Sparse Resultants

Peter F. Stiller*

Introduction

The classical Macaulay resultant of a system of $n+1$ equations in n variables

$$\begin{aligned} f_1(x_1, \dots, x_n) &= 0 \\ &\vdots \\ f_{n+1}(x_1, \dots, x_n) &= 0 \end{aligned} \tag{1}$$

is a polynomial expression in the symbolic coefficients of the f_i which will evaluate to zero if and only if the f_i have a common solution (possibly at infinity). To compute this resultant, we first homogenize the system to get:

$$\begin{aligned} F_1(X_0, \dots, X_n) &= 0 \\ &\vdots \\ F_{n+1}(X_0, \dots, X_n) &= 0 \end{aligned}$$

where

$$F_i(X_0, \dots, X_n) = X_0^{d_i} f_i\left(\frac{X_1}{X_0}, \dots, \frac{X_n}{X_0}\right)$$

and where d_i = degree f_i. We then form a certain square matrix A of size $\binom{n+d}{d}$ where $d = 1 + \sum_{i=1}^{n+1}(d_i - 1)$. The Macaulay resultant divides the determinant of this matrix A. (In fact the extraneous factor is the determinant of a submatrix M of A.)

A key observation here is that the size of the matrix A, and hence the size/complexity of the computation involved, is related to the degrees d_i of the polynomials involved. In effect, the form of A allows for the possibility that every monomial of degree less than or equal to d_i actually occurs in f_i. Of course in practice many of these monomials will be absent. The theory of sparse resultants attempts to exploit the pattern of missing terms to construct resultants via determinants of much smaller matrices. Such special forms for resultants are not new, for example the Dixon resultant is a special type of sparse resultant.

To explain what lies behind this smaller matrix, recall that the classical Bezout bound for a system of n equations in n variables

$$\begin{aligned} g_1(x_1, \dots x_n) &= 0 \\ &\vdots \\ g_n(x_1, \dots, x_n) &= 0 \end{aligned} \tag{2}$$

*Professor of Mathematics and Computer Science, Texas A&M University, College Station, TX 77843-3368

states that the system has either an infinite number of solutions (including solutions at infinity) or it has no more than $\tilde{d} = \tilde{d}_1\tilde{d}_2 \dots \tilde{d}_n$ complex solutions (where $\tilde{d}_i$ = degree g_i). Moreover, if we count solutions at infinity and with appropriate multiplicity, then $\tilde{d} = \tilde{d}_1\tilde{d}_2 \dots \tilde{d}_n$ is the exact number of solutions in complex projective space, $\mathbb{P}^n_{\mathbb{C}}$. Of course, we are usually not interested in solutions at infinity, and one often finds in practice that relatively few of the solutions predicted by the Bezout count lie in $\mathbb{C}^n$; the majority lie at infinity.

The sparse resultant techniques exploit an alternative bound, the so-called BKK bound, which bounds the number of solutions of a system like (2) in $(\mathbb{C}^*)^n$. Here $\mathbb{C}^* = \mathbb{C} - \{0\}$ is the set of nonzero complex numbers. Notice that $(\mathbb{C}^*)^n$ is $\mathbb{P}^n_{\mathbb{C}}$ minus all the coordinate hyperplanes $x_0 = 0$, $x_1 = 0, \dots,$ and $x_n = 0$. Since $x_0 = 0$ is the hyperplane at infinity, the BKK bound counts solutions to (2) in $\mathbb{C}^n$ which have all coordinates non-zero. This bound is computed from the so-called mixed volume of a sum of polytopes constructed from the system (2).

Two remarks are in order. First, because we have removed zero from consideration, we can consider more general systems of equations; namely systems where the g_i's are polynomials in the x_i's and their reciprocals the x_i^{-1}'s. In other words, we can work in the coordinate ring $\mathbb{C}[x_1, x_1^{-1}, \dots, x_n, x_n^{-1}]$ of $(\mathbb{C}^*)^n$. This is completely analogous to working with n polynomials in $2n$ variables, $\mathbb{C}[x_1, \dots, x_n, y_1, \dots, y_n]$, where we also impose the additional n conditions $x_iy_i - 1 = 0$ for $i = 1, \dots, n$. Any solution to all $2n$ equations necessarily has nonzero values. Second, we can easily get around the problem of having excluded those solutions where some of the x_i's are zero. For example, if in

$$\begin{aligned} g_1(x_1, \dots, x_n) &= 0 \\ &\vdots \\ g_n(x_1, \dots, x_n) &= 0, \end{aligned} \tag{3}$$

we want to find solutions with $x_1 = 0$, we simply consider the system

$$\begin{aligned} g_1(0, x_2, \dots, x_n) &= 0 \\ &\vdots \\ g_n(0, x_2, \dots, x_n) &= 0, \end{aligned} \tag{4}$$

For the generic system (i.e. symbolic coefficients) we can use this to count the number of solutions N in $\mathbb{C}^n$. Namely,

$$N = N_0 + N_1 + N_2 + N_3 + N_4 + \cdots$$

where N_0 is the number of solutions to (3) with no zero values (i.e. the number predicted by the BKK bound), N_1 is the number of solutions where exactly one of the x_i is zero, N_2 is the number of solutions where exactly two of the x_i are zero, etc. The only problem with this approach is that once zero is substituted for one of the x_i's, the system (e.g. (4)) becomes overdetermined. However, for systems with no more than three variables this is not a serious impediment to carrying out the count in the manner indicated above.

1 Newton Polytopes and the BKK Bound

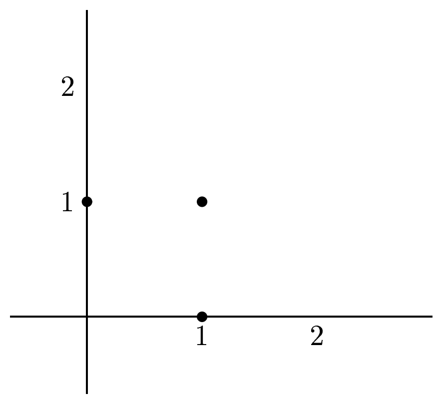
We begin by considering a single polynomial $f(x_1, \dots, x_n)$ in n-variables. To reduce notation, we use x^e to denote the monomial $x_1^{e_1}x_2^{e_2} \dots x_n^{e_n}$ where $e = (e_1, \dots, e_n) \in \mathbb{Z}^n$ is a multi-exponent. Let

$E_f = \{m_1, \ldots, m_{i_f}\} \subset \mathbb{Z}^n$ denote the set of all multi-exponents occurring in f, so that

$$f = \sum_{j=1}^{i_f} c_j x^{m_j}$$

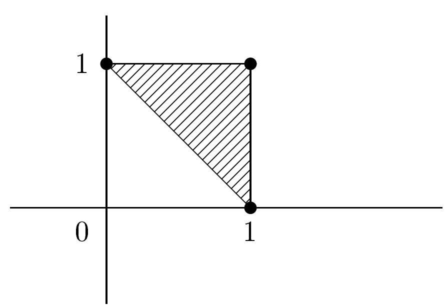
where $c_j \neq 0$. Thus E_f is exactly the set of integer lattice points in $\mathbb{R}^n$ that occur as multi-exponents of nonzero terms in f.

Example 1: If $f(x, y) = 3x - 2y - xy$ then $E_f = \{(1, 0), (0, 1), (1, 1)\} \subset \mathbb{Z}^2 \subset \mathbb{R}^2$. Graphically this is



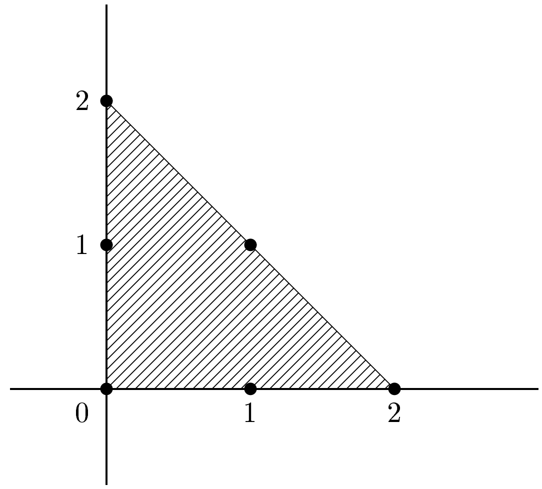
Definition 1: *The Newton polytope of f, denoted Q_f, is the convex hull of the set E_f in $\mathbb{R}^n$.*

Example 2: For f as in Example 1, the Newton polytope Q_f is a triangle:



What makes sparse resultants "sparse" is that the matrix we use to compute the resultant is based on the specific Newton polytopes (i.e. the specific multi-exponents) that occur, rather than all terms up to some particular degree.

Example 3: The polynomial in Example 1, $f(x, y) = 3x - 2y - xy$, has degree 2. The general polynomial of degree 2 is $c_0 + c_1x + c_2y + c_3x^2 + c_4xy + c_5y^2$ which has a much larger Newton polytope:



Now consider a system of n equations in n unknowns

$$\begin{aligned} g_1(x_1, \ldots, x_n) &= 0 \\ &\vdots \\ g_n(x_1, \ldots, x_n) &= 0. \end{aligned}$$

Each equation has its own Newton polytope Q_{g_i}. Polytopes can be "added" to produce new polytopes by a process called Minkowski sum.

Definition 2: *The Minkowski sum $P_1 + P_2$ of two convex polytopes P_1 and P_2 in $\mathbb{R}^n$ is the set*

$$P = P_1 + P_2 = \{p_1 + p_2 \mid p_1 \in P_1 \text{ and } p_2 \in P_2\}$$

which is again a convex polytope. We let Vol(P) denote the usual volume of P in $\mathbf{R}^n$.

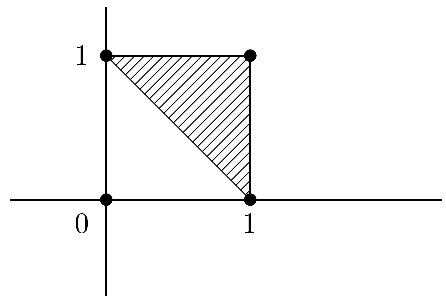
In order to state the BKK bound, we must introduce the notion of mixed volume. Specifically, given convex polytopes $P_1, \ldots, P_n \subseteq \mathbb{R}^n$, there is a unique real-valued function $MV(P_1, \ldots, P_n)$ called the mixed volume which is multilinear with respect to Minkowski sum, and has the property that $MV(P_1, P_1, \ldots, P_1) = n!\,\mathrm{Vol}(P_1)$. Equivalently, if $\lambda_1, \ldots, \lambda_n$ are scalars, then $MV(P_1, \ldots, P_n)$ is precisely the coefficient of $\lambda_1\lambda_2 \ldots \lambda_n$ in $\mathrm{Vol}(\lambda_1 P_1 + \cdots + \lambda_n P_n)$ expanded as a polynomial in $\lambda_1, \ldots, \lambda_n$.

Theorem 1 (BKK Bound): *Let $g_1, \ldots, g_n \in \mathbb{C}[x_1, x_1^{-1}, \ldots, x_n, x_n^{-1}]$, then the number of common zeros in $(\mathbb{C}^*)^n$ is either infinite, or is less than or equal to $MV(Q_{g_1}, \ldots, Q_{g_n})$. If the g_i's are given symbolic coefficients, then for almost all choices of numerical values for those coefficients the number of solutions is exactly $MV(Q_{g_1}, \ldots, Q_{g_n})$.*

Example 4: Consider the following system of equations

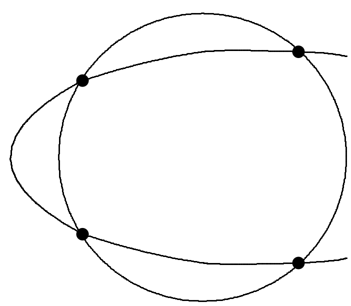
$$\begin{aligned} g_1(x, y) &= c_1 x + c_2 y + c_3 xy = 0 \\ g_2(x, y) &= d_1 x + d_2 y = d_3 xy = 0 \end{aligned}$$

with the c_i's and d_j's $\neq 0$. Note that the only common solution with either x or y equal to zero is $(x, y) = (0, 0)$. The BKK bound in this case is $MV(Q_{g_1}, Q_{g_2})$ where Q_{g_1} and Q_{g_2} are both the triangle Q pictured below:



As $MV(Q,Q) = 2!\ \mathrm{Vol}(Q) = 2 \cdot \frac{1}{2} = 1$, the BKK bound assures us in this case that there is at most one solution with $x \neq 0$ and $y \neq 0$ (unless there are infinitely many, which can only occur if g_1 and g_2 have a common factor, which in this case forces g_1 and g_2 to be scalar multiples of one another).

Note that since both g_1 and g_2 have degree 2, the Bezout bound is 4. To compare the two bounds, one can easily see that this system always has 2 solutions at infinity. Thus the four solutions of the Bezout bound are accounted for by (0,0), two solutions at infinity, and the one solution from the BKK bound. Of course, the general pair of quadratic equations in two variables can have up to four solutions in $(\mathbb{C}^*)^2 \subset \mathbb{P}^2_{\mathbb{C}}$:



but because our two equations are "sparse", i.e. omit certain monomials, the BKK bound improves on the Bezout bound.

2 Sparse Resultants

Consider now a system of $n+1$ polynomials in n variables:

$$\begin{aligned} f_1(x_1, \ldots, x_n) &= 0 \\ &\vdots \\ f_{n+1}(x_1, \ldots, x_n) &= 0. \end{aligned} \tag{5}$$

In general such a system is over-determined and has no solutions. Thus, in order for a solution to exist, some relationship must hold among the coefficients of the f_i's. In fact, there will be a single such polynomial relationship, known as the resultant of the system. This resultant is often calculated using the determinant of a matrix involving the coefficients of the f_i. Unfortunately, it is known that resultants, in particular sparse resultants, are not in general *equal* to a determinant.

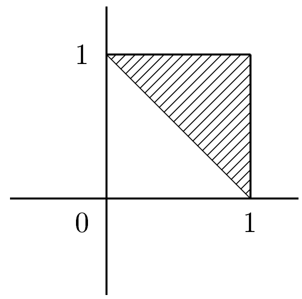
They will however divide a determinant. Thus we can calculate them up to some extraneous factor as a determinant. For sparse resultants the method is described below.

To calculate the sparse resultant, we make use of the combinatorial data contained in a mixed subdivision of the polytope $Q = Q_{f_1} + \cdots + Q_{f_{n+1}} \subset \mathbb{R}^n$, the Minkowski sum of the Newton polytopes of the polynomials $f_1, \ldots, f_{n+1}$. We therefore begin with a definition of mixed subdivision.

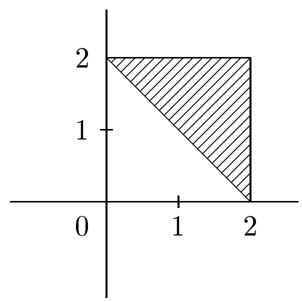
Consider m polytopes $Q_1, \ldots, Q_m \subset \mathbb{R}^n$ and their Minkowski sum $Q = Q_1 + \cdots + Q_m \subset \mathbb{R}^n$.

Definition 3: *A <u>mixed subdivision</u> Δ of $Q = Q_1 + \cdots + Q_m$ is a polyhedral subdivision of Q such that every polyhedron $F \in \Delta$ is of the form $F = F_1 + \cdots + F_m$ where F_i is a face of Q_i and* $\dim F = \sum_{i=1}^{m} \dim F_i$.

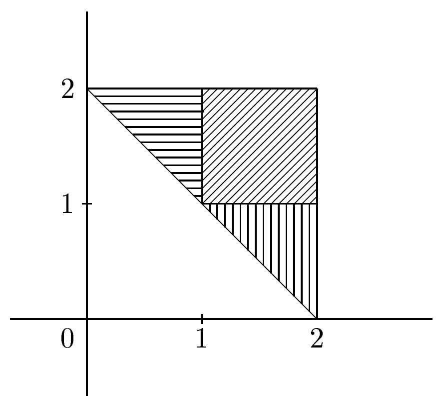
Example 5: If $Q_1 = Q_2 =$



then $Q = Q_1 + Q_2$ is



A mixed subdivision is:



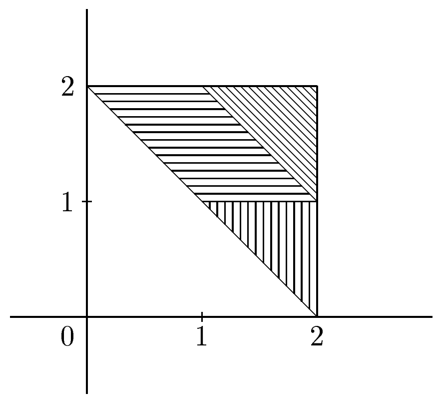
where:

(1) the triangle on the upper left is the sum of the vertex (0,1) regarded as a face of Q_1 with all of Q_2;

(2) the triangle on the lower right is the sum of Q_1 with the vertex (1,0) of Q_2;

(3) the square is the sum of the edge $\{(x,1) \text{ s.t. } 0 \leq x \leq 1\}$ of Q_1 with the edge $\{(1,y) \text{ s.t. } 0 \leq y \leq 1\}$ of Q_2.

Note that these descriptions are not unique. Moreover the subdivision itself is not unique. Another possibility is:



Fast algorithms exist for finding mixed subdivisions.

Among the polyhedra of a mixed subdivision Δ of $Q = Q_1 + \cdots + Q_m \subset \mathbb{R}^n$ are some that are especially important:

Definition 4: *A polyhedron $F = F_1 + \cdots + F_m \in \Delta$ is called a* <u>*mixed facet*</u> *if $\dim F = n$ and every F_i has dimension ≤ 1. (This requires $m \geq n$.)*

If $m = n$ then all the F_i in a mixed facet $F = F_1 + \cdots + F_n$ have $\dim F_i = 1$, and in this case the mixed volume is calculated as the sum of the ordinary volumes in $\mathbb{R}^n$ of the mixed facets of the mixed subdivision Δ:

$$MV(Q_1, \ldots, Q_n) = \sum_{\substack{\text{mixed} \\ \text{facets} \\ F \in \Delta}} \mathrm{Vol}(F).$$

Example 6: In Example 5 above the square is the only mixed facet. Thus $MV(Q_1, Q_2) = \mathrm{Vol}(\text{Square}) = 1$. Note this agrees with our previous observation; since $Q_1 = Q_2$, we had $MV(Q_1, Q_2) = 2!\,\mathrm{Vol}(Q_1) = 1$. Also in the second mixed subdivision in Example 5 the only mixed facet is the parallelogram which also has volume equal to 1.

Now we consider again the system (5) of $n+1$ equations in n variables and compute its resultant (up to extraneous factors) as $\det M$ for a certain square matrix M. The rows and columns of M will essentially be indexed by the integer lattice points in the Minkowski sum $Q = Q_{f_1} + \cdots + Q_{f_{n+1}}$ of the Newton polytopes of the polynomials f_i, $i = 1, \ldots, n+1$ in the system (5). The specific entries m_{ij} of M are determined with the help of a mixed subdivision Δ of Q.

Explicitly, we must first perturb the Minkowski sum Q slightly so that each integer lattice point will lie in the interior of a polyhedron F in the mixed subdivision Δ. So pick a small vector δ and consider the shifted polyhedron $(\delta + Q)$.

The rows and columns of M will be indexed by

$$I = \mathbb{Z}^n \cap (\delta + Q).$$

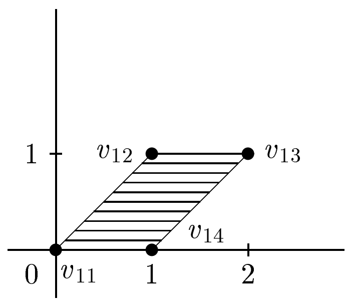
Example 7 (see Canny "A Toolkit for Non-linear Algebra"):

Let

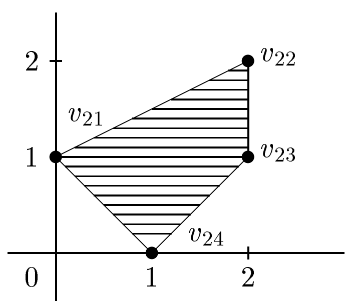
$$\begin{aligned}
f_1(x,y) &= c_{11} + c_{12}xy + c_{13}x^2y + c_{14}x \\
f_2(x,y) &= c_{21}y + c_{22}x^2y^2 + c_{23}x^2y + c_{24}x \\
f_3(x,y) &= c_{31} + c_{32}y + c_{33}xy + c_{34}x.
\end{aligned}$$

The Newton polytopes are

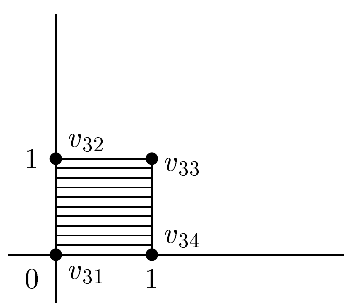
$Q_{f_1} =$



$Q_{f_2} =$

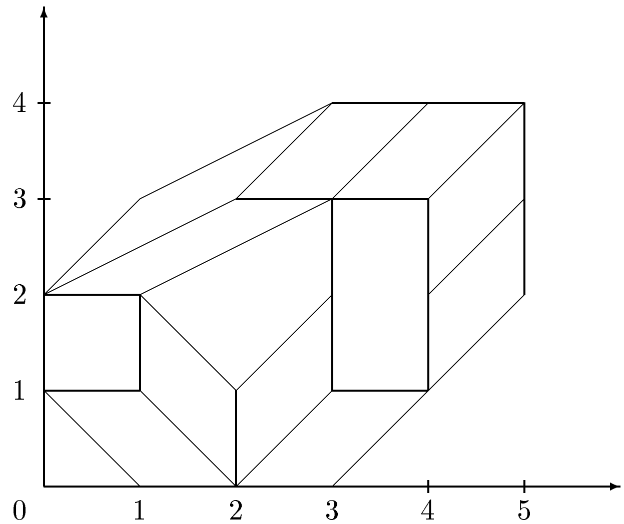


$Q_{f_3} =$

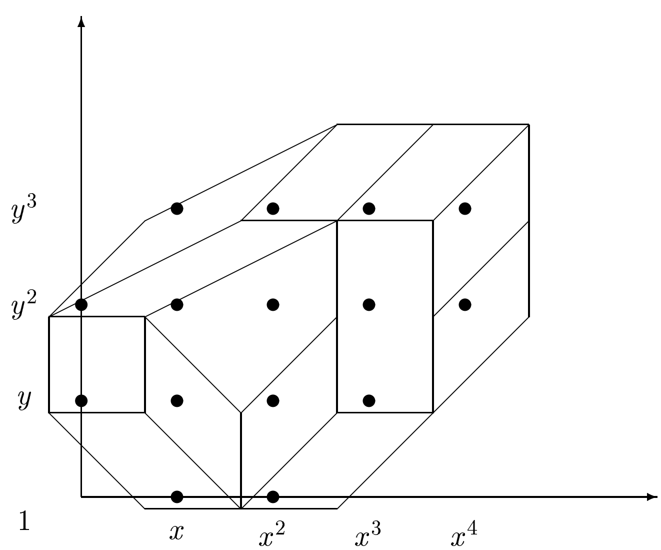


where we have labeled the vertices to correspond with the respective coefficients of the f_i's.

The Minkowski sum Q with a particular mixed subdivision Δ is shown below:



We perturb this slightly by shifting it to the left and down. This leaves 15 lattice points $\{(1{,}0)$ $(2{,}0)$ $(0{,}1)$ $(1{,}1)$ $(2{,}1)$ $(3{,}1)$ $(0{,}2)$ $(1{,}2)$ $(2{,}2)$ $(3{,}2)$ $(4{,}2)$ $(1{,}3)$ $(2{,}3)$ $(3{,}3)$ $(4{,}3)\}$ inside $\delta + Q$. Note that 7 lattice points in Q are dropped when we shift.



M will thus be a 15×15 matrix of the form:

$$
\begin{array}{cc}
 & \begin{array}{ccccccccccccccc} 1,0 & 2,0 & 0,1 & 1,1 & 2,1 & 3,1 & 0,2 & 1,2 & 2,2 & 3,2 & 4,2 & 1,3 & 2,3 & 3,3 & 4,3 \end{array} \\
\begin{array}{c} 1,0 \\ 2,0 \\ 0,1 \\ 1,1 \\ 2,1 \\ 3,1 \\ 0,2 \\ 1,2 \\ 2,2 \\ 3,2 \\ 4,2 \\ 1,3 \\ 2,3 \\ 3,3 \\ 4,3 \end{array} & \left[\begin{array}{ccccccccccccccc}
 &&&&&&&&&&&&&& \\
 &&&&&&&&&&&&&& \\
 &&&&&&&&&&&&&& \\
 &&&&&&&&&&&&&& \\
 &&&&&&&&&&&&&& \\
 &&&&&&&&&&&&&& \\
 &&&&&&&&&&&&&& \\
 &&&&&&&&&&&&&& \\
 &&&&&&&&&&&&&& \\
 &&&&&&&&&&&&&& \\
 &&&&&&&&&&&&&& \\
 &&&&&&&&&&&&&& \\
 &&&&&&&&&&&&&& \\
 &&&&&&&&&&&&&& \\
 &&&&&&&&&&&&&&
\end{array}\right]
\end{array}
$$

Its rows and columns will be indexed by the lattice points inside $\delta + Q$.

To illustrate how to fill in a row of the matrix M, we take a particular case from the example. Consider the row labeled $(1,2)$, which is the eighth row as shown above. The point $(1,2)$ lies in a certain polytope $\delta + F$ where $F = F_1 + F_2 + F_3$ is a polytope of the mixed subdivision Δ of Q. In this case F_1 is the vertex v_{11} of Q_{f_1}, F_2 is all of Q_{f_2}, and F_3 is the vertex v_{33} of Q_{f_3}. The rule is that we select the vertex associated with the F_i of dimension zero having the largest i, in this case $i = 3$. Thus v_{33} is the relevant vertex. Now $v_{33} = (1,1)$. We take the row index (1,2) and subtract (1,1) to get (0,1) which is the monomial y. We multiply y with f_3 (because we are using a vertex in Q_{f_3}):

$$y f_3 = c_{31} y + c_{32} y^2 + c_{33} x y^2 + c_{34} x y.$$

We enter c_{31} in the (0,1) column of row (1,2), c_{32} in the (0,2) column, c_{33} in the (1,2) column, and c_{34} in the (1,1) column. All other entries are zero. The final result would be

$$
\begin{array}{cc}
 & \begin{array}{ccccccccccccccc} 1,0 & 2,0 & 0,1 & 1,1 & 2,1 & 3,1 & 0,2 & 1,2 & 2,2 & 3,2 & 4,2 & 1,3 & 2,3 & 3,3 & 4,3 \end{array} \\
\begin{array}{c} 1,0 \\ 2,0 \\ 0,1 \\ 1,1 \\ 2,1 \\ 3,1 \\ 0,2 \\ 1,2 \\ 2,2 \\ 3,2 \\ 4,2 \\ 1,3 \\ 2,3 \\ 3,3 \\ 4,3 \end{array} & \left[\begin{array}{ccccccccccccccc}
c_{11} & c_{14} & 0 & 0 & c_{12} & c_{13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
c_{31} & c_{34} & 0 & c_{32} & c_{33} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
c_{24} & 0 & c_{21} & 0 & c_{23} & 0 & 0 & 0 & c_{22} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & c_{11} & c_{14} & 0 & 0 & 0 & c_{12} & c_{13} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & c_{11} & c_{14} & 0 & 0 & 0 & c_{12} & c_{13} & 0 & 0 & 0 & 0 \\
0 & c_{24} & 0 & c_{21} & 0 & c_{23} & 0 & 0 & 0 & c_{22} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & c_{11} & c_{14} & 0 & 0 & 0 & c_{12} & c_{13} & 0 & 0 \\
0 & 0 & c_{31} & c_{34} & 0 & 0 & c_{32} & c_{33} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & c_{31} & c_{34} & 0 & 0 & c_{32} & c_{33} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & c_{31} & c_{34} & 0 & 0 & c_{32} & c_{33} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & c_{24} & 0 & 0 & c_{21} & 0 & c_{23} & 0 & 0 & 0 & c_{22} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{31} & c_{34} & 0 & 0 & c_{32} & c_{33} & 0 & 0 \\
0 & 0 & 0 & c_{24} & 0 & 0 & c_{21} & 0 & c_{23} & 0 & 0 & 0 & c_{22} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{31} & c_{34} & 0 & 0 & c_{32} & c_{33} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{31} & c_{34} & 0 & 0 & c_{32} & c_{33}
\end{array}\right]
\end{array}
$$

Note that the labeling and indexing choices could affect the placing of the matrix entries, but the desired resultant will always be a factor of $\det M$ independent of the choices made.

3 Solving Systems of Equations

Consider now a system of n equations in n variables

$$\begin{aligned} f_1(x_1, \ldots, x_n) &= 0 \\ &\vdots \\ f_n(x_1, \ldots, x_n) &= 0 \end{aligned} \tag{6}$$

and let's assume that the set of common solutions (including the solutions at infinity) consists of a finite set of isolated points. In other words, we assume that the homogenized system,

$$\begin{aligned} F_1(X_0, \ldots, X_n) &= 0 \\ &\vdots \\ F_n(X_0, \ldots, X_n) &= 0 \end{aligned} \tag{7}$$

where

$$F_i(X_0, \ldots, X_n) = X_0^{d_i} f_i\left(\frac{X_1}{X_0}, \ldots, \frac{X_n}{X_0}\right)$$

and where $d_i =$ degree f_i, has a zero dimensional set of solutions in complex projective n-space, $\mathbb{P}^n_{\mathbb{C}}$.

As we have already mentioned, Bezout's Theorem tells us that the number of common solutions is bounded by the product of the degrees $d = \prod_{i=1}^n d_i$, and in fact is equal to d if we count solutions (including those at infinity) with appropriate multiplicity. The question now is to find those solutions, and to do it by making use of the sparse resultant.

Recall that the resultant tells us, loosely speaking, whether or not a system of $n+1$ equations in n variables (or $n+1$ homogeneous equations in $n+1$ variables) has a solution (non-trivial solution). Thus to use the resultant, we must augment the system (6) (or (7)) by one equation. The appropriate choice of equation to add is a generic linear equation:

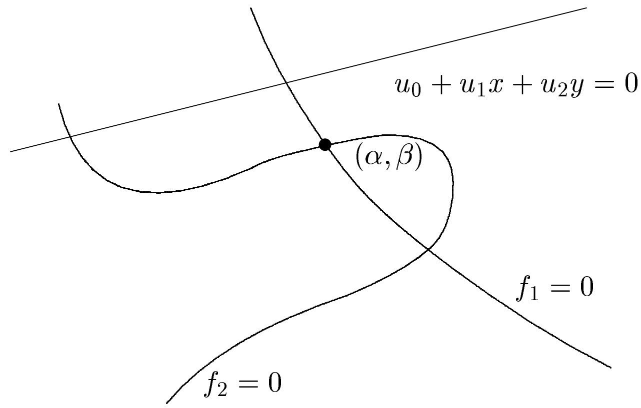
$$u_0 + u_1 x_1 + \cdots + u_n x_n = 0$$

(or $u_0 X_0 + u_1 X_1 + \cdots + u_n X_n = 0$).

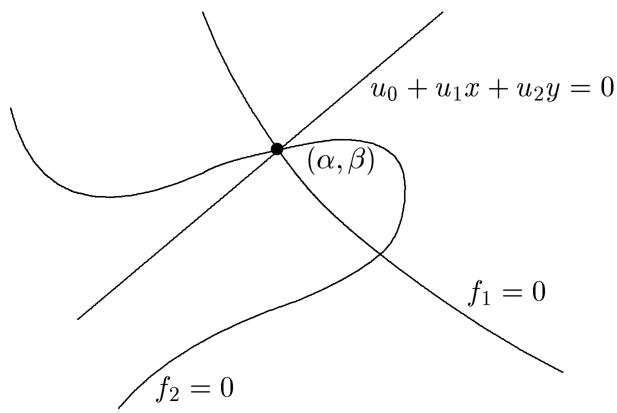
The idea behind this choice is best illustrated in the case of two variables. Consider a system of 2 equations in 2 variables:

$$\begin{aligned} f_1(x, y) &= 0 \\ f_2(x, y) &= 0. \end{aligned} \tag{8}$$

Geometrically these equations represent two curves in the plane, and the common solutions to the system (8) are just the points where the curves intersect. Say (α, β) is one such point:



If we add a general line $u_0 + u_1x + u_2y = 0$ to the system, the result will be a system of 3 equations in two variables which will have *no* common solution. Only special choices of u_0, u_1, and u_2 will produce a system with such a solution. On the one hand, this will be the case if the line $u_0 + u_1x + u_2y$ goes through one of the solutions to the original system $f_1 = 0$ and $f_2 = 0$ (possibly at infinity) say for example (α, β):



This will happen precisely when the u_i satisfy $u_0 + \alpha u_1 + \beta u_2 = 0$. On the other hand, we know that

$$\begin{aligned} f_1(x, y) &= 0 \\ f_2(x, y) &= 0 \\ u_0 + u_1x + u_2y &= 0 \end{aligned}$$

will have a common solution (possibly at infinity) if and only if the resultant $R(u_0, u_1, u_2) = 0$. Here we regard the resultant as a polynomial in the variables u_0, u_1, u_2. It follows from this that $u_0 + \alpha u_1 + \beta u_2$ must be a factor of $R(u_0, u_1, u_2)$. In fact, we see that $R(u_0, u_1, u_2)$ must factor completely into linear factors of the form $\lambda u_0 + \theta u_1 + \tau u_2$ where $(\lambda : \theta : \tau) \in \mathbb{P}^2_{\mathbb{C}}$ is a solution of $F_1(X, Y, Z) = 0$ and $F_2(X, Y, Z) = 0$ where $F_i(X, Y, Z) = Z^{d_i} f_i\left(\frac{X}{Z}, \frac{Y}{Z}\right)$. In particular $\alpha = \frac{\theta}{\lambda}$ and $\beta = \frac{\tau}{\lambda}$ will be a solution to the original system

$$\begin{aligned} f_1(x, y) &= 0 \\ f_2(x, y) &= 0. \end{aligned}$$

(Note that those factors with $\lambda = 0$ correspond to solutions at infinity.)

In higher dimensions the linear equation

$$u_0 + u_1 x_1 + \cdots + u_n x_n = 0$$

represents a "variable" hyperplane, but the arguments above work the same way to show that $R(u_0, \ldots, u_n)$ factors into linear factors corresponding to the roots of (6) (really of (7)).

A potential problem occurs if the set of solutions to the original system is not a discrete set of points, but contains a component of dimension one or more. In that case, every hyperplane $u_0 + u_1 x_1 + \cdots + u_n x_n$ will intersect that component (in projective n-space) so that the system

$$\begin{aligned} f_1(x_1, \ldots, x_n) &= 0 \\ &\vdots \\ f_n(x_1, \ldots, x_n) &= 0 \\ u_0 + u_1 x_1 + \cdots + u_n x_n &= 0 \end{aligned}$$

will always have a solution. This means that the resultant $R(u_0, \ldots, u_n)$ will be identically zero and of no use. Note that this can occur even when the solution set in $\mathbb{C}^n$ is a discrete set of points if the equations have a positive dimensional set of solutions at infinity. Canny in his paper "Generalized Characteristic Polynomials" details a way to handle this problem.

Finally, in lieu of actually factoring $R(u_0, \ldots, u_n)$, we can use a modification of the above ideas, coupled to some numerical linear algebra techniques (specifically computing eigenvalues), to produce an effective algorithm for solving the original system (6). This is discussed in the next section.

4 Numerical Methods

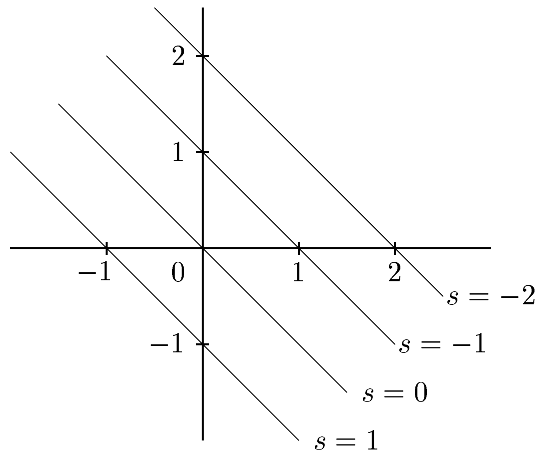
Currently there are two methods which are the primary candidates for use in solving systems of non-linear polynomial equations. The first is based on resultants as discussed above and uses numerical linear algebra techniques, specifically eigenvalue computations, to find the solution. This approach is discussed below. The other method is homotopy continuation. Homotopy methods are faster in practice for problems involving a very large number of equations and variables. Unfortunately, if the problem is singular in a certain sense, homotopy methods may run very slowly or even diverge. The resultant/eigenvalue method can handle small to medium sized problems and generally is more robust in the singular cases.

The approach begins with a version of the method discussed in section 3 above. We augment the system (6) of n equations in n variables by adding a linear equation of the form

$$s + a_1 x_1 + \cdots + a_n x_n$$

where $a_1, \ldots, a_n$ are specified constants and only s (and $x_1, \ldots, x_n$ of course) is an unknown.

As s varies, we are sweeping out a family of parallel hyperplanes that fill n-space. For example when $n = 2$ the lines $s + x + y = 0$ look like:



Again, the resultant $R(s)$, viewed as a polynomial in s, will have a root $s = s_0$ precisely when the hyperplane

$$s_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0$$

passes through one of the points corresponding to a solution of the original system (6) (actually (7)). If we do this for n independent systems of hyperplanes (by choosing the a_i so that the $n \times n$ matrix A, formed by the coefficients a_1 to a_n for each of the n systems is invertible), then we can solve for the roots by computing

$$A^{-1}\begin{pmatrix} s_0^{(1)} \\ \vdots \\ s_0^{(n)} \end{pmatrix}$$

where $s_0^{(i)}$ is a root of the ith resultant $R^{(i)}(s)$ obtained by augmenting (6) with the i^{th} linear equation which includes s as a variable in the constant term (with respect to $x_1, \ldots, x_n$).

Example 8: Augmenting by the linear form $s - x_1$ will yield a resultant $R(s)$ whose roots are the x_1 coordinates of the common solutions to $f_1 = 0, \ldots, f_n = 0$.

Now the (sparse) resultant $R(s)$ of the system

$$\begin{aligned} f_1(x_1, \ldots, x_n) &= 0 \\ &\vdots \\ f_n(x_1, \ldots, x_n) &= 0 \\ s + a_1x_1 + \cdots + a_nx_n &= 0 \end{aligned}$$

is computed using the determinant of a matrix which can be put in block form:

$$\begin{bmatrix} R_{11} & R_{12} \\ R_{21}(s) & R_{22}(s) \end{bmatrix}.$$

Here $R_{21}(s)$ and $R_{22}(s)$ have entries that are linear in s. Using elementary row operations we can get

$$\begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22}(s) - R_{21}(s)R_{11}^{-1}R_{12} \end{bmatrix}$$

which is a matrix whose determinant (up to the constant factor $\det R_{11}$) is $\det M(s)$ where $M(s) = R_{22}(s) - R_{21}(s) R_{11}^{-1} R_{12}$ has entries that are linear in s. Thus $M(s) = sM_1 + M_0$. Finally multiplying by M_1^{-1}, we get a matrix

$$sI + M_0 M_1^{-1}$$

whose determinant has roots in s that are the same as the roots of the resultant $R(s)$. But the roots of

$$\det(sI + M_0 M_1^{-1})$$

are just the eigenvalues of $M_0 M_1^{-1}$ and these can be computed by well known methods.

Note when using the sparse resultant, one can define the matrix used to compute it in such a way that M_1 will be the identity. Thus inversion will not be required. Moreover, if the linear form $s + a_1 x_1 + \cdots + a_n x_n$ has been selected so that it never goes through two solutions of (6) (which is the case with probability one), then as long as the common solutions to (6) have multiplicity one, the eigenvalues will have multiplicity one, and the roots can be computed from the eigenvector. One thus avoids doing the computation n times, at the expense of computing eigenvectors.