathbf{p}) = 0 \text{ for all } \mathbf{p} \in S\}, \\ \mathbf{V}(I) &:= \{\mathbf{p} \in \mathbb{F}^n : f(\mathbf{p}) = 0 \text{ for all} f \in I\}. \end{aligned}$$

To check that $\mathbf{I}(S) \subseteq \mathbb{F}[\mathbf{x}]$ is an ideal, let $f, g \in \mathbf{I}(S)$ and $h \in \mathbb{F}[\mathbf{x}]$. Then we have $(f+gh)(\mathbf{p}) = f(\mathbf{p}) + g(\mathbf{p})h(\mathbf{p}) = 0 + 0h(\mathbf{p}) = 0$ for all $\mathbf{p} \in S$, and hence $f + gh \in \mathbf{I}(S)$. The functions $\mathbf{I}, \mathbf{V}$ are a Galois connection with respect to inclusion, since for all $S \subseteq \mathbb{F}^n$ and $I \subseteq \mathbb{F}[\mathbf{x}]$ we have

$$S \subseteq \mathbf{V}(I) \quad \Leftrightarrow \quad \forall \mathbf{p} \in S, \forall f \in I, f(\mathbf{p}) = 0 \quad \Leftrightarrow \quad I \subseteq \mathbf{I}(S).$$

Note that the subsets of $\mathbb{F}^n$ form a lattice with $\vee = \cup$ and $\wedge = \cap$, and the ideals of $\mathbb{F}[\mathbf{x}]$ form a lattice with $\vee = +$ and $\wedge = \cap$. Thus from property (f) of Galois connections we have the following identities for any subsets $\{S_i\}$ and any ideals $\{I_i\}$:

$$\begin{array}{lll} \sum_i \mathbf{I}(S_i) \subseteq \mathbf{I}\left(\cap_i S_i\right) & \text{and} & \cup_i \mathbf{V}(I_i) \subseteq \mathbf{V}\left(\cap_i I_i\right), \\ \cap_i \mathbf{I}(S_i) = \mathbf{I}\left(\cup_i S_i\right) & \text{and} & \cap_i \mathbf{V}(I_i) = \mathbf{V}\left(\sum_i I_i\right). \end{array}$$

Furthermore, from property (e) of Galois connections we have an order-reversing bijection between the "closed' elements" on each side:

$$\mathbf{I} : (\text{closed subsets of } \mathbb{F}^n) \xleftrightarrow{\sim} (\text{closed ideals of } \mathbb{F}[\mathbf{x}]) : \mathbf{V}$$

The Galois closure $\mathbf{VI}$ on subsets of $\mathbb{F}^n$ is called the *Zariski closure* and subsets $S \subseteq \mathbb{F}^n$ satisfying $\mathbf{V}(\mathbf{I}(S)) = S$ are called *Zariski closed*. The Zariski closure automatically satisfies properties (T1)–(T4). I claim that it also satisfies the properties (T5),(T6) and hence it defines topology on $\mathbb{F}^n$, called the *Zariski topology*.

Proof. To show (T5) we observe that $\emptyset = \mathbf{V}(\mathbb{F}[\mathbf{x}])$, which is a closed set. To prove (T6) it is sufficient to show that the union of two closed sets is closed, then the general result follows by induction. Observe from property (d) that any two closed sets have the form $\mathbf{V}(I_1)$ and $\mathbf{V}(I_2)$ for some ideals $I_1, I_2 \subseteq \mathbb{F}[\mathbf{x}]$. Then I claim that that the union of these sets is given by

$$\mathbf{V}(I_1) \cup \mathbf{V}(I_1) = \mathbf{V}(I_1 \cap I_2),$$

which is a closed set. Indeed, we already know that $\mathbf{V}(I_1) \cup \mathbf{V}(I_2) \subseteq \mathbf{V}(I_1 \cap I_2)$. For the other direction, suppose that $\mathbf{p} \notin \mathbf{V}(I_1) \cup \mathbf{V}(I_1)$. Since $\mathbf{p} \notin \mathbf{V}(I_1)$ there exists $f_1 \in I_1$ such that $f_1(\mathbf{p}) \neq 0$, and since $\mathbf{p} \notin \mathbf{V}(I_2)$ there exists $f_2 \in I_2$ such that $f_2(\mathbf{p}) \neq 0$. But then the function $f_1 f_2 \in I_1 \cap I_2$ satisfies $(f_1 f_2)(\mathbf{p}) = f_1(\mathbf{p}) f_2(\mathbf{p}) \neq 0$ and hence $\mathbf{p} \notin \mathbf{V}(I_1 \cap I_2)$. $\square$

The Zariski-closed subsets of affine space $\mathbb{F}^n$ are also called *affine varieties*. We observe that a general variety $\mathbf{V}(I)$ is the intersection of infinitely many hypersurfaces: $\mathbf{V}(I) = \cap_{f \in I} V_f$. Furthermore, from the Hilbert Basis Theorem we observe that only **finitely many hypersurfaces** are necessary.

Proof. From (d) we know that every Zariski-closed set has the form $\mathbf{V}(I)$ for some ideal $I \subseteq \mathbb{F}[\mathbf{x}]$. Then from the HBT we know that $I = f_1\mathbb{F}[\mathbf{x}] + \cdots + f_m\mathbb{F}[\mathbf{x}]$ for some finite set of polynomials $f_1, \ldots, f_m \in \mathbb{F}[\mathbf{x}]$. It follows that

$$\mathbf{V}(I) = V_{f_1} \cap V_{f_2} \cap \cdots \cap V_{f_m}. \qquad \square$$

And what about the closed ideals? In order to get a clean description we must assume that the field $\mathbb{F}$ is **algebraically closed**. Then the **Strong Nullstellensatz** tells us that the Galois closure **IV** on ideals of $\mathbb{F}[\mathbf{x}]$ is the so-called *radical closure*:

$$\mathbf{I}(\mathbf{V}(I)) = \sqrt{I} := \{g \in \mathbb{F}[\mathbf{x}] : g^r \in I \text{ for some } r \geq 0\}.$$

Proof. One direction is easy: Suppose that $g \in \sqrt{I}$, so that $g^r \in I$ for some $r \geq 0$. Then for any point $\mathbf{p} \in \mathbf{V}(I)$ we have $g(\mathbf{p})^r = 0$ and hence $g(\mathbf{p}) = 0$. In other words, $g \in \mathbf{I}(\mathbf{V}(I))$. For the other direction, suppose that $g \in \mathbf{I}(\mathbf{V}(I))$, so that g vanishes on the set $\mathbf{V}(I)$. Then from the Strong Nullensatz we have $g^r \in I$ for some $r \geq 0$. In other words, $\mathbf{I}(\mathbf{V}(I)) \subseteq \sqrt{I}$. $\square$

Simple Corollary. The radical closure $\sqrt{I} = \mathbf{I}(\mathbf{V}(I))$ is also an ideal. [In fact, this holds for an arbitrary ring (use the binomial theorem).] ///

Ideals satisfying $I = \sqrt{I}$ are called *radical ideals*. Thus we obtain an order-reversing bijection:

$$\mathbf{I} : (\text{varieties in } \mathbb{F}^n) \xleftrightarrow{\sim} (\text{radical ideals of } \mathbb{F}[\mathbf{x}]) : \mathbf{V}$$

Recall that any hypersurface can be expressed uniquely as a union of irreducible hypersurfaces. The original proof used Study's Lemma and the fact that $\mathbb{F}[\mathbf{x}]$ is a UFD. To complete our description of the affine Zariski topology we will extend this unique decomposition to general varieties. There are three parts. The first part is purely combinatorial and the second part holds over any field. Only the third part uses the Nullstellensatz.

(1) Every variety V has a unique minimal decomposition into irreducible varieties:

$$V = V_1 \cup V_2 \cup \cdots \cup V_k.$$

The adjective *minimal* means that $V_i \not\subseteq V_j$ for all $i \neq j$, hence no V_i can be omitted.

(2) As with hypersurfaces, we say that a variety V is *reducible* if it can be expressed as a union of varieties $V = V_1 \cup V_2$ with $V_1, V_2 \neq V$. Otherwise it is *irreducible*. I claim that the maps $\mathbf{I}, \mathbf{V}$ restrict to a bijection between **irreducible** varieties and **prime** ideals:[33]

$$\mathbf{I} : (\text{irreducible varieties in } \mathbb{F}^n) \xleftrightarrow{\sim} (\text{prime ideals in } \mathbb{F}[\mathbf{x}]) : \mathbf{V}.$$

[33] Observe that every prime ideal $P \subseteq \mathbb{F}[\mathbf{x}]$ is radical, since $g \notin P$ implies $g^r \notin P$ for all $g \in \mathbb{F}[\mathbf{x}]$ and $r \geq 1$.

(3) It follows from (1) and (2) that every radical ideal $I \subseteq \mathbb{F}[\mathbf{x}]$ has a unique expression as a minimal intersection of prime ideals:

$$I = P_1 \cap P_2 \cap \cdots \cap P_k.$$

The adjective *minimal* means that $P_i \not\subseteq P_j$ for all $i \neq j$, hence no P_i can be omitted.

Proof. (1): *Existence.* If V cannot be expressed as a union of irreducibles then we obtain an infinite descending chain of varieties, which maps to an infinite ascending chain of ideals in $\mathbb{F}[\mathbf{x}]$. This contradicts the Hilbert Basis Theorem. *Uniqueness.* Suppose that

$$V_1 \cup V_2 \cup \cdots \cup V_k = V_1' \cup V_2' \cup \cdots \cup V_\ell', \qquad (*)$$

where $V_i \not\subseteq V_j$ and $V_i' \not\subseteq V_j'$ for all $i \neq j$. Now observe that

$$\begin{aligned} V_i &\subseteq (V_1' \cup V_2' \cup \cdots \cup V_\ell') \\ V_i &= V_1 \cap (V_1' \cup V_2' \cup \cdots \cup V_\ell') \\ V_i &= (V_i \cap V_1') \cup (V_i \cap V_2') \cup \cdots \cup (V_i \cap V_\ell'). \end{aligned}$$

Since V_i is irreducible this implies that $V_i \cup V_j' = V_i$ for some j, and hence $V_j' \subseteq V_i$. A symmetric argument shows that $V_k \subseteq V_j'$ for some k, and hence

$$V_k \subseteq V_j' \subseteq V_i.$$

Finally, since $V_k \not\subseteq V_i$ for all $k \neq i$ we must have $k = i$ and hence $V_i = V_j'$. We have shown that each summand on the left side of $(*)$ is equal to a summand on the right, and vice versa. Since no two summands on the same side are equal, this implies uniqueness.

[Remark: The same result holds for any lattice in which descending chains stabilize and $\wedge$ distributes over $\vee$.]

(2): **(I prime $\Rightarrow$ V irreducible).** Let $V = V_1 \cup V_2$ be a proper decomposition and define the ideals $I_1 = \mathbf{I}(V_1)$ and $I_2 = \mathbf{I}(V_2)$. We will show that there exist $f_1, f_2 \in \mathbb{F}[\mathbf{x}] \setminus I$ such that $f_1 f_2 \in I$, and hence I is not prime. To do this, we first observe that that $V_1, V_2 \neq V$ implies $I_1, I_2 \neq I$, since the map $\mathbf{I}$ is injective on closed ideals.[34] Thus we may choose polynomials $f_1 \in I \setminus I_1$ and $f_2 \in I \setminus I_2$. Then for all $\mathbf{p} \in V = V_1 \cup V_2$ we must have $\mathbf{p} \in V_1$, and hence $f_1(\mathbf{p}) = 0$, or $\mathbf{p} \in V_2$, and hence $f_2(\mathbf{p}) = 0$. In either case we have $(f_1 f_2)(\mathbf{p}) = f_1(\mathbf{p}) f_2(\mathbf{p}) = 0$, and it follows that $f_1 f_2 \in I = \mathbf{I}(V)$.

(V irreducible $\Rightarrow$ I prime). Suppose that I is not prime, so there exist $f_1, f_2 \notin I$ with $f_1 f_2 \in I$. We will show that there exist varieties $V_1, V_2 \neq V$ such that $V = V_1 \cup V_2$. Indeed, I claim that $V_1 := \mathbf{V}(I + f_1\mathbb{F}[\mathbf{x}])$ and $V_2 := \mathbf{V}(I + f_2\mathbb{F}[\mathbf{x}])$ will work. To see this, we first observe that $I \subseteq I + f_1\mathbb{F}[\mathbf{x}]$ and $I \subseteq I + f_2\mathbb{F}[\mathbf{x}]$ imply $V_1 \subseteq V$ and $V_2 \subseteq V$,[35] hence $V_1 \cup V_2 \subseteq V$. To show that $V_1, V_2 \neq V$ we use the fact that $f_1, f_2 \notin I$. This implies that there exist $\mathbf{p}_1, \mathbf{p}_2 \in V$

[34] This is just a property of Galois connections. We do not need to know that the closed ideals are radical.

[35] Here we use the fact that $\mathbf{V}$ is order-reversing.

with $f_1(\mathbf{p}_1) \neq 0$ and $f_2(\mathbf{p}_2) = 0$, hence $\mathbf{p}_1 \notin V_1$ and $\mathbf{p}_2 \notin V_2$. To show that $V \subseteq V_1 \cup V_2$ we use the fact that $f_1 f_2 \in I$. This implies that for all $\mathbf{p} \in V$ we have $f_1(\mathbf{p}) f_2(\mathbf{p}) = (f_1 f_2)(\mathbf{p}) = 0$, which implies that $f_1(\mathbf{p}) = 0$ or $f_2(\mathbf{p}) = 0$. If $f_1(\mathbf{p}) = 0$ then for all $\varphi = g + f_1 h \in I + f_1 \mathbb{F}[\mathbf{x}]$ we have $\varphi(\mathbf{p}) = 0$, hence $\mathbf{p} \in V_1$. And a similar argument shows that $f_2(\mathbf{p}) = 0$ implies $\mathbf{p} \in V_2$. Thus for any $\mathbf{p} \in V$ we have $\mathbf{p} \in V_1$ or $\mathbf{p} \in V_2$ as desired.

(3): *Existence.* Let $I \subseteq \mathbb{F}[\mathbf{x}]$ be a radical ideal with corresponding variety $V = \mathbf{V}(I)$. From part (1) we can write $V = V_1 \cup \cdots \cup V_k$ where the V_i are irreducible and $V_i \subsetneq V_j$ for all $V_i \neq V_j$. Let $P_i := \mathbf{I}(V_i)$ be the corresponding ideals, which are prime by part (2). Now we invoke the Strong Nullstellensatz to observe that $I = \sqrt{I} = \mathbf{I}(\mathbf{V}(I)) = \mathbf{I}(V)$, and it follows from property (f) of Galois connections that

$$\begin{aligned} V &= V_1 \cup V_2 \cup \cdots \cup V_k \\ \mathbf{I}(V) &= \mathbf{I}(V_1) \cap \mathbf{I}(V_2) \cap \cdots \cap \mathbf{I}(V_k) \\ I &= P_1 \cap P_2 \cap \cdots \cap P_k. \end{aligned}$$

This intersection is minimal since if $P_i \subseteq P_j$ for some $i \neq j$ then by applying $\mathbf{V}$ we obtain the contradiction $V_j = \mathbf{V}(P_j) \subseteq \mathbf{V}(P_i) = V_i$.

Uniqueness. Suppose that $P_1 \cap \cdots \cap P_k = P'_1 \cap \cdots \cap P'_\ell$ for some prime ideals with $P_i \not\subseteq P_j$ and $P'_i \not\subseteq P'_j$ for all $i \neq j$. Applying the map $\mathbf{V}$ to both sides and using the fact that $\mathbf{V}(I_1 \cap I_2) = \mathbf{V}(I_1) \cup \mathbf{V}(I_2)$ (proved above) shows that

$$\mathbf{V}(P_1) \cup \cdots \cup \mathbf{V}(P_k) = \mathbf{V}(P'_1) \cup \cdots \cup \mathbf{V}(P'_\ell).$$

Assume for contradiction that we have $\mathbf{V}(P_i) \subseteq \mathbf{V}(P_j)$ for some $i \neq j$. Then since prime ideals are radical we can apply $\mathbf{I}$ and use the Nullstellensatz to obtain the contradiction $P_j = \mathbf{I}(\mathbf{V}(P_j)) \subseteq \mathbf{I}(\mathbf{V}(P_i)) = P_i$. Similarly we have $\mathbf{V}(P'_i) \not\subseteq \mathbf{V}(P'_j)$ for all $i \neq j$. Thus it follows from part (1) that the sets $\{\mathbf{V}(P_i)\}$ and $\{\mathbf{V}(P'_i)\}$ are equal, and then applying $\mathbf{I}$ shows that the sets $\{P_i\}$ and $\{P'_i\}$ are equal, as desired. □

Corollary (Classification of subvarieties of $\mathbb{F}^2$). If $\mathbb{F}$ is algebraically closed then we have seen (on Oct 28) that the non-zero prime ideals of $\mathbb{F}[x, y]$ are the just the principal prime ideals $f\mathbb{F}[x, y]$ and the maximal ideals $M_{a,b}$. It follows that any subvariety of the affine plane $\mathbb{F}^2$ is a finite union of points and irreducible curves. ///

Remark: As with linear subspaces and hyperplanes, the expression of a variety as an intersection of hypersurfaces is not unique. But the situation is even more complicated than the linear case might suggest. That is, it might be the case that a "d-dimensional" subvariety of $\mathbb{F}^n$ or $\mathbb{FP}^n$ cannot be expressed as an intersection of $n - d$ hypersurfaces. Furthermore, the minimal number of hypersurfaces required to represent a given variety might not coincide with the minimal number of polynomials needed to generate the corresponding ideal. Dimension theory is quite subtle.

Nov 11,13: The Projective Zariski Topology

Based on the example of Study's Lemma, we expect that the Zariski topology on affine space $\mathbb{F}^n$ should extend in a straightforward way to projective space $\mathbb{FP}^n$. In the case of hypersurfaces this was done via the homogenization of polynomials. For general varieties we need some sort of "homogenization of ideals."

To warm up, we try to find the ideal of a point in homogeneous space. Recall that a point $\mathbf{p} = (p_1 : p_2 : \cdots : p_{n+1}) \in \mathbb{FP}^n$ corresponds to the line $L = t(p_1, \ldots, p_{n+1})$ in $\mathbb{F}^{n+1}$, which can be expressed (non-uniquely) as an intersection of n linear hyperplanes:

$$L = H_{x_1p_2 - x_2p_1} \cap H_{x_2p_3 - x_3p_2} \cap \cdots \cap H_{x_np_{n+1} - x_{n+1}p_n}.$$

Since each of the polynomials $x_ip_{i+1} - x_{i+1}p_i$ is homogeneous, we can also view $\mathbf{p}$ as the intersection of the corresponding projective hyperplanes. Thus we define the ideal of the projective point $\mathbf{p} \in \mathbb{FP}^n$ to be the ideal of the corresponding line $L \subseteq \mathbb{F}^{n+1}$:

$$\mathbf{I}(\mathbf{p}) := \mathbf{I}(L) = \sum_{i=1}^{n} (x_ip_{i+1} - x_{i+1}p_i)\mathbb{F}[\mathbf{x}].$$

Note that we can also express this ideal in a more symmetric form:

$$\mathbf{I}(\mathbf{p}) = \sum_{i<j} (x_ip_j - x_jp_i)\mathbb{F}[\mathbf{x}].$$

This ideal has the important property that it is **generated by homogeneous polynomials**.

More generally, we say that $V \subseteq \mathbb{FP}^n$ is a *projective variety* if it equal to the intersection of (finitely many) projective hypersurfaces:

$$V = V_{F_1} \cap V_{F_2} \cup \cdots \cap V_{F_m},$$

where $F_1, \ldots, F_m \in \mathbb{F}[\mathbf{x}]$ are homogeneous polynomials. In order to define the ideal of V we consider the affine cone $\mathrm{Cone}(V) \subseteq \mathbb{F}^{n+1}$ defined as follows:

$$\mathrm{Cone}(V) := \{(p_1, \ldots, p_{n+1}) \text{ such that } (p_1 : \cdots : p_{n+1}) \in V\} \cup \{\mathbf{0}\}.$$

Then we define the *ideal of the projective set* $\mathbf{I}(V) \subseteq \mathbb{F}[x_1, \ldots, x_{n+1}]$ as the ideal of the cone. Note that this ideal is also **generated by homogeneous polynomials**:

$$\mathbf{I}(V) := \mathbf{I}(\mathrm{Cone}(V)) = F_1\mathbb{F}[\mathbf{x}] + F_2\mathbb{F}[\mathbf{x}] + \cdots + F_m\mathbb{F}[\mathbf{x}].$$

Since not every ideal has this form, we make the following definition.

Homogeneous Ideals. Let $I \subseteq \mathbb{F}[\mathbf{x}]$ be an ideal. The following are equivalent:

(1) If $f \in I$ then $f^{(k)} \in I$ for all k.

(2) I is generated by (finitely many) homogeneous polynomials.

When these properties hold we say that $I \subseteq \mathbb{F}[\mathbf{x}]$ is a *homogeneous ideal*.

Proof. (1)⇒(2): From the Hilbert Basis Theorem we know that $I = f_1\mathbb{F}[\mathbf{x}] + \cdots + f_m\mathbb{F}[\mathbf{x}]$ for some (possibly non-homogeneous) polynomials $f_1, \ldots, f_m \in \mathbb{F}[\mathbf{x}]$. From (1) we know that each of the (finitely many) homogeneous parts $f_i^{(k)}$ is in I, so that

$$\sum_{i,k} f_i^{(k)}\mathbb{F}[\mathbf{x}] \subseteq I.$$

Conversely, every element $f = \sum_i f_i g_i \in I$ has the form $f = \sum_{i,k} f_i^{(k)} g_i$, so that

$$I \subseteq \sum_{i,k} f_i^{(k)}\mathbb{F}[\mathbf{x}].$$

(2)⇒(1): Suppose that $I = F_1\mathbb{F}[\mathbf{x}] + \cdots + F_m[\mathbf{x}]$ for some homogeneous polynomials with $\deg(F_i) = d_i \geq 0$, and consider some (possibly non-homogeneous) polynomial $f \in I$. By hypothesis we can write $f = \sum F_i g_i$ for some (possibly non-homogeneous) polynomials $g_i \in \mathbb{F}[\mathbf{x}]$. By considering homogeneous parts of the g_i we have

$$f = \sum_i F_i \sum_\ell g_i^{(\ell)} = \sum_{i,\ell} F_i g_i^{(\ell)}.$$

Finally, since the polynomial $F_i g_i^{(\ell)}$ is homogeneous of degree $d_i + \ell$ we have

$$f^{(k)} = \sum_{i,\ell:\, d_i+\ell=k} F_i g_i^{(\ell)} \in I.$$

[Note that this sum may be empty, in which case $f^{(k)} = 0 \in I$.] □

Corollary. It follows from property (1) that intersections of homogeneous ideals are homogeneous. It follows from property (2) that sums of homogeneous ideals are homogeneous. Thus the collection of homogeneous form a lattice with least upper bound $\vee = +$ and greatest lower bound $\wedge = \cap$. ///

We have seen that the ideal of a projective set is homogeneous. Conversely, we will show that the variety of a homogeneous ideal in $\mathbb{F}[x_1, \ldots, x_{n+1}]$ corresponds to some projective variety in $\mathbb{FP}^n$, i.e., an intersection of finitely many hypersurfaces in $\mathbb{FP}^n$. For this purpose, it is convenient to define the notion of a conical set.

Projective vs Conical Sets. We say that an affine set $C \subseteq \mathbb{F}^{n+1}$ is *conical* when it is closed under scalar multiplication:

$$\mathbf{p} \in C \quad \Rightarrow \quad \lambda\mathbf{p} \in C \text{ for all } \lambda \in \mathbb{F}.$$

Note that any conical set $C \subseteq \mathbb{F}^{n+1}$ determines a subset $C/(\text{scalars})$ of projective space $\mathbb{FP}^n = \mathbb{F}^{n+1}/(\text{scalars})$. Conversely, any subset $S \subseteq \mathbb{FP}^n$ of projective space determines a conical set, called its *affine cone*:

$$\text{Cone}(S) := \{(p_1, \ldots, p_{n+1}) \text{ such that } (p_1 : \cdots : p_{n+1}) \in S\} \cup \{\mathbf{0}\},$$

Note that the cone of the empty projective set is non-empty: $\text{Cone}(\emptyset) = \{\mathbf{0}\}$. Thus we obtain a bijection between projective sets and non-empty conical affine sets:

$$\text{Cone} : (\text{subsets of } \mathbb{FP}^n) \leftrightarrow (\text{non-empty conical subsets of } \mathbb{F}^{n+1}) : (-)/(\text{scalars}).$$

The empty set $\emptyset \subseteq \mathbb{F}^{n+1}$ is conical, but it does not correspond to any projective set.

Now we can state and prove the basic properties of the projective Zariski topology.

The Projective Zariski Topology (Projective Nullstellensatz). Let $\mathbb{F}$ be any infinite field and let $\mathbf{x} = (x_1, \ldots, x_{n+1})$. By restricting the function $\mathbf{I}$ to conical sets, I claim that we obtain a Galois connection between non-empty conical sets and non-unit homogeneous ideals:

$$\mathbf{I} : (\text{non-empty conical subsets of } \mathbb{F}^{n+1}) \rightleftarrows (\text{non-unit homogeneous ideals } \mathbb{F}[\mathbf{x}]) : \mathbf{V}$$

Proof. We only need to show that (1) $\mathbf{I}$ sends conical sets to homogeneous ideals, and (2) that $\mathbf{V}$ sends homogeneous ideals to conical sets. (1): Let S be a conical set and let $f \in \mathbf{I}(S)$. If $\mathbf{p} \in S$ then for all $\lambda \in \mathbb{F}$ we have $\lambda\mathbf{p} \in S$ and hence $f(\lambda\mathbf{p}) = 0$. Now let $f = \sum_k f^{(k)}$ be the homogeneous filtration of f, so that

$$0 = f(\lambda\mathbf{p}) = \sum_k f^{(k)}(\lambda\mathbf{p}) = \sum_k \lambda^k f^{(k)}(\mathbf{p}).$$

Let y be another variable and consider the polynomial $g(y) = \sum_k y^k f^{(k)}(\mathbf{p}) \in \mathbb{F}[y]$. Since this polynomial has infinitely many roots $\lambda \in \mathbb{F}$ we conclude that it must be the zero polynomial. That is, we must have $f^{(k)}(\mathbf{p}) = 0$. Finally, since this holds for any $\mathbf{p} \in S$ we conclude that $f^{(k)} \in \mathbf{I}(S)$ for all k, hence $\mathbf{I}(S)$ is a homogeneous ideal. (2): Let $I \neq \mathbb{F}[\mathbf{x}]$ be a homogeneous ideal, so that $I = \sum F_i\mathbb{F}[\mathbf{x}]$ for some homogeneous polynomials $F_1, \ldots, F_m \in \mathbb{F}[\mathbf{x}]$. To show that $\mathbf{V}(I)$ is a conical set, consider any point $\mathbf{p} \in \mathbf{V}(I)$. Since $F_i \in I$ we have $F_i(\mathbf{p}) = 0$ and since F_i is homogeneous, we have for all $\lambda \in \mathbb{F}$ that

$$F_i(\lambda\mathbf{p}) = \lambda^{\deg F_i} F_i(\mathbf{p}) = 0.$$

Finally, for any $f = \sum F_i g_i \in I$ we have

$$f(\lambda\mathbf{p}) = F_1(\lambda\mathbf{p})g_1(\lambda\mathbf{p}) + \cdots + F_m(\lambda\mathbf{p})g_m(\lambda\mathbf{p}) = 0g_1(\lambda\mathbf{p}) + \cdots + 0g_m(\lambda\mathbf{p}) = 0,$$

and hence $\lambda\mathbf{p} \in \mathbf{V}(I)$. □

By identifying subsets of $\mathbb{FP}^n$ with non-empty conical subsets of $\mathbb{F}^{n+1}$ we can define a map $\mathbf{VI}$ on projective sets, called the *projective Zariski closure*. Since every homogeneous ideal is generated by finitely many homogeneous polynomials, we conclude that Zariski closed projective sets are the same as intersections of (finitely many) projective hypersurfaces. In other words:

$$(\text{Zariski-closed subsets of } \mathbb{FP}^n) = (\text{projective varieties in } \mathbb{FP}^n).$$

Now the following properties are inherited from the affine Galois connection $\mathbf{I}, \mathbf{V}$:

- An arbitrary intersection of projective varieties is a projective variety.
- A finite union of projective varieties is a projective variety.
- Each projective variety has a unique minimal expression as a union of finitely many irreducible projective varieties.

Thus we obtain a *projective Zariski topology* on $\mathbb{FP}^n$. The main difference with the affine Zariski topology is that the empty set $\emptyset \subseteq \mathbb{FP}^n$ does not correspond to the unit ideal $\mathbb{F}[\mathbf{x}]$. Instead, it corresponds to the maximal ideal of origin:

$$\mathbf{I}(\emptyset \subseteq \mathbb{FP}^n) = \mathbf{I}(\{\mathbf{0}\} \subseteq \mathbb{F}^{n+1}) = \{f \in \mathbb{F}[\mathbf{x}] : f(\mathbf{0}) = 0\} = M_{\mathbf{0}}.$$

For this reason, $M_{\mathbf{0}} = x_1\mathbb{F}[\mathbf{x}] + \cdots + x_{n+1}\mathbb{F}[\mathbf{x}]$ is called the *irrelevant ideal*. Furthermore, we already know that $M_{\mathbf{0}} \subsetneq \mathbb{F}[\mathbf{x}]$ is a maximal ideal. I claim that it is **the unique maximal homogeneous ideal**.

Proof. Let $M \subsetneq \mathbb{F}[\mathbf{x}]$ be maximal among homogeneous ideals. Since $\mathbf{I}(\mathbf{V}(M)) \supseteq M$ is also homogeneous, we must have $\mathbf{I}(\mathbf{V}(M)) = M$. Now I claim that $\mathbf{V}(M)$ is minimal among projective varieties. To see this, let $V \subseteq \mathbf{V}(M)$ be a projective variety. By definition we have $V = \mathbf{V}(\mathbf{I}(V))$. Applying $\mathbf{I}$ to the contaiment $V \subseteq \mathbf{V}(M)$ gives

$$M = \mathbf{I}(\mathbf{V}(M)) \subseteq \mathbf{I}(V),$$

which implies that $M = \mathbf{I}(V)$ because $\mathbf{I}(V)$ is homogeneous. Then applying $\mathbf{V}$ gives $\mathbf{V}(M) = \mathbf{V}(\mathbf{I}(V)) = V$. Since $\emptyset$ is the unique minimal projective variety, it follows that $\mathbf{V}(M) = \emptyset$, and hence $M = \mathbf{I}(\mathbf{V}(M)) = \mathbf{I}(\emptyset) = M_{\mathbf{0}}$. □

Now let us suppose that $\mathbb{F}$ is algebraically closed. Then the following properties are inherited from the affine Zariski topology:

- The radical closure of a homogeneous ideal is homogeneous.[36]
- The maps $\mathbf{I}, \mathbf{V}$ define order-reversing bijections:

$$(\text{projective varieties}) \overset{\sim}{\longleftrightarrow} (\text{radical homogeneous ideals}),$$
$$(\text{irreducible projective varieties}) \overset{\sim}{\longleftrightarrow} (\text{prime homogeneous ideals}).$$

[36] Actually, this holds over any field.

- Each radical homogeneous ideal has a unique minimal expression as an intersection of finitely many prime homogeneous ideals.

It follows that every radical (and, in particular, prime) homogeneous ideal is contained in the irrelevant ideal $M_{\mathbf{0}}$. Homogeneous ideals whose radical closure is strictly contained in $M_{\mathbf{0}}$ are called *relevant*, since they correspond to non-empty subsets of projective space.

Examples:

Minimal Prime Homogeneous Ideals (Study's Lemma). Since $\mathbb{F}[\mathbf{x}]$ is a UFD, we already know that the minimal non-empty prime ideals are precisely the principal ideals $f\mathbb{F}[\mathbf{x}]$ with f irreducible. I claim that the minimal prime homogeneous ideals are precisely the principal ideals $F\mathbb{F}[\mathbf{x}]$ with F homogeneous and irreducible. In other words, they correspond to irreducible projective hypersurfaces.[37]

Proof. Let $P = F\mathbb{F}[\mathbf{x}]$ for some irreducible (non-zero, non-unit) homogeneous polynomial F. Then we already know that P is minimal among all prime ideals, hence minimal among homogeneous prime ideals. Conversely, let $P \supsetneq 0$ be minimal among prime homogeneous ideals. Choose some non-zero, non-unit homogeneous polynomial $F \in P$, with prime factorization $F = F_1 \cdots F_k$. Since F is homogeneous we know that its factors are homogeneous, and since P is prime we know that $F_i \in P$ for some i, so that $F_i\mathbb{F}[\mathbf{x}] \subseteq P$. Finally, since $F_i\mathbb{F}[\mathbf{x}]$ is a prime homogeneous ideal we must have $P = F_i\mathbb{F}[\mathbf{x}]$. □

Maximal Prime Homogeneous Ideals (Weak Nullstellensatz). Recall that we have an order-reversing bijection between irreducible projective varieties and prime homogeneous ideals. Since the points of $\mathbb{FP}^n$ are the minimal non-empty (irreducible) varieties, it follows that the each maximal prime homogeneous ideal $P \subsetneq M_{\mathbf{0}}$ has the form

$$P = \sum_{i<j} (x_i p_j - x_j p_i)\mathbb{F}[\mathbf{x}]$$

for some projective point $\mathbf{p} = (p_1 : \cdots : p_{n+1}) \in \mathbb{FP}^n$.

Nov 13: Homogenization and Dehomogenization

Look at this: `https://www3.risc.jku.at/education/courses/ss2017/caag/05-proj.pdf`

We have some maps **I**: sets to ideals, conical sets to homogeneous ideals. **V**: ideals to affine varieties, homogeneous ideals to conical varieties. **C**: affine sets to conical sets (cone over zero). Don't even bother with projective space. But we will deal with two rings: $\mathbb{F}[\mathbf{x}'] \subseteq \mathbb{F}[\mathbf{x}]$. Given $I' \subseteq \mathbb{F}[\mathbf{x}']$ we have the variety $\mathbf{V}'(I') := \mathbf{V}(I') \cap H$ and given $V' \subseteq H$ we have $\mathbf{I}'(V') =$

To complete our discussion of the Zariski topology, we must determine the relationship between affine and projective varieties. Let $\mathbb{F}$ be an algebraically closed field.

[37] This result holds over any field.

On the geometric side, let $V \subseteq U = \mathbb{F}^n \subseteq \mathbb{FP}^n$ be a variety in an affine chart of projective space. In general, the set $V \subseteq \mathbb{FP}^n$ is **not** a projective variety, therefore we let $\overline{V} \subseteq \mathbb{FP}^n$ denote the projective Zariski closure, and we call this the *projective completion* of V. Conversely, given a projective variety $V \subseteq \mathbb{FP}^n$ not contained in the hyperplane $H = \mathbb{FP}^n \setminus \mathbb{F}^n = H_{n+1}$, I claim that the intersection $V \cap U$ is an affine variety in U. Indeed, we already know from the projective Study's Lemma that this holds for hypersurfaces, hence it must also hold for intersections of hypersurfaces. We will see that the two operations $V \mapsto \overline{V}$ and $V \mapsto V \cap U$ are inverses.

Cone over an affine variety. Given $V \subseteq \mathbb{F}^{n+1}$, we define $\text{Cone}(V) = \{\lambda \mathbf{p} : \mathbf{p} \in V, \lambda \in \mathbb{F}\} \subseteq \mathbb{F}^{n+1}$. This is the smallest conical set containing V. If $V = \mathbf{V}(I)$ is a variety then I claim that

$$\mathbf{I}(\text{Cone}(V)) = I' := (\text{ideal generated by homogeneous elements of } I) \subseteq I.$$

Furthermore, I claim that I' is the largest homogeneous ideal contained in I', hence I' is radical, and it follows that the Zariski closure $\mathbf{VI}(\text{Cone}(V)) = \mathbf{V}(I')$ is the smallest conical variety containing V.

[Remark: The set $\text{Cone}(V)$ need not be a variety. For example, the plane minus a line through the origin, plus the origin, is not a variety.

Proof 1: The closure of a basic open set $U : (f(\mathbf{x}) \neq 0)$ is everything. Indeed, suppose that $g \in \mathbf{I}(U)$ so that $f(\mathbf{p}) \neq 0$ implies $g(\mathbf{p}) = 0$. This implies that $f(\mathbf{p})g(\mathbf{p}) = 0$ for all $\mathbf{p}$, hence $f(\mathbf{x})g(\mathbf{x}) = 0$. Since $f(\mathbf{x}) \neq 0$ this implies that $g(\mathbf{x}) = 0$. It follows that $\mathbf{I}(U) = 0$ and hence $\mathbf{VI}(U) = \mathbb{F}^n$.

Proof 2: We know that the whole space is irreducible. Let V be the closure of $U : (f(\mathbf{x}) \neq 0)$ and note that $\mathbb{F}^n = V \cup V_f$. Since $\mathbb{F}^n$ is irreducible this implies that $\mathbb{F}^n = V$ or $\mathbb{F}^n = V_f$. Since the second is false, the first must be true. In fact, this proof shows that any non-empty open set is dense.

]

Proof. First we show that $\mathbf{I}(\text{Cone}(V)) = I'$. Indeed, if $f \in \mathbf{I}(\text{Cone}(V))$ then since $\text{Cone}(V)$ is a conical set we have $f^{(k)} \in \mathbf{I}(\text{Cone}(V))$ and hence $f = \sum f^{(k)} \in I'$. Conversely, if $f \in I'$ then we can write $f = \sum F_i g_i$ for some homogeneous F_i. If $\mathbf{p}$ then we want to show that $f(\lambda \mathbf{p}) = 0$ for all λ, so that $f \in \mathbf{I}(\text{Cone}(V))$. Well, since $F_i \in I$ and $V = \mathbf{V}(I)$ we have $F_i(\mathbf{p}) = 0$, and since F_i is homogeneous this implies that $F_i(\lambda \mathbf{p}) = 0$ for all λ. Finally, we have

$$f(\lambda \mathbf{p}) = \sum F_i(\lambda \mathbf{p}) g_i(\lambda \mathbf{p}) = 0.$$

Next we show that $I' \subseteq I$ is the largest homogeneous ideal contained in I. Indeed, suppose $J \subseteq I$ is homogeneous, so that $J = F_1 \mathbb{F}[\mathbf{x}] + \cdots + F_m \mathbb{F}[\mathbf{x}]$ for some homogeneous F_i. But then we have $F_i \in I'$ and hence $J \subseteq I'$ as desired. [Corollary: I' is radical.]

To finish the proof, let $W \supseteq V = \mathbf{V}(I)$ be a conical variety containing V. Applying $\mathbf{I}$ gives $\mathbf{I}(W) \subseteq \mathbf{IV}(I) = I$. Since $\mathbf{I}(W)$ is homogeneous we conclude that $\mathbf{I}(W) \subseteq I'$ and hence $\mathbf{V}(I') \subseteq \mathbf{VI}(W) = W$. $\square$

Projective Closure of an Affine Variety. Let $\mathbf{x}' = (x_1, \ldots, x_n)$ and $\mathbf{x} = (x_1, \ldots, x_{n+1})$. Let $V = \mathbf{V}(I) \subseteq \mathbb{F}^n$ be an affine variety with $I \subseteq \mathbb{F}[\mathbf{x}']$. Let $\overline{V} = \mathbf{VI}(\text{Cone}(V)) \subseteq \mathbb{F}^{n+1}$ be the smallest conical variety in $\mathbb{F}^{n+1}$ containing the set V. We call this the *projective closure*. Claim:

Then the geometric operations above define maps between radical ideals of the ring $\mathbb{F}[\mathbf{x}'_i]$ and radical homogeneous ideals of the ring $\mathbb{F}[\mathbf{x}]$. We will determine the exact nature of these maps.

Let $I \subseteq \mathbb{F}[\mathbf{x}']$ be the ideal of $V \subseteq U$, so that $I + (x_{n+1} - 1)$ is the ideal of $V \subseteq \mathbb{F}^{n+1}$. (Indeed, recall that $\cap_i \mathbf{V}(I_i) = \mathbf{V}(\sum_i I_i)$. We want to describe the homogeneous ideal $I^* := \mathbf{I}(\text{Cone}(V)) \subseteq \mathbb{F}[\mathbf{x}]$. From above we knows that this is the sub-ideal of $I + (x_{n+1} - 1)$ generated by the homogeneous elements.

Claim: I^* is generated by the homogenizations of elements of I. One direction: If $f = \sum (f_i)^* g_i$ for some $f_i \in I$, then each $(f_i)^*$ vanishes on $\text{Cone}(V)$, hence f vanishes on $\text{Cone}(V)$. Conversely, let F be a homogeneous generator of $\mathbf{I}(\text{Cone}(V))$, so F vanishes on $\text{Cone}(V)$. Then F_* vanishes on V, hence $F_* \in I$. We can assume that there exists a generator that is not divisible by x_{n+1}, so $x_{n+1} \nmid F$. Hence $F = (F_*)^*$ and so $I^* = \mathbf{I}(\text{Cone}(V))$ is generated by homogenizations of elements of I. [Warning: If $f_1, \ldots, f_m$ generate I then we cannot assume that $(f_1)^*, \ldots, (f_m)^*$ generate I^*.

Theorem $\mathbf{VI}(\text{Cone}(V)) = \overline{V}$. Wait, this is just the definition.

Homogenization and Dehomogenization of Ideals. Given a homogeneous ideal $I \subseteq \mathbb{F}[\mathbf{x}]$ we define the dehomogenization $I_* \subseteq \mathbb{F}[\mathbf{x}']$ by substituting $x_{n+1} = 1$ in every element of the ideal:

$$I_* := \{f_* : f \in I\}.$$

To see that this is an ideal, consider any elements $f_*, g_* \in I_*$ and $h \in \mathbb{F}[\mathbf{x}']$. By thinking of h as an element of $\mathbb{F}[\mathbf{x}] \supseteq \mathbb{F}[\mathbf{x}']$ we trivially have $h_* = h$. Then since evaluation is a ring homomorphism we conclude that $f_* + g_* h = f_* + g_* h_* = (f + gh)_* \in I_*$. Furthermore, if $I = \sum F_k \mathbb{F}[\mathbf{x}]$ then I claim that $I_* = \sum (F_k)_* \mathbb{F}[\mathbf{x}'_i]$. Indeed, since $(F_k)_* \in I_*$ we have $\sum (F_k)_* \mathbb{F}[\mathbf{x}'] \subseteq I_*$. Conversely, for any $\sum F_k g_k \in I$ we have $(\sum_k F_k g_k)_* = \sum (F_k)_* (g_k)_* \in I_*$.

Next we define the ith homogenization of an ideal $I \subseteq \mathbb{F}[\mathbf{x}']$:

$$I^* := \text{ ideal generated by } \{f^* : f \in I\} \subseteq \mathbb{F}[\mathbf{x}].$$

This ideal is homogeneous because it is generated by homogeneous polynomials (see the proof above), hence it can also be generated by finitely many homogeneous polynomials.

I claim that this is the largest homogeneous ideal of $\mathbb{F}[\mathbf{x}]$ contained in I. Indeed, To see that $I^* \subseteq I$, consider any $f \in I^*$, so that $f^{(k)} \in I^*$ for all k. Thus $f^{(k)} = \sum g^* h$ for some $g \in I$. Now consider any homogeneous ideal $J \subseteq \mathbb{F}[\mathbf{x}]$ such that $J \subseteq I$. Note that J is genreated by finitely many homogeneous elements $F_i \in I$, then since $(F_i)^* = F_i$ this implies that $J \subseteq I^*$ as desired. ///

More generally, any variety defined parametrically by polynomials is irreducible.

Theorem.

(1) Given a variety $V = \mathbf{V}(I) \subseteq U$ for some $I \subseteq \mathbb{F}[\mathbf{x}']$ we have $\overline{V} = \mathbf{V}(I^*)$. [Remark: I should want to define $\overline{V} = \mathrm{Cone}(V)$. But for this I need to check that $\mathrm{Cone}(V)$ is a variety.]

(2) Given a variety $V = \mathbf{V}(I) \subseteq \mathbb{FP}^n$ we have $V \cap U = \mathbf{V}(I_*)$.

Proof. (2): Cox-Little-O'Shea page 372.

(1): Note that $\mathbf{V}(I^*)$ contains V. Indeed, if $\mathbf{p} \in V$ then for every

To see this, recall that $\mathbf{V}$ is the smallest projective variety containing V. Since $I^* \subseteq I$ we know that $V = \mathbf{V}(I) \subseteq \mathbf{V}(^*)$, so $\mathbf{V}(I^*)$ is a projective variety containing V, hence $\overline{V} \subseteq \mathbf{V}(I^*)$. For the other direction, if W is a projective variety containing $\mathbf{V}(I^*)$ then we will show that $\mathbf{V}(I^*) \subseteq W$.

COX LITTLE O'SHEA PROOF: Let $I = \mathbf{I}(W)$ and define $\overline{W} := \mathbf{V}(I^*)$. We will show that $\overline{W}$ is the smallest projective variety containing W. So let V be a projective variety containing W and let $V = \mathbf{V}(\sum F_i\mathbb{F}[\mathbf{x}])$ with F_i homogeneous. Then F_i vanishes on V, so it vanishes on W, so its dehomogenization $(F_i)_*$ vanishes on W. Thus $(F_i)_* \in I = \mathbf{I}(W)$ and $(F_i)_*^* \in I^*$. By definition, this means that $(F_i)_*^*$ vanishes on $\overline{W} = \mathbf{V}(I^*)$. But $x_{n+1}^e(F_i)_*^* = F_i$. Since x_{n+1} does not vanish on $\overline{W}$ (because $\overline{W} \not\subseteq H_{n+1}$) we conclude that F_i vanishes on $\overline{W}$. Since this holds for all F_i we know that $\sum F_i\mathbb{F}[\mathbf{x}]$ vanishes on $\overline{W}$. But V is the place where this ideal vanishes, hence $\overline{W} \subseteq V$. ///

Thus we need to show that $\mathbf{V}(I^*)$ is the smallest projective variety containing $\mathbf{V}(I) \subseteq U_i$. On the one hand, suppose that $\mathbf{p} = (\mathbf{p}', 1) \in \mathbf{V}(I)$, so that $f(\mathbf{p}') = 0$ for all $f \in I$. Then for any $f = \sum (f_k)^* g_k \in I^*$, where $f_k \in I$, we have

$$f(\mathbf{p}) = \sum (f_k)^*(\mathbf{p}) g_k(\mathbf{p}) = \sum f_k(\mathbf{p}') g_k(\mathbf{p}) = 0,$$

and hence $\mathbf{p} \in \mathbf{V}(I^*)$. Conversely, suppose that $\mathbf{V}(J)$ is a projective variety containing $\mathbf{V}(I) \subseteq U_i$.

$\overline{V}$ is the smallest projective variety containing affine variety V. Therefore $\mathbf{I}(\overline{V})$ is the largest homogeneous ideal contained in $\mathbf{I}(V)$.

LOOK AT THIS

Identify U_i with the subspace $x_i = 1$ in $\mathbb{F}^{n+1}$. Given a variety $V \subseteq U_i$, we have an ideal $I \subset \mathbb{F}[\mathbf{x}_i']$. Homogenizing any element of this ideal gives a conical hypersurface containing the cone $\mathrm{Cone}(V) \subseteq \mathbb{F}^{n+1}$. The intersection of all such hypersurfaces equals the cone $\mathrm{Cone}(V)$.

But to conclude this I need to know that $\mathrm{Cone}(V)$ is Zariski-closed.

LOOK AT THIS

Examples:

- **Hypersurfaces.** Let $V_f \subseteq \mathbb{F}^n$ be a hypersurface with $f \in \mathbb{F}[\mathbf{x}']$ square-free, so that $I = \mathbf{I}(V_f) = f\mathbb{F}[\mathbf{x}']$. I claim that $I^* = f^*\mathbb{F}[\mathbf{x}]$, and hence

$$\overline{V_f} = \mathbf{V}(I^*) = \mathbf{V}(f^*\mathbb{F}[\mathbf{x}]) = V_{f^*}.$$

In other words, the projective closure of the affine hypersurface V_f is the projective hypersurface defined by the homogeneous polynomial f^*, as we already knew.

- **Points.** Consider the ideal $M_{\mathbf{p}} = \sum(x_i - p_i)\mathbb{F}[\mathbf{x}']$ of the affine point $\mathbf{p} = (p_1, \ldots, p_n) \in \mathbb{F}^n$. By homogenizing the generators of this ideal we obtain

$$\sum(x_i - p_i x_{n+1})\mathbb{F}[\mathbf{x}] \subseteq M_{\mathbf{p}}^*.$$

In fact, I claim that this is an equality. To see this, recall that the projective closure of the point $\mathbf{p} \in \mathbb{F}^n$ is the line $\overline{\mathbf{p}} = \{(\lambda p_1, \ldots, \lambda p_n, \lambda) : \lambda \in \mathbb{F}\} \subseteq \mathbb{F}^{n+1}$. Since this line is the intersection of the projective hyperplanes $x_i = p_i x_{n+1}$ we have

$$M_{\mathbf{p}}^* = \mathbf{I}(\overline{\mathbf{p}}) = \mathbf{I}\left(\cap H_{x_i - p_i x_{n+1}}\right) = \sum(x_i - p_i x_{n+1})\mathbb{F}[\mathbf{x}].$$

Warning: If $f_1, \ldots, f_m$ is a generating set for I then it is not necessarily true that $(f_1)^*, \ldots, (f_m)^*$ is a generating set for I^*. We will see an example in the next section.

Theorem. If $V \subseteq \mathbb{F}^n$ is irreducible then $\overline{V} \subseteq \mathbb{FP}^n$ is irreducible.

Proof. If $V = V_1 \cup V_2$ is reducible then I claim $\overline{V} = \overline{V_1} \cup \overline{V_2}$ is reducible. Indeed, if $\overline{V} = \overline{V_1}$ then $V = \overline{V} \cap U = \overline{V_1} \cap U = V_1$. Conversely,

if $\overline{V} = V_1 \cup V_2$ is reducible then $V = (V_1 \cap U) \cup (V_2 \cap U)$ is reducible. If V were irreducible then we would have $V = V_1 \cap U$ hence $V \subseteq V_1$. Since V_1 is closed this implies $\overline{V} \subseteq V_1$. Contradiction.

□

Nov 18,20: The Twisted Cubic Curve

The *twisted cubic curve* is defined as follows:

$$C = \{(t, t^2, t^3) : t \in \mathbb{F}\} \subseteq \mathbb{F}^3.$$

Intuitively, this is a one-dimensional curve living in three-dimensional space. However, it is not so clear how this picture relates to the ideal theory of the ring $\mathbb{F}[x, y, z]$.

Claim. The twisted cubic is a variety.

Proof. I claim that $C = \mathbf{V}(I)$, where the ideal I is defined as

$$I = (x^2 - y)\mathbb{F}[x, y, z] + (x^3 - z)\mathbb{F}[x, y, z].$$

To see this, first note that for any $(t, t^2, t^3) \in C$ and $f = (x^2 - y)g + (x^3 - z)h \in I$ we have

$$f(t, t^2, t^3) = (t^2 - t^2)g(t, t^2, t^3) + (t^3 - t^3)h(t, t^2, t^3) = 0,$$

and hence $f \in \mathbf{V}(I)$. Conversely, if $(a,b,c) \in \mathbf{V}(I)$ then since $x^2 - y$ and $x^3 - z$ are in I we must have $a^2 - b = 0$ and $a^3 - c = 0$, which implies that $(a,b,c) = (a, a^2, a^3) \in C$. $\square$

Picture:

Claim. The ideal I is prime, hence $C = \mathbf{V}(I)$ is an irreducible variety.

Proof. Consider any $f, g \in \mathbb{F}[x,y,z]$ with $fg \in I$. We want to show that $f \in I$ or $g \in I$. To do this, we first divide f by $x^3 - z$ in the ring $\mathbb{F}[x,y][z]$ to get

$$f(x,y,z) = (x^3 - z)f_1(x,y,z) + r(x,y,z),$$

for some $f_1, r \in \mathbb{F}[x,y,z]$ with $\deg_z(r) < \deg_z(x^2 - z) = 1$. But this implies that $r(x,y,z)$ has degree zero as a polynomial in z, hence $r(x,y,z) = r(x,y)$ is a polynomial in x and y alone. Then we divide $r(x,y)$ by $x^2 - y$ in the ring $\mathbb{F}[x][y]$ to get

$$r(x,y) = (x^2 - y)f_2(x,y) + f_3(x,y),$$

for some $f_2, f_3 \in \mathbb{F}[x,y]$ with $\deg_y(f_3) < \deg_y(x^2 - y) = 1$. But again, this implies that f_3 has degree zero in y, hence $f_3(x,y) = f_3(x)$ is a polynomial in x alone. By applying the same arguments to g we have the following expressions:

$$\begin{aligned} f &= (x^3 - z)f_1(x,y,z) + (x^2 - y)f_2(x,y) + f_3(x), \\ g &= (x^3 - z)g_1(x,y,z) + (x^2 - y)g_2(x,y) + g_3(x). \end{aligned}$$

Our goal now is to show that f_3 or g_3 is the zero polynomial, so that $f \in I$ or $g \in I$. To do this, we evaluate each expression at the point $(t, t^2, t^3) \in C$ to obtain

$$\begin{aligned} f(t,t^2,t^3) &= (t^3 - t^3)f_1(t,t^2,t^3) + (t^2 - t^2)f_2(t,t^2) + f_3(t) = f_3(t), \\ g(t,t^2,t^3) &= (t^3 - t^3)g_1(t,t^2,t^3) + (t^2 - t^2)g_2(t,t^2) + g_3(t) = g_3(t). \end{aligned}$$

If we define the polynomial $h(x) := f_3(x)g_3(x) \in \mathbb{F}[x]$, then since $fg \in I$ we have

$$h(t) = f_3(t)g_3(t) = f(t,t^2,t^3)g(t,t^2,t^3) = (fg)(t,t^2,t^3) = 0.$$

Since this holds for **infinitely many** $t \in \mathbb{F}$, Descartes' theorem says that $h(x)$ is the zero polynomial. Finally, since $\mathbb{F}[x]$ is a domain we have $f_3(x) = 0$ or $g_3(x) = 0$, as desired. $\square$

Corollary. The ideal I is radical, hence $\mathbf{I}(C) = \mathbf{IV}(I) = I$ by the Nullstellensatz.

Proof. Indeed, every prime ideal is radical. $\square$

So far everything makes sense. The twisted cubic $C \subseteq \mathbb{F}^3$ can be expressed as an intersection of two (hyper-)surfaces in $\mathbb{F}^3$, which agrees with our intuition that C is one-dimensional. Furthermore, the ideal of C is generated by the minimal polynomials of these two hypersurfaces.

Now we consider the projective closure $\overline{C} \subseteq \mathbb{FP}^3$. It turns out that this is more complicated.

Claim. Let $\mathbb{F}^3 = \{(x, y, z)\}$ be embedded in $\mathbb{FP}^3 = \{(w : x : y : z)\}$ as the complement of the plane $w = 0$. Then the projective closure of the twisted cubic is equal to the set

$$T = \{(s^3 : s^2t : st^2 : t^3) : s, t \in \mathbb{F}\} \subseteq \mathbb{FP}^3.$$

Proof. It is more convenient to work with conical affine varieties. Using this language, the projective closure is the smallest conical affine variety $\overline{C} \subseteq \mathbb{F}^4$ containing the set $C = \{(1, t, t^2, t^3) : t \in \mathbb{F}\} \subseteq \mathbb{F}^4$. I claim that $\overline{C}$ is equal to

$$T = \{(s^3, s^2t, st^2, t^3) : s, t \in \mathbb{F}\} \subseteq \mathbb{F}^4.$$

First we observe that the set T is conical. Indeed, for any λ we can write $\lambda = \omega^3$ and then

$$(\lambda s^3, \lambda s^2 t, \lambda s t^2, \lambda t^3) = ((\omega s)^3, (\omega s)^2(\omega t), (\omega s)(\omega t)^2, (\omega t)^3) \in T.$$

Next we will prove that T is contained in $\overline{C}$. To see this, we observe that the Zariski closure of $\mathrm{Cone}(C) = \{(\lambda, \lambda t, \lambda t^2, \lambda t^3) : \lambda, t \in \mathbb{F}\}$ is a conical variety containing C, and hence is contained in $\overline{C}$.[38] To see this, consider any point $(s^3, s^2t, st^2, t^3) \in T$ with $s \neq 0$, so that

$$(s^3, s^3(t/s), s^3(t/s)^2, s^3(t/s)^3) \in \mathrm{Cone}(C) \subseteq \overline{C}.$$

Finally, we will show that T is a variety, hence it will follow from the minimality of $\overline{C}$ that $T = \overline{C}$. Indeed, we will show that $T = \mathbf{V}(J)$ for the following homogeneous ideal:

$$J = (x^2 - wy)\mathbb{F}[w, x, y, z] + (xy - wz)\mathbb{F}[w, x, y, z] + (xz - y^2)\mathbb{F}[w, x, y, z].$$

Equivalently, T is the intersection of the three corresponding conical hypersurfaces in $\mathbb{F}^4$. To see this we first observe that any point $(w, x, y, z) = (s^3, s^2t, st^2, t^3)$ is in the intersection of these hypersurfaces because

$$\begin{array}{rcccl} x^2 - wy & = & (s^2t)^2 - (s^3)(st^2) & = & 0, \\ xy - wz & = & (s^2t)(st^2) - (s^3)(t^3) & = & 0, \\ xz - y^2 & = & (s^2t)(t^3) - (st^2)^2 & = & 0. \end{array}$$

Conversely, suppose that the point $(a, b, c, d) \neq (0, 0, 0, 0)$ is in the intersection of the hypersurfaces, so that $b^2 - ac = 0$, $bc - ad = 0$ and $bd - c^2 = 0$. Note that we must have either $a \neq 0$ or $d \neq 0$, otherwise $(a, b, c, d) = (0, 0, 0, 0)$. From the symmetry $a \leftrightarrow d$ and $b \leftrightarrow c$ we may assume that $d \neq 0$ and hence $d = t^3$ for some $t \neq 0$. Since $d \neq 0$ we observe that a, b, c are either all zero or all non-zero. In the former case we can take $s = 0$ to obtain

$$(a, b, c, d) = (0, 0, 0, d) = (s^3, s^2t, st^2, t^3) \in T.$$

In the latter case we observe that $a/b = b/c = c/d$. Then by defining $s = (a/b)t$ we obtain

$$(a, b, c, d) = (s^3, s^2t, st^2, t^3) \in T. \qquad \square$$

[38] In fact, $\overline{C}$ is equal to the Zariski of $\mathrm{Cone}(C)$.

Corollary. The ideal $J = (x^2 - wy, xy - wz, xz - y^2)$ satisfies $J \subseteq \mathbf{I}(\overline{C})$.

Proof. We showed that $\mathbf{V}(J) = \overline{C}$, and it follows that $J \subseteq \mathbf{IV}(J) = \mathbf{I}(\overline{C})$. □

The proof of the next result is similar to our proof that I is prime, since it involves tricky computations with generators. The theory of Gröbner bases provides a general machinery for these types of proofs. However, since we have not developed that machinery we will use an ad hoc method.[39] There is a more conceptual approach using *Hilbert functions* that we might encounter next semester.

Claim. We also have $\mathbf{I}(\overline{C}) \subseteq J$, and hence $J = \mathbf{I}(\overline{C})$.

Proof. Consider an element $f \in \mathbf{I}(\overline{C}) \subseteq \mathbb{F}[w, x, y, z]$ satisfying

$$f(s^3, s^2t, st^2, t^3) = 0 \quad \text{for all } s, t \in \mathbb{F}.$$

There are many ways to proceed. First we divide f by $x^2 - wy$ with respect to x to obtain

$$\begin{aligned} f(w, x, y, z) &= (x^2 - wy)f'(w, x, y, z) + xp(w, y, z) + q(w, y, z), \\ f(s^3, s^2t, st^2, t^3) &= 0f'(s^3, s^2t, st^2, t^3) + s^2tp(s^3, st^2, t^3) + q(s^3, st^2, t^3), \\ 0 &= s^2tp(s^3, st^2, t^3) + q(s^3, st^2, t^3). \end{aligned}$$

Next we wish to divide $p(w, y, z)$ and $q(w, y, z)$ by a polynomial in $\mathbb{F}[w, y, z] \cap J$. Unfortunately none of the generators have this property so we use the auxiliary polynomial $y^3 - wz^2 \in \mathbb{F}[w, y, z]$, which is in J because

$$y^3 - wz^2 = z(xy - wz) - y(xz - y^2) \in J.$$

Dividing p and q by $y^3 - wz^2$ gives

$$\begin{aligned} p &= (y^3 - wz^2)p'(w, y, z) + y^2p_1(w, z) + yp_2(x, z) + p_3(x, z), \\ q &= (y^3 - wz^2)q'(w, y, z) + y^2q_1(w, z) + yq_2(x, z) + q_3(x, z), \end{aligned}$$

and substituting $(w, x, y, z) = (s^3, s^2t, st^2, t^3)$ gives

$$\begin{aligned} 0 &= (st^2)^2p_1(s^3, t^3) + (st^2)p_2(s^3, t^3) + p_3(s^3, t^3), \\ 0 &= (st^2)^2q_1(s^3, t^3) + (st^2)q_2(s^3, t^3) + q_3(s^3, t^3). \end{aligned}$$

Next is a trick. We multiply the first of these equations by $x = s^2t$ to obtain

$$\begin{aligned} 0 &= (s^4t^5)p_1(s^3, t^3) + (s^3t^3)p_2(s^3, t^3) + (s^2t)p_3(s^3, t^3), \\ 0 &= (s^2t^4)q_1(s^3, t^3) + (st^2)q_2(s^3, t^3) + q_3(s^3, t^3). \end{aligned}$$

[39] See Cox-Little-O'Shea page 389 for a proof using Gröbner bases.

Since these equations hold for all values of $s, t \in \mathbb{F}$ we can think of s and t as variables. Then by collecting monomials $s^i t^j$ according to the exponents (i, j) modulo 3 we obtain

$$\begin{aligned} 0 &= (s^4 t^5) p_1(s^3, t^3) + (st^2) q_2(s^3, t^3), \\ 0 &= (s^3 t^3) p_2(s^3, t^3) + q_3(s^3, t^3), \\ 0 &= (s^2 t) p_3(s^3, t^3) + (s^2 t^4) q_1(s^3, t^3). \end{aligned}$$

In other words:

$$\begin{aligned} 0 &= wz p_1(w, z) + q_2(w, z), \\ 0 &= wz p_2(w, z) + q_3(w, z), \\ 0 &= p_3(w, z) + z q_1(w, z). \end{aligned}$$

Finally, we can substitute these expressions into f to obtain

$$\begin{aligned} f &= (x^2 - wy) f' + x[(y^3 - wz^2) p' + y^2 p_1 + y p_2 + p_3] + (y^3 - wz^2) q' + y^2 q_1 + y q_2 + q_3 \\ &= (x^2 - wy) f' + x[(y^3 - wz^2) p' + y^2 p_1 + y p_2 - z q_1] + (y^3 - wz^2) q' + y^2 q_1 - wyz p_1 - wz p_2 \\ &= (x^2 - wy) f' + (y^3 - wz^2)(xp' + q') + (xy^2 - wyz) p_1 + (xy - wz) p_2 - (y^2 - xz) q_1 \end{aligned}$$

The proof is completed by observing that each of the coefficients in this expression is an element of J, and the only one left to check is

$$xy^2 - wyz = z(x^2 - wy) - x(xz - y^2) \in J. \qquad \square$$

To summarize, we have shown that

$$\begin{aligned} \mathbf{I}\{(t, t^2, t^3) : t \in \mathbb{F}\} &= \mathbf{I}(C) = I = (x^2 - y, x^3 - z), \\ \mathbf{I}\{(s^3, s^2 t, st^2, t^3) : s, t \in \mathbb{F}\} &= \mathbf{I}(\overline{C}) = J = (x^2 - wy, xy - wz, xz - y^2). \end{aligned}$$

Corollaries.

- Let $I^* \subseteq \mathbb{F}[w, x, y, z]$ be the homogenization of $I \subseteq \mathbb{F}[x, y, z]$. From the previous section we know that $I^* = \mathbf{I}(\overline{C})$, and hence $I^* = J$.
- From the previous section, we know that the projective closure of an irreducible variety is irreducible. It follows that $J = \mathbf{I}(\overline{C})$ is a prime ideal.

Generalization: The Rational Normal Curve. Consider the following set:

$$C = \{(t, t^2, \ldots, t^n) : t \in \mathbb{F}\} \subseteq \mathbb{F}^n.$$

This is an affine variety with corresponding ideal

$$\mathbf{I}(C) = (x_2^2 - x_1, x_3^3 - x_1, \ldots, x_n^n - x_1) \subseteq \mathbb{F}[x_1, \ldots, x_n].$$

This ideal is prime, hence the curve C is irreducible. The projective closure of C is given by

$$\overline{C} = \{(s^n, s^{n-1}t, \ldots, st^{n-1}) : s, t \in \mathbb{F}\} \subseteq \mathbb{F}^{n+1},$$

which is irreducible because C is. The corresponding homogeneous ideal $\mathbf{I}(\overline{C}) \subseteq \mathbb{F}[x_0, \ldots, x_n]$ is generated by the 2×2 minors of the following $2 \times n$ matrix:

$$\begin{pmatrix} x_0 & x_1 & \cdots & x_{n-1} \\ x_1 & x_2 & \cdots & x_n \end{pmatrix}.$$

This ideal is prime because $\overline{C}$ is irreducible. ///

Most of this can be proved by straightforward generalization of the proofs in this section. The only result that is more difficult is to find the generators of the ideal $\mathbf{I}(\overline{C})$.[40] To end this section we illustrate an unfortunate property of the twisted cubic.

Corollary. We have shown that $I^* = J$, and it follows that

$$(x^2 - y, x^3 - z) = I = (I^*)_* = J_* = (x^2 - y, xy - z, xz - y^2).$$

Hence the affine curve C is the intersection of three affine surfaces:

$$C = \mathbf{V}(x^2 - y) \cap \mathbf{V}(xy - z) \cap \mathbf{V}(xz - y^2).$$

I claim that this description is **redundant**. Indeed, one can easily check that any two of these surfaces suffice to define C. Similarly, the projective twisted cubic is the intersection of the corresponding projective surfaces:

$$\overline{C} = \mathbf{V}(x^2 - wy) \cap \mathbf{V}(xy - wz) \cap \mathbf{V}(xz - y^2).$$

But I claim that this description is **not redundant**. That is, no two of these surfaces suffice to define the curve $\overline{C}$. For example, the intersection of the first two surfaces contains $\overline{C}$, but it also contains the projective line $L = \{(0 : 0 : u : v) : (u : v) \in \mathbb{FP}^1\}$. On the other hand, this line is not contained in $\overline{C}$ because any point $(s^3, s^2t, st^2, t^3) = (0, 0, u, v)$ must have $s = 0$ and hence $u = st^2 = 0$. More generally, one can show that the homogeneous ideal $J = (x^2 - xy, xy - wz, xz - y^2)$ **cannot be generated by two elements**.[41]

[40] And the case $n = 3$ was already difficult!

[41] This is not easy. It can be proved, for example, with the theory of Hilbert functions. There is also a subtlety here because the curve $\overline{C}$ **can** be expressed as the intersection of two projective surfaces; for example, as $\mathbf{V}(x^2 - wy) \cap \mathbf{V}(y^3 - 2xyz + w^z)$. It's just that these two polynomials do not generate the ideal of the curve. *Hartshorne's conjecture* implies that any projective curve in $\mathbb{FP}^n$ can be expressed as an intersection of $n - 1$ projective hypersurfaces. One says that a d-dimensional variety $V \subseteq \mathbb{FP}^n$ is a *set-theoretic complete intersection* if it is an intersection of $n - d$ projective hypersurfaces, and an *ideal-theoretic complete intersection* if its ideal can be generated by $n - d$ homogeneous polynomials. Of course these concepts depend on the concept of *dimension*, which we have not defined. Slogan: "Varieties can be arbitrarily complicated." Next semester we will focus on one-dimensional varieties, which already have a very rich theory.

This is our first indication of the disctinction between local and global properties of projective varieties. Locally (i.e., in any affine chart) the twisted cubic can be viewed as an intersection of two surfaces. Moreover, we can choose these surfaces so that the tangent space at any point (defined as the intersection of the tangent planes of the surfaces at this point) is one-dimensional. This agrees with our intuition that the twisted cubic is a "one-dimensional variety." In fact, one could use this as the **definition** of the dimension. However, the global picture is more complicated because these local descriptions do not patch together in a trivial way. Next semester we will develop a language to describe this phenomenon.