SYMMETRIC POLYNOMIALS

1. Definition of the Symmetric Polynomials

Let n be a positive integer, and let $r_1, \cdots, r_n$ be indeterminates over $\mathbb{Z}$ (they are *algebraically independent*, meaning that there is no nonzero polynomial relation among them).

The monic polynomial $g \in \mathbb{Z}[r_1, \cdots, r_n][X]$ having roots $r_1, \cdots, r_n$ expands as

$$g(X) = \prod_{i=1}^{n}(X - r_i) = \sum_{j \in \mathbb{Z}}(-1)^j \sigma_j X^{n-j}$$

whose coefficients are (up to sign) the **elementary symmetric functions** of $r_1, \cdots, r_n$,

$$\sigma_j = \sigma_j(r_1, \cdots, r_n) = \begin{cases} \sum_{1 \le i_1 < \cdots < i_j \le n} \prod_{k=1}^{j} r_{i_k} & \text{for } j \ge 0 \\ 0 & \text{for } j < 0. \end{cases}$$

Note the special cases $\sigma_0 = 1$ and $\sigma_j = 0$ for $j > n$. For example, if $n = 4$ then the nonzero elementary symmetric functions are

$$\begin{aligned}
\sigma_0 &= 1, \\
\sigma_1 &= r_1 + r_2 + r_3 + r_4, \\
\sigma_2 &= r_1r_2 + r_1r_3 + r_1r_4 + r_2r_3 + r_2r_4 + r_3r_4, \\
\sigma_3 &= r_1r_2r_3 + r_1r_2r_4 + r_1r_3r_4 + r_2r_3r_4, \\
\sigma_4 &= r_1r_2r_3r_4.
\end{aligned}$$

It seems clear that because $r_1, \cdots, r_n$ are algebraically independent, so are $\sigma_1, \cdots, \sigma_n$, but a small argument is required to show this. The problem is that although an integer polynomial relation $f(\sigma_1, \cdots, \sigma_n) = 0$ expands to an integer polynomial relation $F(r_1, \cdots, r_n) = 0$, forcing F to be the trivial polynomial, it is not immediate that consequently f is the trivial polynomial as well. So, suppose a relation

$$f(\sigma_1, \cdots, \sigma_n) = 0, \quad f \in \mathbb{Z}[X_1, \cdots, X_n].$$

Any nonzero term of $f(X_1, \cdots, X_n)$ takes the form

$$aX_1^{d_1} X_2^{d_2} \cdots X_n^{d_n}.$$

Set

$$\begin{aligned}
e_n &= d_n \\
e_{n-1} &= d_{n-1} + e_n \\
e_{n-2} &= d_{n-2} + e_{n-1} \\
&\vdots \\
e_1 &= d_1 + e_2.
\end{aligned}$$

Then the nonzero term of f is now

$$aX_1^{e_1-e_2}X_2^{e_2-e_3}\cdots X_n^{e_n}, \quad e_1 \geq e_2 \geq \cdots \geq e_n \geq 0.$$

Sort the nonzero terms lexicographically, i.e., first by total degree, then by X_1-exponent, then X_2-exponent, and so on. In the lex-initial term, substituting the σ_i for the X_i gives

$$a\sigma_1^{e_1-e_2}\sigma_2^{e_2-e_3}\cdots\sigma_n^{e_n} = a(r_1^{e_1}r_2^{e_2}\cdots r_n^{e_n} + \cdots).$$

Now $ar_1^{e_1}r_2^{e_2}\cdots r_n^{e_n}$ is the lex-initial nonzero term of $g(r_1, \cdots, r_n)$, sorting here by r_i-exponents rather than X_i-exponents. Thus no other term can cancel it in the relation $g(r_1, \cdots, r_n) = 0$. Therefore, no nonzero term of $f(X_1, \cdots, X_n)$ exists.

Give the ring of polynomials in $r_1, \cdots, r_n$ a name,

$$R = \mathbb{Z}[r_1, \cdots, r_n].$$

The symmetric group S_n acts on R,

$$\sigma f(r_1, \cdots, r_n) = f(r_{\sigma 1}, \cdots, r_{\sigma n}), \quad \sigma \in S_n,\ f \in \mathbb{Z}[r_1, \cdots, r_n].$$

The polynomials in R that are invariant under the action form a subring of R,

$$R_o = \{S_n\text{-invariant polynomials in } R\}.$$

The product form in the earlier equality

$$g(X) = \prod_{i=1}^{n}(X - r_i) = \sum_{j\in\mathbb{Z}}(-1)^j\sigma_j X^{n-j}$$

shows that the σ_j are invariant under the action, and hence

$$\mathbb{Z}[\sigma_1, \cdots, \sigma_n] \subset R_o.$$

In fact the containment is an equality.

Theorem 1.1 (Fundamental Theorem of Symmetric Polynomials)**.** *The subring of polynomials in $\mathbb{Z}[r_1, \cdots, r_n]$ that are fixed under the action of S_n is $\mathbb{Z}[\sigma_1, \cdots, \sigma_n]$.*

Proof. Consider a nonzero polynomial $f \in \mathbb{Z}[r_1, \cdots, r_n]$ that is fixed under the action of S_n. Sort its nonzero terms lexicographically, first by total degree, then by r_1-exponent, then r_2-exponent, and so on. Consider its lex-initial term,

$$ar_1^{e_1}\cdots r_n^{e_n}.$$

For any $\sigma \in S_n$ the polynomial f contains a term having the same coefficient but with the variables permuted by σ. Thus the lex-initial term takes the form

$$t = ar_1^{e_1}\cdots r_n^{e_n}, \quad e_1 \geq \cdots \geq e_n \geq 0.$$

Now consider the coefficient of t times a product of elementary symmetric functions,

$$g_t = a\sigma_1^{e_1-e_2}\sigma_2^{e_2-e_3}\cdots\sigma_n^{e_n} \in \mathbb{Z}[\sigma_1, \cdots, \sigma_n]$$

(the exponents are all nonnegative because of the conditions on the e_i). This polynomial's lexicographically-highest term is exactly t. Thus, recalling that f is our S_n-invariant polynomial and noting that g_t is certainly S_n-invariant as well, we see that the polynomial $f - g_t$ is also S_n-fixed, and it has a smaller lex-initial term than f. Replace f by $f - g_t$ and continue in this fashion until the original f is expressed as a polynomial in the σ_i. $\square$

The **discriminant** of $r_1, \cdots, r_n$ (also called the discriminant of g) is

$$\Delta = \Delta(r_1, \cdots, r_n) = \Delta(g) = \prod_{1 \le i < j \le n} (r_i - r_j)^2.$$

Being visibly invariant under S_n, the discriminant lies in the coefficient field of g. For example, if $n = 2$ then

$$\Delta = (r_1 - r_2)^2 = (r_1 + r_2)^2 - 4r_1 r_2 = \sigma_1^2 - 4\sigma_2.$$

Trying similarly to analyze the case $n = 3$ quickly shows that expressing Δ in terms of the σ_j is not easy, although the proof of the Fundamental Theorem shows us how to do it. (Answer: $\sigma_1^2\sigma_2^2 - 4\sigma_2^3 - 4\sigma_1^3\sigma_3 - 27\sigma_3^2 + 18\sigma_1\sigma_2\sigma_3$.) Soon we will develop a general discriminant algorithm.

The square root of the discriminant,

$$\sqrt{\Delta} = \prod_{1 \le i < j \le n} (r_i - r_j),$$

changes its sign when any two of the r's are exchanged, i.e., $(k\,\ell)\sqrt{\Delta} = -\sqrt{\Delta}$ for any transposition $(k\,\ell) \in S_n$. That is, $\sqrt{\Delta}$ is fixed by A_n but not by S_n.

2. Guided example: Solving the Cubic Equation

To solve the general cubic equation, the task is to express r_1, r_2, r_3 in terms of $\sigma_1, \sigma_2, \sigma_3$. Let

$$r = r_1 + \zeta_3 r_2 + \zeta_3^2 r_3.$$

Show that r^3 is invariant under the alternating group A_3. Let S_3 act on $\mathbb{Z}[r_1, r_2, r_3]$. Then we have

$$(2\,3)r = r_1 + \zeta_3 r_3 + \zeta_3^2 r_2.$$

Show that $((2\,3)r)^3 \ne r^3$ and hence that $(2\,3)(r^3) \ne r^3$. Thus r^3 is not invariant under the full symmetric group S_3. Since a set of coset representatives for S_3/A_3 is $\{1, (2\,3)\}$, the polynomial

$$R_{r^3}(X) = (X - r^3)(X - (2\,3)(r^3)) = X^2 - (r^3 + (2\,3)(r^3))X + r^3 \cdot (2\,3)(r^3)$$

lies in $\mathbb{Z}[\sigma_1, \sigma_2, \sigma_3]$. (This polynomial is the *resolvent* of r^3.) Use the proof of the Fundamental Theorem of Symmetric Functions for $n = 3$ to show that

$$r \cdot (2\,3)r = \sigma_1^2 - 3\sigma_2,$$
$$r^3 + (2\,3)(r^3) = 2\sigma_1^3 - 9\sigma_1\sigma_2 + 27\sigma_3,$$

so that the resolvent expands as

$$R_{r^3}(X) = X^2 - (2\sigma_1^3 - 9\sigma_1\sigma_2 + 27\sigma_3)X + (\sigma_1^2 - 3\sigma_2)^3.$$

Taking a square root over the coefficient field gives r^3 and $(r^3)^{(2\,3)}$. (We don't know which is which because there is no canonical labeling of r_1, r_2, r_3, so just designate one as r^3.) Now r is a root of

$$R_r(X) = X^3 - r^3$$

(there are three roots, but again they are indistinguishable under relabeling of the r_i), and $r^{(2\,3)} = (\sigma_1^2 - 3\sigma_2)/r$ as computed above. Now that we have r and $r^{(2\,3)}$, find r_1, r_2, r_3 by solving the linear system

$$\begin{aligned} r_1 + \zeta_3 r_2 + \zeta_3^2 r_3 &= r \\ r_1 + \zeta_3^2 r_2 + \zeta_3 r_3 &= r^{(2\,3)} \\ r_1 + \quad r_2 + \quad r_3 &= \sigma_1. \end{aligned}$$

Use these methods to solve the cubic polynomial $X^3 - 3X + 1$.

The strategy of this example is very general. Suppose that a polynomial

$$g(X) = \prod_{i=1}^{n} (X - r_i)$$

has roots $r_1, \cdots, r_n$ that need not be algebraically independent, and suppose that a group G acts on the roots, fixing some underlying ring A. If we can find some polynomial expression in the roots,

$$s = s(r_1, \cdots, r_n), \quad s \in A[X_1, \cdots, X_n],$$

that is invariant under the action of a subgroup H of G, then the associated resolvent polynomial is

$$f_s(X) = \prod_{gH \in G/H} (X - gs).$$

(The name g for group-elements in the formula for the resolvent has no connection to the name g of the original polynomial from a moment ago.) The resolvent has degree $[G : H]$, and it has s as a root, and it is invariant under the action of the full group G because the map $gH \mapsto \gamma gH$ permutes the coset space G/H,

$$(\gamma f_s)(X) = \prod_{gH \in G/H} (X - \gamma gs) = \prod_{\gamma gH \in G/H} (X - \gamma gs) = f_s(X).$$

Thus, the coefficients of f_s are G-invariant. An algorithm might consequently be available to compute them, and then perhaps we can find the roots of f_s, one of which is s. Thus the problem of finding the roots of g given only the elementary symmetric functions of the roots would be reduced to finding the roots of g given also the roots of f_s, those roots being $\{gs : gH \in G/H\}$.

Depending on the context, one can bring various artfulnesses to bear on choosing a subgroup H of G and then finding an H-invariant expression s.

3. Guided Example: Solving the Quartic Equation

Let $n = 4$. Let

$$\begin{aligned} r &= r_1 - r_2 + r_3 - r_4, \\ s &= r^2. \end{aligned}$$

Show that the subgroup of S_4 leaving s invariant is the dihedral group

$$D = \langle (1\,2\,3\,4), (1\,3) \rangle,$$

and that a set of coset representatives for S_4/D is $\{1, (1\,2), (1\,4)\}$. Show that the Fundamental Theorem of Symmetric Functions gives

$$\begin{aligned} r \cdot (1\,2)r \cdot (1\,4)r &= \sigma_1^3 - 4\sigma_1\sigma_2 + 8\sigma_3 \\ s + (1\,2)s + (1\,4)s &= 3\sigma_1^2 - 8\sigma_2 \\ s \cdot (1\,2)s + s \cdot (1\,4)s + (1\,2)s \cdot (1\,4)s &= 3\sigma_1^4 - 16\sigma_1^2\sigma_2 + 16\sigma_1\sigma_3 + 16\sigma_2^2 - 64\sigma_4. \end{aligned}$$

To solve the quartic, take the cubic resolvent of s,

$$\begin{aligned} R_s(X) &= (X - s)(X - (1\,2)s)(X - (1\,4)s) \\ &= X^3 - (3\sigma_1^2 - 8\sigma_2)X^2 + (3\sigma_1^4 - 16\sigma_1^2\sigma_2 + 16\sigma_1\sigma_3 + 16\sigma_2^2 - 64\sigma_4)X \\ &\quad - (\sigma_1^3 - 4\sigma_1\sigma_2 + 8\sigma_3)^2. \end{aligned}$$

The three roots are s, $(1\,2)s$, and $(1\,4)s$; taking square roots of the first two gives r and $(1\,2)r$, so as computed above, $(1\,4)r = (\sigma_1^3 - 4\sigma_1\sigma_2 + 8\sigma_3)/(r \cdot (1, 2)r)$. Now to solve the original quartic, solve the linear system

$$\begin{aligned} r_1 - r_2 + r_3 - r_4 &= r \\ -r_1 + r_2 + r_3 - r_4 &= r^{(1\,2)} \\ -r_1 - r_2 + r_3 + r_4 &= r^{(1\,4)} \\ r_1 + r_2 + r_3 + r_4 &= \sigma_1. \end{aligned}$$

4. Newton's identities

Retaining the notation from before, now define the **power sums** of $r_1, \cdots, r_n$ to be

$$s_j = s_j(r_1, \cdots, r_n) = \begin{cases} \sum_{i=1}^n r_i^j & \text{for } j \geq 0 \\ 0 & \text{for } j < 0 \end{cases}$$

including $s_0 = n$. The power sums are clearly invariant under the action of S_n. We want to relate them to the elementary symmetric functions σ_j. Start from the general polynomial,

$$g(X) = \prod_{i=1}^n (X - r_i) = \sum_{j \in \mathbb{Z}} (-1)^j \sigma_j X^{n-j}.$$

Certainly

$$g'(X) = \sum_{j \in \mathbb{Z}} (-1)^j \sigma_j (n - j) X^{n-j-1}.$$

But also, the logarithmic derivative and geometric series formulas,

$$\frac{g'(X)}{g(X)} = \sum_{i=1}^n \frac{1}{X - r_i} \qquad \text{and} \qquad \frac{1}{X - r} = \sum_{k=0}^\infty \frac{r^k}{X^{k+1}},$$

give

$$\begin{aligned} g'(X) &= g(X) \cdot \frac{g'(X)}{g(X)} = g(X) \sum_{i=1}^{n} \sum_{k=0}^{\infty} \frac{r_i^k}{X^{k+1}} = g(X) \sum_{k \in \mathbb{Z}} \frac{s_k}{X^{k+1}} \\ &= \sum_{k,\ell \in \mathbb{Z}} (-1)^\ell \sigma_\ell s_k X^{n-k-\ell-1} \\ &= \sum_{j \in \mathbb{Z}} \left[\sum_{\ell \in \mathbb{Z}} (-1)^\ell \sigma_\ell s_{j-\ell} \right] X^{n-j-1} \quad \text{(letting } j = k + \ell\text{)}. \end{aligned}$$

Equate the coefficients of the two expressions for $g'(X)$ to get

$$\sum_{\ell=0}^{j-1} (-1)^\ell \sigma_\ell s_{j-\ell} + (-1)^j \sigma_j n = (-1)^j \sigma_j (n-j).$$

Newton's identities follow,

$$\sum_{\ell=0}^{j-1} (-1)^\ell \sigma_\ell s_{j-\ell} + (-1)^j \sigma_j j = 0 \qquad \text{for all } j.$$

Explicitly, Newton's identities are

$$\begin{aligned} & s_1 - \sigma_1 = 0 \\ & s_2 - s_1\sigma_1 + 2\sigma_2 = 0 \\ & s_3 - s_2\sigma_1 + s_1\sigma_2 - 3\sigma_3 = 0 \\ & s_4 - s_3\sigma_1 + s_2\sigma_2 - s_1\sigma_3 + 4\sigma_4 = 0 \\ & \text{and so on.} \end{aligned}$$

The identities show (exercise) that for any $j \in \{1, \cdots, n\}$, the power sums $s_1, \cdots, s_j$ are integer polynomials (with constant terms zero) in the elementary symmetric functions $\sigma_1, \cdots, \sigma_j$, while the elementary symmetric functions $\sigma_1, \cdots, \sigma_j$ are *rational* polynomials with constant terms zero) in the power sums $s_1, \cdots, s_j$. Consequently,

Proposition 4.1. *The first j coefficients $a_1, \cdots, a_j$ of the polynomial $f(X) = X^n + a_1 X^{n-1} + \cdots + a_n$ are zero exactly when the first j power sums of its roots are zero.*

5. Resultants

Given polynomials p and q, we can determine whether they have a root in common without actually finding their roots.

Let m and n be nonnegative integers. Let

$$a_0, \cdots, a_m, \quad b_0, \cdots, b_n, \quad (a_0 \neq 0,\ b_0 \neq 0)$$

be symbols (possibly elements of the base field $\mathbb{Q}$). Let the coefficient field be

$$k = \mathbb{Q}(a_0, \cdots, a_m, b_0, \cdots, b_n).$$

The polynomials

$$p(X) = \sum_{i=0}^{m} a_i X^{m-i}, \qquad q(X) = \sum_{i=0}^{n} b_i X^{n-i}$$

in $k[X]$ are utterly general when the a_i's and the b_i's form an algebraically independent set, or conversely they can be explicit polynomials when all the coefficients lie in $\mathbb{Q}$ or in $\mathbb{R}$ or in $\mathbb{C}$ or in some other extension field of $\mathbb{Q}$. It is an exercise to show that the polynomials p and q share a nonconstant factor in $k[X]$ if and only if there exist nonzero polynomials in $k[X]$,

$$P(X) = \sum_{i=0}^{n-1} c_i X^{n-1-i}, \qquad Q(X) = \sum_{i=0}^{m-1} d_i X^{m-1-i},$$

having respective degrees less than n and m, such that $pP = qQ$. Such P and Q exist if and only if the system

$$vM = 0$$

of $m+n$ linear equations over k in $m+n$ unknowns has a nonzero solution v, where

$$v = [c_0, c_1, \cdots, c_{n-1}, -d_0, -d_1, \cdots, -d_{m-1}]$$

lies in k^{m+n}, and M is the **Sylvester matrix**

$$M = \begin{bmatrix} a_0 & a_1 & \cdots & \cdots & a_m & & \\ & \ddots & \ddots & & & \ddots & \\ & & a_0 & a_1 & \cdots & \cdots & a_m \\ b_0 & b_1 & \cdots & b_n & & & \\ & b_0 & b_1 & \cdots & b_n & & \\ & & \ddots & \ddots & & \ddots & \\ & & & b_0 & b_1 & \cdots & b_n \end{bmatrix}$$

(n staggered rows of a_i's, m staggered rows of b_j's, all other entries 0), an $(m+n)$-by-$(m+n)$ matrix. Such a nonzero solution exists in turn if and only if $\det M = 0$. This determinant is called the **resultant** of p and q,

$$R(p, q) = \det M \in \mathbb{Z}[a_0, \cdots, a_m, b_0, \cdots, b_n].$$

The condition that p and q share a factor in $k[X]$ is equivalent to their sharing a root in the splitting field over k of pq. Thus the result is

Theorem 5.1. *The polynomials p and q in $k[X]$ share a nonconstant factor in $k[X]$, or equivalently, share a root in the splitting field over k of their product, if and only if $R(p, q) = 0$.*

When the coefficients of p and q are algebraically independent, $R(p, q)$ is a master formula that applies to all polynomials of degrees m and n. At the other extreme, if the coefficients lie in some numerical superfield of $\mathbb{Q}$ then $R(p, q)$ is a number that is zero or nonzero depending on whether the particular polynomials p and q share a factor.

Taking the resultant of p and q to check whether they share a root can also be viewed as eliminating the variable X from the pair of equations $p(X) = 0$ and $q(X) = 0$, leaving one equation $R(p, q) = 0$ in the remaining variables $a_0, \cdots, a_m$, $b_0, \cdots, b_n$.

In principle, evaluating $R(p, q) = \det M$ can be carried out via a process of row and column operations. (Using only row operations encompasses computing the greatest common divisor of p and q by the Euclidean algorithm.) In practice, evaluating a large determinant is an error-prone process by hand. The next theorem will supply as a corollary a more efficient method to compute $R(p, q)$. In any case, since any worthwhile computer symbolic algebra package is equipped with a resultant function, nontrivial resultants can often be found by machine.

In their splitting field over k, the polynomials p and q factor as

$$p(X) = a_0 \prod_{i=1}^{m}(X - r_i), \qquad q(X) = b_0 \prod_{j=1}^{n}(X - s_j).$$

To express the resultant $R(p,q)$ explicitly in terms of the roots of p and q introduce the quantity $\tilde{R}(p,q) = a_0^n b_0^m \prod_{i=1}^m \prod_{j=1}^n (r_i - s_j)$. This polynomial vanishes if and only if p and q share a root, so it divides $R(p,q)$. Note that $\tilde{R}(p,q)$ is homogeneous of degree mn in the r_i and s_j. On the other hand, each coefficient $a_i = a_0(-1)^i\sigma_i(r_1, \cdots, r_m)$ of p has homogeneous degree i in $r_1, \cdots, r_m$, and similarly for each b_j and $s_1, \cdots, s_n$. Thus in the Sylvester matrix the (i,j)th entry has degree

$$\begin{cases} j - i \text{ in the } r_i & \text{if } 1 \le i \le n,\ i \le j \le i+m, \\ j - i + n \text{ in the } s_j & \text{if } n+1 \le i \le n+m,\ i-n \le j \le i. \end{cases}$$

It quickly follows that any nonzero term in the determinant $R(p,q)$ has degree mn in the r_i and the s_j, so $\tilde{R}(p,q)$ and $R(p,q)$ agree up to multiplicative constant. Matching coefficients of $(s_1 \cdots s_n)^m$ shows that the constant is 1. This proves

Theorem 5.2. *The resultant of the polynomials*

$$p(X) = \sum_{i=0}^{m} a_i X^{m-i} = a_0 \prod_{i=1}^{m}(X - r_i), \quad q(X) = \sum_{j=0}^{n} b_j X^{n-j} = b_0 \prod_{j=1}^{n}(X - s_j)$$

is given by the formulas

$$R(p,q) = a_0^n b_0^m \prod_{i=1}^{m}\prod_{j=1}^{n}(r_i - s_j) = a_0^n \prod_{i=1}^{m} q(r_i) = (-1)^{mn} b_0^m \prod_{j=1}^{n} p(s_j).$$

A special case of this theorem gives the efficient formula for the discriminant promised earlier. See the exercises.

Computing resultants can now be carried out via a Euclidean algorithm procedure: repeatedly do polynomial division with remainder and apply formula (4) in

Corollary 5.3. *The following formulas hold:*

(1) $R(q,p) = (-1)^{mn} R(p,q)$.
(2) $R(p\tilde{p}, q) = R(p,q)R(\tilde{p},q)$ *and* $R(p, q\tilde{q}) = R(p,q)R(p,\tilde{q})$.
(3) $R(a_0, q) = a_0^n$ *and* $R(a_0X + a_1, q) = a_0^n q(-a_1/a_0)$.
(4) *If* $q = Qp + \tilde{q}$ *with* $\deg(\tilde{q}) < \deg(p)$ *then*

$$R(p,q) = a_0^{\deg(q) - \deg(\tilde{q})} R(p, \tilde{q}).$$

The proof of the corollary is an exercise.