Higher Order Fréchet Derivatives of Matrix Functions and their Applications

Samuel Relton
samuel.relton@maths.man.ac.uk,
www.maths.man.ac.uk/~srelton

Joint work with Nicholas J. Higham
higham@maths.man.ac.uk
www.maths.man.ac.uk/~higham

University of Manchester, UK

July 9, 2013

Overview

- Matrix Functions
- 1st Fréchet derivatives and condition numbers
- Higher order Fréchet derivatives
- Application 1: The level-2 condition number
- Application 2: Conditioning of Fréchet derivatives

Matrix Functions

What is a matrix function $f: \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$?

$$e^A = \sum_{k=0}^{\infty} A^k/k!$$

$$\log(I + A) = \sum_{k=1}^{\infty} (-1)^{k+1} A^k/k, \qquad \rho(A) < 1.$$

- Analytic functions: defined by power series (λ in convergence region).
- If $A = XDX^{-1}$ then $f(A) = Xf(D)X^{-1}$.
- Differential equations: $\frac{du}{dt} = Au, \quad u(t) = e^{At}u_0$.
- For 2nd order, use $\cos(A)$ and $\sin(A)$.

1st Fréchet derivative

Let $f : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ be a matrix function

Definition (Fréchet derivative)

The Fréchet derivative of f at A in the direction E is the unique linear function $L_f(A, \cdot)$ such that

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

1st Fréchet derivative

Let $f : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ be a matrix function

Definition (Fréchet derivative)

The Fréchet derivative of f at A in the direction E is the unique linear function $L_f(A, \cdot)$ such that

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

- Guaranteed to exist when f has $2n-1$ derivatives on $\Lambda(A)$
- Applications include manifold optimization and Markov models

Condition numbers

There are two (level-1) condition numbers.

Definition (Absolute condition number)

The absolute condition number of a matrix function is

$$\mathrm{cond}_{\mathrm{abs}}(f, A) = \lim_{\epsilon \to 0} \sup_{\|E\| \le \epsilon} \frac{\|f(A+E) - f(A)\|}{\epsilon} = \max_{\|E\|=1} \|L_f(A, E)\|$$

Definition (Relative condition number)

The relative condition number of a matrix function is

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

Higher Order Derivatives

The second Fréchet derivative $L_f^{(2)}(A, E_1, E_2)$ satisfies

$$L_f(A+E_2, E_1) - L_f(A, E_1) - L_f^{(2)}(A, E_1, E_2) = o(\|E_2\|).$$

Higher Order Derivatives

The second Fréchet derivative $L_f^{(2)}(A, E_1, E_2)$ satisfies

$$L_f(A+E_2, E_1) - L_f(A, E_1) - L_f^{(2)}(A, E_1, E_2) = o(\|E_2\|).$$

- kth Fréchet derivative exists when f has $2^k n - 1$ derivatives on $\Lambda(A)$ (H & R '13).
- E_i are interchangable: $L_f^{(2)}(A, E_1, E_2) = L_f^{(2)}(A, E_2, E_1)$.
- We also have

$$L_f^{(k)}(A, E_1, \dots, E_k) = \left.\frac{\partial}{\partial s_1 \cdots \partial s_k}\right|_{\mathbf{s}=0} f(A + s_1 E_1 + \cdots + s_k E_k).$$

Computing kth Fréchet derivatives

Given A and E_i this algorithm gives $L = L_f^{(k)}(A, E_1, \dots, E_k)$ (H & R '13).

$X_0 = A$
for $i = 1\colon k$
    $X_i = I_2 \otimes X_{i-1} + \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right] \otimes I_{2^{i-1}} \otimes E_i$
end
$F = f(X_k)$
Take L to be the upper-right $n \times n$ block of F.

Cost: 2^{3k} times more expensive than computing $f(A)$
(ignoring the structure of X_k).

Computing kth Fréchet derivatives

Given A and E_i this algorithm gives $L = L_f^{(k)}(A, E_1, \dots, E_k)$ (H & R '13).

$X_0 = A$
for $i = 1\colon k$
$\quad X_i = I_2 \otimes X_{i-1} + \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right] \otimes I_{2^{i-1}} \otimes E_i$
end
$F = f(X_k)$
Take L to be the upper-right $n \times n$ block of F.

Cost: 2^{3k} times more expensive than computing $f(A)$
(ignoring the structure of X_k).

For example:

$$L_f^{(2)}(A, E_1, E_2) = f\left(\left[\begin{array}{cc|cc} A & E_1 & E_2 & 0 \\ 0 & A & 0 & E_2 \\ \hline 0 & 0 & A & E_1 \\ 0 & 0 & 0 & A \end{array}\right]\right)_{1,4}$$

Application 1: Level-2 condition number

New algorithms give $f(A)$ and $\mathrm{cond}(f, A)$ but how accurate is $\mathrm{cond}(f, A)$?

Definition (Absolute level-2 condition number)

The absolute level-2 condition number of a matrix function is

$$\mathrm{cond_{abs}}^{[2]}(f, A) = \lim_{\epsilon \to 0} \sup_{\|Z\| \leq \epsilon} \frac{|\,\mathrm{cond_{abs}}(f, A + Z) - \mathrm{cond_{abs}}(f, A)|}{\epsilon}$$

Application 1: Level-2 condition number

New algorithms give $f(A)$ and $\text{cond}(f, A)$ but how accurate is $\text{cond}(f, A)$?

Definition (Absolute level-2 condition number)

The absolute level-2 condition number of a matrix function is

$$\text{cond}_{\text{abs}}{}^{[2]}(f, A) = \lim_{\epsilon \to 0} \sup_{\|Z\| \leq \epsilon} \frac{|\text{cond}_{\text{abs}}(f, A + Z) - \text{cond}_{\text{abs}}(f, A)|}{\epsilon}$$

- Complicated to analyze analytically for general f.
- Some results for special f:

$$\text{cond}_{\text{abs}}{}^{[2]}(x^{-1}, A) = 2\, \text{cond}_{\text{abs}}(x^{-1}, A)^{3/2},$$

$$\text{cond}_{\text{abs}}{}^{[2]}(e^{x}, A) = \text{cond}_{\text{abs}}(e^{x}, A), \text{ for normal matrices.}$$

- Promising numerical experiments for other f (not shown here)

Application 2: Conditioning of Fréchet derivatives

How accurately can we compute $L_f(A, E)$?

A bad example for the matrix logarithm:

$$A = \begin{bmatrix} e^{(\pi - 10^{-7})i} & 1000 \\ 0 & e^{(\pi + 10^{-7})i} \end{bmatrix}, \quad E = \begin{bmatrix} 0.23 & 0.05 \\ 0.41 & 0.49 \end{bmatrix},$$

$$V = \begin{bmatrix} 0.1535 + 0.1535i & 0.1535 + 0.1535i \\ 0.1535 + 0.7677i & 0.1535 + 0.1535i \end{bmatrix}.$$

Application 2: Conditioning of Fréchet derivatives

How accurately can we compute $L_f(A, E)$?

A bad example for the matrix logarithm:

$$A = \begin{bmatrix} e^{(\pi - 10^{-7})i} & 1000 \\ 0 & e^{(\pi + 10^{-7})i} \end{bmatrix}, \quad E = \begin{bmatrix} 0.23 & 0.05 \\ 0.41 & 0.49 \end{bmatrix},$$

$$V = \begin{bmatrix} 0.1535 + 0.1535i & 0.1535 + 0.1535i \\ 0.1535 + 0.7677i & 0.1535 + 0.1535i \end{bmatrix}.$$

With $u \approx 1.1 \times 10^{-16}$ we have (using 250 digit precision)

$$\frac{\|L_{\log}(A + uV, E) - L_{\log}(A, E)\|_1}{\|L_{\log}(A, E)\|_1} = 1.0318.$$

- How can we detect the Fréchet derivative is so sensitive?

The condition number of a Fréchet derivative

The absolute and relative condition number of a Fréchet derivative are

$$\mathrm{cond}_{\mathrm{abs}}(L_f, A, E) = \lim_{\epsilon \to 0} \sup_{\substack{\|\Delta A\| \le \epsilon \\ \|\Delta E\| \le \epsilon}} \frac{\|L_f(A + \Delta A, E + \Delta E) - L_f(A, E)\|}{\epsilon},$$

$$\mathrm{cond}_{\mathrm{rel}}(L_f, A, E) = \lim_{\epsilon \to 0} \sup_{\substack{\|\Delta A\| \le \epsilon \|A\| \\ \|\Delta E\| \le \epsilon \|E\|}} \frac{\|L_f(A + \Delta A, E + \Delta E) - L_f(A, E)\|}{\epsilon \|L_f(A, E)\|}.$$

The condition number of a Fréchet derivative

The absolute and relative condition number of a Fréchet derivative are

$$\mathrm{cond}_{\mathrm{abs}}(L_f, A, E) = \lim_{\epsilon \to 0} \sup_{\substack{\|\Delta A\| \le \epsilon \\ \|\Delta E\| \le \epsilon}} \frac{\|L_f(A + \Delta A, E + \Delta E) - L_f(A, E)\|}{\epsilon},$$

$$\mathrm{cond}_{\mathrm{rel}}(L_f, A, E) = \lim_{\epsilon \to 0} \sup_{\substack{\|\Delta A\| \le \epsilon \|A\| \\ \|\Delta E\| \le \epsilon \|E\|}} \frac{\|L_f(A + \Delta A, E + \Delta E) - L_f(A, E)\|}{\epsilon \|L_f(A, E)\|}.$$

They are related since

$$\mathrm{cond}_{\mathrm{rel}}(L_f, A, E) = \frac{\mathrm{cond}_{\mathrm{abs}}(L_f, A, sE)\|E\|}{\|L_f(A, E)\|},$$

where $s = \|A\| / \|E\|$.

Computing the condition number of the Fréchet derivative

Theorem (H & R '13)

$$\mathrm{cond}_{\mathrm{abs}}(L_f, A, sE) \le \lim_{\epsilon \to 0} \sup_{\|\Delta A\| \le \epsilon} \|L_f^{(2)}(A, sE, \Delta A)\| + \mathrm{cond}_{\mathrm{abs}}(f, A).$$

- Estimate first half by maximizing $\|L_f^{(2)}(A, sE, \Delta A)\|$ using `normest1` algorithm (H & Tisseur '00), calculate ≈ 8 second Fréchet derivatives.
- Use existing methods for second half.
- Now use $\mathrm{cond}_{\mathrm{rel}}(L_f, A, E) = \mathrm{cond}_{\mathrm{abs}}(L_f, A, sE)\|E\|/\|L_f(A, E)\|$.

Numerical Experiments

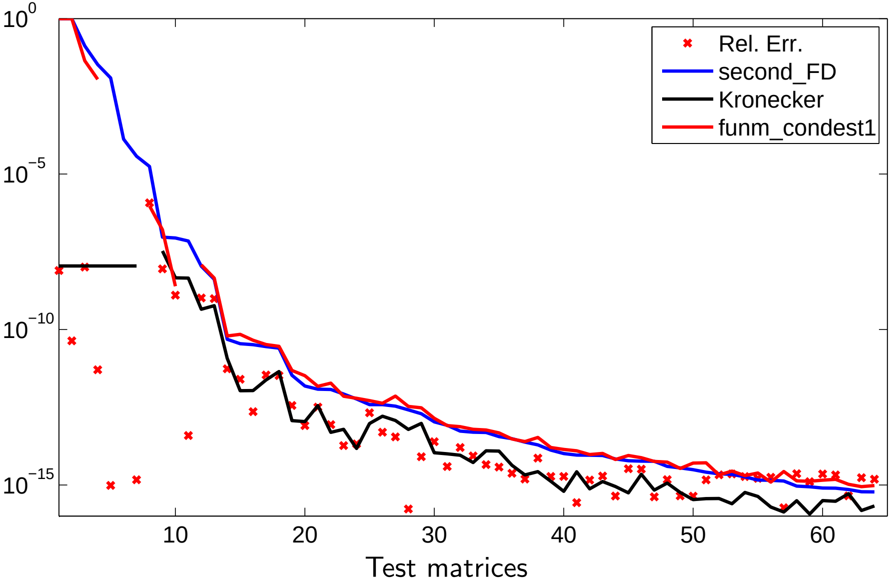
The competing methods are:

- second_FD - Our new method.
- Kronecker - Heuristic method to estimate $\text{cond}_{\text{rel}}(L_f, A, E)$.
- funm_condest1 - Computes upper bound using finite differences.

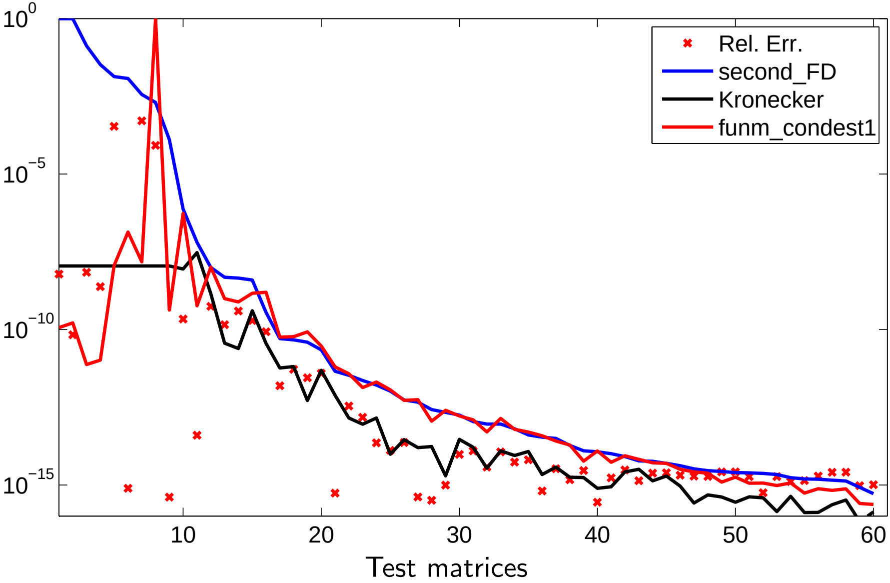
The test matrices are:

- mainly 10×10.
- taken from the Matrix Computation Toolbox and the literature.

Performance on the matrix logarithm



Performance for matrix powers - $A^{1/15}$



Conclusions for application 2

- `second_FD` cannot fail (unlike `Kronecker` and `funm_condest1`).
- The seemingly pessimistic estimates from `second_FD` can be attained (recall our "bad example").
- `second_FD` gives good condition number estimates even for ill conditioned problems.

Summary

- Higher order derivatives can be generalized to matrix functions
- Currently expensive to compute
- Level-2 condition number of matrix functions
- Estimate $\mathrm{cond}_{\mathrm{rel}}(L_f, A, E)$ with new `second_FD` method

Future Work:

- Cheaper computation of $L_f^{(k)}(A, E_1, \dots, E_k)$
- Additional applications

References

[1] - Higham and Relton, *Higher Fréchet derivatives of matrix functions and their level-2 condition numbers.* MIMS EPrint. In Preparation.

[2] - Higham and Relton, *The condition number of the Fréchet derivative of a matrix function.* MIMS Eprint. In Preparation.

[3] - Mathias, *A chain rule for matrix functions and applications.* SIAM J. Matrix Anal. Appl., 1996, 17:610–620.

[4] - Bhatia, *Matrix Analysis.* Springer 1996

[5] - Higham and Tisseur, *A block algorithm for matrix 1-norm estimation, with an application to 1-norm pseudospectra.* SIAM J. Matrix Anal. Appl., 2000, 21(4):1185–1201.