Functions of A Matrix

Dar-Veig Ho

December 8, 2003

Abstract. Definition of functions of a matrix given in Gantmacher's text is attained from a different aspect.

1. Functions of a matrix. After I have derived formulas in [1, 2] for power, exponential, and logarithm functions of a matrix A, it is only natural that one wishes to do so for more matric functions. From those results, a pattern will emerge for a formula or a definition of a wider class of functions of A.

Suppose that the characteristic polynomial of a given matrix A is

$$p(\lambda) = \prod_{i=1}^{s} (\lambda - \lambda_i)^{m_i}$$

where m_i is the multiplicity of the eigenvalue λ_i, and $\lambda_i \neq \lambda_j$ if $i \neq j$. Recall that in [1], the Herod's projection matrices $P_i(A)$ can be obtained readily from the characteristic polynomial of A. For readers' convenience, we mention some of the properties of these matrices:

$$\begin{aligned} \sum_{i=1}^{s} P_i(A) &= I, \\ P_i(A)P_j(A) &= P_i(A)\delta_{ij}, \\ P_i(A)(A - \lambda_i I)^{m_i} &= [P_i(A)(A - \lambda_i I)]^{m_i} = 0 \ , \ m_i \geq 1. \end{aligned}$$

Let $N_i(A)$ denote the nilpotent $P_i(A)(A - \lambda_i I)$. The nonzero matrices of the set $\{P_i(A), N_i(A), \cdots, N_i^{m_i-1}(A)$, for $i = 1, 2, 3, \cdots, s,\}$ will be used to serve as a basis for all valid $f(A)$ of a given matrix A. We first identify those nonzero elements and then show their linear independence. Note that $P_i(A)$ can not be a zero matrix. Suppose that

$$P_i(A)(A - \lambda_i I) \neq 0$$

then

$$P_i(A) \neq 0.$$

If , on the other hand,

$$P_i(A)(A - \lambda_i I) = 0,$$

then

$$AP_i(A) = \lambda_i P_i(A)$$

and a nonzero column of $P_i(A)$ is an eigenvector corresponding to the eigenvalue λ_i so $P_i(A)$ can not be a zero matrix. However the nilpotent matrix $N_i^j(A) = P_i(A)(A - \lambda_i I)^j$ may become a zero matrix for $j < m_i$. Let n_i $(1 \leq n_i \leq m_i)$ be such that $N_i^{n_i}(A) = 0$ but $N_i^{n_i-1}(A) \neq 0$. (n_i is the so-called index of nilpotency.) Thus none of the set

$$\{P_i(A), N_i(A), \cdots, N_i^{n_i-1}(A),\ 1 \leq i \leq s\} \tag{1}$$

is a zero matrix and now we can show the set to be linearly independent. Suppose that

$$\sum_{i=1}^{s}[c_{i0}P_i(A) + \sum_{j=1}^{n_i-1} c_{ij}N_i^j(A)] = 0$$

which means

$$\sum_{i=1}^{s}\sum_{j=0}^{n_i-1} c_{ij}P_i(A)(A - \lambda_i I)^j = 0. \tag{2}$$

Multiply the above equation by $P_k(A)$ and use the property

$$P_k(A)P_i(A) = P_k(A)\delta_{ki}$$

to find

$$\sum_{j=0}^{n_k-1} c_{kj}P_k(A)(A - \lambda_k I)^j = 0. \tag{3}$$

Multiplying the above equation by $(A - \lambda_k I)^{n_k-1}$, we get

$$c_{k0}P_k(A)(A - \lambda_k I)^{n_k-1} = 0$$

because $P_k(A)(A - \lambda_k I)^{n_k} = 0$. But $P_k(A)(A - \lambda_k I)^{n_k-1} \neq 0$ so c_{k0} must be zero. Repeatedly multiplying (3) by $(A - \lambda_k I)^{n_k-\alpha}$ with $\alpha = 2, 3, 4, \cdots, n_k - 1$ and using the same argument, we find all $c_{kj} = 0$. The set (1) is thus linearly independent and can be used as a basis for $f(A)$.

2. A definition of $f(A)$. In the following we summarize some results from [1, 2], from which a pattern for a formula of $f(A)$ will emerge.

(i) Polynomials of A

The formula for positive integral powers of A

$$A^n = \sum_{i=1}^{s} \sum_{j=0}^{n_i-1} \binom{n}{j} P_i(A) \lambda_i^{n-j} (A - \lambda_i I)^j$$

was obtained in [1]. Note here we made use of $P_i(A)(A - \lambda_i I)^{n_i} = 0$. We can thus formulate a polynomial of A. Let $Q(\lambda)$ be a polynomial of λ,

$$\begin{aligned}
Q(A) &= \sum_{n=0}^{N} c_n A^n \\
&= \sum_{n=0}^{N} c_n \sum_{i=1}^{s} \sum_{j=0}^{n_i-1} \binom{n}{j} P_i(A) \lambda_i^{n-j} (A - \lambda_i I)^j \qquad \text{by (9) of [1]} \\
&= \sum_{i=1}^{s} \sum_{j=0}^{n_i-1} \frac{1}{j!} [\sum_{n=0}^{N} c_n n(n-1)(n-2)\cdots(n-j+1)\lambda_i^{n-j}] P_i(A)(A - \lambda_i I)^j \\
&= \sum_{i=1}^{s} \sum_{j=0}^{n_i-1} \frac{1}{j!} Q^{(j)}(\lambda_i) P_i(A)(A - \lambda_i I)^j.
\end{aligned}$$

(ii) Taylor series

Suppose that

$$u(\lambda) = \sum_{j=0}^{\infty} \frac{1}{j!} u^{(j)}(\lambda_i)(\lambda - \lambda_i)^j.$$

We assert

$$\begin{aligned}
u(A) &= \sum_{j=0}^{\infty} \frac{1}{j!} u^{(j)}(\lambda_i)(A - \lambda_i I)^j \\
P_i(A)u(A) &= \sum_{j=0}^{n_i-1} \frac{1}{j!} u^{(j)}(\lambda_i) P_i(A)(A - \lambda_i I)^j \qquad \text{since } P_i(A)(A - \lambda_i I)^{n_i} = 0 \\
u(A) &= \sum_{i=1}^{s} P_i(A)u(A) \qquad \text{by } \sum_{i=1}^{s} P_i(A) = I, \\
&= \sum_{i=1}^{s} \sum_{j=0}^{n_i-1} \frac{1}{j!} u^{(j)}(\lambda_i) P_i(A)(A - \lambda_i I)^j.
\end{aligned}$$

In fact, there is no need to require $u(\lambda)$ to be analytic for each λ_i. We need only to assume that the first m_i derivatives exist and plus a Lagrange type remainder, i.e.,

$$u(\lambda) = \sum_{j=0}^{m_i-1} \frac{1}{j!} u^{(j)}(\lambda_i)(\lambda - \lambda_i)^j + R(\lambda; \lambda_i)(\lambda - \lambda_i)^{m_i}.$$

That remainder will be annihilated upon multiplying it by $P_i(A)$.

(iii) The inverse of A

A^{-1} does not belong to the above categories, yet it is perhaps one of the most commonly used functions of A. In [2], we found

$$\begin{aligned} A^{-1} &= \sum_{i=1}^{k} P_i(A) \sum_{j=0}^{m_i-1} \binom{-1}{j} \lambda_i^{-j-1}(A-\lambda_i I)^j \\ &= \sum_{i=1}^{k} P_i(A) \frac{1}{j!} \sum_{j=0}^{m_i-1} (-1)(-2)\dots(-j)\lambda_i^{-j-1}(A-\lambda_i I)^j \\ &= \sum_{i=1}^{k} \sum_{j=0}^{n_i-1} \frac{1}{j!} [\frac{d}{d\lambda}(\lambda^{-1})]_{\lambda=\lambda_i} P_i(A)(A-\lambda_i I)^j \end{aligned}$$

These results are sufficient to motivate the following definition:

$$\begin{aligned} f(A) &= \sum_{i=1}^{s} \sum_{j=0}^{n_i-1} \frac{1}{j!} f^{(j)}(\lambda_i) P_i(A)(A-\lambda_i I)^j \\ &= \sum_{i=1}^{s} [f(\lambda_i)P_1(A) + \sum_{j=1}^{n_i-1} \frac{1}{j!} f^{(j)}(\lambda_i) N_i^j(A) \end{aligned} \tag{4}$$

for every $f(A)$ defined on the spectrum of A, by which we mean that the set

$$f^{(j)}(\lambda_i) \text{ for } i = 1, 2, 3, \cdots, s \text{ and } j = 0, 1, 2, \cdots, n_i - 1 \tag{5}$$

exist. Each $f(A)$ is a linear combination of the set of the basis (1) so that the collection of all functions of a given A forms a finite dimensional space. I wish that I could claim that I have discovered this definition of $f(A)$. Actually I find it in the treatise on matrices by Gantmakher [3]. It also appeared in other texts such as those by Lancaster [4] and Lancaster and Tismenetsky [5]. Gantmakher defined $f(A)$ through interpolation polynomials $r(\lambda)$, which coincides with a given function $f(\lambda)$ at the spectrum points of A. $r(A)$ thus becomes $f(A)$. Somehow I feel more comfortable to arrive at it by generalization from those special cases.

Notice that Gantmakher built the definition of $f(A)$ on the minimum polynomial of A whereas we constructed the basis (1) by using $n_{i,}$ in place of m_i, so that the set $\{N_i^j(A), i = 1, \cdots, s,\ j = 1, \cdots, n_i - 1\}$ consists of nonzero matrices. The following theorem will show they are equivalent.

Theorem

The polynomial

$$\prod_{i=1}^{s}(\lambda-\lambda_i)^{k_i},$$

where $n_i \leq k_i \leq m_i$, is the minimum polynomial of A if, and only if, the nilpotent $N_i^j(A)$ in the expression

$$f(A)=\sum_{i=1}^{s}[f(\lambda_i)P_i(A)+\sum_{j=1}^{k_i-1}\frac{1}{j!}f^{(j)}(\lambda_i)N_i^j(A)]\ , \tag{6}$$

are nonzero..

Proof

If $N_i^j(A)\neq 0$, let

$$f(\lambda)=\prod_{i=1}^{s}(\lambda-\lambda_i)^{k_i}$$

so that

$$f(\lambda_i)=f'(\lambda_i)=\cdots=f^{(k_i-1)}(\lambda_i)=0.$$

Substituting into (6), we get

$$f(A)=\prod_{i=1}^{s}(A-\lambda_i I)^{k_i}=0$$

However if $f(\lambda)$ is of lower degree, say

$$f(\lambda)=(\lambda-\lambda_1)^{k_1-1}\prod_{i=2}^{s}(\lambda-\lambda_i)^{k_i},$$

then

$$f^{(k_1-1)}(\lambda_1)=(k_1-1)!\prod_{i=2}^{s}(\lambda_1-\lambda_i)^{k_i}\neq 0, \qquad \text{but all other} \qquad f^{(j)}(\lambda_i)=0.$$

By (6),

$$f(A)=(A-\lambda_1 I)^{k_1-1}\prod_{i=2}^{s}(A-\lambda_i I)^{k_i}=[\prod_{i=2}^{s}(\lambda_1-\lambda_i)^{k_i}]N_1^{k_1-1}(A), \tag{7}$$

which is not a zero matrix by our hypothesis $N_i^j(A)\neq 0$. The polynomial $\prod_{i=1}^{s}(\lambda-\lambda_i)^{k_i}$ is thus the unique minimum polynomial of A. Conversely, if any $N_i^j(A)$ appeared in (6) is a zero matrix, say $N_1^j(A)=0$ for $1\leq j\leq k_i-1$, then $N_1^{k_1-1}(A)=$

$[N_i^j(A)]^{k_1-1-j} = 0$. The above equation shows that the polynomial $(\lambda-\lambda_1)^{k_1-1}\prod_{i=2}^{s}(\lambda-\lambda_i)^{k_i}$, of lower degree, annihilates A, so $\prod_{i=1}^{s}(\lambda-\lambda_i)^{k_i}$ can not be the minimum polynomial.

The theorem demonstrates that our using n_i in the definition of $f(A)$ is equivalent to Gantmakher's use of the minimum polynomial. It is known that finding the minimum polynomial can be tedious. The following example illustrates a scheme to find the index n_i for each nilpotent $N_i(A)$ and thereby find $f(A)$.

Example.
Let

$$A = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

The characteristic polynomial is $p(\lambda) = \lambda^2(\lambda-2)^2$, by which (4) becomes

$$f(A) = f(0)P_1(A) + f'(0)N_1(A) + f(2)P_2(A) + f'(2)N_2(A). \tag{8}$$

Notice that the formula for $f(A)$ is valid for all $f(\lambda)$ defined on the spectrum of A. Therefore we can choose $f(\lambda)$ judiciously to determine $P_i(A)$ and $N_i^j(A)$. Naturally one should start with the $N_i^j(A)$ of the highest degree in the formula, not only because it can be done readily also if it is zero, the formula will be simplified. So we start with $N_1(A)$ in the current example. Choose

$$f(\lambda) = \lambda(\lambda-2)^2, \text{ so that } f(0) = f(2) = f'(2) = 0,\ f'(0) = 4.$$

Thus substituting all these into (8), we get

$$N_1(A) = \frac{1}{4}A(A-2I)^2 = 0.$$

As we mentioned before that $P_1(A) \neq 0$, therefore $n_1 = 1$. The formula (8) is simplified to

$$f(A) = f(0)P_1(A) + f(2)P_2(A) + f'(2)N_2(A).$$

Next determine $N_2(A)$ by choosing

$$\begin{aligned} f(\lambda) &= \lambda(\lambda-2) \\ f(0) &= f(2) = 0, f'(2) = 2 \\ N_2(A) &= \frac{1}{2}A(A-2I) = 0, \end{aligned}$$

Thus $n_2 = 1$ and the minimum polynomial is $m(\lambda) = \lambda(\lambda - 2)$, and

$$f(A) = f(0)P_1(A) + f(2)P_2(A).$$

For $P_1(A)$, we choose

$$\begin{aligned} f(\lambda) &= \lambda - 2 \\ A - 2I &= -2P_1(A) \\ P_1(A) &= -\frac{1}{2}(A - 2I). \end{aligned}$$

Finally let

$$f(\lambda) = \lambda,$$

which gives us

$$P_2(A) = \frac{1}{2}A.$$

Therefore

$$\begin{aligned} f(A) &= f(0)P_1(A) + f(2)P_2(A) \\ &= -\frac{1}{2}f(0)(A - 2I) + \frac{1}{2}f(2)A. \end{aligned}$$

Consider some simple matric functions:

(i) If $f(\lambda) = \sqrt{\lambda}$, the above formula gives us

$$\sqrt{A} = \frac{1}{2}\sqrt{2}A = \frac{A}{\sqrt{2}}.$$

(ii) If $f(\lambda) = \cos(\frac{\pi}{2}\lambda)$, then

$$\cos(\frac{\pi}{2}A) = -\frac{1}{2}(1)(A - 2I) + (\frac{1}{2})(-1)A = I - A.$$

The last equation seems to indicate that the cosine of A is not an even function of A. In fact if we compute $\cos(\frac{\pi}{2}B)$ with $B = -A$, we will find

$$\begin{aligned} f(B) &= \frac{1}{2}f(0)(B + 2I) - \frac{1}{2}f(-2)B \\ \cos(\frac{\pi}{2}B) &= (1)\frac{1}{2}(B + 2I) - \frac{1}{2}(-1)B = I + B \\ &= I - A = \cos(\frac{\pi}{2}A). \end{aligned}$$

3. A different basis.

Ting in [6] uses the set
$$\{I, A, A^2, \cdots, A^{m-1}\}$$
to formulate all analytic functions, $f(A)$, of an m by m matrix A, and defines

$$f(A) = \sum_{k=0}^{m-1} a_{k+1} A^k$$

as a consequence of Cayley-Hamilton Theorem. The coefficients a_k can be determined by the spectral values of $f(\lambda)$. For the matrix A in the last example, the four coefficients $a's$ are to be solved from the 4 equations

$$\begin{aligned} f(\lambda_1) &= a_1 + a_2\lambda_1 + a_3\lambda_1^2 + a_4\lambda_1^3 \\ f'(\lambda_1) &= a_2 + 2a_3\lambda_1 + 3a_4\lambda_1^2 \\ f(\lambda_2) &= a_1 + a_2\lambda_2 + a_3\lambda_2^2 + a_4\lambda_2^3, \\ f'(\lambda_2) &= a_2 + 2a_3\lambda_2 + 3a_4\lambda_2^2. \end{aligned}$$

It is complicated. However if, at the beginning, we know the minimum polynomial for A is
$$m(\lambda) = \lambda(\lambda - 2),$$
so that
$$A(A - 2I) = 0.$$
No power higher than one is needed for $f(A)$ and the formula becomes simply

$$f(\lambda) = a_1 + a_2\lambda.$$

(i) For the case $f(\lambda) = \sqrt{\lambda}$, we have

$$\begin{aligned} f(\lambda_1) &= 0 = a_1 \\ f(\lambda_2) &= \sqrt{2} = a_1 + 2a_2 \end{aligned}$$

so
$$\sqrt{A} = \frac{A}{\sqrt{2}}.$$

(ii) For the case $f(\lambda) = \cos(\frac{\pi}{2}\lambda)$,

$$\begin{aligned} f(\lambda_1) &= 1 = a_1 \\ f(\lambda_2) &= -1 = a_1 + 2a_2 \end{aligned}$$

from which we find

$$a_1 = 1,\ a_2 = -1,$$

and

$$\cos(\frac{\pi}{2}A) = I - A.$$

References

[1] D.V. Ho, *The power & exponential of a matrix,* http://www.math.gatech.edu/~ho.
[2] D.V. Ho, *Real powers & logarithm of a matrix,* http://www.math.gatech.edu/~ho.
[3] Gantmacher,F.R.,*Matrix Theory,*Chelsea Publishing Company. 1959
[4] Lancaster,P.,*Theory of Matrices,*Academic Press. 1969
[5] Lancaster,P.&Tismenetsky,M.,*Theory of Matrices,Ed.2,*Academic Press. 1985
[6] T.C.T. Ting, *Determination of* $\mathbf{C}^{1/2,}$ $\mathbf{C}^{-1/2}$ *and More General Isotropic Tensor Functions of* $\mathbf{C}$, J. Elasticity, 15, pp.319-323 (1985).