Computing matrix functions

Note Title 2025-03-27

Last time:

$$A = \begin{bmatrix} 3 & -1 \\ 1 & 1+\varepsilon \end{bmatrix}$$

- $\kappa(V)$ large
- λ_1, λ_2 close

Inaccurate computation of $f(A)$ via diagonalization

Accurate via Taylor expansion

Now:

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

- $\kappa(V) = 1$
- $\lambda_1, \lambda_2 = \pm 30i$

$$\exp(A) = \begin{bmatrix} \cos(30) & \sin(30) \\ -\sin(30) & \cos(30) \end{bmatrix} \qquad \text{(radians!)}$$

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

$$\frac{1}{(k+1)!}A^{k+1} = \frac{1}{k+1} A \left(\frac{1}{k!}A^k\right)$$

Intermediate terms $\approx 10^{12}$ error $\approx 10^{12} \cdot 10^{-16} \approx 10^{-4}$.

Computing matrix functions with the Schur form:

$$A = QUQ^* \qquad U = \begin{bmatrix} t_{11} & t_{12} \\ 0 & t_{22} \end{bmatrix}$$

$f(A) = Q f(U) Q^*$ How do we find $f(U)$?

$f(U) = p(U)$, so it must be upper triangular

$$f(U) = p(U) = \begin{bmatrix} s_{11} & s_{12} \\ 0 & s_{22} \end{bmatrix} \qquad \begin{aligned} s_{11} &= p(u_{11}) = f(u_{11}) \\ s_{22} &= p(u_{22}) = f(u_{22}) \end{aligned}$$

To compute s_{12}, we use the fact that $f(U)U = Uf(U)$.

$$\begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix}\begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix} = \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix}\begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix}$$

(1,2) entry: $\quad S_{11}U_{12} + S_{12}U_{22} = U_{11}S_{12} + U_{12}S_{22}$

$$S_{12}U_{22} - U_{11}S_{12} = U_{12}S_{22} - U_{12}S_{11} \Rightarrow S_{12} = \frac{U_{12}(S_{22}-S_{11})}{U_{22}-U_{11}}$$

$$= U_{12}\,\frac{f(U_{22})-f(U_{11})}{U_{22}-U_{11}}$$

Also in larger dimension:

$$\begin{bmatrix} S_{11} & \cdots & S_{1n} \\ & \ddots & \vdots \\ 0 & & S_{nn} \end{bmatrix}\begin{bmatrix} U_{11} & \cdots & U_{1n} \\ & \ddots & \vdots \\ 0 & & U_{nn} \end{bmatrix} = \begin{bmatrix} U_{11} & \cdots & U_{1n} \\ & \ddots & \vdots \\ 0 & & U_{nn} \end{bmatrix}\begin{bmatrix} S_{11} & \cdots & S_{1n} \\ & \ddots & \vdots \\ 0 & & S_{nn} \end{bmatrix}$$

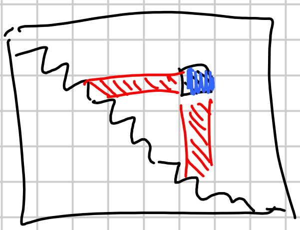
(i,j) entry: $\quad i<j$

$$\sum_{k=i}^{j} S_{ik}U_{kj} = \sum_{k=i}^{j} U_{ik}S_{kj}$$

$$S_{ij}U_{jj} - U_{ii}S_{ij} = \sum_{k=i+1}^{j} U_{ik}S_{kj} - \sum_{k=i}^{j-1} S_{ik}U_{kj}$$

$$\boxed{S_{ij}} = \frac{\boxed{\sum_{k>i} U_{ik}S_{kj} - \sum_{k<j} S_{ik}U_{kj}}}{U_{jj}-U_{ii}}$$

We can compute S_{ij} from S_{kj} for $k\in\{i+1,\dots j\}$ and S_{ik} for $k\in\{i,\dots j-1\}$



Same recurrence as the Bartels-Stewart algorithm to solve Sylvester equations.

Parlett recurrence $\qquad$ Cost: $O(n^3)$.

Matlab implementation: snober $\qquad S = \begin{bmatrix} 1 & 3 & 6 & 10 \\ & 2 & 5 & 9 \\ & & 4 & 8 \\ & & & 7 \end{bmatrix}$

Problem: close-by or equal eigenvalues

Solution: work blockwise.

$$U = \begin{bmatrix} 1 & \times & \times & \times & \times \\ & 1.01 & \times & \times & \times \\ & & 2.01 & \times & \times \\ & & & 1.99 & \times \\ & & & & 3 \end{bmatrix}$$

$$S = f(U)$$

- compute the Schur form
- reorder the Schur form so that close-by eigenvalues are consecutive
- compute $f(U_{ii})$ on each triangular block U_{ii} using Taylor series
- use a block version of the Parlett recurrence to compute blocks S_{ij} with $i \neq j$
- $f(A) = Q S Q^*$

Block Parlett recurrence:

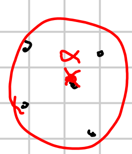
$$S_{ij} U_{jj} - U_{ii} S_{ij} = \boxed{\sum_{k>i} U_{ik} S_{kj} - \sum_{k<j} S_{ik} U_{kj}}$$

computed at the previous steps

It is a Sylvester equation with triangular coefficients U_{jj}, U_{ii}.

How to split eigenvalues into clusters?
Many edge cases!

How to choose the Taylor series for the diagonal blocks?
Center?



Best radius of convergence if we choose $\alpha = \dfrac{\lambda_1 + \dots + \lambda_k}{k}$,
where $(\lambda_1, \dots \lambda_k) = \Lambda(U_{ii})$ $= \dfrac{\mathrm{Tr}(U_{ii})}{k}$

How many terms to choose?

Stop when $(U_{ii}-\alpha I)^d$ is small enough

How to compute the coefficients of the Taylor series?

$$C_d = \frac{f^{(d)}(\alpha)}{d!}$$

We need to be able to compute all derivatives of f.

State of the art method, despite few accuracy guarantees.

- clustering might just return a single cluster,
- Taylor series might be slow to converge $O(n^3 d)$
- The Sylvester equations might be ill-conditioned:

$$\operatorname{sep}(U_{ii}, U_{jj}) \leq \min\{|\lambda - \mu| : \lambda \in \Lambda(U_{ii}), \mu \in \Lambda(U_{jj})\}$$

is only an inequality: for non-normal matrices, the separation can be small even if U_{ii}, U_{jj} have no close eigenvalues.