Algorithmic Elimination Methods

Deepak Kapur
Institute for Programming and Logics
Department of Computer Science
State University of New York
Albany, NY 12222
kapur@cs.albany.edu

Very rough draft: June 19, 1995

Abstract

This tutorial provides a gentle introduction to different methods for eliminating variables from a finite system of polynomials and for solving polynomial equations. Elimination methods have applications in many fields.

The first part of the tutorial will discuss resultant-based methods investigated in the 19th century and early 20th century. The main idea is to generate from a system of nonlinear polynomial equations, possibly a larger system of independent polynomial equations. The larger system has as many equations as distinct terms in the polynomials. Each term is treated as an unknown and the theory of linear algebra can be applied. Three major resultant methods due to Euler, Bezout, Sylvester, Cayley, Dixon and Macaulay, and a sparse reformulation of Macaulay will be discussed and compared.

In the second part, two other methods – Gröbner basis method based on polynomial ideal theory and the characteristic set method based on algebraic geometry will be reviewed. The use of these methods for elimination and equation solving will be discussed.

Examples from application domains such as computer vision, geometric reasoning and solid modeling will be used.

These notes discuss resultants and characteristic set methods as written material on these topics is not easily available. Discussion on Gröbner basis methods is omitted because there are now good books introducing Gröbner basis methods as well as discussing Gröbner basis theory in depth.
1 Introduction

Nonlinear polynomials are used to model many physical phenomena in engineering applications. Often there is a need to solve nonlinear polynomial equations, or if solutions do not have to be explicitly computed, it is necessary to study the nature of solutions. Many engineering problems can be easily formulated using polynomials with the help of extra variables. It then becomes necessary to eliminate some of those extra variables. Examples include implicitization of curves and surfaces, invariants with respect to transformation groups, robotics and kinematics. Projection of a variety (which is the set of solutions of polynomial equations in an algebraically closed field) is yet another useful tool in algebraic geometry. These applications require the use of elimination techniques on polynomials, and symbolically studying the solutions of the corresponding equations. To get an idea about the use of polynomials for modeling in different application domains, the reader may consult books by Hoffman [14], Abhyankar [2], Morgan [24], Kapur and Mundy [17], Donald, Kapur and Mundy [11].

In these notes, we discuss different symbolic methods for studying the solutions of polynomial equations. Because of lack of time and space, we are unable to discuss numerical techniques for solving polynomial equations. An interested reader may want to look at [28] for a simple approach based on interval methods and branch and prune techniques, that has been found to be quite effective. The method is compared with homotopy based continuation methods [24] as well as other interval based techniques.

In the first part of these notes, we focus on resultant methods. Resultant methods were developed in the 18th century by Newton, Euler and Bezout, in the 19th century by Sylvester and Cayley, and the early parts of 20th century by Dixon and Macaulay. These constructivists were attempting to extend simple linear algebra techniques developed for linear equations to nonlinear polynomial equations. A number of books on the theory of equations were written which are now out of print. Some very interesting subsections were omitted by the authors in revised editions of some of these books because abstract methods in algebraic geometry and invariant theory gained dominance, and constructive methods began to be viewed as too concrete to provide any useful structural insights. A good example is the 1970 edition of van der Waerden’s book on algebra which does not include a beautiful chapter on elimination theory that appeared as the first chapter in the second volume of its 1940 edition. For an excellent discussion of the history of constructive methods in algebra, the reader is referred to an article by Professor Abhyankar [1].

In the second part of this paper, we discuss Ritt’s characteristic set construction and its popularization by Wu Wen-tsuen. Despite its tremendous success in geometry theorem proving, characteristic set methods do not seem to have gotten as much exposure as they should. For many problems, characteristic set construction is quite efficient and powerful.

In these notes, we have focussed on the problems of computing zeros of polynomials, elimination of variables and projection of the the zeros of polynomials along certain dimensions.

We have attempted to provide information that is not easily available; otherwise, we have assumed that an interested reader can consult various references for further details. For instance, we have chosen not to include a section on Gröbner basis constructions and the related theory. For an introductory reading, an interested reader may consult Buchberger’s survey articles [4, 5], a section in our introductory paper on elimination methods [16],
Hoffman’s book [14], or Cox et al’s book [9]. For a comprehensive treatment, the reader can consult Weispfenning’s book [29].

For a discussion of characteristic set construction, the reader may consult Wu’s articles [31, 32], Chou’s book [8] or a section in our introductory article [16]. In these notes, we have added some new material on characteristic sets which could not be included in the introductory article because of lack of space.

Most of the material in these notes is on resultants since except for Sylvester’s resultants, very little material is included in text books. In order to conserve space, we have been selective in including examples to illustrate different concepts. In the oral presentation at the tutorial, we plan to use many examples – simple as well as nontrivial ones from different application domains. Examples have not been included to illustrate Sylvester resultants for the univariate case, Dixon’s resultants for the bivariate case, or Macaulay’s multivariate resultants as these examples are given in detail in our introductory paper [16]; an interested reader can consult [16] for examples on these topics. We have, however, included examples to illustrate sparse resultants as well as the new material on Dixon resultants.

1.1 Overview

In the next subsection, we provide basic background information about zeros of polynomials, distinction between affine zeros and projective zeros, and the zero set of a set of polynomials.

The next section is about resultants. We first review Euler’s method for computing the resultant of two polynomials by eliminating one variable, Sylvester’s reformulation using di-alytic expansion, Macaulay’s generalization for simultaneously eliminating many variables. $U$-resultants are reviewed to discuss how common zeros can be extracted using resultants. The second part starts with Bezout’s formulation of the resultant of two polynomials by eliminating one variable. This is followed by Cayley’s remarkable formulation of Bezout’s method which naturally generalizes nicely to a formulation for simultaneously eliminating many variables. Dixon’s formulation of the resultant for three polynomials in two variables is first discussed. Then this formulation is generalized. Two methods - generalized characteristic polynomial method based on perturbation and the rank submatrix method, are reviewed for extracting a projection operator from singular resultant matrices. An experimental comparison of different resultant formulations is given in the form of a table presenting performance of different methods on examples from different application domains. Some recent results on theoretical complexity of Dixon resultants are included.

The reader would note that the formulations in this section are expressed in terms of appropriate determinants which correspond to solving a linear system of equations in which the unknowns are the power products appearing in the polynomials. This linear system is obtained by transforming the original set of polynomial equations. This is in contrast to Gröbner basis and characteristic methods, where no system of linear equations is explicitly and a priori set up. Instead, these methods appear to be similar to completion algorithms in which new polynomials are successively constructed until a subset of polynomials with the desired property is generated.

The third section discusses Ritt’s characteristic sets. We have pointed fundamental differences in Ritt’s definition and the definition used by Wu in his work. We discuss the use of characteristic set method for elimination. As an alternative to decomposition of a zero set into irreducible zero sets based on Rit-Wu’s algorithm, a different method using Duval’s D5 algorithm is presented for decomposing a zero set.
1.2 Preliminaries

Let \( \mathbb{Q} \), \( \mathbb{R} \), \( \mathbb{C} \), respectively, denote the fields of rational numbers, real numbers, and complex numbers. Unless specified otherwise, by a polynomial, we mean a multivariate polynomial with rational coefficients. A multivariate polynomial can be viewed as a sum of products with nonzero rational coefficients, where each product is called a term or power product; together with its coefficient, it is called a monomial.

A univariate polynomial \( p(x) \) is said to vanish at a value \( a \) if the polynomial \( p \) evaluates to zero when \( a \) is uniformly substituted for \( x \) in \( p \). The value of \( p \) at \( a \) is denoted by \( p(a) \); if \( p(a) = 0 \), then \( a \) is called a zero of \( p \). Equivalently, the polynomial equation \( p(x) = 0 \) is said to have a solution \( a \). The domain from which the values are picked to be checked for zeros of \( p \) is very important. It is possible for \( p \) to have a zero in one domain and not in another domain. For instance, \( x^2 + 1 \) does not have a zero in the reals, but it does have a zero in the complex numbers.

Given a multivariate polynomial \( p(x_1, \ldots, x_n) \), an \( n \)-tuple \( (a_1, a_2, \ldots, a_n) \in \mathbb{C}^n \), the affine \( n \)-space over complex numbers, is a zero of \( p \) if \( p \) evaluates to zero when, for each \( 1 \leq i \leq n \), \( a_i \) is uniformly substituted for \( x_i \), i.e., \( p(a_1, a_2, \ldots, a_n) = 0 \). Equivalently, \( (a_1, a_2, \ldots, a_n) \) is called a solution of \( p(x_1, \ldots, x_n) = 0 \) if \( p(a_1, a_2, a_n) = 0 \). Given a set \( \{ f_1(x_1, x_2, \ldots, x_n), f_2(x_1, x_2, \ldots, x_n), \ldots, f_r(x_1, x_2, \ldots, x_n) \} \) of polynomials (equivalently a system \( \{ f_1(x_1, x_2, x_n) = 0, f_2(x_1, x_2, x_n) = \ldots, f_r(x_1, x_2, x_n) = 0 \} \) of polynomial equations), \( (a_1, a_2, \ldots, a_n) \) is a common zero (respectively, a common solution) of the system, if, for each \( 1 \leq i \leq r \), \( f_i(a_1, a_2, \ldots, a_n) = 0 \), i.e., \( (a_1, a_2, \ldots, a_n) \) is a zero of every polynomial in the system. The \( n \)-tuple \( (a_1, a_2, \ldots, a_n) \) is also called a common root of these polynomials. Henceforth, we will abuse the terminology: by a system, we will mean either a set of polynomials or a set of polynomial equations, with the hope that the context can resolve the ambiguity.

Given a system \( S \) of polynomials, the set

\[
\text{Zero}(S) = \{(a_1, a_2, \ldots, a_n) \in \mathbb{C}^n \mid \forall f \in S, f(a_1, a_2, \ldots, a_n) = 0\}
\]

is called the zero set of \( S \). The zero set defined by a system of polynomials is also called an algebraic set. Throughout this paper, we have focussed on zeros in the field of complex numbers (in general, an algebraically closed field).

The zero set of a system of polynomials could be infinite. In that case, it is possible, like in linear algebra, to talk about the dimension of the zero set. The system is said to have positive dimension. If the zero set is finite, the system is called zero-dimensional. If the zero set is empty, then the system is said to be of dimension -1.

The familiar \( n \)-space \( \mathbb{C}^n \) is known as the affine \( n \)-space over the complex numbers. The projective \( n \)-space over the complex numbers, denoted by \( \mathbb{P}^n \), consists of all \( (n + 1) \)-tuples over the complex numbers except \((0, 0, \ldots, 0)\). Each \((n + 1)\)-tuple in \( \mathbb{P}^n \) is said to represent a point in the projective \( n \)-space. However, the tuples \((a_0, a_1, \ldots, a_n)\) and \((\lambda a_0, \lambda a_1, \ldots, \lambda a_n)\) are said to represent the same point for any non-zero complex number \( \lambda \) and the two \((n + 1)\)-tuples are said to be equivalent. Thus, each point in \( \mathbb{P}^n \) is represented by any member of an infinite set of proportionate (equivalent) \( n \)-tuples. Any \((n + 1)\)-tuple \( B = (b_0, b_1, b_2, \ldots, b_n) \) in which \( b_0 \neq 0 \) is associated with the unique point in affine \( n \)-space whose coordinates are

\[
\tilde{B} = (b_1/b_0, b_2/b_0, \ldots, b_n/b_0).
\]
Those \((n+1)\)-tuples which have \(b_0 = 0\) are said to represent \textit{points at infinity}. The set of all points at infinity in \(P^n\) is known as the \textit{hyper plane at infinity}. The affine \(n\)-space can be embedded into the projective \(n\)-space. To each point \((c_1, c_2, \ldots, c_n) \in \mathbb{C}^n\), associate the \((n+1)\)-tuple \((1, c_1, c_2, \ldots, c_n)\).

A polynomial \(f(x_0, x_1, \ldots, x_n)\) is said to be homogeneous of degree \(d\) if each term in \(f\) has degree \(d\). If \((a_0, a_1, \ldots, a_n)\) is a zero of the homogeneous polynomial \(f(x_0, x_1, \ldots, x_n)\), i.e., \(f(a_0, a_1, \ldots, a_n) = 0\), then any other \((n+1)\)-tuple \((\lambda a_0, \lambda a_1, \ldots, \lambda a_n)\) is also a zero of \(f(x_0, x_1, \ldots, x_n)\).

Given a non-homogeneous polynomial \(f(x_1, x_2, \ldots, x_n)\), of degree \(d\), it can be \textit{homogenized} as follows. Consider

\[
^h f(x_0, x_1, x_2, \ldots, x_n) = x_0^d f(x_1/x_0, x_2/x_0, \ldots, x_n/x_0)
\]

where \(x_0\) is a new variable. Polynomial \(^h f(x_0, x_1, \ldots, x_n)\) is homogeneous and of degree \(d\) such that

\[
^h f(1, x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n).
\]

Let \((a_1, a_2, \ldots, a_n) \in \mathbb{C}^n\) be a zero of \(f\), i.e., \(f(a_1, a_2, \ldots, a_n) = 0\). Then, \((1, a_1, a_2, \ldots, a_n) \in P^n\) (or any \((n+1)\)-tuple equivalent to it) is a zero of \(^h f\). Conversely, if \((a_0, a_1, a_2, \ldots, a_n) \in P^n\) is a zero of \(^h f\), and \(a_0 \neq 0\), then, \((a_1/a_0, a_2/a_0, \ldots, a_n/a_0) \in \mathbb{C}^n\) is a zero of \(f\). If \(a_0 = 0\), then there is no corresponding point in the affine space that is a zero of \(f\). Such zeros of \(^h f\) are called \textit{zeros at infinity} of \(f\).

## 2 Resultants

Resultant means the result of elimination. It is also the projection of intersection. Let us start with a simple example. Given two polynomials \(f(x)\), \(g(x) \in \mathbb{Q}[x]\) of degrees \(m\) and \(n\) respectively, i.e.,

\[
\begin{align*}
  f(x) &= f_n x^n + f_{n-1} x^{n-1} + \ldots + f_1 x + f_0, \quad \text{and}, \\
  g(x) &= g_m x^m + g_{m-1} x^{m-1} + \ldots + g_1 x + g_0.
\end{align*}
\]

when do \(f\) and \(g\) have common roots?

According to the western literature on the history of mathematics [21], Newton in his \textit{Arithmetica Universalis} discussed rules for eliminating a variable from two equations of degree two, three and four. It would be quite surprising to find that ancient Indian, Chinese and Arab mathematicians did not know methods for solving such equations even before Newton (at least for the restricted cases).

### 2.1 Euler’s method

Here is a derivation due to Euler (1764). Let \((x - \alpha)\) be a common root of \(f\) and \(g\). Then \(f = (x - \alpha)f_1\) and \(g = (x - \alpha)g_1\). In other words, there exist polynomials \(\phi(x), \psi(x) \neq 0\) with \(\deg(\phi) < m\) and \(\deg(\psi) < n\) such that

\[
\phi f + \psi g = 0
\]
if and only if \( f \) and \( g \) have a common root. Let
\[
\phi(x) = \phi_{m-1}x^{m-1} + \phi_{m-2}x^{m-2} + \ldots + \phi_1x + \phi_0, \quad \text{and,}
\psi(x) = \psi_{n-1}x^{n-1} + \psi_{n-2}x^{n-2} + \ldots + \psi_1x + \psi_0.
\]

The above implies that for every \( x \), the polynomial \( \phi f + \psi g \) is identically zero, i.e., each of the coefficients of \( \phi f + \psi g \) is zero. Collecting coefficients of \( 1, x, x^2, \ldots, x^{m+n-1} \) from \( \phi f + \psi g \), we get,
\[
\phi_0 f_0 + \psi_0 g_0 = 0,
\phi_1 f_0 + \phi_0 f_1 + \psi_1 g_0 + \psi_0 g_1 = 0,
\phi_2 f_0 + \phi_1 f_1 + \phi_0 f_2 + \psi_2 g_0 + \psi_1 g_1 + \psi_0 g_2 = 0,
\vdots
\phi_{m-1} f_{n-1} + \phi_{m-2} f_n + \psi_{n-1} g_{m-1} + \psi_{n-2} g_m = 0,
\phi_{m-1} f_n + \psi_{n-1} g_m = 0.
\]

This constitutes a square system of \((m+n)\) linear equations for \((m+n)\) unknowns \( \phi_0, \phi_1, \ldots, \phi_{m-1}, \psi_0, \psi_1, \ldots, \psi_{n-1} \) which can be written in matrix form as
\[
R \bar{a} = 0
\]
where
\[
R = \begin{pmatrix}
    f_0 & 0 & 0 & \ldots & 0 & g_0 & 0 & 0 & \ldots & 0 \\
    f_1 & f_0 & 0 & \ldots & 0 & g_1 & g_0 & 0 & \ldots & 0 \\
    f_2 & f_1 & f_0 & \ldots & 0 & g_2 & g_1 & g_0 & \ldots & 0 \\
    \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
    f_n & f_{n-1} & f_{n-2} & \ldots & f_{n-m+1} & g_n & g_{n-1} & g_{n-2} & \ldots & g_0 \\
    0 & f_n & f_{n-1} & \ldots & f_{n-m+2} & g_{n+1} & g_n & g_{n-1} & \ldots & g_1 \\
    \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & 0 & \ldots & f_n & 0 & 0 & 0 & \ldots & g_m
\end{pmatrix}
\quad \text{and} \quad \bar{a} = \begin{pmatrix}
    \phi_0 \\
    \phi_1 \\
    \phi_2 \\
    \vdots \\
    \phi_{m-1} \\
    \psi_0 \\
    \psi_1 \\
    \psi_2 \\
    \vdots \\
    \psi_{n-1}
\end{pmatrix}
\]

(it is assumed that \( g_i = 0 \) if \( i > m \)). The existence of a non-zero solution implies that the determinant of the matrix \( R \) is zero.

### 2.2 Sylvester’s Dialytic Method

This matrix can also be obtained by using Sylvester’s dialytic method of elimination in which the equation \( f(x) = 0 \) is successively multiplied by \( x^{m-1}, \ldots, x, 1 \) to generate \( m \) equations and the equations \( g(x) = 0 \) is successively multiplied by \( x^{n-1}, \ldots, x, 1 \) to generate \( n \) equations. We have:
\[
R^\text{Tr} \bar{X}^\text{Tr} = 0,
\]
where
\[
\bar{X} = \begin{pmatrix}
    x^{n+m-1} & \ldots & x & 1
\end{pmatrix}.
\]
There are thus \( m + n \) equations in which the \( m + n \) power products \( x^{n+m-1}, \ldots, x, 1 \) act as unknowns. The determinant of \( R \) has come to be known as the \textit{Sylvester resultant} of \( f \) and \( g \), and \( R \) is called the \textit{Sylvester matrix}.

Notice that the Sylvester resultant is a homogeneous polynomial in the coefficients of \( f \) and \( g \); it is of degree \( m \) in the coefficients of \( f \), and of degree \( n \) in the coefficients of \( g \), i.e., each power product in the resultant has \( m \) \( f_i \)'s and \( n \) \( g_j \)'s.

Assuming that at least one of \( f_n, g_m \) is non-zero, the vanishing of the Sylvester resultant is a necessary and sufficient condition for \( f \) and \( g \) to have common roots.

Let \( \alpha_i, 1 \leq i \leq n \), be the roots of \( f \) and similarly \( \beta_i, 1 \leq i \leq m \), be the roots of \( g \). Then the resultant of \( f \) and \( g \) is

\[
\prod_{i=1}^{n} g(\alpha_i) = (-1)^{mn} \prod_{i=1}^{m} f(\beta_i) = \prod_{i=1}^{m} \prod_{j=1}^{n} (\alpha_i - \beta_j).
\]

See [16] for an example illustrating this construction, where Collins’ and Brown’s beautiful subresultant theory is also reviewed for efficiently computing determinant of the resultant matrix.

The Sylvester resultant can be used successively to eliminate several variables, one at a time. That is what one ends up doing while using commercially available computer algebra packages such as Maple, Mathematica, Reduce etc. since these systems do not seem to provide support for simultaneously eliminating several variables from a polynomial system. One soon discovers that the method suffers from an explosive growth in the degrees of the polynomials generated in the successive elimination steps. If one starts out with \( n \) or more polynomials in \( \mathbb{Q}[x_1, x_2, \ldots, x_n] \), whose degrees are bounded by \( d \), polynomials of degree doubly exponential in \( n \), i.e., \( d^{2^n} \), can get generated in the successive elimination process. The technique is impractical for eliminating more than three variables.

### 2.3 Macaulay’s Multivariate Resultants

Macaulay generalized the resultant construction for eliminating one variable to eliminating several variables simultaneously from a system of homogeneous polynomial equations. His construction is also a generalization of the determinant of a system of homogeneous linear equations. For solving non-homogeneous polynomial equations, the polynomials must be homogenized first. Methods based on Macaulay’s matrix give out zeros in \( \mathbb{P}^n \) for the homogenized system of equations and they can include zeros at infinity.

The key idea is to show which power products are sufficient in the dialytic method to be used as multipliers for the polynomials so as to produce a square system of linear equations in as many unknowns as the equations.

Let \( f_1, f_2, \ldots, f_n \) be \( n \) \textit{homogeneous} polynomials in \( x_1, x_2, \ldots, x_n \). Let \( d_i = \deg(f_i) \), \( 1 \leq i \leq n \), and \( d_M = 1 + \sum_{i=1}^{n} (d_i - 1) \). Let \( T \) denote the set of all terms of degree \( d_M \) in the \( n \) variables \( x_1, x_2, \ldots, x_n \), i.e.,

\[
T = \left\{ x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n} | \alpha_1 + \alpha_2 + \ldots + \alpha_n = d_M \right\}
\]

and

\[
|T| = \binom{d_M + n - 1}{n - 1}.
\]
Polynomial $f_1$ is viewed to introduce the variable $x_1$; similarly, $f_2$ is viewed to introduce $x_2$, and so on. For $f_1$, the power products used to multiply $f_1$ to generate equations are all terms of degree $d_M - d_1$, where $d_1$ is the degree of $f_1$. For $f_2$, the power products used to multiply $f_2$ to generate equations are all terms of degree $d_M - d_2$ that are not multiple of $x_1^{d_1}$; power products that are multiples of $x_1^{d_1}$ are considered to be taken care of while generating equations from $f_1$. Similarly, for $f_3$, the power products used to multiply $f_3$ to generate equations are all terms of degree $d_M - d_3$ that are not multiple of $x_1^{d_1}$ or $x_2^{d_2}$, and so on. The order in which polynomials are considered for selecting multipliers results in different (but equivalent) systems of linear equations.

Macaulay showed that the above construction would result in $|T|$ equations in which power products in $x_i$’s of degree $d_M$ (these are the terms in $T$) are unknowns. A square matrix $A$ is constructed with $|T|$ columns and rows in which the columns of $A$ are labeled by the terms in $T$ in some order, and rows are labeled by the power products used as multipliers for generating equations from different $f_i$’s.

Let $\det(A)$ denote the determinant of $A$, which is a polynomial in the coefficients of the $f_i$’s. It is easy to see that $\det(A)$ contains the resultant, denoted as $R$, as a factor. This polynomial $\det(A)$ is homogeneous in the coefficients of each $f_i$. From the above analysis of the way power products to multiply $f_i$’s are computed to generate additional equations, it is easy to see that there are $d_1d_2\ldots d_{n-1}$ power products used to multiply $f_n$, so the degree of $\det(A)$ in the coefficients of $f_n$ is $d_1d_2\ldots d_{n-1}$.

Macaulay discussed two ways to extract a resultant from the determinants of $A$ matrices. The resultant $R$ can be computed by taking the gcd of all possible determinants of different $A$ matrices that can be constructed by ordering $f_i$’s in different ways (i.e., viewing them as introducing different variables). This is however quite an expensive way of computing a resultant. Macaulay also constructed the following formula relating $R$ and $\det(A)$:

$$R \det(B) = \det(A)$$

where $\det(B)$ is the determinant of a submatrix of $A$. The submatrix $B$ is obtained from $A$ by deleting those columns labeled by terms not divisible by any $n-1$ of the power products $\{x_1^{d_1}, x_2^{d_2}, \ldots, x_n^{d_n}\}$, and deleting those rows which contain at least one non-zero entry in the deleted columns.

An example illustrating Macaulay’s formulation is discussed in [16].

2.4 U-Resultant

If we wish to extract common zeros of a system $\mathcal{F} = \{f_1(x_1, x_2, \ldots, x_n), \ldots, f_n(x_1, x_2, \ldots, x_n)\}$ of $n$ polynomials in $n$ variables, this can be done using the construction of $u$-resultants discussed in [27]. Introduce a new homogenizing variable $x_0$ and let $R_u$ denote the Macaulay resultant of the $(n+1)$ homogeneous polynomials $^h f_1, ^h f_2, \ldots, ^h f_n, f_u$ in $(n+1)$ variables $x_0, x_1, \ldots, x_n$ where $f_u$ is the linear form

$$f_u = x_0u_0 + x_1u_1 + \ldots + x_nu_n,$$

and $u_0, u_1, \ldots, u_n$ are new unknowns. $R_u$ is a polynomial in $u_0, u_1, \ldots, u_n$, homogeneous in the $u_i$ of degree $B = \prod_{i=1}^{n} d_i$. $R_u$ is known as the $u$-resultant of $\mathcal{F}$. It can be shown that

1That is why polynomial $f_1$, for instance, is viewed as introducing variable $x_1$. 8
of a set of polynomials, one does so by computing specializations of the resultant of $g_i$, $i = 1, \ldots, n$. The converse can also be proved, i.e., if $(\beta_0, \beta_1, \ldots, \beta_n)$ is a common zero of $g_i$, $i = 1, \ldots, n$ then $(u_0\beta_0 + u_1\beta_1 + \ldots + u_n\beta_n)$ divides $R_u$. This gives an algorithm for finding all the common zeros of $g_1, g_2, \ldots, g_n$.

Recall that $B$ is precisely the Bezout bound on the number of common zeros of $R_u$ when zeros at infinity are included and their multiplicity is also taken into consideration.

Constructing the full $u$-resultant, however, is an almost impossible task (it is a polynomial of degree $\prod_{i=1}^{n} d_i$ in $n$ variables). So, if one is interested in computing the common zeros of a set of polynomials, one does so by computing specializations of $R_u$. For example, the univariate polynomial $R_1(u_0)$ obtained by substituting $u_i = 0$ for $i = 2, \ldots, n$ and $u_1 = -1$ has as its roots the $x_1$ coordinates of the common zeros. $R_1(u_0)$ can be computed from the Macaulay matrices by evaluation of determinants with rational (or complex) entries and interpolation and without constructing the full $u$-resultant.

### 2.5 Reality Strikes

The above methods based on Macaulay’s formulation do not always work, however. For a specific polynomial system in which the coefficients of terms are specialized, matrix $A$ may be singular or matrix $B$ may be singular. In fact, there are examples of polynomial systems for which one ordering of polynomials gives singular matrices $A$ and/or $B$, but a different ordering of polynomials gives nonsingular matrices. And, of course, there are examples for which no matter what ordering of polynomials is used, matrix $A$ is singular.

If $F$ has infinitely many common zeros, then its $u$-resultant $R_u$ is identically zero since for every zero, $R_u$ has a factor. Even if we assume that $f_i$’s have only finitely affine common zeros, that is not sufficient since the $u$-resultant vanishes whenever there are infinitely many common zeros of the homogenized polynomials $h f_i$’s – finitely many affine zeros but infinitely many zeros at infinity (when this happens, the zero set is said to have an excess component at infinity). And, that is often the case. The $u$-resultant algorithm cannot be used as it is in this situation.

One can try ad hoc perturbations of the polynomial system in the hope of eliminating excess components at infinity. Randomly perturb the system using a perturbation variable. Compute the projection operator for this perturbed system, and substitute zero for the perturbation variable in the result to get the projection operator for the original system. Grigoriev and Chistov, and Canny [6] suggested a modification of the above algorithm that will give all the affine zeros of the original system (as long as they are finite in number) even in the presence of excess components at infinity.

Let $f_1, f_2, \ldots, f_n$ be as before. Let $g_i = h f_i + \lambda x_i^d_i$, for $i = 1, \ldots, n$, and $g_u = (u_0 + \lambda)x_0 + u_1x_1 + \ldots + u_n x_n$, where $\lambda$ is a new unknown. Let $R_u(\lambda, u_0, \ldots, u_n)$ be the Macaulay resultant of $g_1, g_2, \ldots, g_n$ and $g_u$, regarded as homogeneous polynomials in $x_0, x_1, \ldots, x_n$. Now, look at $R_u(\lambda, u_0, \ldots, u_n)$ as a polynomial in $\lambda$ whose coefficients are polynomials in $u_0, u_1, \ldots, u_n$. 

$/
where $k \geq 0$ and the $R_k$ are polynomials in $u_0, u_1, \ldots, u_n$ (if $k = 0$, $R_k$ will be the same as the $u$-resultant $R_u$; however, if there are excess components at infinity, then $k > 0$). The trailing coefficient $R_k$ is called the \textit{generalized characteristic polynomial}. Just like $u$-resultant, it can be factored and the affine common zeros can be extracted from these factors even in the presence of excess components at infinity. Again, in practice, one never constructs the complete generalized characteristic polynomial of a system of polynomials. As with the $u$-resultant, one can recover the affine common zeros by computing specializations of the generalized characteristic polynomial.

Later we discuss Dixon’s formulation of multivariate resultants. For extracting a projection operator, we discuss a linear algebra construction, called \textit{rank submatrix}. This construction can be used for Macaulay matrices also.

2.6 Sparse Resultants

The reader might have noticed that the Macaulay matrix constructed above is sparse with most entries being 0. This is so because the size $T$ of the matrix is much larger as compared to the number of terms in a polynomial $f_i$. The size of the Macaulay matrix is determined by the total degree of the input polynomials. However it should be obvious that the number of affine roots of polynomial does not directly depend upon the total degree of polynomials, and that this number decreases as certain terms are deleted from polynomials. This observation has led to the development of sparse elimination theory and sparse resultants. The theoretical basis of this approach is the so called \textit{BKK bound} [13] on the number of zeros of a polynomial system in a toric variety (in which no variable in a zero can take a zero value), which depends only on the volume of the Newton polytopes associated with polynomials irrespective of their degrees. The BKK bound is much tighter than the Bezout bound used in Macaulay’s formulation.

The main idea is to refine the subset of multipliers needed using \textit{Newton Polytopes}. Given a polynomial $f_i$, the exponents vector corresponding to every term in $f_i$ defines a point in $n$-dimensional Euclidean space $\mathbb{R}^n$. The Newton polytope of $f_i$ is the convex hull of the set of points corresponding to exponent vectors of all terms in $f_i$. The \textit{Minkowski sum} of two Newton polytopes $Q$ and $S$ is the set of all vector sums, $q + s$, $q \in Q$, $s \in S$. Let $Q_i \subset \mathbb{R}^n$ be the Newton polytope (wrt. $X$) of the polynomial $p_i$ in $F$. Let $Q = Q_1 + \cdots + Q_{n+1} \subset \mathbb{R}^n$ be the Minkowski sum of Newton polytopes of all the polynomials in $F$. Let $E$ be the set of exponents (lattice points of $\mathbb{Z}^n$) in $Q$ obtained after applying a small perturbation to $Q$ to move as many boundary lattice points as possible outside $Q$.

Construction of $\hat{F}$ here is similar to Macaulay formulation - each polynomial, $p_i$ in $F$ is multiplied with certain terms to generate $\hat{F}$ with $|E|$ equations in $|E|$ unknowns (which are terms in $E$), and its coefficient matrix is the \textbf{sparse resultant matrix}. A projection operator for $F$ is simply the determinant of this matrix. Columns of the sparse resultant matrix are labeled by the terms in $E$ in some order, and each row corresponds to a polynomial in $F$ multiplied by a certain term.

In contrast to the size of the Macaulay matrix ($|T|$), the size of the sparse resultant matrix is $|E|$ and is typically smaller than $|T|$, especially for polynomial systems for which
the BKK bound is tighter than the Bezout bound. Canny and his colleagues have developed algorithms to construct matrices using some greedy heuristics which may result in smaller matrices, but in the worst case, the size can still be $|E|$.

**Example 1:** Let $\mathcal{F} = \{f_1, f_2, f_3\} \subset \mathfrak{g}[a, b, c, x, y]$ where

\[
\begin{align*}
\rho &= -2a^2 (a - b - c) \left( a^2 c + abc - 4ab + 4b^3 \right) \rho (a, b, c),
\end{align*}
\]

and the variables $x$ and $y$ are to be eliminated to obtain the resultant which is a polynomial in the parameters $a, b$ and $c$.

Construction of the Newton polytopes and their Minkowski sum (after a perturbation in the positive direction) reveals that $\mathcal{E} = \{xy, xy^2, xy^3, x^2y, x^2y^2, x^3y, x^3y^2, x^4y\}$. The $9 \times 9$ sparse resultant matrix is:

\[
\begin{pmatrix}
xy & x^2y & x^3y & x^2 & x^3 & x^2y^2 & x^3y^2 & x^4y & x^4
x^2 & 3(c-1) & a & 0 & a + c & b & 0 & a & 0
x^3y & f_1 & 0 & 3(c-1) & a & 0 & b + a + c & 0 & b & a
x^4y & f_2 & 0 & 0 & 0 & 3(c-1) & a & 0 & b + a + c & b & a
x^2 & f_1 & b^4 & ab & 0 & 0 & 2ab & 0 & 2b^2 & 0 & 0
x^2 & f_2 & 0 & b^4 & ab & 0 & 0 & 2ab & 0 & 2b^2 & 0 & 0
x^2 & f_3 & 0 & 0 & 0 & b^4 & ab & 0 & 2ab & 0 & 2b^2 & 0
x^2 & f_4 & 0 & 4ab & (a+b) & 0 & 0 & 2b & 0 & 2ab & 0 & 0
x^2 & f_5 & 0 & 4ab & (a+b) & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\]

The determinant of the above matrix is a projection operator with 205 terms (in expanded form):

\[
\rho(a, b, c) = 144ca^2b^3 + 28a^2c^2b^5 - 144a^4b^4 + 36c^2a^2b^5 - 72a^3c^2b^5 - 144abc - 16b^4c + 96ab^5b^2 + 240a^4b^2 - 264a^4c^2b + 336a^4c^2b - 144a^4cb - 72a^5c^2 - 72a^5c^3 + 36a^5c^2
\]

\[
+ 112a^5b^3 - 8ca^5b^2 - 80a^5b^4 + 20ca^4b^4 - 12c^2a^2b^5 + 48a^3b^6 + 288a^2b^7
\]

\[
- 96a^3b^5 + 192a^3b^3 - 48a^2b^4 - 192a^5b^2 - 32a^5b^2 + 16a^5b^3c + 16a^5b^5c
\]

\[
- 8a^3b^6c - 8a^2b^6c^2 + 2a^2c^2b^6c - 12a^2c^2b^5c + 80a^3b^6c + 21a^2c^2b^6c - 16a^2c^2b^7
\]

\[
- 96a^3b^5c - 80a^3b^5c^2 + 320a^3b^5c^2 - 40a^3ab^5c - 16a^3b^5c^3 + 44a^2c^2b^6c + 4cb^9
\]

\[
+ 2a^3b^9 + 16a^3b^4c + 108a^3b^4c^2 - 20a^3b^5c - 28a^3b^5c - 8a^2c^3b^5 - 2a^5c^2b^3 + 72a^4c^3b^4 + 4a^2c^4b^4 + 36a^3c^2b^4 + 2ac^4b^5 - 24c^2a^2b^7 + 120c^2a^2b^7 + 64a^2b^7
\]

\[
+ 72ca^2b^4 - 4cab^7 + 4c^3ab^6 - 52c^2a^2b^7 + 2c^2a^2b^7 + 48c^ab^7 + 16a^2b^7 + 16c^ab^7
\]

\[
+ 96a^6cb - 24a^5c^2b - 24a^4b^2c^2 + 216a^4b^2c^2 + 40a^4c^3b^2 + 40a^3c^3b^3 + 36a^5c^4
\]

\[
- 216a^5c^3b^3 - 384a^5c^3b^3 - 264c^2a^3b^7 - 24a^2c^2b^9.
\]

$\rho(a, b, c)$ can be shown to be the resultant. So the projection operator computed by the sparse resultant method has 3 extraneous factors. □

Much like Macaulay matrices, for specialization of coefficients, the sparse resultant matrix can be singular (Example 2 in Table 1) – even though this happens a lot rarely than in the case of Macaulay’s formulation
So far, we have used the dialytic method for setting up the resultant matrix and computing the projection operator. To summarize, the main idea in this method is to identify enough power products which can be used as multipliers for polynomials to generate a square system with as many equations generated from the polynomials as the power products in the resulting equations.

In the next few sections, we discuss a related but different approach for setting up the resultant matrix based on the methods of Bezout and Cayley.

2.7 Bezout’s Method

In 1764, Bezout also developed a method for computing the resultant which is quite different from Euler’s method discussed in the previous subsection. According to Kline [21] (Vol. II, pp. 607-608), Bezout’s method was apparently most widely accepted one until Sylvester proposed his dialytic method which as seen above, is a reformulation of Euler’s method. In Bezout’s method, instead of power products, polynomials are multiplied by polynomials constructed using coefficients of terms in \( f(x) \) and \( g(x) \). Further, instead of generating \( m + n \) equations, Bezout constructed \( n \) equations (assuming \( n \geq m \)) as follows: (i) First multiply \( g(x) \) by \( x^{n-m} \) to make the result a polynomial of degree \( n \). (ii) From \( f(x) \) and \( g(x)x^{n-m} \), construct equations in which the degree of \( x \) is \( n-1 \), by multiplying:

1. \( f(x) \) by \( g_m \) and \( g(x)x^{n-m} \) by \( f_n \) and subtracting,
2. \( f(x) \) by \( g_mx + g_{m-1} \), and \( g(x)x^{n-m} \) by \( f_nx + f_{n-1} \) and subtracting,
3. \( f(x) \) by \( g_mx^2 + g_{m-1}x + g_{m-2} \), and \( g(x)x^{n-m} \) by \( f_nx^2 + f_{n-1}x + f_{n-2} \) and subtracting, and so on.

This construction yields \( m \) equations. Additional \( n - m \) equations are obtained by multiplying \( g(x) \) by \( 1, x, \ldots, x^{n-m-1} \), respectively.

\[
\begin{align*}
g_m f(x) - f_n g(x)x^{n-m} &= 0 \\
(g_m x + g_{m-1}) f(x) - (f_n x + f_{n-1}) g(x)x^{n-m} &= 0 \\
\vdots
\end{align*}
\]

\[
\begin{align*}
(g_m x^{m-1} + g_{m-1} x^{m-2} + \cdots + g_1) f(x) - (f_n x^{m-1} + f_{n-1} x^{m-2} + \cdots + f_{n-m+1}) g(x) &= 0 \\
g(x) &= 0 \\
g(x)x &= 0 \\
\vdots
\end{align*}
\]

There are \( n \)-equations and \( n \) unknowns, the power products, \( 1, x, \ldots, x^{n-1} \).

In contrast to Euler’s construction in which the coefficients of the power products in equations are the coefficients of power products in \( f \) and \( g \), the coefficients in this construction are sums of \( (2 \times 2) \) determinants of the form \( (f_i g_j - f_j g_i) \). According to [3], Cauchy presented Bezout’s method in a somewhat different form from which the coefficient of each power product in the final system of \( n \) equations is quite apparent.
the maximum degree of these polynomials is $n$

Consider the following three generic bi-degree polynomials which have all the power products

$$f_n = \frac{f_n}{g_m} = \frac{f_{n-1}}{g_m} \frac{g_{n-1}}{g_m} = \frac{f_{n-1}x^{n-1} + f_{n-2}x^{n-2} + \cdots + f_1x + f_0}{(g_{n-1}x^{n-1} + g_{m-2}x^{m-2} + \cdots + g_1x + g_0)x^{n-m}},$$

$$f_{n+m} = \frac{f_{n+m}}{g_{m+1}} = \frac{f_{n+m-1}}{g_{m+1}} \frac{g_{n+m-1}}{g_{m+1}} = \frac{f_{n+m-1}x^{n+m-1} + f_{n+m-2}x^{n+m-2} + \cdots + f_1x + f_0}{(g_{m+1}x^{m+1} + g_{m-2}x^{m-2} + \cdots + g_1x + g_0)x^{n+m}}.$$

### 2.8 Cayley’s Method

Cayley reformulated Bezout’s method. A remarkable aspect of Cayley’s formulation is that it generalizes naturally to the multivariate case as discussed in the next subsection. Cayley proposed viewing the resultant of $f(x)$ and $g(x)$ as follows: Replace $x$ by $\alpha$ in both $f(x)$ and $g(x)$ and we get polynomials $f(\alpha)$ and $g(\alpha)$. The determinant $\Delta(x, \alpha)$ of the matrix

$$\begin{vmatrix} f(x) & g(x) \\ f(\alpha) & g(\alpha) \end{vmatrix}$$

is a polynomial in $x$ and $\alpha$, and is obviously equal to zero if $x = \alpha$. This implies that the determinant has $(x - \alpha)$ as a factor. The polynomial

$$\delta(x, \alpha) = \frac{\Delta(x, \alpha)}{(x - \alpha)}$$

is an $n - 1$ degree polynomial in $\alpha$ and is symmetric in $x$ and $\alpha$. It vanishes at every common zero $x_0$ of $f(x)$ and $g(x)$ no matter what values $\alpha$ has. So, at $x = x_0$, the coefficient of every power product of $\alpha$ in $\delta(x, \alpha)$ is 0. This gives $n$ equations which are polynomials in $x$, and the maximum degree of these polynomials is $n - 1$. Any common zero of $f(x)$ and $g(x)$ is a solution of these polynomial equations, and, these polynomial equations have a common solution if the determinant of their coefficients is 0.

Notice that there is some price to be paid by this formulation; the result computed using Cayley’s method has an additional extraneous factor of $f_n^{n-m}$. This factor arises because Cayley’s formulation is set up assuming both polynomials are of the same degree.

### 2.9 Dixon’s Formulation

Dixon showed how to extend this formulation to three polynomials in two variables. Consider the following three generic bi-degree polynomials which have all the power products of the type $x^i y^j$, where $0 \leq i \leq m, 0 \leq j \leq n$, i.e.,

$$f(x, y) = \sum_{i,j} a_{ij} x^i y^j, \quad g(x, y) = \sum_{i,j} b_{ij} x^i y^j, \quad h(x, y) = \sum_{i,j} c_{ij} x^i y^j.$$

Just as in the single variable case, Dixon observed that the determinant

$$\Delta(x, y, \alpha, \beta) = \begin{vmatrix} f(x, y) & g(x, y) & h(x, y) \\ f(\alpha, y) & g(\alpha, y) & h(\alpha, y) \\ f(\alpha, \beta) & g(\alpha, \beta) & h(\alpha, \beta) \end{vmatrix}$$

13
vanishes when $\alpha$ is substituted for $x$ or $\beta$ is substituted for $y$, implying that $(x-\alpha)(y-\beta)$ is a factor of the above determinant. The expression

$$\delta(x, y, \alpha, \beta) = \frac{\Delta(x, y, \alpha, \beta)}{(x - \alpha)(y - \beta)}$$

is a polynomial of degree $2m - 1$ in $\alpha$, $n - 1$ in $\beta$, $m - 1$ in $x$ and $2n - 1$ in $y$. Since the above determinant vanishes when we substitute $x = x_0, y = y_0$ where $(x_0, y_0)$ is a common zero of $f(x, y), g(x, y), h(x, y)$, into the above matrix, $\delta(x, y, \alpha, \beta)$ must vanish no matter what $\alpha$ and $\beta$ are. The coefficients of each power product $\alpha^j \beta^i, 0 \leq i \leq 2m - 1, 0 \leq j \leq n - 1$, have common zeros which include the common zeros of $f(x, y), g(x, y), h(x, y)$. This gives $2mn$ equations in power products of $x, y$, and the number of power products $x^i y^j, 0 \leq i \leq m - 1, 0 \leq j \leq 2n - 1$ is also $2mn$. The determinant of the coefficient matrix from these equations is a multiple of the resultant. Using a simple geometric argument, Dixon proved that in this case, the determinant is in fact the resultant up to a constant factor.

For three arbitrary polynomials, Dixon developed some special methods which selected some coefficients of $\alpha^j \beta^i$ from $\delta$, and used diatonic expansion of $f(x, y), g(x, y), h(x, y)$ to come up with a system of $k$ “independent” polynomial equations expressed using $k$ power products in $x, y$.

If polynomials $f(x, y), g(x, y), h(x, y)$ are not bidegree, it can be the case that the resulting Dixon matrix is not square or even if square, is singular. This phenomenon often happens when the number of variables to be eliminated increases. This issue is addressed in the next subsection where a generalization of Dixon’s formulation for eliminating arbitrarily many variables is discussed.

### 2.10 Generalizing Dixon’s Formulation

Cayley’s construction for two polynomials generalizes for eliminating $n - 1$ variables from a system of $n$ non-homogeneous polynomials $f_1, \ldots, f_n$. A matrix similar to the above can be set up by introducing new variables $\alpha_1, \ldots, \alpha_{n-1}$, and its determinant vanishes whenever $x_i = \alpha_i, 1 \leq i < n$ implying that $(x_1 - \alpha_1) \cdot \ldots \cdot (x_{n-1} - \alpha_{n-1})$ is a factor. The polynomial $\delta$, henceforth called the Dixon polynomial, can be expressed directly as the determinant of the matrix:

$$\delta(X, \Gamma) = \begin{vmatrix}
    f_{1,1} & \cdots & f_{1,n} \\
    f_{2,1} & \cdots & f_{2,n} \\
    \vdots & \vdots & \vdots \\
    f_{n-1,1} & \cdots & f_{n-1,n} \\
    f_1(\alpha_1, \alpha_2, \ldots, \alpha_{n-1}) & \cdots & f_n(\alpha_1, \alpha_2, \ldots, \alpha_{n-1})
\end{vmatrix},$$

where $\Gamma = \{\alpha_1, \alpha_2, \ldots, \alpha_{n-1}\}$, for $1 \leq j \leq n - 1$ and $1 \leq i \leq n$,

$$f_{j,i} = \frac{f_i(\alpha_1, \ldots, \alpha_{j-1}, x_j, x_{j+1}, \ldots, x_n) - f_i(\alpha_1, \ldots, \alpha_{j-1}, \alpha_j, x_{j+1}, \ldots, x_n)}{(x_j - \alpha_j)}$$

and $f_i(\alpha_1, \ldots, \alpha_k, x_{k+1}, \ldots, x_n)$ stands for uniformly replacing $x_j$ by $\alpha_j$ for $1 \leq j \leq k$ in $f_i$.

Let $\mathcal{F}$ be the set of all coefficients (which are polynomials in $X$) of terms in $\delta$, when viewed as a polynomial in $\Gamma$. Terms in $\Gamma$ can be used to identify polynomials in $\mathcal{F}$ based on
their coefficients. A matrix $D$ is constructed from $\mathcal{F}$ in which rows are labeled by terms in $\Gamma$ and columns are labeled by the terms in polynomials in $\mathcal{F}$. An entry $d_{i,j}$ is the coefficient of $j$-th term in the $i$-th polynomial, where polynomials in $\mathcal{F}$ and terms appearing in the polynomials can be ordered in any manner. We have called $D$ as the **Dixon matrix**. Dixon argued that for *generic ndegree* polynomials, $D$ is square - hence its determinant is a projection operator.

The Dixon polynomial for example 1 discussed in the subsection on sparse resultants is

$$
\delta(x, y, \alpha, \beta) = 
\begin{align*}
&\alpha(8ca^2b^2 - 4b^5 + 24cab^2 + 12a^2bx - 30ca^2b + 24a^2b - 24ab^2 - cb^4a - cb^7a^2 + 8a^3b^3 - 4a^4b + 4ab^4 - 6a^2c + 6\alpha^3c^2) \\
&-8a^4b + 6a^2c^2 + 2a^3c^2x - 2xa^2c + 12a^2y^2 - 8a^4yb + 2a^3cy + 2a^2cy^2 - 4a^2cy + 2a^2y^2 + 2a^2c^2y + 2\alpha^2cy + 2\beta^2a^2c
\\
&-8xa^2 + 2a^2y^2 + 4b^2y^2 + 12cab^2 - 12ca^2b - 6a^2c^2y - 8a^4yb^2 - 6\alpha^2c^2y + 12a^2b - 12ab^2 - 2c^2b^3
\\
&+ cb^3a^2 - c^2y^4 - cb^5 + 8a^4b^3 - 4a^3b^2 - 4a^4y + 4a^2c^2x + 6a^2c^2y - b^4y^2 - xac^4 + 6a^2y^2 + 6a^2z^2 c + 6a^2ce + xay^2b
\\
&- 6b^4x - 6a^2cx - 8a^3bx + 4b^3y^2).
\end{align*}
$$

The Dixon matrix is of dimension $2 \times 3$:

\[
\begin{pmatrix}
\begin{array}{ccc}
4ca^2b^2 - c^2b^4 + 12cab^2 - 12ca^2b \\
+ 12a^2b - 12ab^2 - c^2b^2a + cb^2a^2 \\
- 4ab^4 - c^2b^2 + 8a^2b^2
\end{array}
\end{pmatrix}
\begin{pmatrix}
\begin{array}{c}
1 \\
1 \\
1
\end{array}
\end{pmatrix}
\begin{pmatrix}
\begin{array}{c}
4a^3b^2 - 8a^4b - 6a^6c + 6a^3c^2 \\
+ 6a^5c^2 - 6ca^2b - cb^3a^2 + cb^3a \\
- 6abc^2b - 6cab^2 + 4a^2b^3 - 8a^3b^2
\end{array}
\end{pmatrix}
\begin{pmatrix}
\begin{array}{c}
\begin{cases}
\alpha \\
\beta
\end{cases} \\
x \\
y
\end{array}
\end{pmatrix}
\]

Since the Dixon matrix is rectangular, its determinant cannot be computed. There are examples in which the Dixon matrix is square, but it could be singular. The Dixon matrix being rectangular is essentially the same as a singular square matrix; additional zero rows (or columns) can be appended to a rectangular matrix to make it square.

As we saw earlier, this phenomenon of resultant matrices being singular is observed in all the resultant methods when more than one variable is eliminated. In the next section, we discuss a linear algebra construction, **rank submatrix**, which we have found very useful for extracting projection operators from singular resultant matrices. As shown in the table in the subsection on empirical comparison, this construction seems to outperform other techniques based on perturbation, for example, the generalized characteristic polynomial construction proposed by Canny for singular Macaulay matrices.

### 2.11 Rank Submatrix (RSC)

If a resultant matrix is singular (or rectangular), a projection operator can be extracted by computing the determinants of its submatrices. This method is applicable to Macaulay matrices, sparse matrices as well as Dixon matrices. The general construction is as follows:

1. Set up the Resultant matrix ($R$) of $\mathcal{F}$.
2. If there is any column in $R$ which is linearly independent of the other columns, then return the determinant of any *rank submatrix* of $R$.
3. Else, this heuristic fails.
By rank submatrix we mean a maximal nonsingular submatrix of $R$, and its determinant may be computed by performing Gaussian elimination and returning the product of the (non-zero) pivots [20]. Though there exist examples where the condition in step (2) is not true, they seem to be rare. In all the examples we tried, $R$ was degenerate many times, but the condition of step (2) was satisfied. It was shown in [20, 18] that

**Theorem 1** If Dixon matrix has a column which is linearly independent of others then determinant of any rank submatrix is a non-trivial multiple of resultant.

For example 1, this construction can be used to compute a projection operator. From the rectangular Dixon matrix $D$ shown above, it can be checked that the first column of $D$ is linearly independent of the other columns and hence the rank of the matrix (which is 2) decreases (becomes 1) on its removal. We do Gaussian elimination on $D$ and the projection operator is the product of the 2 pivots. The projection operator in this example is found to be the following polynomial with 75 terms (in expanded form):

$$q = -b \rho (a, b, c).$$

Again, besides the resultant, there is one extraneous factor.

### 2.11.1 Implicitization

We illustrate the use of Dixon resultants for generating an implicit representation of a torus from its explicit representation. The constraints defining the torus are:

1. $x = r \cos(u) \cos(t) + R \cos(t),$
2. $y = r \cos(u) \sin(t) + R \sin(t),$
3. $z = r \sin(u).$

The parameters $r$ and $R$ are the inner and outer radii, respectively. The two trigonometric identities relating $\sin$ and $\cos$ are:

4. $\cos(u)^2 + \sin(u)^2 = 1,$
5. $\cos(t)^2 + \sin(t)^2 = 1.$

The objective is to compute an implicit representation of the torus in terms of $x, y, z, r, R$, the outer radius. This can be achieved by successively eliminating $\cos(t), \sin(t), \sin(u)$ and $\cos(u)$ from the above equations.

The Dixon polynomial for this example is given below. Symbols $c_u, c_t, s_u, s_t$ are used instead of $\cos(u), \cos(t), \sin(u), \sin(t)$, respectively, and $\bar{c}_u, \bar{c}_t, \bar{s}_u, \bar{s}_t$ are the new variables substituted for $c_u, c_t, s_u, s_t$, respectively, for formulating the Dixon resultant.

$$
\begin{align*}
&\left( r^3 c_t^2 - r^3 + r^3 s_t^2 \right) \bar{c}_u^3 \\
&\quad + \left( -r^3 c_u c_t + x r^2 - R c_t r^2 \right) \bar{c}_u^2 \bar{c}_t \\
&\quad + \left( -r^3 c_u s_t + y r^2 - R s_t r^2 \right) \bar{c}_u \bar{s}_t \\
&\quad + \left( -r^3 c_t s_u + r^2 c_t z \right) \bar{s}_u \bar{s}_t \bar{c}_t \\
&\quad + \left( -r^3 s_t s_u + r^2 s_t z \right) \bar{s}_u \bar{s}_t \bar{s}_t
\end{align*}
$$
+ (x r^2 c_t + r^2 c_t^2 R + r^2 y s_t - 2 r^2 R - r^3 c_u + R r^2 s_t^2) c_u^2 \\
+ (-r^3 c_t^2 s_u + r^2 s_t^2 z - r^2 c_t^2 z - s^3 s_t^2 s_u) c_u s_u \\
+ (r x R + r^2 c_t z s_u + x^2 c_u - r R^2 c_t - r^3 c_t - 2 r^2 c_t c_t R) c_u c_t \\
+ (-r^3 s_t + r R^2 s_t - 2 r^2 R s_t c_u + R r y c_u + r^2 s_t z s_u) c_u s_t \\
+ (r c_t z - r^2 c_t R s_u) s_u c_t \\
+ (R r s_t z - R^2 s_t s_u) s_u s_t \\
+ (-r^3 s_t^2 + r^2 c_t^2 z s_u + x R c_t + R r y s_t - 2 r^2 R c_u - r^3 c_t^2 + x^2 c_u c_t + r^2 y c_u s_t - r R^2 + r^2 s_t^2 z s_u) c_u \\
+ (R r s_t z + r c_t^2 R z - r^2 c_t^2 R s_u - R R^2 s_t^2 s_u) s_u \\
+ (r c_t R z s_u + r x R c_u - R c_t r^2 - r R^2 c_t c_u) c_t \\
+ (-r^3 R s_t + R r y c_u + R r s_t z s_u - r R^2 s_t c_u) s_t \\
+ (R r s_t^2 s_u + R r y c_u s_t - R^2 s_t^2 + r^2 c_t^2 R z s_u + r x R c_u c_t - r^2 c_t^2 R - R^2 c_u) \\

The (16 × 12) Dixon matrix is:

\[
\begin{pmatrix}
1 & c_u & c_t & s_t & c_t^2 & s_t^2 & c_u c_t & s_u c_t & s_t c_u & s_u s_t & s_u^2 & s_u s_t^2
\end{pmatrix}
\]

<table>
<thead>
<tr>
<th>1</th>
<th>c_u</th>
<th>c_t</th>
<th>s_t</th>
<th>c_t^2</th>
<th>s_t^2</th>
<th>c_u c_t</th>
<th>s_u c_t</th>
<th>s_t c_u</th>
<th>s_u s_t</th>
<th>s_u^2</th>
<th>s_u s_t^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-r R^2</td>
<td>0</td>
<td>0</td>
<td>-r^2 R</td>
<td>-r^2 R</td>
<td>r x R</td>
<td>R R^2</td>
<td>r R</td>
<td>R R^2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>R R^2</td>
<td>-r^2 R</td>
<td>x^2 R</td>
<td>-r^2 R</td>
<td>x^2 R</td>
<td>r R</td>
<td>R R^2</td>
<td>-r^2 R</td>
<td>r R</td>
<td>-r R</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Its rank is 10, and the projection operator is

\[ r^{23} R (-z^2 - R^2 + r^2)(4 R^2 (r^2 - z^2) - (x^2 + y^2 + z^2 - R^2 - r^2)^2) \]

and the factor \[ 4 R^2 (r^2 - z^2) - (x^2 + y^2 + z^2 - R^2 - r^2)^2 \] is the implicit representation of torus. In contrast, the best sparse matrix is of dimension 27 \times 27.

2.12 Empirical Comparison

The following table gives some empirical data comparing different methods on a suite of examples taken from different application domains as well as including some randomly generated examples. Macaulay/GCP stands for Macaulay’s method augmented with generalized characteristic polynomial construction for handling singular matrices. Macaulay/RSC, Dixon/RSC and Sparse/RSC, respectively, stand for, Macaulay’s method, Dixon’s method and sparse resultant method augmented with rank submatrix computation for handling singular matrices. Emiris’s C implementation was used for computing sparse resultant matrices. For Macaulay’s method and Dixon’s method, matrices were set up on Maple. For RSC method, the linear independence of any column in these matrices was checked probabilistically by substituting random numbers for parameters. Zippel’s sparse interpolation algorithm implemented in C++ was used to interpolate the rank subdeterminant. Further details are given in our ISSAC-95 paper [18].
Table 1: Empirical Data

<table>
<thead>
<tr>
<th>Ex</th>
<th>Application</th>
<th>No. of Polys</th>
<th>Dixon</th>
<th>Macaulay</th>
<th>Sparse</th>
<th>Gröbner</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Matrix</td>
<td>Constr.</td>
<td>RSC</td>
<td>Matrix</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Size</td>
<td>Time</td>
<td>Time</td>
<td>Size</td>
</tr>
<tr>
<td>1</td>
<td>Geom. Reason</td>
<td>3</td>
<td>13</td>
<td>0.18</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Formula Deriv</td>
<td>5</td>
<td>14</td>
<td>0.36</td>
<td>76</td>
<td>330</td>
</tr>
<tr>
<td>3</td>
<td>Formula Deriv</td>
<td>3</td>
<td>5</td>
<td>0.08</td>
<td>*</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>Geometry</td>
<td>3</td>
<td>4</td>
<td>0.08</td>
<td>102</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>Random</td>
<td>4</td>
<td>6</td>
<td>0.06</td>
<td>N/R</td>
<td>84</td>
</tr>
<tr>
<td>6</td>
<td>Implicitization</td>
<td>3</td>
<td>10</td>
<td>0.4</td>
<td>88.5</td>
<td>45</td>
</tr>
<tr>
<td>7</td>
<td>Implicitization</td>
<td>3</td>
<td>18</td>
<td>9.33</td>
<td>128</td>
<td>153</td>
</tr>
<tr>
<td>8</td>
<td>Implicitization</td>
<td>3</td>
<td>18</td>
<td>9.33</td>
<td>128</td>
<td>153</td>
</tr>
<tr>
<td>9</td>
<td>Random</td>
<td>5</td>
<td>36</td>
<td>1.1</td>
<td>8130</td>
<td>1365</td>
</tr>
<tr>
<td>10</td>
<td>Implicitization</td>
<td>3</td>
<td>11</td>
<td>0.06</td>
<td>3.01</td>
<td>36</td>
</tr>
<tr>
<td>11</td>
<td>Implicitization</td>
<td>3</td>
<td>14</td>
<td>0.45</td>
<td>47</td>
<td>66</td>
</tr>
<tr>
<td>12</td>
<td>Equilibrium</td>
<td>4</td>
<td>5</td>
<td>0.117</td>
<td>220</td>
<td>56</td>
</tr>
<tr>
<td>13</td>
<td>Vision</td>
<td>6</td>
<td>6</td>
<td>0.06</td>
<td>0.27</td>
<td>28</td>
</tr>
</tbody>
</table>

Fourteen examples used in Table 1 can be obtained in MAPLE format by anonymous ftp from ftp://ftp.cs.albany.edu/pub/saxena/examples.14. All timings are in seconds on a 64Mb SUN Sparc10. A (*) in a Time column either means that the resultant cannot be computed even after running for more than a day or the program ran out of memory. An N/R in the GCP column means there exists a polynomial ordering for which the Macaulay and the denominator matrix are nonsingular and therefore GCP computation is not required. We have also included timings for computing resultant using Gröbner basis construction with block ordering (variables in one block and parameters in another) using the Macaulay system. Gröbner basis computations are done in a finite field with a much smaller characteristic (31991) than in the resultant computations. Examples 3, 4 and 5 consist of generic polynomials with numerous parameters and dense resultants. Interpolation methods are not appropriate for such examples and timings using straightforward determinant expansion in MAPLE are in Table 2.

Table 2: Timings - direct determinant expansion in MAPLE.

<table>
<thead>
<tr>
<th>Ex</th>
<th>Dixon/ RSC</th>
<th>*</th>
<th>Mac/ GCP</th>
<th>Mac/ RSC</th>
<th>Sparse/ RSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>6.56</td>
<td>N/R</td>
<td>26.5</td>
<td>76.2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.6</td>
<td>*</td>
<td>84</td>
<td>290</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.9</td>
<td>N/R</td>
<td>850</td>
<td>23.4</td>
<td></td>
</tr>
</tbody>
</table>

As can be seen from the above tables, all the examples were completed using the Dixon/RSC method. Sparse/RSC also solved all the examples, but always took much longer than Dixon/RSC. Macaulay/RSC could not complete two problems and took much longer than Sparse/RSC (for most problems) and Dixon/RSC (for all problems).

3See Table 2 for timing using direct determinant expansion.
2.13 Theoretical Comparison

The following theoretical results about Dixon matrix suggest that the Dixon formulation is able to exploit the sparsity of the input polynomial system [19].

A system \( \mathcal{F} \) of polynomials is called unmixed if every polynomial in \( \mathcal{F} \) has the same Newton polytope. Let \( \mathcal{N}(P) \) denote the \textbf{Newton Polytope} of any polynomial in \( \mathcal{F} \). Let \( \pi_i(A) \) be the \textbf{Projection} of an \( n \)-dimensional set, \( A \), to \( n - i \) dimension, obtained by substituting 0 for all the first \( i \) dimensions. Let \( \text{mvol}(\mathcal{F}) \) be the \textbf{\( n \)-fold Mixed Volume} of \( \mathcal{F} \) (= \( n! \text{vol}(\mathcal{F}) \) in unmixed case).

**Theorem 2** The number of columns in the Dixon matrix of an unmixed set \( \mathcal{F} \) of \( n + 1 \) polynomials in \( n \) variables is

\[
O \left( \text{vol} \left( \sum_{i=0}^{n} \pi_i(\mathcal{N}(P)) \right) \right).
\]

To study the relationship between the size of the Dixon matrix and to the volume of \( \mathcal{F} \), the class of multi-homogeneous polynomials is considered.

An multi-homogeneous system of polynomials of type \( (1, \ldots, l_r; d_1, \ldots, d_r) \) is defined to be an unmixed set of \( \sum_{i=1}^{r} l_i + 1 \) polynomials in \( \sum_{i=1}^{r} l_i \) variables, where

1. \( r \) = number of partitions of variables.
2. \( l_i \) = number of variables in \( i^{th} \) partition.
3. \( d_i \) = total degree of each polynomial in the variables of the \( i^{th} \) partition.

The size of the Dixon matrix for such a system can be derived to be:

\[
O \left( \prod_{i=1}^{r} \frac{\left( \sum_{j=1}^{l_i} l_j + 1 \right)^{l_i}}{n + 1} \prod_{i=1}^{r} \frac{d_i^{l_i}}{l_i!} \right).
\]

For asymptotically large number of partitions, the size of the Dixon matrix is

\[
O \left( \frac{\text{mvol}(P)}{n + 1} \right).
\]

It can be seen that in the multi-homogeneous case, the size of the Dixon matrix, for asymptotically large number of partitions, varies at a slightly smaller rate than the mixed volume of the system.

The size of the Dixon matrix is much smaller than the size of Macaulay matrix. It also turns out to be much smaller than the size of the sparse resultant matrix which is bounded by the number of integral points inside the Minkowski sum of the Newton polytopes (not their successive projections) of the input polynomials. This is analogous to the Dixon matrix whose size is bounded by the number of integral points inside the Minkowski sum of the \textit{successive projections} of Newton polytopes of input polynomials. However, since the projections are of successively lower dimension than the polytopes themselves, the Dixon matrix is smaller. Specifically, the size of the Dixon matrix of multi-homogeneous polynomials is of smaller order than their \( n \)-fold mixed volume whereas the size of the sparse
resultant matrix is larger than the $n$-fold mixed volume by an exponential multiplicative factor. If the $n$-fold mixed volume of a multi-homogeneous polynomial system is $M$ and the number of variables being eliminated is $n$ then for asymptotically large number of partitions of variables (approaching $n$), size of its Dixon matrix is $O\left(\frac{1}{n+1} M\right)$, whereas the size of its sparse resultant matrix is $O\left(e^{n+1} M\right)$.
3 Characteristic Set Construction

In this section, we briefly discuss Ritt’s characteristic set construction. The discussion is based on Ritt’s presentation in his book *Differential Algebra* [25] (chapter 4) and Wu’s exposition of Ritt’s approach as discussed in [31]. We first give an informal explanation of Ritt’s characteristic set construction, outline how it can be used for elimination and then give the technical details. This section includes some new material comparing Ritt’s definitions with Wu’s definitions, as well as a new way to decompose a zero set of a polynomial system.

Given a system $S$ of polynomials, the characteristic set algorithm transforms $S$ into a triangular form $S'$ so that the zero set of $S$ is “roughly equivalent” to the zero set of $S'$. A total ordering on variables is assumed. Unlike in Gröbner basis computations, polynomials are treated as univariate polynomials in their highest variable. The primitive operation used in the transformation is that of *pseudo-division* of a polynomial by another polynomial. The algorithm is similar in flavor to Gaussian elimination for linear equations. It proceeds by considering polynomials in the lowest variables. For each variable $x_i$, there may be several polynomials of different degrees with $x_i$ as the highest variable. Multivariate gcd like computations are performed first to obtain the lowest degree polynomial, say $h_i$, in $x_i$. If $h_i$ is linear in $x_i$, then it can be used to eliminate $x_i$ from the rest of the polynomials. Otherwise, polynomials in which the degree of $x_i$ is higher, are pseudo-divided by $h_i$ to give polynomials that are of lower degrees in $x_i$. This process is repeated until for each variable, a minimal degree polynomial is generated such that every polynomial generated thus far pseudo-divides to 0.

If the number of equations in $S$ is less than the number of variables (which is typically the case when elimination or projection needs to be computed), then the variable set $\{x_1, \ldots, x_n\}$ is typically classified into two subsets: independent variables (also called parameters) and dependent variables, and a total ordering on the variables is so chosen that the dependent variables are all higher in the ordering than the independent variables. For elimination, the variables to be eliminated must be classified as dependent variables. We denote the independent variables by $u_1, \cdots, u_k$ and dependent variables by $y_1, \cdots, y_l$, and the total ordering is $u_1 \prec \cdots \prec u_k \prec y_1 \prec \cdots \prec y_l$, where $k + l = n$.

To check whether another equation, say $f = 0$, follows from $S$, i.e., $f$ vanishes on (most) of the common zeros of $S$, $f$ is pseudo-divided using the polynomials in a triangular form $S'$ of $S$. If the remainder of pseudo-division is 0, then $f = 0$ is said to follow from $S$ under the condition that the initials of polynomials in $S'$ are not zero. This use of characteristic set construction is extensively discussed in [30, 31, 8] for geometry theorem proving. In [32], Wu also discussed how the characteristic set method can be used to study the zero set of polynomial equations. These uses of characteristic set method were also discussed in detail in our introductory paper [16].

For elimination, this construction can be employed with a goal to generate a polynomial free of variables to be eliminated.

3.1 Preliminaries

Given a polynomial $p$, the highest variable of $p$ is $y_i$ if $p \in \mathbb{Q}[u_1, \ldots, u_k, y_1, \ldots, y_i]$ and $p \notin \mathbb{Q}[u_1, \ldots, u_k, y_1, \ldots, y_{i-1}]$ (there is a similar definition for $u_i$). The class of $p$ is then called $i$.\footnote{It is possible to devise an algorithm which considers the variable in descending order.}
A polynomial $p$ is $\geq$ another polynomial $q$ if and only if

1. the highest variable of $p$, say $y_i$, is $\succ$ the highest variable of $q$, say $y_j$, i.e., the class of $p$ is higher than the class of $q$, or
2. the class of $p = \text{the class of } q$, and the degree of $p$ in $y_i$ is $\geq$ the degree of $q$ in $y_i$.

If $p \geq q$ and $q \geq p$, then $p$ and $q$ are said to be equivalent, written as $p \equiv q$; this means that $p$ and $q$ are of the same class $i$, and the degree of $y_i$ in $p$ is the same as the degree of $y_i$ in $q$.

A polynomial $p$ is reduced with respect to another polynomial $q$ if (a) the highest variable, say $y_i$, of $p$ is $\prec$ the highest variable of $q$, say $y_j$ (i.e., $p \prec q$), or (b) $y_i \succ y_j$ and the degree of $y_j$ in $q$ is $\succ$ the degree of $y_j$ in $p$.

A sequence $C$ of polynomials, $(p_1, \ldots, p_m)$ is called a chain if either (i) $m = 1$ and $p_1 \neq 0$, or (ii) $m > 1$ and the class of $p_1$ is $> 0$, and for $j > i$, $p_j$ is of higher class than $p_i$ and reduced with respect to $p_i$; we thus have $p_1 \prec p_2 \prec \cdots \prec p_m$.

Given two chains $C = (p_1, \ldots, p_m)$ and $C' = (p_1', \ldots, p_m')$, $C \succ C'$ if (i) there is a $j \leq m$ as well as $j \leq m'$ such that $p_i \equiv p_i'$ for all $i < j$ and $p_j \succ p_j'$, or (ii) $m' > m$ and for $i \leq m$, $p_i \equiv p_i'$.

A set $G = \{g_1, \ldots, g_m\}$ of polynomials is said to be in triangular form if and only if $g_1, g_2, \ldots, g_m$ are respectively, polynomials in $\{u_1, \ldots, u_k, y_1\}$, $\{u_1, \ldots, u_k, y_2\}$, $\ldots$, $\{u_1, \ldots, u_k, y_1, \ldots, y_m\}$. If $m = l$, then $G$ is said to be fully triangular. It is easy to see that every chain is in triangular form.

### 3.1.1 Pseudo-Division

Given two multivariate polynomials $p$ and $q$, viewed as polynomials in the main variable $x$, with coefficients that are polynomials in the other variables: $p$ can be pseudo-divided by $q$ if the degree of $q$ is less than or equal to the degree of $p$. Let $I_q$ denote the initial of $q$ and $e = \text{degree}(p) - \text{degree}(q) + 1$. Then

$$I_q^e p = s \ q + r,$$

where $s$ and $r$ are polynomials in $x$ with coefficients that are polynomials in the other variables and the degree of $r$ in $x$ is lower than the degree of $q$. The polynomial $r$ is called the remainder (or pseudo-remainder) of $p$ obtained by dividing by $q$. It is easy to see that the common zeros of $p$ and $q$ are also zeros of the remainder $r$, and that $r$ is in the ideal of $p$ and $q$.

For example, if $p = xy^2 + y + (x + 1)$ and $q = (x^2 + 1)y + (x + 1)$, then $p$ cannot be divided by $q$ but can be pseudo-divided as follows:

$$(x^2 + 1)^2(xy^2 + y + x + 1) = ((x^3 + x)y + (1-x))(x^2 + 1)y + (x + 1) + x^5 + x^4 + 2x^3 + 3x^2 + x,$$

with the polynomial $(x^5 + x^4 + 2x^3 + 3x^2 + x)$ as the pseudo-remainder.

If $p$ is not reduced with respect to $q$, then $p$ reduces to $r$ using $q$ by pseudo-dividing $p$ by $q$ giving $r$ as the remainder of the result of pseudo-division. A polynomial $p$ reduces to $p'$ with respect to a chain $C = (p_1, \ldots, p_m)$ if there exist nonnegative integers $i_1, \cdots, i_m$ such that

$$I_1^{i_1} \cdots I_m^{i_m} p = q_1 p_1 + \cdots + q_m p_m + p',$$

\footnote{A chain is the same as an ascending set defined by Wu [31].}
where \( p' \) is reduced with respect to each of \( p_i, 1 \leq i \leq m \). Typically pseudo-division is successively done using \( p_i \)'s starting with \( p_m \), the polynomial in the highest variable.

3.2 Characteristic Sets a la Ritt

A characteristic set for a set \( \Sigma \) of polynomials was defined by Ritt (p. 5 in [25]) to be the lowest chain in \( \Sigma \). After introducing the definition, Ritt remarked that a necessary and sufficient condition for a chain \( C \) in a set \( \Sigma \) of polynomials to be a characteristic set of \( \Sigma \) is that there is no non-zero polynomial \( p \) in \( \Sigma \) that is reduced with respect to \( C \), or equivalently, every polynomial \( p \) in \( \Sigma \) can be reduced using \( C \).

**Theorem 3 (Ritt)** A chain \( C \) in a set \( \Sigma \) of polynomials is a characteristic set of \( \Sigma \) if and only if every non-zero polynomial in \( \Sigma \) can be reduced by \( C \).

For a polynomial ideal \( \Sigma \), Ritt gave a proof of existence for characteristic sets. Given a total ordering on variables, \( C \) is a finite maximal subset of \( \Sigma \) consisting of minimal reduced polynomials in triangular form. In other words, pick a minimal polynomial \( c_1 \) in \( \Sigma \) and include it in \( C \). Delete from \( \Sigma \) all polynomials which can be reduced by \( c_1 \); from the remaining subset of \( \Sigma \), called \( \Sigma_1 \), pick a minimal polynomial \( c_2 \). The class of \( c_2 \) is necessarily higher than the class of \( c_1 \). Delete from \( \Sigma_1 \) all polynomials which can be reduced using \( c_1 \) and \( c_2 \). From the remaining subset \( \Sigma_2 \), pick a minimal polynomial \( c_3 \), and so on. This process terminates since there are only finitely many indeterminates and a chain cannot have more than \( n \) elements.

Ritt was apparently interested in associating characteristic sets only with prime ideals. A prime ideal is an ideal with the property that if an element \( h \) of the ideal can be factored as \( h = h_1h_2 \), then, either \( h_1 \) or \( h_2 \) must be in the ideal. For a prime ideal, its characteristic set \( C \) has the nice property that a polynomial pseudo-divides to 0 using \( C \) if and only if the polynomial is in the ideal, a property much like of a Gröbner basis except that in case of an ideal, a polynomial reduces by its Gröbner basis to 0 if and only if the polynomial is in the ideal. For arbitrary ideals, the following holds:

**Theorem 4** Every polynomial in an ideal \( I \) pseudo-divides to 0 using a characteristic set \( C \) of \( I \).

As said earlier, the converse of the above theorem holds only for prime ideals.

In chapter 4 of his book in the section *Components of Finite Systems* in [25] (pp. 95-96), Ritt gave a method for computing a characteristic set from a finite basis \( \Sigma \) of an ideal. A characteristic set \( \Phi \) is computed from \( \Sigma \) by successively adjoining \( \Sigma \) with remainder polynomials obtained by pseudo-division. Starting with \( \Sigma_0 = \Sigma \), we extract a minimal chain from \( \Sigma_i \), and compute non-zero remainders of polynomials in \( \Sigma_i \) with respect to the minimal chain. If this remainder set is nonempty, we adjoin it to \( \Sigma_i \) to obtain \( \Sigma_{i+1} \) and repeat the computation until we have \( \Sigma_j \) such that every polynomial in \( \Sigma_j \) pseudo-divides to 0 with respect to its minimal chain. The set \( \Sigma_j \) is called a saturation of \( \Sigma \), and that minimal chain of \( \Sigma_j \) is a characteristic set of \( \Sigma_j \) as well as \( \Sigma \). The above construction terminates since the minimal chain of \( \Sigma_i \) is \( \succ \) the minimal chain of \( \Sigma_{i+1} \) and the ordering on chains is well-founded.

The above algorithm can be viewed as augmenting \( \Sigma \) with additional polynomials from the ideal generated by \( \Sigma \) until a set \( \Delta \) is generated such that...
1. \( \Sigma \subseteq \Delta \),
2. \( \Sigma \) and \( \Delta \) generate the same ideal, and
3. a minimal chain \( \Phi \) of \( \Delta \) pseudo-divides every polynomial in \( \Delta \) to 0.

There can be many ways to compute such a \( \Delta \). For detailed description of some of the algorithms, the reader may consult [25, 30, 31, 8, 16].

In order to use characteristic sets for elimination, variables being eliminated are made greater than other variables in the ordering, and a characteristic set is computed. The least polynomial in a characteristic set can be used as a projection operator.

In his Ph.D. thesis, Kandri-Rody gave a method for extracting a characteristic set from a Gröbner basis of a polynomial ideal. Computing a characteristic set of a polynomial ideal directly, without having to compute its Gröbner basis, is, however, not as easy as one would think. An interested reader may consult [15] for a discussion of this topic.

### 3.3 Characteristic Sets a la Wu

Wu [31] altered Ritt’s notation somewhat to make it more useful. Wu associated a characteristic set with the zero set of an arbitrary set of polynomials, instead of the ideal generated by the polynomials. He called a characteristic set associated with a prime ideal (equivalently, an irreducible zero set) as irreducible. A zero set is irreducible if it cannot be expressed as a union of proper algebraic subsets.\(^6\)

A characteristic set \( \{g_1, \cdots, g_l\} \) as defined by Wu could be irreducible or reducible, whereas a characteristic set defined by Ritt is always irreducible. Below, we reproduce the definition of a characteristic set due to Wu [31].

**Definition** Given a finite set \( \Sigma \) of polynomials in \( u_1, \ldots, u_k, y_1, \ldots, y_l \), a characteristic set \( \Phi \) of \( \Sigma \) is defined to be either

1. \( \{p_1\} \), where \( p_1 \) is a polynomial in \( u_1, \ldots, u_k \) or
2. a chain \( \langle p_1, \ldots, p_l \rangle \), where \( p_1 \) is a polynomial in \( y_1, u_1, \ldots, u_k \) with initial \( I_1 \), \( p_2 \) is a polynomial in \( y_2, y_1, u_1, \ldots, u_k \), with initial \( I_2 \), \ldots, \( p_l \) is a polynomial in \( y_l, \ldots, y_1, u_1, \ldots, u_k \), with initial \( I_l \), such that
   
   (a) any zero of \( \Sigma \) is a zero of \( \Phi \), and
   
   (b) any zero of \( \Phi \) that is not a zero of any of the initials \( I_i \), is a zero of \( \Sigma \).

We thus have \( \text{Zero}(\Sigma) \subseteq \text{Zero}(\Phi) \) as well as \( \text{Zero}(\Phi/\prod_{i=1}^l I_i) \subseteq \text{Zero}(\Sigma) \), where using Wu’s notation, \( \text{Zero}(\Phi/I) \) stands for \( \text{Zero}(\Phi) - \text{Zero}(I) \). We also have

\[
\text{Zero}(\Sigma) = \text{Zero}(\Phi/\prod_{i=1}^l I_i) \cup \bigcup_{i} \text{Zero}(\Sigma \cup \{I_i\}).
\]

**Definition** A characteristic set \( \Phi = \{p_1, \cdots, p_l\} \) is irreducible over \( \mathbb{Q}[u_1, \ldots, u_k, y_1, \ldots, y_l] \) if for \( i = 1 \) to \( l \), \( p_i \) is irreducible (i.e., cannot be factored) over \( \mathbb{Q}_{i-1} \) where \( \mathbb{Q}_0 = \mathbb{Q}(u_1, \cdots, u_k) \)

\(^6\)Recall that a zero set is algebraic if it is the zero set of a set of a polynomials.
and $Q_j = Q_{j-1}(\alpha_j)$ is an algebraic extension of $Q_{j-1}$, obtained by adjoining a root $\alpha_j$ of $p_j = 0$ to $Q_{j-1}$, i.e., $p_j(\alpha_j) = 0$ in $Q_j$ for $1 \leq j < i$.

If a characteristic set $\Phi$ of $\Sigma$ is irreducible, then $\text{Zero}(\Sigma) = \text{Zero}(\Phi)$ since the initials of $\Phi$ do not have a common zero with $\Phi$.

Different orderings on dependent variables can not only result in different characteristic sets but one ordering may generate an reducible characteristic set whereas another one might give an irreducible characteristic set. For example, consider $\Sigma = \{(x^2 - 2x + 1) = 0, (x - 1)z - 1 = 0\}$. Under the ordering $x < z$, $\Sigma$ is a characteristic set. The two polynomials do not have a common zero, and under the ordering $z < x$, the characteristic set of $\Sigma$ includes $1$.

In the process of computing a characteristic set from $\Sigma$, if an element of the coefficient field (a rational number if $l = n$) is generated as a remainder, this implies that $\Sigma$ does not have a solution, or is inconsistent.

If $\Sigma$ does not have a solution, either a characteristic set $\Phi$ of $\Sigma$ includes a constant (in general, an element $Q(u_1, \cdots, u_k)$), or $\Phi$ is reducible and each of the irreducible characteristic sets includes constants (respectively, elements of $Q(u_1, \cdots, u_k)$).

**Theorem 5** Given a finite set $\Sigma$ of polynomials, if (i) its characteristic set $\Phi$ is irreducible, (ii) $\Phi$ does not include a constant, and (iii) the initials of the polynomials in $\Phi$ do not have a common zero with $\Phi$, then $\Sigma$ has a common zero.

### 3.4 Elimination using Characteristic Sets: An example of implicitization

We illustrate the use of characteristic set construction for elimination. We use the example discussed in subsection 2.11.1 for computing an implicit representation of a torus from its explicit representation in terms of $x, y, z, r, R$.

Characteristic set construction can be used for elimination. Just like resultant methods, the result may have extraneous factors which are typically initials of polynomials generated during the computation of a characteristic set. For example, a characteristic set can be computed using the ordering $\cos(u) < \sin(u) < \sin(t) < \cos(t)$, and the last variable can be any of $r, R, x, y, z$ as we are interested in a polynomial in these variables.

The variable $\sin(t)$ can be eliminated by pseudo-dividing the polynomial from constraint 5 by the polynomial from constraint 1, which gives as remainder:

6. $r^2 \cos(u)^2 \sin(t)^2 + 2 R r \cos(u) \sin(t)^2 + R^2 \sin(t)^2 - r^2 \cos(u)^2 - 2 R r \cos(u) - R^2 + x^2$.

The variable $\sin(t)$ is eliminated by pseudo-dividing polynomial 6 by the polynomial from constraint 2. The remainder is:

7. $-r^4 \cos(u)^4 - 4 R r^3 \cos(u)^3 + (\cos(u)^3 + r^2 x^2 + r^2 y^2 - 6 R^2 r^2) \cos(u)^2 + (2 R r x^2 + 2 R r y^2 - 4 R^3 r) \cos(u) + (R^2 x^2 + R^2 y^2 - R^4)$.

Similarly, $\sin(u)$ can be easily eliminated from the polynomial in constraint 4 by the polynomial in constraint 3; the remainder is a polynomial with 3 terms. Then $\cos(u)$ can be eliminated using the result of the last step from polynomial 7 above. The result of these eliminations are two polynomials in $x, y, z, r, R$. In order to get the result, this polynomial has to be pseudo-divided using the other polynomial. The result given by the characteristic
set computation for the torus has an extraneous factor. After removing this factor, the implicit representation of a torus is:

\[ 4R^2(r^2 - z^2) - (x^2 + y^2 + z^2 - R^2 - r^2)^2 = 0. \]

If a different ordering is used on the dependent variables, the computation proceeds much faster without generating very large intermediate polynomials. The ordering \( \sin(t) \prec \cos(t) \prec \cos(u) \prec \sin(u) \) particularly works well.

### 3.5 Proving Conjectures from a System of Equations

A direct way to check whether an equation \( c = 0 \) follows (under certain conditions) from a system \( S \) of equations is to compute a characteristic set \( \Phi = \{ p_1, \ldots, p_l \} \) from \( S \) and check whether \( c \) pseudo-divides to 0 with respect to \( \Phi \). If \( c \) has a zero remainder with respect to \( \Phi \), then the equation \( c = 0 \) follows from \( \Phi \) under the conditions that none of the initials used to multiply \( c \) is 0. The algebraic relation between the conjecture \( c \) and the polynomials in \( \Phi \) can be expressed as:

\[
I_1^{i_1} \cdots I_l^{i_l} c = q_1p_1 + \cdots + q_lp_l,
\]

where \( I_j \) is the initial of \( p_j \), \( j = 1, \ldots, l \).

This approach is used by Wu and Chou for geometry theorem proving. A characteristic set is is computed from the hypotheses of a geometry problem. For plane Euclidean geometry, most hypotheses can be formulated as linear polynomials in dependent variables, so a characteristic set can be computed efficiently. A conjecture is pseudo-divided by the characteristic set to check whether the remainder is 0. If the remainder is 0, the conjecture can be declared to be not generically valid. Otherwise, the zero set of the hypotheses must be decomposed and then the conjecture is checked for validity on each of the components or some of the components.

A refutational way to check whether \( c = 0 \) follows from \( S \) is to compute a characteristic set of \( S \cup \{ cz - 1 = 0 \} \), where \( z \) is a new variable.

### 3.6 Decomposing a zero set into irreducible zero sets

To deal with a reducible characteristic set, Ritt and Wu advocated the use of factorization over algebraic extensions of \( \mathbb{Q} (u_1, \ldots, u_k) \), which is an expensive operation. In the case that any of the polynomials in a characteristic set can be factored, there is a branch for each irreducible factor as the zeros of \( p_j \) are the union of the zeros of its irreducible factors. Suppose we compute a characteristic set \( \Phi = \{ p_1, \ldots, p_l \} \) from \( \Sigma \) such that for \( i > 0 \), \( p_1, \ldots, p_i \) are irreducible over \( Q_0, \ldots, Q_{i-1} \), respectively, but \( p_{i+1} \) can be factored over \( Q_i \). It can be assumed that

\[
g \cdot p_{i+1} = p_{i+1}^1 \cdots p_{i+1}^j,
\]

where \( g \) is in \( \mathbb{Q} [u_1, \ldots, u_k, y_1, \ldots, y_i] \), and \( p_{i+1}^1 \cdots p_{i+1}^j \in \mathbb{Q} [u_1, \ldots, u_k, y_1, \ldots, y_i, y_{i+1}] \) and these polynomials are reduced with respect to \( p_1, \ldots, p_i \). Wu [31] proved that

\[
\text{Zero}(\Sigma) = \text{Zero}(\Sigma^{11}) \cup \cdots \cup \text{Zero}(\Sigma^{1j}) \cup \text{Zero}(\Sigma^{21}) \cup \cdots \cup \text{Zero}(\Sigma^{2i}),
\]
where $\Sigma^1 = \Sigma \cup \{p^h_{i+1}\}$, $1 \leq h \leq j$, and $\Sigma^2 = \Sigma \cup \{I_h\}$, where $I_h$ is the initial of $p_h$, $1 \leq h \leq i$. So characteristic sets are instead computed from new polynomials sets to give a system of characteristic sets. (To make full use of the intermediate computations already performed, $\Sigma$ above can be replaced by a saturation $\Delta$ of $\Sigma$ used to compute the reducible characteristic set $\Phi$.) The final result of this decomposition is a system of irreducible characteristic sets:

$$\text{Zero}(\Sigma) = \bigcup_i \text{Zero}(\Phi_i/J_i),$$

where $\Phi_i$ is an irreducible characteristic set and $J_i$ is the product of the initials of all the polynomials in $\Phi_i$.

**Example:** Consider the polynomial set

$$H = \{13z^3 + 12x^2 - 23xz + 16z^2 - 49x + 39z + 40, 13xz^2 - 12x^2 + 10xz - 29z^2 + 49x - 26z - 53, 13x^3 - 88x^2 + 4xz + 4z^2 + 173x - 94, yz - y - z + 1, x^2z - x^2 - 3xz + 3x + 2z - 2, y^2 - 1, yx^2 - 6yx + 9y - x^2 + 6x - 9\}.$$

The following characteristic set can be computed from it:

$$C = \{p_1 = x^6 - 12x^5 + 58x^4 - 144x^3 + 193x^2 - 132x + 36, \quad p_2 = (x^2 - 6x + 9)y - x^2 + 6x - 9, \quad p_3 = (x^2 - 3x + 2)z - x^2 + 3x - 2\}.$$

The polynomial $p_1$ in $C$ can be factored as $(x - 1)^2(x - 2)^2(x - 3)^2$, giving the following characteristic sets:

$$C_1 = \{x - 1, y - 1, z^2 + z + 1\}, \quad C_2 = \{x - 2, y - 1, z^2 + 2z + 1\}, \quad C_3 = \{x - 3, y^2 - 1, z - 1\}.$$

Characteristic sets $C_2$ and $C_3$ can be further decomposed to give:

$$C_{21} = \{x - 2, y - 1, z + 1\}, \quad C_{31} = \{x - 3, y + 1, z - 1\}, \quad C_{32} = \{x - 3, y - 1, z - 1\}.$$

The final system thus consists of irreducible characteristic sets $C_1, C_{21}, C_{31}$ and $C_{32}$.

In [32], Wu generalized the above result and gave a structure theorem as well as a method for decomposing $\text{Zero}(\Sigma/G)$. He showed that $\text{Zero}((\Sigma)/G)$ can also be decomposed into a finite union of the zero sets of irreducible characteristic sets written as:

$$\text{Zero}(\Sigma/g) = \bigcup_i \text{Zero}(\Phi_i/r_i),$$

where $r_i$ is the non-zero remainder of $J_i g_i$, $J_i$ is the product of the initials of polynomials in $\Phi_i$ and $g_i$ are certain non-zero polynomials.
3.6.1 Integrating D5 with characteristic sets

An alternative to using factorization over extension fields is to use Duval’s D5 method [12]. A reducible zero set can be decomposed into a union of zero sets of chains in which initials are invertible as defined below.

Invertibility and other computations with respect to an ideal

Let $I$ be an ideal generated by a basis $(p_1, \cdots, p_i)$. A polynomial $q$ is said to be invertible with respect to an ideal $I$ if there exist polynomials $a_1, \cdots, a_i$ and $q^{-1}$ such that

$$q^{-1}q = a_1p_1 + \cdots + a_ip_i + r,$$

where $r$ is a non-zero polynomial in the $u_j$’s. The polynomial $q^{-1}$ is called an inverse of $q$ with respect to $I$ in the quotient ring $Q[u_1, \cdots, u_k, y_1, \cdots, y_l]/I$.

If a polynomial $q$ is not invertible with respect to $(p_1, \cdots, p_i)$, then it is possible to find a polynomial $d$ such that

$$dq = a_1p_1 + \cdots + a_ip_i.$$

The polynomial $d$ is a zero divisor in the quotient ring $k[x_1, \cdots, x_n]/I$, and will be called an annihilator of $q$ with respect to $I$.

Similarly, a polynomial $r$ divides another polynomial $s$ with respect to $I$ if and only if there exist $q$ and $a_i$’s such that

$$qr = a_0s + a_1p_1 + \cdots + a_ip_i,$$

where $a_0$ is a polynomial in the $u_j$’s, if any.

It is easy to see that $q$ being invertible with respect to $I$ means that $q$ and $I$ do not have any common zero in any algebraic closed extension of the field $k(u_1, \cdots, u_k)$ (follows from Hilbert’s Nullstellensatz). A polynomial $q$ is not invertible with respect to $I$ if there do not exist $r, q^{-1}, a_1, \cdots, a_i$ satisfying the above properties, in which case, $q$ and $I$ have a common zero (again follows from Hilbert’s Nullstellensatz).

If a basis of $I$ is a chain $C$, then the class of $p_i$ above is lower than or equal to the class of $q$, and the inverse $q^{-1}$ of $q$ is of lower rank than $p_i$. If $q$ is is not invertible with respect to $C$, then also the class of $p_i$ is lower than or equal to the class of $q$, and $d$, the annihilator of $q$, is of lower rank than $p_i$. Also, if $r$ divides $s$ with respect to $C$ then the class of $p_i$ is not higher than the class of $s$. Henceforth by a polynomial $q$ being invertible (or not invertible) with respect to $C$, we mean that $q$ is invertible (not invertible, respectively) with respect to the ideal generated by $C$.

A chain $C = (p_1, \cdots, p_n)$ is said to be invertible if the initial of every polynomial $p_i$ in $C$ is invertible with respect to $C$ (really, all polynomials smaller than $p_i$ in $C$).

Let $C = \{p_1, p_2, \ldots, p_l\}$ be a chain with $y_i$ being the main variable of $p_i$. Without any loss of generality, we can assume that $f$ whose inverse is being attempted, is already reduced with respect to $C$ (if $f$ is not already reduced with respect to $C$, it is reduced using $C$ first). In case $f$ does not have an inverse with respect to $C$, the process of trying to invert $f$ may lead to a splitting of $C$.

Let $y_i$ be the highest variable appearing in $f$. Treating $f$ and $p_i$ as polynomials with coefficients that are polynomials in $u_1, \ldots, u_k, y_1, \ldots, y_{i-1}$, i.e., $f, p_i \in \mathbb{Q}[u_1, \ldots, u_i, y_1, \ldots, y_{i-1}][y_i]$, 28
compute their greatest common divisor using the extended Euclidean algorithm. The crucial point here is that we proceed as though the leading coefficients of the remainders that appear in performing Euclid’s algorithm on \( f \) and \( p_i \) are invertible in \( C \). Let \( g \) be the gcd of \( f \) and \( p_i \). We have

\[
g = fa + pq
\]

where \( p,q \) are the multipliers produced by the extended Euclidean algorithm. There are three possibilities:

- if \( g \) is a polynomial in \( u_j \)'s, then \( a \) is the inverse of \( f \) with respect to \( C \);
- if \( g = p_i \), then \( p_i \) divides \( f \) and \( f \) is not invertible with respect to \( C \);
- otherwise, \( p_i \) has factors \( g \) and \( p_i/g \), and the chain \( C \) is equivalent to the union of the two chains

\[
C^{(1)} = \{p_1, p_2, \ldots, g, p_i+1, \ldots, p_l\}, \text{ and } C^{(2)} = \{p_1, p_2, \ldots, p_i/g, p_i+1, \ldots, p_l\}.
\]

Polynomial \( f \) is not invertible with respect to \( C^{(1)} \) and the inverse of \( f \) with respect to \( C^{(2)} \) is given by \( g^{-1}a \) where \( g^{-1} \) denotes the inverse of \( g \) with respect to \( C^{(2)} \).

The crucial point is that the chain \( C \) can split at lower levels. This is because the extended Euclidean algorithm used to compute the gcd of \( f \) and \( p_i \) needs to compute inverses of polynomials with respect to \( p_j, j < i \), which is done recursively.

The operation of inverting a polynomial \( f \) with respect to a chain \( C \) can thus produces a family of chains \( C^{(1)}, C^{(2)}, \ldots, C^{(k)} \) and polynomials \( h_1, \ldots, h_k \) such that

- \( C \) is equivalent to the union of the chains \( C^{(1)}, C^{(2)}, \ldots, C^{(k)} \);
- if \( h_i \neq 0 \), then \( h_i \) is the inverse of \( f \) with respect to \( C^{(i)} \), i.e., the remainder of \( fh_i \) with respect to \( C^{(i)} \) is a polynomial in \( u_j \)'s; if \( h_i = 0 \), then \( f \) is not invertible with respect to \( C^{(i)} \).

We illustrate the operation of inversion below; for complete details on the D5 technique, we refer the reader to [12].

**Example:** Let us invert the polynomial \( f = (x - 1)z + 1 \) with respect to the chain \( C = (x^2 - 3x + 2, y - 1, z^2 + xz + 1) \) with the ordering \( x < y < z \). The first step is to compute the gcd of \( f \) and \( z^2 + xz + 1 \), treating them as polynomials in \( z \). At this point, we are required to invert \( x - 1 \) (the leading coefficient of \( f \)) with respect to \( C \). We now try to compute the gcd of \( y - 1 \) and \( x - 1 \), treating them as polynomials in \( y \) and we are lead to invert \( x - 1 \). We then try to compute the gcd of \( x^2 - 3x + 2 \) and \( x - 1 \) treating them as polynomials in \( x \) and we find that \( x - 1 \) is the gcd.

This leads to a split of \( C \) into \( C^{(1)} = (x - 1, y - 1, z^2 + xz + 1) \) and \( C^{(2)} = (x - 2, y - 1, z^2 + xz + 1) \). Polynomial \( x - 1 \) is not invertible with respect to \( C^{(1)} \) and the inverse of \( x - 1 \) with respect to \( C^{(2)} \) is 1 (since \((x - 1),1 \equiv 1 \mod (x - 2)\)). \( C^{(1)}, C^{(2)} \) can be reduced to give \( C^{(1)} = (x - 1, y - 1, z^2 + z + 1) \) and \( C^{(2)} = (x - 2, y - 1, z^2 + 2z + 1) \). Note that we may have to reduce the chains as we propagate the split upwards. Our task now is to invert \( f \) with respect to \( C^{(1)} \) and \( C^{(2)} \) separately. The polynomial \( f \) has remainder 1 with respect to \( C^{(1)} \) and 1 is its own inverse; \( f \) has remainder \( z + 1 \) with respect to \( C^{(2)} \). We find that the gcd of \( z + 1 \) and \( z^2 + 2z + 1 \) is \( z + 1 \). Therefore, \( f \) is not invertible with respect to \( C^{(2)} \).
Decomposing a zero set

After a characteristic set $\Phi$ from $\Delta$ is computed as discussed earlier, we try to check whether the initial of each polynomial $p_i$ in $\Phi$ has a zero common with polynomials smaller than $p_i$ in $\Phi$. This can be done by attempting to invert the initial with respect to $p_i$’s using the D5 method. For instance, we try to invert the initial $I_2$ of $p_2$ with respect to $p_1$; since the initial of $p_1$ is a polynomial in $u_j$’s, it is invertible. If $I_2$ is invertible, multiply $p_2$ by the inverse of $I_2$ which will make the result monic, i.e., the initial of the result is a polynomial in $\mathbb{Q}[u_1, \ldots, u_k]$ (this step is optional and is unnecessary as we only need to ensure that the initial is invertible; the initial does not have to be made monic). If $I_2$ is not invertible, then $C$ is split by factoring $p_1$ into $p_1'$ and $p_1''$, where $p_1'$ is a common factor of $p_1$ and $I_2$. Decompose the characteristic set $\Phi$ as follows: Each of $p_1'$ and $p_1''$ is augmented to $\Delta$ separately and two new characteristic sets are computed. After $p_2$ has been considered, $p_3$ is processed and so on. The result in this case also is a system of characteristic sets which need not be irreducible (since a polynomial in a characteristic set so generated could be still factored), but their initials are invertible.

For the polynomial set $H$ discussed above, we try to invert the initial of $p_2$ in $C$ with respect to $p_1$; since $x^2 - 6x + 9$ is a factor of $p_1$, we split and compute a characteristic set of $H \cup \{x^2 - 6x + 9\}$, which is,

$$D_1 = \{x^2 - 6x + 9, y^2 - 1, (3x - 7)z - 3x + 7\},$$

and a characteristic set of $H \cup \{x^4 - 6x^3 + 13x^2 - 12x + 4\}$, which is

$$D_2 = \{x^4 - 6x^3 + 13x^2 - 12x + 4, (x^2 - 6x + 9)y - x^2 + 6x - 9, (x^2 - 3x + 2)z - x^2 + 3x - 2\}.$$

Now, $x^2 - 6x + 9$ is invertible with respect to $x^4 - 6x^3 + 13x^2 - 12x + 4$. However, $x^4 - 6x^3 + 13x^2 - 12x + 4$ is square of $(x^2 - 3x + 2)$, so we compute a characteristic set from $H \cup \{x^4 - 6x^3 + 13x^2 - 12x + 4\} \cup \{x^2 - 3x + 2\}$, which is:

$$D_{21} = \{x^2 - 3x + 2, (3x - 7)y - 3x + 7, (13x - 29)z^2 + (10x - 26)z + 13x - 29\}.$$

Each of the initials is now invertible, giving us a decomposition consisting of $D_1$ and $D_{21}$. Notice that neither of the two characteristic sets is irreducible. If the initials are made monic, we get:

$$D_1 = \{x^2 - 6x + 9, y^2 - 1, z - 1\},$$

$$D_{21} = \{x^2 - 3x + 2, y - 1, z^2 + xz + 1\}.$$

Compare this output with the one obtained using Wu’s method in which there are 4 irreducible characteristic sets.

Acknowledgement: I got interested in resultants while writing an introductory article on elimination methods with Lakshman Y.N. [16]. I thank him for many discussion on this subject. Yang Lu’s visit from Academia Sinica to SUNY, Albany, in the fall of 1993 acted as a catalytic agent for our work on Dixon resultants. Results on Dixon resultants reported in these notes were obtained jointly with Yang Lu and my student Tushar Saxena. Some of material included in these notes is taken from the introductory paper with Lakshman [16] and our two papers in ISSAC’94 [20] and ISSAC’95 [18], and another draft paper [19].
References


