Symmetric Polynomials: Polys

Jonathan L.F. King
University of Florida, Gainesville FL 32611-2082, USA
squash@ufl.edu
Webpage http://squash.1gainesville.com/
23 August, 2016 (at *00:07*)

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.)

Preliminaries. Fix a posint N and let "perm" mean a permutation of $[1..N]$. Given an N-variable poly $Y(z_1,\dots,z_N)$ and a perm π, say that Y is ***π-invariant*** if

$$Y(z_1,\dots,z_N) = Y\big(z_{\pi(1)}, z_{\pi(2)},\dots,z_{\pi(N)}\big),$$

where the equality is "as polynomials"; that is, *corresponding coefficients are equal.* Say that Y is a "***symmetric polynomial*** in $z_1,\dots,z_N$" if Y is invariant[♡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 $z_1,\dots,z_N$ or $(z_1,\dots,z_N)$ as appropriate.

[♡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.

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

$$\begin{aligned}\sigma_1(\vec{z}) &:= z_1+z_2+z_3\,;\\ \sigma_2(\vec{z}) &:= z_1z_2+z_1z_3+z_2z_3\,;\\ \sigma_3(\vec{z}) &:= z_1\cdot z_2\cdot z_3\ .\end{aligned}$$

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

2: $$\sigma_{j,N}(\vec{z}) := \sum_S\Big[\prod_{j\in S} z_j\Big],$$

where S, here, ranges over the cardinality-j subsets of $[1..N]$. Henceforth, the subscript "N" will be implicit and I will write σ_j in place of $\sigma_{j,N}$.

These sympolys $\sigma_1,\dots,\sigma_N$ are called the ***elementary symmetric polynomials***, or ***esp***s, for short.

The goal of this note is the following result.

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*

3′: $$F\Big(\sigma_1(\vec{z}),\sigma_2(\vec{z}),\dots,\sigma_N(\vec{z})\Big) = Y(\vec{z}).$$

Additionally, $\mathrm{Deg}(F) \leqslant \mathrm{Deg}(Y)$. ◊

Convention. Typically I will let σ_j also mean $\sigma_j(\vec{z})$. In particular, I will usually write LhS(3′) as $F(\sigma_1,\dots,\sigma_N)$ or just as $F(\vec{\sigma})$.

When I use a specific value of N in an example, I may use variable names $a,b,\dots$ in place of $z_1,z_2,\dots$. □

Remark. With alternating signs, the coefficients of a monic 1-var poly $g()$ appear as the esps of its zeros $\vec{z}$ (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

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

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

$$4\text{:}\qquad \operatorname{Discr}(g) := \prod_{1\leqslant j<k\leqslant N} [z_k - z_j]^2 .$$

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

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

so we see that $\operatorname{Discr}(g)$ does *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. □

Tools

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

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

> *There exists an index $j \in [1\,..\,N]$ with $v_j \neq w_j$. For the smallest such j, furthermore, $v_j < w_j$.*

Easily, dictionary order[♡2] is *transitive* ($\vec{u} \preccurlyeq \vec{v}$ and $\vec{v} \preccurlyeq \vec{w}$ together imply that $\vec{u} \preccurlyeq \vec{w}$). And dictionary order is a *total* order. That is, for each pair $\vec{v}$ and $\vec{w}$, one of them is $\preccurlyeq$ the other one.

Multinomial coefficients. Given natnums $d_1 + d_2 = N$, we know that the number of ways of choosing, from N distinct objects, d_1 of them to put in my lefthand pocket, is $\binom{N}{d_1}$.

I can put the other d_2 objects in my righthand pocket. Writing $\binom{N}{d_1}$ as $\binom{N}{d_1,d_2}$, then, I can interpret this last number as: *The number of ways of partitioning N objects into two distinguished pockets, putting d_1 objects in the first pocket and d_2 in the second.*

Now suppose that I have P distinguished pockets and natnums $d_1 + \cdots + d_P = N$. Define the ***multinomial coefficient***[♡3]

$$5\text{:}\qquad \binom{N}{d_1, d_2, \ldots, d_P}$$

to be the *number of ways of partitioning N distinguished objects into P many distinguished pockets, putting d_j objects in the j^{th} pocket, for each $j \in [1\,..\,P]$.*

This multinomial coefficient equals

$$5'\text{:}\qquad \frac{N!}{d_1! \cdot d_2 \cdot \ldots \cdot d_P!} .$$

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 $a^9 \cdot b^4 \cdot c$. Necessarily $Y(a, b, c)$ equals this sum of six terms:

$$a^9b^4c + a^9bc^4 + a^4b^9c + a^4bc^9 + ab^9c^4 + ab^4c^9$$

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, $\boldsymbol{\alpha}$, can always be put in decreasing (well... non-increasing, actually) order as $\boldsymbol{\alpha} = (h_1, \ldots, h_N)$ with $h_1 \geqslant \ldots \geqslant h_N$, with each of the exponents a natnum. Furthermore, courtesy (1), necessarily $h_1 \geqslant 1$. We take this as our definition of a ***profile***.

Let's refer to the N natnums h_n as ***heights***. They are indeed shown as such in Figure 10, below. Let $\operatorname{Deg}(\boldsymbol{\alpha})$ denote the sum $h_1 + \ldots + h_N$.

[♡2] Also known as ***lexicographic order***.

[♡3] Pronounced "N choose $d_1, d_2, \ldots, d_P$".

The Canonical Rep. Let $S^{\boldsymbol{\alpha}}(\vec{z})$ be the sympoly having the fewest terms and having term $z_1^{h_1} \cdot z_2^{h_2} \cdots z_N^{h_N}$. Letting π, below, range over all permutations of $[1 .. N]$, then,

$$6: \quad S^{\boldsymbol{\alpha}}(\vec{z}) \;:=\; \tfrac{1}{B} \cdot \sum_{\pi} z_1^{h_{\pi(1)}} \cdot z_2^{h_{\pi(2)}} \cdots z_N^{h_{\pi(N)}} ,$$

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 $\boldsymbol{\alpha}$. Let $v_1, \ldots, v_P$ be the values of $\boldsymbol{\alpha}$, and write

$$\boldsymbol{\alpha} \;=\; \Big(\underbrace{v_1, \ldots, v_1,}_{d_1} \underbrace{v_2, \ldots, v_2,}_{d_2} \ldots, \underbrace{v_P, \ldots, v_P,}_{d_P}\Big)$$

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^{th} value.) It is straightforward to show that B equals the product $[d_1!] \cdot [d_2!] \cdots [d_P!]$. Thus $S^{\boldsymbol{\alpha}}$ has a multinomial-coeff number of terms, namely $(5')$.

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

$$7: \; Y \;=\; \sum_{k=1}^{K} \mathsf{q}_k S^{\boldsymbol{\alpha}_k} , \qquad \text{where } \boldsymbol{\alpha}_1 \succ \boldsymbol{\alpha}_2 \succ \ldots \succ \boldsymbol{\alpha}_K \text{ and each } \mathsf{q}_k \text{ is a non-zero integer.}$$

The natnum K, coefficients q_k, and profile sequence $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_K$ are uniquely determined.

Establishing the ESP Theorem

An N-profile can be viewed as in Figure 10. 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 $L \leqslant N$, with equality IFF $\boldsymbol{\alpha}$ 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.

8: Width Lemma. *Fix a profile $\boldsymbol{\alpha}$ and let $w_1 \leqslant \ldots \leqslant w_L$ be its tuple of widths. Then there is a posint J so that the esp-product*

$$\sigma_{w_1} \cdot \sigma_{w_2} \cdots \sigma_{w_L} , \quad (\text{evidently a sympoly})$$

has canonical representation

$$9: \qquad \sigma_{w_1} \cdot \sigma_{w_2} \cdots \sigma_{w_L} \;=\; S^{\boldsymbol{\alpha}} + \sum_{j=2}^{J} \mathsf{q}_j S^{\boldsymbol{\beta}_j} ,$$

for some coefficients q_j and profiles $\boldsymbol{\beta}_J \prec \boldsymbol{\beta}_{J-1} \prec \ldots \prec \boldsymbol{\beta}_2 \prec \boldsymbol{\alpha}$. In other words, the original profile $\boldsymbol{\alpha}$ is the highest profile in the esp-product (9), and it occurs with a coefficient of 1. ♢

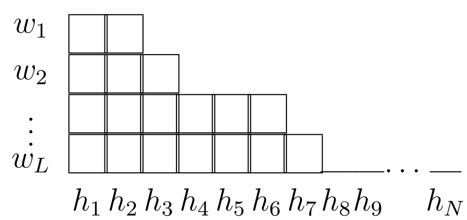

FIG. 10: *This shows a profile $\boldsymbol{\alpha}$ having heights $h_1 \geqslant \ldots \geqslant h_N$. In this example, heights h_8 through h_N are zero. The picture also defines L many widths $w_1 \leqslant \ldots \leqslant w_L$, where $L := h_1$. In this example $L = 4$ and the widths are $w_1 = 2$, $w_2 = 3$, $w_3 = 6$ and $w_4 = 7$.*

11: *Example.* Let $\boldsymbol{\alpha} := \big(2, 1, 1, 0\big)$; so $N = 4$. The width tuple for $\boldsymbol{\alpha}$ 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 "dictionarily" below $\boldsymbol{\alpha}$ is $\boldsymbol{\beta} := \big(1, 1, 1, 1\big)$. Collecting terms in $(*)$ gives that

$$9': \qquad \sigma_1 \cdot \sigma_3 \;=\; S^{\boldsymbol{\alpha}} \;+\; 4 S^{\boldsymbol{\beta}} .$$

So $(9')$ is (9) with $J = 2$ and $\mathsf{q}_2 = 4$.

There is an alternative way to derive $(9')$, without multiplying out. Courtesy the lemma, we will get one copy of $\boldsymbol{\alpha}$ and some unknown number, q, of copies of $\boldsymbol{\beta}$. But the repetition tuple for $\boldsymbol{\alpha}$ is $\mathbf{(}1,2,1\mathbf{)}$, so $S^{\boldsymbol{\alpha}}$ has $\binom{4}{1,2,1}$=12 terms. Thus 12 of the $16 = 4 \cdot 4$ products of $(*)$ will be terms of the single copy of $S^{\boldsymbol{\alpha}}$.

In consequence, the remaining $16-12 = 4$ products, when divided by the number of terms in $S^{\boldsymbol{\beta}}$, must equal q. Since $S^{\boldsymbol{\beta}}$ has only a single term, *abcd*, we conclude that $\mathsf{q} = 4/1$; 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 (7), let

12: $$\mathsf{p} \cdot S^{\boldsymbol{\alpha}'} + \sum_{k=2}^{K} \mathsf{q}_k S^{\boldsymbol{\beta}_k}$$

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 $\boldsymbol{\alpha}'$ equals $\boldsymbol{\alpha}$ and that $\mathsf{p} = 1$. It is implicit in the argument below (Exercise!) that no profile dictionarily greater than $\boldsymbol{\alpha}$ can occur in (12). So, I will content myself with showing that the coeff of $\boldsymbol{\alpha}$ in (12) is 1.

The coeff of $\boldsymbol{\alpha}$ 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 esps, 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 σ_{w_1} was in fact the term *abcde*.

Rest of proof is missing, 17Sep2001. ♦

An implementation

The algorithm will manipulate lists OTForm

13: $$\mathsf{y} = \Big\langle \mathbf{(}\boldsymbol{\alpha}_1, \mathsf{c}_1\mathbf{)}, \mathbf{(}\boldsymbol{\alpha}_2, \mathsf{c}_2\mathbf{)}, \ldots, \mathbf{(}\boldsymbol{\alpha}_K, \mathsf{c}_K\mathbf{)} \Big\rangle,$$

where $\boldsymbol{\alpha}_1 \succ \ldots \succ \boldsymbol{\alpha}_K$ are profiles and the coeffs c_k are non-zero integers.

Given a second such list

$$\mathsf{x} = \Big\langle \mathbf{(}\boldsymbol{\beta}_1, \mathsf{q}_1\mathbf{)}, \ldots, \mathbf{(}\boldsymbol{\beta}_K, \mathsf{q}_K\mathbf{)} \Big\rangle,$$

let `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 $\boldsymbol{\alpha} \succ \boldsymbol{\beta} \succ \boldsymbol{\gamma} \succ \boldsymbol{\delta} \succ \boldsymbol{\varepsilon}$ are profiles and our lists are

$$\mathsf{y} := \Big\langle \mathbf{(}\boldsymbol{\beta}, 20\mathbf{)}, \mathbf{(}\boldsymbol{\gamma}, \text{-}30\mathbf{)}, \mathbf{(}\boldsymbol{\varepsilon}, 40\mathbf{)} \Big\rangle;$$
$$\mathsf{x} := \Big\langle \mathbf{(}\boldsymbol{\alpha}, 5\mathbf{)}, \mathbf{(}\boldsymbol{\beta}, 6\mathbf{)}, \mathbf{(}\boldsymbol{\gamma}, 30\mathbf{)}, \mathbf{(}\boldsymbol{\delta}, 7\mathbf{)}, \mathbf{(}\boldsymbol{\varepsilon}, 8\mathbf{)} \Big\rangle.$$

Then `merge-sort`(y, x) equals

$$\Big\langle \mathbf{(}\boldsymbol{\alpha}, 5\mathbf{)}, \mathbf{(}\boldsymbol{\beta}, 26\mathbf{)}, \mathbf{(}\boldsymbol{\delta}, 7\mathbf{)}, \mathbf{(}\boldsymbol{\varepsilon}, 48\mathbf{)} \Big\rangle.$$

The setup. The given sympoly has canonical representation

IN: $$Y = \sum_{k=1}^{K} \mathsf{c}_k S^{\boldsymbol{\alpha}_k}.$$

The desired output poly F has form

OUT: $$F = \sum_{j=1}^{J} \mathsf{d}_j \Upsilon_j,$$

where each Υ_j is OTForm $\sigma_{w_1} \cdots \sigma_{w_L}$ for numbers L and $w_1 \leqslant \ldots \leqslant w_L$ which depend on j.

An algorithm

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

Initialization: Init $\mathtt{y}$ to list (13) from (IN). Init $\mathtt{f}$ to the empty list $\langle\rangle$.

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

Otherwise, *pop-off* of $\mathtt{y}$ its leftmost pair and call this pair $(\boldsymbol{\alpha}, \mathsf{c})$. Compute L and widths $w_1 \leqslant \ldots \leqslant w_L$ for profile $\boldsymbol{\alpha}$.

Step 2: Courtesy the Width Lemma, we can write the product $\sigma_{w_1} \cdots \sigma_{w_L}$ as

$$\Big\langle (\boldsymbol{\alpha}, 1), (\boldsymbol{\beta}_2, \mathsf{q}_2) \ldots, (\boldsymbol{\beta}_J, \mathsf{q}_J) \Big\rangle$$

where $\boldsymbol{\alpha} \succ \boldsymbol{\beta}_2 \succ \ldots \succ \boldsymbol{\beta}_J$ are profiles.

Push pair $\Big([w_1, w_2, \ldots, w_L], \mathsf{c}\Big)$ onto $\mathtt{f}$.

Step 3: Define a list

$$\mathtt{x} := \Big\langle (\boldsymbol{\beta}_2, \text{-}\mathsf{c}\mathsf{q}_2), (\boldsymbol{\beta}_3, \text{-}\mathsf{c}\mathsf{q}_3), \ldots, (\boldsymbol{\beta}_J, \text{-}\mathsf{c}\mathsf{q}_J) \Big\rangle;$$

note the negated coefficients. Now assign

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

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

Efficiency considerations

An expensive part is *Step 2*, computing the canon-rep of the esp-product.

To be written.

Filename: `Problems/Polynomials/symmetric_polys.latex`
As of: *Sunday 02Apr2006.* Typeset: *23Aug2016 at 00:07.*