ROOTS AND SYMMETRIC POLYNOMIALS

DAVID SMYTH

1. From finding roots to factoring.

To see the connection between finding roots and factoring the polynomial, we begin with the following easy lemma. It says that finding a root α of $f(x)$ is the same as factoring $f(x)$ into $(x-\alpha)$ and a lower factor.

Lemma 1.1 (Remainder Theorem)**.** *Let k be a field, and let $f(x) \in k[x]$ be a polynomial. For any $\alpha \in k$, we can write*

$$f(x) = (x-\alpha)g(x) + f(\alpha),$$

where $g(x)$ is a polynomial of degree $n-1$. In particular if $f(\alpha) = 0$, then $f(x)$ admits a factorization as $(x-\alpha)g(x)$.

Proof. Using the usual division algorithm for polynomials, just divide $f(x)$ by $(x-\alpha)$. We will get

$$f(x) = (x-\alpha)g(x) + c$$

where $c \in k$ is some constant. By plugging α into both sides of this equation, we see that $c = f(\alpha)$. □

If we use this factoring procedure inductively, we get two useful corollaries.

Corollary 1.2. *If $f(x) \in k[x]$ is a degree n polynomial, then f has at most n roots.*

Proof. If f has no roots, then there is nothing to prove, so we may assume that f has a root α. By the Remainder Theorem, we may factor f as

$$f(x) = (x-\alpha)g(x).$$

By induction on the degree of f, we may assume that g has no more than $n-1$ roots. Since any root of f must be either a root of $(x-\alpha)$ (namely α) or a root of $g(x)$, it follows that $f(x)$ has no more than n roots. □

Corollary 1.3. *If $f(x) \in k[x]$ is a degree n polynomial with n distinct roots $\alpha_1, \ldots, \alpha_n \in k$, then f can be factored as:*

$$f(x) = \prod_{i=1}^{n}(x-\alpha_i)$$

Proof. We can factor $f(x)$ as:

$$f(x) = (x-\alpha_1)g(x).$$

Since the roots are distinct $(\alpha_i - \alpha_1) \neq 0$ for all $i = 2, \ldots, n$. Thus, $\alpha_2, \alpha_3, \ldots, \alpha_n$ must be roots of $g(x)$. By induction on the degree of f, we may assume $g(x) = \prod_{i=2}^{n}(x-\alpha_i)$, and the desired result follows. □

Now let us assume for the time being that $f(x)$ actually has n distinct roots, so that we can factor

$$f(x) = \prod_{i=1}^{n} (x - \alpha_i)$$

Then we can view the roots α_i as variables, and the coefficients of the polynomial as giving equations involving these variables. At least in the case $n = 2$, this idea should be familiar from high school, i.e. if we write

$$(x - \alpha_1)(x - \alpha_2) = x^2 + a_1 x + a_2,$$

where a_1 and a_2 are given to begin with, then we see the the problem of finding the roots of f is just the same as finding two numbers α_1, α_2 such that

$$\begin{aligned} -(\alpha_1 + \alpha_2) &= a_1 \\ \alpha_1 \alpha_2 &= a_2. \end{aligned}$$

In the next lecture, we will generalize this system of equations to higher n.

2. Symmetric Functions

Definition 2.1 (Elementary Symmetric Functions). *Let $k[x_1, \ldots, x_n]$ be a polynomial ring in n variables. For $i = 1, \ldots, n$, we define the following special polynomials $s_i \in k[x_1, \ldots, x_n]$:*

$$\begin{aligned} s_1 &= x_1 + x_2 + \cdots + x_n \\ s_2 &= x_1 x_2 + x_1 x_3 + \ldots + x_{n-1} x_n \\ &\vdots \\ s_k &= \sum_{1 \leq i_1 \leq i_2 \leq \ldots \leq i_k \leq n} x_{i_1} x_{i_2} \ldots x_{i_k} \\ &\vdots \\ s_n &= x_1 x_2 \cdots x_n \end{aligned}$$

In words, we can say that the k^{th} *symmetric function* is simply the sum of all degree k monomials with no repeated variables.

The point of this definition is that the functions s_i precisely encode the relationship between the roots of a polynomial and its coefficients. By some straight-forward high school algebra, you can check:

$$\prod_{i=1}^{n} (x - \alpha_i) = x^n - s_1(\alpha_1, \ldots, \alpha_n) x^{n-1} + s_2(\alpha_1, \ldots, \alpha_n) x^{n-2} - \ldots + (-1)^n s_n(\alpha_1, \ldots, \alpha_n),$$

This means that finding the roots of a given polynomial $f(x) = x^n + a_1 x^{n-1} + \ldots + a_n$ (at least under the assumption that $f(x)$ has n distinct roots - later we'll see this assumption is unnecessary) is precisely equivalent to finding $\alpha_1, \dots, \alpha_n$ which satisfy the following equations.

$$\begin{aligned} \alpha_1 + \ldots + \alpha_n &= a_1 \\ \alpha_1\alpha_2 + \ldots + \alpha_{n-1}\alpha_n &= -a_2 \\ &\vdots \\ \alpha_1\alpha_2 \cdots \alpha_n &= (-1)^n a_n \end{aligned}$$

In these equations, you should think of a_i as given, and α_i as being unknown numbers that you are trying to find.

We started off with a single equation in one variable, and now we have n equations in n variables. How on earth could this be any easier than the original problem? The point is that these are not just any random old equations; the elementary symmetric functions have very special properties that will make them easier to work with than arbitrary functions. For starters, they are symmetric. What exactly does that mean?

Definition 2.2. *Let S_n act on $k[x_1, \dots, x_n]$ by permuting the variables, i.e. $\sigma(f(x_1, \dots, x_n)) = f(x_{\sigma(1)}, \dots, x_{\sigma(n)})$. We say that a function is* symmetric *if $\sigma(f) = f$.*

Example 2.3 ($n = 3$)**.** *If $\sigma = (123)$ is the cyclic permutation of 3 variables, then $\sigma(x_1^3x_2^2x_3) = x_2^3x_3^2x_1$. Evidently, $x_1^3x_2^2x_3$ is not a symmetric function. On the other hand, $x_1^3 + x_2^3 + x_3^3$ is a symmetric function.*

The elementary symmetric functions s_i are all symmetric. While there are many symmetric functions besides the elementary ones, it turns out that they are all generated as polynomial combinations of the elementary symmetric functions. This is an astounding fact!

Theorem 2.4 (Fundamental Theorem of Symmetric Functions)**.** *Let $f(x_1, \dots, x_n)$ be any symmetric polynomial. Then, f can be expressed as polynomial in the symmetric function, i.e $f = g(s_1, \dots, s_n)$ for some polynomial g.*

Example 2.5. *The theorem says that one can express $x_1^3+x_2^3+x_3^3$ as a polynomial in s_1, s_2, s_3. One can easily check that*

$$x_1^3 + x_2^3 + x_3^3 = s_1^3 - 3s_1s_2 + 3s_3$$

Is there a way to derive this formula systematically? Yes, there is - we shall spell out the algorithm in full detail when we prove the fundamental theorem, but it may be useful to sketch the idea informally in the context of this example. To begin with, it's easy to see that s_1^3, s_1s_2, s_3 are the only monomials in s_1, s_2, s_3 that give rise to degree 3 monomials in x_1, x_2, x_3, so these are the only monomials that can appear in our formula. In other words, we must have a formula like

$$x_1^3 + x_2^3 + x_3^3 = as_1^3 + bs_1s_2 + cs_3,$$

for some coefficients a, b, c, and the question is how to figure out these coefficients.

First, focus attention on the x_1^3 term. On the left, it occurs with coefficient 1. On the right, it's easy to see that only s_1^3 contains an x_1^3 term and it occurs with coefficient one. Thus, we must have $a = 1$.

Next, let's subtract s_1^3 from both sides, to get:

$$x_1^3 + x_2^3 + x_3^3 - (x_1 + x_2 + x_3)^3 = bs_1s_2 + cs_3,$$

If we expand out the left hand side, the x_1^3 term cancels, so let's examine the next lowest order term, i.e. $x_1^2x_2$. On the left, it occurs with coefficient -3. On the right, it's easy to see that only s_1s_2 contains an $x_1^2x_2$ term and it occurs with coefficient 1. Thus, we must have $b = -3$.

Next, let's add $3s_1s_2$ to both sides to get:

$$x_1^3 + x_2^3 + x_3^3 - (x_1 + x_2 + x_3)^3 + 3(x_1 + x_2 + x_3)(x_1x_2 + x_1x_3 + x_2x_3) = cs_3,$$

If we expand out the left hand side, we see that x_1^3 and $x_1^2x_2$ terms cancel - in fact everything cancels except the $x_1x_2x_3$ term which occurs with coefficient 3. Since $s_3 = x_1x_2x_3$, we must have $c = 3$, and we are done.

In the example, I made use of the idea of the "lowest" monomial without actually explaining what I meant. The key technical tool in proving the general theorem is the introduction of an ordering on monomials which makes this concept precise.

Definition 2.6 (Lexicographic Ordering)**.** *The* lexicographic ordering *is a total ordering on all degree m monomials in n variables, which can be defined as follows. Given any monomial, we can write it as $x_{i_1}x_{i_2}\dots x_{i_m}$ with $i_1 \leq i_2 \leq \dots \leq i_m$. In other words, we can write it as a product of m variables whose subscripts go from lowest to highest. To compare two monomials, we then just look at the first subscript were two monomials differ. More formally, we say that*

$$x_{i_1}x_{i_2}\dots x_{i_m} < x_{j_1}x_{j_2}\dots x_{j_m}$$

if $i_1 = j_1$, $i_2 = j_2$, $\dots$, $i_{k-1} = j_{k-1}$ and $i_k < j_k$ for some $k \in \{1, \dots, m\}$.

Example 2.7. *The lexicographic ordering for degree 3 monomials in 3 variables goes like this:*

$$x_1^3 < x_1^2x_2 < x_1^2x_3 < x_1x_2^2 < x_1x_2x_3 < x_1x_3^2 < x_2^3 < x_2^2x_3 < x_2x_3^2 < x_3^3$$

Definition 2.8. *If f is a homogeneous polynomial of degree m (homogeneous means that every monomial in f has the same degree), we let $L(f)$ be the "lowest" monomial of f, i.e. the monomial of f which is least with respect to the lexicographic ordering.*

Example 2.9. *If $f = 2x_1^2x_2 + x_1x_2x_3 + 3x_3^3$, then $L(f) = 2x_1^2x_2$, because $x_1^2x_2 < x_1x_2x_3 < x_3^3$ in the lexicographic ordering.*

If you think about it, you will see that certain monomials cannot occur as $L(f)$ for a symmetric function f. For example, $x_1x_2^2$ could never be the lowest monomial of a symmetric function. Why not? Because if f contains the monomial $x_1x_2^2$, then it must also (by symmetry) contain the monomial $x_1^2x_2$ and $x_1^2x_2 < x_1x_2^2$. More generally, we have the following lemma.

Lemma 2.10. *If f is a symmetric function, and $L(f) = cx_1^{k_1}x_2^{k_1}\dots x_n^{k_n}$, then $k_1 \geq k_2 \geq \dots \geq k_n$.*

Proof. Let f be a symmetric function with $L(f) = x_1^{k_1}x_2^{k_1}\dots x_n^{k_n}$, and suppose the statement of the lemma fails, i.e. suppose that the k_i's are not ordered from largest to smallest. Let $\sigma \in S_n$ be a permutation that orders the k_i's correctly, i.e. such that

$$k_{\sigma(1)} \geq k_{\sigma(2)} \geq \dots \geq k_{\sigma(n)}.$$

Since f is symmetric, f must contain the monomial $x_1^{k_{\sigma(1)}}x_2^{k_{\sigma(2)}}\dots x_n^{k_{\sigma(n)}}$. By the definition of the lexicographic ordering, we have $x_1^{k_{\sigma(1)}}x_2^{k_{\sigma(2)}}\dots x_n^{k_{\sigma(n)}} < x_1^{k_1}x_2^{k_1}\dots x_n^{k_n}$. But this is a contradiction, since we started by assuming that $x_1^{k_1}x_2^{k_1}\dots x_n^{k_n}$ was the lowest monomial in f. □

Now we are ready to prove the fundamental theorem of symmetric functions. The idea, as demonstrated in Example 2.5, is to focus on the lowest monomial of our symmetric function, and then find a monomial in the elementary symmetric functions which matches it. By successively subtracting off appropriate multiples of monomials in the symmetric functions, we can work our way up the lexicographic ordering until there are monomials left! At that point, we have expressed f as a polynomial in the symmetric functions.

Proof. First, we reduce to the case that f is homogenous. We claim that if we know the fundamental theorem for homogenous symmetric functions, then we know the fundamental theorem for all symmetric functions. To see this, let f be any symmetric function and write $f = f_1 + \ldots + f_m$, where each f_i is homogenous of degree i. If f is symmetric, then each f_i must be as well (because the action of S_n preserves the degree of each monomial of f). If we know the fundamental theorem for homogenous symmetric functions, then we can write each f_i as a polynomial in elementary symmetric function. But then we clearly get a representation of f as a polynomial in elementary symmetric functions as desired.

Now let f be a homogeneous symmetric function and let $L(f) = cx^{k_1} \ldots x^{k_n}$. By Lemma 2.10, we know that $k_1 \geq k_2 \geq \ldots \geq k_n$. We claim that there exists a monomial in the symmetric functions, say $cs_1^{l_1} s_2^{l_2} \ldots s_n^{l_n}$ such that

$$L(f) = L(cs_1^{l_1} s_2^{l_2} \ldots s_n^{l_n}).$$

To check this, we need to investigate the lowest monomials of the elementary symmetric functions. By the definition of the elementary symmetric functions, one easily checks that:

$$L(s_i) = x_1 x_2 \ldots x_i.$$

It follows that

$$L(cs_1^{l_1} s_2^{l_2} \ldots s_n^{l_n}) = cx_1^{l_1+l_2+\ldots+l_n} x_2^{l_1+l_2+\ldots+l_{n-1}} \ldots x_n^{l_n}.$$

Thus, in order to get $L(f) = L(cs_1^{l_1} s_2^{l_2} \ldots s_n^{l_n})$, we simply need to find non-negative integers $l_1, l_2, \ldots, l_n$ such that

$$\begin{aligned} l_1 + \ldots + l_n &= k_1 \\ l_1 + \ldots + l_{n-1} &= k_2 \\ &\vdots \\ l_n &= k_n. \end{aligned}$$

Happily, the condition $k_1 \geq k_2 \geq \ldots \geq k_n$ guarantees that we can do this. Indeed, we simply set $l_n = k_n$ and $l_i = k_i - k_{i+1}$ for $i = 1, \ldots, n-1$. With this choice of l_i, we have $L(f) = L(cs_1^{l_1} s_2^{l_2} \ldots s_n^{l_n})$ as desired.

Now we are basically done. If we let $f' := f - cs_1^{l_1} s_2^{l_2} \ldots s_n^{l_n}$, then f' is a symmetric function with $L(f') > L(f)$. Thus, we can simply replace f by f' and repeat this procedure. As we do this, we will subtract off multiples of monomials of the symmetric functions to get a sequence of functions $f, f', f'', \ldots$ with higher and higher lowest monomials. The only way this process can terminate is to have $f^k = 0$ for some k. At that point, we have an equation expressing f as a sum of monomials of elementary symmetric functions, i.e. a polynomial in the elementary symmetric functions. □