6.S897 Algebra and Computation April 22, 2015

Ideal Membership Problem and Gröbner Basis

Instructor: Madhu Sudan *Scribe: Chennah Heroor*

1 Overview

Today we will discuss an algorithm for solving the ideal membership problem: the method of Gröbner bases. Gröbner bases were introduced in the last lecture, but today we will see the math behind them. The Gröbner bases algorithm was due to Buchberger. It solves the ideal membership question, and suggests the notion of a canonical remainder.

2 Ideal Membership Problem

We begin by defining the Ideal Membership Problem

Definition 2.1. Ideal Membership Problem: Given a target polynomial f and polynomials $f_1, \ldots, f_m \in \mathbb{K}[x_1, \ldots, x_n]$, is $f \in \langle f_1, \ldots, f_m \rangle$, where $\langle f_1, \ldots, f_m \rangle$ denotes the ideal generated by the f_i

We note that an equivalent formulation of the problem is: Does there exist polynomials $g_1, ...g_m$ such that $f = \sum f_i g_i$.

We will solve this question by using Gröbner bases. The Gröbner basis gives us a representation of an ideal, that allows us to easily decide membership, and the basis is constructed via Buchberger's algorithm. An important question regarding Buchberger's algorithm is its complexity, or given a set of n polynomials with degree m, what is the running time of the algorithm. In order to learn the complexity of this problem, it becomes necessary to determine the degree of the polynomials g_i. However, we leave the discussion of the algorithm's complexity to the next lecture.

First, we focus on the idea of finding the remainder of a polynomial modulo an ideal. This problem is simple if we restrict ourselves to univariate polynomials. We order the polynomials by the highest degree and repeatedly take the remainder modulo the highest degree we can remove.

However, with multivariate polynomials, we have a notion of preference: which terms would we prefer to reduce? For example, consider the polynomial $x^2y + y^2x$ divided by $x + y$. It is unclear what answer should be, as the remainder can be written as either a polynomial purely in x or purely in y, or we can reduce the total degree of the polynomial.

The Gröbner basis seeks to solve this question by imposing an order on the division for multivariate polynomials. We first order the monomials in $\mathbb{K}[x_1, \ldots, x_n]$ by some total order (called an *admissible order*) such that

- $m_1 < m_2 \implies m_1 * m < m_2 * m$
- $1 \leq m \ \forall m$, so 1 is the canonical smallest element

We note that there are many possible orders, but we must fix a single order. An example of an admissible total ordering is lexicographic ordering in which $x_i \geq x_{i-1}^d \ \forall d$. Thus we prioritize all powers of x_i before any power of x_{i-1}. Another admissible total ordering is one where $\sum d_i \leq \sum e_i$ implies $x^d < x^e$, and ties are broken in lexicographic order.

The choice of monomial orderings is important to the complexity of the solution. Certain orderings might quickly reveal that a function is not in the ideal, while other orderings will reveal this after more work has been completed. Furthermore, there exist ideals such that even an optimal ordering of the monomials cannot be used to solve the membership problem faster than doubly exponential time. Thus Buchberger's algorithm attempts to limit the amount of work done given a particular monomial ordering.

However, even with the monomial ordering, we have not completed the notion of dividing a polynomial by an ideal. This problem can be tricky to solve, even in the univariate case. Consider x(x+1) and x(x+2). This is the ideal generated by x. However, until we realize this, it is difficult to answer what the remainder should be when we divide x^2. We can report a remainder of x or 2x, but neither is the correct answer, since x^2 is in the ideal generated by x.

Thus we need to understand the relationship between the polynomials. This brings us to the notion of GCD. If we are able to compute the GCD of x(x+1) and x(x+2), we can learn that x is generating these polynomials.

However, the GCD is defined primarily for univariate polynomials because they can belong to a principal ideal domain. This is not true for multivariate polynomials.

For example, consider the set $\{x^i y^{d-i} | 0 \leq i \leq d\}$ which generates polynomials of total degree at least d. It is clear that this is a minimal generating set. Every other monomial $x^i y^j$ for $i + j < d$ is not in there, and no member of this set can be generated by any other. Thus d monomials are necessary for ideal, even if the polynomial has at most 2 variables. This illustrates that there is no a priori bound on the number of elements necessary to generate the ideal that is based on the number of variables. Thus every ideal has a finite generating set, but the size can be arbitrarily large.

Thus when we are discussing ideals, we need to describe the notion of the canonical remainder, and we want the remainder calculation to be "not too variable". The Gröbner basis gives us a very nice, unique remainder, once we fix the ordering of the polynomials.

The Gröbner basis is useful by reducing the complexity of the polynomials by not focusing on the entire polynomial at once. Instead, it focuses on the leading term in polynomial first, and considers the ideal formed by the leading terms of the polynomials.

Thus, we describe the useful properties of the Gröbner basis in more detail.

3 Gröbner Basis

We begin by formally defining the leading term of a polynomial

Definition 3.1. Leading Term: Given a polynomial f = sum $c_d x^d$, the leading term or LT(f) = $c_d x^d$ where x^d is maximal monomial according to the total ordering. The leading monomial, denoted LM(f), is thus x^d.

Then we can use the leading terms to define a weak notion of a canonical remainder. Given a monomial f and polynomials $< g_1, ...g_t >$, we want to repeatedly reduce the power

of f by scaling out the powers of g_i. This gives us a remainder $r = f - \sum \alpha_i g_i$ such that for every monomial m $\in$ f $\forall i \in [t]$, m is not divisible by $LT(g_i)$. Due to the total ordering of the g_i, we guarantee that the remainder has a strictly decreasing leading term.

Considering our earlier example, with x^2, and the ideal of x(x+1) and x(x+2), it is clear that x^2 should not be a remainder because its power has not strictly decreased. However, r = x and r = 2x are valid remainders.

Now, we consider monomial ideals, which are simply ideals generated by monomials. If we have a collection of monomials M = $m_1, ...m_t$ it is trivial to determine whether a particular monomial m is in the ideal generated by (M) since $m \in (M) \iff \exists$i such that $m_i \mid m$

This gives us an intuition for Dickenson's Lemma.

Lemma 3.2 (Dickson's Lemma)**.** *Let $J \subseteq \mathbb{K}[x_1, \ldots, x_n]$ be a monomial ideal, that is, an ideal generated be a (possibly infinite) set of monomials. Then J is finitely generated, by a finite set of monomials.*

Proof Sketch. The proof is by induction on the number of variables. We will sketch how the proof goes for $n = 2$. Consider an ideal J, with monomial $x^i y^j$. Now observe that if we have another monomial $x^k y^l$ which is a multiple $x^i y^j$, then the monomial $x^k y^l$ is "covered" by $x^i y^j$. In particular, in the set of monomial generators for J, we can discard any such $x^k y^l$.

Now consider those monomials of the form $x^k y^l$ for any fixed $k < i$. These monomials are only really on one variable, y, so we can appeal to induction to show that for any fixed $k < i$ that set of monomials is also finitely generated. We can also make the same argument for any fixed $l < j$. As there are a finite number of $k < i$, and a finite number of $l < j$, we can simply union all of these generators together, and thus generate the entire space of monomials. □

Thus, we can finally define a Gröbner basis.

Definition 3.3. $\langle g_1, \ldots, g_t \rangle$ is a **Gröbner basis** for the ideal $J := \langle f_1, \ldots, f_m \rangle$ if: $g_1, \ldots, g_t \in J$, and $\langle LT(g_1), \ldots, LT(g_t) \rangle = \langle LT(J) \rangle$.

Thus the ideal formed by $\langle LT(g_1), \ldots, LT(g_t) \rangle$ is a monomial ideal. If we take the ideal generated by the leading terms of all of the polynomials in J, then this ideal is also generated by the leading terms of the g_i alone. They have a special structure, and in particular the following, easily proven, fact holds. Given a monomial $\vec{x}^{\vec{a}}$ in a monomial ideal generated by $\{\vec{x}^{\vec{b}}\}_{\vec{b} \in S}$, where S is possibly infinite, it must be that there is some $\vec{b} \in S$ such that $\vec{x}^{\vec{b}}$ divides $\vec{x}^{\vec{a}}$. Thus, these conditions imply that our above reduction step $f \mapsto f - g \cdot LT(f)/LT(g)$ can always make progress when working on a Gröbner basis, as there will always some g so that $LT(g)$ divides $LT(f)$. We will show shortly that these conditions on the Gröbner basis also imply that the g_i also generate J.

Note that we have not yet shown that $\langle g_1, \ldots, g_t \rangle$ form a basis of J. However, this tells us g_i are sufficiently representative of J. There may be multiple Gröber bases for an ideal J, but modulo simple translations, the Gröber basis is unique.

Additionally, remainders are unique with respect to a Gröber basis. This also relies on the idea of the leading terms.

Lemma 3.4. *Let $g_1, \ldots, g_t$ be a Gröbner basis for the ideal $J = \langle g_1, \ldots, g_t \rangle$. Then for any f, the weak remainder with respect to the g_i's is unique.*

Proof. Suppose $f = r_1 + \sum a_i g_i = r_2 + \sum b_i g_i$ are two weak remainders of f. Then $r_1 - r_2 = \sum (a_i - b_i) g_i \in J$. As the g_i form a Gröbner basis, it follows that if $r_1 - r_2$ is non-zero then its leading monomial of $r_1 - r_2$ must be divisible by some $LT(g_i)$. But the monomials of $r_1 - r_2$ are a subset of the union of the monomials of r_1 and r_2, and none of those monomials are divisible by any g_i. Thus, it follows that $r_1 - r_2$ must be zero, so $r_1 = r_2$. Thus the remainder is unique. □

Finally, we note that the remainder is unique given an ideal and a total ordering, without requiring the basis as a condition of uniqueness.

As a special case of this, if f is in the ideal, the remainder should be 0. If f is in J, then the of remainder f by $\langle g_1, ...g_t \rangle$ is 0. We know that LT(f) ∈ LT(J). Thus we would get some reduction in degree since $\exists$ g_i such that $LT(g_i) | LT(f)$ so we subtract $a_i g_i$ from f to get lower degree. We continue until have 0, with sequentially lower degree polynomials along the way.

As a consequence of this, we have proved Hilbert's Basis Theorem.

Theorem 3.5. *If $J \subset K[x_1, ..., x_n]$ is an ideal, then there are finitely many polynomials $f_1, ..., f_n \in J$ so that $J = \langle f_1, ...f_n \rangle$.*

Therefore, if we have a Gröbner basis, we can complete membership testing.

4 Construction of Gröbner Bases

Having shown that Gröbner bases solve the ideal membership problem, we now show an algorithm, that runs in finite time, for constructing these objects. The essential problem is that given the set of polynomials $f_1, ...f_n$, can we construct the basis $g_1, ...g_m$. Thus we need to do nontrival work to find polynomials that were not previously represented. For example, given $x^2 + x, x^2 + 2x$,we need to complete cancellations to realize the common x term. But how can we to do this cancellation within a time bound?

This is the essence of Buchberger's algorithm, in which shows how to transform any set of generators to the Gröbner basis. The algorithm to solve this problem will require the *syzygy*, which will allow us to cancel the high-degrees of polynomials.

Definition 4.1. Let f and g be two monic polynomials. Let $m = \text{LCM}(LT(f), LT(g))$. Define $S(f, g)$, the *syzygy* of f and g, to be $S(f, g) = mf/LT(f) - mg/LT(g)$.

Note that this produces the desired cancellation, though the remainder r may be larger than f or g.

We now give the Gröbner basis algorithm, starting with the polynomials $f_1, \ldots, f_t$. We use the operator mod to denote the canonical weak remainder.

- $B \leftarrow \{f_1, \ldots, f_t\}$
- iterate until no additions: if $\exists g_i, g_j \in B$ so $r := S(g_i, g_j) \bmod B$ has $r \neq 0$, then $B \leftarrow B \cup \{r\}$.

- output B.

We note that the syzgy S(f, g) is in the ideal generated by f and g.

The algorithm terminates in a finite amount of time. Consider the ideal $\langle LT(B)\rangle$ over the course of the algorithm, which grows in size, and the ideals generated by LT(B). Then the first ideal $I_1 = \langle LT(f_1), ..., LT(f_k)$, $I_2 = \langle LT(f_1), ..., LT(f_k), LT(r_1)$, and so forth. Thus the sequence of ideals is $I_1 \subset I_2 \subset ... \subset I$ is finitely generated. Then I is a monomial ideal, and by Dickson's lemma, it has a finite basis, and therefore the algorithm terminates finitely.

Next we prove correctness. At termination, the syzgyies have a remainder of 0. But we still need to show that LT(B) will generate our ideal.

Lemma 4.2. *Let* $J = \langle g_1, \ldots, g_t\rangle$. *If* $S(g_i, g_j) \bmod \{g_1, \ldots, g_t\} = 0$ *for all* i, j, *then* $\forall f \in J = (f_1, ..., f_k)$, $\langle LT(f)\rangle = \langle LT(g_1), \ldots, LT(g_t)\rangle$.

Proof. Consider any $f \in J$. Then $f = \sum_{j=1}^{k} m_j g_{i_j}$, where each m_j is a monomial. Thus following condition holds

- $\text{degree}(m_i g_{i_j}) \geq \text{degree}(m_{i+1} g_{i_{j+1}}))$

Then we want to show that $\langle LT(f)\rangle = \langle LT(g_1), \ldots, LT(g_t)\rangle$. Suppose for a contradiction that this statement was false. Then we want to find the polynomial with the smallest $m_1 g_{i_1}$ that violates the statement. Now, $m_1 g_{i_1}$ is divisible by the claim. This isn't sufficient because there could be a cancellation between $m_1 g_{i_1}$ and $m_2 g_{i_2}$, and a lower order term is the leading term of f.

Then consider the following f with leading terms $m_1 g_1$ and $m_2 g_2$ with the same leading terms. They we must have taken the syzgy of $m_1 g_1$ and $m_2 g_2$ and multiplied it by some polynomial. Then $m_1 = m' a_1$ and $m_2 = m' a_2$. Thus $a_1 g_1 - a_2 g_2 = syz(g_1, g_2)$. Note that $m_1 g_{i_1} - m_2 g_{i_2}$ has cancellation at the leading term, as these two polynomials have the same multi-degree. Thus, there exists a monomial w such that $m_1 g_{i_1} - m_2 g_{i_2} = wS(g_{i_1}, g_{i_2})$. As $S(g_{i_1}, g_{i_2}) \bmod \{g_1, \ldots, g_t\} = 0$ we have that $S(g_{i_1}, g_{i_2}) = \sum q_i g_i$, and as this is by the weak remainder algorithm, we have $\text{mdeg}(w) + \text{mdeg}(q_i g_i) \leq \text{mdeg}(w) + \text{mdeg}(S(g_{i_1}, g_{i_2})) < \text{mdeg}(m_1 g_{i_1})$. So we can express

$$m_1 g_{i_1} - m_2 g_{i_2} = \sum (a_i w) g_i$$

as desired, as the right hand side has lower multi-degree than the $\text{mdeg}(m_2 g_{i_2})$. □

So putting this all together, the algorithm must terminate with a set of polynomials, whose syzygies have zero weak remainder on this set. This then implies the set is a Gröbner basis for itself, and as it contains the f_i, is a basis for the f_i. We can then use this basis for testing membership in the ideal $\langle f_1, \ldots, f_m\rangle$.

Acknowledgements Many of the proofs have been drawn from Lecture 15 ST12 scribe notes by Michael Forbes.