MATH 497A: INTRODUCTION TO APPLIED ALGEBRAIC GEOMETRY

These are notes from the Penn State 2015 MASS course Introduction to Applied Algebraic Geometry. This class is taught by Jason Morton and the notes are those of Sara Jamshidi, the teaching assistant.

The material is an amalgam of many sources, which are cited throughout these notes. If you discover any typos or mistakes, please email me at sxj937@psu.edu.

Date: December 12, 2015.

Lecture 1

Assigned Reading

The reading for this lecture is §1 of chapter 1 in [**?**].

Within commutative algebra and algebraic geometry, it is common to denote a field with a k.

Recall 1.1. Recall from abstract algebra[1] that a field is a set F closed under addition $+$ and multiplication $\cdot$, such that

(1) $(F, +)$ is an abelian group,
(2) $(F^*, \cdot)$ is an abelian group, and
(3) the distributive property holds: $a(b + c) = ab + ac$, where $F^* = F \setminus \{0\}$.

One need not think about the technical definition of a field, however. Since elementary school, we have worked over fields. These are merely sets where one can add, subtract, multiply and divide (except for 0) without any issues. The fields we will study in the class are

- The **complex numbers**, denoted by $\mathbb{C}$.
- The **real numbers**, denoted by $\mathbb{R}$.
- The **rational numbers**, denoted by $\mathbb{Q}$.
- **Finite fields**, denoted by $\mathbb{F}_{p^n}$ where p is prime.

Example 1. An example of a finite field is $\mathbb{F}_3 = \{0, 1, 2\}$, the finite field of three elements. Notice that $\{0, 1, 2\}$ is a cyclic group of order three over addition and $\{0, 1, 2\}$ is also a multiplicative group (modulo 3) with the following table:

$\times$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

We leave it as an exercise to check that the distributive property holds.

The field $\mathbb{C}$ is particularly important to the field of algebraic geometry. Because we only have access to three dimensions, we cannot draw even the simplest complex valued functions. So we create pictures based on $\mathbb{R}$. Computers, however, do not do well over uncountable fields. So many computational methods use $\mathbb{Q}$ and $\mathbb{F}_{p^n}$.

In this class, the term "polynomial" will often denote multivariate polynomials. These are linear combinations of monomials (i.e. products of variables) over k. For example,

$$3xy^2 + 2x^2 + \pi i z^4 y^2$$

is a polynomial with three variables, x, y, z, and is a linear combination over $\mathbb{C}$.

Although we can use variables like x, y, z, w, we will more often use the variables $x_1 \ldots, x_n$.[2]

[1] The definition of a field is often presented using rings to account for the field of one element $\{0\}$, which this definition excludes. We use this definition, however, since most undergraduate curriculum studies groups more heavily than rings.

[2] For projective spaces (which will be discussed later), we will use x_0 as the first variable.

Multivariate monomials, like

$$x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n},$$

have a **multidegree[?]**, namely $\alpha = (\alpha_1, \ldots, \alpha_n) \in (\mathbb{Z}_{\geq 0})^n$.The **degree** of this term is found by taking the ℓ^1-norm of α.[3] Alternatively, we can calculate the degree by taking the dot product of α with the vector 1_n, or the n dimensional vector all of whose entries are 1. Hence,

$$\deg(\alpha) = \alpha_1 + \cdots + \alpha_n.$$

Notation 1.2. We often condense the notation. For a multivariate polynomial, $f \in k[x_1, \ldots, x_n]$[4] we often express f as

$$f = \sum_{\alpha \in \mathcal{A}} C_\alpha x^\alpha$$

where α is a multidegree, x^α represents $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$, and $C_\alpha \in k$.

The degree of a polynomial is the maximum monomial degree. That is,

$$\deg(f) = \max_{\alpha \in \mathcal{A}} |\alpha|.$$

Example 2. $\deg(7x_1^2 + \pi x_2 x_3^4 + x_1^5 x_3^3 x_7^2) = 10$. Why? Because the monomial $7x_1^2$ has degree 2, $\pi x_2 x_3^4$ has degree 5, and $x_1^5 x_3^3 x_7^2$ has degree 10. So the maximum monomial degree is 10.

Notice that the multidegree of a monomial is a vector. For example,

$$\text{mdeg}\left(x_1^2 x_3 x_5^3\right) = \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \\ 3 \end{pmatrix}.$$

The degree of this monomial is 6.

Question 1.3. Can we find a multidegree for a polynomial that is not a monomial?

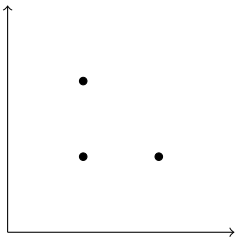
This is not well defined. However, the set of multidegrees for the monomials of a function will turn out to be a very informative set. In other words, the expression

$$\text{mdeg}(2x^2y + xy + 3xy^2) = \textit{undefined},$$

but the set

$$\left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$$

can tell us something about the *variety* carved out by the function.

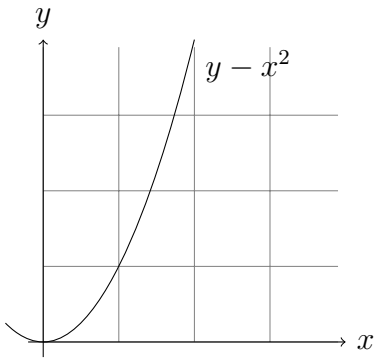


[3]The ℓ_1-norm of a vector v is the sum of the absolute values of its entries: $\sum_{i=1}^n |v_i|$.

[4]This means f is a polynomial with coefficients coming from k in the variables $x_1, \ldots, x_n$. For example, $ax_1^{\alpha_1} + bx_7^{\alpha_7} x_{12}^{\alpha_{12}}$ is such a polynomial, so long as $a, b \in k$.

At this point, you might be asking, "What is a variety?" If so, good question! A variety, or more specifically an **algebraic variety**, is the set of solutions of a system of polynomial equations.

So the set consisting of the polynomial $y - x^2$ is a variety whose zero locus (the solution to $y - x^2 = 0$) is the parabola (if w assume $k = \mathbb{R}$).



Remark 1.4. In algebraic geometry, we often want to study these varieties intrinsically; that is, without thinking about some ambient space (like $\mathbb{R}^2$) where they can live. For this class, we will understand varieties in the context of some ambient space.

Generally, there are two kinds of ambient spaces in which a variety lives:

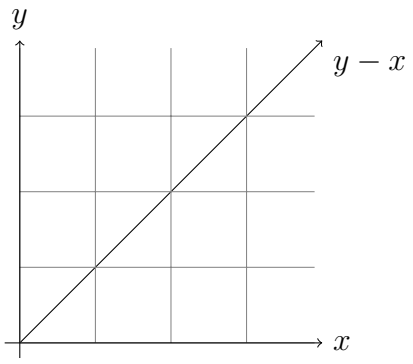
(1) projective space
(2) affine space

Definition 1.5. *The n-dimensional affine space over a field k is the set $k^n = \{(a_1, \ldots, a_n) : a_i \in k\}$.*

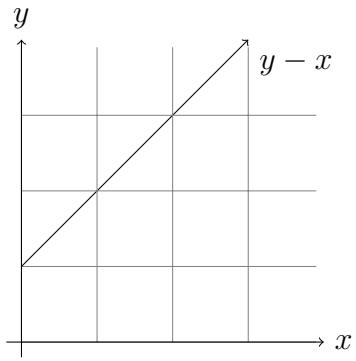
Example 3. $k^1, k^2, \ldots k^n$ are all affine spaces.

Question 1.6. How is an affine space different from a vector space?

A vector space is a special kind of affine space; however, an affine space is not necessarily a vector space. Instead, it can be an *affine transformation* of a vectors (sub)space. For example, the variety $y - x$ is a vector subspace of $\mathbb{R}2$.



Another variety is $y - x - 1$. This is *not* a vector subspace.



Aside 1.7. You may think about polynomials as related to dual spaces. That is, $f \in k[x_1, \ldots, x_n]$ defines a function $k^n \to k$ by evaluation (i.e. plugging into the polynomial). This f is an element in $\mathscr{L}(k^n, k)$, so it lives in the dual space of k^n.

We had briefly mentioned finite fields in the beginning of this lecture. One has to be very careful when working with them as they can be counterintuitive. For example, suppose a function f is such that

$$f(a_1, \ldots, a_n) = 0 \quad \forall a \in k^n$$

where k is a finite field. It is not necessarily true that f is the zero function. For example, consider $k = \mathbb{F}_2 = \{0, 1\}$ and $f(x) = x(x-1)$. Notice that $f(0) = 0 \cdot 1 = 0$ and $f(1) = 1 \cdot 0 = 0$. But, of course, $x(x-1)$ is not the zero function.

This, however, can't happen over infinite fields, which are often called **fields of characteristic zero**).

Lecture 2

Assigned Reading

The reading for this lecture is §2 of chapter 1 in [**?**].

In Lecture **??**, we essentially ended with the following theorem:

Theorem 2.1. *Let k be a field of characteristic zero and let $f \in k[x_1, \ldots, x_n]$. Then $f = 0$ in $k[x_1, \ldots, x_n]$ if and only if $f : k^n \to k$ is the zero function.*[5]

A nice corollary to this fact is that if two functions $f, g : k^n \to k$ such that $f(x) = g(x)$ for each $x \in k^n$ and k is characteristic zero, then f and g are the same function. We can see this by observing that $f - g = 0$ over an infinite field k. By the above theorem, $f - g$ is precisely the zero function, so f and g must be the same function.

Counterexample 1. Let $k = \mathbb{F}_3$ and consider the following two polynomials in $k[x]$:

- $f(x) = x(x-1)(x-2)$
- $g(x) = x^2(x-1)(x-2)$

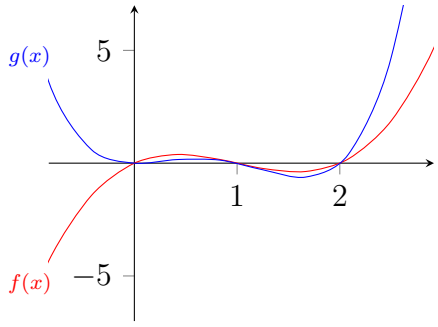
These two polynomials are not the same. Let us now plug in each point in $\mathbb{F}_3$:

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

So when k is not characteristic zero (here it is characteristic 3), this fact does not hold.

Remark 2.2. For the counterexample above, recall that $-1 = 2$ and $-2 = 1$ in $\mathbb{F}_3$.

Notice that the reason finite fields fail is because we can construct polynomials whose roots *cover* the field.



Because we are studying the null space of a set of polynomials, it is important for us to recall the fundamental theorem of algebra (FTA).

Theorem 2.3. *(**Fundamental Theorem of Algebra**) Every nonconstant polynomial $p(x) \in \mathbb{C}[x]$ has at least one complex root.*

[5] The proof of the above theorem can be found on page 3 of [**?**].

Corollary 2.4. *A polynomial over $\mathbb{C}$ of degree n has n roots (counting multiplicity).*

Let $p_n(x)$ have degree n. It has root α_n. Thus, we can divide $p(x)$ by $(x - \alpha_n)$ to get a new polynomial of degree $n-1$, called $p_{n-1}(x)$. This can continue so long as the degree is greater than or equal to 1. Hence, the polynomial will have n (possibly not unique) roots: $\alpha_n, \alpha_{n-1}, \ldots, \alpha_1$.

Definition 2.5. *A field k is said to be* **algebraically closed** *if every nonconstant polynomial $p(x) \in k[x]$ has at least one root.*

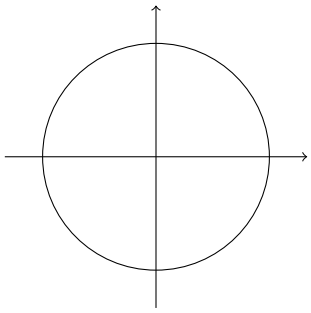
Within algebraic geometry, we prefer to work over algebraically closed fields because roots do not vanish.

Definition 2.6. *Let k be a field and let $f_1, \ldots, f_s$ comprise the set of polynomials $\mathscr{F} \subset k[x]$. Then we say $\mathscr{F}$ defines a variety:*

$$V(f_1, \ldots, f_s) = V(\mathscr{F}) := \{(a_1, \ldots, a_n) \in k^n \mid f_i(a_1, \ldots, a_n) = 0, \ \ \forall i \in 1, \ldots, s\}$$

We may also write $Z(\mathscr{F})$, which stands for the "zero locus" of $\mathscr{F}$. This is yet another term to refer to the roots, zeroes, nullspace, etc. of $\mathscr{F}$.

Example 4. The function $f(x) = x^2 + y^2 - 1$ defines a variety. If we consider $f(x) \in \mathbb{R}[x]$, we get the unit circle.



More examples of varieties like this exist in [**?**].

Remark 2.7. Although we will often consider these functions as existing over $\mathbb{C}[x]$, we will often draw them over $\mathbb{R}$, assuming such a depiction is even possible.

A family of polynomials have no restrictions on them and, therefore, may contain redundancies.[6] To avoid this problem of redundancy, we will work with *ideals* in the polynomial ring. We will formally discuss ideals on Monday.

2.1 Biased Coins. Suppose we have to distinct coins, say a quarter and a time. Also suppose these coins can be biased. Let

- q_H be the probability the quarter comes up heads,
- q_T be the probability the quarter comes up tails,
- d_H be the probability the dime comes up heads and
- d_T be the probability the dime comes up tails.

There are four possible outcomes

[6] This is analogous to a system of linear equations whose matrix is not of full rank.

	quarter	dime
x_{HH}	H	H
x_{HT}	H	T
x_{TH}	T	H
x_{TT}	T	T

Because these are probabilities, we have the property

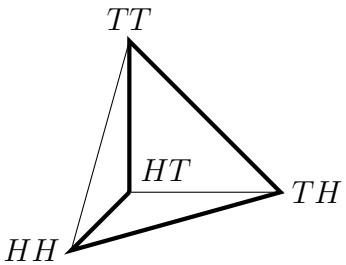
$$x_{HH} + x_{HT} + x_{TH} + x_{TT} = 1.$$

For the moment, let us ignore the requirement that these values be nonnegative.

Questions about this problem:

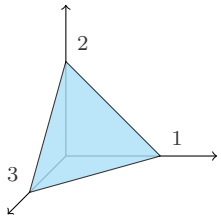
(1) Can I get any 4 numbers with this set up?
(2) Does there exist a nonzero polynomial in $\mathbb{C}[x_{HH}, x_{HT}, x_{TH}, x_{TT}]$ which is always zero?
(3) Now add the restriction that all the values of $x_{HH}, x_{HT}, x_{TH}, x_{TT}$ are nonnegative. Draw the set of possible probability distributions.

A hint for problem 3 is to consider the tetrahedron with the following labeling:



Notice that a particular path in this tetrahedron has been highlighted. This is the path where only one of the two coins has a changing probability.

For some motivation regarding how the above picture is a hint, let us consider a lower-dimensional example. Suppose we have three variables x_1, x_2, x_3 such that $x_1 + x_2 + x_3 = 1$. This represents a plan in $\mathbb{R}^3$. Now consider only the portion of this plane for which $x_i \geq 0$ for $i \in \{1, 2, 3\}$. This is a triangle whose vertices represent certainty in the outcomes $1, 2, 3$.

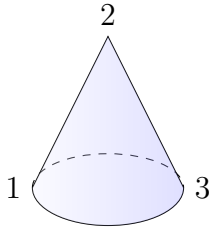


Remark 2.8. A triangle is called a 2-simplex. A tetrahedron is a 3-simplex. We will often be working with higher-dimensional versions of these shapes. To avoid worrying about the dimension, we will simply refer to these shapes as *simplicies*.

Fact 2.9. *The (complex) variety associated to a single polynomial, $V(f)$, is a hypersurface, meaning it has dimension 1 less than the ambient space.*

So in the above examples, we have a plane ($x_1 + x_2 + x_3 = 1$) in $\mathbb{R}^3$ and a three-dimensional "surface" ($x_{HH} + x_{HT} + x_{TH} + x_{TT} = 1$) in $\mathbb{R}^4$.

Example 5. Consider the function $p_1p_3 - p_2^2$. We know from the above fact that this is a two-dimensional surface in $\mathbb{R}^3$. This function is an finite elliptical cone. Intersecting this with the restrictions we expect from probability (e.g. $0 \leq p_i \leq 1$), we should get something very close to a probability simplex. How might we label the object?



A variety of a family of functions $V(\mathscr{F})$ is the set of zeros the functions in $\mathscr{F}$ share. As a result, we can say the following.

Lemma 2.10. *Let $\mathscr{F}$ be a finite set of polynomials in $k[x_1, \ldots, x_n]$. The variety $V(\mathscr{F})$ is the intersection of the hypersurfaces*

$$\bigcap_{f \in \mathscr{F}} V(f).$$

Question 2.11. What happens if $\mathscr{F}$ contains an infinite number of polynomials?

2.2 Redundancy. We mentioned redundancy in the beginning of the lecture. Notice that with redundant functions (for example, functions which are constant multiples of each other), it is possible that $\mathscr{F}_1 \neq \mathscr{F}_2$ but $V(\mathscr{F}) = V(\mathscr{F})$. It turns out to be very difficult in general to know when this is the case. For this issue, it helps to know the following:

Fact 2.12. *Every variety is the intersection of finitely my of hypersurfaces.*

This result is known as Hilbert's Basis Theorem, which we will discuss in more detail later in the course. In particular, it states,

Theorem 2.13. *Every (possibly infinite) set $\mathscr{F}$ of polynomials in $\mathbb{C}[x_1, \ldots, x_n]$ has a finite subset $\mathscr{F}' \subset \mathscr{F}$ such that $\langle \mathscr{F} \rangle = \langle \mathscr{F}' \rangle$ and (therefore) $V(\mathscr{F}) = V(\mathscr{F}')$.*

Notation 2.14.

$$\langle \mathscr{F} \rangle = \{h_1 f_f + \ldots + h_r f_r \ : \ f_i \in \mathscr{F} \text{ and } h_i \in \mathbb{C}[x_1, \ldots, x_n]\}$$

2.2.1 Dimension. We expect that for each f we add to the set $\mathscr{F}$ will cut our dimension by 1. Hilbert's basis theorem inspires the idea that we can somehow reduce or eliminate redundancies altogether.

When the f_i are linear, this idea is precisely the rank-nullity theorem. The rank tells us the redundancy among the f_i. When there is no redundancy, the dimension of the variety is the dimension of the ambient space minus the number of polynomials.

$$\dim(V(\mathscr{F})) = \dim W - \text{\# of polys}$$

Check out the discussion on dimension on pages 8 - 10 in [**?**].

Lecture 3

Assigned Reading

The reading for this lecture is §3 of chapter 1 in [**?**].

Another application of algebraic geometry is optimization. Typically, optimization problems are of the form

$$\min f(x_1, \ldots, x_n) \qquad \text{s. t. } g(x_1, \ldots, x_n) = 0$$

The method arising from multivariate calculus is Lagrange multipliers where $\nabla f - \lambda \nabla g = 0$. Together with g this defines a variety with $n+1$ equations and $n+1$ unknowns. We expect a zero-dimensional variety, which contains all the optimized points.

In lecture **??**, lemma **??** stated that the variety of a finite set of polynomials can be thought of as the intersection of the individual function's varieties. A similar statement is the following.

Lemma 3.1. *Let V, W be affine varieties in k^n. Then $V \cap W$ and $V \cup W$ are variaties in k^n. In particular, let $V = V(f_1, \ldots, f_s)$ and $W = V(g_1, \ldots, g_t)$. Then,*

- $V \cap W = V(f_1, \ldots, f_s, g_1, \ldots, g_t)$
- $V \cup W = V(\{f_i g_i\}_{i,j}, i \in [s]\ j \in [t])$

Remark 3.2. A subset U of a space X corresponds to an *indicator function*, 1_U where

$$1_U(x) = \begin{cases} 1 & \text{if } x \in U \\ 0 & \text{otherwise} \end{cases}.$$

Given two sets, U and V, how do we define the indicator function of their intersection? We take the product:

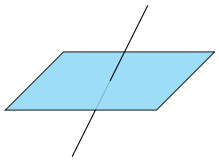
$$1_{U \cap V}(x) = 1_U(x) 1_V(x).$$

So the intersection is expressed with a product. Lemma **??** shows us that an "or" (i.e. a union) is calculated as an "and" (i.e. an intersection).

Here are some questions to consider:

(1) When does a system $f_1, \ldots, f_s$ have at least 1 solution? Put another way, when is it true that
$$V(f_1, \ldots, f_s) \neq \emptyset.$$
(2) When is the solution finite?
(3) What is the dimension of $V(f_1, \ldots, f_s)$?
(4) How do we tell if a variety is a union?

With regard to the last question on that list, we may construct a union of varieties (which is also a variety) from those with different dimensions. How might factoring be important here?



3.1 Implicitization and Parameterization. Two common ways of representing a variety is via the *implicit representation* and the *parametric representation*. The implicit representation is precisely the set of polynomials whose zero locus defines the variety. The parametric representation is a set of equations whose image (almost) defines the variety.

Let us begin with a example from linear algebra.

Example 6. Consider the following linear transformation from $\mathbb{R}^3 \to \mathbb{R}^3$:

$$N = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The image of this transformation is a linear subspace. From the homework, we know it is an affine variety.

Let (x', y', z') be the variables for the domain and (x, y, z) be the variables of the codomain. Then the transformation is equivalent to the set of polynomials:

$$\begin{cases} x' = x \\ y' = y \\ z' = 0 \end{cases}.$$

The image Im N the xy-plane. So a parametric equation for the image could trivially be written as

$$\begin{cases} x(s,t) = s \\ y(s,t) = t \end{cases}.$$

This is the *polynomial parametric representation*. The affine variety $V(z = 0)$ is equivalent to the image. The function $z = 0$ is the *implicit representation* of the variety.

Example 7. Consider the following linear transformation from $\mathbb{R}^3 \to \mathbb{R}^3$:

$$M = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

This gives rise to the set of equations

$$\begin{cases} x' = x + y + z \\ y' = z \\ z' = x + y \end{cases}.$$

Based on the above equations, we can quickly recognize an implicitization:

$$\text{Im } (M) = V(x' - y' - z').$$

Example 8. Consider the variety defined by the following parameterization of $\mathbb{R}^3$:

$$\begin{cases} x = -1 - 3t \\ y = 2 + 2t \\ z = t \end{cases}.$$

From the parameterization, we see that this is a line! The implicitization is

$$V(x + y + z - 1, x + 2y - z - 3).$$

Remark 3.3. Notice that the variety, which is a line, has dimension 1 in a 3-dimensional space. Also notice that the implicitization consists of two equations. This should remind you of the rank-nullity theorem.

Let us compare the features of each type of representation.

Parametric Representation	**Implicit Representation**
easy to draw and sample	easy to test membership
often how it is given	generally difficult to find
better for gradient-based optimization	better for computing properties

If a parameterization is rational[7] (as often needs to be the case), there will be some points missing from the variety. For example, consider

$$x = \frac{1-t^2}{1+t^2} \quad y = \frac{2t}{1+t^2}$$

which gives rise to the implicitization

$$x^2 + y^2 = 1.$$

The parameterization, however, misses the point $(-1, 0)$.

If we are given a rational parameterization, which we know will be missing points, how can we recover the full variety containing it? We need to find the smallest variety containing the rational parameterization. This is a closure in the Zariski topology.

The closed sets of a Zariski topology of an affine space are generated from the zeros of polynomials over that space. Hence, all the closed sets of the Zariski topology are varieties.

Recall 3.4. Let us take a moment to recall the definition of a topology. It is generally defined using open sets. Let X be a set and $\mathscr{T}$ be a family of subsets. We call $\mathscr{T}$ a (open set) topology if

(1) $X, \emptyset \in \mathscr{T}$

(2) Given a (possibly uncountable) subset $\mathscr{S} \subset \mathscr{T}$,

$$\bigcup_{V \in \mathscr{S}} V \in \mathscr{T}.$$

(3) Given a finite subset $\{V_1, \ldots, V_n\} \subseteq \mathscr{T}$,

$$\bigcap_{i=1}^{n} V_i \in \mathscr{T}.$$

The Zariski topology is defined with closed sets. So the analogous definition is:

(2) the intersection of a (possibly uncountable) set of closed sets is closed in the topology, and

(3) the union of a finite set of closed sets is closed in the topology.

We close this lecture with some questions:

(1) Does every variety have a rational parameterization?

Answer: No, but many of the varieties we study will have a rational parameterization.

(2) Given the parametric representation, can we always find the implicit representation?

Answer: Technically, yes. It is always possible, but it is often not computationally feasible.

[7]Recall that a rational function is a quotient of polynomial functions. That is, for $f, g \in k[x_1, \ldots, k_n]$, $\frac{f}{g}$ is a f

Lecture 4

Assigned Reading

The reading for this lecture is §4 of chapter 1 in [**?**].

Two important topics will be ideals and term order. In this lecture, we focus on ideals.

Definition 4.1. *A subset $I \subset k[x_1, \ldots, x_n]$ is an* **ideal** *if*

(1) $0 \in I$,
(2) *for* $f_1, f_2 \in I$, $f_1 + f_2 \in I$, *and*
(3) *for* $f \in I$ *and* $g \in k[x_1, \ldots, x_n]$, $fg \in I$.

Given a set of polynomials in $k[x_1, \ldots, x_n]$, we can consider the smallest ideal containing those polynomials. That ideal is referred to as the ideal generated by them.

Definition 4.2. *Let* $\{f_1, \ldots, f_s\} \subset k[x_1, \ldots, x_n]$. *The* **ideal generated by** $\{f_1, \ldots, f_s\}$ *is*

$$\langle f_1, \ldots, f_s \rangle = \left\{ \sum_{i=1}^{s} h_i f_i \mid h_i \in k[x_1, \ldots, x_n] \right\}$$

We can easily show that $\langle f_1, \ldots, f_s \rangle$ is an ideal according to Definition **??**.

(1) $0 \in \langle f_1, \ldots, f_s \rangle$ since $0 \in k[x_1, \ldots, x_n]$ and $0 = 0 \cdot f_1 + \ldots + 0 \cdot f_n$.
(2) For $g_1, g_2 \in \langle f_1, \ldots, f_s \rangle$, we know

$$g_1 = h_1 f_1 + \ldots h_s f_s \text{ and } g_2 = k_1 f_1 + \ldots + k_s f_s.$$

A very useful fact to know is that every ideal in a polynomial ring can finitely generated. This is known as **Hilbert's Basis Theorem**.

Within linear algebra, where we have drawn much of our intuition so far, the subsets of interest have been vector subspaces. These are subsets that maintain important algebraic structures that are of interest in linear algebra. Here, we are working within the field of commutative algebra, which is the backbone of algebraic geometry. The subsets of interest are ideals. So there exists a rough analogy between these two subsets.

Commutative Algebra	**Linear Algebra**
Ideal	*Vector Subspace*
Closed under addition	Closed under addition
Closed under polynomial multiplication	Closed under field element multiplication
Generated by "linear" combos of polys over poly ring[8]	Generated by linear combos of vecs over field
A kernel of a ring homomorphism	A null space of a linear transformation

There can be more than one set of generators for an ideal. This makes the notion of a minimal generating set require some thought.

For example, we could ask that a minimal generating set be "inclusion minimal," meaning that if I delete any one generator, the set no longer generates the same ideal. Unfortunately, an ideal can have multiple inclusion minimal sets, each having a different number of elements.

When you are lucky, the number of generators of an ideal will equal the codimension of the variety. That is, every time you add an equation, we reduce the dimension of the variety by one. Think back to the rank-nullity theorem for intuition regarding why.

Not only does minimality require some thought, "best" generating sets will as well. One can pick different generating sets which allow one to do different kinds of calculations

Example 9. Let $f_1 = 2x^2 + 3y^2 - 11$, $f_2 = x^2 + y^2 - 3$, $g_1 = x^2 + 2$ and $g_2 = y^2 - 5$. Consider the ideal $\langle f_1, f_2 \rangle$. It turns out to be equivalent to $\langle g_1, g_2 \rangle$. Why? Well, if we can use one set of generators to construct the other and vice versa, then we know the two ideals contain one another and, hence, are equal. For example, $f_1 - 2f_2 = g_1$. The remaining are left up to the reader.

One could compare two ideals using SAGE. For the above example, the code would be the following:

```
1 ------------------------------------------------------------
2 P.<x,y> = PolynomialRing(QQ,2);
3 I=ideal(2*x^2 + 3*y^2 - 11, x^2 + y^2 - 3);
4 J=ideal(x^2 + 2, y^2 - 5);
5 I == J
6 True
7 ------------------------------------------------------------
```

Two generating sets for the same ideal will have the same zero set. That is, if

$$\langle f_1, \ldots, f_s \rangle = \langle g_1, \ldots, g_t \rangle = I$$

then

$$V(f_1, \ldots, f_s) = V(g_1, \ldots, g_t) = V(I).$$

4.1 Zariski Closure. Let us recall the idea of Zariski closure from the last lecture. We begin with a set S in an affine space k^n. In particular, we are thinking of S as an image of some parameterization map, φ. We can think of this map as $\varphi : k^m \to k^n$.

This set may be a variety or may not (in the case of a rational parameterization).

Definition 4.3. *Let $S \subset k^n$. Define*

$$I(S) = \{f \in k[x_1, \ldots, x_n] \mid f(a_1, \ldots, a_n) = 0 \textit{ for all } (a_1, ldots, a_n) \in S\}$$

to be the ideal corresponding to S.

We can easily show that this is an ideal. With this ideal, we can define the ZAriski closure (can be found on page 193 in [**?**]).

Definition 4.4. *Let $S \subset k^n$. The Zariski closure $\bar{S}$ of S is the variety of the ideal of S:*

$$V(I(S))$$

In other words, we consider all the polynomials whose zero set contains S. Then we consider the variety associated to that ideal. This may pick up more points.

Example 10. If our starting set is a variety V, then

$$V(I(V)) = V.$$

Recall that a variety has a defining set of polynomials. These polynomials generate the ideal $I(V)$. So their corresponding variety will be the same set. No extra points are picked up.

Example 11. $I(V(I)) \neq I$ in general. Consider $\langle x^2 \rangle$. This is the ideal of polynomials divisible by x^2. Notice that $x \notin \langle x^2 \rangle$. The variety corresponding to this idea is $\{0\}$.

Now consider all the polynomials which vanish on this set. It is easy to see that x is one of them. In fact, $I(V(I)) = \langle x \rangle$. The proof is left to the reader.

Finally, we observe that over an algebraically closed field, the relationship between $\langle f_1, \ldots, f_s \rangle$ and $I(V(f_1, \ldots, f_s))$ is easily characterized.

Lemma 4.5. $\langle f_1, \ldots, f_s \rangle \subseteq I(V(f_1, \ldots, f_s))$

We can characterize this relationship further using the Nullstellensatz Theorem, which will be discussed in a later lecture.

Lecture 5

Assigned Reading

The reading for this lecture is §5 of chapter 1 in [**?**].

Let us begin with some questions from the last class.

(1) Can every ideal in a polynomial ring be written as $\langle f_1, \ldots, f_s \rangle$?
(2) If so, how do I find this finite set?
(3) How can I determine ideal membership? (e.g. $f \in \langle f_1, \ldots, f_s \rangle$?)

This third question turns out to be incredibly hard.

Recall from the last lecture the Nullstellensatz theorem. Given $f_1, \ldots, f_s \in k[x_1, \ldots, x_n]$ this theorem relates $\langle f_1, \ldots, f_s \rangle$ and $I(V(f_1, \ldots, f_s))$. We already know one is contained in the other, but the Nullstellensatz will provide a more precise characterization of their relationship. In particular, what more do we need in $\langle f_1, \ldots, f_s \rangle$ to get $I(V(f_1, \ldots, f_s))$?

In this lecture, we will not answer this question. Instead, we will discuss Buchberger's algorithm that can help motivate the ideas behind the Nullstellensatz. This algorithm has two familiar special cases:

(1) The Euclidean division algorithms for univariate polynomials.
(2) Gaussian elimination.

Today, we will focus on the division algorithm. Let $f = a_0x^m + a_1x^{m-1} + \ldots + a_{m-1}x + a_m$. We call the first term in this polynomial the leading term because it has highest degree and a nonzero coefficient.

$$LT(f) = a_0x^m$$

Notice that in the multivariate case, it is not always obvious what term should be the leading term. This is because in the univariate case, it is completely characterized by degree. That is,

$$\deg(f) \leq \deg(\hat{f}) \iff LT(f) \big| LT(\hat{f}).$$

To do this, we are using the underlying field structure of the coefficients.

Univariate polynomial division works in the following way. Every $f \in k[x]$ given g can be written as

$$f(x) = q(x)g(x) + r(x)$$

such that $\deg(r) < \deg(g)$. In this case, q and r are unique and there is an algorithm for finding both of them. Below is the pseudocode for this process.

$$\underline{\texttt{Input}}\texttt{:} \quad \texttt{g, f}$$

$$\underline{\texttt{Output}}\texttt{:} \quad \texttt{q, r}$$

$$\texttt{Set q} := \texttt{0, r} := \texttt{f.} \quad \texttt{While r} \neq \texttt{0 and LT(g)} \big| \texttt{LT(r)}$$

$$\texttt{q} =: \texttt{q} + \frac{\texttt{LT(r)}}{\texttt{LT(g)}}$$

$$\mathtt{r} =: \mathtt{r} - \frac{\mathtt{LT(r)}}{\mathtt{LT(g)}}\mathtt{g}$$

This algorithm slowly chips away at the remainder. Let us see an example of this process.

$$\underline{\texttt{Input}}\texttt{:} \quad \mathtt{g} = \mathtt{x}^2 + 5, \mathtt{f} = 3\mathtt{x}^3 + 2\mathtt{x} + 7$$

$$\texttt{Let } \mathtt{q} := 0, \ \mathtt{r} := 3\mathtt{x}^3 + 2\mathtt{x} + 7. \quad \texttt{Since } \mathtt{r} \neq 0 \texttt{ and } \mathtt{LT(g)} \big| \mathtt{LT(r)}$$

$$\mathtt{q} =: 0 + \frac{3\mathtt{x}^3}{\mathtt{x}^2} = 3\mathtt{x}$$

$$\mathtt{r} =: 3\mathtt{x}^3 + 2\mathtt{x} + 7 - (3\mathtt{x})(\mathtt{x}^2 + 5) = -13\mathtt{x} + 7$$

$$\texttt{STOP } \underline{\texttt{Output}}\texttt{:} \quad \mathtt{q(x)} = 3\mathtt{x}, \mathtt{r(x)} = -13\mathtt{x} + 7$$

It is left to the reader to check that r, q are unique,

Remark 5.1. If you have never seen the division algorithm in an upper-division math course, know that it is precisely the same algorithm taught in k-12 schools but written differently.

$$\begin{array}{r} \qquad\qquad 3x \\ x^2+5\big) \overline{\quad 3x^3 \ +2x+7} \\ \underline{-\,3x^3 \ -15x} \\ \qquad -\,13x \end{array} \qquad \begin{array}{l} \big(\quad 3x^3 \ +2x+7 \big) \div \big(x^2+5 \big) = 3x + \dfrac{-13x+7}{x^2+5} \\ \underline{-\,3x^3 \ -15x} \\ \qquad\qquad -\,13x \end{array}$$

In the "traditional" style of division, we focus on the leading term and push remainders over. A moments thought will convince you that the same process is happening with the algorithm above.

With this algorithm in mind, consider an arbitrary ideal in $k[x]$. How many polynomials does one need to represent a given ideal? The answer is one.

Corollary 5.2. *Every ideal in $k[x]$ is of the form $\langle f \rangle$.*

A ring in which every ideal is generated by one element is called a principle ideal domain (often called a PID for short). Another example of such a ring is the integers.

Proof. Fix $I \in k[x]$. Take $f \in I$ a nonzero polynomial of minimal degree. It is clear that $\langle f \rangle \subset I$. Now consider some $h \in I$. Consider $f|h$. In other words, consider

$$h(x) = q(x)f(x) + r(x).$$

If $r(x) = 0$, then we are done. Suppose not. Since $f, h \in I$, it follows that $r \in I$. But according to the division algorithm $\deg(r) < \deg(f)$, which contradicts the assumption that f has minimal degree. □

Given an ideal in $k[x]$ with multiple generators, the minimal degree polynomial may be smaller in degree than the other generators. Consider the following examples

Example 12. Consider $I = \langle 3x^5 - x^2, x^5 \rangle$. Notice that $3(x^5) - (3x^5 - x^2) = x^2 \in I$. In fact, $I = \langle x^2 \rangle$.

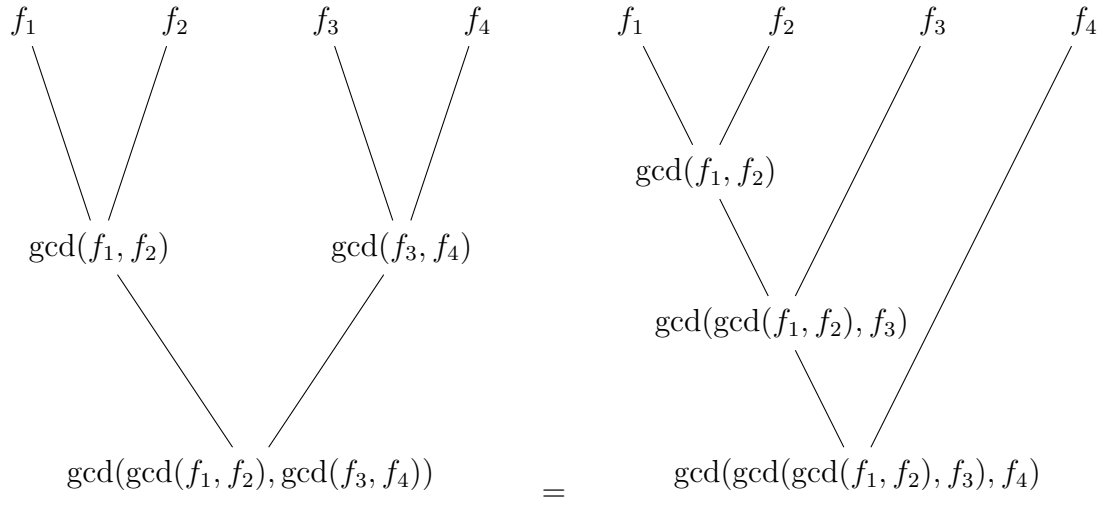
Example 13. Consider $I = \langle x^4 - 1, x^6 - 1 \rangle$. Notice that $(x^6 - 1) - x^2(x^4 - 1) = x^2 - 1 \in I$. In fact, $I = \langle x^2 - 1 \rangle$.

The idea behind these examples is the greatest common divisor. The greatest common divisor of a set of polynomials

$$\gcd(f_1, \ldots, f_s),\ f_i \in k[x]$$

is a polynomial $h \in k[x]$ such that $h \big| f_i$ for all $1 \leq i \leq s$ and if there exists a $p \in k[x]$ such that $p \big| f_i$ for all i, then $h \big| p$. This h will be unique up to a field constant and will generate the idea $\langle f_1, \ldots, f_s \rangle$.

To find the GCD of a set of functions, it is enough to consider them pairwise. The actual ordering does not matter. Both approaches pictures below will yield the same result.



Here is how the GCD algorithm works:

$$\underline{\texttt{Input}}\texttt{:}\quad \texttt{f}, \texttt{g}$$

$$\underline{\texttt{Output}}\texttt{:}\quad \gcd \texttt{f}, \texttt{g}$$

$\texttt{h} := \texttt{f}\ \ \texttt{s} := \texttt{g}$ `While` $\texttt{s} \neq 0$ `set` $\texttt{rem} :=$ `remainder` `r` `when we write` $\texttt{h} + \texttt{qs} + \texttt{r}$

$\texttt{h} := \texttt{s}\ \ \texttt{s} := \texttt{rem}$

Why does this process work? Consider an ideal generated by f, g. Let q, r be such that $f = qg + r$. Then

$$\langle f, g \rangle = \langle f - qg, g \rangle = \langle r, g \rangle .$$

If g is not a multiple of r, we can continue this process.

Lecture 6

Assigned Reading

The reading for this lecture is §1 & §2 of chapter 2 in [**?**].

Now we plan to develop some of the computation tools that generalize the division algorithm and come up with nice bases. In higher dimensions, our ideals won't necessarily be principle so our basis will be more complex. Therefore, we will need a division algorithm that has multiple divisors as well as a technique for ordering monomials.

Given the ring $k[x_1, \ldots, x_m]$, we can think of this as a vector space. A possible spanning set might be the set of monomials, i.e. elements of the form

$$x_1^{i_1} \ldots x_m^{i_m}, \quad (i_1, \ldots, i_n) \in \mathbb{Z}_{\geq 0}$$

What is a possible order for this set?

Definition 6.1. *A* **monomial order** *is a total order $<$ on these monomials such that*

(1) $1 = x_1^0 \ldots x_m^0$ *is smaller than every other monomial*

(2) *for some positive shift* $(k_1, \ldots, k_m)$*,*

$$x_1^{i_1} \ldots x_1^{i_m} < x_1^{j_1} \ldots x_1^{j_m} \implies x_1^{i_1+k_1} \ldots x_1^{i_m+k_m} < x_1^{j_1+k_1} \ldots x_1^{j_m+k_m}.$$

There are many orders which satisfy this criteria. One such way is *lexicographic order*, which discussed in the homework. It states that that

$$x_1^{i_1} \ldots x_1^{i_m} < x_1^{j_1} \ldots x_1^{j_m}$$

if the left-most entry in

$$(i_1 - j_1, \ldots, i_m - j_m)$$

is positive.

Example 14. $x_1 > x_2^{10}$ since $x_1 = x_1 x_2^0$ and $x_2^{10} = x_1^0 x_2^{10}$.

Example 15. $x_3 x_4^2 x_5 < x_2 x_3$ for similar reason. That is, $x_3 x_4^2 x_5 = x_1^0 x_2^0 x_3^1 x_4^2 x_5^1$ and $x_2 x_3 = x_1^0 x_2^1 x_3^1 x_4^0 x_5^0$.

This particular order is useful when trying to eliminate a particular variable; however, it is very computationally expensive.

Given a method of term order, every polynomial has a leading term, denoted by $LT_<(f)$. For a polynomial like

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

the leading term is $LT_<(f) = 3x_1^2 x_2$. The initial monomial is the leading term divided by its coefficient. Here, it is $in_<(f) = x_1^2 x_2$.

Given a polynomial ideal I in $k[x_1, \ldots, x_n]$, we can define the **initial ideal** of I with respect to the ordering chosen is

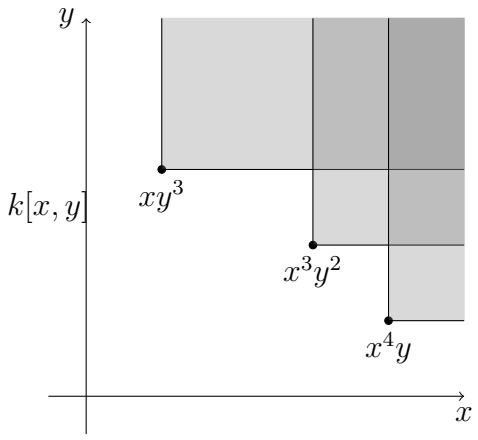
$$in_<(I) = \{in_<(f) \ : \ f \in I\}.$$

Notice that we construct this ideal not just from the generators. This ideal gives us a lot of information about the variety.

Another type of ideal is a **monomial ideal**. As the name suggests, it is an ideal generated by monomials. In particular, let $A \subset \mathbb{Z}^n_{\geq 0}$ and let

$$I = \langle x^\alpha \ : \ \alpha \in A \rangle = \left\{ \sum_{\alpha \in A} h_\alpha x^\alpha \ \middle| \ h_\alpha \in k[x_1, \ldots, x_n] \right\}.$$

We can represent monomial ideals by plotting the monomials and considering the lattice points above.



For a monomial ideal I, the following are equivalent:

- $f \in I$
- every term of f is in I
- f is a k-linear combination of lattice points

Suppose $f_1, \ldots, f_s \in k[x_1, \ldots, x_n]$ generate an ideal $I = \langle f_1, \ldots, f_s \rangle$. A natural question to ask is when does $in_< I$ equal $\langle in_< f_1, \ldots, in_< f_s \rangle$?

Exercise 1. Find a counter example.

A **Gröbner basis** is a set of generators for an ideal I such that $in_< I = \langle in_< f_1, \ldots, in_< f_s \rangle$ for I.

Example 16. $\langle x^2, y^2 \rangle$

Example 17. $\langle x^2, y^2, x^2 + y^2 \rangle$

Definition 6.2. *A* **reduced Gröbner basis** *is a Gröwner basis G such that*

(1) *The coefficients of each $LT_<(g)$ is 1 for each $g \in G$.*
(2) *The generating set of $in_< I$, $\{in_< g \ : \ g \in G\}$, is (inclusion) minimal.*
(3) *No trailing term of some $g \in G$ lies in the initial ideal.*

Lemma 6.3. (Dickson's Lemma) *Every monomial ideal has a finite monomial basis.*

Proof. Let $I \subset k[x_1, \ldots, x_n]$. For $n = 1$, then the ring is a principal ideal domain.

Now assume this holds for $n - 1$. Define $I \subset k[x_1, \ldots, x_{n-1}, y]$. Let

$$J = \langle x^\alpha \ : \ x^\alpha y^m \in I, \text{ for some } m \geq 0 \rangle.$$

By assumption $J = \left\langle x^{\alpha(i)}, \ldots, x^{\alpha(s)} \right\rangle$. Then, there is some M such that $x^{\alpha(i)} y^M \in I$ for all $1 \leq i \leq s$. Let J_k be the ideals

$$J_k = \left\langle x^\beta \ : \ x^\beta y^k \in I \right\rangle$$

for $0 \leq k \leq M-1$. Each J_k is finitely generated. That is,

$$J_k = \left\langle x^{\alpha_k(1)}, \ldots, x^{\alpha_k(n)} \right\rangle.$$

Claim: I is generated by

$$\begin{array}{ccccc} x^{\alpha(1)} y^M, & \ldots, & x^{\alpha(s)} y^M & & \text{(from } J\text{)} \\ x^{\alpha_0(1)} y^M, & \ldots, & x^{\alpha_0(s)} y^M & & \text{(from } J_1\text{)} \\ \vdots & & \vdots & & \\ x^{\alpha_{M-1}(1)} y^M, & \ldots, & x^{\alpha_{M-1}(s)} y^M & & \text{(from } J_M\text{)} \end{array}$$

Consider $x^\alpha y^p \in I$.

- If $p \geq M$, then some $x^{\alpha(i)} y^M \big| x^\alpha y^p$.
- If $p \leq M-1$, then some $x^{\alpha_p(i)} y^p \big| x^\alpha y^p$.

This, I is finitely generated. □

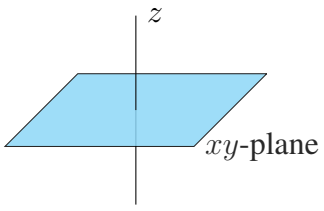
The set generated for the proof above is far bigger than necessary. See [**?**] for more on this.

We have enough information for a working definition for a dimension of a variety. Here are some the answers you gave in your homework regarding dimension:

- Consider the dimension of a related linear space.
  This is a good idea. In particular, the related linear space we want is the tangent space, which make sense for ideals over $\mathbb{C}^n$.
- A notion of a basis for the ideal.
  This is also a good idea and is essentially the Krull dimension.
- The number of unique/distinct nonredundant parameters.
  There are ways to make this equivalent to the definition.
- The dimension of the largest component
  This is correct, although it avoids the important discussion on how one determines the dimension of the largest component.
- Parameterize the variety and count the number of parameters
  This is a good idea and will more or less work.
- The highest dimensional ball that fits in the variety
  This is very similar to the first one listed.

Basically, all of these answers are correct. They are just different ways of looking at the dimension.

To motivate the discussion, let us consider the union $V(x,y) \cup V(z) = V(xz, yz)$.



Any polynomial that only contains the variables x and y will have zeros completely contained in the variety. If, however, a polynomials involves z, some restrictions are needed.

It turns out an idea related to this gives rise to the first working definition of dimension. Consider an ideal with a chosen term order.

Definition 6.4. *A monomial x^α is called* **standard for** *I if $x^\alpha \notin in_< I$.*

Question 6.5. When is the number of standard monomials finite?

Now let us consider how we construct these standard polynomials.

Definition 6.6. *A subset $s \subset \{x_1, \ldots, x_n\}$ is a standard set for I under a chosen order $>$ if every monomial constructible from S,*

$$\prod_{x_i \in S} x_i^{a_i},$$

is not in the initial ideal.

By considering the standard set, we can construct a definition for dimension.

Theorem 6.7. *The dimension of an affine algebraic variety $V(I) \subset \mathbb{C}^m$ us the maximal cardinality of a standard set for the ideal I.*

We will discuss this theorem further next time.

Lecture 7

Assigned Reading

The reading for this lecture is §2 & §3 of chapter 2 in [**?**].

Recall from Lecture **??** the definition of monomial ordering, which required that it satisfy the following:

(1) 1 is the smallest monomial,
(2) $\geq$ is a total order, and
(3) the ordering is invariant under shifts,

$$x_1^{i_1} \dots x_1^{i_m} \leq x_1^{j_1} \dots x_1^{j_m} \implies x_1^{i_1+k_1} \dots x_1^{i_m+k_m} \leq x_1^{j_1+k_1} \dots x_1^{j_m+k_m}.$$

The definition in [**?**] requires that $>$ is a *well ordering* on $\mathbb{Z}_{\geq 0}^n$.

Exercise 2. Show that the well-ordering requirement is not needed with the definition provided in class.

Remark 7.1. In order to do this exercise, it benefits us to know the definitions of these ordering terms.

Definition 7.2. *Let X be* **totally ordered** *under $\leq$. Then for all $a, b, c \in X$, the following statements hold:*

- *(***antisymmetry***) If $a \leq b$ and $b \leq a$, then $a = b$.*
- *(***transitivity***) If $a \leq b$ and $b \leq c$, then $a \leq c$.*
- *(***totality***) For $a, b \in X$, then $a \leq b$ or $b \leq a$.*

Definition 7.3. *A* **well-ordering** *on a set X is a total order on X such that every nonempty subset has a least element (by the ordering).*

In Lecture **??**, the definition of lexicographic order was introduced. This was an ordering such that $x^\alpha >_{lex} x^\beta$ if and only if $\alpha - \beta$ has a positive number as its left-most nonzero entry.[9]

Question 7.4. How many possible lexicographic orderings are there?

The answer is $n!$. Consider when there are only three variables, x, y, and z. We can order them as (x, y, z), but we can also order by some permutation, (y, z, x) or (z, y, x). The total number of ways to order three variables is $3!$.

There exist order methods of term order. For example, the **graded lex order** uses the total degree as an initial (partial order). If the total degree of two terms is equal, then a lexicographic order is used. That is, $\alpha >_{grlex} \beta$ if and only if

$$|\alpha| > |\beta| \quad \text{or} \quad |\alpha| = |\beta| \text{ and } \alpha >_{lex} \beta.$$

Another monomial ordering is **reverse lexicographic ordering**. As the name would suggest, this is a dual notion of the lexicographic ordering. Here, $x^\alpha >_{revlex} x^\beta$ if and only if $\alpha - \beta$ has a negative number as its right-most nonzero entry.

[9] Here, by x^α, we mean $x_1^{a_1} \cdot \ldots \cdot x_n^{alpha_n}$.

At first glance, $>_{lex}$ and $>_{revlex}$ may seem to be an identical ordering, but there are a few cases in which they differ.

Example 18. Consider the monomials $x4y^2z^4$ and x^5z^5. Observe that

$$x^5z^5 >_{lex} x^4y^2z^4$$

but

$$x^4y^2z^4 >_{revlex} x^5z^5.$$

Just as there is a graded lexicographic order, there is a graded reverse lexicographic ordering which works as you expect–order first by total degree, then by reverse lexicographic order.

Remark 7.5. For the initial monomial term, there are two ways we can denote this: $in_>(f)$, which was discussed during the last lecture, and $LM(f)$, which is used in the book to denote the leading monomial.

Each of these choices of monomial order impact how the division algorithm works. As we saw earlier in the course, the division algorithm gives us a solution to the ideal membership problem for ideals in $k[x]$.

7.1 Division Algorithm for Linear Systems. Before we define the division algorithm in high dimensional space, let us first restrict our discussion to linear polynomials. This is just Gaussian elimination.

Example 19. Consider the linear system

$$\begin{aligned} 2x_1 + 3x_2 - x_3 &= 0 \\ x_1 + x_2 - 1 &= 0 \\ x_1 + x_3 - 3 &= 0 \end{aligned}$$

Here, we're given a system of linear equations and we're searching for a solution; in other words, we are looking for a parameterization of the solution set. So we write the system as a matrix.

$$\left(\begin{array}{ccc|c} 2 & 3 & -1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 3 \end{array}\right) \mapsto \left(\begin{array}{ccc|c} 1 & 0 & 1 & 3 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Row reduction yields the parameterization $x_3 = t$, $x_1 + t = 3$ and $x_2 - t = -2$ or

$$\begin{cases} x_1 = 3 - t \\ x_2 = t - 2 \\ x_3 = t \end{cases}$$

The above example tells us how we go from the implicit form of the variety to the parametric form. As we saw in the problem 1 of homework 2, it is important to know how to go from the parametric form to the implicit form. The method for the linear case works similarly.

Input:

$$\begin{array}{ccccccccc} x_1 & = & a_{11}t_1 & + & \ldots & + & a_{1m}t_m & + & b_1 \\ & & \vdots & & & & & & \\ x_n & = & a_{n1}t_1 & + & \ldots & + & a_{nm}t_m & + & b_n \end{array}$$

Output: Implicit equations for the space. Algorithm: Write input in augmented matrix form.

$$\left(\begin{array}{ccc|cccc|c} a_{11} & \dots & a_{1m} & -1 & 0 & \dots & 0 & b_1 \\ a_{21} & \dots & a_{2m} & 0 & -1 & \dots & 0 & b_2 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & & \vdots \\ a_{n1} & \dots & a_{nm} & 0 & 0 & \dots & -1 & b_n \end{array}\right)$$

Row reductions are performed until it is in reduced echelon form. The implicitization (highlighted in yellow) is returned.

$$\left(\begin{array}{cccc|cccc|ccc|c} 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & * & \dots & * & * \\ 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 & * & \dots & * & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & * & \dots & * & * \\ 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & * & \dots & * & * \\ 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & * & \dots & * & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & * & \dots & * & * \end{array}\right)$$

Remark 7.6. For example of this process see warm up problem 3 of Session 3.

Question 7.7. When $n = m$ but a nontrivial implicitization can be found with this method, what does that tell us about the matrix $A = (a_{ij})$?

The matrix A has rank less than n (i.e. it is not invertible).

7.2 Division Algorithm for Polynomials. Now let us consider how the division algorithm should work in general. Let $R = k[x_1, \dots, x_n]$. Let $f \in R$ and $\{f_1, \dots, f_s\} \subset R$ (usually a generating set of an ideal). When performing the division algorithm, we wish to find $a_1, \dots, a_s, r \in k[x_1, \dots, x_n]$ such that no term of r is divisible by $LT(f_i)$ and

$$f = a_1 f_1 + \dots + a_s f_s + r.$$

Here is the pseudocode for the algorithm:

Input: $\mathtt{f}, \mathtt{f_1}, \dots, \mathtt{f_s} \in \mathtt{k}[\mathtt{x_1}, \dots, \mathtt{x_n}]$

Output: $\mathtt{a_1}, \dots, \mathtt{a_s}, \mathtt{r} \in \mathtt{k}[\mathtt{x_1}, \dots, \mathtt{x_n}]$

Initialize: Set $\mathtt{a_1} = 0, \dots, \mathtt{a_s} = 0, \mathtt{r} = 0$. Define p and let $\mathtt{p} = \mathtt{f}$.

While $\mathtt{p} \neq 0$
 for $\mathtt{i} = 1, \dots, \mathtt{s}$
  if $\mathtt{LT(f_i)} \big| \mathtt{LT(p)}$,
   set $\mathtt{a_i} := \mathtt{a_i} + \dfrac{\mathtt{LT(p)}}{\mathtt{LT(f_i)}}$
   $\mathtt{p} := \mathtt{p} - \dfrac{\mathtt{LT(p)}}{\mathtt{LT(f_i)}} \mathtt{f_i}$
   set flag = True
   break

```
if flag = False
  r := r + LT(p)
  p := p − LT(p)
```

In the next homework (4) assignment, you will be asked to implement this code in the language of your choice. If you use SAGE, you will need to set this up for an arbitrary polynomial ring. If you use another language, you can fix $n = 4$.

We now compute a small example where $n = 2$.

Example 20. Let $f = x^2y + xy^2 + xy + x + y + 1$ and let our generating set be $\langle x^2 + 1, y^2 + 1\rangle$. Let us use the graded lexicographic ordering on terms where $x > y$. We begin by dividing by $x^2 + 1$ and get

$$f = (y)(x^2 + 1) + (xy^2 + xy + x + 1).$$

Next, we divide by $y^2 + 1$ and get

$$f = (y)(x^2 + 1) + (x)(y^2 + 1) + (xy + 1).$$

Thus, $a_1 = y, a_2 = x$, and $r = xy + 1$.

Exercise 3. Construct an example where changing the choice of term order changes the resulting answers for $a_1, \ldots, a_s$ and r.

Lecture 8

Assigned Reading

The reading for this lecture is §4, §5 and §6 of chapter 2 in [**?**].

We begin with a challenging problem to consider (see [**?**] for similar problems).

Question 8.1. Consider problem 1 of homework 2. What is the implicitization when there are three or four coins?

Now suppose there are 2 piles of coins plus a bonus coin. Suppose you flip the bonus coin to determine which pile you randomly draw from. What is the implicitization of this set-up when there are three coins per pile? What is the implicitization for four coins per pile?

Here is a hint for the following problem: There is a trick to implicitization when considering the properties of the matrix of exponents. For example, for problem 2 of homework 2, this matrix would be

$$\begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix}$$

for the parametric equations $x_1 = \theta_1^2$, $x_2 = \theta_1\theta_2$, and $x_3 = \theta_2^2$.

Recall from Lecture **??** Dickson's Lemma, which states that every monomial ideal $I = \langle x^\alpha \ : \ \alpha \in A \rangle \subseteq k[x_1, \ldots, x_n]$ is finitely generated $I = \left\langle x^{\alpha(1)}, \ldots, x^{\alpha(s)} \right\rangle$ where $\alpha(i) \in A$.

For a given ideal I, its initial ideal is a monomial ideal (by definition)

$$in_< I = \langle LT(f) \ : \ f \in I \rangle .$$

From Dickson's lemma, it is easy to see that $in_< I$ must be finitely generated.

It is always true that

$$in_< \langle f_1, \ldots, f_s \rangle \subset \langle LT(f_1), \ldots, LT(f_s) \rangle$$

but often

$$in_< \langle f_1, \ldots, f_s \rangle \neq \langle LT(f_1), \ldots, LT(f_s) \rangle .$$

When equality does hold, then the set of generators is called a Gröbner basis.

Notation 8.2. Recall that $LT(f)$ is the leading term of the polynomial f according to a chosen term ordering. $in_<(f)$ is also the leading term of f according to some ordering, but with the field element coefficient of 1. That if, if $f = 3x^2 + 2x + 5$, then $LT(f) = 3x^2$ and $in_<(f) = x^2$. In the discussion above, using $LT(f)$ and $in_<(f)$ is equivalent when generating a polynomial ideal.

It is common to also use $LM(f)$ in place of $in_<(f)$ to refer to the *initial term* as the *leading monomial*.

In Lecture **??**, recall Hilbert's Basis Theorem, which states that every ideal I in $k[x_1, \ldots, x_n]$ is finitely generated. One of the goals for this lecture is to prove this theorem using Dickson's Lemma. Before we can, however, we need one more fact.

Lemma 8.3. *Let I be a (finitely generated) monomial ring: $I = \left\langle x^{\alpha(1)}, \ldots, x^{\alpha(s)} \right\rangle$. Then for any $x^\beta \in I$, there exists a $x^{\alpha(i)}$ such that $x^{\alpha(i)} \big| x^\beta$.*

We now prove Hilbert's Basis Theorem:

Proof. Let $I \in k[x_1, \ldots, x_n]$ and fix a term ordering. Then

$$in_< I = \langle \{LT(f) \ : \ f \in I\} \rangle .$$

By Dickson's Lemma, we can choose $g_1, \ldots, g_t \in I$ suc that

$$in_< I = \langle LT(g_1), \ldots, LT(g_t)\} \rangle .$$

Claim: $I = \langle g_1, \ldots, g_t \rangle$. To show this, consider an arbitrary $f \in I$. By the division algorithm,

$$f = a_1 f_1 + \ldots + a_t g_t + r$$

such that no term of r is divisible by $LT(g_i)$ for any i. If $r = 0$, we are done. So suppose not. Then

$$r = f - a_1 f_1 - \ldots - a_t g_t \in I.$$

Since $r \neq 0$, we know $LT(r) \in in_< I$. By Lemma **??**, there exists a g_i in the set such that $LT(g_i) \big| LT(r)$, which is a contradiction. □

Hilbert's proof was drastically different, primarily because many of the facts we used here were found much later. The paper [**?**] is worth reading. It not only introduces Hilbert's Basis Theorem, but also the Hilbert function, Hilbert polynomial, and Hilbert series–all of which are still used today.

An equivalent condition to the Hilbert basis theorem is known as the ascending chain condition (ACC), which is commonly discussed in commutative algebra.

Definition 8.4. *A set of nested ideals*

$$I_1 \subseteq I_2 \subseteq \ldots$$

satisfy the **ascending chain condition** *if there exists an n such that $I_n = I_{n+k}$ for all $k \in \mathbb{N}$.*

Remark 8.5. Rings which satisfy ACC are called **Noetherian**. By proving Hilbert's basis theorem, we have show that all polynomial rings over a field (with finitely many variables) is Noetherian.

8.1 The Geometry Behind the Division Algorithm. Let f be an arbitrary polynomial in $k[x_1, \ldots, x_n]$ and let $V(\langle f_1, \ldots, f_s \rangle)$ where $\{f_1, \ldots, f_s\}$ is a Gröbner basis for the ideal it generates.

A natural question to ask is: are the zeros of f contained in the variety? Put another way, is f in the ideal $\langle f_1, \ldots, f_s \rangle$?

The way to understand "how far" f is from the ideal is by considering its remainder when dividing by the generating set:

$$f = a_1 f_1 + \ldots + a_s f_s + r.$$

If $r = 0$, then $f \in \langle f_1, \ldots, f_s \rangle$. If $r \neq 0$, then this tells us f is not in the ideal. Moreover, $f \equiv r$ in the quotient ring $k[x_1, \ldots, x_n]/I$. For this reason, the remainder r is often referred to as the **normal form** of f and is denoted $NF(f, I)$.

Remark 8.6. It is important in this set up that the generating set be a Gröwner basis. Otherwise, if $r \neq 0$ it is still possible that $f \in I$.

For example, consider the ideal we saw in homework 2: $\langle x + xy, y + xy, x^2, y^2 \rangle = \langle x, y \rangle$. Suppose we use the former generating set (which is clearly not a Gröbner basis) and consider the division algorithm for $f = x$. Assume we use the graded lexicographic ordering with $x > y$. Then

$$x = 0(x + xy) + 0(y + xy) + 0(x^2) + 0(y^2) + x,$$

so the remainder $r = x$ is nonzero. But we know x is in the ideal.

Proposition 8.7. *Let I have $\{g_1, \ldots, g_t\}$ as a (reduced) Gröbner basis. For all $f \in k[x_1, \ldots, x_n]$, there exists a unique normal form $r \in k[x_1, \ldots, x_n]$ such that*

- *no term of r is visible by any $LT(g_i)$, and*
- *there exists a $g \in I$ such that $f = g + r$.*

We call g the **projection** *of f onto I.*

In particular, when we use the division algorithm, we get an expression

$$f = \underbrace{a_1 g_1 + \ldots + a_t g_t}_{g} + r$$

Key to all this discussion is the construction of a Gröbner basis. By definition, this is a generating set such that $in_> I = \langle LT(f_1), \ldots, LT(f_s) \rangle$. We can see that an arbitrary generating set is *not* a Gröbner when there exists a lower order leading term than any of those in the set $\{LT(f_1), \ldots, LT(f_s)\}$. In such a case, there is some kind of cancelation possible with monomials that reduces the degree of a leading term in the generating set:

$$\text{mdeg}(ax^\alpha f_i - bx^\beta f_j) < \min\{\text{mdeg}(f_i), \text{mdeg}(f_j)\}.$$

This corresponds to the least common multiple (LCM).

Definition 8.8. *Let f, g be nonzero polynomials.*

- *If* $\text{mdeg}(f) = \alpha$ *and* $\text{mdeg}(g) = \beta$ *then* $\gamma = (\gamma_1, \ldots, \gamma_n)$, *where* $\gamma_i = \max(\alpha_i, \beta_i)$ *for each i. We call $x^\gamma = LCM(in_>(f), in_>(g))$.*
- *The corresponding* **S-polynomial** *is*

$$\boxed{S(f,g) = \frac{x^\gamma}{LT(f)} \cdot f - \frac{x^\gamma}{LT(g)} \cdot g}$$

Notice the S-polynomial cancels out the leading term and finds the "hidden" generator.

Example 21. Consider the ideal $\langle x + xy, y + xy, x^2, y^2 \rangle$ and suppose we are uses the graded lexicographic ordering where $x > y$. Then $LCM(xy, xy) = xy$ and

$$\begin{aligned} S(xy + y, xy + x) &= \frac{xy}{xy} \cdot (xy + y) - \frac{xy}{xy} \cdot (xy + x) \\ &= y - x \end{aligned}$$

So we can change the basis to $\langle xy + y, x - y, x^2, y^2 \rangle$.

Question 8.9. Can we use S-polynomials to construct a Gröbner basis from a generating set?

Buchberger's Criteria tells us the connection between Gröbner bases and the S-polynomial.

Theorem 8.10. (Buchberger's Criterion) *Let I be a polynomial ideal. Then a basis $G = \{g_1, \ldots, g_t\}$ for I is a Gröbner basis for I if and only if for all pairs $i \neq j$, $S(g_i, g_j) = 0$.*

We will discuss them further on Monday.

Lecture 9

Assigned Reading

The reading for this lecture is §4, §5 and §6 of chapter 2 in [**?**].

We begin with a challenging problem to consider (see [**?**] for similar problems).

Question 9.1. Consider problem 1 of homework 2. What is the implicitization when there are three or four coins?

Now suppose there are 2 piles of coins plus a bonus coin. Suppose you flip the bonus coin to determine which pile you randomly draw from. What is the implicitization of this set-up when there are three coins per pile? What is the implicitization for four coins per pile?

Here is a hint for the following problem: There is a trick to implicitization when considering the properties of the matrix of exponents. For example, for problem 2 of homework 2, this matrix would be

$$\begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix}$$

for the parametric equations $x_1 = \theta_1^2$, $x_2 = \theta_1\theta_2$, and $x_3 = \theta_2^2$.

Recall from Lecture **??** Dickson's Lemma, which states that every monomial ideal $I = \langle x^\alpha \ : \ \alpha \in A \rangle \subseteq k[x_1, \ldots, x_n]$ is finitely generated $I = \left\langle x^{\alpha(1)}, \ldots, x^{\alpha(s)} \right\rangle$ where $\alpha(i) \in A$.

For a given ideal I, its initial ideal is a monomial ideal (by definition)

$$in_< I = \langle LT(f) \ : \ f \in I \rangle .$$

From Dickson's lemma, it is easy to see that $in_< I$ must be finitely generated.

It is always true that

$$in_< \langle f_1, \ldots, f_s \rangle \subset \langle LT(f_1), \ldots, LT(f_s) \rangle$$

but often

$$in_< \langle f_1, \ldots, f_s \rangle \neq \langle LT(f_1), \ldots, LT(f_s) \rangle .$$

When equality does hold, then the set of generators is called a Gröbner basis.

Notation 9.2. Recall that $LT(f)$ is the leading term of the polynomial f according to a chosen term ordering. $in_<(f)$ is also the leading term of f according to some ordering, but with the field element coefficient of 1. That if, if $f = 3x^2 + 2x + 5$, then $LT(f) = 3x^2$ and $in_<(f) = x^2$. In the discussion above, using $LT(f)$ and $in_<(f)$ is equivalent when generating a polynomial ideal.

It is common to also use $LM(f)$ in place of $in_<(f)$ to refer to the *initial term* as the *leading monomial*.

In Lecture **??**, recall Hilbert's Basis Theorem, which states that every ideal I in $k[x_1, \ldots, x_n]$ is finitely generated. One of the goals for this lecture is to prove this theorem using Dickson's Lemma. Before we can, however, we need one more fact.

Lemma 9.3. *Let I be a (finitely generated) monomial ring: $I = \left\langle x^{\alpha(1)}, \ldots, x^{\alpha(s)} \right\rangle$. Then for any $x^\beta \in I$, there exists a $x^{\alpha(i)}$ such that $x^{\alpha(i)} \big| x^\beta$.*

We now prove Hilbert's Basis Theorem:

Proof. Let $I \in k[x_1, \ldots, x_n]$ and fix a term ordering. Then

$$in_< I = \langle \{LT(f) \ : \ f \in I\} \rangle .$$

By Dickson's Lemma, we can choose $g_1, \ldots, g_t \in I$ suc that

$$in_< I = \langle LT(g_1), \ldots, LT(g_t)\} \rangle .$$

Claim: $I = \langle g_1, \ldots, g_t \rangle$. To show this, consider an arbitrary $f \in I$. By the division algorithm,

$$f = a_1 f_1 + \ldots + a_t g_t + r$$

such that no term of r is divisible by $LT(g_i)$ for any i. If $r = 0$, we are done. So suppose not. Then

$$r = f - a_1 f_1 - \ldots - a_t g_t \in I.$$

Since $r \neq 0$, we know $LT(r) \in in_< I$. By Lemma **??**, there exists a g_i in the set such that $LT(g_i) \big| LT(r)$, which is a contradiction. □

Hilbert's proof was drastically different, primarily because many of the facts we used here were found much later. The paper [**?**] is worth reading. It not only introduces Hilbert's Basis Theorem, but also the Hilbert function, Hilbert polynomial, and Hilbert series–all of which are still used today.

An equivalent condition to the Hilbert basis theorem is known as the ascending chain condition (ACC), which is commonly discussed in commutative algebra.

Definition 9.4. *A set of nested ideals*

$$I_1 \subseteq I_2 \subseteq \ldots$$

satisfy the **ascending chain condition** *if there exists an n such that $I_n = I_{n+k}$ for all $k \in \mathbb{N}$.*

Remark 9.5. Rings which satisfy ACC are called **Noetherian**. By proving Hilbert's basis theorem, we have show that all polynomial rings over a field (with finitely many variables) is Noetherian.

9.1 The Geometry Behind the Division Algorithm. Let f be an arbitrary polynomial in $k[x_1, \ldots, x_n]$ and let $V(\langle f_1, \ldots, f_s \rangle)$ where $\{f_1, \ldots, f_s\}$ is a Gröbner basis for the ideal it generates.

A natural question to ask is: are the zeros of f contained in the variety? Put another way, is f in the ideal $\langle f_1, \ldots, f_s \rangle$?

The way to understand "how far" f is from the ideal is by considering its remainder when dividing by the generating set:

$$f = a_1 f_1 + \ldots + a_s f_s + r.$$

If $r = 0$, then $f \in \langle f_1, \ldots, f_s \rangle$. If $r \neq 0$, then this tells us f is not in the ideal. Moreover, $f \equiv r$ in the quotient ring $k[x_1, \ldots, x_n]/I$. For this reason, the remainder r is often referred to as the **normal form** of f and is denoted $NF(f, I)$.

Remark 9.6. It is important in this set up that the generating set be a Gröwner basis. Otherwise, if $r \neq 0$ it is still possible that $f \in I$.

For example, consider the ideal we saw in homework 2: $\langle x + xy, y + xy, x^2, y^2 \rangle = \langle x, y \rangle$. Suppose we use the former generating set (which is clearly not a Gröbner basis) and consider the division algorithm for $f = x$. Assume we use the graded lexicographic ordering with $x > y$. Then

$$x = 0(x + xy) + 0(y + xy) + 0(x^2) + 0(y^2) + x,$$

so the remainder $r = x$ is nonzero. But we know x is in the ideal.

Proposition 9.7. *Let I have $\{g_1, \ldots, g_t\}$ as a (reduced) Gröbner basis. For all $f \in k[x_1, \ldots, x_n]$, there exists a unique normal form $r \in k[x_1, \ldots, x_n]$ such that*

- *no term of r is visible by any $LT(g_i)$, and*
- *there exists a $g \in I$ such that $f = g + r$.*

We call g the **projection** *of f onto I.*

In particular, when we use the division algorithm, we get an expression

$$f = \underbrace{a_1 g_1 + \ldots + a_t g_t}_{g} + r$$

Key to all this discussion is the construction of a Gröbner basis. By definition, this is a generating set such that $in_> I = \langle LT(f_1), \ldots, LT(f_s) \rangle$. We can see that an arbitrary generating set is *not* a Gröbner when there exists a lower order leading term than any of those in the set $\{LT(f_1), \ldots, LT(f_s)\}$. In such a case, there is some kind of cancelation possible with monomials that reduces the degree of a leading term in the generating set:

$$\text{mdeg}(ax^\alpha f_i - bx^\beta f_j) < \min\{\text{mdeg}(f_i), \text{mdeg}(f_j)\}.$$

This corresponds to the least common multiple (LCM).

Definition 9.8. *Let f, g be nonzero polynomials.*

- *If* $\text{mdeg}(f) = \alpha$ *and* $\text{mdeg}(g) = \beta$ *then* $\gamma = (\gamma_1, \ldots, \gamma_n)$, *where* $\gamma_i = \max(\alpha_i, \beta_i)$ *for each i. We call $x^\gamma = LCM(in_>(f), in_>(g))$.*
- *The corresponding* **S-polynomial** *is*

$$S(f, g) = \frac{x^\gamma}{LT(f)} \cdot f - \frac{x^\gamma}{LT(g)} \cdot g$$

Notice the S-polynomial cancels out the leading term and finds the "hidden" generator.

Example 22. Consider the ideal $\langle x + xy, y + xy, x^2, y^2 \rangle$ and suppose we are uses the graded lexicographic ordering where $x > y$. Then $LCM(xy, xy) = xy$ and

$$\begin{aligned} S(xy + y, xy + x) &= \frac{xy}{xy} \cdot (xy + y) - \frac{xy}{xy} \cdot (xy + x) \\ &= y - x \end{aligned}$$

So we can change the basis to $\langle xy + y, x - y, x^2, y^2 \rangle$.

Question 9.9. Can we use S-polynomials to construct a Gröbner basis from a generating set?

Buchberger's Criteria tells us the connection between Gröbner bases and the S-polynomial.

Theorem 9.10. (Buchberger's Criterion) *Let I be a polynomial ideal. Then a basis $G = \{g_1, \ldots, g_t\}$ for I is a Gröbner basis for I if and only if for all pairs $i \neq j$, $S(g_i, g_j) = 0$.*

We will discuss them further on Monday.

Lecture 10

Assigned Reading

The reading for this lecture is §1 of chapter 3 in [**?**].

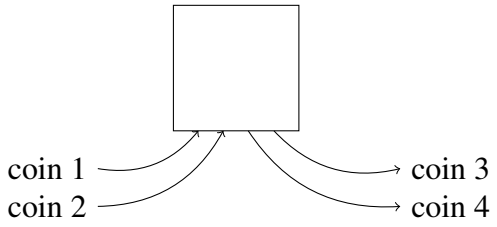
10.1 About Homework 5. Today's homework contains the coins problems discussed in class and in the last TA session. A hint for how to think of this problems is to approach them graphically. One problem you'll face is how to organize the information you have in order to get the implicitization. One way is to consider a *flattening* of the four coins problem in the following way:

	HH	HT	TH	TT
HH	x_{HHHH}	x_{HHHT}	x_{HHTH}	x_{HHTT}
HT	x_{HTHH}	x_{HTHT}	x_{HTTH}	x_{HTTT}
TH	x_{THHH}	x_{THHT}	x_{THTH}	x_{THTT}
TT	x_{TTHH}	x_{TTHT}	x_{TTTH}	x_{TTTT}

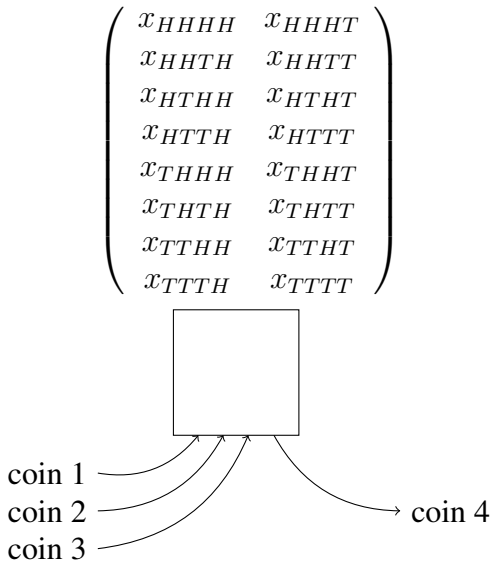
When we write out the possibilities in this flattened matrix (a 4×4 array)

$$\begin{pmatrix} x_{HHHH} & x_{HHHT} & x_{HHTH} & x_{HHTT} \\ x_{HTHH} & x_{HTHT} & x_{HTTH} & x_{HTTT} \\ x_{THHH} & x_{THHT} & x_{THTH} & x_{THTT} \\ x_{TTHH} & x_{TTHT} & x_{TTTH} & x_{TTTT} \end{pmatrix}$$

we are identifying "inputs" and "outputs" in this process. That is, every outcome, we are thinking of two coins as an input and two coins as an output. Graphically, we can represent where the array is represented as a box and the inputs are wires being fed in while outputs are being pushed out. We can read this diagram going from left to right. We add the arrows to emphasize this idea, but we will generally leave these off.



We could of course organize this in different ways. For example, we could flatten this array by feeding three coins and returning the outcomes of the remaining coin.



This graphical language is a very convenient way to think about these problems. What is the language? Generally, we draw vectors as distinct from higher dimensional arrays by sketching the former as a triangle with one wire and the latter as a box with multiple wires. For example, the vector $\vec{x} = (x_1, \ldots, x_n)^T$ can be drawn as

If we wish to encode the particular orientation, i.e. $\vec{v}$ as a column vector versus a row vector[10], we can consider bending the wire.

$$\vec{x}^* = \begin{pmatrix} x_1 & \ldots & x_n \end{pmatrix} \qquad \vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

These are equivalent whether we are considering the diagrams or the vectors themselves. For this reason, we won't bother bending the wires unless we're interested in communicating something about a particular flattening.

Let's think back to the two coin problem. We showed that the variety could be realized by the outer product of the two coin flips:

$$\begin{pmatrix} H_1 \\ T_1 \end{pmatrix} \cdot \begin{pmatrix} H_2 & T_2 \end{pmatrix} = \begin{pmatrix} x_{HH} & x_{HT} \\ x_{TH} & x_{TT} \end{pmatrix}.$$

This outer product is a *tensor product*

[10]This is known as the dual.

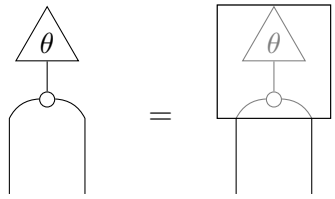
$$\begin{pmatrix} H_1 \\ T_1 \end{pmatrix} \otimes \begin{pmatrix} H_2 \\ T_2 \end{pmatrix} = \begin{pmatrix} x_{HH} & x_{HT} \\ x_{TH} & x_{TT} \end{pmatrix}$$

and graphically, we can represent the tensor product of vectors as drawing triangles in parallel.

Also within this graphical language, we can use a copy dot to indicate that a vector is being copied and tensored with itself. For example, consider problem 2 of homework 2.

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \otimes \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} \theta_1^2 & \theta_1\theta_2 \\ \theta_1\theta_2 & \theta_2^2 \end{pmatrix} := \begin{pmatrix} x & y \\ y & z \end{pmatrix}.$$

The corresponding picture is the following



How might these pictures appear for the problems in homework 5? What insights can they provide?

10.2 Elimination Theory. When we use the lex ordering, we end up with a process that is a generalization of Gaussian elimination from linear algebra.

Theorem 10.1. *(***Elimination Theorem***) Let G be a Gröbner basis of $I \subseteq k[x_1, \ldots, x_n]$ with respect to the lexicographic order where $x_1 > x_2 > \ldots > x_n$. Then for every $0 \leq \ell \leq n$, the set*

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

is a Gröbner basis of the ℓ^{th} eliminiation ideal

$$I_\ell := I \cap k[x_{>\ell}].$$

It is surprising that this reduction should still produce a Gröbner basis, but it does. Let us sketch out the proof.

Proof. Define $\langle G \rangle = I$ where $G = \{g_1, \ldots, g_s\}$. Fix ℓ. Since $G_\ell \subset I_\ell$, we know $\langle G_\ell \rangle \subset I_\ell$ and, therefore, $\langle in_> G \rangle \subset in_> I_\ell$.

To show the opposite inclusion, we need to show that for $f \in I_\ell$, then $in_> f$ is divisible by $in_> g$ for some $g \in G_\ell$.

Consider $f \in I_\ell \subset I$. Then there exists some $g \in G$ such that $in_> g$ divides $in_> f$. Since $f \in k[x_{>\ell}]$ then $in_> g \in k[x_{>\ell}]$. By the lex ordering, $g \in k[x_{>\ell}]$. Otherwise, there exists a term with a variable in $\{x_1, \ldots, x_\ell\}$ and $LT(g) \notin k[x_{>\ell}]$. Hence $g \in G_\ell$. □

Example 23. $I = \langle xy - 1, xz - 1 \rangle$. Is this a Gröbner basis? We can check this using the S-polynomial. $LCM(xy, xz) = xyz$.

$$S(f_1, f_2) = \frac{xyz}{xy}(xy - 1) - \frac{xyz}{xz}(xz - 1) = y - z =: f_3.$$

We now need to check f_3 against f_1. $LCM(xy, y) = xy$.

$$S(f_1, f_3) = \frac{xy}{xy}(xy - 1) - \frac{xy}{y}(y - z) = xy - 1 = f_2.$$

Similarly, we need to check f_3 against f_2. $LCM(xz, y) = xyz$.

$$S(f_2, f_3) = \frac{xyz}{xz}(xz - 1) - \frac{xyz}{y}(y - z) = y - xz^2 = -z(xz - 1) + (y - z).$$

Hence, Buchberger's criterion is met, so we now have a Göbner basis with respect to the Lex ordering $x > y > z$: $\langle xy - 1, xz - 1, y - z \rangle$.

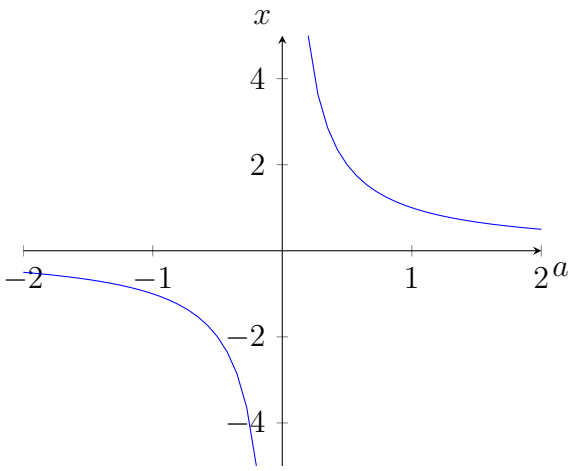
Remark 10.2. Notice that any two functions can generate the ideal; however, we cannot combine the leading terms of any two to get the third.

Observe $I_2 = I \cap k[z] = \langle 0 \rangle$ and $I_1 = I \cap k[y, z] = \langle f_3 \rangle$.

Consider, then, $y = z$ in $k[y, z]$. The corresponding variety is the set of points $(a, a) \in \mathbb{C}^2$. Within $k[x, y, z]$ the restriction $y = a = z$ implies the Gröbner basis:

$$ax - 1, ax - 1, a - a$$

So $x = \frac{1}{a}$ except when $a = 0$.



Consider $xy - 1, xz - 1$ as elements in $k[y, z][x]$. These are polynomials in x with coefficients in $k[y, z]$. We are on the variety when $(y, z) = (a, a)$ such that $a \neq 0$. If $a = 0$, then

$$0x - 1 = -1, 0x - 1 = -1$$

and the point is undefined.

This is the only kind of problem one can have with this construction. The reasons can be read about in [**?**], page 118.

Theorem 10.3. *(***The Extension Theorem***) Let k be an algebraically closed field. Let $I = \langle f_1, \ldots, f_s \rangle$ and write f_i as elements of $\mathbb{C}[x_2, \ldots, x_n][x_1]$. If $(a_2, \ldots, a_n) \in V(I_1)$ and the "coefficients" g_i don't all vanish at that point, then it can be extended. That is, there exists $a_1 \in \mathbb{C}$ such that $(a_1, \ldots, a_n) \in V(I)$.*

Note that even if we get a univariate polynomial, it may not be solvable in $\mathbb{Q}$ or even with radical extensions. One difficulty is dealing with this numerically is the problem of floating point numbers.

For more on this, check out Bertini (`https://bertini.nd.edu/`), which untilizes continuation-based numerics (homotopy continuation) to work around this problem.

We finish with a simple definition:

Definition 10.4. *A* **coordinate projection** *is a map*

$$\pi_\ell : \mathbb{C}^n \to \mathbb{C}^{n-\ell}$$

which drops the first ℓ variables.

Lecture 11

Assigned Reading

The reading for this lecture is §2 of chapter 3 in [**?**].

We continue where we let off last time with the discussion of the projection map. Let $\{f_1, \ldots, f_s\} \subset \mathbb{C}[x_1, \ldots, x_n]$ and let $V = \mathbb{V}(f_1, \ldots, f_s) \subset \mathbb{C}^n$ be an affine variety. The projection map,

$$\pi_\ell : \mathbb{C}^n \to \mathbb{C}^{n-\ell},$$

"forgets" the first ℓ coordinates. In other words, it sends $(a_1, \ldots, a_n)$ to $(a_{\ell+1}, \ldots, a_n)$. Hence $\pi_\ell(V) \subseteq \mathbb{C}^{n-\ell}$.

A fundamental question in algebraic geometry is, How can we relate the projection to the $\ell-$elimination ideal with the original variety? The lemma below tells us the projection of the original variety must be contained in the variety generated by the ℓ-elimination ideal.

Lemma 11.1.

$$\pi_\ell(\underbrace{V}_{V(I)}) \subset V(\underbrace{I_\ell}_{I \cap \mathbb{C}[x_{\ell+1}, \ldots]})$$

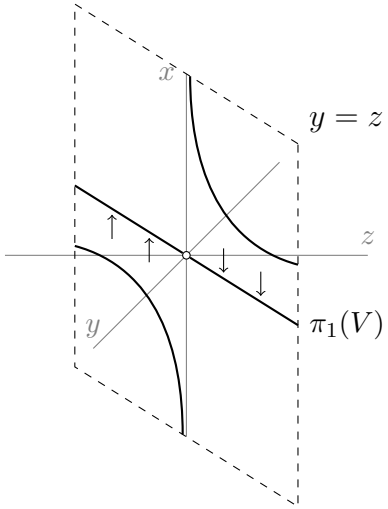
Proof. Let $f \in I_\ell$. Then f only has $x_{\ell+1}, \ldots, x_n$ as variables and vanishes at $(a_{\ell+1}, \ldots, a_n)$ where $(a_1, \ldots, a_n) \in V$. That is,

$$f(a_{\ell+1}, \ldots, a_n) = f(\pi_\ell(a_1, \ldots, a_n)) = 0$$

for $\vec{a} \in V$. Therefore, f vanishes on all points in $V(I_\ell)$. □

It is important to note that the projection of V isn't always a variety.

Example 24. Consider the previously discussed example, $I = \langle xy - 1, xz - 1 \rangle$. The variety of this ideal is $\{(a, b, b) \mid ab = 1\}$. The projection of this variety is $\pi_1(V) = \{(b, b) \mid b \neq 0\}$. In other words, it is the line $y = z$ in the yz-plane, excluding the point $y = z = 0$. Hence it is not a variety.



Let's understand I_1. This is the ideal $\langle xy-1, xz-1\rangle \cap \mathbb{R}[y,z]$, the ideal where x has been eliminated. When we eliminate x, we get the relation $y=z$, the Zariski closure of $\pi_1(V)$.

Theorem 11.2.

$$V(I_1) = \pi_1(V) \cup \left(V(\underbrace{g_1, \ldots, g_s}_{\substack{\textit{from extension} \\ \textit{theorem}}}) \cap V(I_1) \right)$$

where $g_i \in k[x_2, \ldots x_n][x_1]$.

When considering the variety associated to the first elimination ideal I_1, we can look at the projection of $V(I)$; however, this may not capture the full ideal. Whatever missing points might exist are captured by the set of coefficients from the extension theorem. This set, however, might capture many additional points, hence the intersection.

Theorem 11.3. *(***The Closure Theorem***) Let* $V = \mathbb{V}(f_1, \ldots, f_s) \subset \mathbb{C}^n$. *Let* I_ℓ *be the* ℓ^{th} *elimination ideal of* $\langle f_1, \ldots, f_s\rangle$ *. Then*

(1) $V(I_\ell)$ *is the smallest variety containing* $\pi_\ell(V)$ *in* $\mathbb{C}^{n-\ell}$ *(i.e. it is the Zariski closure)*
(2) *When* $V \neq \emptyset$, *there exists an affine* $W \subsetneq V(I_\ell)$ *such that* $V(I_\ell) \setminus W \subset \pi_\ell(V)$

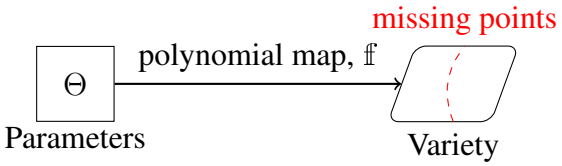
One might hope is process of defining $\pi_\ell(V)$ has only three steps: you take the ℓ^{th} projection of V, find its Zariski closure and take out "bad" points. Unfortunately, that is not how it is in general. The best we can say is the true image of the variety is *constructible*, but may require many larger and smaller varieties to define $\pi_\ell(V)$.

Definition 11.4. *A set* $X \subset \mathbb{C}^n$ *is called* **constructible** *if there exists affine varieties* $Z_i \subset W_i \subset \mathbb{C}^n$ *where* $1 \leq i \leq m$ *such that*

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

Hence, $\pi_\ell(V) = \cup_{i=1}^m (W_i - Z_i)$ for some varieties W_i and Z_i.

11.1 Implicitization. The general set up of an implicitization problem is set up as follows:



The models we study have inputs arising from some parameter space, Θ. The model is (in this class) a set of polynomial maps $\mathbb{f}(f_1(t_1, \ldots, t_m), \ldots, f_n(t_1, \ldots, t_m))$, where $x_1 = f_1(\theta), \ldots, x_n = f_n(\theta)$. The image of $\mathbb{f}$ lands in some variety, $\overline{\text{Im }(\mathbb{f})}$. Inside this variety are points not directly arising from the model. These "missing" points are themselves form varieties of smaller codimension.

We'd like to answer some basic questions about this map, $\mathbb{f}$.

Some Basic Questions

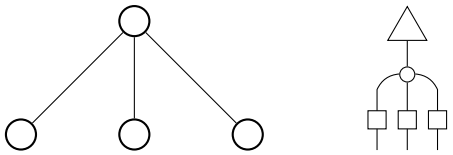
- Given $\mathbb{f}$, what is $\overline{\text{Im }(\mathbb{f})} = V$?
- What is the dimension of $\overline{\text{Im }\mathbb{f}}$?

- Does the parametrization fill all of V? If not, characterize the missing points (a constructible set).

Problems are often complicated by the applications themselves. When restricting to

- real nonnegative parameters (e.g. probabilities),
- $z = |z|$, or
- inequalities $\mathbb{R}_{\geq 0}$

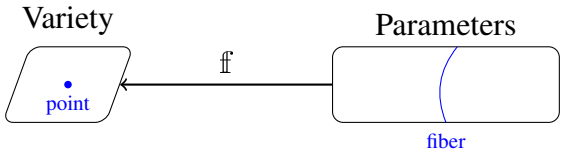
means we are restricting to a "semi-algebraic set." In the homework, you showed the model below, a bayesian network with three binary coins and a hidden coin, is defined by no equations. In other words, in the seven dimensional probability simplex, this model is full dimensional.



Question 11.5. Does this mean I can get any probability distribution from this set up?

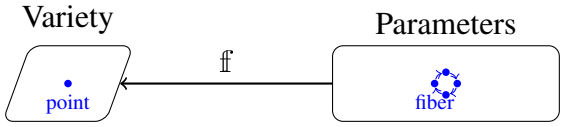
In terms of algebraic geometry, this statement says every point can be described by this model. In terms of the physical situation of coin flipping, this is **not** true. That is, I cannot always find a set of coins which can give me an arbitrary distribution in this space. Instead, it means that any point in the image can be described as a *mixture* of physical probability distributions. These missing points are ubiquitous throughout the variety; they even exist arbitrarily close to the uniform distribution.

Suppose $\dim(\text{Im } \mathbb{f}) < \dim \Theta$. In other words, I wasted parameters. So the preimage of a point may be higher than zero dimensional. These preimages are called **fibers**.

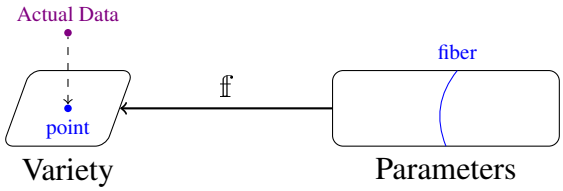


When this happens—and this is common—need to describe these fibers. These may be lines, circles or some other algebraic shape. The fibers can be described as a family of varieties

Even when $\dim \Theta = \dim(\text{Im } \mathbb{f})$, we can still have finite fibers.

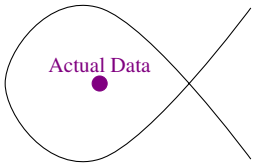


Often, science is interested in comparing a given model to an observed data set. When comparing a point to its closed point on the variety, if we have fibers, we cannot *identify* which set of parameters that could have given rise to the situation.



In other words, fibers a very bad for any model. In statistics, this is called a **failure of identifiability**. All of the modern standard theory of statistics relies on models being injective, smooth maps. When those assumptions fail, we can lose our ability to do statistical analysis.

Another potential problem arises from **singularities**, meaning the Fisher information matrix is singular. This can arise because there is no *unique* closest point to observed data.



There are other issues that can arise, all of which would make good topics for a class project.

11.2 Projective Space. In your homework, you have been asked to think about a hypersurface in $\mathbb{P}^{15}$. There are details of projective space in [**?**] in chapter 8. This is a topic in any algebraic geometry book.

There are multiple ways of thinking of projective space. One way is to think of it as sets of values. Consider a point $(x_1, x_2) \in \mathbb{C}^2$. In projective space, we consider a ratio shared between the values: $(x_0 : x_1) \in \mathbb{P}^1_{\mathbb{C}}$. The colon notation is meant to indicate a point in projective space.

We can think of points in $\mathbb{P}^1_{\mathbb{C}}$ to be an equivalence class of points,

$$[\vec{x}] = (x_0 : x_1) = (\lambda x_0 : \lambda x_1)$$

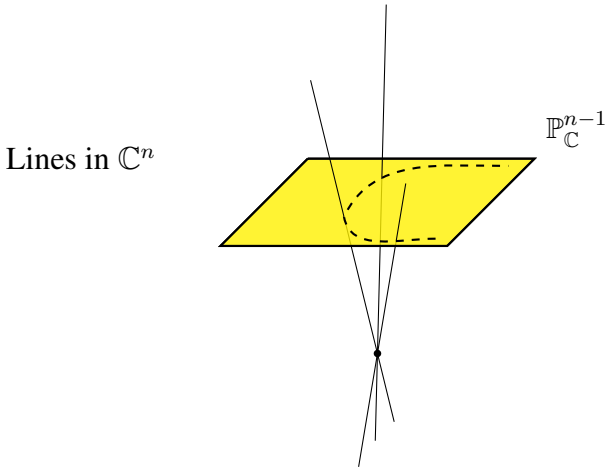
where

- not all entries are zero, and
- $\lambda \in \mathbb{C} \setminus \{0\}$.

This is allows us to add points at infinity. We can define the map

$$\mathbb{C} \to \mathbb{P}^1_{\mathbb{C}}$$

where $z \mapsto (z : 1)$. The reverse mapping is $(x_0 : x_1) \mapsto \frac{x_0}{x_1}$. What happens if $x_1 = 0$? Then the point is ∞.



Question 11.6. Does there exist a negative infinity?

In the case of $\mathbb{P}^1_{\mathbb{C}}$, there is no negative infinity since $(-x : 0) = (x : 0)$.

Question 11.7. How many points at infinity are there for $\mathbb{P}^2_{\mathbb{C}}$?

There is an entire line at infinity. These are the points $(x_0 : x_1 : 0)$ for which there are many ratios for the first two coordinates. Continuing the pattern, when $\mathbb{P}^n_{\mathbb{C}}$ is out space, the dimension of infinity is $n-1$.

Benefits of Projective Space

- Representations are invariant under rescaling.
- We can study points at ∞.
- Things that are almost true are become true. For example, any two lines in a plane intersect at a point.

In the setting of probability, we might have three outcomes (x_1, x_2, x_3) with normalizing conditions:

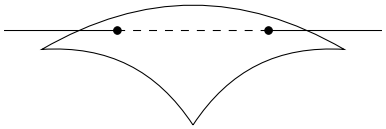
$$(x_1, x_2, x_3) \mapsto \left(\frac{x_1}{z}, \frac{x_2}{z}, \frac{x_3}{z}\right)$$

where $z = x_1 + x_2 + x_3$.

If we replace the condition "sums to one" with "don't care about the constant," we essentially can work over projective space. For each point in projective space, there will only be one point in affine space that is a probability distribution. When we introduced the problem of three coins, we used the notation (d_H, d_T) instead of $(d_H, 1 - d_H)$ so that we could transition to $(d_H : d_T)$, which is the problem in projective space.

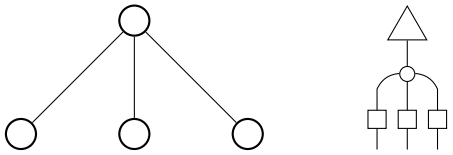
For example, the Zariski closure of the three coins problem is contained in $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$. This is called a **Segre product**.

The **secant** of a segre is the following. Take a segre product variety. The union of the lines connecting two points on the segre variety comprises the secant variety.



Secant varieties give us descriptions of *mixture models* and *quantum entanglement*. They sounds simple, but in even seemingly small examples, we do not know the implicitizations of secant and segre varieties.

Exercise 4. True or false: The model below is a secant variety of a segre variety.

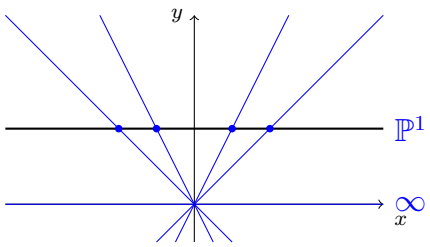


11.3 The Geometry of Projective Space. Consider a point $\mathbb{P}^1_{\mathbb{C}}$

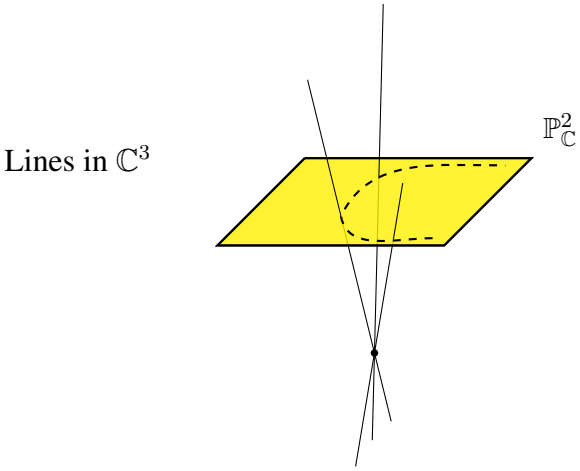
$$(x_0, x_1) = (\lambda x_0, \lambda x_1).$$

This point in $\mathbb{C}^2$, this is a line.

We can pick a representation via a line in $\mathbb{R}^2$. The space $\mathbb{P}^1_{\mathbb{R}}$ is represented with a line which does not pass through $(0, 0)$. The set of lines through $(0, 0)$ each represent a point in $\mathbb{P}^1_{\mathbb{R}}$, So $\mathbb{P}^1$ is the space of all lines through the origin. The line parallel to our choice represents ∞.



Here is the same interpretation in three dimensions. Notice that ∞ is the set of lines parallel to the plane (not depicted).



Next time, we will think about what changes are needed in the definition of variety to make it work in projective space.

Lecture 12

Assigned Reading

The reading for this lecture is §2 of chapter 8 in [**?**].

Last time, we discussed projective spaces. We mentioned that one can think of projective space $\mathbb{P}^n$ as having coordinates $(x_0 : x_1 : \ldots : x_n) \sim (\lambda x_0 : \lambda x_1 : \ldots : \lambda x_n)$. We finished the lecture with the observation that $\mathbb{C}^n \subset \mathbb{P}^n_{\mathbb{C}} = (\mathbb{C}^{n+1} \setminus \{0\})/\sim$. In other words, a point $(x_0 : \ldots : x_n)$ in $\mathbb{P}^n$ can be understood as all the scalar multiples of the point $(x_0, \ldots, x_n)$ in $\mathbb{C}^{n+1}$. We do have a choice in how we understand these equivalence classes based on our choice of orientation of the hypersurface relative to the origin (see the end of lecture 12).

Projective space is particularly useful in the field of probability since we often have some kind of relationship like

$$x_0 + \ldots + x_n = 1.$$

This relationship is well captures with a "smart" choice of orientation of the hypersurface in $\mathbb{C}^{n+1}$. For example, in the three coin case, we can choose the hyperplane at infinity to be

$$x_{HHH} + x_{HHT} + \ldots + x_{TTT} = 0.$$

This is projective space. But what about projective varieties?

12.1 Projective Varieties. Let us consider the basic case of $\mathbb{P}^1$ where points are $(x_0 : x_1)$.

Question 12.1. In this space, can we define the variety $V(\langle x_0^2 - 1\rangle)$?

The answer is no. The obvious solution for the equation is $x_0 = \pm 1$; however, this cannot be characterized by a point in $\mathbb{P}^1$ since $(1, x_1) \sim (\lambda, \lambda x_1)$. Hence, varieties in $\mathbb{P}^n$ must be homogeneous.

Definition 12.2. *A polynomial is said to be* **homogeneous** *if every term has the same degree.*

Example 25. $f = x_1 x_0 + x_0^2$

The zeros of f must have the relationship $x_0(x_1 + x_0) = 0$. Hence all points of the form $(0 : 0)$ or $(1 : -1)$ in $\mathbb{P}^1$ are in the variety $V(\langle f\rangle)$.

Homogeneous functions are necessary precisely because of their property $f(\lambda x_0, \ldots, \lambda x_1) = \lambda^d f(x_0, \ldots, x_n)$ where d is the degree of f. Thus, homogeneous functions respect the equivalence on $\mathbb{C}^{n+1}$.

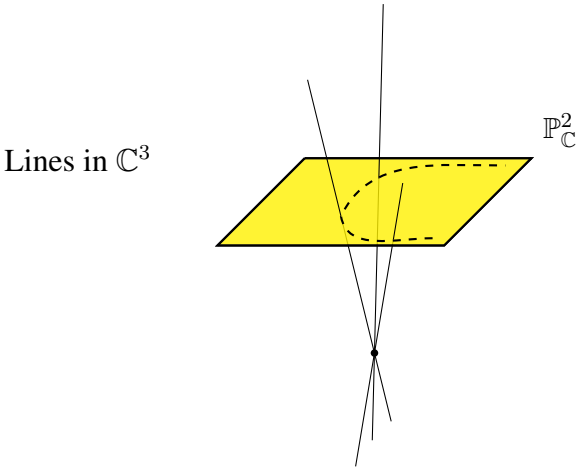
Although ideals can have homogeneous generators, they can contain nonhomogeneous elements. For example, the ideal $\langle xy - zw, x - z\rangle$ contains non-homogenous polynomials, like $xy - zw + x - z$.

We cannot require that an ideal (corresponding to a projective variety) contain only homogeneous functions. Such a requirement isn't compatible with the definition of an ideal. Instead, all functions in the ideal must be representable as a sum of homogeneous functions and each of those homogeneous pieces has to separately vanish. This is what we'll refer to as a **homogeneous ideal**.

Exercise 5. Prove that a homogeneous parameterization gives rise to a homogeneous implicitization.

As we mentioned in the last class, the coin problems from the homework were written as homogeneous parameterizations: $x_{HT} = d_H q_T$ instead of $x_{HT} = d_H(1 - q_H)$. When it was implicitized, we found homogeneous polynomials. This is an example of what's being asked in the exercise.

A projective variety can be fairly complicated. As we change the choice of embedding, the descriptions will change. In a graduate level algebraic geometry course, a fair amount of energy is spent studying how this works and what can go wrong. For this class' applications, this is rarely an issue. The projective varieties we study are generally simple and are mostly described by one affine space. Put another way, for our purposes, given a projective variety $V \subset \mathbb{P}^n$, there is a "natural" cone over V in $\mathbb{C}^{n+1}$. This cone is an affine variety containing all the points in V, plus the origin, with the equivalence relation being removed/forgotten.



The "cone" is the lines in $\mathbb{C}^{n+1}$ corresponding to all the points in $\mathbb{P}^n$.

12.2 Elimination Theory. We now return to elimination theory over affine varieties.

Theorem 12.3. *(**Polynomial Implicitization Theorem**) Let $\mathbb{f} : \mathbb{C}^m \to \mathbb{C}^n$ be a polynomial parameterization*

$$x_i = \sum_{\vec{\alpha}(i)} A_{\alpha(i)} t^{\alpha(i)}$$

for $\vec{\alpha}(i) \in \mathbb{Z}^m_{\geq 0}$ with $I = \langle x_1 - f_1, \ldots, x_n - f_n \rangle$ and $I_m = I \cap k[x_1, \ldots, x_n]$

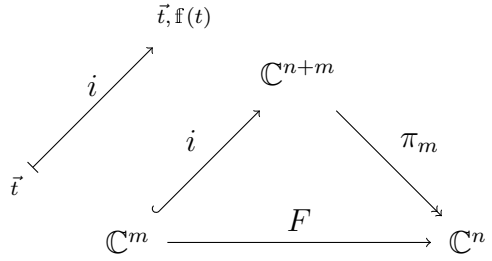
$$V(I_m) = \overline{\mathbb{f}(\mathbb{C}^m)}$$

(Zariski closure).

This is a result of the fact that $V(I) = \text{graph}(\mathbb{f})$ where

$$(\vec{t}, \mathbb{f}(t)) \subset \mathbb{C}^{n+m}.$$

The result follows by using the coordinate projection π_m. The diagram below summarizes this idea. The map i is injective (denoted using the hook arrow) and the map π_m is surjective (denoted with the double arrow). The composition of those two maps is equal to the parameterization map $\mathbb{f}$.



The proof of this theorem is a straightforward application of the closure theorem, which relies on $\mathbb{C}$ being algebraically closed.

Theorem 12.4. *(***Closure Theorem***)* $V(f_1, \ldots, f_s) \subseteq \mathbb{C}^n$, I_ℓ *the* ℓ^{th} *elmination ideal, then* $V(I_\ell) = \overline{\pi_\ell(V)} \subset \mathbb{C}^{n-d}$.

Therefore, $V(I)$ is graph of $\mathbb{f}$ and the closure theorem gives us these equalities: $\mathbb{f}(\mathbb{C}^n) = \pi_m(i(k^m)) = \pi_m(V)$.

When $k \subset \mathbb{C}$, infinite field, like $\mathbb{Q}$ or $\mathbb{R}$, this theorem still holds. This is great since we often compute over $\mathbb{Q}$ and grape over $\mathbb{R}$. To see why this is true, we need to show that $V_k(I_m) = V_{\mathbb{f}}(I_m)$ is the smallest (inclusion minimal) variety containing the projection of the image.

Lemma 12.5. $V_k(I_m)$ *is the smallest (inclusion minimal) variety containing the image.*

Remark 12.6. If the polynomials from the parameterization have coefficients in k $f_1, \ldots, d_n \in k[t_1, \ldots, t_n]$, then the Gr obner basis coefficients will also be in k. Therefore

$$(I_\ell)_k = (I_\ell)_{\mathbb{C}},$$

meaning I can think about k or $\mathbb{C}$.

Let Z_k be a variety containing $\mathbb{f}(k^m)$. By Hilbert's Basis theorem, let $Z_k = V(g_1, \ldots, g_s)$. We know $Z_k \supset \mathbb{f}(k^n)$, so for each g_i, $g_i \circ \mathbb{f} : k^m \to 0$. $\mathbb{f}$ is made up of n polynomials over the parameters (t_i) and g_i describes a polynomial way of combining these polynomials. So $g_i \circ \mathbb{f}$ is a polynomial in $k[t_1, \ldots, t_m]$ vanishing on all of k^m. It must therefore vanish on $\mathbb{C}^m$ since it has infinitely many zeros.

For the reserve, consider the complex variety $Z_{\mathbb{C}} = V_{\mathbb{C}}(g_1, \ldots, g_s) \subset \mathbb{C}^n$ has to contain $\mathbb{f}(\mathbb{C}^n)$. By the closure theorem, $V_{\mathbb{C}}(I_m) \subset Z_{\mathbb{C}}$. If we restrict to k, then $V_k(I_m) \subset Z_k$.

This is important to applications since many physical applications happen over $\mathbb{Q}$ or $\mathbb{R}$ instead of $\mathbb{C}$.

12.3 Rational Parameterization. A probability distribution might be of the form,

$$\begin{cases} x_1 = \frac{s^2}{s^2+st+t^2} \\ x_2 = \frac{st}{s^2+st+t^2} \\ x_3 = \frac{t^2}{s^2+st+t^2} \end{cases}$$

where outcomes are modeled with rational functions over some set of hidden parameters (s, t). An obvious thing to do is to turn these expressions into polynomials. When we eliminate the

denominators, but we pick up extra "garbage." For example, consider the following rational parameterization:

$$\begin{cases} x = \frac{u^2}{v} \\ y = \frac{v^2}{u} \\ z = u \end{cases}$$

Clearing the denominators of this parameterization yields the parameterization ideal

$$I = \left\langle vx - u^2, uy - v^2, z - u \right\rangle.$$

Then, $I_1 = \langle vx - z^2, zy - v^2, xyz - vz^2 \rangle$ and $I_2 = \langle x^2yz - z^4 \rangle$, meaning

$$V(I_2) = V(x^2y - z^3) \cup V(z).$$

The "garbage" in this case is the plane $V(z)$, which is not meaningful for the rational parameterization.

One "cleans things up" via saturation. Time permitting, we will discuss this more in detail later in the course.

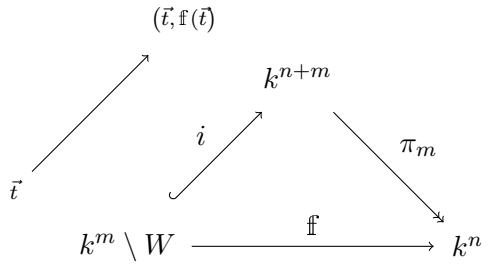
In general, a rational parameterization will be of this form:

$$\begin{cases} x_1 = \frac{f_1(\vec{t})}{g_1(\vec{t})} \\ \vdots \\ x_n = \frac{f_n(\vec{t})}{g_n(\vec{t})} \end{cases}$$

where it is not defined for any point $\vec{t}$ such that there exists an i for which $g_i(\vec{t}) = 0$. So the variety must be contained in

$$\underbrace{k^m \setminus V(\underbrace{g_1 g_2 \cdots g_n}_{g})}_{W}.$$

Although true, this fact is computationally expensive, so it is not always the best approach. Conceptually, however, it is very similar to what we discussed earlier. The rational paramaterization map $\mathbb{f}$ can be constructed as a composition of the injective map from $k^m \setminus W$ to a projective map onto k^n.



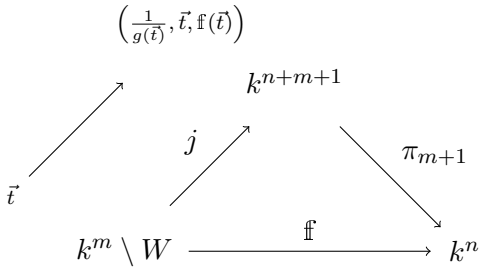
where

$$i(k^m \setminus W) \subset V(\langle g_1 x_1 - f_1, \ldots, g_n x_n - f_n \rangle + \text{"garbage"})$$

To address these extra points we don't want, we introduce what is called a "slack variable," which forces the requirement that $g_i \neq 0$ for every i. We do this by taking the ideal above and

adding the equation $1 - gy$, where g is the product $g_1 g_2 \cdots g_n$. We order y as highest since we wish to eliminate it with ease.

$$J = \langle g_1 x_1 - f_1, \ldots, g_n x_n - f_n, 1 - gy \rangle \subset k[y, t_1, \ldots, t_m, x_1, \ldots, x_m]$$



With this approach, $j(k^n \setminus W) = V(J)$ and $\overline{\mathbb{f}(k^n \setminus W)} = \overline{\pi_{m+1}(V(I))}$. No "garbage" to clean up.

Remark 12.7. The more g_i functions we have, the harder this technique is to implement.

Theorem 12.8. *(***Rational Implicitization***) Let k a field of characteristic 0. Do the process discussed above, then $V(J_{m+1}) = \overline{\mathbb{f}(k^m \setminus W)}$.*

Lecture 13

Assigned Reading

The reading for this lecture is §1 of chapter 4 in [**?**].

Lecture began with a discussion of the homework.

13.1 Nullstellensatz and Ideal-Variety Correspondence. We have been using this in an intuitive way already, but let's be a little more explicit.

Consider the affine space $\mathbb{C}^n$.

Question 13.1. Let $I = \langle 0 \rangle$, then $V(\langle 0 \rangle) = \mathbb{C}^n$. What ideal I has $V(I) = \emptyset$?

The answer is any ideal generated from constant polynomials, like $I = \langle 1 \rangle$. This is true for any algebraically closed field like $\mathbb{C}$. For ideals over fields without this property, we can include elements whose roots lie outside the field, for example, if $I_{\mathbb{R}} = \langle x^2 + 1 \rangle$, then $V_{\mathbb{R}}(x^2 + 1) = \emptyset$.

The bigger question is, given $I = \langle f_1, \ldots, f_s \rangle$, when is it true that $V(I) = \emptyset$? When working over $\mathbb{C}$, this is true whenever $1 \in I$. For example, the ideal $\langle 1 + x^2, x^2 \rangle$ contains 1. It's reduced Gr obner basis is $\{1\}$. The corresponding variety is then empty.

Lemma 13.2. *An ideal over $\mathbb{C}$ corresponds to an empty variety if and only if its reduced Gr obner basis is $\{1\}$.*

Over a non-algebraically closed field, the exist many different ideals which correspond to the empty variety. Even over $\mathbb{C}$, the correspondence between ideals and varieties is complicated by the fact that different ideals can correspond to the same variety, for example $V(\langle x^2 \rangle) = V(\langle x \rangle)$.

Theorem 13.3. ***(Weak Nullstellensatz)*** *If k is algebraically closed, $I \subset k[x_1, \ldots, x_n]$ with $V(I) = \emptyset$ then $I = k[x_1, \ldots, x_n]$.*

In other words, $1 \in I$.

Proof. If $n = 1$, then I is principal. Then $V(f)$ equals the set of roots. If it has no roots, then f is a constant.

Assume this holds for up to $n - 1$. Let I be some ideal $\langle f_1, \ldots, f_s \rangle \subset k[x_1, \ldots, x_n]$ with $V(I) = \emptyset$. Suppose $\deg f_i = N$. Pick a linear change of coordinates such that

$$f = \alpha \hat{x}_1^N + \text{lower order terms in } \hat{x}_1$$

Then

$$\begin{aligned} x_1 &= \hat{x}_1 \\ x_2 &= \hat{x}_2 + a_2 \hat{x}_1 \\ &\vdots \\ x_n &= \hat{x}_n + a_n \hat{x}_1 \end{aligned}$$

and

$$f(\hat{x}_1, \hat{x}_2 + a_2\hat{x}_1, \ldots, \hat{x}_n + a_n\hat{x}_1) = c(a_2, \ldots, a_n)\hat{x}_1^N + \text{ lower degree terms}$$

We can choose $\vec{a}$ so that $c(\vec{a}) \neq 0$. We call this an open condition since it is the complement of some variety. A typical way of saying this is choosing a "generic point" (something on a set

minus some smaller variety). Now, the induction step we want follows from the extension theorem discussed earlier.

Theorem 13.4. *(**Extension Theorem**) Given $V = V(f_1, \ldots, f_s) \subset \mathbb{C}^n, g_i \in k[x_1, \ldots, x_n]$ and $f_i = g_i(x_2, \ldots, x_n)x_1^{N_i} +$ lower degree terms. Then*

$$V(I_1) = \pi_1(V) \cup [V(g_1, \ldots, g_s) \cap V(I_1)]$$

and there is a g_i which is a nonzero constant. Hence, $V(I_1) = \pi_1(V)$ and every partial solution extends into a solution for V.

This constant g_i is the coefficient $c(a_2, \ldots, a_n)$.
By induction $V(I_1) = \emptyset$. So $\pi_1(V) = \emptyset$, and by the extension theorem $V = \emptyset$.
Thus, $V(\hat{I}_1) = \pi_1(V(\hat{I})) = \emptyset$. So $1 \in \hat{I}_1 \subset \hat{I}$, $V(\hat{I}) = \emptyset$. □

Geometrically, the property of being empty isn't coordinate-dependent. So our computational solution to the question, "Does $f_1, \ldots, f_s$ have a solution?" depends on this extension technology and tells us 1 is in the reduced Gröbner basis is necessary and sufficient.

What if k is not an algebraically closed? One direction of implication works. If 1 is the reduced Gröbner basis of $\langle f_1, \ldots, f_s \rangle$, then there does not exist a solution. The converse, however, is false. The counterexample was mentioned earlier: $V(\langle x^2 + 1 \rangle) = \emptyset$ over $\mathbb{R}$.

13.2 Hilbert's Nullstellensatz. Now, let us discuss the related question, when is it true that $V(I) = V(J)$? The example of this given earlier was $V(\langle x^2 \rangle) = V(\langle x \rangle)$.

Theorem 13.5. *(**Hilbert's Nullstellensatz**) Let $f, f_i \in k[x_1, \ldots, x_n]$, k is algebraically closed. If $f \in I(V(f_1, \ldots, f_s))$ then $\exists m \geq 1$ such that $f^m \in \langle f_1, \ldots, f_s \rangle$.*

The radical ideal $\sqrt{I} = \{f : f^m \in I \text{ for some } m\}$ is the radical of I. If $I = \sqrt{I}$, then I is called a **radical ideal**.

$$I(V(I)) = \sqrt{I}.$$

This is an algebraic closure operation and is sometimes called the strong Nullstellensatz.

Proof. Let $f^m \in \langle f_1, \ldots, f_s \rangle$. Show that $f^m = \sum_{i=1}^s h_i f_i$.

$$J = I + \langle 1 - yf \rangle \subseteq 0k[y, \vec{x}].^{11}$$

So suppose we have that $1 \in J$. Then $1 = q(1 - yf) + \sum_{i=1}^s p_i f_i$ where q, p_i are polynomials. This implies $y = \frac{1}{f(x_1, \ldots, x_n)}$,

$$1 = \sum_{i=1}^{s} p_i(x_1, \ldots, x_n, \frac{1}{f}) f_i,$$

and hence

$$f^m = \sum_{i=1}^{s} f^m p_i(\vec{x}, \frac{1}{p}) f_i = \sum_{i=1}^{s} h_i f_i.$$

Next time we justify that $1 \in J$. □

[11] This is the trick we did in the previous lecture.

Lecture 14

Assigned Reading

The reading for this lecture is §2 of chapter 4 in [**?**].

14.1 Nullstellensatz and Ideal-Variety.

Theorem 14.1. *(**Hilbert's (Strong) Nullstellensatz**) Let $f, f_1, \ldots, f_s \in k[\vec{x}]$. If $f \in I(V(\langle f_1, \ldots, f_s\rangle))$ then $f^m \in \langle f_1, \ldots, f_s\rangle$ for some m.*

Proof. Suppose f vanishes at points where $f_1, \ldots, f_s$ do as well. Then $f \in I(V(\langle f_1, \ldots, f_s\rangle))$ and we wish to show there exists an m such that $f^m = \sum h_i f_i$.

Just as we did in the previous lecture, we use the trick of increasing the dimension. Let $J = \langle f_1, \ldots, f_s, 1 - yf\rangle \subset k[\vec{x}, y]$. If $1 \in J$, then

$$1 = q(1 - yf) + \sum p_i f_i$$

$$y = \frac{1}{f(\vec{x})}$$

$$1 = \sum p_i \left(x_1, \ldots, x_n, \frac{1}{f}\right) f_i$$

$$f^m = \sum f^m p_i \left(\vec{x}, \frac{1}{f}\right) f_i$$

It is enough to show that $V(J) = \emptyset$. Let $\vec{a} \in k^{n+1}$.

(1) If $a_1, \ldots, a_n \in V(f_1, \ldots, f_s)$, then $1 - a_m f = 1 - 0 \neq 0$. Hence $(a_1, \ldots, a_n, a_{n+1}) \notin V(J)$.
(2) If $a_1, \ldots, a_n \notin V(f_1, \ldots, f_s)$, for some $f_i(\vec{a}_{\leq n}) \neq 0$, meaning for some $f_i \in k[\vec{x}] \subset k[\vec{x}, y]$. Therefore, $(a_1, \ldots, a_n, a_{n+1}) \notin V(J)$

So by the the weak Nullstellensatz, $1 \in J$. □

The proof above tells us that $I(V)$ is the radical ideal, $\sqrt{I}$. Explicitly,

$$I(V(I)) = \sqrt{I} = \{f : f^m \in I \text{ for some } m\}.$$

Recall the inclusion-reversing corresponding between ideals and varieties.

(1) $I_1 \subset I_2 \implies V(I_1) \supset V(I_2)$
(2) $V_1 \subset V_2 \implies I(V_1) \supset I(V_2)$

The second one tells us how $V(I)$ and $I(V(I))$ are related. Note that $V(I(V)) = V$. (Why?) Furthermore, $I(V(I)) = I$ if and only if I is radical (meaning $I = \sqrt{I}$).

Some natural questions to ask:

(1) Is I radical?
(2) Find generators for $\sqrt{I}$.
(3) Determine if $f \in \sqrt{I}$.

These can be answered in SAGE using the techniques we've discussed in the class.

Consider the following proposition.

Proposition 14.2. *Let* $I = \langle f_1, \ldots, f_s \rangle \subset k[\vec{x}]$. *Then* $f \in \sqrt{I}$ *if and only if* $1 \in \hat{I} = \langle f_1, \ldots, f_s, 1 - yf \rangle \subset k[\vec{x}, y]$.

Proof. This is a consequence of HIlbert's Nullstellensatz. If $1 \in \hat{I}$, then $f^m \in I$ for some m whenever $f \in \sqrt{I}$.

Alternatively, if $f \in \sqrt{I}$ then $f^m \in I \subset \hat{I}$ and $1 - yf \in \hat{I}$ where

$$1 = y^m f^m + (1 - y^m f^m) = y^m f^m + (1 - yf)(1 + yf + \ldots + y^{m-1} f^{m-1}) \in \hat{I}$$

since $f^m \in \hat{I}$ and $(1 - yf) \in \hat{I}$. □

Based on the above proposition, the algorithm for answering question 3 will take the approach for $I = \langle f_1, \ldots, f_s \rangle$ and compute the reduced Gr obner Basis for $\langle f_1, \ldots, f_s, 1 - yf \rangle$ and check if it is equivalent to $\{1\}$.

14.2 The Algebra of Ideals and Varieties. Suppose $V(I) \cdots V(J)$. Then $V(I + J) =$? Recall that $I + J = \{f + g \mid f \in I, g \in J\}$. So $\langle f_1, \ldots, f_s \rangle + \langle g_1, \ldots, g_t \rangle = \langle f_1, \ldots, f_s, g_1, \ldots, g_t \rangle$. Hence,

$$V(I + J) = V(I) \cap V(J).$$

What about the union of two varieties? We learned previously that $V(f) \cup V(g) = V(fg)$. With more generators, it works similarly:

$$V(f_1, \ldots, f_s) \cup V(g, \ldots, g_t) = V(f_i g_j \ : \ i \in [s], j \in [t]).$$

This is precisely the produce of two ideals:

$$IJ = \{f_1 g_1 + \ldots f_r g_r \mid f_i \in I, g_j \in J\}.$$

Hence $V(IJ) = V(I) \cup V(J)$.

This make sense. Consider $v \in V(IJ)$ then for all $g \in I, h \in J$

$$gh(v) = 0$$

$g(v) = o$ or $h(v) = 0$. On the other hand, suppose $v \in V(I) \cup V(J)$. Then either $g(v) =$ for all $g \in I$ or $h(v) = 0$ for all $h \in J$ so $gh(v) = 0$ for all $g_i h_j$.

What is the intersection of two ideals? It is precisely the intersection as we understand it for sets.

Definition 14.3. $I \cap J = I \cap_{set} J$

This, unfortunately, is not the product of the two. If $I = J = \langle x, y \rangle$, then

$$IJ = \left\langle x^2, y^2, xy \right\rangle.$$

The generators of the intersection should be the least common multiple:

$$I \cap J = \langle LCM(f, g) \rangle.$$

It is also true that

$$V(I \cap J) = V(I) \cup V(J).$$

We leave this fact as an exercise.

If $I = \sqrt{I}$, $J = \sqrt{J}$, this does not imply $IJ = \sqrt{IJ}$. This idea does apply to intersections of ideals.

Proposition 14.4. $\sqrt{I \cap J} = \sqrt{I} \cap \sqrt{J}$

Proof. Suppose $f \in \sqrt{I \cap J}$. Then $f^m \in I \cap J$. Hence $f^m \in \sqrt{I}$ and $\sqrt{J}$.

Let $f \in \sqrt{I} \cap \sqrt{J}$. THen $f^m \in I$ and $f^n \in J$. Then $f^{m+n} \in I \cap J$. Hence $f \in \sqrt{I \cap J}$. □

14.3 Detecting Unions. Sometimes you may be given a variety which is best expressed as a union of two or more familiar varieties. How can we determine when this is the case and what those varieties are?

Definition 14.5. *Affine variety $V \subset k^n$ is* **irreducible** *if whenever $V = V_1 \cup V_2$, then either $V_1 = V$ or $V_2 = V$.*

In other words, a variety is irreducible when any union must be a trivial one. On the algebraic side, this relates to the idea of the polynomial being prime.

Definition 14.6. *$I \subset k[\vec{x}]$ is* **prime** *if $fg \in I \implies f \in I$ or $g \in I$.*

Theorem 14.7. *$V \subset k^n$ an affine variety. V is irreducible if and only if $I(V)$ is prime.*

Corollary 14.8. *k algebraically closed. Then $\mathbb{I}, \mathbb{V}$ are bijections between irreducible varieties and prime ideals.*

Here, $\mathbb{I} : V \to I(V)$ and $\mathbb{V} : I \to V(I)$. This is true simply because when k is algebraically closed, we can factor completely.

Corollary 14.9. *Suppose k is an infinite field. Let $V = \overline{\mathbb{f}(k^m)}$. Then*

$$x_i = f_i(t_1, \ldots, t_n).$$

Then V is irreducible.

Proof. $I(V) = I(\mathbb{f}(k^n))$. $g \in k[x], g \circ \mathbb{f} \in k[\vec{t}]$.

$I(V) = I(\mathbb{f}(k^n)) =$ set of polynomials in $k[\vec{x}]$ whose composition with $\mathbb{f}$ is the zero polynomials in $k[\vec{t}]$.

$$I(v) = \{g \in k[\vec{x}] \ : \ g \circ \mathbb{f} = 0\}.$$

$gh \in I(V)$. $gh \circ \mathbb{f} = 0$. $(g \circ \mathbb{f})(h \circ \mathbb{f}) = 0$. Since this is an integral domain, one of these terms is zero. So V is irreducible. □

Lecture 15

Assigned Reading

The reading for this lecture is §3 of chapter 4 in [**?**].

15.1 Irreducible Varieties and Prime Ideals.

Theorem 15.1. *Let $V \subset k^n$ be an affine variety. Then there exists a unique minimal decomposition into irreducible*

$$V = V_1 \cup \ldots \cup v_m.$$

The trick to the proof is understanding how the process stops. Consider

$$V = V_1 \cup \ldots, \cup V_m$$

.

We get a chain

$$W_1 \supseteq W_2 \supseteq W_3 \ldots$$

if and only if

$$I(W_1) \subsetneq I(W_2) \subsetneq \ldots$$

The ascending chain condition.

$$V = V_1 \cup \ldots \cup V_m = V_1' \cup \ldots \cup V_\ell'$$

irreducible
$V_i = V_i \cap V = \bigcup(V_i \cap V_j')$. Then $V_i = V_i \cap V_j'$. $V_i \subset V_j' \subset V_k$. Squeese by minimality $V_j = V_k = V_1$.
The commutative algebra version of this ideal.

Theorem 15.2. *k an algebraically closed field. Every radical ideal in $k[\vec{x}]$ can be written uniquely as a finite intersection of prime ideals*

$$I = P_1 \cap \ldots \cap P_r,$$

$P_i \not\subset P_j$ for $i \neq j$.

Theorem 15.3. *Every ideal $I \subset k[\vec{x}]$ can be written as a finite intersection of primary ideals.*

Definition 15.4. *An ideal I is primary if $fg \in I$ implies $f \in I$ or $g^m \in I$.*

Theorem 15.5. *(***Lasker-Noether***) Every ideal $I \subset k[\vec{x}]$ has a minimal primary decomposition*

$$I = \cap Q_i$$

with $\sqrt{Q_i}$ distinct.

$$Q_i \not\supset \cap_{i \neq j} Q_j$$

Algorithms for primary decomposition.

Algebra	Geometry
radical ideals	varieties $V(I)$
Add ideals	$\cap$ varieties
prod ideals	$\cup$ varietieis
$I(V(IJ)) = \sqrt{IJ}$	”
intersection $I \cap J$	”
quotients of ideals $I : J$	difference of varieties
elimination of vars	projection of
$\sqrt{I \cap k[x_{\ell+1}], \ldots, x_n]}$	$\overline{\pi(V(I))}$
$k[\vec{x}]/I$	ring of functions
prime ideals	irreducible variety
maximal ideal	variety pts

Next time we discuss tensor networks.

Lecture 16 (Guest Lecture)

Assigned Reading

No reading

This was a guest lecture discussing applications of algebraic geometry in neuroscience. Slides can be found on the website.

Lecture 17 (Guest Lecture)

Assigned Reading

No reading

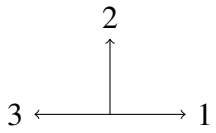
This was a guest lecture on matroids. *VERY* rough notes can be found below.

17.1 Algebraic Matroids. What is a matroid? The idea of a matroid is to generalize ideas of "independence" from different areas of mathematics.

The definition isn't so enlightening. So we will begin with a couple of examples.

What is the first kind of independence you can think of? Linear independence.

Example 26. Linear Matroid $E = \{0, 1, 2, 3\}$ is called our ground set.



- Independent sets $\mathcal{I} = \{\emptyset, \{1\}, \{2\}, \{1, 2\}, \{2, 3\}\}$
- Bases $\mathcal{B} = \{\{1, 2\}, \{2, 3\}\}$
- Circuits $\mathcal{C} = \{\{0\}, \{1, 3\}\}$
- Rank $\rho(s) = \dim \text{span } \mathcal{S}$.

Example 27. Graphic Matroid. Let G be a graph.

- $E = E(G) = \{0, 1, 2, 3\}$
- $\mathcal{I} = \{$ acyclic subragphs$\} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}\}$
- $\mathcal{B} = \{\{1, 2\}, \{2, 3\}\}$
- $\mathcal{C} = \{\{0\}, \{1, 3\}\}$
- $\rho =$ size of largest acyclic subgraph

We now define a matroid using the independence set.

Definition 17.1. *Let E be a finite set. Let $\mathcal{I} \subseteq 2^E$ satisfying*

(1) $\emptyset \in \mathcal{I}$

(2) $I_1 \subseteq I_2$, $I_2 \in \mathcal{I} \implies I_1 \in \mathcal{I}$

(3) *(Augmentation)* $I_1, I_2 \in \mathcal{I}$, $|I_1| > |I_2| \implies \exists x \in I$, *such that* $I_2 \cup \{x\} \in \mathcal{I}$.

Then $(E, \mathcal{I})$ is a matroid.

The main focus of this talk is an algebraic matroid.

Definition 17.2. *Let $k \subset K$ be a field extension. $E \subset K$ finite set.*

$$\mathcal{I} = \{ \text{ algebraic independent subsets of } E\}$$
$$\mathcal{B} = \{\text{transcendence bases of } K\}$$
$$\mathcal{C} = \{\text{minimial algebraic dependent subsets of } E\}$$
$$\rho(S) = \{ \text{ transcendence degree } k(S)/k\}$$

$(E, \mathcal{I})$ is a matroid.

Definition 17.3. *(Matroid of a prime ideal) Let $I \subset k[x_1, \ldots, x_n]$ be a prime ideal. Let $R := k[x]/I$ is an integral domain (sometimes called the coordinate ring). $K = Frac(R)$.*

$$E = \{\bar{x_1}, \ldots, \bar{x_n}\}.$$

$\mathscr{M}(I)$ is the matroid.

Example 28. Let $k = \mathbb{R}$ and $I = \langle x^2 - y \rangle$. This is a prime ideal. What is the associated algebraic matroid? $K = \text{Frac}(\mathbb{R}[x, y]/\langle x^2 - y\rangle)$ $E = \{\bar{x}, \bar{y}\}$ $\mathcal{I} = \{\emptyset, \{x\}, \{y\}\}$ $\rho(M) = 1$. $\mathcal{B} = \{\{x\}, \{y\}\}$ $\mathcal{C} = \{\{x, y\}\}$

Proposition 17.4. *If C is a circuit of the algebraic matroid $\mathcal{M}(I)$ the elimination ideal $I \cap k[C]$ is principal.*

Proof. $I \cap k[C] \subset k[C]$, which has height 1 and codimension 1. UFD implies $I \cap k[C]$ principal. □

Definition 17.5 (KRT)**.** *the circuit polynomial P_c is the unique up to scalar generator of $I \cap k[C]$.*

Proposition 17.6. *The projection of $V(I)$ onto the subspace $k^{|B|}$ for $B \in \mathcal{B}$ is a dominant map with finite fibers.*

Proof. $k(E)$ from $k(B)$ algebraic field extension. What happens when I add in $x \notin B$? Then $B \cup \{x\}$ is dependent. □

Definition 17.7. *The degree of the projection onto $k^{|B|}$ is called the base degree.*

17.2 Computation. In terms of computation, there are two main strategies. The first is differentials.

Theorem 17.8. *An algebraic matroid over a field of characteristic 0 is representable as a linear matroid over $k(T)$ where T is some finite set of transcendentals.*

Example 29.

This method transitions us from algebraic to linear.
The other strategy is Gröbner bases.

Proposition 17.9. *For every circuit C, there is a monomial ordering where the GB contains P_c.*

The ordering we want in particular is the elimination ordering – C-variables are smallest.

Corollary 17.10. *The Universal Gröbner basis contains every circuit polynomial.*

The universal Gröbner basis is the union of all Gröbner bases under every ordering. This set is finite, but very large.

Computing the base degree.

Proposition 17.11. *Base degree is the degree of $I + \ell$ where ℓ is a generic line of projection.*

17.3 Application.

(1) Low-rank matrix completion

$$\begin{pmatrix} 1 & 2 \\ 3 & ? \end{pmatrix}.$$

If we know this is rank 1, then $? = 6$. Then $x_{11}x_{22} = x_{12}x_{21}$. Now consider an $m \times m$ matrix of rank r. Then we consider the $(r+1) \times (r+1)$ minors of the matrix.

What are the bases of the matroid? These are the subsets of the entries that only have a finite number of completions. The circuits of this matroid are minimal subsets that have an algebraic dependency.

The rank of this matroid is minimal number of entries needed to complete the matrix.

Whenever you have partial information about a variety, you want a matroid.

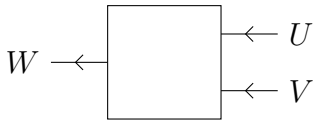
Lecture 18

Assigned Reading

Please see "What is a tensor?" notes.

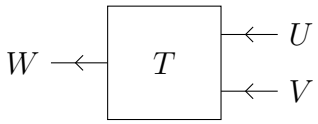
We have occasionally mentioned diagrams which are called "tensor networks." It's a network in the sense that it is a graph whose edges represent vector spaces and whose nodes represent "tensors" (i.e. n-way tables or multi-dimensional arrays).

Suppose we have vector spaces U, V, W. Consider the tensor represented by the box below.



This box (which is unlabeled) indicates *any* tensor in the space $U^* \otimes V^* \otimes W$ since it takes in vectors from $U \otimes V$ and produces vectors in W.[12]

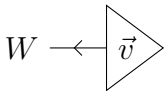
If we wish to discuss a particular tensor $T \in U^* \otimes V^* \otimes W$, we indicate this by labeling the diagram.



If $\dim(U) = 2, \dim(V) = 3$, and $\dim(W) = 4$, then T is a 4 by 6 matrix. We can assign letters i, j, k for the vector spaces (respectively). So the entries of the matrix are t_{ijk}. The matrix is a flattening of this tensor by $ij \mid k$.

While the wires represent vector spaces, they also represent the number of indices one can use to describe the tensor. The number of indices used is referred to as the *order*[13] The tensor T above is order 3.

An order 1 tensor is a vector since it is indexed with one letter. Although this is not necessary, we often distinguish vectors with triangles.[14]



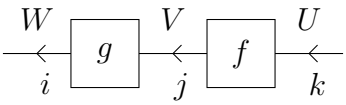
18.1 Combination Rules. Composition, which is also called contraction, tells us how we can compose two tensors.

Let $f : U \to V$ and $g : V \to W$, then $g \circ f : U \to W$. We represent this by writing the tensors in series.

[12]Although we will use the dual notation to indicate the vectors spaces that are being "fed" into the tensor, this is not super important to keep track of. In finite dimensions, we know $U^* \cong U$. The congruence is easy to define whenever we fix a basis.

[13]In online resources, you'll see this referred to as the "rank." This is not the best word to use in math since the rank can mean the matrix rank or, as we will see later, the tensor rank.

[14]This emphasizes that there are no inputs and matches up nicely with bra-ket notation.

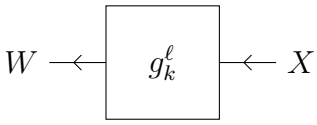


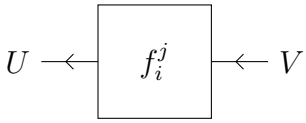
Note that the direction we choose to draw these diagrams does not matter so long we are consistent. We will choose to read the diagram from right to left (to match up with tensor composition) and up to down (to indicate tensor product order[15]).

The maps f and g are tensors. To differentiate which indices correspond to the codomain or domain, we will write them above or below the tensor. For example, f^j_k and g^i_j. The composition is

$$[g \circ f]^i_k = \sum_j f^j_k g^i_j.$$

Writing wires in parallel indicates that the spaces are being tensored. Writing tensors in parallel indicates the tensors themselves are being tensored.





The image above is

$$[f \otimes g]^{j\ell}_{ik} = f^j_i g^\ell_k.$$

In other words, if

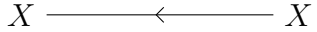
$$f = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{and} \quad g = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

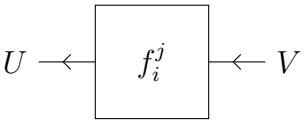
then

$$g \otimes f = \begin{pmatrix} Aa & Ab & Ba & Bb \\ Ac & Ad & Bc & Bd \\ Ca & Cb & Da & Db \\ Cc & Cd & Dc & Dd \end{pmatrix}$$

This representation of the tensor requires a fixed basis and is called the **Kronecker product.** This is the tensor product on linear transformations with a chosen basis.

If we have the identity matrix, we simply draw the line. Hence, the diagram





[15]For example, the previous diagram was interpreted as $U^* \otimes V^*$ instead of the reverse.

corresponds to $Id_X \otimes f_i^j$. Its Kronecker product, assuming $X = U = W = \mathbb{C}^2$ is

$$Id_X \otimes f_i^j = \begin{pmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & a & b \\ 0 & 0 & c & d \end{pmatrix}$$

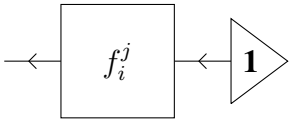
The reverse tensor product is not the same, for obvious reasons.

$$f_i^j \otimes Id_X = \begin{pmatrix} a & 0 & b & 0 \\ 0 & a & 0 & b \\ c & 0 & d & 0 \\ 0 & c & 0 & d \end{pmatrix}$$

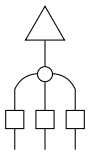
18.2 Special Tensors. There are a set of special tensors that come up often. For example, let

$$\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

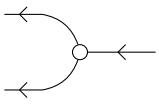
and let f be a matrix. Then the diagram below produces a vector which is the sum of all the columns of f.



Another special tensor is the copy tensor, often referred to as a "copy dot." We've seen it already in the one pocket, three coin model:



We can think of this as taking a vector and producing a copy of it. Let us consider the case where one vector space becomes two.

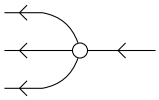


For a two dimensional vector space, the copy dot is represented as the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$$

since it sends $e_1 \mapsto e_1 \otimes e_1$ and $e_2 \mapsto e_2 \otimes e_2$.

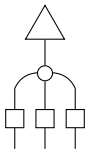
This is generalizable to larger dimensions.



With this picture, the copy dot is

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$$

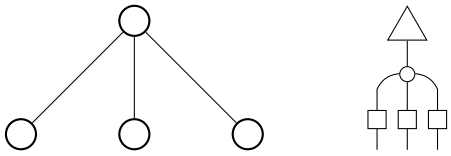
There is a lot more you can do, but what we've covered is enough to explore the main models in statistics and physics. These diagrams represent polynomial parameterizations where the unknown values are the parameters in the tensors. For example, the diagram (which is a segre variety) below is represented by the polynomials



$$x^{ijk} = \sum_{\ell=1}^{d} v^{\ell} a_{\ell}^{i} b_{\ell}^{j} c_{\ell}^{k}$$

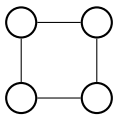
where $a_\ell^i, b_\ell^j, c_\ell^k$ are the parameters from the matrix inputs, r^ℓ is the parameters from the vector input, and x^{ijk} are the variable outputs. The summation was gotten by reading the diagram downward. Since vector spaces and their duals can be identified with one another, we will avoid writing indices are superscripts in the future.

18.3 Graphical Models in Statistics. In a bayesian network, a dot represents a variable–some of which might be hidden. For example, the graph on the left is the bayesian network of the tensor network on the right.

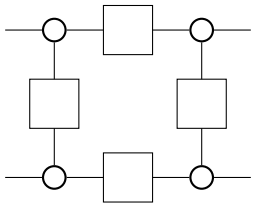


In the case of the pocket model, we might think of the vector as yielding L or R to indicate which pocket.

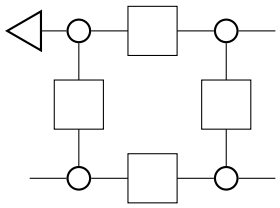
The translation can be understood as the following: the circles must represent vectors. If the vector consists of hidden parameters, draw a triangle. If the vectors are observable, produce a hanging wire. On internal wires, place tensors to indicate the transitions. These transitions will generally be matrices for bayesian networks. For example, suppose the nodes in the graph below are all observable states.



As a tensor network, we draw this as



Now suppose one of the states was a hidden variable. Then we change the tensor network to be



Remember that we have no arrows, so we can choose to read this diagram however we wish. Some, all or none of the variables can be treated as inputs. Since we work in finite dimensions, we do not need to be so careful.

18.4 Probabilistic Independence and Condition Independence. For small cases (maybe four or five statistical variables), these problems are tractable. There are tricks–graphical methods–for finding the implicitization of the parameterization problems we see from the diagrams. To understand these tricks, we need to know something about *probabilistic independence* and *conditional independence*.

Let A, B, C be discrete random variables. Suppose that A is independence of B, written as $A \perp\!\!\!\perp B$. Suppose A takes on states $a \in \{1, 2\}$ and B takes on states $b \in \{1, 2, 3\}$. Since they are independent, then $p_{ab} = p_a p_b$. This is a polynomial equation: $p_{ab} - p_a p_b$. Remember, this is a *set* of polynomial equations for each of the values a and b take on.

To represent this as a bayesian network, we would draw two circles. For tensor networks, we take the tensor product of two vectors (i.e. the independence model). The first is a vector (or state) in k^2 and the second is a vector in k^3.

The ring we are working over is $k[p_{11}, p_{12}, p_{13}, p_{21}, p_{22}, p_{23}]$. For a fixed choice of a and b, our six equations are of the form:

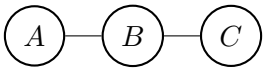
$$p_{ab} - (p_{a1} + p_{a2} + p_{a3})(p_{1b} + p_{2b}).$$

When A and B are both two dimensional, this turns out the be the independent coin problem. There is only one polynomial:

$$p_{11}p_{22} - p_{21}p_{12}$$

This is the determinant of the Kronecker product of the tensor product of two, two-dimensional vectors.

Now let's move on to conditional independence. Suppose I have an undirected model of three observable states.



Exercise 6. Draw the tensor network for this model.

Although it is not true that A and C are independent, I do know that, given B, A and C can be treated as independent. This is written as

$$(A \perp\!\!\!\perp B)\big|C.$$

This means, $p_{ac|b} = p_{a|b}p_{c|b}$. The power of this technique is that we can use this conditional independence as a short-cut.

Exercise 7. Suppose A, B, C take on two states. Based on the tensor network set-up, show that $p_{ac|b} = p_{a|b}p_{c|b}$.

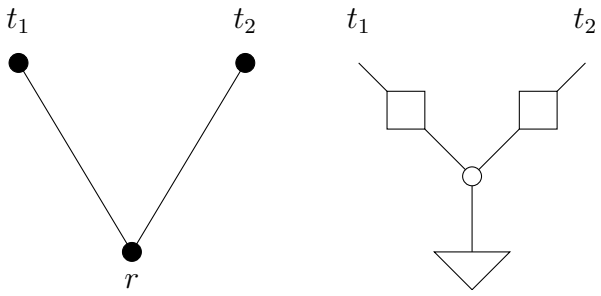
Lecture 19

Assigned Reading

The reading for this lecture is [**?**].

19.1 Algebraic Phylogenetics. There are different ways we can look at the genetic information mathematically. A common approach is to look at important subsequences, also known as *sites*, of the DNA. The assumption is that all sites independently evolve according to the same model (i.i.d.). We can describe each of these subsequences with a vector, $\boldsymbol{\pi} = (\pi_G, \pi_C, \pi_A, \pi_T)$, where π_X is the proportion that X appears. Alternatively, under the i.i.d. assumption, π_X can be the relative frequency X appears across all sites.

Let us consider the simplest example: two organisms springing from an ancestor. Let $\boldsymbol{\pi}_r = (\pi_1, \pi_2, \pi_3, \pi_4)$ be the ancestral sequence that transition to two extant organisms, which we will call *taxa*. Below is the bayesian network and the tensor network describing this. The parameter vector is dimension 4 and the transition matrices are 4×4.



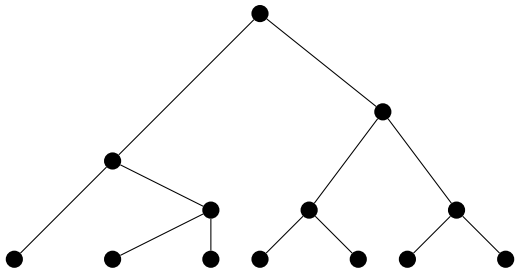
Then the joint probability distribution is:

$$P(t_1 = i, t_2 = j) = p_{ij} = \sum_{r=1}^{4} \pi_r A_{ri} B_{rj}$$

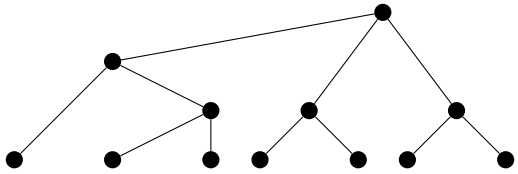
where A_{ij} and B_{ij} are the entries of the transformations A and B, respectively.

These tree models were among the first applications for algebraic statistics, so we have a lot of information on them. In some sense, they are the simplest example of (directed) bayesian networks because they take the shape of a tree. Because so much is known about trees, we often reduce more complex problems to trees.

19.2 Trees. The top most node of a tree is called the **root**. The bottom-most nodes are **leaves**. In the case of phylogenetics, the leaves consist of observable data (DNA of extant species). The internal nodes (which represent extinct species) are hidden parameters.



This is called a tree because as an undirected model, it has no cycles. Note that mathematically, we can simplify this picture by eliminating the root and it is still equivalent to the original problem.

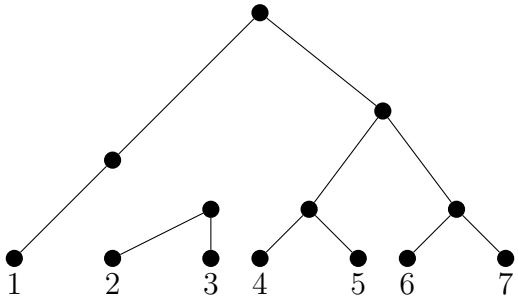


In the bayesian model, this is not an obvious fact; however, if you draw the tensor network version of this tree and contract the root (just a vector) into one of the matrices, we get an equivalent variety.

Exercise 8. Draw the tensor network version of this diagram and justify why the same variety arises even when we contract the root.

Because of this fact, we can always assume that every vertex has degree three except the leaves.

Each edge in the Bayesian network defines a split for the leaf labels. That is, let V be the set of leaves of the tree. A deletion of an edge partitions V into two sets, the leaves of one subtree and the leaves of the second subtree. For example, the split below partitions V into a set $A = \{2, 3\}$ of two leaves and a set $B = \{1, 4, 5, 6, 7\}$ of 5 leaves.



Given two edges e_1 and e_2, we can consider the compatibility of their **split systems**, the partitions of V arising from each edge.

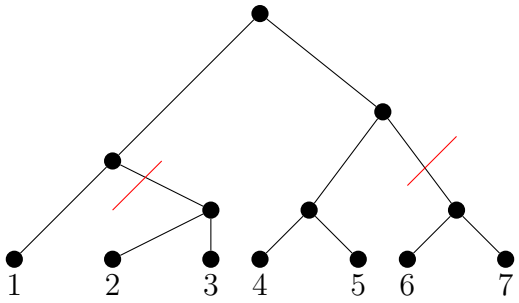
- What should compatibility mean here?
- Can the tree be recovered from the possible split systems?

Definition 19.1. *Two splits, $A|B$, $C|D$ are* **compatible** *if at least one of the four possible intersections:*

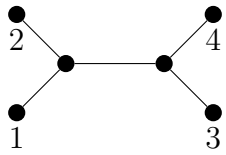
$$A \cap C, A \cap D, B \cap C, B \cap D$$

is empty.

For example, consider the tree below and the two splits (indicated by the red lines). Then $A = \{2, 3\}$, $B = \{1, 4, 5, 6, 7\}$, $C = \{1, 2, 3, 4, 5\}$ and $D = \{6, 7\}$. Hence, $A \cap D = \emptyset$ and the splits are compatible.



Now consider the tree below.



Suppose we have the splits $A = \{1, 2\}$, $B = \{3, 4\}$ and $C = \{2\}$, $D = \{1, 3, 4\}$. Then $C \cap B = \emptyset$.

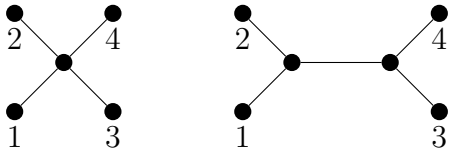
Lemma 19.2. *Given two splits $A|B$ and $C|D$ arising from two edges of a tree, then they are compatible.*

The proof is a homework exercise.

Now, given a compatible split system, can you recover the tree? The answer is yes if we add some assumptions.

Theorem 19.3. *(***Splits Equivalent Theorem***) A collection of splits S is pairwise compatible if and only if there exists a tree T such that $S = Splits(T)$. If T exists, then it is unique.*

If there are no splits, then we let all labels be attached to one node. Furthermore, if my tree is not trivalent, we can create an equivalent trivalent tree (which is what we will restrict to). For example, we can turn the tree on the left into the tree on the right.



We have one additional requirement. Suppose we have a function called $label$ which maps labels to the tree T. Then the "unique" tree that is produced from this theorem is $(T, label)$, which is sometimes called an x-tree. Otherwise uniqueness can be a problem by simply permuting labels.

The proof is by induction.

Proof. Suppose $Splits = \emptyset$. Then all labels in our set are assigned to a tree of one node (and no edges). Now suppose the set $Splits$ has $n + 1$ splits. Consider $S' = Splits \setminus \{A|B\}$. By assumption, there exists T' for S'.

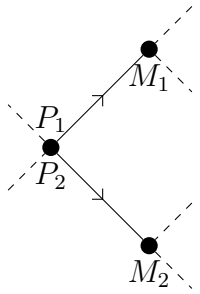
Definition 19.4. *Let $A|B$ and $C|D$ be compatible splits. A is* **mixed** *with respect to $C|D$ if $A \cap C \neq \emptyset$ and $A \cap D \neq \emptyset$. Otherwise, A is called* **pure***.*

Claim: For compatible splits, exactly one is pure and one is mixed.

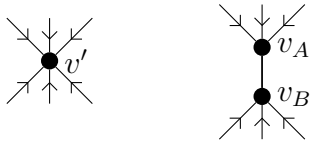
Exercise 9. Justify this claim.

Consider every split in S'. Every split has one pure side and one mixed side with respect to $A|B$. Call $M_i|P_i$. Orient e_i in T' so that $P \to M$. In other words, orient e_i so that pure leaf set is entering into e_i and the mixed leaf set is exiting e_i. This essentially means we draw arrows and turn T' into a directed graph.

We claim that no node of T' has out degree bigger than or equal to 2. If such a thing existed, then there would be a split when a partition that is both mixed and pure.



Hence, every node has out degree 0 or 1. A tree is a graph whose number of nodes equals its number of edges plus 1. So T' has a unique sink v' (Why?). Replace v' with an edge e whose split gives rise to $A|B$.



The resulting new tree, T, is such that $S = Splits(T)$ □

Remark 19.5. The set $Splits(T)$ will contain *every* split system.

Now that we've talked about some of the combinatorial properties of trees, let us discuss the **General Markov model** associated to T. This is the three model used in algebraic phylogenetics.

Let T be a rooted tree that is directed from the root to the leaves. Associate to each vertex a random variable with κ states. And consider the internal vertices, including the root, as hidden parameters. The leaves are observable (i.e. treated as variables).

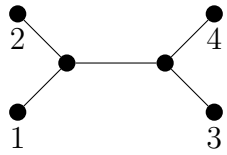
If everything were visible, then the probability of the outcome would be calculated via the tensor network contraction. The matrices in the model are Markov (meaning they represent a probabilistic transformation). So in the A model at the beginning of this lecture's notes, the matrices A and B are Markov.

The Bayesian network model focuses more on states (i.e. DNA vectors) than on transformations (i.e. Markov matrices). So another way to think about this tree is to say that the probability of an outcome is based on the probability of the root p_r multiplied by the probabilities of each extinct ancestor p_u (internal node) given that species ancestors $p_a(u)$ (the node's parents). So the probability at the leaves can be written as the following.

$$p_{\ell_1,\ldots,\ell_n} = \sum_{v \in int(V(T))} \sum_{S_v} p_r \prod_{u \in T \setminus \{r\}} p_u | p_a(u)$$

For example, suppose instead of four states G,C,A,T we consider only two states GC, AT (the pairings of DNA). This is the binary model. In the tree below, the total possible states of the leaves

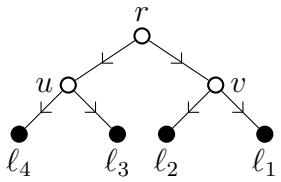
are 2^4 (e.g. $\langle GC, GC, GC, GC\rangle$, or $\langle AT, GC, AT, AT\rangle$, etc). There are 2^6 states for all nodes (leaves and interior). And there are 2^2 states for the internal nodes.



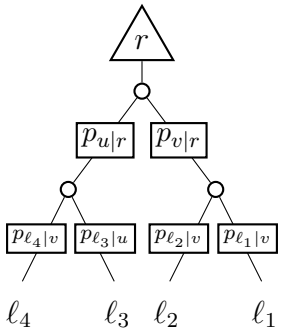
This is equivalent to the tensor network description, which focuses on the evolutionary transition. It is of the opinion of this writer, that the tensor network model is much easier to understand and define.

Remark 19.6. At this point, it is helpful to review the tensor network diagram discussion in the previous lecture. Make sure you know what a copy dot tensor is. Remember that it maps $e_1 \mapsto e_1 \otimes e_1$ and $e_2 \mapsto e_2 \otimes e_2$. Also review how to convert a Bayesian network picture to the tensor network picture. As may already be apparent, the tensor network picture is often easier to parametrize.

In the tensor network model, these Markov matrices tell us the condition probability distributions of the internal nodes. For example, the tree below



has a tensor network diagram depicted below. We have labeled the Markov transformations according to what the represent: conditional probabilities.



The implicitization problem built into this set up answers the following question: What are all the probability distributions we can get from this set up?

When the model is a two-state model, we can derive this information using only the flattenings according to the edge-splits of the tree.[16] Since the possible probability distributions P are tensors in $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \ldots \otimes \mathbb{C}^2$, we can pick how we flatten that tensor into a matrix.

For example, in the tree depicted above, then tensor is an array parametrized by elements p_ijkl where $i, j, k, l \in \{0, 1\}$ where perhaps $0 = GC$ and $1 = AT$. Let i correspond to the index of ℓ_1, j to ℓ_2, etc. Consider the split according to the edge between u and r. Then the corresponding

[16]When there are more than two states, the flattenings only give a partial generating set of the ideal in question.

flattening gives the matrix

$$\begin{pmatrix} p_{0000} & p_{0001} & p_{0010} & p_{0011} \\ p_{0100} & p_{0101} & p_{0110} & p_{0111} \\ p_{1000} & p_{1001} & p_{1010} & p_{1011} \\ p_{1100} & p_{1101} & p_{1110} & p_{1111} \end{pmatrix}$$

One dimension of the matrix considers the possible state combinations of $\ell_1 \otimes \ell_2$ and the other $\ell_3 \otimes \ell_4$. To formalize this process, we assume trees are trivalent (meaning internal nodes have degree three (with the exception of the root) using the process discussed earlier.

Hence the split system gives rise to a set of matrices of the variables. The rank of these matrices is precise the number of states, κ. When studying the two-state model, the rank of these matrices is 2. Hence all the 3×3 minors of all flattenings must lie in the ideal corresponding to the model's variety.

A theorem we will reference a lot is the Allman-Rhodes-Draisma-Kuttler Theorem, which states that for the two state model, these 3×3 minors generate the ideal. This was only proved in the last ten years.

Lecture 20

Assigned Reading

The reading for this lecture is the "Group Action" article on wikipedia.

Remark 20.1. These notes will only cover the important information from this lecture.

The elements of a group are often functions. For example, the symmetric group consists of functions which permute elements. So often we think of a group G as *acting* on a set X.

Example 30. The group $GL_2(\mathbb{R})$ acts on the vector space $\mathbb{R}^2$. For each $A \in GL_2(\mathbb{R})$, the action of A on $\mathbb{R}^2$ is defined by $v \mapsto Av$.

For an element $g \in G$ and $x \in X$, the action of g on x is often written $g.x$ or just gx. The "." is helpful notation when the set and the group are hard to distinguish.

Example 31. The group $GL_2(\mathbb{R})$ acts on the vector space $\mathbb{R}^2 \otimes \mathbb{R}^2$ (i.e. the space of 2×2 matrices. In this case the group is the set, so the distinction can be fuzzy.

When thinking of group actions, we're generally interested in what subsets of X "respect" the action. By that, we mean, is there an $A \subsetneq X$ such that $G \curvearrowright A \subseteq A$? Here $G \curvearrowright A$ is the set of all possible elements $g.a$ for $g \in G$ and $a \in A$.

For example, the span of an eigenvector respects the action of a particular matrix. Consider the group

$$G = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\}.$$

This group acts on $\mathbb{R}^2$. A subspace of $\mathbb{R}^2$ that respects this action is

$$A = \text{span} \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}.$$

Another is

$$B = \text{span} \left\{ \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}.$$

We say that A is **invariant** under the group action of G. We can also define functions of the entries of vectors in $\mathbb{R}^2$ which are **invariants**. For example, let $\vec{v} = (v_1, v_2)^T$. A function $f(v_1, v_2) = v_1 + v_2$ is an invariant since $f(v) = f(g.v)$ for any $g \in G$ and $v \in V$. A function which is not invariant is $h(v_1, v_2) = 2v_1 + v_2$.

Exercise 10. Try to characterize the invariant polynomials for this example. Make sure to consider higher degrees.

These invariants often provide insight into the geometry of these invariant subspaces of X. In the above example, what is the variety of the set of invariants?

20.1 Orbits and Stabilizers. Occasionally, we will reference orbits and stabilizers.

An **orbit** is the set of all possible elements to which an element $x \in X$ can be sent according to the group action.

$$G.x = \{g.x \mid g \in G\}$$

A **stabilizer** is the set of elements $g \in G$ that fix a particular point $x \in X$.

$$G_x = \{g \in G \mid g.x = x\}$$

Let $G = GL_2(\mathbb{R})$ and $X = \mathbb{R}^2$. The orbit of $(0,0)$ is $\{(0,0)\}$. The orbit of $\{(1,0)\}$ is $\mathbb{R}^2 \setminus \{(0,0)\}$. (This actually tells us that X has only two invariant sets. The stabilizer of $(1,0)$ is the set of matrices

$$\left\{ \begin{pmatrix} 1 & x \\ 0 & y \end{pmatrix} \middle| \, y \neq 0 \right\}.$$

Lecture 21

Assigned Reading

Suggested reading for this lecture is Blake and Kohli's introduction to Markov Random Fields (`https://mitpress.mit.edu/sites/default/files/titles/content/9780262015776_sch_0001.pdf`).

21.1 Graphical Models. Today we'd like to discuss undirected graphical models, which will overlap heavily with the last two lectures. These undirected graphical models, which are also called Markov random fields, illustrate a number of important concepts that work in other settings of Applied Algebraic Geometry. In particular, toric varieties are one such setting, which are widely used.

Fully observed graphical models (with no hidden variables) are generally harder than the hidden variable models that we've seen before since implicitization is easier.

There are two ways to think about these models:

(1) (The usual way.) As the image of a polynomial parameterization

$$\mathbb{f} : \mathbb{R}^d \to \mathbb{R}^m \supset \Delta_{m-1}$$

where $\mathbb{R}^d$ is a parameter space.

(2) (The new way.) A collection of conditional independence statements.

Conditional independence is important because it is a way to formalize locality in the model. These are statements that say one part of a model is unrelated to another part of the model except for some notion of "in between." So in probability, and, in particular the undirected fully observed case, conditional independence is almost a complete description of what it means to be local in the model. We'll see later that we don't really have this notion for quantum models. There are plenty of open problems on how to extend this idea of conditional independence to models with hidden variables.

Suppose I have three random variables x_1, x_2, x_3 with finitely many states, $d_1, d_2, d_3 \in \mathbb{Z}$ (respectively). Furthermore, suppose $x_1 \perp\!\!\!\perp x_2 \big| x_3$, meaning x_1 and x_2 are independent given x_3. This means for each state $x \in X_3$, we have a table where rows are indexed by possible states of X_1 and columns are indexed by possible states of X_2. The table itself is indexed by one possible states of X_3. So if there are five possible states of X_3, then we have five such tables. Inside this table are probabilities corresponding to each possible state of X_1 and X_2 given the state of X_3 that indexes the table.

For example, suppose X_3 has two states, X_2 has three states and X_1 has four states. Then our set up yields two 4×3 matrices.

$$\begin{pmatrix} P(X_1=1,X_2=1|X_3=1) & P(X_1=1,X_2=2|X_3=1) & P(X_1=1,X_2=3|X_3=1) \\ P(X_2=1,X_2=1|X_3=1) & P(X_2=1,X_2=2|X_3=1) & P(X_2=1,X_2=3|X_3=1) \\ P(X_1=3,X_2=1|X_3=1) & P(X_1=3,X_2=2|X_3=1) & P(X_1=3,X_2=3|X_3=1) \\ P(X_1=4,X_2=1|X_3=1) & P(X_1=4,X_2=2|X_3=1) & P(X_1=4,X_2=3|X_3=1) \end{pmatrix},$$

$$\begin{pmatrix} P(X_1=1,X_2=1|X_3=2) & P(X_1=1,X_2=2|X_3=2) & P(X_1=1,X_2=3|X_3=2) \\ P(X_2=1,X_2=1|X_3=2) & P(X_2=1,X_2=2|X_3=2) & P(X_2=1,X_2=3|X_3=2) \\ P(X_1=3,X_2=1|X_3=2) & P(X_1=3,X_2=2|X_3=2) & P(X_1=3,X_2=3|X_3=2) \\ P(X_1=4,X_2=1|X_3=2) & P(X_1=4,X_2=2|X_3=2) & P(X_1=4,X_2=3|X_3=2) \end{pmatrix}$$

For each of these tables, there are a set of equations which are satisfied by this table. This is precisely because the probabilities are independent! That is,

$$P(X_1=i,X_2=j|X_3=k)=P(X=i|X_3=k)P(X_2=j|X_3=k).$$

Hence, these tables are derived as outer product. For example, the first table above can be written as

$$\begin{pmatrix} P(X_1=1|X_3=1) \\ P(X_1=2|X_3=1) \\ P(X_1=3|X_3=1) \\ P(X_1=4|X_3=1) \end{pmatrix} \otimes \begin{pmatrix} P(X_2=1|X_3=1) \\ P(X_2=2|X_3=1) \\ P(X_2=3|X_3=1) \\ P(X_2=4|X_3=1) \end{pmatrix}.$$

In full generality, we're given a set of random variables $X_1,\ldots,X_n$ each with the number of states $d_1,\ldots,d_n$. A conditional independence statement about this collection of states is a statement of the form $(A \perp\!\!\!\perp B)|C$ where A,B,C are pairwise disjoint subsets of the random variables. This "chunking" of the random variables translates into the cartesian product. So if $A=\{X_1,X_2,X_3\}$, we treat the random variable for A as $X_A=X_1\times X_2\times X_3$.

So each conditional independence statement defines a set of quadratics (degree two equations since the matrices are rank 1) in the unknown variables representing these joint probabilities.

$$P(A=a,B=b,C=c)P(A=a',B=b',C=c)-P(A=a',B=b,C=c)P(A=a,B=b',C=c)$$

These are the same kinds of invariants we saw from the independent coin problem. For example, suppose $X_1 \perp\!\!\!\perp \{X_2,X_3\}$ where each has two states. Then the frame is

$$\begin{pmatrix} p_{000} & p_{001} & p_{010} & p_{011} \\ p_{100} & p_{101} & p_{110} & p_{111} \end{pmatrix},$$

where $p_{000}=P(A=1,B=1,C=1)$, etc. The ideal of this variety is generated by the 2×2 minors of this matrix.

Example 32. Suppose $(X_2 \perp\!\!\!\perp X_3)\big|x_1$. Then the ideal is generated by

$$I=\langle p_{000}p_{011}-p_{001}p_{010}, p_{100}p_{111}-p_{101}p_{110}\rangle$$

$V(I)$ is the variety cut out by the ideal. Unfortunately, this contains points not in the model (i.e. points which are negative or complex, points which sum to more or less than 1, etc).

Hence, we do not study $V(I)$ but the intersection of $V(I)$ with the probability simplex, denoted $V(I)\cap\Delta$. This intersection will not perfectly fit the parametrization map, but will contain it. It should also contain limit points which may not directly arise form the model. So $V(I)\cap\Delta$ is a larger set than Im $(\mathbb{f})$. It is the Zariski closure of the model restricted to the simplex.

Definition 21.1. *The* **probability simplex** *in $\mathbb{R}^n$ is the space of vectors such that $\sum v_i=1$ and $0\le v_i\le 1$. In $\mathbb{R}^2$, this is a line segment. In $\mathbb{R}^3$, this is a triangle. In $\mathbb{R}^4$, this is the tetrahedron (see Lecture* **??***).*

A conditional independence model is a big set of conditional independences.

$$\mu-\{(A^{(i)} \perp\!\!\!\perp B^{(i)})\big|C^{(i)}, i=1,\ldots,s\}$$

Given this collection, we get an ideal which is the sum of the ideals generated by the conditional indepdences.

$$I_\mu = \sum_{i=1}^{s} I_{(A^{(i)} \perp\!\!\!\perp B^{(i)})|C^{(i)}}$$

The variety is gotten by taking the intersection with the probability simplex. $V(I_\mu)$,

$$V_\Delta(\mu) = V(I_\mu) \cap \Delta.$$

Example 33. For 3 binary random variables suppose

$$\mu = \{(X_1 \perp\!\!\!\perp X_2)\big|X_3, (X_1 \perp\!\!\!\perp X_3)\big|X_2\}$$

This gives rise to the invariants

- $p_{000}p_{110} - p_{010}p_{100}$,
- $p_{001}p_{111} - p_{011}p_{101}$,
- $p_{000}p_{101} - p_{001}p_{100}$,
- $p_{010}p_{111} - p_{011}p_{110}$,

The intersection of this model with the probability simplex $V_\Delta(\mu)$ has three components. Two are tetrahedra: faces on the seven dimensional simplex. The third is a component lies in the interior of the seven dimensional simplex (denoted Δ_7°).

(1) $\{p \in \Delta \ : \ p_{001} = p_{010} = p_{101} = p_{110} = 0\}$
(2) $\{p \in \Delta \ : \ p_{000} = p_{011} = p_{100} = p_{111} = 0\}$
(3) $V_\Delta(X_1 \perp\!\!\!\perp \{X_2, X_3\}) \subset \Delta_7^\circ$

How do you determine these three components? One way is to reason it out as you did on the midterm. The other is to type it into SAGE or an analogous software.

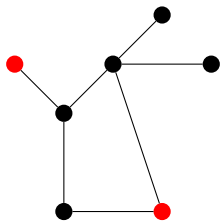
This last piece is the most interesting. It says if all $p_{ijk} > 0$, then $(X_1 \perp\!\!\!\perp X_2)\big|X_3$ and $(X_1 \perp\!\!\!\perp X_3)\big|X_2$ implies $X_1 \perp\!\!\!\perp \{X_2, X_3\}$

The implication in the above example–where the conditional independence statements gave rise to another conditional independence statement–is generally pretty subtle. There is some ways to axiomatize them via a graphoid or a semigraphoid. When these were first discovered, it was thought that these implications could be completely axiomitized; however, this turns out to be false (a result proven in the 1970s by Studeny). So this sort of implication is not guaranteed and requires a little work to find when possible.

21.2 Graphiod and Graph Separation. Let G be an undirected graph. Label the vertices by random variables $X_1, \ldots, X_n$. Let μ_G be a set of conditional independence statements, like

$$(X_i \perp\!\!\!\perp X_j)\big|\{X_1, \ldots, X_n\} \setminus \{X_i, X_j\}.$$

For example, if the two random variables represented by red nodes are independent of each other given all the other nodes.



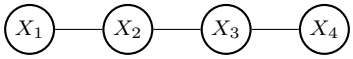
These conditional independence statements are often called "atomic" or "pairwise." This says that two random variables are conditionally independent if their respective nodes are not connected by an edge.

Definition 21.2. *The* **implicit undirected graphical model** *defined by a graph G is the set of probability distributions*

$$V_{\Delta^\circ}(\mu_G) = \Delta^\circ \cap V(I_{\mu_G})$$

If you take a class on graphical models or machine learning, this is not what you would see as the definition. It is typically set up in a parametric way.

Example 34. Suppose our graph is a chain.



Then

$$\mu_G = \{(1 \perp\!\!\!\perp 3)\big|\{2,4\}, (1 \perp\!\!\!\perp 4)\big|\{2,3\}, (2 \perp\!\!\!\perp 4)\big|\{1,3\}\}.$$

This yields 12 quadratic equations total, four coming from each conditional statement.

Why? Because each statement is only a pair of binary random variables. There are 4 states for each conditioning set. So that gives rise to the four equations per conditioning statement. For example, the statement $(1 \perp\!\!\!\perp 3)\big|\{2,4\}$ gives rise to the statements

$$p_{0010}p_{1000} - p_{0000}p_{1010},\ p_{0011}p_{1001} - p_{0001}p_{1011}.$$

These models are examples of

- a Toric model,
- a discrete exponential family, or
- a log-linear models.

The key structure of this model is that it can always be described by a matrix of nonnegative integers such that all the column sums are equal.

$$\mathcal{A} \in \mathbb{Z}_{\geq 0}^{d \times m}$$

We think about this matrix as organizing the exponent vectors. That is, the j^{th} column vector, a_j, corresponds to the monomial

$$\theta^{a_j} = \prod_{i=1}^{d} \theta_i^{a_{ij}}.$$

In other words, the matrix is encoding the monomial parametrization. Because of this condition that all column sums are equal, the monomials all have same degree.

Example 35. Let the rows be indexed by $\theta_1, \theta_2, \ldots, \theta_5$ and the columns indexed by $p_{11}, p_{12}, p_{13}, p_{21}, p_{22}, p_{23}$. Lines are added to highlight the blocks within the matrix.

$$\mathcal{A} = \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ \hline 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right)$$

The toric model or exponential family model from this matrix $\mathcal{A}$ can be described by a polynomial parametrization from $\mathbb{f} : \mathbb{R}^d \to \mathbb{R}^m$ such that

$$\theta \mapsto \frac{1}{z}(\theta^{a_1}, \ldots, \theta^{a_m})$$

where θ is the vector of parameters and

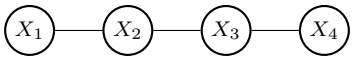
$$p_j = \frac{\theta^{a_j}}{z} \quad \text{where} \quad z = \sum_{j=1}^{n} \theta^{a_j}.$$

Essentially z is whatever normalizes the answer.

Once you have the $\mathcal{A}$ matrix of the model, you have all the information you need.

Theorem 21.3. $\dim(\mathbb{f}(\mathbb{R}^d_{>0})) = \textit{rank}(\mathcal{A}) - 1$.

Let us consider the chain example once more.



We label the columns based on the possible states $0000, 0001, \ldots, 1111$. The rows are factors represented by the lines, $\theta_{12}, \theta_{23}, \theta_{14}$. Each of these θ_{ij} is a2×2 matrix whose entries correspond to states of the two indices. The matrix $\mathcal{A}$ is a series of zeros and ones–ones are entered when the entry in θ_{ij} matches the global state. For this particular problem we get a 12×16 matrix:

$$\mathcal{A}_G = \left(\begin{array}{cccccccccccccccc}
1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\
\hline
1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
\hline
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1
\end{array}\right)$$

Theorem 21.4. *(Undirected Hammersley Clifford) The implicit model $V_{\Delta^\circ}(\mu_G)$ equals the image of the parametrization given by $\mathcal{A}_G$.*

Essentially, we can read off the implicitization by using conditional independence to construct $\mathcal{A}$.

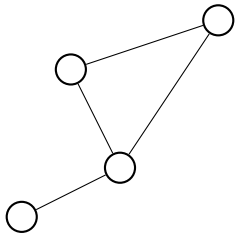
Lecture 22

Assigned Reading

Suggested reading for this lecture is ... I don't know.

Last time, we ended on the implicitization theorem by Hammersley and Clifford. A requirement for this theorem is that all the entries were positive; that is, $p_{ijk\ell} > 0$. When the probabilities are strictly positive, then there exists a set of parameters that maps to them.

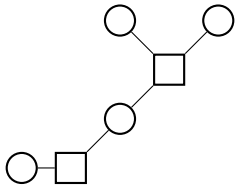
Suppose you have an undirected graph.



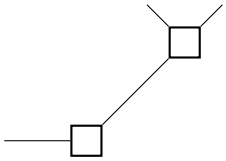
You can create the factor graph or clique graph.

Definition 22.1. *A* **clique** *of an undirected graph is an subset of vertices C such that any two vertices in C are connected.*

The above graph has two cliques. And for every clique, we draw a box with with the number of edges equaling the size of the clique. For this image, the clique graph is the following:

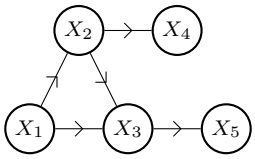


Notice that the factor graph is a lot like a tensor network and, it turns out, there is a way to interpret them as such. The tensor network is depicted below and it gives rise to the parameterization from $\mathcal{A}_G$ discussed in the last lecture.



Unfortunately, to allow include $p_{ijk\ell} = 0$ turned out to be a difficult problem. The extension to the case where $p_{ijk\ell} \geq 0$ came about 20 years later and utilizes topics in toric geometry that will not be covered in this class.

For directed graphical models (or Baysian networks), there is a similar theorem which does allow for $p_{ijk\ell} \geq 0$. Let D be a directed, acyclic graph. As before, let the nodes represent random variables $X_1, \ldots, X_n$ with finitely many states, $d_1, \ldots, d_n$, where d_i is the number of states for X_i.



Parents of a node X_i will be denoted as $pa(i)$. For example, $pa(3) = \{X_1, X_2\}$. The nodes which are *not* descended from X_i minus $pa(i)$, we will denote as $nd(i)$. For example, $nd(3) = \{X_4\}$.

To D we associate a conditional independence model. By a conditional independence model, we mean a set of conditional independence statements (similar to those we saw in last lecture). This model is the set of statements

$$\mu_D = \left\{ [X_i \perp\!\!\!\perp nd(X_i)] \middle| pa(X_i) \ : \ \forall i \in [n] \right\}.$$

For example, for the above picture, we get the set of statements:

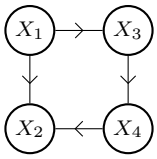
- $(X_3 \perp\!\!\!\perp X_4) \big| \{X_1, X_2\}$
- $(X_4 \perp\!\!\!\perp \{X_1, X_3, X_5\}) \big| X_2$
- $X_5 \perp\!\!\!\perp \{X_1, X_2, X_4\}) \big| X_3$

Much like for undirected graphs, the probabilistic variety cut out by this model is

$$V_\Delta(\mu_\Delta) := \Delta \cap V(I_{\mu_D})$$

where, as before, each conditional independence statement gives rise to a degree two equation (or set of equations) that defines the variety $V(I_{\mu_D})$.

Example 36. Suppose there are four random variables, each with two states and modeled by the directed graph D:



Then the independence statements are

- $(X_2 \perp\!\!\!\perp X_3) \big| X_1$
- $(X_4 \perp\!\!\!\perp X_1) \big| \{X_2, X_3\}$

Using the method from last time, we can translate these statements into the following equations. The first conditional statement has no mention of X_4. As a result, the entries corresponding to $X_2 = j, X_3 = k, X_1 = i$ will be

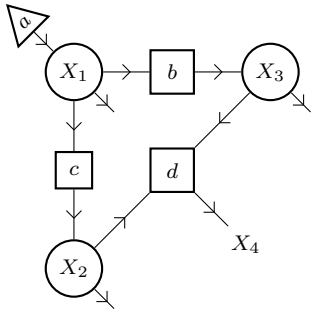
$$p_{ijk0} + p_{ijk1}.$$

- $(p_{0000} + p_{0001})(p_{0110} + p_{0111}) - (p_{0010} + p_{0011})(p_{0100} + p_{0101})$
- $(p_{1000} + p_{1001})(p_{1110} + p_{1111}) - (p_{1010} + p_{1011})(p_{1100} + p_{1101})$
- $p_{0000}p_{1001} - p_{0001}p_{1000}$
- $p_{0010}p_{1011} - p_{0011}p_{1010}$
- $p_{0100}p_{1101} - p_{0101}p_{1100}$
- $p_{0110}p_{1111} - p_{0111}p_{1110}$

From SAGE, one can see that $V_\Delta(I_{\mu_D})$ is a 9-dimensional variety inside the simplex $\Delta_1 5$. The parameters correspond to conditionals, which are of the form

$$P(X_i = a \big| pa(i) = \vec{c}).$$

Instead of writing out all the parameters, we can simply write the tensor network diagram, which communicates what the parameters should be. *Remember that the matrices in this diagram must have columns that sum to one, otherwise it is not probabilistic.*



where

$$c = \begin{pmatrix} c_1 & c_2 \\ 1-c_1 & 1-c_2 \end{pmatrix}, \quad b = \begin{pmatrix} b_1 & b_2 \\ 1-b_1 & 1-b_2 \end{pmatrix}, \quad d = \begin{pmatrix} d_{00} & d_{01} & d_{10} & d_{11} \\ 1-d_{00} & 1-d_{01} & 1-d_{10} & 1-d_{11} \end{pmatrix}.$$

The map described by the diagram above yields a $2 \times 2 \times 2 \times 2$ tensor for each set of parameters. So its image is isomorphic to k^{16}.

According to the image, $p_{0010} = a(1-b_1)c_1d_{01}$. We calculate this by considering what each state gives us. Explicitly, here is how we choose the coefficients:

- $\vec{a}$
  - If $X_1 = 0$, then a.
  - If $X_1 = 1$, we choose $1 - a$.
- b
  - If $X_1 = 0$ and $X_3 = 0$, then b_1.
  - If $X_1 = 0$ and $X_3 = 1$, then b_2.
  - If $X_1 = 1$ and $X_3 = 0$, then $1 - b_1$.
  - If $X_1 = 1$ and $X_3 = 1$, then $1 - b_2$.
- c
  - If $X_1 = 0$ and $X_2 = 0$, then c_1.
  - If $X_1 = 0$ and $X_2 = 1$, then c_2.
  - If $X_1 = 1$ and $X_2 = 0$, then $1 - c_1$.
  - If $X_1 = 1$ and $X_2 = 1$, then $1 - c_2$.
- d
  - If $X_2 = 0$, $X_3 = 0$ and $X_4 = 0$, then d_1.
  - If $X_2 = 0$, $X_3 = 1$ and $X_4 = 0$, then d_2.
  - If $X_2 = 1$, $X_3 = 0$ and $X_4 = 0$, then d_3.
  - If $X_2 = 1$, $X_3 = 1$ and $X_4 = 0$, then d_4.
  - If $X_2 = 0$, $X_3 = 0$ and $X_4 = 1$, then $1 - d_1$.
  - If $X_2 = 0$, $X_3 = 1$ and $X_4 = 1$, then $1 - d_2$.
  - If $X_2 = 1$, $X_3 = 0$ and $X_4 = 1$, then $1 - d_3$.
  - If $X_2 = 1$, $X_3 = 1$ and $X_4 = 1$, then $1 - d_4$.

We prioritize according to which variables feed in versus those that are produced.

There are a total of 9 parameters: $q = 2^0 + 2^1 + 2^1 + 2^2$, each term corresponding to a, b, c, d respectively.

To make this projective, we often define parameters to replace those of the form $1 - \#$. For example, $\vec{a} = (a_1, a_2)^T$ instead of $(a, 1 - a)$.

A natural question is, what is the dimension of the model? Like the undirected version, this is fairly easy to define.

$$\dim(V_\Delta(\mu_D)) = \sum_{i=1}^{n}(d_i - 1) \prod_{d_j \in p(i)} d_j.$$

In the above example, we consider the following facts.

- X_1 has two states and no parents.
- X_2 has two states and one parent.
- X_3 has two states and one parent.
- X_4 has two states and two parents.

Thus,

$$\dim(V_\Delta(\mu_D)) = 1 + 1(2) + 1(2) + 1(2 + 2) = 9.$$

Another question we may be, what is the monomial map? This is precisely the construction we used earlier, but with the homogeneous version of the problem:

$$p_{\vec{x}} = \prod_{i=1}^{n} \theta_{\vec{x},\vec{x}_p(i)}$$

So, in the previous example, suppose we replace variables like $1 - b_1$ with $\hat{b}_1$. Then one of our monomials is

$$p_{0010} = a\hat{b}_1 c_1 d_{01}.$$

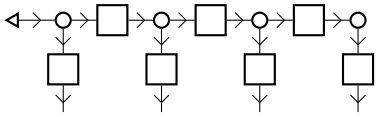
Exercise 11. Write out the monomials for $p_{0000}, p_{1111}, p_{0101}, p_{1010}$, and p_{0011}.

It is important to remember that these monomials represent conditional probabilities. For example, b_1 is the probability that $X_3 = 0$ given that $X_1 = 0$.

Theorem 22.2. *(Directed Hammersley-Clifford) The directed graphical model $V_\Delta(\mu_D)$ equals the image of the parameter space Θ.*

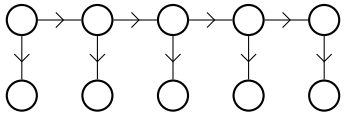
For trees, the directed and undirected models get the same answer. This means that moving the location of the root does not change the tree.

In the tree example, the undirected graphical model example, and the directed graphical model example, we have always allowed the parameters in every tensor to be different. In many models, this is not the case. For example, a hidden Markov model and its corresponding string diagram have repeated items.



In this model, we have a root distribution which is a hidden variable that is never observed. This could be, for example, the state of the economy. Then, there are emission matrices which measures, for example, the probabilities of changes in a stock given the state of the economy. In addition, there are transition transformations (internal) that produces the hidden state in the next time step. Then the same emission matrix produces the influence of the new state on the stock.

The directed graphical model version of the same model is depicted below. Remember, the T and E transformations are the same.



Lecture 23

Assigned Reading

Suggested reading for this lecture is [**?**].

Today, we get a crash course in quantum information, which is not to be confused with quantum mechanics. We will say little about the dynamics of quantum mechanics.

Up to this point, we've covered many computational tools regarding how to compute the variety. We've also found that for examples which have no Hammersley-Clifford model, the varieties cannot be computed by SAGE since even small examples are computationally expensive. After we introduce these problems (the next few examples) we will need to cover more tools. One of those tools will be to use group actions to remove those symmetries.

Before we explore these tools, let us consider the quantum models that relate to the statistical models covered in this class.

Much like in the statistical case, we will have states, or outcomes, like heads or tails and up or down. Here, this outcome is spin orientation–spin up or spin down–or polarization. The precise meaning of these states will not matter much at this level of abstraction. We are simply getting some number of states based on the physical situation, which we use to index an orthonormal basis.

When we have two (pure) states, we will call these orthonormal bases $|0\rangle$ and $|1\rangle$. These can represent the standard basis or some other orthonormal basis.

In the classical setting, a probability distribution is some convex combination of the possible states. A quantum state need not be convex. Instead, a probability distribution (which is often called a state as well) is of the form

$$\Psi = \alpha_0 |0\rangle + \alpha_1 |1\rangle$$

where $\alpha_0, \alpha_1 \in \mathbb{C}$ with the condition that $\Psi = (\alpha_0, \alpha_1)^T$ has norm 1 over $\mathbb{C}$. This means we are working over a *Hilbert space*, a complex vector space with an inner product. We denote the inner product with angle brackets and a vertical line (instead of a comma) separating the vectors, so

$$\langle \Psi | \Psi \rangle = 1.$$

This is why we use the notation $|0\rangle$ or $|1\rangle$, which looks similar to the right-hand side of this inner product. Additionally, the triangles in tensor network diagrams are also meant to mirror this notation.

Since Ψ has norm one, then so does $e^{i\theta}\Psi$. A common term, "up to phase" refers to the equivalence class induced by $e^{i\theta}$. That is, $\hat{\Psi} \sim \Psi$ if there exists a θ such that $\hat{\Psi} = e^{i\theta}\Psi$. Because of this equivalence class, we often think of Ψ as living in complex projective space.

If the particles have two possible states $|0\rangle, |1\rangle$, then we call it a *qubit*. If there are more states, $|0\rangle, \ldots, |d-1\rangle$, then we call it a *qudit*. Each particle lives in $\mathbb{C}^2$ or $\mathbb{C}^d$. The state Ψ for a system of, for example, 3 qubit particles lives in the space arising from the tensor product $\mathbb{C}^2 \times \mathbb{C}^2 \times \mathbb{C}^2$. We represent the basis elements as $|000\rangle, |001\rangle, |010\rangle, |100\rangle, |011\rangle, |101\rangle, |110\rangle, |111\rangle$. Hence, a vector Ψ is written as

$$\Psi = a_0 |000\rangle + a_1 |001\rangle + \ldots + a_8 |111\rangle .$$

Because this state lives inside a tensor product, we have multiple ways or representing it, like flattening it into a matrix or into a $2 \times 2 \times 2$ dimensional array.

In this ket notation, we keep track of an individual particle's state based on the ordering of the numbers. So $|010\rangle$ tells us that particle 1 is in the down state, particle 2 is in the up state, and particle 3 is in the down state. Each of these kets is a rank 1 tensor.

Tensor rank 1 tensors are those of the form $v \otimes w \in V \otimes W$ where $v \in V, w \in W$. In statistics, this is the independence model (two independent coins). Ih physics, this is called a *product state*. In Geometry, this is a *Segre variety*. Although we are looking at these (tensor) rank 1 tensors of the form $v \otimes w$, the same is true with more factors (i.e. $v_1 \otimes v_2 \otimes \ldots \otimes v_k$).

The tensor rank of an arbitrary tensor T is the smallest number of rank tensors which add to T. It turns out determining tensor rank is very hard. There are many open problems regarding computing the tensor rank of small tensors.

A state is *entangled* if it has tensor rank greater than 1. For example, the state

$$\Psi = |00\rangle + |11\rangle$$

is an entangled state. This particular state is called a *Bell state*.

23.1 Measurement. An **observable** is a measurement we can make that is represented mathematically by a Hermitian (self-adjoint) operator M. By the Spectral theorem, this is a diagonalizable matrix with an orthonormal eigenbasis and real eigenvalues. This means

$$M = \sum_i c_i e_i e_i^T$$

where e_i are the orthonormal eigenvectors. Using the bra-ket notation, we would write this as

$$M = \sum_i c_i \left|e_i\right\rangle \left\langle e_i\right|$$

where $\left|e_i\right\rangle \left\langle e_i\right|$ represents the outer product.

When we have a state,

$$\Psi = \alpha_0 \left|0\right\rangle + \alpha_1 \left|1\right\rangle$$

we can find probability of observing e_i by projecting the state onto the basis element

$$p_i = \left\langle \Psi | e_i \right\rangle \left\langle e_i | \Psi \right\rangle .$$

Exercise 12. Prove that $\sum p_i = 1$. (Hint: use that e_i is an orhtonormal basis.)

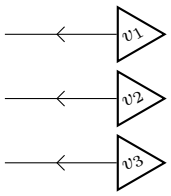
That means we still get a classical distribution for probabilities. Although we can extract probabilities, because of the nature of this problem–the fact that we can take an inner product and have entangled states–the notion of correlation is very different. The possible correlations you can get cannot be explained by classical probability.

Example 37. Consider one qubit with state $\alpha \left|0\right\rangle + \beta \left|1\right\rangle$. Consider the operator

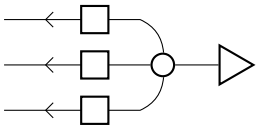
$$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The probability of getting $|0\rangle$ is $p_0 = |\alpha|^2 = p$ and $|1\rangle$ is $p_1 = |\beta|^2 = 1 - p$.

Just like in the case of undirected and directed graphical models, there are graphical models associated to certain types of entanglements. The pictures are precisely tensor networks. For example, the matrix product state (i.e. the segre variety or the indepdendence model) is represented by the diagram below:



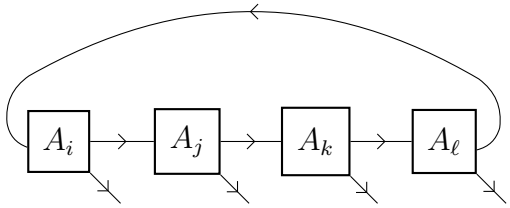
The mixture model, which we have called the *two pocket model* but is also called the *secant variety of the segre variety* by geometers, is another type of entanglement tensor.



Another type of entanglement model is called a **matrix product state**, denoted

$$\Psi_{ijk\ell} = \sum_{i,j,k,\ell \in \{0,1\}} tr(A_i A_j A_k A_\ell) \, |ijk\ell\rangle$$

where, assuming 2 dimensions, A_0 and A_1 matrices in $\mathbb{C}^2 \times \mathbb{C}^2$. Then the coefficients are $\alpha_{0110} = tr(A_0 A_1 A_1 A_0)$. The tensor network for the diagram is depicted below.



In this set up, the coefficients $a_{ikj\ell}$ are the variables and the entries of A_0 and A_1 are the parameters. So a particular state Ψ is a point in the image of a polynomial map

$$\mathbb{C}[a_{ijk\ell}] \to \mathbb{C}[a_{00}, a_{01}, a_{10}, a_{11}, b_{00}, b_{01}, b_{10}, b_{11}]$$

where

$$A_0 = \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}, A_1 = \begin{pmatrix} b_{00} & b_{01} \\ b_{10} & b_{11} \end{pmatrix}.$$

The ideal, of course, is the zero locus of this set of polynomials.

The motivation behind the matrix product state is to approximate a large line of qubits where each A_0 and A_1 is the density matrix of the state (i.e. the outer product).

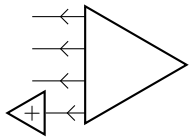
Lecture 24

Assigned Reading

Suggested reading for this lecture are the notes "What is a Tensor?"

Today, we discuss marginalization and methods to "ignore" variables. So far, we have looked at classical and quantum models and have had to deal with unobserved (i.e. hidden) information. In the coin model, the two pocket model (the secant variety of the segre variety) had the choice of pocket as the hidden variable. Computationally, we are summing over what is missed–we sum over these hidden parameters.

In other words, if we have a table of probabilites–a joint probability distribution–we sum over the possible values for hidden parameters. Take the state Ψ written below with entries $p_{ijk\ell}$ in the tensor/vector. In the diagram below, we marginalize the ℓ dimension by composing it with a vector $\vec{1}$ consisting of all 1's. We denote this as plus to indicate that we will be summing over the corresponding dimension.

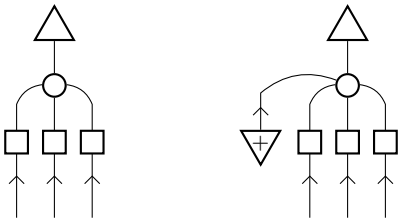


If this last wire corresponds to the ℓ dimension, then we have

$$p_{ijk} = \sum_{\ell} p_{ijk\ell}.$$

Summing over the states we do not have access to, like the possible values of ℓ, is a **marginal probability**.

We often omit wires where this marginalization is happening. For example, the the diagram on the left is simplified to the diagram on the right.



In this sense, marginalization is the process of ignoring a dimension by summing over it.

24.1 Quantum Marginalization. Formally, marginalization looks similar to the classical case. Let A and B be Hilbert spaces. A tensor $t \in A \otimes B$ is

$$t = \sum_{i=1}^{r} a_i \otimes b_i$$

for $a_i \in A, b_i \in B$. Suppose we only had access to information related to A. Based on the classical case, we would want to sum over the dimension of B; however, this would be wrong. This does not

agree with the experiments of quantum theorem. Instead, we need to take what is called a *partial trace*.

In order to find the partial trace, we need to construct the density matrix from the state. Given $\Psi = (\psi_1, \psi_2, \ldots, \psi_n) \in \mathbb{C}^n$ such that $\langle \Psi | \Psi \rangle = 1$, we can construct a density matrix by taking the outer product.

$$|\Psi\rangle \langle \Psi| = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_n \end{pmatrix} \begin{pmatrix} \psi_1 & \psi_2 & \ldots & \psi_n \end{pmatrix}$$

In the space of density matrices, we not only have these states like Ψ, which are rank 1 matrices, but we also have ensemble states:

$$\rho = \sum_i c_i |\Psi_i\rangle \langle \Psi_i| .$$

These density matrices have the following properties:

- Hermitian
- $Tr(\rho) = \sum_i \rho_{ii} = 1$
- ρ is positive semi-definite
- All eigenvalues are nonnegative.
- $Tr(\rho^2) \geq 1$
- $Tr(\rho^2) = 1$ if and only if there exists a Ψ such that $\rho = |\Psi\rangle \langle \Psi|$ (we call such matrices "pure states")

Given a density matrix, we an marginalize in a quantum sense by taking a partial trace. Hence, ρ is constructed from multipartite states (i.e. $\Psi_i \in V_1 \otimes \ldots V_n$), we can sum over the hidden states via the partial trace.

The remaining discussion of partial trace from this lecture has been omitted. You are encouraged to look through the recommended reading for a refresher on this topic.

24.2 Tensor Rank. In the last lecture, we mentioned tensor rank.

Definition 24.1. *A tensor $t \in V_1 \otimes \ldots \otimes V_n$ is called* **(tensor) rank one** *if it can be written as $v_1 \otimes \ldots \otimes v_n$ for $v_i \in V_i$.*

In the language of this class, rank 1 tensors...

- arise from the independence model,
- are product states (not to be confused with matrix product states), or
- live in the affine Segre variety.

Although we have mentioned the Segre variety, we will define this variety for this class in the following way:

Definition 24.2. *The segre variety is the Zariski closure of the space of rank one tensors.*

Question 24.3. Is the Segre variety mapped back to itself when we change the bases of the V_i in the tensor product?

The answer is yes! This action of the change of basis is often called the "local general linear group action." In the language of lecture 20, the Segre variety is *closed* under the action of

$$GL(V_1) \times \ldots \times GL(V_n).$$

In the language of Lecture 20, the Segre variety is an *orbit* when $GL(V_1) \times \ldots \times GL(V_n) \curvearrowright V_1 \otimes V_2 \ldots \otimes V_n$.

To unpack this statement, it helps to recall how the change of basis works as an action of the general linear group. Given a vector space V, an element $M \in GL(V)$ changes a basis via conjugation: $v \mapsto MvM^{-1}$. So rank 1 elements in the tensor product $V_1 \otimes \ldots \otimes V_n$ are mapped based on this "local" (as in acting coordinate-wise) action:

$$v_1 \otimes \ldots \otimes v_n \mapsto M_1 v_1 M_1^{-1} \otimes \ldots \otimes M_n v_n M_n^{-1},$$

where $M_i \in GL(V_i)$.

With this, we can define tensor rank in general.

Definition 24.4. *The* **tensor rank** *of a tensor T is the least integer r such that*

$$T = \sum_{i=1}^{r} t_i$$

such that each t_i is (tensor) rank 1.

Implicit in this definitions are two facts:

- Every tensor can be written as a sum of rank 1 tensors.
- Tensor rank is not basis dependent (i.e. r is the same regardless of the choice of basis).

This second fact follows from the discussion above. That is, since the space of rank 1 tensors are an orbit under the action of changing the basis, it follows that the choice of basis has no impact on r. (This is not immediately obvious and requires work to show.)

Tensor rank may seem like a natural and, perhaps, simple extension of the matrix rank. However, there are *many* tensors, some of which are fairly small, for which we do not know the tensor rank.

Question 24.5. What is the tensor rank of

- $|001\rangle + |011\rangle$,
- $|000\rangle + |111\rangle$, and
- $|001\rangle + |010\rangle + |100\rangle + |000\rangle + |110\rangle$?

The first is has rank 1 since it can be written as $|0\rangle\,(|0\rangle + |1\rangle)\,|1\rangle$. Explicitly, this is

$$|001\rangle + |011\rangle = (0,0,0,0,1,1,0,0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Lecture 25

Assigned Reading

Suggested reading for this lecture is....

Today, we continue our discussion of tensor rank. Last time, we defined the tensor rank. We also stated that this definition is invariant under a local change of basis (i.e. a basis change on the individual vector spaces). We ended with a few questions regarding tensor rank.

Here are some examples of tensors and their rank.

$$\begin{array}{rl} |000\rangle & \text{rank 1} \\ |001\rangle + |011\rangle & \text{rank 1} \\ |001\rangle + |010\rangle\,|100\rangle & \text{rank 3} \\ |000\rangle + |110\rangle & \text{rank 2} \end{array}$$

The second example can be written as $|0\rangle\,(|0\rangle + |1\rangle)\,|1\rangle$. Remember that $|0\rangle$ and $|1\rangle$ correspond to basis elements. For example, if we are considering the space $W \otimes V \otimes U$ with respective basis sets, $\{w_0, w_1\}, \{v_0, v_1\}, \{u_0, u_1\}$, the expression $|011\rangle = w_0 \otimes v_1 \otimes u_1$. A tensor is of rank one if we can find (in this case) 3 vectors such that $t = v \otimes w \otimes u$ for $v \in V$, etc. The vector

$$|001\rangle + |011\rangle = v_0 \otimes w_0 \otimes u_1 + v_0 \otimes w_1 \otimes u_1 = v_0 \otimes (w_0 + w_1) \otimes u_1.$$

This highlights the following fact: a rank one tensor is not necessarily one which is an induced basis element. We can construct a basis of the tensor product space by using the bases of each vector space used. In the case of two dimensional vector spaces V, W, U, the basis for the space $V \otimes W \otimes U$ is

$$\{|000\rangle, |001\rangle, |010\rangle, |100\rangle, |011\rangle, |101\rangle, |110\rangle, |111\rangle\}.$$

These basis elements, however, are not the only rank 1 tensors in the tensor product space. Hence, tensor rank does not arise from the basis.

The basis is not completely useless as it does provide a convenient upper bound for all the tensor ranks in the space

$$rank(t) \leq \prod_i \dim(V_i)$$

since every tensor can be written as a combination of these vectors. Very often, however, we can find better upper bounds. For example, let A, B, C be two dimensional complex vector spaces. Define S to be the space of rank two tensors in $A \otimes B \otimes B$. What is the Zariski closure of S? This example is the set of "mixtures" of rank one tensors. Recalling the coin problem, this is the set of one-pocket problems. We know this is the tensor network of the squid of three tentacles, which has an ideal $I = \langle 0 \rangle$. This tells us that every tensor in $A \otimes B \otimes C$ is either rank 2 or arbitrarily close to a tensor of rank 2 in the Zariski topology. The generic rank is 2.

The above variety is one constructed from an existing variety. Let V be a variety. Consider points $v, w \in V$. The set of elements $\alpha v + \beta w$ is the collection of *secant lines* of this space. The Zariski closure of this collection of lines is called the **secant variety of** V. Because the above

example is using the Segre variety (the variety of rank one tensors), it is the **secant variety of the segre variety**. In this case, elements of this secant variety of the form

$$s = (x_1a_0 + x_2a_1) \otimes (x_3b_0 + x_4b_1) \otimes (x_5c_0 + x_5c_1) + (y_1a_0 + y_2a_1) \otimes (y_3b_0 + y_4b_1) \otimes (y_5c_0 + y_5c_1).$$

Remark 25.1. Tangent lines are technically not in the original collection of lines; however, these tangent lines are limits of secant lines.

The set T_r or tensors of rank at most r is not closed in either the Euclidean or Zariski topology. Notice that for analogous set with matrix rank closure is automatic. So this is another way in which tensor rank and matrix rank differ.

The larger set containing T_r and its tangent lines are sometimes called the tensors of *border rank* r.

Example 38. Let $t = |000\rangle + |001\rangle + |010\rangle + |100\rangle$. This turns out to have tensor rank at most 3.

$$t = |00\rangle (|0\rangle + |1\rangle) + |010\rangle + |100\rangle .$$

This tensor turns out to have border rank at most 2–meaning it is a limit point in the set T_2. In the Euclidean topology, this tensor is a limit point for the family of tensors

$$t_\epsilon = \frac{1}{\epsilon} \left[(\epsilon - 1) |000\rangle + (|0\rangle + \epsilon |0\rangle) \otimes (|0\rangle + \epsilon |1\rangle) \otimes (|0\rangle + |1\rangle) \right]$$

Finding tensor rank and border rank turns out to be a challenging problem. When set up appropriately, this problem can be embedded in the 3-SAT problem, meaning that it is NP-hard. More over, there are specific tensors we can write down for which we do not know the tensor rank.

A famous such tensor is the tensor of "matrix muliplication." For 3×3 matrices, we do not know the rank.

Remark 25.2. Discussion of what this tensor is has been omitted from these notes.

One of the nice properties of the matrix multiplication tensor that sends $(M, N) \to NM$ is invariant under "local" change of basis, $GL(A) \times GL(B) \times GL(C)$, where $M : A \to B$ and $N : B \to C$. This is not surprising since the process of multiplying two matrices is not dependent on the chosen basis.

Another interesting tensor is the one which maps a rank 1 tensor $v \otimes w \mapsto v \otimes w$. This is a symmetrization tensor. This tensor lives in the space $(V \otimes W)^* \otimes (W \otimes V)$. In the tensor network language, this tensor is represented by a wire swap.

A third interesting tensor is the symmetrizing operator. This maps rank one tensors

$$v_0 \otimes v_1 \to v_0 \otimes v_1 + v_1 \otimes v_0.$$

The image of this map is denoted S^2V and is called the space of symmetric tensors. A similar map,

$$v_0 \otimes v_1 \to v_0 \otimes v_1 - v_1 \otimes v_0,$$

has the image of skew-symmetric tensors, and is denoted $\wedge^2 V$. Both of these maps are basis-independent (i.e. invariant under the local GL action).

Interestingly enough,

$$V \otimes V = S^2V \oplus \wedge^2 V,$$

which allows us to represent matrices in a basis-independent way.

We end with the following claim:

$$(\wedge^2 V) \otimes (\wedge^2 W) = \text{ the space of } 2 \times 2 \text{ minors of } m \times n \text{ matrices.}$$

Lecture 26

Assigned Reading

Suggested reading for this lecture is....

Last lecture, we discussed particular varieties are invariant under actions like local GL. This turns out to be a useful fact in the area of *entanglement classification*.

The state of three particles is represented by elements in the space $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$. The action of local $GL_2(\mathbb{C}) \times GL_2(\mathbb{C}) \times GL_2(\mathbb{C})$ defines a set of orbits. Each of these orbits is a type of entanglement. The orbit classes are written out below:

Orbit Name	One Representative	Rank
Zero (not physical)	0	0
Unentangled	$\|000\rangle$	1
Bipartite AB-C	$(\|00\rangle + \|11\rangle)\|0\rangle$	2
Bipartite A-BC	$\|0\rangle(\|00\rangle + \|11\rangle)$	2
Bipartite AC-B	$\|000\rangle + \|101\rangle$	2
GHZ	$\|000\rangle + \|111\rangle$	2
W	$\|001\rangle + \|010\rangle + \|100\rangle$	3

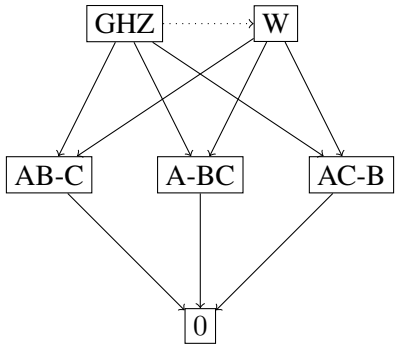
The bipartite states are those where precisely two of the three particles are entangled. The GHZ state are tensors of rank two such that no one particle can be separated from the other two. The W state contains rank 3 tensors.

Remember that these representatives are just one of the elements in the orbit. Based on the equivalence relation arising from $GL_2(\mathbb{C}) \times GL_2(\mathbb{C}) \times GL_2(\mathbb{C}) \curvearrowright \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$, any other element in the same orbit will be equivalent to the representative. Put another way, given two states of the same orbit, there exists a a group element in $GL_2(\mathbb{C}) \times GL_2(\mathbb{C}) \times GL_2(\mathbb{C})$ that sends one state to the other.

One way we can distinguish the W state from the GHZ is based on rank. We can use the invariant polynomial called the hyperdeterminant that produces zero whenever a tensor of rank 2 or less is plugged in. For tensors of rank 3, we get a nonzero answer (think determinant and matrix rank).

26.1 Polynomial Invariants. Polynomial invariants like determinants and hyperdeterminants can only define the Zariski closure of an action. Hence, the polynomial invariants will actually produce *equivalence classes* of orbits. In the case of the hyperdeterminant above will only make the six lower-rank orbits equivalent to each other and distinct from the W state.

In the case of the three-qubit problem, we see that these orbits have limit points in other orbits in the Zariski topology. The diagram below, we draw arrows pointing an orbit to other orbits that contain its limit points. The one dotted arrow indicates a mathematical limit point that is considered nonphysical for the three qubit problem.



26.2 Group Actions. Let's pick up where we left off in the last lecture. We discussed the group action local change of basis as one that is often invariant for a model. Continuing this idea, let us think about the Bayesian networks we discussed a few lectures previous (both directed and indirected graphical models).

The naive Bayes model is the "squid" model with n tentacles (i.e. the secant variety of the product state of n states). A famous conjecture by Garera, Stillman, and Sturmfels in the paper the Algebraic Geometry of Bayesian Networks was the following:

Conjecture 26.1. *The prime ideal of a naive bayes model with two hidden classes and is generated by the* 3×3 *minors of the flattenings.*

This conjecture was presented around 2005 and was solved in two different ways recently using invariant theory.

When we want to think about the space of 2×2 minors without choosing a basis, we mentioned that we can study

$$\wedge^2 A \otimes \wedge^2 B \subset S^2(A \otimes B),$$

where $S^2(A \times B)$ is the space of symmetric tensors inside $A \otimes B$. These are really the space of degree two polynomials (see "What is a tensor?" for more on why this is).

To make sense of this statement, let $\{a_i\}_{i=1}^n$ and $\{b_i\}_{i=1}^m$ be bases for A and B respectively. The induced basis on $A \otimes B$ is every combination $a_i \otimes b_j$. If we understand elements of $A \otimes B$ as $s \times n$ matrices in the obvious representation, then $a_i \times b_j$ is the matrix of all zeros except for the ij^{th} entry, where it is 1.

$$a_i \otimes b_j = \begin{array}{c} \\ \\ \\ \\ j \end{array} \begin{array}{c} \begin{array}{ccccc} & & i & & \end{array} \\ \begin{pmatrix} 0 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{pmatrix} \end{array}$$

A matrix $X \in \mathbb{C}^{n \times m}$ can be written as

$$X = \sum_{ij} x_{ij} a_i \otimes b_j.$$

Then the 2×2 minors of this matrix can be read off my the coefficients:

$$x_{jt}x_{ku} - x_{kt}x_{ju}.$$

Then this expression arises from considering

$$(a_j \otimes b_t) \circ (a_k \otimes b_u) - (a_k \otimes b_t) \circ (a_j \otimes b_u)$$

where $\circ$ is the symmetrizing product:

$$(a_j \otimes b_t) \circ (a_k \otimes b_u) = \frac{1}{2}[a_j \otimes b_t \otimes a_k \otimes b_u + a_k \otimes b_u \otimes a_j \otimes b_t].$$

Expanding this out and grouping elements, we end up with a wedge product on A and B.

Exercise 13. Write this out and justify it to yourself.

Now, let us show the same fact–that $\wedge^2 A \otimes \wedge^2 B$–gives rise to the 2×2 minors using bra-ket notation. Consider

$$(|ij\rangle - |ji\rangle)(|k\ell\rangle - |\ell k\rangle)$$

where $|i\rangle, |j\rangle$ are basis elements of$A$. We can expand this to the following expression:

$$|ijk\ell\rangle - |ij\ell k\rangle - |jik\ell\rangle + |ji\ell k\rangle .$$

In its current form, these elements are in $A \otimes A \otimes B \otimes B$. Since this space is isomorphism to $A \otimes B \otimes A \otimes B$, this statement becomes

$$|ik\rangle |j\ell\rangle - |i\ell\rangle |jk\rangle - |jk\rangle |i\ell\rangle + |j\ell\rangle |ik\rangle .$$

We treat this as an operator that computes the corresponding 2×2 minor.

In many of the models we have seen, the implicitization has been gotten by studying the minors of the tensor. Researchers often study the $k \times k$ minors in a basis-independent way via the space $\wedge^k A \otimes \wedge^k B$.

In summation, the space $V \otimes V$ breaks up into two irreducible modules $S^2V \oplus \wedge^2 V$. These two subspaces are invariant under the action of $GL(V)$.

A natural next question is whether this decomposition applies to $V \otimes V \otimes V$. After all, this is the space where the three coin problem's joint probability distribution lives (both the independent and the hidden coin version).

Under the action of $GL(V)$, $V \otimes V \otimes V$ decomposes into more spaces.

$$V \otimes V \otimes V = S^3V \oplus \wedge^3 V \oplus S_{\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}}V \oplus S_{\begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}}V.$$

These two new spaces are best understood as images of maps. Consider two maps $\rho_{\begin{array}{|c|c|}\hline 1&3\\\hline\end{array}}$ and $\rho_{\begin{array}{|c|}\hline 1\\\hline 2\\\hline\end{array}}$ as

$$\rho_{\begin{array}{|c|c|}\hline 1&3\\\hline\end{array}} : V \otimes V \otimes V \to V \otimes S^2V$$

and

$$\rho_{\begin{array}{|c|}\hline 1\\\hline 2\\\hline\end{array}} : V \otimes V \otimes V \to \wedge^2 V \otimes V.$$

Then we define the map $\rho_{\begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}} = \rho_{\begin{array}{|c|c|}\hline 1&3\\\hline\end{array}} \circ \rho_{\begin{array}{|c|}\hline 1\\\hline 2\\\hline\end{array}}$. The image of this map is the space $S_{\begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}}V$.

Exercise 14. Write out the maps that compose to produce the image $S_{\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}}V$.

REFERENCES

[1] E. S. Allman and J. A. Rhodes. Phylogenetic ideals and varieties for the general markov model. Adv. Appl. Math., 40(2):127–148, Feb. 2008.

[2] D. A. Cox, J. Little, and D. O'Shea. Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra, 3/e (Undergraduate Texts in Mathematics). Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.

[3] D. Hilbert. Ueber die theorie der algebraischen formen. Mathematische Annalen, 36(4):473–534, 1890.

[4] L. Pachter and B. Sturmfels. Algebraic Statistics for Computational Biology. Cambridge University Press, New York, NY, USA, 2005.

[5] B. Zeng, X. Chen, D.-L. Zhou, and X.-G. Wen. Quantum information meets quantum matter – from quantum entanglement to topological phase in many-body systems. ArXiv, 2015.

PENN STATE UNIVERSITY, UNIVERSITY PARK, PA 16802