MANCHESTER 1824

The University of Manchester

Computing the Fréchet Derivative of e^A with an Application to Condition Number Estimation

Nick Higham
School of Mathematics
The University of Manchester

higham@ma.man.ac.uk
http://www.ma.man.ac.uk/~higham/

Joint work with
Awad Al-Mohy

Householder Symposium XVII
Zeuthen, Germany, 2008

Fréchet Derivative

Fréchet derivative of $f : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ at $A \in \mathbb{C}^{n\times n}$

A linear mapping $L : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ s.t. for all $E \in \mathbb{C}^{n\times n}$

$$f(A+E) - f(A) - L(A,E) = o(\|E\|).$$

Special formula for the exponential:

$$L_{\exp}(A,E) = \int_0^1 e^{A(1-s)} E e^{As} \, ds.$$

Condition Number

$$\|L(A)\| := \max_{E \neq 0} \frac{\|L(A,E)\|}{\|E\|}.$$

$$\mathrm{cond}(f,A) := \lim_{\epsilon \to 0} \sup_{\|E\| \le \epsilon \|A\|} \frac{\|f(A+E) - f(A)\|}{\epsilon \|f(A)\|}.$$

Lemma

$$\mathrm{cond}(f,A) = \frac{\|L(A)\| \, \|A\|}{\|f(A)\|}.$$

Computational Framework: Outline

- Padé approximant $r_m(x) = p_m(x)/q_m(x)$ to f $[f(x) - r_m(x) = O(x^{2m+1})]$. Then $f(A) \approx r_m(A)$.
- $L_f(A, E) \approx L_{r_m}(A, E)$.

Fréchet differentiate initial transformations on A *and* efficient schemes for evaluating $p_m(A)$ and $q_m(A)$.

Lemma

The Fréchet derivative L_{r_m} of $r_m(x) = p_m(x)/q_m(x)$ satisfies

$$q_m(A) L_{r_m}(A, E) = L_{p_m}(A, E) - L_{q_m}(A, E) r_m(A).$$

Scaling and Squaring Algorithm for e^A

- $B \leftarrow A/2^s$ so $\|B\|_\infty \approx 1$
- $X = r_m(B)^{2^s} \approx e^A$

Differentiating $e^A = (e^{A/2})^2$ gives

$$\begin{aligned} L_{\exp}(A, E) &= L_{x^2}\big(e^{A/2}, L_{\exp}(A/2, E/2)\big) \\ &= e^{A/2} L_{\exp}(A/2, E/2) + L_{\exp}(A/2, E/2) e^{A/2}. \end{aligned}$$

Method:

$$\begin{aligned} X_s &= r_m(2^{-s}A), \\ L_s &= L_{r_m}(2^{-s}A, 2^{-s}E), \end{aligned}$$

$$\left.\begin{aligned} L_{i-1} &= X_i L_i + L_i X_i \\ X_{i-1} &= X_i^2 \end{aligned}\right\} \quad i = s\colon -1\colon 1.$$

Numerical Stability

Suppose $\|e^{-2^{-s}A} r_m(2^{-s}A) - I\| < 1$. $g_m(x) := \log(e^{-x} r_m(x))$.

Theorem (H, 2005)

$r_m(2^{-s}A)^{2^s} = e^{A+\Delta A}$, where $\Delta A = 2^s g_m(2^{-s}A)$ and

$$\frac{\|\Delta A\|}{\|A\|} \le \frac{-\log(1 - \|e^{-2^{-s}A} r_m(2^{-s}A) - I\|)}{\|2^{-s}A\|}.$$

Theorem

L_0 from the method satisfies

$$L_0 = L_{\exp}\big(A + \Delta A, E + L_{g_m}(2^{-s}A, E)\big).$$

Details

- Minor mods to choice of parameters s, m of H (2005) ensure $\|\Delta A\| \le u\|A\|$, $\|\Delta E\| \le u\|E\|$.
  Degree $m = 13$ still optimal.
- Differentiating poly evaluation schemes of H (2005) provides L_{r_m} at twice the cost of r_m.
- $L_{\exp}(A, \alpha E) = \alpha L_{\exp}(A, E)$. Alg is not influenced by $\|E\|$.
  Alg based on

$$f\left(\begin{bmatrix} X & E \\ 0 & X \end{bmatrix}\right) = \begin{bmatrix} f(X) & L(X, E) \\ 0 & f(X) \end{bmatrix}$$

is sensitive to $\|E\|$.

Cost Comparison

Kronecker–Sylvester alg (Kenney & Laub; see H, 2008):
$\mathrm{vec}(L(A, E)) = \frac{1}{2}(e^{A^T} \oplus e^A)\tau\big(\frac{1}{2}[A^T \oplus (-A)]\big)\mathrm{vec}(E)$, where $\tau(x) = \tanh(x)/x$ and $\|\frac{1}{2}[A^T \oplus (-A)]\| < \pi/2$.

Assume $\|A\| = 9$.

- e^A only by S&S: $16n^3$ flops.
- e^A and $L(A, E)$ by new algorithm: $48n^3$ flops.
- $L(A, E)$ by Kronecker–Sylvester $538n^3$ flops.

Tests show new alg superior to Kronecker–Sylvester in accuracy, too.

Condition Number Estimation (1)

$$\kappa_{\exp}(A) = \frac{\|L_{\exp}(A)\| \, \|A\|}{\|e^A\|}.$$

We have

$$\mathrm{vec}(L_{\exp}(A, E)) = K(A)\mathrm{vec}(E),$$

where $K(A) \in \mathbb{C}^{n^2 \times n^2}$ and $\|L(A)\|_F = \|K(A)\|_2$.

Lemma (H, 2008)

For $A \in \mathbb{C}^{n \times n}$ and any function f,

$$\frac{\|L(A)\|_1}{n} \le \|K(A)\|_1 \le n\|L(A)\|_1.$$

Condition Number Estimation (2)

- Apply block 1-norm estimation algorithm of H & Tisseur (2000)—`normest1` in MATLAB.
- Need to evaluate $L_{\exp}(A, E)$ and $L_{\exp}(A^*, E)$ for fixed A and several E.
- Store matrices accrued during computation of e^A and re-use them.
- Use the parameters for e^A: $\|\Delta E\|/\|E\| \leq 28u$.
- **Total cost**: cost of e^A plus about 8 $L_{\exp}$ evaluations.
- So e^A and $\kappa_{\exp}$ about 17 times the cost of just e^A.
- **Note**: *exact* computation of $\kappa_{\exp}$ is $O(n^5)$ flops.

Conclusions

- New alg for $L_{\exp}(A, E)$ with supporting backward error analysis.
- Order of magnitude more efficient than Kronecker–Sylvester alg.
- New alg for simultaneously computing e^A and estimating $\kappa_{\exp}(A)$.
- Currently applying our Padé framework to log, cos, sin, ...

MIMS EPrint 2008.26

http://www.manchester.ac.uk/mims/eprints

References I

A. H. Al-Mohy and N. J. Higham.
Computing the Fréchet derivative of the matrix exponential, with an application to condition number estimation.
MIMS EPrint 2008.26, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, Feb. 2008.
20 pp.

References II

N. J. Higham.
Functions of Matrices: Theory and Computation.
Society for Industrial and Applied Mathematics,
Philadelphia, PA, USA, 2008.
ISBN 978-0-898716-46-7.
xx+425 pp.

C. S. Kenney and A. J. Laub.
A Schur–Fréchet algorithm for computing the logarithm and exponential of a matrix.
SIAM J. Matrix Anal. Appl., 19(3):640–663, 1998.

References III

I. Najfeld and T. F. Havel.
Derivatives of the matrix exponential and their computation.
Advances in Applied Mathematics, 16:321–375, 1995.