Gröbner Bases and Their Applications: A Simplified Review

Elias T. Sink

Bunker Hill Community College, Computer Science Department

CSC-237: C++ Programming

Professor Elizabeth Miller

May 5, 2022

Gröbner Bases and Their Applications: A Simplified Review

Polynomials are among the simplest and most common entities in all of mathematics. Built from the familiar operations of addition and multiplication, they undergird large swaths of modern algebra and geometry and show up everywhere from physics to computer science. One of the most ubiquitous questions one needs to answer when working with polynomials is as follows: given polynomials $f_1, \dots, f_n$ and g, are there polynomials $u_1, \dots, u_n$ such that $g = u_1 f_1 + \cdots + u_n f_n$? This purely algebraic problem, known as the ideal membership problem, is interesting in its own right, and it has deep and surprising connections with other areas of mathematics as well as with various practical domains.

Given the concrete, finite nature of polynomial expressions, one might wonder if such questions can be answered by computers. Indeed, it turns out that the ideal membership problem is quite amenable to algorithmic solutions, as are a host of related questions about polynomials. However, the techniques required to tackle these questions are not at all trivial. Developed by Wolfgang Gröbner and his student Bruno Buchberger in the 1950s and 60s, the theory of Gröbner bases is usually taught in the context of abstract algebra and ring theory. While the formal details of the techniques and proofs of their effectiveness are surely out of reach for those without significant background in mathematics, it is our belief that the core ideas can be meaningfully understood by anyone confident with elementary algebra and mathematical reasoning. Gröbner bases provide an elegant and effective algorithmic framework for understanding polynomial equations, and the breadth of their relevance across a multitude of disciplines makes them worthy of study even by nonexperts.

History

The history of Gröbner bases is rooted in commutative algebra and ring theory, which are fundamentally concerned with the abstraction and generalization of the usual properties of addition and multiplication. While we won't need the full power of ring theory for our exposition of Gröbner bases, their history is inextricable from the technical setting in which they were developed.

There were two main lines of research which culminated in modern ring theory: arithmetic and geometric. On the arithmetic side of early commutative algebra was the work of C. F. Gauss on algebraic number theory in the early 19th century, where he investigated small algebraic extensions of the integers. This was followed by Dedekind's formulation of orders (early rings) and ideals in the context of algebraic number fields (Kleiner, 1998).

In geometry, Max Noether's study of the function fields of algebraic curves led to David Hilbert's work on polynomial rings, including his landmark proof of the Hilbert Basis Theorem in 1890. In the early 20th century, these ideas were unified into Fraenkel's axiomatic definition of rings, which is almost identical to the one we use today. Finally, Emmy Noether's heroic adaptation and generalization of her predecessor's results into this fully abstract context cemented ring theory as a pillar of modern mathematics (Kleiner, 1998).

Several decades later, Wolfgang Gröbner assigned his student Bruno Buchberger the problem of finding a basis for certain quotients of polynomial rings. While Gröbner had a process for doing so as early as 1939, he could not prove that it would always terminate (Abramson, 2009). Buchberger was able to modify Gröbner's technique. In his 1965 thesis, he

introduced what would later be called Gröbner bases (named for his advisor) and described an algorithm to compute them along with a proof of correctness (Buchberger, 2006).

Theory

With the history out of the way, we turn to the mathematics. We begin with a few preliminaries.

Preliminaries

First, a *polynomial* (with rational coefficients[*]) is any expression that can be built from the rational numbers (such as 0, 3, 1/2, or -3/5) and some finite set of variables (say, x, y, and z) through addition and multiplication according to the usual rules of algebra. Typical examples are $\frac{1}{2}xy^2 - 3x + 5$ and $x^3 + 2xyz$. Products of just the variables, such as xy or xy^2z^5 are known as *power products*. (Note that $1 = x^0y^0z^0$ is also considered a power product, while $\frac{1}{x} = x^{-1}$ is not). A power product times a rational number, like $\frac{1}{3}x$ or $-\frac{3}{2}x^3yz^2$, is called a *term*, where the number is called the *coefficient* of that term. A polynomial is then the sum of its terms.

Given nonzero polynomials $f_1, f_2, \ldots, f_n$, we define the *ideal* generated by them, written $\langle f_1, \ldots, f_n \rangle$, to be the collection of polynomials g that can be written $g = u_1f_1 + \cdots + u_nf_n$ for some polynomials $u_1, \ldots, u_n$. Such a g is said to be a *member* of the ideal. For example, the ideal $\langle x, y \rangle$ is the set of polynomials whose terms are all multiples of x or y. $x^2 + 3y$ is a member, while $x + 2$ and $x^2y + z$ are not. With these definitions in mind, a natural question is, given

[*] The choice of rational numbers as coefficients is largely arbitrary; we could have chosen real numbers or complex numbers without affecting the discussion at all. Indeed, for one of the applications we'll make a different choice.

some g, how can we determine if g is a member of $\langle f_1, \ldots, f_n \rangle$? This is easy for $\langle x, y \rangle$, but what about $\langle x^2 - 2xy, y + x - 3 \rangle$? How can we decide if, say, $x^2y + 5$ is a member of this ideal?

This is the ideal membership problem, and its solution is the main objective of this section. Following (Adams & Loustaunau, 1994), we introduce two simple special cases to guide our discussion: polynomials in one variable (e.g., $x^3 - 2x + 1$), and linear polynomials in several variables (e.g., $x + 3y - z$). In both instances, the problem can be solved with relative ease, and our approach to the general case will mirror aspects of both.

Term Orders and Division

Since our goal is to decide whether $g = u_1 f_1 + \cdots + u_n f_n$ for some u_i, we might simply try to find these u_i. Let's think about how this works in one of the simplest cases: $n = 1$, with single-variable polynomials. That is, given polynomials $f(x)$ and $g(x)$, we want to decide if $g(x) = u(x)f(x)$ for some polynomial $u(x)$. This can be solved by a straightforward division algorithm: we write $g(x) = u(x)f(x) + r(x)$, where the *degree* (the highest power of x present) of $r(x)$ is less than that of $f(x)$. The algorithm is as follows: let $r(x) = g(x)$, $u(x) = 0$. (Note that our target equation is satisfied.) If the degree of r is less than that of f, then we're done. Otherwise, let ax^m and bx^n be the highest degree terms of $r(x)$ and $f(x)$ respectively. Since $m \geq n$, $\frac{a}{b}x^{m-n}$ is a valid term. Thus, we can add $\frac{a}{b}x^{m-n}$ to $u(x)$ and subtract $\frac{a}{b}x^{m-n}f(x)$ from $r(x)$. This cancels the highest degree term of r while still satisfying $g(x) = u(x)f(x) + r(x)$. This process is repeated until the degree of r is less than that of f.

For example, take $g(x) = x^3 - 5x^2 + 2$ and $f(x) = x - 3$. We begin with $r(x) = x^3 - 5x^2 + 2$, $u(x) = 0$. The highest degree term of $f(x)$ is x, and that of $r(x)$ is x^3.

We add $\frac{x^3}{x} = x^2$ to $u(x)$ and subtract $x^2(x-3)$ from $r(x)$, leaving us with $u(x) = x^2$ and $r(x) = -2x^2 + 2$. The degree of $r(x)$ is still too big, so we repeat: add $\frac{-2x^2}{x} = -2x$ to $u(x)$ and subtract $(-2x)(x-3)$ from $r(x)$. We get $u(x) = x^2 - 2x$ and $r(x) = -6x + 2$. We repeat this one more time to get $u(x) = x^2 - 2x - 6$ and $r(x) = -16$. Our degree constraint is met, so we're done, having obtained

$$x^3 - 5x^2 + 2 = (x^2 - 2x - 6)(x - 3) - 16.$$

After this division, we can see $g(x)$ is not a member of $\langle f(x) \rangle$, since if it were, the remainder would clearly be zero.

In summary, we subtract multiples of $f(x)$ from $g(x)$ so that the highest degree terms cancel until this is no longer possible. If what's left over is zero, g is a member. This same philosophy can be applied to several multivariate polynomials. However, a key component of the preceding algorithm was choosing the highest degree terms of r and f; which has the "higher degree", x^2y or xy^2? It turns out not to matter very much which we choose, but we do need to make a choice about how to compare power products with multiple variables. Such a choice is called a *term order*. For this discussion, we will choose *lexicographical order*: we'll choose an order for the variables (x, then y, then z). Given two terms, we'll compare powers of x. In case of a tie, we compare powers of y, and so on. Thus x^2y is higher than xy^2, while x^2y^3 is higher than x^2y^2. We emphasize that, subject to a few technical considerations, this is an arbitrary choice. Having decided on an ordering, we call the largest power product appearing in a polynomial f its *leading power product*, denoted $lp(f)$, and the corresponding term is its *leading*

term $lt(f)$. Finally, given power products p and q, we say that p *divides* q if $\frac{q}{p}$ is a valid power product.

With this established, we proceed in close analogy with the single variable case. Given polynomials $f_1, \dots, f_n$ and g, we seek $u_1, \dots, u_n$ and r such that $g = u_1 f_1 + \cdots + u_n f_n + r$, where $lp(f_i)$ does not divide $lp(r)$ for any i. The algorithm to do this is as follows (Adams & Loustaunau, 1994): start with $u_1 = \cdots = u_n = r = 0$, and let $h = g$. If some $lp(f_i)$ divides $lp(h)$, add $\frac{lt(h)}{lt(f_i)}$ to u_i and subtract $\frac{lt(h)}{lt(f_i)} f_i$ from h. If there is no such f_i, simply subtract the highest term from h and add it to r. Repeat this until h is 0.

As an example, take $f_1 = xy + 1$ and $f_2 = y + 1$, with $g = xy^2 + 1$. We start with $h = xy^2 + 1$. According to our term order, the leading power of f_1 is xy, which divides the leading power xy^2 of h. We add $\frac{x^2y}{xy} = y$ to u_1 and subtract $y(xy + 1)$ from h, leaving $u_1 = y$, $h = -y + 1$. xy doesn't divide y, but the y from f_2 clearly does. We add $\frac{-y}{y} = -1$ to u_2 and subtract $-1(y + 1)$ from h. We now have $u_1 = y$, $u_2 = -1$, $h = 2$. Neither xy nor y divides 2, so we add it to r, leaving $u_1 = x, u_2 = -1, r = 2, h = 0$. We're done:

$$xy^2 + 1 = y(xy + 1) + (-1)(y + 1) + 2$$

(adapted from Cox et al, 2015).

Since the remainder wasn't zero, we would like to conclude from this calculation that x^2y is not a member of $\langle xy - x, x^2 - y \rangle$, as we did in the simple case above. Unfortunately, this is not possible. While a remainder of 0 does demonstrate membership, a nonzero remainder is

not generally enough to refute it (Adams & Loustaunau, 1994). To draw the negative conclusion, we need Gröbner bases.

Gröbner Bases

A key technique from linear algebra is row reduction, where a system of linear equations is transformed into a simpler system with the same solution set. In a similar sense, the basis of an ideal is not unique; that is, $\langle f_1, \dots, f_n \rangle$ may be the same ideal as $\langle g_1, \dots, g_m \rangle$, even if the f_i are different from the g_i. Indeed, just as $x + 3y = -2x + y = 0$ is equivalent to $x = y = 0$, we have $\langle x + 3y, -2x + y \rangle = \langle x, y \rangle$ as ideals. From both perspectives, the latter basis is clearly preferable. Our solution to the weakness of the division algorithm is to construct a better basis for $\langle f_1, \dots, f_n \rangle$ such that division with respect to the new basis gives a remainder of zero if and only if the dividend is a member of the ideal. This desire is captured by the following definition: a *Gröbner basis* for an ideal consists of nonzero polynomials $g_1, \dots, g_m$ such that if h is any member of the ideal, then the leading power of h is divisible by the leading power of some g_i. This guarantees that the division of f by the g_i gives a remainder of 0.

Given this definition, how can one compute a Gröbner basis for $\langle f_1, \dots, f_n \rangle$? Buchberger's fundamental insight is this: the obstruction to $f_1, \dots, f_n$ being a Gröbner basis is the cancellation of their leading terms. For instance, if $f_1 = x^2$ and $f_2 = x^3 - x$, then $(x)f_1 + (-1)f_2 = x$ is a member of $\langle f_1, f_2 \rangle$, but x is not divisible by $lp(f_1) = x^2$ or $lp(f_2) = x^3$, as the larger terms cancelled out. It turns out that a Gröbner basis can be constructed by adding to $f_1, \dots, f_n$ extra polynomials to account for such cancellations. More specifically, we define the *S-polynomial*

$$S(f, g) = \frac{P}{lt(f)} f - \frac{P}{lt(g)} g$$

where P is the least common multiple of $lp(f)$ and $lp(g)$. (That is, P is the power product of lowest degree in each variable that is divisible by both $lp(f)$ and $lp(g)$.) Intuitively, the $S(f,g)$ is the simplest way that the leading terms of f and g can cancel out. Buchberger's Theorem says that $g_1, \dots, g_n$ is a Gröbner basis for the ideal it generates if and only if the division of $S(g_i, g_j)$ by $g_1, \dots, g_n$ always leaves a remainder of zero (Adams & Loustaunau, 1994). In other words, the S-polynomials completely capture the failure of a basis to be a Gröbner basis. This suggests the following algorithm for computing Gröbner bases for $\langle f_1, \dots, f_n \rangle$, known as Buchberger's algorithm (Adams & Loustaunau, 1994): for each pair f_i, f_j, divide $S(f_i, f_j)$ by $f_1, \dots, f_n$. If the remainder is nonzero, add it to the end of the list. Repeat until the remainder is zero for all pairs.

That this algorithm terminates is not trivial; see (Adams & Loustaunau, 1994) or (Cox et al., 2015) for proof. However, once this is established, we have a simple, efficient procedure for computing Gröbner bases. With some simple reduction steps to eliminate redundancies, we can obtain a unique reduced Gröbner bases for the ideal. This also allows us to finally solve the ideal membership problem: to determine if g is a member of $\langle f_1, \dots, f_n \rangle$, compute a Gröbner basis for that ideal and divide g. The remainder will be zero if and only if the answer is yes.

Applications

The ideal membership problem is, of course, not the only use for Gröbner bases. Within pure mathematics, the methods we've discussed can be used to solve systems of polynomial equations symbolically, perform quotient ring arithmetic by computer, and facilitate quantitative reasoning about varieties in algebraic geometry. Outside of mathematics proper, Gröbner bases find applications in the study of error-correcting codes, robotics, and more.

Error-Correcting Codes

Suppose we want to send a message, encoded as a string of 1s and 0s (bits), over a noisy channel where errors may occur. How can we ensure that our message is intelligible despite these errors? Error correcting codes provide an answer: we use more bits than necessary to encode our message, but only certain bit strings are considered valid messages. These valid strings are called *codewords.* For instance, suppose we want to send a four-bit message. We send five bits, where the first four bits are our message, and the fifth bit is equal to the parity (0 if even, 1 if odd) of the sum of the first four. Our codewords are the bit strings satisfying that constraint. If a message is received that isn't a codeword, the receiver knows an error has occurred, and with a well-designed code, they can determine what the message was supposed to be with good probability. This is called *decoding*.

Cyclic codes are a special kind of code with two key properties. First, they are *linear*, meaning the 0-string is a codeword and the sum of two codewords (as binary numbers, dropping the extra carry bit) is also a codeword. Second, the cyclic shift of any codeword is also a codeword. That is, if 1011 is a codeword, then so are 0111, 1110, and 1101. The codewords of cyclic codes can be profitably represented as polynomials in one variable with coefficients 0 or 1, added modulo 2. That is, 1011 is $x^3 + x + 1$, where we take $1 + 1 = 0$. Furthermore, multiplication by x is nearly the same as a cyclic shift; all we need to do is set $x^n = 1$, where n is the number of bits in the codeword. With $n = 4$, we have

$$x(x^3 + x + 1) = x^4 + x^2 + x = x^2 + x + 1 \rightarrow 0111.$$

(For those familiar with quotient rings, we are working in $\mathbf{Z_2}[x]/(x^n + 1)$.) Since ideals are closed under addition and multiplication by x, an ideal* gives a cyclic code under this correspondence. Indeed, it turns out that all cyclic codes arise in this way (MacWilliams & Sloane, 1977).

Given this correspondence between cyclic codes and polynomials, its not too large a leap to imagine that Gröbner bases are useful in this context. Indeed, Chen et al. show that the decoding process for cyclic codes is equivalent to a system of polynomials called syndrome polynomials (1994). Gröbner bases can then be used to simplify this system down to just one polynomial, called the error-locator polynomial, which allows cyclic codes to be decoded accurately even after many errors. This method is very general and powerful, and it represents a significant contribution to the study of error-correcting codes.

Robotics

Consider a robotic arm with several rotating joints and a hand at the end. We want to position the hand at some coordinates by setting the angles of the joints. By choosing appropriate parameters for those angles, the mapping from joint configurations to hand positions can be made polynomial. According to Cox et al., Gröbner bases can be used to invert this mapping and obtain all the configurations that place the hand at the desired point (2015). This can be used to plan the motion of the arm such that it avoids obstacles and minimizes total movement and joint velocity.

* Strictly speaking, an ideal containing $x^n + 1$ due to our stipulation $x^n = 1$

Other Applications

A few other applications bear brief mention. Graph coloring is a well-known problem in mathematics and computer science. Given a collection of nodes, some connected by edges, assign one of, say, three colors to each node such that no two connected nodes share a color. By associating each color with a 3rd root of unity (e.g., $e^{2\pi i/3}$) and writing the constraints out as polynomials in the color of each node, the problem can be solved (or shown to have no solution) by computing the Gröbner basis of the resulting system (Adams & Loustaunau, 1994).

In a similar vein, Gröbner bases can be used to reason about Sudoku boards. Just as with graph coloring, the rules of Sudoku can be realized as a large system of polynomial equations in the value of each box. The corresponding Gröbner bases can, in principle, be used to find solutions or count the number of boards (Arnold et al., 2010). While some of these calculations may not be feasible with modern computers and full-size Sudoku boards, the approach remains promising.

Conclusion

Pure mathematics is often derided as excessively abstract and disconnected from practical considerations. However, time and again, the loftiest and most theoretical mathematics proves immensely useful in the real world. Differential geometry with general relativity, functional analysis with quantum mechanics, and number theory with cryptography all serve to demonstrate the utility of pure math. The theory of Gröbner bases is yet another example for this list. Developed by mathematicians working on commutative algebra, Gröbner bases have found application in coding theory, robotics, and more. The elegance and efficiency of the technique make it a critical tool for anyone interested in computational symbolic algebra.

References

Abramson, M. P. (2009). Historical background to Gröbner's paper. *ACM Communications in Computer Algebra*, *43*(1/2), 22–23. https://doi.org/10.1145/1610296.1610301

Adams, W. W., & Loustaunau, P. (1994). *An Introduction to Gröbner Bases.* American Mathematical Society.

Arnold, E., Lucas, S., & Taalman, L. (2010). Gröbner Basis Representations of Sudoku. *The College Mathematics Journal*, *41*(2), 101–112. https://doi.org/10.4169/074683410x480203

Buchberger, B. (2006). Bruno Buchberger's PhD thesis 1965: An algorithm for finding the basis elements of the residue class ring of a zero dimensional polynomial ideal. *Journal of Symbolic Computation*, *41*(3-4), 475–511. https://doi.org/10.1016/j.jsc.2005.09.007

Chen, X., Reed, I. S., Helleseth, T., & Truong, T. K. (1994). Use of Gröbner bases to decode binary cyclic codes up to the true minimum distance. *IEEE Transactions on Information Theory*, *40*(5), 1654–1661. https://doi.org/10.1109/18.333885

Cox, D. A., Little, J., & O'Shea, D. (2015). *Ideals, Varieties, and Algorithms*. Cham Springer International Publishing.

Kleiner, I. (1998). From Numbers to Rings: The Early History of Ring Theory. *Elemente Der Mathematik*, *53*(1), 18–35. https://doi.org/10.1007/s000170050029

MacWilliams, F. J., & Sloane, N. J. A. (1977). *The Theory of Error-Correcting Codes.* North-Holland Publishing Company.