Note Title 2020-03-26

Padé approximants to $\exp(x)$

$$\exp(x) = \frac{N(x)}{D(x)} + O(x^{p+q+1})$$

$N(x)$ of degree p

$D(x)$ poly. of degree q

Ex: $p = q = 2$

$$\exp(x) = \frac{ax^2+bx+c}{x^2+dx+e} + O(x^{p+q+1})$$

$$\updownarrow$$

$$D(x)\exp(x) = N(x) + O(x^{p+q+1}) \quad \text{linear in } a,b,c,d,e$$

$$(x^2+dx+e)\left(1+x+\frac{x^2}{2}+\frac{x^3}{3!}+\frac{x^4}{4!}+O(x^5)\right) - ax^2-bx-c = O(x^5)$$

$$\underbrace{e-c}_{=0} + \underbrace{(e+d-b)}_{=0}x + \underbrace{\left(1+d+\frac{e}{2}-a\right)}_{=0}x^2 + \dots \underbrace{(\quad)}_{=0}x^3 + \underbrace{(\quad)}_{=0}x^4 + O(x^5) = 0$$

Solution: $N = x^2+6x+12$ $\qquad \frac{N}{D} = \frac{x^2+6x+12}{x^2-6x+12}$

$D = x^2-6x+12$

Backward error of Padé approximants:

$$f(x) = \log\left(e^{-x}\frac{N(x)}{D(x)}\right) = O(x^{p+q+1})$$

$$e^{-x}\frac{N(x)}{D(x)} = 1 + O(x^{p+q+1})$$

$$e^{f(x)} = e^{-x}\frac{N(x)}{D(x)} \iff e^{x+f(x)} = \frac{N(x)}{D(x)}$$

"backward error" setting: $\frac{N(x)}{D(x)} = e^{\tilde{x}}$ $\qquad \tilde{x} = x + f(x)$

Let us apply everything to a matrix A: $H = f(A)$

$AH = HA$ because H is a function of A

$$e^H = e^{-A}N(A)D(A)^{-1} \iff e^{A+H} = e^A e^H = N(A)D(A)^{-1}$$

If $\|H\| = O(u)\|A\|$ then $N(A)D(A)^{-1}$ approximates e^A with an error that is comparable to intrinsic error (error induced by the approximation $A \approx fl(A)$) $\qquad u \approx 10^{-16}$

$\tilde{A} \quad fl(A) \qquad \tilde{A}_{ij} = A_{ij}(1+\varepsilon_{ij}) \quad |\varepsilon_{ij}| \le u \approx 10^{-16}$

$\|\tilde{A} - A\| = \|A\| \cdot O(u)$

$\Rightarrow$ the algorithm is "as good as it could get" on a machine

(still does not ensure that it is accurate: e^{A+H} could be very far from e^A, if the problem is ill-conditioned)

$f(x) = c_1 x^{p+q+1} + c_2 x^{p+q+2} + c_3 x^{p+q+3} + \dots$ (Taylor expansion)

$H = f(A) = c_1 A^{p+q+1} + c_2 A^{p+q+2} + c_3 A^{p+q+3} + \dots$

$\|H\| = |c_1| \cdot \|A\|^{p+q+1} + |c_2| \cdot \|A\|^{p+q+2} + |c_3| \cdot \|A\|^{p+q+3} + \dots$

If $p = q = 13$ and $\|A\| \le 5.4$, then $\frac{\|H\|}{\|A\|} \le u \approx 10^{-16}$

[Higham book p. 244]

---

for $p = q$, $N(x) = r(x^2) + xs(x^2)$

$D(x) = r(x^2) - xs(x^2)$ for some polynomials r, s of degree 6

you can compute D, N with Paterson-Stockmeyer with 6 matrix multiplications

$r(x^2) = (r_0 + r_1 x^2 + r_2 x^4 + r_3 x^6) + x^2 (r_4 x^2 + r_5 x^4 + r_6 x^6)$

3 matmuls to form x^2, x^4, x^6

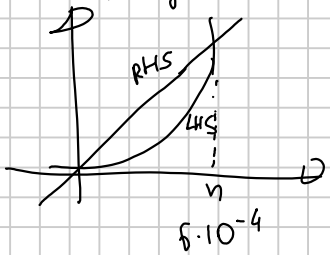
1 for $r(x^2)$ 1 for $xs(x^2)$

1 for $s(x^2)$

$\eta = \|A\|$

$\|H\| \le |c_5| \eta^5 + |c_6| \eta^6 + \dots + |c_{30}| \eta^{30} \le 10^{-16} \cdot \eta$

we need to find η s.t. this inequality holds



$\Rightarrow$ for $\|A\| \le 5.4$, the Padé approximant of degree $(13,13)$ has backward error smaller than machine precision

---

What happens if $\|A\| \ge 5.4$

Trick: $e^x = \left(e^{\frac{1}{2}x}\right)^2 \qquad \exp(A) = \left(\exp\left(\frac{1}{2}A\right)\right)^2$

Algorithm: 1. Find k s.t. $\left\|\frac{1}{2^k}A\right\| \le 5.4$

"Scaling and squaring"

2. Compute $\exp\left(\frac{1}{2^k}A\right) \approx N\left(\frac{1}{2^k}A\right) D\left(\frac{1}{2^k}A\right)^{-1} = F$ (Padé of degree $p = q = 13$)

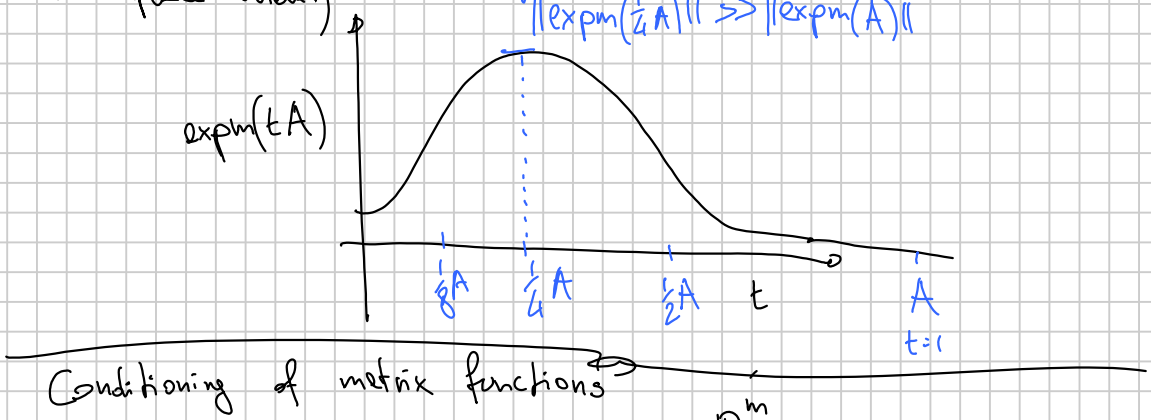
3. $\exp(A) = F^{2^k}$ by repeated squaring

(This is what Matlab's expm does)

(If $\|A\|$ is already much smaller than 5.4, it may use Padé expansions with degrees < 13)

(Other trick: if A is triangular, expm(A) is triangular, too. On its diagonal, we should have $\exp(A_{ii})$; we can compute them and replace them.)

It might happen that $\|\text{expm}(\frac{1}{4}A)\| > \|\text{expm}(A)\|$



Conditioning of matrix functions

Condition number of $y = f(x)$: $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$

$\Leftrightarrow$ how far away can $\tilde{y} = f(\tilde{x})$ be from y if $\tilde{x} - x$ is small?

$$\limsup_{\tilde{x} \to x} \frac{\|f(\tilde{x}) - f(x)\|}{\|\tilde{x} - x\|}$$

If f differentiable, $f(\tilde{x}) = f(x) + J_{f,x} \cdot (\tilde{x} - x) + o(\|\tilde{x} - x\|)$

Absolute c.n.

$$\limsup_{\|\tilde{x} - x\| \to 0} \frac{\|J_{f,x}(\tilde{x} - x)\|}{\|\tilde{x} - x\|} + o(1) = \|J_{f,x}\|$$

(matrix norm)

Relative cond. number

$$K_{rel}(f,x) = \frac{\frac{\|f(\tilde{x}) - f(x)\|}{\|f(x)\|}}{\frac{\|\tilde{x} - x\|}{\|x\|}} = K_{abs}(f,x) \cdot \frac{\|x\|}{\|f(x)\|} = \|J_{f,x}\| \cdot \frac{\|x\|}{\|f(x)\|}.$$

If $f(A)$ is a matrix function, what is its Jacobian as a function $\mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ ($J_{f,A} \in \mathbb{R}^{n^2 \times n^2}$)

Def: the **Fréchet derivative**:

The Fréchet derivative of a matrix function f is the linear operator $L_{f,X}: \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ such that

$$f(X+E) = f(X) + L_{f,X}(E) + o(\|E\|)$$

Example: $f(x) = x^2$ → matrix function $f(X) = X^2$

$$(X+E)^2 = \underbrace{X^2}_{f(X)} + \underbrace{XE+EX}_{\text{linear in } E} + \underbrace{E^2}_{o(\|E\|)}$$

The Fréchet derivative of f in X is the map $L_{f,X}$ s.t.

$$L_{f,X}(E) = XE+EX$$

(It is linear in E: $L_{f,X}(E+F) = L_{f,X}(E) + L_{f,X}(F)$)

$$L_{f,X}(\alpha E) = \alpha L_{f,X}(E)$$

$\mathbb{R}^{n\times n} \simeq \mathbb{R}^{n^2}$ $\quad \mathrm{vec}(X) \in \mathbb{R}^{n^2}$ different point of view:

$\hat{L}$ is the Jacobian of $\mathrm{vec}\, f(\mathrm{vec}^{-1}(x))$

$\hat{L}_{f,X}$ is the map $\mathrm{vec}\, E \to \mathrm{vec}\, L_{f,X}(E) = \mathrm{vec}(XE+EX)$

$$= \mathrm{vec}(XE) + \mathrm{vec}(EX)$$
$$= (I \otimes X)\,\mathrm{vec}\, E + (X^T \otimes I)\,\mathrm{vec}\, E$$
$$= (I \otimes X + X^T \otimes I)\,\mathrm{vec}\, E$$

(recall that $\mathrm{vec}(AXB) = (B^T \otimes A)\,\mathrm{vec}(X)$)

$$\hat{L}_{f,X} = I \otimes X + X^T \otimes I \in \mathbb{R}^{n^2 \times n^2}$$