Numerical Solution of Systems of Polynomial Equations

Diyang Ma, u5626686

October 24, 2018

Acknowledgements

To my supervisor, Professor Markus Heglend: thank you for the invaluable advice and guide.

To my parents, Yueyong and Haimei: for the unending support whenever I need.

To Abhishek: the hours of discussion and advice builds the foundation of my research. To Yuancheng: for always sharing information with me.

To Zhidi: for your company whenever sad or glad.

To Yihan: for hours of discussion and game.

To Bolin: for the first job you introduced.

To Dingqian: for the best roommate I have ever met.

To Shiqiu: for preparing meals for me when I am alone.

To Yossi and Shuai: the hours of conversation and listening helped me push through.

And finally, to all my friends and family, thank you for long-time support, both mathematical and emotional.

Contents

Chapter 1

Introduction

Polynomials appear in many areas of mathematics and science. The structure of polynomials is relative simple as it only involves four kinds of operations which are multiplication, addition, subtraction and non-negative integer exponent. One can even treat non-negative integer exponent as a multiplication of same variables. And in fields like $\mathbb{R}$ and $\mathbb{C}$ where 0 is the additive identity, subtraction is another form of addition. It seems we are working with some simple object that only involves addition and multiplication, which we learned from very young age. However as the study of polynomials proceeds, problems occur all the time. In this thesis we are going to have a glimpse of the problem that how to find roots of polynomials.

We can evaluate a univariate polynomial by replacing its variable by any number and carrying out the computation. For a trivial case like $f(x) = x + 3$, replacing x by 2 gives $f(2) = 2 + 3 = 5$. Sometime we are interested in another kind of question, *i.e.* replacing x by which number gives $x + 3 = 7$. It is easy to check $f^{-1}(y) = y - 3$. The inverse problem is not obvious in general. Most polynomial equations are not bijective like x^2. And as a result, its inverse is not even a function. In the case above x^2 is still a univariate polynomial with only one term. the problem becomes tricky as the polynomial becomes complicated.

Given a polynomial $p : K^s \to K$ and $y \in K$, finding the set

$$\{x \in K^s \mid p(x) = y\},$$

is equivalent to find the set

$$\{x \in L^s \mid p(x) - y = 0\}.$$

Moreover as y is a constant in K, set $q(x) = p(x) - y$ then q is also a polynomial from K^s to K. Then to find the inverse image of a polynomial function is equivalent to find roots of a some other polynomial. And that starts our thesis to find the roots of a polynomial.

1.1 Finding the roots of a polynomial

The cost to find the roots of a polynomial depends on the polynomials. For a basic case where the polynomial is linear, the roots of the polynomial is just a hyperplane in its domain. For example $f_1 : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f_1(x, y) = x + y - 1.$$

And the set of roots is the line $y = -x + 1$ in $\mathbb{R}^2$. Although f is a polynomial which is not injective and has uncountable roots, the roots form a line in $\mathbb{R}^2$. Now consider a quadratic polynomial $f_2 : \mathbb{R}^2 \to \mathbb{R}$

$$f_2(x, y) = x^2 + y^2 + 2xy - 1.$$

By factorization $f_2(x, y) = (x + y + 1)(x + y - 1)$. Then the roots of f_2 are two parallel lines in $\mathbb{R}^2$. Comparing f_2 and f_1 we find f_2 is a multiplication of f_1 and another linear polynomial. However the roots of f_2 is not as easy as f_1 to identify without factorization. Moreover factorization of a polynomial is not always obvious and sometimes finding factorization of a polynomial is equivalent to find roots as in example above. If we change f_2 a little by subtracting xy and gives

$$f_3(x, y) = x^2 + y^2 + xy - 1.$$

The roots of f_3 is the ellipse in Figure 1. One can also generate it by clockwise rotating the ellipse $2x^2 + 6y^2 = 3$ by $\frac{\pi}{4}$. Although the set of roots of f_3 is regular, it is not easy to find by just inspection. As the degree of polynomial increases, the pattern of roots could be more complicated and irregular.

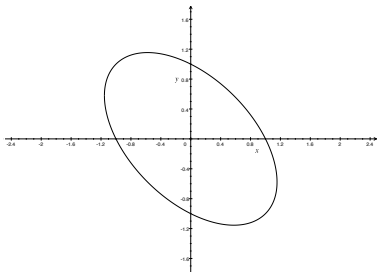

Figure 1.1: $f_3(x, y) = 0$

The problem of finding the roots of a polynomial is a major topic of algebraic geometry. There exists various methods to find roots. For a univariate quadratic equation of the form $x^2 + px + q = 0$, one can find its roots immediately if factorization is possible. Quadratic formula is an alternative which gives a formula for solution directly. Inspired by factorization, it is always possible to

transform the equation to a system of two variables [1]

$$\begin{cases} x_1 + x_2 + p = 0 \\ x_1 x_2 - q = 0. \end{cases}$$

This method transforms an univariate polynomial to an equivalent multi-variable polynomial system. The logic behind is that if one can factorize $x^2 + px + q$ as $(x - x_1)(x - x_2)$, it then leads to the system above.

The methods above all solve quadratic univariate polynomials. Unfortunately, there is no specified method for general polynomials. Actually quadratic equation is one of relatively basic polynomials and easy to solve. For general cases, scientists developed several numerical methods like bisection method based on the fact that polynomial is continuous or Newton–Raphson method which considering the derivative of a function. These methods are effective to find an approximate root.

1.2 Dimension of roots

An univariate polynomial equation always have finite roots both in $\mathbb{R}$ and $\mathbb{C}$, which is a result of the fundamental theorem of algebra. Real univariate polynomial are functions from $\mathbb{R}$ to $\mathbb{R}$ and the roots is a set of discrete points in $\mathbb{R}$. However for multi-variable polynomial over field K. the function is a map from $K^s \to K$, $s > 1$, a map from higher dimension space to a single line. In such the set of roots is not discrete in general.

One can see the roots of $x^2 + y^2 + 2xy - 1 = 0$ are two parallel lines and the roots of $x^2 + y^2 + xy - 1 = 0$ is a ellipse. The functions above both maps $\mathbb{R}^2$ to $\mathbb{R}$ and the set of roots are one dimensional. For a non-zero polynomial f maps K^s to K where K is a field, the dimension of roots is always less than s. In this thesis we are more interested in polynomial systems of zero dimension root.

1.3 System of polynomial equations

Let S be the set of all polynomials map from K^s to K over a field K. Then for any polynomials f in S, the roots of $f = 0$ is a subset of K^s which has dimension less than s. Now consider the a system of polynomials $f_1, ..., f_n \in S$. The roots of $f_1 = ... = f_n = 0$ are just the intersection of roots of $f_1 = 0, ..., f_n = 0$. Consider the case

$$\begin{cases} 2x_1 + 3x_2 - 1 = 0 \\ x_1 + 2x_2 - 1 = 0, \end{cases}$$

over field $\mathbb{R}$. The roots of two polynomial equations are lines in $\mathbb{R}^2$

$$\begin{cases} x_1 = -\frac{3}{2}x_2 + \frac{1}{2} \\ x_1 = -2x_2 + 1. \end{cases}$$

And the root of the polynomial system is the intersection of two lines where

$$-\frac{3}{2}x_2 + \frac{1}{2} = -2x_2 + 1.$$

However to find roots of a system of polynomial equations, one can not always calculate the roots of each polynomial equation and take the intersection. Methods like factorization fails in general as the roots of one single polynomial equation may be uncountable. The first step to solve a system of polynomial equations often involves elimination. One can see even from the example above, the variable x_1 is eliminated in the process. In linear algebra, we can write the system of polynomial equations as matrix multiplication

$$\begin{bmatrix} 2x_1 + 3x_2 \\ x_1 + 2x_2 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$

With Gaussian elimination the second row is eliminated to

$$\begin{bmatrix} 2 & 3 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ -\frac{1}{2} \end{bmatrix}.$$

We see from that process the system of polynomial equations is transformed to a triangular system. With similar idea of variable elimination, Gaussian elimination eliminates variables one by one and ends up with a polynomial of only one variable. The direct advantage of triangular system is that one can solve an univariate polynomial equation each time and the roots are always finite.

1.4 Linear spaces of polynomials

In the example above, we find that the system of polynomial equation can be represented as a multiplication of matrices in the form

$$\begin{bmatrix} 2 & 3 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ -\frac{1}{2} \end{bmatrix}.$$

One question is that if there is a general form that every system of polynomial could be represented in this way. Before that we first give a formal definition of polynomial [2].

We notice that each polynomial is a sum of monomials times some coefficients. In later chapter one may see that monomial is a fundamental unit of polynomial and hence we start with monomial.

Definition 1.1 A *monomial* in s variables $x_1, ..., x_s$ is a product if powers of the variables, *i.e.*

$$x_1^{j_1} x_2^{j_2} \cdots x_s^{j_s},$$

with $j_i \in \mathbb{N}_0$. We abbreviate it as x^j, where $j = (j_1, ..., j_s) \in N_0^s$ is a tuple

$$x^j = x_1^{j_1} x_2^{j_2} \cdots x_s^{j_s}.$$

The total degree of a monomial x^j is $|j| = j_1 + \cdots + j_s$. The set of all monomials in s variables will be denoted by $\mathcal{T}^s$ and $\mathcal{T}_d^s \subset \mathcal{T}^s$ is the set with total degree less than or equal to d. For instance, $\mathcal{T}_1^s$ is the set of monomials with degree 1

$$\{1, x_1, ..., x_s\},$$

and the set $\mathcal{T}_d^1$ is

$$\{1, x, ..., x_d\}.$$

We can see from the example above $\mathcal{T}_d^s$ contains the monomial 1.

Proposition 1.2 $\mathcal{T}_d^s$ contains $\binom{d+s}{d}$ monomials; $\binom{d+s-1}{d}$ of these have total degree d [2].

The proposition follows from the fundamental result of combinatorics. One can see that the number of monomials grows rapidly with the number of variables s and degree d and the rapid growth is a major reason for the high computational complexity of many polynomial algorithm.

Definition 1.3 A *complex (real) polynomial* in s variables is a finite linear combination of monomials from $\mathcal{T}^s$ with coefficients in $\mathbb{C}$ or $\mathbb{R}$,

$$p(x) = p(x_1, ..., x_s) = \sum_{(j_1,...,j_s) \in J} \alpha_{j_i,...,j_s} x_1^{j_1} \cdots x_s^{j_s} = \sum_{j \in J} \alpha_j x^j.$$

The set $J \subset \mathbb{N}_0^s$ which contains the exponents of those monomials which are present in the polynomial p is called the *support* of p. One can express a polynomial as $p(x) = a^T v(x)$. where $a \in \mathbb{C}^{|J|} (\mathbb{R}^{|J|})$ and $b(x) = \left[x^j\right]_{j \in J}$ are vectors of scalars and monomials respectively. Consider the example,

$$p = x_1 x_2 + 2x_2 + 1,$$

can be written as

$$p(x) = \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 x_2 \\ x_2 \\ 1 \end{bmatrix} = a^T b(x).$$

Here we have coefficients in vector a and monomials in vector $b(x)$.

Chapter 2

Polynomials and Algebra

The product of two monomials $x^j, x^{\tilde{j}} \in \mathcal{T}^s$ is

$$x^j \cdot x^{\tilde{j}} = x^{j+\tilde{j}}.$$

The multiplication is commutative and associative.

Let $\mathcal{P}^s$ or $k[x_1, ..., x_s]$ denote the set of polynomials in s variables where k is the field ($\mathbb{R}$ or $\mathbb{C}$). $\mathcal{P}^s$ is also commutative under multiplication. One can see for any two polynomials $p_1, p_2 \in \mathcal{P}^s$, let $p_i = \sum_{j \in J_i} \alpha_{i,j} x^j$ where $i \in \{1, 2\}$ and

$$p_1 \cdot p_2 = \sum_{j \in J_1} \sum_{\tilde{j} \in J_2} \alpha_{1,j} \alpha_{1,\tilde{j}} x^{j+\tilde{j}} = p_2 \cdot p_1 \in \mathcal{P}^s.$$

As $\mathcal{P}^s$ is also distributive and associative under multiplication with identity element $f(x) = 1$ [2], $\mathcal{P}^s$ is a commutative ring. And we can now consider some algebraic properties of $\mathcal{P}$.

Proposition 2.1. The multiplication by a fixed polynomial $p \in \mathcal{P}^s, p \neq 0$ is a regular linear mapping in $\mathcal{P}^s$, i.e. $q \in \mathcal{P}^s, p \cdot q = 0 \Rightarrow q = 0$.

We can give a sketch of proof as following. When considering a non-zero polynomial $p \in \mathcal{P}^s$, let its leading term be $\alpha_j x^j$. Although we are still not giving a definition of leading term, one can assume the leading term is the monomial in f with largest total degree or the largest term with some specified order. Then for any $q \in \mathcal{P}^s$, the leading term of q is not zero if q is not zero and hence the product of leading terms of p and q will not be canceled. Then the corollary follows.

Corollary 2.2. The polynomial rings $\mathcal{P}^s$ are *integral domains, i.e.* $p_1 \cdot p_2 = 0$ implies $p_1 = 0$ or $p_2 = 0$ for $p_1, p_2 \in \mathcal{P}^s$

As $\mathcal{P}^s$ is a ring, for a set of polynomials $\{f_1, ..., f_n\} \subset \mathcal{P}^s$, we can define a ideal generated by $f_1, ..., f_n$.

Definition 2.3 A set of polynomials in $\mathcal{P}^s$ which is closed under linear combination is a *polynomial ideal* in $\mathcal{P}^s$. The ideal which consists of the linear combinations of the polynomials $p_\nu \in \mathcal{P}^s, 1 \leq \nu \leq n$ is denoted by

$$\langle f_1, ..., f_n \rangle.$$

System of polynomial equations has close relation with polynomial ideal. Consider a system of polynomial equation $f_1 = ... = f_n = 0$ with $f_i \in \mathcal{P}^s$ and the polynomial ideal

$$I = \langle f_1, ..., f_n \rangle.$$

$p = 0$ for any polynomial $p \in I$ since

$$p = \sum_{i=1}^{n} c_i f_i, \text{ with } c_i \in \mathcal{P}^s.$$

Or in other words any solution of the system of polynomial equations is also a solution of p. Conversely, if z is a root of any polynomial $p \in I$, z is then a root for each f_i and hence is solution of the system of polynomial equations.

Finally we give a definition for the set of points where all polynomials in I vanish.

Proposition 2.4. Consider a set Z of points in $\mathbb{C}^s$. The set

$$\mathcal{I}_Z := \{f \in \mathcal{P}^s : f(z) = 0 \ \ \forall z \in Z\}$$

of all polynomials which vanish at each $z \in Z$ is an ideal in $\mathcal{P}^s[2]$ and

Definition 2.5. $z \in \mathbb{C}^s$ is a *zero of the polynomial ideal* $I \subset \mathcal{P}^s$ iff $f(z) = 0$ for all $f \in I$. The *zero set of* I is denoted by

$$Z[I] := \{z \in \mathbb{C}^s : z \text{ is a zero of } I\}$$

To find roots of a system of polynomial equations $f_1 = f_2 = \cdots = f_n = 0$, it is equivalent to find zeros of the polynomial ideal $I = \langle f_1, \cdots, f_n \rangle$. Note if a set of polynomials $\{p_1, \cdots, p_t\}$ is also a generator of I, then the roots of $p_1 = p_2 = \cdots = p_t = 0$ are same as zeros of I as a consequence of definition **2.3**. Generally the generator of an polynomial ideal is unique. For instance consider following ideals $\langle f_1, f_2 \rangle$ and $\langle f_2, f_3 \rangle$

$$\begin{cases} f_1 = xy + 1 \\ f_2 = x + y \\ f_3 = y^2 - 1. \end{cases}$$

Here we have $f_3 = yf_2 - f_1$ and $f_1 = yf_2 - f_3$. As f_3 is an element of $\langle f_1, f_2\rangle$ and conversely f_1 is an element of $\langle f_2, f_3\rangle$, $\langle f_1, f_2\rangle$ and $\langle f_2, f_3\rangle$ are actually same. It follows from proposition 2.4 that the zero sets of $\langle f_2, f_3\rangle$ and $\langle f_1, f_2\rangle$ are same.

Although the two systems above have the same solution, the generator $\{f_2, f_3\}$ is easier to work with as f_3 is a univariate polynomial. As a polynomial ideal could have many generators and finding zero set of the polynomial ideal is equivalent to find roots of a system polynomial equations consists of elements in generator, we would like to work with generator with simpler structure.

2.1 Polynomial Division

Given an polynomial ideal I, it is natural to ask if a polynomial p is one of its elements. One may even want to work out a quotient ring based on I. These questions rely on a definition of division in polynomial.

2.1.1 Univariate polynomial

Given two polynomials $f, g \in k[x]$, we can divide f by g, producing a unique quotient q and remainder r such that

$$f = qg + r,$$

with r has degree strictly less than g or $r = 0$. For a polynomial ideal $\langle g\rangle$ with $g \in k[x]$, $f \in k[x]$ is an element of $\langle g\rangle$ when $f = qg + r$ with $r = 0$. This is equivalent to say that f is a multiple of g.

Consider two polynomials

$$\begin{aligned} f &= x^3 + 2x^2 + 3 \\ g &= x + 2. \end{aligned}$$

Here f and g are univariate polynomials and division algorithm gives

$$f = x^2 g + 3.$$

2.1.2 Ordering

Now consider multivariate cases. For two polynomials f and g in more than one variable, the choice to divide f by g may not be unique. Consider the following example with two variables.

Example 2.6.

$$\begin{cases} f = x^2y + xy^3 \\ g = x^2 + xy. \end{cases}$$

Directly we have at least two ways to divide f,

$$f = yg + xy^3 - xy^2$$
$$f = (y^2 - xy + x^2 + x)g - x^4 - x^3.$$

In the first equation $xy^3 - xy^2$ is not divisible by x^2 and in the second one $-x^4 - x^3$ is not divisible by xy. On the other hand $xy^3 - xy^2$ is further divisible by xy in g and $-x^4 - x^3$ is also divisible by x^2. It seems end up with an endless division. Note in univariate polynomial, one can divide f by the term of g with largest degree and the remainder is unique. However it is not clear which term has the largest degree in multi-variable polynomial.

It is then critical to ask if there is a leading term in multi-variable polynomial analogous to the term in univariate polynomial that has largest degree. Here we introduce the definition of ordering.

Definition 2.7. A *monomial order* on $k[x_1, ..., x_n]$ is any relation $<$ on the set of monomials x^α in $k[x_1, ..., x_n]$ satisfying: [3]

1. $<$ is a *total ordering* relation.
2. $<$ is *compatible with multiplication* in $k[x_1, ..., x_n]$, in the sense that if $x^j < x^{\tilde{j}}$ and x^k is any monomial, then $x^j x^k < x^{\tilde{j}} x^k$.
3. $<$ is *well-ordering*. That is, every non-empty collection of monomials has a smallest element under $<$.

For instance $1 < x < x^2 < \cdots < x^n$ is an order in $\mathcal{P}^1$. One can check that $x^i < x^j$ implies $x^{i+k} < x^{j+k}$ where $i, j, k \in \mathbb{N}_0$.

Definition 2.8. Let x^j and $x^{\tilde{j}}$ be monomials in $\mathcal{T}^s$. We say $x^j >_{lex} x^{\tilde{j}}$ if the difference $j - \tilde{j} \in \mathbb{Z}^n$, the left-most nonzero entry is positive [3] and the order is called *Lexicographic order*.

Lexicographic order is also called dictionary order which is analogous to the ordering of words in dictionary.

Definition 2.9. For a polynomial $f \in \mathcal{P}^s$ in the form

$$f = \sum_{j \in J} \alpha_j x^j,$$

and an ordering $<$, the *leading term* $LT_<(f)$ of f is the term $\alpha_{\tilde{j}} x^{\tilde{j}}$, $\alpha_{\tilde{j}} \neq 0$ such that $x^j < x^{\tilde{j}}$ if $\alpha_j x^j$ is a term of f where $\alpha_j \neq 0$. And *leading monomial* $LM_<(f)$ is the monomial part of $LT_<(f)$.

By the definition of ordering, a polynomial has unique leading term given an order. In most cases we use Lexicographic order in this thesis and except explicit notation specified, otherwise Lexicographic order is default. Now consider (2.6) again. Once we specify that $x >_{lex} y$, then $LT(f) = x^2y$ and $LT(g) = x^2$, the division algorithm gives

$$f = yg + xy^3 - xy^2.$$

In such case $xy^2 < xy^3 < x^2$ and the division algorithm terminates. One can also check that

$$f = (y^2 - xy + x^2 + x)g - x^4 - x^3,$$

if $y >_{lex} x$.

Proposition 2.10. Given $f, g \in \mathcal{P}^s$ and $<_{lex}$, $q, r \in \mathcal{P}^s$ are unique where $f = qg + r$ by division algorithm.

Assume that there are at least two different pairs of q_1, r_1 and q_2, r_2, and

$$f = q_1g + r_1 = q_2g + r_2.$$

If $q_1 \neq q_2$ one have

$$r_1 = (q_2 - q_1)g + r_2.$$

The equation above indicates either r_1 or r_2 is further reducible by g, which is a contraction. Now consider the case $q_1 = q_2$ but $r_1 \neq r_2$ which is a direct contradiction that $q_1g + r_1 = q_2g + r_2$. Then the pair of $q, r \in \mathcal{P}^s$ such that $f = qg + r$ is unique.

Definition 2.11. Let x^j and $x^{\tilde{j}}$ be monomials in $\mathcal{T}^s$. *Graded Reverse Lexicographic Order* is the order that $x^j <_{grevlex} x^{\tilde{j}}$ if $\sum_{i=1}^s j_i < \sum_{i=1}^s \tilde{j}_i$, or if $\sum_{i=1}^s j_i = \sum_{i=1}^s \tilde{j}_i$, and in the difference $j - \tilde{j} \in \mathbb{Z}^n$, the right-most nonzero entry is positive [3].

Let $x_1 > x_2 > x_3$. For monomials in $\mathcal{T}_2^3$ with Lexicographic Order one have

$$x_1^2 > x_1x_2 > x_1x_3 > x_2^2 > x_2x_3 > x_3^2.$$

But with Graded Reverse Lexicographic Order

$$x_1^2 > x_1x_2 > x_2^2 > x_1x_3 > x_2x_3 > x_3^2.$$

Here all monomials above has total degree of 2. In the case above the difference between two ordering is that $x_1x_3 >_{lec} x_2^2$ while $x_2^2 >_{grevlec} x_1x_3$.

Remember a polynomial $f \in \mathcal{P}_d^s$ can be written as a product of two vectors. One

can sort elements in $b(x)$ with a specified order. For instance consider $f \in \mathcal{P}_2^2$

$$f = x_1^2 + 2x_1x_2 + x_2 = \begin{bmatrix} 1 & 2 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1^2 \\ x_1x_2 \\ x_2^2 \\ x_1 \\ x_2 \\ 1 \end{bmatrix}.$$

Here monomials in $\mathcal{T}_2^2$ are sorted by Graded Reverse Lexicographic Order.

2.1.3 Multi-variable polynomial

For an univariate polynomial f, g and corresponding polynomial ideal $\langle g \rangle$, one can check that if $f \in \langle g \rangle$ by division. For instance

$$\begin{cases} f = x^3 - x \\ g = x^2 + x. \end{cases}$$

As $f = qg + r$ with $q = x - 1$ and $r = 0$, $f \in \langle g \rangle$.

Proposition 2.12. $f \in \langle g \rangle$ iff $f = qg + r$ with $f, g, q \in k[x]$ which indicates $r = 0$.

One can check if $r = 0$ and then $f = qg$, which further indicates $f \in \langle g \rangle$. Conversely if $f \in \langle g \rangle$, then $f = qg$ for some $q \in k[x]$. By proposition **2.10**, $f = qg + r$ is unique and $r = 0$.

For polynomial ideal I generated by multi-variable polynomials in $\mathcal{P}^s$, we also want to know if a polynomial $f \in \mathcal{P}^s$ is an element of I. Especially is that possible to determine if $f \in I$ by division just like how we did with univariate polynomials.

Division of multi-variate polynomials comes as following.

Definition 2.13. Given a monomial order $<$ in $\mathcal{P}^s$, and a set of polynomials $\{f_1, ..., f_n\}$ where $f_i \in \mathcal{P}^s$. For every $f \in \mathcal{P}^s$, one can write f by *Division algorithm* as [3]

$$f = \sum_{i=1}^{n} a_i f_i + r,$$

where $a_i, r \in \mathcal{P}^s$ and either $r = 0$ or r is a linear combination of monomials with constant coefficients and none of those is divisible by $LT_<(f_i)$ for $1 \leq i \leq n$. In such case r is called a remainder of f on division by I.

2.2 Gröbner basis

2.2.1 Uniqueness of remainder

It is natural to ask if $f \in I$ is equivalent to the remainder of f on division by I is 0, where $I = \langle f_1, ..., f_n \rangle$ and $f, f_i \in \mathcal{P}^s$. It is easy to show that if $r = 0$, then f is a linear combination of $f_1, ..., f_n$ and hence $f \in I$. However the converse is not always true.

Example 2.14 Given a polynomial ideal $\langle g_1, g_2 \rangle$ where

$$\begin{cases} g_1 = xy + x \\ g_2 = x^2 y. \end{cases}$$

and $x >_{lex} y$, we want to determine whether $x^2 \in \langle g_1, g_2 \rangle$. As x^2 is neither divisible by xy or $x^2 y$, it leads to the conclusion that $x^2 \notin \langle g_1, g_2 \rangle$ from division algorithm. However one can find $x^2 \in \langle g_1, g_2 \rangle$ where

$$x^2 = xg_1 - g_2.$$

Unlike univariate polynomial ideal, one can not determine if a polynomial is an element of a polynomial ideal only by division algorithm. As $\langle f_1, ..., f_n \rangle$ is a fixed ideal with $f_1, ..., f_n \in \mathcal{P}^s$ and that leads to a question. How to tell if a polynomial is in an ideal.

Example 2.15 Even the remainder itself is not unique. Consider the case

$$\begin{cases} f_1 = xy^2 + y \\ f_2 = x^2 y + y, \end{cases}$$

and $f = x^2 y^2$ with $<_{lex}$. Then

$$f = xf_1 - xy = yf_2 - y^2.$$

One can see the remainders of f are $-xy$ and $-y^2$ separately. Considering a nonzero polynomial $f \in I$ where I is a polynomial ideal, it is always possible to find an element g in I where f is divisible by g since f itself is a polynomial in I satisfies the criteria. And that leads to the following definition.

2.2.2 Gröbner basis

Definition 2.16 Fix a monomial order $<$ on $\mathcal{T}^s$, and let $I \subset \mathcal{P}^s$ be an ideal. A *Gröbner basis* for I (with respect to $<$) is a finite collection of polynomials $G = \{g_1, ..., g_n\} \subset I$ with the property that for every nonzero $f \in I$, $LT_<(f)$ is divisible by $LT(g_i)$ for some i [3].

One can see a Gröbner basis of an ideal I is a subset of I and it has some nice properties we will show next.

Proposition 2.17 Fix a monomial order $<$ on $\mathcal{T}^s$, and let $I \subset \mathcal{P}^s$ be an ideal. Let $G = \{g_1, ..., g_n\} \subset I$ be a *Gröbner basis* for I. Then for any $f \in I$, the remainder of f on division by G is 0.

By the definition of Gröbner basis G of I, for any nonzero $f \in I$, f is divisible by some $g_i \in G$. Or equivalently

$$f = \sum_{i=1}^{n} a_i g_i + r,$$

Immediately $r \in I$ as $f, g_i \in I$. Here r is a remainder and is not reducible by any g_i. Consider the case that $r \neq 0$, by definition of Gröbner basis it implies there exists $g' \in G$ that $LT(r)$ is divisible by $LT(g')$, which is a contradiction that $r \in I$ is not reducible by any $g_i \in G$. Then r must be 0 and finish the proof of proposition **2.17**.

Proposition 2.18 For an ideal I, a Gröbner basis $G = \{g_1, ..., g_n\} \subset I$ is also a basis of I

This is a direct result from proposition 2.17 as for any $f \in I$, it is an linear combination of $g_1, ..., g_n$ with polynomial coefficients.

Proposition **2.17** says that for an ideal I one can determine if a polynomial is in an ideal with its Gröbner basis. However one question still remains. Does Gröbner basis exists for every polynomial ideal.

The answer is yes. Consider an polynomial ideal $I = \langle f_1, ..., f_n \rangle \subset \mathcal{P}^s$ and an order $<$. Then define the ideal:

$$LM(I) := \langle LM(f) \mid f \in I, f \neq 0 \rangle,$$

Then $LM(I)$ is the set of leading monomial of all elements in I. Note here $LM(I)$ is an ideal with generators of monomials. By Hilbert Basis Theorem, $LM(I)$ is Noetherian and hence finitely generated by polynomials. As each polynomial is also finitely generated by monomials, it implies that $LM(I)$ is finitely generated by a set of monomials $x^{j_1}, ..., x^{j_m}$.

One may see that Gröbner basis is not unique even the order is fixed. Note in vector spaces, basis is often a linear independent subset. However for any two polynomials $f, g \in \mathcal{P}^s$,

$$fg - gf = 0.$$

And hence f, g are not linear independent.

Proposition 2.19. Given an polynomial ideal $I = \langle f_1, ..., f_n \rangle \subset \mathcal{P}^s$ and an order $<$. For any monomial $x^j \in LM(I)$, there exist $g \in I$ such that $LT(g) = cx^j$ with some constant c.

One can see if $x^j \in LM(I)$ is a monomial, then it is a multiple of $LM(f)$ for some $f \in I$ and hence there exists $g \in I$ such that $LT(g) = cx^j$. Then for each monomial x^{j_i} in generator set, there exists g_i such that $LT(g_i) = cx^{j_i}$. It then shows that Gröbner basis exists for any polynomial ideal.

Proposition 2.20. Fix a monomial order $<$ and $I \subset \mathcal{P}^s$ with a Gröbner basis $G = \{g_1, ..., g_n\}$. Division of $f \in \mathcal{P}^s$ by $\{g_1, ..., g_n\}$ for I produces an expression $f = \sum_{i=1}^n c_i g_i + r$ where $c_i, r \in \mathcal{P}^s$ and no term in r is divisible by any element of $LT(I)$. If $f = \sum_{i=1}^n c_i' g_i + r'$ is another such expression, then $r = r'$ [3].

The proof is direct. Assume $r \neq r'$ then

$$r - r' = \sum_{i=1}^{n} (c_i' - c_i) g_i \neq 0.$$

The equations above shows that $r - r'$ is a linear combination of elements in G with polynomial coefficients, which indicates $r - r' \in I$. As G is a Gröbner basis, $r - r' \neq 0$ indicates it can be further divided by some $g \in G$ until $r - r' = 0$, which is a contradiction that $r - r' \neq 0$.

2.3 Polynomial ideal and Gröbner basis

Now we are going to have a look of the relations between polynomial ideals $\langle f_1, ..., f_n \rangle$ and the roots of system of polynomial equations $f_1 = \cdots = f_n = 0$. Before that we first give a couple of definitions.

2.3.1 Affine variety

Definition 2.21. For a set of polynomials $f_1, ..., f_n \in \mathcal{P}^s$ over field k, the set of all simultaneous solutions $a \in k^s$ of a system equation

$$f_1(a) = f_2(a) = \cdots = f_n(a) = 0,$$

is known as the *affine variety* defined by $f_1, ..., f_n$, and is denoted by $\mathbf{V}(f_1, ..., f_n)$ or $\mathbf{V}(I)$ where $I = \langle f_1, ..., f_n \rangle$. A subset $V \subset k^s$ is said to be an affine variety if $V = \mathbf{V}(f_1, ..., f_n)$ for some collection of polynomials $f_i \in \mathcal{P}^s$ [3].

Given an affine variety, one can define a set of all polynomial equations that vanish at points in affine variety.

Definition 2.22. Let $V \subset k^s$ be a variety. We denote by $\mathbf{I}(V)$ the set

$$\{f \in \mathcal{P}^s : \ f(a) = 0 \text{ for all } a \in V\}.$$

$\mathbf{I}(V)$ is an ideal as one can check for a variety V, if $f \in V$ then $pf = fp \in V$ for any $p \in \mathcal{P}^s$. Similarly for $f_1, f_2 \in \mathbf{I}(V)$, $f_1 + f_2 \in \mathbf{I}(V)$.

Note by definition of Affine variety, directly we have $V = \mathbf{V}(\mathbf{I}(V))$. One may ask is it still true that

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

As an ideal I determines $\mathbf{V}(I)$, the question is equivalent to whether $\mathbf{V}$ is a bijective map. The answer is no. Consider the example that

$$I = \langle x^2 \rangle \in \mathbb{R}.$$

Then $\mathbf{V}(I)$ is just $\{0\}$. However if we consider the polynomial x, it also vanishes at point 0 and then $x \in \mathbf{I}(\mathbf{V}(I))$. However the polynomial $x \notin I$ since any element in I is a multiple of x^2 or 0. One can see in such case $\mathbf{I}(\mathbf{V}(I))$ is strictly larger than I. Actually for any polynomial ideal $I \subset \mathbf{I}(\mathbf{V}(I))$ direct from definition. Then one may wonder what is $\mathbf{I}(\mathbf{V}(I))$ and that leads to a new definition.

Definition 2.23. Let $I \subset \mathcal{P}^s$ be an ideal. The *radical of* I is the set

$$\sqrt{I} = \{g \in \mathcal{P}^s : g^m \in I \text{ for some } m \geq 1\}.$$

An ideal is said to be a *radical ideal* if $\sqrt{I} = I$.

For instance

$$x + y \in \sqrt{\langle x^2 + 2xy, y^2 \rangle},$$

in $\mathbb{R}[x, y]$ since $(x + y)^2 = x^2 + 2xy + y^2 \in \langle x^2 + 2xy, y^2 \rangle$. One can see for any polynomial ideal I, it is always true that $I \subset \sqrt{I}$ since

$$I = \{g \in \mathcal{P}^s : g^m \in I, \quad m = 1\}.$$

Strong Nullstellensatz Theorem 2.24. If k is an *algebraically closed* field, and I is an ideal in $k[x_1, ..., x_s]$, then

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

With Strong Nullstellensatz Theorem, we are now able to solve a system of polynomial equations by Gröbner basis.

Definition 2.25. For $a = \begin{pmatrix} a_1 \\ : \\ a_n \end{pmatrix} \in \mathbb{R}^n$, a_i is the ith element of a.

Keep that in mind we then have the following proposition.

Proposition 2.26. For an ideal $I \subset k[x_1, ..., x_s]$ where k is an algebraically closed field, like $\mathbb{C}$, with finite $\mathbf{V}(I)$

$$\Pi_{a \in \mathbf{V}(I)}(x_s - a_s) \in \sqrt{I}.$$

Note $\Pi_{a \in \mathbf{V}(I)}(x_s - a_s)$ above defines a univariate polynomial with variable x_s. For instance, considering polynomial ideal $I = \langle xy + y, x + y^2 \rangle$ over field $\mathbb{C}$ and

$$\mathbf{V}(I) = \{(-1, 1), (0, 0), (1, 1)\}.$$

The proposition said that

$$(y + 1)y(y - 1) \in \sqrt{I},$$

which is equivalent to $((y+1)y(y-1))^m \in I$ for some $m \geq 1$. By definition**2.22** of $\mathbf{I}(V)$, $\mathbf{I}(\mathbf{V}(I))$ contains all polynomials in $\mathbb{C}[x, y]$ that vanish at $a, a \in \mathbf{V}(I)$. As $\Pi_{a \in \mathbf{V}(I)}(x_s - a_s)$ vanishes at any $a \in \mathbf{V}(I)$ directly by definition, then $\Pi_{a \in \mathbf{V}(I)}(x_s - a_s) \in \sqrt{I}$.

2.3.2 Variable elimination and extension

Definition 2.27. If I is an ideal in $k[x_1, ..., x_s]$ where k, then the lth *elimination ideal* is

$$I_l = I \ \cap \ k[x_{l+1}, ..., x_s].$$

One can see an elimination ideal I_l is just the polynomials in I that only contains variables $x_{l+1}, ..., x_s$. Or equivalently if $I = \langle f_1, ..., f_n \rangle \subset k[x_1, ..., x_s]$, the elements of I_l are the linear combinations of $f_1, ..., f_n$, with polynomial coefficients, that eliminates $x_1, ..., x_l$ from the equations $f_1 = \cdots = f_n = 0$.

Note one may solve a system of multi-variable polynomial equations by elimination, and the proposition **2.26** said that for a multi-variable polynomial ideal $I \subset k[x_1, ..., x_s]$ over algebraically closed field with nonempty finite solutions, I_{s-1} is always nonempty.

The Elimination Theorem 2.28. If G is a Gröbner basis for I with respect to the *lex* order $(x_1 > x_2 > \cdots > x_s)$(or any order where monomials involving at least one of $x_1, ..., x_l$ are greater than all monomials involving only the remaining variables), then

$$G_l = G \ \cap \ k[x_{l+1}, ..., x_s],$$

is a Gröbner basis of lth elimination ideal I_l.

The proof of the theorem is direct from the definition of Gröbner basis **2.16**. Let G be a Gröbner basis of ideal I and I_l an elimination ideal. For any polynomial $f \in I_l$, there is a $g \in G$ such that $LT(f)$ is divisible by $LT(g)$. As $f \in I_l$, then $LT(f) \in k[x_{l+1,...,x_s}]$ and hence $LT(g)$ is also an element of I_l. Then prove the theorem.

Now consider an ideal I satisfies the proposition **2.26**. As $\Pi_{a\in\mathbf{V}(I)}(x_s - a_s)$ is just a polynomial in $k[x_s]$,

$$\Pi_{a\in\mathbf{V}(I)}(x_s - a_s) \in \sqrt{I} \ \cap \ k[x_s].$$

Then for some $m \geq 1$

$$\Pi_{a\in\mathbf{V}(I)}(x_s - a_s)^m \in I_{s-1}.$$

By the elimination theorem, there is some $g \in G$ such that $LT(\Pi_{a\in\mathbf{V}(I)}(x_s - a_s)^m)$ is divisible by $LT(g)$, then $\mathbf{V}(I_{s-1}) \subset \{a_s \mid a \in \mathbf{V}(I)\}$. Furthermore as $\mathbf{I}(\mathbf{V}(I))$ is the set of all polynomials that vanish at any $a \in \mathbf{V}(I)$, then any $f \in I_{s-1}$ is a multiple of $\Pi_{a\in\mathbf{V}(I)}(x_s - a_s)$ otherwise it does not vanish at some $a_s, a \in \mathbf{V}(I)$. Then it shows $\{a_s \mid a \in \mathbf{V}(I)\} \subset \mathbf{V}(I_{s-1})$ which gives

$$\mathbf{V}(I_{s-1}) = \{P_s(a) \mid a \in \mathbf{V}(I)\}$$

Definition 2.29. A point $(a_{l+1}, ..., a_s) \in \mathbf{V}(I_l) \subset k^{s-l}$ is called a *partial solution.*

One can see partial solution is a projection of solution to subspace. Or equivalently, it is a solution of a system of polynomials after eliminating $x_1, \cdots, x_l$.

Every solution $(a_1, ..., a_n) \in \mathbf{V}(I) \subset k^s$ truncates to a partial solution as intuitively a projection from higher dimension to lower dimension often exists. However the converse is not always true and a partial solution may not extend to solutions.

Example 2.30. Consider the following ideal $\langle x^2 + y, y^2 - y\rangle$ over field $\mathbb{R}$. $y = 1$ and $y = 0$ are two partial solutions but $y = 1$ does not extend to solutions since $x^2 + 1 = 0$ has no solution over $\mathbb{R}$.

The Extension Theorem 2.31. If k is algebraically closed (e.g., $k = \mathbb{C}$), then a partial solution $a_{l+1}, ..., a_s$ in $\mathbf{V}(I_l)$ extends to $(a_l, ..., a_s)$ in $\mathbf{V}(I_{l-1})$ provided the leading coefficient polynomials of the elements of a lex Gröbner basis for I_{l-1} do not all vanish at $(a_{l+1}, ..., a_s)$.

Note in the example **2.30**, $y = 1$ is a solution of $y^2 - y = 0$ but does not extend to a full solution over $\mathbb{R}$. The Extension Theorem says that if we consider the field $\mathbb{C}$ then the partial solution extends to $(i, 1)$ in such case.

Now given a polynomial ideal $I = \langle f_1, ..., f_n \rangle \subset k[x_1, ..., x_s]$ and we want to find the affine variety $\mathbf{V}(I)$ which is also the set of solutions where $f_1 = \cdots = f_n = 0$. One can start with elimination and the Elimination theorem guarantees that for each elimination ideal I_l, one can always find $\mathbf{V}(I_l)$ by Gröbner basis G_l. Finally by the Extension Theorem, if k is algebraically closed, i.e. $\mathbb{C}$, one can then extend partial solutions in $\mathbf{V}(I_l)$ to solutions.

In the case the polynomial ideal $I \subset \mathbb{R}[x_1, ..., x_s]$ where $\mathbb{R}$ is not algebraically closed, we can still extend the field to $\mathbb{C}$ and then take intersection. For instance let $I_{\mathbb{R}} = \langle x^2 + y, y^2 - y \rangle \subset \mathbb{R}[x, y]$ and $I_{\mathbb{C}} = \langle x^2 + y, y^2 - y \rangle \subset \mathbb{C}[x, y]$. Then

$$\mathbf{V}(I_{\mathbb{C}}) = \{(i, 1), (0, 0)\},$$

and

$$\mathbf{V}(I_{\mathbb{R}}) = \mathbf{V}(I_{\mathbb{C}}) \ \cap \ \mathbb{R}^2 = \{(0, 0)\}.$$

2.4 Buchberger's algorithm

For any polynomial ideal $I \subset k[x_1, ..., x_s]$, the Gröbner basis always exist for I as shown before. The problem is then given a specified order, i.e. $x_1 >_{lex} \cdots >_{lex} x_s$, how to find a Gröbner basis of I? As one can eliminate variables of a polynomial system based on its Gröbner basis, this is equivalent to solve the polynomial system.

2.4.1 S-polynomial

Example 2.32. Consider the case $I = \langle x^2 + 2xy + 1, xy + y^2 \rangle$ with $x >_{lex} y$, then

$$\begin{cases} x^2 + 2xy + 1 \\ xy + y^2 \\ y^3 - y, \end{cases}$$

is a Gröbner basis of I. As Gröbner basis always exists then is that possible to find Gröbner basis for each I.

Definition 2.33. Let $f, g \in k[x_1, ..., x_s]$ be nonzero. Fix a polynomial order and let

$$\mathrm{LT}(f) = \alpha x^j \qquad \text{and} \qquad \mathrm{LT}(g) = \beta x^{\tilde{j}},$$

where $\alpha, \beta \in k$. Let x^m be the least common multiple of x^j and $x^{\tilde{j}}$. The $S - polynomial$ of f and g, denoted $S(f, g)$, is the polynomial [3]

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

Example 2.34. For instance $f = 2x^2y + x$ and $g = xy^2 - 2xy$ with order $x >_{lex} y$. Then

$$LT(f) = 2x^2y$$
$$LT(g) = xy^2,$$

and x^2y^2 is then the least common multiple of x^2y and xy^2. In such case

$$S(f,g) = \frac{y}{2}f - xg = 2x^2y + \frac{1}{2}xy.$$

For a polynomial ideal $I = \langle f_1, ..., f_n \rangle$, as any $S(f_a, f_b)$, $1 \leq a, b \leq n$, is a linear combination of f_a and f_b with polynomial coefficients, then $S(f_a, f_b) \in I$. However, $S(f_a, f_b)$ may not be reduced to 0 by $f_1, ..., f_n$. Continued with example **2.34**, $S(f,g) = 2x^2y + \frac{1}{2}xy$ is further reducible by f which gives

$$S(f,g) - f = \frac{1}{2}xy - x.$$

$S(f,g) - f \in I$ since both $S(f,g)$ and f are elements of I but xy is not divisible by either $LT(f)$ or $LT(g)$. As $\{f_1, ..., f_n\}$ is not a Gröbner basis of I, there is no guarantee that the remainder of $S(f,g)$ by $\{f_1, ..., f_n\}$ is zero.

Definition 2.35. The remainder of a polynomial f on division by a set of polynomials $G = g_1, \cdots, g_n$ will be denoted as $\overline{f}^G$ and is obtained by repeatedly dividing by $g \in G$ until the remainder is stationary.

For the case $S(f_i, f_j) \neq 0$, there may be more than one $g \in \{f_1, ..., f_n\}$ such that $LT(S(f_i, f_j))$ is divisible by $LT(g)$.

Proposition 2.36. For the ideal $I = \langle f_1, ..., f_n \rangle$, $S(f_i, f_j) \in I$ and $\overline{S(f_i, f_j)}^G \in I$.

This is a direct result as both $S(f_i, f_j)$ and $\overline{S(f_i, f_j)}^G$ are linear combinations of $f_1, ..., f_n$ with polynomial coefficients. The remainder of a S-polynomial on division by some polynomial set is not always unique.

Example 2.37. Let $I = \langle f_1, f_2, f_3 \rangle \in \mathcal{P}^3$ with $x >_{lex} y >_{lex} z$ where

$$f_1 = x^2y + y$$
$$f_2 = xy^2 + y^2z^2$$
$$f_3 = y^2z + z.$$

In such case

$$S(f_1, f_2) = xy^2z^2 - y^2.$$

Here $LT(S(f_1, f_2)) = xy^2z^2$ which is both divisible by $LT(f_2) = xy^2$ and $LT(f_3) = y^2z$. And divide by f_2 and f_3 give different remainder.

$$S(f_1, f_2) = z^2 f_2 - y^2z^4 - y^2$$
$$S(f_1, f_2) = xzf_3 - xz^2 - y^2.$$

As the remainder divided by f_2 is $-y^2z^4 - y^2$ and $-xz^2 - y^2$ by f_3, one can see even with a specified order the remainder is not always same.

2.4.2 Buchberger's criterion

Buchberger's criterion 2.38. A finite set $G = \{g_1, ..., g_t\} \subset I$ is a Gröbner basis of I if and only if $\overline{S(g_i, g_j)}^G = 0$ for all pairs $i \neq j$ [3].

One fact is that as long as a set $G = \{g_1, ..., g_t\}$ satisfies Buchberger's criterion, then no matter what order of g_i it takes to perform division algorithm, $\overline{S(f_i, f_j)}^G$ is always 0. Consider Example **2.37** above, $G = \{g_1, g_2, g_3, g_4, g_5\}$ is a Gröbner basis of $I = f_1, f_2, f_3$ where

$$g_1 = x^2y + y$$
$$g_2 = xy^2 - z^2$$
$$g_3 = xz + z^3$$
$$g_4 = y^2 - z^4$$
$$g_5 = z^5 + z.$$

One can check that

$$f_1 = g_1$$
$$f_2 = g_2 + z^2g_4 + zg_5$$
$$f_3 = zg_4 + g_5.$$

As f_1, f_2, f_3 are linear combinations of $g_1, ..., g_5$, $I \subset G$, we omit the proof that $G \subset I$ and one can check that by finding Gröbner basis of I online. Consider the $S(g_1, g_2)$, one have

$$S(g_1, g_2) = xz^2 + y^2.$$

Then applying division algorithm with $<_{lex}$, one have

$$S(g_1, g_2) = zg_3 + g_4.$$

$\overline{S(g_1, f_2)}^G = 0$ as it is a linear combinations of g_3, g_4 with polynomial coefficients. For the polynomial $f_2 \in I$, its remainder on division by G is always 0.

$LT(f_2) = xy^2$ is divisible by both $LT(g_2) = xy^2$ and $LT(g_4) = y^2$. One have

$$f_2 = g_2 + z^2 g_4 + z g_5$$
$$f_2 = x g_4 + z^3 g_3 + z^2 g_4.$$

The remainder is 0 either it is first divided by g_2 or g_4. Alternatively one can write $f_2 = (x + z^2)g_4 + z^3 g_3$.

2.4.3 Algorithm review

Buchberger's criterion provides a method to check if we have a Gröbner basis. Using Buchberger's algorithm, we also obtain a rudimentary procedure for producing a Gröbner basis of a given ideal.

Data: $\{f_1, f_2, ..., f_n\}$
Result: a Gröbner basis $G = \{g_1, ..., g_t\}$ for $I = \langle F \rangle$ and $F \subset G$
$G := F$;
$G' := \emptyset$;
while $G \neq G'$ **do**
  $G' = G$;
  forall $g_i, g_j \in G', i \neq j$ **do**
    $S := \overline{S(g_i, g_j)}^{G'}$;
    if $S \neq 0$ **then**
      $G := G \cup S$
    end
  end
end

Algorithm 1: Buchberger's algorithm

One can see the algorithm terminates when $G = G'$, or equivalently $\overline{S(g_i, g_j)}^{G'} = 0$ for any $g_i, g_j \in G', i \neq j$. Then Buchberger's algorithm generates a set of Gröbner basis immediately as it satisfies Buchberger's criterion when it terminates.

Termination of the algorithm is shown by considering the leading term $LT(g)$ of $g \in G$. A leading term $LT(f)$ of some new polynomial f cannot be a multiple of a leading term $LT(g)$ of a polynomial in $g \in G$, because the polynomials f are completely reduced with respect to G, *i.e.* the leading monomial of f is not divisible by any other polynomial's leading monomial in G. If on some input the algorithm did not terminate, this would give rise to an infinite sequence of leading term in which none is a multiple of any leading term appearing earlier in the sequence. Dickson's lemma [4] says that such a sequence cannot exist.

It is hard to determine the complexity of Buchberger's algorithm. Actually

different choices of $S-polynomial$ change the complexity.

The example next shows that how Buchberger's algorithm works.

Example 2.39. Let $G = \{f_1, f_2\}$ and $I = \langle G\rangle \subset \mathcal{P}^3$ be an ideal where

$$f_1 = xyz + xy$$
$$f_2 = x^2y^2z + 2yz.$$

Here as we only have two polynomials and there is only one choice at first

$$S(f_1, f_2) = x^2y^2 - 2yz.$$

Here $S(f_1, f_2)$ is neither reducible by f_1 nor f_2 and then set

$$f_3 = x^2y^2 - 2yz,$$

and $G_1 = G \cup \{f_3\}$. Here if we choose f_1 and f_3.

$$S(f_1, f_3) = x^2y^2 + 2yz^2,$$

Further reduced by f_3 gives

$$\overline{S(f_1, f_3)}^{G_1} = \overline{S(f_2, f_3)}^{G_1} = 2yz^2 + 2yz$$

Add $f_4 = 2yz^2 + 2yz$ and $G_2 = G_i \cup \{f_4\}$. After checking that $\overline{S(f_i, f_4)}^{G_2} = 0$ for $1 \leq i \leq 3$, G_2 now satisfies Buchberger's criterion and then the algorithm terminates.

2.5 Reduced Gröbner basis

Note for any polynomial ideal I, its Gröbner basis may not be unique. Consider the following ideal in $\mathcal{P}^2$

$$I = \langle x, y\rangle$$

Then $G = \{x, y\}$ is a Gröbner basis of I as it satisfies Buchberger's criterion. Actually the affine variety $\mathbf{V}(I)$ is the set $\{(0, 0)\}$. Similarly

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

is also a Gröbner basis of I. The definition does not specify that Gröbner basis is unique. In the example above, although $\{xy, x, y\}$ is still a Gröbner basis, however the element xy is not necessary. Actually if one have a look of Buchberger's algorithm, it only adds element the to set G', where G' is a Gröbner basis when terminated. Buchberger's algorithm generates a Gröbner basis but may contain some redundant elements. As in the example above one can exclude xy in the Gröbner basis and $\{x, y\}$ is still a Gröbner basis.

From algorithm's view, redundant elements could increase complexity. Considering Buchberger's criterion that a set of basis G is Gröbner basis when any two elements $p, q \in G$ satisfies

$$S := \overline{S(g_1, f_2)}^{G'}.$$

For the basis $G_1 = \{x, y\}$, the algorithm terminates by just checking

$$\overline{S(x, y)}^{G_1} = 0.$$

However when it comes to $G_2 = \{xy, x, y\}$, the algorithm needs to check both

$$\begin{aligned}\overline{S(x, y)}^{G_2} &= 0\\ \overline{S(xy, y)}^{G_2} &= 0\\ \overline{S(x, xy)}^{G_2} &= 0\end{aligned}$$

In the trivial case above, one single redundant element triple the times to generate S-polynomial. Then it is nature to ask is there a more efficient way to generate Gröbner basis.

Definition 2.40. A *reduced Gröbner basis* for an ideal $I \subset k[x_1, ..., x_n]$ is a Gröbner basis for I such that for all distinct $p, q \in G$, no monomial appearing in p is a multiple of $LT(q)$. A *monic Gröbner basis* is a reduced Gröbner basis in which the leading coefficient of every polynomial is 1, or $\emptyset$ if $I = \langle 0 \rangle$[3].

Again consider the case $I = \langle x, y \rangle$. $G_2 = \{xy, x, y\}$ is then not a Gröbner basis of I as $xy, x \in G$ and xy is a multiple of x.

Actually for a Gröbner basis G and two elements $p, q \in G$, if $LT(p)$ is a multiple of $LT(q)$, then one can always exclude p from G. Consider the following case.

Example 2.41. Let $I = \langle xyz + xy, x^2y^2z + 2yz \rangle$. Then by Buchberger's algorithm, the Gröbner basis is given by

$$\begin{aligned}g_1 &= x^2y^2z + 2yz\\ g_2 &= x^2y^2 - 2yz\\ g_3 &= xyz + xy\\ g_4 &= yz^2 + yz.\end{aligned}$$

One may notice that $LT(g_1) = x^2y^2z$ which is a multiple of $LT(g_2) = x^2y^2$. Then one can exclude g_1 and the reduced Gröbner basis is just

$$\{g_2, g_3, g_4\}.$$

Proposition 2.42. For a Gröbner basis G of a polynomial ideal I, if $p, q \in G$ and $LT(p)$ is a multiple of $LT(q)$, $G\backslash\{p\}$ is still a Gröbner basis of I.

The proposition is direct from definition of Gröbner basis. As for any $f \in I$, if $LT(f)$ is divisible by $LT(p)$ and then divisible $LT(q)$ and then one can exclude p from G. On the other case if $LT(f)$ is not divisible by $LT(q)$ and one can still exclude p from G.

Chapter 3

Algorithm and improvement

For a polynomial ideal I, one can generate a Gröbner basis using Buchberger's algorithm. The complexity of Buchberger's algorithm is somehow related to the number of variables and structure of polynomials. It is often hard to estimate the algorithm's complexity [6]. Consider Gaussian quadrature points problem with two points or equivalently finding Gröbner basis for $I = \langle f_1, f_2, f_3, f_4 \rangle$ with polynomials in w_0, x_0, w_1, x_1

$$\begin{aligned} f_1 &= w_0 + w_1 - 2 \\ f_2 &= w_0 x_0 + w_1 x_1 \\ f_3 &= w_0 x_0^2 + w_1 x_1^2 - 2/3 \\ f_4 &= w_0 x_0^3 + w_1 x_1^3. \end{aligned}$$

When running Buchberger's algorithm, it first set $G_0 := \{f_1, f_2, f_3, f_4\}$ and one can check that

$$\overline{S(f_i, f_j)}^{G_0} \neq 0 \qquad \forall\ 1 \leq i < j \leq 4.$$

The algorithm then adds all non-zero $\overline{S(f_i, f_j)}^{G_0}$ into the set G_0 which gives

$$G_1 = G_0 \cup \{\overline{S(f_i, f_j)}^{G_0} \mid 1 \leq i < j \leq 4\}.$$

Note in **Algorithm 1** it uses notation G' and keeps updating G'. However to distinguish G' in each iteration, we use G_i to represent G' in ith iteration. Then G_0 is the initialized state of G' and G_1 is the G' after first iteration. In the case above $|G_0| = 4$ as there are four polynomials in G_0. One can also check $|G_1| = 10$ as for any pair of $f_i, f_j \in G_0$, $\overline{S(f_i, f_j)}^{G_0}$ is non-zero.

Suppose after i iterations $|G_i| = t$, there are then t different elements in G_i and $O(t^2)$ different pairs of $S(p, q)$ with $p, q \in G'$. And for each $S(p, q) \neq 0$, the

complexity of calculating the remainder $\overline{S(p,q)}^{G'}$ is at least $O(t)$ as it needs to traverse each element in G_i.

As the complexity of Buchberger's algorithm increases fast with the size of set G, we would rather use a computer to run the algorithm than by hand. There are already some packages to calculate a Gröbner basis on both Maple and Python. However most of them are symbolic. In this chapter we focus on some improvement of Buchberger's algorithm to reduce the complexity and give another two algorithms based on Buchberger's algorithm in section 2 and 3 separately.

3.1 Polynomials and sparse matrix

We can arrange the monomials in Lexicographic order where

$$x_1 >_{lex} x_2 >_{lex} \cdots >_{lex} x_s.$$

Since a polynomial f_i is sum of monomials with coefficients, an intuitive way to represent the polynomial is a product of two vectors.

Definition 3.1 Given a set of monomials $x = \{x^{j_1}, x^{j_2}, \cdots, x^{j_n}\}$, a monomial vector $v(x)$ is then

$$\begin{pmatrix} x^{j_1} \\ x^{j_2} \\ \vdots \\ x^{j_n} \end{pmatrix}.$$

Considering a trivial case

$$f = x_1x_2 + 2x_2 + 1.$$

Let $\alpha = (1, 0, 2, 1)$ and $v(x) = (x_1x_2, x_1, x_2, 1)^T$, we have

$$f = \begin{pmatrix} 1 & 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1x_2 \\ x_1 \\ x_2 \\ 1 \end{pmatrix} = \langle c, v \rangle$$

Here we have coefficients in vector c and monomials in vector v. Specifically, we arrange v in Lexicographic oder where $x_1 >_{lex} x_2$.

3.1.1 Monomial vector b

Now consider a set of polynomials $\{f_1, \cdots, f_s\}$, for each polynomial f_i, one can write it in the way below

$$f_i = \sum_{j \in J} \alpha_{i,j} x^j,$$

Note if one can find a large enough monomial vector v that for any polynomial f_i, f_i can be represented as a inner product of a coefficient vector $c^{(i)}$ and v, then the set of polynomials is then

$$\begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_s \end{pmatrix} = \begin{bmatrix} c^(1) \\ c^{(2)} \\ \vdots \\ c^{(s)} \end{bmatrix} v.$$

Intuitively this v exists as long as it contains all monomials x^j where $\alpha_j x^j$ is a term of some polynomial f_i. So we generate such v by let it contain all such x^j. Considering monomial x^j with $j = (j_1, ..., j_n)$. Let $m_j = \max\{j_1, ..., j_n\}$. Then m_j is the largest degree of variables in monomial x^j and then

$$x^j \in \{x^k \mid k = (k_1, ..., k_n) \text{ and } k_i \leq m_j\}.$$

With that definition we can now define m to be the maximum degree of all variables in all polynomials $f_1, ..., f_s$.

$$m := \max\{\ m_j \mid \alpha_j x^j \text{ is a term of } f_i \text{ for some } i\}.$$

Since m is the largest degree of variables appears in all terms of the set of polynomials $\{f_1, ..., f_s\}$, then for any term $c_j x^j$ of some f_i with $j = (j_1, ..., j_n)$, $j_a \leq m$ for $1 \leq a \leq s$. Then we define set M

$$M := \{\ x^j \mid j = (j_1, ..., j_n),\ j_i \leq m \text{ for } 1 \leq i \leq n\}.$$

Here M is the set of monomials that has maximum degree less than or equal to m, or equivalently, it contains all elements we need in monomial vector v. Now for any term $\alpha_j x^j$ of f_i for some i, let $j = (j_1, ..., j_n)$. Since $j_i \leq m$ for $1 \leq i \leq s$ by how we define m, $x^j \in M$. Then for any $1 \leq i \leq s$

$$f_i = \sum_{j \in J} \alpha_{i,j} x^j, \quad x^j \in M.$$

where $\alpha_{i,j}$ is the coefficient of the term x^j in f_i. The equations above shows that all terms of f_i is a production of an monomial in M and some coefficient. Then let monomial vector v contain all elements in M. Then for any $f_i \in \{f_1, ..., f_s\}$, there is a coefficient vector c_i that

$$f_i = \langle c_i, v \rangle.$$

Specifically we can arrange elements in monomial vector v with Lexicographic order. For instance, consider a set of polynomials

$$\begin{cases} f_1 = x^2yz + yz \\ f_2 = xz + y^2 \\ f_3 = x + yz + z. \end{cases}$$

The largest degree of variable $m = 2$. Then the set M is now

$$M = \{\ x^{j_1}y^{j_2}z^{j_3} \mid j_i \leq 2 \text{ for } 1 \leq i \leq 3\}.$$

M is a set of size 3^3 since each j_i can take 3 different values, $0, 1, 2$. Then the number of elements in M is $3^3 = 27$. With Lexicographic order $x >_{lex} y >_{lex} z$, v is a vector of size 27 that

$$v = \begin{pmatrix} x^2y^2z^2 \\ x^2y^2z \\ \vdots \\ x^2 \\ xy^2z^2 \\ \vdots \\ z \\ 1 \end{pmatrix}.$$

3.1.2 Coefficient matrix C

With polynomial set $\{f_1, ..., f_s\}$ and corresponding monomial matrix v, we can easily define the Coefficient matrix C that

$$C := \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_s \end{pmatrix}.$$

C is a $n \times |M|$ matrix where s is the number of polynomials and

$$|M| = (m+1)^n.$$

Here n is the number of variables $\{x_1, ..., x_n\}$ and m is the largest degree of all variables. If we go back to check how we set M, one can see that for each variable $x_i^{j_i}$ where $1 \leq i \leq n$, there are $(m+1)$ different values for j_i which are $0, 1, ..., m$. Since there are n such variables, $|M| = (m+1)^n$.

C could be a huge matrix as it increases exponentially as n increases. And it seems that

$$\begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_s \end{pmatrix} = Cv.$$

could be an expensive way to store the polynomial set. However here C is a sparse matrix and only the non-zero entries need to be stored. In such way the size of C is linear to the total number of terms in $\{f_1, ..., f_s\}$. Now consider the way we generate monomial vector v. Once m in (3) and the number of variables $x_1, ..., x_n$ fixed, then set of monomials M is then fixed. And by Lexicographic order there is only one possible arrangement for elements in M. So instead of the monomial vector v, a pair of integers (m, n) is enough to restore v.

3.1.3 Generating coefficient matrix C

As in section 2.3, C is a sparse matrix and v is stored as a pair of integers (m, n). One question remains unanswered. Which entry of C should be set given a new term of some polynomial. Consider the example at the beginning of section 2.

$$f = \begin{pmatrix} 1 & 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1x_2 \\ x_1 \\ x_2 \\ 1 \end{pmatrix}.$$

Assume now we are going to add a new polynomial g

$$g = 4x_1 + x_2.$$

Here g has two terms $4x_1$ and x_2. Without two much work we find

$$\begin{pmatrix} f \\ g \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 & 1 \\ 0 & 4 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1x_2 \\ x_1 \\ x_2 \\ 1 \end{pmatrix}.$$

Since x_1 is the second element of v then the coefficient of x_1, 4, is at the second column of C. Similarly, the coefficient of x_2, 1, is at the third column of C. We find that the column of a coefficient in C is determined by the monomial's position in v. Then given a monomial vector v we define the position of a monomial x^jin v by

$$p(j) = i \text{ where } x^j \text{ is the } i\text{th element of } v.$$

For a monomial vector generated in **3.1.1** with pairs of integers (m, n), a monomial x^j where $j = (j_1, \cdots, j_n)$

$$p(j) = \sum_{i=0}^{n-1} (m - j_{n-i})(m + 1)^i + 1.$$

One can prove as following. For some monomial vector v with corresponding pairs of integers (m, n). Considering two monomials

$$\begin{cases} x^j = x_1^{j_1} \cdots x_k^{j_k} \cdots x_n^{j_n} \\ x^{\tilde{j}} = x_1^{\tilde{j}_1} \cdots x_k^{\tilde{j}_k} \cdots x_n^{\tilde{j}_n}, \end{cases}$$

where $j_i, \tilde{j}_i \leq m$ and $j_i = \tilde{j}_i$ if $i \neq k$. Here $j, \tilde{j}$ are sequences with only one element different which are j_k and $\tilde{j}_k$. It is easy to check

$$p(j) - p(\tilde{j}) = -(j_k - \tilde{j}_k)(m+1)^{n-k}.$$

First consider the case $j_k - \tilde{j}_k = 1$. Define the set

$$S = \{x^\gamma \mid x^{\tilde{j}} <_{lex} x^\gamma \leq_{lex} x^j\}.$$

Here S is the set of all elements in v that less than or equal to x^j and greater than $x^{\tilde{j}}$ by Lexicographic order. Then for $x^\gamma = x_1^{\gamma_1} \cdots x_n^{\gamma_n} \in S$, $\gamma_i = j_i = \tilde{j}_i$ for $i < k$ since otherwise $x^\gamma > x^j > x^{\tilde{j}}$ or $x^\gamma < x^{\tilde{j}} < x^j$ which is a contradiction of how we set S. The statement above says that for any element x^γ in S, the first $(k-1)$ elements of γ must be same with the first $(k-1)$ elements of j and $\tilde{j}$. For similar reason either $\gamma_k = j_k$ or $\gamma_k = \tilde{j}_k$.

Now consider $(\gamma_{k+1}, ..., \gamma_n)$. There are $(m+1)^{n-k}$ different combinations of the sequence since for each γ_i with $k < i \leq n$, it can take $(m+1)$ different values. And for each combination of $(\gamma_{k+1}, ..., \gamma_n)$, either $x_1^{j_1} \cdots x_k^{j_k} x_{k+1}^{\gamma_{k+1}} \cdots x_n^{\gamma_n} \in S$ or $x_1^{\tilde{j}_1} \cdots x_k^{\tilde{j}_k} x_{k+1}^{\gamma_{k+1}} \cdots x_n^{\gamma_n} \in S$. Then

$$|S| = (m+1)^{n-k}.$$

Since the size of S is $(m+1)^{n-k}$, $|p(j) - p(\tilde{j})| = (m+1)^{n-k}$. As $j_k > \tilde{j}_k$, by Lexicographic order $p(j) < p(\tilde{j})$.

$$p(j) - p(\tilde{j}) = -(j_k - \tilde{j}_k)(m+1)^{n-k}.$$

That leads to the general case that

$$p(j) = \sum_{i=0}^{n-1} (m - j_{n-i})(m+1)^i + 1.$$

Or equivalently for some term $\alpha_j x^j$ of the ith polynomial f_i, α_j is the $(i, p(j))$th entry of coefficient matrix C. Note here we can create a coefficient matrix C even without the monomial vector v but just a pair of integers (m, n).

3.1.4 The set $\mathcal{T}_d^s$ of monomials

Note we define $\mathcal{T}_d^s$ to be the set of all monomials that has s variables and total degree less than d and $\mathcal{T}_d^s$ contains exactly $\binom{d+s}{d}$ elements. One can also generate a monomial vector v contains all elements of $\mathcal{T}_d^s$.

Here we have another way to generate monomial vector. Consider the set $\mathcal{T}_3^2$ and let x, y be two variables. Then one have

$$\mathcal{T}_3^2 = \{1, x, xy, xy^2, x^2, x^2y, x^3, y, y^2, y^3\}.$$

One can check it contains exactly 10 elements. Given a order that $x >_{lex} y$, one have

$$x^3 > x^2y > \cdots > y > 1.$$

And one can now generate a monomial vector with all elements in $\mathcal{T}_3^2$ which is now $v = (x^3, x^2y, \cdots, y, 1)^T$. Again a polynomial can be written as a product of coefficient vector c and monomial vector v as we did before. consider a polynomial

$$f = 2x^3 + xy.$$

There is not doubt that x^3 is the first element of $(x^3, x^2y, \cdots, y, 1)^T$. However it is not that obvious when considering xy. This is actually a critical step when we store a polynomial in sparse matrix as one only need to store a matrix with its nonzero elements and their location. And $2x^3$ is stored as $(2, 1)$ where the first element 2 is the coefficient of $2x^3$ and the second element 1 indicates it multiplies the first element in $v = (x^3, x^2y, \cdots, y, 1)^T$.

Now consider the index of xy in v. One can use a star and bar model to represent all elements in $\mathcal{T}_d^s$. There is a bijection between elements in $\mathcal{T}_d^s$ and arrangement of s stars and d bars. Consider the case $\mathcal{T}_3^2$, then the following arrangement

$$* \ * \ | \ * \ |$$

is mapped to x^2y. The bijection works like following. For an element $x^ay^b \in \mathcal{T}_3^2$, as $a + b \leq 3$, one can rewrite x^ay^b as $x^ay^b1^c$ where $a, b, c \geq 0$ and $a + b + c = 3$. Then it is then mapped to

$$\underbrace{* \ \cdots \ *}_{\text{a}} \ | \ \underbrace{* \ \cdots \ *}_{\text{b}} \ | \ \underbrace{* \ \cdots \ *}_{\text{c}}.$$

The map is bijective as each monomial has unique star/bar arrangement and a star/bar arrangement corresponds to one monomial. One can check there are exactly $\binom{5}{2}$ different arrangement which is same as Proposition **1.2**. To determine the index of xy with $x >_{lex} y$, one can compute the number of all elements $x^ay^b \in \mathcal{T}_3^2$ that $x^ay^b >_{lex} xy$, or the size of $\{x^ay^b >_{lex} xy \mid x^ay^b \in \mathcal{T}_3^2\}$. By definition of $>_{lex}$, $x^ay^b >_{lex} xy$ iff $a > 1$ or $a = 1$ and $b > 1$. Considering the case that $a > 1$, which means there are at least two stars in star/bar arrangement

$$* \ * \ \cdots$$

As there are only one star and two bars left, there are $\binom{3}{1}$ elements that satisfies $a > 1$, which are x^3, x^2y, x^2. Then considering the case that $a = 1$ and $b > 1$. Similarly it must starts with

$$* \ | \ * \ * \ \cdots$$

There is no star left and only one bar remained, and hence $\binom{1}{1}$ element which is xy^2. Then there are 3 elements in the case $a > 1$ and 1 element in the case

$a = 1$ and $b > 1$. Then xy is the fifth element in $\mathcal{T}_3^2$ with $x >_{lex} y$.

Consider the general case $\mathcal{T}_d^s$. The total degree is bounded by d and hence there are d stars and s bars. Let $x_1, ..., x_s$ be s variables of monomials in $\mathcal{T}_d^s$ and v be a monomial vector that contains all elements of $\mathcal{T}_d^s$ with order that $x_1 >_{lex} x_2 >_{lex} \cdots >_{lex} x_s$. Then for a monomial $x^j = x_1^{j_1} x_2^{j_2} \cdots x_s^{j_s}$, we want to find the index of x^j in monomial vector v.

One can find number of elements $x^{\tilde{j}} = x_1^{\tilde{j}_1} x_2^{\tilde{j}_2} \cdots x_s^{\tilde{j}_s} \in \mathcal{T}_d^s$ such that $x^{\tilde{j}} >_{lex} x^j$, or equivalently the size of the set

$$E := \{x^{\tilde{j}} >_{lex} x^j \mid x^{\tilde{j}} \in \mathcal{T}_3^2\}.$$

Again by definition of $>_{lex}$, $x^{\tilde{j}} >_{lex} x^j$ iff for some $1 \le k \le s$,

$$\begin{cases} \tilde{j}_k > j_k \\ \tilde{j}_i = j_i \quad i < k \end{cases}.$$

Set

$$E_k := \{x^{\tilde{j}} = x_1^{\tilde{j}_1} x_2^{\tilde{j}_2} \cdots x_s^{\tilde{j}_s} \in \mathcal{T}_d^s \mid \tilde{j}_k > j_k \text{ and } \tilde{j}_i = j_i \text{ for } i < k\}.$$

E_k is the set of all elements that has exactly the kth variable's degree larger than x^j and for all $i < k$. Explicitly, let $x^j = x_1^{j_1} x_2^{j_2} \cdots x_s^{j_s}$, then for any $x^{\tilde{j}} \in E_k$,

$$x^{\tilde{j}} = x_1^{j_1} x_2^{j_2} \cdots x_{k-1}^{j} x_k^{\tilde{j}_j} \cdots x_s^{\tilde{j}_s}.$$

The first $k-1$ variables' degree of x^j and $x^{\tilde{j}}$ are the same for fixed k then for $E_k, 1 \le k \le s$ is a partition of all elements $x^{\tilde{j}} \in \mathcal{T}_d^s$ such that $x^{\tilde{j}} > x^j$. Then

$$|E| = \sum_{i=1}^{s} |E_k|.$$

For each E_k, one can compute its size by star/bar model again. Let $x^j = x_1^{j_1} x_2^{j_2} \cdots x_s^{j_s}$ and consider the case $k = 1$ and $|E_1|$. Then for any $x^{\tilde{j}} > x^j$, its corresponding star/bar arrangement must start with $j + 1$ stars

$$\underbrace{* \cdots *}_{j_1 + 1} \cdots$$

Then as there are s stars and d bars in total, $s - j_1 - 1$ stars and d bars are left and hence

$$|E_1| = \binom{s + d - j_1 - 1}{d}.$$

For any $1 \le k \le s$, all elements in E_k must start with

$$\underbrace{* \cdots *}_{j_1} \mid \underbrace{* \cdots *}_{j_2} \mid \cdots \mid \underbrace{* \cdots *}_{j_{j-1}} \mid \underbrace{* \cdots *}_{j_k + 1} \cdots$$

Here there are $(\sum_{i=1}^{j} j_j) + 1$ stars and $k-1$ bars used, then one have

$$|E_k| = \binom{s+d-(\sum_{i=1}^{k} j_i)-k}{d-k+1}.$$

Then one have

$$\begin{aligned} |E| &= \sum_{k=1}^{s} |E_j| \\ &= \sum_{k=1}^{s} \binom{s+d-(\sum_{i=1}^{k} j_i)-k}{d-k+1}. \end{aligned}$$

3.2 Algorithm implementation

Buchberger's algorithm terminates when the set G satisfies Buchberger's criterion **2.38**. In algorithm **1** it checks if $\overline{S(g_i, g_j)}^{G'} = 0$ for all $g_i, g_j \in G'$. Consider example **2.39** again, where $G = \{f_1, f_2\}$, $I = \langle G \rangle \subset \mathcal{P}^3$ and

$$\begin{aligned} f_1 &= xyz + xy \\ f_2 &= x^2y^2z + 2yz. \end{aligned}$$

Buchberger's algorithm sets $G' := G$ and then checks if G satisfies Buchberger's criterion. Here as G contains only two polynomials f_1, f_2, it only remains to check if $\overline{S(f_1, f_2)}^{G} = 0$. As in our case, we already see that

$$\overline{S(f_1, f_2)}^{G} = x^2y^2 - 2yz \neq 0.$$

Then we introduce $G_1 = G \cup \{f_3\}$ where

$$f_3 = x^2y^2 - 2yz.$$

Here the first loop ends and $G_1 \neq G'$ and the algorithm continues. In the second iteration, there are three polynomials in G_1 which are f_1, f_2, f_3. It seems we still need to check that if

$$\overline{S(f_1, f_2)}^{G_1} = 0$$

One may notice that this is actually a redundant step. As in first iteration

$$S(f_1, f_2) = xyf_1 - f_2 - f_3,$$

and f_3 is immediately added in G_1, hence $\overline{S(f_1, f_2)}^{G_1} = 0$. After checking $\overline{S(f_1, f_2)}^{G_1}$ we then compute $S(f_1, f_3)$ which gives

$$S(f_1, f_3) = f_3 + 2yz^2 + 2yz.$$

Let

$$f_4 = 2yz^2 + yz,$$

and f_4 is not divisible by either f_1, f_2 and f_3, then set $G_2 = G_1 \cup \{f_4\}$. One may note $\overline{S(f_2, f_3)}^{G_1} = f_4$ which ends the second iteration. In the third iteration, we again have the knowledge that

$$\begin{aligned} S(f_1, f_2) &= xyf_1 - f_2 - f_3 \\ S(f_1, f_3) &= f_3 + f_4 \\ S(f_2, f_3) &= f_4. \end{aligned}$$

Then we only need to check remainders of $\{S(f_1, f_4), S(f_2, f_4), S(f_3, f_4)\}$.

For a set $G = \{f_1, \cdots, f_n\}$ and two polynomials $f_i, f_j \in G$, if

$$\overline{S(f_i, f_j)}^{G} = 0,$$

then one can express $S(f_i, f_j)$ as

$$S(f_i, f_j) = \sum \gamma_k f_k, \qquad f_k \in G.$$

Here γ_k are polynomial coefficients. As a consequence, for any set G' that $G \subset G'$, as $f_1, \cdots, f_n \in G'$, it is then still possible to write

$$S(f_i, f_j) = \sum \gamma_k f_k, \qquad f_k \in G'.$$

Note here we are not saying that $\overline{S(f_i, f_j)}^{G'} = 0$. The reason is that $\overline{S(f_i, f_j)}^{G'}$ may not be unique. Consider the case below where

$$\begin{aligned} f_1 &= x^2y + xy \\ f_2 &= xy \\ f_3 &= x + 1. \end{aligned}$$

Here one have

$$S(f_1, f_2) = f_1 - xf_2 = xy.$$

xy is divisible by both $LT(f_2)$ and $LT(f_3)$. However the remainder divided by f_2 and f_3 are different as one can see

$$\begin{aligned} S(f_1, f_2) &= f_1 - xf_2 + f_2 \\ S(f_1, f_2) &= f_1 - xf_2 + yf_3 - y. \end{aligned}$$

The remainder divided by f_2 is 0 while is $-y$ when divided by f_3. It then leads to the fact that when we say $\overline{S(f_i, f_j)}^{G} = 0$, it implicitly specify a sequence of polynomials in division algorithm. In the case above, $\overline{S(f_1, f_2)}^{G} = 0$ indicates we divide $S(f_1, f_2)$ by f_2 and $\overline{S(f_1, f_2)}^{G} = -y$ indicates the divisor is f_3.

Then for a S-polynomial $S(f_i, f_j)$, we say its remainder $\overline{S(f_1, f_2)}^G = 0$ if there exists a sequence of polynomials in G such that the remainder of $S(f_i, f_j)$ divided by such sequence is 0. With that modification, we then have the following proposition.

Proposition 3.1. For a set $G = \{f_1, \cdots, f_n\}$ and two polynomials $f_i, f_j \in G$, if

$$\overline{S(f_i, f_j)}^G = 0,$$

then for any set G' that $G \subset G'$,

$$\overline{S(f_i, f_j)}^{G'} = 0,$$

Now have a review on Buchberger's algorithm. For some iteration where $G = \{g_1, \cdots, g_n\}$ and two polynomials $g_i, g_j \in G$, if $\overline{S(g_i, g_j)}^G$ is non-zero then add it to G and let it be $\bar{G} = G \cup \{\overline{S(g_i, g_j)}^G\}$, or otherwise do nothing. By the proposition above, for any set G' that $\bar{G} \subset G'$, $\overline{S(g_i, g_j)}^{\bar{G}} = 0$, which indicates that one do not need to check $\overline{S(g_i, g_j)}^{\bar{G}'}$ in the following iterations.

Then one can modify **Algorithm 1** as below

Data: $\{f_1, f_2, ..., f_n\}$
Result: a Gröbner basis $G = \{g_1, ..., g_t\}$ for $I = \langle F \rangle$ and $F \subset G$
$G := F$;
$i := 2$;
$index := n + 1$;
while $i \neq index$ **do**
  forall $j < i$ **do**
    $S := \overline{S(f_i, f_j)}^G$;
    if $S \neq 0$ **then**
      $f_{index} = S$;
      $G := G \cup \{f_{index}\}$;
      $index = index + 1$
    end
  end
  $i = i + 1$
end

Algorithm 2: Modified Buchberger's algorithm

Considering the Example **2.39** again, the order of S-polynomials that the algorithm checks is as the following

$$S(f_1, f_2)$$
$$S(f_1, f_3), S(f_2, f_3)$$
$$S(f_1, f_4), S(f_2, f_4), S(f_3, f_4).$$

One can see each S-polynomial is only checked once. By proposition **3.1**, the set G also satisfies Buchberger's criterion when terminated and hence generate a Gröbner basis.

3.3 Reduced basis

Buchberger's algorithm generates a Gröbner basis. One may notice if G is a Gröbner basis of polynomial ideal I, any super set $G' \subset I$ is again a Gröbner basis of I. And the size of Gröbner basis generated by Buchberger's algorithm could be large.

Example 3.2. Consider the ideal $I = \langle f_1, f_2, f_3, f_4 \rangle$ again where

$$\begin{aligned} f_1 &= w_0 + w_1 - 2 \\ f_2 &= w_0 x_0 + w_1 x_1 \\ f_3 &= w_0 x_0^2 + w_1 x_1^2 - 2/3 \\ f_4 &= w_0 x_0^3 + w_1 x_1^3 \end{aligned}$$

One may note finding affine variety of I is equivalent to find Gaussian Quadrature points and weights for the case of two points [7]. If we set $w_0 >_{lex} x_0 >_{lex} w_1 >_{lex} x_1$, then Gröbner basis of I generated by Buchberger's algorithm is then the set $G = \{g_1, g_2, \cdots, g_{11}\}$ where

$$\begin{aligned} g_1 &= w_0 + w_1 - 2 \\ g_2 &= w_0 x_0 + w_1 x_1 \\ g_3 &= w_0 x_0^2 + w_1 x_1^2 - 2/3 \\ g_4 &= w_0 x_0^3 + w_1 x_1^3 \\ g_5 &= x_0 w_1 - 2x_0 - w_1 x_1 \\ g_6 &= x_0 x_1 + \frac{1}{3} \\ g_7 &= x_0 + x_1 \\ g_8 &= w_1 x_1^2 + \frac{1}{3} w_1 - \frac{2}{3} \\ g_9 &= w_1 x_1 - x_1 \\ g_{10} &= x_1^2 - \frac{1}{3} \\ g_{11} &= w_1 - 1 \end{aligned}$$

Here f_1, f_2, f_3, f_4 are relatively simple polynomials where total degree of all monomials in f_i are no greater than 4 and each f_i has at most 3 terms. Contrary to the simple form of polynomial ideal, there are 11 polynomials in the Gröbner basis generated by Algorithm **2**. Although different choices of $S(f_i, f_j)$ in Buchberger's algorithm may change the Gröbner basis generated, *i.e.* we start

with $S(f_1, f_4)$ instead of $S(f_1, f_2)$, the size of Gröbner basis is not changed too much. And the size of Gröbner basis all increase fast.

Here G is not a reduced Gröbner basis. And by proposition **2.42**, one may exclude g_j from G if $LM(g_j)$ is a multiple of $LM(g_i)$ for some $g_i \in G$ and $G\backslash\{g_j\}$ is still a Gröbner basis of I. If we delete all $g \in G$ if its leading monomial is a multiple of some polynomial's leading monomial in G, it gives $\tilde{G}$ and

$$\tilde{G} = \{g_1, g_7, g_{10}, g_{11}\}.$$

$\tilde{G}$ is again a Gröbner basis of I. One can check that as $LM(g_2), LM(g_3), LM(g_4)$ are multiples of $LM(g_1)$; $LM(g_5), LT(g_6)$ are multiples of $LM(g_7)$, and $LM(g_8), LM(g_9)$ are multiples of $LM(g_{11})$. $\tilde{G}$ is a Gröbner basis directly from proposition **2.42**. One can even generate a reduced Gröbner basis G_R of I where

$$G_R = \{lw_0 - 1, x_0 + x_1, x_1^2 - \frac{1}{3}, w_1 - 1\}$$

One may notice that the size of G_R is only 4.

3.3.1 Reduced basis in algorithm

The size of G has critical impact on the efficiency of algorithm. Consider example **3.2**. Buchberger's algorithm terminates when the set G satisfies Buchberger's criterion. Assume we need to check

$$\overline{S(g_4, g_8)}^G = 0.$$

Then in the first step one needs to compute

$$S(g_4, g_8) = \frac{1}{3}w_0x_0^3w_1 - \frac{2}{3}w_0x_0^3 - w_1^2x_1^5.$$

Then $LM(S(g_4, g_8))$ is divisible by $LM(g_1) = w_0$ and hence

$$\begin{aligned}
S(g_4, g_8) &= \frac{1}{3}x_0^3w_1g_1 + \frac{2}{3}w_0x_0^3 + \frac{1}{3}x_0^3 * w_1^2 - \frac{2}{3}x_0^3w_1 + w_1^2x_1^5 \\
&= \frac{1}{3}x_0^3w_1g_1 - \frac{2}{3}x_0^3g_1 + \frac{1}{3}x_0^3w_1^2 - \frac{4}{3}x_0^3w_1 + \frac{4}{3}x_0^3 + w_1^2x_1^5 \\
&= \frac{1}{3}x_0^3w_1g_1 - \frac{2}{3}x_0^3g_1 - \frac{1}{3}x_0^2w_1g_5 - \frac{2}{3}x_0^3w_1 + \frac{4}{3}x_0^3 + \frac{1}{3}x_0^2w_1^2x_1 + w_1^2x_1^5 \\
&\quad\vdots \\
&= \frac{1}{3}x_0^3w_1g_1 - \frac{2}{3}x_0^3g_1 - \frac{1}{3}x_0^2w_1g_5 + \frac{2}{3}x_0^2g_5 - \frac{1}{3}x_0w_1x_1g_5 - \frac{1}{3}w_1x_1^2g_5 \\
&\quad - \frac{2}{3}x_1^2g_5 - \frac{4}{3}x_1g_6 - w_1x_1^3g_8 - \frac{4}{3}x_1g_8 + \frac{4}{9}g_9.
\end{aligned}$$

The equation above shows that $S(g_4, g_8)$ is a linear combination of $g_i \in G$ and hence $\overline{S(g_4, g_8)}^G = 0$. Furthermore, one can find it takes 11 steps to find the

remainder of $S(g_4, g_8)$ and the sequence of polynomial divisors are

$$g_1, g_1, g_5, g_5, g_5, g_5, g_5, g_6, g_8, g_8, g_9.$$

To find the remainder of $S(g_4, g_8)$, or $\overline{S(g_4, g_8)}^G$, the algorithm traverses G to find a polynomial $g_i \in G$ such that $LM(S(g_4, g_8)) = w_0 x_0^3 w_1$ is divisible by $LM(g_i)$ and in such case $LM(g_1) = w_0$ divides $LT(S(g_4, g_8))$. Subtract $S(g_4, g_8)$ by $\frac{1}{3}x_0^3 w_1 g_1$ gives a new polynomial

$$S(g_4, g_8) - \frac{1}{3}x_0^3 w_1 g_1 = \frac{2}{3}w_0 x_0^3 + \frac{1}{3}x_0^3 * w_1^2 - \frac{2}{3}x_0^3 w_1 + w_1^2 x_1^5.$$

We repeat the step that find $g_i \in G$ such that $LM(g_i)$ divides

$$LM(S(g_4, g_8) - \frac{1}{3}x_0^3 w_1 g_1) = w_0 x_0^3.$$

Again to find such g_i we need to traverse the set G and hence the complexity is $O(|G|)$. Moreover for a set G, there are $O(|G|^2)$ different pairs $S(g_i, g_j)$ where $g_i, g_j \in G$. then the complexity to check G satisfies Buchberger's criterion is at least $O(|G|^3)$.

The complexity is not only affected by the size of Gröbner basis $|G|$ and is often hard to give an accurate bound of complexity. In the example above, it takes 11 iterations to check $\overline{S(g_4, g_8)}^G = 0$. So far there is no hint of how many iterations we need to find a S-polynomial' remainder. Considering the first iteration that $LM(S(g_4, g_8)) = w_0 x_0^3 w_1$, the polynomial $g_i \in G$ that $LM(g_i)$ divides $LT(S(g_4, g_8))$ is not unique. The leading term of both g_1, g_2, g_3, g_4 both divide $LT(S(g_4, g_8))$. And different choice of g_i may change the number of iterations to find the remainder.

Some S-polynomials need much less effort to find their remainders. Consider the case

$$S(g_7, g_{11}) = x_0 + w_1 x_1.$$

It only takes two iterations to find $\overline{S(g_7, g_{11})}^G$ where

$$S(g_7, g_{11}) = g_7 + g_9.$$

One may expect that the S-polynomial of a pair of polynomials with less terms and lower degree would be easier to reduce. Many aspects affect the number of iterations to find $\overline{S(g_i, g_j)}^G$ for $g_i, g_j \in G$. The polynomial ideal $I = \langle x^{2000} - 1, x^{1000} - 1\rangle$ takes only one step to reduce as $x^{2000} - 1 = (x^{1000} + 1)(x^{1000} - 1)$, although the total degree of x^{2000} is 2000.

The problem is that can we reduce the size of G as we know some elements in G are not necessary. That immediately remind us of reduced Gröbner basis. As we mentioned before,

$$G_R = \langle w_0 - 1, x_0 + x_1, x_1^2 - \frac{1}{3}, w_1 - 1\rangle,$$

is a reduced Gröbner basis of I and only has size 4. Not only the size of G_R is small, elements in G_R have lower degree. The property is a consequence of the definition of reduced Gröbner basis where all polynomials $g_i \in G$ that $LM(g_i)$ is divisible by some $LM(g_j), g_j \in G$ can not appear in reduced Gröbner basis. If we let

$$\begin{aligned}\bar{g}_1 &= w_0 - 1\\ \bar{g}_2 &= x_0 + x_1\\ \bar{g}_3 &= x_1^2 - \frac{1}{3}\\ \bar{g}_4 &= w_1 - 1,\end{aligned}$$

and hence $G_R = \langle \bar{g}_1, \bar{g}_2, \bar{g}_3, \bar{g}_4 \rangle$. One can check that

$$\begin{aligned}S(\bar{g}_1, \bar{g}_2) &= x_1\bar{g}_1 + \bar{g}_2\\ S(\bar{g}_1, \bar{g}_3) &= -\frac{1}{3}\bar{g}_1 + \bar{g}_3\\ S(\bar{g}_1, \bar{g}_4) &= -\bar{g}_1 + \bar{g}_4\\ S(\bar{g}_2, \bar{g}_3) &= \frac{1}{3}\bar{g}_2 + x_1\bar{g}_3\\ S(\bar{g}_2, \bar{g}_4) &= \bar{g}_2 + x_1\bar{g}_4\\ S(\bar{g}_3, \bar{g}_4) &= \frac{1}{3}\bar{g}_4 - \bar{g}_3.\end{aligned}$$

It takes only two iterations to check $\overline{S(\bar{g}_i, \bar{g}_j)}^{G_R} = 0$ for each $\bar{g}_i, \bar{g}_j \in G_R$. One may notice that $\bar{g}_1, \bar{g}_2, \bar{g}_3, \bar{g}_4$ are fairly simple polynomials where each polynomial has two terms and each term has total degree no greater than 2. Elements in reduced Gröbner basis tends to have simple form and lower degree. With these two features, a reduced Gröbner basis's size could be much smaller than a Gröbner basis. As the size is often small, it is more efficient to check if Buchberger's criterion holds on a reduced Gröbner basis.

The problem is now how to get a reduced Gröbner basis. In the case above we first generate a Gröbner basis, and reduce that to a reduced Gröbner basis. However, the process contribute no improvement in efficiency when considering the algorithm. Then one may ask if one can generate a reduced Gröbner basis during the algorithm. The answer is yes.

Consider Example **3.2** again. The ideal $I = \langle f_1, f_2, f_3, f_4 \rangle$ where

$$\begin{aligned}f_1 &= w_0 + w_1 - 2\\ f_2 &= w_0x_0 + w_1x_1\\ f_3 &= w_0x_0^2 + w_1x_1^2 - 2/3\\ f_4 &= w_0x_0^3 + w_1x_1^3\end{aligned}$$

Immediately we find $LT(f_2) = w_0x_0$ is divisible by $LT(f_1) = w_0$. Then one can see

$$f_2 = x_0f_1 - x_0w_1 + 2x_0 + w_1x_1 = x_0f_1 - g_5.$$

Here g_5 is the element in Gröbner basis G as in Example **3.2**. One can see that

$$I = \langle f_1, f_2, f_3, f_4\rangle = \langle f_1, f_3, f_4, g_5\rangle,$$

which is a consequence of $f_2 = x_0f_1 - g_5$. Then to find Gröbner basis of I with Buchberger's algorithm, one can start with the set $S' = \{f_1, f_3, f_4, g_5\}$ instead of the set $S = \{f_1, f_2, f_3, f_4\}$ where we replace f_2 by g_5. The leading term of g_5 is x_0x_1 while the leading term of f_2 is w_0x_0. As $w_0x_0 >_{lex} x_0x_1$, this replacement somehow simplifies the task.

Proposition 3.3 For a set of polynomial monomials $S = \{f_1, \cdots, f_n\}$ and $I = \langle G\rangle$, if $LM(f_i)$ is a multiple of $LM(f_j)$ for some $f_i, f_j \in S$, then

$$I = \{f_1, \cdots, f_{i-1}, \overline{S(f_i, f_j)}^S, f_{i+1}, \cdots, f_n\}.$$

It says that I dose not change if we replace f_i by $\overline{S(f_i, f_j)}^S$. This a direct result from the fact that f_i is a linear combination of elements in

$$\langle f_1, \cdots, f_{i-1}, \overline{S(f_i, f_j)}^S, f_{i+1}, \cdots, f_n\rangle,$$

with polynomial coefficients. Again If we have a review of Example **3.2** and running Buchberger's algorithm, the first step of both **Algorithm 1** and **Algorithm 2** is to compute

$$\overline{S(f_1, f_2)}^S = g_5,$$

and set $S_1 = S \cup \{g_5\}$. And by Proposition **3.43**, one can then delete f_2 from S_1 or equivalently it set $S_1' = S_1\backslash\{f_2\}$. Note when applying **Algorithm 1**, Buchberger's algorithm repeats the **for** loop which then calculate $\overline{S(f_1, f_3)}^S$. Here

$$\overline{S(f_1, f_3)}^S = x_0^2w_1 - 2x_0^2 - w_1x_1^2 + \frac{2}{3}.$$

However when running **Algorithm 2**, it calculates $\overline{S(f_1, f_3)}^{S_1}$ instead of $\overline{S(f_1, f_3)}^S$. And $x_0^2w_1$ is further divisible by $LT(g_5) = -x_0w_1$ which gives

$$\overline{S(f_1, f_3)}^{S_1} = x_0x_1 + \frac{1}{3}.$$

Still one can replace f_3 by $\overline{S(f_1, f_3)}^{S_1}$. Set $g_6 = \overline{S(f_1, f_3)}^{S_1}$ then the polynomial ideal is

$$I = \langle f_1, g_5, g_6, f_4\rangle.$$

In the process above, we do not only add elements to the set but also remove redundant elements. And the algorithm can be modified as following

Data: $\{f_1, f_2, ..., f_n\}$
Result: a Gröbner basis $G = \{g_1, ..., g_t\}$ for $I = \langle F \rangle$ and $F \subset G$
$G := F$;
$i := 2$;
$index := n + 1$;
while $i < index$ **do**
  forall $j < i$ **do**
    if $f_i = 0$ *or* $f_j = 0$ **then**
      continue;
    end
    $S := \overline{S(f_i, f_j)}^{G}$;
    if $LM(f_i)$ *is a multiple of* $LM(f_j)$ **then**
      $f_i := 0$;
    end
    if $LM(f_j)$ *is a multiple of* $LM(f_i)$ **then**
      $f_j := 0$;
    end
    if $S \neq 0$ **then**
      $f_{index} = S$;
      $G := G \cup \{f_{index}\}$;
      $index = index + 1$
    end
  end
  $i = i + 1$
end

Algorithm 3: Modified Buchberger's algorithm with reduction

Algorithm 3 is same as **Algorithm 2** except it set $f_i = 0$, if $LM(f_i)$ is a multiple of $LM(f_j)$. Or similarly set $f_j = 0$, if $LM(f_j)$ is a multiple of $LM(f_i)$. Note set $f_i = 0$ is equivalent to delete f_i, as for any non-zero polynomial f_k, $S(f_i, f_k)$ is always 0 when $f_i = 0$. The reason that we do not just remove f_i from the set is to avoid the **for** loop collapsing.

Note when considering Proposition **3.2**, it says we only need to check each S-polynomial $S(f_i, f_j)$ once for algorithm's termination. The Proposition still holds here.

One may also notice that we are not getting a strictly reduced Gröbner basis by the algorithm. For a Gröbner basis G, we only guarantee that for any $p, q \in G$, $LM(p)$ is not a multiple of $LM(q)$ while Reduced Gröbner basis has a further requirement that any term of p is not a multiple of $LM(q)$. However as the second requirement does not further reduce the size of G, or $|G|$, we may omit that for only the concern of simplicity.

Algorithm 3 has outstanding improvement compared with **Algorithm 2**. For

the problem in Example **3.2**, **Algorithm 3** takes 27 ms and **Algorithm 2** takes 764 ms. To have a taste of how the complexity increases, consider the following Gaussian Quadrature point problem with only one point, then the polynomial ideal is $I = \langle f_1, f_2 \rangle$ where

$$\begin{aligned} f_1 &= w_0 - 2 \\ f_2 &= w_0 x_0. \end{aligned}$$

Here there are only two variables, two polynomials and maximum total degree equals 2. To find Gröbner basis for such polynomial ideal, **Algorithm 2** takes 9.27 ms while **Algorithm 3** takes 4.74 ms.

For the case of harder problem, or specifically

$$\begin{aligned} f_1 &= w_0 + w_1 + w_2 - 2 \\ f_2 &= w_0 x_0 + w_1 x_1 + w_2 x_2 \\ f_3 &= w_0 x_0^2 + w_1 x_1^2 + w_2 x_2^2 - \frac{2}{3} \\ f_4 &= w_0 x_0^3 + w_1 x_1^3 + w_2 x_2^3 \\ f_5 &= w_0 x_0^4 + w_1 x_1^4 + w_2 x_2^4 - \frac{2}{5} \\ f_6 &= w_0 x_0^5 + w_1 x_1^5 + w_2 x_2^5, \end{aligned}$$

which corresponds to Gaussian Quadrature problem with 3 points. **Algorithm 3** takes 191 ms to terminate while it is hard to terminate with **Algorithm 2**. One can see from the experiment above, **Algorithm 3** is less sensitive to the number of variables, total degree and number of input polynomials.

Chapter 4

Floating point arithmetic

One question remains is if Buchberger's algorithm stable? The process to compute S-polynomial remainder involves many additions and multiplications. Roundoff error could occur in each step. Unfortunately the algorithm is especially vulnerable to these errors. Consider the single case below

Example 4.1

$$f_1 = x^2y + 2x$$
$$f_2 = \frac{1}{3}xy + \frac{2}{3}.$$

Obviously one can check $f_1 = 3f_2$ and the S-polynomial $S(f_1, f_2) = 0$. For the case of decimal system in application, $\frac{1}{3}$ has infinite digits, then to compute it in a numerical way, one needs to round it to some rational number. One rounding function maps number to nearest one. For the case of n valid digits, $\frac{1}{3}$ is then mapped to

$$0.\underbrace{3\ \cdots\ 3}_{n}.$$

Apparently this number is less than $\frac{1}{3}$ and let it be $\frac{1}{3} - \epsilon_1$ where $\epsilon_1 > 0$. Similarly $\frac{2}{3}$ is rounded to some number $\frac{2}{3} + \epsilon_2$ with $\epsilon_2 > 0$. Then the polynomial f_2 after perturbation becomes

$$\hat{f}_2 = (\frac{1}{3} - \epsilon_1)xy + (\frac{2}{3} + \epsilon_2).$$

One can see

$$S(f_1, \hat{f}_2) = \frac{3(\epsilon_1 + \epsilon_2)}{1 - 3\epsilon_1}x.$$

For any polynomial ideal I that contains f_1 and f_2, $S(f_1, f_2) \in I$. However the little perturbation of f_2 leads to some new element, $S(f_1, \hat{f}_2)$, that does not belong to I before. In the case above as $S(f_1, \hat{f}_2)$ is non zero, Buchberger's

algorithm adds $S(f_1, \hat{f}_2)$ into ideal I. Multiplying $S(f_1, \hat{f}_2)$ by a constant gives

$$\frac{1-3\epsilon_1}{3(\epsilon_1+\epsilon_2)} S(f_1, \hat{f}_2) = x.$$

It is then equivalent to add x into I. And again $LT(f_2)$ is divisible by x and adding x into I.

$$\frac{2}{3} = f_2 - \frac{1}{3}yx \in I.$$

It ends up with $1 \in I$ and all elements in I is only 0. In the case above a little rounding error leads a failure of the whole system.

Unfortunately, this is a common problem in Buchberger's algorithm algorithm. It is not always possible to have exact polynomial in implementation. For perturbed polynomials around original polynomial, we give the following definition.

Definition 4.1. For a polynomial p where

$$p(x) = \sum_{j \in J} \alpha_j x^j,$$

an *empirical polynomial* of p is defined as

$$p_\theta(x) = \sum_{j \in J} (\alpha_j + \theta_j) x^j,$$

where θ is a vector. One can see an empirical polynomial's coefficients are perturbed by some θ_j.

A set G is a Gröbner basis if and only if for any $f, g \in G$, $\overline{S(f,g)}^G = 0$. However if f, g is replaced by some empirical polynomials $f_{\theta_1}, g_{\theta_2}$, any non-zero elements in θ_1, θ_2 could make $\overline{S(f,g)}^G \neq 0$. Again if $\overline{S(f,g)}^G \neq 0$, it is then added to set G which means it adds some new element the ideal.

Setting tolerance is an intuitive way to solve the problem. It simply waves aside terms with small coefficients. Consider the example above again, if we have 10 valid digits, then $\epsilon_1, \epsilon_2 < 0.5 \times 10^{-10}$ [8]. The upper bound of ϵ_1 and ϵ_2 is a direct result of how we choose round function. One can check then

$$\frac{3(\epsilon_1+\epsilon_2)}{1-\epsilon_1} < 10^{-9}.$$

In such case the coefficients of $S(f_1, \hat{f}_2)$ are relatively small. If we set a tolerance equals 10^{-9}, it means eliminate all terms if the absolute value of their coefficients are less than the tolerance.

Tolerance is effective to eliminate small rounding errors. However it may also eliminate coefficients which are small themselves. For instance,

$$x^2 - 10^{-10}x = 0.$$

Setting tolerance could transform the problem to $x^2 = 0$. Large tolerance often means higher likelihood to cancel useful term in polynomial. Here useful means the term is not generated by rounding error. On the other hand, roundoff errors accumulates under multiplication and small tolerance does not work for complicated system in general. It is then often hard to find a trade-off between large or small tolerance.

4.1 Pivoting

In Chapter **3**, we see a system of polynomials can be written as a product of coefficient matrix C and monomial vector v. As a division algorithm often involves elimination of leading term, it shares some similarity with Gaussian elimination. Consider a following example [8],

$$A = \begin{bmatrix} 0.0001 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 10,000 & 1 \end{bmatrix} \begin{bmatrix} 0.0001 & 1 \\ 0 & -9999 \end{bmatrix} = LU.$$

It shows a LU-decomposition of A. One relative small entry in A leads to some large entry in L and U. Consider a polynomial system $I = \langle f_1, f_2 \rangle$ where

$$\begin{aligned} f_1 &= 0.0001x + y \\ f_2 &= x + y. \end{aligned}$$

Specify $x >_{lex} y$ and the division algorithm gives

$$f_2 - 10000f_1 = -9999y.$$

No surprisingly, the system of polynomial equations can be written as

$$A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0.0001 & 1 \\ 0 & -9999 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

And finding Gröbner basis of a polynomial ideal is somehow similar as Gaussian elimination of matrix A. Partial pivoting is one simple and effective trick to enhance numerical stability [2][8]. For a set of polynomials $\{f_1, \cdots, f_n\}$, if the leading monomial of $\{f_1, \cdots, f_n\}$ are same, let it be x^j. And one can write $LT(f_i) = k_i x^j$ for $i = 1, \cdots, n$ where k_i is the the coefficient of the leading term of polynomial f_i. Partial pivoting then works as choosing the polynomial f_i with largest $|k_i|$ as divisor. In the case above, partial pivoting swaps the first row and second row which gives

$$\begin{bmatrix} 1 & 1 \\ 0.0001 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Note when swapping two rows, it is equivalent to swap two polynomials in the matrix, which has no effect of polynomial ideal. And Gaussian elimination now gives

$$\begin{bmatrix} 1 & 1 \\ 0 & 0.9999 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Not only works for elimination of polynomials with same leading monomial, one can also apply partial pivoting in division algorithm. Consider the case

$$\begin{aligned} p_1 &= 0.0001x + y \\ p_2 &= x^2 + y. \end{aligned}$$

$LT(p_2)$ is then divisible by $LT(p_1)$. One can first set

$$xp_1 = 0.0001x^2 + xy.$$

In such case xp_1 and p_2 have same leading monomial and one can again apply partial pivoting. The case above demonstrates the use of partial pivoting in division algorithm. Although it is studied many times in Gaussian elimination, here we still give an example of how pivoting could improves the accuracy of Buchberger's algorithm.

Example 4.2 Consider a polynomial ideal $I = \langle p_1, p_2 \rangle$ where [2]

$$\begin{aligned} p_1(x,y) = &\ 0.02467x^3 - 0.02053x^2y + 2.82741y^3 + 0.68701x^2 \\ &+ 3.51842xy^2 + xy + \cdots \\ p_2(x,y) = &- 4.58163x^2 + 3.83952xy + 2.44073y^2 + x + y + \cdots \end{aligned}$$

Note here $LT(p_1)$ is a multiple of $LT(p_2)$, then the reduction gives the following matrix

	x^3	x^2y	y^3	x^2	xy^2	xy	$\cdots$
p_1	0.02467	−0.02053	2.82741	0.68701	3.51842	1	$\cdots$
xp_2	−4.58163	3.83952		1	2.44073	1	$\cdots$
yp_2		−4.58163	2.44073		3.83952	1	$\cdots$
p_2				−4.58163		3.83952	$\cdots$

If we apply Gaussian elimination, either with partial pivoting or not, and then backward substitution, it then ends with a matrix with following structure

x^3	x^2y	y^3	x^2	xy^2	xy	$\cdots$
×				×	×	$\cdots$
	×			×	∘	$\cdots$
		×		×	×	$\cdots$
			×	×	×	$\cdots$

If we have a system using 32-bit floating point numbers, which means there are 7 valid digits. The number at the circle in the matrix above only has two valid digits without pivoting and 6 valid digits with partial pivoting. Accuracy improves a lot by partial pivoting in the case above.

Partial pivoting does not only work for division algorithm. When computing S-polynomial, it still involves leading term reduction which generates some polynomial with new leading monomial in the ideal.

4.2 Empirical polynomial

For a polynomial $p(x) = \sum_{j\in J} \alpha_j x^j$, its empirical polynomial has the form

$$p(x;\theta) = \sum_{j\in J} (\alpha_j + \theta_j) x^j.$$

In many cases we are dealing with empirical polynomials rather than its original form. For a root x of polynomial $p(x)$, one has

$$p(x) = \sum_{j\in J} \alpha_j x^j = 0.$$

Now considering some empirical polynomial $p(x;\theta)$ of polynomial p,

$$p(x;\theta) = \sum_{j\in J} \alpha_j x^j + \sum_{j\in J} \theta_j x^j = p(x) + \sum_{j\in J} \theta_j x^j,$$

which then gives

$$|p(x;\theta)| = |\sum_{j\in J} \theta_j x^j| \leq \sum_{j\in J} |\theta_j x^j|.$$

If we have the some knowledge or assumption of θ, for example $|\theta_j| \leq \epsilon_j$ for $j \in J$, then when dealing with empirical polynomial $p(x;\theta)$, x is a solution of $p(x)$ only if

$$|p(x;\theta)| \leq \sum_{j\in J} |\theta_j x^j| \leq \sum_{j\in J} \epsilon_j |x^j|.$$

The inequality above gives us some feeling if x is a root of p when we only have its empirical polynomial. Consider example **4.1** again where $G = \{x^2y + 2x\}$ is a Gröbner basis. In such case any for any pairs of (x, y) where $xy = -2$, (x, y) is then a root. $f_2 = \frac{1}{3}xy + \frac{2}{3}$ is also an element in that ideal, and its empirical polynomial has the form

$$(\frac{1}{3} + \theta_1)xy + \frac{2}{3} + \theta_2.$$

Again assume that we are dealing with $|\theta_1|, |\theta_2| \leq 5 \times 10^{-11}$. Consider a pair $x = 2, y = -1$ which is an root of the polynomial system. Then

$$(\frac{1}{3} + \theta_1)xy + \frac{2}{3} + \theta_2 = -2\theta_1 + \theta_2.$$

If one has

$$|-2\theta_1+\theta_2| \leq 5\times 10^{-11}|xy|+5\times 10^{-11} = 1.5\times 10^{-10},$$

$x=2, y=-1$ is then a potential solution of the polynomial system.

4.3 Loss of accuracy

There are some methods to avoid redundant computation and hence reduce roundoff error [2]. However for complicated polynomial ideal, the procedure to find Gröbner basis is often hard too. Roundoff error accumulates in this procedure and grows fast.

To handle these errors, setting coefficients with absolute value less than a given tolerance ϵ to 0 is one intuitive way. However, the elements in polynomial ideal are also perturbed by this way. One can see that in a long run of Buchberger's algorithm, the elements in polynomial are perturbed once and once again and carry less information of the original polynomial ideal after each perturbation. At some stage, we are no longer working on the original polynomial any more. That happens when the coefficients of polynomials have only a few valid digits.

To handle this situation, one may always need to keep track of the original polynomial. For a polynomial set $S=\{f_1,\cdots,f_n\}$ ideal $I=\langle S\rangle$, each element p in I has the form

$$I=\sum_{i=1}^{n} k_i f_i,$$

where k_i is a polynomial coefficient. Of course when running Buchberger's algorithm, we keep adding elements to S.

Example 4.3 Consider a set $F=\{f_1,f_2\}$ and corresponding polynomial ideal $I\langle G\rangle$ where

$$f_1 = 3xyz+xy$$
$$f_2 = 7x^2y^2z+2xy.$$

To find a Gröbner basis of I, the first step is

$$S(f_1,f_2)=\frac{7}{3}x^2y^2-2xy.$$

Set $f_3=S(f_1,f_2)$ and add it to the set F. In numerical system, $\frac{7}{3}$ is rounded to some number that has the form 2.3...3, and hence f_3 is a little perturbed. The next step is to compute

$$\overline{S(f_1,f_3)}^F, \overline{S(f_1,f_3)}^F.$$

No surprisingly, $\overline{S(f_1,f_3)}^F$ is a linear combination of f_1,f_2,f_3 with polynomial coefficients. However as in numerical system, f_3 is already perturbed, and

$\overline{S(f_1, f_3)}^F$ will be annihilated even more. It keeps losing accuracy during this process. An intuitive way to avoid that is to further write f_3 as a linear combination of f_1 and f_2. And $\overline{S(f_1, f_3)}^F$ can be written as only a linear combination of f_1 and f_2.

Bibliography

[1] Stillwell, John (2004). *Mathematics and Its History (2nd ed.)*. Springer. ISBN 0-387-95336-1

[2] Hans J. Stetter (2004). *Numerical Polynomial Algebra*. ISBN: 0-89871-557-6

[3] Cox, Little, O'Shea (1998). *Using Algebra Geometry*. ISBN: 0-387-98487-9

[4] L. E. Dickson (1913). *Finiteness of the odd perfect and primitive abundant numbers with n distinct prime factors*. American Journal of Mathematics, 35:413–422.

[5] Buchberger, B. (1976). *Theoretical Basis for the Reduction of Polynomials to Canonical Forms*. ACM SIGSAM Bulletin. ACM. 10 (3): 19–29. doi:10.1145/1088216.1088219. MR 0463136

[6] W.W. Adam, P. Loustaunau, *An introduction to Gröbner bases* , Graduate Studies in Math. , 3 , Amer. Math. Soc. and Oxford Univ. Press (1994) MR1287608

[7] C.F. Gauss, *Methodus nova integralium valores per approximationem inveniendi* , Werke , 3 , K. Gesellschaft Wissenschaft. Göttingen (1886) pp. 163–196

[8] Golub, Van Loan, *Matrix Computation(2nd ed.)*. ISBN: 0-8018-3739-1