GRÖBNER BASES

DYLAN ZWICK

1. Overview

Gröbner bases are a powerful computational tool in commutative algebra that were only developed in the last fifty years and for which applications are being rapidly discovered today. Gröbner bases address fundamental questions about computational commutative algebra, but are relatively easy to understand. In fact, *why* they're useful can be readily understood by any undergraduate math major, and *how* they work can be understood by undergraduates with relatively little additional effort. In this talk I'll attempt to address the "why" part, and encourage you to find our more about the "how" part on your own. Discovering ways of improving Gröbner bases methods and applying them to new fields, not just in pure and applied mathematics but also in biology and statistics, is an active area of research today.

2. Rings, Ideals, and Polynomials

The first step in understanding Gröbner bases is understanding mathematical rings. A ring R is a set of elements upon which two binary operations, addition (denoted $+$), and multiplication (denoted $\cdot$ or $\times$ when denoted at all), are defined. The ring R must satisfy the following requirements:

(a) $a + b \in R$ for all $a, b \in R$. This is called being *closed under addition.*

(b) $a + (b + c) = (a + b) + c$ for all $a, b, c \in R$. This is called *associativity of addition.*

(c) There exists a unique element, denoted 0, satisfying $a + 0 = 0 + a = a$ for all $a \in R$. This is called the existance of an *additive identity.*

(d) For every element $a \in R$ there is a unique element, denoted $-a$, such that $a + (-a) = 0$. This element is the *additive inverse* of a.

(e) $a + b = b + a$ for all $a, b \in R$. This is called *commutativity of addition.*

(f) $a \times b = a \cdot b = ab \in R$ for all $a, b \in R$. This is called being *closed under multiplication.*

(g) $a(bc) = (ab)c$ for all $a, b, c \in R$. This is called *associativity of multiplication.*

(h) $a(b+c) = ab+ac$ and $(a+b)c = ac+bc$ for all $a, b, c \in R$. This is called the *distributive property.*

Now, in general these are the only properties that are required of a ring. However, we're going to be focusing in on a specific class of rings, namely commutative rings with a 1. So, we'll require two additional properties:

(i) There exists an element $1 \in R$ such that $1r = r1 = r$ for all $r \in R$. This is called the existance of a *multiplicative identity.*

(j) $ab = ba$ for all $a, b \in R$. This is called *commutativity of multiplication.*

Some examples of rings that you may be familiar with (even if you didn't know they're rings) are:

- The integers $\mathbb{Z}$ form a ring.
- The rational numbers $\mathbb{Q}$ form a ring. In fact, they form a very special type of ring called a *field.* A field is a commutative ring with identity in which every non-zero element has a *multiplicative* inverse.
- The set of polynomials in one variable over $\mathbb{R}$, denoted $\mathbb{R}[x]$ form a ring.
- The set of polynomials in a finite number of variables over $\mathbb{C}$, denoted $\mathbb{C}[z_1, \ldots, z_n]$, form a ring.

Now, when studying rings one of the primary concepts is that of an *ideal.* An ideal I is a subset of a ring such that:

(1) For all $a, b \in I$, $a + b \in I$.

(2) For all $r \in R$ and $a \in I$, $ra \in I$.

Some examples of ideals are:

- The even integers, denoted $2\mathbb{Z}$, are an ideal of the ring of integers $\mathbb{Z}$.
- The set of functions in $\mathbb{R}[x]$ that vanish at 0 are an ideal of the ring $\mathbb{R}[x]$.

Now, if we have a set of elements $s_1, s_2, \ldots, s_n \in I$ we say that these elements *generate* I if every element $s \in I$ can be written as a linear combination of these elements:

$$s = r_1 s_1 + r_2 s_2 + \cdots + r_n s_n, \text{ for } r_i \in R.$$

Similarly, for *any* set of elements $t_1, \ldots, t_m \in R$ the ideal generated by these elements will be the set of all elements $t \in R$ such that:

$$t = r_1 t_1 + r_2 t_2 + \cdots + r_m t_m, \text{ for } r_j \in R.$$

It's quick to verify that this set is an ideal, and in fact it's the smallest ideal of R that contains the elements $t_1, \ldots, t_m$. We say it's the ideal generated by these elements, and denote it $(t_1, \ldots, t_m)$.

Now, there are many, many questions mathematicians can (and do) ask about rings and their ideals, but today we're going to focus in on one question in particular, and we're only going to address this question to a particular important class of rings.

3. The Big Question

The big question for this talk is:

Given an ideal I, and an element $a \in R$,
how can we determine if $a \in I$?

We call this the "ideal membership problem", or IMP. Our main focus in this lecture will be on addressing this problem for the ring $R = k[x_1, \ldots, x_n]$, or in words the ring of polynomials in variables $x_1, \ldots, x_n$ over the field k. We use the letter k to represent *any* arbitrary field, and we note that all of our methods here will work for any field. If you don't like this kind of generality, you can just view k as being $\mathbb{R}$, or $\mathbb{C}$, or $\mathbb{Q}$ or whatever your favorite particular field may be. Now, before we address this question in a more difficult context, let's first discuss it with some simpler examples. These examples should also reinforce the concepts of ideals and rings.

3.1. **The Integers.** - In $\mathbb{Z}$, ideals are pretty simple. The reason for this is based upon Euclid's algorithm. If we're given two elements $a, b \in \mathbb{Z}$, we can find their *greatest common denominator*, or $gcd(a, b)$. The ideal in $\mathbb{Z}$ generated by (a, b) will be the same as the ideal generated by $(gcd(a, b))$. In this way an ideal generated by any finite set of elements $(n_1, \ldots, n_l) \subseteq \mathbb{Z}$ can be written as an ideal generated by only one element, $(gcd(n_1, \ldots, n_l)$. . Using this idea we can prove that $\mathbb{Z}$ is something called a *principle ideal domain*, or even better a *Euclidean domain*. Now, what domain means[1] shouldn't concern us now, and the important thing here is that every ideal can be generated by just one element. So, to figure out if a number $n \in \mathbb{Z}$ is in our ideal, we just find the generating element for our ideal, and see if it divides n.

For example, suppose we want to find out if 15 is an element within the ideal generated by $12, 30$ and 9. Well, $gcd(12, 30, 9) = 3$. So,

$$(12, 30, 9) = (3)$$

and $3|15$, so $15 \in (12, 30, 9)$. We could show this explicitly by writing:

$$15 = 5 \times (9 - (30 - 2 \times 12)) = 5 \times 9 - 5 \times 30 + 10 \times 12.$$

Note that this representation isn't unique, and there may be others. For example,

$$15 = 5 \times 12 - 5 \times 9.$$

This illustrates a general fact about ideals. Namely, that if you're given a generating set, there may be many distinct ways of writing the same element as a linear combination of the elements in the generating set.

3.2. **Polynomials in One Variable.** - The situation with polynomials in $k[x]$ is more or less the same as with the integers. For any two polynomials we have a well defined least common denominator, and using an analogue of Euclid's algorithm we can find it without too much trouble. So, for any ideal $I \subseteq k[x]$ we know I will be generated by only one element, say f, and then to determine if $g \in I$ we just need to determine if f divides g.

[1] A domain is a ring in which if $ab = 0$ either $a = 0$ or $b = 0$.

3.3. **Polynomials in Many Variables.** - This is where things get interesting. First, we remark that if we're only dealing with one or two variables, we'll write $k[x, y]$ or $k[x, y, z]$ in place of $k[x_1, x_2]$ or $k[x_1, x_2, x_3]$, respectively. Now, we note that, unlike with the integers and polynomials in one variable, polynomials in more than one variable are not a PID. For example, the ideal generated by x and y in $k[x, y]$, denoted (x, y), cannot be generated by a single polynomial $f \in k[x, y]$. Also, there's no obvious division algorithm for polynomials in more than one variable like there is when dealing with integers or single variable polynomials. So, for the rest of the talk we'll devote ourselves to addressing our big question for the more difficult situation of polynomial rings in more than one variable.

4. Multivariable Polynomial Division

In order to figure out the IMP for multivariable polynomials, we have to first talk about how we'd perform division on them. In particular, we have to define something called a *monomial order*. A monomial order is, basically, a way of comparing monomial terms in our polynomial ring and figuring out which one is "bigger". Now, when we're doing division on a single-variable polynomial, there's an order so obvious we usually don't even mention it. For example, when we're dividing $x^2 - 1$ by $x + 1$ we get[2]:

$$\begin{array}{rrrr} & x & -1 & \\ \cline{2-4} x+1 \,\big) & x^2 - & & 1 \\ & -(x^2 & +x) & \\ & & -x & -1 \\ & -(& -x & -1) \\ & & & 0 \end{array}$$

and inherent within this division was an ordering on the terms (the monomials) within our polynomial. To be specific, x^2 terms were considered before x terms, and x terms before constant terms. In general, we've got a monomial ordering on $k[x]$ determined by the degree of the monomial.

[2] For our specific examples we'll be assuming our field k is $\mathbb{C}$.

With more than one variable, things become a bit tricky. For example, it's not immediately obvious which of the following three terms in $k[x, y]$:

$$\{x^2, xy, y^2\}$$

would be the greatest. In fact, this depends upon the monomial order we choose.

Before we formally define a monomial ordering, we note that any monomial in $k[x_1, \ldots, x_n]$ can be viewed as an element in $\mathbb{Z}^n_{\geq 0}$. For example, in $k[x, y, z]$ we'd have:

$$x^2y^3z = (2, 3, 1),$$

and so a monomial ordering is the same as an ordering on $\mathbb{Z}^n_{\geq 0}$.

Without further ado, let's define what a monomial ordering is:

Definition - A monomial ordering on $k[x_1, \ldots, x_n]$ is any relation $>$ on $\mathbb{Z}^n_{\geq 0}$, or equivalently, any relation on the set of monomials x^α, $\alpha \in \mathbb{Z}^n_{\geq 0}$ satisfying:

(1) $>$ is a total ordering on $\mathbb{Z}^n_{\geq 0}$
(2) If $\alpha > \beta$ and $\gamma \in \mathbb{Z}^n_{\geq 0}$, then $\alpha + \gamma > \beta + \gamma$.
(3) $>$ is a well-ordering on $\mathbb{Z}^n_{\geq 0}$. This means that every nonempty subset of $\mathbb{Z}^n_{\geq 0}$ has a smallest element under $>$.

Basically, what this ordering does is that it lets us compare any two monomials, assures us that if an element is greater than another, this doesn't change if we divide them both by the same thing, and finally that our division algorithm will eventually terminate.

Three examples of monomial orderings, and the only monomial ordering we'll be using in our examples, are:

- **Lexicographic (Lex) Ordering** : Let $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $\beta = (\beta_1, \ldots, \beta_n) \in \mathbb{Z}^n_{\geq 0}$. We say $\alpha >_{lex} \beta$ if, in the vector difference $\alpha - \beta \in \mathbb{Z}^n$, the leftmost nonzero entry is positive. We will write $x^\alpha >_{lex} x^\beta$ if $\alpha >_{lex} \beta$.

- **Graded Lex Order** : Let $\alpha, \beta \in \mathbb{Z}^n_{\geq 0}$. We say $\alpha >_{grlex} \beta$ if

$$|\alpha| = \sum_{i=1}^{n} \alpha_i > |\beta| = \sum_{i=1}^{n} \beta_i$$

or

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

- **Graded Reverse Lex Order** : Let $\alpha, \beta \in \mathbb{Z}^n_{\geq 0}$. We say $\alpha >_{grevlex} \beta$ if

$$|\alpha| = \sum_{i=1}^{n} \alpha_i > |\beta| = \sum_{i=1}^{n} \beta_i$$

or

$|\alpha| = |\beta|$ and the rightmost nonzero entry of $\alpha - \beta \in \mathbb{Z}^n$ is negative.

We note that if our variables are, for example, $k[x, y, z]$, then lex order is almost, but not quite, the way things are ordered in the dictionary. The difference is that, in lex order, $xx > x$, while in the dictionary it's the other way around.

Now that we've established a monomial order, we can then perform division on our multivariable polynomials. That is to say, we can divide a polynomial $f \in k[x_1, \ldots, x_n]$ by an ordered sequence of polynomials $f_1, \ldots, f_n \in k[x_1, \ldots, x_n]$. I won't go through the formal details, but instead just give some illustrative examples.

Example 1 Let's divide $f = xy^2 + 1$ by $f_1 = xy + 1$ and $f_2 = y + 1$ using lex order.

$$\begin{array}{rl}
a_1 & y \\
a_2 & -1 \\
\hline
xy+1 & \big)\; xy^2+1 \\
y+1 & \big)\; -(xy^2+y) \\
 & \quad -y+1 \\
 & \quad -(-y-1) \\
 & \qquad 2 \leftarrow \text{remainder}
\end{array}$$

So,

$$\boxed{xy^2+1 = y(xy+1) - (y+1) + 2}$$

Example 2 To illustrate a subtlety that does not come up with single variable division, let's divide $x^2y + xy^2 + y^2$ by $f_1 = xy - 1$ and $f_2 = y^2 - 1$.

$$\begin{array}{ll} a_1 : & x + y \\ a_2 : & \end{array}$$

$$\begin{array}{r|l} xy - 1 & x^2y + xy^2 + y^2 \\ y^2 - 1 & -(x^2y - x) \\ & \qquad xy^2 + x + y^2 \\ & \quad -(xy^2 - y) \\ & \qquad\qquad x + y^2 + y \end{array}$$

Note that neither the leading term of f_1 nor f_2 divides the leading term of $x + y^2 + y$. However, $x + y^2 + y$ is *not* the remainder, since the leading term of f_2 does divide y^2 Thus, if we move x to the remainder, we can continue dividing[3].

$$\begin{array}{ll} a_1 = & x + y \\ a_2 = & 1 \end{array} \qquad\qquad \begin{array}{c} \underline{r} \\ x \end{array}$$

$$\begin{array}{r|l} xy - 1 & y^2 + y \\ y^2 - 1 & -(y^2 - 1) \\ & \qquad \underline{y + 1} \\ & \qquad\quad \underline{1} \\ & \qquad\quad 0 \end{array} \qquad\qquad \begin{array}{c} x + y \\ x + y + 1 \end{array}$$

$$\Rightarrow x^2y + xy^2 + y^2 = (x+y)(xy-1) + 1 \cdot (y^2-1) + (x+y+1).$$

[3] This is something that never happens in the one-variable case

So, now we've solved our ideal membership problem! To determine if $f \in (f_1, \ldots, f_n)$ we just pick a monomial order, divide f by $f_1, \ldots, f_n$, and see if our remainder is zero. Right?

5. Gröbner Bases

Not. So. Fast. There are two issues here. First, we don't know yet that for any ideal $I \subset k[x_1, \ldots, x_n]$ there exists a finite set of generators $(f_1, \ldots, f_n)$. This questions was actually a big deal in the late nineteenth century, and it was solved by Hilbert. The Hilbert basis theorem, which is usually the first big theorem in an algebraic geometry or commutative algebra class, tells us that we can indeed always find a finite basis. So, that issue is solved, and that problem doesn't come up.

The other problem is more substantial. We note that if we divide $f = xy^2 - x$ by $f_1 = xy + 1$ and $f_2 = y^2 - 1$ we get:

$$\begin{array}{ll}
a_1 : & y \\
a_2 : & \\
\begin{array}{l} xy+1 \\ y^2-1 \end{array} & \begin{array}{|l} \overline{\;xy^2 - x\;} \\ -(xy^2 + y) \\ \qquad -x - y \end{array}
\end{array}$$

$$\Rightarrow \boxed{xy^2 - x = y(xy+1) + (-x-y)}$$

So, we'd be tempted to conclude $f \notin (f_1, f_2)$. However, if we reverse our order we get:

$$\begin{array}{ll}
a_1 : & x \\
a_2 : & \\
\begin{array}{l} y^2-1 \\ xy+1 \end{array} & \begin{array}{|l} \overline{\;xy^2 - x\;} \\ -(xy^2 - x) \\ \qquad 0 \end{array}
\end{array}$$

$$\Rightarrow \boxed{xy^2 - x = x(y^2-1)}$$

So, in fact, $f \in (f_1, f_2)$. This is a big problem. If our remainder upon division is 0, then we know the element is in our ideal. However, the other way isn't true.

If you're of a more theoretical bent, you'd argue that you've actually solved the problem. To determine if an element is in your ideal, you just divide by *every possible ordering*! Problem solved. Unfortunately, this is, in some sense, a very hard solution. This is because if you've got n polynomials defining your ideal, there are $n!$ possible orderings, and factorial functions grow very quickly. So, if n is fairly large, the number of possible combinations is unmanageably huge.

This is where a Gröbner basis comes in. A Gröbner basis is a basis $(g_1, \ldots, g_n)$ where, for a given monomial ordering, $f \in (g_1, \ldots, g_n)$ if *and only if* the remainder of f after division by $g_1, \ldots, g_n$ is zero. So, we only have to run through *one* division, not $n!$, which for even relatively small values of n is a huge savings in computational time.

Now, as I said at the beginning, we're not going to get too deep into the theory or the practice here, as it would take too long. So, we'll just note that for a given monomial ordering and a given polynomial f we can define its leading term, $LT(f)$, as its greatest monomial term for the given ordering. For an ideal I we can define $< LT(I) >$ to be the ideal generated by the leading terms of all the polynomials in I. A Gröbner basis $(g_1, \ldots, g_n)$ for an ideal I is a set of polynomials satisfying:

$$< LT(g_1), \ldots, LT(g_n) >=< LT(I) >.$$

Now, it will turn out that this Gröbner basis will indeed be an actual basis for I, and that it will satisfy our divison requirement. It turns our furthermore that we can define a *reduced* Gröbner basis, and that for any polynomial ideal I and given monomial ordering, there will exist a unique reduced Gröbner basis. Finally, Buchburger's algorithm gives us a way of finding this reduced Gröbner basis for a given set of generators $(f_1, \ldots, f_n)$ for any polynomial ideal I[4]. Most computer math programs like Maple, Matlab, Mathematica, etc... have built-in Gröbner bases functions.

As I said, we don't have time to go over all of this now, but none of it is very difficult to understand, and in fact it's all explained very nicely, at a level understandable to undergraduate math majors, in the book "Ideals, Varieties, and Algorithms" by Cox, Little, and O'Shea[5].

[4] The discovery of these Gröber bases, the proof that they work, and the algorithm for finding them, were all given in Buchburger's Ph.D. thesis. He named the bases "Gröbner bases" after his thesis adviser.

[5] An excellent book, and the source of the material for this lecture. In fact, this talk is a substantially trimmed (some would say butchered) version of chapter 2 from the book. The book could be readily understood by a junior or senior math undergraduate. Be warned, though, that the current printing of the third edition has an unforgiveably large number of typos. It's really quite ridiculous.

6. Sneak Peak at an Application

Finally, as an example of where these methods can be applied, we note that frequently when dealing with Lagrange multipliers we have to find the zeros of some polynomial equations in many variables, and these equations are not all linear, so we can't just use linear algebra. For example, consider the equations:

$$\begin{gathered} x^2 + y^2 + z^2 = 1, \\ x^2 + z^2 = y, \\ x = z. \end{gathered}$$

Now, frequently the Lagrange multiplier problems from textbooks are easy to solve, but that's because they're particularly chosen. In general, things might not be so nice.

Well, suppose we take a look at the ideal:

$$I =< x^2 + y^2 + z^2 - 1, x^2 + y^2 - y, x - z > \subset \mathbb{C}[x, y, z].$$

If we find a Gröbner basis for these polynomials, we get:

$$\begin{gathered} g_1 = x - z, \\ g_2 = -y + 2z^2, \\ g_3 = z^4 + \frac{1}{2}z^2 - \frac{1}{4}. \end{gathered}$$

Well, what's interesting about these? First, we note that as these polynomials in our Gröbner basis generate the same ideal as our original set of polynomials, the zeros of these Gröbner basis polynomials will be exactly the same as the zeros of the polynomials in our original set. The polynomial g_3 depends *only* on z, and so we can find its roots, in this case using the quadratic equation. Once we know z, we can solve for y in g_2, and once we know z and y (actually, in this case we only need z), we can solve for x in g_1. So, we've got a method for finding the roots (the solutions) to our original set of polynomials! This is just one example, and one swallow does not make a spring, but as you might imagine this method can be generalized to a large set of problems, and is just one situation where Gröbner basis methods can help substantially in practical computations.