POLYNOMIALS

Polynomials of one or more variables are likely to be familiar to most readers. Expressions such as

$$3t^2 - 7t + 2 \quad \text{or} \quad x^2 + y^2 - 1$$

are easily recalled from high school mathematics. In general, polynomials involve some number n of variables, call them $x_1, \ldots, x_n$, and a set of allowable coefficients usually taken to lie in particular field or ring. Common fields are the field of rational numbers $\mathbb{Q}$, the field of real numbers $\mathbb{R}$, or the field of complex numbers $\mathbb{C}$. The ring of ordinary integers $\mathbb{Z}$ provides an example of a coefficient ring that is not a field.

A monomial in the variables $x_1, \ldots, x_n$ is a power product of the form

$$x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n}$$

where the exponents $\alpha_1, \ldots, \alpha_n$ are nonnegative integers. The total degree of the monomial is the sum $\alpha_1, + \cdots + \alpha_n$.

Because of the number of variables involved, a shorthand notation is needed. We let $\alpha = (\alpha_1, \ldots, \alpha_n)$ be an n-tuple of nonnegative integers, and we define

$$x^{\alpha} = x_1^{\alpha_1} \ldots x_n^{\alpha_n}$$

where x represents $(x_1, \ldots, x_n)$. The total degree is the denoted by $|\alpha| = \alpha_1 + \cdots + \alpha_n$.

A polynomial $f(x_1, \ldots, x_n)$ in the variables $x_1, \ldots, x_n$ with coefficients in a field K (or ring R) is a finite sum of terms of the form

$$f(x_1, \ldots, x_n) = \sum_{\alpha} a_{\alpha} x^{\alpha} = \sum a_{\alpha_1, \ldots, \alpha_n} x_1^{\alpha_1} \ldots x_n^{\alpha_n}$$

where $a_\alpha \in K$ (or R). The set of all such polynomials is written $K[x_1, \ldots, x_n]$. We call a_α the coefficient of the monomial x^α and call $a_\alpha x^\alpha$ a term in the polynomial when $a_\alpha \neq 0$.

The total degree (or just degree) of $f(x_1, \ldots, x_n)$, denoted deg f, is the maximum of the degrees $|\alpha|$ of the monomials that occur in the terms of f, that is, the maximum over the $|\alpha| = \alpha_1 + \cdots + \alpha_n$ such that a_α is not zero.

A polynomial is said to be homogeneous of degree d if every monomial occurring in a term of f has degree equal d. Thus $y^3 + x^2y + zw^2$ is homogeneous of degree 3 in four variables, whereas $x^3y + 3xwz$ is of degree 4 in four variables but is not homogeneous.

One central problem that frequently arises is the need to solve a system of m polynomial equations in n variables:

$$\begin{gathered} f_1(x_1, \ldots, x_n) = 0 \\ \vdots \\ f_m(x_1, \ldots, x_n) = 0 \end{gathered}$$

where the f_i are in $K[x_1, \ldots, x_n]$. Solutions are sought in K^n or in some larger field E^n where $K \subset E$. (The example to keep in mind is finding complex solutions to equations with real coefficients.) K^n in this case is just the set of n-tuples of elements of K, which we call n-space:

$$K^n = \{(a_1, \ldots, a_n) \text{ with } a_i \in K\}$$

We say that $(a_1, \ldots, a_n) \in K^n$ is a solution to the system above if $f_i(a_1, \ldots, a_n) = 0$ for all $i = 1, \ldots, n$.

Naively, we expect that a system of n equations in n variables will have a finite number of solutions. This, however, need not be the case. Consider three equations in three variables (coefficients in $\mathbb{R}$ say):

$$\begin{gathered} f_1(x, y, z) = 0 \\ f_2(x, y, z) = 0 \\ f_3(x, y, z) = 0 \end{gathered}$$

Each represents a surface in three-space. If those surfaces should all contain a common curve, then the set of solutions to the system would be infinite.

For example, the system

$$\begin{aligned} x^2 + y^2 + z^2 - 1 &= 0 \\ x^2 + y^2 + \frac{z^2}{4} - 1 &= 0 \\ x^2 + y^2 - 1 &= 0 \end{aligned}$$

has as solutions the unit circle in the (x, y) plane, that is, all points $(a, b, 0)$ where $a^2 + b^2 = 1$.

Numerical methods to solve systems of polynomial equations (when those systems have isolated point solutions) are known and discussed elsewhere. Here we take up some perhaps less well known techniques for dealing with and understanding systems of polynomial equations. Later we discuss an important use of what are called invariant polynomials in image understanding applications.

OVERVIEW OF RESULTANTS

Resultants are used to solve systems of polynomial equations, to determine whether or not solutions exist, or to reduce a given system to one with fewer variables and/or fewer equations.

Input

The typical input will be a system of m equations in n variables:

$$\begin{gathered} f_1(x_1, \ldots, x_n) = 0 \\ \vdots \\ f_m(x_1, \ldots, x_n) = 0 \end{gathered} \tag{1}$$

Each equation has an associated degree $d_i \geq 1$. Recall that $f_i(x_1, \ldots, x_n)$ has degree d_i if all monomials $x_1^{e_1} x_2^{e_2} \cdots x_n^{e_n}$ appearing in f_i have $\sum_{i=1}^n e_i \leq d_i$ and at least one monomial has $\sum_{i=1}^n e_i = d_i$. As an example, $f(x_1, x_2, x_3) = 3x_1^2x_3 + 4x_1x_2 - x_2 + 7x_3 - 1$ has degree $d = 3$. The integers $m, n, d_1, \ldots, d_m$ are important indicators of the specific techniques that will need to be employed.

Output

There are two essentially different cases.

Case 1: $m > n$ (overdetermined). This is the case where we have more equations than unknowns and where we generally expect to have no solutions. The resultant will be a system of equations (one equation when $m = n + 1$) in the symbolic coefficients of the f_i that has the following property: when we substitute the specific numerical coefficients of the f_i, we will get zero in every equation in the resultant system if and only if the original overdetermined system has a solution.

Case 2: $m \leq n$ (exact and underdetermined). In this case the number of equations is less than or equal to the number of variables, and we expect to have solutions. In fact, if we allow complex solutions and solutions at infinity, we are guaranteed to have solutions.

Of course, only when $m = n$ do we expect a finite number s of solutions. Bezout's theorem then provides a count of $s = d_1d_2 \cdots d_m$ solutions (counting complex solutions, solutions at infinity, and counting with appropriate multiplicities). Unfortunately, as mentioned, the possibility also exists (even when $m = n$) that there will be an infinite number of solutions.

In general, for $m \leq n$, the resultant will be one equation in $n - m + 1$ of the variables. In effect, the resultant eliminates $m - 1$ of the variables. For example, if we choose to eliminate $x_{n-m+2}, \ldots, x_n$, then the resultant R will be a polynomial $R(x_1, \ldots, x_{n-m+1})$ in the remaining variables. If $(a_1, \ldots, a_{n-m+1})$ is a solution to $R = 0$, then there will exist values $a_{n-m+2}, \ldots, a_n$ such that $(a_1, a_{n-m+1}, a_{n-m+2}, \ldots, a_n)$ is a solution to the original system. [One must be a little careful here. The system should be modified to make it homogeneous with respect to $x_{n-m+2}, \ldots, x_n$ by adding appropriate powers of a variable w. The values $a_{n-m+2}, \ldots, a_n$ should be regarded as the coordinates of a point $(a_{n-m+2} : \cdots : a_n : 1)$ in projective $(m - 1)$-space $\mathbb{P}^{m-1}$. We must allow for the possibility that this point will be at infinity where $w = 0$. In that case, a solution to $R = 0$ would not necessarily give rise to a solution of the original system.]

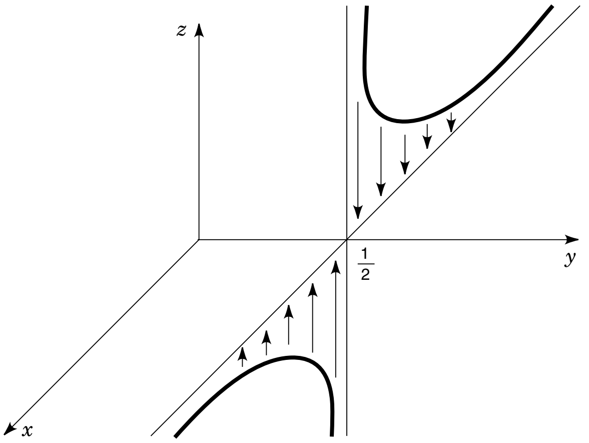

Figure 1. Resultant as projection.

For example, looking at Fig. 1, consider the system of $m = 2$ equations in $n = 3$ variables: $4xyz - 1 = 0$ and $y + xz - 1 = 0$. The resultant eliminating z is $R(x, y) = x(4y^2 - 4y + 1)$. When $x = 0$ we will have $R = 0$, but clearly our system has no solution when $x = 0$. However, homogenizing with respect to z gives the system

$$4xyz - w = 0 \quad \text{and} \quad (y-1)w + xz = 0$$

Now when we look at the condition $x = 0$, we find that $(z:w) = (1:0)$ is a solution. This is a point at infinity.

Notice that we also have solutions to $R = 0$ when $x \neq 0$ by taking $y = \frac{1}{2}$. This yields $z = 1/2x$. Geometrically the solution set is a hyperbola in the plane $y = \frac{1}{2}$ in space. The resultant "projects" that hyperbola to the line $y = \frac{1}{2}$ in the (x, y) plane, except that $(x, y) = (0, \frac{1}{2})$ is not hit.

In this context (the underdetermined case) the resultant can be viewed as a projection of the nominally $(n - m)$-dimensional locus of solutions in $\mathbb{R}^n$ to an $(n - m)$-dimensional locus (hypersurface) in $\mathbb{R}^{n-m+1}$. Note that in our example $n = 3$, $m = 2$, and we are projecting the one-dimensional locus of solutions in $\mathbb{R}^3$ to a one-dimensional locus in $\mathbb{R}^2$ which is described by one equation $y - \frac{1}{2} = 0$.

Approach in This Article

We begin with the first major distinction in methods, namely the one based on the number of variables n. The case $n = 1$ of a single variable is discussed first. We then move on to the multivariate case $n \geq 2$. See Table 1 (1).

RESULTANTS OF POLYNOMIALS IN ONE VARIABLE

The Basic Case: Two Polynomials and the Sylvester Matrix

Given two positive integers r, $s \geq 1$ and two polynomials in one variable

$$f(x) = a_r x^r + \cdots + a_1 x + a_0 \quad \text{and} \quad g(x) = b_s x^s + \cdots + b_1 x + b_0$$

of degree less than or equal to r and s, respectively, we define their resultant $R_{r,s}(f, g)$ by Sylvester's formula:

$$R_{r,s}(f,g) = \det \begin{pmatrix} a_0 & a_1 & & \cdots & a_r & 0 & \cdots & 0 \\ 0 & a_0 & a_1 & \cdots & \cdots & a_r & 0 \cdots & 0 \\ \vdots & & & & & & & \\ 0 & 0 & \cdots & 0 & a_0 & a_1 & \cdots & a_r \\ b_0 & b_1 & \cdots & b_{s-1} & b_s & 0 & \cdots & 0 \\ 0 & b_0 & b_1 & \cdots & & b_s & \cdots & 0 \\ \vdots & & & & & & & \\ 0 & 0 & \cdots & 0 & b_0 & b_1 & \cdots & b_s \end{pmatrix}$$

which is the determinant of an $r + s$ by $r + s$ matrix with s rows involving the a's and r rows involving the b's.

Example 1

$$\begin{aligned} R_{2,2}(a_2x^2 + a_1x + a_0, b_2x^2 + b_1x + b_0) \\ = a_0^2b_2^2 + a_0a_2b_1^2 - a_0a_1b_1b_2 + a_1^2b_0b_2 \\ - a_1a_2b_0b_1 + a_2^2b_0^2 - 2a_0a_2b_0b_2 \end{aligned}$$

Note that in this example each monomial in the resultant has total degree $r + s = 4$ and is bihomogeneous of bidegree $(s, r) = (2, 2)$ in the a's and b's respectively. This is true in general.

Basic Properties of the Resultant $R_{r,s}(f, g)$

1. *Relationship to common roots.*

   $$R_{r,s}(f,g) = a_r^s b_s^r \prod_{i,j}(x_i - y_j)$$

   where $x_1, \ldots, x_r$ are the roots of f and $y_1, \ldots, y_s$ are the roots of g. (Here we are assuming $a_r \neq 0$ and $b_s \neq 0$.) Thus $R_{r,s}(f, g)$ will be zero if and only if f and g have a root in common.
2. *Irreducibility.* $R_{r,s}(f, g) \in \mathbb{Z}[a_0, \ldots, a_r, b_0, \ldots, b_s]$ is irreducible, that is, the resultant is an irreducible polynomial with integer ($\mathbb{Z}$) coefficients in $(r + 1)(s + 1) = rs + s + r + 1$ variables.
3. *Symmetry.* $R_{r,s}(f, g) = (-1)^{rs}R_{r,s}(g, f)$
4. *Factorization.* $R_{r_1+r_2,s}(f_1f_2, g) = R_{r_1,s}(f_1, g)R_{r_2,s}(f_2, g)$.

Discriminants and Resultants

The discriminant $\Delta(f)$ of a polynomial $f = a_rx^r + \cdots + a_0$, $a_r \neq 0$, is essentially the resultant of f and its derivative f'. The exact relationship is

$$\Delta(f) = \frac{1}{a_r} R_{r,r-1}(f, f')$$

which is a homogeneous polynomial of degree $2r - 2$ in the $r + 1$ variables $a_0, \ldots, a_r$. It provides a test for multiple roots.

Just as the discriminant can be defined in terms of the resultant, the resultant can be defined in terms of the discriminant:

$$[R_{r,s}(f,g)]^2 = (-1)^{rs} \frac{\Delta(fg)}{\Delta(f)\Delta(g)}$$

when $a_r \neq 0$ and $b_s \neq 0$.

Table 1. Table of Resultants

n	m	Type of Resultant to Use	Notes
1	2	Determinant of the Sylvester matrix	This is what is most commonly thought of as *the* resultant.
1	≥ 3	Requires a system of equations	See the discussion in van der Waerden (1).
≥ 2	$m = n + 1$	Macaulay resultant	This is computed as the quotient of two determinants. It is a polynomial in the symbolic coefficients and is zero if and only if the system has a solution.
≥ 2	$m \geq n + 2$	Requires a system of equations	See van der Waerden (1).
≥ 2	$m = n$	U resultant or generalized characteristic polynomial	This resultant is designed to find the finite set of all solutions to the system of equations.
≥ 2	$m < n$	Macaulay resultant using $m - 1$ variables, while treating the other $n - m + 1$ variables as included in the coefficients	The result is a single polynomial in the remaining $n - m + 1$ variables.

Note: One can also employ the standard Sylvester resultant in the multivariate case, using it iteratively to successively eliminate variables. For example, with three equations in three unknowns $f(x, y, z) = 0$, $g(x, y, z) = 0$, and $h(x, y, z) = 0$, we can take the resultant of f and g treating z as the only variable to get $R_1(x, y)$. Likewise we can take the resultant of g and h again treating z as the only variable to get $R_2(x, y)$. Finally, the resultant of R_1 and R_2 with y as the variable yields $R(x)$, whose roots can then be found using standard root-finding methods.

Finding the Common Roots: Subresultants

Again, suppose we are given two polynomials in a single variable x, say

$$f(x) = a_r x^r + \cdots + a_1 + a_0 \quad \text{and} \quad g(x) = b_s x^s + \cdots + b_1 x + b_0$$

of degrees $r \geq 1$ and $s \geq 1$, respectively. (We assume that $a_r \neq 0$ and $b_s \neq 0$.) As we saw previously, the resultant $R_{r,s}(f, g)$ of f and g will be zero if and only if f and g have a common root. Two questions immediately occur.

> *Question 1.* Suppose $R_{r,s}(f, g) = 0$, so that f and g have at least one common root. Can we determine how many roots they have in common? This is the same as asking for the degree $1 \leq d \leq \min(r, s)$ of the greatest common divisor $h(x)$ of $f(x)$ and $g(x)$.
>
> *Question 2.* Can we find the common roots?

The answer to Question 2 is more subtle. In general, we cannot expect to be able to express the common roots of f and g (assuming they have a root or roots in common) as rational expressions in the coefficients $a_r, \ldots, a_0, b_s, \ldots, b_0$. For example, if f and g have rational coefficients, that is, $a_r, \ldots, a_0, b_s, \ldots, b_0 \in \mathbb{Q}$, the field of rational numbers, then any polynomial expression in the coefficients would be a rational number. But polynomials with rational coefficients can have common roots that are not rational.

Example 2. $f(x) = 3x^4 + x^3 + 4x^2 + x + 1 = (x^2 + 1)(3x^2 + x + 1)$ and $g(x) = x^2 - 1 = (x^2 + 1)(x^2 - 1)$ both have rational coefficients, but the common roots $\pm i$ are not rational numbers.

We can however answer Question 2 in a special case. If $R_{r,s}(f, g) = 0$ and at least one partial derivative of the resultant computed symbolically

$$\frac{\partial R}{\partial a_0}, \ldots, \frac{\partial R}{\partial a_r}, \frac{\partial R}{\partial b_0}, \ldots, \frac{\partial R}{\partial b_s} \tag{2}$$

is nonzero when the coefficients of f and g are substituted, then f and g have exactly one common root α and it can be found via the proportions:

$$(1 : \alpha : \alpha^2 : \cdots : \alpha^r) = \left(\frac{\partial R}{\partial a_0}(f,g) : \frac{\partial R}{\partial \alpha_1}(f,g) : \cdots : \frac{\partial R}{\partial \alpha_r}(f,g)\right)$$

$$(1 : \alpha : \alpha^2 : \cdots : \alpha^s) = \left(\frac{\partial R}{\partial b_0}(f,g) : \frac{\partial R}{\partial b_1}(f,g) : \cdots : \frac{\partial R}{\partial b_s}(f,g)\right)$$

In particular the common root α can be computed as

$$\alpha = \frac{\dfrac{\partial R}{\partial a_1}(f,g)}{\dfrac{\partial R}{\partial a_0}(f,g)} = \frac{\dfrac{\partial R}{\partial b_1}(f,g)}{\dfrac{\partial R}{\partial b_0}(f,g)}$$

This result also has a geometric interpretation. The space of all pairs of polynomials (f, g) where the degree of f is less than or equal to r and the degree of g is less than or equal to s can be identified with $\mathbb{R}^{r+s+2}$ having coordinates $(a_r, \ldots, a_0, b_s, \ldots, b_0)$. The symbolic resultant R is a polynomial in these variables, and the locus $R = 0$ in $\mathbb{R}^{r+s+2}$ is an irreducible hypersurface (of dimension $r + s + 1$) consisting of pairs (f, g) with a root in common. A point on this hypersurface where at least one of the partial derivatives in Eq. (2) is nonzero is a smooth point. At such points we have exactly one common root. Moreover, that root can be expressed as a quotient of polynomial expressions in $a_r, \ldots, a_0, b_s, \ldots, b_0$. We remind the reader that "most" points on the locus $R = 0$ are smooth points. Those that are not are called singular points and they occur in dimension $r + s$ or less.

RESULTANT METHODS FOR SYSTEMS OF POLYNOMIAL EQUATIONS IN SEVERAL VARIABLES

Theory

The linear algebra techniques discussed next can be used to solve systems of polynomial equations in several variables. If there are only two equations, then the Sylvester technique (discussed earlier) can be employed by treating all but one variable as part of the coefficients. However, when the number of equations exceeds two, the Sylvester approach can be misleading. For example, taking the equations two at a time using the Sylvester determinant can lead the user to the conclusion that there is a common solution, when in fact there are no common solutions for the system of equations taken as a whole.

What it means to "solve" a given set of polynomial equations depends upon the number of variables and the number of equations. Assuming the equations are inhomogeneous, let n be the number of variables and m be the number of equations. The expected dimensionality of the set of solutions is $n - m$ when viewed over the complex numbers. For example, if there are three equations ($m = 3$) and five variables ($n = 5$), then the space of solutions is expected to have dimension $n - m = 5 - 3 = 2$. Geometrically, the set of solutions forms a surface. Sometimes, however, components of excess dimension occur in the set of solutions. These are geometric loci of higher dimension than the expected dimension. They occur because, in a very loose sense, the equations have certain dependencies.

Finally, a note is given about homogeneous equations. Recall that a set of polynomial equations is considered homogeneous if in each equation all the terms have the same degree. If this is not the case, even for only one of the equations, the set is regarded as inhomogeneous. For systems of homogeneous equations the number n of variables should be taken as one less than the actual number of variables when computing expected dimensions. This is because we want to regard the solutions as lying in an $(n - 1)$-dimensional projective space.

The Macaulay Resultant, the *U* Resultant, and the Generalized Characteristic Polynomial

The Macaulay resultant is the ratio of two determinants formed from the coefficients of the given polynomials in a manner to be described later in this section. If the number of equations exceeds the number of variables by one ($n - m = -1$), then the Macaulay resultant tests whether or not a common solution exists. [For systems of homogeneous equations in which the number of equations equals the number of variables, the expected dimension is still -1, and the Macaulay resultant tests for a nontrivial common solution, that is, a solution other than $(0, . . ., 0)$]

If there are as many inhomogeneous equations as unknowns ($n - m = 0$), then the equations can often be solved by adding the U equation (explained later in this section) to the homogenized set and forming the Macaulay resultant. The Macaulay resultant is then called the U resultant.

In some cases, however, there will be a component of excess dimension (≥ 1) which masks some or all of the desired solutions. In this case Canny's generalized characteristic polynomial (GCP) approach is useful (see Ref. 2).

In order to illustrate the various methods, the following system of three polynomial equations will be used:

$$f_1 = y - 3x + 5 = 0$$
$$f_2 = x^2 + y^2 - 5 = 0$$
$$f_3 = y - x^3 + 3x^2 - 3x + 1 = 0$$

Here we have three inhomogeneous equations in two variables ($n - m = 2 - 3 = -1$). The multiresultant techniques described below can be used to test for the existence of a solution.

Step 1: Homogenization

The equations must first be homogenized. This is done by adding a third variable z. Specifically x is replaced by x/z and y is replaced by y/z, and the factors of z are cleared from the denominators. In the previous example this leads to three equations:

$$f_1 = y - 3x + 5z = 0$$
$$f_2 = x^2 + y^2 - 5z^2 = 0$$
$$f_3 = yz^2 - x^3 + 3x^2z - 3xz^2 + z^3 = 0$$

This is the homogenized version of the original system.

Step 2: Degree Determination

Each of the multiresultants being considered involves the coefficients of various monomials that appear in the equations. The variables involved in the monomials are the variables that appear in the homogeneous form of the polynomial equations. For example, the homogeneous polynomial equations above have the variables x, y, and z. All the monomials in a given equation are constrained to have the same degree because we have homogenized. The "overall degree" of the system is determined from the degrees of the individual homogeneous equations by the following rule:

$$d = 1 + \sum_{i=1}^{m} (d_i - 1)$$

where m is the number of equations and d_i the degree of the ith equation.

For the homogeneous polynomials given previously (f_1, f_2, and f_3) the degrees are

Equation	Degree
f_1	$d_1 = 1$
f_2	$d_2 = 2$
f_3	$d_3 = 3$

Therefore,

$$d = 1 + (1 - 1) + (2 - 1) + (3 - 1) = 4$$

Step 3: Matrix Size Determination

Each of the multiresultants to be discussed involves the ratio of two determinants. The numerator is the determinant of a matrix, the formation of which will be discussed in subsequent sections. The denominator determinant is formed from a submatrix of the numerator matrix.

The number of variables in the inhomogeneous equations is n. Since one additional variable has to be added to homogenize the equations, the number of variables in the homogeneous equations is $n + 1$. The size of the numerator matrix equals the number of monomials in the $n + 1$ variables that have overall degree d (discussed in the previous step).

$$\text{Numerator matrix size} = \binom{n+d}{d}$$

For the three polynomial equations (f_1, f_2, f_3) we have already calculated that $d = 4$. Since the original set of inhomogeneous variables consisted of x and y, we have that n equals 2. Thus for our example,

$$\text{Numerator matrix size} = \binom{2+4}{4} = \binom{6}{4} = \frac{6!}{(2!)(4!)} = 15$$

that is, it is a 15 × 15 matrix.

Step 4: Determining "Big" versus "Small" Exponents

A few of the 15 monomials involving the variables x, y, and z with an overall degree of 4 include:

$$yz^3 \quad \text{and} \quad x^2y^2$$

In the next section we will discuss whether certain of these monomials are reduced. This will be determined by whether the exponents are "big" or "small." In this section we discuss how bigness is defined.

Each variable will be associated with a particular equation. For example, the first variable, x, will be associated with the first equation, f_1. The second variable, y, will be associated with the second equation, f_2, etc. The degrees of the associated equations define bigness for the exponents of that variable. Specifically, since d_1 (the degree of f_1) is 1, if the exponent of x is greater than or equal to 1, it is considered big. Since $d_2 = 2$, whenever the exponent of y is greater than or equal to 2, it is considered big. The degree of f_3 is 3, therefore, whenever the exponent of z is greater than or equal to 3, it is considered big.

For example, consider the monomial yz^3. The exponent of y is 1. This is *less than* d_2, and is considered small. The exponent of z is 3. This is *equal to* d_3, and is therefore big. On the other hand, consider the monomial x^2y^2. The exponent of x is 2. This is *greater than* d_1 and is *big*. The exponent of y is 2. This is *equal to* d_2 and is *big*.

Step 5: Determining the Reduced Monomials

If for a particular monomial of degree d the exponent of *only one* variable is big, the monomial is said to be reduced. In the previous step the monomial yz^3 is reduced. For that monomial only the exponent of z is big, whereas for x^2y^2, both the exponent of x and the exponent of y are big. Thus the monomial x^2y^2 is not reduced.

Step 6: Creating the *A* Matrix

The Macaulay resultant is the ratio of two determinants. The numerator is the determinant of a matrix that we will call the A matrix. The denominator is the determinant of a matrix that we will call the M matrix

$$R = \frac{\det|A|}{\det|M|}$$

We have discussed above how the size of the A matrix is determined. In this section we will show how the matrix entries are obtained.

Each row and column of the matrix should be thought of as being labeled by one of the monomials of degree d. This labeling can be done in any desired order. Recall that for f_1, f_2, and f_3 in our example there were 15 possible monomials of degree 4 in x, y, z, and therefore the A matrix would be 15×15.

There are three rules for determining the elements of the A matrix. After presenting the rules, the example involving f_1, f_2, and f_3, will be used to illustrate the process. The reader may find it helpful to read the example simultaneously with the rules.

Rules for inputting the elements of each column of the A matrix:

1. Search the monomial labeling that column from left to right for the *first* variable with a big exponent. Such a variable must exist. Call it the marker variable.
2. Form a new polynomial from the polynomial associated with this marker variable by multiplying the associated polynomial by the monomial and dividing by the marker variable raised to the degree of the associated polynomial.
3. The coefficients of the new polynomial are the elements of the columns. Each coefficient goes in the row labeled by the monomial it multiples. All the other rows get zeros.

Example 3. Recall that for the system of equations f_1, f_2, f_3 there are 15 monomials of degree 4 that can be formed from x, y, and z. Two of these were considered above, namely yz^3 and x^2y^2.

- For the column labeled by yz^3:
  1. The first variable with a big exponent is z, so z is the marker variable.
  2. The polynomial associated with z is f_3. Multiply f_3 by the monomial yz^3, and divide this product by z^3.

$$\begin{aligned}\frac{f_3(yz^3)}{z^3} &= \frac{(yz^2 - x^3 + 3x^2z - 3xz^2 + z^3)(yz^3)}{z^3} \\ &= y^2z^2 - x^3y + 3x^2yz - 3xyz^2 + yz^3\end{aligned}$$

  3. The coefficient of y^2z^2 is $+1$. Therefore the element of the row labeled y^2z^2 is $+1$. The coefficient of x^3y is -1. Therefore the element of the row labeled x^3y is -1. The coefficient of x^2yz is $+3$. Therefore the element of the row labeled x^2yz is $+3$. The coefficient of xyz^2 is -3. Therefore the element of the row labeled xyz^2 is -3. The coefficient of yz^3 is $+1$. Therefore the element of the row labeled yz^3 is $+1$. All other entries in the column are zero.
- For the column labeled by x^2y^2:
  1. The first variable with a big exponent is x, so x is the marker variable.
  2. The polynomial associated with x is f_1. Multiply f_1 by the monomial x^2y^2, and divide this product by x:

$$\frac{f_1(x^2y^2)}{x} = \frac{(y - 3x + 5z)(x^2y^2)}{x} = xy^3 - 3x^2y^2 + 5xy^2z$$

  3. The coefficient of xy^3 is $+1$. Therefore the element of the row labeled xy^3 is $+1$. The coefficient of x^2y^2 is -3. Therefore the element of the row labeled x^2y^2 is -3. The coefficient of xy^2z is $+5$. Therefore the element of the row labeled xy^2z is $+5$.

When all the columns are determined, the A matrix in our example takes the form:

$$\begin{array}{l|ccccccccccccccc}
 & x^4 & x^3y & x^3z & x^2y^2 & x^2yz & x^2z^2 & xy^3 & xy^2z & xyz^2 & xz^3 & y^4 & y^3z & y^2z^2 & yz^3 & z^4 \\
\hline
x^4 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
x^3y & 1 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\
x^3z & 5 & 0 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\
x^2y^2 & 0 & 1 & 0 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
x^2yz & 0 & 5 & 1 & 0 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 3 & 0 \\
x^2z^2 & 0 & 0 & 5 & 0 & 0 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 3 \\
xy^3 & 0 & 0 & 0 & 1 & 0 & 0 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
xy^2z & 0 & 0 & 0 & 5 & 1 & 0 & 0 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
xyz^2 & 0 & 0 & 0 & 0 & 5 & 1 & 0 & 0 & -3 & 0 & 0 & 0 & 0 & -3 & 0 \\
xz^3 & 0 & 0 & 0 & 0 & 0 & 5 & 0 & 0 & 0 & -3 & 0 & 0 & 0 & 0 & -3 \\
y^4 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
y^3z & 0 & 0 & 0 & 0 & 0 & 0 & 5 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
y^2z^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 & 1 & 0 & -5 & 0 & 1 & 1 & 0 \\
yz^3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 & 1 & 0 & -5 & 0 & 1 & 1 \\
z^4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 & 0 & 0 & -5 & 0 & 1
\end{array}$$

The determinant of the above A matrix is zero. If the determinant of the M matrix is nonzero, this would imply that the system has a solution.

Step 7: Creating the M Matrix

The denominator of the Macaulay resultant is the determinant of the M matrix. The M matrix is a submatrix of the A matrix. It consists of the elements that have row and column monomial labels which are *not reduced.* Recall that a monomial is not reduced if it has more than one variable with a big exponent.

The size of the M matrix equals the size of the A matrix minus D, where

$$D = \sum_{i=1}^{m} \prod_{i \neq j} d_j$$

In our example,

$$D = d_2d_3 + d_1d_3 + d_1d_2 = (2)(3) + (1)(3) + (1)(2) = 11$$

so that the size of the M matrix is $15 - 11 = 4$. The actual M matrix for f_1, f_2, and f_3 is

$$\begin{array}{r|cccc}
 & x^2y^2 & xy^3 & xy^2z & xz^3 \\
\hline
x^2y^2 & -3 & 0 & 0 & 0 \\
xy^3 & 1 & -3 & 0 & 0 \\
xy^2z & 5 & 0 & -3 & 0 \\
xz^3 & 0 & 0 & 0 & -3
\end{array}$$

The determinant of this M matrix yields a value of 81. Since the determinant of the A matrix was zero, the Macaulay resultant is zero, which implies that there is a solution to our system. The following plot of the three polynomials (f_1, f_2, and f_3) confirms that there is a common point at $x = 2$ and $y = 1$ (see Fig. 2).

Sometimes both the A matrix and the M matrix have zero determinants. This indeterminacy can often be circumvented if the polynomials are first written with symbolic coefficients. The determinants of the A and M matrices are obtained, polynomial division is performed, and then at the end, the symbolic coefficients are replaced by their numerical values to check if the resultant is zero. Since one does not know ahead of time whether or not this "division by zero" condition will arise, the symbolic coefficient approach is the best strategy. It is also often sufficient to treat just a subset of the coefficients symbolically—sometimes as few as a single symbolic coefficient will remove the indeterminacy.

The U Resultant

For problems with as many inhomogeneous equations as variables, the U resultant can often be used to solve for the point solutions. The three polynomial equations f_1, f_2, f_3, do not satisfy these conditions, since there are three equations in two

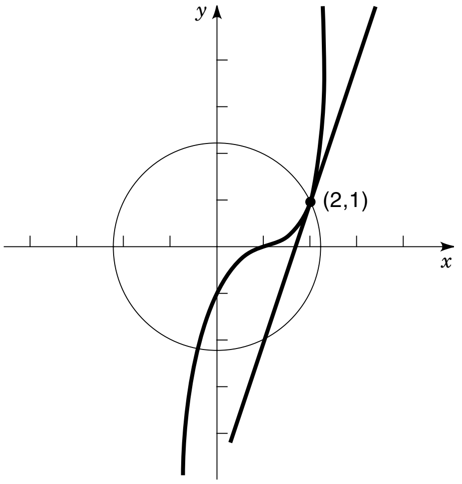

Figure 2. Common solution in our example.

inhomogeneous variables, x and y. However, if we take just the first two equations, namely f_1 and f_2, we would have a system with as many equations as variables.

The given equations must first be homogenized. This adds one additional variable. We then add one additional equation to the system. This equation is called the U equation. If x and y are the given variables and z is the homogenizing variable, then the U equation takes the form

$$u_1x + u_2y + u_3z = 0$$

The Macaulay resultant R is then computed for these $m + 1$ equations, treating the u_i as symbolic coefficients. The result is called the U resultant. Notice that R will be a polynomial in the u_i's and the coefficients of the original equations.

After R is determined, it is factored into linear factors. For each linear factor there is a point solution of the original system of equations. The coordinates of each solution are given as ratios of the coefficients of the u_i's. The denominator is always the coefficient of the u_i associated with the homogenizing variable. In our example this is the coefficient of u_3. Thus

$$x = \frac{\text{coefficient of } u_1}{\text{coefficient of } u_3} \quad \text{and} \quad y = \frac{\text{coefficient of } u_2}{\text{coefficient of } u_3}$$

For example, if a linear factor turned out to be

$$u_1 - u_2 - u_3$$

then the coordinates of the associated solution would be

$$x = \frac{+1}{-1} = -1 \quad \text{and} \quad y = \frac{-1}{-1} = +1$$

Example 4. As mentioned above, the polynomial equation system f_1, f_2, f_3 is overdetermined ($n - m = -1$). However, we can use the U resultant to solve f_1 and f_2 for x and y ($n - m = 0$). In this example, we will also demonstrate the symbolic approach alluded to in the previous section. Recall that the homogenized form of f_1 and f_2 is

$$f_1 = y - 3x + 5z = 0$$
$$f_2 = x^2 + y^2 - 5z^2 = 0$$

Rewriting these two equations with symbolic coefficients and including the U equation yields

$$f_1 = a_1x + b_1y + c_1z = 0$$
$$f_2 = a_2x^2 + b_2y^2 + c_2z^2 = 0$$
$$U = u_1x + u_xy + u_3z = 0$$

where $a_1 = -3$, $b_1 = 1$, $c_1 = 5$, $a_2 = 1$, $b_2 = 1$, and $c_2 = -5$.

The U resultant is calculated in the same way as the Macaulay resultant, that is, with the A matrix and the M matrix, except now we are using symbolic coefficients.

$$\begin{array}{c|cccccc}
 & x^2 & xy & xz & y^2 & yz & z^2 \\
\hline
x^2 & a_1 & 0 & 0 & a_2 & 0 & 0 \\
xy & b_1 & a_1 & 0 & 0 & u_1 & 0 \\
xz & c_1 & 0 & a_1 & 0 & 0 & u_1 \\
y^2 & 0 & b_1 & 0 & b_2 & u_2 & 0 \\
yz & 0 & c_1 & b_1 & 0 & u_3 & u_2 \\
z^2 & 0 & 0 & c_1 & c_2 & 0 & u_3
\end{array}$$

The corresponding M matrix is a single element, namely a_1.

The determinant of M is divided into the determinant of A to obtain the U resultant. Finally, the symbolic coefficients are replaced by their numeric equivalents. (This could have been done from the outset, unless a_1 had been zero.) The result is

$$10(u_1 - 2u_2 + u_3)(2u_1 + u_2 + u_3)$$

This yields two solutions

Solution 1

$$x = \frac{\text{coefficient of } u_1}{\text{coefficient of } u_3} = \frac{+1}{-1} = -1 \quad \text{and}$$
$$y = \frac{\text{coefficient of } u_2}{\text{coefficient of } u_3} = \frac{-2}{+1} = -2$$

Solution 2

$$x = \frac{\text{coefficient of } u_1}{\text{coefficient of } u_3} = \frac{+2}{+1} = +2 \quad \text{and}$$
$$y = \frac{\text{coefficient of } u_2}{\text{coefficient of } u_3} = \frac{+1}{+1} = +1$$

We remark that the U resultant will be identically zero and give no information, if the set of common solutions contains a component of excess dimension one or more. Moreover, this component may be at infinity where the homogenizing variable is zero.

The Generalized Characteristic Polynomial Approach

The generalized characteristic polynomial (GCP) approach (2) avoids the problem of components of excess dimension in the set of solutions. It can be used together with the U resultant, which was discussed previously. If the U resultant leads to an indeterminant (0/0) form even when symbolic coefficients are used, an "excess" solution exists. The GCP takes the form

$$R = \frac{\det|A - sI|}{\det|M - sI|}$$

where A and M are the matrices defined earlier, s is a perturbation parameter, and I is the identity matrix.

One way to carry out the above operation is the following:

1. Set up the A matrix (as described previously). Subtract s along the diagonal. Evaluate the determinant. Retain the coefficient of the lowest surviving power of s.
2. Repeat step 1 for the M matrix.
3. Divide the result of step 1 by the result of step 2.

All of these multiresultant techniques have one characteristic in common. They require that there be one more equation than variable, $n - m = -1$. If there are as many equations as variables $n - m = 0$, the U equation is added and the effective situation is again $n - m = -1$. If there are more variables than equations ($n - m$ is a positive integer), then enough of these variables must be regarded as parameters in the coefficients so that effectively $n - m = -1$. Geometrically this amounts to projecting the locus of solutions to a hypersurface in a lower-dimensional space. Finally, if the number of equations exceeds the number of variables by more than one ($n - m \leq -2$), then some technique other than the multiresultant techniques noted earlier (e.g., a system of multiresultants) must be employed to determine if a solution exists.

INVARIANT POLYNOMIALS

In this section we will consider polynomials invariant under various transformation groups. Such polynomials have important applications in computer vision and image understanding. We will consider several specific cases rather than develop the general theory.

Affine Invariants of Point Sets in the Plane

Let $P_i = (x_i, y_i)$, $i = 1, \ldots, n$, be a set of n points in the plane $\mathbb{R}^2$. We will assume that these points are in general position, which means that no three are collinear. The group of affine transformations of the plane can be represented by a group of 3×3 matrices:

$$\mathrm{AFF}(2, \mathbb{R}) = \left\{ \begin{pmatrix} a & b & \xi_1 \\ c & d & \xi_2 \\ 0 & 1 & 1 \end{pmatrix} ; ad - bc \neq 0; a, b, c, d, \xi_1, \xi_2 \in \mathbb{R} \right\}$$

These affine transformations act on the plane by sending the point (x, y) to the point $(ax + by + \xi_1, cx + dy + \xi_2)$. In matrix terms this is

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} a & b & \xi_1 \\ c & d & \xi_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

AFF(2, $\mathbb{R}$) is a six-dimensional Lie group that is precisely the group of all transformations of $\mathbb{R}^2$ that preserve collinearity, that is, transform straight lines to straight lines.

The set of all ordered n-tuples of points in $\mathbb{R}^2$ is parametrized by $\mathbb{R}^2 \times \cdots \times \mathbb{R}^2 \cong \mathbb{R}^{2n}$ with coordinates $(x_1, y_1, x_2, y_2, \ldots, x_n, y_n)$. Those ordered n-tuples which are in general position form a dense open subset U of $\mathbb{R}^{2n}$.

The group AFF(2, $\mathbb{R}$) acts diagonally on $\mathbb{R}^{2n}$ and on U. We are interested in rational expressions that are invariant under the group action

$$\frac{p(x_1, y_1, x_2, y_2, \ldots, x_n, y_n)}{q(x_1, y_1, x_2, x_2, \ldots, x_n, y_n)}$$

Here p and q are polynomials with real coefficients. The invariant expressions will take the same value if (x_i, y_i) is replaced by $(ax_i + by_i + \xi_1, cx_i + dy_\mathrm{i} + \xi_2)$ for every $i = 1, \ldots, n$ and for every choice of a, b, c, d, ξ_1, and ξ_2 with $ad - bc \neq 0$.

Example 5. Consider the area of a triangle formed by three of our points—say (x_1, y_1), (x_2, y_2), and (x_3, y_3). This area is

$$\frac{1}{2} \left| \det \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \right|$$

After applying an affine transformation this area becomes

$$\frac{1}{2} \left| \det \left[\begin{pmatrix} a & b & \xi_1 \\ c & d & \xi_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \right] \right| = \frac{1}{2} |ad - bc| \left| \det \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \right|$$

Note that the absolute value signs are not necessary because we can permute the columns of the matrix to change sign. Also note that an affine transformation has a constant Jacobian determinant, namely $|ad - bc|$, which measures the "distortion" of areas.

It is clear that the ratio of the areas of two such triangles, or the ratio of two such determinants, is an invariant:

$$\frac{\det \begin{pmatrix} x_i & x_j & x_k \\ y_i & y_j & y_k \\ 1 & 1 & 1 \end{pmatrix}}{\det \begin{pmatrix} x_\ell & x_m & x_s \\ y_\ell & y_m & y_s \\ 1 & 1 & 1 \end{pmatrix}}$$

We can form

$$\binom{n}{3}$$

such triangles and after dividing by the area of one of them—say the one formed by the first three points or the one of largest area—we can get

$$\binom{n}{3} - 1$$

invariants. These, however, are not all independent. For example, consider the case of four points P_1, P_2, P_3, P_4 in general position in $\mathbb{R}^2$. We can find a unique affine transformation that takes P_1 to $(0, 0)$, P_2 to $(1, 0)$, and P_3 to $(0, 1)$, detailed later. The fourth point P_4 will go to some point not on the triangle of lines:

$$x = 0, \quad y = 0, \quad x + y = 1$$

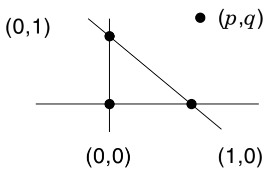

Figure 3. Transformed points.

For simplicity assume that P_4 goes to (p, q) with $p > 0$, $q > 0$, and $p + q > 1$ (see Fig. 3). Our

$$\binom{4}{3} = 4$$

triangles have areas 1/2, $p/2$, $q/2$, and $(p + q - 1)/2$. We see that the

$$\binom{4}{3} - 1 = 3$$

ratios p, q, and $p + q - 1$ are not independent—although any two of them are.

The mathematical interpretation of these invariants is straightforward, although somewhat abstract. They are functions on the quotient space

$$U/\mathrm{AFF}(2, \mathbb{R})$$

obtained by identifying those n-tuples of points in general position that can be transformed into each other by an affine transformation.

We can specifically determine this quotient because on each orbit there is a unique n-tuple with $P_1 = (0, 0)$, $P_2 = (1, 0)$, and $P_3 = (0, 1)$. This can be seen by noting that there is a unique affine transformation which carries (x_1, y_1) to (0, 0), (x_2, y_2) to (1, 0) and (x_3, y_3) to (0, 1). The uniqueness is clear because the only affine transformation

$$\begin{pmatrix} a & b & \xi_1 \\ c & d & \xi_2 \\ 0 & 0 & 1 \end{pmatrix}$$

which fixes (0, 0), (1, 0), and (0, 1) is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Existence is also easy. We translate (x_1, y_1) to (0, 0) by

$$\begin{pmatrix} 1 & 0 & -x_1 \\ 0 & 1 & -y_1 \\ 0 & 0 & 1 \end{pmatrix}$$

This carries (x_2, y_2) to $(x_2 - x_1, y_2 - y_1)$ and (x_3, y_3) to $(x_3 - x_1, y_3 - y_1)$. We then construct a 2×2 invertible matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

which carries these two vectors to (1, 0) and (0, 1) respectively. Specifically, we need

$$\begin{aligned} a(x_2 - x_1) + b(y_2 - y_1) &= 1 \\ c(x_2 - x_1) + d(y_2 - y_1) &= 0 \end{aligned}$$

and

$$\begin{aligned} a(x_3 - x_1) + b(y_3 - y_1) &= 0 \\ c(x_3 - x_1) + d(y_3 - y_1) &= 1 \end{aligned}$$

Solving this system of four equations in four unknowns yields

$$a = \frac{y_3 - y_1}{\det\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix}} \qquad c = \frac{-(y_2 - y_1)}{\det\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix}}$$

$$b = \frac{-(x_3 - x_1)}{\det\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix}} \qquad d = \frac{x_2 - x_1}{\det\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix}}$$

The composition

$$\begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -x_1 \\ 0 & 1 & -y_1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} a & b & -x_1 a - y_1 b \\ c & d & -x_1 c - y_1 d \\ 0 & 0 & 1 \end{pmatrix}$$

is the desired transformation.

A simple calculation shows that this transformation sends (x, y) to

$$\left(\frac{\det\begin{pmatrix} x_1 & x & x_3 \\ y_1 & y & y_3 \\ 1 & 1 & 1 \end{pmatrix}}{\det\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix}}, \quad \frac{\det\begin{pmatrix} x_1 & x_2 & x \\ y_1 & y_2 & y \\ 1 & 1 & 1 \end{pmatrix}}{\det\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix}} \right)$$

which makes it obvious that (x_1, y_1) goes to (0, 0), (x_2, y_2) goes to (1, 0), and (x_3, y_3) goes to (0, 1). Notice also that the remaining $n - 3$ points $(x_4, y_4), \ldots, (x_n, y_n)$ are sent to points whose coordinates are invariants. These $2n - 6$ invariant values serve as coordinate functions on the quotient space, which is clearly isomorphic to an open set W of $\mathbb{R}^{2n-6}$:

$$U/\mathrm{AFF}(2, \mathbb{R}) \cong W$$

The central theorem is the following.

Theorem. Any affine invariant expression

$$\frac{p(x_1, y_1, \ldots, x_n, y_n)}{q(x_1, y_1, \ldots, x_n, y_n)}$$

is a rational function of the invariant coordinate functions noted previously.

An equivalent formulation is that every invariant is a rational function of $2n - 6$ ratios of areas of triangles, for example:

$$\frac{\text{area}(P_i, P_1, P_2)}{\text{area}(P_1, P_2, P_3)} \quad \text{and} \quad \frac{\text{area}(P_i, P_1, P_3)}{\text{area}(P_1, P_2, P_3)}$$

for $i = 4, \ldots, n$. Note that we do not need to consider the ratio

$$\frac{\text{area}(P_i, P_2, P_3)}{\text{area}(P_1, P_2, P_3)}$$

because, as shown above, for the four points P_1, P_2, P_3, P_i, the areas of the

$$\binom{4}{3} = 4$$

triangles are linearly related and therefore the three ratios are linearly dependent.

Affine Invariants of Two Points and Two Lines in the Plane

Consider two lines L_1 and L_2 and two points P_1 and P_2 in the plane in general position. For our purposes general position means that L_1 and L_2 are not parallel and that P_1 is not on either L_1 or L_2. Given another set of two lines L_1' and L_2' and two points P_1' and P_2', we would like to know if there is an affine transformation of the plane that carries L_i to L_i' and P_i to P_i' for $i = 1, 2$. As we shall see, this will be true if and only if the two invariants constructed below have the same value for both of the geometric configurations.

The geometric configurations of interest (an ordered pair of lines and an ordered pair of points in general position in the plane) are parametrized by an open subset U of $\mathbb{P}^2_{\mathbb{R}} \times \mathbb{P}^2_{\mathbb{R}} \times \mathbb{R}^2 \times \mathbb{R}^2$. (Recall that lines $ax + by + c = 0$ in the plane are parametrized by points $(a : b : c) \in \mathbb{P}^2_{\mathbb{R}}$, real projective two-space.) The affine group AFF(2, $\mathbb{R}$) acts on $\mathbb{R}^2$ in a way that preserves lines, and so acts on U. Note that

$$M = \begin{pmatrix} \tilde{a} & \tilde{b} & \xi_1 \\ \tilde{c} & \tilde{d} & \xi_2 \\ 0 & 0 & 1 \end{pmatrix}$$

acts on points by sending

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \text{ to } M \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

but it acts on lines by sending

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} \text{ to } (M^T)^{-1} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

Since $\dim_{\mathbb{R}} U = 8$ and $\dim_{\mathbb{R}}$ AFF(2, $\mathbb{R}$) $= 6$, we expect a two-dimensional quotient AFF(2, $\mathbb{R}$)$\backslash U$. This quotient is, as we shall see, diffeomorphic to $\mathbb{R}^2$.

Let Q be the point of intersection of L_1 and L_2. We can find an affine transformation that moves Q to the origin, L_1 to the x axis, and L_2 to the y axis. In fact, if L_1 is given by $a_1x + b_1y + c_1 = 0$ and L_2 is given by $a_2x + b_2y + c_2 = 0$, then

$$M = \begin{pmatrix} a_2 & b_2 & c_2 \\ a_1 & b_1 & c_1 \\ 0 & 0 & 1 \end{pmatrix}$$

is one such transformation. Having moved L_1 and L_2 to the x and y axes, respectively, we can still act by transformations of the form:

$$N = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

If P_1 originally had coordinates (x_1, y_1), then after applying M, we will have the point $(a_2x_1 + b_2y_1 + c_2, a_1x_1 + b_1y_1 + c_1)$. Our general position assumption implies that neither coordinate is zero. Setting

$$\lambda_1 = \frac{1}{a_2x_1 + b_2y_1 + c_1}$$
$$\lambda_2 = \frac{1}{a_1x_1 + b_1y_1 + c_2}$$

in N will move this point to $(1, 1)$. Putting M and N together yields an affine transformation

$$T = NM$$
$$= \begin{pmatrix} \dfrac{a_2}{a_2x_1 + b_2y_1 + c_2} & \dfrac{b_2}{a_2x_1 + b_2y_1 + c_2} & \dfrac{c_2}{a_2x_1 + b_2y_1 + c_2} \\ \dfrac{a_1}{a_1x_1 + b_1y_1 + c_1} & \dfrac{b_1}{a_1x_1 + b_1y_1 + c_1} & \dfrac{c_1}{a_1x_1 + b_1y_1 + c_1} \\ 0 & 0 & 1 \end{pmatrix}$$

which moves L_1 to the x axis, L_2 to the y axis, and P_1 to $(1, 1)$. No degrees of freedom remain, so this must be the unique such affine transformation.

Suppose P_2 originally had coordinates (x_2, y_2); then the coordinates of P_2 after transformation by T parametrize the quotient AFF(2, $\mathbb{R}$)$\backslash U$ and are the essential invariants

$$I_1 = \frac{a_2x_2 + b_2y_2 + c_2}{a_2x_1 + b_2y_1 + c_2} \quad \text{and} \quad I_2 = \frac{a_1x_2 + b_1y_2 + c_1}{a_1x_1 + b_1y_1 + c_1}$$

In general an invariant takes the form

$$\frac{p(a_1, b_1, c_1, a_2, b_2, c_2, x_1, y_1, x_2, y_2)}{q(a_1, b_1, c_1, a_2, b_2, c_2, x_1, y_1, x_2, y_2)}$$

where p and q are polynomials that are homogeneous of the same degree in a_1, b_1, c_1 and in a_2, b_2, c_2. It can be shown that every such invariant expression is a rational function of the two fundamental invariants I_1 and I_2.

Example 6

$$L_1: x - y + 1 = 0$$
$$L_2: 2x - y = 0$$
$$P_1: (1, 0)$$
$$P_2: (0, 2)$$
$$T = \begin{pmatrix} 1 & -1/2 & 0 \\ 1/2 & -1/2 & 1/2 \\ 0 & 0 & 1 \end{pmatrix}$$
$$I_1 = -1 \quad I_2 = -1/2$$

Projective Invariants of Five Points in the Plane

Consider an ordered set of five points P_1, P_2, P_3, P_4, P_5 in the plane $\mathbb{R}^2$. We regard $\mathbb{R}^2$ as an open dense subset of the projective plane $\mathbb{P}^2_{\mathbb{R}}$. We assume that these points are in general position, that is, that no three are collinear. Notice that our geometry is parametrized by a 10-dimensional space, while the group of projective transformations has dimension 8. Thus we expect a two-dimensional quotient, and therefore two fundamental invariants.

To determine these invariants, observe that there is a unique projective transformation taking P_1 to $(1:0:0) \in \mathbb{P}^2_{\mathbb{R}}$, P_2 to $(0:1:0)$, P_3 to $(0:0:1)$, and P_4 to $(1:1:1)$. The point P_5 will be sent to some point $(a:b:c)$ under this transformation; moreover none of a, b, or c will be zero by the general position assumption. The ratios

$$I_1 = \frac{a}{c} \quad \text{and} \quad I_2 = \frac{b}{c}$$

will be the basic invariants. Any other invariant we might construct will be a rational function of these two.

The matrix

$$M = \frac{1}{\det\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix}} \begin{pmatrix} y_2 - y_3 & x_3 - x_2 & x_2y_3 - x_3y_2 \\ y_3 - y_1 & x_1 - x_3 & x_3y_1 - y_3x_1 \\ y_1 - y_2 & x_2 - x_1 & x_1y_2 - x_2y_1 \end{pmatrix}$$

sends P_1 to $(1:0:0)$, P_2 to $(0:1:0)$, and P_3 to $(0:0:1)$. However it sends $P_4 = (x_4 : y_4 : 1)$ to

$$Q_4 = \begin{pmatrix} \det\begin{pmatrix} x_4 & x_2 & x_3 \\ y_4 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \Big/ \det\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \\ \det\begin{pmatrix} x_1 & x_4 & x_3 \\ y_1 & y_4 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \Big/ \det\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \\ \det\begin{pmatrix} x_1 & x_2 & x_4 \\ y_1 & y_2 & y_4 \\ 1 & 1 & 1 \end{pmatrix} \Big/ \det\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \end{pmatrix}$$

(Note that none of these determinants are zero by the general position assumption.) Multiplying M by the diagonal matrix whose entries are the reciprocals of the components of Q_4 gives

$$\begin{pmatrix} \frac{y_2 - y_3}{\det\begin{pmatrix} x_4 & x_2 & x_3 \\ y_4 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix}} & \frac{x_3 - x_2}{\det\begin{pmatrix} x_4 & x_2 & x_3 \\ y_4 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix}} & \frac{x_2y_2 - x_3y_2}{\det\begin{pmatrix} x_4 & x_2 & x_3 \\ y_4 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix}} \\ \frac{y_3 - y_1}{\det\begin{pmatrix} x_1 & x_4 & x_3 \\ y_1 & y_4 & y_3 \\ 1 & 1 & 1 \end{pmatrix}} & \frac{x_1 - x_3}{\det\begin{pmatrix} x_1 & x_4 & x_3 \\ y_1 & y_4 & y_3 \\ 1 & 1 & 1 \end{pmatrix}} & \frac{x_3y_1 - y_3x_1}{\det\begin{pmatrix} x_1 & x_4 & x_3 \\ y_1 & y_4 & y_3 \\ 1 & 1 & 1 \end{pmatrix}} \\ \frac{y_1 - y_2}{\det\begin{pmatrix} x_1 & x_2 & x_4 \\ y_1 & y_2 & y_4 \\ 1 & 1 & 1 \end{pmatrix}} & \frac{x_2 - x_1}{\det\begin{pmatrix} x_1 & x_2 & x_4 \\ y_1 & y_2 & y_2 \\ 1 & 1 & 1 \end{pmatrix}} & \frac{x_1y_2 - x_2y_1}{\det\begin{pmatrix} x_1 & x_2 & x_4 \\ y_1 & y_2 & y_4 \\ 1 & 1 & 1 \end{pmatrix}} \end{pmatrix}$$

This is the desired projective transformation in the group PGL(3, $\mathbb{R}$) of all projective transformations of the projective plane (essentially 3×3 invertible matrices modulo scalars. It takes $P_5 = (x_5, y_5)$ to

$$\left(\frac{\det\begin{pmatrix} x_5 & x_2 & x_3 \\ y_5 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix}}{\det\begin{pmatrix} x_4 & x_2 & x_3 \\ y_4 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix}} : \frac{\det\begin{pmatrix} x_1 & x_5 & x_3 \\ y_1 & y_5 & y_3 \\ 1 & 1 & 1 \end{pmatrix}}{\det\begin{pmatrix} x_1 & x_4 & x_3 \\ y_1 & y_4 & y_3 \\ 1 & 1 & 1 \end{pmatrix}} : \frac{\det\begin{pmatrix} x_1 & x_2 & x_5 \\ y_1 & y_2 & y_5 \\ 1 & 1 & 1 \end{pmatrix}}{\det\begin{pmatrix} x_1 & x_2 & x_4 \\ y_1 & y_2 & y_4 \\ 1 & 1 & 1 \end{pmatrix}} \right)$$

This yields the invariants

$$I_1 = \frac{\det\begin{pmatrix} x_5 & x_2 & x_3 \\ x_5 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \det\begin{pmatrix} x_1 & x_2 & x_4 \\ y_1 & y_2 & y_4 \\ 1 & 1 & 1 \end{pmatrix}}{\det\begin{pmatrix} x_4 & x_2 & x_3 \\ y_4 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \det\begin{pmatrix} x_1 & x_2 & x_5 \\ y_1 & y_2 & y_5 \\ 1 & 1 & 1 \end{pmatrix}}$$

$$I_2 = \frac{\det\begin{pmatrix} x_1 & x_5 & x_3 \\ y_1 & y_5 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \det\begin{pmatrix} x_1 & x_2 & x_4 \\ y_1 & y_2 & y_4 \\ 1 & 1 & 1 \end{pmatrix}}{\det\begin{pmatrix} x_1 & x_4 & x_3 \\ y_1 & y_4 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \det\begin{pmatrix} x_1 & x_2 & x_5 \\ y_1 & y_2 & y_5 \\ 1 & 1 & 1 \end{pmatrix}}$$

Other ratios would also be just as good. Moreover, any invariant will be a rational expression in these.

Notice that the individual determinants are (up to sign and a factor of $\frac{1}{2}$) the areas of certain triangles. Thus our projective invariants are ratios of products of areas of certain pairs of triangles and are affine invariants as they should be.

Affine Invariants of Five Points in Space

Let P_1, P_2, P_3, P_4, P_5 be an ordered 5-tuple of distinct points in space $\mathbb{R}^3$. Say the coordinates of $P_i = (x_i, y_i, z_i)$. We will assume that the points are in general position, so that no four are coplanar (implying no three are collinear).

The fundamental affine invariants for any number of points in the space are formed from the ratios of the volumes of two tetrahedrons in space. If P_i, P_j, P_k, P_l are the vertices, the volume is

$$\det\begin{pmatrix} x_i & x_j & x_k & x_\ell \\ y_i & y_j & y_k & y_\ell \\ z_i & z_j & z_k & z_\ell \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

up to a factor of $\pm 1/6$.

Five points in general position yield

$$\binom{5}{4} = 5$$

tetrahedrons. Under an affine transformation of $\mathbb{R}^3$ these five volumes all scale by the same constant factor. Thus we can regard the volumes as giving a well-defined point in $\mathbb{P}^4_{\mathbb{R}}$. However, the points we get in $\mathbb{P}^4_{\mathbb{R}}$, as we run through all 5-tuples of points in $\mathbb{R}^3$, lie in a hyperplane, that is, they all satisfy a fixed linear relation. This can be seen by expanding the following determinant

$$0 = \det\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ y_1 & y_2 & y_3 & y_4 & y_5 \\ z_1 & z_2 & z_3 & z_4 & z_5 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

along the bottom row.

Thus we have only four independent volumes. Normalizing one of them to one yields three ratios of volumes of tetrahedra, which are the fundamental affine invariants of our five points. (This squares with the fact that our geometry (5 points in general position in space) is parametrized by a 15-dimensional space while AFF(3, $\mathbb{R}$) has dimension 12.

BIBLIOGRAPHY

1. B. L. van der Waerden, *Modern Algebra,* City: Frederick Ungar, 1949 and 1950, Vols. 1 and 2.
2. J. Canny, Generalized characteristic polynomials, *J. Symbolic Comput.* **9**: 241–250, 1990.

Reading List

S. Barnett, *Polynomials and Linear Control Systems,* New York: Marcel Dekker, 1983.

E. J. Barbeau, *Polynomials,* New York: Springer-Verlag, 1989.

D. Cox, J. Little, and D. O'Shea, *Ideals, Varieties, and Algorithms,* Undergraduate Texts in Mathematics, New York: Springer, 1992.

W. Fulton, *Algebraic Curves,* Menlo Park, CA: Benjamin, 1969.

I. M. Gelfand, et al., *Discriminants, Resultants, and Multidimensional Determinants,* Boston: Birkhäuser, 1994.

R. F. Gleeson and R. M. Williams, *A Primer on Polynomial Resultants,* Naval Air Development Center Tech. Rep., 1991, ADA 246 883.

J. Harris, *Algebraic Geometry: A First Course,* Graduate Texts in Mathematics, Vol. 133, New York: Springer, 1992.

A. P. Morgan, *Solving Polynomial Systems Using Continuation for Engineering and Scientific Problems,* Englewood Cliffs, NJ: Prentice Hall, 1987.

B. Roth, Computation in kinematics, in J. Angeles et al. (eds.), *Computational Kinematics,* Norwell, MA: Kluwer, 1993.

G. Salmon, *Lessons Introductory to the Modern Higher Algebra,* 5th ed., New York: Chelsea, 1932.

J. P. Serre, *A Course in Arithmetic,* Graduate Texts in Mathematics, Vol. 7, New York: Springer, 1971.

PETER F. STILLER
Texas A & M University

PONTRYAGIN MAXIMUM PRINCIPLE. See OPTIMAL CONTROL.

PORTABLE COMPUTERS. See LAPTOP COMPUTERS.