Incremental condition estimation with inverse triangular factors

Jurjen Duintjer Tebbens

joint work with

Miroslav Tůma

Institute of Computer Science

Academy of Sciences of the Czech Republic

ALA 2010, Novi Sad, May 25, 2010.

1. Motivation: BIF

- This work is motivated by the recently introduced Balanced Incomplete Factorization (BIF) method for LDU decomposition [Bru, Marín, Mas, Tůma - 2008] - see the after-next talk ...

- The method is remarkable, among others, in that it computes the *inverse triangular factors* simultaneously during the factorization process.

- Can the presence of the inverse factors in BIF be exploited ?

Perhaps the first thing that comes to mind, is to use the inverse triangular factors for improved condition estimation.

We will see that exploiting the inverse factors for better condition estimation is possible, but not as straightforward as it may seem.

2. Incremental condition estimation (ICE)

We assume A is real and positive definite symmetric. If

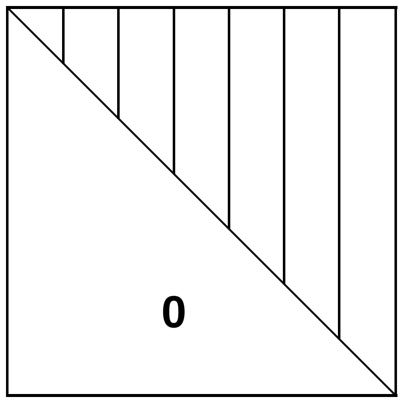
$$A = LL^T$$

is the Cholesky decomposition of A, the condition number of A satisfies

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

We focus on estimation of the 2-norm condition number of L^T. This can be done cheaply with a technique called *incremental* condition number estimation. Main idea: Subsequent estimation of leading submatrices:

L^T =



$\Rightarrow$ all columns are accessed only once.

2. Incremental condition estimation (ICE)

We will call the original incremental technique, introduced by Bischof [Bischof - 1990], simply incremental condition estimation (ICE):

Let R be upper triangular with a given approximate maximal singular value $\sigma_{maxICE}(R)$ and corresponding approximate singular vector y, $\|y\| = 1$,

$$\sigma_{maxICE}(R) = \|y^T R\| \approx \sigma_{max}(R) = \max_{\|x\|=1} \|x^T R\|.$$

ICE approximates the maximal singular value of the extended matrix

$$R' = \begin{pmatrix} R & v \\ 0 & \gamma \end{pmatrix}$$

by maximizing

$$\left\| \begin{pmatrix} sy, & c \end{pmatrix} \begin{pmatrix} R & v \\ 0 & \gamma \end{pmatrix} \right\|, \quad \text{over all } s, c \text{ satisfying} \quad c^2 + s^2 = 1.$$

2. Incremental condition estimation (ICE)

We have

$$\max_{s,c,c^2+s^2=1} \left\| \begin{pmatrix} sy, & c \end{pmatrix} \begin{pmatrix} R & v \\ 0 & \gamma \end{pmatrix} \right\|^2 = \max_{s,c,c^2+s^2=1} \begin{pmatrix} sy, & c \end{pmatrix} \begin{pmatrix} R & v \\ 0 & \gamma \end{pmatrix} \begin{pmatrix} R^T & 0 \\ v^T & \gamma \end{pmatrix} \begin{pmatrix} sy \\ c \end{pmatrix}$$

$$= \max_{s,c,c^2+s^2=1} \begin{pmatrix} s, & c \end{pmatrix} \begin{pmatrix} \sigma_{maxICE}(R)^2 + (y^T v)^2 & \gamma(v^T y) \\ \gamma(v^T y) & \gamma^2 \end{pmatrix} \begin{pmatrix} s \\ c \end{pmatrix}.$$

The maximum is obtained with the normalized eigenvector corresponding to the maximum eigenvalue $\lambda_{max}(B_{ICE})$ of

$$B_{ICE} \equiv \begin{pmatrix} \sigma_{maxICE}(R)^2 + (y^T v)^2 & \gamma(v^T y) \\ \gamma(v^T y) & \gamma^2 \end{pmatrix}.$$

We denote the normalized eigenvector by $\begin{pmatrix} \hat{s} \\ \hat{c} \end{pmatrix}$.

2. Incremental condition estimation (ICE)

Then with $\hat{y}^T = \begin{pmatrix} \hat{s}y, & \hat{c} \end{pmatrix}$ we define the approximate maximal singular value of the extended matrix as

$$\sigma_{maxICE}(R') \equiv \|\hat{y}^T R'\| \approx \sigma_{max}(R').$$

Similarly, if for some y with unit norm,

$$\sigma_{minICE}(R) = \|y^T R\| \approx \sigma_{min}(R) = \min_{\|x\|=1} \|x^T R\|,$$

then ICE uses the *minimal* eigenvalue $\lambda_{min}(B_{ICE})$ of

$$B_{ICE} = \begin{pmatrix} \sigma_{minICE}(R)^2 + (y^T v)^2 & \gamma(v^T y) \\ \gamma(v^T y) & \gamma^2 \end{pmatrix}$$

The corresponding eigenvector of B_{ICE} yields the new vector $\hat{y}^T$ and

$$\sigma_{minICE}(R') \equiv \|\hat{y}^T R'\| \approx \sigma_{min}(R').$$

2. Incremental condition estimation (ICE)

Experiment:

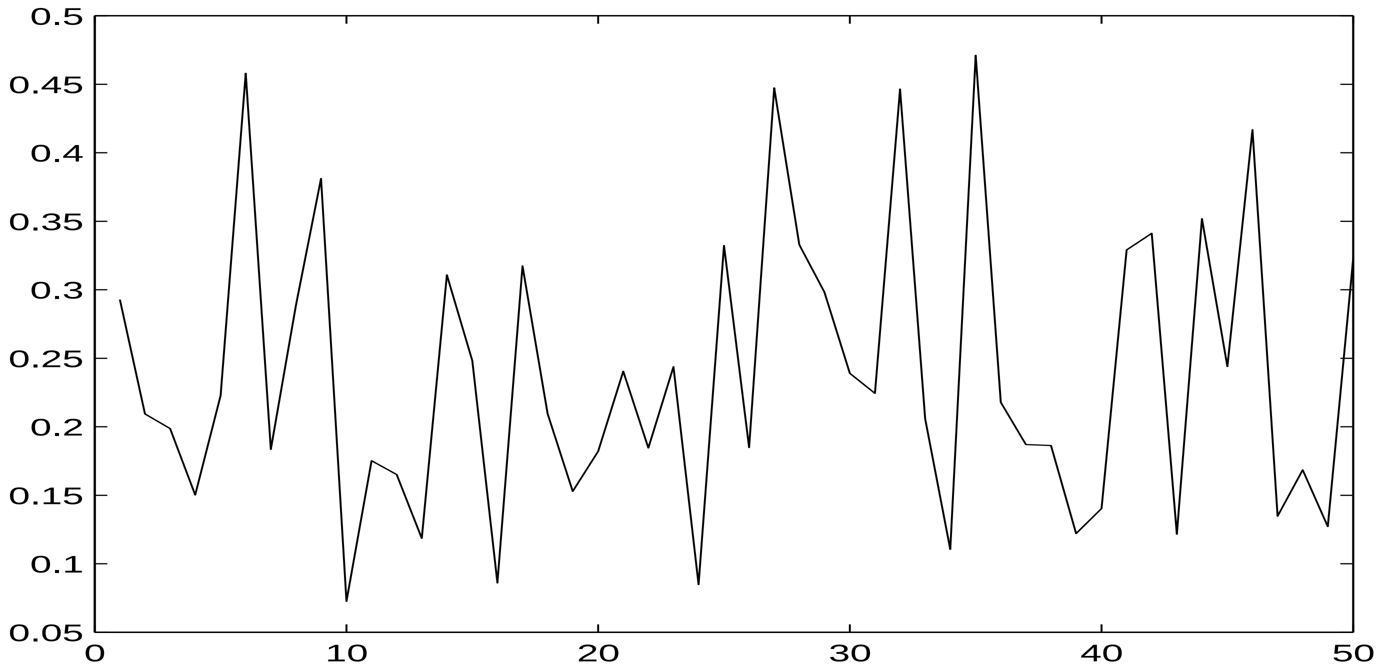
- We generate 50 random matrices A of dimension 100 with the Matlab command A $= \mathtt{randn}(100, 100)$
- We compute the Cholesky decompositions LL^T of the 50 symmetric positive definite matrices AA^T
- We compute the estimations $\sigma_{maxICE}(L^T)$ and $\sigma_{minICE}(L^T)$
- In the following graph we display the quality of the estimations through the number

$$\frac{\left(\frac{\sigma_{maxICE}(L^T)}{\sigma_{minICE}(L^T)}\right)^2}{\kappa(AA^T)},$$

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

$$\left(\frac{\sigma_{maxICE}(L^T)}{\sigma_{minICE}(L^T)}\right)^2 \leq \kappa(AA^T).$$

2. Incremental condition estimation (ICE)



Quality of the estimator ICE for 50 random s.p.d. matrices of dimension 100.

2. Incremental condition estimation (ICE)

Now assume we have to our disposal not only the Cholesky decomposition of AA^T,

$$AA^T = LL^T$$

but also the inverse Cholesky factors as is the case in BIF, i.e. we have

$$(AA^T)^{-1} = L^{-T}L^{-1}.$$

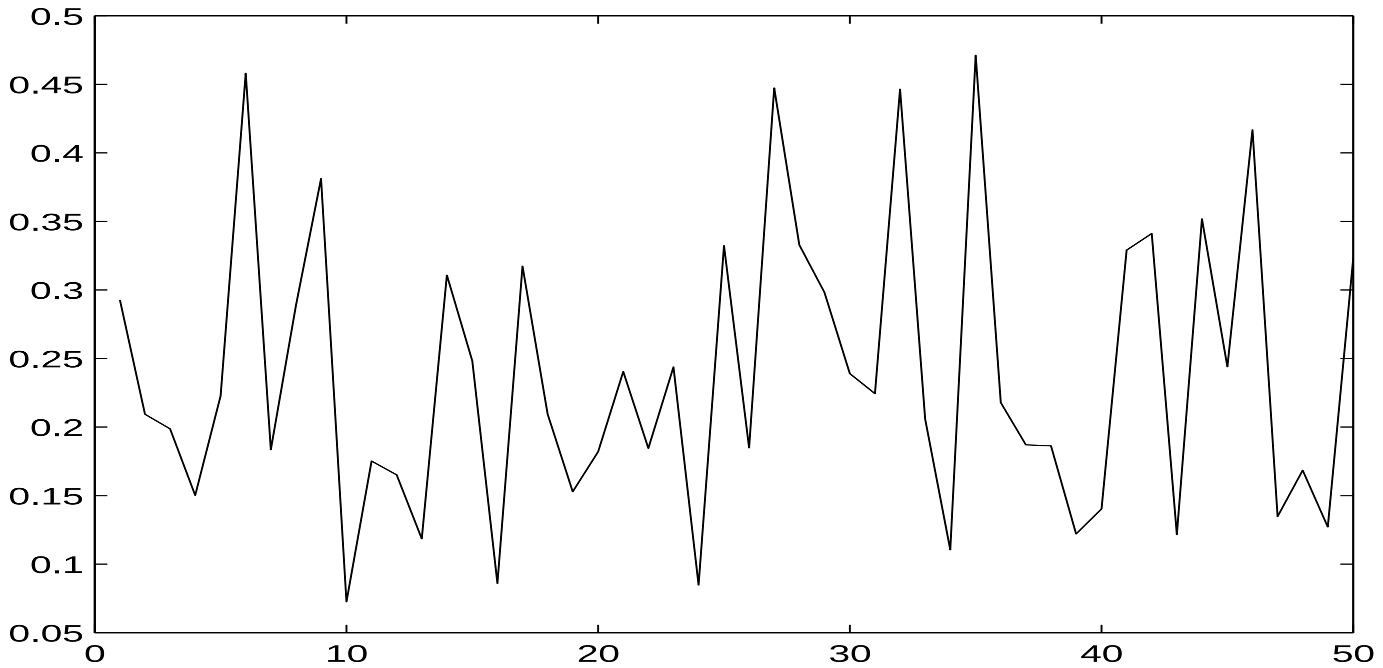
Then we can run ICE on L^{-T} and use the additional estimations

$$\frac{1}{\sigma_{maxICE}(L^{-T})} \approx \sigma_{min}(L^T), \qquad \frac{1}{\sigma_{minICE}(L^{-T})} \approx \sigma_{max}(L^T).$$

In the following graph we take the best of both estimations, we display

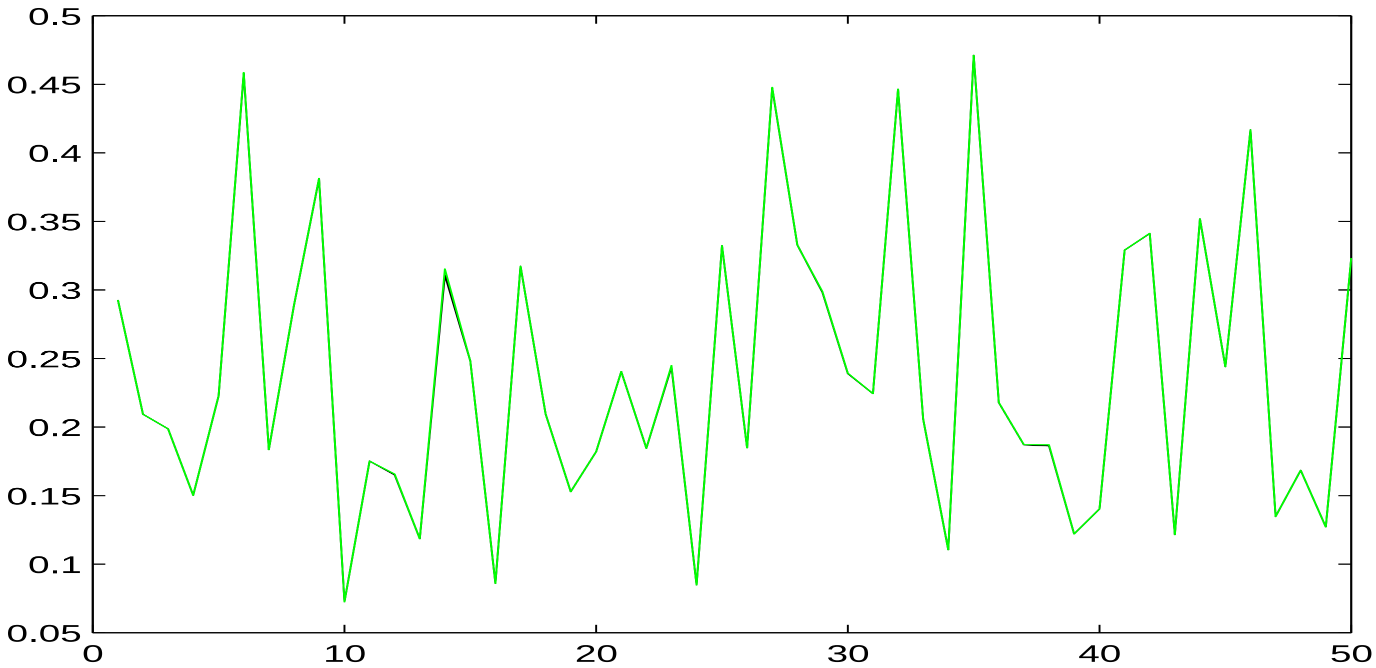
$$\frac{\left(\frac{\max(\sigma_{maxICE}(L^T),\sigma_{minICE}(L^{-T})^{-1})}{\min(\sigma_{minICE}(L^T),\sigma_{maxICE}(L^{-T})^{-1})}\right)^2}{\kappa(AA^T)}.$$

2. Incremental condition estimation (ICE)



Quality of the estimator ICE for 50 random s.p.d. matrices of dimension 100.

2. Incremental condition estimation (ICE)



Quality of the estimator ICE without (black) and with exploiting (green) the inverse for 50 random s.p.d. matrices of dimension 100.

3. Incremental norm estimation (INE)

An alternative technique called incremental norm estimation (INE) was proposed in [Duff, Vömel - 2002]:

Let R be upper triangular with given approximate maximal singular value $\sigma_{maxINE}(R)$ and corresponding approximate *right* singular vector z, $\|z\| = 1$,

$$\sigma_{maxINE}(R) = \|Rz\| \approx \sigma_{max}(R) = \max_{\|x\|=1} \|Rx\|.$$

INE approximates the maximal singular value of the extended matrix

$$R' = \begin{pmatrix} R & v \\ 0 & \gamma \end{pmatrix}$$

by maximizing

$$\left\| \begin{pmatrix} R & v \\ 0 & \gamma \end{pmatrix} \begin{pmatrix} sz \\ c \end{pmatrix} \right\|, \quad \text{over all } s, c \text{ satisfying } c^2 + s^2 = 1.$$

3. Incremental norm estimation (INE)

We have

$$\max_{s,c,c^2+s^2=1} \left\| \begin{pmatrix} R & v \\ 0 & \gamma \end{pmatrix} \begin{pmatrix} sz \\ c \end{pmatrix} \right\|^2 = \max_{s,c,c^2+s^2=1} \begin{pmatrix} sz \\ c \end{pmatrix} \begin{pmatrix} R^T & 0 \\ v^T & \gamma \end{pmatrix} \begin{pmatrix} R & v \\ 0 & \gamma \end{pmatrix} \begin{pmatrix} sz \\ c \end{pmatrix}$$

$$= \max_{s,c,c^2+s^2=1} \begin{pmatrix} s & c \end{pmatrix} \begin{pmatrix} z^T R^T R z & z^T R^T v \\ z^T R^T v & v^T v + \gamma^2 \end{pmatrix} \begin{pmatrix} s \\ c \end{pmatrix}.$$

The maximum is obtained for the normalized eigenvector corresponding to the maximum eigenvalue $\lambda_{max}(B_{INE})$ of

$$B_{INE} \equiv \begin{pmatrix} z^T R^T R z & z^T R^T v \\ z^T R^T v & v^T v + \gamma^2 \end{pmatrix}.$$

We denote the normalized eigenvector by $\begin{pmatrix} \hat{s} \\ \hat{c} \end{pmatrix}$.

3. Incremental norm estimation (INE)

Then with $\hat{z} = \left(\hat{s}z, \quad \hat{c}\right)^T$ we define the approximate maximal singular value of the extended matrix as

$$\sigma_{maxINE}(R') \equiv \|R'\hat{z}\| \approx \sigma_{max}(R').$$

Similarly, if for a unit vector z,

$$\|Rz\| \approx \sigma_{min}(R) = \min_{\|x\|=1} \|Rx\|,$$

then INE uses the *minimal* eigenvalue $\lambda_{min}(B_{INE})$ of

$$B_{INE} = \begin{pmatrix} z^T R^T R z & z^T R^T v \\ z^T R^T v & v^T v + \gamma^2 \end{pmatrix}.$$

The corresponding eigenvector of B_{INE} yields the new vector $\hat{z}$ and

$$\sigma_{minINE}(R') \equiv \|R'\hat{z}\| \approx \sigma_{min}(R').$$

3. Incremental norm estimation (INE)

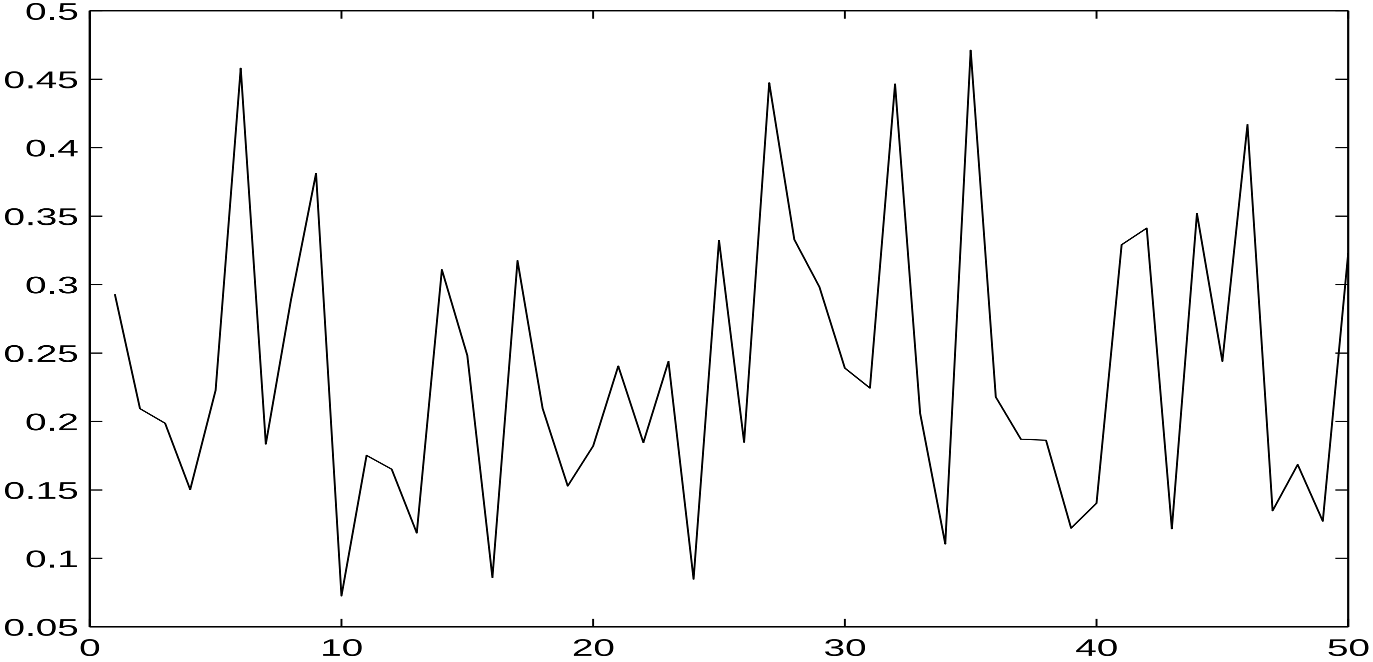
Consider the same experiment as before.

We can combine the estimations from ICE and INE to improve the estimator.

In the following graph we take the best of both estimations and display

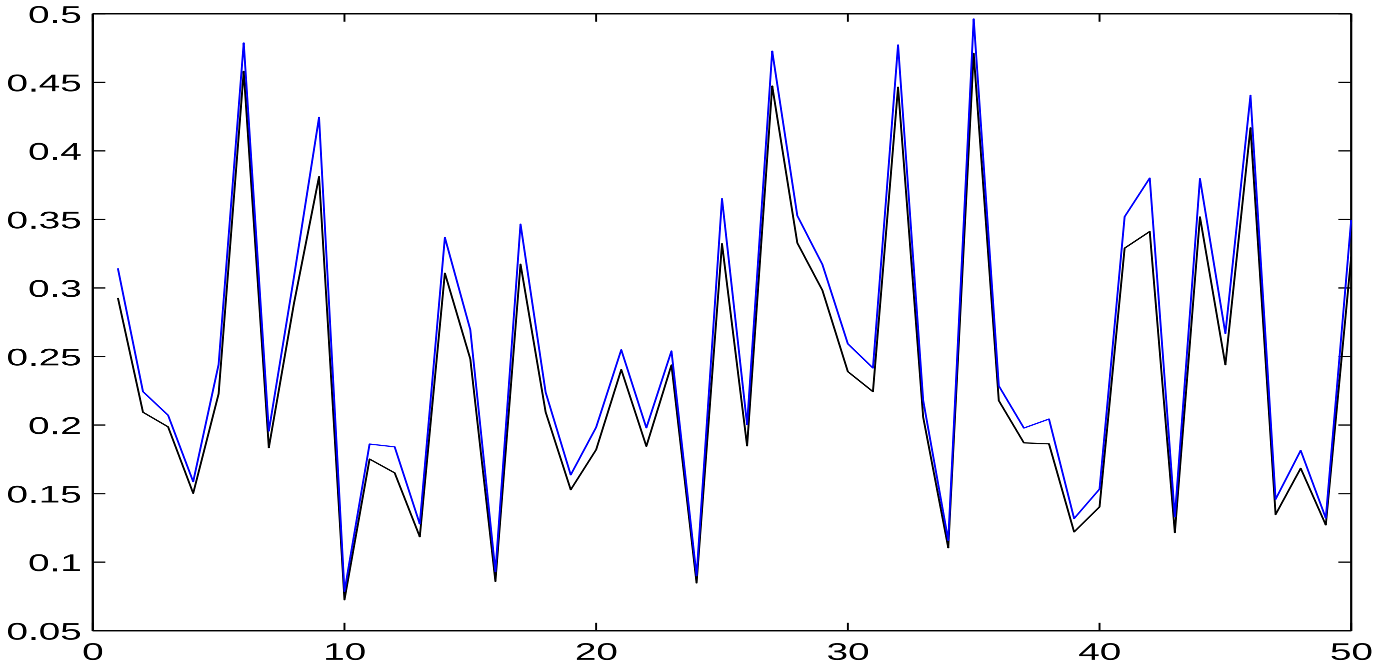
$$\frac{\left(\frac{\max(\sigma_{maxICE}(L^T),\sigma_{maxINE}(L^T))}{\min(\sigma_{minICE}(L^T),\sigma_{minINE}(L^T))}\right)^2}{\kappa(AA^T)}\,.$$

3. Incremental norm estimation (INE)



Quality of the estimator ICE for 50 random s.p.d. matrices of dimension 100.

3. Incremental norm estimation (INE)



Quality of the estimator ICE (black) and of ICE combined with INE (blue) for 50 random s.p.d. matrices of dimension 100.

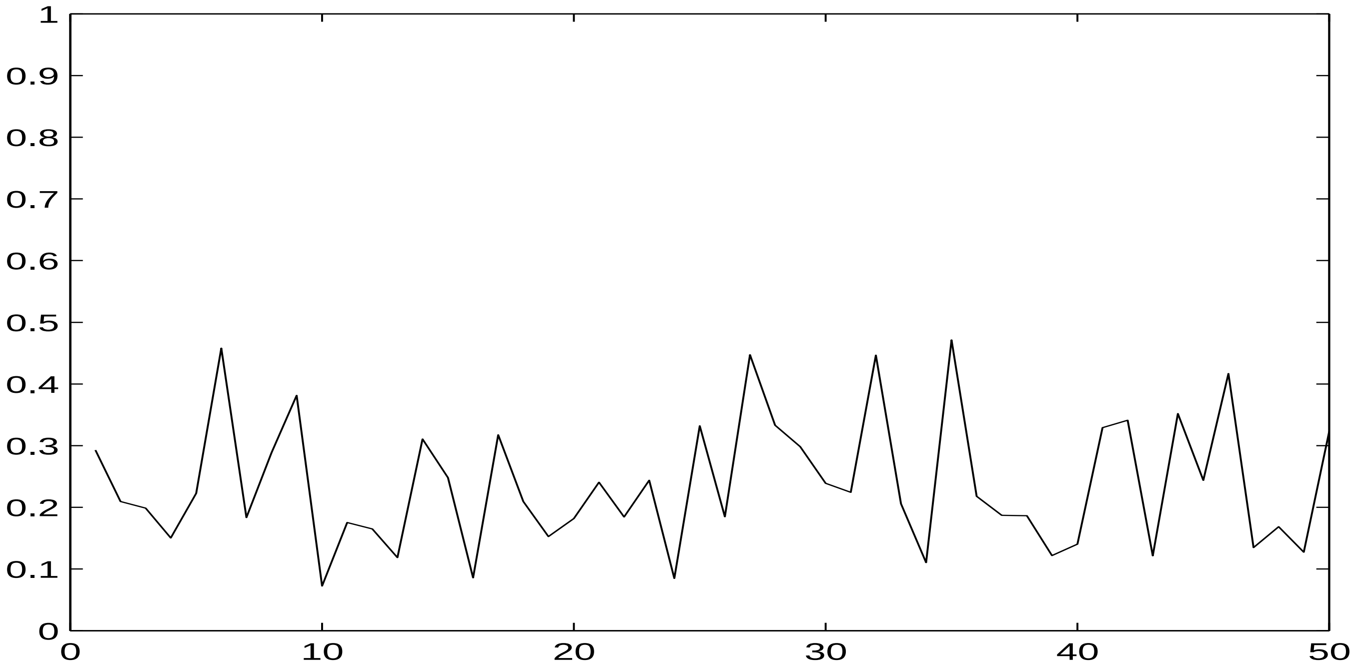
3. Incremental norm estimation (INE)

Finally, if we assume we have to our disposal the inverse factors, we can combine ICE with INE for *both* L^T and L^{-T}.

In the following graph we take the best of four estimations and display

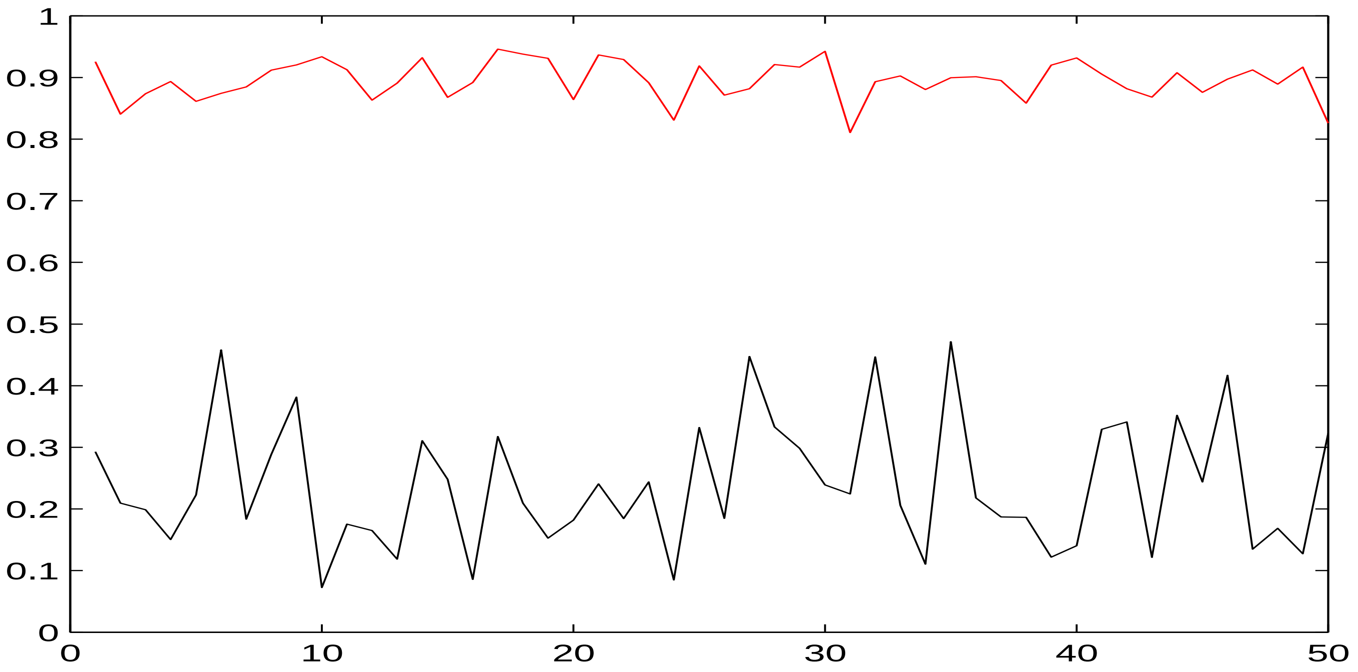
$$\frac{\left(\frac{\max(\sigma_{maxICE}(L^T),\sigma_{maxINE}(L^T),\sigma_{minICE}(L^{-T})^{-1},\sigma_{minINE}(L^{-T})^{-1})}{\min(\sigma_{minICE}(L^T),\sigma_{minINE}(L^T),\sigma_{maxICE}(L^{-T})^{-1},\sigma_{maxINE}(L^{-T})^{-1})}\right)^2}{\kappa(AA^T)}.$$

3. Incremental norm estimation (INE)



Quality of the estimator ICE for 50 random s.p.d. matrices of dimension 100.

3. Incremental norm estimation (INE)



Quality of the estimator ICE (black) and of ICE combined with INE and exploiting the inverse (red) for 50 random s.p.d. matrices of dimension 100.

4. Superiority of INE for σ_{max}

Why this improvement ?

- In general, both ICE and INE give a satisfactory approximation of $\sigma_{max}(A)$, though INE tends to be better.
- The problem is to approximate $\sigma_{min}(A)$, for ICE as well as for INE.
- The trick is to translate to the problem of finding the maximal singular value $\sigma_{max}(A^{-1})$ of A^{-1}, which is in general done *better* with INE than with ICE.
- This has an important impact on the estimate because $\sigma_{min}(A)$ is typically small and appears in the denominator of $\frac{\sigma_{max}(A)}{\sigma_{min}(A)}$,

We see that the main reason for the improvement is that INE tends to give a better estimate of maximal singular values. And why is that ?

4. Superiority of INE for σ_{max}

Note: INE does not *always* give a better estimate of the maximal singular value. But we have the following rather technical result.

Theorem. Consider condition estimation of the matrix

$$R' = \begin{pmatrix} R & v \\ 0 & \gamma \end{pmatrix},$$

where both ICE and INE start with the same approximation of $\sigma_{max}(R)$ denoted by δ. Let y, $\|y\| = 1$ be the approximate singular vector for ICE,

$$\delta = \|y^T R\| \approx \sigma_{max}(R),$$

and let z, $\|z\| = 1$ be the approximate singular vector for INE,

$$\delta = \|Rz\| \approx \sigma_{max}(R).$$

4. Superiority of INE for σ_{max}

Theorem (continued). Then we have superiority of INE,

$$\sigma_{maxINE}(R') \quad \geq \quad \sigma_{maxICE}(R'),$$

if

$$(v^T R z)^2 \quad \geq \quad \delta^2 (v^T y)^2 + \frac{1}{2}\left(v^T v - (v^T y)^2\right)\left(\alpha - \sqrt{\alpha^2 + 4\delta^2 (v^T y)^2}\right).$$

where $\alpha = \delta^2 - \gamma^2 - (v^T y)^2$.

Hence if $\delta^2 (v^T y)^2 + \frac{1}{2}\left(v^T v - (v^T y)^2\right)\left(\alpha - \sqrt{\alpha^2 + 4\delta^2 (v^T y)^2}\right) \leq 0$, then INE is unconditionally superior to ICE (i.e. regardless of the approximate singular vector z). Let us use the notation

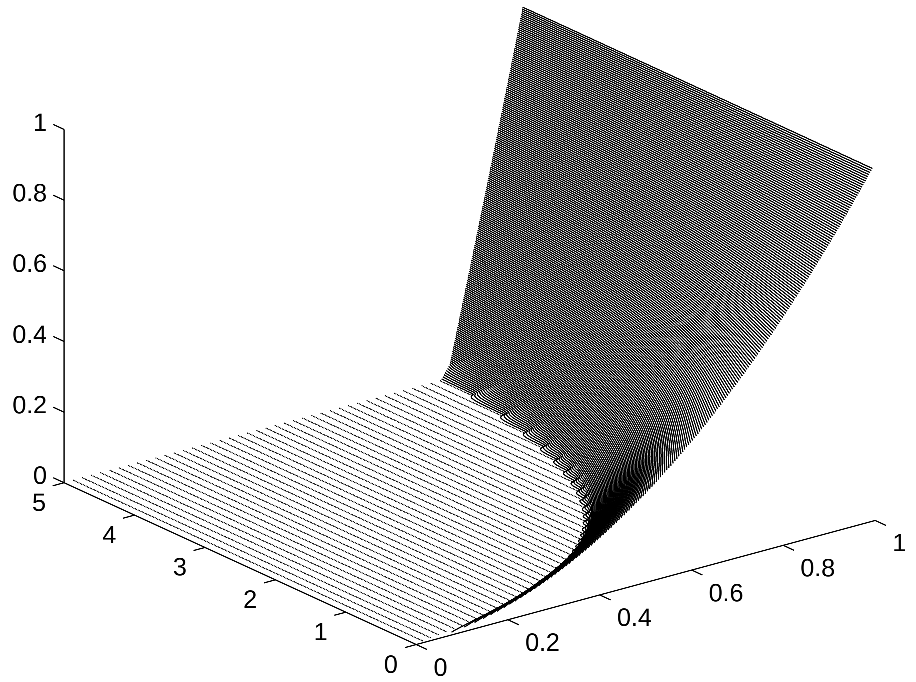
$$\rho \equiv \delta^2 (v^T y)^2 + \frac{1}{2}\left(v^T v - (v^T y)^2\right)\left(\alpha - \sqrt{\alpha^2 + 4\delta^2 (v^T y)^2}\right).$$

4. Superiority of INE for σ_{max}

To conclude we demonstrate the previous theorem.

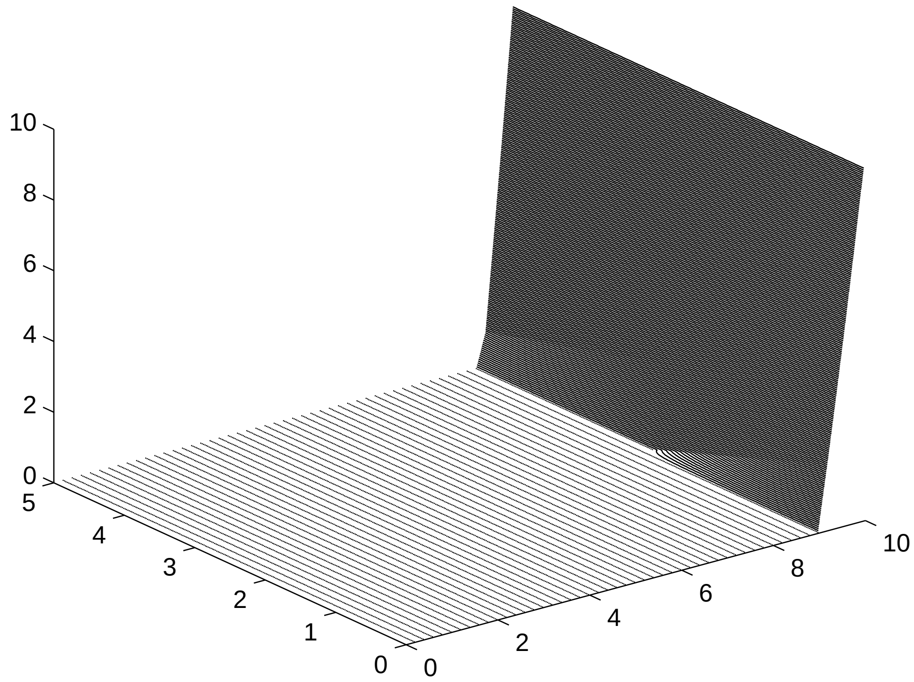
- Assume that at some stage of an incremental condition estimation process we have $\sigma_{maxICE}(R) = \sigma_{maxINE}(R) = 1$.
- Consider possible new columns v of R' that have unit norm, i.e. $v^T v = 1$.
- Then $(v^T y)^2 \leq 1$. The x-axes of the following figures represent the possible values of $(v^T y)^2 < 1$.
- The y-axes represent values of γ^2, i.e. the square of the new diagonal entry.
- The superiority criterion for INE expressed by the value of ρ is given by the z-axes.

4. Superiority of INE for σ_{max}



Value of ρ in dependence of $(v^T y)^2$ (x-axis) and γ^2 (y-axis) with $\|v\|^2 = 1$.

4. Superiority of INE for σ_{max}



Value of ρ in dependence of $(v^T y)^2$ (x-axis) and γ^2 (y-axis) with $\|v\|^2 = 10$.

5. Conclusion and future work

- Exploiting the presence of inverse factors *combined* with INE gives a significant improvement of incremental condition estimation.

- This may be an important advantage of methods like BIF where inverse triangular factors are just a by-product of the factorization.

- We did not consider sparse Cholesky factors, which ask for modified ICE [Bischof, Pierce, Lewis - 1990].

- We did not consider exploiting the inverse in estimation of the 1-norm and other non-Euclidean condition number.

Thank you for your attention!

Supported by project number IAA100300802 of the grant agency of the Academy of Sciences of the Czech Republic.