A way to improve incremental 2-norm condition estimation

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

26th Biennial Conference of Numerical Analysis,
June 23, Glasgow, 2015

The condition number

The matrix condition number: an important quantity in matrix theory and computations. We consider square nonsingular matrices:

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

The condition number

The matrix condition number: an important quantity in matrix theory and computations. We consider square nonsingular matrices:

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

The condition number is used, among others, to

- assess the quality of computed solutions
- estimate the sensitivity to perturbations
- monitor and control adaptive computational processes.

Introduction: Estimators for adaptive processes

- Applications involving adaptive computational processes include: adaptive filters, recursive least-squares, ACE for multilevel PDE solvers.

Introduction: Estimators for adaptive processes

- Applications involving adaptive computational processes include: adaptive filters, recursive least-squares, ACE for multilevel PDE solvers.
- We are particularly interested in the adaptive process of incomplete LU factorization using dropping and pivoting.

Introduction: Estimators for adaptive processes

- Applications involving adaptive computational processes include: adaptive filters, recursive least-squares, ACE for multilevel PDE solvers.
- We are particularly interested in the adaptive process of incomplete LU factorization using dropping and pivoting.
- It is important to monitor the condition number of the submatrices that are computed progressively in the incomplete factorization process:

Introduction: Estimators for adaptive processes

- Applications involving adaptive computational processes include: adaptive filters, recursive least-squares, ACE for multilevel PDE solvers.
- We are particularly interested in the adaptive process of incomplete LU factorization using dropping and pivoting.
- It is important to monitor the condition number of the submatrices that are computed progressively in the incomplete factorization process:
- If A is incompletely factorized as

  $$A \approx LU,$$

  then the preconditioned matrix is, e.g.

  $$L^{-1}AU^{-1}$$

  (or $AU^{-1}L^{-1}$ or other variants) and the norms of L^{-1} and U^{-1} directly influence the stability of the preconditioned system.

Introduction: Dropping rules

- In fact, it has been shown that dropping rules based on the sizes of L^{-1} and U^{-1} lead, with appropriate pivoting, to robust ILU methods [Bollhöfer 2001, 2003, Bollhöfer & Saad 2006].

Introduction: Dropping rules

- In fact, it has been shown that dropping rules based on the sizes of L^{-1} and U^{-1} lead, with appropriate pivoting, to robust ILU methods [Bollhöfer 2001, 2003, Bollhöfer & Saad 2006].
- More precisely, dropping of new entries for the kth leading submatrix constructed in the ILU process is done according to the rule

$$|L_{jk}| \cdot \|e_k^T L^{-1}\|_\infty \leq \tau,$$

  and similarly for new entries of U.

Introduction: Dropping rules

- In fact, it has been shown that dropping rules based on the sizes of L^{-1} and U^{-1} lead, with appropriate pivoting, to robust ILU methods [Bollhöfer 2001, 2003, Bollhöfer & Saad 2006].
- More precisely, dropping of new entries for the kth leading submatrix constructed in the ILU process is done according to the rule
$$|L_{jk}| \cdot \|e_k^T L^{-1}\|_\infty \leq \tau,$$
and similarly for new entries of U.
- The information the size of $\|e_k^T L^{-1}\|_\infty$ is obtained by using a cheap condition estimator for the ∞-norm.

Introduction: Dropping rules

- In fact, it has been shown that dropping rules based on the sizes of L^{-1} and U^{-1} lead, with appropriate pivoting, to robust ILU methods [Bollhöfer 2001, 2003, Bollhöfer & Saad 2006].
- More precisely, dropping of new entries for the kth leading submatrix constructed in the ILU process is done according to the rule

$$|L_{jk}| \cdot \|e_k^T L^{-1}\|_\infty \leq \tau,$$

  and similarly for new entries of U.
- The information the size of $\|e_k^T L^{-1}\|_\infty$ is obtained by using a cheap condition estimator for the ∞-norm.
- In recently introduced mixed direct/inverse decomposition methods called Balanced Incomplete Factorization (BIF) for Cholesky (or LDU) decomposition [Bru & Marín & Mas & Tůma 2008, 2010] similar dropping rules are used, but in this type of incomplete decomposition the inverse triangular factors are available as a by-product of the factorization process.

Introduction: Dropping rules

- In the mixed direct/inverse BIF method, the main idea is to balance the growth of both the direct and the inverse factors by exploiting the natural relation between the dropping rules

$$|L_{jk}| \cdot \|e_k^T L^{-1}\|_\infty \le \tau, \quad |L_{jk}^{-1}| \cdot \|e_k^T L\|_\infty \le \tau,$$

and similarly for U and U^{-1}.

Introduction: Dropping rules

- In the mixed direct/inverse BIF method, the main idea is to balance the growth of both the direct and the inverse factors by exploiting the natural relation between the dropping rules

$$|L_{jk}| \cdot \|e_k^T L^{-1}\|_\infty \leq \tau, \quad |L_{jk}^{-1}| \cdot \|e_k^T L\|_\infty \leq \tau,$$

  and similarly for U and U^{-1}.

- But if the inverses of the triangular factors are available, perhaps even more robust dropping rules can be obtained from information on the size of the entire submatrix $\|L_k^{-1}\|$ instead of its kth row $\|e_k^T L^{-1}\|$.

Introduction: Dropping rules

- In the mixed direct/inverse BIF method, the main idea is to balance the growth of both the direct and the inverse factors by exploiting the natural relation between the dropping rules

$$|L_{jk}| \cdot \|e_k^T L^{-1}\|_\infty \le \tau, \quad |L_{jk}^{-1}| \cdot \|e_k^T L\|_\infty \le \tau,$$

  and similarly for U and U^{-1}.

- But if the inverses of the triangular factors are available, perhaps even more robust dropping rules can be obtained from information on the size of the entire submatrix $\|L_k^{-1}\|$ instead of its kth row $\|e_k^T L^{-1}\|$.

- In this talk we present a relatively accurate 2-norm condition estimator which is very suited for use during incomplete factorization and which assumes that inverses of triangular factors are available (or can be computed cheaply).

Introduction: Incremental estimation

Traditionally, 2-norm condition number estimators assume a triangular decomposition and compute estimates for the factors.

Introduction: Incremental estimation

Traditionally, 2-norm condition number estimators assume a triangular decomposition and compute estimates for the factors. E.g., if A is symmetric positive definite with Cholesky decomposition

$$A = R^T R,$$

then the condition number of A satisfies

$$\kappa(A) = \kappa(R)^2 = \kappa(R^T)^2.$$

Introduction: Incremental estimation

Traditionally, 2-norm condition number estimators assume a triangular decomposition and compute estimates for the factors. E.g., if A is symmetric positive definite with Cholesky decomposition

$$A = R^T R,$$

then the condition number of A satisfies

$$\kappa(A) = \kappa(R)^2 = \kappa(R^T)^2.$$

$\kappa(R)$ can be cheaply estimated with a technique called *incremental* condition number estimation, which is suited for incomplete factorization.

Introduction: Incremental estimation

Traditionally, 2-norm condition number estimators assume a triangular decomposition and compute estimates for the factors. E.g., if A is symmetric positive definite with Cholesky decomposition

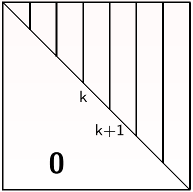
$$A = R^T R,$$

then the condition number of A satisfies

$$\kappa(A) = \kappa(R)^2 = \kappa(R^T)^2.$$

$\kappa(R)$ can be cheaply estimated with a technique called *incremental* condition number estimation, which is suited for incomplete factorization. Main idea: Subsequent estimation for all principal leading submatrices:

$R =$



$\Rightarrow$ every column is accessed only once.

Introduction: Motivation

- The introduction of incremental techniques by Bischof in 1990 was a milestone for 2-norm estimators.

Introduction: Motivation

- The introduction of incremental techniques by Bischof in 1990 was a milestone for 2-norm estimators.
- Other papers on incremental condition estimation include [Bischof 1991], [Bischof & Pierce & Lewis 1990], [Bischof & Tang 1992], [Ferng & Golub & Plemmons 1991], [Pierce & Plemmons 1992], [Stewart 1998], [Duff & Vömel 2002].

Introduction: Motivation

- The introduction of incremental techniques by Bischof in 1990 was a milestone for 2-norm estimators.

- Other papers on incremental condition estimation include [Bischof 1991], [Bischof & Pierce & Lewis 1990], [Bischof & Tang 1992], [Ferng & Golub & Plemmons 1991], [Pierce & Plemmons 1992], [Stewart 1998], [Duff & Vömel 2002].

- The starting point for our method: the methods by Bischof (1990) (incremental condition number estimation - ICE, denoted with a superscript C) and Duff, Vömel (2002) (incremental norm estimation - INE, denoted with a superscript N).

ICE - Bischof (1990)

Consider two leading principal submatrices R and $\hat{R}$ such that

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

ICE - Bischof (1990)

Consider two leading principal submatrices R and $\hat{R}$ such that

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

Let the SVD of R be

$$R = U\Sigma V^T,$$

then with a left minimum singular vector u_-, clearly

$$\|u_-^T R\| = \|u_-^T U \Sigma V^T\| = \sigma_-(R).$$

ICE - Bischof (1990)

Consider two leading principal submatrices R and $\hat{R}$ such that

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

Let the SVD of R be

$$R = U\Sigma V^T,$$

then with a left minimum singular vector u_-, clearly

$$\|u_-^T R\| = \|u_-^T U\Sigma V^T\| = \sigma_-(R).$$

Bischof (1990): If y_- is an approximate left minimum singular vector, then

$$\|y_-^T R\| \equiv \sigma_-^C(R) \approx \sigma_-(R)$$

ICE - Bischof (1990)

Consider two leading principal submatrices R and $\hat{R}$ such that

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

Let the SVD of R be

$$R = U\Sigma V^T,$$

then with a left minimum singular vector u_-, clearly

$$\|u_-^T R\| = \|u_-^T U\Sigma V^T\| = \sigma_-(R).$$

Bischof (1990): If y_- is an approximate left minimum singular vector, then

$$\|y_-^T R\| \equiv \sigma_-^C(R) \approx \sigma_-(R)$$

and we get an incremented approximate left minimum singular vector $\hat{y}_-$ for $\hat{R}$ from y_- putting

$$\|\hat{y}_-^T \hat{R}\| = \mathrm{min}_{s^2+c^2=1} \left\| \begin{bmatrix} s\, y_-^T, & c \end{bmatrix} \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix} \right\|.$$

ICE - Bischof (1990)

This minimization problem is easily solved by taking s and c as the entries of the eigenvector corresponding to the minimum eigenvalue of

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

ICE - Bischof (1990)

This minimization problem is easily solved by taking s and c as the entries of the eigenvector corresponding to the minimum eigenvalue of

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

Then the incremented estimate for $\hat{R}$ is defined as

$$\|\hat{y}_-^T \hat{R}\| = \|[\ s\, y_-^T, \quad c\ ]\hat{R}\| \equiv \sigma_-^C(\hat{R}) \approx \sigma_-(\hat{R}).$$

ICE - Bischof (1990)

This minimization problem is easily solved by taking s and c as the entries of the eigenvector corresponding to the minimum eigenvalue of

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

Then the incremented estimate for $\hat{R}$ is defined as

$$\|\hat{y}_-^T \hat{R}\| = \|[\ s\, y_-^T, \quad c\ ]\hat{R}\| \equiv \sigma_-^C(\hat{R}) \approx \sigma_-(\hat{R}).$$

To find an estimate for $\sigma_+(\hat{R})$ one applies the same technique but starting with an approximate left maximum singular vector y_+ and incrementing it using the maximum eigenvector of

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

INE - Duff, Vömel (2002)

Considering again

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

Duff and Vömel (2002) compute estimates to extremal (minimum or maximum) singular values and right singular vectors: Starting from

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

INE - Duff, Vömel (2002)

Considering again

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

Duff and Vömel (2002) compute estimates to extremal (minimum or maximum) singular values and right singular vectors: Starting from

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

$$\|\hat{R}\hat{z}_{ext}\| = \mathsf{opt}_{s^2+c^2=1} \left\| \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} s\, z_{ext} \\ c \end{bmatrix} \right\| .$$

INE - Duff, Vömel (2002)

Considering again

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

Duff and Vömel (2002) compute estimates to extremal (minimum or maximum) singular values and right singular vectors: Starting from

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

$$\|\hat{R}\hat{z}_{ext}\| = \mathsf{opt}_{s^2+c^2=1} \left\| \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} s\, z_{ext} \\ c \end{bmatrix} \right\| .$$

Again, s and c are the components of the eigenvector corresponding to the extremal (minimum or maximum) eigenvalue of

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

ICE versus INE

- In both ICE and INE the main computational costs come from forming inner products needed to define the size two matrices whose eigenvectors are needed.

ICE versus INE

- In both ICE and INE the main computational costs come from forming inner products needed to define the size two matrices whose eigenvectors are needed.
- This gives for both ICE and INE computational costs of the order n^2 to estimate the condition number of a dense uppper triangular matrix of size n.

ICE versus INE

- In both ICE and INE the main computational costs come from forming inner products needed to define the size two matrices whose eigenvectors are needed.
- This gives for both ICE and INE computational costs of the order n^2 to estimate the condition number of a dense uppper triangular matrix of size n.
- Based on their definitions, it is very hard to guess which technique will perform better.

ICE versus INE

- In both ICE and INE the main computational costs come from forming inner products needed to define the size two matrices whose eigenvectors are needed.
- This gives for both ICE and INE computational costs of the order n^2 to estimate the condition number of a dense uppper triangular matrix of size n.
- Based on their definitions, it is very hard to guess which technique will perform better.
- For dense matrices ICE seems to be superior in general, but INE has been advocated for sparse matrices.

ICE versus INE

- In both ICE and INE the main computational costs come from forming inner products needed to define the size two matrices whose eigenvectors are needed.
- This gives for both ICE and INE computational costs of the order n^2 to estimate the condition number of a dense uppper triangular matrix of size n.
- Based on their definitions, it is very hard to guess which technique will perform better.
- For dense matrices ICE seems to be superior in general, but INE has been advocated for sparse matrices.
- But if we need only estimates of the maximum singular value $\sigma_+(R)$, INE usually does better. This is why INE is called incremental *norm* estimation.

Experiment

- We generate 50 random matrices B of dimension 100 with the Matlab command $\mathsf{B} = \mathsf{randn}(100, 100)$

Experiment

- We generate 50 random matrices B of dimension 100 with the Matlab command $\mathsf{B} = \mathsf{randn}(100, 100)$
- We compute the Cholesky decompositions $R^T R$ of the 50 symmetric positive definite matrices $A = BB^T + I_{100}$

Experiment

- We generate 50 random matrices B of dimension 100 with the Matlab command $\mathsf{B} = \mathsf{randn}(100, 100)$
- We compute the Cholesky decompositions $R^T R$ of the 50 symmetric positive definite matrices $A = BB^T + I_{100}$
- We compute the estimations $\sigma_+^C(R)$ and $\sigma_-^C(R)$

Experiment

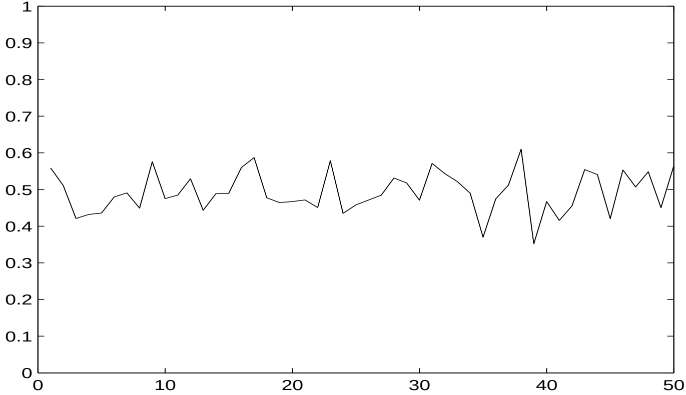
- We generate 50 random matrices B of dimension 100 with the Matlab command $\mathsf{B} = \mathsf{randn}(100, 100)$
- We compute the Cholesky decompositions $R^T R$ of the 50 symmetric positive definite matrices $A = BB^T + I_{100}$
- We compute the estimations $\sigma_+^C(R)$ and $\sigma_-^C(R)$
- In the following graph we display the quality of the estimations through the number

$$\frac{\left(\frac{\sigma_+^C(R)}{\sigma_-^C(R)}\right)^2}{\kappa(A)},$$

where $\kappa(A)$ is the true condition number. Note that we always have

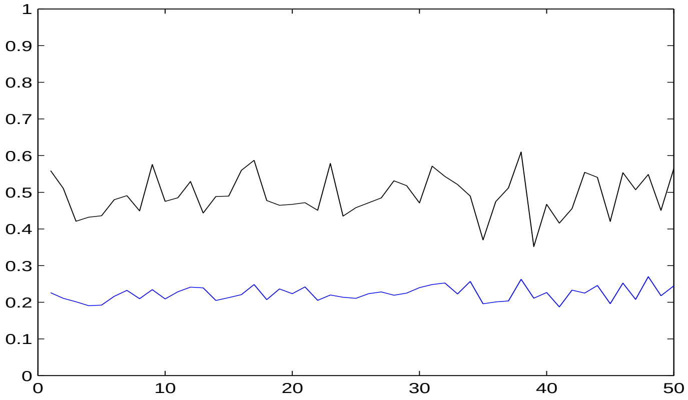
$$\left(\frac{\sigma_+^C(R)}{\sigma_-^C(R)}\right)^2 \leq \kappa(A).$$

Experiment with ICE



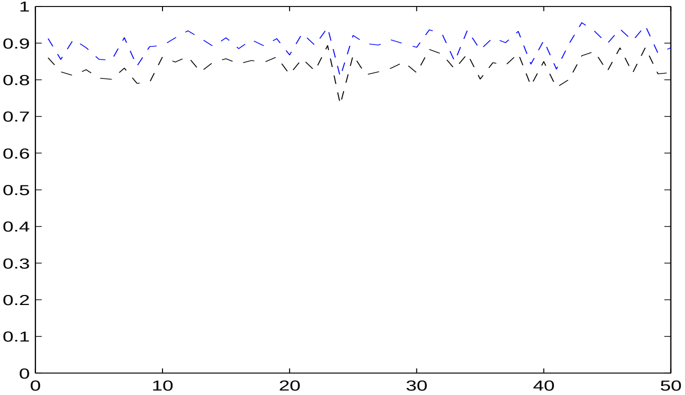
Quality of the estimator ICE for 50 random upper triangular matrices of dimension 100.

Experiment with ICE and INE



Quality of the estimator ICE (black) and of the estimator INE (blue) for 50 random upper triangular matrices of dimension 100.

Experiment with ICE and INE: Only norm estimates



Quality of the ICE technique used to estimate the largest singular value (black) and of the INE technique used to estimate the largest singular value (blue) for 50 random upper triangular matrices of dimension 100.

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

We now assume we have both R and R^{-1}

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

We now assume we have both R and R^{-1}

Then we can for instance run ICE on R^{-1} and use the additional estimations

$$\frac{1}{\sigma_+^C(R^{-1})} \approx \sigma_-(R), \qquad \frac{1}{\sigma_-^C(R^{-1})} \approx \sigma_+(R).$$

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

We now assume we have both R and R^{-1}

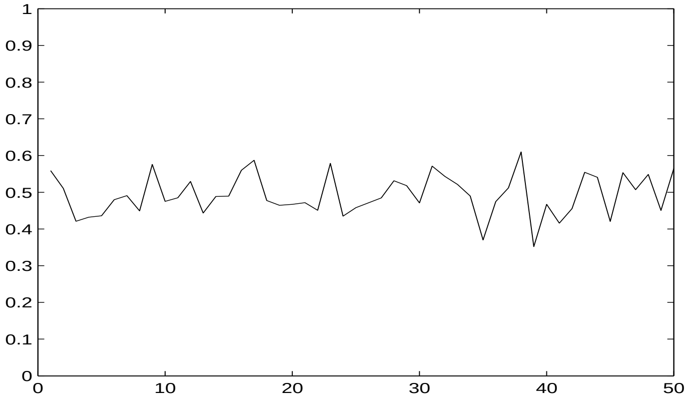
Then we can for instance run ICE on R^{-1} and use the additional estimations

$$\frac{1}{\sigma_+^C(R^{-1})} \approx \sigma_-(R), \qquad \frac{1}{\sigma_-^C(R^{-1})} \approx \sigma_+(R).$$

In the following graph we use the same data as before and take the best of both estimations, we display

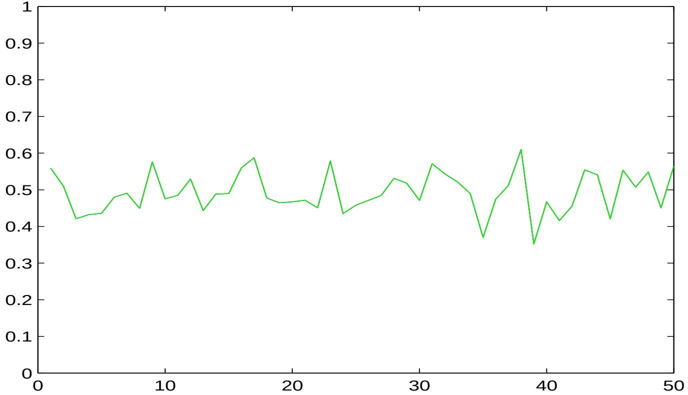
$$\frac{\left(\frac{\max(\sigma_+^C(R),\sigma_-^C(R^{-1})^{-1})}{\min(\sigma_-^C(R),\sigma_+^C(R^{-1})^{-1})}\right)^2}{\kappa(A)}.$$

Experiment with ICE



Quality of the estimator ICE for 50 random upper triangular matrices of dimension 100.

Experiment with ICE when both direct and inverse factors



Quality of the estimator ICE without (black) and with exploiting the inverse (green).

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

Theorem

Computing the inverse factor R^{-1} in addition to R **does not give any improvement for ICE***:*

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

Theorem

Computing the inverse factor R^{-1} in addition to R **does not give any improvement for ICE***:* *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: ICE

Theorem

Computing the inverse factor R^{-1} in addition to R **does not give any improvement for ICE***: 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

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

Similarly, one can prove $\sigma_+^C(R) = 1/\sigma_-^C(R^{-1})$.

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} are exact:*

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

Then the INE estimates of the singular values related to the incremented 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

An analogue of the previous theorem for estimates of the maximum singular value shows that

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

In this sense, for INE maximization performs better than minimization.

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

An analogue of the previous theorem for estimates of the maximum singular value shows that

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

In this sense, for INE maximization performs better than minimization.

In case the assumption is relaxed to $1/\sigma_+^N(R^{-1}) \leq \sigma_-^N(R)$ we obtain a rather technical theorem, saying essentially that maximization with R^{-1} is in most cases superior to minimization with R.

Small example: ICE and INE with and without inverse

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

Small example: ICE and INE with and without inverse

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

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

Small example: ICE and INE with and without inverse

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

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

Small example: ICE and INE with and without inverse

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

$\sigma_-^C(R) = 1 = 1/\sigma_+^C(R^{-1})$,
$\sigma_-^N(R) = 1$, but $1/\sigma_+^N(R^{-1}) = 0.8944$.

Small example: ICE and INE with and without inverse

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

$\sigma_-^C(R) = 1 = 1/\sigma_+^C(R^{-1})$,
$\sigma_-^N(R) = 1$, but $1/\sigma_+^N(R^{-1}) = 0.8944$.

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

Small example: ICE and INE with and without inverse

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

$\sigma_-^C(R) = 1 = 1/\sigma_+^C(R^{-1})$,
$\sigma_-^N(R) = 1$, but $1/\sigma_+^N(R^{-1}) = 0.8944$.

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

$\sigma_-^C(\hat{R}) = 0.618 = 1/\sigma_+^C(\hat{R}^{-1})$,

Small example: ICE and INE with and without inverse

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

$\sigma_-^C(R) = 1 = 1/\sigma_+^C(R^{-1})$,
$\sigma_-^N(R) = 1$, but $1/\sigma_+^N(R^{-1}) = 0.8944$.

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

$\sigma_-^C(\hat{R}) = 0.618 = 1/\sigma_+^C(\hat{R}^{-1})$,
$\sigma_-^N(\hat{R}) = 0.835$,

Small example: ICE and INE with and without inverse

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

$\sigma_-^C(R) = 1 = 1/\sigma_+^C(R^{-1})$,
$\sigma_-^N(R) = 1$, but $1/\sigma_+^N(R^{-1}) = 0.8944$.

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

$\sigma_-^C(\hat{R}) = 0.618 = 1/\sigma_+^C(\hat{R}^{-1})$,
$\sigma_-^N(\hat{R}) = 0.835$, but $1/\sigma_+^N(\widehat{R^{-1}}) = 0.5381$.

Small example: ICE and INE with and without inverse

However, a counterexample is given by

$$\hat{R} = \begin{bmatrix} 2 & 0 & 1 & 0 \\ & 1 & 0 & 1 \\ & & 1 & 0 \\ & & & 1 \end{bmatrix}, \qquad \sigma_-(\hat{R}) = 0.618$$

Small example: ICE and INE with and without inverse

However, a counterexample is given by

$$\hat{R} = \begin{bmatrix} 2 & 0 & 1 & 0 \\ & 1 & 0 & 1 \\ & & 1 & 0 \\ & & & 1 \end{bmatrix}, \qquad \sigma_{-}(\hat{R}) = 0.618$$

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

Small example: ICE and INE with and without inverse

However, a counterexample is given by

$$\hat{R} = \begin{bmatrix} 2 & 0 & 1 & 0 \\ & 1 & 0 & 1 \\ & & 1 & 0 \\ & & & 1 \end{bmatrix}, \qquad \sigma_-(\hat{R}) = 0.618$$

$\sigma_-^C(\hat{R}) = 1 = 1/\sigma_+^C(\hat{R}^{-1}),$
$\sigma_-^N(\hat{R}) = 0.618,$

Small example: ICE and INE with and without inverse

However, a counterexample is given by

$$\hat{R} = \begin{bmatrix} 2 & 0 & 1 & 0 \\ & 1 & 0 & 1 \\ & & 1 & 0 \\ & & & 1 \end{bmatrix}, \qquad \sigma_-(\hat{R}) = 0.618$$

$\sigma_-^C(\hat{R}) = 1 = 1/\sigma_+^C(\hat{R}^{-1})$,
$\sigma_-^N(\hat{R}) = 0.618$, and $1/\sigma_+^N(\widehat{R^{-1}}) = 0.7071$.

Small example: ICE and INE with and without inverse

However, a counterexample is given by

$$\hat{R} = \begin{bmatrix} 2 & 0 & 1 & 0 \\ & 1 & 0 & 1 \\ & & 1 & 0 \\ & & & 1 \end{bmatrix}, \qquad \sigma_-(\hat{R}) = 0.618$$

$\sigma_-^C(\hat{R}) = 1 = 1/\sigma_+^C(\hat{R}^{-1})$,
$\sigma_-^N(\hat{R}) = 0.618$, and $1/\sigma_+^N(\hat{R^{-1}}) = 0.7071$.

Nevertheless, in *all* performed numerical experiments we found that $\sigma_-^N(\hat{R})$ gives an estimate which is worse than $1/\sigma_+^N(\hat{R}^{-1})$.

Small example: ICE and INE with and without inverse

However, a counterexample is given by

$$\hat{R} = \begin{bmatrix} 2 & 0 & 1 & 0 \\ & 1 & 0 & 1 \\ & & 1 & 0 \\ & & & 1 \end{bmatrix}, \qquad \sigma_-(\hat{R}) = 0.618$$

$\sigma_-^C(\hat{R}) = 1 = 1/\sigma_+^C(\hat{R}^{-1})$,
$\sigma_-^N(\hat{R}) = 0.618$, and $1/\sigma_+^N(\widehat{R^{-1}}) = 0.7071$.

Nevertheless, in *all* performed numerical experiments we found that $\sigma_-^N(\hat{R})$ gives an estimate which is worse than $1/\sigma_+^N(\hat{R}^{-1})$.

We now give a striking example.

An example showing the possible gap between INE with and without using the inverse

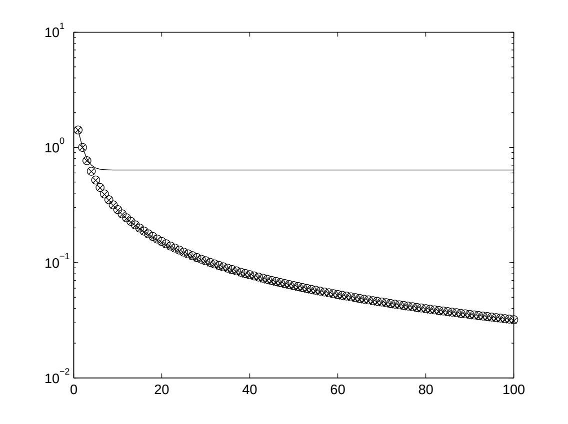


Figure : INE estimation of the smallest singular value of the 1D Laplacians of size one until hundred: INE with minimization (solid line), INE with maximization (circles) and exact minimum singular values (crosses).

An example showing the possible gap between INE with and without using the inverse

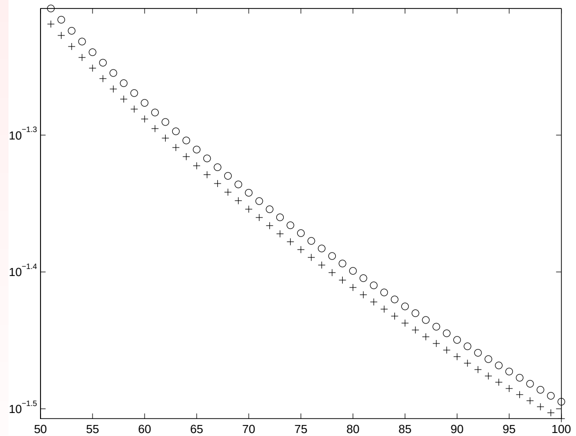
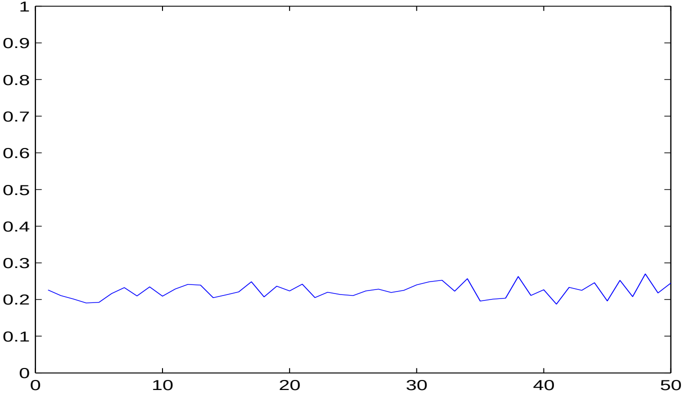


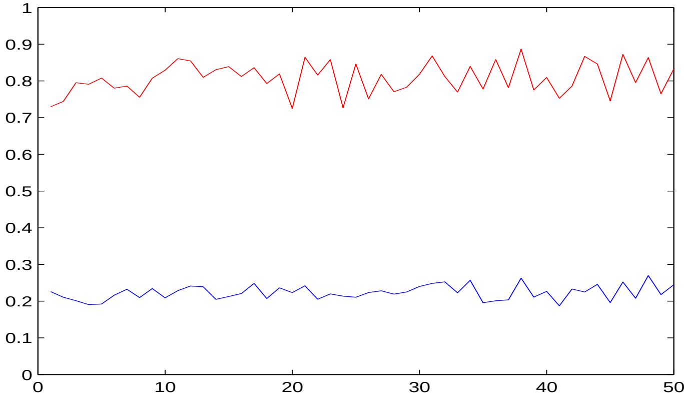
Figure : INE estimation of the smallest singular value of the 1D Laplacians of size fifty until hundred (zoom of previous figure for INE with maximization and exact minimum singular values).

Experiment with INE



Quality of the estimator INE for 50 random upper triangular matrices of dimension 100.

Experiment with INE when both direct and inverse factors available



Quality of the standard INE estimator (blue) and of INE using maximization and R^{-1} to estimate the smallest singular value (red).

Why such an improvement?

- This significant improvement is partly explained by the fact that a moderate improvement of the estimate for $\sigma_{min}(R)$ (from using the inverse) has an important impact because $\sigma_{min}(R)$ is typically small and appears in the denominator in

$$\kappa(R) = \frac{\sigma_{max}(R)}{\sigma_{min}(R)} \approx \frac{\sigma_+^N(R)}{\sigma_-^N(R)}.$$

Why such an improvement?

- This significant improvement is partly explained by the fact that a moderate improvement of the estimate for $\sigma_{min}(R)$ (from using the inverse) has an important impact because $\sigma_{min}(R)$ is typically small and appears in the denominator in

$$\kappa(R) = \frac{\sigma_{max}(R)}{\sigma_{min}(R)} \approx \frac{\sigma_+^N(R)}{\sigma_-^N(R)}.$$

- Similarly, if $\sigma_{min}(R)$ is slightly better estimated with INE than with ICE (expoiting the inverse factor), the improvement for the condition number estimate will be more important.

Why such an improvement?

- This significant improvement is partly explained by the fact that a moderate improvement of the estimate for $\sigma_{min}(R)$ (from using the inverse) has an important impact because $\sigma_{min}(R)$ is typically small and appears in the denominator in

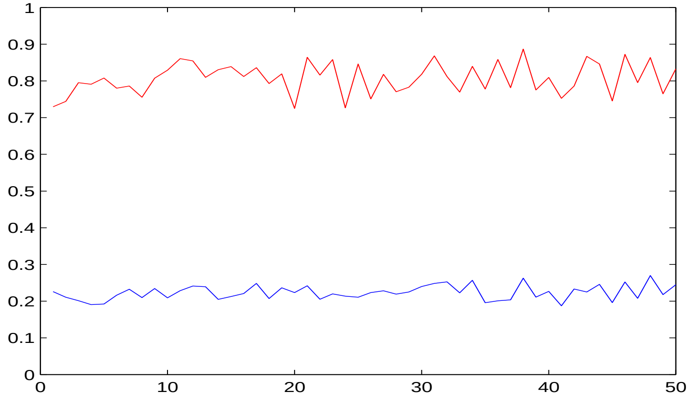
$$\kappa(R) = \frac{\sigma_{max}(R)}{\sigma_{min}(R)} \approx \frac{\sigma_+^N(R)}{\sigma_-^N(R)}.$$

- Similarly, if $\sigma_{min}(R)$ is slightly better estimated with INE than with ICE (expoiting the inverse factor), the improvement for the condition number estimate will be more important.

- This can be expected because we have observed that INE gives better estimates of maximum singular values than ICE, in particular

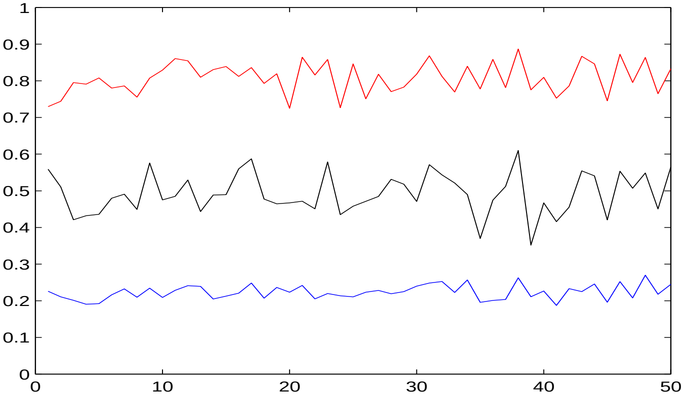
$$1/\sigma_+^N(\hat{R}^{-1}) < 1/\sigma_+^C(\hat{R}^{-1}).$$

Experiment with INE when both direct and inverse factors available



Quality of the standard INE estimator (blue) and of INE using maximization and R^{-1} to estimate the smallest singular value (red).

Experiment with INE and ICE when both direct and inverse factors available



Quality of INE (blue), of INE using maximization and R^{-1} to estimate the smallest singular value (red) and of standard ICE (black).

Short summary

Summarizing, we showed that

- ICE cannot profit from the presence of the inverse factor.

Short summary

Summarizing, we showed that

- ICE cannot profit from the presence of the inverse factor.
- INE can profit from the presence of the inverse factor when it used in a maximization process.

Short summary

Summarizing, we showed that

- ICE cannot profit from the presence of the inverse factor.
- INE can profit from the presence of the inverse factor when it used in a maximization process.
- This does not yet explain why INE using maximization (for the inverse factor) is more powerful than ICE using maximization (for either the direct or the inverse factor). This was observed in the experiments.

Short summary

Summarizing, we showed that

- ICE cannot profit from the presence of the inverse factor.
- INE can profit from the presence of the inverse factor when it used in a maximization process.
- This does not yet explain why INE using maximization (for the inverse factor) is more powerful than ICE using maximization (for either the direct or the inverse factor). This was observed in the experiments.
- We now give theoretical results which make it plausible that INE maximization will tend to perform better than ICE maximization.

A superiority condition for INE

Theorem

Consider norm estimation of the incremented matrix

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

let ICE and INE start with $\sigma_+ \equiv \sigma_+^C(R) = \sigma_+^N(R)$*; let* y *be the ICE approximate LSV,* z *be the INE approximate RSV and* $w = Rz/\sigma^+$*. Then*

$$\sigma_+^N(\hat{R}) \geq \sigma_+^C(\hat{R}) \qquad \text{if} \qquad (v^T w)^2 \geq \rho,$$

where the critical value ρ *is the smaller root of the quadratic equation*

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

Graphical demonstration of potential INE superiority

In the next figures,

- The superiority criterion for INE expressed by the value $\max(\rho, 0)$ is given by the z-axes.

Graphical demonstration of potential INE superiority

In the next figures,

- The superiority criterion for INE expressed by the value $\max(\rho, 0)$ is given by the z-axes.
- Without loss of generality we can assume $\sigma_{+}^{C}(R) = \sigma_{+}^{N}(R) = 1$.

Graphical demonstration of potential INE superiority

In the next figures,

- The superiority criterion for INE expressed by the value $\max(\rho, 0)$ is given by the z-axes.
- Without loss of generality we can assume $\sigma_+^C(R) = \sigma_+^N(R) = 1$.
- Then the coefficients of the quadratic equation depend on the sizes of v, γ and $(v^T y)^2$ only. We fix $v^T v$.

Graphical demonstration of potential INE superiority

In the next figures,

- The superiority criterion for INE expressed by the value $\max(\rho, 0)$ is given by the z-axes.
- Without loss of generality we can assume $\sigma_+^C(R) = \sigma_+^N(R) = 1$.
- Then the coefficients of the quadratic equation depend on the sizes of v, γ and $(v^T y)^2$ only. We fix $v^T v$.
- The x-axes of the following figures represent the possible values of $(v^T y)^2$.

Graphical demonstration of potential INE superiority

In the next figures,

- The superiority criterion for INE expressed by the value $\max(\rho, 0)$ is given by the z-axes.
- Without loss of generality we can assume $\sigma_+^C(R) = \sigma_+^N(R) = 1$.
- Then the coefficients of the quadratic equation depend on the sizes of v, γ and $(v^T y)^2$ only. We fix $v^T v$.
- The x-axes of the following figures represent the possible values of $(v^T y)^2$.
- The y-axes represent values of γ^2, i.e. the square of the new diagonal entry.

Graphical demonstration of potential INE superiority

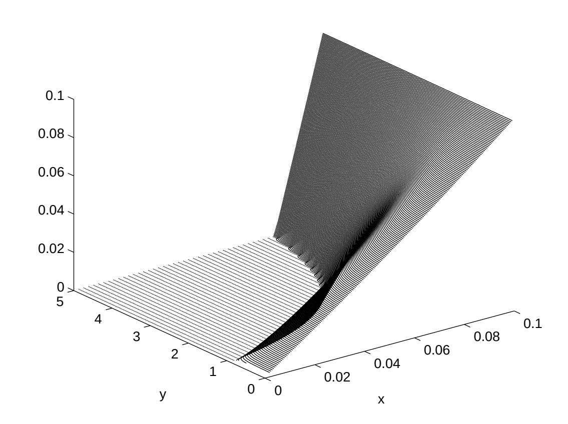


Figure : Critical value ρ in dependence of $(v^T y)^2$ (x-axis) and γ^2 (y-axis) with $\sigma_+ = 1$, $\|v\|^2 = 0.1$.

Graphical demonstration of potential INE superiority

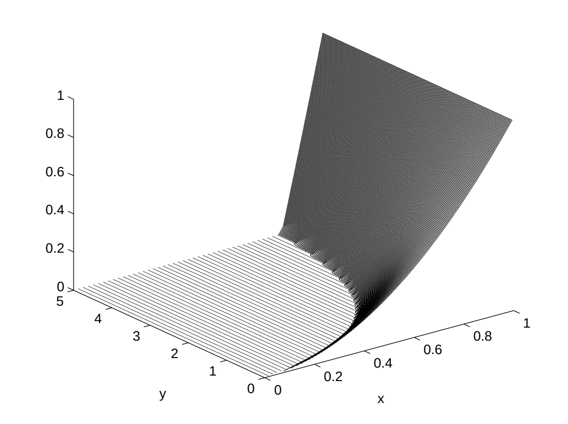


Figure : Critical value ρ in dependence of $(v^T y)^2$ (x-axis) and γ^2 (y-axis) with $\sigma_+ = 1$, $\|v\|^2 = 1$.

Graphical demonstration of potential INE superiority

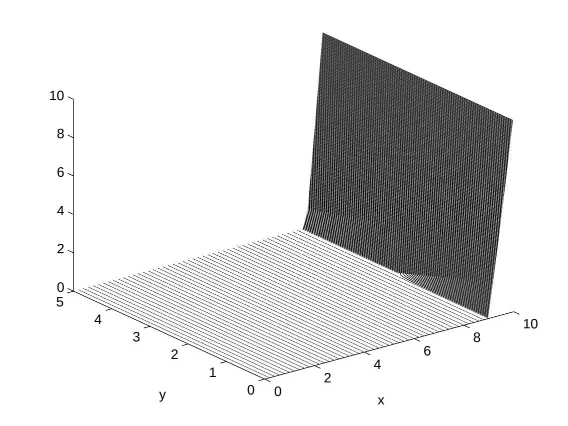


Figure : Critical value ρ in dependence of $(v^T y)^2$ (x-axis) and γ^2 (y-axis) with $\sigma_+ = 1$, $\|v\|^2 = 10$.

Graphical demonstration of potential INE superiority

- The previous theorem can also be formulated when

$$\sigma_+^C(R) \neq \sigma_+^N(R).$$

Graphical demonstration of potential INE superiority

- The previous theorem can also be formulated when

$$\sigma_+^C(R) \neq \sigma_+^N(R).$$

- Let

$$\Delta \equiv \sqrt{(\sigma_+^N)^2 - (\sigma_+^C)^2}, \qquad \sigma_+^N \geq \sigma_+^C.$$

Graphical demonstration of potential INE superiority

- The previous theorem can also be formulated when

$$\sigma_+^C(R) \neq \sigma_+^N(R).$$

- Let

$$\Delta \equiv \sqrt{(\sigma_+^N)^2 - (\sigma_+^C)^2}, \qquad \sigma_+^N \geq \sigma_+^C.$$

- Intuitively we expect $\Delta > 0$ to even increase the potential superiority of INE over ICE.

Graphical demonstration of potential INE superiority

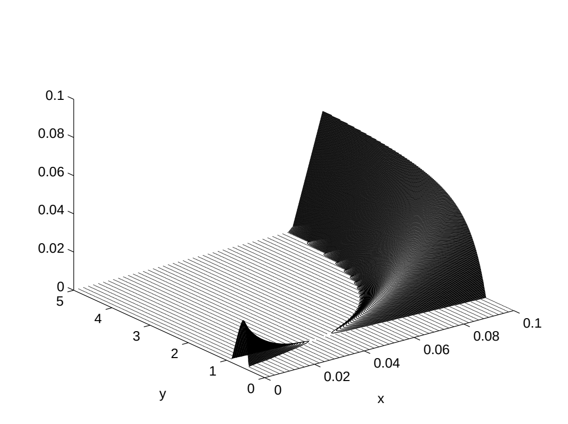


Figure : Critical value ρ in dependence of $(v^T y)^2$ (x-axis) and γ^2 (y-axis) with $\sigma_+ = 1$, $\Delta = 0.6$, $\|v\|^2 = 0.1$.

Graphical demonstration of potential INE superiority

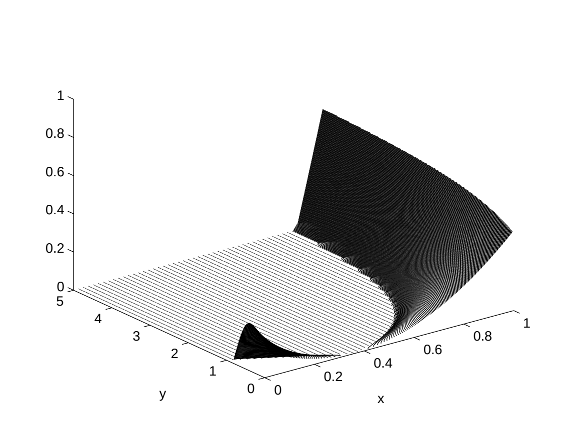


Figure : Critical value ρ in dependence of $(v^T y)^2$ (x-axis) and γ^2 (y-axis) with $\sigma_+ = 1$, $\Delta = 0.6$, $\|v\|^2 = 1$.

Graphical demonstration of potential INE superiority

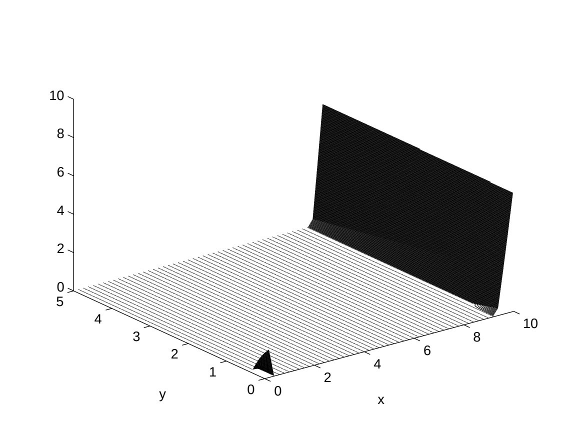


Figure : Critical value ρ in dependence of $(v^T y)^2$ (x-axis) and γ^2 (y-axis) with $\Delta = 0.6$, $\|v\|^2 = 10$.

The compared estimators

We will compare the following estimators:

- The original ICE technique with the estimates defined as

$$\sigma_+^C(R)/\sigma_-^C(R).$$

The compared estimators

We will compare the following estimators:

- The original ICE technique with the estimates defined as

$$\sigma_+^C(R)/\sigma_-^C(R).$$

- The original INE technique with the estimates defined by

$$\sigma_+^N(R)/\sigma_-^N(R).$$

The compared estimators

We will compare the following estimators:

- The original ICE technique with the estimates defined as

$$\sigma_+^C(R)/\sigma_-^C(R).$$

- The original INE technique with the estimates defined by

$$\sigma_+^N(R)/\sigma_-^N(R).$$

- The INE technique based on maximization only, i.e. estimates defined as

$$\sigma_+^N(R)/\left(\sigma_+^N(R^{-1})\right)^{-1}.$$

The compared estimators

We will compare the following estimators:

- The original ICE technique with the estimates defined as
$$\sigma_+^C(R)/\sigma_-^C(R).$$
- The original INE technique with the estimates defined by
$$\sigma_+^N(R)/\sigma_-^N(R).$$
- The INE technique based on maximization only, i.e. estimates defined as
$$\sigma_+^N(R)/\left(\sigma_+^N(R^{-1})\right)^{-1}.$$
- The INE technique based on minimization only which uses the matrix inverse as well, that is
$$\left(\sigma_-^N(R^{-1})\right)^{-1}/\sigma_-^N(R).$$

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].

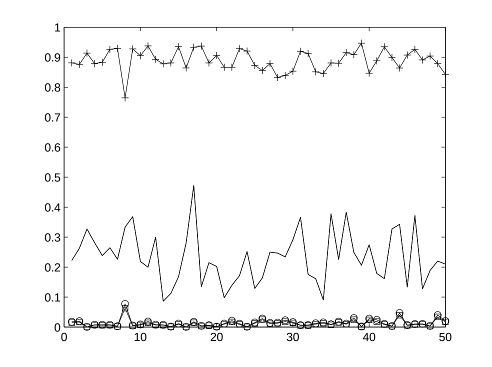


Figure : Ratio of estimate to real condition number for the 50 matrices in example 1. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.

Comparison 2

Example 2: 50 matrices $A = U\Sigma V^T$ of size 100 with 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 are random unitary factors, R from the QR decomposition of A with colamd ([Bischof 1990, Section 4, Test 2], [Duff & Vömel 2002, Section 5, Table 5.4]).

With $\kappa(A) = 10$ we obtain:

Comparison 3

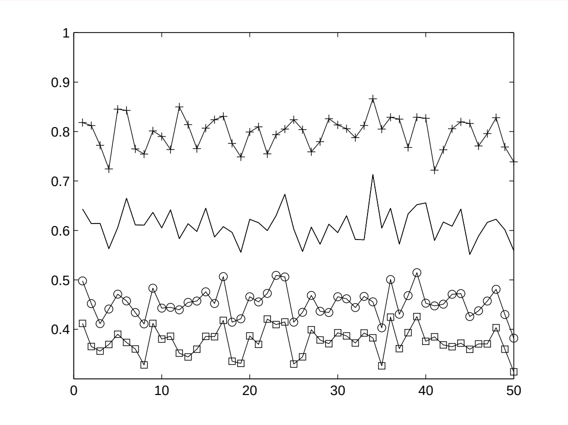


Figure : Ratio of estimate to real condition number for the 50 matrices in example 2 with $\kappa(A) = 10$. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.

Comparison 4

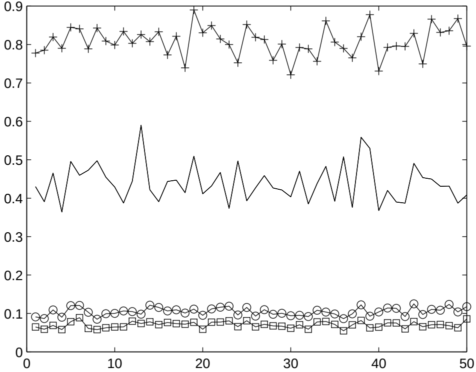


Figure : Ratio of estimate to real condition number for the 50 matrices in example 2 with $\kappa(A) = 100$. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.

Comparison 5

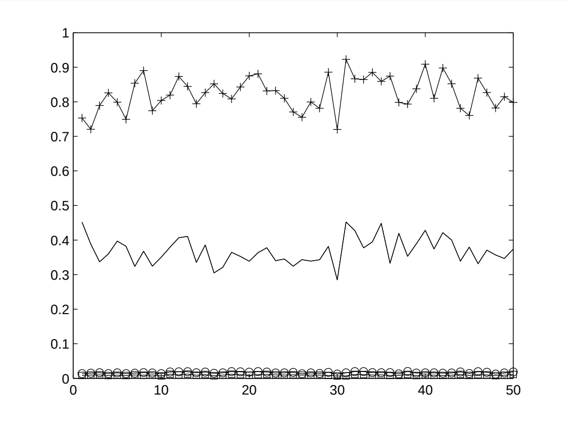


Figure : Ratio of estimate to real condition number for the 50 matrices in example 2 with $\kappa(A) = 1000$. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.

Matrices from MatrixMarket

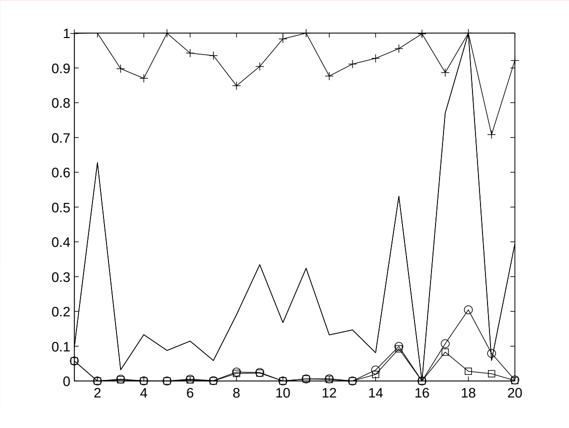


Figure : Ratio of estimate to actual condition number for the 20 matrices from the Matrix Market collection with column pivoting. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.

Conclusions

- The two main 2-norm incremental condition estimators are inherently different - confirmed both theoretically and experimentally.

Conclusions

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

Conclusions

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

Main references

C.H. Bischof: *Incremental Condition Estimation*, SIAM J. Matrix Anal. Appl., vol. 11, pp. 312–322, 1990.

M. Bollhöfer: *A Robust ILU With Pivoting Based on Monitoring the Growth of the Inverse Factors*, Lin. Algebr. Appl., vol. 338, pp. 201–218, 2001.

M. Bollhöfer, Y. Saad: *Multilevel Preconditioners Constructed from Inverse-Based ILU's*, SIAM J. Sci. Comput., vol. 27, pp. 1627–1650, 2006.

I. Duff, Ch. Vömel: *Incremental Norm Estimation for Dense and Sparse Matrices*, BIT, vol. 42, pp. 300–322, 2002.

M. Bollhöfer: *A Robust and Efficient ILU that Incorporates the Growth of the Inverse Triangular Factors*, SIAM J. Sci. Comput., vol. 25, pp. 86–103, 2003.

R. Bru, J. Marín, J. Mas, M. Tůma:*Balanced Incomplete Factorization*, SIAM J. Sci. Comput., vol. 30, pp. 2302–2318, 2008.

R. Bru, J. Marín, J. Mas, M. Tůma:*Improved Balanced Incomplete Factorization*, SIAM J. Matrix Anal. Appl., vol. 31, pp. 2431–2452, 2010.

J. Duintjer Tebbens, M. Tůma: *On Incremental Condition Estimators in the 2-Norm*, SIAM J. Matrix Anal. Appl., vol. 35, no. 1, pp. 174–197, 2014.

Last but not least

Thank you for your attention!