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.

Computational Linear Algebra and Optimization for the Digital Economy
Edinburgh, October 2013

Sources of Error

$f : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$.

Want to compute $f(A)$ but given $A + \Delta A$ not A.

ΔA may come from

- Rounding errors in storing A.
- Measurement errors.
- Errors from an earlier computation.

Sources of Error

$f : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$.
Want to compute $f(A)$ but given $A + \Delta A$ not A.

ΔA may come from

- Rounding errors in storing A.
- Measurement errors.
- Errors from an earlier computation.
- *Photocopying errors!*

Photocopier

http://en.wikipedia.org/wiki/Photocopier

- "Most current photocopiers use a technology called xerography, a dry process that uses electrostatic charges on a light sensitive photoreceptor to first attract and then transfer toner particles (a powder) onto paper in the form of an image"
- "There is an increasing trend for new photocopiers to adopt digital technology, thus replacing the older analog technology. With digital copying, the copier effectively consists of an integrated scanner and laser printer."

Xerox WorkCentre 7535, 7556

August 2013: German computer scientist David Kriesel discovered the machines often change “6” to “8”.

Before	After
7 113569370251 11356937025	7 113569370251 11356937025
113569470251 113569470251	113569470251 113569470251
113669471251 113669471251	113869471251 113669471251
113669571251 113669581251	113669571251 113869581251
113669581261 113669581261	113669581261 113669581261
114669581262 114669581262	114669581262 114869581262
114670581262 114670581262	114670581262 114670581262
115670681262 115670681262	115670681262 115670681262

- Jbig2 compression algorithm implicated.

Maxim

Every code for solving a numerical problem should return an error estimate or bound with the computed result.

Conditioning of $f(x)$

Scalar function $f \in C^2$, and $y = f(x)$, $y + \Delta y = f(x + \Delta x)$. Then

$$\frac{\Delta y}{y} = \left(\frac{xf'(x)}{f(x)}\right)\frac{\Delta x}{x} + O(\Delta x^2).$$

The **relative condition number**

$$c(x) = \left|\frac{xf'(x)}{f(x)}\right|.$$

Conditioning of $f(x)$

Scalar function $f \in C^2$, and $y = f(x)$, $y + \Delta y = f(x + \Delta x)$. Then

$$\frac{\Delta y}{y} = \left(\frac{xf'(x)}{f(x)}\right)\frac{\Delta x}{x} + O(\Delta x^2).$$

The **relative condition number**

$$c(x) = \left|\frac{xf'(x)}{f(x)}\right|.$$

Condition number of the condition number: assuming $xf'(x) > 0$ and $f(x) > 0$,

$$c^{[2]}(x) := \left|\frac{xc'(x)}{c(x)}\right| = \left|1 + x\left(\frac{f''(x)}{f'(x)} - \frac{f'(x)}{f(x)}\right)\right|.$$

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 $E \in \mathbb{C}^{n\times n}$

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

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 $E \in \mathbb{C}^{n\times n}$

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

Examples:

$$(X+E)^2 - X^2 = XE + EX + E^2 \Rightarrow L_{x^2}(X,E) = XE + EX.$$

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

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 $E \in \mathbb{C}^{n\times n}$

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

Examples:

$$(X+E)^2 - X^2 = XE + EX + E^2 \Rightarrow L_{x^2}(X,E) = XE + EX.$$

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

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

Gâteaux Derivative

Gâteaux derivative of matrix function f at A in direction E:

$$G_f(A, E) = \lim_{\epsilon \to 0} \frac{f(A + \epsilon E) - f(A)}{\epsilon}.$$

Weaker notion of differentiability.

Applications of Fréchet Derivative

- Computation of correlated choice probabilities (2013),
- registration of MRI images (2013),
- Markov models applied to cancer data (2013),
- matrix geometric mean computation (2012),
- model reduction (2012),
- first and second Fréchet derivative in Halley's method on Banach space (2003).

Componentwise Sensitivity (1)

Let $E = e_i e_j^T$. Then

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

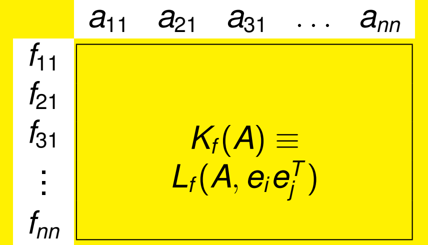
$(L_f(X, e_i e_j^T))_{rs}$ measures the sensitivity of $f(X)_{rs}$ to perturbations in a_{ij}.

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 the Fréchet derivative.



Condition Number

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

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

Lemma

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

Condition Number

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

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

Lemma

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

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

Relative Condition Number

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

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

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: Via $2n \times 2n$ Matrix

Theorem (Mathias, 1996)

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

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

- Note that $L_f(A, \alpha E) = \alpha L_f(A,E)$, but α may effect alg used for the evaluation.

Computing L_f: Complex Step

Assume that $f : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ and $A, E \in \mathbb{R}^{n\times n}$. Then

$$f(A + ihE) - f(A) - ihL_f(A, E) = o(h).$$

Thus (Al-Mohy & H, 2010)

$$f(A) \approx \text{Re } f(A + ihE),$$

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

- Errors $O(h^2)$.
- h not restricted by fl pt arith considerations. Can take $h = 10^{-100}$.
- f alg must not employ complex arith.

Condition Estimation (1)

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 the Fréchet derivative.

Can show

$$\|L_f(A)\|_F = \|K_f(A)\|_2,$$

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

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

Condition Estimation (2)

Use the **block 1-norm estimator** of H & Tisseur (2000).
For $\|B\|_1$ it needs Bx and B^*y for several x and y.
Let $\mathrm{vec}(X) = x$. Then $K_f(A)x = \mathrm{vec}(L_f(A, X))$.

Theorem (H & Lin, 2013)

Let $f \in C^{2n-1}$ and 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* $\mathrm{vec}(X) = x$.

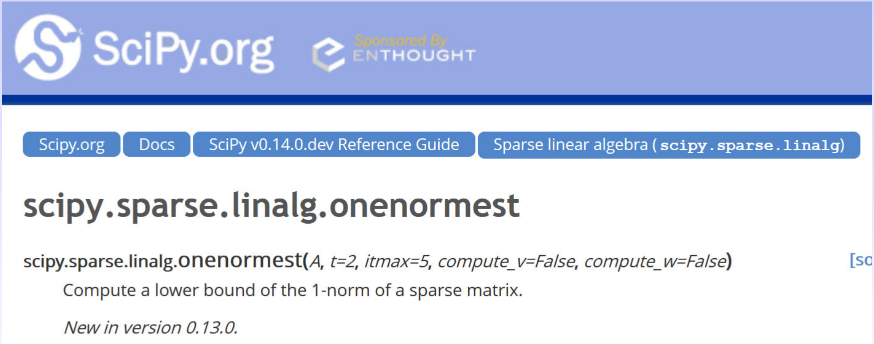
- $\widetilde{f} = f$ for most functions of interest.
- $\widetilde{f} = f \Rightarrow \widetilde{f}(A^*)^* = \widetilde{f}(A^*)$.

Software for 1-Norm Estimator

MATLAB: `normest1`.

NAG library: `F04YD`, `F04ZD`, Mark 24.

Python:


SciPy.org Sponsored By ENTHOUGHT

Scipy.org | Docs | SciPy v0.14.0.dev Reference Guide | Sparse linear algebra (`scipy.sparse.linalg`)

scipy.sparse.linalg.onenormest

scipy.sparse.linalg.onenormest(*A*, *t=2*, *itmax=5*, *compute_v=False*, *compute_w=False*) [so

Compute a lower bound of the 1-norm of a sparse matrix.

New in version 0.13.0.

Python: SciPy 0.13.0

- linalg/_expm_frechet.py
- linalg/_expm_multiply.py
- linalg/_sqrtm.py (blocked)

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 mutlilinear 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 mutlilinear 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)

Let f be $2^k p - 1$ times continuously differentiable on $\mathcal{D}$. Then for $A \in \mathbb{C}^{n\times n}(\mathcal{D},p)$ the kth Fréchet derivative $L_f^{(k)}(A)$ exists and $L_f^{(k)}(A, E_1, \ldots, E_k)$ is continuous in A and $E_1, \ldots, E_k \in \mathbb{C}^{n\times n}$.

- Proof uses Gâteaux derivative.
- $k = 1$: Mathias (1996).

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

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

The Setup

- Wish to estimate norms and condition numbers.
- Knowing how to evaluate exact quantities helps **develop** and **test** the estimates.
- In practice we only ever work with $n \times n$ matrices.

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

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

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 $L_f^{(k)}(A, E_1, \dots, E_k)$.

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)

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

for a Kronecker matrix $K^{(2)}_f(A) \in \mathbb{C}^{n^4\times n^2}$*.*

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: Matrix Inverse

Theorem

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

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

- 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.

Condition Number of the Fréchet Derivative

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

Theorem (H & Relton, 2013)

We have

$$\begin{aligned} \text{cond}_{\text{abs}}(f, A) &\le \text{cond}_{\text{abs}}(L_f, A, E) \\ &\le \max_{\|Z\|=1} \|L_f^{(2)}(A, E, Z)\| + \text{cond}_{\text{abs}}(f, A). \end{aligned}$$

We have developed method to estimate the bounds.

Summary and Open Questions

- Showed that given an $O(n^3)$ flop method for f:
  - Computing $L_f^{(k)}(A, E)$ costs $O(8^k n^3)$ flops.
  - Computing $K_f^{(k)}(A)$ costs $O(8^k n^{3+2k})$ flops.
- Cost of **estimates** is $O(n^3)$ flops, given an $O(n^3)$ flop method for L_f.
- Open questions about 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?
- Looking at componentwise condition numbers.

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 and N. J. Higham.
The complex step approximation to the Fréchet derivative of a matrix function.
Numer. Algorithms, 53(1):133–148, 2010.

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.

References III

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

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, 2013.

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.
The condition number of the Fréchet derivative of a matrix function.
MIMS EPrint, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, 2013.
In preparation.

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.

References VII

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.

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 VIII

- 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.