MANCHESTER 1824

The University of Manchester

M\crNA Manchester Numerical Analysis

How and Why to Estimate Condition Numbers for Matrix Functions

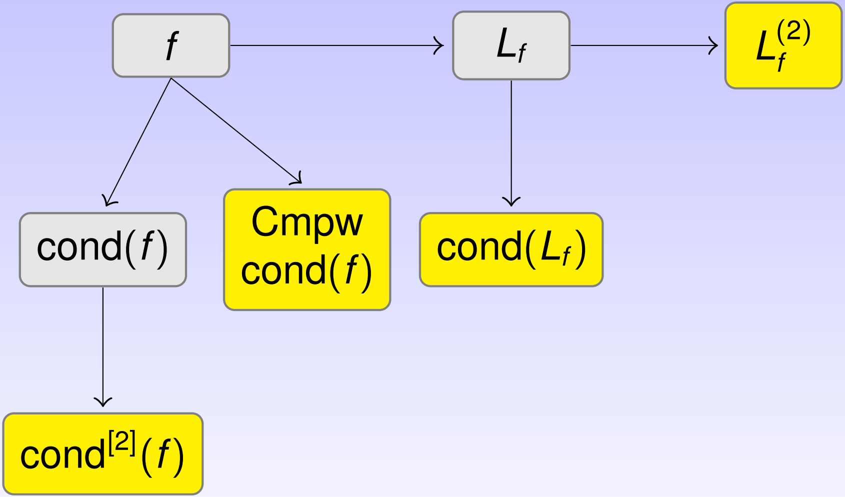
Nick Higham
School of Mathematics
The University of Manchester

http://www.maths.manchester.ac.uk/~higham
@nhigham, nickhigham.wordpress.com

Joint work with Sam Relton and Lijing Lin

Householder Symposium XIX
Spa, Belgium, June 8-13, 2014

The Big Picture



Fréchet Derivative

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

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

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

Fréchet Derivative

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

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

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

Examples:

$$(A+E)^2 - A^2 = AE + EA + E^2 \Rightarrow L_{x^2}(A,E) = AE + EA.$$

$$(A+E)^{-1} - A^{-1} = -A^{-1}EA^{-1} + O(E^2)$$
$$\Rightarrow L_{x^{-1}}(A,E) = -A^{-1}EA^{-1}.$$

Fréchet Derivative

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

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

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

Examples:

$$(A+E)^2 - A^2 = AE + EA + E^2 \Rightarrow L_{x^2}(A,E) = AE + EA.$$

$$(A+E)^{-1} - A^{-1} = -A^{-1}EA^{-1} + O(E^2)$$

$$\Rightarrow L_{x^{-1}}(A,E) = -A^{-1}EA^{-1}.$$

Scalar case: $L_f(x,e) = f('x)e$.

Applications of Fréchet Derivative

- First *and second* Fréchet derivative in Halley's method on Banach space (2003).
- Matrix geometric mean computation (2012).
- Model reduction (2012).
- Computation of correlated choice probabilities (2013).
- Registration of MRI images (2013).
- Markov models applied to cancer data (2014).
- Computing linearized backward errors for matrix functions (2014).

Computing L_f: Methods for Specific f

- **exponential** Kenney & Laub (1998), Al-Mohy & H (2009).
- **logarithm** Al-Mohy, H & Relton (2013).
- **fractional power** H & Lin (2013).

Latter three methods obtained by differentiating alg for f.

Computing L_f: General Methods

Theorem (Mathias, 1996)

If f is $2n-1$ times ctsly diffble,

$$f\left(\begin{bmatrix} A & E \\ 0 & A \end{bmatrix}\right) = \begin{bmatrix} f(A) & L_f(A,E) \\ 0 & f(A) \end{bmatrix}.$$

Complex step method .

Assume $f : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ and $A, E \in \mathbb{R}^{n\times n}$. Then (H & Al-Mohy, 2010):

$$L_f(A,E) \approx \operatorname{Im} \frac{f(A+ihE)}{h}.$$

Componentwise Sensitivity (1)

Let $E = e_i e_j^T$. Then

$$\lim_{\epsilon \to 0} \frac{f(A + \epsilon e_i e_j^T) - f(A)}{\epsilon} = L_f(A, e_i e_j^T).$$

$(L_f(A, e_i e_j^T))_{rs}$ measures the sensitivity of $f(A)_{rs}$ to perturbations in a_{ij}.

Componentwise Sensitivity (2)

Since L_f is a linear operator,

$$\mathrm{vec}(L_f(A, E)) = K_f(A)\mathrm{vec}(E)$$

where $K_f(A) \in \mathbb{C}^{n^2 \times n^2}$ is the **Kronecker form** of $L_f(A)$.

	a_{11}	a_{21}	a_{31}	$\dots$	a_{nn}
f_{11}					
f_{21}					
f_{31}			$K_f(A) \equiv$		
$\vdots$			$L_f(A, e_i e_j^T)$		
f_{nn}					

Componentwise Sensitivity (2)

Since L_f is a linear operator,

$$\mathrm{vec}(L_f(A,E)) = K_f(A)\mathrm{vec}(E)$$

where $K_f(A) \in \mathbb{C}^{n^2\times n^2}$ is the **Kronecker form** of $L_f(A)$.

	a_{11}	a_{21}	a_{31}	$\dots$	a_{nn}
f_{11}					
f_{21}					
f_{31}			$K_f(A) \equiv L_f(A, e_i e_j^T)$		
$\vdots$					
f_{nn}					

In progress: sensitivity of $y' = A(t)y$ (nuclear physics).

Absolute Condition Number

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

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

Lemma

$$\text{cond}_{\text{abs}}(f, A) = \|L_f(A)\|.$$

Absolute Condition Number

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

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

Lemma

$$\mathrm{cond}_{\mathrm{abs}}(f, A) = \|L_f(A)\|.$$

Scalar case: $\mathrm{cond}_{\mathrm{abs}}(f, x) = |f'(x)|$.

Relative Condition Number

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

The Setup

- Assume we have $O(n^3)$ algs for $f(A)$ and $L_f(A, E)$.
- $K_f(A)$ has n^4 entries.
- Computing $K_f(A)$ requires $O(n^5)$ flops.
- Wish to estimate norms and condition numbers in $O(n^3)$ flops.
- In practice we only ever work with $n \times n$ matrices.

Condition Estimation (1)

Can show

$$\frac{\|L_f(A)\|_1}{n} \le \|K_f(A)\|_1 \le n\|L_f(A)\|_1.$$

So problem reduces to **matrix norm estimation**.

Use the **block 1-norm estimator** of H & Tisseur (2000).
For $\|B\|_1$ it needs Bx and B^*y for several x and y. Here,

$$\begin{aligned} K_f(A)x &= \mathrm{vec}(L_f(A,X)), & \mathrm{vec}(X) &= x, \\ K_f(A)^*y &= \mathrm{vec}(L_f^\star(A,Y)), & \mathrm{vec}(Y) &= y. \end{aligned}$$

Condition Estimation (2)

Theorem (H & Lin, 2013)

Let $f \in C^{2n-1}$ and $\widetilde{f}(z) := \overline{f(\overline{z})}$.
If $\widetilde{f}(A)^* = \widetilde{f}(A^*)$ for all $A \in \mathbb{C}^{n\times n}$ then
$K_f(A)^* x = \mathrm{vec}(L_{\widetilde{f}}(A, X^*)^*)$*, where* $\mathrm{vec}(X) = x$.

For most functions of interest, $\widetilde{f} = f$, and we can deduce that

$$K_f(A)^* x = \mathrm{vec}(L_f(A, X^*)^*).$$

Higher Derivatives

Needed for

- level-2 condition number,
- understanding accuracy of algorithms for computing $L_f(A, E)$.

Second Fréchet derivative $L_f^{(2)}(A, E_1, E_2)$ is unique multilinear function of E_1, E_2 satisfying

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

Higher Derivatives

Needed for

- level-2 condition number,
- understanding accuracy of algorithms for computing $L_f(A, E)$.

Second Fréchet derivative $L_f^{(2)}(A, E_1, E_2)$ is unique multilinear function of E_1, E_2 satisfying

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

***k*th** Fréchet derivative defined by

$$L_f^{(k-1)}(A+\mathbf{E_k}, E_1, \ldots, E_{k-1}) - L_f^{(k-1)}(A, E_1, \ldots, E_{k-1}) - L_f^{(k)}(A, E_1, \ldots, \mathbf{E_k}) = o(\|\mathbf{E_k}\|).$$

Existing Literature

- Large literature on Fréchet derivatives in Banach space.
- Need specialized results for matrix functions.

Existence & Continuity of Fréchet Derivatives

- $\mathcal{D}$ = open subset of $\mathbb{C}$.
- $\mathbb{C}^{n\times n}(\mathcal{D}, p)$ = matrices with spectrum in $\mathcal{D}$ and largest Jordan block of size p.

Theorem (H & Relton, 2013)

If $f \in C^{2^k p-1}(\mathcal{D})$ and $A \in \mathbb{C}^{n\times n}(\mathcal{D}, p)$ then the kth Fréchet derivative $L_f^{(k)}(A)$ exists and $L_f^{(k)}(A, E_1, \dots, E_k)$ is continuous in A and $E_1, \dots, E_k \in \mathbb{C}^{n\times n}$.

- Proof uses Gâteaux derivative.
- $k = 1$: Mathias (1996).
- *Open problem*: necessary conditions.

Properties

Assume from now on conditions of theorem satisfied.
Then the E_i are interchangeable:

$$L_f^{(2)}(A, E_1, E_2) = L_f^{(2)}(A, E_2, E_1).$$

Indeed

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

How to Compute $L_f^{(2)}$

$X_1 = \left[\begin{smallmatrix} A & E_1 \\ 0 & A \end{smallmatrix}\right]$. Know $f(X_1) = \begin{bmatrix} f(A) & L_f(A,E) \\ 0 & f(A) \end{bmatrix}$.

Let

$$X_2 = \underbrace{I_2 \otimes X_1 + \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \otimes E_2}_{\text{“nivellateur”}} = \left[\begin{array}{cc|cc} A & E_1 & E_2 & 0 \\ 0 & A & 0 & E_2 \\ \hline 0 & 0 & A & E_1 \\ 0 & 0 & 0 & A \end{array}\right].$$

Then

$$f(X_2) = \left[\begin{array}{cc|cc} f(A) & L_f(A,E_1) & L_f(A,E_2) & L_f^{(2)}(A,E_1,E_2) \\ 0 & f(A) & 0 & L_f(A,E_2) \\ \hline 0 & 0 & f(A) & L_f(A,E_1) \\ 0 & 0 & 0 & f(A) \end{array}\right].$$

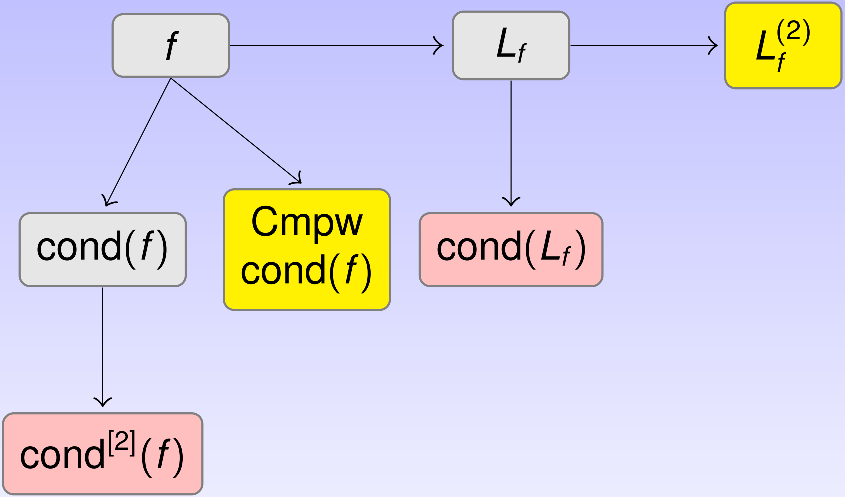
How to Compute $L_f^{(k)}$

Define $X_i \in \mathbb{C}^{2^i n \times 2^i n}$ by

$$X_i = I_2 \otimes X_{i-1} + \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \otimes I_{2^{i-1}} \otimes E_i, \qquad X_0 = A.$$

Theorem (H & Relton, 2013)

The $(1, n)$ block of $f(X_k)$ is $L_f^{(k)}(A, E_1, \ldots, E_k)$.


f
L_f
L_f^(2)
cond(f)
Cmpw
cond(f)
cond(L_f)
cond^[2](f)

Condition Number of the Fréchet Derivative

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

Theorem (H & Relton, 2013)

$$\begin{aligned} \max(\mathrm{cond}_{\mathrm{abs}}(f,A), sM)r &\le \mathrm{cond}_{\mathrm{rel}}(L_f, A, E) \\ &\le (\mathrm{cond}_{\mathrm{abs}}(f,A) + sM)r, \end{aligned}$$

$$s = \frac{\|A\|}{\|E\|}, \quad r = \frac{\|E\|}{\|L_f(A,E)\|}, \quad M = \max_{\|\Delta A\|=1} \|L_f^{(2)}(A, E, \Delta A)\|.$$

Hence we can compute $\mathrm{cond}_{\mathrm{rel}}(L_f, A, E)$ to within a factor 2.

Estimating $\text{cond}_{\text{rel}}(L_f, A, E)$

Need second Kronecker form satisfying $K_f^{(2)}(A) \in \mathbb{C}^{n^4 \times n^2}$

$$\text{vec}(L_f^{(2)}(A, E_1, E_2)) = (\text{vec}(E_1)^T \otimes I_{n^2})K_f^{(2)}(A)\text{vec}(E_2).$$

Can show M is within factor n of $\|(\text{vec}(E)^T \otimes I_{n^2})K_f^{(2)}(A)\|_1$.

Theorem (H & Relton 2013)

For most functions of interest,

$$\left[(\text{vec}(E)^T \otimes I_{n^2})K_f^{(2)}(A)\right]^* = (\text{vec}(E^*)^T \otimes I_{n^2})K_f^{(2)}(A^*).$$

Estimating $\text{cond}_{\text{rel}}(L_f, A, E)$

Need second Kronecker form satisfying $K_f^{(2)}(A) \in \mathbb{C}^{n^4 \times n^2}$

$$\text{vec}(L_f^{(2)}(A, E_1, E_2)) = (\text{vec}(E_1)^T \otimes I_{n^2}) K_f^{(2)}(A) \text{vec}(E_2).$$

Can show M is within factor n of $\|(\text{vec}(E)^T \otimes I_{n^2}) K_f^{(2)}(A)\|_1$.

Theorem (H & Relton 2013)

For most functions of interest,

$$\left[(\text{vec}(E)^T \otimes I_{n^2}) K_f^{(2)}(A)\right]^* = (\text{vec}(E^*)^T \otimes I_{n^2}) K_f^{(2)}(A^*).$$

Get $O(n^3)$ estimate of $\text{cond}_{\text{rel}}(L_f, A, E)$ within factor $\approx 6n$.

Level-2 Condition Number

"Condition number of the condition number".

Demmel (1987) showed that for

- matrix inversion (and **D. J. Higham**, 1995),
- the eigenproblem,
- polynomial zero-finding,
- pole assignment in linear control problems

(relative) level-1 and level-2 cond no's are equivalent.

Cheung & Cucker (2005) show same holds when "condition number = 1 / distance to nearest ill-posed problem".

Level-2 Condition Number

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

Theorem (H & Relton, 2013)

In the Frobenius norm,

$$\mathrm{cond}^{[2]}_{\mathrm{abs}}(f,A) \le \|K^{(2)}_f(A)\|_2.$$

Level-2 Condition Number: Exponential

Theorem

Let $A \in \mathbb{C}^{n\times n}$ be normal. Then in the 2-norm

$$\mathrm{cond}_{\mathrm{abs}}^{[2]}(\exp, A) = \mathrm{cond}_{\mathrm{abs}}(\exp, A).$$

Level-2 Condition Number: Exponential

Theorem

Let $A \in \mathbb{C}^{n\times n}$ be normal. Then in the 2-norm

$$\mathrm{cond}^{[2]}_{\mathrm{abs}}(\exp, A) = \mathrm{cond}_{\mathrm{abs}}(\exp, A).$$

Theorem

Let $A \in \mathbb{C}^{n\times n}$ be normal. Then in the 2-norm

$$1 \le \mathrm{cond}^{[2]}_{\mathrm{rel}}(\exp, A) \le 2\mathrm{cond}_{\mathrm{rel}}(\exp, A) + 1.$$

Level-2 Condition Number: Hermitian A

Theorem (H & Relton, 2013)

Let $A \in \mathbb{C}^{n\times n}$ be Hermitian and let $f : \mathbb{R} \to \mathbb{R}$ have a strictly monotonic derivative. Then in the Frobenius norm,

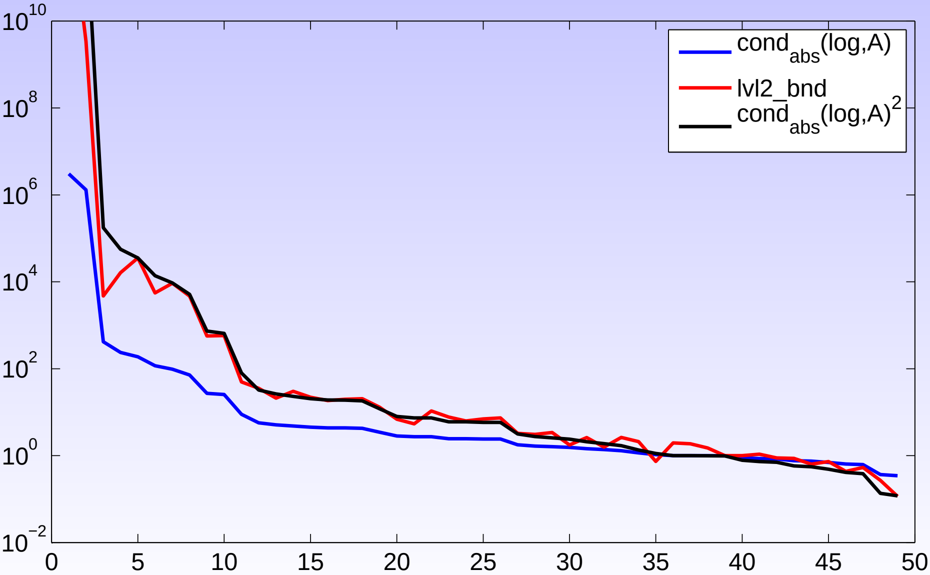
$$\mathrm{cond}_{\mathrm{abs}}(f, A) = \max_i |f'(\lambda_i)|.$$

Moreover, if the maximum is attained uniquely for $i = k$ then

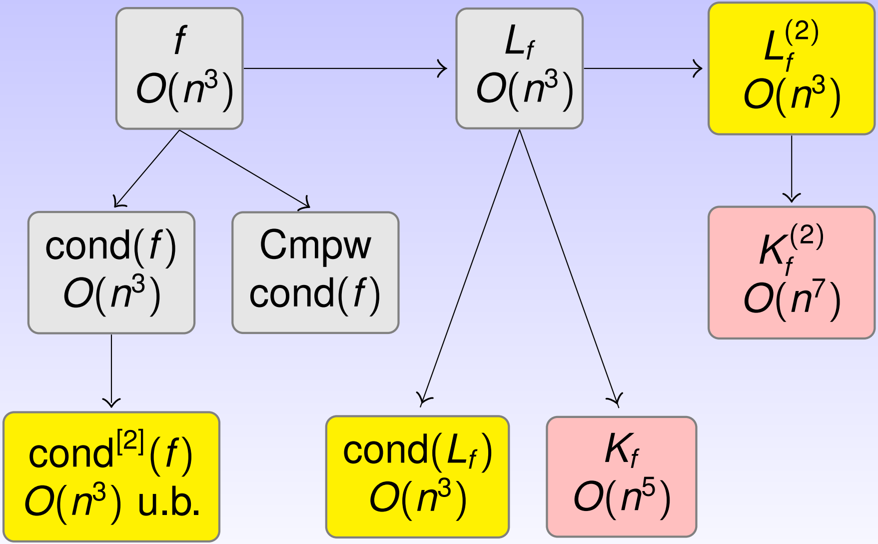
$$\mathrm{cond}^{[2]}_{\mathrm{abs}}(f, A) \geq |f''(\lambda_k)|.$$

- Applies to log, square root, inverse.
- For inverse the inequality is an equality.

Logarithm Experiment



Summary


f
$O(n^3)$
L_f
$O(n^3)$
$L_f^{(2)}$
$O(n^3)$
cond(f)
$O(n^3)$
Cmpw
cond(f)
$K_f^{(2)}$
$O(n^7)$
cond$^{[2]}$(f)
$O(n^3)$ u.b.
cond(L_f)
$O(n^3)$
K_f
$O(n^5)$

Open Questions

- Necessary conditions for existence of L_f?
- Quality of upper bound for $\mathrm{cond}^{[2]}(f)$?
- Relation between level-1 and level-2 condition numbers?
- How can we exploit the **symmetry** of $L_f^{(k)}(E_1, E_2, \ldots, E_k)$ in the E_i?

References I

- S. D. Ahipasaoglu, X. Li, and K. Natarajan.
  A convex optimization approach for computing correlated choice probabilities with many alternatives.
  Preprint 4034, Optimization Online, 2013.
  26 pp.

- A. H. Al-Mohy and N. J. Higham.
  Computing the Fréchet derivative of the matrix exponential, with an application to condition number estimation.
  SIAM J. Matrix Anal. Appl., 30(4):1639–1657, 2009.

References II

A. H. Al-Mohy, N. J. Higham, and S. D. Relton.
Computing the Fréchet derivative of the matrix logarithm and estimating the condition number.
SIAM J. Sci. Comput., 35(4):C394–C410, 2013.

S. Amat, S. Busquier, and J. M. Gutiérrez.
Geometric constructions of iterative functions to solve nonlinear equations.
J. Comput. Appl. Math., 157(1):197–205, 2003.

J. Ashburner and G. R. Ridgway.
Symmetric diffeomorphic modelling of longitudinal structural MRI.
Frontiers in Neuroscience, 6, 2013.
Article 197.

References III

D. Cheung and F. Cucker.
A note on level-2 condition numbers.
J. Complexity, 21:314–319, 2005.

E. Deadman and N. J. Higham.
Testing matrix function algorithms using identities.
MIMS EPrint 2014.13, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, Mar. 2014.
15 pp.

J. W. Demmel.
On condition numbers and the distance to the nearest ill-posed problem.
Numer. Math., 51:251–289, 1987.

References IV

B. García-Mora, C. Santamaría, G. Rubio, and J. L. Pontones.
Computing survival functions of the sum of two independent Markov processes. An application to bladder carcinoma treatment.
Int. Journal of Computer Mathematics, 91(2):209–220, 2014.

D. J. Higham.
Condition numbers and their condition numbers.
Linear Algebra Appl., 214:193–213, 1995.

References V

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.

N. J. Higham and A. H. Al-Mohy.
Computing matrix functions.
Acta Numerica, 19:159–208, 2010.

N. J. Higham and L. Lin.
An improved Schur–Padé algorithm for fractional powers of a matrix and their Fréchet derivatives.
SIAM J. Matrix Anal. Appl., 34(3):1341–1360, 2013.

References VI

N. J. Higham and S. D. Relton.
Estimating the condition number of the Fréchet derivative of a matrix function.
MIMS EPrint 2013.84, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, Dec. 2013.
18 pp.

References VII

- N. J. Higham and S. D. Relton.
  Higher order Fréchet derivatives of matrix functions and the level-2 condition number.
  MIMS EPrint 2013.58, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, Nov. 2013.
  19 pp.
  Revised April 2014. To appear in SIAM J. Matrix Anal. Appl.

- B. Jeuris, R. Vandebril, and B. Vandereycken.
  A survey and comparison of contemporary algorithms for computing the matrix geometric mean.
  Electron. Trans. Numer. Anal., 39:379–402, 2012.

References VIII

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.

R. Mathias.
A chain rule for matrix functions and applications.
SIAM J. Matrix Anal. Appl., 17(3):610–620, 1996.

References IX

D. Petersson.
A Nonlinear Optimization Approach to $\mathcal{H}_2$-Optimal Modeling and Control.
PhD thesis, Department of Electrical Engineering, Linköping University, Sweden, SE-581 83 Linköping, Sweden, 2013.
Dissertation No. 1528.

D. Petersson and J. Löfberg.
Model reduction using a frequency-limited $\mathcal{H}_2$-cost.
Technical report, Department of Electrical Engineering, Linköpings Universitet, SE-581 83, Linköping, Sweden, Mar. 2012.
13 pp.

Extra slides

Level-2 Condition Number: Matrix Inverse

Theorem

For nonsingular $A \in \mathbb{C}^{n\times n}$,

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

- Latter is what is expected from the scalar case: $f(x) = x^{-1} \ \Rightarrow\ |f''| = |2(f')^{3/2}|$.
- Have experimental evidence that matrix case can be similar to scalar case.