# Math 4370. Computational Algebra

Taught by Karola Mészáros

Notes by Linus Setiabrata

This course is an introduction to computational algebra. Please let me know if you spot any mistakes! There are probably lots of typos. Things in [blue font square brackets] are personal comments. Things in [red font square brackets] are (important) announcements.

There is a website for this course, which can be found here. The main textbook for the course is here. There is a rough mapping between lecture material and sections of the book; the map can be found here.

Arthur found lots of typos. Thanks, Arthur.

Perhaps in the far future, I hope to comb through these notes and collate the details left to the reader and do them all in a separate document. If you would like a detail done *now*, feel free to email me.

Choice of font heavily influenced by SL.

## Contents

**1 Ideals and Varieties** **3**
1.1 Aug 29, 2019 . . . 3
1.2 Sept 3, 2019 . . . 6
1.3 Sept 5, 2019 . . . 9

**2 Division Algorithm in $k[x_1, \ldots, x_n]$** **12**
2.4 Sept 10, 2019 . . . 12
2.5 Sept 12, 2019 . . . 15
2.6 Sep 17, 2019 . . . 19
2.7 Sep 19, 2019 . . . 23
2.8 Sep 24, 2019 . . . 26
2.9 Sep 26, 2019 . . . 29

**3 Nullstellensatz** **32**
3.10 Oct 1, 2019 . . . 32
3.11 Oct 3, 2019 . . . 36

**4 Practice Prelim** **39**
4.12 Oct 8, 2019 . . . 39

**5 Nullstellensatz** **41**
5.13 Oct 17, 2019 . . . 41
5.14 Oct 22, 2019 . . . 43

**6 Some Elimination Theory** **47**
6.15 Oct 24, 2019 . . . . . 47
6.16 Oct 29, 2019 . . . . . 50
6.17 Oct 31, 2019 . . . . . 53
6.18 Nov 5, 2019 . . . . . 56
6.19 Nov 7, 2019 . . . . . 61

**7 Invariant Theory of Finite Groups** **66**
7.20 Nov 12, 2019 . . . . . 66

**8 Practice Prelim 2** **69**
8.21 Nov 19, 2019 . . . . . 69

**9 Invariant Theory of Finite Groups** **71**
9.22 Nov 26, 2019 . . . . . 71
9.23 Dec 3, 2019 . . . . . 72
9.24 Dec 5, 2019 . . . . . 75
9.25 Dec 10, 2019 . . . . . 80

# 1 Ideals and Varieties

## 1.1 Aug 29, 2019

[I am grateful to Anmol, whose notes I am copying from.]

Our professor's name is Karola Mészáros. Her website is here, and her email is karola@math.cornell.edu. Our TA is Avery St. Dizier.

In linear algebra, we studied solutions to a set of linear equations. For example, we may have the two equations (in three variables $(x, y, z)$)

$$\begin{cases} x - y + z + 2 = 0 \\ x - 2y + z - 6 = 0 \end{cases}$$

and ask for tuples $(x, y, z)$ satisfying both equations. Hopefully we all remember an algorithm that allows us to decide whether there are solutions (check the rank of a matrix), and how to describe the set of solutions (free variables, ...), and so on.

We're interested in studying systems of *polynomial* equations. For example, we may have the two equations (in three variables $(x, y, z)$)

$$\begin{cases} x^2 + y^2 - 1 = 0 \\ 2x + y^2 - 3z = 0 \end{cases}$$

and, as before, ask for tuples $(x, y, z)$ satisfying both equations. Can we find an algorithm that allows us to decide whether there are solutions? Can we describe the set of solutions?

As in linear algebra, we have to specify whether we are solving our equations for real numbers $(x, y, z)$, or for complex numbers $(x, y, z)$, or perhaps even for rational numbers $(x, y, z)$, and so on. These are all examples of *fields*, and we will usually denote fields by $k$. (It may be useful to read a bit about fields, but if you're uncomfortable with them, just think $k = \mathbb{R}, \mathbb{C}, \mathbb{Q}$.)

We understand how to solve systems of linear equations, from linear algebra. We also understand how to solve "systems" of one polynomial equation, at least over $\mathbb{C}$:

**Theorem 1.1.1** (Fundamental Theorem of Algebra)**.** *Any nonconstant polynomial*

$$f = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0 = 0,$$

*with $a_i \in \mathbb{C}$, has a root.*

Let's make some definitions precise.

**Definition 1.1.2.**

1. A monomial in $x_1, \ldots, x_n$ is a product
$$x_1^{\alpha_1} \ldots x_n^{\alpha_n},$$
where $\alpha_i \in \mathbb{Z}_{\geq 0}$ are nonnegative integers and $x_i$ are variables. We write $\mathbf{x}$ to denote the tuple $(x_1, \ldots, x_n)$ and $\alpha$ to denote the tuple $(\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n_{\geq 0}$. Then we abbreviate
$$\mathbf{x}^\alpha = x_1^{\alpha_1} \ldots x_n^{\alpha_n}$$
and say that $\mathbf{x}^\alpha$ has total degree $|\alpha| = \sum \alpha_i$.
2. A polynomial in $x_1, \ldots, x_n$ over a field $k$ is a *finite* linear combination of monomials with coefficients in a field $k$. We write
$$p(x_1, \ldots, x_n) = \sum_{\alpha \in A} a_\alpha \mathbf{x}^\alpha,$$
where $A \subset \mathbb{Z}^n_{\geq 0}$ is a *finite* subset of the set of $n$-tuples of nonnegative integers.

3. The set of polynomials in $n$ variables is written $k[x_1, \ldots, x_n]$, and it's called the polynomial ring in $n$-variables. [It is a *ring*, which has a specific definition, but that won't be important for us.] Importantly, the sum and product of two polynomials is again a polynomial.
4. When $a_\alpha \neq 0$, then $a_\alpha \mathbf{x}^\alpha$ is called a term of $p$.
5. For a field $k$ and $n \in \mathbb{Z}_{\geq 0}$, the $n$-dimensional affine space over $k$ is denoted

$$k^n = \{(a_1, \ldots, a_n) \colon a_i \in k \text{ for all } i \in \{1, \ldots, n\}\}.$$

We call $k^1$ the affine line and $k^2$ the affine plane. [We also use $[n]$ to denote the set $\{1, \ldots, n\}$.] △

Note that a polynomial

$$f = \sum_\alpha a_\alpha \mathbf{x}^\alpha \in k[x_1, \ldots, x_n]$$

gives rise to a function $k^n \to k$, which we also call $f$. This function $f \colon k^n \to k$ sends $(a_1, \ldots, a_n)$ to $f(a_1, \ldots, a_n)$ [i.e., plug $a_i$ in for $x_i$ in the formula defining $f$].

In light of this discussion, it is ambiguous what we mean when we ask "is $f = 0$?", since we may be asking whether $f = 0$ as a polynomial, meaning that all the coefficients are zero, or whether $f$ is the zero function, meaning that the polynomial vanishes on $k^n$.

Let us describe a field $k$ that has two elements. This field is denoted $\mathbb{F}_2$, and the two elements of $\mathbb{F}_2$ are $0$ and $1$. Then we define

$$\begin{aligned} 0 + 0 = 1 + 1 = 0, \qquad & 0 \cdot 0 = 1 \cdot 0 = 0 \cdot 1 = 0, \\ 0 + 1 = 1 + 0 = 1, \qquad & 1 \cdot 1 = 1. \end{aligned}$$

Let us return to the discussion above (about $f = 0$). It is not too hard to check that the polynomial $f = x^2 + x \in \mathbb{F}_2[x]$ is zero as a function. However, it is not zero as a polynomial. [This is really sad!]

Thankfully, this issue does not arise when $k$ is infinite.

**Proposition 1.1.3.** *If $k$ is an infinite field, and $f \in k[x_1, \ldots, x_n]$, then*

$$f \text{ is the zero polynomial} \iff f \colon k^n \to k \text{ is the zero function.}$$

*Proof.* The forward direction is trivial.

The backward direction asks us to show that if $f(\mathbf{a}) = 0$ for all $\mathbf{a} = (a_1, \ldots, a_n) \in k^n$, then $f$ is the zero polynomial. The proof is by induction on the number of variables $n$. Let us assume, for now, the following fact (we'll prove it at some point!)

**Theorem 1.1.4.** *Every nonzero polynomial $f \in k[x]$ of degree $m$ has at most $m$ distinct roots.*

Now the base case of the induction is that if $f \in k[x]$ such that $f(a) = 0$ for all $a \in k$ then $f$ is the zero polynomial. This follows from Theorem 1.1.4 above, because if a polynomial has infinite roots then it is the zero polynomial. (Here we are using that $k$ is infinite!)

The inductive hypothesis says that for $f \in k[x_1, \ldots, x_n]$, then if $f(a_1, \ldots, a_{n-1}) = 0$ for all $(a_1, \ldots, a_{n-1}) \in k^{n-1}$ then $f(x_1, \ldots, x_{n-1})$ is the zero polynomial.

With this hypothesis, we need to show that if $f \in k[x_1, \ldots, x_n]$ and $f(a_1, \ldots, a_n) = 0$ for all $(a_1, \ldots, a_n) \in k^n$, then $f$ is the zero polynomial. Let us write

$$f(x_1, \ldots, x_n) = \sum_{i=0}^{N} g_i(x_1, \ldots, x_{n-1}) x_n^i,$$

where $g_i \in k[x_1, \ldots, x_{n-1}]$. Our goal is to show that each $g_i$ is the zero polynomial, because this would imply that $f$ is the zero polynomial.

To show that the $g_i$ are zero polynomials, let us fix any $(a_1, \ldots, a_{n-1}) \in k^{n-1}$. Let us consider the *univariate* polynomial $f(a_1, \ldots, a_{n-1}, x_n) \in k[x_n]$. Then $f(a_1, \ldots, a_{n-1}, x_n) = 0$ has infinitely many solutions $x_n \in k$, so Theorem 1.1.4 implies $f(a_1, \ldots, a_{n-1}, x_n)$ is the zero polynomial in $k[x_n]$. This means that $g_i(a_1, \ldots, a_{n-1}) = 0$ for all $i$.

Since this argument holds for every $(a_1, \ldots, a_{n-1}) \in k^{n-1}$, the polynomial $g_i(x_1, \ldots, x_{n-1})$ has infinitely many roots (we showed $g_i(a_1, \ldots, a_{n-1}) = 0$ for any choice of $(a_1, \ldots, a_{n-1})$, after all.)

The inductive hypothesis implies that $g_i$ is the zero polynomial, and hence the coefficients of $f$ are all zero, and hence $f$ is the zero polynomial. $\square$

## 1.2 Sept 3, 2019

[I am grateful to Anmol, whose notes I am copying from.]

[We're going to have two prelims in this course. The first one will happen on October 10, and the second will happen on November 21.]

Let's talk about affine varieties today. This is a familiar concept, but with a different name.

Let $k$ be a field, and let $f_1, \dots, f_s \in k[x_1, \dots, x_n]$. We write

$$\mathbf{V}(f_1, \dots, f_s) \stackrel{\text{def}}{=} \{(a_1, \dots, a_n) \in k^n \colon f_i(a_1, \dots, a_n) = 0 \text{ for all } i \in [s]\}.$$

**Example 1.2.1.** Examples include:

- $\mathbf{V}(x^2 + y^2 - 1)$, which geometrically corresponds to the unit circle,
- The graph of a function $y = f(x)$, which algebraically corresponds to $\mathbf{V}(y - f(x))$, and
- $\mathbf{V}(xz, yz)$, which geometrically corresponds to the $xy$-plane along with the $z$-axis. △

Let's prove some basic properties of affine varieties.

**Lemma 1.2.2.** *If $V, W \subseteq k^n$ are affine varieties, then so are $V \cap W$ and $V \cup W$.*

*Proof.* Let $V = \mathbf{V}(f_1, \dots, f_s)$ and $W = \mathbf{V}(g_1, \dots, g_t)$. We have

$$V \cap W = \mathbf{V}(f_1, \dots, f_s, g_1, \dots, g_t).$$

This is not too hard to prove. We claim that

$$V \cup W = \mathbf{V}(f_i g_j \colon i \in [s], j \in [t]).$$

To prove these two sets are equal, we should show that both inclusions $V \cup W \subseteq \mathbf{V}(f_i g_j \colon i \in [s], j \in [t])$ and $V \cup W \supseteq \mathbf{V}(f_i g_j \colon i \in [s], j \in [t])$ hold.

To show $V \cup W \subseteq \mathbf{V}(f_i g_j \colon i \in [s], j \in [t])$, we take $\mathbf{a} \in V$ and note that $f_i(\mathbf{a}) = 0$ for all $i \in [s]$ implies $f_i g_j(\mathbf{a}) = 0$ for all $i \in [s], j \in [t]$. We can run a similar argument for $\mathbf{a} \in W$, and it shows that $\mathbf{a} \in V \cup W$ implies $\mathbf{a} \in \mathbf{V}(f_i g_j \colon i \in [s], j \in [t])$. In other words, we've shown $V \cup W \subseteq \mathbf{V}(f_i g_j \colon i \in [s], j \in [t])$.

To show $V \cup W \supseteq \mathbf{V}(f_i g_j \colon i \in [s], j \in [t])$. Let's take an $\mathbf{a} \in \mathbf{V}(f_i g_j \colon i \in [s], j \in [t])$; if $\mathbf{a} \in V$, then we are done. So let us suppose $\mathbf{a} \notin V$; we need to show that it's in $W$. If $\mathbf{a} \notin V$, then there exists $i_0 \in [s]$ such that $f_{i_0}(\mathbf{a}) \neq 0$. However, $f_{i_0} g_j(\mathbf{a}) = 0$ for all $j \in [t]$, so $g_j(\mathbf{a}) = 0$ for all $j \in [t]$, and hence $\mathbf{a} \in W$ and $\mathbf{a} \in V \cup W$. In other words, we've shown $\mathbf{V}(f_i g_j \colon i \in [s], j \in [t]) \subseteq V \cup W$. □

Note that an infinite union of varieties is not always a variety. While it is true that an infinite intersection of varieties is always a variety, this is a fairly deep theorem that we can't prove yet.

Once we have a variety, there are two big questions we could ask: is $\mathbf{V}(f_1, \dots, f_s)$ nonempty, and if so, is $\mathbf{V}(f_1, \dots, f_s)$ finite or infinite?

Let us also talk about ideals.

**Definition 1.2.3.** We say $I \subseteq k[x_1, \dots, x_n]$ is an ideal if

(1) $0 \in I$,

(2) $f, g \in I \implies f + g \in I$, and
(3) $f \in I, h \in k[x_1, \ldots, x_n]$ implies $hf \in I$.

△

These conditions in $I$ imply that if $f_1, \ldots, f_s \in I$ then $\sum_{i=1}^{s} h_i f_i \in I$ for every $h_i \in k[x_1, \ldots, x_n]$.

**Definition 1.2.4.** Let $f_1, \ldots, f_s \in k[\mathbf{x}]$. We say

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

is the ideal generated by $f_1, \ldots, f_s$. (We should prove that it actually is an ideal, but this will be left as an exercise.) The set $\{f_1, \ldots, f_s\}$ is called a basis for the ideal. △

The ideal description problem asks whether every ideal has a finite basis. The answer is yes, and this is called Hilbert's basis theorem. We'll prove this theorem later.

**Example 1.2.5.** In the univariate case, so $k[x]$, the ideal

$$\langle x \rangle = \{h(x)x \colon h(x) \in k[x]\}$$

consists of the zero polynomial, along with the polynomials in $k[x]$ without a constant term. △

An ideal is finitely generated if there exist $f_1, \ldots, f_s \in k[\mathbf{x}]$ such that $I = \langle f_1, \ldots, f_s \rangle$. Of course, we call $f_1, \ldots, f_s$ a basis for $I$.

**Proposition 1.2.6.** *If $\{f_1, \ldots, f_s\}$ and $\{g_1, \ldots, g_t\}$ are two bases of an ideal $I \subseteq k[\mathbf{x}]$, then $\mathbf{V}(f_1, \ldots, f_s) = \mathbf{V}(g_1, \ldots, g_t)$.*

*Proof.* This will be left as an exercise; as a starting point, begin with the observation that $\langle f_1, \ldots, f_s \rangle = \langle g_1, \ldots, g_t \rangle$. □

The moral of the Proposition 1.2.6 is that affine varieties are determined by ideals, rather than the functions $f_1, \ldots, f_s$.

**Definition 1.2.7.** Let $V \subseteq k^n$ be an affine variety. We define

$$\mathbf{I}(V) \stackrel{\text{def}}{=} \{f \in k[x_1, \ldots, x_n] \colon f(\mathbf{a}) = 0 \text{ for all } \mathbf{a} \in V\}.$$

This is called the ideal of $V$. △

This naming begs the following lemma:

**Lemma 1.2.8.** *Let $V \subseteq k^n$ be an affine variety. Then $\mathbf{I}(V)$ is an ideal of $k[x_1, \ldots, x_n]$.*

*Proof.* It's straightforward to check that $\mathbf{I}(V)$ satisfies the three conditions required of an ideal:

(1) We have $0 \in \mathbf{I}(V)$ since it vanishes on $V$,
(2) If $f(\mathbf{a}) = g(\mathbf{a}) = 0$ for all $\mathbf{a} \in V$, then $(f + g)(\mathbf{a}) = 0$ as well. Thus $f + g \in \mathbf{I}(V)$.
(3) If $f(\mathbf{a}) = 0$ for all $\mathbf{a} \in V$, then $hf(\mathbf{a}) = 0$ for any $h \in k[x_1, \ldots, x_n]$. Thus $hf \in \mathbf{I}(V)$. □

Recall Proposition 1.1.3 from last lecture, which stated that if $k$ is an infinite field, then $f \in k[x_1, \ldots, x_n]$ being the zero polynomial is equivalent to $f$ being the zero function $k^n \to k$. This proposition is really the statement that $\mathbf{I}(k^n) = \{0\}$. (We answered the question, "Which polynomials take all of $k^n$ to zero?")

Let $f_1, \ldots, f_s \in k[x_1, \ldots, x_n]$. We have functions $\mathbf{V}$ and $\mathbf{I}$, and in particular the composite map

$$\langle f_1, \ldots, f_s \rangle \longrightarrow \mathbf{V}(\langle f_1, \ldots, f_s \rangle) \longrightarrow \mathbf{I}(\mathbf{V}(f_1, \ldots, f_s))$$

sends an ideal $\langle f_1, \ldots, f_s \rangle$ to another ideal $\mathbf{I}(\mathbf{V}(f_1, \ldots, f_s))$. It's natural to wonder what the relationship between $\langle f_1, \ldots, f_s \rangle$ and $\mathbf{I}(\mathbf{V}(\langle f_1, \ldots, f_s \rangle))$ is (if you're ambitious, you might hope that they're equal).

**Lemma 1.2.9.** *We have*

$$\langle f_1, \ldots, f_s \rangle \subseteq \mathbf{I}(\mathbf{V}(f_1, \ldots, f_s)).$$

*Equality isn't always true. (See Example 1.2.10 for an example.)*

*Proof.* Let $f \in \langle f_1, \ldots, f_s \rangle$, so that $f = \sum_{i=1}^{s} h_i f_i$ for some $h_i \in k[\mathbf{x}]$. Note that $f_i(\mathbf{a}) = 0$ for all $\mathbf{a} \in \mathbf{V}(f_1, \ldots, f_s)$, by definition of a variety, so $\sum_{i=1}^{s} h_i f_i(\mathbf{a}) = 0$ for all $\mathbf{a} \in \mathbf{V}(f_1, \ldots, f_s)$. Since this means that $f$ vanishes on $\mathbf{V}(f_1, \ldots, f_s)$, this means that $f \in \mathbf{I}(\mathbf{V}(f_1, \ldots, f_s))$. □

**Example 1.2.10.** Consider $k[x, y]$ and let $f_1(x, y) = x^2$ and let $f_2(x, y) = y^2$. Then we have $\mathbf{V}(x^2, y^2) = \{(0, 0)\}$. Thus $\mathbf{I}(\mathbf{V}(x^2, y^2)) = \mathbf{I}(\{(0, 0)\}) = \langle x, y \rangle$. It's not too hard to show that $\langle x^2, y^2 \rangle \neq \langle x, y \rangle$. △

Let us end with a proposition that we'll prove in the homework.

**Proposition 1.2.11** (cf. HW 2)**.** *Let $V$ and $W$ be affine varieties in $k^n$. Then*

*(1) We have $V \subseteq W \iff \mathbf{I}(V) \supseteq \mathbf{I}(W)$, and*
*(2) We have $V = W \iff \mathbf{I}(V) = \mathbf{I}(W)$.*

The exact relation between $\langle f_1, \ldots, f_s \rangle$ and $\mathbf{I}(\mathbf{V}(f_1, \ldots, f_s))$ is described by the nullstellensatz, which will be a central theorem in this class.

As a reminder, we want to answer the ideal description problem, which asks whether every ideal $I \subseteq k[x_1, \ldots, x_n]$ can be written as $\langle f_1, \ldots f_s \rangle$, and there is the ideal membership problem, which asks for an algorithm to decide whether or not $f \in \langle f_1, \ldots, f_s \rangle$.

## 1.3 Sept 5, 2019

[I am grateful to Anmol, whose notes I am copying from.]

We are studying polynomials in $k[x]$, and in particular we want to build toward a division algorithm in $k[x]$. This division algorithm will answer the question of finite generation of ideals (at least in $k[x]$), and will also solve ideal membership.

**Definition 1.3.1.** Let $f \in k[x]$ be a nonzero polynomial, so

$$f(x) = c_m x^m + c_{m-1}x^{m-1} + \cdots + c_0,$$

where $c_i \in k$ and $c_m \neq 0$.

1. We define $m$ to be the degree of $f$, and denote it $\deg(f)$.
2. We define $c_m x^m$ to be the leading term of $f$, and denote it $\mathrm{LT}(f)$.
3. We define $x^m$ to be the leading monomial of $f$, and denote it $\mathrm{LM}(f)$.
4. We define $c_m$ to be the leading coefficient of $f$, and denote it $\mathrm{LC}(f)$. △

In particular, $\mathrm{LT}(f) = \mathrm{LC}(f) \cdot \mathrm{LM}(f)$.

Note that if $f, g \in k[x]$ are nonzero, then $\mathrm{LT}(f)$ divides $\mathrm{LT}(g)$ if and only if $\deg(f) \leq \deg(g)$.

**Proposition 1.3.2** (The division algorithm in $k[x]$)**.** *Let $k$ be a field, and let $g$ be a nonzero polynomial in $k[x]$. Then for all $f \in k[x]$ we can write $f = qg + r$, where $q, r \in k[x]$ and either $r = 0$ or $\deg(r) < \deg(g)$. Furthermore, $q$ and $r$ is are unique.*

*Proof.* Here is an algorithm to find $q$ and $r$:

```
Input f, g
Output q, r
q := 0, r := f
WHILE(r ≠ 0 and LT(g) divides LT(r)) DO
  q := q + LT(r)/LT(g)
  r := r − LT(r)/LT(g) · g
RETURN q, r.
```

See Example 1.3.3 for an example of this algorithm at work.

We have three tasks in front of us:

(1) We have to show that the algorithm terminates,
(2) That the final $q$ and $r$ have the desired properties, and
(3) That they are unique.

Note that at each step, the equation $f = qg + r$ holds. Then, when $r = 0$ or $\mathrm{LT}(g)$ doesn't divide $\mathrm{LT}(r)$, we terminate. The latter means that $\deg(g) > \deg(r)$, and so if we terminate, the output $q$ and $r$ have the desired properties. This takes care of the second task.

To see that the algorithm terminates, we just observe that the degree of $r$ always gets lowered in the `WHILE` loop, since we've set

$$r := r - \frac{\mathrm{LT}(r)}{\mathrm{LT}(g)}(\mathrm{LT}(g) + \text{lower order terms}),$$

the leading term of $r$ dies and the degree goes down. This takes care of the first task.

We need to show the uniqueness of $q$ and $r$. Suppose that

$$f = qg + r = q_2 g_2 + r_2,$$

with $\deg(r), \deg(r_2) < \deg(g)$. If $r - r_2 = 0$ then $r = r_2$ and hence $q = q_2$ and we'd be done, so let us assume $0 \neq r - r_2$. Rearranging the above equation gives

$$r - r_2 = (q_2 - q)g,$$

and in particular

$$\deg(g) > \deg(r - r_2) = \deg(q - q_2) + \deg(g) \geq \deg(g).$$

This is a contradiction! □

**Example 1.3.3.** Let us take $f = x^3 + 2x^2 + x + 1$ and $g = 2x + 1$.

We initiate $q = 0$ and $r = x^3 + 2x^2 + x + 1$, and in the first iteration of the while loop we replace $q$ with $q + \frac{x^3}{2x}$ and $r$ with $r - \frac{x^3}{2x}(2x + 1)$, so $q = \frac{1}{2}x^2$ and $r = \frac{3}{2}x^2 + x + 1$. Note that $\deg(r)$ has decreased, in this case from 3 to 2, and that $f = qg + r$ still holds, since

$$x^3 + 2x^2 + x + 1 = \left(\frac{1}{2}x^2(2x + 1) + \frac{3}{2}x^2 + x + 1\right).$$

In the second iteration of the while loop, we replace $q$ with $q + \frac{3x^2}{4x}$ and replace $r$ with $r - \frac{3x^2}{4x}(2x + 1)$, so $q = \frac{1}{2}x^2 + \frac{3}{4}x$ and $r = \frac{1}{4}x + 1$. Note that $\deg(r)$ has decreased, in this case from 2 to 1, and that $f = qg + r$ still holds, since

$$x^3 + 2x^2 + x + 1 = \left(\left(\frac{1}{2}x^2 + \frac{3}{4}x\right)(2x + 1) + \frac{1}{4}x + 1\right).$$

Finally, we replace $q$ with $q + \frac{1}{8}$ and $r$ with $r - \frac{1}{8}(2x + 1)$, so that $q = \frac{1}{2}x^2 + \frac{3}{4}x + \frac{1}{8}$ and $r = \frac{7}{8}$. Note that $\deg(r)$ has decreased, in this case from 1 to 0, and that $f = qg + r$ still holds, since

$$x^3 + 2x^2 + x + 1 = \left(\left(\frac{1}{2}x^2 + \frac{3}{4}x + \frac{1}{8}\right)(2x + 1) + \frac{7}{8}\right).$$ △

**Corollary 1.3.4** (cf. Theorem 1.1.4)**.** *If $k$ is a field and $f \in k[x]$ is a nonzero polynomial, then $f$ has at most $\deg(f)$ distinct roots.*

*Proof.* The proof is by induction on $m = \deg(f)$. For $m = 0$, then $f$ is a nonzero constant and has no roots.

The hypothesis states that for all $f$ of degree (at most) $m - 1$, then $f$ has at most $\deg(f)$ distinct roots.

Let us assume $\deg(f) = m$. If $f$ has no roots in $k$, then of course we are done. So suppose $a \in k$ is a root of $f$, so that $f(a) = 0$. Then the division algorithm for $q = x - a$ says that $f = q(x - a) + r$ where $r = 0$ or $\deg(r) < \deg(x - a) = 1$. Hence $r$ must be a constant. However, $f(a) = q(a - a) + r = 0$, which means that $r = 0$.

This means that $f = q(x - a)$. Note that $\deg(q) = m - 1$, since $\deg(f) = \deg(q) + \deg(x - a)$. Now, if $b \neq a$ is also a root of $f$, we have $q(b) = 0$. Since there are at most $m - 1$ roots of $q$, there are at most $m$ roots of $f$. □

**Corollary 1.3.5.** *If $k$ is a field then every ideal of $k[x]$ can be written as $\langle f \rangle$ for some $f \in k[x]$. Moreover, $f$ is unique up to multiplication by a nonzero constant.*

*Proof.* Let $I \subseteq k[x]$. If $I = \{0\}$ then $I = \langle 0 \rangle$. Otherwise, let us pick a nonzero $f \in I$ of minimal degree.

Our claim is that $I = \langle f \rangle$. It's obvious that $\langle f \rangle \subseteq I$, since $f \in I$ [cf. HW 2]. We want to show that $I \subseteq \langle f \rangle$.

For all $g \in I$, we have $g = qf + r$ where $r = 0$ or $\deg(r) < \deg(f)$. Since $I$ is an ideal, $f \in I$ implies $qf \in I$ and hence $g - qf = r \in I$. If $r \neq 0$ then $\deg(r) < \deg(f)$, which is a contradiction (recall we picked $f$ to be of minimal degree). Hence $r = 0$. This means that $g = qf \in \langle f \rangle$, and shows that $I \subseteq \langle f \rangle$.

To show uniqueness, let us suppose $\langle f \rangle = \langle g \rangle$. Since $f \in \langle g \rangle$, this means that $f = hg$ for some $h \in k[x]$. In particular, $\deg(f) = \deg(h) + \deg(g) \geq \deg(g)$. Similarly, $g \in \langle f \rangle$ implies $\deg(g) \geq \deg(f)$, so $\deg(f) = \deg(g)$, and in particular $\deg(h) = 0$ which means $h$ is a nonzero constant. □

**Definition 1.3.6.** An ideal generated by one element is called a principal ideal. Since every ideal of $k[x]$ is principal (Corollary 1.3.5), we say $k[x]$ is a principal ideal domain. △

Suppose $I = \langle x^4 - 1, x^6 - 1 \rangle$. Corollary 1.3.5 asserts that $I = \langle f \rangle$. Who is $f$?

**Definition 1.3.7.** A greatest common divisor of $f, g \in k[x]$ is a polynomial $h \in k[x]$ such that

(1) $h$ divides both $f$ and $g$, and
(2) If $p \in k[x]$ also divides both $f$ and $g$, then $p$ divides $h$.

△

**Proposition 1.3.8.** *Let $f, g \in k[x]$. Then*

*(1) $\gcd(f, g)$ exists and is unique up to multiplication by a nonzero constant,*
*(2) $\langle \gcd(f, g) \rangle = \langle f, g \rangle$, and*
*(3) There's an algorithm to find $\gcd(f, g)$.*

*Proof.* Since $k[x]$ is a PID, then $\langle f, g \rangle = \langle h \rangle$ for some $h \in k[x]$. We claim that $h$ is a gcd of $f$ and $g$; this would prove part (2) of the proposition. Checking the two conditions required of a greatest common divisor, we see that:

(1) Necessarily, $h$ divides $f$ and $g$ since $f, g \in \langle h \rangle$.
(2) If $p \in k[x]$ and $p$ divides $g$ and $g$, then $f = cp$ and $g = dp$ for some $c, d \in k[x]$. Since $h \in \langle f, g \rangle$ there exist $a, b \in k[x]$ so that $h = af + bg = (ac)p + (bd)p = (ac + bd)p$. Since $ac + bd \in k[x]$, this implies $p$ divides $h$.

Since Corollary 1.3.5 says that $h$ is unique up to multiplication by a nonzero constant, this proves part (1) of the proposition.

The proof of part (3) of the proposition will come next lecture. □

# 2 Division Algorithm in $k[x_1, \ldots, x_n]$

## 2.4 Sept 10, 2019

[I am grateful to Anmol, whose notes I am copying from.]

Let us prove part (3) of Proposition 1.3.8 from last time. Let's begin with the basic idea of the algorithm before giving the pseudocode. [Sometimes this is called the "Euclidean algorithm".]

*Proof.* The crucial observation is that if we write $f = qg + r$ as in the division algorithm, then $\gcd(f, g) = \gcd(f - qg, g) = \gcd(g, r)$. Repeating, we use the division algorithm to write $g = q_2 r + r_2$, and hence $\gcd(g, r) = \gcd(r, r_2)$, and so on. We note that if $r = 0$, then $\gcd(g, 0) = g$, and if $r \neq 0$ then we just repeat. The algorithm terminates when $r = 0$, at which point we know that we got $\gcd(f, g)$. Let's write down the pseudocode; see Example 2.4.1 for an explicit example of the algorithm at work.

```
Input f, g
Output gcd(f, g)
h := f
s := g
WHILE s ≠ 0 DO
  rem := remainder(h, s)
  h := s
  s := rem
```

□

**Example 2.4.1.** Let us compute the gcd of $x^4 - 1$ and $x^6 - 1$. We have

$$\langle x^4 - 1, x^6 - 1\rangle = \langle x^6 - 1, x^4 - 1\rangle = \langle x^4 - 1, x^2 1\rangle = \langle x^2 - 1\rangle,$$

where the first equality is just switching the order of the generators so that the degree of the first one is bigger than that of the second, and the second equality is $x^6 - 1 = x^2(x^4 - 1) + (x^2 - 1)$, and the third equality is $x^4 - 1 = (x^2 + 1)(x^2 - 1) + 0$. [I guess one should write $\langle x^2 - 1, 0\rangle$, but this is equal to $\langle x^2 - 1\rangle$.] △

**Definition 2.4.2.** The greatest common divisor of $f_1, \ldots, f_s \in k[x]$ is a polynomial $h \in k[x]$ such that

1. $h$ divides $f_1, \ldots, f_s$,
2. If $p$ divides $f_1, \ldots, f_s$,then $p$ divides $h$.

We write $h = \gcd(f_1, \ldots, f_s)$. △

**Proposition 2.4.3.** *Let $f_1, \ldots, f_s \in k[x]$ be polynomials. (Assume $s \geq 2$.)*

*(1) Then $\gcd(f_1, \ldots, f_s)$ exists and is unique up to multiplication by a nonzero constant,*
*(2) $\langle \gcd(f_1, \ldots, f_s)\rangle = \langle f_1, \ldots, f_s\rangle$,*
*(3) For $s \geq 3$ we have $\gcd(f_1, \ldots, f_s) = \gcd(f_1, \gcd(f_2, \ldots, f_s))$, and*
*(4) There exists an algorithm to compute $\gcd(f_1, \ldots, f_s)$.*

This follows from the $s = 2$ case (Proposition 1.3.8). The first three parts of 2.4.3 also hold when we have (countably) infinitely many polynomials $f_1, f_2, \ldots$, but our algorithm no longer works. [I don't think there exists an algorithm that is guaranteed to terminate in finite time.]

Note that Proposition 2.4.3 solves the ideal membership problem: for $f_1, \ldots, f_s \in k[x]$, deciding whether $f \in \langle f_1, \ldots, f_s\rangle$ boils down to computing $\gcd(f_1, \ldots, f_s)$ and checking whether the division algorithm for $f = q \gcd(f_1, \ldots, f_s) + r$ yields $r = 0$.

**Warning.** This won't go through in $k[x_1, \ldots, x_n]$. We are really using the PID-ness of $k[x]$ here.

**Example 2.4.4.** We can write

$$xy^2 - x = y(xy-1) + 0(y^2-1) + (-x+y) \quad \text{and} \quad xy^2 - x = 0(xy-1) + x(y^2-1) + 0.$$

It's not clear from the first equation whether or not $xy^2 - x \in \langle xy-1, y^2-1\rangle$, although it is clear from the second equation that it is. △

We will later show (Theorem 2.5.5) that given $f, f_1, \ldots, f_s \in k[x_1, \ldots, x_n]$, we will be able to write

$$f = \sum_{i=1}^{s} q_i f_i + r,$$

where either $r = 0$ or "deg"$(r)$" < "deg"$(f_i)$. (I use quotes because these haven't been defined for multivariate polynomials!) But even in the univariate case, we cannot guarantee a unique remainder! (As an exercise, find an example of this in the univariate case – there, deg and $<$ really *are* defined!)

Let us discuss orderings of monomials in $k[x_1, \ldots, x_n]$. The idea is to axiomatize (and hence generalize) the notion of degree, and with it to axiomatize (and hence generalize) the idea of " $>$ ".

**Definition 2.4.5.** A <u>total order</u> on $\mathbb{Z}^n_{\geq 0}$, denoted $>$, is a rule so that for every $\alpha \neq \beta \in \mathbb{Z}^n_{\geq 0}$, we have

$$\text{either } \alpha > \beta \quad \text{or} \quad \beta > \alpha,$$

and furthermore for $\alpha, \beta, \gamma \in \mathbb{Z}^n_{\geq 0}$, we have

$$\alpha > \beta \text{ and } \beta > \gamma \implies \alpha > \gamma.$$

Sometimes we use the symbol $<$, e.g. $\alpha < \beta$, to mean $\beta > \alpha$. △

An example of a total order, for $n = 1$, is the "usual" $>$. [Somewhat annoyingly, the $\geq$ in $\mathbb{Z}^n_{\geq 0}$ is the "usual" $>$, but the monomial ordering $>$ is just some rule satisfying some conditions...]

**Definition 2.4.6.** A monomial ordering $>$ on $k[x_1, \ldots, x_n]$ is a total order on $\mathbb{Z}^n_{\geq 0}$ with two additional conditions:

(1) If $\alpha > \beta$ and $\gamma \in \mathbb{Z}^n_{\geq 0}$, then $\alpha + \gamma > \beta + \gamma$, and
(2) $>$ is a well-ordering on $\mathbb{Z}^n_{\geq 0}$: every nonempty subset $A \subseteq \mathbb{Z}^n_{\geq 0}$ has a smallest element with respect to $>$. △

Given a monomial ordering on $\mathbb{Z}^n_{\geq 0}$, and two monomials $\mathbf{x}^\alpha$ and $\mathbf{x}^\beta$ of $k[x_1, \ldots, x_n]$, we sometimes write $\mathbf{x}^\alpha > \mathbf{x}^\beta$ when $\alpha > \beta$. We also may write $\alpha \geq \beta$ (or $\mathbf{x}^\alpha \geq \mathbf{x}^\beta$) to mean $\alpha > \beta$ or $\alpha = \beta$.

**Lemma 2.4.7.** *An order relation $>$ on $\mathbb{Z}^n_{\geq 0}$ is a well ordering if and only if any sequence $\alpha(1) > \alpha(2) > \alpha(3) > \ldots$ of elements in $\mathbb{Z}^n_{\geq 0}$ terminates.*

*Proof.* The contrapositive of this statement is that $>$ is not a well ordering if and only if there exists an infinite strictly decreasing sequence in $\mathbb{Z}^n_{\geq 0}$. Let's prove this (equivalent) statement instead.

To prove the forward direction, we assume that $>$ is not a well ordering, so that there is a nonempty subset $S \subseteq \mathbb{Z}^n_{\geq 0}$ which has no smallest element. Then pick any $\alpha(1) \in S$; since it's not the smallest element there exists an $\alpha(2) \in S$ such that $\alpha(1) > \alpha(2)$. But now there exists an $\alpha(3) \in S$ such that $\alpha(2) > \alpha(3)$. Since $S$ has no smallest element we can keep finding smaller elements and get our infinite strictly decreasing sequence.

Conversely, if $\alpha(1) > \alpha(2) > \alpha(3) > \ldots$, then $S = \{\alpha(1), \alpha(2), \ldots\}$ has no least element. □

In general, there are many monomial ideals. However, there are three that are commonly used:

**Definition 2.4.8.** The lexicographic order, denoted $>_{\text{lex}}$, is a monomial order defined by

$$\alpha >_{\text{lex}} \beta \text{ if the leftmost nonzero element in } \alpha - \beta \in \mathbb{Z}^n_{\geq 0} \text{ is positive.}$$ △

For example, we have $x^2y >_{\text{lex}} xy^4$, since $(2,1) - (1,4) = (1,-3)$. Note that we have implicitly ordered $x > y$ here. (If we had instead chosen $y > x$, then $y^4x >_{\text{lex}} yx^2$, since $(4,1) - (1,2) = (3,-1)$; however, when we write variables $x, y, z$ then we mean $x > y > z$ unless stated otherwise, and similarly when we write $x_1, \ldots, x_n$ then we mean $x_1 > \cdots > x_n$ unless stated otherwise.)

**Definition 2.4.9.** The graded lexicographic order, denoted $>_{\text{grlex}}$, is a monomial order defined by

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

[Some people write this as $>_{\text{deglex}}$. In particular, I think this is what the programming software `sage` uses.] △

**Definition 2.4.10.** The graded reverse lexicographic order, denoted $>_{\text{grevlex}}$, is a monomial order defined by

$$\alpha >_{\text{grevlex}} \beta \text{ if } |\alpha| > |\beta| \text{ or } |\alpha| = |\beta| \text{ and the rightmost nonzero entry of } \alpha - \beta \text{ is negative.}$$

Importantly, this is not the same as saying $|\alpha| = |\beta|$ and $\alpha <_{\text{lex}} \beta$. [Some people write this as $>_{\text{degrevlex}}$. In particular, I think this is what the programming software `sage` uses.] △

## 2.5 Sept 12, 2019

Recently, we have been thinking about multivariate polynomials

$$f = \sum_{\alpha} a_\alpha \mathbf{x}^\alpha \in k[\mathbf{x}] = k[x_1, \ldots, x_n].$$

Let us fix a monomial order, which we denote by $>$. We begin with some definitions which will be useful later.

**Definition 2.5.1** (see Example 2.5.2)**.** We have

1. The multidegree of $f$, denoted multidegree$(f)$ (or sometimes multideg$(f)$) is defined to be
$$\max_> \{\alpha \in \mathbb{Z}^n_{\geq 0} \colon a_\alpha \neq 0\},$$
where $\max_>$ is the maximum element as dictated by $>$.
2. The leading coefficient of $f$, denoted LC$(f)$, is defined to be
$$a_{\text{multideg}(f)} \in k.$$
3. The leading monomial of $f$, denoted LM$(f)$, is defined to be $\mathbf{x}^{\text{multideg}(f)}$.
4. The leading term of $f$, denoted LT$(f)$, is defined to be LT$(f) =$ LC$(f)$LM$(f)$. △

**Example 2.5.2.** Let $f = ix^2y + \pi y^4 \in \mathbb{C}[x, y]$, and let our monomial ordering be $>_{\text{lex}}$. In this case, the multidegree of $f$ is $(2, 1)$, because $(2, 1) >_{\text{lex}} (0, 4)$. The leading coefficient is $i$, the leading monomial is $x^2y$, and the leading term is $ix^2y$. [Everything would be different if we picked, for example, our monomial to be $>_{\text{grlex}}$.] △

The words in Definition 2.5.1 will be useful for us because of the following reason. Recall how we solved the ideal membership problem for univariate polynomials: to decide whether $f \in \langle f_1, \ldots, f_s \rangle$, we write

$$\langle f_1, \ldots, f_s \rangle = \langle \gcd(f_1, \ldots, f_s) \rangle,$$

and showed $f \in \langle h \rangle$ if and only if $f = ph$ for some $p \in k[x]$. To find such a $p$, we use the division algorithm to write $f = qh + r$ and check that $r = 0$.

We would like to solve the ideal membership problem for multivariate polynomials. Unfortunately, the crucial step of writing our ideal $\langle f_1, \ldots, f_s \rangle = \langle h \rangle$ as a principal ideal fails in general (in HW 2, we'll see that $\langle x, y \rangle$ is not a principal ideal). Thus we would have to make some division algorithm that can "divide" $f$ by the tuple of polynomials $(f_1, \ldots, f_s)$.

When we try to do the naive thing (i.e., try to use the division algorithm to divide $f$ by $f_1$ and then by $f_2$), we would need to choose to "prefer" $f_1$ over $f_2$ (or vice versa). This choice will cause problems: as we will see in Examples 2.5.3 and 2.5.4, the choice of ordering $f_1$ over $f_2$ (or vice versa) will change the remainder that we get at the end (!).

We will soon build up to the notion of a Gröbner basis, which is designed to fix this issue. The basic idea is the following: for an ideal $\langle f_1, \ldots, f_s \rangle$, we will find polynomials $g_1, \ldots, g_t$ so that $\langle g_1, \ldots, g_t \rangle = \langle f_1, \ldots, f_s \rangle$. Importantly, the $g_1, \ldots, g_t$ will have special properties which make it so that the division algorithm gives the same remainder *no matter which ordering* of $g_1, \ldots, g_t$ we choose.

**Example 2.5.3.** Let's "divide" $f = x^2y + xy^2 + y^2$ by $f_1 = xy - 1$ and $f_2 = y^2 - 1$, and let us pick the monomial ordering $>_{\text{lex}}$. That is to say, we want to write $f = q_1f_1 + q_2f_2 + r$. To begin, we set $q_1, q_2, r = 0$.

Note that LT$(f_1) = xy$ divides LT$(f) = x^2y$; in particular, LT$(f)$/LT$(f_1) = x$, so we replace $q_1$ by $q_1 + x$ and replace $f$ with $f - xf_1 = xy^2 + x + y^2$.

The new leading term $\mathrm{LT}(f) = xy^2$ is still divisible by $\mathrm{LT}(f_1)$, in particular with $\mathrm{LT}(f)/\mathrm{LT}(f_1) = y$, so we replace $q_1$ by $q_1 + y$ and replace $f$ with $f - yf_1 = x + y^2 + y$.

Now the leading term $\mathrm{LT}(f) = x$ is no longer divisible by either of $\mathrm{LT}(f_1)$ or $\mathrm{LT}(f_2)$, so we replace $r$ with $r + x$ and replace $f$ with $f - x = y^2 + y$.

Observe that $\mathrm{LT}(f_1) = xy$ does not divide $\mathrm{LT}(f) = y^2$, but $\mathrm{LT}(f_2) = y^2$ does. Since $\mathrm{LT}(f)/\mathrm{LT}(f_2) = 1$, we replace $q_2$ by $q_2 + 1$ and replace $f$ with $f - f_2 = y + 1$.

The leading term $\mathrm{LT}(f) = y$ is no longer divisible by either of $\mathrm{LT}(f_1)$ or $\mathrm{LT}(f_2)$, so we replace $r$ with $r + y$ and replace $f$ with $f - y = 1$.

Finally, the leading term $\mathrm{LT}(f) = 1$ is still not divisible by either of $\mathrm{LT}(f_1)$ or $\mathrm{LT}(f_2)$, so we replace $r$ with $r + 1$ and replace $f$ with $f - 1 = 0$.

Hence, the result is

$$f = (x + y)f_1 + f_2 + (x + y + 1).$$

△

To check whether $f \in \langle f_1, f_2 \rangle$, the above example, we chose to "prefer" $f_1$ over $f_2$. Let's see what happens if we do it the other way around:

**Example 2.5.4.** Let's "divide" $f = x^2y + xy^2 + y^2$ by $f_1 = y^2 - 1$ and $f_2 = xy - 1$, again with the monomial ordering $>_{\text{lex}}$. As before, we set $q_1, q_2, r = 0$.

We see that $\mathrm{LT}(f_1) = y^2$ does not divide $\mathrm{LT}(f) = x^2y$, but $\mathrm{LT}(f_2) = xy$ does. Since $\mathrm{LT}(f)/\mathrm{LT}(f_2) = x$, we replace $q_2$ with $q_2 + x$ and replace $f$ with $f - xf_2 = xy^2 + x + y$.

Now $\mathrm{LT}(f_1) = y^2$ does divide $\mathrm{LT}(f) = xy^2$, in particular with $\mathrm{LT}(f)/\mathrm{LT}(f_1) = x$. Hence, we replace $q_1$ with $q_1 + x$, and replace $f$ with $f - xf_1 = 2x + y^2$.

Neither $\mathrm{LT}(f_1) = y^2$ nor $\mathrm{LT}(f_2) = xy$ divides $\mathrm{LT}(f) = 2x$, so we replace $r$ with $r + 2x$ and replace $f$ with $f - 2x = y^2$.

Since $\mathrm{LT}(f_1) = y^2$ divides $\mathrm{LT}(f) = y^2$, with $\mathrm{LT}(f)/\mathrm{LT}(f_1) = 1$, we replace $q_1$ with $q_1 + 1$ and replace $f$ with $f - f_1 = 1$.

Finally, $\mathrm{LT}(f) = 1$ is not divisible by either of $\mathrm{LT}(f_1)$ or $\mathrm{LT}(f_2)$, so we replace $r$ with $r + 1$ and replace $f$ with $f - 1 = 0$.

Hence, the result is

$$f = (x + 1)f_1 + xf_2 + (2x + 1).$$

Hence, the remainder is not unique! △

**Theorem 2.5.5** (Division algorithm in $k[x_1, \ldots, x_n]$; cf. Proposition 1.3.2)**.** *Let $>$ be a monomial order on $\mathbb{Z}^n_{\geq 0}$, and let $F = (f_1, \ldots, f_s)$ be an ordered $s$-tuple of polynomials in $k[x_1, \ldots, x_n]$. Then every $f \in k[x_1, \ldots, x_n]$ can be written as $f = q_1f_1 + \cdots + q_nf_n + r$, where $q_i, r \in k[x_1, \ldots, x_n]$, and either $r = 0$ or no term of $r$ is divisible by any of the $\mathrm{LT}(f_1), \ldots, \mathrm{LT}(f_s)$. Furthermore, if $q_if_i \neq 0$, then $\mathrm{multideg}(f) \geq \mathrm{multideg}(q_if_i)$.*

The algorithm is given by the following pseudocode:

`Input` $f_1, \ldots, f_s, f$
`Output` $q_1, \ldots, q_s, r$
$q_1 := 0; \ldots; q_s := 0; r := 0$
$p := f$

```
WHILE p ≠ 0 DO
    i := 1
    divisionoccurred := false
    WHILE (i ≤ s AND divisionoccurred = false) DO
        IF LT(f_i) divides LT(p) THEN
            q_i := q_i + LT(p)/LT(f_i)
            p := p − (LT(p)/LT(f_i))f_i
            divisionoccurred := true
        ELSE
            i := i + 1
    IF divisionoccurred = false THEN
        r := r + LT(p)
        p := p − LT(p)
RETURN q_1, ..., q_s, r.
```

The proof that this algorithm does what Theorem 2.5.5 asks for is the same as the univariate case. Let us emphasize that if $r = 0$ then $f \in \langle f_1, \ldots, f_s \rangle$, but even if $f \in \langle f_1, \ldots, f_s \rangle$ it does not necessarily mean that $r = 0$.

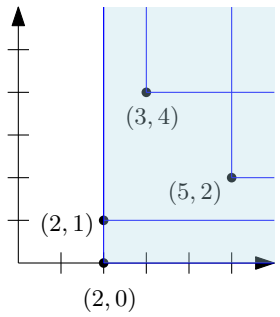
Let's now think about "monomial ideals" in $k[x_1, \ldots, x_n]$; understanding these will allow us to understand ideals of $k[x_1, \ldots, x_n]$ in general.

**Definition 2.5.6.** An ideal $I \subseteq k[x_1, \ldots, x_n]$ is called a monomial ideal if there is a subset $A \subseteq \mathbb{Z}^n_{\geq 0}$ (possibly infinite) such that $I$ is the set

$$I = \left\{ \sum_{\alpha \in A} h_\alpha x^\alpha \colon h_\alpha \in k[x_1, \ldots, x_n], \text{only finitely many } h_\alpha \neq 0 \right\}.$$

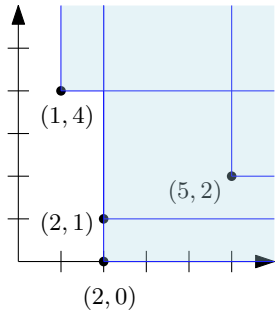
We write $I = \langle x_\alpha \colon \alpha \in A \rangle$. △

**Example 2.5.7.** Let $I \subseteq k[x, y]$ be given by $\langle x^3y^4, x^2y, x^2, x^5y^2 \rangle$. Note that $I = \langle x^2 \rangle$, as the picture below "shows":




△

**Example 2.5.8.** Let $I \subseteq k[x, y]$ be given by $\langle xy^4, x^2y, x^2, x^5y^2 \rangle$. Then $I$ is not a principal ideal, as the picture below "shows".




△

**Lemma 2.5.9.** *Let $I = \langle x^\alpha : \alpha \in A \rangle$ be a monomial ideal. Then*

$$x^\beta \in I \iff x^\beta \text{ is divisible by } x^\alpha \text{ for some } \alpha \in A.$$

*Proof.* The backward direction asks us to prove $x^\beta$ is divisible by $x^\alpha$ implies $x^\beta \in I$. This is clear from the definitions.

The forwards direction asks us to prove that $x^\beta \in I$ implies $x^\beta$ is divisible by $x^\alpha$ for some $\alpha \in A$. Let us write $x^\beta$ as the finite sum

$$x^\beta = \sum_i h_i x^{\alpha(i)},$$

for some $\alpha(i) \in A$ and $h_i \in k[x_1, \ldots, x_n]$. Then

$$x^\beta = \sum_{i=1}^{s} \left( \sum_j c_{ij} x^{\beta(i,j)} \right) x^{\alpha(i)} = \sum_{i,j} c_{i,j} x^{\beta(i,j)} x^{\alpha(i)}.$$

Note that this is an equality of polynomials. Since each term on the right hand side is divisible by some $x^{\alpha(i)}$, so is the left hand side. In other words, $x^\beta$ is divisible by some $x^{\alpha(i)}$. □

**Exercise-Lemma 2.5.10.** *Let $I$ be a monomial ideal, and let $f \in k[x_1, \ldots, x_n]$. The following are equivalent:*

*(1) $f \in I$*
*(2) Each term of $f$ is in $I$*
*(3) $f$ is a $k$-linear combination of monomials in $I$.*

(The only nontrivial direction is (1) $\implies$ (2), or alternatively (1) $\implies$ (3).)

## 2.6 Sep 17, 2019

We begin with recalling some Lemmas from last time:

**Lemma 2.6.1** (cf. Lemma 2.5.9)**.** *Let $I = \langle x^\alpha : \alpha \in A \rangle$ be a monomial ideal. Then*

$$x^\beta \in I \iff x^\beta \text{ is divisible by } x^\alpha \text{ for some } \alpha \in A.$$

**Lemma 2.6.2** (cf. Exercise-Lemma 2.5.10)**.** *Let $I$ be a monomial ideal, and let $f \in k[x_1, \ldots, x_n]$. The following are equivalent:*

*(1) $f \in I$*
*(2) Each term of $f$ is in $I$*
*(3) $f$ is a $k$-linear combination of monomials in $I$.*

It's not too hard to show that Lemma 2.6.2 implies the following Corollary:

**Corollary 2.6.3.** *Two monomial ideals are the same if and only if they contain the same monomials.*

We now have enough lemmas to prove a big result:

**Theorem 2.6.4** (Dickson's Lemma)**.** *Let $I = \langle x^\alpha : \alpha \in A \rangle \subseteq k[x_1, \ldots, x_n]$ be a monomial ideal. Then $I$ is finitely generated by monomials, so*

$$I = \langle \mathbf{x}^{\alpha(1)}, \ldots, \mathbf{x}^{\alpha(s)} \rangle,$$

*where $\alpha(1), \ldots, \alpha(s) \in A$. In particular, it has a finite basis.*

(We'll basically be describing an algorithm to find it; see Example 2.6.5 for an explicit example of this algorithm at work.)

*Proof.* We use induction on $n$, the number of variables. For $n = 1$, we have a monomial ideal $I = \langle x_1^\alpha : \alpha \in A \rangle$. Pick the smallest element $\beta \in A \subseteq \mathbb{Z}_{\geq 0}$. Since $\beta \leq \alpha$ for all $\alpha \in A$, we have $x_1^\beta | x_1^\alpha$ and $I = \langle x_1^\beta \rangle$, as desired.

Let us assume the inductive hypothesis for $k[x_1, \ldots, x_{n-1}]$, and use the variables $x_1, \ldots, x_{n-1}, y$ (just to distinguish the "new" variable $y = x_n$ over the "inductive hypothesis" variables $x_1, \ldots, x_{n-1}$.)

Take $I \subseteq k[x_1, \ldots, x_{n-1}, y]$, and define the ideal

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

Note here that $\mathbf{x} = (x_1, \ldots, x_{n-1})$ and $\alpha = (\alpha_1, \ldots, \alpha_{n-1})$ are vectors. Since $J$ is a monomial ideal in $k[x_1, \ldots, x_{n-1}]$, then $J = \langle \mathbf{x}^{\alpha(1)}, \ldots, \mathbf{x}^{\alpha(s)} \rangle$ for some $\alpha(i) \in \mathbb{Z}_{\geq 0}^{n-1}$. By definition, for all $i \in [s]$ there is $m_i \in \mathbb{Z}_{\geq 0}$ so that $\mathbf{x}^{\alpha(i)} y^{m_i} \in I$. Importantly, let us take $m_i$ to be the *minimal* such $m_i$ for which this is true. (Again, see Example 2.6.5 for an explicit example of this at work.)

Let $m = \max(m_1, \ldots, m_s)$. For $\ell \in [0, m-1]$ let us define an ideal $J_\ell \subseteq k[x_1, \ldots, _{n-1}]$ by

$$J_\ell = \langle \mathbf{x}^\beta : \mathbf{x}^\beta y^\ell \in I \rangle.$$

[Unlike our definition for $J$, here the power of $y$ is given to us.]

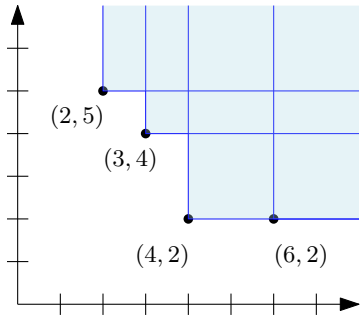
Our claim is that the set of monomials $\{\mathbf{x}^{\alpha(1)} y^m, \ldots, \mathbf{x}^{\alpha(s)} y^m\}$ ("coming from $J$"), along with the sets $\{\mathbf{x}^{\alpha_\ell(1)} y^\ell, \ldots, \mathbf{x}^{\alpha_\ell(s_\ell)} y^\ell\}$ ("coming from $J_\ell$") for each $\ell \in [0, m-1]$, generate $I$. Indeed, every monomial in $I$ is divisible by a monomial in our list above, since if

$$\mathbf{x}^\alpha y^p \in I, \text{ then } \begin{cases} \text{if } p \geq m, & \text{it's in the list from } J \\ \text{if } p \leq m-1, & \text{it's in the list from } J_p. \end{cases}$$

In particular, the monomials from the list above generate an ideal having the same monomials as $I$. By Corollary 2.6.3, they are the same ideal.

We still need to show that these finite set of generators can be chosen from $A$. So far, we've only shown that $I = \langle \mathbf{x}^{\beta(1)}, \ldots, \mathbf{x}^{\beta(s')} \rangle$ (with $s'$ possibly not equal to $s$). Note that since $\mathbf{x}^{\beta(i)} \in I = \langle x^\alpha \colon \alpha \in A \rangle$, Lemma 2.6.1 implies that $\mathbf{x}^{\beta(i)}$ is divisible by $\mathbf{x}^\alpha$ for some $\alpha \in A$. With this in mind, it's easy (an *exercise*, even!) to show that $I = \langle \mathbf{x}^{\alpha(1)}, \ldots, \mathbf{x}^{\alpha(s)} \rangle$. □

**Example 2.6.5.** Suppose $n = 2$ and our monomial ideal was $I = \langle x^2y^5, x^3y^4, x^4y^2, x^6y^2 \rangle$. (Obligatory picture below.)



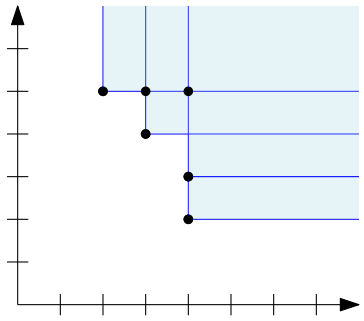

Note in the above picture that points correspond to monomials in $A$, and we shade the cone because if $x^\alpha \in I$ then $x^\beta \in I$ for every $\beta$ that is "northeast" of $\alpha$.

Our ideal $J$ lives in the polynomial ring with $n - 1 = 1$ variable, and is equal to $\langle x^2, x^3, x^4, x^6 \rangle$. By induction, we have a finite subset generating it. Suppose we were "silly" and picked $\langle x^2, x^3, x^4 \rangle$ as generators. (of course, we know $k[x]$ is a PID, but this isn't true in general, so...)

Thus, $\alpha(1) = 2$, $\alpha(2) = 3$, and $\alpha(3) = 4$. Correspondingly, $m_1 = 5$, $m_2 = 4$, and $m_3 = 2$. So we take $m = 5$. Unravelling definitions, we see $J_0 = J_1 = \{0\}$, $J_2 = J_3 = \langle x^4 \rangle$, and $J_4 = \langle x^3 \rangle$. Putting this all together, we see that

$$\{\underbrace{x^2y^5, x^3y^5, x^4y^5}_{\text{"from } J\text{"}}, \underbrace{x^4y^2}_{\text{"from } J_2\text{"}}, \underbrace{x^4y^3}_{\text{"from } J_3\text{"}}, \underbrace{x^3y^4}_{\text{"from } J_4\text{"}}\}$$

generate $I$. (Here's another obligatory picture.)



△

Note that Theorem 2.6.4 answers the ideal description problem for monomial ideals. (!)

We'll see (hopefully soon!) that any ideal of $k[x_1, \ldots, x_n]$ will be finitely generated. Let's turn to the ideal membership problem fist.

**Theorem 2.6.6** (Ideal membership for monomial ideals)**.** *Let $I = \langle \mathbf{x}^{\alpha(1)}, \ldots, \mathbf{x}^{\alpha(s)} \rangle$ be a monomial ideal. Then, $f \in I$ if and only if the remainder of $f$ on division by $\mathbf{x}^{\alpha(1)}, \ldots, \mathbf{x}^{\alpha(s)}$ is 0. [In particular, the division algorithm we described last time will give a unique remainder, regardless of ordering. (!)]*

*Proof.* The backward direction is trivial.

Let us take $f \in I$. By Lemmas 2.6.1 and 2.6.2, we see

$$f = \sum_{\substack{i \in [s] \\ j \in [p_i]}} c_{ij} \mathbf{x}^{\beta(i,j)} \mathbf{x}^{\alpha(i)},$$

for $c_{ij} \in k$. In particular, every term of $f$ is divisible by some $\mathbf{x}^{\alpha(i)}$. Recall our division algorithm theorem (Theorem 2.5.5), which says that we can write

$$f = \sum h_i \mathbf{x}^{\alpha(i)} + r,$$

where $h_i \in k[\mathbf{x}]$ and $r$ is such that $r = 0$ or no term of $r$ is divisible by any of the leading terms of (in this case) $\mathbf{x}^{\alpha(i)}$. These leading terms are just $\mathbf{x}^{\alpha(i)}$.

Rearranging the above equation we get

$$f - \sum h_i \mathbf{x}^{\alpha(i)} = r,$$

and note that every term on the left side is divisible by some $\mathbf{x}^{\alpha(i)}$ (since $f \in I$). Hence every term on the right hand side is also divisible by some $\mathbf{x}^{\alpha(i)}$, so we must be in the $r = 0$ case of the division algorithm. (!) □

**Proposition 2.6.7.** *A monomial ideal $I \subseteq k[x_1, \ldots, x_n]$ has a basis $\{\mathbf{x}^{\alpha(1)}, \ldots, \mathbf{x}^{\alpha(s)}\}$ so that $\mathbf{x}^{\alpha(i)}$ does not divide $\mathbf{x}^{\alpha(j)}$ for every $i \neq j$. Furthermore, such a basis is unique and is called the minimal basis of $I$.*

*Proof.* We know by Dickson's lemma (Theorem 2.6.4) that $I$ has a finite basis consisting of monomials. If one monomial in this basis divides the other, then we can remove the "other" and still have a basis. Repeating this, we get a basis with the desired properties.

Let us show uniqueness of this basis. Assume that $\{\mathbf{x}^{\alpha(1)}, \ldots, \mathbf{x}^{\alpha(s)}\}$ and $\{\mathbf{x}^{\beta(1)}, \ldots, \mathbf{x}^{\beta(t)}\}$ are two minimal bases. Note that $\mathbf{x}^{\alpha(1)} \in \langle \mathbf{x}^{\beta(1)}, \ldots, \mathbf{x}^{\beta(t)} \rangle$ implies $\mathbf{x}^{\alpha(1)}$ is divisible by $\mathbf{x}^{\beta(i)}$ for some $i \in [t]$. Now $\mathbf{x}^{\beta(i)} \in I = \langle \mathbf{x}^{\alpha(1)}, \ldots, \mathbf{x}^{\alpha(s)} \rangle$ implies $\mathbf{x}^{\beta(i)}$ is divisible by $\mathbf{x}^{\alpha(j)}$ for some $j \in [s]$. Since the $\{\mathbf{x}^{\alpha(1)}, \ldots, \mathbf{x}^{\alpha(s)}\}$ basis was minimal, we have $j = 1$ and $\beta(i) = \alpha(1)$. Removing these two from our bases, we repeat the above argument, and conclude that

$$\{\mathbf{x}^{\alpha(1)}, \ldots, \mathbf{x}^{\alpha(s)}\} \subseteq \{\mathbf{x}^{\beta(1)}, \ldots, \mathbf{x}^{\beta(t)}\}.$$

We can also interchange the role of $\alpha$ and $\beta$ in the proof above, and this gives the other inclusion. □

We now have a very satisfactory picture of monomial ideals. They have a unique minimal basis and we basically know how to find them. Furthermore, we can solve the ideal membership problem. Our next goals are to solve the ideal description problem in general, and to solve the ideal membership problem in general.

To recap, we have the following. In $k[x]$, we were in a PID, so every ideal is generated by a single element. Furthermore, dividing by a single polynomial gives a unique remainder.

In a monomial ideal, we have finite generation and that dividing by multiple monomials, regardless of order, gives a unique remainder.

Our goal is to find good generators for $I = \langle f_1, \ldots, f_s \rangle \subseteq k[x_1, \ldots, x_n]$, call them $\{g_1, \ldots, g_t\}$, so that dividing by these good generators in any order gives a unique remainder.

We'll see that the case of monomial ideals paves the way for the general case.

To get to our goal, we begin with a key definition.

**Definition 2.6.8.** Let $I \subseteq k[x_1, \ldots, x_n]$ be an ideal that is not $\{0\}$. Fix a monomial ordering $>$. Denote by $\mathrm{LT}(I)$ the set of leading terms of nonzero elements in $I$, that is,

$$\mathrm{LT}(I) = \{c\mathbf{x}^\alpha : \text{ there exists } f \in I \setminus \{0\} \text{ such that } \mathrm{LT}(f) = c\mathbf{x}^\alpha\}.$$

Denote by $\langle \mathrm{LT}(I) \rangle$ the ideal of $k[x_1, \ldots, x_n]$ generated by the elements in $\mathrm{LT}(I)$. It should be clear that this is a monomial ideal. △

Note that if $I = \langle f_1, \ldots, f_s \rangle$, then

$$\langle \mathrm{LT}(f_1), \ldots, \mathrm{LT}(f_s) \rangle \subseteq \langle \mathrm{LT}(I) \rangle.$$

They need not be equal! (See Example 2.6.9 below.)

**Example 2.6.9.** Let $I = \langle x^3 - 2xy, x^2y - 2y^2 + x \rangle$ and let our monomial order be $>_{\mathrm{grlex}}$. Then, although we have

$$\langle x^3, x^2y \rangle \subseteq \langle \mathrm{LT}(I) \rangle,$$

note that $x(x^3 - 2xy) - y(x^2y - 2y^2 + x) = x^2 \in \mathrm{LT}(I)$. Hence $x^2 \in \langle \mathrm{LT}(I) \rangle$, and the inclusion is strict.

[Arthur also came up with a nice example: $I = \langle x + 1, x - 1 \rangle$ does the trick.] △

## 2.7 Sep 19, 2019

Last time, we defined $\langle \mathrm{LT}(I)\rangle$ (Definition 2.6.8) and saw examples where, for $I = \langle f_1, \ldots, f_s\rangle$, we have the strict inclusion $\langle \mathrm{LT}(f_1), \ldots, \mathrm{LT}(f_s)\rangle \subsetneq \langle \mathrm{LT}(I)\rangle$ (Example 2.6.9).

**Proposition 2.7.1.** *Let $I \subseteq k[x_1, \ldots, x_n]$ be an ideal, and suppose $I \neq \{0\}$. Then:*

*(1) $\langle \mathrm{LT}(I)\rangle$ is a monomial ideal, and*
*(2) There are $g_1, \ldots, g_t \in I$ such that $\langle \mathrm{LT}(I)\rangle = \langle \mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_t)\rangle$.*

*Proof.* We will prove just part (2). (Part (1) follows from the definition of monomial ideals and $\langle \mathrm{LT}(I)\rangle$.)

In light of the fact that $\langle \mathrm{LT}(I)\rangle$ is a monomial ideal, Dickson's Lemma (Theorem 2.6.4) applies and says that

$$\langle \mathrm{LT}(I)\rangle = \langle \mathrm{LM}(g_1), \ldots, \mathrm{LM}(g_t)\rangle$$

for some $g_i \in I$. [It will be crucial for us that $g_i \in I$, and this follows because Dickson says we can pick our finite set of generators from the generating set $A = \mathrm{LM}(I)$.]

The result follows from the observation that

$$\langle \mathrm{LM}(g_1), \ldots, \mathrm{LM}(g_t)\rangle = \langle \mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_t)\rangle. \qquad \square$$

As we have said before, understanding monomial ideals paves the way for the general case. Here is an example of that:

**Theorem 2.7.2** (Hilbert Basis Theorem; cf. Theorem 2.6.4)**.** *Every ideal of $k[x_1, \ldots, x_n]$ is finitely generated.*

(This answers the ideal description problem in general.)

*Proof.* If $I = \{0\} = \langle 0\rangle$, we are done. So let us consider a nonzero ideal $I \subseteq k[x_1, \ldots, x_n]$, and let us pick a monomial order $>$. Proposition 2.7.1 says that there exists $g_1, \ldots, g_t \in I$ so that $\langle \mathrm{LT}(I)\rangle = \langle \mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_t)\rangle$.

Our claim is that in fact, $I = \langle g_1, \ldots, g_t\rangle$.

To see this, let us first note that $I \supseteq \langle g_1, \ldots, g_t\rangle$ follows from $g_1, \ldots, g_t \in I$. Thus, we need to show $I \subseteq \langle g_1, \ldots, g_t\rangle$. Take $f \in I$, and use the division algorithm (Theorem 2.5.5) to divide $f$ by the (ordered tuple of) polynomials $(g_1, \ldots, g_t)$ with respect to the monomial ordering $>$. The division algorithm guarantees quotients $(q_1, \ldots, q_t)$ and a remainder $r$ so that

$$f = q_1 g_1 + \cdots + q_t g_t + r,$$

so that $r = 0$ or no term of $r$ is divisible by any leading term $\mathrm{LT}(g_i)$, $i \in [t]$. (Note that if $r = 0$, then $f \in \langle g_1, \ldots, g_t\rangle$, and we are done.)

Assume $r \neq 0$, so no term of $r$ is divisible by any leading term $\mathrm{LT}(g_i)$. Note that

$$r = f - q_1 g_1 - q_2 g_2 - \cdots - q_t g_t \in I,$$

since $f, g_1, \ldots, g_t \in I$ by assumption. Hence $\mathrm{LT}(r) \in \langle \mathrm{LT}(I)\rangle = \langle \mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_t)\rangle$. Since $\langle \mathrm{LT}(I)\rangle$ is a monomial ideal, Lemma 2.5.9 says that the monomial $\mathrm{LT}(r)$ is divisible by one of the monomials $\mathrm{LT}(g_i)$, contradictory to our assumptions on $r$. Hence, $r = 0$, and $f \in \langle g_1, \ldots, g_t\rangle$. This proves our claim that $I = \langle g_1, \ldots, g_t\rangle$. In particular, $I$ is finitely generated. $\square$

**Definition 2.7.3.** Fix a monomial order $>$ on $k[x_1, \ldots, x_n]$. A finite subset $G = \{g_1, \ldots, g_t\}$ of the elements from the ideal $I \subseteq k[x_1, \ldots, x_n]$ is a Gröbner basis of $I$ if $\langle \mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_t) \rangle = \langle \mathrm{LT}(I) \rangle$.

Sometimes, Gröbner basis is abbreviated to "G.b.". △

Example 2.6.9 shows that not all bases are Gröbner bases.

**Definition 2.7.4.** An ascending chain of ideals is a sequence of ideals

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \ldots. \qquad \triangle$$

Consider the increasing chain of ideals in $k[x_1, \ldots, x_n]$ given by

$$\langle x_1 \rangle \subseteq \langle x_1, x_2 \rangle \subseteq \cdots \subseteq \langle x_1, \ldots, x_n \rangle.$$

Pick an $f \in k[x_1, \ldots, x_n]$, and suppose $f(\mathbf{0}) = 0$. Then $f \in \langle x_1, \ldots, x_n \rangle$, and hence $\langle f, x_1, \ldots, x_n \rangle = \langle x_1, \ldots, x_n \rangle$. Suppose instead that $f(\mathbf{0}) \neq 0$. Then $\langle f, x_1, \ldots, x_n \rangle = k[x_1, \ldots, x_n]$ is the whole ideal. Putting all this together, we've shown that an ascending chain of ideals

$$\langle x_1 \rangle \subseteq \langle x_1, x_2 \rangle \subseteq \cdots \subseteq \langle x_1, \ldots, x_n \rangle \subseteq I_{n+1} \subseteq I_{n+2} \subseteq I_{n+3} \subseteq \ldots$$

necessarily "stabilizes", that is, there exists an integer $N$ so that $I_N = I_{N+1} = I_{N+2} = \ldots$.

**Theorem 2.7.5** (Ascending chain condition)**.** *Let $I_1 \subseteq I_2 \subseteq \ldots$ be an ascending chain of ideals of $k[x_1, \ldots, x_n]$. Then there exists an integer $N \geq 1$ such that $I_N = I_{N+1} = I_{N+2} = \ldots$.*

*Proof.* We claim that

$$I = \bigcup_{i=1}^{\infty} I_i$$

is an ideal. Indeed, the three conditions required of an ideal is that:

(1) $0 \in I$, since $0 \in I_i$ for each $i$,
(2) if $f, g \in I$, then there exist $i, j$ so that $f \in I_i$ and $g \in I_j$. Without loss of generality, we can assume $i \leq j$; the inclusion $I_i \subseteq I_j$ means that $f \in I_j$, and since $I_j$ is an ideal we have $f + g \in I_j$, so $f + g \in I$.
(3) If $f \in I$ and $h \in k[x_1, \ldots, x_n]$, then there exists $i$ so that $f \in I_i$; since $I_i$ is an ideal $hf \in I_i$ and hence $hf \in I$.

Since $I$ is an ideal, Theorem 2.7.2 says that $I$ is finitely generated, so $I = \langle f_1, \ldots, f_s \rangle$. Let us take integers $j_i$ so that $f_i \in I_{j_i}$, and let $N = \max_i \{j_i : i \in [s]\}$. Then $f_i \in I_N$ for each $i \in [s]$, and we have

$$I = \langle f_1, \ldots, f_s \rangle \subseteq I_N \subseteq I_{N+1} \subseteq I_{N+2} \subseteq \cdots \subseteq I.$$

This means each inclusion above is an equality. □

Let's talk about varieties again. Recall that, for $f_1, \ldots, f_s \in k[x_1, \ldots, x_n]$, we have

$$\mathbf{V}(f_1, \ldots, f_s) = \{(a_1, \ldots, a_n) \in k^n : f_i(a_1, \ldots, a_n) = 0 \text{ for all } i \in [s]\}.$$

**Proposition 2.7.6.** *If $I = \langle f_1, \ldots, f_s \rangle$ then $\mathbf{V}(f_1, \ldots, f_s) = \mathbf{V}(I)$.*

*Proof.* For $\mathbf{a} \in \mathbf{V}(I)$, we have $f(\mathbf{a}) = 0$ for all $f \in I$; this implies $f_i(\mathbf{a}) = 0$ for every $i \in [s]$, and it follows that $\mathbf{a} \in \mathbf{V}(f_1, \ldots, f_s)$.

Conversely, let $\mathbf{a} \in \mathbf{V}(f_1, \ldots, f_s)$ and let $f \in I$. Then

$$f = \sum_{i=1}^{s} h_i f_i \text{ for some } h_i \in k[x_1, \ldots, x_n],$$

and hence $f(\mathbf{a}) = \sum_i h_i f_i(\mathbf{a}) = 0$. This implies $\mathbf{a} \in \mathbf{V}(I)$. □

Let's prove some properties about Gröbner bases.

**Proposition 2.7.7.** *Let $I \subseteq k[x_1, \ldots, x_n]$ be an ideal, and let $G = \{g_1, \ldots, g_t\}$ be a Gröbner basis with respect to a fixed monomial ordering $>$. Given $f \in k[x_1, \ldots, x_n]$, there is a unique $r \in k[x_1, \ldots, x_n]$ with the following properties:*

*(1) No term of $r$ is divisible by any of $\mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_t)$.*
*(2) There is $g \in I$ so that $f = g + r$.*

*In particular, $r$ is the remainder of $f$ upon division by $G$ (in any order!).*

*Proof.* The existence is just the division algorithm (Theorem 2.5.5).

Let us prove uniqueness. Suppose $f = g + r = g_2 + r_2$. Then note that $r - r_2 = g_2 - g \in I$, so if $r \neq r_2$ then

$$\mathrm{LT}(r - r_2) \in \langle \mathrm{LT}(I) \rangle = \langle \mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_t) \rangle,$$

which implies $\mathrm{LT}(r - r_2)$ is divisible by some monomial $\mathrm{LT}(g_i)$ (like earlier, this is Lemma 2.5.9). This contradicts property (1), since no terms of $r$ or $r_2$ can be divisible by any of $\mathrm{LT}(g_i)$. It follows that $r = r_2$, and hence $g = g_2$. □

**Corollary 2.7.8.** *Let $G = \{g_1, \ldots, g_t\}$ be a Gröbner basis for an ideal $I \subseteq k[x_1, \ldots, x_n]$. Then $f \in I$ if and only if the remainder upon division of $f$ by $G$ is zero.*

*Proof.* If $r = 0$ then certainly $f \in I$, since $G$ is a basis for $I$. Conversely, if $f \in I$, then $f = f + 0$, and $(f, 0)$ satisfies the two properties required of $(g, r)$ in Proposition 2.7.7. The uniqueness of $r$ in Proposition 2.7.7 says that $0 = r$ must be the remainder of $f$ upon division by $G$. □

We want to know how to tell if $\{f_1, \ldots, f_s\}$ is a Gröbner basis for $I$. Equivalently, we want to understand when

$$\langle \mathrm{LT}(f_1), \ldots, \mathrm{LT}(f_s) \rangle \subsetneq \langle \mathrm{LT}(I) \rangle.$$

**Definition 2.7.9.** Let $f, g \in k[x_1, \ldots, x_n]$ be nonzero polynomials, and let $\alpha = \mathrm{multideg}(f)$ and $\beta = \mathrm{multideg}(g)$. Also denote by $\gamma_i = \max_i(\alpha_i, \beta_i)$, and by $\gamma = (\gamma_1, \ldots, \gamma_n)$. We define

1. The least common multiple of $\mathrm{LM}(f)$ and $\mathrm{LM}(g)$ is defined to be $\mathbf{x}^\gamma$, and is denoted $\mathrm{lcm}(\mathrm{LM}(f), \mathrm{LM}(g))$.
2. The $S$-polynomial of $f$ and $g$ is defined to be

$$\frac{\mathbf{x}^\gamma}{\mathrm{LT}(f)} f - \frac{\mathbf{x}^\gamma}{\mathrm{LT}(g)} g,$$

and is denoted $S(f, g)$. △

**Example 2.7.10.** Let $f = x + 1$ and $g = x - 1$, and let $I = \langle f, g \rangle$. Then $S(f, g) = 2$. Note that $2 = \mathrm{LT}(S(f, g)) \in \langle \mathrm{LT}(I) \rangle$, whereas $2 \notin \langle \mathrm{LT}(f), \mathrm{LT}(g) \rangle = \langle x \rangle$. △

**Exercise:** (HW) We have $\mathrm{multideg}(S(f, g)) < \gamma$, where $\gamma = \mathrm{lcm}(\mathrm{LM}(f), \mathrm{LM}(g))$.

## 2.8 Sep 24, 2019

Let us recall Definition 2.7.9 again:

**Definition 2.8.1.** Let $f, g \in k[x_1, \ldots, x_n]$ be nonzero polynomials, and let $\alpha = \text{multideg}(f)$ and $\beta = \text{multideg}(g)$. Also denote by $\gamma_i = \max_i(\alpha_i, \beta_i)$, and by $\gamma = (\gamma_1, \ldots, \gamma_n)$. We define

1. The least common multiple of $\text{LM}(f)$ and $\text{LM}(g)$ is defined to be $\mathbf{x}^\gamma$, and is denoted $\text{lcm}(\text{LM}(f), \text{LM}(g))$.
2. The $S$-polynomial of $f$ and $g$ is defined to be

$$\frac{\mathbf{x}^\gamma}{\text{LT}(f)} f - \frac{\mathbf{x}^\gamma}{\text{LT}(g)} g,$$

and is denoted $S(f, g)$. △

We'll be using the following innocent-looking lemma a lot today:

**Lemma 2.8.2.** *Let $p_1, \ldots, p_s \in k[x_1, \ldots, x_n]$ be polynomials, all of the same multidegree $\text{multideg}(p_i) = \delta \in \mathbb{Z}^n_{\geq 0}$ for each $i \in [s]$. Then if*

$$\text{multideg}\left(\sum_{i=1}^s p_i\right) < \delta,$$

*then $\sum_{i=1}^s p_i$ can be written as a $k$-linear combination of $S(p_i, p_j)$. Moreover, each $S(p_i, p_j)$ has multidegree strictly less than $\delta$.*

*Proof.* We denote by $d_i \stackrel{\text{def}}{=} \text{LC}(p_i)$, so that $\text{LT}(p_i) = d_i \mathbf{x}^\delta$. Since $\text{multideg}(\sum p_i) < \delta$, the coefficient of $\mathbf{x}^\delta$ in $\sum p_i$ is zero, that is, we have

$$\sum_{i=1}^s d_i = 0.$$

Since $S(p_i, p_j) = \frac{1}{d_i} p_i - \frac{1}{d_j} p_j$, we write

$$\begin{aligned}
\sum_{i=1}^{s-1} d_i S(p_i, p_j) &= \sum_{i=1}^{s-1} d_i \left(\frac{1}{d_i} p_i - \frac{1}{d_s} p_s\right) \\
&= \sum_{i=1}^{s-1} p_i - \frac{p_s}{d_s} (\underbrace{d_1 + \cdots + d_{s-1}}_{=-d_s}) \\
&= \sum_{i=1}^{s-1} p_i + p_s = \sum_{i=1}^{s} p_i,
\end{aligned}$$

as desired. The fact that $\text{multideg}(S(p_i, p_j)) < \delta$ follows from the observation that the leading terms of $p_i$ and $p_j$ will cancel. □

We now have two big theorems back to back.

**Theorem 2.8.3** (Buchberger's Criterion)**.** *Let $I \subseteq k[x_1, \ldots, x_n]$ be an ideal. Then, a basis $G = \{y_1, \ldots, y_t\}$ is a Gröbner basis of $I$ if and only if the remainder of $S(g_i, g_j)$ upon division by $G$ (in any order) is zero.*

*Proof.* For the forwards direction, we note that if $G$ is a Gröbner basis, then since $S(g_i, g_j) \in I$ we must have $\overline{S(g_i, g_j)}^G = 0$. [Recall the notation $\overline{f}^G$, which denotes the remainder of $f$ upon division by the Gröbner basis $G$ in any order.]

Hence we need to prove the backwards direction. For each $f \in I$ with $f \neq 0$; we want to show that $\mathrm{LT}(f) \in \langle \mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_t) \rangle$. Since $G$ is a basis of $I$, we can express $f$ as

$$f = \sum_{i=1}^{t} h_i g_i, \quad h_i \in k[x_1, \ldots, x_n].$$

In particular, we have

$$\mathrm{multideg}(f) \leq \max_{i \in [t]} \{\mathrm{multideg}(h_i g_i) \colon h_i g_i \neq 0\}.$$

Let us pick a way of writing $f$ as $f = \sum h_i g_i$ so that

$$\delta \stackrel{\mathrm{def}}{=} \max_{i \in [t]} \{\mathrm{multideg}(h_i g_i) \colon h_i g_i \neq 0\}$$

is *minimal* among all ways of writing $f = \sum h_i g_i$. (This minimality will be crucial later. Intuitively, we're trying to minimize cancellations among leading terms of the $g_i$'s.)

Note that we have $\mathrm{multideg}(f) \leq \delta$. Furthermore, if $\mathrm{multideg}(f) = \delta$, then $\mathrm{multideg}(f) = \mathrm{multideg}(h_i g_i)$ for some $i \in [t]$. Then $\mathrm{LT}(f)$ is divisible by $\mathrm{LT}(g_i)$, and $\mathrm{LT}(f) \in \langle \mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_t) \rangle$. In other words, if $\mathrm{multideg}(f) = \delta$ for each $f \in I$, then $G$ is a Gröbner basis.

Let us use the fact that $\overline{S(g_i, g_j)}^G = 0$ for all $i \neq j$ to show that indeed $\mathrm{multideg}(f) = \delta$ for all $f \in I$. Suppose instead that there exists $f$ so that $\mathrm{multideg}(f) < \delta$. We have an expression

$$\begin{aligned} f = \sum_{i=1}^{t} h_i g_i &= \sum_{\substack{i \in [t] \\ \mathrm{mdeg}(h_i g_i) = \delta}} h_i g_i + \sum_{\substack{i \in [t] \\ \mathrm{mdeg}(h_i g_i) < \delta}} h_i g_i \\ &= \sum_{\substack{i \in [t] \\ \mathrm{mdeg}(h_i g_i) = \delta}} \mathrm{LT}(h_i) g_i + \sum_{\substack{i \in [t] \\ \mathrm{mdeg}(h_i g_i) = \delta}} (h_i - \mathrm{LT}(h_i)) g_i + \sum_{\substack{i \in [t] \\ \mathrm{mdeg}(h_i g_i) < \delta}} h_i g_i; \end{aligned}$$

here we use mdeg in the summation to mean multideg. Rearranging the terms in the above equation, we get

$$\sum_{\substack{i \in [t] \\ \mathrm{mdeg}(h_i g_i) = \delta}} \mathrm{LT}(h_i) g_i = f - \sum_{\substack{i \in [t] \\ \mathrm{mdeg}(h_i g_i) = \delta}} (h_i - \mathrm{LT}(h_i)) g_i - \sum_{\substack{i \in [t] \\ \mathrm{mdeg}(h_i g_i) < \delta}} h_i g_i, \tag{1}$$

where each term in the right hand side consists of terms whose multidegrees are strictly less than $\delta$.

The goal is to express the left side of Equation (1) using only multiples of $g_i$ with multidegree strictly less than $\delta$.

Let us denote, for each $i \in [t]$ with $\mathrm{multideg}(h_i g_i) = \delta$ by

$$p_i \stackrel{\mathrm{def}}{=} \mathrm{LT}(h_i) g_i.$$

Each $p_i$ has the same multidegree $\delta$, and yet the sum of the $p_i$ is the right hand side of Equation (1), all of whose terms have multidegree strictly less than $\delta$. We are allowed to apply Lemma 2.8.2 [there are inconsequential indexing issues, but never mind those] to obtain an expression

$$\sum_{\substack{i \in [t] \\ \mathrm{mdeg}(h_i g_i) = \delta}} \mathrm{LT}(h_i) g_i = \sum_{\substack{i,j \in [t] \\ i \neq j}} c_{i,j} S(p_i, p_j). \tag{2}$$

In our homework we will check that $S(p_i, p_j) = \mathbf{x}^{\delta - \gamma_{i,j}} S(g_i, g_j)$, where $\gamma_{i,j} = \mathrm{lcm}(\mathrm{LM}(g_i), \mathrm{LM}(g_j))$.

Since $\overline{S(g_i,g_j)}^{G} = 0$, the division algorithm (Theorem 2.5.5) gives

$$S(g_i,g_j) = \sum_{\ell=1}^{t} A_\ell^{ij} g_\ell,$$

where (crucially!) $\text{multideg}(A_\ell^{ij} g_\ell) \leq \text{multideg}(S(g_i,g_j))$.

In particular, we can write, for every $i,j \in [t]$ with $i \neq j$,

$$c_{i,j} S(p_i,p_j) = c_{i,j} \sum_{\ell=1}^{t} \mathbf{x}^{\delta-\gamma_{ij}} A_\ell^{ij} g_\ell;$$

every term on the right hand side has degree at most $S(p_i,p_j) < \delta$. In light of equation (2), we can express

$$\sum_{\substack{i\in[t] \\ \text{mdeg}(h_i g_i)=\delta}} \text{LT}(h_i) g_i$$

as a sum $\sum h_i' g_i$ where each $h_i' g_i$ has multidegree strictly less than $\delta$. This achieves our goal. □

**Theorem 2.8.4** (Buchberger's Algorithm)**.** *Let $I = \langle f_1,\ldots,f_s\rangle$ be an ideal, with $I \neq \{0\}$. A Gröbner basis can be computed using the following algorithm.*

```
Input:  F = (f_1, ..., f_s)
Output:  a Gröbner basis G = (g_1, ..., g_t) for I, with F ⊆ G
G := F
REPEAT
  G' := G
  FOR each pair {p,q}, p ≠ q in G' DO
    r := S(p,q)^{G'}
    IF r ≠ 0 THEN G := G ∪ {r}
UNTIL G = G'
RETURN G
```

Let's prove that this works.

*Proof.* Let $G = \{g_1,\ldots,g_t\}$, and define $\langle G\rangle = \langle g_1,\ldots,g_t\rangle$, as well as $\langle \text{LT}(G)\rangle = \langle \text{LT}(g_1),\ldots,\text{LT}(g_t)\rangle$. By induction, observe that at each step, we have $G \subseteq I$: it's true at the first step (when $G = F$), and enlarging $G$ consists of adding $\overline{S(p,q)}^{G'}$ with $p,q \in I$ (hence $S(p,q) \in I$, and hence $\overline{S(p,q)}^{G'} \in I$ too). Also observe that $G$ is always a basis for $I$, since $G \supseteq F$.

If the algorithm terminates, then $G' = G$. This means that $S(p,q)^{G'} = 0$ for all $p,q \in G'$, which means that $G' = G$ is a Gröbner basis by Buchberger's Criterion (Theorem 2.8.3).

It is left to show that the algorithm terminates. To see this, observe that $\langle \text{LT}(G')\rangle \subseteq \langle \text{LT}(G)\rangle$. We claim that if $G \neq G'$, then $\langle \text{LT}(G')\rangle \subsetneq \langle \text{LT}(G)\rangle$. This will show that the algorithm terminates, by the ascending chain condition (Theorem 2.7.5). [(!)]

Indeed, suppose $0 \neq r = \overline{S(p,q)}^{G'}$ was adjoined to $G$. Since $r$ is a remainder obtained by the division algorithm, $\text{LT}(r)$ is not divisible by any leading term of any element in $G'$. In particular, we have $\text{LT}(r) \notin \langle \text{LT}(g_i)\rangle$. Yet $r \in G$, so $\text{LT}(r) \in \langle \text{LT}(G)\rangle$. □

## 2.9 Sep 26, 2019

Recall from last class the Buchberger criterion, Theorem 2.8.3:

**Theorem 2.9.1** (Buchberger's Criterion). *Let $I \subseteq k[x_1, \ldots, x_n]$ be an ideal. Then, a basis $G = \{y_1, \ldots, y_t\}$ is a Gröbner basis of $I$ if and only if the remainder of $S(g_i, g_j)$ upon division by $G$ (in any order) is zero.*

We saw how to use this to construct a Gröbner basis (Theorem 2.8.4). Today we're going to prove some theoretical properties about these bases. We begin with

**Lemma 2.9.2.** *Let $G$ be a Gröbner basis of an ideal $I \subseteq k[x_1, \ldots, x_n]$. Let $p \in G$ be such that $\mathrm{LT}(p) \in \langle \mathrm{LT}(G \setminus \{p\}) \rangle$. Then $G \setminus \{p\}$ is a Gröbner basis of $I$.*

*Proof.* Since $\langle \mathrm{LT}(G) \rangle = \langle \mathrm{LT}(I) \rangle$, and $\mathrm{LT}(p) \in \langle \mathrm{LT}(G \setminus \{p\}) \rangle$, then $\langle \mathrm{LT}(G) \rangle = \langle \mathrm{LT}(G) \rangle = \langle \mathrm{LT}(G \setminus \{p\}) \rangle$, so if $G \setminus \{p\}$ is a basis then it is a Gröbner basis.

To see that $G \setminus \{p\}$ is a basis, we refer to the proof of Hilbert's Basis Theorem (see Theorem 2.7.2), where we showed that if $\langle \mathrm{LT}(I) \rangle = \langle \mathrm{LT}(G \setminus \{p\}) \rangle$, then $I = \langle G \setminus \{p\} \rangle$. □

Notice that we can scale elements of a Gröbner basis by scalars in $k$, so we can ask for our generators $g_i$ to have leading coefficient 1.

**Definition 2.9.3.** A Gröbner basis $G$ of an ideal $I$ is called a minimal Gröbner basis if $\mathrm{LC}(g_i) = 1$ and there doesn't exist $p \in G$ such that $\mathrm{LT}(p) \in \langle \mathrm{LT}(G \setminus \{p\}) \rangle$. △

We leave the following observation as an exercise:

**Observation 2.9.4.** Given a minimal Gröbner basis $G$, note that $\mathrm{LT}(G)$ is the (unique!) minimal basis for $\langle \mathrm{LT}(I) \rangle$. (See Proposition 2.6.7.) △

**Definition 2.9.5.** A reduced Gröbner basis $G$ for an ideal $I \subseteq k[x_1, \ldots, x_n]$ is a Gröbner basis such that

1. For all $p \in G$, we have $\mathrm{LC}(p) = 1$, and
2. For all $p \in G$, no monomial of $p$ is in $\langle \mathrm{LT}(G \setminus \{p\}) \rangle$. △

**Theorem 2.9.6.** *Let $I \neq \{0\}$ be an ideal in $k[x_1, \ldots, x_n]$. Then for a given monomial order $>$, $I$ has a unique reduced Gröbner basis.*

*Proof.* Note that all minimal Gröbner bases have the same leading term. With this in mind, let $G$ denote a minimal Gröbner basis for $I$. We say $f \in G$ is fully reduced for $G$ if no monomial of $f$ is in $\langle \mathrm{LT}(G \setminus \{f\}) \rangle$. (We want to produce a basis where each element is fully reduced.)

Note that if $f \in G$ is fully reduced with respect to $G$, then it is fully reduced with respect to any other minimal Gröbner basis (since all the minimal bases have the same ideal of leading terms.)

Given $f \in G$, we define $f_2 = \overline{f}^{G \setminus \{f\}}$, with the set $G \setminus \{f\}$ ordered however you like. Set

$$G_2 = (G \setminus \{f\}) \cup \{f_2\},$$

that is, let us replace $f$ with $f_2$ in $G$. We claim that $G_2$ is a minimal Gröbner basis for $I$.

Indeed, since $\mathrm{LT}(f)$ is not divisible by any of $\mathrm{LT}(G \setminus \{f\})$ (by assumption that $G$ was a minimal Gröbner basis), the term $\mathrm{LT}(f)$ goes into the remainder upon division of $f$ by $G \setminus \{f\}$; hence $\mathrm{LT}(f) = \mathrm{LT}(f_2)$. This implies that $\langle \mathrm{LT}(G_2) \rangle = \langle \mathrm{LT}(G) \rangle$, and $G_2 \subseteq I$. This means that $G_2$ is a Gröbner basis, and furthermore $G_2$ is minimal. Furthermore, $f_2$ is fully reduced with respect to $G_2$.

In light of the observation that an element $f \in G$ is fully reduced with respect to $G$ if and only if it is fully reduced with respect to any other minimal Gröbner basis, we can repeat this process to every element of $G$ to replace each $g_i \in G$ with a fully reduced version. Note that we do not have to revisit any already-fully-reduced element.

Let's prove the uniqueness of reduced Gröbner bases. Suppose $G$ and $\widetilde{G}$ are reduced Gröbner bases for $I$ with respect to a fixed monomial order. Since $G, \widetilde{G}$ are both minimal Gröbner bases, we have $\mathrm{LT}(G) = \mathrm{LT}(\widetilde{G})$ [there is no $\langle\cdot\rangle$; see Observation 2.9.4]. In particular, or $g \in G$, there exists $\widetilde{g} \in \widetilde{G}$ so that $\mathrm{LT}(g) = \mathrm{LT}(\widetilde{g})$. We claim that $g = \widetilde{g}$ (uniqueness follows).

Consider $g - \widetilde{g} \in I$. Since $G$ is a Gröbner basis, we have $\overline{g - \widetilde{g}}^{G} = 0$, but $\mathrm{LT}(g) = \mathrm{LT}(\widetilde{g})$ implies all terms of $g - \widetilde{g}$ only has terms of multidegree strictly less than $\mathrm{LT}(g)$. Furthermore, by condition (2) of reduced Gröbner bases, no terms of $g - \widetilde{g}$ are divisible by any elements of $\mathrm{LT}(G \setminus \{g\})$. Since none of the terms can be divisible by $\mathrm{LT}(g)$ either (they all have multidegree less than $\mathrm{LT}(g)$), we see that no term of $g - \widetilde{g}$ is divisible by any element of $\mathrm{LT}(G)$; hence $g - \widetilde{g} = \overline{g - \widetilde{g}}^{G} = 0$. □

Theorem 2.9.6 gives rise to the ideal equality algorithm: to tell if $\langle f_1, \ldots, f_s\rangle = \langle g_1, \ldots, g_t\rangle$, we fix a monomial order, compute the reduced Gröbner bases of the two sets of generators, and check if the resulting reduced Gröbner bases are equal.

We'll omit some of the next proofs; they're not too hard and will be optional HW. (In particular, they're variations of Buchberger's Criterion, and we've already seen all of the relevant ideas).

In the next few lectures, we'll start seeing some results about the Nullstellensatz (Chapter 4 in the book) which will use these specific variations of Buchberger's Criterion.

**Definition 2.9.7.** Fix a monomial order and let $G = \{g_1, \ldots, g_t\} \subseteq k[x_1, \ldots, x_n]$. Given $f \in k[x_1, \ldots, x_n]$ we say that $f$ reduces to 0 modulo $G$, denoted $f \to_G 0$, if $f$ has a standard representation

$$f = A_1 g_1 + \cdots + A_t g_t,$$

such that whenever $A_i g_i \neq 0$ we have $\mathrm{multideg}(f) \geq \mathrm{multideg}(A_i g_i)$. △

With this language, we'll "say a few things – nothing that we wouldn't be able to prove on an exam, for example."

**Theorem 2.9.8.** *A basis $G = \{g_1, \ldots, g_t\}$ for an ideal $I \subseteq k[x_1, \ldots, x_n]$ is a Gröbner bases if and only if $S(g_i, g_j) \to_G 0$ for all $i \neq j$.*

*Proof.* Optional HW. □

**Definition 2.9.9.** Given nonzero polynomials $F = (f_1, \ldots, f_s)$, we say that

$$S(f_i, f_j) = \sum_{\ell=1}^{s} A_\ell f_\ell$$

is a least common multiple representation (sometimes LCM rep) if

$$\mathrm{lcm}(\mathrm{LM}(f_i), \mathrm{LM}(f_j)) > \mathrm{LM}(A_\ell f_\ell)$$

whenever $A_\ell f_\ell \neq 0$. △

**Proposition 2.9.10.** *Every standard representation is a LCM representation. The converse doesn't hold.*

*Proof.* For any $S$-polynomial, we have $\mathrm{LM}(S(f_i, f_j)) < \mathrm{lcm}(\mathrm{LM}(f_i), \mathrm{LM}(f_j))$. If we have a standard representation

$$S(f_i, f_j) = \sum_{\ell=1}^{s} A_\ell f_\ell$$

then by definition we have $\mathrm{LM}(S(f_i, f_j)) \geq \mathrm{LM}(A_\ell f_\ell)$ whenever $A_\ell f_\ell \neq 0$. Chaining these two inequalities we get

$$\mathrm{lcm}(f_i, f_j) > \mathrm{LM}(S(f_i, f_j)) \geq \mathrm{LM}(A_\ell, f_\ell).$$

That the converse doesn't hold will be in the HW. □

**Theorem 2.9.11.** *A basis $G = \{g_1, \ldots, g_\ell\}$ for an ideal $I \subseteq k[x_1, \ldots, x_n]$ is a Gröbner basis if and only if for all $i \neq j$, $S(g_i, g_j)$ has an LCM representation.*

*Proof.* Optional HW. □

(This is less obvious than Theorem 2.9.8 above.)

# 3 Nullstellensatz

[Although this is chapter 3 in my notes, this is chapter 4 in the book.]

## 3.10 Oct 1, 2019

Let's restart and continue, in the following sense: we'll continue what we've been talking about in Section 2, but we'll also recall our original setup from long ago in Section 1. Recall that for a variety $V \subseteq k^n$ we defined

$$\mathbf{I}(V) \stackrel{\text{def}}{=} \{f \in k[\mathbf{x}] \colon f(\mathbf{a}) = 0 \text{ for all } \mathbf{a} \in V\}.$$

Now $\mathbf{I}$ gives a map

$$\begin{aligned} \mathbf{I} \colon \{\text{Varieties}\} &\to \{\text{Ideals}\} \\ V &\mapsto \mathbf{I}(V). \end{aligned}$$

We also defined

$$\mathbf{V}(I) \stackrel{\text{def}}{=} \{\mathbf{a} \in k^n \colon f(\mathbf{a}) = 0 \text{ for all } f \in I\}$$

giving a map

$$\begin{aligned} \mathbf{V} \colon \{\text{Ideals}\} &\to \{\text{Varieties}\} \\ I &\mapsto \mathbf{V}(I). \end{aligned}$$

We asked ourselves what the relationship between $I$ and $\mathbf{I}(\mathbf{V}(I))$ is; we saw Lemma 1.2.9 and Example 1.2.10. Consider also the univariate ideals $\langle x \rangle, \langle x^2 \rangle, \langle x^3 \rangle, \cdots \subseteq \mathbb{C}[x]$. They all give the same variety,

$$\mathbf{V}(\langle x \rangle) = \mathbf{V}(\langle x^2 \rangle) = \mathbf{V}(\langle x^3 \rangle) = \cdots = \{0\}.$$

Suppose our field is not algebraically closed, e.g. take $k = \mathbb{R}$. Then the ideals $\langle 1 \rangle, \langle 1 + x^2 \rangle$, and $\langle 1 + x^2 + x^4 \rangle$ all give the same variety, namely

$$\mathbf{V}(\langle 1 \rangle) = \mathbf{V}(\langle 1 + x^2 \rangle) = \mathbf{V}(\langle 1 + x^2 + x^4 \rangle) = \emptyset.$$

This might reasonably be called "bad behavior" of the function $\mathbf{V}$. The weak Nullstellensatz says that when $k$ is algebraically closed, this behavior does not occur. Let's first state it for the univariate polynomial ring $k[x]$, with $k$ algebraically closed.

**Lemma 3.10.1** (Weak Nullstellensatz for $k[x]$)**.** *If $k$ is an algebraically closed field, and $I \subseteq k[x]$ is such that $\mathbf{V}(I) = \emptyset$, then $I = \langle 1 \rangle = k[x]$.*

*Proof.* Recall that $k[x]$ is a PID. So $I = \langle f \rangle$ for $f \in k[x]$. Now recall

$$\mathbf{V}(I) = \mathbf{V}(\langle f \rangle) = \{a \in k \colon f(a) = 0\}$$

is nonempty whenever $f$ is nonconstant, by definition of algebraically closed. Hence $f$ would need to be constant, and $1 \in I$. This implies $I = k[x]$. □

**Theorem 3.10.2** (Weak Nullstellensatz for $k[x_1, \ldots, x_n]$)**.** *If $k$ is an algebraically closed field, and $I \subseteq k[x_1, \ldots, x_n]$ is such that $\mathbf{V}(I) = \emptyset$, then $I = \langle 1 \rangle = k[x_1, \ldots, x_n]$.*

*Proof.* We'll prove the equivalent statement that if $I \subsetneq k[x_1, \ldots, x_n]$, then $\mathbf{V}(I) \neq \emptyset$. Let us pick $a \in k$ and $f \in k[x_1, \ldots, x_n]$. We define $\overline{f} \in k[x_1, \ldots, x_{n-1}]$ by

$$\overline{f} \stackrel{\text{def}}{=} f(x_1, \ldots, x_{n-1}, a).$$

Also define

$$I_{x_n=a} \stackrel{\text{def}}{=} \{\overline{f} \colon f \in I\} \subseteq k[x_1, \ldots, x_{n-1}].$$

We'll show in a homework set that $I_{x_n=a}$ is an ideal.

Here's the central claim to the proof.

**Claim 1.** If $k$ is an algebraically closed field, and $I \subsetneq k[x_1, \ldots, x_n]$, then there exists $a \in k$ so that $I_{x_n=a} \subsetneq k[x_1, \ldots, x_{n-1}]$.

Let's first see why Claim 1 implies the Nullstellensatz, and then prove Claim 1. Indeed, assuming claim 1, there exist $a_1, \ldots, a_n \in k$ so that

$$(\ldots(((I_{x_n=a_n})_{x_{n-1}=a_{n-1}})_{x_{n-2}=a_{n-2}})\ldots)_{x_1=a_1}. \subsetneq k.$$

Since the only ideals of a field $k$ are $\{0\}$ and $k$ itself (exercise, if you don't see it!), the above ideal must be the zero ideal. This says that for all $f \in I$, we have $f(a_1, \ldots, a_n) = 0$. In particular, $(a_1, \ldots, a_n) \in \mathbf{V}(I)$ and hence $\mathbf{V}(I) \neq \emptyset$. This accomplishes what we said we wanted to accomplish in the first sentence of this proof, that is, we've shown that $\mathbf{V}(I) \neq \emptyset$.

Let's prove Claim 1. There are two cases, which we will consider separately:

- Case (1) concerns the case where $I \cap k[x_n] \neq \{0\}$, and
- Case (2) concerns the case where $I \cap k[x_n] = \{0\}$.

**Let's tackle Case (1) first.** If there exists $f \in I$ nonconstant that only depends on $x_n$, then

$$f = c\prod_{i=1}^{r}(x_n - b_i)^{m_i},$$

with $c, b_1, \ldots, b_r \in k$. [Note that here we are using algebraically-closed-ness of $k$!] We'll use later the observation that

$$\prod_{i=1}^{r}(x_n - b_i)^{m_i} = c^{-1}f \in I,$$

so keep this in mind.

If $I_{x_n=b_i} \neq k[x_1, \ldots, x_{n-1}]$ for any $i \in [r]$, we've proven our claim, so let us suppose otherwise. We have $r$ many ideals $I_{x_n=b_1}, \ldots, I_{x_n=b_r}$, all equal to $k[x_1, \ldots, x_{r-1}]$. We'll obtain a contradiction in the following way:

Note that $1 \in I_{x_n=b_i}$ for all $i \in [r]$ if and only if there exists $B_i \in I$ so that $B_i(x_1, \ldots, x_{n-1}, b_i) = 1$ for all $i \in [r]$. Observe that

$$B_i(x_1, \ldots, x_{n-1}, b_i) = B_i(x_1, \ldots, x_{n-1}, x_n - (x_n - b_i)),$$

and also [here's a tricky part]

$$B_i(x_1, \ldots, x_n - (x_n - b_i)) = B_i(x_1, \ldots, x_n) + A_i(x_1, \ldots, x_n)(x - b_i)$$

for some $A_i(x_1, \ldots, x_n) \in k[x_1, \ldots, x_n]$, by the binomial theorem: for every $b_n^\ell$ term in $B_i(x_1, \ldots, x_{n-1}, b_i)$, the binomial theorem says we can expand

$$(x_n - (x_n - b_i))^\ell = x_n^\ell + \binom{\ell}{1}x_n^{\ell-1}(-(x_n - b_i)) + \binom{\ell}{2}x_n^{\ell-2}(-(x_n - b_i))^2 + \binom{\ell}{3}x_n^{\ell-3}(-(x_n - b_i))^3 + \ldots$$

and we can collect all the terms involving an $x_n - b_i$ to get $A_i(x_1, \ldots, x_n)$, and the remaining term is the $x_n^\ell$ term "replaces" the $b_i^\ell$ term in $B_i(x_1, \ldots, b_i)$. So the resulting polynomial is $B_i(x_1, \ldots, x_n)$.

Chaining all these equalities together we get

$$1 = B_i(x_1, \ldots, x_{n-1}, b_i) = B_i(x_1, \ldots, x_n) + A_i(x_1, \ldots, x_n)(x_n - b_i),$$

for every $i \in [r]$. Let's raise everything to the $m_i$th power and multiply everything together, to get

$$1 = \prod_{i=1}^{r} 1^{m_i} = \prod_{i=1}^{r} \left( \Big( B_i(x_1, \ldots, x_n) + A_i(x_1, \ldots, x_n)(x_n - b_i) \Big)^{m_i} \right), \tag{3}$$

and expand in the following way. We will write the right hand side as

$$1 = \underbrace{B(x_1, \ldots, x_n)}_{\text{"everything else"}} + \Big( \prod_{i=1}^{r} (A_i(x_1, \ldots, x_n))^{m_i} \Big) \underbrace{\prod_{i=1}^{r} (x_n - b_i)^{m_i}}_{c^{-1} f \in I};$$

the summand on the right is obtained from always taking the $A_i(x_1, \ldots, x_n)(x - b_i)$ term in the product in Equation (3); observe that it is in $I$. The summand on the left is the "everything else", namely, those terms that are obtained from taking some $B_i(x_1, \ldots, x_n)$ in the product in Equation (3). Since $B_i(x_1, \ldots, x_n) \in I$, the whole $B(x_1, \ldots, x_n) \in I$ as well.

Finally, we've written 1 as a sum of two elements of $I$. This means that $1 \in I$, contrary to our assumption that $I \subsetneq k[x_1, \ldots, x_n]$ from the very first sentence of this proof. **This proves that in case (1) we always have some $i \in [r]$ so that $I_{x_n = b_i} \neq k[x_1, \ldots, x_{n-1}]$.**

**Let's tackle Case (2) now.** In this case, we have $I \cap k[x_n] = \{0\}$. Let us take a Gröbner basis $G = \{g_1, \ldots, g_t\}$ for $I$ with respect to the lex order, with $x_1 > \cdots > x_n$. We write $\mathbf{x} = (x_1, \ldots, x_{n-1})$ for the first $n-1$ variables and $\alpha(i) = (\alpha(i)_1, \ldots, \alpha(i)_{n-1}) \in \mathbb{Z}_{\geq 0}^{n-1}$ for the exponent vector so that $g_i \in G$ can be written

$$g_i = \mathbf{x}^{\alpha(i)} c_i(x_n) + \left( \sum (\text{monomials} <_{\text{lex}} \mathbf{x}^{\alpha(i)}) \cdot (\text{polynomials in } k[x_n]) \right). \tag{4}$$

Note that this isn't quite the same as pulling out the leading term of $g_i$; it's as if we're pulling out the leading term among the monomials in the first $n-1$ variables. For example, if $n = 3$ and

$$g = 2x_1^3 x_2 x_3^4 - 11 x_1^3 x_2 x_3 + x_1^3 x_3 + 3 x_1^2 x_2^7 + x_1 x_2 x_3^{13} + x_1,$$

then we'd write

$$g = x_1^3 x_2 (2x_3^4 - 11 x_3) + \Big( (x_1^3) \cdot (x_3) + (x_1^2 x_2^7) \cdot (3) + (x_1 x_2) \cdot (x_3^{13}) + (x_1) \cdot (1) \Big),$$

so $\alpha = (3, 1)$ and $c(x_3) = 2x_3^4 - 11 x_3$.

Let's return to the proof of Case (2); we've written $g_i$ as in Equation (4). Note that in this equation, we necessarily have $c_i(x_n) \neq 0$ as a polynomial.

Let us pick $a \in k$ such that $c_i(a) \neq 0$ for all $i \in [t]$. The existence of this will be left as an exercise. [At the very least, one would need to prove that algebraically closed fields are infinite. This follows from observing, for example, that algebraically closed fields should contain all the roots of unity.]

Let us consider the polynomials

$$\overline{g}_i(x_1, \ldots, x_{n-1}) \stackrel{\text{def}}{=} g_i(x_1, \ldots, x_{n-1}, a) \in k[x_1, \ldots, x_{n-1}].$$

We claim that $\overline{g}_i$ form a basis of $I_{x_n=a}$. This is not terribly difficult and will be left as an exercise.

Let us substitute $x_n = a$ into Equation (4), and take the leading terms of both sides. We arrive at

$$\mathrm{LT}(\overline{g}_i) = c_i(a)\mathbf{x}^{\alpha(i)}$$

where $c_i(a) \neq 0$ by assumption on $a$. Observe also that $\mathbf{x}^{\alpha(i)} \neq 1$ because then $g_i \in I \cap k[x_n] = \{0\}$, which is a contradiction. Hence $\mathrm{LT}(\overline{g}_i)$ is nonconstant for all $i$. For time reasons, let us show the following claim next lecture:

**Claim 2.** The set $\{\overline{g}_i\}$ is a Gröbner basis for $I_{x_n=a}$.

Note that since none of the $\mathrm{LT}(\overline{g}_i) = 1$, we have $1 \notin \mathrm{LT}(I_{x_n=a}) = \langle \mathrm{LT}(\overline{g}_1), \ldots, \mathrm{LT}(\overline{g}_t)\rangle$, by Lemma 2.5.9. Then if $1 \in I_{x_n=a}$, then $1 \in \mathrm{LT}(I_{x_n=a})$, so that's impossible. Hence $I_{x_n=a} \neq k[x_1, \ldots, x_{n-1}]$, which proves **in Case (2), the conclusion of Claim 1 is always satisfied**. □

## 3.11 Oct 3, 2019

[Reminder: There's a prelim in a week! It'll happen in class, on October 10. The emphasis will be on Chapters 1 and 2, although it might touch on Chapter 4. We should definitely know why all of the algorithms introduced in this class work (i.e., why they terminate, and so on.)]

Let's recall where we left off last time.

**Theorem 3.11.1** (The Weak Nullstellensatz, cf. Theorem 3.10.2)**.** *Let $k$ be an algebraically closed field, and let $I \subseteq k[x_1, \dots, x_n]$ be an ideal. Then $\mathbf{V}(I) = \emptyset$ implies that $I = \langle 1 \rangle$.*

*Proof.* We were partway through the proof last time. For $f \in k[x_1, \dots, x_n]$ and $a \in k$, we set

$$\overline{f} \stackrel{\text{def}}{=} f(x_1, \dots, x_{n-1}, a) \quad \text{and} \quad I_{x_n = a} \stackrel{\text{def}}{=} \{\overline{f} \colon f \in I\}.$$

We also took a Gröbner basis $\{g_1, \dots, g_t\}$ for $I$ with respect to the lexicographic order $>_{\text{lex}}$ with $x_1 > x_2 > \cdots > x_n$, and we had:

**Claim 2.** The polynomials $\overline{g}_i$ is a Gröbner basis for $I_{x_n = a}$.

The basis part is an easy exercise, and the Gröbner part will be proven. In order to prove this, we had written

$$g_i = \mathbf{x}^{\alpha(i)} c_i(x_n) + \left( \sum (\text{monomials} <_{\text{lex}} \mathbf{x}^{\alpha(i)}) \cdot (\text{polynomials in } k[x_n]) \right). \tag{5}$$

We picked $a \in k$ to be so that $c_i(a) \neq 0$ for all $i \in [t]$, and left the existence of such an $a$ as an exercise. We observed that $\text{LT}(\overline{g}_i) = c_i(a)\mathbf{x}^{\alpha(i)}$.

Let us now recall Theorem 2.9.11, which says that a basis $G = \{g_i\}$ is a Gröbner basis if and only if each $S(g_i, g_j)$ has an LCM representation with the $\{g_i\}$; recall that an LCM representation is an expression

$$S(g_i, g_j) = \sum_{\ell=1}^{t} A_\ell g_\ell$$

such that $\text{lcm}(\text{LM}(g_i), \text{LM}(g_j)) > \text{LM}(A_\ell g_\ell)$. (Just in case, see Definition 2.7.9 for the definition of $S$-polynomial.)

Let's prove Claim 2 now. We want to produce LCM representatives for $S(\overline{g}_i, \overline{g}_j)$ for all $i \neq j$. Let us define the polynomials

$$S_{ij}(x_1, \dots, x_n) \stackrel{\text{def}}{=} c_j(x_n) \frac{\mathbf{x}^{\gamma_{ij}}}{\mathbf{x}^{\alpha(i)}} g_i - c_i(x_n) \frac{\mathbf{x}^{\gamma_{ij}}}{\mathbf{x}^{\alpha(j)}} g_j,$$

where $\gamma_{ij} = \text{lcm}(\mathbf{x}^{\alpha(i)}, \mathbf{x}^{\alpha(j)})$. (As a reminder, note that $\alpha(i)$ is an $(n-1)$ tuple, and $\mathbf{x} = (x_1, \dots, x_{n-1})$.)

Here we make the crucial claim that $\mathbf{x}^{\gamma_{ij}} > \text{LM}(S_{ij})$. To see this, observe that in light of Equation (5), the $(x_1, \dots, x_{n-1})$ part of any monomial appearing in $S_{ij}$ has exponent vector strictly less than $\gamma_{ij}$. (This is similar to what we did in the homework.) Then, lex has the nice property that if $(\alpha_1, \dots, \alpha_{n-1}) >_{\text{lex}} (\beta_1, \dots, \beta_{n-1})$, then $(\alpha_1, \dots, \alpha_n) >_{\text{lex}} (\beta_1, \dots, \beta_n)$ for any $\alpha_n$ and $\beta_n$.

Thus

$$\overline{S}_{ij} = S_{ij}(x_1, \dots, x_{n-1}, a) = c_j(a) \frac{\mathbf{x}^{\gamma_{ij}}}{\mathbf{x}^{\alpha(i)}} \overline{g}_i - c_i(a) \frac{\mathbf{x}^{\gamma_{ij}}}{\mathbf{x}^{\alpha(j)}} \overline{g}_j = c_i(a) c_j(a) S(\overline{g}_i, \overline{g}_j).$$

Now $S_{ij} \in I$, and $G$ a Gröbner basis for $I$, implies that we can take an LCM representation

$$S_{ij} = \sum_{\ell=1}^{t} A_\ell g_\ell,$$

where $\mathrm{LM}(S_{ij}) \geq \mathrm{LM}(A_\ell g_\ell)$ whenever $A_\ell g_\ell \neq 0$. "Barring" both sides we get

$$c_i(a)c_j(a)S(\overline{g}_i, \overline{g}_j) = \overline{S}_{ij} = \sum_{\ell=1}^{t} \overline{A}_\ell \overline{g}_\ell,$$

with the leading monomial of the left hand side being $\mathbf{x}^{\gamma_{ij}}$. In light of the chain of inequalities

$$\mathbf{x}^{\gamma_{ij}} > \mathrm{LM}(S_{ij}) \geq \mathrm{LM}(A_\ell g_\ell) \geq \mathrm{LM}(\overline{A}_\ell g_\ell),$$

we see that we have obtained an LCM representation for $S(\overline{g}_i, \overline{g}_j)$. This shows that $\overline{g}_i$ form a Gröbner basis for $I_{x_n=a}$. □

The Weak Nullstellensatz (Theorem 3.11.1) says that any system of polynomial equations which generates an ideal strictly smaller than $\mathbb{C}[x_1, \ldots, x_n]$ has a common zero in $\mathbb{C}^n$.

We can thus answer the question of when polynomials over algebraically closed fields have a common solution. For example, take $k = \mathbb{C}$, and let $f_1, \ldots, f_s \in \mathbb{C}[x_1, \ldots, x_n]$. Then $f_1, \ldots, f_s = 0$ fail to have a common solution if and only if $\mathbf{V}(f_1, \ldots, f_s) = \emptyset$, if and only if $\langle f_1, \ldots, f_s \rangle = \mathbb{C}[x_1, \ldots, x_n]$, if and only if $1 \in \langle f_1, \ldots, f_s \rangle$. To answer the question of when $1 \in I$ is in some ideal, we have the following claim:

**Proposition 3.11.2.** *The set $\{1\}$ is the only reduced Gröbner basis of $\langle 1 \rangle = I$, for any monomial order.*

*Proof.* Suppose $\{g_1, \ldots, g_t\}$ is a Gröbner basis for $I$. Then, since $1 \in I$ we have $1 \in \mathrm{LT}(I) = \langle \mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_t) \rangle$. By Lemma 2.5.9, 1 is divisible by some $\mathrm{LT}(g_i)$, so $\mathrm{LT}(g_i)$ must be constant. This means that every other $\mathrm{LT}(g_j)$, when $j \neq i$, is also divisible by $\mathrm{LT}(g_i)$. Thus we remove these $g_j$ from our Gröbner basis when we construct a reduced Gröbner basis.

Furthermore, if $\mathrm{LT}(g_i)$ is constant, then so is $g_i$, because we've proven in HW 4 that $\mathbf{0} \in \mathbb{Z}^n_{\geq 0}$ must be the minimal element of any monomial ordering. Also, because elements of a reduced Gröbner basis have leading coefficient 1, it follows that $g_i = 1$ is the constant 1 function. □

Let us recall we had two ideals $\langle x \rangle$ and $\langle x^2 \rangle$, such that $\mathbf{V}(\langle x \rangle) = \mathbf{V}(\langle x^2 \rangle) = \{0\}$. More generally, if $f \in k[x_1, \ldots, x_n]$, then $\mathbf{V}(\langle f \rangle) = \mathbf{V}(\langle f^r \rangle)$ for any $r \geq 1$, essentially because for any $\alpha \in k$, we have $\alpha = 0$ if and only if $\alpha^r = 0$.

**Theorem 3.11.3** (Hilbert Nullstellensatz)**.** *Let $k$ be algebraically closed. If $f, f_1, \ldots, f_s \in k[x_1, \ldots, x_n]$, then*

$$f \in \mathbf{I}(\mathbf{V}(f_1, \ldots, f_s)) \quad \text{if and only if} \quad f^m \in \langle f_1, \ldots, f_s \rangle \text{ for some } m \geq 1.$$

This theorem underpins the dictionary between algebra and geometry, which we'll see bits of in this course. Take a class in algebraic geometry to learn more!

*Proof.* The backwards direction is not terribly difficult. (We'll also see essentially a proof of this when we try to prove the other direction.)

Let's take $f \in \mathbf{I}(\mathbf{V}(f_1, \ldots, f_s))$; our goal is to show that $f^m \in \langle f_1, \ldots, f_s \rangle$ for some $m \geq 1$. The statement $f \in \mathbf{I}(\mathbf{V}(f_1, \ldots, f_s))$ means that $f$ vanishes on the set of common zeros of $f_1, \ldots, f_s$. We proceed with the following trick, which will allow us to deduce the theorem from the Weak Nullstellensatz (Theorem 3.10.2).

Consider

$$\widetilde{I} = \langle f_1, \ldots, f_s, 1 - yf \rangle \subseteq k[x_1, \ldots, x_n, y]. \tag{6}$$

Note that the $f_i$ are considered as elements of $k[x_1, \ldots, x_n, y]$. Our claim is that $\mathbf{V}(\widetilde{I}) = \emptyset$.

Indeed, let us consider a point $(a_1, \ldots, a_n, a_{n+1}) \in k^{n+1}$. We want to show it is not in $\mathbf{V}(\widetilde{I})$. There are two cases: in the first case, we have $(a_1, \ldots, a_n) \in \mathbf{V}(f_1, \ldots, f_s) \subseteq k^n$, and in the second, we have $(a_1, \ldots, a_n) \notin \mathbf{V}(f_1, \ldots, f_s) \subseteq k^n$.

In the first case, we have $f_1(a_1, \ldots, a_n) = \cdots = f_s(a_1, \ldots, a_n) = 0$. Since $f$ vanishes on $\mathbf{V}(f_1, \ldots, f_s)$ too, we have $f(a_1, \ldots, a_n) = 0$ as well. Then the polynomial $1 - yf$, evaluated at the point $(a_1, \ldots, a_n, a_{n+1})$, is equal to

$$(1 - yf)(a_1, \ldots, a_n, a_{n+1}) = 1 - a_{n+1} \underbrace{f(a_1, \ldots, a_n)}_{=0} = 1 \neq 0,$$

so $(a_1, \ldots, a_{n+1}) \notin \mathbf{V}(f_1, \ldots, f_s, 1 - yf)$, since $1 - yf$ does not vanish on that point.

In the second case, we have $f_i(a_1, \ldots, a_n) \neq 0$ for some $i$. Then, thinking of $f_i \in k[x_1, \ldots, x_n, y]$, we still have $f_i(a_1, \ldots, a_n, a_{n+1}) \neq 0$, hence $(a_1, \ldots, a_n, a_{n+1}) \notin \mathbf{V}(\widetilde{I})$ either.

Thus we have proven $\mathbf{V}(\widetilde{I}) = \emptyset$. The Weak Nullstellensatz (Theorem 3.10.2) implies $1 \in \widetilde{I}$. This means that there exists $p_1, \ldots, p_s, q \in k[x_1, \ldots, x_n, y]$ such that

$$1 = \left( \sum_{i=1}^{s} p_i f_i \right) + q(x_1, \ldots, x_n, y)(1 - yf).$$

Let us set $y = 1/f$. Because we're (only temporarily!) going outside of the world of polynomials with this substitution, this step is a little shady; to make everything rigorous we should say that our computations will be done in the field of rational functions. But allow us to shove this under the rug.

We get the expression

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

where the right hand side is a ratio of polynomials. Let us clear denominators, by multiplying both sides by $f^m$ until the $p_i(x_1, \ldots, x_n, \frac{1}{f})$ become polynomials; for some $m \geq 1$, we get

$$f^m = \sum_{i=1}^{s} A_i f_i,$$

as desired. □

**Definition 3.11.4.** Let $I \subseteq k[x_1, \ldots, x_n]$ be an ideal. Define the radical of $I$ as

$$\sqrt{I} \stackrel{\text{def}}{=} \{f \colon f^m \in I \text{ for some } m \geq 1.\}$$ △

Observe that $I \subseteq \sqrt{I}$, where the equality isn't always true.

# 4 Practice Prelim

## 4.12 Oct 8, 2019

We discussed a "practice prelim" today. Prof Mészáros also gave a teaser problem for us to think about if we were done early. I'll give my solutions, although it's much easier to do this when I can cite theorems at will! Without further ado:

**Problem 1.**

(a) Let $V \subseteq k^n$ be an affine variety. Define $\mathbf{I}(V)$.
(b) Prove that $\mathbf{I}(V)$ is an ideal.
(c) For affine varieties $V, W \subseteq k^n$, prove $V \subseteq W \iff \mathbf{I}(V) \subseteq \mathbf{I}(W)$.
(d) Let $f_1, \dots, f_s \in k[x_1, \dots, x_n]$. What is the relationship between $\langle f_1, \dots, f_s \rangle$ and $\mathbf{I}(\mathbf{V}(\langle f_1, \dots, f_s \rangle))$?

**Problem 2.** Choose the lexicographic order with $x >_{\text{lex}} y$ on $k[x, y]$. Let $f_1 = x^3y - xy^2 + 1$ and $f_2 = x^2y^2 - y^3 - 1$. Let $I = \langle f_1, f_2 \rangle$. Show that $\{f_1, f_2\}$ is not a Gröbner basis.

**Problem 3.** State and prove Buchberger's algorithm.

**Problem 4.**

(a) State what it means for $f = 0 \in k[x_1, \dots, x_n]$ and for $f \colon k^n \to k$ to be the zero function.
(b) $k = \mathbb{F}_2$?
(c) Let $k$ be an infinite field. Prove $f = 0$ as a polynomial if and only if it's 0 as a function.

**Teaser.** Let $G = (V, E)$ be a finite graph (so $V = \{v_1, \dots, v_n\}$ is a set of vertices, and $E = \{\{v_i, v_j\} \subseteq V\}$ is a set of edges). Reduce the task of deciding whether $G$ has an $m$-coloring to the Nullstellensatz.

*Proof of Problem 1.* For part a), see Definition 1.2.7.

For part b), see Lemma 1.2.8.

For part c), see Proposition 1.2.11.

For part d), see Lemma 1.2.9 and Example 1.2.10. □

*Proof of Problem 2.* By Buchberger's Criterion (Theorem 2.8.3) it suffices to check whether or not $S(f_1, f_2)$ gives remainder 0 upon division by $\{f_1, f_2\}$. Observe $S(f_1, f_2) = x + y$ and neither term is divisible by $x^3y$ or $x^2y^2$, both terms $x$ and $y$ go into the remainder. Thus it's not a Gröbner basis. □

*Proof of Problem 3.* See Theorem 2.8.4. □

*Proof of Problem 4.* For part a), the zero polynomial of $k[x_1, \dots, x_n]$ is the polynomial whose coefficients are all zero. A polynomial $f \in k[x_1, \dots, x_n]$ gives rise to a function $f \colon k^n \to k$ given by sending $\mathbf{a} \in k^n$ to $f(\mathbf{a}) \in k$; it is the zero function if $f(\mathbf{a}) = 0$ for all $\mathbf{a} \in k^n$.

For part b), consider $x^2 + x \in \mathbb{F}_2[x]$.

For part c), see Proposition 1.1.3. □

[I feel like I'm writing an article for nlab.]

*Proof of Teaser.* [Proof is intentionally left terse.]

We set $V(G) = [n]$. For each $i \in [n]$, define $f_i \stackrel{\text{def}}{=} x_i^m - 1$ [got this idea from Isaac]. For each $e = \{i, j\} \in E$, define

$$f_e \stackrel{\text{def}}{=} \prod_{\substack{\ell_1, \ell_2 \in [m] \\ \ell_1 \neq \ell_2}} (x_i - x_j - e^{2\pi i \ell_1/m} + e^{2\pi i \ell_2/m}).$$

[The insight is that $f_e$ has finite degree, now that $f_1, \ldots, f_n$ have been defined.] Define

$$I \stackrel{\text{def}}{=} \langle f_i, f_e \colon i \in [n], e \in E \rangle \subseteq \mathbb{C}[x_1, \ldots, x_n].$$

Observe that $p = (p_1, \ldots, p_n) \in \mathbf{V}(I)$ if and only if

$$\underbrace{p_j = e^{2\pi i \ell_j/m} \text{ for some } \ell_j \in [m]}_{f_j = 0}, \text{ with } \underbrace{\ell_i \neq \ell_j \text{ for each } \{i, j\} \in E}_{f_{\{i,j\}} = 0}.$$

This is true if and only if $j \mapsto \ell_j$ is an $m$-coloring. Thus, the problem reduces to computing a reduced Gröbner basis for $I$; see Proposition 3.11.2. □

# 5 Nullstellensatz

## 5.13 Oct 17, 2019

[We got our prelims back today.]

Let's talk for a bit about the teaser problem posed in the practice prelim.

Let $G = (V, E)$ be a finite graph. This means that $V$ (sometimes denoted $V(G)$) is a set of vertices $V = \{1, \dots, m\}$ and $E$ (sometimes denoted $E(G)$) is a set of edges $E = \{\{i, j\} \colon i, j \in V\}$. (Not every pair of vertices $i, j$ must form an edge of $G$.)

An $n$-coloring of $G$ is an assignment $f \colon V(G) \to [n]$ so that $f(u) \neq f(v)$ if $\{u, v\} \in E(G)$. We want to decide (algorithmically) whether $G$ has an $n$-coloring.

In order to answer this question, let us recall the Weak Nullstellensatz:

**Theorem 5.13.1** (The Weak Nullstellensatz, cf. Theorems 3.10.2 and 3.11.1)**.** *Let $k$ be an algebraically closed field and let $I \subseteq k[x_1, \dots, x_n]$ be an ideal satisfying $\mathbf{V}(I) = \emptyset$. Then $I = k[x_1, \dots, x_n] = \langle 1 \rangle$.*

This theorem gives a solution to the consistency problem, which asks us to decide whether there exists a solution to $f_1 = \cdots = f_s = 0$, in light of Proposition 3.11.2.

Let's get back to coloring. If you're interested in related problems, you may be interested in this paper of Lovász, and this paper of De Loera et al, among others.

We will apply Theorem 5.13.1 to this problem. We'll be considering ideals of $k[x_1, \dots, x_m]$; recall that $m = |V(G)|$. The idea is to define "vertex polynomials" $f_v$ for each $i \in V(G)$ and "edge polynomials" $e_{ij}$ for each $e \in E(G)$. The $f_v$ will encode the colors given to $v$, while the $f_e$ ascertain that adjacent vertices do not get the same color. For clarity of notation we'll denote, for $i, j \in V(G)$,

$$v_i(x_1, \dots, x_m) \stackrel{\text{def}}{=} f_i(x_1, \dots, x_m) \quad \text{and} \quad e_{ij}(x_1, \dots, x_m) \stackrel{\text{def}}{=} f_e(x_1, \dots, x_m), \text{ where } e = \{i, j\} \in E(G).$$

Pick $n$ distinct elements $\alpha_1, \dots, \alpha_n$ of $k$. For concreteness, we'll pick $k = \mathbb{C} = \{a + bi \colon a, b \in \mathbb{R}\}$ and $\alpha_j = e^{2\pi i j/n}$ the $n$th roots of unity. [Sorry, I know I used $i$ to mean an element of $V(G)$, but we need it here to denote $i = \sqrt{-1}$...] Let us define

$$v_j = \prod_{k=1}^{n} (x_j - \alpha_k) = x_j^n - 1$$

and

$$e_{jk} = \frac{v_j - v_k}{x_j - x_k} = \sum_{\ell=0}^{n-1} x_j^\ell x_k^{n-1-\ell},$$

where the second equality in both equations above is our concrete special case but in general any $\alpha_k$ will do and $e_{jk}$ will be defined from $v_j$ and $v_k$ as displayed above. The $n$ colors basically correspond to the $n$ many $\alpha_j$.

Observe that, for every $i \in [m]$, we have

$$v_i(\mathbf{x}) = 0 \quad \text{if and only if} \quad x_i = \alpha_j \text{ for some } j \in [n].$$

Furthermore (this needs a little proof!), observe also that if $\mathbf{x}$ satisfies $v_j(\mathbf{x}) = 0$ for every $j \in [m]$, then

$$e_{jk}(\mathbf{x}) = 0 \quad \text{if and only if} \quad x_j = \alpha_{j'} \text{ and } x_k = \alpha_{k'} \text{ for some } j' \neq k' \in [n].$$

Thus we have proven

**Theorem 5.13.2.** *The graph $G$ is $m$-colorable if and only if $\{v_i, e_{ij} : i \in V(G), \{i,j\} \in E(G)\}$ have a common root $\mathbf{x}$. The coloring is given in the following way. If $\mathbf{x} = (\alpha_{i_1}, \dots, \alpha_{i_m})$ is a solution, then $f : V(G) \to [n]$ given by $v \mapsto i_v$ is a coloring.*

Okay, let's recall also Hilbert's Nullstellensatz:

**Theorem 5.13.3** (Hilbert Nullstellensatz, cf. Theorem 3.11.3)**.** *Let $k$ be algebraically closed. Then if $f_1, \dots, f_s \in k[x_1, \dots, x_n]$, then*

$$f \in \mathbf{I}(\mathbf{V}(f_1, \dots, f_s)) \quad \text{if and only if} \quad f^m \in \langle f_1, \dots, f_s \rangle$$

*for some $m \in \mathbb{Z}_{\geq 1}$.*

**Definition 5.13.4.** An ideal $I$ is radical if $f^m \in I$ for $m \geq 1$ implies $f \in I$. △

**Corollary 5.13.5.** *For any variety $V$, the ideal $\mathbf{I}(V)$ is radical.*

**Exercise:** The ideal $I$ is radical if and only if $I = \sqrt{I}$.

**Lemma 5.13.6.** *If $I$ is an ideal, then $\sqrt{I}$ is also an ideal. Moreover, $\sqrt{I}$ is a radical ideal.*

*Proof.* Observe, of course, that $0 \in \sqrt{I}$. We need to check that if $f, g \in \sqrt{I}$ then $f + g \in \sqrt{I}$, and that if $f \in \sqrt{I}$ and $h \in k[x_1, \dots, x_n]$, then $hf \in \sqrt{I}$.

Let's begin with our first task. If $f, g \in \sqrt{I}$, then there exist $m, \ell \geq 1$ so that $f^m, g^\ell \in I$. We want to show that $f + g \in \sqrt{I}$, that is, that there exists an integer $M$ so that $(f+g)^M \in I$. Our claim is that $M = m + \ell - 1$ works: observe that

$$(f+g)^{m+\ell-1} = \sum_{i=0}^{m+\ell-1} \binom{m+\ell-1}{i} f^i g^{m+\ell-1-i},$$

and note that every term in the sum either has $i \geq m$ or $m + \ell - 1 - i \geq \ell$: this is because if instead $i \leq m-1$ and $m + \ell - 1 - i \leq \ell - 1$, then adding these two inequalities gives $m + \ell - 1 \leq m + \ell - 2$, which is a contradiction. Since every term in the sum either has $i \geq m$ or $m + \ell - 1 - i \geq \ell$, each term is divisible by either $f^m$ or $g^\ell$; in either case, each term is in $I$, hence $(f+g)^{m+\ell-1} \in I$ as well. We have shown that $f + g \in \sqrt{I}$.

The second task is less painful: let us take $f \in \sqrt{I}$ and $h \in k[x_1, \dots, x_n]$. Since $f \in \sqrt{I}$ there exists $m \geq 1$ so that $f^m \in I$; but now $h^m f^m = (hf)^m \in I$ since $I$ is an ideal, so we have shown $hf \in \sqrt{I}$.

We'll leave the "moreover" part as an exercise. [If you've done the previous exercise, which asks you to show $I$ is radical if and only if $I = \sqrt{I}$, then this "moreover" part of the theorem equivalently asks you to show that $\sqrt{I} = \sqrt{\sqrt{I}}$.] □

The following theorem has some very nice consequences, some of which we'll see in the near future. (As a teaser: we'll see exactly in what way $\mathbf{I}$ and $\mathbf{V}$ are inverses to each other.)

**Theorem 5.13.7** (Strong Nullstellensatz)**.** *Let $k$ be an algebraically closed field. If $I$ is an ideal of $k[x_1, \dots, x_n]$, then*

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

*Proof.* Let us take generators $\langle f_1, \dots, f_s \rangle = I$. By the Hilbert Nullstellensatz (Theorem 5.13.3) and the definition of the radical of an ideal, we have

$$f \in \mathbf{I}(\mathbf{V}(I)) = \mathbf{I}(\mathbf{V}(f_1, \dots, f_s)) \iff f^m \in \langle f_1, \dots, f_s \rangle \iff f \in \sqrt{\langle f_1, \dots, f_s \rangle} = \sqrt{I}.$$

(The first equality is Proposition 2.7.6.) □

## 5.14 Oct 22, 2019

Last time we ended with the Strong Nullstellensatz:

**Theorem 5.14.1** (Strong Nullstellensatz; cf. Theorem 5.13.7)**.** *Let $k$ be an algebraically closed field. If $I \subseteq k[x_1, \ldots, x_n]$ is an ideal, then*

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

As hinted before, we have some very nice consequences:

**Theorem 5.14.2** (The Ideal-Variety Correspondence)**.** *Let $k$ be an arbitrary field.*

1. *The maps*

$$\{\text{affine varieties}\} \xrightarrow{\mathbf{I}} \{\text{ideals}\}$$
$$\{\text{ideals}\} \xrightarrow{\mathbf{V}} \{\text{affine varieties}\}$$

   *are inclusion reversing:*

$$V_1 \subseteq V_2 \implies \mathbf{I}(V_1) \supseteq \mathbf{I}(V_2)$$
$$I_1 \subseteq I_2 \implies \mathbf{V}(I_1) \supseteq \mathbf{V}(I_2).$$

2. *For any affine variety $V$, $\mathbf{V}(\mathbf{I}(V)) = V$. For any ideal $I$, $\mathbf{V}(\sqrt{I}) = \mathbf{V}(I)$.*
3. *For $k$ algebraically closed, the image of $\mathbf{I}$ is the subset $\{\text{radical ideals}\} \subseteq \{\text{ideals}\}$. Then*

$$\{\text{affine varieties}\} \xrightarrow{\mathbf{I}} \{\text{radical ideals}\}$$
$$\{\text{radical ideals}\} \xrightarrow{\mathbf{V}} \{\text{affine varieties}\}$$

   *are inclusion reversing bijections, which are inverses of each other.*

*Proof.* We've seen part (1) in various homeworks, lectures, and exams.

Let's prove part (2). Let $V = \mathbf{V}(f_1, \ldots, f_s) \subseteq k^n$. To prove $\mathbf{V}(\mathbf{I}(V)) \supseteq V$, observe that every $f \in \mathbf{I}(V)$ vanishes on $V$. To prove $\mathbf{V}(\mathbf{I}(V)) \subseteq V$, note that $f_1, \ldots, f_s \in \mathbf{I}(V)$ implies $\langle f_1, \ldots, f_s \rangle \subseteq \mathbf{I}(V)$. Part (1) of the theorem implies

$$V = \mathbf{V}(\langle f_1, \ldots, f_s \rangle) \supseteq \mathbf{V}(\mathbf{I}(V)).$$

That $\mathbf{V}(\sqrt{I}) = \mathbf{V}(I)$ will be left as an exercise.

For part (3), we know that $\mathbf{I}(V)$ is radical for any variety $V$. We also know, from (2), that $\mathbf{V}(\mathbf{I}(V)) = V$, so we just need $\mathbf{I}(\mathbf{V}(I)) = I$ for radical ideals $I$. This follows from the Strong Nullstellensatz (Theorem 5.14.1), which says that $\mathbf{I}(\mathbf{V}(I)) = \sqrt{I}$. On the HW, we'll prove that when $I$ is a radical ideal, $I = \sqrt{I}$. Note that we use algebraic closedness of $k$ because we appealed to the Nullstellensatz. □

The ideal-variety correspondence (Theorem 5.14.2) leads to natural questions:

1. Radical Generation: Is there an algorithm, which when given $I$, produces a basis for $\sqrt{I}$?
2. Is there an algorithm to decide if $I$ is radical?
3. Radical membership: Given $f \in k[x_1, \ldots, x_n]$ and $I = \langle f_1, \ldots, f_s \rangle$, is there an algorithm to determine if $f \in \sqrt{I}$?

**Proposition 5.14.3** (Radical membership)**.** *Let $k$ be an algebraically closed field, and let $I = \langle f_1, \ldots, f_s \rangle \subseteq k[x_1, \ldots, x_n]$. Then,*

$$f \in \sqrt{I} \quad \iff \quad 1 \in \widetilde{I} = \langle f_1, \ldots, f_s, 1 - yf \rangle \subseteq k[x_1, \ldots, x_n, y].$$

You may recall $\widetilde{I}$ from the proof of Hilbert Nullstellensatz (see Equation (6)). [If you're interested, this proof is called the Rabinowitz trick. As Qiaochu notes, the "trick" is quite natural in the language of commutative algebra. We use that localization with respect to $f$ is trivial iff $f$ is nilpotent. Incidentally, this also proves the above proposition.]

*Proof.* In the proof of the Hilbert Nullstellensatz (Theorem 5.13.3), we showed that $1 \in \widetilde{I}$ means $f^m \in I$ for some $m \geq 1$, hence $f \in \sqrt{I}$. [post prelim 2 addendum: note that algebraically closedness of $k$ is not necessary – that part of the proof of the Nullstellensatz didn't need it.] Conversely, if $f \in \sqrt{I}$, then $f^m \in I$, and hence when thought of as an element of $k[x_1, \ldots, x_n, y]$ we have $f^m \in \widetilde{I}$. Also, $1 - yf \in \widetilde{I}$, so

$$1 = y^m f^m + (1 - y^m f^m) = \underbrace{y^m f^m}_{\in \widetilde{I}} + \underbrace{(1 - yf)}_{\in \widetilde{I}}(1 + yf + \cdots + y^{m-1} f^{m-1}).$$

Hence $1 \in \widetilde{I}$. □

Proposition 5.14.3 gives a solution to the radical membership problem: given $f \in k[x_1, \ldots, x_n]$ and $\langle f_1, \ldots, f_s \rangle = I$, we can compute the reduced Gröbner basis of $\langle f_1, \ldots, f_s, 1 - yf \rangle \subseteq k[x_1, \ldots, x_n, y]$ with respect to any monomial ordering. This reduced Gröbner basis is $\{1\}$ if and only if $f \in \sqrt{I}$.

Let us also compute the radical of a principal ideal.

**Definition 5.14.4.** A nonconstant polynomial $f \in k[x_1, \ldots, x_n]$ is irreducible if when we write $f = gh$ with $g, h \in k[x_1, \ldots, x_n]$, then either $g$ or $h$ is constant. △

Note that the notion of irreducibility depends on $f$ and $k$, i.e., the polynomial $x^2 + 1$ is irreducible as a polynomial in $\mathbb{R}[x]$, but not irreducible as a polynomial in $\mathbb{C}[x]$. We won't prove the following fact. [Note the resemblance to the fundamental theorem of arithmetic!]

**Fact 5.14.5.** *A nonconstant polynomial $f \in k[x_1, \ldots, x_n]$ can be written as*

$$f = c f_1^{a_1} \ldots f_r^{a_r},$$

*where $f_1, \ldots, f_r$ are distinct irreducible polynomials with leading coefficient 1, $c$ is a nonzero constant in $k$, and $a_1, \ldots, a_r$ are positive integers. This factorization is unique up to permuting the $f_i$.*

**Proposition 5.14.6.** *Let $f \in k[x_1, \ldots, x_n]$, and let $I = \langle f \rangle$. If*

$$f = c f_1^{a_1} \ldots f_r^{a_r}$$

*is the factorization of $f$ into irreducibles, then*

$$\sqrt{\langle f \rangle} = \langle f_1 \ldots f_r \rangle.$$

(Notice that this is the product $f_1 \cdot f_2 \cdot \ldots \cdot f_r$, i.e., there is only one generator.)

*Proof.* As usual we have two inclusions to verify. Let us begin by showing $\sqrt{I} = \sqrt{\langle f \rangle} \supseteq \langle f_1 \ldots f_r \rangle$. Let $N > \max(a_1, \ldots, a_n)$, and note that

$$(f_1 \ldots f_r)^N = \frac{1}{c} f_1^{N - a_1} \ldots f_r^{N - a_r} f \in I,$$

so $f_1 \ldots f_r \in \sqrt{I}$ To show that $\sqrt{I} \subseteq \langle f_1 \ldots f_r \rangle$, we take $g \in \sqrt{I}$, so that $g^m \in I$ for some $m \geq 1$. Then $g^m = hf$ for some $h \in k[x_1, \ldots, x_n]$. In particular, $f_1, \ldots, f_r$ are irreducible factors of $g^m$ (there may be more coming from $h$, but never mind those). Note that $g$ has a factorization into irreducibles, by Fact 5.14.5. Furthermore, $g^m$ also has a factorization into irreducibles, by Fact 5.14.5. Since a factorization for $g^m$ is obtained by taking the factorization of $g$ and raising it to the $m$, the uniqueness of Fact 5.14.5 says that this is *the* unique factorization of $g^m$. We said earlier that $f_1, \ldots, f_r$ are irreducible factors of $g^m$; it follows that $f_1, \ldots, f_r$ are also irreducible factors of $g$. Hence $g \in \langle f_1 \ldots f_s \rangle$. □

**Definition 5.14.7.** Let $f \in k[x_1, \ldots, x_n]$. A reduction of $f$ is a polynomial $f_{\text{red}}$ such that

$$\langle f_{\text{red}} \rangle = \sqrt{\langle f \rangle}.$$ △

By Proposition 5.14.6, note that $f_{\text{red}}$ is unique up to multiplication by a scalar.

**Definition 5.14.8** (cf. Definition 1.3.7)**.** Let $f, g \in k[x_1, \ldots, x_n]$. A greatest common divisor of $f$ and $g$, denoted $\gcd(f, g)$, is a polynomial $h \in k[x_1, \ldots, x_n]$ such that $h$ divides $f$ and $g$ so that whenever $p \in k[x_1, \ldots, x_n]$ also divides both $f$ and $g$, then $p$ divides $h$. △

**Exercise:** The polynomial $\gcd(f, g)$ exists and is unique up to multiplication by a nonzero constant.

**Exercise-Proposition 5.14.9.** *Let $I = \langle f \rangle \subseteq \mathbb{C}[x_1, \ldots, x_n]$. Then*

$$\sqrt{I} = \langle f_{\text{red}} \rangle, \quad \text{and} \quad f_{\text{red}} = \frac{f}{\gcd(f, \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n})}.$$

Note that Exercise-Proposition 5.14.9 is not just a naive generalization of Euclidean Algorithm (which worked for us in the univariate case), since it might "get stuck", e.g. as it would do with $xy$ and $xz$.

We're going to talk about operations on ideals and see what they do in geometry. [Let me remark that there are very natural algebraic operations that are mysterious/deep in geometry, and many natural geometric operations that are mysterious/deep in algebra.]

**Definition 5.14.10.** Let $I$ and $J$ be ideals in $k[x_1, \ldots, x_n]$. Then their sum is

$$I + J = \{f + g \colon f \in I, g \in J\}.$$

Their product is

$$IJ = \left\{ \sum_{i=1}^{r} f_i g_i \colon f_i \in I, g_i \in J \right\}.$$

Note that $r \geq 1$ is arbitrary but finite. (Finiteness is crucial for various theorems to hold.) Their intersection is

$$I \cap J = \{f \colon f \in I, f \in J\}.$$

△

**Proposition 5.14.11.** *Let $I$ and $J$ be ideals in $k[x_1, \ldots, x_n]$, then $I + J$, $IJ$, and $I \cap J$ are all ideals. Furthermore, $I + J$ is the smallest ideal containing both $I$ and $J$, where smallest is in the sense of containment. If $I = \langle f_1, \ldots, f_s \rangle$ and $J = \langle g_1, \ldots, g_t \rangle$, then*

$$I + J = \langle f_1, \ldots, f_s, g_1, \ldots, g_t \rangle, \quad \text{and} \quad IJ = \langle f_i g_j \colon i \in [s], j \in [t] \rangle.$$

Note that this implies every ideal is a finite sum of principal ideals, since $\langle f_1, \ldots, f_s \rangle = \langle f_1 \rangle + \cdots + \langle f_s \rangle$. Finding a basis for $I \cap J$ is a bit harder and will need more work.

**Theorem 5.14.12.** *We have*

1. $\mathbf{V}(I + J) = \mathbf{V}(I) \cap \mathbf{V}(J)$,
2. $\mathbf{V}(IJ) = \mathbf{V}(I) \cup \mathbf{V}(J)$, *and*
3. $\mathbf{V}(I \cap J) = \mathbf{V}(I) \cup \mathbf{V}(J)$.

*Proof.* The first two parts follow from combining Proposition 5.14.11 and (the proof of) Lemma 1.2.2.

The third part, as usual, follows from verifying the two inclusions: we have $\mathbf{V}(I \cap J) \supseteq \mathbf{V}(I) \cup \mathbf{V}(J)$ because if $\mathbf{a} \in \mathbf{V}(I)$, then $f(\mathbf{a}) = 0$ for all $f \in I$, in particular for all $f \in I \cap J$, hence $\mathbf{a} \in \mathbf{V}(I \cap J)$, and for exactly the same reason if $\mathbf{a} \in \mathbf{V}(J)$ then $\mathbf{a} \in \mathbf{V}(I \cap J)$ too. We also have $\mathbf{V}(I \cap J) \subseteq \mathbf{V}(I) \cup \mathbf{V}(J)$ because if $\mathbf{a} \in \mathbf{V}(I \cap J)$ but $\mathbf{a} \notin \mathbf{V}(I)$ and $\mathbf{a} \notin \mathbf{V}(J)$, then we could find $f \in I$ and $g \in J$ so that $f(\mathbf{a}) \neq 0$ and $g(\mathbf{a}) \neq 0$; then $fg(\mathbf{a}) \neq 0$ too even though $fg \in I \cap J$. □

In light of Theorem 5.14.2, we like radical ideals. Note that the product of two radical ideals need not be radical: take, for example, $I = J = \langle x \rangle$ (then $IJ = \langle x^2 \rangle$). However, the intersection of two radicals *is* radical, and that's why we want to do the hard work required to find a basis for $I \cap J$. Specifically,

**Proposition 5.14.13.** *Let $I$ and $J$ be ideals. We have*

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

*In particular, the intersection of two radical ideals is still radical.*

*Proof.* If $f \in \sqrt{I \cap J}$, then $f^m \in I \cap J$, that is, $f^m \in I$ and $f^m \in J$. Then $f \in \sqrt{I}$ and $f \in \sqrt{J}$.

If $f \in \sqrt{I} \cap \sqrt{J}$, then there exist $m_1, m_2 \geq 1$ so that $f^{m_1} \in I$ and $f^{m_2} \in J$. Then $f^{m_1+m_2} \in I \cap J$, that is, $f \in \sqrt{I \cap J}$. □

# 6 Some Elimination Theory

## 6.15 Oct 24, 2019

Last time we talked about operations of ideals: if $I, J \subseteq k[x_1, \ldots, x_n]$ are ideals, then we defined the ideals $I + J$, $IJ$, and $I \cap J$. Given bases for $I$ and $J$, we described bases for $I + J$ and $IJ$. We also showed that given radical ideals $I$ and $J$, the ideal $I \cap J$ is also radical. Thus it is of interest to find a basis for $I \cap J$.

As a warmup, suppose $I = \langle f \rangle \subseteq \mathbb{Q}[x, y]$ and $J = \langle g \rangle \subseteq \mathbb{Q}[x, y]$ where

$$f = (x + y)^4(x^2 + y)^2(x - 5y) \quad \text{and} \quad g = (x + y)(x^2 + y)^3(x + 3y).$$

Then it's not too hard to show that $I \cap J = \langle h \rangle$, where

$$h = (x + y)^4(x^2 + y)^3(x - 5y)(x + 3y).$$

One might reasonably want to call $h$ a least common multiple of $f$ and $g$. We'll make this formal later.

Our goal today will be to describe an algorithm that will compute bases for $I \cap J$. A crucial stepping stone will be the following theorem:

**Theorem 6.15.1.** *Let $I$ and $J$ be ideals of $k[x_1, \ldots, x_n] = k[\mathbf{x}]$. Then*

$$I \cap J = \underbrace{(tI + (1 - t)J)}_{\text{ideal of } k[\mathbf{x}, t]} \cap k[\mathbf{x}].$$

*(See Definition 6.15.2.)*

**Definition 6.15.2.** Let $I$ be an ideal in $k[x_1, \ldots, x_n]$, and let $f(t) \in k[t]$. Then

$$f(t)I \stackrel{\text{def}}{=} \langle f(t)h \colon h \in I \rangle.$$

It is an ideal of $k[x_1, \ldots, x_n, t]$. △

Note that the set $\{f(t)h \colon h \in I\}$ is not an ideal of $k[x_1, \ldots, x_n, t]$, one really has to take the ideal generated by it.

We'll need some elimination theory today (this is chapter 3 in the book); after covering some basic results we'll be able to compute least common multiples and greatest common divisors of multivariate polynomials.

In elimination theory we want to solve polynomial equations $f_1 = \cdots = f_r = 0$ in $k[x_1, \ldots, x_n]$. The idea is as follows: there is an elimination step, where we find polynomial consequences of $f_1 = \cdots = f_r = 0$ which involve a proper subset of the initial variables, and there is the extension step, where our solutions of the smaller system (called *partial* solutions of $f_1 = \cdots = f_r = 0$) are extended to complete solutions of the original solution. This is not always possible.

**Example 6.15.3.** We'll see this example later, but suppose $f_1 = xy - 1$ and $f_2 = xz - 1$ in $k[x, y, z]$. A polynomial consequence of $f_1 = f_2 = 0$ is $zf_1 - yf_2 = z - y = 0$, which involves only $y$ and $z$. So this is an elimination step.

An extension step consists of taking solutions of the smaller system (in this case $z - y = 0$) and extending it to solutions of the original solution. The solutions of the smaller system $z - y = 0$ is precisely $\{(a, a) \colon a \in k\}$. These can be extended to a solution $(x, a, a)$ to $f_1 = f_2 = 0$ precisely when $a \neq 0$; in this case it extends to $(\frac{1}{a}, a, a)$. △

In the language of ideals, the following notion will prove to be important:

**Definition 6.15.4.** Given $I = \langle f_1, \ldots, f_s \rangle \subseteq k[x_1, \ldots, x_n]$. The $\ell$-th elimination ideal $I_\ell$ is the ideal in $k[x_{\ell+1}, \ldots, x_n]$ defined by

$$I_\ell \stackrel{\text{def}}{=} I \cap k[x_{\ell+1}, \ldots, x_n]. \qquad \triangle$$

Thus $I_\ell$ consists of polynomial consequences of $f_1 = \cdots = f_s = 0$ involving only $x_{\ell+1}, \ldots, x_n$.

Bases of elimination ideals behave particularly nicely.

**Theorem 6.15.5** (Elimination Theorem)**.** *Let $I \subseteq k[x_1 \ldots, x_n]$ be an ideal and let $G$ be a Gröbner basis of $I$ with respect to lex order, with $x_1 > \cdots > x_n$. Define, for every $0 \leq \ell \leq n$, the set*

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

*Then $G_\ell$ is a Gröbner basis of $I_\ell$.*

*Proof.* Fix an $\ell$. Note that $G_\ell \subseteq I_\ell$, hence $\langle \mathrm{LT}(G_\ell) \rangle \subseteq \langle \mathrm{LT}(I_\ell) \rangle$. We should check that $\langle \mathrm{LT}(I_\ell) \rangle \subseteq \langle \mathrm{LT}(G_\ell) \rangle$.

Indeed, let us take $f \in I_\ell$, and let us show that $\mathrm{LT}(f)$ is divisible by $\mathrm{LT}(g)$ for some $g \in G_\ell$; this suffices because Lemma 2.5.9 would imply the containment. To show this, observe that $f \in I$ so $\mathrm{LT}(f)$ is divisible by $\mathrm{LT}(g)$ for some $g \in G$. Furthermore, since $\mathrm{LT}(f)$ only involves the variables $x_{\ell+1}, \ldots, x_n$, we see that $\mathrm{LT}(g)$ only involves the variables $x_{\ell+1}, \ldots, x_n$ too. Furthermore, every term of $g$ is at most (in lex order) $\mathrm{LT}(g)$, so it can only use the variables $x_{\ell+1}, \ldots, x_n$. Thus $g \in k[x_{\ell+1}, \ldots, x_n]$. Since $g \in G$ as well, we obtain $g \in G_\ell$. $\square$

*Proof of Theorem 6.15.1.* Let us note that $tJ$, $(1-t)J$, and hence $tJ + (1-t)J$, are ideals of $k[x_1 \ldots, x_n, t]$. We have an equality of sets, so we better prove the two inclusions.

To see that $I \cap J \subseteq (tI + (1-t)J) \cap k[x_1, \ldots, x_n]$, let us take $f \in I \cap J$. Then $f \in I$ means $tf \in tI$, and $f \in J$ means $(1-t)f \in (1-t)J$. Hence

$$f = \underbrace{tf}_{\in tI} + \underbrace{(1-t)f}_{\in (1-t)J} \in tI + (1-t)J.$$

To see that $I \cap J \supseteq (tI + (1-t)J) \cap k[x_1, \ldots, x_n]$, let us take $f(\mathbf{x}) \in tI + (1-t)J \cap k[x_1, \ldots, x_n]$. By definition, we have

$$f = \underbrace{g(\mathbf{x}, t)}_{\in tI} + \underbrace{h(\mathbf{x}, t)}_{\in (1-t)J}. \tag{7}$$

Note that every element of $tI$ is divisible by $t$ and every element of $(1-t)J$ is divisible by $(1-t)$. In particular, $g(\mathbf{x}, 0) = 0$ and $h(\mathbf{x}, 1) = 0$; substituting $t = 1$ into the equality (7) gives

$$f(\mathbf{x}) = g(\mathbf{x}, 1) \in (tI)_{t=1},$$

where similar to the proof of the Weak Nullstellensatz (Theorem 5.13.1) the ideal $(tI)_{t=1}$ consists of polynomials of the form $\{g(\mathbf{x}, 1) \colon g \in tI\}$. We'll see (in Lemma 6.15.6) that $(tI)_{t=1} = I$, and this proves $f \in I$. Similarly, substituting $t = 0$ into the equality (7) gives

$$f(\mathbf{x}) = h(\mathbf{x}, 0) \in ((1-t)J)_{t=0},$$

where again $((1-t)J)_{t=0}$ consists of polynomials of the form $\{h(\mathbf{x}, 1) \colon h \in (1-t)J\}$. We'll also see (in Lemma 6.15.6) that $((1-t)J)_{t=0} = J$, and this proves $f \in J$. This shows $f \in I \cap J$, as desired. $\square$

**Lemma 6.15.6.** *Let $I = \langle p_1(\mathbf{x}), \ldots, p_r(\mathbf{x}) \rangle \subseteq k[\mathbf{x}]$ be an ideal. Then*

$$f(t)I = \langle f(t)p_1(\mathbf{x}), \ldots, f(t)p_r(\mathbf{x}) \rangle.$$

*Proof.* Chase definitions. [There are two containments...] □

Theorem 6.15.1, Theorem 6.15.5, and Lemma 6.15.6 give an algorithm to compute a basis of the intersection of two ideals. If $I = \langle f_1, \dots, f_r \rangle$ and $J = \langle g_1, \dots, g_s \rangle$, then $tI + (1-t)J = \langle tf_1, \dots, tf_r, (1-t)g_1, \dots, (1-t)g_s \rangle$ by Proposition 5.14.11. We compute a Gröbner basis for $tI + (1-t)J$ with respect to lex order with $t > x_1 > \cdots > x_n$, and take elements of $G$ that do not contain $t$. Theorem 6.15.5 asserts that this forms a Gröbner basis for $I \cap J$.

**Definition 6.15.7.** Let $f, g \in k[x_1, \dots, x_n]$. A least common multiple for $f$ and $g$, denoted $\operatorname{lcm}(f, g)$, is a polynomial $h$ so that $f$ and $g$ both divide $h$, and if $f$ and $g$ both divide $p$, then $h$ divides $p$ too. △

**Exercise-Proposition 6.15.8.** *We have:*

1. *The intersection of two principal ideals is again principal, and*
2. *If $I = \langle f \rangle$, $J = \langle g \rangle$, and $I \cap J = \langle h \rangle$, then $h$ is an lcm of $f$ and $g$. (Hence, the lcm exists and is unique up to multiplcation by a nonzero constant.)*

**Exercise-Lemma 6.15.9.** *For any two polynomials $f, g \in k[x_1, \dots, x_n]$, we have*

$$fg = \gcd(f, g) \cdot \operatorname{lcm}(f, g).$$

Since we have an algorithm to compute the lcm, we obtain an algorithm to compute the gcd.

Let's go back to the problem of solving $f_1 = \cdots = f_s = 0$. Although we've seen this in Example 6.15.3, let's recast what we did there in today's new language.

**Example 6.15.10** (cf. Example 6.15.3)**.** Let us consider $xy - 1 = 0$ and $xz - 1 = 0$. Then $I_1 = \langle y - z \rangle \subseteq k[y, z]$. In other words, $\mathbf{V}(I_1)$ consists of the $y = z$ line in $yz$-space. Not all solutions in $\mathbf{V}(I_1)$ extend to a solution in $\mathbf{V}(I) \subseteq k^3$, where $k^3$ is now $xyz$-space. As it turns out, $\mathbf{V}(I) \cap \{yz\text{-space}\}$ is $\mathbf{V}(I_1) \setminus \{(0, 0)\}$. △

## 6.16 Oct 29, 2019

We were thinking about Example 6.15.10, which I'll reproduce below.

**Example 6.16.1.** Consider $xy-1=0$ and $xz=1=0$. We computed $\mathbf{V}(I_1)=\{(a,a)\colon a\in k\}\subseteq\{yz\text{-space}\}=k^2\subseteq k^3$; that is, $I_1\subseteq k[y,z]$. The point in $\mathbf{V}(I_1)$ which extend to a point in $\mathbf{V}(I)$ is the set $\{(a,a)\colon a\in k, a\neq 0\}$. △

We'll discuss today the question of extending a partial solution. A first question is: does the set of "extendable" points always a variety? The answer is a resounding no: in Example 6.16.1 we already saw that the extendable points form a "punctured line". It's not too hard to show that this is not a variety, for example, we essentially did this on HW 2 (Exercise 1.2.8 in the book).

**Theorem 6.16.2** (Extension Theorem)**.** *Let $I=\langle f_1,\dots,f_s\rangle\subseteq\mathbb{C}[x_1,\dots,x_n]$ and let $I_1$ be the first elimination of $I$. Write, for each $i\in[s]$,*

$$f_i=c_i(x_2,\dots,x_n)x_1^{N_i}+\{\text{terms with } x_1 \text{ having degree} < N_i\},$$

*where $N_i\geq 0$ and $c_i\in k[x_2,\dots,x_n]$ is nonzero. Suppose we have a solution $(a_2,\dots,a_n)\in\mathbf{V}(I_1)$. If $(a_2,\dots,a_n)\notin\mathbf{V}(c_1,\dots,c_s)$, then there exist $a\in\mathbb{C}$ such that $(a_1,\dots,a_n)\in\mathbf{V}(I)$.*

In Example 6.16.1, we'd have $f_1=xy-1=xc_1(y,z)-1$, and $f_2=xz-1=xc_2(y,z)-1$, where $c_1(y,z)=y$ and $c_2(y,z)=z$. Then the Extension theorem asks us to consider $\mathbf{V}(c_1,c_2)=\mathbf{V}(y,z)=\{(0,0)\}$. Thus it guarantees that everything in $\{(a,a)\colon a\in k,a\neq 0\}\subseteq\mathbf{V}(I_1)$ extends to a solution, and it says nothing about $(0,0)\in\mathbf{V}(I_1)$ (it turns out $(0,0)$ doesn't extend, but the Extension Theorem doesn't know this).

**Example 6.16.3** (Avery's Example)**.** Let $f_1=-2z+z^2$, $f_2=2y-z$, and $f_3=-z+2xz$. We obtain $c_1(y,z)=-2z+z^2$, $c_2=2y-z$, and $c_3=2z$. According to Maple (to Avery), this is a Gröbner basis with respect to the $x>y>z$ lex order. Then $I_1=\langle f_1,f_2\rangle$ and $(0,0)\in\mathbf{V}(I_1)$. Note also that $(0,0)\in\mathbf{V}(c_1,c_2,c_3)$, so the Extension Theorem (Theorem 6.16.2) doesn't know anything. In this case, $(0,0,0)\in\mathbf{V}(I)$, so it does extend. △

**Example 6.16.4.** Consider the system of equations

$$\begin{aligned}f_1&=x^2+y^2+z^2-1\\f_2&=xyz-1\end{aligned}$$

A Gröbner basis with respect to $x>y>z$ lex is given by

$$\begin{aligned}g_1&=y^4z^2+y^2z^4-y^2z^2+1\\g_2&=x+y^3z+yz^3-yz.\end{aligned}$$

The elimination ideals are $I_1=I\cap\mathbb{C}[y,z]=\langle g_1\rangle$, and $I_2=I\cap\mathbb{C}[z]=\{0\}$. Now $I_2=\{0\}$ means $\mathbf{V}(I_2)=\mathbb{C}$. So $c\in\mathbb{C}$ is a partial solution.

We'll show in HW that $I_2$ is the first elimination ideal of $I_1$, that is, $I_2=(I_1)_1$.

When does $(c)\in\mathbf{V}(I_2)=\{z\text{-space}\}=\mathbb{C}^1$ extends to a point $(a,b,c)\in\mathbf{V}(I)\subseteq\{xyz\text{-space}\}=\mathbb{C}^3$? Well, let us apply the Extension Theorem (Theorem 6.16.2) from $I_2=(I_1)_1$ to $I_1$. The coefficient of $y^4$ in $g_1$ is $z^2$ (that is, $c_1(z)=z^2$). Thus the Extension Theorem (Theorem 6.16.2) guarantees the extension of $(c)$ to $(b,c)$ if $c\neq 0$.

The next step is going fro $I_1$ to $I$. Now, $c_1(y,z)=1$ and $c_2(y,z)=yz$ Since now $\mathbf{V}(c_1,c_2)=\emptyset$, the Extension Theorem (Theorem 6.16.2) guarantees every solution $(b,c)$ extends to $(a,b,c)$. △

Let us highlight that $c_1(y,z)=1$ automatically allowed the Extension Theorem (Theorem 6.16.2) to always extend. More precisely:

**Corollary 6.16.5.** *Let $I = \langle f_1, \dots, f_s \rangle \subseteq \mathbb{C}[x_1, \dots, x_n]$, and assume some $f_i$ is of the form*

$$f_i = c_i x_1^{N_i} + \{\text{terms with } x_1 \text{ having degree} < N_i\},$$

*where $c_i \in \mathbb{C}$ is nonzero and $N_i > 0$. If $I_1$ is the first elimination ideal of $I$ and $(a_2, \dots, a_n) \in \mathbf{V}(I_1)$, then there is $a_1 \in \mathbb{C}$ such that $(a_1, \dots, a_n) \in \mathbf{V}(I)$.*

We come to a crucial algebra-geometry relationship [See my remark just above Definition 5.14.10]. Essentially, elimination of variables corresponds to projections of varieties into lower dimensional subspaces. Let's make this precise.

To eliminate the first $\ell$ variables, let's consider the projection map

$$\begin{aligned} \pi_\ell \colon \mathbb{C}^n &\to \mathbb{C}^{n-\ell} \\ (a_1, \dots, a_n) &\mapsto (a_{\ell+1}, \dots, a_n). \end{aligned}$$

Thus for every $V \subseteq \mathbb{C}^n$, we have $\pi_\ell(V) \subseteq \mathbb{C}^{n-\ell}$.

**Lemma 6.16.6.** *Let $I_\ell$ be the $\ell$th elimination ideal for $I = \langle f_1, \dots, f_s \rangle \subseteq \mathbb{C}[x_1, \dots, x_n]$. As subsets of $\mathbb{C}^{n-\ell}$, we have*

$$\pi_\ell(\mathbf{V}(I)) \subseteq \mathbf{V}(I_\ell).$$

*Proof.* Let $f \in I_\ell \subseteq I$. If $(a_1, \dots, a_n) = \mathbf{a} \in \mathbf{V}(I)$, then $f(\mathbf{a}) = 0$ since $f \in I$. But $f$ only uses the variables $x_{\ell+1}, \dots, x_n$, so it makes sense to just plug in $\pi_\ell(\mathbf{a}) = (a_{\ell+1}, \dots, a_n)$ into $f$, now thought of as an element in $\mathbb{C}[x_{\ell+1}, \dots, x_n]$. In particular, $f(\pi_\ell(\mathbf{a})) = 0$, that is, $f \in \mathbf{V}(I_\ell)$. Thus $f$ vanishes on $\pi_\ell(\mathbf{V}(I))$. □

Note that $\pi_\ell(\mathbf{V}(I))$ is the set of extendable partial solutions to $f_1 = \cdots = f_s = 0$ (here $I = \langle f_1, \dots, f_s \rangle$). We've seen, e.g. in Example 6.16.1 that equality in Lemma 6.16.6 does not always hold, and that $\pi_\ell(\mathbf{V}(I))$ does not have to be a variety.

**Theorem 6.16.7** (Closure Theorem)**.** *Let $V = \mathbf{V}(f_1, \dots, f_s) \subseteq \mathbb{C}^n$, and let $I_\ell$ be the $\ell$th elimination ideal of $\langle f_1, \dots, f_s \rangle$. Then $\mathbf{V}(I_\ell)$ is the smallest affine variety containing $\pi_\ell(V)$.*

**Corollary 6.16.8.** *Let $I = \langle f_1, \dots, f_s \rangle \subseteq \mathbb{C}[x_1, \dots, x_n]$, and assume some $f_i$ is of the form*

$$f_i = c_i x_1^{N_i} + \{\text{terms with } x_1 \text{ having degree} < N_i\},$$

*where $c_i \in \mathbb{C}$ is nonzero and $N_i > 0$. Then $\pi_1(V) = \mathbf{V}(I_1)$.*

*Proof.* Combine Corollary 6.16.5 and Theorem 6.16.7. □

For any set $S \subseteq k^n$ (think $S = \pi_\ell(V)$), we define

$$\mathbf{I}(S) \stackrel{\text{def}}{=} \{f \in k[\mathbf{x}] \colon f(\mathbf{a}) = 0 \text{ for all } \mathbf{a} \in S\}.$$

We leave as an exercise to show that $\mathbf{I}(S)$ is a radical ideal. By Theorem 5.14.2, we see that $\mathbf{V}(\mathbf{I}(S))$ is an affine variety.

**Proposition 6.16.9.** *The smallest variety containing $S \subseteq k^n$ is $\mathbf{V}(\mathbf{I}(S))$.*

*Proof.* We want to show that if $W \subseteq k^n$ is a variety satisfying $S \subseteq W$, then $\mathbf{V}(\mathbf{I}(S)) \subseteq W$ as well. To see this, note that the order-reverisng-ness of $\mathbf{I}()$ says that $S \subseteq W$ implies $\mathbf{I}(W) \subseteq \mathbf{I}(S)$; furthermore, the order-reversing-ness of $\mathbf{V}$ says that $\mathbf{V}(\mathbf{I}(S)) \subseteq \mathbf{V}(\mathbf{I}(W))$. Since $W$ is already a variety, Theorem 5.14.2 says $\mathbf{V}(\mathbf{I}(W)) = W$. □

**Definition 6.16.10.** The Zariski closure of $S \subseteq k^n$ is the smallest affine variety containing $S$. We denote it by $\overline{S}$; by Proposition 6.16.9, we we have $\overline{S} = \mathbf{V}(\mathbf{I}(S))$. △

**Proposition 6.16.11.** *Let $S, T \subseteq k^n$. We have:*

*(1)* $\mathbf{I}(\overline{S}) = \mathbf{I}(S)$,
*(2)* $S \subseteq T$ *implies* $\overline{S} \subseteq \overline{T}$, *and*
*(3)* $\overline{S \cup T} = \overline{S} \cup \overline{T}$.

*Proof.* To prove part (1), observe first that since $S \subseteq \overline{S}$ we have $\mathbf{I}(S) \supseteq \mathbf{I}(\overline{S})$, so it suffices to prove the other inclusion. Let us take $f \in \mathbf{I}(S)$. Then $S \subseteq \mathbf{V}(f)$; since $\mathbf{V}(f)$ is an affine variety then (by definition of Zariski closure!) we have $\overline{S} \subseteq \mathbf{V}(f)$ as well. Thus $\overline{S} \subseteq \mathbf{V}(f)$, and now $f \in \mathbf{I}(\overline{S})$. This proves the other inclusion.

We'll leave the other two parts as an exercise. □

## 6.17 Oct 31, 2019

Let us recall the Closure Theorem (Theorem 6.16.7) from last lecture; note that the "Closure Theorem" in the textbook, Theorem 3.2.3, has two parts and this is just part 1 of that theorem. [Part 2 in your book is secretly Theorem 4.7.7 in your book, which is a special case of one of my favorite theorems (Chevalley's Theorem), but read that at your own risk.]

**Theorem 6.17.1** ((Part 1 of) Closure Theorem, cf. Theorem 6.16.7)**.** *Assume $k$ is an algebraically closed field. Let $V = \mathbf{V}(f_1, \dots, f_s) \subseteq k^n$. Define the projection map*

$$\begin{aligned} \pi_\ell \colon k^n &\to k^{n-\ell} \\ (a_1, \dots, a_n) &\mapsto (a_{\ell+1}, \dots, a_n). \end{aligned}$$

*If $I_\ell$ is the $\ell$th elimination ideal, i.e. $I_\ell = \langle f_1, \dots, f_s \rangle \cap k[x_{\ell+1}, \dots, x_n]$, then $\overline{\pi_\ell(V)} = \mathbf{V}(I_\ell)$.*

(Recall Definition 6.16.10, where we said $\overline{S}$ was $\mathbf{V}(\mathbf{I}(S))$; we proved in Proposition 6.16.9 that this is the smallest affine variety containing $S$.)

*Proof.* We have to show two sets $\mathbf{V}(\mathbf{I}(\pi_\ell(V)))$ and $\mathbf{V}(I_\ell)$ are equal, so let us prove the two inclusions.

We've shown $\pi_\ell(\mathbf{V}(I)) \subseteq \mathbf{V}(I_\ell)$; this was Lemma 6.16.6. Since $\mathbf{V}(\mathbf{I}(\pi_\ell(V)))$ is *the smallest* variety containing $\pi_\ell(V)$, and $\mathbf{V}(I_\ell)$ is *some* variety containing $\pi_\ell(V)$, we obtain $\mathbf{V}(\mathbf{I}(\pi_\ell(V))) \subseteq \mathbf{V}(I_\ell)$.

Conversely, we shall show $\mathbf{I}(\pi_\ell(V)) \subseteq \sqrt{I_\ell}$. This suffies, since we would get (by inclusion-reversingness of $\mathbf{V}$) that $\mathbf{V}(\mathbf{I}(\pi_\ell(V))) \supseteq \mathbf{V}(\sqrt{I_\ell})$; the ideal-variety correspondence (specifically, part 2 of Theorem 5.14.2) says that $\mathbf{V}(\sqrt{I_\ell}) = \mathbf{V}(I_\ell)$, giving us the desired inclusion.

Let's see why $\mathbf{I}(\pi_\ell(V)) \subseteq \sqrt{I_\ell}$. Take $f \in \mathbf{I}(\pi_\ell(V))$, so $f(a_{\ell+1}, \dots, a_n) = 0$ for all $\mathbf{a} = (a_1, \dots, a_n) \in V$. Although $f \in k[x_{\ell+1}, \dots, x_n]$ we may consider it as an element of $k[x_1, \dots, x_n]$; observe that $f(\mathbf{a}) = 0$ for all $\mathbf{a} \in V$. Hilbert's Nullstellensatz (Theorem 5.13.3) says $f^N \in \langle f_1, \dots, f_s \rangle$ for some $N \geq 1$. But also $f \in k[x_{\ell+1}, \dots, x_n] \subseteq k[x_1, \dots, x_n]$ implies $f^N \in k[x_{\ell+1}, \dots, x_n] \subseteq k[x_1, \dots, x_n]$, too, and we arrive at $f^N \in I_\ell$. In other words, $f \in \sqrt{I_\ell}$. □

Understanding the set $\mathbf{V}(I_\ell) \setminus \pi_\ell(V)$ is what part 2 of the closure theorem does for us, but part 1 is satisfying enough for us. Let's move on.

We've seen before (*ages* before; Example 1.2.1) that $\mathbf{V}(xz, yz) = \mathbf{V}(z) \cup \mathbf{V}(x, y)$ is the union of the $xy$-plane and the $z$-axis. Intuitively it should feel natural to "decompose" $\mathbf{V}(xz, yz)$ into the two subvarieties $\mathbf{V}(z)$ and $\mathbf{V}(x, y)$, but that the $xy$-plane and $z$-axis should probably not be "decomposable" (at least, over an infinite field like $\mathbb{C}$). [thanks to whoever corrected me during lecture! i mistakenly assumed that lines were *always* irreducible, even over finite fields.]

**Definition 6.17.2.** An affine variety $V \subseteq k^n$ is <u>irreducible</u> if whenever $V = V_1 \cup V_2$, with $V_1, V_2$ also affine varieties, then $V_1 = V$ or $V_2 = V$. △

Note that if an irreducible variety $V$ is a *finite* union of affine varieties $V = V_1 \cup \dots \cup V_n$, then necessarily some $V_i = V$, since you can write $V = V_1 \cup$ (union the other $V_i$) and $V_1 = V$ or (union the other $V_i$) $= V$; in the latter case we write $V = V_2 \cup \dots V_n$ and repeat.

**Definition 6.17.3.** An ideal $I \subseteq k[x_1, \dots, x_n]$ is <u>prime</u> if, whenever $f, g \in k[\mathbf{x}]$ and $fg \in I$, we have $f \in I$ or $g \in I$. △

For example, $\langle x, y^2 \rangle$ is not prime. Note that prime ideals are automatically radical.

**Proposition 6.17.4.** *Let $V$ be an affine variety. Then $V$ is irreducible if and only if $\mathbf{I}(V)$ is prime.*

[A principal ideal $\langle f \rangle$ is prime if and only if $f$ is irreducible (Definition 5.14.4). So in this sense, irreducible varieties in $k^n$, or prime ideals in $k[\mathbf{x}]$, "generalize" the concept of irreducible elements $f \in k[\mathbf{x}]$.]

*Proof.* Let's begin with the forward direction; suppose $V$ is irreducible, and consider $f, g \in k[x_1, \ldots, x_n]$ satisfying $fg \in \mathbf{I}(V)$. Let us define

$$V_1 = V \cap \mathbf{V}(f)$$
$$V_2 = V \cap \mathbf{V}(g);$$

we leave as an exercise to show $V = V_1 \cup V_2$. Since $V$ is irreducible, either $V_1 = V$ or $V_2 = V$. Let's assume $V_1 = V$; the case $V_2 = V$ is treated analogously. Since $V = V_1 = V \cap \mathbf{V}(f)$, we obtain $V \subseteq \mathbf{V}(f)$. In other words, $f$ vanishes on all of $V$, that is, $f \in \mathbf{I}(V)$. This shows $\mathbf{I}(V)$ is prime.

Let's do the backward direction now. Suppose $\mathbf{I}(V)$ is prime, and let us write $V = V_1 \cup V_2$ for some affine varieties $V_1, V_2$. If $V_1 = V$ then we'd be done, so let us assume $V_1 \neq V$. We'd need to show that $V_2 = V$; observe that it suffices to show $\mathbf{I}(V_2) = \mathbf{I}(V)$ by the Ideal-Varieties correspondence (Theorem 5.14.2). Furthermore, the inclusion-reversingness of $\mathbf{I}$, along with $V_2 \subseteq V$, give the inclusion $\mathbf{I}(V_2) \supseteq \mathbf{I}(V)$, so we should show $\mathbf{I}(V_2) \subseteq \mathbf{I}(V)$.

Since $V_1 \subsetneq V$, we have $\mathbf{I}(V_1) \supsetneq \mathbf{I}(V)$. Thus we may pick $f \in \mathbf{I}(V_1) \setminus \mathbf{I}(V)$. Take any $g \in \mathbf{I}(V_2)$; our goal is to show that $g \in \mathbf{I}(V)$, since this would give the inclusion $\mathbf{I}(V_2) \subseteq \mathbf{I}(V)$.

Since $V = V_1 \cup V_2$, the polynomial $fg$ vanishes on $V$ (since on the $V_1$ part, $f = 0$ and on the $V_2$ part, $g = 0$). By assumption, $\mathbf{I}(V)$ is prime, so $fg \in \mathbf{I}(V)$ implies either $f \in \mathbf{I}(V)$ or $g \in \mathbf{I}(V)$. But $f$ was chosen so that $f \notin \mathbf{I}(V)$, so $g \in \mathbf{I}(V)$. This proves what we wanted! □

**Proposition 6.17.5** (Descending Chain Condition, cf. (HW 5, Ex 2.5.13))**.** *Any descending chain of varieties $V_1 \supseteq V_2 \supseteq \ldots$ must stabilize.*

*Proof.* Suppose $V_1 \supseteq V_2 \supseteq \ldots$. The inclusion-reversingness of $\mathbf{I}$ implies that $\mathbf{I}(V_1) \subseteq \mathbf{I}(V_2) \subseteq \ldots$; by Ascending Chain Condition (Theorem 2.7.5) we have an $N$ so that $\mathbf{I}(V_N) = \mathbf{I}(V_{N+1}) = \ldots$. Thus $\mathbf{V}(\mathbf{I}(V_N)) = \mathbf{V}(\mathbf{I}(V_{N+1})) = \ldots$; since $\mathbf{V}(\mathbf{I}(V)) = V$ for every affine variety $V$, we obtain the desired conclusion $V_N = V_{N+1} = \ldots$. □

**Theorem 6.17.6.** *Let $V \subseteq k^n$ be an affine variety. Then we can write $V$ as a finite union $V = V_1 \cup \cdots \cup V_n$ of irreducible affine varieties.*

*Proof.* Assume for the sake of contradiction that $V$ is an affine variety that can't be written as a finite union of irreducible affine varieties. Thus $V$ itself is not irreducible, and we may write $V = V_1 \cup V_1'$ with $V_1 \neq V$ and $V_1' \neq V$. Furthermore, $V_1$ or $V_1'$ also has the property of not being a finite union of irreducible affine varieties: if they both had a finite decomposition into irreducible affines, then we'd get a finite decomposition of $V$ into irreducible affines. Let us assume without loss of generality that $V_1$ is not a finite union. Then we can write $V_1 = V_2 \cup V_2'$, with $V_2 \neq V_1$ and $V_2' \neq V_1$; by the same reason as above we may assume without loss of generality that $V_2$ also has no finite decomposition into irreducible affines, and keep continuing in this way.

We would thus get an infinite descending chain of varieties

$$V \supsetneq V_1 \supsetneq V_2 \supsetneq \ldots,$$

contradictory to Proposition 6.17.5. □

**Definition 6.17.7.** Let $V \subseteq k^n$ be an affine variety. A decomposition $V = V_1 \cup \cdots \cup V_m$ into irreducible affine varieties is called a minimal decomposition if $V_i \not\subseteq V_j$ for any $\neq j$. We call $\{V_i\}$ *the* irreducible components of $V$. △

Calling $\{V_i\}$ *the* irreducible components of $V$ is justified by the following result.

**Theorem 6.17.8.** *Let $V \subseteq k^n$ be an affine variety. Then $V$ has a minimal decomposition, and this decomposition is unique up to reordering.*

*Proof.* We showed that $V$ can be written as a finite union of irreducible varieties in Theorem 6.17.6. From such a decomposition $V = V_1 \cup \cdots \cup V_m$, we may obtain a minimal decomposition by throwing out $V_i$ whenever there exists $j \in [m]$ so that $V_i \subseteq V_j$.

It remains to show uniqueness, which is the main content of this theorem. Suppose we have two minimal decompositions

$$V = V_1 \cup \cdots \cup V_m,$$
$$V = V_1' \cup \cdots \cup V_\ell'.$$

We obtain

$$V_i = V_i \cap V = V_i \cap (V_1' \cup \cdots \cup V_\ell') = (V_i \cap V_1') \cup (V_i \cap V_2') \cup \cdots \cup (V_i \cap V_\ell').$$

We've proven that the intersection of two affine varieties before is another variety (e.g. see Theorem 5.14.12); in particular for every $j \in [\ell]$, the set $V_i \cap V_j'$ is an affine variety. Since $V_i$ is irreducible, we obtain $V_i = V_i \cap V_j'$ for some $j$; in other words, $V_i \subseteq V_j'$.

By the same argument as above, but applied to $V_j'$ (i.e., consider the decomposition of $V_j'$ into

$$V_j' = V_j' \cap (V_1 \cup \cdots \cup V_m) = (V_j' \cap V_1) \cup (V_j' \cap V_2) \cup \cdots \cup (V_j' \cap V_m)$$

and note that for some $k$ we have $V_j' = V_j' \cap V_k$) we obtain $V_j' \subseteq V_k$. In particular, we have the chain of inclusions

$$V_i \subseteq V_j' \subseteq V_k.$$

But the $\{V_i\}$ are minimal, so we can't have $V_i \subseteq V_k$ unless $i = k$ and hence $V_i = V_j' = V_k$. Ripping out $V_i$ and $V_j'$ from the decomposition, we obtain two more minimal decompositions for $V \setminus V_i = V \setminus V_j'$, by

$$V \setminus V_i = \bigcup_{k \in [m], k \neq i} V_k V \setminus V_j' = \bigcup_{k \in [\ell], k \neq j} V_k'.$$

We can repeat this argument, pulling out one irreducible component at a time; we conclude that the two minimal decompositions of $V$ were the same up to reordering. □

From the point of view of ideals, Theorem 6.17.8 says the following:

**Theorem 6.17.9.** *Let $k$ be algebraically closed. Then every radical ideal of $k[x_1, \ldots, x_n]$ can be written uniquely as a finite intersection of prime ideals.*

*Proof.* Apply the ideal-variety correspondence (Theorem 5.14.2) to Theorem 6.17.8, and use the fact that $\mathbf{V}(I \cap J) = \mathbf{V}(I) \cup \mathbf{V}(J)$ from Theorem 5.14.12. □

Notice that since prime ideals are radical, intersections of prime ideals are also automatically radical. There is a notion of *primary* ideal, which somehow capture "priminess" but are also non-radical, and it will turn out that every ideal is an intersection of primary ideals. [I've noted before that principal prime ideals $\langle f \rangle$ correspond to "prime" (=irreducible) polynomials $f$ in Defintinion 5.14.4. Primary ideals correspond to "prime powers" $f^n$, so you might imagine there will be some unique factorization stuff going on.]

## 6.18 Nov 5, 2019

[Avery lectured today.]

[Sorry I've been dropping the ball a lot recently.]

Our main goal this week is to prove the Extension Theorem:

**Theorem 6.18.1** (Extension Theorem, cf. Theorem 6.16.2)**.** *Let $I = \langle f_1, \dots, f_s \rangle \subseteq \mathbb{C}[x_1, \dots, x_n]$. Let $I_1 = I \cap k[x_2, \dots, x_n]$. Write, for each $i$,*

$$f_i = c_i(x_2, \dots, x_n)x_1^{N_i} + (\text{terms of lower } x_1\text{-degree}),$$

*with $N_i \geq 0$ and $c_i \not\geq 0$. If $\mathbf{a} = (a_2, \dots, a_n) \in \mathbf{V}(I_1) \setminus \mathbf{V}(c_1, \dots, c_s)$, then there exists $a_1 \in \mathbb{C}$ so that $(a_1, a_2, \dots, a_n) \in \mathbf{V}(I)$.*

Here's a false proof; hopefully, the failure of this proof will motivate the techniques we will develop to fix it.

*"Proof".* Let $\mathbf{a} = (a_2, \dots, a_n) \in \mathbf{V}(I_1)$. Let us consider the set

$$J = \{f(x_1, \mathbf{a}) \colon f \in I\} \subseteq k[x_1].$$

Note that the set $J$ is an ideal of $k[x_1]$, since if $f(x_1), g(x_1) \in J$ then $f = \widetilde{f}(x_1, \mathbf{a})$ and $g = \widetilde{g}(x_1, \mathbf{a})$ for some $\widetilde{f}, \widetilde{g} \in k[x_1, \dots, x_n]$; then $(f+g)(x_1) = (\widetilde{f} + \widetilde{g})(x_1, \mathbf{a})$ is of the form "take an element of $I$ and evaluate it at $\mathbf{a}$", since $\widetilde{f} + \widetilde{g} \in I$. Similarly, if $h \in k[x_1]$, we may consider $h \in k[x_1, \dots, x_n]$ as a polynomial that doesn't depend on $x_2, \dots, x_n$, and now $h(x_1)f(x_1) = h(x_1, \mathbf{a}) \cdot \widetilde{f}(x_1, \mathbf{a}) = (h\widetilde{f})(x_1, \mathbf{a})$ and $h\widetilde{f} \in I$ as well.

Great, so $J$ is an ideal. Since $\mathbf{a} \notin \mathbf{V}(c_1, \dots, c_s)$, there exists $j$ so that $c_j(\mathbf{a}) \neq 0$, hence for this $j$ we also have $f_j(x_1, \mathbf{a}) \neq 0$. Thus $J$ is not the zero ideal.

Since $J \subseteq k[x_1]$, it is a principal ideal, generated by some $g(x_1, \mathbf{a})$; since $J$ is nonzero $g$ is also nonzero. Now the univariate polynomial $g(x_1, \mathbf{a})$ has a root $a_1 \in \mathbb{C}$, and $g(a_1, \mathbf{a}) = 0$, hence every element of $J$ vanishes at $(a_1, \mathbf{a})$. In particular, for every $f \in I$, we have $f(a_1, \mathbf{a}) = 0$, and now $(a_1, \mathbf{a}) \in \mathbf{V}(I)$. □

Of course, the underlined part is shady because $g$ might be a constant, i.e., $J$ may be all of $k[x_1]$. We want to fix this by detecting when $g$ has a root, i.e., when it has a nontrivial factor $x - a_1$ for some $a_1 \in \mathbb{C}$.

Let's talk about resultants. They are a wonderful algebraic device: given two polynomials

$$f(x) = a_0x^\ell + \cdots + a_\ell$$
$$g(x) = b_0x^m + \cdots + b_m,$$

the resultant $\operatorname{res}(f, g)$ (sometimes denoted $\operatorname{res}_x(f, g)$ or $\operatorname{res}(f, g, x)$) is a polynomial in $\mathbb{Z}[a_0, \dots, a_\ell, b_0, \dots, b_m]$; when $f(x), g(x) \in k[x]$ (i.e. we "plug in" values of $a_0, \dots, a_\ell, b_0, \dots, b_m$ from $k$ to get $f$ and $g$), then $\operatorname{res}(f, g) = 0$ if and only if $f$ and $g$ have a common root in an algebraic closure of $k$. Thus, resultants are a tool for detecting common factors This is a mouthful, and in what follows, we'll define/state/prove the assertions above.

**Lemma 6.18.2.** *Let $f, g \in k[x]$ with $\deg f = \ell$ and $\deg g = m$. Then $f$ and $g$ have a common factor if and only if there exist $A, B \in k[x]$ so that:*

1. *$A$ and $B$ are nonzero,*
2. *$\deg A \leq m - 1$ and $\det B \leq \ell - 1$, and*

3. $Af + Bg = 0$.

*Proof.* We have two directions to prove. Let us begin by assuming that $f$ and $g$ have a common factor $h$; we may write $f = f_1 h$ and $g = g_1 h$ for $f_1, g_1 \in k[x]$. Then we have

$$fg_1 = f\frac{g}{h} = \frac{fg}{h} = \frac{f}{h}g = f_1 g.$$

Then we may pick $A = g_1$ and $B = -f_1$, so that $Af + Bg = fg_1 - f_1 g = 0$. Note that $A$ and $B$ are nonzero, and $\deg A = \deg g - \deg h \leq m - 1$ and $\deg B = \deg f - \deg h \leq \ell - 1$.

Conversely, let us assume there exist $A$ and $B$ so that $Af + Bg = 0$, with $A, B \neq 0$ and $\deg A \leq m-1$ and $\deg B \leq \ell - 1$. Suppose for the sake of contradiction that $f$ and $g$ have no common factor, so $\gcd(f, g) = 1$. Then $1 \in \langle \gcd(f, g) \rangle = \langle f, g \rangle$, so there exist $\widetilde{A}$ and $\widetilde{B}$ so that $\widetilde{A}f + \widetilde{B}g = 1$. Since $B \neq 0$, we have

$$B = B(\widetilde{A}f + \widetilde{B}g) = \widetilde{A}Bf + \widetilde{B}Bg = \widetilde{A}Bf - \widetilde{B}Af = f(\widetilde{A}B - \widetilde{B}A);$$

here we used that $Bg = -Af$. Since $B$ is nonzero, neither is $\widetilde{A}B - \widetilde{B}A$. Now we obtain $\deg B = \deg f + \deg(\widetilde{A}B - \widetilde{B}A) \geq \deg f = \ell$, which is a contradiction to our assumption that $\deg B \leq \ell - 1$. □

Given that we eventually want to detect common factors, we'd probably be using Lemma 6.18.2 in the backwards direction, i.e., we'd want to understand when there are $A$ and $B$ satisfying the conditions of Lemma 6.18.2 and using the lemma to conclude that $f$ and $g$ have a common factor.

How does one possibly check when there are $A$ and $B$ doing this? Let's begin with an example.

**Example 6.18.3.** Suppose

$$f(x) = 2x^2 + 3x + 1$$
$$g(x) = 7x^2 + x + 3.$$

Then, for $A = u_0 x + u_1$ and $B = v_0 x + v_1$, we have

$$A(x)f(x) + B(x)g(x) = (2u_0 + 7v_0)x^3 + (3u_0 + 2u_1 + v_0 + 7v_1)x^2 + (u_0 + 3u_1 + 3v_0 + v_1)x + (u_1 + 3v_1) = 0.$$

When are there $u_0, u_1, v_0, v_1 \in k$ satisfying this? Well, we'd need to solve the linear system of equations given by

$$\begin{bmatrix} 2 & 0 & 7 & 0 \\ 3 & 2 & 1 & 7 \\ 1 & 3 & 3 & 1 \\ 0 & 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ v_0 \\ v_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Okay, when are there nontrivial $(u_0, u_1, v_0, v_1) \in k^4$ satisfying this? Well, we'd take the determinant of the matrix, and check whether it's zero.

Note that everything here is if and only if; there are nontrivial solutions if and only if determinant is zero, and there is a common factor if and only if there is a nontrivial solution to this system of equations, by Lemma 6.18.2. △

We can work out in general this will work out to be. For $f = c_0 x^\ell + \cdots + c_\ell$ and $g = d_0 x^m + \cdots + d_m$, we search for solutions to the linear system of equations described by the Sylvester matrix:

**Definition 6.18.4.** The Sylvester matrix is formed by taking the coefficients of $f$ and $g$, arranging them into a $\ell + 1 = \deg f + 1$ and $m + 1 = \deg g + 1$ column vectors respectively, and then arranging them in a

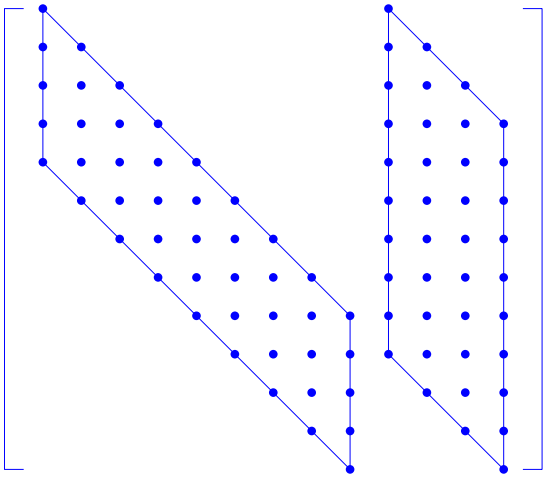
parallelogram fashion $m = \deg g$ and $\ell = \deg f$ times respectively gives an $(m+\ell)$-by-$(m+\ell)$ square matrix, as below:

$$\operatorname{Syl}(f,g,x) \stackrel{\text{def}}{=} \begin{bmatrix} c_0 & 0 & \dots & 0 & d_0 & 0 & \dots & 0 \\ c_1 & c_0 & \dots & 0 & d_1 & d_0 & \dots & 0 \\ c_2 & c_1 & \ddots & 0 & d_2 & d_1 & \ddots & 0 \\ \vdots & \vdots & \ddots & c_0 & \vdots & \vdots & \ddots & d_0 \\ c_\ell & c_{\ell-1} & \dots & \ddots & d_m & d_{m-1} & \dots & \vdots \\ 0 & c_\ell & \ddots & \vdots & 0 & d_m & \ddots & \vdots \\ \vdots & \vdots & \ddots & c_{\ell-1} & \vdots & \vdots & \ddots & d_{m-1} \\ 0 & 0 & \dots & c_\ell & 0 & 0 & \dots & d_m \end{bmatrix}.$$

△

[The "shape" of the Sylvester matrix is a little misleading. I find it easier to argue geometrically about the entries of the matrix by considering a slightly more extreme case, such as below:

$\operatorname{Syl}(\text{small deg}, \text{big deg}, x) \stackrel{\text{def}}{=}$



Here the entries which get coefficients are marked with a bullet; everything else is a zero. The two parallelograms are *always* next to each other for two rows, which is more than what I always expect them to be; equivalently, the parallelograms are always more-vertical/less-diagonal than what I expect them to be; both of these are because a polynomial of degree $d$ has $d+1$ coefficients.]

[I was today years old when I learn that rhombi have four equal side lengths. I've changed my wording accordingly. Oops]

**Definition 6.18.5.** The resultant of two polynomials $f, g \in k[x]$ is defined to be

$$\operatorname{res}(f,g) \stackrel{\text{def}}{=} \det \operatorname{Syl}(f,g,x).$$

△

**Proposition 6.18.6.** *The polynomials $f$ and $g$ have a common factor if and only if* $\operatorname{res}(f,g) = 0$.

*Proof.* Let $\ell = \deg f$ and $m = \deg g$, and let $\mathcal{P}_d$ denote the $(d+1)$-dimensional vector space of polynomials of degree at most $d$. The Sylvester matrix $\operatorname{Syl}(f,g)$ is the matrix of the linear transformation $\varphi\colon \mathcal{P}_{\ell-1} \times \mathcal{P}_{m-1} \to \mathcal{P}_{\ell+m-1}$ between $(\ell+m)$-dimensional vector spaces given by sending $\varphi\colon (A,B) \mapsto Af + Bg$. The linear transformation $\varphi$ has a nontrivial kernel if and only if $\det \operatorname{Syl}(f,g) = 0$. But Lemma 6.18.2 says that $\varphi$ has a nontrivial kernel if and only if $f$ and $g$ have a common factor. □

**Corollary 6.18.7.** *The polynomials $f \in \mathbb{C}[x]$ and $g \in \mathbb{C}[x]$ have a common root if and only if* $\operatorname{res}(f, g) = 0$.

*Proof.* Since $\mathbb{C}$ is algebraically closed, the common factor will have a root. Conversely, $f$ and $g$ have a common root $r$, then $(x - r)$ is a common factor between $f$ and $g$. □

Here's one typical application of Proposition 6.18.6.

**Example 6.18.8.** Let $f = xy - 1$ and $g = x^2 + y^2 - 4$ be polynomials in $\mathbb{C}[x, y]$. Think of $f$ and $g$ as univariate polynomials in $x$, with coefficients not in $\mathbb{C}[y]$, but in the *field of rational functions* $\mathbb{C}(y)$. (The elements of this field are rational functions, i.e., quotients of polynomials. We've seen them before, in our proof of the Hilbert Nullstellensatz (cf. Theorem 5.13.3).)

Thus we consider $f, g \in (\mathbb{C}(y))[x]$. Then, following Definition 6.18.5

$$\operatorname{res}(f, g, x) = \det \begin{bmatrix} y & 0 & 1 \\ -1 & y & 0 \\ 0 & -1 & y^2 - 4 \end{bmatrix} = y^4 - 4y^2 + 1 \in \mathbb{C}(y).$$

In fact we see that $y^4 - 4y^2 + 1 \in \mathbb{C}[y]$ (i.e. it's an honest polynomial rather than a rational function); this is not a coincidence. △

**Proposition 6.18.9.** *The resultant is an integer polynomial in the coefficients of $f$ and $g$. In particular, if the coefficients of $f, g \in (\mathbb{C}(\mathbf{y}))[x]$ are polynomials in $\mathbb{C}[\mathbf{y}] \subseteq \mathbb{C}(\mathbf{y})$, then the resultant* $\operatorname{res}(f, g) \in \mathbb{C}(\mathbf{y})$ *is actually an element of* $\mathbb{C}[\mathbf{y}]$.

*Proof.* The resultant is the determinant of the Sylvester matrix, hence it is a polynomial in the entries of the Sylvester matrix. The entries of these matrices are coefficients of $f$ and $g$. □

**Proposition 6.18.10.** *Given $f, g \in k[x]$, we can find $A, B \in k[x]$ such that $Af + Bg = \operatorname{res}(f, g, x)$. In other words,* $\operatorname{res}(f, g) \in \langle f, g \rangle$. *We may pick $A$ and $B$ so that it is a polynomial in the coefficients of $f$ and $g$.*

(Note that $\operatorname{res}(f, g)$ doesn't depend on $x$, so we're really saying that $\operatorname{res}(f, g)$ is in the elimination ideal!)

(The application to keep in mind is when $k = \mathbb{C}(\mathbf{y})$ is a field of rational functions, but $f$ and $g$ really have coefficients in $\mathbb{C}[\mathbf{y}] \subseteq \mathbb{C}(\mathbf{y})$. Then Proposition 6.18.10 states that $A$ and $B$ also live in $\mathbb{C}[\mathbf{y}]$.)

*Proof.* We treat some easy cases first.

If $\operatorname{res}(f, g, x) = 0$ then we may pick $A = B = 0$.

If $f = c_0$ is a constant and $g$ is a nonzero polynomial, then we compute

$$\operatorname{res}(c_0, g, x) = \det \begin{bmatrix} c_0 & 0 & \dots & 0 \\ 0 & c_0 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & & 0 \\ 0 & 0 & \dots & c_0 \end{bmatrix} = c_0^m$$

by definition. Thus we may pick $A = c_0^{m-1}$ and $B = 0$.

We can do a similar argument when $g = d_0$ is a constant polynomial.

So let's assume $f$ and $g$ are both nonconstant, with $\deg f = \ell$ and $\deg g = m$, and that $\operatorname{res}(f, g) \neq 0$. Let us solve $Af + Bg = 1$ with $\deg A \leq m - 1 = \deg g - 1$ and $\deg B \leq \ell - 1 = \deg f - 1$; cf. Example 6.18.3. Let

us denote the coefficients of $A$ and $B$ by

$$A = u_0 x^{m-1} + \cdots + u_{m-1}$$
$$B = v_0 x^{\ell-1} + \cdots + v_{\ell-1}.$$

We are now trying to solve

$$\text{Syl}(f,g,x) \begin{bmatrix} u_0 \\ \vdots \\ u_{m-1} \\ v_0 \\ \vdots \\ v_{\ell-1} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

The assumption $\text{res}(f,g,x) \neq 0$ means that the Sylvester matrix is invertible, hence there exists a unique solution for $(u_0, \ldots, u_{m-1}, v_0, \ldots, v_{\ell-1}) \in k^{m+\ell}$. This gives a solution $Af + Bg = 1$.

Cramer's rule (see Aside 6.18.11 for a precise statement) says that the solution is of the form

$$u_i = \frac{\det(\text{something})}{\det(\text{Syl}(f,g,x))}; \qquad v_i = \frac{\det(\text{something})}{\det(\text{Syl}(f,g,x))},$$

where all the somethings are matrices obtained from the Sylvester matrix and replacing a column with $(0, \ldots, 0, 1)$. In particular, the entries of the somethings are either 0, 1, or coefficients of $f$ and $g$, hence the determinant is a polynomial in the coefficients of $f$ and $g$.

Since $\det(\text{Syl}(f,g,x)) = \text{res}(f,g,x)$, we see that

$$\underbrace{\text{res}(f,g,x)A}_{\text{poly in coeffs of } f,g} f + \underbrace{\text{res}(f,g,x)B}_{\text{poly in coeffs of } f,g} g = \text{res}(f,g,x).$$

□

**Aside 6.18.11** (Cramer's Rule). *Let $A$ be an invertible $n \times n$ matrix, and fix $\mathbf{b} \in \mathbb{R}^n$. Then the unique solution to $A\mathbf{x} = \mathbf{b}$ is given by*

$$x_i = \frac{\det A_i}{\det A}$$

*for all $i \in [n]$, where $A_i$ is the matrix obtained by replacing the $i$th column of $A$ by $\mathbf{b}$.*

For example, in the 2-by-2 case, the solution to

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

is given by

$$x_1 = \frac{\det \begin{bmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{bmatrix}}{\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}} = \frac{b_1 a_{22} - b_2 a_{12}}{a_{11} a_{22} - a_{12} a_{21}} \qquad x_2 = \frac{\det \begin{bmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{bmatrix}}{\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}} = \frac{a_{11} b_2 - b_1 a_{12}}{a_{11} a_{22} - a_{12} a_{21}}.$$

## 6.19 Nov 7, 2019

[Avery lectured today.]

Recall from last class that we defined the Sylvester matrix $\mathrm{Syl}(f, g, x)$ (Definition 6.18.4) and the resultant (Definition 6.18.5) $\det \mathrm{Syl}(f, g, x)$. We proved Proposition 6.18.6, which said that $f \in \mathbb{C}[x]$ and $g \in \mathbb{C}[x]$ have a common root if and only if $\mathrm{res}(f, g) = 0$. The point is that we had a "proof" of the Extension Theorem (see the underlined part of "proof" of Theorem 6.18.1) and needed to understand when polynomials were nonconstant, and we'll be doing that by understanding resultants.

We also proved Proposition 6.18.10, which is important:

**Proposition 6.19.1.** *Let $f, g \in k[x]$. There are polynomials $A$ and $B$ such that $Af + Bg = \mathrm{res}(f, g, x)$. If $\deg f > 0$ or $\deg g > 0$, the coefficients of $A$ and $B$ are integer polynomials in the coefficients of $A$ and $B$.*

Let's do a reality check. Let $k = \mathbb{C}$ in the proposition above. We have $\mathrm{res}(f, g, x) \in \mathbb{C}$. But Proposition 6.19.1 says that $\mathrm{res}(f, g, x) \in \langle f, g \rangle$, so if $f$ and $g$ don't have a common root then this is saying some nonzero constant is in $\langle f, g \rangle$, hence that $\langle f, g \rangle = \mathbb{C}[x]$. But if $f$ and $g$ don't have a common root, then its gcd is 1, so everything is consistent.

Last time we considered (Example 6.18.8) the polynomials $f = xy - 1$ and $g = x^2 + y^2 - 4$ as univariate polynomials in $(\mathbb{C}(y))[x]$. We computed that its resultant was $y^4 - 4y^2 + 1 \in \mathbb{C}(y)$; this is indeed a constant in our base field. In particular, $f$ and $g$ don't have a common factor, so their gcd is 1. Note that $(\mathbb{C}(y))[x]$ is a univariate polynomial ring over a field, hence we can still do the (extended!) Euclidean Algorithm to find a solution to $h_1 f + h_2 g = \gcd(f, g) = 1$ if we want. Avery did that for us, and asserted

$$\Bigg( \underbrace{\frac{-y}{y^4 - 4y^2 + 1} x + \frac{1}{y^4 - 4y^2 + 1}}_{\text{univariate polynomial; coefficients in } \mathbb{C}(y)} \Bigg) f + \Bigg( \underbrace{\frac{y^2}{y^4 - 4y^2 + 1}}_{\text{constant in } \mathbb{C}(y)} \Bigg) g = 1.$$

Note that the coefficients of $h_1$ and $h_2$ are in $\mathbb{C}(y)$, but their denominators are precisely the resultant. This is Cramer's rule (Aside 6.18.11) at work.

In this sense one can think of resultants as a "denominator-free gcd", in light of Proposition 6.19.1.

To Extension and Beyond Let $f, g \in k[x_1, \ldots, x_n]$. Let us write

$$f = c_0 x_1^\ell + \cdots + c_\ell$$
$$g = d_0 x_1^m + \cdots + x_m,$$

so that $c_i, d_j \in k[x_2, \ldots, x_n]$. We defined the resultant $\mathrm{res}(f, g, x) \in k(x_2, \ldots, x_n)$, and proved that they are polynomials in the variables $c_1, \ldots, c_\ell, d_1, \ldots, d_m$, which are themselves polynomials in the variables $x_2, \ldots, x_n$. It follows that $\mathrm{res}(f, g, x) \in k[x_2, \ldots, x_n]$.

To prove the extension theorem, we need to know how resultants behave with respect to evaluation. In other words, let us pick $\mathbf{a} = (a_2, \ldots, a_n) \in \mathbb{C}^{n-1}$; we want to know the relationship between

$$\mathrm{res}(f, g, x_1)(\mathbf{a}) \quad \text{and res} \quad (f(x_1, \mathbf{a}), g(x_1, \mathbf{a}), x_1).$$

Let's see some examples.

**Example 6.19.2.** Set

$$f(x, y) = x^2 y + 3x - 1,$$
$$g(x, y) = 6x^2 + y^2 - 4.$$

Then

$$\operatorname{res}(f,g,x) = \det \begin{bmatrix} y & 0 & 6 & 0 \\ 3 & y & 0 & 6 \\ -1 & 3 & y^2-4 & 0 \\ 0 & -1 & 0 & y^2-4 \end{bmatrix} \in \mathbb{C}[y].$$

Let us plug in $y = 0$ into our univariate polynomial $\operatorname{res}(f,g,x) \in \mathbb{C}[y]$. We obtain

$$\operatorname{res}(f,g,x)(0) = \det \begin{bmatrix} 0 & 0 & 6 & 0 \\ 3 & 0 & 0 & 6 \\ -1 & 3 & -4 & 0 \\ 0 & -1 & 0 & -4 \end{bmatrix} = -180.$$

Let us, on the other hand, plug in $y = 0$ first and then take the resultant. We have

$$f(x,0) = 3x - 1$$
$$g(x,0) = 6x^2 - 4.$$

Then

$$\operatorname{res}(f(x,0),g(x,0),x) = \det \begin{bmatrix} 3 & 0 & 6 \\ -1 & 3 & 0 \\ 0 & -1 & -4 \end{bmatrix},$$

which has determinant $-30$. △

A lot of things changed between $\operatorname{res}(f,g,x)(0)$ and $\operatorname{res}(f(x,0),g(x,0),x)$. But actually, it's not a complete disaster. The Sylvester matrices

$$\operatorname{Syl}(f,g,x)(0) = \begin{bmatrix} 0 & 0 & 6 & 0 \\ 3 & 0 & 0 & 6 \\ -1 & 3 & -4 & 0 \\ 0 & -1 & 0 & -4 \end{bmatrix} \quad \text{and} \quad \operatorname{Syl}(f(x,0),g(x,0),x) = \begin{bmatrix} 3 & 0 & 6 \\ -1 & 3 & 0 \\ 0 & -1 & -4 \end{bmatrix}$$

are related by the fact that the smaller one is a minor of the bigger one (delete the first row and third column, and see for yourself!). We'll see that this is not a coincidence, and that it is really due to the fact that the leading term $x^2y$ of $f$ died upon setting $y = 0$. We also could have predicted that the resultant would change by a factor of 6, because it is the leading term of of $g(x,0)$; indeed, if $M$ is a matrix with only one nonzero element $m_{1,3}$ in the first row, and $M_{1,3}$ is the minor of $M$ obtained by deleting the first row and third column, then $\det M = m_{1,3} \det M_{1,3}$.

We'll make this argument precise after seeing what goes wrong with a more badly behaved example.

**Example 6.19.3.** This time, let

$$f = x^2y + x - 1,$$
$$g = x^2y + x + y^2 - 4.$$

Then

$$\operatorname{res}(f,g,x) = \det \begin{bmatrix} y & 0 & y & 0 \\ 1 & y & 1 & y \\ -1 & 1 & y^2-4 & 1 \\ 0 & -1 & 0 & y^2-4 \end{bmatrix} \in \mathbb{C}[y]$$

Plugging in $y = 0$ into this univariate polynomial we obtain

$$\operatorname{res}(f,g,x)(0) = \det \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ -1 & 1 & -4 & 1 \\ 0 & -1 & 0 & -4 \end{bmatrix} = 0,$$

since there's a row of zeros. Plugging in $y = 0$ first and taking the resultant, we obtain

$$f(x, 0) = x - 1,$$
$$g(x, 0) = x - 4.$$

We obtain

$$\text{res}(f(x, 0), g(x, 0), x) = \det \begin{bmatrix} 1 & 1 \\ -1 & -4 \end{bmatrix},$$

which has determinant $-3$. △

Instead of being predictably off by a factor of 6, we're now (unpredictably?) off by, like, a factor of $\infty$. Terrible!

**Proposition 6.19.4.** *Let $f, g \in k[x_1, \ldots, x_n]$, and let $\deg_{x_1}(f) = \ell$ and $\deg_{x_1}(g) = m$. Let $\mathbf{a} \in k^{n-1}$, and suppose that:*

1. *$\deg_{x_1} f(x_1, \mathbf{a}) = \ell$, and*
2. *$g(x_1, \mathbf{a}) \neq 0$ and has degree $p \leq m$.*

*Then if $f = c_0 x_1^\ell + \cdots + c_\ell$, and $g = d_0 x^m + \cdots + d_m$, with $c_i, d_j \in k[x_2, \ldots, x_n]$, the equality*

$$\text{res}(f, g, x_1)(\mathbf{a}) = c_0(\mathbf{a})^{m-p} \text{res}(f(x_1, \mathbf{a}), g(x_1, \mathbf{a}), x_1)$$

*holds.*

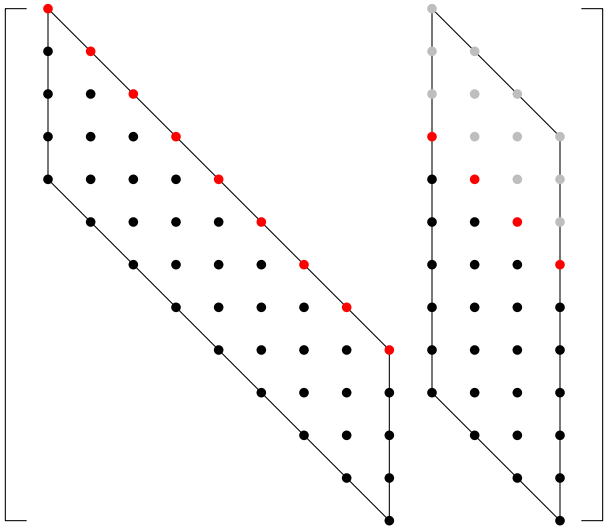
*Proof.* Let $h = \text{res}(f, g, x_1)$. We have

$$h(\mathbf{a}) = \det \begin{bmatrix} c_0(\mathbf{a}) & 0 & \ldots & 0 & d_0(\mathbf{a}) & 0 & \ldots & 0 \\ c_1(\mathbf{a}) & c_0(\mathbf{a}) & \ldots & 0 & d_1(\mathbf{a}) & d_0(\mathbf{a}) & \ldots & 0 \\ c_2(\mathbf{a}) & c_1(\mathbf{a}) & \ddots & 0 & d_2(\mathbf{a}) & d_1(\mathbf{a}) & \ddots & 0 \\ \vdots & \vdots & \ddots & c_0(\mathbf{a}) & \vdots & \vdots & \ddots & d_0(\mathbf{a}) \\ c_\ell(\mathbf{a}) & c_{\ell-1}(\mathbf{a}) & \ldots & \ddots & d_m(\mathbf{a}) & d_{m-1}(\mathbf{a}) & \ldots & \vdots \\ 0 & c_\ell(\mathbf{a}) & \ddots & \vdots & 0 & d_m(\mathbf{a}) & \ddots & \vdots \\ \vdots & \vdots & \ddots & c_{\ell-1}(\mathbf{a}) & \vdots & \vdots & \ddots & d_{m-1}(\mathbf{a}) \\ 0 & 0 & \ldots & c_\ell(\mathbf{a}) & 0 & 0 & \ldots & d_m(\mathbf{a}) \end{bmatrix}. \tag{8}$$

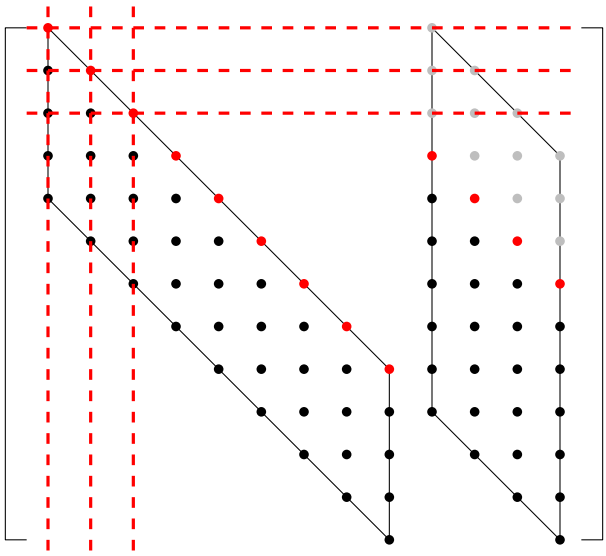
If $p = m$, that is, $d_0(\mathbf{a})$ really was the leading coefficient of $g(x_1, \mathbf{a})$, then the matrix in Equation (8) is actually $\text{Syl}(f(x_1, \mathbf{a}), g(x_1, \mathbf{a}), x_1)$; thus we obtain $h(\mathbf{a}) = \det \text{Syl}(f(x_1, \mathbf{a}), g(x_1, \mathbf{a}), x_1) = \text{res}(f(x_1, \mathbf{a}), g(x_1, \mathbf{a}), x_1)$, as we wanted to prove.

In general, though, if $p < m$ then the matrix in Equation (8) is not $\text{Syl}(f(x_1, \mathbf{a}), g(x_1, \mathbf{a}), x_1)$. It's closely related, though.

Let's think geometrically about the entries of this matrix. Let's assume $\ell, m \neq 0$ so that we actually have "parallelograms" in the matrix in Equation (8). If $g(x_1, \mathbf{a})$ has degree $p$, that means that the coefficients $d_0(\mathbf{a}), \ldots, d_{m-p-1}(\mathbf{a})$ are all equal to zero, and $d_{m-p}(\mathbf{a}) \neq 0$. Also $c_0(\mathbf{a}) \neq 0$. So geometrically, we are in the following situation, where grey bullets mean the entry is most definitely zero, red bullets mean the entry is most definitely not zero, and black dots can be anything (and won't really be relevant in this argument):



Note that there are $m-p$ grey bullets in each column of the right parallelogram. Also note that there is only one nonzero element, in the first row. It's equal to $c_0(\mathbf{a})$, so the determinant is $c_0(\mathbf{a})$ multiplied by the minor obtained by removing the first row and column. As long the right parallelogram is greyed out, we still only have one nonzero element, it's still equal to $c_0(\mathbf{a})$, so we keep removing the first row and column and multiplying the determinant by $c_0(\mathbf{a})$ until we've reached the first red bullet in the right parallelogram:



So we've taken $m-p$ minors, taking care each time to remember to multiply the determinant of the resulting matrix with $c_0(\mathbf{a})$, and arrived at the above matrix. Well, this resulting matrix *is* the Sylvester matrix $\mathrm{Syl}(f(x_1,\mathbf{a}),g(x_1,\mathbf{a}),x_1)$ [(!)]. So we've concluded that the determinant in Equation (8) is equal to $c_0(\mathbf{a})^{m-p}\mathrm{res}(f(x_1,\mathbf{a}),g(x_1,\mathbf{a}),x_1)$, which is exactly what we wanted to prove.

We only have some degenerate cases left to clean up; note that we assumed $\ell, m \neq 0$ to get the above argument to work. But if $\ell = 0$, so $f = c_0$, we've already computed $\mathrm{res}(c_0,g,x_1) = c_0^{\deg g}$ (see the proof of Proposition 6.18.10),hence

$$\mathrm{res}(f,g,x_1)(\mathbf{a}) = c_0(\mathbf{a})^m = c_0(\mathbf{a})^{m-p}c_0(\mathbf{a})^p = c_0(\mathbf{a})^{m-p}\mathrm{res}(f(x_1,\mathbf{a}),g(x_1,\mathbf{a}),x_1),$$

as desired. The case $m = 0$ can be treated analogously. □

**Corollary 6.19.5.** *Let $f, g$ be nonzero polynomials in $\mathbb{C}[x_1, \dots, x_n]$; suppose $\deg_{x_1}(f) = \ell$ or $\deg_{x_1}(g) = m$. Let $\mathbf{a} \in \mathbb{C}^{n-1}$. If $\operatorname{res}(f, g, x_1) = 0$, then $f(a_1, \mathbf{a}) = g(a_1, \mathbf{a}) = 0$ for some $a_1 \in \mathbb{C}$.*

*Proof.* Let us assume without loss of generality that $\deg_{x_1}(f) = \ell$, since the other case is completely analogous. In this case, Proposition 6.19.4 says that

$$0 = \operatorname{res}(f, g, x_1)(\mathbf{a}) = \underbrace{c_0(\mathbf{a})^{m - \deg_{x_1}(g(x_1, \mathbf{a}))}}_{\neq 0} \underbrace{\operatorname{res}(f(x_1, \mathbf{a}), g(x_1, \mathbf{a}), x_1)}_{=0}.$$

We win. □

We can now prove what we've set out to prove!

**Theorem 6.19.6** (Extension Theorem, cf. Theorem 6.16.2 and 6.18.1)**.** *Let $I = \langle f_1, \dots, f_s \rangle \subseteq \mathbb{C}[x_1, \dots, x_n]$ and write*

$$f_i = c_i(x_2, \dots, x_n) x_1^{N_i} + (\text{lower } x_1\text{-degree terms})$$

*for each $i \in [s]$. If $\mathbf{a} \in \mathbb{C}^{n-1}$ and $\mathbf{a} \in \mathbf{V}(I_1) \setminus \mathbf{V}(c_1, \dots, c_s)$ then $(a_1, \mathbf{a}) \in \mathbf{V}(I)$ for some $a_1 \in \mathbb{C}$.*

*Proof.* We had a "proof" with a detail that needed fixing, which we reproduce below.

Let $\mathbf{a} = (a_2, \dots, a_n) \in \mathbf{V}(I_1)$. Let us consider the set

$$J = \{f(x_1, \mathbf{a}) \colon f \in I\} \subseteq k[x_1].$$

Note that the set $J$ is an ideal of $k[x_1]$, since if $f(x_1), g(x_1) \in J$ then $f = \widetilde{f}(x_1, \mathbf{a})$ and $g = \widetilde{g}(x_1, \mathbf{a})$ for some $\widetilde{f}, \widetilde{g} \in k[x_1, \dots, x_n]$; then $(f + g)(x_1) = (\widetilde{f} + \widetilde{g})(x_1, \mathbf{a})$ is of the form "take an element of $I$ and evaluate it at $\mathbf{a}$", since $\widetilde{f} + \widetilde{g} \in I$. Similarly, if $h \in k[x_1]$, we may consider $h \in k[x_1, \dots, x_n]$ as a polynomial that doesn't depend on $x_2, \dots, x_n$, and now $h(x_1) f(x_1) = h(x_1, \mathbf{a}) \cdot \widetilde{f}(x_1, \mathbf{a}) = (h\widetilde{f})(x_1, \mathbf{a})$ and $h\widetilde{f} \in I$ as well.

Great, so $J$ is an ideal. Since $\mathbf{a} \notin \mathbf{V}(c_1, \dots, c_s)$, there exists $j$ so that $c_j(\mathbf{a}) \neq 0$, hence for this $j$ we also have $f_j(x_1, \mathbf{a}) \neq 0$. Thus $J$ is not the zero ideal.

Since $J \subseteq k[x_1]$, it is a principal ideal, generated by some $g(x_1, \mathbf{a})$; since $J$ is nonzero $g$ is also nonzero. Now the univariate polynomial $g(x_1, \mathbf{a})$ has a root $a_1 \in \mathbb{C}$, and $g(a_1, \mathbf{a}) = 0$, hence every element of $J$ vanishes at $(a_1, \mathbf{a})$. In particular, for every $f \in I$, we have $f(a_1, \mathbf{a}) = 0$, and now $(a_1, \mathbf{a}) \in \mathbf{V}(I)$.

We need to find a root $a_1 \in \mathbb{C}$ of $g(x_1, \mathbf{a})$. Note that $\mathbf{a} \notin \mathbf{V}(c_1, \dots, c_s)$, we have $f_j(x_1, \mathbf{a})$ has degree $N_j$ for some $j \in [s]$. Thus, we may apply Proposition 6.19.1 to obtain

$$h = \operatorname{res}(f_j, g, x_1) \in \langle f_j, g, x_1 \rangle \cap \mathbb{C}[x_2, \dots, x_n] \subseteq I_1,$$

and plugging in $\mathbf{a} \in \mathbf{V}(I_1)$ says that $h(\mathbf{a}) = 0$. On the other hand, Proposition 6.19.4 says that

$$0 = h(\mathbf{a}) = (\text{some constant}) \cdot \underbrace{\operatorname{res}(f_j(x_1, \mathbf{a}), g(x_1, \mathbf{a}), x_1)}_{=0}$$

and since the resultant is zero, Corollary 6.19.5 says that there is $a_1$ that makes both $f_j$ and $g$ vanish. In particular, $g$ is nonconstant! □

# 7 Invariant Theory of Finite Groups

## 7.20 Nov 12, 2019

[The second prelim is next Thursday. There will be more information on Thursday and a review session next Tuesday. It will focus on things that were not on the previous prelim, but it's important/useful to know that material. She doesn't intend us to know the proof of the Extension Theorem (Theorem 6.18.1), but we should definitely know how to use it.]

We'll be talking about the invariant theory of finite groups for the next little while. Although it is certainly helpful to know the theory of finite groups, it is certainly not necessary. See the appendix of the textbook for a formal definition, but we'll work with concrete examples here.

The basic goal of invariant theory is to describe the polynomials that are *invariant*, or *unchanged*, when we change variables according to permutations (or, as we'll see exactly how later, according to finite groups of matrices).

A first step towards this goal is understanding symmetric polynomials. For those familiar with the group theoretic language, we are trying to understand invariants of the symmetric group [for which I'll tentatively reserve the notation $\mathfrak{S}_n$, although I don't intend on using this much].

It turns out that symmetric polynomials arise naturally when studying roots of polynomials. As an example:

**Example 7.20.1.** Let $f(x) = x^3 + bx^2 + cx + d$ with $b, c, d, \in \mathbb{C}$. Since $\deg f = 3$, there are three roots $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{C}$ of $f$, and we obtain

$$f(x) = (x-\alpha_1)(x-\alpha_2)(x-\alpha_3) = x^3 - (\alpha_1+\alpha_2+\alpha_3)x^2 + (\alpha_1\alpha_2+\alpha_1\alpha_3+\alpha_2\alpha_3)x - \alpha_1\alpha_2\alpha_3.$$

Thus the formulas

$$\begin{aligned} b &= -(\alpha_1+\alpha_2+\alpha_3) \\ c &= \alpha_1\alpha_2+\alpha_1\alpha_3+\alpha_2\alpha_3 \\ d &= -\alpha_1\alpha_2\alpha_3 \end{aligned}$$

pop out. These expressions for $b$, $c$, and $d$ as polynomials in the roots $\alpha_i$ are highly symmetric: if you interchange any two variables, e.g. switch $\alpha_1 \leftrightarrow \alpha_2$ in the formula for $c$, we obtain $\alpha_2\alpha_1 + \alpha_2\alpha_3 + \alpha_1\alpha_3$, but this *is* the same polynomial as $\alpha_1\alpha_2 + \alpha_1\alpha_3 + \alpha_2\alpha_3$. In this sense, the polynomial is invariant under the switch $\alpha_1 \leftrightarrow \alpha_2$. △

**Definition 7.20.2.** A polynomial $f \in k[x_1, \ldots, x_n]$ is symmetric if $f(x_{i_1}, \ldots, x_{i_n}) = f(x_1, \ldots, x_n)$ for every permutation $(i_1, \ldots, i_n)$ of $(1, \ldots, n)$. (This is an equality of polynomials.) △

In this language, our discussion in Example 7.20.1 says that the polynomial expressions for $b$, $c$, and $d$ as in the roots $\alpha_i$ are symmetric polynomials.

Let $f = x^2 + y^2 \in k[x, y, z]$. Is $f$ symmetric? (No, since $y \leftrightarrow z$ turns $x^2 + y^2$ into $x^2 + z^2$, and these aren't equal.)

Let $f = x^2 + y^2 + z^2 \in k[x, y, z]$. Is $f$ symmetric? (Yes, no matter what permutation of the variables we pick, we'll always end up with the sum of the squares of the three variables.)

**Observation 7.20.3.** Note that if $f$ is a monic degree $d$ polynomial over an algebraically closed field, then as in Example 7.20.1 we may write

$$f(x) = (x-\alpha_1)\ldots(x-\alpha_d) = (x-\alpha_{i_1})\ldots(x-\alpha_{i_d})$$

for any permutation $(i_1, \dots, i_d)$ of $(1, \dots, d)$. Teasing out what this means for the polynomial expressions for the coefficients of $f$ in terms of roots of $f$ will give us that the coefficients of $f$ can be expressed as symmetric polynomials in the roots. △

**Definition 7.20.4.** The elementary symmetric polynomials are denoted $\sigma_1, \dots, \sigma_n \in k[x_1, \dots, x_n]$, and they are defined, for every $r \in [n]$, by

$$\sigma_r = \sum_{1 \le i_1 < \dots < i_r \le n} x_{i_1} \dots x_{i_r}.$$

Thus, for example, $\sigma_1(x_1, \dots, x_n) = x_1 + \dots + x_n$ and $\sigma_n(x_1, \dots, x_n) = x_1 x_2 \dots x_n$. (As another example, $\sigma_2(x_1, x_2, x_3) = x_1x_2 + x_1x_3 + x_2x_3$.) △

**Observation 7.20.5** (cf. Observation 7.20.3)**.** If $f$ is a monic degree $d$ polynomial over an algebraically closed field, then

$$f(x) = x^n - \sigma_1(\alpha_1, \dots, \alpha_n)x^{n-1} + \sigma_2(\alpha_1, \dots, \alpha_n)x^{n-2} - \dots.$$

These are sometimes called *Vieta's formulas*. △

**Observation 7.20.6.** Let $g(y_1, \dots, y_n) \in k[y_1, \dots, y_n]$ be any polynomial in $n$ variables. The polynomial $g(\sigma_1(x_1, \dots, x_n), \dots, \sigma_n(x_1, \dots, x_n)) \in k[x_1, \dots, x_n]$ obtained by replacing each $y_i$ with the $i$th elementary symmetric polynomial in the $x_i$'s is a symmetric polynomial. △

A very key theorem is that every symmetric polynomial can be uniquely in this way. This theorem will be like a golden standard for us.

**Theorem 7.20.7** (Fundamental Theorem of Symmetric Polynomials)**.** *Every symmetric polynomial $f \in k[x_1, \dots, x_n]$ can be written uniquely as a polynomial in the elementary symmetric polynomials.*

We'll present a proof using Gröbner basis, although there are alternative (e.g. combinatorial) proofs of this.

*Proof.* Let us fix the lex order with $x_1 > \dots > x_n$. Let $f \in k[x_1, \dots, x_n]$ be a nonzero symmetric polynomial, and let $\mathrm{LT}(f) = a\mathbf{x}^\alpha$; recall here that $\mathbf{x} = (x_1, \dots, x_n)$ and $\alpha = (\alpha_1, \dots, \alpha_n)$ is an $n$-tuple of nonnegative integers. Convince yourself that $\alpha_1 \ge \dots \ge \alpha_n$ (this is because of lex order business – e.g. since $\mathbf{x}^\alpha$ is a leading term we must have $x_1^{\alpha_1}x_2^{\alpha_2}x_3^{\alpha_3}\dots x_n^{\alpha_n} >_{\text{lex}} x_2^{\alpha_1}x_1^{\alpha_2}x_3^{\alpha_3}\dots x_n^{\alpha_n}$; this implies $\alpha_1 \ge \alpha_2$). Thus we may define the polynomial

$$h^\alpha(x_1, \dots, x_n) = \sigma_1(x_1, \dots, x_n)^{\alpha_1 - \alpha_2}\sigma_2(x_1, \dots, x_n)^{\alpha_2 - \alpha_3} \dots \sigma_n\sigma_n(x_1, \dots, x_n)^{\alpha_n}.$$

Note that $\mathrm{LT}(\sigma_r) = x_1 \dots x_r$, hence $\mathrm{LT}(h^\alpha) = \mathrm{LT}(\sigma_1)^{\alpha_1 - \alpha_2}\mathrm{LT}(\sigma_2)^{\alpha_2 - \alpha_3} \dots = x_1^{\alpha_1}x_2^{\alpha_2} \dots = \mathbf{x}^\alpha$.

Thus we've constructed $h^\alpha$ so that $\mathrm{LM}(f) = \mathrm{LM}(h^\alpha)$. In particular, we may consider the symmetric polynomial $f - ah^\alpha$, which is either zero (then we have written $f$ has a polynomial in the elementary symmetric polynomial) or satisfies $\mathrm{multideg}(f) > \mathrm{multideg}(f - ah^\alpha)$. In case $f - ah^\alpha$ is nonzero, we may apply the same reasoning as above to find $a_1, h^{\alpha_1}$ so that either $(f - ah^\alpha) - a_1h^{\alpha_1} = 0$ or satisfies $\mathrm{multideg}(f - ah^\alpha) > \mathrm{multideg}(f - ah^\alpha - a_1h_1^{\alpha_1})$. Indeed, as long as we are nonzero we may repeat this argument, successively obtaining polynomials of smaller multidegree. The well-orderedness of $>_{\text{lex}}$ guarantees that we must terminate after finitely many steps. This shows existence of $g$.

Let's show uniqueness of $g$. Precisely, suppose that there are two polynomials $g_1, g_2 \in k[y_1, \dots, y_n]$, so that $g_1(\sigma_1(x_1, \dots, x_n), \dots, \sigma_n(x_1, \dots, x_n)) = g_2(\sigma_1(x_1, \dots, x_n), \dots, \sigma_n(x_1, \dots, x_n)) \in k[x_1, \dots, x_n]$ as polynomials. We want to show that $g_1 - g_2 = 0$ as polynomials in $k[y_1, \dots, y_n]$ as well.

Let us write $g_1 - g_2$ as a sum $g_1 - g_2 = \sum_\beta a_\beta \mathbf{y}^\beta$. Then observe that $(g_1 - g_2)(\sigma_1(x_1, \dots, x_n), \dots, \sigma_n(x_1, \dots, x_n))$ is a sum of polynomials $g_\beta \overset{\text{def}}{=} a_\beta\sigma_1(x_1, \dots, x_n)^{\beta_1} \dots \sigma_n(x_1, \dots, x_n)^{\beta_n}$. We have

$$\mathrm{LT}(g_\beta(\sigma_1, \dots, \sigma_n)) = x_1^{\beta_1 + \dots + \beta_n}x_2^{\beta_2 + \dots + \beta_n} \dots x_n^{\beta_n}.$$

We claim that the $\mathrm{LT}(g_\beta(\sigma_1, \ldots, \sigma_n))$ have pairwise distinct leading terms. Indeed, the leading term has exponent vector $(\beta_1 + \cdots + \beta_n, \beta_2 + \cdots + \beta_n, \ldots, \beta_n)$. From this vector, we may recover the vector $(\beta_1, \ldots, \beta_n)$ (just take pairwise differences), so the map $\beta \mapsto \mathrm{LT}(g_\beta(\sigma_1, \ldots, \sigma_n))$ is an injective map.

Among all $\beta$, let us pick the one with the largest $\mathrm{LT}(g_\beta(\sigma_1, \ldots, \sigma_n))$. Then for every other $\gamma$, we have $\mathrm{LT}(g_\beta(\sigma_1, \ldots, \sigma_n))$ strictly greater than every term of $g_\gamma(\sigma_1, \ldots, \sigma_n)$. In other words, there is nothing to cancel the leading term of $g_\beta$. If $\beta \neq 0$, this contradicts the assumption that $(g_1 - g_2)(\sigma_1, \ldots, \sigma_n)$ is the zero polynomial. It follows that $g_1 - g_2$ must be the zero polynomial in $k[y_1, \ldots, y_n]$. □

**Proposition 7.20.8.** *Fix a monomial order $>$ in the ring $k[x_1, \ldots, x_n, y_1, \ldots, y_n]$ so that every monomial involving any $x$ variable is greater than any monomial in $k[y_1, \ldots, y_n]$. Let $G$ be a Gröbner basis for $\langle \sigma_1(x_1, \ldots, x_n) - y_1, \ldots, \sigma_n(x_1, \ldots, x_n) - y_n \rangle$. Then*

1. *$f$ is symmetric if and only if $g \stackrel{\text{def}}{=} \overline{f}^G$ is an element of $k[y_1, \ldots, y_n]$,*
2. *If $f$ is symmetric then $f = g(\sigma_1, \ldots, \sigma_n)$ is the unique polynomial guaranteed by Theorem 7.20.7.*

We'll see more of this next time!

# 8 Practice Prelim 2

## 8.21 Nov 19, 2019

We discussed a "practice prelim" today. I'll give my solutions, although it's much easier to do this when I can cite theorems at will! Without further ado:

**Problem 1.**

(a) Define elementary symmetric polynomials.
(b) State the fundamental theorem of symmetric polynomials.
(c) Is $f = x^3y + x^3z + xy^3 + xz^3 + y^3z + yz^3 \in k[x, y, z]$ symmetric? If so, express it in terms of elementary symmetric polynomials.

**Problem 2.** Let $k$ be an arbitrary field. For $f = xy - 1$, $g = x^2y + y^2 - 4$ in $k[x, y]$, let $I = \langle f, g\rangle$. What are generators of $I_1 = I \cap k[y]$?

**Problem 3.** For any ideal $I \subseteq k[x_1, \ldots, x_n]$, prove that $\sqrt{I} \subseteq \mathbf{I}(\mathbf{V}(I))$.

**Problem 4.** State and prove Hilbert's Nullstellensatz.

**Problem 5.** State an algorithm for gcd computation of two multivariate polynomials.

**Problem 6.**

(a) Define irreducible varieties and prime ideals.
(b) Prove that if $\mathbf{I}(V)$ is prime, then the affine variety $V$ is irreducible.

*Proof of Problem 1.* For part a., see Definition 7.20.4.

For part b., see Theorem 7.20.7.

For part c., it is symmetric; I feel that the least painful way of checking is $f(x, y, z) = f(y, x, z) = f(y, z, x)$ and then applying (HW 9, Ex 7.1.1). To write this as a polynomial in the elementary symmetric polynomials, we use the proof of Theorem 7.20.7. We notice that

$$f - \sigma_1^2\sigma_2 \quad \text{has leading term} \quad -2x^2y^2,$$

and that

$$f - \sigma_1^2\sigma_2 + \sigma_2^2 \quad \text{has leading term} \quad -x^2yz,$$

and that

$$f - \sigma_1^2\sigma_2 + \sigma_2^2 + \sigma_1\sigma_3 = 0. \qquad \square$$

*Proof of Problem 2.* We use the Elimination Theorem (Theorem 6.15.5). We need to find a Gröbner basis $G$ of $I$ with respect to lex order, with $x > y$, and then take the elements of $G$ which are in $k[y]$. Using Buchberger's algorithm (Theorem 2.8.4) we obtain a Gröbner basis $G = \{x + y^2 - 4, y^3 - 4y + 1\}$ and hence $I \cap k[y] = \langle y^3 - 4y + 1\rangle$. $\square$

*Proof of Problem 3.* Let $f \in k[x_1, \ldots, x_n]$ be such that $f^N \in I$ (i.e., $f \in \sqrt{I}$). Then let $\mathbf{a} \in \mathbf{V}(I)$. We know $f^N(\mathbf{a}) = 0$, since $f^N \in I$. But then $f(\mathbf{a}) = 0$. This holds for every $\mathbf{a} \in \mathbf{V}(I)$; we've shown $f$ vanishes on $\mathbf{V}(I)$, i.e., $f \in \mathbf{I}(\mathbf{V}(I))$. $\square$

*Proof of Problem 4.* Hilbert's Nullstellensatz (Theorem 5.13.3) is the middle one (there are 3 Nullstellensatzes: there's the Weak (Theorem 5.13.1), the Hilbert (Theorem 5.13.3), and the Strong (Theorem 5.14.1)). □

*Proof of Problem 5.* We have $\operatorname{lcm}(f,g)\gcd(f,g) = fg$, so it suffices to give an algorithm to compute $\operatorname{lcm}(f,g)$. Now note that $\langle f\rangle \cap \langle g\rangle = \langle \operatorname{lcm}(f,g)\rangle$. But Theorem 6.15.1 says that $\langle f\rangle \cap \langle g\rangle = (t\langle f\rangle + (1-t)\langle g\rangle) \cap k[x_1,\ldots,x_n]$. Thus, our algorithm is as follows:

- Compute a Gröbner basis for $\langle tf, (1-t)g\rangle \subseteq k[t,x_1,\ldots,x_n]$ with respect to lex order with $t > x_1 > \cdots > x_n$.
- There is necessarily exactly one element of this Gröbner basis that is in $k[x_1,\ldots,x_n]$. This element is $\operatorname{lcm}(f,g)$.
- Compute $fg/\operatorname{lcm}(f,g)$.

Please excuse my lack of pseudocode. I guess I should learn that before the prelim...? □

*Proof of Problem 6.* For part (a), see Definitions 6.17.2 and 6.17.3.

For part (b), see Proposition 6.17.4. □

# 9 Invariant Theory of Finite Groups

## 9.22 Nov 26, 2019

We stated

**Proposition 9.22.1** (cf. Proposition 7.20.8)**.** *Fix a monomial order $>$ in the ring $k[x_1, \ldots, x_n, y_1, \ldots, y_n]$ so that every monomial involving any $x$ variable is greater than any monomial in $k[y_1, \ldots, y_n]$. Let $G$ be a Gröbner basis for $I = \langle \sigma_1(x_1, \ldots, x_n) - y_1, \ldots, \sigma_n(x_1, \ldots, x_n) - y_n \rangle$. Then*

1. *$f$ is symmetric if and only if $g \stackrel{\text{def}}{=} \overline{f}^G$ is an element of $k[y_1, \ldots, y_n]$,*
2. *If $f$ is symmetric then $f = g(\sigma_1, \ldots, \sigma_n)$ is the unique polynomial guaranteed by Theorem 7.20.7.*

*Proof.* Say $G = \{g_1, \ldots, g_t\}$, with $g_i \neq 0$ for $i \in [t]$. If $g = \overline{f}^G$, then we have

$$f = \sum_{i=1}^{t} A_i g_i + g, \tag{9}$$

where no term of $g$ is divisible by any $\mathrm{LT}(g_i)$. Let's begin by proving part 1 of the claim.

For the backwards direction, we suppose $g = \overline{f}^G$ is an element of $k[y_1, \ldots, y_n]$. When we substitute $y_i = \sigma_i$ for $i \in [n]$ in Equation (9) we get

$$f = \sum_{i=1}^{t} A_i(\mathbf{x}, \sigma) g_i(\mathbf{x}, \sigma) + g(\sigma);$$

here $\mathbf{x} = (x_1, \ldots, x_n)$ and $\sigma = (\sigma_1, \ldots, \sigma_n)$. Since $g_i \in I = \langle y_i - \sigma_i \rangle$, we have $g_i(\mathbf{x}, \sigma) = 0$, and we obtain $f = g(\sigma_1, \ldots, \sigma_n)$. We obtain that $f$ is symmetric.

For the forwards direction, we suppose $f$ is symmetric. The Fundamental Theorem of Symmetric Polynomials (Theorem 7.20.7) says that there exists a unique polynomial $g$ so that $f = g(\sigma_1, \ldots, \sigma_n)$. We claim that $g = \overline{f}^G$.

Let us observe that

$$\sigma_1^{\alpha_1} \ldots \sigma_n^{\alpha_n} = (y_1 + (\sigma_1 - y_1))^{\alpha_1} \ldots (y_n + (\sigma_n - y_n))^{\alpha_n} = y_1^{\alpha_1} \ldots y_n^{\alpha_n} + B_1(\mathbf{x}, \mathbf{y})(\sigma_1 - y_1) + \cdots + B_n(\mathbf{x}, \mathbf{y})(\sigma_n - y_n)$$

for $B_1, \ldots, B_n \in k[\mathbf{x}, \mathbf{y}]$. Thus, we may "replace" each monomial $\sigma^{\mathbf{a}}$ of $g(\sigma)$ with the monomial $\mathbf{y}^{\alpha}$, with the caveat that we introduce some multiples of $\sigma_i - y_i$ in the process. Specifically, we have

$$g(\sigma_1, \ldots, \sigma_n) = g(y_1, \ldots, y_n) + \sum_{i=1}^{n} C_i(\mathbf{x}, \mathbf{y})(\sigma_i - y_i)$$

for $C_i(\mathbf{x}, \mathbf{y}) \in k[\mathbf{x}, \mathbf{y}]$. In particular, we get

$$f = g(\sigma_1, \ldots, \sigma_n) = \underbrace{\sum C_i(\mathbf{x}, \mathbf{y})(\sigma_i - y_i)}_{\in I} + g(y_1, \ldots, y_n).$$

To get to our claim that $g = \overline{f}^G$, we need to show that no terms of $g(y_1, \ldots, y_n)$ are divisible by $\mathrm{LT}(g_i)$, $g_i \in G$.

Suppose that for some $i \in [t]$, we had $\mathrm{LT}(g_i)$ dividing some term of $g$. By assumption on the monomial order $>$, we see that $g_i \in k[y_1, \ldots, y_n]$ depends only on the $y$ variables. We also have $g_i \in I = \langle y_i - \sigma_i$, so substituting each $y_i$ with $\sigma_i$ gives $g_i(\sigma_1, \ldots, \sigma_n) = 0$. By the uniqueness part of the Fundamental Theorem of Symmetric Polynomials (Theorem 7.20.7), we get that $g_i$ is the zero polynomial. This is a contradiction, and we obtain $g = \overline{f}^G$. □

## 9.23 Dec 3, 2019

Let's state this proposition a third time.

**Proposition 9.23.1** (cf. Propositions 7.20.8 and 9.22.1)**.** *Fix a monomial order $>$ in the ring $k[x_1, \dots, x_n, y_1, \dots, y_n]$ so that every monomial involving any $x$ variable is greater than any monomial in $k[y_1, \dots, y_n]$. Let $G$ be a Gröbner basis for $I = \langle \sigma_1(x_1, \dots, x_n) - y_1, \dots, \sigma_n(x_1, \dots, x_n) - y_n \rangle$. Then*

1. *$f$ is symmetric if and only if $g \stackrel{\text{def}}{=} \overline{f}^G$ is an element of $k[y_1, \dots, y_n]$,*
2. *If $f$ is symmetric then $f = g(\sigma_1, \dots, \sigma_n)$ is the unique polynomial guaranteed by Theorem 7.20.7.*

Great.

**Definition 9.23.2.** Given variables $u_1, \dots, u_s$, we define the homogeneous symmetric polynomials

$$h_j(u_1, \dots, u_s) = \sum_{|\alpha|=j} \mathbf{u}^\alpha.$$

(These are not the elementary symmetric polynomials; for example $h_2(u_1, u_2) = u_1^2 + u_1 u_2 + u_2^2$.) △

**Proposition 9.23.3.** *Fix $>_{\text{lex}}$ on $k[x_1, \dots, x_n, y_1, \dots, y_n]$ with $x_1 >_{\text{lex}} \cdots >_{\text{lex}} x_n >_{\text{lex}} y_1 >_{\text{lex}} \cdots >_{\text{lex}} y_n$. Then, for each $j \in [n]$, the functions*

$$g_j \stackrel{\text{def}}{=} h_j(x_1, \dots, x_n) + \sum_{i=1}^{j} (-1)^i h_{j-i}(x_1, \dots, x_n) y_i$$

*form a Gröbner basis of $\langle \sigma_i - y_i \rangle$.*

*Proof.* Will be left as an exercise. □

**Definition 9.23.4.** Given variables $x_1, \dots, x_n$, we define the power sums $s_j = x_1^j + \cdots + x_n^j \in k[x_1, \dots, x_n]$. △

**Theorem 9.23.5.** *If $k$ is a field containing $\mathbb{Q}$, then every elementary symmetric polynomial in $k[x_1, \dots, x_n]$ can be written as a polynomial in the power sums.*

To see why we need $k \supseteq \mathbb{Q}$, consider the example $xy \in k[x, y]$. Then the expression

$$xy = \frac{(x+y)^2 - (x^2 + y^2)}{2} = \frac{s_1^2 - s_2}{2}$$

needs a $\frac{1}{2} \in k$. There are similar examples to see why we need the other $\frac{1}{p} \in k$ as well.

Let's talk about finite matrix groups and rings of invariants.

**Definition 9.23.6.** The set $\mathrm{GL}(n, k)$ will denote the set of all invertible $n \times n$ matrices with entries in the field $k$. △

Consider a matrix $A \in \mathrm{GL}(n, k)$. We can think of this as an invertible linear map $L_A \colon k^n \to k^n$ given by $\mathbf{v} \mapsto A\mathbf{v}$. In general, any invertible linear map arises in this way (this is why we call GL the "general linear group").

**Definition 9.23.7.** A finite subset $G \subseteq \mathrm{GL}(n, k)$ is called a finite matrix group if it is nonempty and is closed under matrix multiplication. We denote by $|G|$ the size of $G$. (The finiteness guarantees that $G$ contains an identity and inverse.) △

**Example 9.23.8.** The symmetric group $\mathfrak{S}_n$ is a finite matrix group. Indeed, associated to each permutation is a permutation matrix; for example we have

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_3 \\ x_1 \end{bmatrix}.$$

In general, the matrix sending $(x_1, \dots, x_n) \leftrightarrow (x_{\sigma(1)}, \dots, x_{\sigma(n)})$ is the matrix $M = (M_{ij})_{i,j \in [n]}$ with $M_{ij} = 1$ if $j = \sigma(i)$ and $M_{ij} = 0$ otherwise. △

**Example 9.23.9.** The cyclic groups are also finite matrix groups. For an integer $m$, we may take a matrix $A$ such that $A^m = I$ (and $m$ is the minimal power of $A$ so that $A^m = I$). Then $\{I, A, \dots, A^{m-1}\}$ is a matrix group. △

**Definition 9.23.10.** Let $G \subseteq \mathrm{GL}(n, k)$ be a finite matrix group. A polynomial $f \in k[\mathbf{x}]$ is invariant under $G$ if $f(\mathbf{x}) = f(A\mathbf{x})$ for all $A \in G$. The set is denoted $k[\mathbf{x}]^G$. △

**Example 9.23.11.** The set $k[\mathbf{x}]^{\mathfrak{S}_n}$ is the set of symmetric polynomials. Theorem 7.20.7 asserts that $k[\mathbf{x}]^{\mathfrak{S}_n} = k[\sigma_1, \dots, \sigma_n]$, and furthermore that there is a unique expression for each symmetric polynomial $f \in k[\mathbf{x}]^{\mathfrak{S}_n}$. △

We'll state two results which are easy exercises.

**Proposition 9.23.12.** *Let $G \subseteq \mathrm{GL}(n, k)$ be a finite matrix group. Then $k[\mathbf{x}]^G$ is closed under addition and multiplication of polynomials, and contains all the constant polynomials.*

[in fancy language, this says $k[\mathbf{x}]^G$ is a $k$-algebra.]

**Proposition 9.23.13.** *Let $G \subseteq \mathrm{GL}(n, k)$ be a finite matrix group. Then $f \in k[\mathbf{x}]^G$ if and only if the homogeneous components of $f$ are all in $k[\mathbf{x}]^G$.*

(A polynomial $f \in k[\mathbf{x}]$ is homogeneous of total degree $m$ if every term of $f$ has total degree $m$. Every polynomial can be written uniquely as a sum of homogeneous polynomials of different degrees; these are the homogeneous components of $f$.)

**Lemma 9.23.14.** *Let $G \subseteq \mathrm{GL}(n, k)$ be a finite matrix group. Suppose there exist $A_1, \dots, A_n \in G$ such that all $A \in G$ can be written as $A = B_1 \dots B_t$ for $B_i \in \{A_1, \dots, A_m\}$ [in fancy language, $G = \langle A_i \rangle$ is generated by $\{A_i\}$]. Then*

$$f \in k[\mathbf{x}]^G \quad \text{if and only if} \quad f(\mathbf{x}) = f(A_1\mathbf{x}) = \cdots = f(A_m\mathbf{x}).$$

(Note that for $G = \mathfrak{S}_3$, $A_1 = (2, 1, 3)$, and $A_2 = (2, 3, 1)$, this result is Exercise 7.1.1 in HW 9.)

*Proof.* The forward direction is trivial, so let us prove the backwards direction. Specifically, our claim is that if $f$ is invariant under $B_1, B_2, \dots, B_t$, then it is invariant under their product. We prove this by induction on $t$: for $t = 1$ this is obvious; assuming the result for $t - 1$ we see that

$$f(B_1 \dots B_t\mathbf{x}) = f((B_1 \dots B_{t-1})B_t\mathbf{x}) = f(B_t\mathbf{x}) = f(\mathbf{x}).$$ □

**Example 9.23.15** (Klein four-group)**.** Let's consider the finite matrix group

$$V_4 \stackrel{\text{def}}{=} \left\{ \begin{bmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{bmatrix} \right\} \subseteq \mathrm{GL}(2, k).$$

[$V$ stands for *Vierergruppe*, which means four-group.] Note that

$$A_1 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

generate the group $V_4$. Thus Lemma 9.23.14 says

$$f \in k[x,y]^{V_4} \quad \text{if and only if} \quad f(x,y) = f(-x,y) = f(x,-y).$$

If $f = \sum_{i,j} a_{i,j} x^i y^j$, the condition $f(x,y) = f(-x,y)$ says that $a_{i,j} = (-1)^i a_{i,j}$, i.e., for $i$ odd we have $a_{i,j} = 0$. Similarly $f(x,y) = f(x,-y)$ says that $a_{i,j} = 0$ for $j$ odd. In other words, if $a_{i,j} \neq 0$ then $x$ and $y$ both have an even power. We have shown that $f(x,y)$ can be uniquely written as $f(x,y) = g(x^2, y^2)$ for some $g \in k[x,y]$. In other words,

$$k[x,y]^{V_4} = k[x^2, y^2],$$

and every invariant in $k[x,y]^{V_4}$ can be written uniquely as a polynomial in the two homogeneous invariants $x^2, y^2$. △

**Example 9.23.16** (Cyclic group)**.** Let us consider the cyclic group

$$C_2 = \left\{ \pm \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\} \subseteq \mathrm{GL}(2,k).$$

Now Lemma 9.23.14 says that

$$f \in k[x,y]^{C_2} \quad \text{if and only if} \quad f(x,y) = f(-x,-y).$$

Then, for such an $f \in k[x,y]^{C_2}$, we have

$$\sum_{i,j} a_{i,j} x^i y^j = f(x,y) = f(-x,-y) = \sum_{i,j} (-1)^{i+j} a_{i,j} x^i y^j,$$

so we have $a_{i,j} = 0$ when $i + j$ is odd.

In other words, $f$ is invariant under $C_2$ if and only if whenever the monomial $x^i y^j$ appear with a nonzero coefficient in $f$, then either $i$ and $j$ are both even, or $i$ and $j$ are both odd. In the first case, then $x^i y^j = (x^2)^m (y^2)^n$ for some $m, n \in \mathbb{N}$, whereas in the second case we have $x^i y^j = (x^2)^m (y^2)^n xy$ for some $m, n \in \mathbb{N}$. We thus have $k[x,y]^{C_2} = k[x^2, xy, y^2]$, and the generating set $\{x^2, xy, y^2\}$ is *minimal* (in the sense that we can't get rid of any of them and get the entire $k[x,y]^{C_2}$ ring). Thus, $k[x,y]^{C_2}$ is generated by finitely many elements (namely, $x^2, xy, y^2$), but the expressions for polynomials in $k[x,y]^{C_2}$ in terms of the generators are not unique (since, for example, $(xy)^2 = (x^2) \cdot (y^2)$). △

Thus we have two important questions.

**Question 9.23.17** (Finite generation)**.** *Can we find finitely many invariants $f_1, \ldots, f_m$ such that every invariant is a polynomial in the $f_i$?*

**Question 9.23.18** (Uniqueness)**.** *In how many ways can we write an invariant in terms of the $f_1, \ldots, f_m$?*

## 9.24 Dec 5, 2019

We want to zoom towards a second proof of the Fundamental Theorem of Symmetric Polynomials (Theorem 7.20.7), leaving some loose ends along the way.

We've been thinking about generators for the ring of invariants $k[\mathbf{x}]^G$ for a finite matrix group $G$. We've been using the following notation a lot, so to be precise:

**Definition 9.24.1.** Given $f_1, \dots, f_m \in k[x_1, \dots, x_n]$, let $k[f_1, \dots, f_m]$ denote the subset of $k[x_1, \dots, x_n]$ consisting of polynomial expressions in these $f$'s with coefficients in $k$. In other words, it is the set

$$k[f_1, \dots, f_m] = \{g(f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})), \text{ where } g \in k[y_1, \dots, y_m]\} \subseteq k[\mathbf{x}]. \qquad \triangle$$

**Definition 9.24.2.** Let $G \subseteq \mathrm{GL}(n, k)$ be a finite matrix group. The Reynolds operator of $G$ is a map $R_G \colon k[x_1, \dots, x_n] \to k[x_1, \dots, x_n]$ sending $f \mapsto R_G(f)$, where $R_G(f)$ is the polynomial

$$(R_G(f))(\mathbf{x}) = \frac{1}{|G|} \sum_{A \in G} f(A\mathbf{x}). \qquad \triangle$$

(Note that it's important that $G$ is finite, because we want to divide by $|G|$.) [Also, when $k$ has characteristic $p$, there are some issues if $p$ divides $|G|$, i.e., if $k = \mathbb{F}_2$ then we don't want to take

$$G = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\} \subseteq \mathrm{GL}(n, k).$$

]

Here is why the Reynolds operator is important:

**Proposition 9.24.3.** *We have:*

1. *For all $f \in k[\mathbf{x}]$, we have $R_G(f) \in k[\mathbf{x}]^G$.*
2. *If $f \in k[\mathbf{x}]^G$, then $R_G(f) = f$.*

(So the Reynolds operator is like a "projection" onto the ring of invariants.)

*Proof of Proposition 9.24.3.* We'll use the fact that for all $B \in G$, we have an equality of sets $\{A\}_{A \in G} = \{AB\}_{A \in G}$. (In fancy language, this says right multiplication by $B$ is a bijection from a group to itself, and it's because the inverse is given by right multiplication by $B^{-1}$.)

Let's prove Proposition 9.24.3. To prove (1), we need to show that for all $B \in G$, that $(R_G(f))(B\mathbf{x}) = (R_G(f))(\mathbf{x})$. Well,

$$(R_G(f))(B\mathbf{x}) = \frac{1}{|G|} \sum_{A \in G} f(A(B\mathbf{x})) = \frac{1}{|G|} \sum_{A \in G} f(A\mathbf{x}) = (R_G(f))(\mathbf{x}),$$

where the second equality is the $\{AB\}_{A \in G} = \{A\}_{A \in G}$ fact that we assumed.

We also need to show that if $f \in k[\mathbf{x}]^G$ then $R_G(f) = f$. Well, for such an $f$, we have

$$(R_G(f))(\mathbf{x}) = \frac{1}{|G|} \sum_{A \in G} f(A\mathbf{x}) = \frac{1}{|G|} \sum_{A \in G} f(\mathbf{x}) = f(\mathbf{x}). \qquad \square$$

We now have the tools required to state the following Theorem, due to Emmy Noether:

**Theorem 9.24.4** (Noether). *Let $G \subseteq \mathrm{GL}(n,k)$ be a finite matrix group. Let $\mathbf{x}^{\beta_1}, \dots, \mathbf{x}^{\beta_n}$ be the monomials of total degree at most $|G|$ in $k[\mathbf{x}]$. Then*

$$k[\mathbf{x}]^G = k[R_G(\mathbf{x}^{\beta_1}), \dots, R_G(\mathbf{x}^{\beta_n})].$$

This gives an answer to the Finite Generation question (Question 9.23.17).

Let's talk about symmetric functions now. As a word of caution, the name "function" is a misnomer. They're not functions at all; in fact, they're formal power series. (Basically, there will be (infinitely many!) variables $x_i$ floating around, but these variables shouldn't be thought of as "things for which we can plug in an element of $k$" as with polynomials. The $x_i$ are just bookkeeping devices.)

It will take lots of time to build the theory rigorously, so please excuse a little bit of hand-waviness. From now on, $\mathbf{x} = (x_1, x_2, x_3, \dots)$ will denote infinitely many indeterminates.

**Definition 9.24.5.** A homogeneous symmetric function of degree $n$ over $\mathbb{Q}$ is a formal power series

$$f(\mathbf{x}) = \sum_{\alpha} c_\alpha \mathbf{x}^\alpha,$$

where:

(a) $\alpha$ ranges over all $\alpha = (\alpha_1, \alpha_2, \alpha_3, \dots)$ where each $\alpha_i \in \mathbb{Z}_{\geq 0}$ is a nonnegative integer and $\sum_i \alpha_i = n$.
(b) $c_\alpha \in \mathbb{Q}$. [To define the homogeneous symmetric function of degree $n$ over a field $k$, we change this condition to $c_\alpha \in k$. The other two conditions stay the same.]
(c) For any bijection $\sigma\colon \mathbb{N} \to \mathbb{N}$, we have $f(x_{\sigma(1)}, x_{\sigma(2)}, \dots) = f(x_1, x_2, \dots)$.

The set of homogeneous symmetric functions of degree $n$ over $\mathbb{Q}$ is denoted $\Lambda^n_{\mathbb{Q}}$. △

[Pedantic: For technical reasons, it may be useful to note that condition (c) may be replaced with "for any bijection $\sigma\colon \mathbb{N} \to \mathbb{N}$ that is eventually constant, that $f(x_{\sigma(1)}, x_{\sigma(2)}, \dots) = f(x_1, x_2, \dots)$". (Very pedantic: With this definition, the ring of homogeneous symmetric functions are the direct limit of $k[x_1, \dots, x_n]^{\mathfrak{S}_n} \hookrightarrow k[x_1, \dots, x_{n+1}]^{\mathfrak{S}_{n+1}}$ given by $f \mapsto R_{\mathfrak{S}_{n+1}}(f)$.)]

Note that if $f \in \Lambda^n_{\mathbb{Q}}$ and $g \in \Lambda^m_{\mathbb{Q}}$ then $fg \in \Lambda^{n+m}_{\mathbb{Q}}$.

**Definition 9.24.6.** A symmetric function over $\mathbb{Q}$ is an element of the vector space direct sum

$$\Lambda_{\mathbb{Q}} \stackrel{\text{def}}{=} \Lambda^0_{\mathbb{Q}} \oplus \Lambda^1_{\mathbb{Q}} \oplus \dots.$$

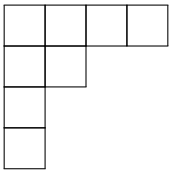
In other words, an element $f \in \Lambda_{\mathbb{Q}}$ is a sum $f = f_0 + f_1 + f_2 + \dots$, with each $f_i \in \Lambda^i_{\mathbb{Q}}$, such that *all but finitely many of the $f_i$ are zero*. △

So a symmetric function has bounded "degree", but may have infinitely many monomials.

A central theme of this business is to find various bases of the vector space $\Lambda^n_k$, and to understand the various transition matrices (=change-of-base matrices) between them.

**Definition 9.24.7.** A partition is a vector $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_m) \in \mathbb{Z}^m_{\geq 0}$ such that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$. If $\lambda_1 + \dots + \lambda_m = n$, we say that $\lambda$ is a partition of $n$, and we write $\lambda \vdash n$, and $|\lambda| = n$. △

**Example 9.24.8.** The vector $\lambda = (4,2,1,1)$ is a partition of $8$. The vector $\lambda = (4,2,1,1,0,0,0,0)$ is also a partition of $8$. We like to draw partitions by Young diagrams. In this case, the Young diagram of $\lambda$ is



(The $(4,2,1,1)$ correspond rows of length four, two, one, and one. They're left justified. The partitions $(4,2,1,1)$ and $(4,2,1,1,0,0,0,0)$ have the same Young diagram.) $\triangle$

We denote by $\text{Par}(n)$ the set of partitions of size $n$. Thus

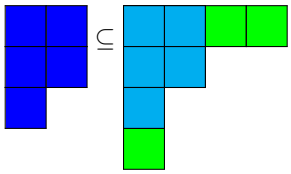
$$\begin{aligned}\text{Par}(1) &= \{(1)\},\\ \text{Par}(2) &= \{(2),(1,1)\},\\ \text{Par}(3) &= \{(3),(2,1),(1,1,1)\}.\end{aligned}$$

We denote by Par the set of all partitions, so

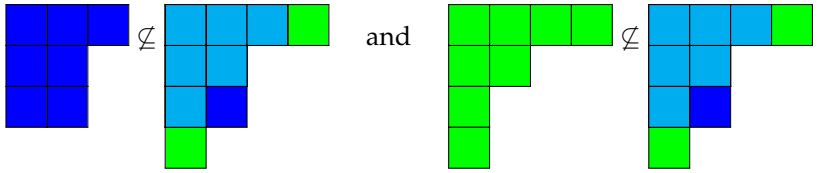
$$\text{Par} = \bigcup_{n\geq 1} \text{Par}(n).$$

**Definition 9.24.9.** There are two important partial orders on the set of partitions:

- Containment order: we say $\mu \subseteq \lambda$ if the Young diagram of $\mu$ is contained in the Young diagram of $\lambda$. For example, $\mu = (2,2,1,0) \subseteq (4,2,1,1) = \lambda$ because



  On the other hand, $\mu = (3,2,2,0)$ and $\lambda = (4,2,1,1)$ are incomparable, because



- Domination order: we say $\mu \leq \lambda$ if $|\mu| = |\lambda|$ and for every $i$ we have $\mu_1 + \cdots + \mu_i \leq \lambda_1 + \cdots + \lambda_i$. For example, $\mu = (3,2,2,1) \leq (4,2,1,1) = \lambda$. On the other hand, $\mu = (3,3,2,0)$ and $\lambda = (4,2,1,1)$ are incomparable. $\triangle$

Here is a **very awesome [but somewhat involved] exercise**.

**Exercise:** Let $\lambda = (\lambda_1, \ldots, \lambda_n) \vdash n$. Let $\mathcal{P}_\lambda$ be the convex hull in $\mathbb{R}^n$ of points obtained from $\lambda$ by permuting the coordinates, that is, let $\mathcal{P}_\lambda = \text{conv}(\mathfrak{S}_n \cdot \lambda)$. Show that for $\lambda, \mu \in \text{Par}(n)$, we have

$$\mu \leq \lambda \quad \text{ if and only if } \quad \mathcal{P}_\mu \subseteq \mathcal{P}_\lambda.$$

In fancy language, this is the statement that there exists an isomorphism of posets

$$(\text{Par}(n), \leq) \cong (\{\mathcal{P}_\lambda\}, \subseteq).$$

(Hint and solution)

Recall the central theme: we wanted to find bases for $\Lambda^n$ and understand the various transition matrices between them.

**Definition 9.24.10** (Monomial symmetric functions)**.** Given a partition $\lambda \vdash n$, define the symmetric function $m_\lambda(\mathbf{x}) \in \Lambda^n$ by

$$m_\lambda(\mathbf{x}) = \sum_\alpha \mathbf{x}^\alpha,$$

where $\alpha$ ranges over distinct permutations of $\lambda$. $\triangle$

**Example 9.24.11.** We have

$$\begin{aligned} m_{(0)} &= 1 \\ m_{(1)} &= \sum_{j=1}^{\infty} x_i \\ m_{(2)} &= \sum_{j=1}^{\infty} x_i^2 \\ m_{(1,1)} &= \sum_{i,j} x_i x_j. \end{aligned}$$

△

Here is a key observation.

**Observation 9.24.12.** Let

$$f = \sum_{\alpha} c_\alpha \mathbf{x}^\alpha \in \Lambda^n.$$

Let us consider $\alpha = (\alpha_1, \alpha_2, \dots)$ and $\beta = (\beta_1, \beta_2, \dots)$, and let us sort $\alpha$ and $\beta$ so that the components are weakly decreasing; that is, let $\text{sort}(\alpha) = (\alpha_{\sigma(1)}, \alpha_{\sigma(2)}, \dots)$ where $\alpha_{\sigma(1)} \geq \alpha_{\sigma(2)} \geq \dots$, and $\text{sort}(\beta) = (\beta_{\tau(1)}, \beta_{\tau(2)}, \dots)$ where $\beta_{\tau(1)} \geq \beta_{\tau(2)} \geq \dots$. Suppose $\text{sort}(\alpha) = \text{sort}(\beta)$. Then $c_\alpha = c_\beta$. Furthermore, for any $\alpha$, $\text{sort}(\alpha)$ is a partition, and $\text{sort}(\alpha) \vdash n$.

Since the coefficients $c_\alpha$ only depend on the partition $\text{sort}(\alpha)$, we can write

$$f = \sum_{\alpha} c_{\text{sort}(\alpha)} \mathbf{x}^\alpha = \sum_{\lambda \in \text{Par}(n)} \left( c_\lambda \sum_{\alpha \,:\, \text{sort}(\alpha) = \lambda} \mathbf{x}^\alpha \right) = \sum_{\lambda \in \text{Par}(n)} c_\lambda m_\lambda(\mathbf{x}).$$

△

We have observed that every $f \in \Lambda^n$ can be written as a $k$-linear combination of the $m_\lambda(\mathbf{x})$. In other words, we have observed that $m_\lambda$ is a basis of the vector space $\Lambda^n$. In particular, $\Lambda^n$ is a vector space of dimension $\#\text{Par}(n)$.

**Definition 9.24.13** (Elementary symmetric functions)**.** Given a partition $\lambda \vdash n$, define the elementary symmetric function $e_\lambda$ in the following way. For $n \in \mathbb{Z}_{\geq 0}$, let $\mathbf{1}_n$ denote the $n$-many-1's vector, i.e. $(1, 1, \dots, 1) \in \mathbb{Z}^n_{\geq 0}$.

$$e_n \stackrel{\text{def}}{=} m_{\mathbf{1}_n} = \sum_{1 \leq i_1 < \dots < i_n} x_{i_1} \dots x_{i_n}.$$

Then we define

$$e_\lambda = e_{(\lambda_1, \lambda_2, \dots)} \stackrel{\text{def}}{=} e_{\lambda_1} \cdot e_{\lambda_2} \cdot \dots.$$

△

If $A = (a_{i,j})_{i,j \geq 1}$ be a matrix with finitely many nonzero integer entries. Because only finitely many entries are nonzero, we may define row and column sums

$$r_i = \sum_j a_{i,j} \quad \text{and} \quad c_j = \sum_i a_{i,j}$$

and the vectors $\text{row}(A) = (r_1, r_2, \dots)$ and $\text{col}(A) = (c_1, c_2, \dots)$.

**Proposition 9.24.14.** *Let $\lambda \vdash n$ and let $(\alpha_1, \alpha_2, \dots)$ be a weak composition of $n$. Write*

$$e_\lambda = \sum_{\alpha} M_{\lambda\alpha} \mathbf{x}^\alpha,$$

*so that $M_{\lambda\alpha}$ is the coefficient of $\mathbf{x}^\alpha$ in $e_\lambda$. Then $M_{\lambda\alpha}$ is equal to the number of $(0,1)$-matrices $(A_{i,j})_{i,j \geq 1}$ so that $\text{row}(A) = \lambda$ and $\text{col}(A) = \mu$.*

Note that since $M_{\lambda\alpha}$ depends only on $\lambda$ and sort$(\alpha)$ (as in Observation 9.24.12), i.e. since we have

$$e_\lambda = \sum_{\mu \vdash n} M_{\lambda\mu} m_\mu,$$

this shows that the number of $(0,1)$-matrices with col$(A) = \alpha$ depends only on sort$(\alpha) = \mu$.

*Proof.* Let's think about the $\infty \times \infty$ matrix

$$X = \begin{bmatrix} x_1 & x_2 & x_3 & \dots \\ x_1 & x_2 & x_3 & \dots \\ x_1 & x_2 & x_3 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

and let's consider $e_\lambda$, where $\lambda = (\lambda_1, \lambda_2, \dots)$. So by definition

$$e_\lambda = e_{\lambda_1} \cdot e_{\lambda_2} \cdot \dots .$$

To obtain a term of $e_\lambda$, we choose terms of $e_{\lambda_i}$ and multiply them together. This is the same as choosing $\lambda_1$ distinct entries of the first row of $X$ (i.e., the product of these $\lambda_1$ entries is a term of $e_{\lambda_1}$), and choosing $\lambda_2$ distinct entries of the second row of $X$ (i.e., the product of these $\lambda_2$ entries is a term of $e_{\lambda_2}$), and so on.

Suppose we pick various entries of $X$, and the corresponding product of the entries we've chosen is $\mathbf{x}^\alpha$. We make the matrix $A = (A_{i,j})_{i,j \geq 1}$ whose entries are all zero, except when the $ij$-th entry of $X$ was picked, in which case we set $A_{i,j} = 1$. Then by construction row$(A) = \lambda$ and col$(A) = \alpha$. Conversely, any matrix $A$ determines a term of $e_\lambda$ that is equal to $\mathbf{x}^\alpha$. □

## 9.25 Dec 10, 2019

We're going to zoom through more stuff today. The goal is to excite you, and not to scare you. A very nice reference for this is Chapter 7 of Stanley's *Enumerative Combinatorics, Volume 2*.

**Proposition 9.25.1.** *Let $\lambda \vdash n$ be a partition. We may express the elementary symmetric function $e_\lambda$ (Definition 9.24.13) in terms of monomial symmetric functions (Definition 9.24.10) $m_\mu$ as*

$$e_\lambda = \sum_{\mu \vdash n} M_{\lambda\mu} m_\mu,$$

*where $M_{\lambda\nu}$ is the number of $\infty \times \infty$ $(0,1)$-matrices $A = (a_{ij})_{i,j \geq 1}$ with row vector $\text{row}(A) = \lambda$ and $\text{col}(A) = \mu$.*

The numbers $M_{\lambda\mu}$ has the following alternative combinatorial interpretation. Suppose we have $n$ balls, of which $\lambda_i$ of them are lebeled $i$. Suppose we also have (infinitely many) boxes. Then $M_{\lambda\mu}$ is the number of ways of placing the balls into our boxes such that:

1. No box contains more than one ball with the same label
2. Box $i$ contains exactly $\mu_i$ balls.

We'll leave the proof of this as a nice combinatorial exercise.

**Corollary 9.25.2.** *Let $M_{\lambda\mu}$ be given as above, i.e., write*

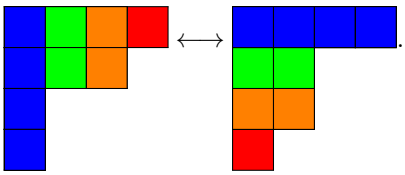
$$e_\lambda = \sum_{\mu \vdash n} M_{\lambda\mu} m_\mu.$$

*Then $M_{\lambda\mu} = M_{\mu\lambda}$*

*Proof.* For a $(0,1)$-matrix $A = (a_{ij})_{i,j \geq 1}$ with row vector $\text{row}(A) = \lambda$ and $\text{col}(A) = \mu$, the matrix $A^T$ is a $(0,1)$-matrix with row vector $\text{row}(A^T) = \mu$ and $\text{col}(A^T) = \lambda$. □

**Theorem 9.25.3.** *Let $\lambda, \mu \vdash n$. Then $M_{\lambda\mu} = 0$ unless $\mu \leq \lambda^t$ in dominance order, while $M_{\lambda\lambda^t} = 1$. Hence $\{e_\lambda \colon \lambda \vdash n\}$ is a basis of $\Lambda^n_{\mathbb{Q}}$, and $\{e_\lambda \colon \lambda \in \text{Par}\}$ is a basis of $\Lambda_{\mathbb{Q}}$. Equivalently, the $\{e_1, e_2, \dots\}$ are algebraically independent and generate $\Lambda_{\mathbb{Q}} = \mathbb{Q}[e_1, e_2, \dots]$.*

(Here $\lambda^t$ is the *conjugate* of $\lambda$: it is the partition whose Young diagram is the transpose of the Young diagram of $\lambda$. As an example, if $\lambda = (4,3,1,1)$ then $\lambda^t = (4,2,2,1)$:



Also, the *generate* in Theorem 9.25.3 is as a $\mathbb{Q}$-algebra. One can interpret the $\Lambda_{\mathbb{Q}} = \mathbb{Q}[e_1, e_2, \dots]$ as the infinite-variable version of the fundamental theorem of symmetric polynomials (Theorem 7.20.7).)

*Proof of Theorem 9.25.3.* Let $M_{\lambda\mu} = 0$. There exists a $(0,1)$-matrix $A = (a_{ij})_{i,j \geq 1}$ with $\text{row}(A) = \lambda$ and $\text{col}(A) = \mu$. Let $A'$ be the matrix obtained from $A$ by left-justifying all the 1's. Thus the 1's in the matrix $A'$ trace out the Young diagram for $\lambda$, i.e., we have $\text{row}(A') = \lambda$ and $\text{col}(A') = \lambda^t$. On the other hand, because we left justified the 1's, the number of 1's in the leftmost $i$ columns of $A'$ is at least the number of 1's in the leftmost $i$ columns of $A$. This is precisely the same as saying $\lambda^t = \text{col}(A') \geq \text{col}(A) = \mu$ in dominance order.

It turns out that the matrix $A'$ is the unique matrix with row sum $\lambda$ and column sum $\lambda^t$, so $M_{\lambda\lambda^t} = 1$.

To attack the claim about $\{e_\lambda : \lambda \vdash n\}$ being a basis, we assume the following exercise:

**Exercise:** There is a total order $<_{\text{TO}}$ on $\text{Par}(n)$ that is compatible with dominance order, so

$$\lambda^1 <_{\text{TO}} \lambda^2 <_{\text{TO}} \cdots <_{\text{TO}} \lambda^{\rho(n)} \quad \text{with} \quad \lambda^i <_{\text{dom}} \lambda^j \text{ implies } \lambda^i <_{\text{TO}} \lambda^j$$

such that the reverse conjugate order

$$(\lambda^{\rho(n)})^t <_{\text{TO}} (\lambda^{\rho(n)-1})^t <_{\text{TO}} < \cdots <_{\text{TO}} (\lambda^1)^t$$

is also compatible with the dominance order. Then our result that $M_{\lambda\mu} = 0$ unless $\mu \leq \lambda^t$ in dominance order precisely says that the matrix $M = (M_{\lambda\mu})$ with rows ordered $\lambda^1 <_{\text{TO}} \lambda^2 <_{\text{TO}} \cdots <_{\text{TO}} \lambda^{\rho(n)}$ and columns ordered $(\lambda^1)^t >_{\text{TO}} (\lambda^2)^t >_{\text{TO}} \cdots >_{\text{TO}} (\lambda^{\rho(n)})^t$ is upper triangular with 1's on the diagonal. Thus $M$ is invertible. Since

$$e_\lambda = \sum_{\mu \vdash n} M_{\lambda\mu} m_\mu,$$

it follows that invertible map $M$ maps the basis $m_\lambda$ of $\Lambda^n$ to the set $e_\lambda$. Thus $e_\lambda$ is a basis of $\Lambda^n$. It follows that $\Lambda_{\mathbb{Q}} = \mathbb{Q}[e_1, \ldots, e_n]$. □

There are various other bases for $\Lambda^n$ other than $m_\lambda$ and $e_\lambda$, which we'll briefly talk about now:

**Definition 9.25.4** (Complete homogeneous symmetric functions)**.** The complete homogeneous symmetric functions $h_\lambda$ are indexed by partitions $\lambda = (\lambda_1, \lambda_2, \ldots) \vdash n$. We first define, for $n \in \mathbb{N}$,

$$h_n \stackrel{\text{d}}{=} \sum_{\lambda \vdash n} m_\lambda.$$

Then as usual we set $h_\lambda = h_{\lambda_1} h_{\lambda_2} \cdots$. △

**Theorem 9.25.5.** *The $\{h_\lambda : \lambda \vdash n\}$ form a basis of $\Lambda^n$. Thus the $\{h_\lambda : \lambda \in \text{Par}\}$ form a basis of $\Lambda$. Equivalently, the $\{h_1, h_2, \ldots\}$ are algebraically independent and generate $\Lambda = \mathbb{Q}[h_1, h_2, \ldots]$.*

We know, from Theorem 9.25.3 that $\Lambda = \mathbb{Q}[e_1, \ldots, e_n]$. Note that to specify a map (a "ring homomorphism") $f : \Lambda \to \Lambda$ it is enough to specify the values of $f(e_i)$.

Let us define a map $\omega : \Lambda \to \Lambda$ by setting $\omega(e_n) = h_n$. (Thus $\omega(e_\lambda) = h_\lambda$.)

**Theorem 9.25.6.** *The map $\omega$ is an involution, that is, $\omega(h_\lambda) = e_\lambda$. In other words, $\omega$ is an isomorphism $\Lambda \to \Lambda$, with inverse equal to $\omega$.*

We won't prove Theorem 9.25.6, but this definitely proves Theorem 9.25.5, since $\omega$ is an automorphism.

Now another basis:

**Definition 9.25.7** (Power sum symmetric functions)**.** The power sum symmetric functions $p_\lambda$ are indexed by partitions $\lambda = (\lambda_1, \lambda_2, \ldots)$. As usual we first define, for $n \in \mathbb{N}$,

$$p_n = m_n = \sum_i x_i^n.$$

Then as usual we set $p_\lambda = p_{\lambda_1} p_{\lambda_2} \cdots$. △

And, as usual, we have

**Theorem 9.25.8.** *The $\{p_\lambda : \lambda \vdash n\}$ form a basis of $\Lambda^n$. Thus the $\{p_\lambda : \lambda \in \text{Par}\}$ form a basis of $\Lambda$. Equivalently, the $\{p_1, p_2, \ldots\}$ are algebraically independent and generate $\Lambda = \mathbb{Q}[p_1, p_2, \ldots]$.*

We will now move towards a very exciting family of polynomials that come up in both combinatorics and representation theory. First, some definitions:

**Definition 9.25.9.** Let $\lambda \vdash n$. Then a semistandard Young tableau $T$ of shape $\lambda$ is a filling of the Young diagram of $\lambda$ with positive integers such that the numbers on the rows are weakly increasing and the numbers on the columns are strictly increasing. To a Young tableau we may wish to associate the monomial $\mathbf{x}^T$, where the variable $x_i$ is raised to the number of times $i$ appears in $T$. △

**Example 9.25.10.** For example,

$$T = \begin{array}{|c|c|c|c|}\hline 1 & 1 & 10 & 13 \\ \hline 2 & 4 \\ \cline{1-2} 5 \\ \cline{1-1} 11 \\ \cline{1-1}\end{array}$$

is a semistandard Young tableau. In this case, the monomial $\mathbf{x}^T$ is $x_1^2x_2x_4x_5x_{10}x_{11}x_{13}$. △

To get to our exciting family of polynomials we first do an example:

Let us fix $\lambda = (2,1)$ and restrict to $x_1, x_2, x_3$. What are the semistandard Young tableau of shape $\lambda$ which only use the numbers $\{1,2,3\}$ to fill them up? Well, we have

$$\begin{array}{|c|c|}\hline 1 & 1 \\ \hline 2 \\ \cline{1-1}\end{array} \text{ and } \begin{array}{|c|c|}\hline 1 & 1 \\ \hline 3 \\ \cline{1-1}\end{array} \text{ and } \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 2 \\ \cline{1-1}\end{array} \text{ and } \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array} \text{ and } \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1}\end{array} \text{ and } \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 3 \\ \cline{1-1}\end{array} \text{ and } \begin{array}{|c|c|}\hline 2 & 2 \\ \hline 3 \\ \cline{1-1}\end{array} \text{ and } \begin{array}{|c|c|}\hline 2 & 3 \\ \hline 3 \\ \cline{1-1}\end{array}$$

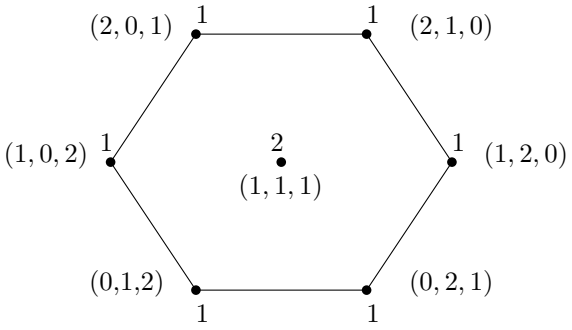
so there are 8 of them. These correspond to the monomials

$$x_1^2x_2 \quad \text{and} \quad x_1^2x_3 \quad \text{and} \quad x_1x_2^2 \quad \text{and} \quad x_1x_2x_3 \quad \text{and} \quad x_1x_2x_3 \quad \text{and} \quad x_1x_3^2 \quad \text{and} \quad x_2^2x_3 \quad \text{and} \quad x_2x_3^2$$

and when we add them up we get some polynomial which we'll call $s_{(2,1)}(x_1,x_2,x_3)$. To be explicit,

$$s_{(2,1)}(x_1,x_2,x_3) = x_1^2x_2 + x_1^2x_3 + x_1x_2^2 + 2x_1x_2x_3 + x_1x_3^2 + x_2^2x_3 + x_2x_3^2.$$

Notice, by the way, that the monomials appearing in $s_{(2,1)}(x_1,x_2,x_3)$, are precisely the integer points of the permutahedron $\mathcal{P}_{(2,1,0)} = \text{conv}(\mathfrak{S}_n \cdot (2,1,0))$. [As they say, "Hm."] For a picture, see below:




There was some discussion to connections to particle physics (the eightfold way). Although we don't understand anything, this (joint work of many superstars) might be relevant. [As they say, *"Hm."*]

**Definition 9.25.11.** The Schur function is

$$s_\lambda(x_1,x_2,\dots) = \sum_{T \in \text{SSYT of shape } \lambda} \mathbf{x}^T.$$ △

The following is extremely deep and beyond the scope of this course:

**Theorem 9.25.12.** *The Schur functions $\{s_\lambda : \lambda \vdash n\}$ form a basis of $\Lambda^n$.*

We will end the course with:

**Theorem 9.25.13.** *The Schur function $s_\lambda$ is a symmetric function.*

*Proof.* It suffices to show that $s_\lambda$ is invariant under the switches $x_i \leftrightarrow x_{i+1}$ (this uses the fact that condition (c) in Definition 9.24.5 can be replaced with "bijection that is eventually constant"). Specifically, for a weak composition $\alpha = (\alpha_1, \alpha_2, \dots)$ of $n$, let us define $T_{\lambda,\alpha}$ as the set of semistandard Young tableaux of shape $\lambda$ and $\alpha_j$ many $j$'s in its filling. Define

$$\widetilde{\alpha} = (\alpha_1, \dots, \alpha_{i-1}, \alpha_{i+1}, \alpha_i, \alpha_{i+2}, \dots)$$

the weak composition obtained from $\alpha$ by switching the $i$ and $i+1$-th coordinates. We want to find a bijection showing $|T_{\lambda,\alpha}| = |T_{\lambda,\widetilde{\alpha}}|$.

To do this, let us consider a semistandard Young tableau of shape $\lambda$ and $\alpha_j$ many $j$'s in its filling. We want a way to turn this into a semistandard Young tableau of shape $\lambda$ and $\alpha_{i+1}$ many $i$'s, $\alpha_i$ many $i+1$'s, and $\alpha_j$ many $j$'s for other $j$. Notice that the columns of a semistandard Young tableau are strictly increasing, so there can only be at most one $i$ and at most one $i+1$. If there are no $i$'s or $i+1$'s, or there is one $i$ and one $i+1$, then we leave those columns alone. Let us focus on the columns for which exactly one of $\{i, i+1\}$ is present. For these $i$'s and $i+1$'s, we are free to replace $i$'s with $i+1$'s or $i+1$'s with $i$'s since the columns are strictly increasing. It's "readily" seen that this is a bijection (i.e., this will be left as an exercise). :) □