Note Title 2020-03-19

Dalla volta scorsa:

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

⚠ se f olomorfa

+ usarlo per dimostrare che $A_n \to A \quad f(A_n) \to f(A)$

Se f non è olomorfa, sketch dimostrazione:

$f(A) = p(A)$ $\quad p$ interpol. polynomial in spectrum of A

$f(A_n) = p_n(A_n)$ $\quad p_n$ " " " " " A_n

$\boxed{\text{fact}}$: coefficients of interp. polynomials are continuous in the nodes, that is $p_n \to p$

(not obvious from our proof)

For n large enough $\|A_n - A\| \le \varepsilon \quad \|p_n - p\| \le \varepsilon$

$$\|f(A) - f(A_n)\| = \|p(A) - p_n(A_n)\| \le \underbrace{\|p(A) - p_n(A)\|}_{\text{bounded}} + \underbrace{\|p_n(A) - p_n(A_n)\|}_{\text{bounded}}$$

Back to methods to compute general matrix functions

$A = VDV^{-1}$ may not work (error $\approx \kappa(V)$)

Better: use Schur form. Can we use $A = QTQ^*$ to compute matrix functions?

Example:

$$A = \begin{bmatrix} t_{11} & t_{12} \\ 0 & t_{22} \end{bmatrix} \qquad f(A) = \begin{bmatrix} s_{11} & s_{12} \\ 0 & s_{22} \end{bmatrix}$$

$f(A)$ is a polynomial in A, so it is also upper triang.

$s_{11} = f(t_{11})$, $s_{22} = f(t_{22})$

(again, because this holds for $p(A) = f(A)$)

Trick: to compute s_{12}, use $A f(A) = f(A) A$

$$\begin{bmatrix} t_{11} & t_{12} \\ 0 & t_{22} \end{bmatrix}\begin{bmatrix} s_{11} & s_{12} \\ 0 & s_{22} \end{bmatrix} = \begin{bmatrix} t_{11}s_{11} & t_{11}s_{12} + t_{12}s_{22} \\ 0 & t_{22}s_{22} \end{bmatrix} \overset{!}{=} \begin{bmatrix} s_{11}t_{11} & s_{11}t_{12} + s_{12}t_{22} \\ 0 & s_{22}t_{22} \end{bmatrix}$$

$t_{11}s_{12}+t_{12}s_{22}=s_{11}t_{12}+s_{12}t_{22}$ $\qquad s_{12}=\dfrac{t_{12}(s_{11}-s_{22})}{t_{11}-t_{22}}$

if $t_{11}\neq t_{22}$.

If $t_{11}\to t_{22}$, then $s_{12}=t_{12}\dfrac{f(t_{11})-f(t_{22})}{t_{11}-t_{22}}\to t_{12}\cdot f'(t_{11})$

so by continuity $f\left(\begin{bmatrix} t_{11} & t_{12} \\ 0 & t_{11}\end{bmatrix}\right)=\begin{bmatrix} f(t_{11}) & t_{12}f'(t_{12}) \\ 0 & f(t_{11})\end{bmatrix}$

The same idea works for larger blocks.

$$A=\begin{bmatrix} t_{11} & t_{12} & t_{13} \\ 0 & t_{22} & t_{23} \\ 0 & 0 & t_{33}\end{bmatrix} \qquad f(A)=\begin{bmatrix} s_{11} & s_{12} & \boxed{s_{13}} \\ 0 & s_{22} & s_{23} \\ 0 & 0 & s_{33}\end{bmatrix}$$

From entry $(1,3)$ of $Af(A)=f(A)A$, one gets

$$t_{11}s_{13}+t_{12}s_{23}+t_{13}s_{33}=s_{11}t_{13}+s_{12}t_{23}+s_{13}t_{33} \quad \Rightarrow$$

$$\boxed{s_{13}}(t_{11}-t_{33})=s_{11}t_{13}+s_{12}t_{23}-t_{12}s_{23}-t_{13}s_{33}$$

If one has computed entries <u>below</u> and <u>to the left</u> of s_{ij}, one can compute s_{ij} as well.

We can compute entries recursively like in Sylv. equations (if eigenvalues are different)

$$f(T)=\begin{bmatrix} \boxed{s_{11}} & \boxed{s_{12}} & \boxed{s_{13}} & \cdots & s_{1n} \\ & \boxed{s_{22}} & \boxed{s_{23}} & & \\ & & \ddots & \ddots & \boxed{s_{n-2\,n}} \\ & & & \ddots & \boxed{s_{n-1\,n}} \\ & & & & \boxed{s_{nn}}\end{bmatrix}$$

$s_{ii}=f(t_{ii})$

$s_{i,i+1}=t_{i,i+1}\dfrac{s_{i+1,i+1}-s_{ii}}{t_{i+1,i+1}-t_{ii}}$

$s_{i,i+2}=\ldots$

can be computed recursively one superdiag. at a time

The same idea works blockwise: (Parlett recurrence) if diagonal blocks are square,

$T_{11}S_{12}-S_{12}T_{22}=S_{11}T_{12}-T_{12}S_{22}$ $\Leftarrow$ Sylvester equation for S_{12} (if one has already computed S_{11}, S_{22})

$T_{11}S_{13}-S_{13}T_{33}=$ other entries $\Leftarrow$ Sylvester equation

The Sylv. equation for S_{ij} is solvable if $\Lambda(T_{ii})\cap\Lambda(T_{jj})=\emptyset$

If $A = QTQ^*$, then $f(A) = p(A) = p(QTQ^*)$

$$= Q\,p(T)\,Q^*$$

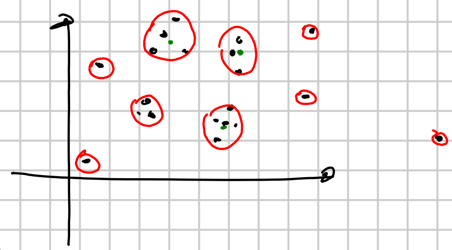
$$= Q\,f(T)\,Q^*$$

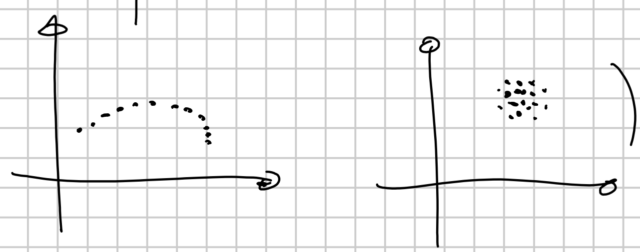
<u>Algorithm</u> (Schur-Parlett algorithm):

1) $A = QTQ^*$ (Schur fact.)

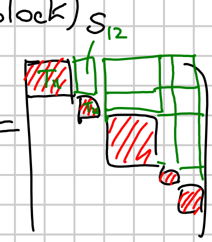
2) Divide T into blocks (reordering the Schur form if necessary)

such that:
- all eigenvalues in the same diagonal block T_{ii} are close to each other
- all eigenvalues in distinct diagonal blocks are "well separated"



(not clear what it means for )

3) Compute $\boxed{f(T_{ii})}$ via Taylor expansion
(in the centroid of the eigenvalues of each block)

$f(T) =$ 

4) Compute off-diagonal blocks with Parlett recurrence (blockwise) e.g.

$$T_{11}S_{12} - S_{12}T_{22} = T_{12}S_{22} - S_{11}T_{12} \quad \text{(Sylv. equation)}$$

5) Return $f(A) = Q\,f(T)\,Q^*$

If the eigenvalues aren't easy to divide into clusters, numerical problems. (Ill-cond. Sylvester equations, or slow Taylor expansion, depending on clustering thresholds)

<u>Computational cost</u>: if all blocks are $O(1)$, one just

need to solve Sylvester equations.

Sum of $2(j-i)$ terms

$$\boxed{T_{ii}} S_{ij} - S_{ij} \boxed{T_{jj}} = \underbrace{-T_{i\,i+1} S_{i+1\,j} - \dots - T_{ij} S_{jj} + S_{ii} T_{ij} + S_{i\,i+1} T_{i+1\,j} + \dots + S_{i\,j-1} T_{j-1\,j}}$$

already triangular! $\quad \mathcal{O}(\text{block size}^3)$

$\mathcal{O}(n^3)$ (putting everything together) for the Sylvester solutions.

If you have large blocks, Taylor expansions might be expensive:

$$f(T_{ii}) = a_0 I + a_1 T_{ii} + a_2 T_{ii}^2 + \dots + a_d T_{ii}^d \qquad \mathcal{O}(d n_i^3)$$

$$a_d = \frac{f^{(d)}(x_0)}{d!}$$

Relation between Schur-Parlett and block diagonalization

$$T = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}$$ can be reduced to block diag. form by a Sylvester equation

$$\underbrace{\begin{bmatrix} I & -X \\ 0 & I \end{bmatrix}}_{W^{-1}} \underbrace{\begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}}_{T} \underbrace{\begin{bmatrix} I & X \\ 0 & I \end{bmatrix}}_{W} = \underbrace{\begin{bmatrix} T_{11} & \boxed{T_{11}X - XT_{22} + T_{12}} \\ 0 & T_{22} \end{bmatrix}}_{D}$$

$= 0$

If X solves $\underline{T_{11}X - XT_{22} = -T_{12}}$, then RHS is block diagonal.

$$T = WDW^{-1}$$

$$f(T) = Wf(D)W^{-1} = W \begin{bmatrix} f(T_{11}) & 0 \\ 0 & f(T_{22}) \end{bmatrix} W^{-1}$$

$$= \begin{bmatrix} I & X \\ 0 & I \end{bmatrix} \begin{bmatrix} f(T_{11}) & 0 \\ 0 & f(T_{22}) \end{bmatrix} \begin{bmatrix} I & -X \\ 0 & I \end{bmatrix} = \begin{bmatrix} f(T_{11}) & \boxed{Xf(T_{22}) - f(T_{11})X} \\ 0 & f(T_{22}) \end{bmatrix}$$

$=: S_{12}$

cfr. with Schur-Parlett: get S_{12} by solving the Sylv. equation

$$\underline{T_{11} S_{12} - S_{12} T_{22}} = -T_{12} f(T_{22}) + f(T_{11}) T_{12} \qquad (*)$$

(one can check that if X solves $T_{11}X - XT_{22} = -T_{12}$ then $S_{12} = Xf(T_{22}) - f(T_{11})X$ solves $(*)$:

$$T_{11}\left(Xf(T_{22}) - f(T_{11})X\right) - \left(Xf(T_{22}) - f(T_{11})X\right)T_{22} \overset{?}{=} T_{12}f(T_{22}) - f(T_{11})T_{12}$$

$$\to = \underline{T_{11}X} f(T_{22}) - f(T_{11})\underline{T_{11}X} - \underline{XT_{22}} f(T_{22}) + f(T_{11})\underline{XT_{22}}$$

$= -T_{12} f(T_{22}) + f(T_{11}) T_{12}$ ✓)

How do compute derivatives on a computer?

1) Symbolically (Mathematica, Wolfram Alpha, ...: take an expression for your function, and compute derivatives explicitly)

One needs to be careful with "exploding number of terms": you do not want to expand $(1+x)^{1000}$

2) $$f'(x) \approx \frac{f(x+h)-f(x)}{h} + O(h)$$

Pitfalls: h too large => not close to the derivative

h too small => subtractive cancellation in the numerator

"optimal" value of h: consider both sources of error:

approximation error in the numerator: $O(f(x)\cdot u)$

Taylor truncation error:

$$f(x+h) = f(x) + h f'(x) + O(h^2) \Rightarrow f'(x) = \frac{f(x+h)-f(x)+O(h^2)}{h}$$
$$= \frac{f(x+h)-f(x)}{h} + O(h)$$

Total error: $O\left(\frac{f(x)\,u}{h} + h\right)$

If $f(x) = O(1)$, optimal value $= h = u^{1/2}$, error $O(u^{1/2})$

"numerical differentiation".

3) Automatic differentiation techniques

Ex: $f(x) = \dfrac{x^2+5}{1+\exp(x)}$

Write "matrix-friendly code" to compute it:

```
n = size(X0);
f = (X0^2 + 5*eye(n))*inv(eye(n)+expm(X0))
```

If we compute $f\left(\begin{bmatrix} x_0 & 1 & 0 \\ 0 & x_0 & 1 \\ 0 & 0 & x_0 \end{bmatrix}\right) =$

$$= \begin{bmatrix} f(x_0) & f'(x_0) & \frac{f''}{2}(x_0) \\ 0 & f(x_0) & f'(x_0) \\ 0 & 0 & f(x_0) \end{bmatrix}$$

This is <u>not</u> numerical differentiation.