# On Determining Functions of Matrices

Charles D. Allison
Brigham Young University

Fall 1978

**Abstract**

This note shows how to compute analytic functions of matrix arguments. It first appeared in the Fall 1978 issue of the Pi Mu Epsilon Journal. I wondered why even though we often spoke of the matrix exponential, $e^{At}$, none of my professors could tell me how to compute it. So I did some digging and came up with this. I was just finishing my first masters degree at the time.

Since the time of Cayley and Sylvester there has been great interest in the computation of matrix functions. For example, to compute the matrix exponential $e^{At}$, which satisfies the matrix differential equation with constant coefficients

$$\dot{X}(t) = AX(t),$$

methods have been developed which rely upon properties of differential equations, the Jordan canonical form, or results from linear algebra such as normality, diagonalizability, etc.[2][4][5] Most techniques for calculating a function $f$ of a matrix $A$ express $f(A)$ as a polynomial in $A$.

Of all such methods, the simplest in concept are those based on an interpolation formula introduced by Sylvester[7],

$$f(A) = \sum_{i=1}^{n} \prod_{j=1, j\neq i}^{n} \frac{A - \lambda_j I}{\lambda_i - \lambda_j} f(\lambda_i) \tag{1}$$

which holds when $A$ has distinct eigenvalues, $\lambda_1, \ldots, \lambda_n$, lying within the circle of convergence of $f(z)$.

The notion of a matrix function is usually seen for the first time in a matrix analysis course or in a course on the theory of ordinary differential equations, which are graduate courses at most schools. The purpose of this note is to give a development of Sylvester's formula accessible to the mathematics undergraduate.

A proof of (1) follows from the following generalization of the division algorithm, which is a modification of a theorem of Friedman[3].

**Theorem 1** *Let $p(z)$ be a polynomial with distinct roots, and let $f(z)$ be a function analytic in a domain $D$, which contains the roots of $p(z)$. Then there exists a unique polynomial $r(z)$, where $deg(r) = deg(p) - 1$, and a function $h(z)$, analytic in $D$, such that*

$$f(z) = p(z)h(z) + r(z) \tag{2}$$

**Proof** Denote the roots of $p(z)$ by $\lambda_i$, $i = 1, \ldots, n$, with $\lambda_i = \lambda_j \Leftrightarrow i = j$. Let $r(z)$ be the unique polynomial of degree $n-1$ that agrees with $f(z)$ at each $\lambda_i$ (this is the unique interpolating polynomial, which can be expressed by the LaGrange formula:

$$r(z) = \sum_{i=1}^{n} \prod_{j=1, j\neq i}^{n} \frac{z-\lambda_j}{\lambda_i - \lambda_j} f(\lambda_i) \quad )$$

and define

$$h(z) = \frac{f(z) - r(z)}{p(z)} \tag{3}$$

Since each zero of the denominator in (3) is also a zero of the numerator, the singularities of $h(z)$ are removable, i.e.

$$\lim_{z\to\lambda_i} (z - \lambda_i) h(z) = 0$$

hence, $h(z)$ is analytic and the result follows. ■

To compute $f(A)$, we shall let $p(z)$ in the above theorem be the characteristic polynomial of $A$ and consider equation (2). By the Cayley-Hamilton theorem, $f(A) = r(A)$ and

$$f(\lambda_i) = r(\lambda_i), i = i, \ldots, n \tag{4}$$

The equations (4) represent a linear system which can be solved for the $n$-coefficients of $r(z)$, and the calculation of $f(A) = r(A)$ is straightforward.

Notice that the LaGrange interpolating polynomial for $r(z)$ satisfying (4) shows that $r(A)$ coincides with (1), by which $f(A)$ may be computed directly. We illustrate the two processes.

## Example 1

1. Compute $f(A) = e^A$, where $A = \begin{pmatrix} 1 & 3 \\ 0 & 2 \end{pmatrix}$. $A$ has characteristic polynomial $p(z) = (1-z)(2-z)$, hence $\lambda_1 = 1, \lambda_2 = 2$. Since $r(z)$ is of the form $a_1 z + a_0$, we obtain the system:

$$\begin{aligned} e^1 &= a_1 + a_0 \\ e^2 &= 2a_1 + a_0 \end{aligned}$$

with solutions $a_1 = e^2 + e, a_0 = 2e - e^2$. Thus,

$$\begin{aligned} e^A &= r(A) = a_1 A + a_0 I \\ &= (e^2 - e)\begin{pmatrix} 1 & 3 \\ 0 & 2 \end{pmatrix} + (2e - e^2)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} e & 3(e^2 - e) \\ 0 & e^2 \end{pmatrix} \end{aligned}$$

2. By equation (1),

$$
\begin{aligned}
f(A) &= \sum_{i=1}^{2} \prod_{j=1, j\neq i}^{2} \frac{A-\lambda_j I}{\lambda_i - \lambda_j} f(\lambda_i) \\
&= \frac{\begin{pmatrix} -1 & 3 \\ 0 & 0 \end{pmatrix}}{-1} e + \frac{\begin{pmatrix} 0 & 3 \\ 0 & 1 \end{pmatrix}}{1} e^2 \\
&= \begin{pmatrix} e & 3(e^2-e) \\ 0 & e^2 \end{pmatrix}
\end{aligned}
$$

Note that method 2 is more efficient for machine computation.

The procedure for the general case follows from Theorem 2, which is based on an extension of (1), first given by Bucheim[1](cf. [5][6]).

**Theorem 2** *Let $p(z)$ be a polynomial of degree $n$ with $k$ distinct roots, $k \leq n$, and let $f(z)$ be a function analytic in a domain $D$ containing the roots of $p(z)$. Then $r(z)$ and $h(z)$ exist as in Theorem 1 and (2) holds.*

**Proof** Let $m_i$ denote the multiplicity of each root $\lambda_i$ of $p(z)$, so that $\sum_{i=1}^{k} m_i = n$. Let $r(z)$ be the polynomial of degree $n-1$ that agrees with $f(z)$ at each $\lambda_i$, and whose derivatives of all orders up to $m_i - 1$ agree with those of $f(z)$ at each $\lambda_i$, i.e.

$$
f^{(j)}(\lambda_i) = r^{(j)}(\lambda_i), j = 0, 1, \ldots, m_i - 1; i = 1, \ldots, k \tag{5}
$$

The polynomial $r(z)$ exists and is unique, being merely a form of the general Hermite osculating polynomial[2][6].

We again form the quotient (3) and notice that if $\lambda_i$ has multiplicity $m_i$, then $\lambda_i$ is a zero of order at least $m_i - 1$ of the denominator of $h(z)$, and we apply L'Hospital's rule $m_i - 1$ times to obtain a finite limit,

$$
\lim_{z \to \lambda_i} h(z) = \lim_{z \to \lambda_i} \frac{f^{m_i-1}(z) - r^{m_i-1}(z)}{p^{m_i-1}(z)} < \infty, i = 1, \ldots, k,
$$

hence $h(z)$ is analytic in $D$ and (2) follows. ■

Again we notice that (5) is a system of equations yielding the coefficients of $r(z)$, and we compute $f(A) = r(A)$ as before.

## Example 2

Compute $\sin(A)$, where $A = \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}$. $A$ has characteristic polynomial $p(z) = (1-z)^2$, hence $\lambda = 1$, and $m = 2$. Since $r(z)$ and $r'(z)$ have respectively the forms $a_1 z + a_0$ and $a_1$, we obtain the system

$$
\begin{aligned}
\sin(1) &= a_1 + a_0 \\
\cos(1) &= a_1
\end{aligned}
$$

with solutions $a_1 = \cos(1), a_0 = \sin(1) - \cos(1)$. Then,

$$\begin{aligned} \sin(A) &= \cos(1) \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix} + [\sin(1) - \cos(1)] \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \sin(1) & 3\cos(1) \\ 0 & \sin(1) \end{pmatrix} \end{aligned}$$

For large matrices, it would be computationally more efficient to evaluate $f(A)$ directly from the Hermite formula.

### Remarks

Notice that the foregoing development is valid if the minimum polynomial of $A$ is used in place of the characteristic polynomial. The results likewise hold for any scalar function $f(z)$ provided that the right side of (2) is well defined for each characteristic root. The purpose for the requirement of analyticity here was to maintain an elementary exposition by avoiding the subtleties involved in shifting from a scalar to a matrix argument in $f(z)$ (see [6]). It is clear that any analytic function can support a matrix argument by virtue of its power series.

## References


[1] A. Bucheim. An Extension of a Theorem of Prof. Sylvester's Relating to Matrices. *Philos. Mag.*, 22:173–174, 1886.

[2] E. Coddington and N. Levinson. *Theory of Ordinary Differential Equations*. McGraw-Hill, 1955.

[3] B. Friedman. *Principles and Techniques of Applied Mathematics*. Wiley, 1956.

[4] E.P. Fulmer. Computation of the Matrix Exponential. *American Math Monthly*, 82:156–159, 1975.

[5] M. Marcus and H. Ming. *A Survey of Matrix Theory and Matrix Inequalities*. Allyn and Bacon, 1964.

[6] R.F. Rinehart. The Equivalence of Definitions of a Matrix function. *American Math Monthly*, 62:395–414, 1955.

[7] J.J. Sylvester. On the Equation to the Secular Inequalities in the Planetary Theory. *Philos. Mag.*, 16:267–269, 1883.