ON INCREMENTAL CONDITION ESTIMATORS IN THE 2-NORM

JURJEN DUINTJER TEBBENS AND MIROSLAV TŮMA*

Abstract. The paper deals with estimating the condition number of triangular matrices in the Euclidean norm. The two main incremental methods, based on the work of Bischof and on the later work of Duff and Vömel, are compared. The paper presents new theoretical results revealing their similarities and differences. As typical in condition number estimation, there is no universal always-winning strategy, but theoretical and experimental arguments show that the clearly preferable approach is the algorithm of Duff and Vömel when appropriately applied to both the triangular matrix itself and its inverse. This leads to a highly accurate incremental condition number estimator.

Key words. condition number estimation, matrix inverses, incremental condition estimator, incremental norm estimator

1. Introduction.

The condition number

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

of a nonsingular matrix is a very important quantity in numerical linear algebra. While its computation is typically as expensive as solving a corresponding system of linear equations, there exist efficient procedures for condition number estimation. Proper use of the computed estimates can often save a lot of computational effort.

First of all, matrix condition number estimates may be used in the basic tasks of numerical linear algebra, that is, in solving systems of linear algebraic equations and solving eigenvalue problems, to assess the quality of the computed solutions and their sensitivity to perturbations. Further, there are specific fields in scientific computing that are strongly linked with condition number estimation. The estimated condition number may be used to monitor and control adaptive computational processes, sometimes using the terminology adaptive condition estimators (ACE). Such adaptive processes may include evaluation of adaptive filters [24], [31] and recursive least squares in signal processing [22] or solving nonlinear problems by linearization methods [24], [36]. Standard algebraic approaches are used for tracking the condition number in a sequence of modified matrices of the same dimension as well as when matrices are subsequently constructed by augmentation [34], [35], [37], [21], [22]. ACE based on properties of model and grid hierarchies is a standard tool in multilevel PDE solvers [29]. Another type of problem-oriented ACE in recursive least squares measured with a norm close to the Frobenius norm is represented in [1], [2]. An emerging application is the use of condition number estimates for dropping and pivoting in incomplete matrix decompositions which we will mention later.

In order to have useful condition estimators, they should be cheap. At the same time, they should provide condition number approximations which are reasonably accurate, and this may mean different things in different applications. Sometimes, relatively rough estimates are satisfactory, e.g., it is sufficient in many cases that the estimates stay within a reasonable multiplicative factor from the exact condition number, see, e.g. [17]. In other cases, more precise estimates may be needed [28].

Condition number estimators typically provide lower bounds on the condition number of a nonsingular matrix A by estimating a lower bound on the norm of A and an upper bound on the norm of A^{-1}. The most popular general approaches compute approximations of the condition

*Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 2, 18207 Praha 8, Czech Republic. This work was supported by the institutional support RVO:67985807 and by the grant P201/13-06684S of the Grant Agency of the Czech Republic.

number in the 1-norm [23], [26], [25], [27]. An important milestone in the development of estimators in the 2-norm was the incremental condition estimation (ICE) of a triangular matrix that was introduced in a series of papers by Bischof [4], [6] and further generalized for solving related tasks [7], [37]. This strategy is naturally connected to adaptive techniques and contains clearly visible links to matrix decompositions. As mentioned by Stewart [38], the approach can be viewed as a special case of the framework in [15]. A closely related approach called incremental norm estimation (INE) was developed by Duff and Vömel in [19] to get an estimation of the norm of a triangular matrix. While a slight reformulation of this algorithm similarly as in [4] can be used to estimate the minimum singular value as well, we will see in this paper that this does not work well in practice and we will give a partial explanation for this. Nevertheless, when the inverses of the triangular factors of A are available, INE could be used to get a useful estimate for the minimum singular value of A [19]. A similar conclusion follows for the recent iterative procedure to get a lower bound for the minimum singular value given in the interesting paper [30]. The actual strategy is based on an improvement of the algorithm in [20] and applied to symmetric and positive definite Toeplitz matrices. Incremental condition estimation is also closely related to rank-revealing decompositions, see, e.g., [33].

A strong motivation to study and further develop incremental condition estimators is their applicability in incomplete decompositions. In particular, a part of recent advances in preconditioning of systems of linear algebraic equations is based on monitoring the conditioning of the partially computed factors via a condition estimator. The incremental nature of the estimator enables to monitor and control both dropping and pivoting of the decomposition. This is done in strategies developed by Bollhöfer and Saad [8, 9, 11] and implemented in the package ILUPACK [10], see also their use in the multilevel framework [12]. Both perturbation arguments and experiments point out that preconditioners from incomplete decompositions using dropping criteria based on conditioning control are rather robust, but we believe that more accurate incremental strategies may help to push the approach even further. Note that the use of ICE for pivoting in decompositions was considered also much earlier, see, e.g., [5] but the significant progress in this research direction is connected with the work of Bollhöfer and Saad.

Recently, incomplete decompositions that compute both direct and inverse factors were introduced. That is, they compute not only the standard Cholesky or LU factors but also their inverses. In [3] the authors propose to compute the inverse of the incomplete factor once the direct factor is computed. The mixed direct-inverse decompositions in [13], [14] obtain the direct and inverse factors simultaneously, enabling to exploit information from the partial inverse factor for the computation of the direct factor and vice-versa. It was shown that despite rather sophisticated implementation, typical computational costs of the decomposition may still be low. Moreover, condition estimators can be applied to both the direct and the inverse factor, thus enabling to use the more accurate condition estimators discussed in this paper.

This paper presents some new theoretical and practical results leading to an improved incremental condition estimator in the 2-norm. As it is well-known that the strengths of different condition estimators are often complementary and any one of them can sometimes fail, we do not propose a strategy that is always better than all the other approaches, but we have rather strong theoretical and computational evidence to propose a choice based on INE. In the paper, we will show some theoretical results related to the condition estimators introduced in [4] and [19] as well as the mutual relation of these estimators. In particular, we will show that the best strategy should be based on the computed factor as well as its inverse. We will remind that factorizations that could be used for this task are readily available. The paper is organized as follows. In Section 2 the two basic strategies for incremental condition estimation in the 2-norm are introduced. Section 3 provides new theoretical results on the ICE and INE estimators. In particular, it reveals the strong potential of the INE algorithm using the factor as well as its inverse. Section 4 then analyzes reasons for the superiority of INE over ICE that is clear from both the graphical demonstration in this section and from the numerical experiments in Section 5. In the sections to follow, we will assume that A is real and $\|\cdot\|$ will denote the 2-norm. With "direct factor" we will mean a triangular Cholesky, L, or U factor of a given input matrix, as opposed to its inverse, the "inverse factor".

2. Incremental condition estimators in the 2-norm. This section presents a brief overview of the two basic incremental strategies to estimate the 2-norm condition number of a triangular matrix. The idea is to find an upper bound estimate σ_{min}^{est} of its smallest singular value and a lower bound estimate σ_{max}^{est} of its largest singular value. The condition number estimate is then $\sigma_{max}^{est}/\sigma_{min}^{est}$. Without loss of generality we assume our matrix to be upper triangular. By the *incremental* nature of the estimates we mean that the estimate for the leading principal submatrix $\hat{R}$ of dimension $k+1$ is obtained from the estimate for its leading principal submatrix R of dimension k in a simple way, without explicitly accessing the entries of R. In order to be able to do this, we also keep estimates of the corresponding singular vectors. Note that the basic matrix decompositions like Cholesky or LU reveal the triangular factors just in this incremental way and the incremental estimates may be used not only to form the final condition number estimate but they may be exploited throughout the decomposition.

Let us use the following notation

$$\hat{R} = \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix}. \tag{2.1}$$

As mentioned above, the first incremental estimation strategy of this kind was proposed by Bischof [4] in 1990 and called incremental condition estimation (ICE). This method computes approximations to the extremal singular values and to left singular vectors of triangular leading principal submatrices. Note that if $R = U\Sigma V^T$ is the singular value decomposition of R, an extremal left singular vector u_{ext} satisfies $\|u_{ext}^T R\| = \|u_{ext}^T U\Sigma V^T\| = \sigma_{ext}(R)$ with σ_{ext} denoting the extremal singular value. The ICE method computes

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

where ext is substituted for either min or max and y_{ext} denotes a left singular vector approximation. The superscript C here means the considered ICE incremental strategy that can be described as follows. Consider the submatrix $\hat{R}$. The algorithm computes the approximation $\sigma_{ext}^C(\hat{R})$ from the optimization problem

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

where the approximation $\hat{y}_{ext}$ of the left singular vector of $\hat{R}$ is

$$\hat{y}_{ext} \equiv \begin{bmatrix} s_{ext}\, y_{ext} \\ c_{ext} \end{bmatrix}.$$

It can be easily verified that s_{ext} and c_{ext} are the components of the eigenvector corresponding to the extremal (minimum or maximum) eigenvalue of the matrix

$$B_{ext}^C \equiv \begin{bmatrix} \sigma_{ext}^C(R)^2 + (y_{ext}^T v)^2 & \gamma(y_{ext}^T v) \\ \gamma(y_{ext}^T v) & \gamma^2 \end{bmatrix}. \tag{2.2}$$

If B^C_{ext} has two identical eigenvalues, the algorithm of [4] puts $s_{ext} = 0$ and $c_{ext} = 1$. Further,

$$\sigma^C_{ext}(\hat{R}) \equiv \|\hat{y}^T_{ext}\hat{R}\| = \sqrt{\lambda_{ext}(B^C_{ext})},$$

where λ_{ext} denotes the extremal (minimum or maximum) eigenvalue. Clearly, the involved eigenvectors are computed without accessing R. Note that the original derivation in [4] uses a lower triangular matrix and it is slightly different from the one presented here, see [19].

Another incremental strategy was proposed in 2002 by Duff and Vömel [19] and used only for norm estimation based on a maximization problem, although it is possible to formulate the dual minimization problem as well. We will denote it here by the acronym INE (incremental norm estimation) using the superscript N. It computes approximations $\sigma^N_{ext}(R)$ of the extremal singular values

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

as well as the corresponding INE approximations z_{ext} to the *right* singular vectors. Similarly as above, $\sigma^N_{ext}(\hat{R})$ is obtained from the following optimization problem

$$\|\hat{R}\hat{z}_{ext}\| = \mathrm{ext}_{\|[s,c]\|=1} \left\| \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} s\, z_{ext} \\ c \end{bmatrix} \right\| = \left\| \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} s_{ext} z_{ext} \\ c_{ext} \end{bmatrix} \right\|$$

where the approximation $\hat{z}_{ext}$ of the right singular vector of $\hat{R}$ is

$$\hat{z}_{ext} \equiv \begin{bmatrix} s_{ext}\, z_{ext} \\ c_{ext} \end{bmatrix}.$$

The scalars s_{ext} and c_{ext} are then the entries of the eigenvector corresponding to the extremal (minimum or maximum) eigenvalue of the matrix

$$B^N_{ext} \equiv \begin{bmatrix} \sigma^N_{ext}(R)^2 & z^T_{ext}R^T v \\ z^T_{ext}R^T v & v^T v + \gamma^2 \end{bmatrix}, \tag{2.3}$$

with the convention that $s_{ext} = 0$ and $c_{ext} = 1$ when B^N_{ext} has two identical eigenvalues. Then

$$\sigma^N_{ext}(\hat{R}) \equiv \|\hat{R}\hat{z}_{ext}\| = \sqrt{\lambda_{ext}(B^N_{ext})}.$$

In the remaining text we will further simplify the notation as follows. The subscripts *min* or *max* denoting minimum or maximum, respectively, such as s_{max} or y_{min} will be replaced by plus or minus signs which gives in this example $s_+ \equiv s_{max}$ and $y_- \equiv y_{min}$.

Note that the main costs involved in both techniques come from the inner products needed to get the entries of the matrices B^C_{ext} and B^N_{ext}. For a dense triangular matrix of dimension n the total costs to obtain its estimate are of the order n^2. Further, the above descriptions give no clear indication about whether one technique is superior to the other. In [19] the authors conclude, based on their experiments, that there is no general superiority of one technique. They explain that INE is more suited for sparse matrices and they show experimentally that INE is slightly superior for finding the largest singular value of dense triangular matrices. The following sections contain, among others, new theoretical comparisons of the quality of the two described techniques and a strong numerical confirmation of our findings.

3. ICE and INE estimates using both direct and inverse factors. Let us consider ICE and INE in the situation when we have both the direct triangular factor and its inverse available. In this section we are interested to know whether exploiting the inverse factor may help to improve accuracy of the estimates. At first sight this may seem trivial since the hard part in the estimation is often to find a good approximation of the minimum singular value. If the inverse is available, the problem can be circumvented by estimating the maximum singular value of the inverse. However, we will see that the two considered techniques behave differently in this respect.

Note that the inverse or its approximation is naturally available in the mixed direct-inverse decompositions [13], [14] mentioned in the introduction. In addition, information on rows and/or columns of the inverse is computed when applying the techniques of [8, 9, 11]. In some other applications, for example in signal processing [16, 32], it is necessary to compute the matrix inverses explicitly and this is traditionally done via their triangular factors. Further, the inversion of a triangular factor can be done at costs that are low compared to the computation of the factor. For example, the algorithm in [19, Lemma 3.1] asks for about $n^3/6$ flops, see also the techniques in [39].

First we will show that using the inverse triangular factor does not give any improvement for ICE. Let us start with a simple lemma related to the exact singular values and vectors.

LEMMA 3.1. *Let R be a nonsingular matrix. Then the extremal singular values of R and R^{-1} satisfy $\sigma_-(R) = 1/\sigma_+(R^{-1})$. The corresponding left singular vectors y_- and x_+ of R and R^{-1}, respectively, satisfy*

$$\sigma_-(R)x_+^T = y_-^T R. \tag{3.1}$$

Proof. The first part of the assertion is trivial. Let $R = USW^T$ be the SVD of R with the singular values in S in non-ascending order. Then $R^{-1} = WS^{-1}U^T$ and the left singular vectors y_- and x_+ can be expressed as $y_- = Ue_n$ and $x_+ = We_n$, respectively. Then we can write $x_+^T R^{-1} = e_n^T W^T R^{-1} = e_n^T W^T W S^{-1} U^T = (1/\sigma_-(R))e_n^T U^T = (1/\sigma_-(R))y_-^T$, which implies (3.1). ∎

The main result relating the ICE estimates for R and R^{-1} looks similarly.

THEOREM 3.2. *Let R be a nonsingular upper triangular matrix. Then the ICE estimates of the singular values of R and R^{-1} satisfy*

$$\sigma_-^C(R) = 1/\sigma_+^C(R^{-1}). \tag{3.2}$$

The approximate left singular vectors y_- and x_+ corresponding to the ICE estimates for R and R^{-1}, respectively, satisfy

$$\sigma_-^C(R)x_+^T = y_-^T R. \tag{3.3}$$

Proof. Consider mathematical induction on the dimension n of R. Clearly, the estimates are exact for $n = 1, 2$. Assume that the lemma holds for some $n \geq 2$ and we will prove it for $n+1$. Let us use the notation (2.1) for the upper triangular $\hat{R}$ of dimension $n + 1$. The estimate $\sigma_-^C(\hat{R})$ for the extended matrix $\hat{R}$ is obtained as the square root of the minimum eigenvalue of the matrix B_-^C given above in (2.2) where "$ext \equiv min \equiv -$". Clearly, B_-^C has the following $L^T L$ decomposition.

$$B_-^C = \begin{bmatrix} \sigma_-^C(R)^2 + (y_-^T v)^2 & \gamma(y_-^T v) \\ \gamma(y_-^T v) & \gamma^2 \end{bmatrix} = \begin{bmatrix} \sigma_-^C(R) & y_-^T v \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} \sigma_-^C(R) & 0 \\ y_-^T v & \gamma \end{bmatrix} \equiv (L_-^C)^T L_-^C.$$

Further, the estimate $1/\sigma_+^C(\hat{R}^{-1})$ for

$$\hat{R}^{-1} = \begin{bmatrix} R^{-1} & -R^{-1}v/\gamma \\ 0 & 1/\gamma \end{bmatrix}$$

is the square root of $1/\lambda_+(B_+^C)$ where B_+^C is defined with respect to $\hat{R}^{-1}$. This value is also equal to the square root of $\lambda_-((B_+^C)^{-1})$. Using the assumptions (3.2) and (3.3), from (2.2) we subsequently get

$$\begin{aligned}
(B_+^C)^{-1} &= \begin{bmatrix} \left(\sigma_+^C(R^{-1})\right)^2 + (-x_+^T R^{-1}v/\gamma)^2 & -(x_+^T R^{-1}v)/\gamma^2 \\ -(x_+^T R^{-1}v)/\gamma^2 & 1/\gamma^2 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} 1/(\sigma_-^C(R))^2 + ((y_-^T v)^2/(\sigma_-^C(R))^2\gamma^2) & -y_-^T v/(\sigma_-^C(R)\gamma^2) \\ -y_-^T v/(\sigma_-^C(R)\gamma^2) & 1/\gamma^2 \end{bmatrix}^{-1} \\
&= \left(\begin{bmatrix} 1/\sigma_-^C(R) & -y_-^T v/(\sigma_-^C(R)\gamma) \\ 0 & 1/\gamma \end{bmatrix} \begin{bmatrix} 1/\sigma_-^C(R) & 0 \\ -y_-^T v/(\sigma_-^C(R)\gamma) & 1/\gamma \end{bmatrix} \right)^{-1} \\
&= \begin{bmatrix} \sigma_-^C(R) & 0 \\ y_-^T v & \gamma \end{bmatrix} \begin{bmatrix} \sigma_-^C(R) & y_-^T v \\ 0 & \gamma \end{bmatrix}.
\end{aligned}$$

Clearly, we obtained the LL^T decomposition $(B_+^C)^{-1} = L_+^C(L_+^C)^T$ where L_+^C is the same as the factor L of the $L^T L$ decomposition of B_-^C. That is, we have $L \equiv L_+^C = L_-^C$. It is easy to see from the singular value decomposition $U_L S W_L^T$ of L that B_+^C and $(B_-^C)^{-1}$ have the same eigenvalues. This implies the first part (3.2) of the theorem.

The approximate singular vectors for the extended problems are

$$\hat{y}_- = \begin{bmatrix} s_-\, y_- \\ c_- \end{bmatrix}, \qquad \hat{x}_+ = \begin{bmatrix} s_+\, x_+ \\ c_+ \end{bmatrix},$$

where $\begin{bmatrix} s_-, & c_- \end{bmatrix}^T$ is the eigenvector of $B_-^C = (L_-^C)^T L_-^C$ corresponding to its minimum eigenvalue and $\begin{bmatrix} s_+, & c_+ \end{bmatrix}^T$ is the eigenvector of $B_+^C = (L_+^C)^{-T}(L_+^C)^{-1}$ corresponding to its maximum eigenvalue. Then $\begin{bmatrix} s_-, & c_- \end{bmatrix}^T = W_L e_2$ is also the right singular vector of L_-^C with the singular value $\sigma_-^C(\hat{R})$. Similarly, $\begin{bmatrix} s_+, & c_+ \end{bmatrix}^T$ is equal to $U_L e_2$ and it is also the the right singular vector of $(L_+^C)^{-1}$ with the singular value $\sigma_+^C(\hat{R}^{-1}) = 1/\sigma_-^C(\hat{R})$. Taking all of these into account, we get

$$\begin{aligned}
\hat{y}_-^T \hat{R} &= \begin{bmatrix} s_- y_-^T, & c_- \end{bmatrix} \begin{bmatrix} R & v \\ 0 & \gamma \end{bmatrix} = \begin{bmatrix} s_- y_-^T R, & s_- y_-^T v + c_-\gamma \end{bmatrix} = \begin{bmatrix} \sigma_-^C(R) s_- x_+^T, & s_- y_-^T v + c_-\gamma \end{bmatrix} \\
&= \begin{bmatrix} \sigma_-^C(R) s_-, & s_- y_-^T v + c_-\gamma \end{bmatrix} \begin{bmatrix} x_+^T & \\ & 1 \end{bmatrix} = \begin{bmatrix} s_-, & c_- \end{bmatrix} (L_-^C)^T \begin{bmatrix} x_+^T & \\ & 1 \end{bmatrix} \\
&= e_2^T W_L^T W_L S U_L^T \begin{bmatrix} x_+^T & \\ & 1 \end{bmatrix} = \sigma_-^C(\hat{R}) e_2^T U_L^T \begin{bmatrix} x_+^T & \\ & 1 \end{bmatrix} = \sigma_-^C(\hat{R}) \hat{x}_+^T.
\end{aligned}$$

We remark that the previous equalities also hold in the special case where B_-^C has two identical eigenvalues and where ICE defines $\begin{bmatrix} s_-, & c_- \end{bmatrix}^T = e_2^T$ and $\begin{bmatrix} s_+, & c_+ \end{bmatrix}^T = e_2^T$. ◻

Note that we can prove analogously that $\sigma_+^C(R) = 1/\sigma_-^C(R^{-1})$. Hence the ICE estimate of the condition number of R is always identical with the reciprocal of the ICE estimate of the condition number of R^{-1}. Now let us consider the alternative incremental norm estimation technique. INE deals with the *right* singular vectors of a triangular matrix. The following lemma is just an analogue of Lemma 3.1 for right singular vectors.

LEMMA 3.3. *Let R be a nonsingular matrix. Then the extremal singular values of R and R^{-1} satisfy $\sigma_-(R) = 1/\sigma_+(R^{-1})$. The corresponding right singular vectors z_- and x_+ of R and R^{-1}, respectively, satisfy*

$$\sigma_-(R)x_+ = Rz_-. \tag{3.4}$$

Proof. As above, the first part is trivial. Let $R = USW^T$ be the SVD of R with the singular values in S in non-ascending order. Clearly, $z_- = We_n$. Since $R^{-1} = WS^{-1}U^T$ we also have $x_+ = Ue_n$. Furthermore, $R^{-1}Ue_n = WS^{-1}U^TUe_n = 1/(\sigma_-(R))We_n$. This immediately implies (3.4). ◻

The following theorem shows that INE is inherently different from ICE and it reveals that there is no analogy with Theorem 3.2. In particular, Theorem 3.4 cannot be applied recursively for leading principal submatrices of growing dimension because the assumption $1/\sigma_+^N(R^{-1}) = \sigma_-^N(R)$ will in general cease to hold.

THEOREM 3.4. *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 (2.1) satisfy*

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

with equality if and only if v in (2.1) is collinear with the left singular vector corresponding to the smallest singular value of R.

Proof. Consider the INE process applied to $\hat{R}$. The estimate $\sigma_-^N(\hat{R})$ is given by the square root of the minimum eigenvalue of the matrix B_-^N obtained from (2.3) by setting "$ext \equiv min \equiv -$", which is also equal to the inverse of the square root of the maximum eigenvalue of the matrix $(B_-^N)^{-1}$. The L^TL decomposition of the matrix $(B_-^N)^{-1}$ is derived as follows using also Lemma 3.3 and its notation.

$$\begin{aligned}(B_-^N)^{-1} &= \begin{bmatrix} z_-^TR^TRz_- & v^TRz_- \\ v^TRz_- & v^Tv+\gamma^2 \end{bmatrix}^{-1} = \begin{bmatrix} \sigma_-(R)^2 & \sigma_-(R)v^Tx_+ \\ \sigma_-(R)v^Tx_+ & v^Tv+\gamma^2 \end{bmatrix}^{-1} \\ &= \left(\begin{bmatrix} \sigma_-(R) & 0 \\ v^Tx_+ & \sqrt{v^Tv-(v^Tx_+)^2+\gamma^2} \end{bmatrix} \begin{bmatrix} \sigma_-(R) & v^Tx_+ \\ 0 & \sqrt{v^Tv-(v^Tx_+)^2+\gamma^2} \end{bmatrix} \right)^{-1} = L_-^TL_-\end{aligned}$$

with

$$L_- = \begin{bmatrix} 1/\sigma_-(R) & 0 \\ -v^Tx_+/\left(\sigma_-(R)\sqrt{v^Tv-(v^Tx_+)^2+\gamma^2}\right) & 1/\sqrt{v^Tv-(v^Tx_+)^2+\gamma^2} \end{bmatrix}.$$

Further, the INE estimate for $1/\sigma_+^N(\hat{R}^{-1})$ is obtained from the eigenvalues of the matrix B_+^N which can be put down and represented in the form of a LL^T decomposition. Its derivation uses the fact that $\sigma_-(R)R^{-T}z_- = x_+$, which can be easily seen from the singular value decomposition $R = USW^T$ with $z_- = We_n$ and $x_+ = Ue_n$. Then with Lemma 3.3, $R^{-T}R^{-1}x_+ = x_+/\sigma_-(R)^2$. A few simple steps provide

$$\begin{aligned}B_+^N &= \begin{bmatrix} x_+^TR^{-T}R^{-1}x_+ & -x_+^TR^{-T}R^{-1}v/\gamma \\ -x_+^TR^{-T}R^{-1}v/\gamma & v^TR^{-T}R^{-1}v/\gamma^2+1/\gamma^2 \end{bmatrix} \\ &= \begin{bmatrix} 1/\sigma_-(R)^2 & -v^Tx_+/(\sigma_-(R)^2\gamma) \\ -v^Tx_+/(\sigma_-(R)^2\gamma) & (||R^{-1}v||^2+1)/\gamma^2 \end{bmatrix} = L_+L_+^T\end{aligned}$$

with

$$L_+ = \begin{bmatrix} 1/\sigma_-(R) & 0 \\ -v^T x_+/(\sigma_-(R)\gamma) & \left(\sqrt{||R^{-1}v||^2 - (v^T x_+)^2/\sigma_-(R)^2 + 1}\right)/\gamma \end{bmatrix}.$$

The Cauchy inequality $(v^T x_+)^2 \leq v^T v$ and properties of the singular value decomposition imply

$$\|R^{-1}v\|^2 = \|S^{-1}U^T v\|^2 = \sum_{j=1}^{n} \frac{(e_j^T U^T v)^2}{s_{jj}^2} \geq (v^T x_+)^2/\sigma_-(R)^2. \tag{3.5}$$

This implies the relation

$$\|L_-\| = \left\| \begin{bmatrix} 1 & \\ & \frac{\gamma}{\sqrt{v^T v - (v^T x_+)^2 + \gamma^2}} \end{bmatrix} L_+ \begin{bmatrix} 1 & \\ & \frac{1}{\sqrt{||R^{-1}v||^2 - (v^T x_+)^2/\sigma_-^N(R)^2 + 1}} \end{bmatrix} \right\| \leq \|L_+\|.$$

The involved norms of the triangular factors directly provide

$$\left(\sigma_+^N(\hat{R}^{-1})\right)^{-1} = \|L_+\|^{-1} \leq \|L_-\|^{-1} = \sigma_-^N(\hat{R}). \tag{3.6}$$

Equality in (3.6) is attained if and only if $(v^T x_+)^2/(\sigma_-(R))^2 = \|R^{-1}v\|^2$ and also $(v^T x_+)^2 = v^T v$. These two conditions are equivalent with the collinearity of v with $x_+ = Ue_n$. $\square$

We can obtain the analogue result for the approximate *largest* singular value $\sigma_+^N(\hat{R})$ if we consider in Theorem 3.4 instead of $\hat{R}$ its inverse. Let us denote the inverse of $\hat{R}$ by $\hat{S}$. Using Theorem 3.4 we get $\sigma_-^N(\hat{R}) = \sigma_-^N(\hat{S}^{-1}) \geq 1/\sigma_+^N(\hat{R}^{-1}) = 1/\sigma_+^N(\hat{S})$, i.e. for any upper triangular S with $1/\sigma_-^N(S^{-1}) = \sigma_+^N(S) = \sigma_+(S)$ we have for the extended matrix $\hat{S}$

$$\sigma_+^N(\hat{S}) \geq 1/\sigma_-^N(\hat{S}^{-1}). \tag{3.7}$$

Consequently, under the assumption of starting with exact estimates like in Theorem 3.4, INE will be more accurate when estimating σ_-, respectively σ_+, if one applies incremental *maximization* (using $1/\sigma_+^N$ or σ_+^N, respectively) instead of incremental *minimization* (using σ_-^N or $1/\sigma_-^N$, respectively). This is in contrast with the ICE technique, where maximization and minimization give identical approximations in the sense of (3.2). When the inverse is not available, Theorem 3.4 and (3.7) seem to suggest that the quality of the INE estimate of the largest singular value might in most cases be better than the quality of the estimate for the smallest singular value. Further, Theorem 3.4 and (3.7) assume that the INE estimates of the singular values of R and R^{-1} are exact. Our experiments suggest that even in the more general situation when the assumptions of Theorem 3.4 may not hold, minimization works better than maximization very rarely in practice. In fact, in our tests with various types of matrices traditionally used to asses the quality of incremental condition estimators and with matrices from the Matrix Market collection [18] this never occurred. In order to better understand this behavior, we propose to consider the following expressions for $1/\sigma_+^N(\hat{R}^{-1})$ and $\sigma_-^N(\hat{R})$.

PROPOSITION 3.5. *Let R be a nonsingular upper triangular matrix and let the INE approximate singular vectors for $\sigma_+^N(R^{-1})$ and $\sigma_-^N(R)$ be denoted by x_+ and z_-, respectively. Then the INE estimates of the singular values related to the extended matrix (2.1) satisfy*

$$\sigma_-^N(\hat{R}) = \sigma_-(L_-^N), \qquad L_-^N = \begin{bmatrix} \sigma_-^N(R) & 0 \\ \iota_- & \sqrt{\gamma^2 + v^T v - \iota_-^2} \end{bmatrix}, \quad \iota_- = v^T R z_-/\sigma_-^N(R)$$

and

$$1/\sigma_+^N(\hat{R}^{-1}) = \sigma_-(L_+^N), \qquad L_+^N = \begin{bmatrix} 1/\sigma_+^N(R^{-1}) & 0 \\ \frac{\iota_+}{\sqrt{\|R^{-1}v\|^2-(\frac{\iota_+}{\sigma_+})^2+1}} & \frac{\gamma}{\sqrt{\|R^{-1}v\|^2-(\frac{\iota_+}{\sigma_+})^2+1}} \end{bmatrix},$$

where $\sigma_+ = \sigma_+^N(R^{-1})$, $\iota_+ = v^T R^{-T} R^{-1} x_+ / \sigma_+^2$.

Proof. The estimate $\sigma_-^N(\hat{R})$ is given by the root of the minimum eigenvalue of the matrix B_-^N obtained from (2.3) by setting "*ext* $\equiv$ *min* $\equiv -$". The Cholesky decomposition of the matrix B_-^N is

$$\begin{aligned} B_-^N &= \begin{bmatrix} z_-^T R^T R z_- & v^T R z_- \\ v^T R z_- & v^T v + \gamma^2 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_-^N(R) & 0 \\ \iota_- & \sqrt{\gamma^2 + v^T v - \iota_-^2} \end{bmatrix} \begin{bmatrix} \sigma_-^N(R) & \iota_- \\ 0 & \sqrt{\gamma^2 + v^T v - \iota_-^2} \end{bmatrix} = L_-^N (L_-^N)^T. \end{aligned}$$

This gives $\sigma_-^N(\hat{R}) = \sigma_-(L_-^N)$. Similarly, the estimate $1/\sigma_+(\hat{R}^{-1})^N$ is given by the root of the minimum eigenvalue of the matrix $(B_+^N)^{-1}$ obtained from (2.3) and defined with respect to $\hat{R}^{-1}$ by setting "*ext* $\equiv$ *max* $\equiv +$". The $L^T L$ decomposition of the matrix $(B_+^N)^{-1}$ is

$$\begin{aligned} (B_+^N)^{-1} &= \begin{bmatrix} x_+^T R^{-T} R^{-1} x_+ & -v^T R^{-T} R^{-1} x_+/\gamma \\ -v^T R^{-T} R^{-1} x_+/\gamma & v^T R^{-T} R^{-1} v/\gamma^2 + 1/\gamma^2 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \sigma_+^N(R^{-1})^2 & -\iota_+ \sigma_+^N(R^{-1})^2/\gamma \\ -\iota_+ \sigma_+^N(R^{-1})^2/\gamma & \|R^{-1}v\|^2/\gamma^2 + 1/\gamma^2 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \sigma_+^N(R^{-1}) & -\iota_+ \sigma_+^N(R^{-1})/\gamma \\ 0 & \sqrt{||R^{-1}v||^2 - \iota_+^2 \sigma_+^N(R^{-1})^2 + 1}/\gamma \end{bmatrix}^{-1} \\ &\quad \begin{bmatrix} \sigma_+^N(R^{-1}) & 0 \\ -\iota_+ \sigma_+^N(R^{-1})/\gamma & \sqrt{||R^{-1}v||^2 - \iota_+^2 \sigma_+^N(R^{-1})^2 + 1}/\gamma \end{bmatrix}^{-1} = (L_+^N)^T L_+^N. \end{aligned}$$

The claim follows from

$$L_+^N = \begin{bmatrix} 1/\sigma_+^N(R^{-1}) & 0 \\ \iota_+/\sqrt{||R^{-1}v||^2 - \iota_+^2/(\sigma_+^N(R^{-1}))^2 + 1} & \gamma/\sqrt{||R^{-1}v||^2 - \iota_+^2/(\sigma_+^N(R^{-1}))^2 + 1} \end{bmatrix}.$$

∎

For a partial explanation why maximization seems in general to outperform minimization, let us compare the entries of the matrices L_-^N and L_+^N defined in Proposition 3.5. Since we have $i_-^2 \le v^T v$ and $\iota_+^2/(\sigma_+^N(R^{-1}))^2 \le ||R^{-1}v||^2$, the second diagonal entry of L_+^N is always smaller than that of L_-^N. When the dimension of $\hat{R}$ is two, the first diagonal entries of L_-^N are L_+^N identical at the beginning of the estimation process, because they are exact. When $\hat{R}$ has dimension three, the first diagonal entry of L_+^N is not larger than that of L_-^N from Theorem 3.4. Further, when started with $1/\sigma_+^N(R^{-1}) \le \sigma_-^N(R)$, in order for $1/\sigma_+^N(\hat{R}^{-1}) \le \sigma_-^N(\hat{R})$ to hold it clearly suffices that the off-diagonal entries of L_+^N and L_-^N satisfy the simple inequality stated in the following corollary.

COROLLARY 3.6. *Using the notation of Proposition 3.5 and assuming $1/\sigma_+^N(R^{-1}) \leq \sigma_-^N(R)$, there holds*

$$1/\sigma_+^N(\hat{R}^{-1}) \leq \sigma_-^N(\hat{R}) \qquad \textit{if} \qquad |\iota_-| \leq \left| \frac{\iota_+}{\sqrt{\|R^{-1}v\|^2 - (\frac{\iota_+}{\sigma_+})^2 + 1}} \right|. \tag{3.8}$$

The following example shows that the sufficient condition in the previous corollary may be possibly simplified but it cannot be removed. Let us consider matrices R and R^{-1} defined as follows:

$$R = \begin{bmatrix} 2 & 0 & 1 \\ & 1 & 0 \\ & & 1 \end{bmatrix}, \qquad R^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ & 1 & 0 \\ & & 1 \end{bmatrix}.$$

The ICE estimate $\sigma_-^C(R)$ for the smallest singular value $\sigma_-(R) = 0.874$ is $\sigma_-^C(R) = 1$. The ICE estimate $1/\sigma_+^C(R^{-1})$ is of the same value, i.e. $1/\sigma_+^C(R^{-1}) = 1$, which is in agreement with Theorem 3.2. Note that here we used a matrix in block angular form that does not pass the information in ICE as discussed in [6]. The INE estimate $\sigma_-^N(R)$ for the smallest singular value is also $\sigma_-^N(R) = 1$, but the INE estimate $1/\sigma_+^N(R^{-1})$ is more accurate since $1/\sigma_+^N(R^{-1}) = \sqrt{4/5} \approx 0.8944$. This is what one would expect from Theorem 3.4 (its assumptions are satisfied because the estimates for triangular matrices of size two are always exact).

Consider now an extended matrix $\hat{R}$ with $\gamma = 1$ in (2.1). The choice of v influences the values ι_- and ι_+ in Proposition 3.5, which can be crucial for whether $1/\sigma_+^N(\hat{R}^{-1}) < \sigma_-^N(\hat{R})$ holds, see Corollary 3.6. The INE approximation of the right singular vector z_- corresponding to $\sigma_-^N(R)$ is $z_- = [0,1,0]^T$, hence $\iota_- = (v^T R z_-)/\sigma_-^N(R) = v^T[0,1,0]^T$. Similarly, using the INE approximate right singular vector $x_+ = [0,0,1]^T$ corresponding to $1/\sigma_+^N(R^{-1})$ we arrive at $\iota_+ = (v^T R^{-T} R^{-1} x_+)/\sigma_+^N(R^{-1})^2 = 4/5 \cdot v^T[-1/4, 0, 5/4]^T$. Let us consider the vector $v = [1,1,1]^T$ giving

$$\hat{R} = \begin{bmatrix} 2 & 0 & 1 & 1 \\ & 1 & 0 & 1 \\ & & 1 & 1 \\ & & & 1 \end{bmatrix}, \qquad \sigma_-(\hat{R}) \approx 0.5155, \quad \iota_- = 1, \quad \iota_+ = 4/5, \qquad \text{and}$$

$$0.5381 \approx \left(\frac{17/4 + \sqrt{(17/4)^2 - 11}}{2} \right)^{-\frac{1}{2}} = 1/\sigma_+^N(\hat{R}^{-1}) < \sigma_-^N(\hat{R}) = \sqrt{\frac{5 - \sqrt{13}}{2}} \approx 0.835,$$

which is what one may expect from Proposition 3.5. Just note that the ICE estimates are

$$\sigma_-^C(\hat{R}) = 1/\sigma_+^C(\hat{R}^{-1}) = \sqrt{\frac{3 - \sqrt{5}}{2}} \approx 0.618.$$

We can, however, construct a case where the sufficient condition of Corollary 3.6 is not satisfied and $1/\sigma_+^N(\hat{R}^{-1}) > \sigma_-^N(\hat{R})$ by making ι_+ smaller. For instance, with $v = [0,1,0]^T$ we have $\iota_+ = 0$ and $\iota_- = 1$. The extended matrix is then

$$\hat{R} = \begin{bmatrix} 2 & 0 & 1 & 0 \\ & 1 & 0 & 1 \\ & & 1 & 0 \\ & & & 1 \end{bmatrix}, \qquad \text{with} \quad \sigma_-(\hat{R}) = \sqrt{\frac{3 - \sqrt{5}}{2}},$$

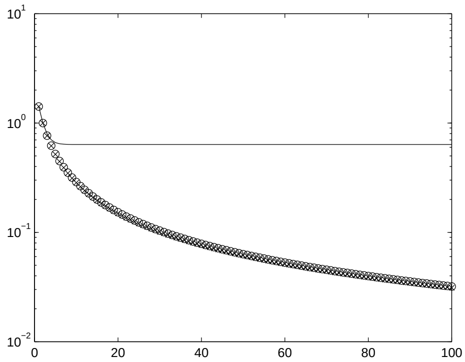

FIG. 3.1. *INE estimation of the smallest singular value of the 1D Laplacians of size one until hundred: INE with minimization (solid line), INE with maximization (circles) and exact minimum singular values (crosses).*

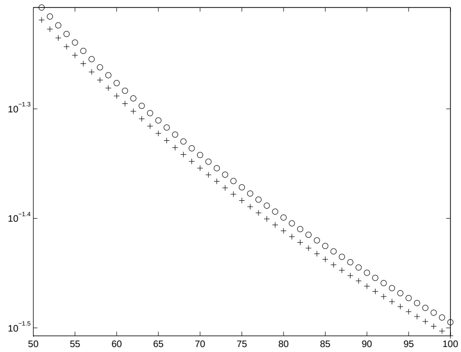

FIG. 3.2. *INE estimation of the smallest singular value of the 1D Laplacians of size fifty until hundred, zoom of Figure 3.1 for INE with maximization and exact minimum singular values.*

and we obtain

$$0.618 \approx \sqrt{\frac{3-\sqrt{5}}{2}} = \sigma_-^N(\hat{R}) < 1/\sigma_+^N(\hat{R}^{-1}) = \sqrt{\frac{1}{2}} \approx 0.7071.$$

The ICE estimates satisfy in this case $\sigma_-^C(\hat{R}) = 1/\sigma_+^C(\hat{R}^{-1}) = 1$.

This example might indicate that it is not too difficult to find academic examples where estimating $\sigma_-(\hat{R})$ by $\sigma_-^N(\hat{R})$ (i.e. with minimization) works better than using $1/\sigma_+^N(\hat{R}^{-1})$ (i.e. maximization). But as we mentioned before, we never observed this is practice. Let us give one striking example. In Figure 3.1 the crosses display the minimum singular value of the one-dimensional Laplacians $\mathcal{L}_i$, $i = 1, \ldots, 100$, of size one until hundred. The circles represent the INE estimates $1/\sigma_+^N(\mathcal{L}_i^{-1})$, $i = 1, \ldots, 100$, and they are very accurate (see also Figure 3.2 which is a zoom of Figure 3.1 for the INE estimates $1/\sigma_+^N(\mathcal{L}_i^{-1})$, $i = 50, \ldots, 100$). The solid line represents the INE estimates $\sigma_-^N(\mathcal{L}_i)$, $i = 1, \ldots, 100$ based on minimization. They stagnate around the value 0.6356.

Summarizing, we presented at the end of this section a number of results suggesting superiority of INE maximization based on the inverse of the triangular factor over INE minimization. A sound theoretical explanation for this phenomenon, which is often observed but for which counterexamples can be constructed (see above), is an open problem.

4. Superiority of INE maximization over ICE maximization. While the previous section concludes that the maximization problem in INE should be preferred for estimating both the maximum and the minimum singular value (exploiting the inverse), this section addresses the question whether the ICE technique can be more efficient than INE when the inverse is available. We did already prove that using the inverse does not improve the ICE technique (Theorem 3.2), but this does not mean that ICE estimates are worse than INE estimates exploiting the inverse. If ICE maximization were more powerful than INE maximization, there would hold, with the assumptions of Theorem 3.4,

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

and in that case also ICE *minimization* would be more powerful than INE minimization. We therefore concentrate on maximization. The subsequent text presents sufficient conditions for the opposite case, that is for superiority of INE maximization over ICE maximization. Extensive numerical experiments confirm that INE maximization is the method of choice. We also graphically demonstrate strength of the introduced sufficient conditions. Let us discuss INE and ICE maximization from the theoretical point of view first.

Similarly to the results of the previous section, we are not able to prove the superiority of INE unconditionally and counterexamples exist. This type of conclusions seems to be present in many areas connected with condition estimators that can sometimes fail. On the other hand we are just interested in proposing a strategy which would give as good results as possible on average and we believe that we are successful in this. We will see that both the theoretical arguments, the figures displayed in this section and also the results in the experimental section support the claim that INE maximization, is preferable over ICE maximization.

The theoretical arguments consist of the two following theorems that provide sufficient conditions for superiority of INE.

THEOREM 4.1. *Consider norm estimation of the extended matrix (2.1) where ICE and INE start with the same approximation $\sigma_+ \equiv \sigma_+^C(R) = \sigma_+^N(R)$. Let y be the corresponding approximate left singular vector, let z be the corresponding approximate right singular vector and let $w = Rz/\sigma_+$. Then the approximation $\sigma_+^N(\hat{R})$ obtained from INE is at least as good as the approximation $\sigma_+^C(\hat{R})$ from ICE if*

$$(v^T w)^2 \geq (v^T y)^2. \tag{4.1}$$

Proof. The largest eigenvalue of B_+^C from (2.2) (with the simplified notation introduced here) corresponds to the rightmost intersection of the parabola $\ell(\lambda) = \left(\lambda - \sigma_+^2 - (v^T y)^2\right)(\lambda - \gamma^2)$ with the horizontal line $h(\lambda) \equiv \gamma^2 (v^T y)^2$. Hence the largest eigenvalue λ_R of B_+^N from (2.3) is larger or equal to the leading eigenvalue of B_+^C if and only if

$$\ell(\lambda_R) \geq \gamma^2 (v^T y)^2. \tag{4.2}$$

The condition (4.2) corresponds to the case when INE maximization for $\hat{R}$ is at least as good as ICE maximization for the same matrix. Substituting

$$\lambda_R \equiv \frac{1}{2}\left(\sigma_+^2 + v^T v + \gamma^2 + S\right), \qquad S \equiv \sqrt{(\sigma_+^2 - \gamma^2 - v^T v)^2 + 4\sigma_+^2 (v^T w)^2} \tag{4.3}$$

into $\ell(\lambda_R)$ we have

$$\begin{aligned}
\ell(\lambda_R) &= \left(\gamma^2 - \lambda_R\right)\left(\sigma_+^2 + (v^T y)^2 - \lambda_R\right) \\
&= \frac{1}{4}\left(\gamma^2 - \sigma_+^2 - v^T v - S\right)\left(\sigma_+^2 + 2(v^T y)^2 - v^T v - \gamma^2 - S\right) \\
&= \frac{1}{4}\left((\gamma^2 - \sigma_+^2 - v^T v)(\sigma_+^2 - \gamma^2 - v^T v + 2(v^T y)^2) - 2((v^T y)^2 - v^T v)S + S^2\right).
\end{aligned}$$

Thus (4.2) is satisfied if and only if

$$(\gamma^2 - \sigma_+^2 - v^T v)(\sigma_+^2 - \gamma^2 - v^T v + 2(v^T y)^2) - 2((v^T y)^2 - v^T v)S + S^2 \;\geq\; 4\gamma^2 (y^T v)^2.$$

Substituting S^2 from (4.3) we can obtain

$$2\left(\sigma_+^2 - \gamma^2 + v^T v + S + 2\gamma^2\right)\left(v^T v - (v^T y)^2\right) - 4\sigma_+^2\left(v^T v - (v^T w)^2\right) \;\geq\; 0,$$

and after some rewriting we arrive at the equivalent condition

$$2\left(\gamma^2 - \sigma_+^2 + v^T v + S\right)\left(v^T v - (v^T y)^2\right) + 4\sigma_+^2\left((v^T w)^2 - (v^T y)^2\right) \geq 0 \tag{4.4}$$

that is equivalent with (4.2). The Cauchy inequality implies that $v^Tv - (v^Ty)^2 \geq 0$. If $v^Tv - (v^Ty)^2 = 0$ then we are done since (4.4) follows directly from (4.1).

Consider $v^Tv - (v^Ty)^2 > 0$. Let $\epsilon \geq 0$ be defined through

$$(v^Tw)^2 - (v^Ty)^2 = \epsilon \left(v^Tv - (v^Ty)^2\right). \tag{4.5}$$

Then (4.4) implies that (4.2) is satisfied if and only if

$$2\left(\gamma^2 + v^Tv + S + (2\epsilon - 1)\sigma_+^2\right) \geq 0, \tag{4.6}$$

that is, if and only if

$$S^2 = (\sigma_+^2 - \gamma^2 - v^Tv)^2 + 4\sigma_+^2(v^Tw)^2 \geq \left(\sigma_+^2 - \gamma^2 - v^Tv - 2\epsilon\sigma_+^2\right)^2.$$

Equivalently, (4.2) is valid with $v^Tv - (v^Ty)^2 > 0$ if and only if

$$\epsilon^2\sigma_+^2 - \epsilon(\sigma_+^2 - \gamma^2 - v^Tv) - (v^Tw)^2 \leq 0. \tag{4.7}$$

This is true for $\epsilon = 0$. But this means, in view of (4.6), that for $\epsilon = 0$

$$\gamma^2 + v^Tv + S + (2\epsilon - 1)\delta^2 \geq 0.$$

Consequently, for *all* $\epsilon \geq 0$,

$$\gamma^2 + v^Tv + S + (2\epsilon - 1)\delta^2 \geq 0.$$

□

The next theorem formulates an even stricter sufficient condition for the superiority of INE. This condition seems to be rather technical but it enables to specify more precisely the areas of parameters where one of the techniques is better than the other one. We will see that based on the input parameters of the condition estimator, there is always a possibility that the INE technique is better than ICE but not vice versa.

THEOREM 4.2. *Using the same notation and assumptions as in Theorem 4.1, the approximation $\sigma_+^N(\hat{R})$ obtained from INE is at least as good as the approximation $\sigma_+^C(\hat{R})$ from ICE if*

$$(v^Tw)^2 \geq \rho_1, \tag{4.8}$$

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

$$\begin{aligned}(v^Tw)^4 &+ \left(\frac{\gamma^2 + (v^Ty)^2}{\sigma_+^2}\left(v^Tv - (v^Ty)^2\right) - v^Tv - (v^Ty)^2\right)(v^Tw)^2 \\ &+ (v^Ty)^2\left(\frac{\gamma^2 + v^Tv}{\sigma_+^2}\left((v^Ty)^2 - v^Tv\right) + v^Tv\right) = 0.\end{aligned} \tag{4.9}$$

Proof. Assume for the moment that $v^Tv - (v^Ty)^2 > 0$. Let us substitute the expression for ϵ from (4.5) into the inequality (4.7). We get directly

$$\begin{aligned}\left(\frac{(v^Tw)^2 - (v^Ty)^2}{v^Tv - (v^Ty)^2}\right)^2\sigma_+^2 &- \frac{\left((v^Tw)^2 - (v^Ty)^2\right)\left(v^Tv - (v^Ty)^2\right)}{\left(v^Tv - (v^Ty)^2\right)^2}(\sigma_+^2 - \gamma^2 - v^Tv) \\ &- \frac{(v^Tw)^2\left(v^Tv - (v^Ty)^2\right)^2}{\left(v^Tv - (v^Ty)^2\right)^2} \leq 0,\end{aligned}$$

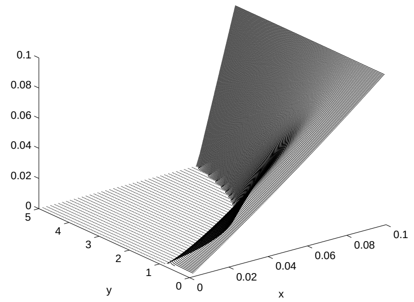

FIG. 4.1. *Value of ρ_1 in (4.8) in dependence of $(v^T y)^2$ (x-axis) and γ^2 (y-axis) with $v^T v = 0.1$, $\sigma_+ = 1$ and with $\Delta = 0$ in (4.13).*

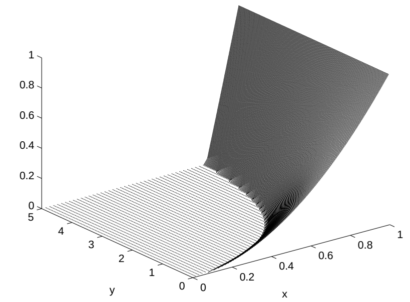

FIG. 4.2. *Value of ρ_1 in (4.8) in dependence of $(v^T y)^2$ (x-axis) and γ^2 (y-axis) with $v^T v = 1$, $\sigma_+ = 1$ and with $\Delta = 0$ in (4.13).*

and after a few simple steps we obtain the sufficient condition for the superiority of INE

$$\rho_1 \le (v^T w)^2 \le \rho_2, \tag{4.10}$$

where $\rho_{1,2}$ are the roots of (4.9). They have the form

$$(v^T y)^2 + \frac{(v^T v - (v^T y)^2)}{2\sigma_+^2} \left(\beta \pm \sqrt{\beta^2 + 4\sigma_+^2 (v^T y)^2} \right), \tag{4.11}$$

where $\beta = \sigma_+^2 - \gamma^2 - (v^T y)^2$. Clearly, we get

$$\rho_1 \le (v^T y)^2 \le \rho_2, \tag{4.12}$$

If $(v^T w)^2 < (v^T y)^2$ then (4.8) and (4.12) imply superiority of INE based on (4.10), otherwise Theorem 4.1 can be applied. Finally, if $v^T v - (v^T y)^2 = 0$, then the roots of (4.9) coincide and take the value $\rho_{1,2} = (v^T y)^2$, see (4.11). Hence the condition (4.8) reduces to (4.1) and again, Theorem 4.1 can be applied. □

An important conclusion of the previous theorems is as follows. Let us divide the possible input vectors v into two sets. The first set contains the v such that $(v^T w)^2 \ge (v^T y)^2$ and the second set contains the other instances of v. Then, the sufficient condition for superiority is *always* valid for all v from the first group and it is possibly valid also for some v from the second group. In particular, INE is never worse than ICE under the assumptions of these theorems whenever $\rho_1 \le 0$. We do not have a similar claim for superiority of ICE.

4.1. Graphical demonstration. In this subsection we graphically demonstrate the relation between ICE and INE maximization that points out the superiority of the latter approach. The presented figures depict on the z-axis the value $\max(0, \rho_1)$, that is, the sufficient condition for the superiority of INE estimation in (4.8), where we display $\max(0, \rho_1)$ because for $\rho_1 \le 0$ the condition is automatically satisfied. If we scale the matrix such that $\sigma_+ = 1$, and this can be always done without loss of generality, the coefficients of the equation (4.9) depend on three variables only. These three variables are $(v^T y)^2$, $v^T v$ and γ^2. Fixing $v^T v$, we can display the dependence on the other variables in the remaining two dimensions of the figures. We plot the values of $(v^T y)^2$ on the x-axis and γ^2 on the y-axis. For practical reasons, we restrict ourselves to $\gamma^2 \le 5$ but the behavior for larger values is more or less the same as for $\gamma^2 = 5$. Figures 4.1-4.3 display the values for three different choices of the norm $v^T v$. We know from Theorem 4.2 that INE is *unconditionally* (regardless of the vector w) superior over ICE for $\rho_1 \le 0$. In our pictures this case corresponds to

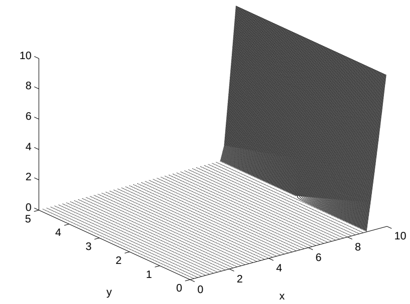

FIG. 4.3. *Value of ρ_1 in (4.8) in dependence of $(v^Ty)^2$ (x-axis) and γ^2 (y-axis) with $v^Tv = 10$, $\sigma_+ = 1$ and with $\Delta = 0$ in (4.13).*

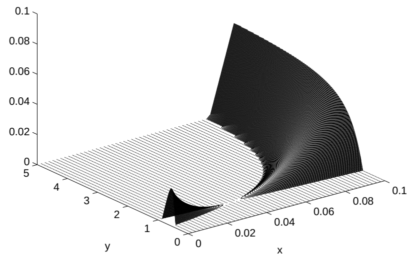

FIG. 4.4. *Value of ρ_1 in (4.8) in dependence of $(v^Ty)^2$ (x-axis) and γ^2 (y-axis) with $v^Tv = 0.1$, $\sigma_+ = 1$ and with $\Delta = 0.6$ in (4.13).*

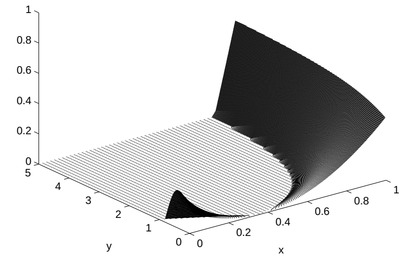

FIG. 4.5. *Value of ρ_1 in (4.8) in dependence of $(v^Ty)^2$ (x-axis) and γ^2 (y-axis) with $v^Tv = 1$, $\sigma_+ = 1$ and with $\Delta = 0.6$ in (4.13).*

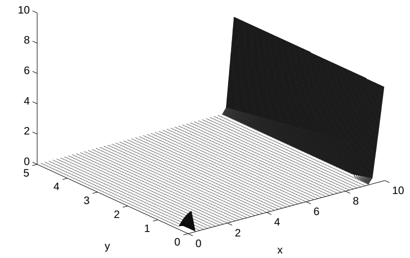

FIG. 4.6. *Value of ρ_1 in (4.8) in dependence of $(v^Ty)^2$ (x-axis) and γ^2 (y-axis) with $v^Tv = 10$, $\sigma_+ = 1$ and with $\Delta = 0.6$ in (4.13).*

its crosshatched part. In the other cases (dark part of the figures), the conclusion whether ICE or INE maximization is better still depends on the mutual relation of $(v^Tw)^2$ and $(v^Ty)^2$ and either of the techniques can be better than the other one.

Let us mention here that also a result similar to Theorem 4.2 could be derived that uses as an additional parameter the distance

$$\Delta \equiv \sqrt{(\sigma_+^N)^2 - (\sigma_+^C)^2}, \qquad \sigma_+^N \geq \sigma_+^C, \tag{4.13}$$

with $\sigma_+^N = \sigma_{max}^N(R)$ and $\sigma_+^C = \sigma_{max}^C(R)$. The previous case corresponds to the case $\Delta = 0$. The claims and proofs are very similar and we omit them here since they would not give an additional insight for our statement that INE maximization is preferable over ICE maximization. Nevertheless, just for illustration, we present here also figures for the same choices of values of $||v||$ and with nonzero Δ, here, $\Delta = 0.6$. In Figures 4.4-4.6 we can see that the results are as we would intuitively expect, $\Delta > 0$ seems to even increase the expectation for the superiority of INE over ICE.

Let us recall the one-dimensional Laplacian example from Section 3. It shows not only that INE maximization based on the inverse matrix may be very accurate, it also points out that the estimate of σ_-^N via INE minimization can be *very* poor. Therfore, if the plain ICE-based strategy is used without the matrix inverse to estimate both singular values, the condition number estimate is often better than if plain INE without inverse is used. In other words, experiments show that

INE minimization is by far the weakest point of the two investigated strategies. The explanation of this observation is an interesting open problem.

5. Numerical experiments. In this section we focus on illustrating the theoretical results in Sections 3 and 4. In particular, we confirm that using just maximization in INE seems to be a better strategy than using minimization as well. Further, we will see that ICE is clearly outperformed by INE using various matrix test sets. The experiments, all run in Matlab, show that the availability of the inverse inside the decomposition is desirable, but, except for the last experiment, we compute the inverse separately with Matlab's backslash command.

Our experiments compare the following four strategies:

1. The original ICE technique from [4] with the estimates defined as $\sigma_+^C(R)/\sigma_-^C(R)$.
2. The INE technique from [19] for estimating both the norm and the minimum singular value with the estimates defined by $\sigma_+^N(R)/\sigma_-^N(R)$. Although INE was originally proposed for norm estimation only, we refer to this estimator as to original INE.
3. The INE technique based on maximization only, that uses also the inverse R^{-1}, that is, estimates defined as $\sigma_+^N(R)\,\sigma_+^N(R^{-1})$.
4. The INE technique based on minimization only which uses the matrix inverse as well, that is $\left(\sigma_-^N(R)\,\sigma_-^N(R^{-1})\right)^{-1}$.

Note that we do not display any results for the estimates $\sigma_-^C(R^{-1})/\sigma_+^C(R^{-1})$ since, as we proved in Theorem 3.2, they are identical with the original ICE estimates.

5.1. Example 1. Using the Matlab command A=rand(100,100) - rand(100,100) we generated 50 matrices A of size 100, computed a column pivoting using colamd and obtained an upper triangular factor R from the QR decomposition of the column permuted matrix A. This is the same type of experiments as in [4, Section 4, Test 1]. The condition estimators were tested on R, see Figure 5.1. When omitting the column pivoting we get qualitatively the same picture.

We can see that the estimate $\sigma_+^N(R)\sigma_+^N(R^{-1})$ which uses maximizing INE processes only, performs by far the best. On the other hand, the estimate $\left(\sigma_-^N(R)\sigma_-^N(R^{-1})\right)^{-1}$ which uses minimizing INE processes only, performs very poorly. This supports experimentally the fact mentioned above that INE is powerful when maximizing and weak when minimizing. The ICE technique performs moderately (and it can not be improved by exploiting the inverse) and the original INE technique performs even worse, again, because of the weak performance when estimating the minimum singular value.

It may be interesting to see a comparison between the theoretically derived sufficient conditions for superiority of INE maximization over ICE maximization. Figures 5.3 and 5.4 display the fraction of cases in which the sufficient conditions for superiority of INE maximization (4.1), (4.8) and (3.8) are satisfied if this superiority is actually achieved. Note that the first two conditions refer to comparison of INE and ICE and the third one just relates INE maximization and minimization. Overall, in about half of the cases the conditions are satisfied and they represent a non-negligible case in the estimation process. We see also verified the fact that condition (4.1) is weaker than (4.8), as mentioned in Section 4.

5.2. Example 2. We generated 50 matrices of the form $A = U\Sigma V^T$ of size 100 with a prescribed condition number κ by choosing $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_{100})$ with

$$\sigma_k = \alpha^k, \quad 1 \le k \le 100, \qquad \text{where} \quad \alpha = \kappa^{-\frac{1}{99}}.$$

U and V are the Q factors of the QR factorizations of matrices B generated using the Matlab command B=rand(100,100) - rand(100,100). Then we computed a column pivoting with the colamd command and obtained an upper triangular factor R from the QR decomposition of the permuted A. This corresponds to the experiments in [4, Section 4, Test 2] and in [19, Section 5, Table

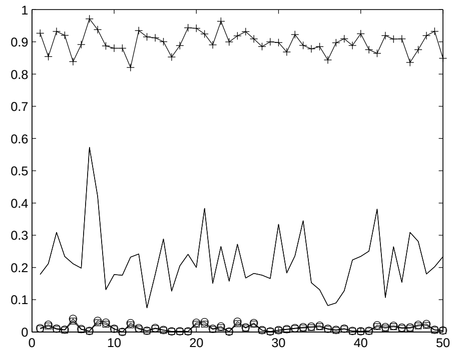

FIG. 5.1. *Ratio of estimate to real condition number for the 50 matrices in example 1. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.*

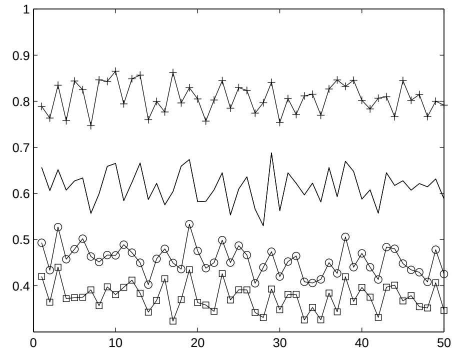

FIG. 5.2. *Ratio of estimate to real condition number for the matrices in example 2 with $\kappa(A) = 10$. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.*

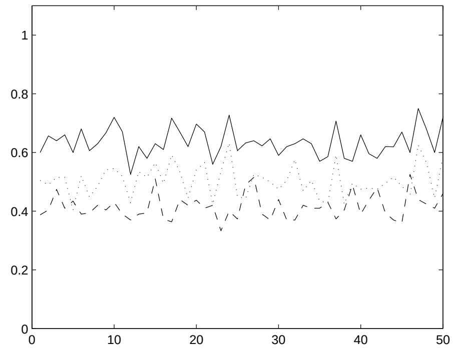

FIG. 5.3. *Ratio of the satisfied sufficient conditions in condition number estimation for the 50 matrices in example 1. Solid line: (4.8), dotted: (4.1), dashed: (3.8)*

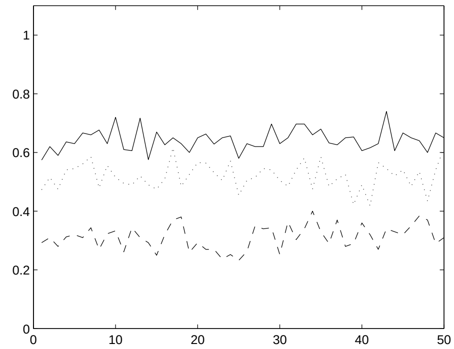

FIG. 5.4. *Ratio of the satisfied sufficient conditions in condition number estimation for the 50 matrices in example 2 with $\kappa(A) = 10$. Solid line: (4.8), dotted: (4.1), dashed: (3.8)*

5.4]. The condition estimators were tested on R, see Figures 5.2, 5.5, 5.6, for $\kappa = 10, 100, 1000$, respectively. When omitting the column pivoting we get qualitatively the same picture.

All the observations from the first example apply. Note that the two better techniques are nearly insensitive to increasing the condition number while the two other are getting worse. Also note that Figures 5.2 and 5.5 seem to suggest a general inferiority of INE using minimization only compared to original INE. This again supports the conjecture that INE is powerful when maximizing and weak when minimizing.

5.3. Example 3. We generated 50 matrices A of size 100 all with the same prescribed Euclidean norm N, by choosing the uniformly distributed singular values

$$\sigma_k = \frac{N}{k}, \quad 1 \leq k \leq 100.$$

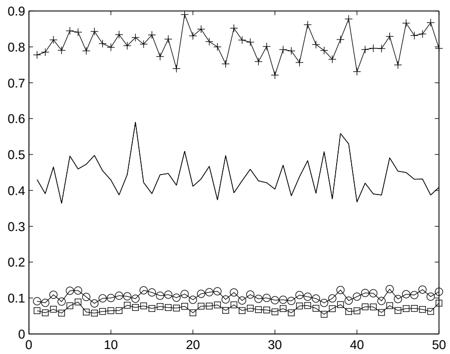

Fig. 5.5. *Ratio of estimate to real condition number for the 50 matrices in example 2 with $\kappa(A) = 100$. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.*

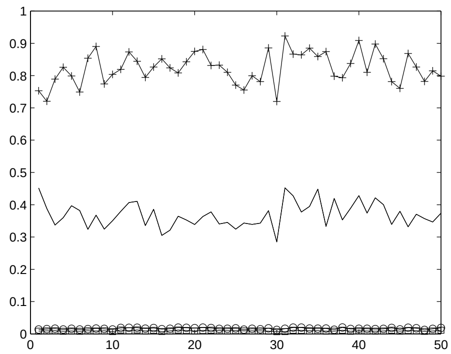

Fig. 5.6. *Ratio of estimate to real condition number for the 50 matrices in example 2 with $\kappa(A) = 1000$. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.*

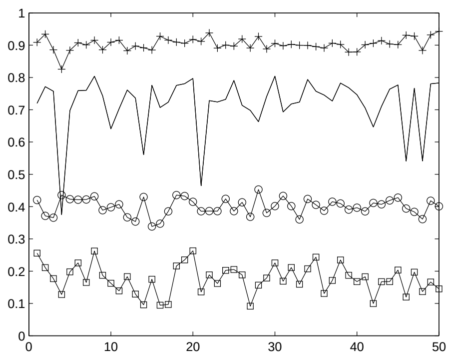

Fig. 5.7. *Ratio of estimate to real condition number for the 50 matrices in example 3 with $N = 10$. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.*

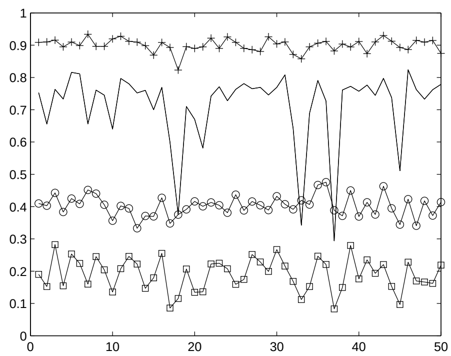

Fig. 5.8. *Ratio of estimate to real condition number for the 50 matrices in example 3 with $N = 10^{12}$. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.*

The matrix A was formed as $A = U\Sigma V^T$ where $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_{100})$ and the matrices U and V are the Q factors of the QR factorizations of matrices B generated using the Matlab command B=rand(100,100) - rand(100,100). Then we computed a column pivoting (using the Matlab command colamd(A)) and obtained an upper triangular factor R from the QR decomposition of the column permuted matrix A. This is the same type of experiments as tested in [19, Section 5, Table 5.3]. The condition estimators were tested on R, see Figures 5.7, 5.8, for, respectively, $N = 10, 10^{12}$. Qualitatively the same pictures are obtained when one omits column pivoting.

Again, INE with maximization only is the best for both cases of N. Also the other techniques keep the same relative superiority as above (exception for one matrix in Figure 5.7 and two matrices in Figure 5.8). Further, all techniques perform overall better than with exponentially distributed singular values, even when the condition number is the same like in Figure 5.5.

5.4. Example 4. We considered 20 small sparse matrices from the Matrix Market collection [18], most of them tested also in [19, Section 5, Table 5.1]. We computed their QR decomposition (with and without column pivoting) and tested the estimators with the factor R. We provide the ratios of the ICE and INE estimates versus the actual condition numbers in Figures 5.9 and 5.10, with and without column pivoting by colamd, respectively. In these figures the x-axis corresponds to the matrix number, where the numbering follows from alphabetical ordering according to matrix name. In order to see the huge differences in the quality of the estimators we also provide the values of these ratios in Table 1. We can see that the differences between the individual techniques do change more among the matrices than in the previous examples, but the basic message is the same: the INE technique with maximization is the clear winner. Column pivoting seems to have a more profound influence. In some situations all techniques do reasonably well (the matrix 'steam' without pivoting) or badly except for INE using only maximization (the matrix 'rw496' with pivoting).

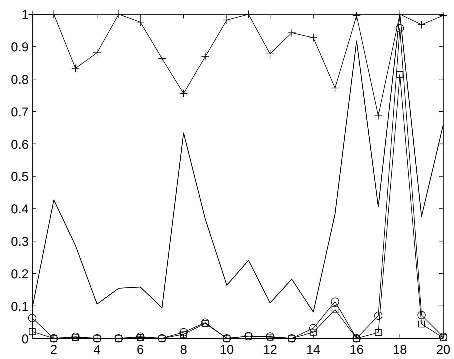

FIG. 5.9. *Ratio of estimate to actual condition number for the 20 matrices from the Matrix Market collection without column pivoting. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.*

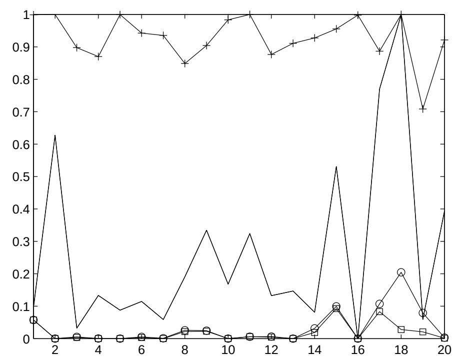

FIG. 5.10. *Ratio of estimate to actual condition number for the 20 matrices from the Matrix Market collection with column pivoting. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.*

As above, we display for the matrices from Matrix Market the fraction of cases in which the sufficient conditions for superiority of INE maximization (4.1), (4.8) and (3.8) are satisfied if this superiority is actually achieved. They are depicted on Figures 5.11 and 5.12. We can see that these conditions often seem to cover even more cases of INE maximization superiority than in the case of the random matrices from Example 1.

5.5. Example 5. The last series of experiments uses the investigated condition estimators inside a mixed direct-inverse matrix decomposition. As we mentioned in the Introduction, we believe that more accurate estimates are also useful in an incomplete decomposition since their values may decide about dropping and pivoting. Here we use the compact BIF decomposition introduced in [13, 14] (see the Matlab code there) that computes the incomplete direct and inverse factor at the same time and their mutual computation can be exploited in monitoring the decomposition. However, to facilitate comparison of the condition estimators, we will use only BIF decomposition without dropping, i.e. both the full direct and inverse factor are computed. Of course, in case of the original ICE method we could use any other implementation of the Cholesky decomposition but for simplicity we stick with the same method also here.

First, we generated 50 dense symmetric positive definite matrices A of size 100 using the Matlab

Table 1
Examples of matrices from Matrix Market: Ratios of the estimates over the actual condition numbers.

Number	Name	dim.	nnz	ICE (orig)	INE (orig)	INE (max)	INE (min)
1	494_bus	494	1666	0.09	0.06	0.99	0.02
1	(colamd)	494	1666	0.09	0.06	1	0.057
2	arc130	130	1037	0.42	4e-06	1	9e-10
2	(colamd)	130	1037	0.63	5e-06	1	5e-6
3	bfw398a	398	3678	0.29	0.005	0.83	0.004
3	(colamd)	398	3678	0.03	0.005	0.9	0.004
4	cavity04	317	5923	0.11	1e-4	0.88	3e-5
4	(colamd)	317	5923	0.13	5e-4	0.87	7e-6
5	ck400	400	2860	0.15	9e-5	0.99	8e-5
5	(colamd)	400	2860	0.09	2e-4	1	2e-5
6	dwa512	512	2480	0.16	0.005	0.97	0.003
6	(colamd)	512	2480	0.11	0.005	0.94	0.003
7	e05r0400	236	5846	0.09	5e-4	0.86	1e-4
7	(colamd)	236	5846	0.06	0.001	0.94	3e-4
8	fidap001	216	4339	0.63	0.02	0.76	0.01
8	(colamd)	216	4339	0.19	0.03	0.85	0.02
9	gre__343	343	1310	0.37	0.05	0.87	0.05
9	(colamd)	343	1310	0.33	0.025	0.9	0.023
10	impcol b	59	271	0.16	2e-4	0.98	5e-5
10	(colamd)	59	271	0.17	2e-4	0.98	5e-5
11	impcol c	137	400	0.24	0.007	0.99	0.007
11	(colamd)	137	400	0.32	0.006	0.99	0.006
12	lshp_406	406	2716	0.11	0.006	0.88	0.004
12	(colamd)	406	2716	0.13	0.006	0.88	0.005
13	lund_a	147	2449	0.18	3e-5	0.94	1e-5
13	(colamd)	147	2449	0.15	2e-4	0.91	1e-4
14	olm500	500	1996	0.08	0.03	0.93	0.019
14	(colamd)	500	1996	0.08	0.03	0.93	0.019
15	pde225	225	1065	0.38	0.11	0.77	0.088
15	(colamd)	225	1065	0.53	0.099	0.96	0.093
16	rw496	496	1859	0.92	3e-8	0.99	3e-8
16	(colamd)	496	1859	1e-5	3e-8	1	2e-8
17	saylr1	238	1128	0.4	0.07	0.69	0.02
17	(colamd)	238	1128	0.77	0.11	0.89	0.08
18	steam	240	2248	1	0.96	1	0.81
18	(colamd)	240	2248	1	0.2	1	0.03
19	str 0	363	2454	0.38	0.07	0.97	0.04
19	(colamd)	363	2454	0.06	0.08	0.71	0.02
20	west0381	381	2134	0.66	0.005	0.99	0.002
20	(colamd)	381	2134	0.4	0.003	0.92	0.002

command B=randn(100,100) and putting $A = B^T B$. The results are displayed in Figure 5.13. Next we generated 50 sparse symmetric positive definite matrices A of size 100 using the Matlab command B=sprandn(100,100,0.02)+speye(100) and putting $A = B^T B$. This gave matrices A with an average of about 850 nonzeros. The results are displayed in Figure 5.14.

As for Example 4, with sparse matrices the differences between the estimators are somehow less regular and sparse matrices seem to be favorable for original ICE. Nevertheless, the overall assessment of the quality of the individual techniques is as in the previous examples.

6. Conclusions and future work. In this paper, we have discussed incremental condition estimators in the 2-norm. In particular, the two main strategies, ICE and INE, were analyzed. It was shown that these two strategies are inherently different and the presented experiments support this claim. Moreover, we accumulated both theoretical and experimental evidence that

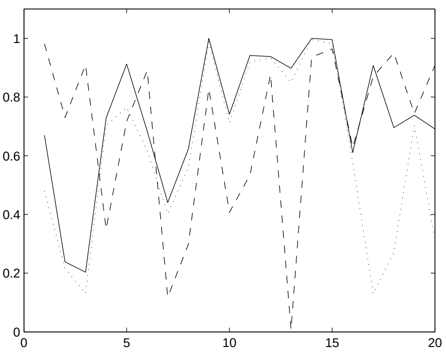

FIG. 5.11. *Ratio of the satisfied sufficient conditions in condition number estimation for the 20 matrices from the Matrix Market. Solid line: (4.8), dotted: (4.1), dashed: (3.8)*

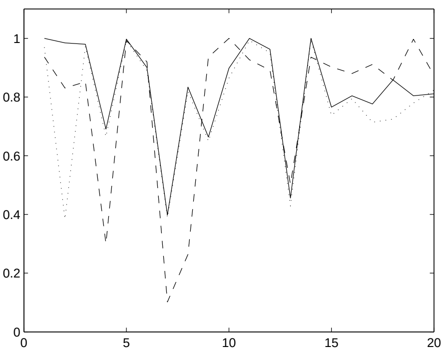

FIG. 5.12. *Ratio of the satisfied sufficient conditions in condition number estimation for the 20 matrices from the Matrix Market. Solid line: (4.8), dotted: (4.1), dashed: (3.8)*

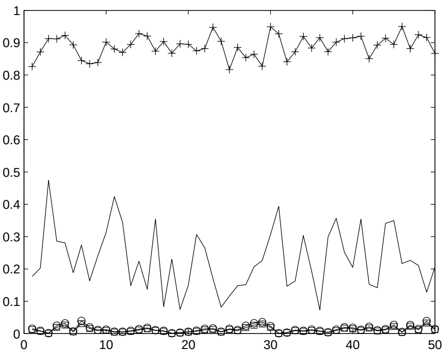

FIG. 5.13. *Ratio of estimate to actual condition number for the 50 dense symmetric positive definite matrices in example 5 decomposed with the BIF method. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.*

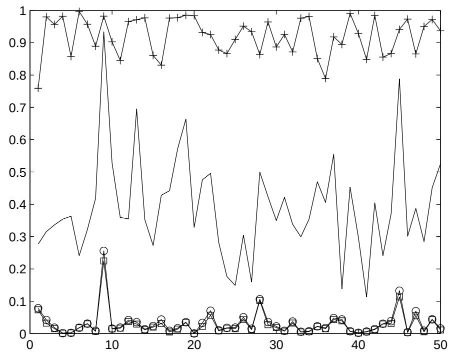

FIG. 5.14. *Ratio of estimate to actual condition number for the 50 sparse symmetric positive definite matrices in example 5 decomposed with the BIF method. Solid line: ICE (original), pluses: INE with inverse and using only maximization, circles: INE (original), squares: INE with inverse and using only minimization.*

the INE strategy using both the direct and inverse factor is a method of choice yielding a highly accurate 2-norm estimator. Our future work will consider the effects of higher accuracy of the condition estimator used inside incomplete factorizations. In particular, we intend to use accurate condition estimation for dropping and pivoting. We also intend to develop a fast block version of the described strategy taking into account several ways to extract the estimates for the diagonal blocks.

Acknowledgement. We thank the two anonymous referees for their recommendations and Gérard Meurant for careful reading of the manuscript.

REFERENCES

[1] J. Benesty and T. Gänsler. New insights into the RLS algorithm. *EURASIP Journal on Applied Signal Processing*, 3:331–339, 2004.
[2] J. Benesty and T. Gänsler. A recursive estimation of the condition number in the RLS algorithm. In *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing, vol. IV*, pages 25–28. IEEE, 2005.
[3] M. Benzi and M. Tůma. A parallel solver for large-scale Markov chains. *Applied Numerical Mathematics*, 41(1):135–153, 2002.
[4] C. H. Bischof. Incremental condition estimation. *SIAM J. on Matrix Analysis and Applications*, 11:312–322, 1990.
[5] C. H. Bischof. A parallel QR factorization algorithm with controlled local pivoting. *SIAM J. on Scientific and Statistical Computing*, 12(1):36–57, 1991.
[6] C. H. Bischof, J. G. Lewis, and D. J. Pierce. Incremental condition estimation for sparse matrices. *SIAM J. on Matrix Analysis and Applications*, 11(4):644–659, 1990.
[7] C. H. Bischof and P. T. P. Tang. Generalizing incremental condition estimation. *J. Numer. Linear Algebra Appl.*, 1(2):149–163, 1992.
[8] M. Bollhöfer. A robust ILU with pivoting based on monitoring the growth of the inverse factors. *Linear Algebra and its Applications*, 338:201–218, 2001.
[9] M. Bollhöfer. A robust and efficient ILU that incorporates the growth of the inverse triangular factors. *SIAM J. on Scientific Computing*, 25(1):86–103, 2003.
[10] M. Bollhöfer. ILUPACK: version 2.4, 2011. `http://www.icm.tu-bs.de/ bolle/ilupack/`.
[11] M. Bollhöfer and Y. Saad. On the relations between ILUs and factored approximate inverses. *SIAM J. on Matrix Analysis and Applications*, 24(1):219–237, 2002.
[12] M. Bollhöfer and Y. Saad. Multilevel preconditioners constructed from inverse-based ILUs. *SIAM J. on Scientific Computing*, 27(5):1627–1650, 2006.
[13] R. Bru, J. Marín, J. Mas, and M. Tůma. Balanced incomplete factorization. *SIAM J. on Scientific Computing*, 30(5):2302–2318, 2008.
[14] R. Bru, J. Marín, J. Mas, and M. Tůma. Improved balanced incomplete factorization. *SIAM J. on Matrix Analysis and Applications*, 31(5):2431–2452, 2010.
[15] A. Cline, A. Conn, and C. Van Loan. Generalizing the LINPACK condition estimator. In J. Hennart, editor, *Numerical Analysis*, volume 909 of *Lecture Notes in Mathematics*, pages 73–83. Springer, Berlin - Heidelberg, 1982.
[16] G. Cybenko. Fast Toeplitz orthogonalization using inner products. *SIAM J. on Scientific and Statistical Computing*, 8(5):734–740, 1987.
[17] J. W. Demmel. *Applied numerical linear algebra*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.
[18] I. S. Duff, R. G. Grimes, and J. G. Lewis. Sparse matrix test problems. *ACM Transactions on Mathematical Software*, 15:1–14, 1989.
[19] I. S. Duff and C. Vömel. Incremental norm estimation for dense and sparse matrices. *BIT Numerical Mathematics*, 42:300–322, 2002.
[20] C. Fassino. On updating the least singular value: a lower bound. *Calcolo*, 40(4):213–229, 2003.
[21] R. Ferng. *Lanczos-based condition estimation in signal processing and optimization*. PhD thesis, Department of Mathematics, North Carolina State University, 1991.
[22] W.R. Ferng, G.H. Golub, and R.J. Plemmons. Adaptive Lanczos methods for recursive condition estimation. *Numerical Algorithms*, 1(1):1–19, 1991.
[23] W. W. Hager. Condition estimates. *SIAM J. on Scientific and Statistical Computing*, 5(2):311–316, 1984.
[24] S. Haykin. *Adaptive Filter Theory. 4th edition.* Prentice-Hall, Englewood Cliffs, NJ., 2002.
[25] N. J. Higham. A survey of condition number estimation for triangular matrices. *SIAM Review*, 29:575–596, 1987.
[26] N. J. Higham. FORTRAN codes for estimating the one-norm of a real or complex matrix, with applications to condition estimation. *ACM Transactions on Mathematical Software*, 14(4):381–396 (1989), 1988.
[27] N. J. Higham. Experience with a matrix norm estimator. *SIAM J. on Scientific and Statistical Computing*, 11(4):804–809, 1990.
[28] N. J. Higham and F. Tisseur. A block algorithm for matrix 1-norm estimation, with an application to 1-norm pseudospectra. *SIAM J. on Matrix Analysis and Applications*, 21(4):1185–1201, 2000.
[29] J. Mandel and B. Sousedík. Adaptive selection of face coarse degrees of freedom in the bddc and the feti-dp iterative substructuring methods. *Comput. Methods Appl. Mech. Engrg.*, 196:1389–1399, 2007.
[30] N. Mastronardi, M. Van Barel, and R. Vandebril. A Schur-based algorithm for computing bounds to the smallest eigenvalue of a symmetric positive definite Toeplitz matrix. *Linear Algebra and its Applications*, 428(2-3):479–491, 2008.
[31] J. Neering. *Optimization and Estimation Techniques for Passive Acoustic Source Localization*. PhD thesis, l'École nationale superieure des mines de Paris, 2009.
[32] C.-T. Pan and R.J. Plemmons. Least squares modifications with inverse factorizations: parallel implications.

J. of Computational and Applied Mathematics, 27(1-2):109–127, 1989.

[33] D. J. Pierce and J. G. Lewis. Sparse multifrontal rank revealing QR factorization. *SIAM J. on Matrix Analysis and Applications*, 18(1):159–180, 1997.

[34] D. J. Pierce and R. J. Plemmons. Fast adaptive condition estimation. *SIAM J. on Matrix Analysis and Applications*, 13(1):274–291, 1992.

[35] D. J. Pierce and R. J. Plemmons. Tracking the condition number for RLS in signal processing. *Math. Control Signals Systems*, 5(1):23–39, 1992.

[36] S. Sastry and M. Bodson. *Adaptive Control: Stability, Convergence, and Robustness*. Prentice-Hall, Englewood Cliffs, NJ, 1989.

[37] G.M. Shroff and C.H. Bischof. Adaptive condition estimation for rank-one updates of QR factorizations. *SIAM J. on Matrix Analysis and Applications*, 13(4):1264–1278, 1992.

[38] G. W. Stewart. *Matrix algorithms. Vol. I. Basic decompositions*. SIAM, Philadelphia, PA, 1998.

[39] A. C. N. van Duin. Scalable parallel preconditioning with the sparse approximate inverse of triangular matrices. *SIAM J. on Matrix Analysis and Applications*, 20(4):987–1006, 1999.