Note Title 2020-03-13

$f: U \subseteq \mathbb{C} \to \mathbb{C}$ $\quad A = SJS$ $\qquad f(A) = S \begin{bmatrix} f(J_1) & & \\ & \ddots & \\ & & f(J_s) \end{bmatrix} S^{-1}$

$$f(J_i) = f\left(\begin{bmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix} \right) = \begin{bmatrix} f(\lambda) & f'(\lambda) & \cdots & \frac{f^{(n_i-1)}(\lambda)}{(n-1)!} \\ & \ddots & \ddots & \vdots \\ & & \ddots & f'(\lambda) \\ & & & f(\lambda) \end{bmatrix} \quad (*)$$

f <u>defined</u> in A iff f regular enough to write $(*)$ for all blocks.

Properties of matrix functions

1) if A has eigenvalues $\lambda_1, \lambda_2 \ldots \lambda_n$, then

$f(A)$ has eigenvalues $f(\lambda_1), f(\lambda_2), \dots f(\lambda_n)$.

<u>Algebraic</u> multiplicities are preserved

<u>Geometric</u> multiplicities may increase

$$f(J) = \begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2} & \cdots \\ & f(\lambda) & \ddots & \vdots \\ & & \ddots & \frac{f''(\lambda)}{2} \\ & 0 & \ddots & f'(\lambda) \\ & & & f(\lambda) \end{bmatrix}$$

geometric multiplicity of $f(\lambda)$ in $f(J)$

$\dim \operatorname{Ker} f(J) - f(\lambda) I =$

$$= \begin{bmatrix} 0 & f'(\lambda) & \frac{f''(\lambda)}{2} & \cdots \\ & \ddots & \ddots & \vdots \\ & & \ddots & \frac{f''(\lambda)}{2} \\ & & & f'(\lambda) \\ & & & 0 \end{bmatrix}$$

If $f'(\lambda) \neq 0$, then $\dim \operatorname{Ker} = 1$

and $m_g(f(\lambda), f(J)) = m_g(\lambda, J) = 1$

If $f'(\lambda)=0$, then $m_g(f(\lambda), f(J)) > 1$

Given $f(x)$, $g(x)$ functions, $h(x) := f(x) + g(x)$

$$h(A) = f(A) + g(A)$$
$$\| \qquad\qquad \| \qquad\qquad \|$$
$$p_h(A) \qquad p_f(A) + p_g(A)$$

And the interpolating polynomial of h is the sum of the two interpolating polynomials of f & g.

Similarly, if $h(x) := f(x)\,g(x) = g(x)\,f(x)$

$$h(A) = f(A)\,g(A) = g(A)\,f(A)$$

In particular, functions of the same matrix commute.

$$\exp(A)\,A^2 = A^2 \exp(A).$$

Similarly, $h(x) = f(g(x))$, $\quad h(A) = f(g(A))$.

Continuity: 1) If $f_1, f_2, \ldots \to f$, is it true that $f_n(A) \to f(A)$?

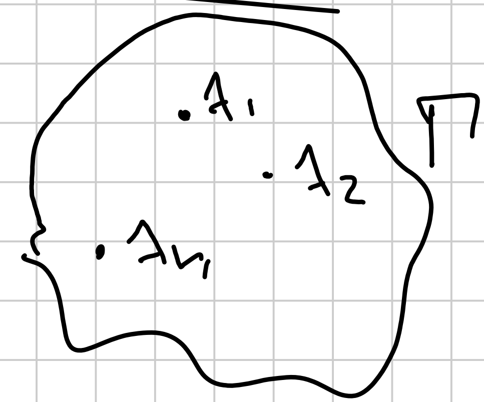
2) If $A_n \to A$, $\quad f(A_n) \overset{?}{\to} f(A)$?

1): If $f_n \to f$, $\quad f_n' \to f'$, $f_n'' \to f''$ with enough derivatives for $f(A)$ to be defined, then $f_n(A) \to f(A)$

$$f_n(J_i) = \begin{bmatrix} f_n(\lambda) & f_n'(\lambda) & \cdots & \frac{f_n^{(k-1)}}{(k-1)!} \\ & \ddots & & \vdots \\ & & \ddots & f_n'(\lambda) \\ 0 & & & f_n(\lambda) \end{bmatrix} \longrightarrow \begin{bmatrix} f(\lambda) & f'(\lambda) & \cdots & \frac{f^{(k-1)}}{(k-1)!} \\ & \ddots & \ddots & \\ & & \ddots & \ddots \\ & & & \end{bmatrix} = f(J_i)$$

Cauchy integral formula

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



where Γ is a contour that encloses the eigenvalues of A

Proof: Jordan form: $A = SJS^{-1}$

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

$$= S \underbrace{\left(\frac{1}{2\pi i} \int_\Gamma f(z) (zI - J)^{-1} dz \right)}_{} S^{-1} = S \begin{bmatrix} f(J_1) & & \\ & \ddots & \\ & & f(J_s) \end{bmatrix} S^{-1} = f(A).$$

block diagonal with blocks $\downarrow$

$$\frac{1}{2\pi i}\int_{\Gamma} f(z)\begin{bmatrix} z-\lambda & -1 & & 0 \\ & z-\lambda & -1 & \\ & & \ddots & \ddots \\ & & & -1 \\ 0 & & & z-\lambda \end{bmatrix}^{-1} dz = \frac{1}{2\pi i}\int_{\Gamma} f(z)\begin{bmatrix} (z-\lambda)^{-1} & (z-\lambda)^{-2} & (z-\lambda)^{-3} & \cdots & (z-\lambda)^{-n} \\ 0 & (z-\lambda)^{-1} & (z-\lambda)^{-2} & \cdots & (z-\lambda)^{-n+1} \\ & & \ddots & \ddots & \\ 0 & & & & (z-\lambda)^{-1} \end{bmatrix} dz$$

$$= \begin{bmatrix} \frac{1}{2\pi i}\int \frac{f(z)}{z-\lambda}dz & \frac{1}{2\pi i}\int \frac{f(z)}{(z-\lambda)^2}dz & \cdots & \frac{1}{2\pi i}\int \frac{f(z)}{(z-\lambda)^n}dz \\ & \ddots & \ddots & \ddots \\ 0 & & & \end{bmatrix}$$

$$= \begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2} & \cdots & \frac{f^{n}(\lambda)}{n!} \\ & \ddots & \ddots & \ddots & \ddots \end{bmatrix} = f(J)$$

Continuity w.r.t. A follows from the Cauchy integral formula:

$A_n \to A$ $\quad \|A_n - A\| \leq \varepsilon$ for suff. large n

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

$$= \left\| \frac{1}{2\pi i} \int_\Gamma f(z) \left[\underbrace{(zI - A_n)^{-1} - (zI-A)^{-1}} \right] dz \right\|$$

identity:

$$M^{-1} - N^{-1} = N^{-1}(N-M)M^{-1}$$

$$\leq \frac{1}{2\pi i}\int_\Gamma \|f(z)\| \, \|(zI-A_n)^{-1}-(zI-A)^{-1}\| \cdot dz \leq$$

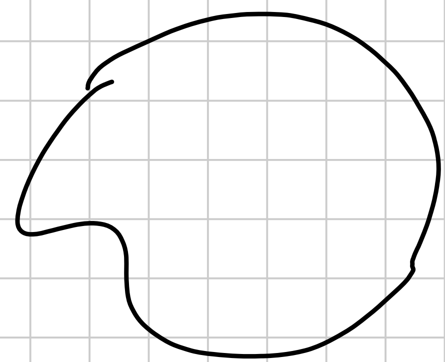
$$\leq \frac{1}{2\pi i}\int_\Gamma \|f(z)\| \cdot \|(zI-A)^{-1}\| \cdot \|A-A_n\| \cdot \|(zI-A_n)^{-1}\| \, dz$$

bounded by $\max_\Gamma |f(z)|$

bounded if all eigenvalues of the A_n are seperated from Γ

$\|A-A_n\| \leq \varepsilon$

Γ



Computing $f(A)$ (for general f)

1) If A diagonalizable, $A = V \cdot \Lambda \cdot V$ $\quad \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$

$$f(A) = V \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix} V^{-1}$$

All ok if V is orthogonal (unitary)

Suppose we compute $\tilde{f}(\lambda_i)$ with error

$|\tilde{f}(\lambda_i) - f(\lambda_i)| \leq \varepsilon$. Then

$$f(A) - \tilde{f}(A) = V\left(\begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix} - \begin{bmatrix} \tilde{f}(\lambda_1) & & \\ & \ddots & \\ & & \tilde{f}(\lambda_n) \end{bmatrix}\right)V^{-1}$$

$$\|f(A) - \tilde{f}(A)\| \leq \|V\| \cdot \underbrace{\left\|\begin{bmatrix} f(\lambda_1)-\tilde{f}(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n)-\tilde{f}(\lambda_n) \end{bmatrix}\right\|} \cdot \|V^{-1}\|$$

$$= \|V\| \cdot \varepsilon \cdot \|V^{-1}\| = \kappa(V) \cdot \varepsilon$$

<u>Ex</u> $f\left(\begin{bmatrix} 3 & -1 \\ 1 & 1 \end{bmatrix}\right)$ $\quad f = \text{sqrt}$

Polynomial s.t. $P(2) = \sqrt{2}$ $\quad f'(x) = \frac{1}{2\sqrt{x}}$

$$P'(2) = \frac{1}{2\sqrt{2}}$$

$$P(x) = \sqrt{2} + (x-2)\frac{1}{2\sqrt{2}}$$

$$= \frac{1}{2\sqrt{2}}x + \sqrt{2} - \frac{1}{\sqrt{2}}$$

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ \frac{1}{2\sqrt{2}} \end{bmatrix}$$

$$p_1 = \frac{1}{2\sqrt{2}} \qquad p_0 = \sqrt{2} - 2 \cdot \frac{1}{2\sqrt{2}} =$$

$$= p_0 = \sqrt{2} - \frac{1}{\sqrt{2}}$$

$$\left(\sqrt{2} - \frac{1}{\sqrt{2}}\right) + \frac{1}{2\sqrt{2}}x$$

---

2) Finding interpolating polynomials and evaluating them.

Question: How does one evaluate a polynomial in a matrix?

1) $p(A) = p_0 I + p_1 A + p_2 A^2 + \cdots + p_d A^d$, after evaluating $A, A^2, \dots A^d$ via successive products $O(dn^2)$ for $p(A)v$

2) $p_0 I + p_1 I \cdots \Big(p_{d-2} I + \big(p_{d-1} I + p_d A\big)A\Big)A \cdots \Big)A$ (Horner rule)

$A(\cdots(A(A p_d v + p_{d-1} v) + p_{d-2} v) \cdots + p_1 v) + p_0 v$ $\quad O(dn^2)$ for $p(A)v$

For <u>scalar arguments</u>, 1) needs $d-1$ products for $A, A^2, \dots A^d$
$d-1$ products & $d-1$ sums for the formula

2) needs $d-1$ sums & $d-1$ products

<u>For matrices</u>, 1) needs $d-1$ <u>matrix</u> products for $A, A^2, \dots A^d$ $\boxed{O(n^3)}$
$d-1$ scalar-matrix products $O(n^2)$
$d-1$ sums of matrices $O(n^2)$

2) needs $d-1$ matrix products $\boxed{O(n^3)}$
1 scalar-matrix product $O(n^2)$
$d-1$ matrix sums $O(n^2)$

Same cost (at least up to lower order terms)

Another method: Paterson-Stockmayer method:
divide terms in chunks of size $\sqrt{d}$: e.g.

$$\left(p_8 A^2 + p_7 A + p_6 I\right) A^6 + \left(p_5 A^2 + p_4 A + p_3 I\right) A^3 + \left(p_2 A^2 + p_1 A + p_0\right)$$

matmuls: 1 to form A^2 | 3 matmuls for degree 8

1 to form $A^2 \cdot A = A^3$

1 to form $A^3 \cdot A^3 = A^6$

$\sqrt{d}$ matmuls for $A^1, A^2, \dots A^{\sqrt{d}}$ $\sqrt{d}$ more for $A^{2\sqrt{d}}, A^{3\sqrt{d}}, \dots A^d$

storage $O(\sqrt{d})$ matrices (larger than the other methods $O(1)$)

The value $\tilde{Y}$ computed with this method satisfies

$$\left|\tilde{Y} - p(A)\right| \lesssim \left(|p_0| I + |p_1| \cdot |A| + |p_2| \cdot |A|^2 + \dots + |p_d| \cdot |A|^d\right) O(d \cdot u)$$

$|X|$ means the square matrix with elements $|X_{ij}|$

$\emptyset$ machine precision

Other part of the question: if I use polynomials to approximate a matrix function, e.g.

$$\exp(A) \approx I + A + \frac{1}{2}A^2 + \dots + \frac{1}{k!}A^k,$$

does the result converge to the right matrix?

Note that if you have a sequence of polynomials p_n s.t.

$p_n(x) \to f(x)$ (for all x), this may not be sufficient to ensure $p(A) \to f(A)$:

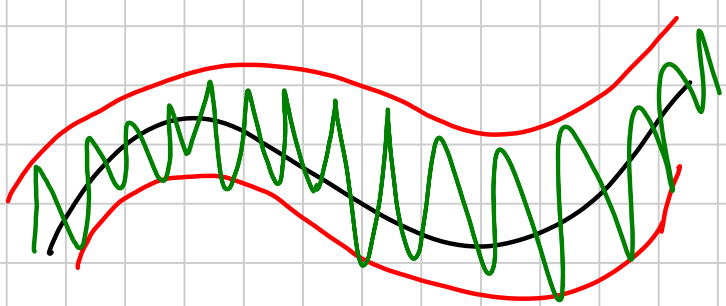
If $|p_n(\lambda)-f(\lambda)|\leq \varepsilon$ for each eigenvalue λ, then

$$\|f(A)-p_n(A)\| \leq \|V\|\cdot\left\|\begin{bmatrix} p_n(\lambda_1)-f(\lambda_1) & & \\ & \ddots & \\ & & p_n(\lambda_s)-f(\lambda_s)\end{bmatrix}\right\|\|V^{-1}\|$$

$$\leq \underbrace{\|V\|\cdot\|V^{-1}\|}\cdot\varepsilon$$

Suppose you have polynomials s.t. $p_n(x)\to f(x)$ $\forall x\in[a,b]$

$p_n'(x) \not\to f'(x)$



$$p_n\left(\begin{bmatrix}\lambda & 1\\ 0 & \lambda\end{bmatrix}\right)=\begin{bmatrix}p_n(\lambda) & p_n'(\lambda)\\ 0 & p_n(\lambda)\end{bmatrix}\not\to\begin{bmatrix}f(\lambda) & f'(\lambda)\\ 0 & f(\lambda)\end{bmatrix}$$

Thm: [Higham book Th. 4.7]

Suppose $f(x) = \sum_{k=0}^{\infty} p_k (x-\alpha)^k \qquad p_k = \frac{f^{(k)}(\alpha)}{k!}$

Taylor series centered in α, with convergence radius r

(converges inside the open ball $\{z \in \mathbb{C} : |z-\alpha| < r\}$)

Then $\lim_{d \to \infty} \sum_{k=0}^{d} p_k (A - \alpha I)^k = f(A)$

for each matrix with eigenvalues s.t. $|\lambda_i - \alpha| < r$

Proof: $\frac{1}{r} = \limsup (p_k)^{1/k}$

Let us call $g_d(x) = \sum_{k=0}^{d} p_k (x-\alpha)^k$

$$g_d^{(h)} = \sum_{k=0}^{d} p_k \, k(k-1)\cdots(k-h+1)(x-\alpha)^{k-h} = \sum_{k=0}^{d} \frac{f^{(k)}(\alpha)}{k!} \cdot k(k-1)\cdots(k-h+1)(x-\alpha)^{k-h}$$

$$= \sum_{k=0}^{d} \frac{f^{(k)}(\alpha)}{(k-h)!}(x-\alpha)^{k-h} = \text{Taylor polynomials of } f^{(h)}(x)$$

with the same radius of convergence: indeed, the first deriv. series

$f'(x) = \sum p_{k-1} k (x-\alpha)^{k-1}$ has radius of conv. given by

$$\limsup \left(k p_{k-1}\right)^{1/k} = \limsup \left(p_k\right)^{1/k}$$

$$\left(\lim k^{1/k} = 1\right)$$

$\Rightarrow$ Taylor polynomials $g_d(x) \to f(x)$

$$g_d'(x) \to f'(x)$$
$$\vdots$$
$$g_d^{(k)}(x) \to f^{(k)}(x) \quad \forall k$$

$\Rightarrow$ $g_d(A) \to f(A)$ when $d \to \infty$

Why can't we use Taylor series for everything?

Problem 1: not everything has a Taylor series, e.g. $f(x) = \sqrt{x}$ in $x = 0$

Problem 2: intermediate growth of coefficients

$f(x) = \exp(x)$

$$A = \begin{bmatrix} 0 & 30 \\ -30 & 0 \end{bmatrix}$$

Intermediate growth: $\frac{A^{30}}{30!} \approx 10^{11}$

Same problem for scalars:

$$\exp(50) = 1 + 50 + \frac{50^2}{2} + \frac{50^3}{3!} + \cdots$$

terms are large, but $\exp(50)$ is even larger

$$\exp(-s\sigma) = 1 - s\sigma + \frac{s\sigma^2}{2} - \frac{s\sigma^3}{3!} + \cdots$$

terms are larger, $\exp(-s\sigma)$ is tiny

$\Rightarrow$ very large (relative) error

For scalars, one can solve the issue by computing

$$\exp(-s\sigma) = \frac{1}{\exp(s\sigma)}$$

(no cancellation any more)

For matrices, you compute the series on "all eigenvalues at the same time" and these tricks do not work.