Some Useful Results in the Theory of Matrix Functions

Nick Higham
School of Mathematics
The University of Manchester

higham@ma.man.ac.uk
http://www.ma.man.ac.uk/~higham/

**Matrix Analysis and Applications,
CIRM Luminy, October 2007**

Sherman–Morrison–Woodbury Formula

If $U, V \in \mathbb{C}^{n\times p}$ and $I_p + V^*A^{-1}U$ is nonsingular then

$$(A + UV^*)^{-1} = A^{-1} - A^{-1}U(I_p + V^*A^{-1}U)^{-1}V^*A^{-1}.$$

Question

Why does this formula hold?

Sherman–Morrison–Woodbury Formula

If $U, V \in \mathbb{C}^{n\times p}$ and $I_p + V^*A^{-1}U$ is nonsingular then

$$(A + UV^*)^{-1} = A^{-1} - A^{-1}U(I_p + V^*A^{-1}U)^{-1}V^*A^{-1}.$$

Question

Why does this formula hold?

Obtained using $A + UV^* = A(I + A^{-1}U \cdot V^*)$ and

$$(I_m + AB)^{-1} = I - A(I_n + BA)^{-1}B \qquad \begin{cases} A \in \mathbb{C}^{m\times n} \\ B \in \mathbb{C}^{n\times m} \end{cases}$$

Identity

$$(I + AB)^{-1} = I - A(I + BA)^{-1}B.$$

$$\begin{aligned} I &= I + AB - (I + AB)A(I + BA)^{-1}B \\ &= I + AB - A(I + BA)(I + BA)^{-1}B \\ &= I + AB - AB \\ &= I \quad \surd \end{aligned}$$

Identity

$$(I + AB)^{-1} = I - A(I + BA)^{-1}B.$$

$$\begin{aligned} I &= I + AB - (I + AB)A(I + BA)^{-1}B \\ &= I + AB - A(I + BA)(I + BA)^{-1}B \\ &= I + AB - AB \\ &= I \quad \surd \end{aligned}$$

Key equation: $(I + AB)A = A(I + BA)$, or

Identity

$$(I + AB)^{-1} = I - A(I + BA)^{-1}B.$$

$$\begin{aligned} I &= I + AB - (I + AB)A(I + BA)^{-1}B \\ &= I + AB - A(I + BA)(I + BA)^{-1}B \\ &= I + AB - AB \\ &= I \quad \surd \end{aligned}$$

Key equation: $(I + AB)A = A(I + BA)$, or

World's Most Fundamental Matrix Equation

$$(AB)A = A(BA).$$

Application of WMFME

$$\begin{aligned} & (AB)A = A(BA) \\ \Rightarrow \quad & (AB)^2 A = ABA(BA) = A(BA)^2. \end{aligned}$$

In general, for any poly p,

$$p(AB)A = Ap(BA).$$

- Does the same hold for arbitrary f?

$f(AB)$ and $f(BA)$

Lemma

Let $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{n\times m}$ and let $f(AB)$ and $f(BA)$ be defined. Then

$$Af(BA) = f(AB)A.$$

$f(AB)$ and $f(BA)$

Lemma

Let $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{n\times m}$ and let $f(AB)$ and $f(BA)$ be defined. Then

$$Af(BA) = f(AB)A.$$

Proof. There is a single polynomial p such that $f(AB) = p(AB)$ and $f(BA) = p(BA)$. Hence

$$Af(BA) = Ap(BA) = p(AB)A = f(AB)A.$$

Special Case

Take $f(t) = t^{1/2}$. When AB (and hence also BA) has no eigenvalues on $\mathbb{R}^-$,

$$A(BA)^{1/2} = (AB)^{1/2}A.$$

Special Case

Take $f(t) = t^{1/2}$. When AB (and hence also BA) has no eigenvalues on $\mathbb{R}^-$,

$$A(BA)^{1/2} = (AB)^{1/2}A.$$

Useful, but

$$Af(BA) = f(AB)A$$

cannot be solved for $f(BA)$ in terms of $f(AB)$.

New Theorem

Theorem (Harris 1993; H 2008)

Let $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{n\times m}$, with $m \geq n$, and assume BA is nonsingular. Then

$$f(\alpha I_m + AB) = f(\alpha)I_m + A(BA)^{-1}\big(f(\alpha I_n + BA) - f(\alpha)I_n\big)B.$$

New Theorem

Theorem (Harris 1993; H 2008)

Let $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{n\times m}$, with $m \geq n$, and assume BA is nonsingular. Then

$$f(\alpha I_m + AB) = f(\alpha)I_m + A(BA)^{-1}\big(f(\alpha I_n + BA) - f(\alpha)I_n\big)B.$$

Proof. Let $g(t) = t^{-1}(f(\alpha + t) - f(\alpha))$.
Then $f(\alpha I + X) = f(\alpha)I + Xg(X)$.
Hence, using the lemma,

$$\begin{aligned} f(\alpha I_m + AB) &= f(\alpha)I_m + ABg(AB) \\ &= f(\alpha)I_m + Ag(BA)B \\ &= f(\alpha)I_m + A(BA)^{-1}\big(f(\alpha I_n + BA) - f(\alpha)I_n\big)B. \end{aligned}$$

Example: Rank 2 Perturbation of I

Consider $f(\alpha I_n + uv^* + xy^*)$, where $u, v, x, y \in \mathbb{C}^n$. Write

$$uv^* + xy^* = [\, u \quad x \,] \begin{bmatrix} v^* \\ y^* \end{bmatrix} \equiv AB.$$

Then

$$C := BA = \begin{bmatrix} v^*u & v^*x \\ y^*u & y^*x \end{bmatrix} \in \mathbb{C}^{2\times 2}.$$

$$f(\alpha I_n + uv^* + xy^*) = f(\alpha) I_n + [\, u \quad x \,]\, C^{-1} \big(f(\alpha I_2 + C) - f(\alpha) I_2 \big) \begin{bmatrix} v^* \\ y^* \end{bmatrix}$$

Example: Rank 2 Perturbation of I

Consider $f(\alpha I_n + uv^* + xy^*)$, where $u, v, x, y \in \mathbb{C}^n$. Write

$$uv^* + xy^* = [\, u \quad x \,] \begin{bmatrix} v^* \\ y^* \end{bmatrix} \equiv AB.$$

Then

$$C := BA = \begin{bmatrix} v^*u & v^*x \\ y^*u & y^*x \end{bmatrix} \in \mathbb{C}^{2\times 2}.$$

$$f(\alpha I_n + uv^* + xy^*) = f(\alpha) I_n + [\, u \quad x \,]\, C^{-1} \big(f(\alpha I_2 + C) - f(\alpha) I_2 \big) \begin{bmatrix} v^* \\ y^* \end{bmatrix}$$

For $A \in \mathbb{C}^{2\times 2}$, $f(A) = f(\lambda_1) I + f[\lambda_1, \lambda_2](A - \lambda_1 I).$

Function of Jordan block

$$A = Z\mathrm{diag}(J_1, \dots, J_p)Z^{-1} \Rightarrow f(A) = Z\mathrm{diag}(f(J_1), \dots, f(J_p))Z^{-1}.$$

$$J_k = \begin{bmatrix} \lambda_k & 1 & & \\ & \lambda_k & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k \end{bmatrix} \in \mathbb{C}^{m_k \times m_k},$$

$$f(J_k) = \begin{bmatrix} f(\lambda_k) & f'(\lambda_k) & \dots & \dfrac{f^{(m_k-1)}(\lambda_k)}{(m_k-1)!} \\ & f(\lambda_k) & \ddots & \vdots \\ & & \ddots & f'(\lambda_k) \\ & & & f(\lambda_k) \end{bmatrix}.$$

Theorem

Let $A \in \mathbb{C}^{n \times n}$ with eigenvalues λ_k.

1. *If $f'(\lambda_k) \neq 0$ then for every $J(\lambda_k)$ in A there is a Jordan block of the* ***same*** *size in $f(A)$ for $f(\lambda_k)$.*

Theorem

Let $A \in \mathbb{C}^{n \times n}$ with eigenvalues λ_k.

1. *If $f'(\lambda_k) \neq 0$ then for every $J(\lambda_k)$ in A there is a Jordan block of the* ***same*** *size in $f(A)$ for $f(\lambda_k)$.*

2. *Let $f'(\lambda_k) = f''(\lambda_k) = \cdots = f^{(\ell-1)}(\lambda_k) = 0$ but $f^{(\ell)}(\lambda_k) \neq 0$, where $\ell \geq 2$, and consider $J(\lambda_k)$ of size r in A.*

   (i) If $\ell \geq r$, $J(\lambda_k)$ splits into r 1×1 ***Jordan blocks*** *for $f(\lambda_k)$ in $f(A)$.*

   (ii) If $\ell \leq r - 1$, $J(\lambda_k)$ splits into Jordan blocks for $f(\lambda_k)$ in $f(A)$ as follows:

   - *$\ell - q$* ***Jordan blocks of size*** *p,*
   - *q* ***Jordan blocks of size*** *$p + 1$,*

   where $r = \ell p + q$ with $0 \leq q \leq \ell - 1$, $p > 0$.

Theorem

Let $A \in \mathbb{C}^{n\times n}$ with eigenvalues λ_k.

1. *If $f'(\lambda_k) \neq 0$ then for every $J(\lambda_k)$ in A there is a Jordan block of the* ***same*** *size in $f(A)$ for $f(\lambda_k)$.*

2. *Let $f'(\lambda_k) = f''(\lambda_k) = \cdots = f^{(\ell-1)}(\lambda_k) = 0$ but $f^{(\ell)}(\lambda_k) \neq 0$, where $\ell \geq 2$, and consider $J(\lambda_k)$ of size r in A.*

   (i) If $\ell \geq r$, $J(\lambda_k)$ splits into r 1×1 ***Jordan blocks*** *for $f(\lambda_k)$ in $f(A)$.*

   (ii) If $\ell \leq r-1$, $J(\lambda_k)$ splits into Jordan blocks for $f(\lambda_k)$ in $f(A)$ as follows:

   - *$\ell - q$* ***Jordan blocks of size*** *p,*
   - *q* ***Jordan blocks of size*** *$p+1$,*

   where $r = \ell p + q$ with $0 \leq q \leq \ell - 1$, $p > 0$.

Application: Matrix Equation

Does the equation

$$\cosh(X) = \begin{bmatrix} 1 & a & a & \dots & a \\ & 1 & a & \dots & a \\ & & 1 & \dots & \vdots \\ & & & \ddots & a \\ & & & & 1 \end{bmatrix} = A \in \mathbb{C}^{n\times n}$$

have a solution for $a \neq 0$ and $n > 1$?

Application: Matrix Equation

Does the equation

$$\cosh(X) = \begin{bmatrix} 1 & a & a & \dots & a \\ & 1 & a & \dots & a \\ & & 1 & \dots & \vdots \\ & & & \ddots & a \\ & & & & 1 \end{bmatrix} = A \in \mathbb{C}^{n\times n}$$

have a solution for $a \neq 0$ and $n > 1$?

No!

A has just one Jordan block.
$\Lambda(X) = \{\cosh^{-1}(1)\} = \{0\}$. But $f'(0) = \sinh(0) = 0$, so $\cosh(X)$ must have more than one Jordan block.

Application: Matrix Square Root

Show that if $A \in \mathbb{C}^{n\times n}$ has a defective zero eigenvalue then A does not have a square root that is a polynomial in A.

Application: Matrix Square Root

Show that if $A \in \mathbb{C}^{n \times n}$ has a defective zero eigenvalue then A does not have a square root that is a polynomial in A.

Solution

Key fact: squaring a matrix leaves intact the number and size of Jordan blocks for $\lambda \neq 0$ but splits blocks for $\lambda = 0$.

Hence defective zero eigenvalue of $A \Rightarrow X$ has Jordan block $\geq 3 \times 3$ for $\lambda = 0$.

But then $X \neq p(A)$, because $A = X^2$ has more Jordan blocks than X.

Function of Block Triangular Matrix (1)

Theorem

$$A = \begin{array}{c c} & \begin{array}{cc} n-1 & 1 \end{array} \\ \begin{array}{c} n-1 \\ 1 \end{array} & \begin{bmatrix} B & c \\ 0 & \lambda \end{bmatrix} \end{array} \quad \Rightarrow \quad f(A) = \begin{bmatrix} f(B) & g(B)c \\ 0 & f(\lambda) \end{bmatrix}$$

where $g(z) = f[z, \lambda]$.

If $\lambda \notin \Lambda(B)$ *then* $g(B) = (B - \lambda I)^{-1}(f(B) - f(\lambda)I)$.

Theorem

$$f\left(\begin{bmatrix} X & E \\ 0 & X \end{bmatrix}\right) = \begin{bmatrix} f(X) & L(X,E) \\ 0 & f(X) \end{bmatrix}.$$

Recall, Fréchet derivative L:

$$f(X+E) - f(X) - L(X,E) = o(\|E\|).$$

Function of Block Triangular Matrix (2)

Theorem

Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \quad D = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}, \quad N = \begin{bmatrix} 0 & A_{12} \\ 0 & 0 \end{bmatrix}.$$

Then $f(A) = f(D) + L(D, N)$ *(i.e., $o(\cdot)$ term in Frechét definition is zero).*

Bibliography I

- Lawrence A. Harris.
  Computation of functions of certain operator matrices.
  Linear Algebra Appl., 194:31–34, 1993.

- Nicholas J. Higham.
  Functions of Matrices: Theory and Computation.
  Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.
  Book to appear.