A Matrix Proof of Newton's Identities

DAN KALMAN
American University
Washington, DC 20016-8050

Newton's identities relate sums of powers of roots of a polynomial with the coefficients of the polynomial. They are generally encountered in discussions of symmetric functions (see **[4, 9]**): a polynomial's coefficients are symmetric functions of the roots, as is the sum of the k^{th} powers of those roots.

Newton's identities also have a natural expression in the context of matrix algebra, where the trace of the k^{th} power of a matrix is the sum of the k^{th} powers of the eigenvalues. In this setting, Newton's identities can be derived as a simple consequence of the Cayley–Hamilton theorem. Presenting that derivation is the purpose of this note.

There are a variety of derivations for Newton's identities in the literature. Berlekamp's derivation **[2]** using generating function methods is short and elegant, and Mead presents a very interesting argument **[7]** using a novel notation. In yet another approach **[1]**, Baker uses differentiation to obtain a nice recursion. Eidswick's derivation **[3]** uses a related application of logarithmic differentiation. All of these proofs are elementary and understandable, but they involve manipulations or concepts that might make them a bit forbidding to students. In contrast, the proof presented here uses only methods that would be readily accessible to most linear algebra students.

Interestingly, the matrix interpretation of Newton's identities is familiar in the linear algebra literature, providing a means of computing the characteristic polynomial of a matrix in terms of the traces of the powers of the matrix (**[1, 8]**). However, using the matrix setting to *derive* Newton's identities doesn't seem to be well known.

Let $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ have roots r_j, $j = 1, \ldots, n$. Define

$$s_k \equiv \sum_{j=1}^{n} r_j^k.$$

Newton's identities are

$$s_k + a_{n-1}s_{k-1} + \cdots + a_0 s_{k-n} = 0 \quad (k > n)$$

$$s_k + a_{n-1}s_{k-1} + \cdots + a_{n-k+1}s_1 = -ka_{n-k} \quad (1 \le k \le n)$$

Now let C be an $n \times n$ matrix with characteristic polynomial equal to p. For example, C might be

$$\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix},$$

the *companion matrix* of p (**[6]**). Then the roots of p are the eigenvalues of C, and more generally, the k^{th} powers of the roots of p are the eigenvalues of C^k. Accordingly, we observe that s_k is the *trace* of C^k, written $\operatorname{tr}(C^k)$. Recall that the trace of a matrix is at once the sum of the eigenvalues and the sum of the diagonal entries. In particular, the trace operation is linear: $\mathrm{tr}(\alpha A + \beta B) = \alpha\, \mathrm{tr}(A) + \beta\, \mathrm{tr}(B)$.

Now for $k > n$, using the trace formulation, Newton's identity becomes

$$\mathrm{tr}(C^k) + a_{n-1}\mathrm{tr}(C^{k-1}) + \cdots + a_0\mathrm{tr}(C^{k-n}) = 0$$

and since the trace function is linear, we can rewrite this as

$$\mathrm{tr}\big(C^k + a_{n-1}C^{k-1} + \cdots + a_0 C^{k-n}\big) = 0, \quad \text{or} \quad \mathrm{tr}\big(C^{k-n}p(C)\big) = 0.$$

Thus, the $k > n$ case follows immediately from the Cayley–Hamilton theorem, which says that $p(C) = 0$.

For $1 \le k \le n$, the trace version of Newton's identity is

$$\mathrm{tr}(C^k) + a_{n-1}\mathrm{tr}(C^{k-1}) + \cdots + a_{n-k+1}\mathrm{tr}(C) = -ka_{n-k}$$

which can again be rewritten as

$$\mathrm{tr}\big(C^k + a_{n-1}C^{k-1} + \cdots + a_{n-k+1}C\big) = -ka_{n-k}.$$

For reasons that will be clear later, we modify this slightly, to

$$\mathrm{tr}\big(C^k + a_{n-1}C^{k-1} + \cdots + a_{n-k+1}C + a_{n-k}I\big) = (n-k)a_{n-k}. \tag{1}$$

This identity can also be derived from the Cayley–Hamilton theorem, in a slightly different way. As is well known, a real number r is a root of a real polynomial $p(x)$ if and only if $(x - r)$ is a factor of $p(x)$, and the complimentary factor can be determined using synthetic division. This situation can be mimicked exactly using matrices: let $X = xI$, and divide $p(X)$ by $X - C$ using synthetic division. Since $p(C) = 0$, the division terminates without remainder, providing the factorization

$$p(X) = (X - C)\,\big[X^{n-1} + (C + a_{n-1}I)X^{n-2} + \big(C^2 + a_{n-1}C + a_{n-2}I\big)X^{n-3} + \cdots + \big(C^{n-1} + a_{n-1}C^{n-2} + \cdots + a_1 I\big)I\big]$$

(see **[5]**).

To relate this to equation (1), we will want to introduce the trace operation. Unfortunately, the trace does not relate well to matrix products, so it is necessary to eliminate the factor of $(X - C)$ on the right. Fortunately, as long as x is not an eigenvalue of C, we know that $(xI - C) = (X - C)$ is non-singular, so we can write

$$(X - C)^{-1}p(X) = X^{n-1} + (C + a_{n-1}I)X^{n-2} + \big(C^2 + a_{n-1}C + a_{n-2}I\big)X^{n-3} + \cdots + \big(C^{n-1} + a_{n-1}C^{n-2} + \cdots + a_1 I\big)I.$$

Taking the trace of each side then leads to

$$\mathrm{tr}\big[(X - C)^{-1}p(X)\big] = nx^{n-1} + \mathrm{tr}(C + a_{n-1}I)x^{n-2} + \cdots + \mathrm{tr}\big(C^{n-1} + a_{n-1}C^{n-2} + \cdots + a_1 I\big) \tag{2}$$

because $\mathrm{tr}(I) = n$ and $\mathrm{tr}(X^k A) = \mathrm{tr}(x^k I A) = x^k \mathrm{tr}(A)$ for any matrix A.

We will next show that the left side of this equation is none other than $p'(x)$. Then, comparing coefficients on either side will complete the proof. Indeed, equating the coefficient of x^{n-k-1} in $p'(x)$ with the corresponding coefficient on the right side of equation (2) gives

$$(n-k)a_{n-k} = \mathrm{tr}\big(C^k + a_{n-1}C^{k-1} + \cdots + a_{n-k+1}C + a_{n-k}I\big)$$

which is exactly the same as equation (1).

So, consider $A = (X - C)^{-1}p(X)$. Observe that $p(X) = p(xI) = p(x)I$, so we can equally well write $A = p(x)(xI - C)^{-1}$. This shows that

$$\mathrm{tr}(A) = p(x)\mathrm{tr}(xI - C)^{-1}.$$

Now the trace of any matrix is the sum of its eigenvalues (with multiplicities as in the characteristic polynomial). And the eigenvalues of $(xI - C)^{-1}$ are simply the fractions $1/(x - r_1), 1/(x - r_2), \cdots, 1/(x - r_n)$. This shows

$$\mathrm{tr}(A) = p(x)\left(\frac{1}{x - r_1} + \frac{1}{x - r_2} + \cdots + \frac{1}{x - r_n}\right)$$

which is immediately recognizable as the derivative $p'(x)$ (using the fact that $p(x) = (x - r_1)(x - r_2) \cdots (x - r_n)$). This completes the proof.

REFERENCES

1. George A. Baker, Jr., A new derivation of Newton's identities and their application to the calculation of the eigenvalues of a matrix, *Journal of the Society for Industrial and Applied Mathematics* 7 (1959), 143–148.
2. Elwyn R. Berlekamp, *Algebraic Coding Theory*, Revised 1984 Edition, Aegean Park Press, Laguna Hills, CA, 1984, page 212.
3. J. A. Eidswick, A proof of Newton's power sum formulas, *American Mathematical Monthly* 75 (1968), 396–397.
4. I. N. Herstein, *Topics in Algebra*, Ginn, Waltham, MA, 1964, page 208.
5. Dan Kalman, An extension of the remainder theorem, *Delta* 8 (1978), 77–80.
6. David C. Lay, *Linear Algebra and Its Applications*, 2nd ed., Addison-Wesley, Reading, MA, 1997, page 365.
7. D. G. Mead, Newton's identities, *American Mathematical Monthly* 99 (1992), 749–751.
8. D. W. Robinson, A matrix application of Newton's identities, *American Mathematical Monthly* 68 (1961), 367–369.
9. B. L. van der Waerden, *Algebra*, volume 1, Frederick Ungar, New York, NY, 1970, page 101.

Boxes for Isoperimetric Triangles

JOHN E. WETZEL
University of Illinois at Urbana-Champaign
Urbana, IL 61801

Introduction A rectangular region *covers* a family of curves if it contains a congruent copy of each curve in the family. We call such a region a *box* for the family. In this note we answer two questions concerning boxes for the family $\mathcal{T}$ of all triangles of perimeter two:

(1) Among all boxes for $\mathcal{T}$, which has least area?
(2) Among all boxes for $\mathcal{T}$ of prescribed shape, which has least area?

Some results are known about *triangular* covers for $\mathcal{T}$. In **[8]** we found the side of the smallest equilateral triangle that can cover $\mathcal{T}$, but the smallest triangular covers for $\mathcal{T}$ of other shapes remain unknown. With Füredi in **[5]**, we found the smallest triangle (without regard to shape) that can cover $\mathcal{T}$, and we showed somewhat surprisingly that