Improved incremental 2-norm condition estimation of triangular matrices

J. Duintjer Tebbens[1] and M. Tůma[1,2]

Institute of Computer Science, Academy of Sciences of the Czech Republic[1] and Faculty of Mathematics and Physics, Charles University in Prague[2]

The matrix condition number for square nonsingular matrices

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

is an important quantity in matrix theory and computations. It is used, among others, to

- assess the quality of computed solutions
- estimate the sensitivity to perturbations
- monitor and control *adaptive* computational processes like adaptive filters, recursive least-squares or ACE for multilevel PDE solvers.

Robust dropping rules

We are particularly interested in the adaptive process of *incomplete LU factorization* using dropping and pivoting where it is important to monitor the condition number of the submatrices that are progressively computed: 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). Therefore the norms of L^{-1} and U^{-1} directly influence the stability of the preconditioned system.

Dropping rules based on the sizes of L^{-1} and U^{-1} lead, with appropriate pivoting, to robust ILU methods [2, 3, 5, 6]. For example, dropping of new entries for the kth leading submatrix of L constructed in the ILU process is done in ILUPACK [4] according to the rule

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

and similarly for new entries of U. Information about the size of $\|e_k^T L^{-1}\|_\infty$ is obtained by using a cheap condition estimator for the ∞-norm.

Mixed direct/inverse factorization

In recently introduced mixed direct/inverse decomposition methods called *Balanced Incomplete Factorization* (BIF) for incomplete Cholesky or LDU decomposition [7, 8] 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. The main idea is to balance the growth of both the direct and the inverse factors by exploiting 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 entries of U and U^{-1}.

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 $\|e_k^T L^{-1}\|$. We present a relatively accurate *2-norm* condition number 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).

Incremental condition number 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

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

$\kappa(R)$ can be cheaply estimated with a technique called *incremental condition number estimation*, introduced in the nineties by Bischof [1]. It computes a sequence of approximate condition numbers of the leading upper left submatrices of growing dimension. The approximation for the current submatrix is obtained from an approximate singular vector constructed without accessing the previous submatrices. This makes the procedure relatively inexpensive and particularly suited when a triangular matrix is computed one row or column at a time.

ICE: Incremental condition estimation as proposed by Bischof [1]

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

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

If y_- is an approximate left minimum singular vector, then this gives an estimate of the minimum singular value

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

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

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

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

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

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 with the entries c and s from 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: An alternative technique proposed by Duff and Vömel [9]

A similar strategy based on approximate *right* singular vectors was proposed later by Duff and Vömel [9] and recommended for norm estimation and for sparse matrices. It estimates extremal (minimum or maximum) singular values and right singular vectors for $\hat{R}$ starting from

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

Then z_{ext} is incremented to $\hat{z}_{ext}$ as

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

where s and c are the entries of the eigenvector corresponding to the extremal eigenvalue of

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

Which incremental technique is superior ?

- In both ICE and INE the main computational costs come from forming inner products in the entries of 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 probably why INE is called incremental *norm* estimation.

ICE and INE if inverse factors are available

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

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

Surprisingly, this gives no better incremental estimator (see [10, Theorem 3.2]):

Theorem 1

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

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

One can prove $\sigma_+^C(R) = 1/\sigma_-^C(R^{-1})$ as well. Using the inverse *does* improve INE estimation [10, Theorem 3.2]:

Theorem 2

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

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

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

In case the assumption is relaxed to $1/\sigma_+^N(R^{-1}) \le \sigma_-^N(R)$ we obtain a rather technical theorem, saying essentially that maximization with R^{-1} is in most cases superior to minimization with R. Nevertheless, in *all* performed numerical experiments we found that $\sigma_-^N(\hat{R})$ gives an estimate which is worse than $1/\sigma_+^N(\hat{R}^{-1})$. For example:

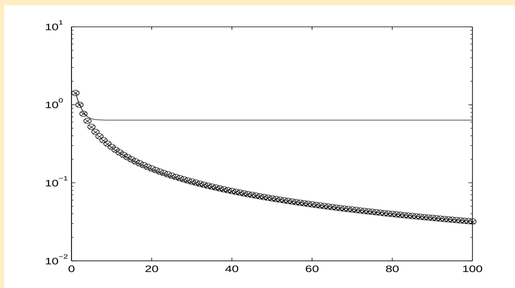

Figure 1 : INE estimation of the smallest singular value for the 1D Laplacians of size one until hundred: INE with minimization (solid line), INE with maximization exploiting the inverse (circles) and exact minimum singular values (crosses).

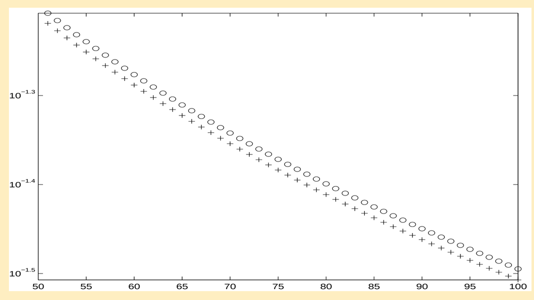

Figure 2 : Zoom of Figure 1. INE estimation of the smallest singular value for the 1D Laplacians of size fifty until hundred for INE with maximization exploiting the inverse (circles) and exact minimum singular values (crosses) .

Hence if the inverse of R is available, it is recommendable to use $1/\sigma_+^N(\hat{R}^{-1})$ instead of $\sigma_-^N(\hat{R})$ to approximate the minimum singular value. An analogue of Theorem 2 for estimates of the maximum singular value shows that

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

Thus for the maximum singular value, it is recommendable to use the original $\sigma_+^N(\hat{R})$ instead of $1/\sigma_-^N(\hat{R}^{-1})$. In this sense, INE performs better when doing maximization than when doing minimization.

Numerical experiments

We will compare the following estimators:

- The original ICE technique with the estimates defined as $\sigma_+^C(R)/\sigma_-^C(R)$: Solid lines.
- The original INE technique with the estimates defined by $\sigma_+^N(R)/\sigma_-^N(R)$: Circles.
- The INE technique based on maximization only, i.e. estimates defined as $\sigma_+^N(R)/\left(\sigma_+^N(R^{-1})\right)^{-1}$: Plusses.
- The INE technique based on minimization only, i.e. estimates defined as $\left(\sigma_-^N(R^{-1})\right)^{-1}/\sigma_-^N(R)$: Squares.

Example 1: The same experiments as in [1, Section 4, Test 2], [9, Section 5, Table 5.4]: 50 matrices $A = U\Sigma V^T$ of size 100 with prescribed condition number κ where

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

U and V are random unitary factors, R is the triangular factor from the QR decomposition of A with colamd.

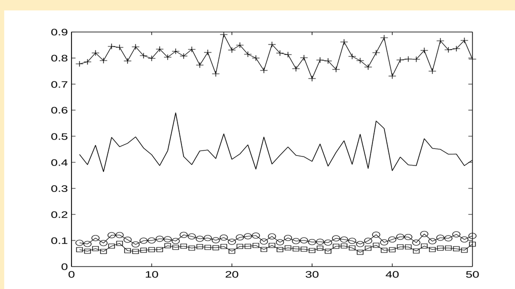

Figure 3 : Ratios of estimate versus exact condition number for Example 1 with $\kappa(A) = 100$.

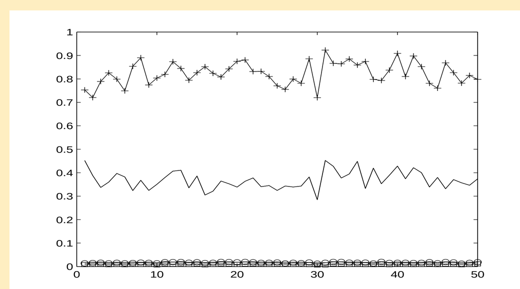

Figure 4 : Ratios of estimate versus exact condition number for Example 1 with $\kappa(A) = 1000$.

Example 2: 20 moderate size matrices from the Matrix Market collection, most of them tested also in [9, Section 5, Table 5.1]. We computed their QR decomposition (with and without column pivoting) and tested the estimators with the factor R.

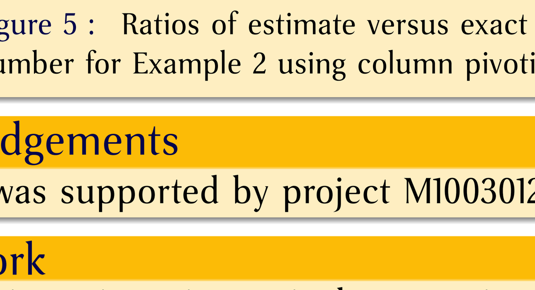

Figure 5 : Ratios of estimate versus exact condition number for Example 2 using column pivoting

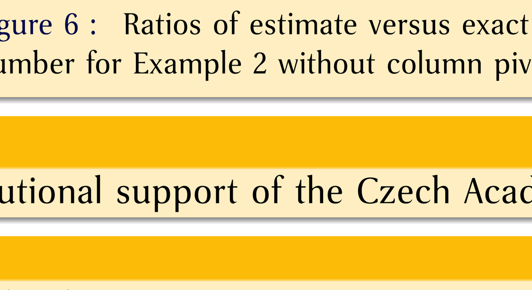

Figure 6 : Ratios of estimate versus exact condition number for Example 2 without column pivoting.

Acknowledgements

This work was supported by project M100301201 of the institutional support of the Czech Academy of Sciences.

Future work

We wish to investigate in particular more in detail the following issues:

- Large sparse matrices: It may be possible to obtain the same accurate estimates without storing the (in general dense) inverse triangular factors. But computation of the inverse factors seems to be unavoidable.
- Block versions based on block factorizations for dense matrices to enable exploitation of fast BLAS techniques.

1. C. H. Bischof, *Incremental condition estimation*, SIAM J. on Matrix Analysis and Applications, 11 (1990), pp. 312–322.
2. M. Bollhöfer, *A robust ILU with pivoting based on monitoring the growth of the inverse factors*, Linear Algebra and its Applications, 338 (2001), pp. 201–218.
3. ——, *A robust and efficient ILU that incorporates the growth of the inverse triangular factors*, SIAM J. on Scientific Computing, 25 (2003), pp. 86–103.
4. ——, *ILUPACK: version 2.4*, 2011. `http://www.icm.tu-bs.de/ bolle/ilupack/`.
5. M. Bollhöfer and Y. Saad, *On the relations between ILUs and factored approximate inverses*, SIAM J. on Matrix Analysis and Applications, 24 (2002), pp. 219–237.
6. ——, *Multilevel preconditioners constructed from inverse-based ILUs*, SIAM J. on Scientific Computing, 27 (2006), pp. 1627–1650.
7. R. Bru, J. Marín, J. Mas, and M. Tuma, *Balanced incomplete factorization*, SIAM J. on Scientific Computing, 30 (2008), pp. 2302–2318.
8. ——, *Improved balanced incomplete factorization*, SIAM J. on Matrix Analysis and Applications, 31 (2010), pp. 2431–2452.
9. I. S. Duff and C. Vömel, *Incremental norm estimation for dense and sparse matrices*, BIT Numerical Mathematics, 42 (2002), pp. 300–322.
10. J. Duintjer Tebbens and M. Tůma, *On incremental condition estimators in the 2-norm*, SIAM J. on Matrix Analysis and Applications, 35 (2014), pp. 174–197.