On Incremental 2-norm Condition Estimators

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

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

Outline

1 Introduction: The Problem

- 2 The two strategies
- ICE and INE with inverse factors
- 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: *A* upper triangular (no loss of generality computations typically based on triangular decomposition)
- Euclidean norm

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

- 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

- 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))

- 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 κ(A) (note that it is often sufficient to have the estimates within a reasonable multiplicative factor from the exact κ(A) - Demmel (1997))
- 1-norm: Hager (1984), Higham (1987, 1988, 1989, 1990) [175], Higham, Tisseur (2000).

- 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))
- 1-norm: Hager (1984), Higham (1987, 1988, 1989, 1990) [175], Higham, Tisseur (2000).
- 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)

- 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 κ(A) (note that it is often sufficient to have the estimates within a reasonable multiplicative factor from the exact κ(A) - Demmel (1997))
- 1-norm: Hager (1984), Higham (1987, 1988, 1989, 1990) [175], Higham, Tisseur (2000).
- 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)
- 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)).
- Starting point: the methods by Bischof (1990) (incremental condition number estimation - ICE) and Duff, Vömel (2002) (incremental norm estimation - INE).

Outline

Introduction: The Problem

- 2 The two strategies
- 3 ICE and INE with inverse factors
- INE maximization versus ICE maximization
- 5 Numerical experiments

6 Conclusions

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

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

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

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

- Using a left extremal (minimum or maximum) singular vector u_{ext} , if $R = U\Sigma V^T \Rightarrow ||u_{ext}^T R|| = ||u_{ext}^T U\Sigma V^T|| = \sigma_{ext}(R)$.
- Bischof (1990): estimates to extremal singular values and left singular vectors:

$$\begin{split} \sigma_{ext}^{C}(R) &= \|y_{ext}^{T}R\| \approx \sigma_{ext}(R), \\ \|\hat{y}_{ext}^{T}\hat{R}\| &= \left. \exp_{\|[s,c]\|=1}^{T} \right\| \left[\begin{array}{cc} s \, y_{ext}^{T}, & c \end{array} \right] \left[\begin{array}{cc} R & v \\ 0 & \gamma \end{array} \right] \right\| \end{split}$$

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

- Using a left extremal (minimum or maximum) singular vector u_{ext} , if $R = U\Sigma V^T \Rightarrow ||u_{ext}^T R|| = ||u_{ext}^T U\Sigma V^T|| = \sigma_{ext}(R)$.
- Bischof (1990): estimates to extremal singular values and left singular vectors:

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

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

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

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

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

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

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

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

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

$$\sigma_{ext}^{N}(R) = \|Rz_{ext}\| \approx \sigma_{ext}(R)$$
$$\|\hat{R}\hat{z}_{ext}\| = \mathsf{ext}_{\|[s,c]\|=1} \left\| \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} s \, z_{ext} \\ c \end{bmatrix} \right\|$$

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

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

$$\sigma_{ext}^{N}(R) = \|Rz_{ext}\| \approx \sigma_{ext}(R)$$
$$\|\hat{R}\hat{z}_{ext}\| = \operatorname{ext}_{\|[s,c]\|=1} \left\| \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} s \, z_{ext} \\ c \end{bmatrix} \right\|$$

• 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

Introduction: The Problem

- 2 The two strategies
- 3 ICE and INE with inverse factors
 - INE maximization versus ICE maximization
 - 5 Numerical experiments

6 Conclusions

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

- 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:

- 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:

- 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

$$\sigma^C_-(R)x^T_+ = y^T_-R.$$

Theorem

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

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

 $1/\sigma^N_+(\hat{R}^{-1}) \le \sigma^N_-(\hat{R})$

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

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.

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

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

$$R = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_{-}(R) = 0.874$$
$$\frac{1/\sigma_{+}^{C}(R^{-1}) = \sigma_{-}^{C}(R) = 1}{0.8944 \approx 1/\sigma_{+}^{N}(R^{-1}) < \sigma_{-}^{N}(R) = 1}$$

$$R = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_{-}(R) = 0.874$$
$$\frac{1/\sigma_{+}^{C}(R^{-1}) = \sigma_{-}^{C}(R) = 1}{0.8944 \approx 1/\sigma_{+}^{N}(R^{-1}) < \sigma_{-}^{N}(R) = 1}$$
$$\hat{R} = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad \sigma_{-}(\hat{R}) \approx 0.5155$$

$$R = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_{-}(R) = 0.874$$
$$\frac{1/\sigma_{+}^{C}(R^{-1}) = \sigma_{-}^{C}(R) = 1}{0.8944 \approx 1/\sigma_{+}^{N}(R^{-1}) < \sigma_{-}^{N}(R) = 1}$$
$$\hat{R} = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad \sigma_{-}(\hat{R}) \approx 0.5155$$
$$\sigma_{-}^{C}(\hat{R}) = 1/\sigma_{+}^{C}(\hat{R}^{-1}) \approx 0.618$$

$$R = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_{-}(R) = 0.874$$
$$\frac{1/\sigma_{+}^{C}(R^{-1}) = \sigma_{-}^{C}(R) = 1}{0.8944 \approx 1/\sigma_{+}^{N}(R^{-1}) < \sigma_{-}^{N}(R) = 1}$$
$$\hat{R} = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad \sigma_{-}(\hat{R}) \approx 0.5155$$
$$\sigma_{-}^{C}(\hat{R}) = 1/\sigma_{+}^{C}(\hat{R}^{-1}) \approx 0.618$$
$$0.5381 \approx 1/\sigma_{-}^{N}(\hat{R}^{-1}) < \sigma_{-}^{N}(\hat{R}) \approx 0.835$$

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

Introduction: The Problem

- 2 The two strategies
- 3 ICE and INE with inverse factors
- INE maximization versus ICE maximization
 - 5 Numerical experiments

6 Conclusions

INE versus ICE

Theorem

Consider norm estimation of the extended 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 $w = Rz/\sigma^+$. We have $\sigma^N_+(\hat{R}) \ge \sigma^C_+(\hat{R})$ if $(v^Tw)^2 \ge \rho_1$,

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

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



Figure : Value of ρ_1 in dependence of $(v^Ty)^2$ (x-axis) and γ^2 (y-axis) with $\sigma_+=1,~\|v\|^2=0.1.$



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



Figure : Value of ρ_1 in dependence of $(v^Ty)^2$ (x-axis) and γ^2 (y-axis) with $\sigma_+ = 1$, $\|v\|^2 = 10$.



Figure : Value of ρ_1 in dependence of $(v^T y)^2$ (x-axis) and γ^2 (y-axis) with $\sigma_+ = 1$, $\Delta = 0.6$, $||v||^2 = 0.1$.



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



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

Introduction: The Problem

- 2 The two strategies
- 3 ICE and INE with inverse factors
- INE maximization versus ICE maximization

5 Numerical experiments

6 Conclusions

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.

Example 2: 50 matrices $A = U\Sigma V^T$ of size 100, prescribed condition number κ choosing

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

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

Introduction: The Problem

- 2 The two strategies
- ICE and INE with inverse factors
- INE maximization versus ICE maximization
- 5 Numerical experiments



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

- 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.

- 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:

J. Duintjer Tebbens, M. Tůma: On Incremental Condition Estimators in the 2-Norm , Preprint NCCM/2013/15, submitted, May 2013.

Thank you for your attention!