$$J = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \lambda & 1 \\ & & & \lambda \end{bmatrix} \qquad f(J) = \begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{1}{2}f''(\lambda) & \frac{1}{6}f'''(\lambda) \\ & \ddots & \ddots & \ddots \\ & 0 & & \end{bmatrix}$$

What is the Jordan structure of $f(J)$ when $f'(\lambda)=0 \quad f''(\lambda)\neq 0$?

We can tell the Jordan structure from ranks

$$rk\left(f(J)-f(\lambda)I\right)^k \qquad f(J)-f(\lambda)I = \begin{bmatrix} 0 & 0 & * & * \\ & 0 & 0 & * \\ & & 0 & 0 \\ & & & 0 \end{bmatrix} \; :$$

$$rk\left(f(J)-f(\lambda)I\right) = 2$$
$$rk\left(\left(f(J)-f(\lambda)I\right)^2\right) = 0$$

These ranks determine the Jordan form of $f(J)$:

$$\begin{bmatrix} \boxed{\begin{matrix} f(\lambda) & 1 \\ & f(\lambda) \end{matrix}} & \\ & \boxed{\begin{matrix} f(\lambda) & 1 \\ & f(\lambda) \end{matrix}} \end{bmatrix}$$

General result: if multiplicity is 2 $(f'(\lambda)=0,\ f''(\lambda)\neq 0)$, then you get two blocks of dimension $\frac{k}{2}, \frac{k}{2}$ $\left(\frac{k-1}{2}, \frac{k+1}{2}\right)$,

for multiplicity d, $(f'(\lambda)=\dots=f^{(d)}(\lambda)=0,\ f^{(d+1)}(\lambda)\neq 0)$ you get d blocks of size $\left\lfloor \frac{k}{d} \right\rfloor, \left\lceil \frac{k}{d} \right\rceil$.

<u>Continuity properties</u>: If $f_n \to f$, is it the case that $f_n(A) \to f(A)$?

Recall that $f(A) = V \operatorname{diag}(f(J_1), \dots f(J_s))V^{-1}$,

$$f_n(J_\lambda) = \begin{bmatrix} f_n(\lambda) & f_n'(\lambda) & \cdots & \frac{1}{(k-1)!} f_n^{(k-1)}(\lambda) \\ & \ddots & & \\ & & & f_n(\lambda) \end{bmatrix}$$

If $f_n(\lambda_i) \to f(\lambda_i)$, and

$f_n^{(j)}(\lambda_i) \to f^{(j)}(\lambda_i)$ for each $j < k_i$,

then $f_n(A) \to f(A)$ $\to$ follows from the definitions.

This allows us to compute matrix functions as limits of Taylor series: e.g.

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

Theorem: let $A \in \mathbb{C}^{n \times n}$, f defined via Taylor series

$$f(z) = \sum_{k=0}^{\infty} c_k (z-\alpha)^k, \quad \text{with convergence radius } r > 0.$$

Then,

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

assuming that all eigenvalues of A satisfy $|\lambda - \alpha| < r$

Proof:

$$f(A) = V \operatorname{diag}(f(J_{\lambda_1}), \dots f(J_{\lambda_s})) V^{-1}$$

$$p_d(A) = V \operatorname{diag}(p_d(J_{\lambda_1}), \dots p_d(J_{\lambda_s})) V^{-1}$$

so we can prove the result on a single Jordan block.

$$p_d(J_\lambda) = \begin{bmatrix} p_d(\lambda) & p_d'(\lambda) & \cdots & \frac{1}{(k-1)!} p^{(k-1)}(\lambda) \\ & \ddots & \ddots & \vdots \\ 0 & & & \vdots \end{bmatrix}$$

On the diagonal, $P_d(\lambda) \to f(\lambda)$ when $d \to \infty$

$$P_d'(\lambda) = \left(\sum_{k=0}^{d} c_k(\lambda-\alpha)^k\right)' = \sum_{k=1}^{d} c_k k(\lambda-\alpha)^{k-1}$$

This is the Taylor series of $f'(z)$:

$$f'(z) = \sum_{k=1}^{\infty} c_k\, k(\lambda-\alpha)^{k-1}$$

truncated to degree $d-1$

This Taylor series has the same conv. radius r as the one for f, so $\lim_{d\to\infty} P_d'(\lambda) = f'(\lambda)$

The same holds for all superdiagonals:

the i-th superdiagonal contains $\frac{1}{i!} P_d^{(i)}(\lambda)$,

this is the Taylor series for $\frac{1}{i!} f^{(i)}(z)$ truncated at degree $d-i$, and this has convergence radius r. $\square$

Cauchy integral formula: $\frac{1}{2\pi i}\int_\Gamma \frac{f(z)}{z-\lambda}\,dz = f(\lambda)$

if $\lambda \in \operatorname{Int}\Gamma$

Theorem: for any $A \in \mathbb{C}^{n\times n}$, f holomorphic, Γ such that $\Lambda(A) \in \operatorname{Int}\Gamma$,

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

(this could also be used to give an alternative definition of $f(A)$!)

Proof: since $f(A) = V \operatorname{diag}(f(J_1), \dots f(J_s))V^{-1}$, and

$$(zI-A)^{-1} = V \operatorname{diag}\left((zI-J_1)^{-1}, \dots (zI-J_s)^{-1}\right)V^{-1}$$

We can reduce to proving the result on a Jordan block.

$$\frac{1}{2\pi i}\int_\Gamma f(z)\left(z1 - J\right)^{-1} dz = \frac{1}{2\pi i}\int_\Gamma f(z)\begin{bmatrix} z-\lambda & -1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & -1 \\ 0 & & & z-\lambda \end{bmatrix}^{-1} dz$$

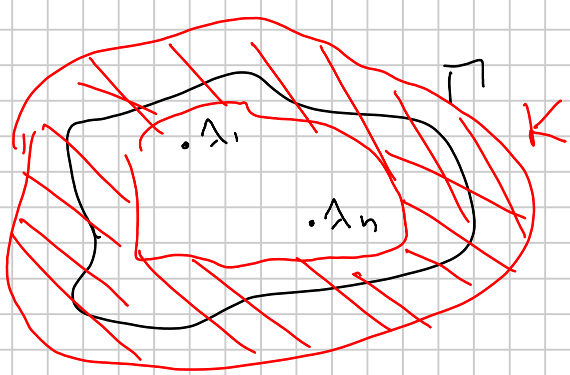
$$= \frac{1}{2\pi i}\int_\Gamma f(z)\begin{bmatrix} \frac{1}{z-\lambda} & \frac{1}{(z-\lambda)^2} & \cdots & \frac{1}{(z-\lambda)^{k-1}} \\ & \ddots & \ddots & \vdots \\ & & \ddots & \\ 0 & & & \frac{1}{z-\lambda} \end{bmatrix} dz$$

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

By the "derivative version" of the Cauchy integral formula, this is equal to $f(J)$ $\square$

__Corollary__: if f holomorphic, and $A_n \to A$, then $f(A_n) \to f(A)$.

__Proof__: $f(A_n) = \frac{1}{2\pi i}\int_\Gamma f(z)(z1 - A_n)^{-1} dz$



Let us take K s.t. $\Gamma \subset K \subset \mathbb{C} \setminus \Lambda(A)$

On K,

$$\max_{z\in K} \|(z1-A)^{-1}\| = M < \infty$$

Thanks to convergence + compactness

$$\max_{z \in K} \| (zI - A_n)^{-1} \| \leq M + \varepsilon \qquad \text{for } n \text{ suff. large,}$$

All functions are bounded + converge point-by-point $\Rightarrow$ the integral converges

$$\int_\Gamma f(z)\,(zI - A_n)^{-1}\,dz \to \int_\Gamma f(z)\,(zI - A)^{-1}\,dz \qquad \square$$

Note: one can also prove $f(A_n) \to f(A)$ under weaker hypotheses on f: $\Lambda(A) \subseteq \mathbb{R}$, $f \in C^{K_{max}}$, where all Jordan blocks of A have dimension $\leq K_{max}$:

take p interpolation polynomial for A s.t. $p(A) = f(A)$, and p_n such that $p_n(A_n) = f(A_n)$.

The difficult part is proving that $p_n \to p$ coefficient-by-coefficient. This follows from finite difference formulas for interpolation.

$$\begin{bmatrix} 0 & 1 \\ \frac{1}{n} & 0 \end{bmatrix} \to \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$p(\lambda_1) = f(\lambda_1)$
$p(\lambda_2) = f(\lambda_2)$

$p(0) = f(0)$
$p'(0) = f'(0)$

different Vandermonde-like matrices!

$$\begin{bmatrix} p_0 \\ p_1 \end{bmatrix} = \begin{bmatrix} 1 & \lambda_1 \\ 1 & \lambda_2 \end{bmatrix}^{-1} \begin{bmatrix} f(\lambda_1) \\ f(\lambda_2) \end{bmatrix} \qquad V \to \begin{bmatrix} 1 & \lambda \\ 1 & \lambda \end{bmatrix} \text{ singular!}$$

$$\begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \lambda_1^3 \\ 0 & 1 & 2\lambda_1 & 3\lambda_1^2 \\ 1 & \lambda_2 & \lambda_2^2 & \lambda_2^3 \\ 0 & 1 & 2\lambda_2 & 3\lambda_2^2 \end{bmatrix} \to \text{singular!}$$

Derivatives of functions of matrices?

Def: The FRÉCHET derivative of a matrix function f in A is the linear operator (unique) (when it exists) such that

$$f(A+H) = f(A) + L_{f,A}[H] + o(\|H\|)$$

$L_{f,A} : \mathbb{C}^{n\times n} \mapsto \mathbb{C}^{n\times n}$ linear operator

Ex: $f(x) = x^2 \qquad f(A) = A^2$

$$f(A+H) = (A+H)^2 = \underbrace{A^2}_{f(A)} + \underbrace{HA+AH}_{L_{f,A}[H]} + \underbrace{H^2}_{o(\|H\|)}$$

The Fréchet derivative $L_{f,A}$ is the linear operator that maps

$$H \mapsto L_{f,A}[H] = HA + AH$$

We can write a $n^2 \times n^2$ matrix associated to $L_{f,A}$ by using vectorization: we are looking for the matrix $K \in \mathbb{C}^{n^2\times n^2}$ s.t.

$$K \cdot \mathrm{vec}(H) = \mathrm{vec}(HA+AH)$$

$$K = I_n \otimes A + A^T \otimes I_n$$ (we have seen this when studying Sylvester equations)

(Note that K is the Jacobian matrix of the map

$$F: \operatorname{vec}(X) \mapsto \operatorname{vec}(X^2) \qquad \mathbb{C}^{n^2} \to \mathbb{C}^{n^2}$$

Ex: $f(z) = \frac{1}{z}$ $\qquad f(A) = A^{-1}$ because $f(z)z = 1$

$\Rightarrow f(A)A = I$

$$f(A+H) = (A+H)^{-1} = \left(A(I+A^{-1}H)\right)^{-1}$$

$$(1-x)^{-1} = 1+x+x^2+\dots$$

$$= (I+A^{-1}H)^{-1}A^{-1}$$

$$= (I - A^{-1}H + A^{-1}HA^{-1}H - A^{-1}HA^{-1}HA^{-1}H + \dots)A^{-1}$$

$$= \underbrace{A^{-1}}_{f(A)} - \underbrace{A^{-1}HA^{-1}}_{L_{f,A}[H]} + \underbrace{A^{-1}HA^{-1}HA - \dots}_{o(\|H\|)}$$

$$L_{f,A}[H] \to -A^{-1}HA^{-1}$$

With vectorization, the associated $n^2 \times n^2$ matrix is

$$K = -(A^{-1})^T \otimes A^{-1} \qquad \operatorname{vec}(AXB) = (B^T \otimes A)\operatorname{vec} X$$

$$K \cdot \operatorname{vec} H = \operatorname{vec}(-A^{-1}HA^{-1}).$$

Properties: those of $n^2 \times n^2$ Jacobians:

- $L_{f+g,A} = L_{f,A} + L_{g,A}$
- $L_{f\circ g,A} = L_{f,g(A)} \circ L_{g,A}$

composition of linear operators, or product of matrices:

$$K_{f,g(A)} \cdot K_{g,A}$$

- $L_{f^{-1},A} = \left(L_{f,f^{-1}(A)}\right)^{-1}$ ~~a~~ inverse of linear operator/ inverse of $n^2 \times n^2$ matrix

EXAMPLE: Derivative of $A^{\frac{1}{2}}$:

$$f(x) = x^2 = y$$
$$g(y) = \sqrt{y} = x$$

(any branch of the square root)

$$L_{f,A}[H] = AH + HA \qquad f(A) = B \quad g(B) = A$$

$L_{g,B}$ is the inverse of this operator:

$X = L_{g,B}[E]$ is the matrix X such that

$$L_{f,A}[X] = E$$
$$AX + XA = E$$

X is the solution of the Sylvester equation $AX + XA = E$.

Given A, we can solve for X and compute the derivative.

$$A = g(B) = B^{\frac{1}{2}}$$

$L_{g,B}[E]$ is the solution of the Sylvester equation $B^{\frac{1}{2}}X + XB^{\frac{1}{2}} = E$

The Sylvester eq. has a unique solution if and only if

$$\Lambda(B^{\frac{1}{2}}) \cap \Lambda(-B^{\frac{1}{2}}) = \emptyset$$

If we take the principal square root, i.e. the one such that $\sqrt{x} \in RHP$ (defined for all x apart from the negative reals)

With this choice,

$\Lambda(B^{\frac{1}{2}}) \subset RHP$ $\Lambda(-B^{\frac{1}{2}}) \subset LHP$, so they never overlap. In fact, this condition is satisfied for ~~any~~ branch of the square root.

Indeed, if $\Lambda(B^{\frac{1}{2}}) \cap \Lambda(-B^{\frac{1}{2}}) \neq \emptyset$, then for some λ both λ and $-\lambda$ are eigenvalues of $B^{\frac{1}{2}}$

$$\lambda = g(\text{something}) = \sqrt{\text{something}}.$$

$$-\lambda = g(\text{something}) = \sqrt{\text{something}}.$$

But when we choose a branch for g, we have to choose either $g(\lambda^2) = \lambda$ or $g(\lambda^2) = -\lambda$.