A MAGIC DETERMINANT FORMULA FOR SYMMETRIC POLYNOMIALS OF EIGENVALUES

JULES JACOBS (JULESJACOBS@GMAIL.COM)

Abstract. Symmetric polynomials of the roots of a polynomial can be written as polynomials of the coefficients, and by applying this theorem to the characteristic polynomial we can write a symmetric polynomial of the eigenvalues a_i of an $n \times n$ matrix A as a polynomial of the entries of the matrix. We give a magic formula for this: symbolically substitute $a \mapsto A$ in the symmetric polynomial and replace multiplication by det. For instance, for a 2×2 matrix A with eigenvalues a_1, a_2,

$$a_1 a_2^2 + a_1^2 a_2 = \det(A_1, A_2^2) + \det(A_1^2, A_2)$$

where A_i^k is the i-th column of A^k and $\det(v_1, \ldots, v_n)$ is the determinant of the matrix with columns $v_1, \ldots, v_n$. One may also take negative powers, allowing us to calculate:

$$a_1 a_2^{-1} + a_1^{-1} a_2 = \det(A_1, A_2^{-1}) + \det(A_1^{-1}, A_2)$$

The magic method also works for multivariate symmetric polynomials of the eigenvalues of a set of commuting matrices, e.g. for 2×2 matrices A and B with eigenvalues a_1, a_2 and b_1, b_2,

$$a_1 b_1 a_2^2 + a_1^2 a_2 b_2 = \det(AB_1, A_2^2) + \det(A_1^2, AB_2)$$

1. Introduction

Let A be an $n \times n$ matrix with eigenvalues $a_1, \ldots, a_n$. It is well known that

$$a_1 + a_2 + \cdots + a_n = \mathrm{tr}(A) = A_{11} + A_{22} + \cdots + A_{nn}$$

$$a_1 a_2 \cdots a_n = \det(A) = \sum_{\sigma \in S_n} (-1)^\sigma A_{1,\sigma(1)} A_{2,\sigma(2)} \cdots A_{n,\sigma(n)}$$

By applying the fundamental theorem of symmetric polynomials to the characteristic polynomial of A, we find that there must be such an equation between any symmetric polynomial in the eigenvalues of A and *some* polynomial in the entries of A. We give an explicit formula for this polynomial in terms of determinants, which generalises the equations above to any symmetric polynomial:

$$\sum_{i \in \mathbb{N}^n} p_i a_1^{i_1} \cdots a_n^{i_n} = \sum_{i \in \mathbb{N}^n} p_i \det(A_1^{i_1}, \ldots, A_n^{i_n})$$

This equation holds when the left hand side is a symmetric polynomial. That is, the coefficients $p_i = p_{(i_1,\ldots,i_n)}$ must satisfy $p_{(i_{\sigma(1)},\ldots,i_{\sigma(n)})} = p_{(i_1,\ldots,i_n)}$ for all permutations σ. At first, this equation may be surprising, since eigenvalues on the left hand side are independent of the choice of basis, whereas taking k-th columns of a matrix on the right hand side is clearly basis dependent. Indeed, each term $p_i \det(A_1^{i_1}, \ldots, A_n^{i_n})$ is on its own not basis independent, only the whole sum is, and only if the p_i are coefficients of a symmetric polynomial. More generally,

$$\sum_{i \in \mathbb{N}^n} p_i \det(A_1^{(i_1)}, \ldots, A_n^{(i_n)}) \tag{1}$$

is basis independent for any family of matrices $A^{(j)}$, not necessarily powers A^j of a single matrix. That is, if we substitue $A^{(j)} \mapsto S^{-1} A^{(j)} S$ for some invertible matrix S, its value does not change. Furthermore, if the $A^{(j)}$ commute, we have the identity

$$\sum_{i \in \mathbb{N}^n} p_i a_1^{(i_1)} \cdots a_n^{(i_n)} = \sum_{i \in \mathbb{N}^n} p_i \det(A_1^{(i_1)}, \ldots, A_n^{(i_n)})$$

where $a_k^{(j)}$ is the k-th eigenvalue of $A^{(j)}$.

Our strategy to prove this is to define a quantity that is basis independent because its definition makes no reference to a basis, and then show that it is equal to (1). Once we have shown that (1) is invariant under basis transformations, we pick a basis in which the determinants become products of eigenvalues, which is possible if the $A^{(j)}$ commute.

By applying this identity to particular families of matrices, we get the fundamental theorem of symmetric polynomials as a corollary, and we are able to deduce various equations between eigenvalues and determinants, such as those in the abstract.

2. Preliminaries

The proof is based on multilinear antisymmetric functions [1].

Definition. A function $f : V^n \to \mathbb{R}$ is

- Multilinear if f is linear in each argument, i.e. $v_k \mapsto f(v_1, \dots, v_k, \dots, v_n)$ is linear, with $v_i \in V$.
- Antisymmetric if applying a permutation $\sigma \in S_n$ to its arguments multiplies it by the sign of the permutation, i.e. $f(v_{\sigma(1)}, \dots, v_{\sigma(n)}) = (-1)^\sigma f(v_1, \dots, v_n)$, with $v_i \in V$.

Multilinear antisymmetric functions form a vector space.

Definition. Let $\bigwedge^n V^* \subset V^n \to \mathbb{R}$ be the space of multilinear antisymmetric functions.

The only property of $\bigwedge^n V^*$ we will need is that $\dim(\bigwedge^n V^*) = 1$ if $n = \dim(V)$. In fact, in general $\dim(\bigwedge^n V^*) = \binom{\dim(V)}{n}$ [1]. Since it is the basis of our main theorem, we will give a proof here.

Lemma. $\dim(\bigwedge^n V^*) = 1$ *if* $\dim(V) = n$.

Proof. Let $f \in \bigwedge^n V^*$ and $v_1, \dots, v_n \in V$. Take a basis $e_1, \dots, e_n$ for V. Writing the v_i in terms of the basis, we have coefficients $a_{ij} \in \mathbb{R}$ for $i, j \in \{1 \dots n\}$ such that $v_i = \sum_j a_{ij} e_j$. The proof proceeds by expanding $f(v_1, \dots, v_n)$ by multilinearity, and then using the fact that f is zero whenever there is a duplicate argument, because applying a permutation that swaps those two positions gives $f(\dots, w, \dots, w, \dots) = -f(\dots, w, \dots, w, \dots)$, which implies $f(\dots, w, \dots, w, \dots) = 0$. This allows us to convert the sum over any pattern of indexing into a sum over permutations.

$$\begin{aligned}
f(v_1, \dots, v_n) &= f(\sum_{i_1=1}^{n} a_{1i_1} e_{i_1}, \dots, \sum_{i_n=1}^{n} a_{1i_n} e_{i_n}) \\
&= \sum_{i_1=1}^{n} \cdots \sum_{i_n=1}^{n} a_{1i_1} \dots a_{ni_n} f(e_{i_1}, \dots, e_{i_n}) \\
&= \sum_{i \in \{1\dots n\} \to \{1 \dots n\}} a_{1i(1)} \dots a_{ni(n)} f(e_{i(1)}, \dots, e_{i(n)}) \\
&= \sum_{\sigma \in S_n} a_{1\sigma(1)} \dots a_{n\sigma(n)} f(e_{\sigma(1)}, \dots, e_{\sigma(n)}) \\
&= \sum_{\sigma \in S_n} (-1)^\sigma a_{1\sigma(1)} \dots a_{n\sigma(n)} f(e_1, \dots, e_e) \\
&= \det(a) f(e_1, \dots, e_n)
\end{aligned}$$

Therefore, the value $f \in \bigwedge^n V^*$ is entirely determined by its value on $f(e_1, \dots, e_n)$, and conversely, any value chosen for $f(e_1, \dots, e_n)$ determines a multilinear antisymmetric function $f \in \bigwedge^n V^*$ via the equation above, so the space is one dimensional. □

3. Proof

Let $\mathbb{I}$ be some index set. We say that $p_i \in \mathbb{R}$ for $i \in \mathbb{I}^n$ are *symmetric coefficients* if only finitely many p_i are nonzero, and $p_{i \circ \sigma} = p_i$ for all $i \in \mathbb{I}^n$ and permutations $\sigma \in S_n$. Let $\vec{A}^{(j)} : V \to V$ for $j \in \mathbb{I}$

be a family of linear maps on a vector space V of dimension n.

Example. A polynomial $p(x_1, \ldots, x_n)$ on n variables can be written as a sum of monomials

$$p(x_1, \ldots, x_n) = \sum_{i \in \mathbb{N}^n} p_i x_1^{i_i} \cdots x_n^{i_n}$$

where $p_i = p_{(i_1, i_2, \ldots, i_n)}$ is the coefficient of $x_1^{i_1} x_2^{i_2} \ldots x_n^{i_n}$. If p is a symmetric polynomial, then p_i are symmetric coefficients with index set $\mathbb{I} = \mathbb{N}$. For instance, if

$$p(x_1, x_2, x_3) = x_1 + x_2 + x_3 = x_1^1 x_2^0 x_3^0 + x_1^0 x_2^1 x_3^0 + x_1^0 x_2^0 x_3^1$$

then $p_i = 1$ for $i = (1,0,0), (0,1,0), (0,0,1)$ and $p_i = 0$ otherwise. Note that $p_{i \circ \sigma} = p_i$ for all permutations $\sigma \in S_3$.

We now define a linear map $p(\vec{A}) : \bigwedge^n V^* \to \bigwedge^n V^*$ in terms of the matrices $\vec{A}^{(j)}$ and symmetric coefficients p_i. The key observation is that a linear map from a one-dimensional vector space to itself is just multiplication by a scalar, so we are in effect defining a scalar here. We shall soon see that this scalar is precisely (1).

Definition 1. If p_i are symmetric coefficients, we define $p(\vec{A}) : \bigwedge^n V^* \to \bigwedge^n V^*$ by

$$p(\vec{A}) f(v_1, \ldots, v_n) = \sum_{i \in \mathbb{I}^n} p_i f(A^{(i_1)} v_1, \ldots, A^{(i_n)} v_n)$$

This does indeed preserve antisymmetry, by change of summation variable $i = j \circ \sigma$, and using $p_{j \circ \sigma} = p_j$:

$$\begin{aligned} p(\vec{A}) f(v_{\sigma(1)}, \ldots, v_{\sigma(n)}) &= \sum_{i \in \mathbb{I}^n} p_i f(A^{(i_1)} v_{\sigma(1)}, \ldots, A^{(i_n)} v_{\sigma(n)}) \\ &= \sum_{j \in \mathbb{I}^n} p_{j \circ \sigma} f(A^{(j_{\sigma(1)})} v_{\sigma(1)}, \ldots, A^{(j_{\sigma(n)})} v_{\sigma(n)}) \\ &= \sum_{j \in \mathbb{I}^n} p_j (-1)^\sigma f(A^{(j_1)} v_1, \ldots, A^{(j_n)} v_n) \\ &= (-1)^\sigma p(\vec{A}) f(v_1, \ldots, v_n) \end{aligned}$$

Example. Given a single matrix $A \in \mathbb{R}^{3 \times 3}$ and index set $\mathbb{I} = \mathbb{N}$, we can pick $\vec{A}^{(j)} = A^j$ to be the powers of that matrix. For the symmetric coefficients p_i from the running example, we have

$$\begin{aligned} p(\vec{A}) f(v_1, v_2, v_3) &= f(A^1 v_1, A^0 v_2, A^0 v_3) + f(A^0 v_1, A^1 v_2, A^0 v_3) + f(A^0 v_1, A^0 v_2, A^1 v_3) \\ &= f(A v_1, I v_2, I v_3) + f(I v_1, A v_2, I v_3) + f(I v_1, I v_2, A v_3) \\ &= f(A v_1, v_2, v_3) + f(v_1, A v_2, v_3) + f(v_1, v_2, A v_3) \end{aligned}$$

We use the notation $[p(\vec{A})] \in \mathbb{R}$ for the scalar corresponding to $p(\vec{A}) \in End(\bigwedge^n V^*)$, so that $p(\vec{A}) f = [p(\vec{A})] f$ for all $f \in \bigwedge^n V^*$. Since $p(\vec{A})$ has been defined in terms of the $A^{(j)}$, the scalar $[p(\vec{A})] \in \mathbb{R}$ is a function of the maps $A^{(j)} : V \to V$. Since we have not used the choice of a basis for V to define $p(\vec{A})$, the scalar $[p(\vec{A})]$ is manifestly invariant under change of basis.

Pick a basis $B : \mathbb{R}^n \to V$ for V (with basis vectors $b_i = B e_i$, where $e_i \in \mathbb{R}^n$ is the standard basis). We have matrix representations $M^{(j)} = B^{-1} A^{(j)} B \in \mathbb{R}^{n \times n}$ for the $A^{(j)} : V \to V$, and we wish to calculate $[p(\vec{A})] \in \mathbb{R}$ explicitly terms of the entries of $M^{(j)} \in \mathbb{R}^{n \times n}$.

Theorem 2. $[p(\vec{A})] = \sum_{i \in \mathbb{I}^n} p_i \det(M_1^{(i_1)}, \ldots, M_n^{(i_n)})$

Example. For the running example, and $B = I$,

$$[p(\vec{A})] = \det(A_1, I_2, I_3) + \det(I_1, A_2, I_3) + \det(I_1, I_2, A_3) = A_{11} + A_{22} + A_{33}$$

Proof. Since $p(\vec{A})$ is multiplication by a scalar $[p(\vec{A})]$,

$$p(\vec{A}) f(v_1, \ldots, v_n) = [p(\vec{A})] \cdot f(v_1, \ldots, v_n)$$

for all $f \in \bigwedge^n V^*$ and vectors $(v_1, \ldots, v_n)$. Taking $f(w_1, \ldots, w_n) = \det(B^{-1}w_1, \ldots, B^{-1}w_n)$ and $(v_1, \ldots, v_n) = (Be_1, \ldots, Be_n)$ to be the basis vectors, on the right hand side

$$f(v_1, \ldots, v_n) = \det(B^{-1}Be_1, \ldots, B^{-1}Be_n) = \det(e_1, \ldots, e_n) = 1$$

and on the left hand side

$$p(\vec{A}) f(v_1, \ldots, v_n) = \sum_{i \in \mathbb{I}^n} p_i \det(B^{-1}A^{(i_1)}Be_1, \ldots, B^{-1}A^{(i_n)}Be_n) = \sum_{i \in \mathbb{I}^n} p_i \det(M_1^{(i_1)}, \ldots, M_n^{(i_n)})$$

giving $[p(\vec{A})] = \sum_{i \in \mathbb{I}^n} p_i \det(M_1^{(i_1)}, \ldots, M_n^{(i_n)})$. □

This shows that the value of $[p(\vec{A})] = \sum_{i \in \mathbb{I}^n} p_i \det(M_1^{(i_1)}, \ldots, M_n^{(i_n)})$ does not depend on the basis, since the left hand side is defined without reference to the basis. By picking a basis in which the $A^{(j)}$ are all upper triangular, we can relate $[p(\vec{A})]$ to the eigenvalues of the $A^{(j)}$.

Theorem 3. *If the $A^{(j)}$ commute, then $[p(\vec{A})] = \sum_{i \in \mathbb{I}^n} p_i a_1^{(i_1)} \cdots a_n^{(i_n)}$, where $a_k^{(j)}$ are the eigenvalues of $A^{(j)}$.*

Example. For the running example, suppose we have a basis in which A is upper triangular (e.g. the Jordan basis), then

$$[p(\vec{A})] = \det(A_1, I_2, I_3) + \det(I_1, A_2, I_3) + \det(I_1, I_2, A_3) = a_1 + a_2 + a_3$$

where a_1, a_2, a_3 are the eigenvalues of A.

Proof. Commuting matrices have a basis in which they are simultaneously upper triangular, by Schur decomposition [2]. The diagonal of those upper triangular matrices $M^{(j)}$ will contain the eigenvalues $a_k^{(j)}$. In this case, $\det(M_1^{(i_1)}, \ldots, M_n^{(i_n)})$ is the determinant of an upper triangular matrix, with eigenvalues $a_k^{(i_k)}$ on the diagonal. Hence $\det(M_1^{(i_1)}, \ldots, M_n^{(i_n)}) = a_1^{(i_1)} \cdots a_n^{(i_n)}$, and substituting this into theorem (2) gives the eigenvalue formula for $p(\vec{A})$. □

The condition that the $A^{(j)}$ commute is not a necessary condition, because there are upper triangular matrices that do not commute.

4. Corollaries

By combining theorems (2) and (3) we can justify the magic formula for converting symmetric expressions involving eigenvalues into symmetric expressions involving determinants. We first show this by example: let A, B be commuting 3×3 matrices with eigenvalues a_1, a_2, a_3 and b_1, b_2, b_3. Formally, we take the index set $\mathbb{I} = \{A, B\}$ in the theorems. Next, we choose symmetric coefficients p_i for $i \in \mathbb{I}^3$; we must choose values for $p_{AAA}, p_{AAB}, p_{ABA}, \ldots, p_{BBB}$ that are symmetric under permutations of the indices. We choose $p_{AAB} = p_{ABA} = p_{BAA} = 1$ and the rest 0. The theorems (2) and (3) give us the equation

$$a_1a_2b_3 + a_1b_2a_3 + b_1a_2a_3 = \det(A_1, A_2, B_3) + \det(A_1, B_2, A_3) + \det(B_1, A_2, A_3)$$

We can now proceed to pick the matrices A, B in this equation, as long as they commute. For instance, given a matrix C, we could pick $A = C^3$ and $B = C^{-1}$. Or, given commuting matrices C, D, we could pick $A = C^2 + D$ and $B = CD$. Or we could pick $A = (C + D)^{-1}$ and $B = \exp(CD)$. We thus obtain various relations between eigenvalues and determinants by substituting, e.g. $A = (C + D)^{-1}$, $B = \exp(CD)$, and $a_i = (c_i + d_i)^{-1}$, $b_i = \exp(c_id_i)$ into the equation.

Rather than trying to capture this general method in a theorem, we present a few special cases.

Corollary 4. *Let $q(b_1, \dots, b_n) = \sum_{i\in\mathbb{N}^n} p_i b_1^{i_1} \dots b_n^{i_n}$ be a symmetric polynomial in the eigenvalues of an $n \times n$ matrix B, then $q(b_1, \dots, b_n) = \sum_{i\in\mathbb{N}^n} p_i \det(B_1^{i_1}, \dots, B_n^{i_n})$.*

Proof. Apply theorems (2) and (3) with index set $\mathbb{I} = \mathbb{N}$ and matrices $A^{(j)} = B^j$. □

Corollary 5. *Let $q(x) = \sum_{k=0}^{n} a_k x^k$ be a polynomial with roots r_k. Then a symmetric polynomial in the roots r_k can be written as a polynomial in the coefficients a_k.*

Proof. Apply the previous corollary with B being the companion matrix of q. The eigenvalues of B are the r_k. The entries of the companion matrix are all 0, or 1, or a_k, so $\det(B_1^{i_1}, \dots, B_n^{i_n})$ is a polynomial in the a_k. □

A symmetric polynomial q can be seen as a function of a vector $x \in \mathbb{R}^n$ satisfying $q(Px) = q(x)$ for permutation matrices P.

Definition 6. A multivariate symmetric polynomial $q(X)$ is a polynomial function of the entries of a matrix $X \in \mathbb{R}^{n\times m}$ satisfying $q(PX) = q(X)$ for permutation matrices P.

The permutation P permutes the rows of X, but keeps each row together. A symmetric polynomial is the special case $m = 1$, when each row consists of a single entry.

Corollary 7. *Let $B^{(j)}$ for $j \in \{1, \dots, m\}$ be commuting matrices with eigenvalues $b_i^{(j)}$, and define $X_{ij} = b_i^{(j)}$ to be the matrix of eigenvalues. Then a multivariate symmetric polynomial $q(X)$ can be written as a polynomial in the entries of the $B^{(j)}$.*

Proof. Take $\mathbb{I} = \mathbb{N}^m$ and $A^{(g)} = (B^{(1)})^{g_1} \cdots (B^{(m)})^{g_m}$ in theorems (2) and (3). □

5. Footnote

The same definition (1) works for $p(\vec{A}) : \bigwedge^k W^* \to \bigwedge^k V^*$ also when $W \neq V$ and $k \neq n$. We can view $p(\vec{A})f$ as a generalised pullback of f along a list of maps $A^{(j)} : V \to W$. The ordinary pullback $A^* : \bigwedge^k W^* \to \bigwedge^k V^*$ is the special case of a single map $A : V \to W$. We have, in general $p(X\vec{A}Y) = Y^* p(\vec{A}) X^*$, where $X\vec{A}Y$ is simultaneous conjugation $(X\vec{A}Y)^{(j)} = XA^{(j)}Y$. This gives us a slightly stronger version of theorem (2): if $k = \dim(V) = \dim(W)$, then $X^* = \det(X)$ and $Y^* = \det(Y)$, so we see that if we multiply the $M^{(j)}$ on the left by X and on the right by Y, the value of $[p(\vec{A})]$ gets multiplied by $\det(X)\det(Y)$. Theorem (2) tells us that the value does not change if we do a basis transformation, but here we get information about the case $X \neq Y^{-1}$. It is also possible to generalise the proof of theorem (2) directly.

References

[1] Nicolas Bourbaki. Elements of mathematics, Algebra I. Springer-Verlag, 1989.
[2] R.A. Horn and C.R. Johnson. Matrix Analysis. Cambridge University Press, 1985.