Matrix Condition Estimators in 2-norm

Jurjen Duintjer Tebbens

Institute of Computer Science

Academy of Sciences of the Czech Republic

duintjertebbens@cs.cas.cz

Miroslav Tůma

Institute of Computer Science

Academy of Sciences of the Czech Republic

tuma@cs.cas.cz

Preconditioning 2013, Oxford

June 20, 2013

Outline

Introduction: The Problem

Matrix condition number: an important quantity used in numerical linear algebra

$$\kappa(A) = \|A\| \cdot \|A^{-1}\|$$

Introduction: The Problem

Matrix condition number: an important quantity used in numerical linear algebra

$$\kappa(A) = \|A\| \cdot \|A^{-1}\|$$

- Assessing quality of computed solutions
- Estimating sensitivity to perturbations
- Monitor and control adaptive computational processes.

Introduction: The Problem

Matrix condition number: an important quantity used in numerical linear algebra

$$\kappa(A) = \|A\| \cdot \|A^{-1}\|$$

- Assessing quality of computed solutions
- Estimating sensitivity to perturbations
- Monitor and control adaptive computational processes.

- Here: A upper triangular (no loss of generality - computations typically based on triangular decomposition)
- Euclidean norm

Introduction: Earlier work

- Condition number estimation is important $\rightarrow$ a lot of excellent previous work
- Part of standard libraries as LAPACK

Introduction: Earlier work

- Condition number estimation is important → a lot of excellent previous work
- Part of standard libraries as LAPACK

- Turing (1948); Wilkinson (1961)
- Gragg, Stewart (1976); Cline, Moler, Stewart, Wilkinson (1979); Cline, Conn, van Loan (1982); van Loan (1987)
- Incremental: Bischof (1990, 1991), Bischof, Pierce, Lewis (1990), Bischof, Tang (1992); Ferng, Golub, Plemmons (1991); Pierce, Plemmons (1992); 2-norm estimator based on pivoted QLP: Stewart (1998); Duff, Vömel (2002)
- 1-norm: Hager (1984), Higham (1987, 1988, 1989, 1990) [175], Higham, Tisseur (2000).
- See also other techniques in various applications: adaptive filters, recursive least-squares in signal processing, ACE for multilevel PDE solvers.
- Typically estimating lower bound for $\kappa(A)$.

Introduction: The goal

- Getting better understanding of incremental estimation methods in 2-norm.

Introduction: The goal

- Getting better understanding of incremental estimation methods in 2-norm.
- Starting point: the methods by Bischof (1990) (incremental condition number estimation - ICE) and Duff, Vömel (2002) (incremental norm estimation - INE).

Introduction: The goal

- Getting better understanding of incremental estimation methods in 2-norm.
- Starting point: the methods by Bischof (1990) (incremental condition number estimation - ICE) and Duff, Vömel (2002) (incremental norm estimation - INE).
- Discussing more accurate estimation techniques and assembling theoretical and experimental evidence about this (note that it is often sufficient to have the estimates within a reasonable multiplicative factor from the exact $\kappa(A)$ - Demmel (1997)); matrix inverse can provide an additional information

Introduction: The goal

- Getting better understanding of incremental estimation methods in 2-norm.
- Starting point: the methods by Bischof (1990) (incremental condition number estimation - ICE) and Duff, Vömel (2002) (incremental norm estimation - INE).
- Discussing more accurate estimation techniques and assembling theoretical and experimental evidence about this (note that it is often sufficient to have the estimates within a reasonable multiplicative factor from the exact $\kappa(A)$ - Demmel (1997)); matrix inverse can provide an additional information
- Motivated also by methods for dropping in preconditioner computation (see Bollhöfer, Saad (2001 - 2006), Bru et al, 2008, 2010; talk by J. Kopal at the Sparse Days (2013))

Outline

ICE - Bischof (1990)

$$\hat{R} = \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix}$$

ICE - Bischof (1990)

$$\hat{R} = \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix}$$

- Bischof (1990): estimates to extremal singular values and left singular vectors: $R = U\Sigma V^T \Rightarrow \|u_{ext}^T R\| = \|u_{ext}^T U\Sigma V^T\| = \sigma_{ext}(R)$
- ICE computes:

$$\sigma_{ext}^C(R) = \|y_{ext}^T R\| \approx \sigma_{ext}(R),$$

ICE - Bischof (1990)

$$\hat{R} = \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix}$$

- Bischof (1990): estimates to extremal singular values and left singular vectors: $R = U\Sigma V^T \Rightarrow \|u_{ext}^T R\| = \|u_{ext}^T U\Sigma V^T\| = \sigma_{ext}(R)$
- ICE computes:

$$\sigma_{ext}^C(R) = \|y_{ext}^T R\| \approx \sigma_{ext}(R),$$

$$\|\hat{y}_{ext}^T \hat{R}\| = \ \text{ext}_{\|[s,c]\|=1} \left\| \begin{bmatrix} s\, y_{ext}^T, & c \end{bmatrix} \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix} \right\|,$$

ICE - Bischof (1990)

$$\hat{R} = \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix}$$

- Bischof (1990): estimates to extremal singular values and left singular vectors: $R = U\Sigma V^T \Rightarrow \|u_{ext}^T R\| = \|u_{ext}^T U\Sigma V^T\| = \sigma_{ext}(R)$
- ICE computes:

$$\sigma_{ext}^C(R) = \|y_{ext}^T R\| \approx \sigma_{ext}(R),$$

$$\|\hat{y}_{ext}^T \hat{R}\| = \text{ ext}_{\|[s,c]\|=1} \left\| \begin{bmatrix} s\, y_{ext}^T, & c \end{bmatrix} \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix} \right\|,$$

- s_{ext} and c_{ext}: components of the eigenvector corresponding to the extremal (minimum or maximum) eigenvalue of B_{ext}^C

$$B_{ext}^C \equiv \begin{bmatrix} \sigma_{ext}^C(R)^2 + (y_{ext}^T v)^2 & \gamma(y_{ext}^T v) \\ \\ \gamma(y_{ext}^T v) & \gamma^2 \end{bmatrix}$$

INE - Duff, Vömel (2002)

$$\hat{R} = \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix}$$

INE - Duff, Vömel (2002)

$$\hat{R} = \left[\begin{array}{cc} R & v \\ 0 & \gamma \end{array} \right]$$

- Duff, Vömel (2002): estimates to extremal singular values and right singular vectors (originally used only to estimate the 2-norm)
- INE computes

$$\sigma_{ext}^{N}(R) = \|Rz_{ext}\| \approx \sigma_{ext}(R)$$

INE - Duff, Vömel (2002)

$$\hat{R} = \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix}$$

- Duff, Vömel (2002): estimates to extremal singular values and right singular vectors (originally used only to estimate the 2-norm)
- INE computes

$$\sigma_{ext}^{N}(R) = \|Rz_{ext}\| \approx \sigma_{ext}(R)$$

$$\|\hat{R}\hat{z}_{ext}\| = \text{ext}_{\|[s,c]\|=1} \left\| \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} s\, z_{ext} \\ c \end{bmatrix} \right\|$$

INE - Duff, Vömel (2002)

$$\hat{R} = \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix}$$

- Duff, Vömel (2002): estimates to extremal singular values and right singular vectors (originally used only to estimate the 2-norm)
- INE computes

$$\sigma_{ext}^{N}(R) = \|R z_{ext}\| \approx \sigma_{ext}(R)$$

$$\|\hat{R}\hat{z}_{ext}\| = \text{ext}_{\|[s,c]\|=1} \left\| \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} s\, z_{ext} \\ c \end{bmatrix} \right\|$$

- Again, s_{ext} and c_{ext}: components of the eigenvector corresponding to the extremal (minimum or maximum) eigenvalue of B_{ext}^{N}

$$B_{ext}^{N} \equiv \begin{bmatrix} \sigma_{ext}^{N}(R)^2 & z_{ext}^T R^T v \\ z_{ext}^T R^T v & v^T v + \gamma^2 \end{bmatrix}$$

Outline

ICE and INE when both direct and inverse factors available: ICE

- Direct and inverse factors: having both R and R^{-1} (mixed direct/inverse (incomplete) decompositions, some other applications)

ICE and INE when both direct and inverse factors available: ICE

- Direct and inverse factors: having both R and R^{-1} (mixed direct/inverse (incomplete) decompositions, some other applications)

Theorem

Computing the inverse factor R^{-1} in addition to R **does not give any improvement for ICE** *(estimation of the extreme singular values and corresponding left singular vectors):*

ICE and INE when both direct and inverse factors available: ICE

- Direct and inverse factors: having both R and R^{-1} (mixed direct/inverse (incomplete) decompositions, some other applications)

Theorem

Computing the inverse factor R^{-1} in addition to R **does not give any improvement for ICE** *(estimation of the extreme singular values and corresponding left singular vectors):*

Let R be a nonsingular upper triangular matrix. Then the ICE estimates of the singular values of R and R^{-1} satisfy

$$\sigma_-^C(R) = 1/\sigma_+^C(R^{-1}).$$

The approximate left singular vectors y_- and x_+ corresponding to the ICE estimates for R and R^{-1}, respectively, satisfy

$$\sigma_-^C(R)x_+^T = y_-^T R$$

ICE and INE when both direct and inverse factors available: INE

Theorem

INE maximization applied to R^{-1} **may provide a better estimate** *than INE minimization applied to* R*:*

ICE and INE when both direct and inverse factors available: INE

Theorem

INE maximization applied to R^{-1} **may provide a better estimate** *than INE minimization applied to* R*:*

Let R *be a nonsingular upper triangular matrix. Assume that the INE estimates of the singular values of* R *and* R^{-1} *satisfy* $1/\sigma_+^N(R^{-1}) = \sigma_-^N(R) = \sigma_-(R)$*. Then the INE estimates of the singular values related to the extended matrix satisfy*

$$1/\sigma_+^N(\hat{R}^{-1}) \leq \sigma_-^N(\hat{R})$$

with equality if and only if v *is collinear with the left singular vector corresponding to the smallest singular value of* R*.*

ICE and INE when both direct and inverse factors available: INE

Theorem

INE maximization applied to R^{-1} **may provide a better estimate** *than INE minimization applied to R:*

Let R be a nonsingular upper triangular matrix. Assume that the INE estimates of the singular values of R and R^{-1} satisfy $1/\sigma_+^N(R^{-1}) = \sigma_-^N(R) = \sigma_-(R)$. Then the INE estimates of the singular values related to the extended matrix satisfy

$$1/\sigma_+^N(\hat{R}^{-1}) \leq \sigma_-^N(\hat{R})$$

with equality if and only if v is collinear with the left singular vector corresponding to the smallest singular value of R.

Rather technical in case the assumption is relaxed to $1/\sigma_+^N(R^{-1}) \leq \sigma_-^N(R)$: the superiority of maximization does not apply always.

An example showing the possible gap between the ICE and INE estimates

$$R = \begin{bmatrix} 2 & 0 & 1 \\ & 1 & 0 \\ & & 1 \end{bmatrix}, \; R^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ & 1 & 0 \\ & & 1 \end{bmatrix}, \sigma_{-}(R) = 0.874$$

$$1/\sigma_{+}^{C}(R^{-1}) = \sigma_{-}^{C}(R) = 1$$

$$0.8944 \approx 1/\sigma_{+}^{N}(R^{-1}) < \sigma_{-}^{N}(R) = 1$$

An example showing the possible gap between the ICE and INE estimates

$$R = \begin{bmatrix} 2 & 0 & 1 \\ & 1 & 0 \\ & & 1 \end{bmatrix}, \ R^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ & 1 & 0 \\ & & 1 \end{bmatrix}, \sigma_-(R) = 0.874$$

$$1/\sigma_+^C(R^{-1}) = \sigma_-^C(R) = 1$$
$$0.8944 \approx 1/\sigma_+^N(R^{-1}) < \sigma_-^N(R) = 1$$

$$\hat{R} = \begin{bmatrix} 2 & 0 & 1 & 1 \\ & 1 & 0 & 1 \\ & & 1 & 1 \\ & & & 1 \end{bmatrix}, \sigma_-(\hat{R}) \approx 0.5155$$

$$\sigma_-^C(\hat{R}) \equiv 1/\sigma_+^C(\hat{R}^{-1}) \approx 0.618$$
$$0.5381 \approx 1/\sigma_+^N(\hat{R}^{-1}) < \sigma_-^N(\hat{R}) \approx 0.835$$

Example: INE with maximization and minimization

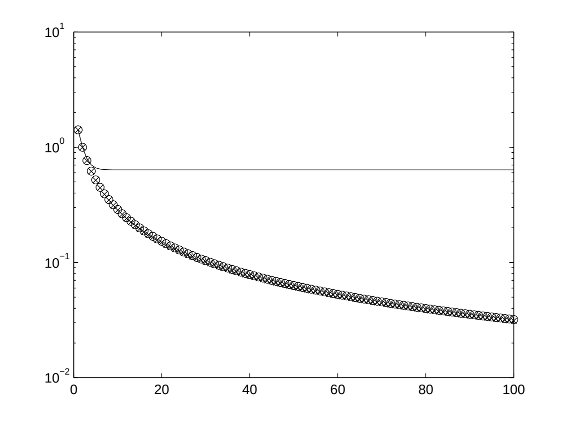


Figure : INE estimation of the smallest singular value of the 1D Laplacians of size one until hundred: INE with minimization (solid line), INE with maximization (circles) and exact minimum singular values (crosses).

Example: INE with maximization and exact smallest singular value

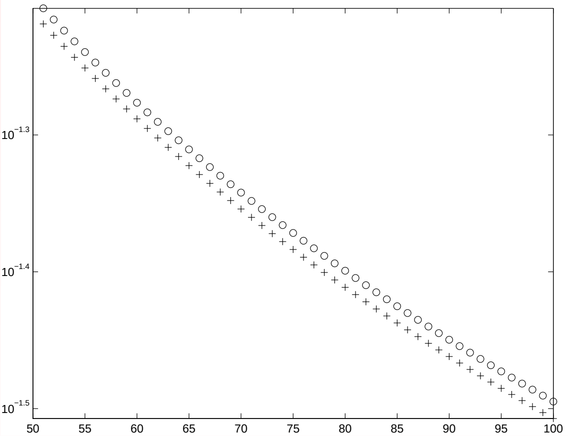


Figure : INE estimation of the smallest singular value of the 1D Laplacians of size fifty until hundred (zoom of previous figure for INE with maximization and exact minimum singular values).

Outline

INE versus ICE

Theorem

Consider norm estimation of the extended matrix

$$\hat{R} = \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix},$$

ICE and INE start with $\sigma_+ \equiv \sigma_+^C(R) = \sigma_+^N(R)$; y *LSV,* z *RSV,* $w = Rz/\sigma^+$. *The approximation* $\sigma_+^N(\hat{R})$ *from INE is at least as good as* $\sigma_+^C(\hat{R})$ *from ICE if*

$$(v^T w)^2 \geq \rho_1, \qquad (1)$$

where ρ_1 *is the smaller root of the quadratic equation in* $(v^T w)^2$*,*

$$(v^T w)^4 \quad + \quad \left(\frac{\gamma^2 + (v^T y)^2}{\sigma_+^2} \left(v^T v - (v^T y)^2 \right) - v^T v - (v^T y)^2 \right) (v^T w)^2$$
$$+ \quad (v^T y)^2 \left(\frac{\gamma^2 + v^T v}{\sigma_+^2} \left((v^T y)^2 - v^T v \right) + v^T v \right) = 0. \qquad (2)$$

Example: ICE versus INE

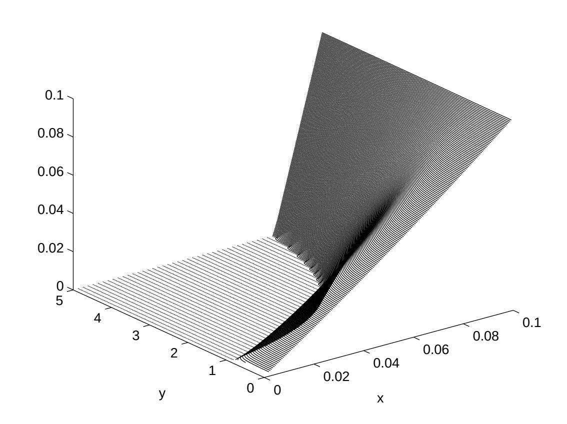


Figure : Value of ρ_1 in dependence of $(v^T y)^2$ (x-axis) and γ^2 (y-axis) with $\Delta = 0$, $\|v\|^2 = 0.1$.

Example: ICE versus INE

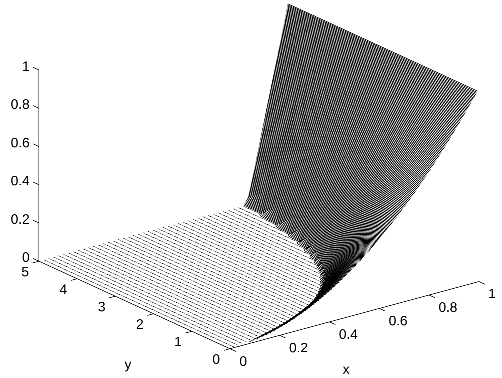


Figure : Value of ρ_1 in dependence of $(v^T y)^2$ (x-axis) and γ^2 (y-axis) with $\Delta = 0$, $\|v\|^2 = 1$.

Example: ICE versus INE

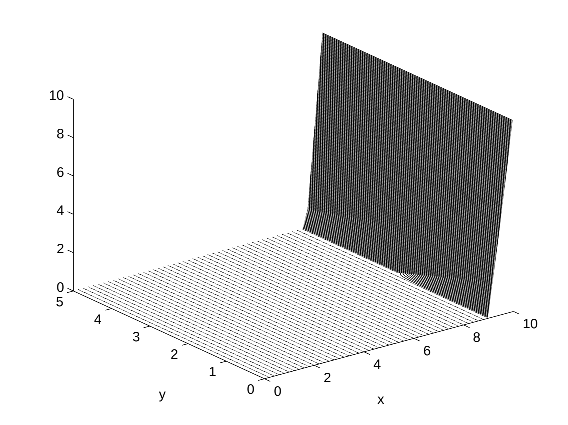


Figure : Value of ρ_1 in dependence of $(v^T y)^2$ (x-axis) and γ^2 (y-axis) with $\Delta = 0$, $\|v\|^2 = 10$.

Example: ICE versus INE

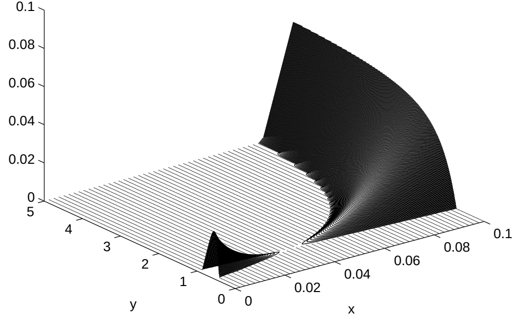


Figure : Value of ρ_1 in dependence of $(v^T y)^2$ (x-axis) and γ^2 (y-axis) with $\Delta = 0.6$, $\|v\|^2 = 0.1$.

Example: ICE versus INE

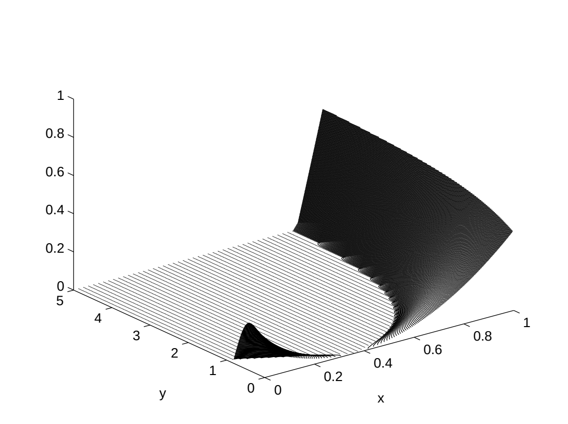


Figure : Value of ρ_1 in dependence of $(v^T y)^2$ (x-axis) and γ^2 (y-axis) with $\Delta = 0.6$, $\|v\|^2 = 1$.

Example: ICE versus INE

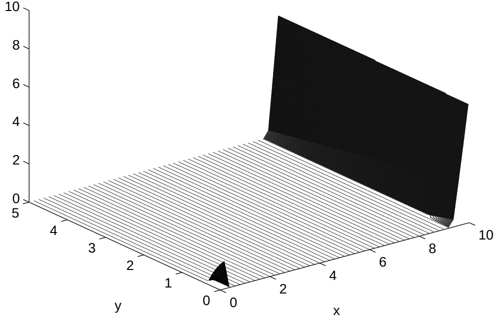


Figure : Value of ρ_1 in dependence of $(v^T y)^2$ (x-axis) and γ^2 (y-axis) with $\Delta = 0.6$, $\|v\|^2 = 10$.

Outline

Comparison 1

Example 1: 50 matrices A=rand(100,100) - rand(100,100), dimension 100, colamd, R from the QR decomposition of A. (Bischof, 1990, Section 4, Test 1).

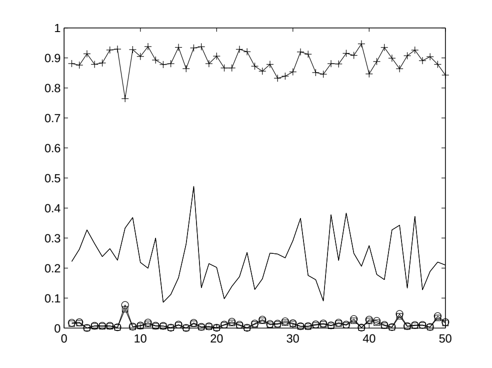
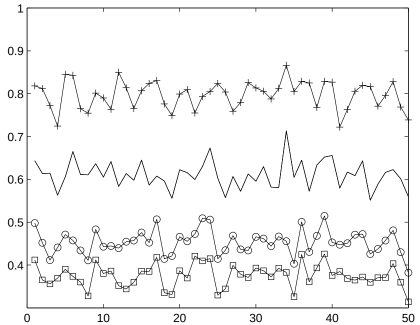


Figure : Ratio of estimate to real condition number for the 50 matrices in example 1. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization

Comparison 2

Example 2: 50 matrices $A = U\Sigma V^T$ of size 100, prescribed condition number κ choosing $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_{100})$ with $\sigma_k = \alpha^k, \quad 1 \leq k \leq 100, \quad \alpha = \kappa^{-\frac{1}{99}}$. U and V: Q factors of the QR factorizations of B=rand(100,100) - rand(100,100), R from the QR decomposition of A with colamd, (Bischof, 1990, Section 4, Test 2; Duff, Vömel, 2002, Section 5, Table 5.4).



Comparison 3

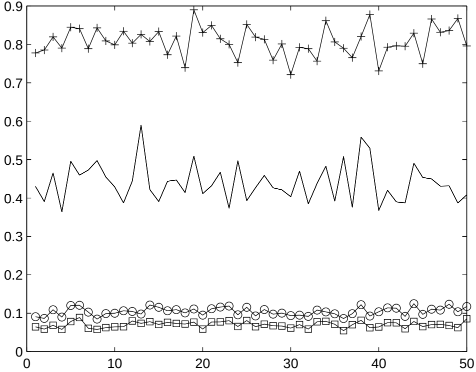


Figure : Ratio of estimate to real condition number for the 50 matrices in example 2 with $\kappa(A) = 100$. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.

Comparison 4

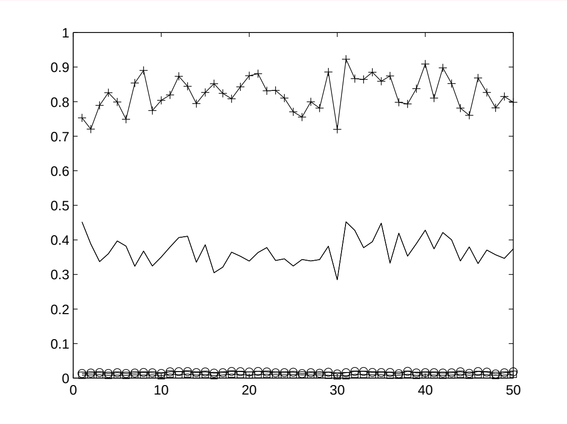


Figure : Ratio of estimate to real condition number for the 50 matrices in example 2 with $\kappa(A) = 1000$. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.

Matrices from MatrixMarket

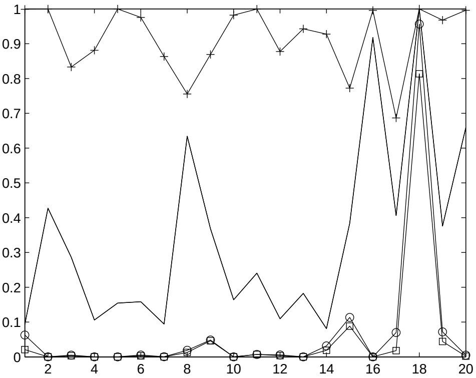


Figure : Ratio of estimate to actual condition number for the 20 matrices from the Matrix Market collection without column pivoting. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.

Matrices from MatrixMarket

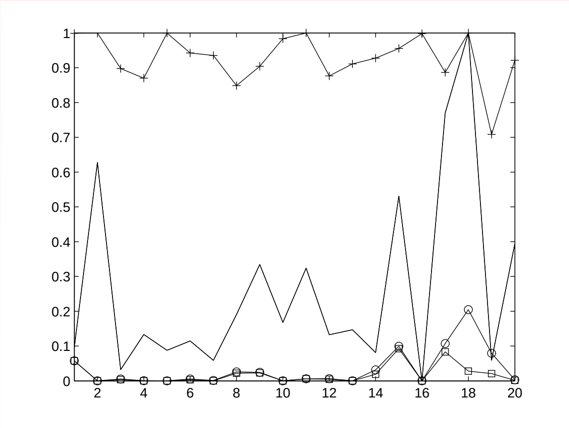


Figure : Ratio of estimate to actual condition number for the 20 matrices from the Matrix Market collection with column pivoting. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.

Matrices from MatrixMarket: Estimates/"Exact" κ

No	Name	dim.	nnz	ICE (org)	INE (orig)	INE (max)	INE (min)
1	494_bus	494	1666	0.09	0.06	0.99	0.02
1	(colamd)	494	1666	0.09	0.06	1	0.057
2	arc130	130	1037	0.42	4e-06	1	9e-10
2	(colamd)	130	1037	0.63	5e-06	1	5e-6
3	bfw398a	398	3678	0.29	0.005	0.83	0.004
3	(colamd)	398	3678	0.03	0.005	0.9	0.004
4	cavity04	317	5923	0.11	1e-4	0.88	3e-5
4	(colamd)	317	5923	0.13	5e-4	0.87	7e-6
5	ck400	400	2860	0.15	9e-5	0.99	8e-5
5	(colamd)	400	2860	0.09	2e-4	1	2e-5
6	dwa512	512	2480	0.16	0.005	0.97	0.003
6	(colamd)	512	2480	0.11	0.005	0.94	0.003
7	e05r0400	236	5846	0.09	5e-4	0.86	1e-4
7	(colamd)	236	5846	0.06	0.001	0.94	3e-4
8	fidap001	216	4339	0.63	0.02	0.76	0.01
8	(colamd)	216	4339	0.19	0.03	0.85	0.02
9	gre___343	343	1310	0.37	0.05	0.87	0.05
9	(colamd)	343	1310	0.33	0.025	0.9	0.023
10	impcol b	59	271	0.16	2e-4	0.98	5e-5
10	(colamd)	59	271	0.17	2e-4	0.98	5e-5

Outline

Conclusions

- Incremental condition estimators in the 2-norm discussed.

Conclusions

- Incremental condition estimators in the 2-norm discussed.
- The two main strategies are inherently different - confirmed both theoretically and experimentally.

Conclusions

- Incremental condition estimators in the 2-norm discussed.
- The two main strategies are inherently different - confirmed both theoretically and experimentally.
- INE strategy using both the direct and inverse factor is a method of choice yielding a highly accurate 2-norm estimator.

Conclusions

- Incremental condition estimators in the 2-norm discussed.
- The two main strategies are inherently different - confirmed both theoretically and experimentally.
- INE strategy using both the direct and inverse factor is a method of choice yielding a highly accurate 2-norm estimator.
- Future work: block algorithm, using the estimator inside a incomplete decomposition.

Many thanks to Gérard Meurant.

Last but not least

Thank you for your attention!

Last but not least

Thank you for your attention!

Last but not least

Thank you for your attention!

Last but not least

Thank you for your attention!