Resultants, Discriminants, Bezout, Nullstellensatz, etc,

Many computational tasks in number theory, algebra, and algebraic geometry can be performed quite efficiently by using a hoary old tool of nineteenth century algebra, called the resultant. This was originally invented in order to solve systems of polynomial equations, but turned out to have many other applications.

Some of these applications have been superseded by such techniques as Gröbner bases, but it turns out that resultants are still important, *both* theoretically and practically, and they have been making something of a resurgence in recent years. The goal of these notes is to give the basic definitions, and to give some sample applications, including: solving systems of polynomial equations, finding discriminants of polynomials, finding norms in algebraic extensions, and proving the Nullstellensatz.

Throughout we work over a ring A that we will assume is a UFD (i.e., there is unique factorization into primes, up to units) . The cases in which $A = \mathbf{Z}$, $A = \mathbf{Z}[x_1, \cdots, x_n]$, or $A = k[x_1, \cdots, x_n]$ are the only real cases of interest to us (where k is a field). Since A is an integral domain it has a fraction field, which we will denote F. Also, it is a standard theorem that if A is a UFD then so is $A[x]$.

Let

$$f(X) = f_0 + f_1 X + f_2 X^2 + \cdots + f_m X^m, \quad g(X) = g_0 + g_1 X + g_2 X^2 + \cdots + g_n X^n$$

be polynomials with coefficients in A. The **resultant** $R(f, g)$ of f and g is defined to be the determinant of the "Sylvester matrix"

$$R(f,g) := \begin{bmatrix} f_0 & f_1 & f_2 & & & \cdots \\ & f_0 & f_1 & f_2 & & \cdots \\ & & f_0 & f_1 & f_2 & \cdots \\ & & \vdots & & & \ddots \\ g_0 & g_1 & g_2 & & & \cdots \\ & g_0 & g_1 & g_2 & & \cdots \\ & & g_0 & g_1 & g_2 & \cdots \\ & & \vdots & & & \ddots \end{bmatrix} .$$

This matrix is square: it has $m+n$ rows and $m+n$ columns; the first n rows contain the coefficients of f staggered one position to the right in each succeeding row and the next m rows contain the coefficients of g similarly placed. All unlabeled entries are zero. For instance, the resultant of a quadratic and cubic polynomial is

$$R(f_0+f_1X+f_2X^2, g_0+g_1X+g_2X^2+g_3X^3) = \det \begin{bmatrix} f_0 & f_1 & f_2 & 0 & 0 \\ 0 & f_0 & f_1 & f_2 & 0 \\ 0 & 0 & f_0 & f_1 & f_2 \\ g_0 & g_1 & g_2 & g_3 & 0 \\ 0 & g_0 & g_1 & g_2 & g_3 \end{bmatrix}$$

$$\begin{aligned} = \quad & f_2^3\, g_0^2 - f_1\, f_2^2\, g_0\, g_1 + f_0\, f_2^2\, g_1^2 + f_1^2\, f_2\, g_0\, g_2 - \\ & 2\, f_0\, f_2^2\, g_0\, g_2 - f_0\, f_1\, f_2\, g_1\, g_2 + f_0^2\, f_2\, g_2^2 - f_1^3\, g_0\, g_3 + \\ & 3\, f_0\, f_1\, f_2\, g_0\, g_3 + f_0\, f_1^2\, g_1\, g_3 - 2\, f_0^2\, f_2\, g_1\, g_3 - f_0^2\, f_1\, g_2\, g_3 + f_0^3\, g_3^2. \end{aligned}$$

By inspection $R(f,g)$ is a polynomial in f_i and g_i in which each term has degree n in the f_i and m in the g_i.

Theorem 1. There are polynomials $u(X)$ and $v(X)$ with $\deg(u) < n$, $\deg(v) < m$ such that

$$u(X)f(X) + v(X)g(X) = R(f,g).$$

Proof. Let

$$u(X) = u_0 + u_1X + ... + u_{n-1}X^{n-1} \qquad v(X) = v_0 + v_1X + ... + v_{m-1}X^{m-1}.$$

be generic polynomials of degrees $n-1$ and $m-1$. The equation $uf+vg = r \in A$ is equivalent to linear equations on the u_i and v_j that can be summarized in matrix form as

$$\begin{bmatrix} f_0 & & & \cdots & g_0 & & & \cdots \\ f_1 & f_0 & & \cdots & g_1 & g_0 & \cdots & \\ f_2 & f_1 & f_0 & \cdots & g_2 & g_1 & g_0 & \cdots \\ \vdots & & & & \vdots & & & \cdots \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ v_0 \\ v_1 \\ \vdots \end{bmatrix} = \begin{bmatrix} r \\ 0 \\ 0 \\ \vdots \\ \\ \end{bmatrix}$$

The coefficient matrix is the transpose of the Sylvester matrix. If $R(f,g) \neq 0$ then by Cramer's rule we can solve these equations in A if $r = R(f,g)$. On the other hand, if $R(f,g) = 0$ then the matrix is singular and there is a nonzero solution in the ring. In either case we find that there are polynomials $u(X), v(X) \in A[X]$ such that $uf + vg = R(f,g)$. □

Remark 2. This theorem says that the resultant lies in the intersection of the ideal (f,g) in $A[x]$ generated by f and g, and the subring A of constant polynomials. It is tempting to conjecture that the ideal $(f,g) \cap A$ is the principal ideal generated by $R(f,g)$ but this turns out to be false in general.

Theorem 3. If at least one of the leading coefficients f_m, g_n is nonzero then $R(f,g) = 0$ if and only if $f(X)$ and $g(X)$ have a common factor of positive degree.

Proof. If $f(X)$ and $g(X)$ have a common factor with positive degree then the formula

$$uf + vg = R(f,g)$$

says that the factor must divide the constant $R(f,g)$ and hence that $R(f,g) = 0$.

Conversely, if $R(f,g) = 0$ then $uf = -vg$. If, for instance, the leading coefficient of f is nonzero, then one of the irreducible factors of f must divide g by unique factorization in $A[X]$ and the fact that v has degree strictly less than n. □

If the polynomials $f(X)$ and $g(X)$ factor completely into linear factors then their resultant can be expressed as a function of their roots. In other words, the resultant can be expressed in terms of the polynomial. Note that we can always find the roots of the polynomials by passing, if necessary, to the splitting field of $f(X)g(X)$.

Theorem 4. If $f(X)$ and $g(X)$ split into linear factors

$$f(X) = f_m \prod_{i=1}^{m} (X - \alpha_i), \qquad g(X) = g_n \prod_{i=1}^{n} (X - \beta_i)$$

then

$$R(f,g) = f_m^n g_n^m \prod_{i=1}^{m} \prod_{j=1}^{n} (\alpha_i - \beta_j) = f_m^n \prod_{i=1}^{m} g(\alpha_i) = (-1)^{mn} g_n^m \prod_{j=1}^{n} f(\beta_j).$$

Proof. First we prove the formula for the ring

$$A_0 = \mathbf{Z}[f_m, g_n, \alpha_1, \ldots, \alpha_d, \beta_1, \ldots, \beta_e]$$

in which the leading coefficients and the roots are indeterminates. Each of the coefficients f_i is $\pm f_m$ times an elementary symmetric function of the roots α_j, and similarly for the coefficients of $g(X)$. For instance,

$$f_{m-1}/f_m = -(\alpha_1 + \cdots + \alpha_m).$$

Factoring out f_m from the first set of rows in the Sylvester matrix and g_n from the second set of rows shows that $R(f, g)$ is equal to $f_m^n g_n^m$ times a function of the roots.

If we think of $R(f, g)$ as a function only of the indeterminate β_j then, by the preceding theorem, the resultant, thought of as a polynomial in β_j, has a root when $\beta_j = \alpha_i$ for some i. This says that $R(f, g)$ is divisible by $\alpha_i - \beta_j$ for all i and j so that

$$R(f,g) = Cf_m^n g_n^m \prod_{i=1}^{m}\prod_{j=1}^{n}(\alpha_i - \beta_j) = Cf_m^n \prod_{i=1}^{m} g(\alpha_i) = C(-1)^{mn} g_n^m \prod_{j=1}^{n} f(\beta_j)$$

for some polynomial C in the roots. In fact C is a symmetric function of the roots and is therefore a polynomial in f_i/f_m and g_j/g_n. Since $f(\beta_j)$ is homogeneous of degree 1 in the f_i the product of all $f(\beta_j)$ is of homogeneous of degree n in the f_i. The resultant is of degree n in the f_i (from the determinantal expression) so C is independent of the f_i. Similarly, C is independent of the g_i. Thus C is a constant (i.e., an integer). The resultant contains the term $f_0^n g_n^m$ (looking down the main diagonal of the Sylvester matrix) and the expression $g_n^m \prod f(\beta_i)$ contains this term. Thus $C = 1$.

The proof for arbitrary rings is now finished by the sort of trick that algebraists love. Namely, if A is any ring and f and g factor as in the statement of the theorem, then there is an obvious homomorphism from A_0 to A (the "evaluation" map that takes each of the generic indeterminates to the corresponding quantity in A). By applying this homomorphism the formulas in the theorem are true in A since they are true in A_0. □

Several very useful facts about resultants follow more or less immediately from the the above the theorems (usually the different expressions in Theorem 3).

Corollary 5.

a. $R(g,f) = (-1)^{mn} R(f,g)$.

b. The resultant is multiplicative; i.e.,

$$R(fg,h) = R(f,h)R(g,h), \qquad R(f,gh) = R(f,g)R(f,h).$$

c. $R(f,a) = a^m$, $R(X-a,f) = f(a)$, $R(aX+b,f) = a^m f(-b/a)$, etc.

d. If we divide f into g to get quotient $q(X)$ and remainder $r(X)$, with degree$(r) = k < d$ then

$$R(f,g) = R(f,qf+r) = f_m^{n-k} R(f,r).$$

e. $R(f(X+a), g(X+a)) = R(f,g)$.

f. $R(f(aX), g(aX)) = a^{mn} R(f,g)$.

g. If $F(X) = \sum f_{m-i} X^i$ is the "reversal" of $f(X)$ and $G(X)$ is the reversal of $g(X)$ then $R(F,G) = R(f,g)$.

h. If f_i is assigned weight $m-i$ and g_j is assigned weight $n-j$ then $R(f,g)$ is homogeneous of weight mn.

Perhaps **d** is the key result in that it gives an efficient algorithm for computing the resultant (much more efficient than computing the original determinant). Namely, one follows a Euclidean algorithm sort of construction that decreases the degrees until we reach the base cases described in **c.**.

It seems a pity to deprive the reader the pleasure of proving the above formulae, so we leave them unproved here.

Example 6. Solving Equations. Resultants can be used to solve equations in several variables by eliminating variables.

Suppose that we want to find all complex numbers x, y, and z such that

$$\begin{aligned} x + y + z &= 6 \\ x^2 + y^2 + z^2 &= 14 \\ x^3 + y^3 + z^3 &= 36 \end{aligned}$$

Although this system is simple enough to solve by hand or by a flash of insight, let's try resultants.

Specifically, think of the equations as three equations in x whose coefficients are in the ring $\mathbf{C}[y,z]$. Then they can have a common root if and only if the resultant of the first two and the resultant of the first and third are zero. So we take two resultants with respect to x. Then we have two polynomials in y with coefficients in $\mathbf{C}[z]$ and we are interested in when they have a common root. So we take their resultant to get a polynomial in z. The possible values of z for which the system has a solution are the roots of this polynomial.

More concretely, we find a Unix prompt somewhere and hope that Magma or Mathematica or Maple is available. For instance, in Mathematica, we implement the above ideas easily:

```
In[1]:= f = x+y+z-6; g = x^2+y^2+z^2-14; h = x^3+y^3+z^3-36;

In[2]:= fg = Resultant[f,g,x]
                        2                      2
Out[2]= 22 - 12 y + 2 y  - 12 z + 2 y z + 2 z

In[3]:= fh = Resultant[f,h,x]
                           2                        2         2        2
Out[3]= 180 - 108 y + 18 y  - 108 z + 36 y z - 3 y  z + 18 z  - 3 y z

In[4]:= Resultant[fg,fh,y]
                       2        3         4         5        6
Out[4]= 1296 - 4752 z + 6948 z  - 5184 z  + 2088 z  - 432 z  + 36 z

In[5]:= Factor[\%]
                     2         2         2
Out[5]= 36 (-3 + z)  (-2 + z)  (-1 + z)
```

Working backwards, one finds 2 possible values of y for each of the three values of z, and finds that the linear equation completely determines x once y and z are known. Thus there are 6 triples all together (the 6 permutations of $1,2,3$), and the algebraic geometer would have guessed that to begin with since the equations are of degrees 1, 2, and 3.

Example 7. The Discriminant of a polynomial. If a polynomial f factors into linear factors

$$f(X) = \sum_{i=0}^{m} f_i X^i = f_m \prod (X - \alpha_i)$$

then the **discriminant** of f is defined to be

$$D(f) = f_m^{2m-2} \prod_{i<j} (\alpha_i - \alpha_j)^2.$$

(The exponent on the leading coefficient is chosen so that the discriminant of a polynomial is equal to the discriminant of the polynomial with coefficients in reverse order.) An easy calculation shows that

$$f_m D(f) = (-1)^{m(m-1)/2} R(f, f').$$

A slightly more difficult calculation shows that

$$D(fg) = D(f)D(g)R(f, g)^2.$$

In particular, the discriminant of a polynomial is a function only of its coefficients and it is zero precisely when the polynomial has a repeated root.

In dealing with elliptic curves of the form $y^2 = x^3 + ax + b$ we will be interested in whether the cubic polynomial on the right-hand side is separable (i.e., has distinct roots in a splitting field). This happens if and only if the discriminant $D(x^3 + ax + b)$ is nonzero. From the above remark we have

$$D(x^3 + ax + b) = -R(x^3 + ax + b, 3x^2 + a).$$

We can use the earlier Corollary to mechanically calculate this result by hand:

$$\begin{aligned} D(x^3 + ax + b) &= -R((3x^2 + a)(x/3) + (2ax/3 + b), 3x^2 + a) \\ &= -3^2 R(2ax/3 + b, 3x^2 + a) \\ &= -3^2 R(3x^2 + a, 2ax/3 + b) \\ &= -3^2 (2a/3)^2 (3(-b/(2a/3))^2 + a) \\ &= -3^2 (3b^2 + a(2a/3)^2) = -27b^2 - 4a^3. \end{aligned}$$

The discriminant of the general monic cubic polynomial $f(X) = X^3 + aX^2 + bX + c$ can be computed by a similar procedure (or just by changing

variables in the above formula and using the Corollary suitably). The result is

$$D(X^3 + ax^2 + bX + c) = -27c^2 - 4b^3 + 18abc + a^2b^2 - 4a^3c.$$

The basic idea in the computation of resultants is to apply Euclid's algorithm while keeping track of the resultant by looking at the leading coefficients that arise. This "gcd" process for computing resultants is significantly easier than computing the defining determinant even for polynomials of small degree as above. In general the resultant of two polynomials of degree approximately n can be computed in time $O(n^2)$ by this method, whereas the computation of the determinant takes time $O(n^3)$ (counting, in both cases, an arithmetic operation as a single step).

Example 8. The irreducible polynomial of the sum of two algebraic numbers. Suppose that x satisfies the equation x^3+x+1 and that y satisfies the equation $y^3-2=0$. Then what is the equation of $z=x+y$? The number z is algebraic and you would probably guess, after a little thought, that its degree is probably nine. In fact the equation can be found just by eliminating x and y from $z-x-y$: $R_x(R_y(z-x-y, y^3-2), x^3+x+1)$, where the subscript is the variable being eliminated. In this case the answer is that z satisfies the equation

$$0 = -3 - 15z - 3z^2 + 58z^3 + 12z^4 + 3z^5 - 3z^6 + 3z^7 + z^9.$$

If the degree of the sum is less than the product of the degrees then the resultant will factor. Similarly, a polynomial satisfied by the product of two algebraic numbers can be found as $R_x(R_y(z-xy, f), g)$. the product of two algebraic numbers.

Example 9. Computing norms in an algebraic extension. If $f(X) \in k[X]$ is an irreducible separable monic polynomial of degree n, then any element in the extension $K = k(\alpha)$, where $f(\alpha) = 0$, can be written as a polynomial $g(\alpha)$ in α of degree strictly less than n. The norm $N_{K/k}(x)$ of $g(\alpha)$ is defined to be the product of all $g(\alpha_i)$ over all conjugates of α (i.e., other roots of $f(X) = 0$). By our formula for the resultant, the norm is

$$N_{K/k}(g(\alpha) = R(g(X), f(X)).$$

Example 10. Tschirnhaus transformations. Let $K = k(\alpha)$ be an algebraic extension as in the preceding example. Then an element $g(\alpha)$ of K satisfies an equation of degree dividing n. One way to find this equation is to compute the characteristic polynomial of the linear transformation "multiply by $g(\alpha)$" on K as a vector space over k, using the obvious basis $1, \alpha, \alpha^2, \ldots$. Another way is to eliminate X between the polynomials $f(X)$ and $y - X$.

For instance, the equation satisfied by $x = \alpha^3 + \alpha$ in the extension $\mathbf{Q}(\alpha)$, $\alpha^5 + \alpha + 1 = 0$ is $R_y(x - y^3 - y, y^5 + y + 1) = x^5 + 4x^3 + 2x^2 - x + 5 = 0$.

Example 11. Bezout's Theorem. If $f(x, y)$ and $g(x, y)$ are polynomials in two variables that have no common factor, then it isn't too hard to work out that in general the resultant $R(f, g)$ with respect to, say, the variable y, is a polynomial in x of degree at most mn. In other words the intersection of two curves in the plane, with no common component, has at most mn points.

Just as one can say that a polynomial of degree n has exactly n roots if they are counted with multiplicity, Bezout's Theorem says that plane curves of degrees m and n intersect in exactly mn points (or infinitely many, if they share a common component).

In order to make this precise, we have to (a) work over an algebraically closed field, like the complex numbers, (b) work in projective space (i.e., add a "line at infinity"), and (c) define "multiplicity" at an intersection of two curves. This is done carefully in Fulton's book on algebraic curves; very briefly, here is a low-tech description in terms of resultants.

Suppose that $f(x, y, z), g(x, y, z) \in \mathbf{C}[x, y, z]$ are homogeneous polynomials of degrees m and n respectively. The zero sets $C = V(f)$ and $D = V(g)$ are curves in the projective plane $\mathbf{P}^2 = (\mathbf{C}^3 - 0)/\equiv$ where $\equiv$ is the equivalence relation "are proportional" on nonzero 3-vectors over the complex numbers. We assume that the polynomials have no common factor, which is the same as saying that the curves C and D have no common component. Choose coordinates so that the point $p = 0\colon 0\colon 1$ isn't in the intersection of these curves, and so that p isn't on any of the (finitely many) lines joining pairs of points in $C \cap D$. Define the intersection multiplicity $I(q, C \cap D)$ to be the order of vanishing of $R_z(f, g)$ at $x\colon y$ where $q = x\colon y\colon z$. The resultant is nonzero since f and g do not share a common factor, and using our earlier results it turns out that $R_z(f, g)$ is a homogeneous polynomial of degree mn in x and y. Such a polynomial factors into mn linear factors, corresponding to intersection points (counted with multiplicity) and this is Bezout's Theo-

rem: curves of degrees m and n in the projective plane over an algebraically closed field intersect in exactly mn points, counted with multiplicity.

Example. The Nullstellensatz. A famous fundamental result of algebraic geometry is called the Nullstellensatz. A very concrete form of the "weak Nullstellensatz" (from which the usual formulation in terms of ideals can be derived quickly) the either a set of polynomials with complex coefficients has a common zero, or else there is a concrete certificate that they can't, namely there is a linear combination of them equal to 1.

More precisely, suppose that we have a collection

$$f_1, \ldots, f_m \in \mathbf{C}[x_1, \ldots, x_n]$$

of m polynomials in n variables. The weak form of the Nullstellensatz asserts that exactly one of the following two things must happen:

- **A:** There is an $a \in \mathbf{C}^n$ such that $f_i(a) = 0$ for all i, $1 \leq i \leq n$.
- **B:** There are polynomials $g_1, \ldots, g_m \in \mathbf{C}[x_1, \ldots, x_n]$ such that

$$\sum g_i f_i = 1.$$

In the language of algebraic geometry, if $I = (f_1, \ldots, f_m)$ is the ideal generated by the f_i then **A** asserts that the mutual zero set $V(I)$ of the ideal is nonempty and **B** asserts that $I = (1) = \mathbf{C}[x_1, \ldots, x_n]$.

An old-fashioned proof of this result can be given fairly easily using resultants and induction on n. If $n = 0$ the result is trivial (a collection of constants either are all zero, or else contain a multiple of one).

Now suppose that the result is true for polynomials with fewer than n variables. Constants c_i can be chosen so that the change of variables of the form

$$x_i' = x_i + c_i x_n, \qquad 1 \leq i \leq n-1$$

leaves f_1 having its highest term in x_n be a constant times x_n^e, while leaving the **A**/**B** dichotomy unchanged. (Indeed, the highest term in x_n in $F_1(x_1, \ldots, x_n) := f_1(x_1', \ldots, x_n')$ is a nonzero polynomial in the c_i; since $\mathbf{C}$ is infinite we can choose constants so that the value of the polynomial is nonzero. If **A** is true for the F_i then by the change of variables the f_i also have a common zero. If **B** is true for the F_i then $1 = \sum g_i F_i$; by reversing

the change of variables we see that $1 = \sum h_i f_i$ so **B** is true for the original polynomials.)

If there is only one polynomial ($m = 1$) then **A** is true since **C** is algebraically closed, unless the polynomial is a nonzero constant in which case **B** is trivially true (this is the only place in the proof where the fact that **C** is algebraically closed is used). If $m > 1$ then introduce indeterminates $u_2, \ldots, u_n$ and compute the resultant

$$R_{x_n}(f_1, \sum_{i=2}^{m} u_i f_i)$$

over the ring $A = \mathbf{C}[x_1, \ldots, x_{n-1}, u_2, \ldots, u_m]$. Collecting terms in this resultant in monomials u^α in the u_i gives

$$R_{x_n}(f_1, \sum_{i=2}^{m} u_i f_i) = \sum_{\alpha} R_\alpha(x_1, \ldots, x_{n-1}) u^\alpha.$$

Apply the induction assumption to the R_α. If case **A** applies to those polynomials then there is an $a \in \mathbf{C}^{n-1}$ such that all R_α vanish on a. Since the leading coefficient of f_1 in x_n doesn't vanish on specializing $(x_1, \ldots, x_{n-1})$ to a we conclude that f_1 and $\sum_{i=2}^{m} u_i f_i$ have a common factor after making this specialization. Since this factor can't involve the u_i we see that the f_i have a common root and there is an n-tuple on which all f_i vanish.

On the other hand, if case **B** applies to the R_α

$$1 = \sum g_\alpha R_\alpha$$

for suitable polynomials $g_\alpha \in \mathbf{C}[x_1, \ldots, x_{n-1}]$. From our first theorem on resultants we know that

$$R_{x_n}(f_1, \sum_{i=2}^{m} u_i f_i) = G \cdot f_1 + H \cdot (\sum_{i=2}^{m} u_i f_i)$$

for suitable polynomials $G, H \in \mathbf{C}[x_1, \ldots, x_{n-1}, u_2, \ldots, u_m]$. Collecting terms in monomials in the u_i gives

$$R_\alpha = \sum_{i=1}^{m} h_i f_i$$

for appropriate $h_i \in \mathbf{C}[x_1, \ldots, x_n]$. Substituting this into $1 = \sum g_\alpha R_\alpha$ shows that an appropriate linear combination of the f_i is equal to 1 and case **B** holds for the f_i as desired. This finishes the proof.