Symmetric Polynomials: Polys

Jonathan L.F. King
University of Florida, Gainesville FL 32611-2082, USA
squash@ufl.edu
Webpage http://squash.1gainesville.com/
14 October, 2017 (at 11:56)

ABSTRACT: Proves the Fund. Thm of Symmetric Polynomials, that each sympoly is determined by the elementary symmetric polynomials; the esps. Indeed, each N-variable sympoly is a polynomial in the N-variable esps.

Along the way, this note provides an introduction to multinomial coefficients.

(See “∼/Elisp/SymPoly/symmetricpoly.el for code.)

Review of multinomial coefficients. For a natnum n, use “n!” to mean “n factorial”; the product of all posints ≤ n. So 3! = 3 · 2 · 1 = 6 and 5! = 120. Also 0! = 1 and 1! = 1.

The binomial coefficient \( \binom{n}{k} \), read “7 choose 3”, means the number of ways of choosing 3 objects from 7 distinguishable objects. If we think of putting these objects in our left pocket, and putting the remaining 4 objects in our right pocket, then we write the coefficient as \( \binom{7}{3,4} \). [Read as “7 choose 3-comma-4.”]

Note that \( \binom{7}{0,7} = 1 \). Observe that \( \binom{N+1}{k+1} = \binom{N}{k} + \binom{N}{k+1} \). Finally, the Binomial theorem says

B1: \[ (x+y)^N = \sum_{j+k=N} \binom{N}{j,k} x^j y^k, \]

where \((j,k)\) ranges over all ordered pairs of natural numbers with sum \(N\).

In general, for natnums \(N = k_1 + \ldots + k_P\), the multinomial coefficient \( \binom{N}{k_1,k_2,\ldots,k_P} \) is the number of ways of partitioning \(N\) objects, by putting \(k_1\) objects in pocket-one, \(k_2\) objects in pocket-two, \ldots putting \(k_j\) objects in the \(j\)th pocket. Easily

B2: \[ \binom{N}{k_1,k_2,\ldots,k_P} = \frac{N!}{k_1! \cdot k_2! \cdot \ldots \cdot k_P!}. \]

And \((x_1 + \cdots + x_P)^N\) indeed equals the sum of terms

\[ \binom{N}{k_1,k_2,\ldots,k_P} \cdot x_1^{k_1} \cdot x_2^{k_2} \cdots x_P^{k_P}, \]

taken over all natnum-tuples \(\vec{k}=(k_1,\ldots,k_P)\) that sum to \(N\).

Preliminaries. Fix a posint \(N\) and let “perm” mean a permutation of \([1..N]\). Given an \(N\)-variable poly \(Y(z_1,\ldots,z_N)\) and a perm \(\pi\), say that \(Y\) is \(\pi\)-invariant if

\[ Y(z_1,\ldots,z_N) = Y(z_{\pi(1)}, z_{\pi(2)}, \ldots, z_{\pi(N)}), \]

where the equality is “as polynomials”; that is, the corresponding coefficients are equal. Say that \(Y\) is a symmetric polynomial in \(z_1,\ldots,z_N\) if \(Y\) is invariant \(^{\heartsuit}1\) under each of the \(N\) many perms. I will use sympoly to abbreviate “symmetric polynomial”.

1: Modification. For notational convenience, henceforth assume that sympolys have no constant term. This does not affect the below ESP Theorem, (3), since the constant of \(Y()\) would simply become the constant term of \(F()\).

ESP. In the sequel, use \(\vec{z}\) to denote \(z_1,\ldots,z_N\) or \((z_1,\ldots,z_N)\) as appropriate.

For \(N=3\), there are three especially simple sympolys:

\[ \sigma_1(\vec{z}) := z_1 + z_2 + z_3 ; \]
\[ \sigma_2(\vec{z}) := z_1 z_2 + z_1 z_3 + z_2 z_3 ; \]
\[ \sigma_3(\vec{z}) := z_1 z_2 z_3 . \]

For a general \(N\) and \(j \in [1..N]\), let

2: \[ \sigma_{j,N}(\vec{z}) := \sum_{S \subseteq [1..N]} [\prod_{j \in S} z_j] , \]

where \(S\), here, ranges over the cardinality-\(j\) subsets of \([1..N]\). Henceforth, the subscript “\(N\)” will be implicit and I will write \(\sigma_j\) in place of \(\sigma_{j,N}\).

These sympolys \(\sigma_1,\ldots,\sigma_N\) are called the elementary symmetric polynomials, or esps, for short.

The goal of this note is the following result.

\(^{\heartsuit}1\)Consider the polynomial \(Y(x,y) := x^2 + y\) as a poly over the integers mod-2. Then \(Y\) is certainly a symmetric function, but it is not a “symmetric polynomial” because its coefficient structure is not invariant under exchanging \(x\) and \(y\).
3: Fundamental ESP Theorem. Fix an \( N \)-variable \( \mathbb{Z} \)-sympoly \( Y \). Then there is a unique \( N \)-variable \( \mathbb{Z} \)-poly \( F \) such that

\[
F(\sigma_1(z), \sigma_2(z), \ldots, \sigma_N(z)) = Y(z).
\]

Additionally, \( \text{Deg}(F) \leq \text{Deg}(Y) \). \( \Box \)

**Convention.** Typically I will let \( \sigma_j \) also mean \( \sigma_j(z) \). In particular, I will usually write \( \text{LHS}(3') \) as \( F(\sigma_1, \ldots, \sigma_N) \) or just as \( F(\sigma) \).

When I use a specific value of \( N \) in an example, I may use variable names \( a, b, \ldots \) in place of \( z_1, z_2, \ldots \).

**Remark.** With alternating signs, the coefficients of a monic 1-var poly \( g() \) appear as the esps of its zeros \( z \) (also called “roots” of \( g \)). I.e., consider

\[
g(X) := [X - z_1][X - z_2] \cdots [X - z_N].
\]

Multiplying this out, \( g(X) \) equals

\[
X^N - \sigma_1(z) \cdot X^{N-1} + \sigma_2(z) \cdot X^{N-2} - \sigma_3(z) \cdot X^{N-3} + \cdots + (-1)^N \sigma_N(z) \cdot X^0.
\]

Consequently, the ESP Theorem has the handy corollary that each symmetric polynomial of the roots of \( g() \) is some nice polynomial in the coefficients of \( g \).

As an application, the “discriminant of \( g \)” is

\[
\text{Discr}(g) := \prod_{1 \leq j < k \leq N} [z_k - z_j]^2.
\]

Letting \( B \) be the binomial coeff \( \binom{N}{2} \), the RhS equals

\[
[(-1)^B] \cdot \prod_{j \neq k, j,k \in [1..N]} [z_k - z_j],
\]

so we see that \( \text{Discr}(g) \) does \textbf{not} depend on some arbitrary ordering of the roots. The ESP Thm tells us that we can compute the discriminant of \( g \) from its coefficients. \( \Box \)

---

**Tools**

Here are the utensils that we will use to obtain the ESP Theorem.

**Dictionary order.** We define \textit{dictionary order}, \( \prec \), on pairs of \( N \)-tuples \( \vec{v} = (v_1, \ldots, v_N) \) and \( \vec{w} = (w_1, \ldots, w_N) \), each a tuple of ntums. Say that \( \vec{v} \prec \vec{w} \) if:

\[
\text{There exists an index } j \in [1..N] \text{ with } v_j \neq w_j. \text{ For the smallest such } j, \text{ furthermore, } v_j < w_j.
\]

Easily, dictionary order \( \prec \) is \textit{transitive} (\( \vec{u} \prec \vec{v} \) and \( \vec{v} \prec \vec{w} \) together imply that \( \vec{u} \prec \vec{w} \)). And dictionary order is a \textit{total} order. That is, for each pair \( \vec{v} \) and \( \vec{w} \), one of them is \( \leq \) the other one.

**Profiles.** A sympoly can be written more concisely than just listing all of its terms. Consider the \( N=3 \) sympoly \( Y() \) with the smallest number of terms and possessing term \( a^9 b^4 c \). Necessarily \( Y(a, b, c) \) equals this sum of six terms:

\[
a^9 b^4 c + a^9 b^4 c + a^4 b^9 c + a^4 b^9 c + a b^9 c^4 + a b^4 c^9
\]

What we have done is take the exponent triple \((9, 4, 1)\) and put it over \((a, b, c)\) in all possible ways.

In general, an exponent \( N \)-tuple, \( \alpha \), can always be put in decreasing (well . . . non-increasing, actually) order as \( \alpha = (h_1, \ldots, h_N) \) with \( h_1 \geq \ldots \geq h_N \), with each of the exponents a ntum. Furthermore, courtesy (1), necessarily \( h_1 \geq 1 \). We take this as our definition of a \textit{profile}.

Let’s refer to the \( N \) ntums \( h_n \) as heights. They are indeed shown as such in Figure 9, below. Let \( \text{Deg}(\alpha) \) denote the sum \( h_1 + \ldots + h_N \).

**The Canonical Rep.** Let \( S^\alpha(z) \) be the sympoly having the fewest terms and having term \( z_{h_1}^{h_1} \cdot z_{h_2}^{h_2} \cdots z_{h_N}^{h_N} \). Letting \( \pi \), below, range over all permutations of \([1..N]\), then,

\[
5: \quad S^\alpha(z) := \frac{1}{B} \cdot \sum_{\pi} z_{\pi(1)}^{h_{\pi(1)}} \cdot z_{\pi(2)}^{h_{\pi(2)}} \cdots z_{\pi(N)}^{h_{\pi(N)}},
\]

\( ^2 \)Also known as \textit{lexicographic order}. 

Filename: Problems/Polynomials/symmetric_polys.latex
where we need to define the posint $B$ appropriately so that, after combining like-terms, each term will have a coefficient of 1.

To compute $B$, we count the number of height-repetitions in our $N$-tuple $\alpha$. Let $v_1, \ldots, v_P$ be the values of $\alpha$, and write

$$\alpha = \left( \frac{v_1, \ldots, v_1, v_2, \ldots, v_2, \ldots, v_P, \ldots, v_P}{d_1, d_2, \ldots, d_P} \right)$$

with strict inequality $v_1 > \cdots > v_P$. (So $v_1$ equals $h_1$ and $v_P = h_N$. There are $d_j$ many copies of the $j^{\text{th}}$ value.) It is straightforward to show that $B$ equals the product $[d_1!] \cdot [d_2!] \cdots [d_P!]$. Thus $S^\alpha$ has a multinomial-coeff number of terms, namely $(\alpha_1, \ldots, \alpha_K)$.

The upshot is this: Using dictionary order on profiles, an arbitrary sympoly $Y()$ has a canonical representation (abbr. canon-rep)

$$6: \quad Y = \sum_{k=1}^{K} q_k S^{\alpha_k},$$

where $\alpha_1 > \alpha_2 > \cdots > \alpha_K$ and each $q_k$ is a non-zero integer.

The natnum $K$, coefficients $q_k$, and profile sequence $\alpha_1, \ldots, \alpha_K$ are uniquely determined.

Establishing the ESP Theorem

An $N$-profile can be viewed as in Figure 9. It determines a decreasing tuple of widths $[w_1, \ldots, w_L]$. Notice that $w_L$ will always be the number of non-zero heights. Thus necessarily $L \leq N$, with equality IFF $\alpha$ is the profile of an ESP. Whoa! Did I mean iff $h_1 = 1$? I.e $L = 1$?

The key step in proving the theorem is this lemma.

7: Width Lemma. Fix a profile $\alpha$ and let $w_1 \leq \cdots \leq w_L$ be its tuple of widths. Then there is a posint $J$ so that the esp-product

$$\sigma_{w_1} \cdot \sigma_{w_2} \cdots \sigma_{w_L},$$

has canonical representation

$$8: \quad \sigma_{w_1} \cdot \sigma_{w_2} \cdots \sigma_{w_L} = S^\alpha + \sum_{j=2}^{J} q_j S^{\beta_j},$$

for some coefficients $q_j$ and profiles $\beta_1 < \beta_{j-1} < \cdots < \beta_2 < \alpha$. In other words, the original profile $\alpha$ is the highest profile in the esp-product (8), and it occurs with a coefficient of 1.\hfill

\begin{figure}[h]
\centering
\begin{tikzpicture}
\draw [very thin, step=0.5] (0,0) grid (5,4);
\draw (0,0) -- (5,0);
\draw (0,1) -- (5,1);
\draw (0,2) -- (5,2);
\draw (0,3) -- (5,3);
\draw (0,4) -- (5,4);
\draw (1,0) -- (1,4);
\draw (2,0) -- (2,4);
\draw (3,0) -- (3,4);
\draw (4,0) -- (4,4);
\draw (0,1) -- (5,1);
\draw (0,2) -- (5,2);
\draw (0,3) -- (5,3);
\draw (0,4) -- (5,4);
\node at (0.5,0.5) {$w_1$};
\node at (1.5,0.5) {$w_2$};
\node at (2.5,0.5) {$\vdots$};
\node at (3.5,0.5) {$w_L$};
\node at (0.5,1.5) {$h_1$};
\node at (1.5,1.5) {$h_2$};
\node at (2.5,1.5) {$h_3$};
\node at (3.5,1.5) {$h_4$};
\node at (4.5,1.5) {$h_5$};
\node at (0.5,2.5) {$h_6$};
\node at (1.5,2.5) {$h_7$};
\node at (2.5,2.5) {$h_8$};
\node at (3.5,2.5) {$h_9$};
\node at (4.5,2.5) {$h_N$};
\end{tikzpicture}
\caption{This shows a profile $\alpha$ having heights $h_1 \geq \ldots \geq h_N$. In this example, heights $h_8$ through $h_N$ are zero. The picture also defines $L$ many widths $w_1 \leq \ldots \leq w_L$, where $L := h_1$. In this example $L = 4$ and the widths are $w_1 = 2$, $w_2 = 3$, $w_3 = 6$ and $w_4 = 7$.}
\end{figure}

10: Example. Let $\alpha := (2, 1, 1, 0)$; so $N = 4$. The width tuple for $\alpha$ is $[1, 3]$, hence the corresponding esp-product is $\sigma_1 \cdot \sigma_3$, i.e., is

$$*: \quad [a + b + c + d][abc + abd + acd + bcd].$$

The only Deg-4 profile which is “dictionaries” below $\alpha$ is $\beta := (1, 1, 1, 1)$. Collecting terms in $(*)$ gives that

$$8': \quad \sigma_1 \cdot \sigma_3 = S^\alpha + 4S^\beta.$$

So $(8')$ is (8) with $J = 2$ and $q_2 = 4$.

There is an alternative way to derive $(8')$, without multiplying out. Courtesy the lemma, we will get one copy of $\alpha$ and some unknown number, $q_2$, of copies of $\beta$. But the repetition tuple for $\alpha$ is $(1, 2, 1)$, so $S^\alpha$ has $(1, 2, 1) = 12$ terms. Thus 12 of the 16 = 4 · 4 products of $(*)$ will be terms of the single copy of $S^\alpha$.

In consequence, the remaining $16 - 12 = 4$ products, when divided by the number of terms in $S^\beta$,
must equal \( q \). Since \( S^\beta \) has only a single term, \( abcd \), we conclude that \( q = 4/1 \); this is also what we saw by multiplying out. 

\[ \]

**Sketch of proof of the Width Lemma.** There is no true loss of generality in assuming that \( N = 5 \); now I can write the variables as \( a, \ldots, e \) rather than \( z_1, \ldots, z_5 \). Referring to (6), let

\[ 11: \quad p \cdot S^{\alpha'} + \sum_{k=2}^{K} q_k S^{\beta_k} \]

be the canon-rep of the esp-product \( \sigma_{w_1} \cdot \sigma_{w_2} \cdots \sigma_{w_L} \).

Our goal is to show that \( \alpha' \) equals \( \alpha \) and that \( p = 1 \). It is implicit in the argument below (Exercise!) that no profile dictionary greater than \( \alpha \) can occur in (11). So, I will content myself with showing that the coeff of \( \alpha \) in (11) is 1.

The coeff of \( \alpha \) is simply the coeff of the term

\[ a^{h_1} b^{h_2} c^{h_3} d^{h_4} e^{h_5} \]

in esp-product \( \sigma_{w_1} \cdots \sigma_{w_L} \). In each of the \( L \) many eps, the degree of “\( a \)” is 1. Thus the highest degree of “\( a \)” in the esp-product is \( L \), which indeed equals \( h_1 \). How many of the variables \( a, b, \ldots \) can have this maximal exponent \( L \) — at most \( w_1 \) of them. So the only way to get term (**) in the esp-product is if the term used from esp \( \sigma_{w_1} \) was in fact the term \( abcd \).

Rest of proof is missing, 17Sep2001.

Given a second such list

\[ x = \langle (\beta_1, q_1), \ldots, (\beta_K, q_K) \rangle, \]

let \( \text{merge-sort}(y, x) \) be the interwoven list of pairs, sorted by \( \succ \). Furthermore, like-terms are combined and terms with coeff zero are dropped. For example, suppose that \( \alpha \succ \beta \succ \gamma \succ \delta \succ \varepsilon \) are profiles and our lists are

\[ y := \langle (\beta, 20), (\gamma, -30), (\varepsilon, 40) \rangle; \]
\[ x := \langle (\alpha, 5), (\beta, 6), (\gamma, 30), (\delta, 7), (\varepsilon, 8) \rangle. \]

Then \( \text{merge-sort}(y, x) \) equals

\[ \langle (\alpha, 5), (\beta, 26), (\delta, 7), (\varepsilon, 48) \rangle. \]

**The setup.** The given sympoly has canonical representation

\[ \text{IN:} \quad Y = \sum_{k=1}^{K} c_k S^{\alpha_k}. \]

The desired output poly \( F \) has form

\[ \text{OUT:} \quad F = \sum_{j=1}^{J} d_j Y_j, \]

where each \( Y_j \) is OTForm \( \sigma_{w_1} \cdots \sigma_{w_L} \) for numbers \( L \) and \( w_1 \leq \ldots \leq w_L \) which depend on \( j \).

**An algorithm**

The algorithm repeats \textit{Steps 1, 2, 3} below until list \( y \) becomes empty. The value of list \( f \) then describes the desired poly \( F \).

\[ \]

**Initialization:** Init \( y \) to list (12) from (IN). Init \( f \) to the empty list \( \langle \rangle \).

**Step 1:** If the current \( y \) is empty then STOP; the current value of \( f \) describes (OUT).

Otherwise, \textit{pop-off} of \( y \) its leftmost pair and call this pair \( (\alpha, c) \). Compute \( L \) and widths \( w_1 \leq \ldots \leq w_L \) for profile \( \alpha \).
Step 2: Courtesy the Width Lemma, we can write the product $\sigma_{w_1} \cdots \sigma_{w_L}$ as

$$\langle (\alpha, 1), (\beta_2, q_2), \ldots, (\beta_J, q_J) \rangle$$

where $\alpha \succ \beta_2 \succ \ldots \succ \beta_J$ are profiles.

Push pair $([w_1, w_2, \ldots, w_L], c)$ onto $f$.

Step 3: Define a list

$$x := \langle (\beta_2, -cq_2), (\beta_3, -cq_3), \ldots, (\beta_J, -cq_J) \rangle;$$

note the negated coefficients. Now assign

$$y := \text{merge-sort}(y, x).$$

(Let $K$ denote the number of pairs in the old $y$. Then the number of pairs in the new $y$ may be as much as $K + J - 2$. But the maximum profile in the new $y$ is strictly $\prec$ the maximum profile in old $y$. So eventually $y$ will become exhausted.)

Efficiency considerations

An expensive part is Step 2, computing the canonrep of the esp-product.

To be written.