MATH 371 – Class notes/outline – November 5, 2013

Intro to Gröbner bases

We return to the study of polynomial rings with several variables over the fields $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$, and their relationship with the (algebraic) geometry of $\mathbb{Q}^n$, $\mathbb{R}^n$ and $\mathbb{C}^n$. We'll use the letter k to denote one of these fields (many things we do here will be applicable to other fields as well).

Up to now, we know a lot about the ring $k[x]$ — owing mostly to the division algorithm, which relies on the fact that $k[x]$ is a Euclidean domain. Unfortunately, although $k[x_1, \ldots, x_n]$ is a unique factorization domain for $n \geqslant 2$, it is not a Euclidean domain, and we have remarked on the apparent difficulty of the following questions:

1. *Ideal description*: Not all ideals in $k[x_1, \ldots, x_n]$ are principal. But can every ideal $I \subset k[x_1, \ldots, x_n]$ be written as $\langle f_1, \ldots, f_s \rangle$ for a *finite* collection of polynomials $f_1, \ldots, f_s \in k[x_1, \ldots, x_n]$?

2. *Ideal membership*: Given an ideal $I \subset k[x_1, \ldots, x_n]$ of the form $\langle f_1, \ldots, f_s \rangle$ and another polynomial $f \in k[x_1, \ldots, x_n]$, determine whether $f \in I$.

3. *Solving polynomial equations*: Find (all of the) solutions in k^n of the system of polynomial equations
$$f(x_1, \ldots, x_n) = \cdots = f_s(x_1, \ldots, x_n) = 0.$$
Another way to say this is to determine the points on the affine variety $\mathbf{V}(f_1, \ldots, f_s)$.

4. *Implicitization* (or *de-parametrization*): Suppose $S \subset k^n$ is given parametrically as
$$\begin{aligned} x_1 &= g_1(t_1, \ldots, t_\ell) \\ x_2 &= g_2(t_1, \ldots, t_\ell) \\ \vdots\ & \quad \vdots \quad \vdots \\ x_n &= g_n(t_1, \ldots, t_\ell) \end{aligned}$$
where the g_i are polynomials or rational functions. Then S is an affine variety (or at least part of one). Find the system of polynomial equations that defines this variety.

Of course, we know the answers to questions 1 and 2 in the single-variable case — every ideal in $k[x]$ is principal, and the division algorithm takes care of question 2: if $I = \langle g \rangle$ and we express f as $f = qg + r$ with $\deg(r) < \deg(g)$, then $f \in I$ if and only if $r = 0$.

A situation where we can answer questions 3 and 4 satisfactorily is when all the polynomials in question are *linear*. For instance, consider the system of linear equations:

$$\begin{array}{rcrcrcr} 2x_1 & + & x_2 & + & x_3 & = & 1 \\ x_1 & + & 2x_2 & - & x_3 & = & -4 \\ x_1 & - & x_2 & + & 2x_3 & = & 5 \end{array}$$

The row-reduced form of the matrix of the system is:

$$\left[\begin{array}{rrr|r} 1 & 0 & 1 & 2 \\ 0 & 1 & -1 & -3 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The usual drill from linear algebra gives us a parametrization of the line (one-dimensional affine variety) of solutions of the equations as follows:

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

We can also consider the ideal of this variety within $k[x_1, x_2, x_3]$, namely $I = \langle 2x_2 + x_2 + x_3 - 1\,,\ x_1 + 2x_2 - x_3 + 4\,,\ x_1 - x_2 + 2x_3 - 5\rangle$. This ideal can also be expressed as $I = \langle x_1 + x_3 - 2\,,\ x_2 - x_3 + 3\rangle$. This latter way of expressing the ideal will be much more convenient for determining whether a polynomial $f \in k[x_1, \ldots, x_n]$ is in I, since we would be able to consider "one variable at a time", i.e., first x_1 and then x_2. We'll have much more to say about this later.

An example for question 4 would be to reverse the process for this example, i.e, to start with the parametrization of the solutions and then work back to a set of equations in x_1, x_2 and x_3 as follows: Rewrite the parametrization equations by putting the t terms *first* on the left side:

$$\begin{array}{ccccccccc} t & + & x_1 & & & & & = & 2 \\ t & & & - & x_2 & & & = & 3 \\ t & & & & & - & x_3 & = & 0 \end{array}$$

Now do row reduction and get the matrix

$$\left[\begin{array}{rrrr|r} 1 & 1 & 0 & 0 & 2 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 & -3 \end{array}\right].$$

Neither of the last two rows has a non-zero entry in the t column, and the equations they represent, namely

$$\begin{array}{ccccccc} x_1 & + & x_2 & & & = & -1 \\ & & x_2 & - & x_3 & = & -3 \end{array}$$

are equivalent to the original system.

Notice that in the single-variable case, we rely on the division algorithm, and in the linear multivariable case the row-reduction procedure relies on keeping careful track of the how the terms are ordered (and we are careful to write the terms in the same order in every step as we solve a problem). The method of Gröbner bases combines these two elements to handle general systems of polynomials in several variables.

Order: In the one-variable polynomial ring $k[x]$, there is a natural order to the terms that we use for long division in algebra — namely, we order the terms from highest to lowest degree (in calculus, we usually do it the other way around). To indicate this, we'll use ">" to mean "comes before" and write $\cdots > x^{m+1} > x^m > \cdots > x^2 > x > 1$.

In the ring $k[x_1, \ldots, x_n]$ for $n \geqslant 2$, there is no obvious notion of order for the terms, and indeed it will turn out that there are many (in fact an infinite number of) possible orderings. To begin, recall our multi-index notation for monomials in several variables: we write $\alpha = (\alpha_1, \dots, \alpha_n)$ for an n-tuple of non-negative integers (so $\alpha \in \mathbb{Z}^n_{\geqslant 0}$) and when we write x^α we mean $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$. We want to define a "comes before" ordering $>$ on $\mathbb{Z}^n_{\geqslant 0}$ that will help us concoct a useful long division algorithm, and we'll need it to have three basic properties:

1. $>$ should be a total or linear ordering on $\mathbb{Z}^n_{\geqslant 0}$, that is, for every $\alpha, \beta \in \mathbb{Z}^n_{\geqslant 0}$ exactly one of $\alpha > \beta$, $\beta > \alpha$ or $\alpha = \beta$ is true.

2. $>$ should respect multiplication, that is, if $\alpha > \beta$ and $\gamma \in \mathbb{Z}^n_{\geqslant 0}$ then $\alpha + \gamma > \beta + \gamma$.

3. $>$ should be a well-ordering on $\mathbb{Z}^n_{\geqslant 0}$, that is, every nonempty subset of $\mathbb{Z}^n_{\geqslant 0}$ has a smallest element.

A relation $>$ that satisfies these three properties will be called a *monomial ordering.*

Note that the condition that $>$ is a well-ordering is equivalent to the fact that any decreasing sequence of multi-indices $\alpha^{(1)} > \alpha^{(2)} > \cdots$ must terminate (or stabilize).

Example: Perhaps the simplest example of a monomial ordering is the *lexicographic ordering*, or **lex** for short. It is basically the "dictionary order", or more precisely, if $\alpha, \beta \in \mathbb{Z}^n_{\geqslant 0}$ then $\alpha >_{lex} \beta$ if, in the vector difference $\alpha - \beta$, the leftmost nonzero entry is positive. We'll also write $x^\alpha >_{lex} x^\beta$ if $\alpha >_{lex} \beta$.

For instance $(1, 2, 0) >_{lex} (0, 3, 4)$ and $(3, 2, 4) >_{lex} (3, 2, 1)$, so we also have $x_1 x_2^2 >_{lex} x_2^3 x_3^4$ and $x_1^3 x_2^2 x_3^4 >_{lex} x_1^3 x_2^2 x_3$. It is straightforward to show that $>_{lex}$ is a monomial ordering. Note that the lexicographic ordering depends on an ordering of the variables, so we can create other (in fact $n!$) different lexicographic orderings by deciding to write the variables in a different order.

One peculiarity of the **lex** ordering is that it doesn't take the total degree of monomials into account, so in the first example above we had a term of total degree 3 coming before a term of total degree 7. For some purposes this is fine, but we frequently want to write our terms in decreasing order of total degree. Doing this will give us the *graded lexicographic order*, or **grlex**. It is defined as follows: If $\alpha, \beta \in \mathbb{Z}^n_{\geqslant 0}$ then $\alpha >_{grlex} \beta$ if $|\alpha| > |\beta$, or if $|\alpha| = |\beta|$ and $\alpha >_{lex} \beta$.

A few definitions: We need a little bit of jargon to describe the multivariable division algorithm and some of its consequences, so here it is: Let $f = \sum_\alpha a_\alpha x^\alpha$ be a nonzero polynomial in $k[x_1, \dots, x_n]$ and let $>$ be a monomial order.

- The *multidegree* of f is $\text{multideg}(f) = \max\{\alpha \in \mathbb{Z}^n_{\geqslant 0} \,|\, a_\alpha \neq 0\}$ where the maximum is taken with respect to the monomial ordering $>$ – in other words multideg(f) is the multi-index which gives the exponents in the "first term" of f (with respect to $>$).

- The *leading coefficient* of f is $\text{LC}(f) = a_{\text{multideg}(f)} \in k$.

- The *leading monomial* of f is $\text{LM}(f) = x^{\text{multideg}(f)}$.

- The *leading term* of f is $\text{LT}(f) = \text{LC}(f) \cdot \text{LM}(f)$.

It is easy to see that if $f, g \in k[x_1, \ldots, x_n]$ are nonzero, then $\text{multideg}(fg) = \text{multideg}(f) + \text{multideg}(g)$ and if $f + g \neq 0$ then $\text{multideg}(f + g) \leqslant \max(\text{multideg}(f), \text{multideg}(g))$ (and equality occurs if $\text{multideg}(f) \neq \text{multideg}(g)$).

The division algorithm. The division algorithm in $k[x_1, \ldots, x_n]$ reflects the fact that not all ideals in $k[x_1, \ldots, x_n]$ are principal, and the fact that an important use of the division algorithm in one variable, where we were given f and d and expressed f as $f = qd + r$ where $\deg(r) < \deg(d)$, was to provide standard representatives of elements of $k[x]/\langle d\rangle$, namely that $[f] = r + \langle d\rangle$. Since ideals in $k[x_1, \ldots, x_n]$ are generated by more than one polynomial, for instance $\langle d_1, d_2, \ldots, d_s\rangle$, we'll try to express a general polynomial $f \in k[x_1, \ldots, x_n]$ as

$$f = q_1 d_1 + q_2 d_2 + \cdots + q_s d_s + r$$

where something about r is "less" than the corresponding thing about any of the d_i's. This will require us to work with a fixed monomial ordering $>$.

Proposition: Fix a monomial ordering $>$ on $\mathbb{Z}^n_{\geqslant 0}$. Let $f \in k[x_1, \ldots, x_n]$ with $f \neq 0$ and suppose that $(d_1, d_2, \ldots, d_s)$ is a sequence of non-zero polynomials in $k[x_1, \ldots, x_n]$. Then there exist $q_1, \ldots, q_s, r \in k[x_1, \ldots, x_n]$ such that

$$f = q_1 d_1 + q_2 d_2 + \cdots + q_s d_s + r$$

and either $r = 0$ or none of the terms in r is divisible by any of $\text{LT}(d_1)$, $\text{LT}(d_2)$,...,$\text{LT}(d_s)$. Furthermore, if $q_i d_i \neq 0$ then $\text{multideg}(f) \geqslant \text{multideg}(q_i d_i)$.

Proof: Here is the algorithm: To begin, set $q_1, \ldots, q_s = 0$, $r = 0$ and $g = f$, so that we have

$$f = q_1 d_1 + q_2 d_2 + \cdots + q_s d_s + (r + g).$$

This expression will be invariant throughout the algorithm, even though the constituent parts will change. Here is the iterative step of the algorithm: If $g = 0$ then we are done. Otherwise,

- If $\text{LT}(g)$ is divisible by some $\text{LT}(d_i)$ then pick the *smallest* i with this property and let

  $$g = g - \frac{\text{LT}(g)}{\text{LT}(d_i)} g_i \quad \text{and} \quad q_i = q_i + \frac{\text{LT}(g)}{\text{LT}(d_i)}.$$

  Note that the above equation for f still holds after these assignments since we have simply added something to the $q_i d_i$ term and subtracted the same thing from the g term.

- On the other hand, if $\text{LT}(g)$ is not divisible by any of the $\text{LT}(d_i)$ then we add the initial term of g to r and subtract it from g:

  $$r = r + \text{LT}(g) \quad \text{and} \quad g = g - \text{LT}(g).$$

Clearly after these assignments $r + g$ is unchanged and the above equality still holds. Now repeat — if $q = 0$ we are done and otherwise, the multidegree of the initial term of g has strictly decreased. Since $>$ is a well-ordering, we know the algorithm must terminate and all the other conclusions of the proposition are satisfied.

Examples: For our first example, we use the **lex** ordering and divide $xy^2 + 1$ by the pair $(xy + 1\,,\, y + 1)$

$$
\begin{array}{rr@{\quad}c@{\quad}l}
q_1: & y\phantom{{}+1} & & \\
q_2: & -1\phantom{{}+1} & & r \\
\cline{2-2} \cline{4-4}
xy+1 & |\, xy^2+1 & & \\
y+1 & | & & \\
 & xy^2+y & & \\
\cline{2-2}
 & -y+1 & & \\
 & -y-1 & & \\
\cline{2-2}
 & 2 & & \\
\cline{2-2}
 & 0 & \rightarrow & 2
\end{array}
$$

This allows us to conclude that $xy^2 + 1 = y(xy + 1) + (-1)(y + 1) + 2$.

A somewhat more complicated example, which illustrates how the remainder can be accumulated over the course of several steps, is to divide $x^2y + xy^2 + y^2$ by the pair $(xy - 1\,,\, y^2 - 1)$ (we continue to use the **lex** ordering):

$$
\begin{array}{rr@{\quad}c@{\quad}l}
q_1: & x+y\phantom{{}+xy^2+y^2} & & \\
q_2: & 1\phantom{{}+xy^2+y^2} & & r \\
\cline{2-2} \cline{4-4}
xy-1 & |\, x^2y+xy^2+y^2 & & \\
y^2-1 & | & & \\
 & x^2y-x\phantom{{}+y^2} & & \\
\cline{2-2}
 & xy^2+x+y^2 & & \\
 & xy^2-y\phantom{{}+y^2} & & \\
\cline{2-2}
 & x+y^2+y & & \\
\cline{2-2}
 & y^2+y & \rightarrow & x \\
 & y^2-1 & & \\
\cline{2-2}
 & y+1 & & \\
\cline{2-2}
 & 0 & \rightarrow & x+y+1
\end{array}
$$

This shows that $x^2y + xy^2 + y^2 = (x + y)(xy - 1) + (1)(y^2 - 1) + (x + y + 1)$.

Next, we repeat this example, but with the divisors reversed. So we're dividing $x^2y + xy^2 + y^2$

by the pair $(y^2-1\,,\,xy-1)$

$$
\begin{array}{rrcl}
q_1: & x+1 & & \\
q_2: & x & & r \\
\cline{2-2}\cline{4-4}
y^2-1 & |\,x^2y+xy^2+y^2 & & \\
xy-1 & | \qquad\qquad\qquad & & \\
 & x^2y-x \qquad\quad & & \\
\cline{2-2}
 & xy^2+x+y^2 & & \\
 & xy^2-x \qquad & & \\
\cline{2-2}
 & 2x+y^2 & & \\
\cline{2-2}
 & y^2 & \rightarrow & 2x \\
 & y^2-1 & & \\
\cline{2-2}
 & 1 & & \\
\cline{2-2}
 & 0 & \rightarrow & 2x+1
\end{array}
$$

This shows that $x^2y+xy^2+y^2=(x+1)(y^2-1)+(x)(xy-1)+(2x+1)$. It is perhaps a little disconcerting that the remainder in this example is different from the previous one.

This phenomenon of different remainders dependent on the order of the divisors can even result in a zero remainder for one order and non-zero for another. For instance, if we divide xy^2-x by the pair $(xy+1\,,\,y^2-1)$, we get

$$
\begin{array}{rrcl}
q_1: & y \qquad\quad & & \\
q_2: & & & r \\
\cline{2-2}\cline{4-4}
xy+1 & |\,xy^2-x & & \\
y^2-1 & | \qquad\quad & & \\
 & xy^2+y & & \\
\cline{2-2}
 & -x-y & & \\
\cline{2-2}
 & 0 & \rightarrow & -x-y
\end{array}
$$

and so $xy^2-x=y(xy+1)+0(y^2-1)+(-x-y)$. On the other hand, if we divide xy^2-x by $(y^2-1\,,\,xy+1)$ we get

$$
\begin{array}{rrcl}
q_1: & x \qquad\quad & & \\
q_2: & & & r \\
\cline{2-2}\cline{4-4}
y^2-1 & |\,xy^2-x & & \\
xy+1 & | \qquad\quad & & \\
 & xy^2-x & & \\
\cline{2-2}
 & 0 & &
\end{array}
$$

and so $xy^2-x=x(y^2-1)+0(xy+1)+0$. This last division shows that $xy^2-x\in\langle y^2-1\,,\,xy+1\rangle$, where the previous division did not. This phenomenon of non-uniqueness of the remainder will go away if we use Gröbner bases for our "divisor ideal", as we shall see.

Ideals generated by monomials: We call $I \subset k[x_1, \ldots, x_n]$ an *ideal generated by monomials* (some authors call these "monomial ideals", which is a bit shorter but also somewhat misleading) if there is a subset $A \subset \mathbb{Z}^n_{\geqslant 0}$ such that I consists of all polynomials which are *finite* sums of the form $\sum_{\alpha \in A} p_\alpha x^\alpha$ where $p_\alpha \in k[x_1, \ldots, x_n]$ (and all but finitely many of the p_α are zero in case the set A is inifinite). We will write $I = \langle x^\alpha \rangle_{\alpha \in A}$.

Clearly, if $I = \langle x^\alpha \rangle_{\alpha \in A}$ is an ideal generated by monomials, then a monomial x^β is in I if and only if it is divisible by x^α for some $\alpha \in A$. And x^β is divisible by x^α if and only if $\beta = \alpha + \gamma$ for some $\gamma \in \mathbb{Z}^n_{\geqslant 0}$. So the set of monomials divisible by a given x^α is the set of points with integer coordinates in the translated version of the positive orthant of $\mathbb{R}^n$ (translated so that the origin moves to the point α).

More generally, if $I = \langle x^\alpha \rangle_{\alpha \in A}$ is an ideal generated by monomials, then a polynomial $p \in k[x_1, \ldots, x_n]$ is in I if and only if each term of f is in I, so two ideals generated by monomials are equal if and only if they contain the same monomials (this is not to say that their generating sets are the same). The main result about these ideals, which is a precursor to Hilbert's basis theorem, is the following:

Dickson's Lemma: Let $I = \langle x^\alpha \rangle_{\alpha \in A}$ be an ideal generated by monomials. Then I has a finite basis (generating set). In particular, $I = \langle x^{\alpha_1}, x^{\alpha_2}, \ldots, x^{\alpha_s} \rangle$, where $\alpha_1, \ldots, \alpha_s \in A$.

Idea of proof. Induct on the number of variables n in $k[x_1, \ldots, x_n]$. If $n = 1$ the result is obvious. So assume the result for $n - 1$ and suppose $I = \langle x^\alpha \rangle_{\alpha \in A} \subset k[x_1, \ldots, x_n]$. Define the projection map $\pi \colon \mathbb{Z}^n_{\geqslant 0} \to \mathbb{Z}^{n-1}_{\geqslant 0}$ vis $\pi(a_1, a_1, \ldots, a_n) = (a_2, a_3, \ldots, a_n)$, and let J be the ideal in $k[x_2, \ldots, x_n]$ generated by the monomials x^β for which $\beta = \pi\alpha$ for some $x^\alpha \in I$. By induction, J is generated by finitely many monomials of the form $x^{\pi(\alpha)}$ with $\alpha \in A$, so assume J is generated by $x^{\pi(\alpha_1)}, x^{\pi(\alpha_2)}, \ldots, x^{\pi(\alpha_s)}$. For each $i = 1, \ldots, s$, there is a non-negative integer k_i such that $\alpha_i = (k_i, \pi(\alpha_i))$ — let K be the largest of these (finitely many) integers.

Next, for each $k = 0, \ldots, K$ let J_k be the ideal in $k[x_2, \ldots, x_n]$ generated by monomials x^β such that $x_1^k x^\beta \in I$. Again, J_k has a finite generating set, say $J_k = \langle x^{\beta_{k,1}}, \ldots, x^{\beta_{k,s_k}} \rangle$. It is now true that I is generated by the totality of generators of $J, J_0, J_1, \ldots, J_K$. And by the observation above, each monomial in this generating set is divisible by some monomial in A, which yields the desired finite generating set coming from $\{x^\alpha \mid \alpha \in A\}$.

One consequence of Dickson's lemma is a simplification of the definition of monomial ordering — rather than insisting on the relation $>$ being a well-ordering, it is enough to insist that $\alpha > 0$ for all $\alpha \neq 0$ in $\mathbb{Z}^n_{\geqslant 0}$. This is because if we are given an ordering that satisfies the first two conditions of the definition (total ordering, respects multiplication) for which $\alpha > 0$ for all $\alpha \neq 0$ in $\mathbb{Z}^n_{\geqslant 0}$, and any subset $A \subset \mathbb{Z}^n_{\geqslant 0}$, we can form the ideal $I = \langle x^\alpha \rangle_{\alpha \in A}$. By Dickson's Lemma, I has a finite generating set, and the multidegree of the smallest element in the generating set will be the smallest element in A.

Ideals of leading terms and the Hilbert Basis Theorem: Let $I \subset k[x_1, \ldots, x_n]$ be an ideal. Write $\mathrm{LT}(I)$ for the set of leading terms of elements of I, so $cx^\alpha \in \mathrm{LT}(I)$ if and only if there exists $f \in I$ with $\mathrm{LT}(f) = cx^\alpha$. Then $\langle \mathrm{LT}(I) \rangle$ is the ideal generated by the leading terms of elements of I

Note that if $I = \langle p_1, \ldots, p_s \rangle$ then necessarily $\langle \mathrm{LT}(p_1), \ldots, \mathrm{LT}(p_s) \rangle \subset \langle \mathrm{LT}(I) \rangle$, but the opposite inclusion need not be true. For example (use $>_{lex}$ to order the terms) $x^2 \in \langle x^3 + y\,,\, x^2y + y^2 + x\rangle$ since $x^2 = x(x^2y + y^2 + x) - y(x^3 + y)$ but $x^2 \notin \langle x^2y, x^3\rangle$.

Clearly, $\langle \mathrm{LT}(I)\rangle$ is an ideal generated by monomials, so by Dickson's Lemma it has a finite generating set. In particular, there are polynomials $g_1, \ldots, g_s \in I$ such that $\langle \mathrm{LT}(I)\rangle = \langle \mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_s)\rangle$. Armed with this observation and the division algorithm, we can prove that *every* ideal $I \subset k[x_1, \ldots, x_n]$ has a finite basis:

Hilbert's Basis Theorem: Every ideal $I \subset k[x_1, \ldots, x_n]$ has a finite generating set; in other words, $I = \langle g_1, \ldots, g_s\rangle$ for some choice of $g_1, \ldots, g_s \in I$.

Proof: Claim that a set $g_1, \ldots, g_s$ such that $\langle \mathrm{LT}(I)\rangle = \langle \mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_s)\rangle$ is a finite generating set for I. To see this, choose $f \in I$ and divide f by $(g_1, \ldots, g_s)$ using the division algorithm (after picking some monomial ordering $>$). The result is an expression of the form

$$f = q_1g_1 + \cdots + q_sg_s + r$$

where no term of r is divisible by any of the $\mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_s)$. But it is clear that $r \in I$ so that if $r \neq 0$ we would have to have $\mathrm{LT}(r) \in \langle \mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_s)\rangle$, which is impossible.

Definition: Given an ideal $I \subset k[x_1, \ldots, x_n]$, a (finite) subset $\{g_1, \ldots, g_s\} \subset I$ which satisfies

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

is called a *Gröbner basis* (or sometimes a *standard basis*) for I

Gröbner bases were invented (discovered?) in 1965 by Buchberger and named by him in honor of his thesis adviser. Independently, Hironaka developed the idea of standard bases for ideals in power series rings in 1964.

An immediate application of Gröbner bases (or of the Hilbert Basis Theorem) is to show that the polynomial ring $k[x_1, \ldots, x_n]$ satisfies the *ascending chain condition* (ACC): if $I_1 \subset I_2 \subset I_3 \subset \cdots$ is an ascending chain of ideals in $k[x_1, \ldots, x_n]$, then there exists an $N \geqslant 1$ such that $I_N = I_{N+1} = I_{N+2} = \cdots$. A ring that satisfies the ACC is called a Nötherian ring.

Another consequence is that even though we defined them as solutions of finite systems of polynomial equations, affine varieties are naturally associated with ideals — we can define $\mathbf{V}(I)$ to be the set of points $x \in k^n$ such that $p(x) = 0$ for all $p \in I$. But we know that I has a finite generating set $\{g_1, \ldots, g_s\}$ so that $\mathbf{V}(I) = \mathbf{V}(g_1, \ldots, g_s)$ is really an affine variety.

Properties of Gröbner bases: So far we have a non-constructive proof of the fact that every ideal $I \subset k[x_1, \ldots, x_n]$ has a Gröbner basis (since the generating set obtained in the proof of the Hilbert Basis Theorem is by definition a Gröbner basis). An important consequence of having a Gröbner basis $\{g_1, \ldots, g_s\}$ for I is that for any $f \in k[x_1, \ldots, x_n]$ we have that there is a *unique* $r \in k[x_1, \ldots, x_n]$ such that no term of r is divisible by any of $\mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_s)$ and there is a $g \in I$ such that $f = g + r$. In particular, g and r are obtained by using the division algorithm to divide f by $(g_1, \ldots, g_s)$, and g and r are independent of what order we write the g_i's in.

To prove this, start from $f = q_1g_1 + \cdots + q_sg_s + r$, which gives a g and an r where no term of r is divisible by any of $\mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_s)$. To prove uniqueness, suppose that $f = g + r = g' + r'$ where both r and r' satisfy this property. If $r \neq r'$, then $\mathrm{LT}(r - r') \in \langle \mathrm{LT}(I) \rangle = \langle \mathrm{LT}(g_1), \ldots, \mathrm{LT}(g_s) \rangle$, so that we would need $\mathrm{LT}(r - r')$ to be divisible by some $\mathrm{LT}(g_i)$, which is impossible.

Proposition: Let $G = (f_1, \ldots, f_s)$ be a Gröbner basis with respect to some ordering and let $f \in k[x_1, \ldots, x_n]$. Then $f \in I = \langle f_1, \ldots, f_s \rangle$ if and only if the remainder of f on division by G is zero.

Proof. If the remainder is zero then clearly $f \in I$. On the other hand, suppose $f \in I$ and that $f = a_1 f_1 + \cdots + a_s f_s + r$ is the output of the division algorithm. Then $r \in I$ and no term of r is divisible by any $\mathrm{LT}(f_i)$, so in particular $\mathrm{LT}(r)$ is not divisible by any $\mathrm{LT}(f_i)$, which contradicts that G was a Gröbner basis unless $r = 0$.

So a Gröbner basis gives us a way to avoid all our long division issues and solve the ideal membership problem. There are two obstacles we need to overcome: the first is to find a more workable criterion for knowing that a given basis for an ideal is in fact a Gröbner basis, and the second more ambitious one is to find a Gröbner basis for an ideal. We turn to these next.

***S*-polynomials and Buchberger's criterion**: Fix a monomial ordering $>$ on $k[x_1, \ldots, x_n]$.

Suppose we want to check whether a set of polynomials $\{f_1, \ldots, f_r\} \subset k[x_1, \ldots, x_n]$ (we assume that none of the f_i's are the zero polynomial) is a Gröbner basis for the ideal generated by these polynomials. According to the definition of Gröbner bases given above, for any polynomial f of the form

$$f = a_1 f_1 + \cdots + a_r f_r \in \langle f_1, \ldots, f_r \rangle$$

(where $a_1, \ldots, a_r \in k[x_1, \ldots, x_n]$) we need to check whether $\mathrm{LT}(f_i)$ divides $\mathrm{LT}(f)$ for at least one value of i. Of course, this criterion is unworkable because we would have to check every $f \in I$, and there are infinitely many such polynomials f.

We begin by understanding further what can be challenging about testing the criterion. To do this, we need some notation.

Let $\alpha_i = \mathrm{multideg}(a_i)$, let $\beta_i = \mathrm{multideg}(f_i)$ and let $\gamma = \mathrm{multideg}(f)$. Also, let $c_i = \mathrm{LC}(a_i)$, let $d_i = \mathrm{LC}(f_i)$ and let $a = \mathrm{LC}(f)$. Finally, set $\delta = \max\{\alpha_i + \beta_i\}$ with respect to the ordering $>$. Clearly, we have $\delta \geqslant \gamma$, and if $\delta = \gamma$ then there are some terms in the sum for f, say m of them (which we can assume are the first m terms without loss of generality) such that $\delta = \alpha_1 + \beta_1 = \cdots = \alpha_m + \beta_m$ and for which $a_i d_i \neq 0$ for $i = 1, \ldots, m$. Then we have

$$a x^\gamma = (c_1 d_1 + \cdots + c_m d_m) x^\delta.$$

In this case $d_1 x_1^\beta = \mathrm{LT}(f_1)$ divides $a x^\gamma = \mathrm{LT}(f)$ and so the Gröbner basis criterion is clearly satisfied.

On the other hand, if $\gamma < \delta$ then there must be cancellation among the leading terms on the right-hand side, in which case $\mathrm{LT}(f)$ is not necessarily divisible by (f_i) for any $i = 1, \ldots, r$. This is what happened in the example on the midterm, and in the following:

Example: Using the lexicographic ordering, consider $I = \langle x^2 - y \,,\, x^2 y + 1 \rangle \subset k[x, y]$. Then $f = y^2 + 1 = -y(x^2 - y) + 1(x^2 y + 1) \in I$, but neither $\mathrm{LT}(x^2 - y) = x^2$ nor $\mathrm{LT}(x^2 y + 1) = x^2 y$ divides $y^2 = \mathrm{LT}(f)$. Thus, $\{x^2 - y \,,\, x^2 y + 1\}$ is *not* a Gröbner basis for I.

So now we've identified the phenomenon that prevents a given basis from being a Gröbner basis, namely the kind of cancellation of leading terms that occurs in this example. So we need to understand this better, and we still need to avoid the need to check every $f \in I$. It turns out that we can reduce the amount of work required to a finite amount by choosing a clever "representative sample" of all the (problematic) $f \in I$. To see how to do this, suppose

$$f = a_1 f_1 + \cdots + a_r f_r \in I$$

and, using the notation above (including the fact that the first m terms in the sum have $\alpha_i + \beta_i = \delta$), we have

$$f = A + (a_1 - \mathrm{LT}(a_1))f_1 + \cdots + (a_m - \mathrm{LT}(a_m))f_m + a_{m+1}f_{m+1} + \cdots + a_r f_r,$$

where $A = \mathrm{LT}(a_1)f_1 + \cdots + \mathrm{LT}(a_m)f_m = c_1 x^{\alpha_1} f_1 + \cdots + c_m x^{\alpha_m} f_m$. Thus f is the sum of A and set of polynomials with leading terms which come strictly after x^δ. As we have already seen, if $c_1 d_1 + \cdots + c_m d_m \neq 0$, then no cancelation among the leading terms occurs and we have that $\mathrm{LT}(f)$ is divisible by $\mathrm{LT}(f_i)$ for some $i = 1, \ldots, r$. On the other hand, if $c_1 d_1 + \cdots + c_m d_m = 0$ then cancellation does occur among the leading terms, so $\delta > \gamma$. In this case, we set $g_i = x^{\alpha_i} f_i / d_i$ and see that

$$\begin{aligned}
A &= c_1 x^{\alpha_1} f_1 + \cdots + c_m x^{\alpha_m} f_m \\
&= c_1 d_1 g_1 + \cdots + c_m d_m g_m \\
&= c_1 d_1 (g_1 - g_2) + (c_1 d_1 + c_2 d_2)(g_2 - g_3) + (c_1 d_1 + c_2 d_2 + c_3 d_3)(g_3 - g_4) \\
&\qquad + \cdots + (c_1 d_1 + \cdots + c_{m-1} d_{m-1})(g_{m-1} - g_m) + (c_1 d_1 + \cdots + c_m d_m) g_m.
\end{aligned}$$

If $c_1 d_1 + \cdots + c_m d_m = 0$, this shows that A is a k-linear combination of $g_i - g_j = x^{\alpha^i} f_i / d_i - x^{\alpha_j} f_j / d_j$ (because the c_i's and d_i's are *constants*). From this observation we get Buchberger's S-polynomials.

Definition: Let $\alpha = (a_1, \ldots, a_n)$ and $\beta = (b_1, \ldots, b_n)$ be elements of $\mathbb{Z}^n_{\geqslant 0}$, and define $\gamma = (c_1, \ldots, c_n)$ via $c_i = \max(a_i, b_i)$. Then x^γ is the least common multiple of x^α and x^β. For $f, g \in k[x_1, \ldots, x_n]$, let x^γ be the least common multiple of $\mathrm{LM}(f)$ and $\mathrm{LM}(g)$. The S-polynomial of f, g is

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

Note that we have a strict inequality $\gamma > \mathrm{multideg}(S(f, g))$ because of cancellation.

Example: With the lexicographic order (with $x > y$) in $k[x, y]$, we have $\mathrm{lcm}(x^2, x^2 y) = x^2 y$ and so

$$\begin{aligned}
S(x^2 + y\,,\, x^2 y + 1) &= \frac{x^2 y}{x^2}(x^2 + y) - \frac{x^2 y}{x^2 y}(x^2 + 1) \\
&= y(x^2 + y) - (x^2 y + 1) \\
&= y^2 - 1
\end{aligned}$$

Before we defined the S-polynomials, we showed that if cancellation occurs among the leading terms in the sum $a_1 f_1 + \cdots + a_r f_r$ then $c_1 d_1 + \cdots + c_m d_m = 0$ and so

$$\begin{aligned}
A &= \mathrm{LT}(a_1) f_1 + \cdots + \mathrm{LT}(a_m) f_m \\
&= b_1 x^{\zeta_1} S(f_1, f_2) + b_2 x^{\zeta_2} S(f_2, f_3) + \cdots + b_{m-1} x^{\zeta_{m-1}} S(f_{m-1}, f_m)
\end{aligned}$$

where the coefficients b_i are constants (elements of k), and $x^{\zeta_i} = x^{\alpha_i+\beta_i}/\operatorname{lcm}(x^{\beta_i}, x^{\beta_{i+1}})$ Because of the nature of the S-polynomials, we know that $\delta > \text{multideg}(x^{\zeta_i}S(f_i, f_{i+1}))$ and the inequality is strict (because of the cancellation). This is the observation that makes the whole theory work:

Theorem (Buchberger's criterion): Let I be an ideal in $k[x_1, \ldots, x_n]$. Then a basis $G = \{f_1, \ldots, f_s\}$ is a Gröbner basis for I we have $S(f_i, f_j) \in I$ for all i and j, i.e., if and only if for all pairs $i \neq j$, the remainder on division of $S(f_i, f_j)$ by G is zero.

Proof: If G is a Gröbner basis for I, then by the proposition at the top of page 9, since the S-polynomials are in I, we will have the remainder of $S(f_i, f_j)$ on division by G is zero.

Now let $f \in I$ be a non-zero polynomial. We must show that if the S-polynomials all have zero remainders on division by G, then $\text{LT}(f) \in \langle \text{LT}(f_1), \ldots, \text{LT}(f_s)\rangle$. We will go about this as follows: since $f \in I$, we can express $f = \sum_{i=1}^{s} a_i f_i$ for some $a_i \in k[x_1, \ldots, x_n]$, and we know that

$$\delta = \max(\text{multideg}(a_i f_i)) \geqslant \gamma = \text{multideg}(f).$$

As we noted above, if $\delta = \gamma$, then $\text{LT}(f)$ is divisible by at least one of the $\text{LT}(f_i)$, so we will have $\text{LT}(f) \in \langle \text{LT}(f_1), \ldots, \text{LT}(f_s)\rangle$. So our objective is to show that it is possible to find such an expression for f so that $\delta = \gamma$.

It's important to realize that there are in fact many possible choices of the polynomials a_i so that $f = \sum_{i=1}^{s} a_i f_i$ — for instance, given one such choice, you could replace a_3 by $a_3 + f_5$ and replace a_5 by $a_5 - f_3$ and get another set of a_i's that work. So among all the possible ways of choosing the a_i's there is (at least) one for which the resulting δ is *minimal* with respect to the monomial ordering $>$. We claim that if G is a Gröbner basis for I, then for such a "minimal" choice of a_i's we will have $\delta = \gamma$.

We will prove this by contradiction, and show that if $\delta > \gamma$ then we can find another choice of a_i's for which δ is strictly smaller (i.e., comes strictly after) the one we had, which would contradict the minimality of the δ we had. But this is the content of the observations made just before the definition of the S-polynomials, and just before the statement of this theorem. Namely, if $\delta > \gamma$, then we know there must be cancellation, and so that $c_1 d_1 + \cdots + c_m d_m = 0$ in the notation we were using before the definition. Therefore, we can express f as

$$f = A + (a_1 - \text{LT}(a_1))f_1 + \cdots + (a_m - \text{LT}(a_m))f_m + a_{m+1}f_{m+1} + \cdots + a_r f_r,$$

where

$$\begin{aligned} A &= \text{LT}(a_1)f_1 + \cdots + \text{LT}(a_m)f_m \\ &= b_1 x^{\zeta_1} S(f_1, f_2) + b_2 x^{\zeta_2} S(f_2, f_3) + \cdots + b_{m-1} x^{\zeta_{m-1}} S(f_{m-1}, f_m) \end{aligned}$$

From the observation just before the statement of this theorem, we have that $\delta > \text{multideg}(A)$. And we can use the fact that the remainder on division of $S(f_i, f_j)$ by G is zero to get an expression for $S(f_i, f_j)$ of the form

$$S(f_i, f_j) = \sum_{k=1}^{s} b_{ijk} f_k$$

where $b_{ijk} \in k[x_1, \ldots, x_n]$ and (from the division algorithm) $\text{multideg}(S(f_i, f_j)) \geqslant \text{multideg}(b_{ijk} f_k)$ for all i, j and k. Therefore, we can rewrite A as a sum of the form $A = \sum_{i=1}^{s} h_i f_i$ where the δ is greater than the multidegree of each term. All the other terms in the expression for f above clearly have multidegrees less than δ, so we have found a new expression for f with a reduced value of δ. This contradiction proves the theorem.

Buchberger's algorithm. Now we turn to the problem of constructing a Gröbner basis of an ideal. From Buchberger's criterion, if we start with a basis $F = (f_1, \ldots, f_s)$ for an ideal $I = \langle f_1, \ldots, f_s \rangle$, we can check whether F is a Gröbner basis for I by calculating the remainders on division by F of all the S-polynomials $S(f_i, f_j)$ for $i \neq j$. If all the remainders are zero, then F is a Gröbner basis. What do we do if some of the remainders are not zero? If we adjoin the non-zero remainders to the basis F, we'll get a (bigger and necessarily redundant) basis F' for the ideal I for which the particular remainders $S(f_i, f_j)$ will now be zero. However, in doing this we introduce some new pairs and so some new remainders and there is no guarantee that the new remainders will be zero. Buchberger's fundamental observation was that iterating this process of adjoining the remainders to the given basis will terminate in a finite number of steps with a basis for which all the remainders are zero, i.e., with a Gröbner basis.

Theorem (Buchberger's algorithm): Let $I = \langle f_1, \ldots, f_s \rangle$ be a non-zero ideal in $k[x_1, \ldots, x_n]$. Then a Gröbner basis for I can be constructed in a finite number of steps by the following algorithm:

```
Input: $F = (f_1, \ldots, f_s)$
Output: a Gröbner basis $G = (g_1, \ldots, g_t)$ for $I$, with $F \subset G$

Set $G = F$
REPEAT
    Set $G' = G$
    For each pair $p, q$ in $G'$ with $p \neq q$:
        Calculate the remainder $r$ on division of $S(p, q)$ by $G$.
        If $r \neq 0$ then let $G = G \cup \{r\}$.
UNTIL $G = G'$
```

Proof. A bit of useful notation to get us started: if $G = \{g_1, \ldots, g_s\}$, then $\langle G \rangle$ will denote the ideal $\langle g_1, \ldots, g_s \rangle$ and $\langle \text{LT}(G) \rangle$ will denote the ideal $\langle \text{LT}(g_1), \ldots \text{LT}(g_s) \rangle$.

Clearly, since G contains F at every step of the algorithm, and since we only append elements of I as we go, we always have that G is a basis for I. The algorithm can terminate only when $G = G'$, which means that the remainder of $S(p, q)$ on division by G is zero for all $p, q \in G$, which means that G is a Gröbner basis of I by Buchberger's criterion. So it remains to show that the algorithm terminates.

Each time through the loop, we adjoin the non-zero remainders of S-polynomials of pairs of elements of G' to the original set G'. So clearly, since $G' \subset G$, we have $\langle \text{LT}(G') \rangle \subset \langle \text{LT}(G) \rangle$. This containment must be strict if $G' \neq G$ — to see this suppose r is a non-zero remainder from one of the S-polynomials from G'. Since r is a remainder on division by G', we have that $\text{LT}(r)$ is not divisible by the leading terms of any element of G', and hence $\text{LT}(r) \notin \langle \text{LT}(G') \rangle$. But $\text{LT}(r) \in \langle \text{LT}(G) \rangle$, which proves the claim.

So now we have that the ideals $\langle \mathrm{LT}(G')\rangle$ from successive iterations of the loop form an ascending chain of ideals in $k[x_1, \ldots, x_n]$. Since $k[x_1, \ldots, x_n]$ is a Noetherian ring, this chain of ideals must stabilize, and so $\langle \mathrm{LT}(G')\rangle = \langle \mathrm{LT}(G)\rangle$ must happen eventually. But then we must have $G' = G$, so the algorithm must terminate after a finite number of steps.

Example: Work in the ring $k[x, y]$ with grlex order, and let $f_1 = x^3 - 2xy$ and $f_2 = x^2y - 2y^2 + x$. Then $\{f_1, f_2\}$ is not a Gröbner basis for $I = \langle f_1, f_2\rangle$ because $\mathrm{LT}(S(f_1, f_2) = -x^2 \notin \langle \mathrm{LT}(f_1), \mathrm{LT}(f_2)\rangle$. The Gröbner basis produced by Buchberger's algorithm is $\{f_1, f_2, f_3, f_4, f_5\}$ where $f_3 = -x^2$, $f_4 = -2xy$ and $f_5 = -2y^2 + x$.

Theorem: Let $f_1, \ldots, f_s \in k[x_1, \ldots, x_n]$. Then let I be the ideal $I = \langle t_1 - f_1, \ldots, t_s - f_s\rangle$ in the ring $k[x_1, \ldots, x_n, t_1, \ldots, t_s]$. Let G be a Gröbner basis of I with respect to the lex ordering with $x_1 > \cdots > x_n > t_1 > \cdots > t_s$. A polynomial $f \in k[x_1, \ldots, x_n]$ can be written as a polynomial in $f_1, \ldots, f_s$ if and only if the remainder r of f on division by G is in $k[t_1, \ldots, t_r]$. In this case $f = r(f_1, \ldots, f_r)$.

Example: Let $f_1 = x + y$, $f_2 = xy$ in $k[x, y]$ (symmetric polynomials). Then a Gröbner basis of $\langle t_1 - x - y\,,\, t_2 - xy\rangle$ is $G = \{t_1 - x\,,\, t_2 - t_1y + y^2\}$.