On Incremental 2-norm Condition Estimators

Jurjen Duintjer Tebbens
Institute of Computer Science
Academy of Sciences of the Czech Republic
duintjer@cs.cas.cz

Miroslav Tůma
Institute of Computer Science
Academy of Sciences of the Czech Republic
tuma@cs.cas.cz

GAMM Workshop Applied and Numerical Linear Algebra,
September 9, Wuppertal, 2013
Outline

1. Introduction: The Problem
2. The two strategies
3. ICE and INE with inverse factors
4. INE maximization versus ICE maximization
5. Numerical experiments
6. Conclusions
Matrix condition number: an important quantity used in numerical linear algebra. We consider square nonsingular matrices:

\[ \kappa(A) = \|A\| \cdot \|A^{-1}\| \]
Matrix condition number: an important quantity used in numerical linear algebra. We consider square nonsingular matrices:

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

- Assessing quality of computed solutions
- Estimating sensitivity to perturbations
- Monitor and control adaptive computational processes.
Matrix condition number: an important quantity used in numerical linear algebra. We consider square nonsingular matrices:

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

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

Here: \emph{A upper triangular} (no loss of generality - computations typically based on triangular decomposition)

- Euclidean norm
Introduction: Earlier work

- Turing (1948); Wilkinson (1961)
- Gragg, Stewart (1976); Cline, Moler, Stewart, Wilkinson (1979); Cline, Conn, van Loan (1982); van Loan (1987)
Introduction: Earlier work

- Turing (1948); Wilkinson (1961)
- Gragg, Stewart (1976); Cline, Moler, Stewart, Wilkinson (1979); Cline, Conn, van Loan (1982); van Loan (1987)
- Condition number estimation is part of standard libraries as LAPACK
Introduction: Earlier work

- Turing (1948); Wilkinson (1961)
- Gragg, Stewart (1976); Cline, Moler, Stewart, Wilkinson (1979); Cline, Conn, van Loan (1982); van Loan (1987)
- Condition number estimation is part of standard libraries as LAPACK
- Typically estimating lower bound for $\kappa(A)$ (note that it is often sufficient to have the estimates within a reasonable multiplicative factor from the exact $\kappa(A)$ - Demmel (1997))
Introduction: Earlier work

- Turing (1948); Wilkinson (1961)
- Gragg, Stewart (1976); Cline, Moler, Stewart, Wilkinson (1979); Cline, Conn, van Loan (1982); van Loan (1987)
- Condition number estimation is part of standard libraries as LAPACK
- Typically estimating lower bound for $\kappa(A)$ (note that it is often sufficient to have the estimates within a reasonable multiplicative factor from the exact $\kappa(A)$ - Demmel (1997))
Introduction: Earlier work

- Turing (1948); Wilkinson (1961)
- Gragg, Stewart (1976); Cline, Moler, Stewart, Wilkinson (1979); Cline, Conn, van Loan (1982); van Loan (1987)
- Condition number estimation is part of standard libraries as LAPACK
- Typically estimating lower bound for $\kappa(A)$ (note that it is often sufficient to have the estimates within a reasonable multiplicative factor from the exact $\kappa(A)$ - Demmel (1997))
Introduction: Earlier work

- Turing (1948); Wilkinson (1961)
- Gragg, Stewart (1976); Cline, Moler, Stewart, Wilkinson (1979); Cline, Conn, van Loan (1982); van Loan (1987)
- Condition number estimation is part of standard libraries as LAPACK
- Typically estimating lower bound for $\kappa(A)$ (note that it is often sufficient to have the estimates within a reasonable multiplicative factor from the exact $\kappa(A)$ - Demmel (1997))
- See also other techniques in various applications: adaptive filters, recursive least-squares, ACE for multilevel PDE solvers.
Motivated by new mixed direct/inverse decomposition methods (see Bru et al, 2008, 2010), we initially wondered whether presence of the matrix inverse can improve incremental condition number estimation.
Motivated by new mixed direct/inverse decomposition methods (see Bru et al, 2008, 2010), we initially wondered whether presence of the matrix inverse can improve incremental condition number estimation. We did find more accurate estimation techniques but the theoretical explanation is not straightforward. As a by-product we have some enhanced insight in incremental condition number estimation methods in the 2-norm.
Motivated by new mixed direct/inverse decomposition methods (see Bru et al, 2008, 2010), we initially wondered whether presence of the matrix inverse can improve incremental condition number estimation.

We did find more accurate estimation techniques but the theoretical explanation is not straightforward. As a by-product we have some enhanced insight in incremental condition number estimation methods in the 2-norm.

An immediate application is dropping and pivoting in preconditioner computation (see Bollhöfer, Saad (2001 - 2006)).
Motivated by new mixed direct/inverse decomposition methods (see Bru et al, 2008, 2010), we initially wondered whether presence of the matrix inverse can improve incremental condition number estimation.

We did find more accurate estimation techniques but the theoretical explanation is not straightforward. As a by-product we have some enhanced insight in incremental condition number estimation methods in the 2-norm.

An immediate application is dropping and pivoting in preconditioner computation (see Bollhöfer, Saad (2001 - 2006)).

Outline

1 Introduction: The Problem

2 The two strategies

3 ICE and INE with inverse factors

4 INE maximization versus ICE maximization

5 Numerical experiments

6 Conclusions
\( \hat{R} = \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix} \)
\[ \hat{R} = \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix} \]

Using a left extremal (minimum or maximum) singular vector \( u_{ext} \), if \( R = U \Sigma V^T \) \( \Rightarrow \| u_{ext}^T R \| = \| u_{ext}^T U \Sigma V^T \| = \sigma_{ext}(R) \).
Using a left extremal (minimum or maximum) singular vector $u_{\text{ext}}$, if $R = U \Sigma V^T \Rightarrow \| u_{\text{ext}}^T R \| = \| u_{\text{ext}}^T U \Sigma V^T \| = \sigma_{\text{ext}}(R)$.

Bischof (1990): estimates to extremal singular values and left singular vectors:

$$
\sigma^C_{\text{ext}}(R) = \| y_{\text{ext}}^T R \| \approx \sigma_{\text{ext}}(R),
\| \hat{y}_{\text{ext}}^T \hat{R} \| = \text{ext}_{\|[s,c]\|=1} \left\| \begin{bmatrix} s y_{\text{ext}}^T & c \end{bmatrix} \begin{bmatrix} R & \nu \\ 0 & \gamma \end{bmatrix} \right\|.
$$
\[ \hat{R} = \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix} \]

- Using a left extremal (minimum or maximum) singular vector \( u_{\text{ext}} \), if \( R = U\Sigma V^T \Rightarrow \| u_{\text{ext}}^T R \| = \| u_{\text{ext}}^T U\Sigma V^T \| = \sigma_{\text{ext}}(R) \).

- Bischof (1990): estimates to extremal singular values and left singular vectors:

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

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

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

\[
B_{\text{ext}}^C \equiv \begin{bmatrix} \sigma_{\text{ext}}^C(R)^2 + (y_{\text{ext}}^T v)^2 & \gamma(y_{\text{ext}}^T v) \\ \gamma(y_{\text{ext}}^T v) & \gamma^2 \end{bmatrix}.
\]
\[
\hat{R} = \begin{bmatrix}
    R & v \\
    0 & \gamma
\end{bmatrix}
\]
Duff, Vömel (2002): estimates to extremal singular values and right singular vectors (originally used only to estimate the 2-norm). INE computes
\[ \sigma_{ext}^N(R) = \| R z_{ext} \| \approx \sigma_{ext}(R) \]
\[
\hat{R} = \begin{bmatrix}
R & v \\
0 & \gamma
\end{bmatrix}
\]

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

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

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

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

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

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

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

1. Introduction: The Problem
2. The two strategies
3. ICE and INE with inverse factors
4. INE maximization versus ICE maximization
5. Numerical experiments
6. Conclusions
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)

- Estimation of $\sigma-(R)$ is often harder than estimation of $\sigma+(R)$. With $R^{-1}$ this can be circumvented using $1/\sigma_+(R^{-1})$. However:
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)
- Estimation of $\sigma_-(R)$ is often harder than estimation of $\sigma_+(R)$. With $R^{-1}$ this can be circumvented using $1/\sigma_+(R^{-1})$.

**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

- Direct and inverse factors: having both $R$ and $R^{-1}$ (mixed direct/inverse (incomplete) decompositions, some other applications)
- Estimation of $\sigma_-(R)$ is often harder than estimation of $\sigma_+(R)$. With $R^{-1}$ this can be circumvented using $1/\sigma_+(R^{-1})$. However:

**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

$$\sigma_C^-(R)x_+^T = y_-^TR.$$
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 $\nu$ 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 $1/\sigma_+^N(R^{-1}) \leq \sigma_-^N(R)$. Superiority of maximization does not apply always, but might explain the name incremental norm estimation.
Small example: ICE and INE with maximization and minimization

\[ R = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \quad \sigma_-(R) = 0.874 \]
Small example: ICE and INE with maximization and minimization

\[ R = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \quad \sigma_-(R) = 0.874 \]

\[ \frac{1}{\sigma_+^C(R^{-1})} = \sigma_-^C(R) = 1 \]
Small example: ICE and INE with maximization and minimization

\[ R = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \quad \sigma_-(R) = 0.874 \]

\[ \frac{1}{\sigma^C_+(R^{-1})} = \sigma^C_-(R) = 1 \]
\[ 0.8944 \approx \frac{1}{\sigma^N_+(R^{-1})} < \sigma^N_-(R) = 1 \]
Small example: ICE and INE with maximization and minimization

\[
R = \begin{bmatrix}
2 & 0 & 1 \\
1 & 0 & 1 \\
1 & 0 & 1 \\
\end{bmatrix}, \quad \sigma_-(R) = 0.874
\]

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

\[
\hat{R} = \begin{bmatrix}
2 & 0 & 1 & 1 \\
1 & 0 & 1 & 1 \\
1 & 0 & 1 & 1 \\
1 & 0 & 1 & 1 \\
\end{bmatrix}, \quad \sigma_-(\hat{R}) \approx 0.5155
\]
Small example: ICE and INE with maximization and minimization

\[
R = \begin{bmatrix}
2 & 0 & 1 \\
1 & 0 & 1 \\
1 & & \\
\end{bmatrix}, \quad \sigma_-(R) = 0.874
\]

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

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

\[
\sigma_+^C(\hat{R}) = 1/\sigma_+^C(\hat{R}^{-1}) \approx 0.618
\]
Small example: ICE and INE with maximization and minimization

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

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

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

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

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

$$0.5381 \approx 1/\sigma_+^N(\hat{R}^{-1}) < \sigma_+^N(\hat{R}) \approx 0.835$$
An example showing the possible gap between the ICE and INE estimates

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

1. Introduction: The Problem
2. The two strategies
3. ICE and INE with inverse factors
4. INE maximization versus ICE maximization
5. Numerical experiments
6. Conclusions
Consider norm estimation of the extended matrix

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

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 \( w = Rz/\sigma^+ \). We have \( \sigma_N^+(\hat{R}) \geq \sigma_C^+(\hat{R}) \) if

\[
(v^T w)^2 \geq \rho_1,
\]

where \( \rho_1 \) is the smaller root of the quadratic equation in \((v^T w)^2\),

\[
(v^T w)^4 + \left( \frac{\gamma^2 + (v^T y)^2}{\sigma_+^2} \right) (v^T v - (v^T y)^2) - v^T v - (v^T y)^2 (v^T w)^2 \\
+ \left( v^T y \right)^2 \left( \frac{\gamma^2 + v^T v}{\sigma_+^2} \right) \left( (v^T y)^2 - v^T v \right) + v^T v \right) = 0.
\]
Figure: Value of $\rho_1$ in dependence of $(v^T y)^2$ (x-axis) and $\gamma^2$ (y-axis) with $\sigma_+ = 1$, $\|v\|^2 = 0.1$. 
Example: ICE versus INE

Figure: Value of $\rho_1$ in dependence of $(v^T y)^2$ (x-axis) and $\gamma^2$ (y-axis) with $\sigma_+ = 1$, $\|v\|^2 = 1$. 
Example: ICE versus INE

Figure: Value of $\rho_1$ in dependence of $(v^T y)^2$ (x-axis) and $\gamma^2$ (y-axis) with $\sigma_+ = 1$, $\|v\|^2 = 10$. 
Example: ICE versus INE

Figure: Value of $\rho_1$ in dependence of $(v^T y)^2$ (x-axis) and $\gamma^2$ (y-axis) with $\sigma_+ = 1$, $\Delta = 0.6$, $\|v\|^2 = 0.1$. 
Figure: Value of $\rho_1$ in dependence of $(v^T y)^2$ (x-axis) and $\gamma^2$ (y-axis) with $\sigma_+ = 1$, $\Delta = 0.6$, $\|v\|^2 = 1$. 
Figure: Value of $\rho_1$ in dependence of $(v^T y)^2$ (x-axis) and $\gamma^2$ (y-axis) with $\Delta = 0.6$, $\|v\|^2 = 10$. 
Outline

1. Introduction: The Problem
2. The two strategies
3. ICE and INE with inverse factors
4. INE maximization versus ICE maximization
5. Numerical experiments
6. Conclusions
Example 1: 50 matrices $A = \text{rand}(100,100) - \text{rand}(100,100)$, 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, prescribed condition number $\kappa$ choosing

$$\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_{100}), \quad \sigma_k = \alpha^k, \quad 1 \leq k \leq 100, \quad \alpha = \kappa^{-\frac{1}{99}}.$$ 

$U$ 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:
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.
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.
Outline

1. Introduction: The Problem
2. The two strategies
3. ICE and INE with inverse factors
4. INE maximization versus ICE maximization
5. Numerical experiments
6. Conclusions
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.
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.
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.

For more details see:

Last but not least

Thank you for your attention!