Polynomials of matrices

What happens to Jordan blocks when we take a scalar polynomial, and apply it to a (square) matrix? E.g.,

$$p(x) = 1 + 3x - 5x^2 \implies p(A) = I + 3A - 5A^2.$$

Lemma

If $A = S \operatorname{blkdiag}(J_1, J_2, \dots, J_s) S^{-1}$ is a Jordan form, then $p(A) = S \operatorname{blkdiag}(p(J_1), p(J_2), \dots, p(J_s)) S^{-1}$, and

$$p(J_i) = \begin{bmatrix} p(\lambda_i) & p'(\lambda_i) & \dots & \frac{1}{k!} p^{(k)}(\lambda_i) \\ & p(\lambda_i) & \ddots & \vdots \\ & & \ddots & p'(\lambda_i) \\ & & & p(\lambda_i) \end{bmatrix}.$$

(Proof: Taylor expansion of p around λ.)

Functions of matrices [Higham book, '08]

We can extend the same definition to arbitrary scalar functions:

Given a function $f : U \subseteq \mathbb{C} \to \mathbb{C}$, we say that f is defined on A if f is defined and differentiable at least $n(\lambda_i) - 1$ times on each eigenvalue λ_i of A.
($n(\lambda_i) =$ size of largest Jordan block with eigenvalue λ_i)

Definition attempt

If $A = S\,\mathrm{blkdiag}(J_1, J_2, \dots, J_s)S^{-1}$ is a Jordan form, then $f(A) = S\,\mathrm{blkdiag}(f(J_1), f(J_2), \dots, f(J_s))S^{-1}$, where

$$f(J_i) = \begin{bmatrix} f(\lambda_i) & f'(\lambda_i) & \dots & \frac{1}{k!}f^{(k)}(\lambda_i) \\ & f(\lambda_i) & \ddots & \vdots \\ & & \ddots & f'(\lambda_i) \\ & & & f(\lambda_i) \end{bmatrix}.$$

(Reasonable doubt: is it independent of the choice of S?)

Alternate definition: via Hermite interpolation

Definition

$f(A) = p(A)$, where p is a polynomial such that
$f(\lambda_i) = p(\lambda_i), f'(\lambda_i) = p'(\lambda_i), \dots, f^{(n(\lambda_i)-1)}(\lambda_i) = p^{(n(\lambda_i)-1)}(\lambda_i)$
for each i.

We may use this as a definition of $f(A)$ (and it does not depend on S).

Obvious from the definitions that it coincides with the previous one.

Remark: if $A \in \mathbb{C}^{m\times m}$ has multiple Jordan blocks with the same eigenvalue, these may be fewer than m conditions.
Remark: be careful when you say "all matrix functions are polynomials", because p depends on A.

Example: square root

$$A = \begin{bmatrix} 4 & 1 & & \\ & 4 & 1 & \\ & & 4 & \\ & & & 0 \end{bmatrix}, \quad f(x) = \sqrt{x}$$

We look for an interpolating polynomial with

$$p(0) = 0,\ p(4) = 2,\ p'(4) = f'(4) = \frac{1}{4},\ p''(4) = f''(4) = -\frac{1}{32}.$$

I.e.,

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 4^3 & 4^2 & 4 & 1 \\ 3 \cdot 4^2 & 2 \cdot 4 & 1 & 0 \\ 6 \cdot 4 & 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_3 \\ p_2 \\ p_1 \\ p_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ \frac{1}{4} \\ -\frac{1}{32} \end{bmatrix},$$

$$p(x) = \frac{3}{256}x^3 - \frac{5}{32}x^2 + \frac{15}{16}x.$$

Example – continues

$$p(A) = \frac{3}{256}A^3 - \frac{5}{32}A^2 + \frac{15}{16}A = \begin{bmatrix} 2 & \frac{1}{4} & -\frac{1}{64} & \\ & 2 & \frac{1}{4} & \\ & & 2 & \\ & & & 0 \end{bmatrix}.$$

(One can check that $f(A)^2 = A$.)

Hermite interpolation

A suitable polynomial always exists:

Theorem

Given distinct points $x_1, x_2, \ldots, x_n$, multiplicities $m_1, m_2, \ldots, m_n$, there exists a unique polynomial of degree $d < m_1 + m_2 + \cdots + m_n$ such that (for all $i = 1, \ldots, n$)

$$p(x_i) = y_{i,0},\ p'(x_i) = y_{i,1},\ \ldots,\ p^{(m_i-1)}(x_i) = y_{i,m_i-1},$$

where the y_{ij} are prescribed values.

Proof (sketch)

- Interpolation conditions $\iff$ square linear system $Vp = y$, where p is the vector of polynomial coefficients.
- We prove that V has no kernel. If $Vz = 0$ for a vector z, then the associated polynomial $z(x)$ has roots at x_i of multiplicity m_i. By degree reasons it must be the zero polynomial.

Example – square root

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad f(x) = \sqrt{x}$$

does not exist (because $f'(0)$ is not defined).

(Indeed, there is no matrix such that $X^2 = A$: every 2×2 nilpotent matrix X has Jordan form equal to A, thus $X^2 = 0$.)

Example – matrix exponential

$$A = S \begin{bmatrix} -1 & & & \\ & 0 & & \\ & & 1 & 1 \\ & & & 1 \end{bmatrix} S^{-1}, \quad f(x) = \exp(x).$$

$$\exp(A) = S \begin{bmatrix} e^{-1} & & & \\ & 1 & & \\ & & e & e \\ & & & e \end{bmatrix} S^{-1}$$

Can also be obtained as $I + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \dots$

(This is not immediate, for Jordan blocks; we will prove later in more generality that Taylor series ‘work’.)

Example – matrix sign

$$A = S \begin{bmatrix} -3 & & & \\ & -2 & & \\ & & 1 & 1 \\ & & & 1 \end{bmatrix} S^{-1}, \quad f(x) = \operatorname{sign}(x) = \begin{cases} 1 & \operatorname{Re} x > 0, \\ -1 & \operatorname{Re} x < 0. \end{cases}$$

$$f(A) = S \begin{bmatrix} -1 & & & \\ & -1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} S^{-1}.$$

Not constant (for general S).

Instead, we can recover stable / unstable invariant subspaces of A as $\ker(f(A) \pm I)$.

If we found a way to compute $f(A)$ without diagonalizing, we could use it to compute eigenvalues via bisection...

Example – complex square root

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad f(x) = \sqrt{x}$$

We can play around with branches: let us say $f(i) = \frac{1}{\sqrt{2}}(1+i)$, $f(-i) = \frac{1}{\sqrt{2}}(1-i)$.

Polynomial: $p(x) = \frac{1}{\sqrt{2}}(1+x)$.

$$p(A) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

(This is the so-called principal square root – we have chosen the values of $f(\pm i)$ in the right half-plane — other choices are possible).

(We get a non-real square root of A, if we choose non-conjugate values for $f(i)$ and $f(-i)$)

Example – nonprimary square root

With our definition, if we have

$$A = S \begin{bmatrix} 1 & & \\ & 1 & \\ & & 2 \end{bmatrix} S^{-1}, \quad f(x) = \sqrt{x}$$

we cannot get

$$f(A) = S \begin{bmatrix} 1 & & \\ & -1 & \\ & & \sqrt{2} \end{bmatrix} S^{-1} :$$

either $f(1) = 1$, or $f(1) = -1 \ldots$

This would also be a solution of $X^2 = A$, though.

Nonprimary matrix functions

If a matrix A has multiple eigenvalues, one could also define a 'square root' by choosing different signs on Jordan blocks with the same eigenvalue: for instance, $\begin{bmatrix} 1 & \\ & -1 \end{bmatrix}$ as a square root of I_2 (or also $V \begin{bmatrix} 1 & \\ & -1 \end{bmatrix} V^{-1}$ for any invertible $V \dots$).

These are called nonprimary matrix functions (and they are not matrix functions with our definition).

They all satisfy $f(A)^2 = A$.

These are not polynomials in A.

Some properties

- If the eigenvalues of A are $\lambda_1, \ldots, \lambda_s$, the eigenvalues of $f(A)$ are $f(\lambda_1), \ldots, f(\lambda_s)$. (Remark: geometric multiplicities may increase)
- $f(A)g(A) = g(A)f(A) = (fg)(A)$ (since they are all polynomials in A). Analogously for sums and compositions.
- If $f_n \to f$ together with 'enough derivatives' (for instance because they are analytic and the convergence is uniform), then $f_n(A) \to f(A)$.
- If a sequence of matrices $A_n \to A$, then $f(A_n) \to f(A)$. Proof: we will see it later.

Cauchy integrals

If f is holomorphic (analytic) on and inside a contour Γ that encloses the eigenvalues of A,

$$f(A) = \frac{1}{2\pi i} \int_\Gamma f(z)(zI - A)^{-1} dz.$$

Generalizes the analogous scalar formula (Cauchy's integral formula).

Proof Use a Jordan form $A = VJV^{-1} \in \mathbb{C}^{m \times m}$; we can pull the integral inside each Jordan block. Then,

$$\frac{1}{2\pi i}\int_\Gamma f(z)(zI-J)^{-1}dz = \frac{1}{2\pi i}\int_\Gamma f(z)\begin{bmatrix} z-\lambda & -1 & & \\ & z-\lambda & -1 & \\ & & \ddots & \ddots \\ & & & z-\lambda \end{bmatrix}^{-1} dz$$

$$= \begin{bmatrix} \frac{1}{2\pi i}\int_\Gamma \frac{f(z)}{z-\lambda}dz & \frac{1}{2\pi i}\int_\Gamma \frac{f(z)}{(z-\lambda)^2}dz & \dots & \frac{1}{2\pi i}\int_\Gamma \frac{f(z)}{(z-\lambda)^{n-1}}dz \\ & \ddots & \ddots & \ddots \end{bmatrix}$$

$$= \begin{bmatrix} f(\lambda) & f'(\lambda) & \dots & \frac{f^{(n-1)}}{(n-1)!} \\ & \ddots & \ddots & \ddots \end{bmatrix}$$

by the scalar version of Cauchy's integral formula (including the version that computes derivatives).

Corollary $f(A)$ is continuous in A (since that integral formula is so).

Continuity

The previous proof works for holomorphic f.

$f(A)$ is continuous in A also in more general settings (as long as there are enough derivatives), but the proof is more complex.

Sketch:

- The coefficients of the interpolating polynomial are continuous in the nodes (not clear at all from our proof!).
- Take a sequence $A_n \to A$, and let p_n be interpolating polynomials of f in the eigenvalues of A_n.
- $\|f(A) - f(A_n)\| = \|p_n(A_n) - p(A)\| \leq \|p_n(A_n) - p_n(A)\| + \|p_n(A) - p(A)\|$, and both terms are bounded.

Methods

Matrix functions arise in several areas: solving ODEs (e.g., $\exp A$), matrix analysis (e.g., square roots), physics, ...

Main methods to compute them:

- Factorizations (eigendecompositions, Schur...),
- Matrix versions of scalar iterations (e.g., Newton on $x^2 = a$),
- Interpolation / approximation,
- Complex integrals + quadrature,
- Arnoldi.