# ON THE ACCURACY OF THE KARLSON-WALDÉN'S ESTIMATE OF THE BACKWARD ERROR FOR LINEAR LEAST SQUARES PROBLEMS*

SERGE GRATTON†, PAVEL JIRÁNEK‡, AND DAVID TITLEY-PELOQUIN§

CERFACS TECHNICAL REPORT TR/PA/11/19

**Abstract.** We consider the backward error associated with a given approximate solution of a linear least squares problem. The backward error can be very expensive to compute, as it involves the minimal singular value of certain matrix that depends on the problem data and the approximate solution. An estimate based on a regularized projection of the residual vector has been proposed in the literature and analyzed by several authors. Although numerical experiments in the literature suggest that it is a reliable estimate of the backward error for any given approximate LS solution, to date no satisfactory explanation for this behavior had been found. We derive new bounds which confirm this experimental observation.

**Key words.** linear least squares problems, backward error

**AMS subject classifications.** 15A06, 65F99

## 1. Introduction.

We consider a linear least squares (LS) problem

$$\text{find } \hat{x} \in \mathbb{R}^n \text{ such that } \quad \|b - A\hat{x}\|_2 = \min_{x \in \mathbb{R}^n} \|b - Ax\|_2, \tag{1.1}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $\|v\|_2 = \sqrt{v^T v}$ denotes the Euclidean norm. It is well known that the vector $\hat{x}$ is a solution of (1.1) if and only if it solves the system of normal equations: $A^T A\hat{x} = A^T b$. Any such $\hat{x}$ can be expressed as $\hat{x} = A^\dagger b + (I - A^\dagger A)z$ for some $z \in \mathbb{R}^n$, where $A^\dagger$ denotes the Moore-Penrose pseudo-inverse of $A$. The unique residual vector corresponding to any LS solution is $\hat{r} \equiv b - A\hat{x} = (I - P_A)b$, where $P_A \equiv AA^\dagger$ is the orthogonal projector onto the range of $A$, hereafter denoted $\mathcal{R}(A)$. For a more complete background, see, e.g., [2, 7, 15].

Let $x \in \mathbb{R}^n$ be a nonzero approximate solution of the LS problem (1.1). We are interested in computing or estimating the backward error associated with the approximate solution $x$. More precisely, we would like to find the "smallest" perturbations $E$ and $f$ for which $x$ is a solution of the perturbed LS problem with matrix $A + E$ and right-hand side vector $b + f$. Such an analysis has many practical applications. For instance, it is used for testing fast, but potentially unstable, algorithms; see, e.g., [10]. It is used to monitor the convergence of iterative solution methods and to design reliable stopping criteria for these methods; see, e.g., [17, 1, 4, 13, 6]. In this context the approximate solution $x$ is an iterate from any chosen iterative method, and we stop the iteration and accept $x$ as a valid computed solution when the backward error (or an estimate of the backward error) is smaller than a chosen (relative) tolerance.

*March 16, 2011.

†INPT-IRIT, University of Toulouse and ENSEEIHT, 2 Rue Camichel, BP 7122, 31071 Toulouse Cedex 7, France (serge.gratton@enseeiht.fr).

‡CERFACS, 42 Avenue Gaspard Coriolis, 31057 Toulouse, France (jiranek@cerfacs.fr). The work of this author was supported by the ADTAO project funded by the foundation STAE, Toulouse, France, within RTRA.

§Mathematical Institute, University of Oxford, 24-29 St-Giles', Oxford, OX1 3LB, UK (David.Titley-Peloquin@maths.ox.ac.uk).

In [19] Waldén, Karlson, and Sun provide an explicit expression for the LS backward error $\mu$ defined by

$$\mu \equiv \min_{E,f}\{\|[E,\theta f]\|_F;\ (A+E)^T(A+E)x = (A+E)^T b\}, \tag{1.2}$$

where $\theta$ is a given positive weighting parameter and $\|\cdot\|_F$ is the Frobenius matrix norm. The parameter $\theta$ balances the norms of the backward perturbations in $A$ and $b$, and can be chosen based on a priori knowledge of the uncertainty in the problem data; see, e.g., [4, Theorem 6.3]. Consider

$$\omega \equiv \min_{E,f}\{\|[E,\theta f]\|_F;\ (A+E)x = b+f\} = \frac{\theta\|r\|_2}{\sqrt{1+\theta^2\|x\|_2^2}}, \tag{1.3}$$

the backward error of $x$ for compatible linear equations (see, e.g., [3, Theorem 2.2], [11, Problem 7.8]). Clearly $\mu \leq \omega$, as the constraint in (1.3) is more restrictive than that in (1.2). Waldén, Karlson, and Sun [19, Corollary 2.1] showed that $\mu = \min\{\omega, \sigma\}$, where $\sigma$ is the smallest singular value of a certain $(n+m)\times m$ matrix that depends on the problem data $A$ and $b$ and the approximate LS solution $x$. Thus, the LS backward error $\mu$ is strictly less than the backward error for compatible linear equations when $\sigma < \omega$. In Theorem 2.2 we give a necessary and sufficient condition for $\mu = \sigma < \omega$.

Because the expression for $\mu$ involves $\sigma$, the minimal singular value of an $(n+m)\times m$ matrix, computing $\mu$ directly can be prohibitively expensive. In Section 3 we recall an estimate $\nu$ of $\mu$ proposed by Karlson and Waldén in [14]. This estimate involves a regularized projection of the residual vector $r = b - Ax$. Methods for computing $\nu$ were considered by Grcar, Saunders, and Su [9] (see also the thesis of Su [18]) and its efficient computation in the iterative method LSQR [17, 16] was proposed in [13].

Properties of the estimate $\nu$ have been studied before by Gu [10] and Grcar [8]. In particular, it is known that $\nu$ is a good approximation to the LS backward error $\mu$ provided the approximate solution $x$ is sufficiently close to a LS solution $\hat{x}$. Numerical experience, however, indicates that $\nu$ is a good approximation to the LS backward error $\mu$ for *any* given $x$; see [9, 18, 13]. In Section 3 we derive new bounds on $\mu$ in terms of the estimate $\nu$ of Karlson and Waldén, and prove that $\nu$ is always within a constant factor of the backward error $\mu$. These results confirm that $\nu$ can always be used as a reliable estimate of $\mu$. In Section 4 we illustrate our theoretical results with numerical experiments.

We often use the singular value decomposition (SVD) of $A$ (see, e.g., [7, Theorem 2.5.2])

$$A = U\begin{bmatrix}\Sigma & 0\\ 0 & 0\end{bmatrix}V^T, \tag{1.4}$$

where $\Sigma = \operatorname{diag}(\sigma_1,\ldots,\sigma_p)$ with $\sigma_1 \geq \cdots \geq \sigma_p > 0$ is a $p\times p$ diagonal matrix ($p = \operatorname{rank}(A)$) containing the nonzero singular values of $A$ on its diagonal, and $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices with $U = [U_1, U_2]$, $U_1 \in \mathbb{R}^{m\times p}$, such that $\mathcal{R}(U_1) = \mathcal{R}(A)$. We use $\sigma_{\min}(M)$ to denote the smallest (possibly zero) singular value of a matrix $M$. In the following $x = \hat{x}$ or $x \neq \hat{x}$ means that $x$ is or is not, respectively, *a* solution of the LS problem (1.1). Throughout the paper we restrict ourselves to real arithmetic; generalization to the complex case is straightforward.

**2. The backward error in LS problems.** We first recall the expression for the LS backward error $\mu$ given in (1.2), which was obtained by Karlson, Waldén, and Sun in [19, Corollary 2.1] (see also [11, Theorem 20.5]).

THEOREM 2.1. *Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $0 \neq x \in \mathbb{R}^n$, $r = b - Ax$, and $\theta > 0$. The backward error $\mu$ defined by* (1.2) *can be expressed as*

$$\mu = \min\{\omega, \sigma_{\min}(M)\}, \tag{2.1}$$

*where $\omega$ is given by* (1.3) *and*

$$M \equiv \begin{bmatrix} A^T \\ \omega(I - rr^\dagger) \end{bmatrix}. \tag{2.2}$$

It can easily be shown that $\mu = 0$ if and only if $x = \hat{x}$; see, e.g., [5, Corollary 3.1]. Thus as long as $x \neq \hat{x}$, the matrix $M$ in (2.2) has full column rank and $\sigma_{\min}(M) > 0$. We will make use this property later.

In [19, Section 2] and [11, Section 20.7] it is pointed out that a sufficient condition for $\mu = \sigma_{\min}(M) < \omega$ is $b \notin \mathcal{R}(A)$. In other words, the LS backward error $\mu$ in (1.2) is always strictly less than the backward error for compatible linear equations $\omega$ in (1.3) if the LS problem (1.1) is incompatible. The converse, however, is not always true. In the following theorem we give a necessary and sufficient condition for $\mu < \omega$.

THEOREM 2.2. *Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$ nonzero, $r = b - Ax$, and $\theta > 0$. Let $\mu$ be the LS backward error defined by* (1.2) *and $\omega$ be the LE backward error defined by* (1.3).

1. *If $b \notin \mathcal{R}(A)$, then $\mu < \omega$.*
2. *If $b \in \mathcal{R}(A)$, then*

$$\mu < \omega \qquad \Leftrightarrow \qquad \frac{\theta\|A^\dagger r\|_2}{\sqrt{1+\theta^2\|x\|_2^2}} > 1. \tag{2.3}$$

*Proof* The case $b \notin \mathcal{R}(A)$ is stated in [19, Section 2] and can easily be proven as follows. If $b \notin \mathcal{R}(A)$, then $\|r\|_2 \geq \|\hat{r}\|_2 > 0$ and $r^T\hat{r} = r^T(I - P_A)r = \|\hat{r}\|_2^2 \neq 0$. Therefore, from (2.2),

$$\sigma_{\min}(M) = \min_{\substack{z\in\mathbb{R}^m \\ z\neq 0}} \frac{\|Mz\|_2}{\|z\|_2} \leq \frac{\|M\hat{r}\|_2}{\|\hat{r}\|_2} = \omega\frac{\|(I - rr^\dagger)\hat{r}\|_2}{\|\hat{r}\|_2} < \omega. \tag{2.4}$$

Thus, if $b \notin \mathcal{R}(A)$, from (2.1) we must have $\mu = \sigma_{\min(M)} < \omega$.

Now suppose that $b \in \mathcal{R}(A)$, so that $r \in \mathcal{R}(A)$. If $x = \hat{x}$ then $r = \hat{r} = 0$, so $\mu = \omega = 0$, $A^\dagger r = 0$, and the theorem holds. We now assume that $x \neq \hat{x}$, so that $r \neq 0$. The minimal singular value of $M$ in (2.2) is the square root of the minimal eigenvalue of the matrix

$$M^TM = AA^T + \omega^2 I - \frac{\omega^2}{\|r\|_2^2} rr^T.$$

Consider the SVD of $A$ given in (1.4). Because $r \in \mathcal{R}(A)$, we can write $r = U_1\bar{r}$. The matrix $M^TM$ has the same eigenvalues as the orthogonally similar matrix

$$U^TM^TMU = \begin{bmatrix} \Sigma^2 + \omega^2 I_p - \frac{\omega^2}{\|r\|_2^2}\bar{r}\bar{r}^T & 0 \\ 0 & \omega^2 I_{m-p} \end{bmatrix}. \tag{2.5}$$

Therefore, $\sigma_{\min}(M) = \min\{\omega, \sqrt{\lambda}\}$, where $\lambda$ is the minimal eigenvalue of

$$\Sigma^2 + \omega^2 I - \frac{\omega^2}{\|r\|_2^2} \bar{r}\bar{r}^T. \tag{2.6}$$

It follows that $\mu = \sigma_{\min}(M) < \omega$ if and only if

$$\lambda < \omega^2 \quad \Leftrightarrow \quad \lambda_{\min}\left(\Sigma^2 - \frac{\omega^2}{\|r\|_2^2}\bar{r}\bar{r}^T\right) < 0 \quad \Leftrightarrow \quad \lambda_{\min}\left(I - \frac{\omega^2}{\|r\|_2^2}\Sigma^{-1}\bar{r}\bar{r}^T\Sigma^{-1}\right) < 0,$$

where the last equivalence is due to Sylvester's law of inertia (see, e.g., [12, Theorem 4.5.8]). Thus if $b \in \mathcal{R}(A)$ and $x \neq \hat{x}$, we have $\mu = \sigma_{\min}(M) < \omega$ if and only if

$$1 - \frac{\omega^2}{\|r\|_2^2}\|\Sigma^{-1}\bar{r}\|_2^2 < 0,$$

which can be rearranged to (2.3) using (1.3) and the fact that

$$\|A^\dagger r\|_2 = \left\| V \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^T r \right\|_2 = \|\Sigma^{-1}\bar{r}\|_2. \qquad \square$$

It follows from Theorem 2.2 that $b \notin \mathcal{R}(A)$ is not only sufficient for $\mu < \omega$ but also necessary provided

$$\frac{\theta\|A^\dagger r\|_2}{\sqrt{1+\theta^2\|x\|_2^2}} > 1.$$

On the other hand, if

$$\frac{\theta\|A^\dagger r\|_2}{\sqrt{1+\theta^2\|x\|_2^2}} \leq 1$$

and $b \in \mathcal{R}(A)$, then $\mu = \omega$. It is often the case in practice that the matrix $A$ has full column rank. Then the above is equivalent to

$$\frac{\theta\|\hat{x} - x\|_2}{\sqrt{1+\theta^2\|x\|_2^2}} \leq 1. \tag{2.7}$$

Since

$$\frac{\theta\|\hat{x} - x\|_2}{\sqrt{1+\theta^2\|x\|_2^2}} \leq \frac{\|\hat{x} - x\|_2}{\|x\|_2},$$

the condition (2.7) can be replaced by a stronger requirement $\|\hat{x} - x\|_2/\|x\|_2 \leq 1$ on the standard relative error with respect to the norm of the approximate solution $x$. In other words, if $b \in \mathcal{R}(A)$, the case $\mu < \omega$ can only occur if $\|\hat{x} - x\|_2/\|x\|_2 > 1$.

Theorem 2.2 gives a necessary and sufficient condition for $\mu$ to be *strictly* less than $\omega$. We remark that if $b \in \mathcal{R}(A)$ and $\operatorname{rank}(A) < m$ (i.e., the block $\omega^2 I_{m-p}$ in (2.5) is a matrix of nonzero order) then $\sigma_{\min}(M) \leq \omega$ and hence from (1.2) $\mu = \sigma_{\min}(M)$. In other words, as long as $\operatorname{rank}(A) < m$, we can always write $\mu = \sigma_{\min}(M)$. The condition $\operatorname{rank}(A) < m$ usually holds in practice, as most practical LS problems (1.1) are overdetermined.

**3. The Karlson-Waldén estimate and its accuracy.** The expression for the LS backward error $\mu$ given in Theorem 2.1 involves the minimal singular value of the $(n+m)\times m$ matrix $M$. Therefore, $\mu$ can be very expensive to compute directly. In this section we give tight bounds on $\mu$ which involve the estimate

$$\nu \equiv \frac{\omega}{\|r\|_2}\|(A^TA+\omega^2 I)^{-1/2}A^Tr\|_2 = \frac{\omega}{\|r\|_2}\left\|\begin{bmatrix}A\\ \omega I\end{bmatrix}\begin{bmatrix}A\\ \omega I\end{bmatrix}^{\dagger}\begin{bmatrix}r\\ 0\end{bmatrix}\right\|_2 \tag{3.1}$$

introduced by Karlson and Waldén in [14]. The quantity $\nu$ can be computed directly more cheaply than $\mu$ (see, e.g., [9, 18]) and can be estimated very efficiently in the iterative method LSQR (see [13]).

Note that $\nu = 0$ if and only if $x = \hat{x}$. (In the trivial case $r = 0$ we explicitly set $\nu = 0$.) In the following we assume that $x \neq \hat{x}$, so $r \neq 0$ and $\nu > 0$. Note also that

$$\nu^2 = \frac{\omega^2}{\|r\|_2^2} r^TA(A^TA+\omega^2 I)^{-1}A^Tr < \omega^2, \tag{3.2}$$

which can easily be verified using the singular value decomposition of $A$ given in (1.4).

Karlson and Waldén [14] give the following lower bound on the ratio $\mu/\nu$:

$$2-\sqrt{2} \leq \frac{\mu}{\nu}. \tag{3.3}$$

In [10] Gu obtains the following bounds:

$$\frac{2}{1+\sqrt{5}} \leq \frac{\mu}{\nu} \leq \frac{\|r\|_2}{\|\hat{r}\|_2} \tag{3.4}$$

(see also [9, Equation (1.5)]). This shows that if $\|r\|_2 \approx \|\hat{r}\|_2$, as we would expect if $x$ is a good approximation to $\hat{x}$, then $\nu$ is a good estimate of $\mu$. In [8] Grcar shows that $\nu$ is asymptotically equal to $\mu$ in the sense that

$$\lim_{x\to\hat{x}} \frac{\mu}{\nu} = 1. \tag{3.5}$$

The above results show that $\nu$ can be an excellent estimate of $\mu$, provided that $x$ is sufficiently close to $\hat{x}$. However, if $x$ is not close to $\hat{x}$ and $\|r\|_2 \gg \|\hat{r}\|_2$, the estimate $\nu$ might severely underestimate $\mu$. This could have important consequences in practice. For example, in the context of iterative methods, a stopping criterion that uses $\nu$ (or an estimate of $\nu$) instead of $\mu$ could be triggered several iterations too early if $\nu \ll \mu$, resulting in a poor computed solution $x$.

Numerical tests performed in [9, 18, 13] suggest that $\nu$ is a good approximation to the LS backward error $\mu$ for *any* given $x$. In the following, we confirm this observation by establishing bounds on $\mu$ in the form

$$\nu \leq \mu \leq \gamma\nu. \tag{3.6}$$

Indeed, $\nu$ is a lower bound on $\mu$ (see Theorem 3.1 below), which slightly improves the lower bounds in (3.3) and (3.4). Using this fact and the upper bound in (3.4) we have

$$\nu \leq \mu \leq \gamma_1\nu, \qquad \gamma_1 \equiv \frac{\|r\|_2}{\|\hat{r}\|_2}. \tag{3.7}$$

In the following theorem we show that the above upper bound can be considerably improved.

THEOREM 3.1. *Let the assumptions of Theorem 2.1 be satisfied and let $\mu$ and $\nu$ be defined by* (1.2) *and* (3.1)*, respectively. Then*

$$\nu \leq \mu \leq \sqrt{2}\,\nu. \tag{3.8}$$

*Proof.* Recall that the backward error $\mu$ and the estimate $\nu$ are both equal to zero if and only if $x = \hat{x}$. Therefore, (3.8) holds when $x = \hat{x}$. We now assume that $x \neq \hat{x}$ (so that $r$, $\omega$, $\mu$, and $\nu$ are nonzero) and denote $\sigma \equiv \sigma_{\min}(M)$, so $\mu = \min\{\omega, \sigma\}$.

First we show that $\nu \leq \mu$. Suppose that $\omega \leq \sigma$, so $\mu = \min\{\omega, \sigma\} = \omega$. In this case it immediately follows from (3.2) that $\nu < \mu$. Now suppose that $\mu = \sigma < \omega$. The minimal singular value $\sigma$ is the square root of the minimal eigenvalue of the matrix

$$M^T M = AA^T + \omega^2 I - \omega^2 \tilde{r}\tilde{r}^T, \tag{3.9}$$

where $\tilde{r} = r/\|r\|_2$ is the normalized residual vector associated with $x$. We denote by $u_*$ the (normalized) eigenvector corresponding to the minimal eigenvalue $\sigma^2$ of $M^T M$ (equivalently, the right singular vector corresponding to the minimal singular value $\sigma$ of $M$). Then

$$(AA^T + \omega^2 I - \omega^2 \tilde{r}\tilde{r}^T)u_* = \sigma^2 u_* \tag{3.10}$$

and

$$\left[AA^T + (\omega^2 - \sigma^2)I\right] u_* = \omega^2(\tilde{r}^T u_*)\tilde{r}. \tag{3.11}$$

The vector $\tilde{r}$ is not orthogonal to $u_*$. Indeed, if $\tilde{r}^T u_* = 0$ then (3.10) gives $0 \leq u_* AA^T u_* = \sigma^2 - \omega^2$, which is in contradiction with the fact that $\sigma < \omega$. Note that the matrix $AA^T + (\omega^2 - \sigma^2)I$ is nonsingular. Multiplying (3.11) with $\tilde{r}^T \left[AA^T + (\omega^2 - \sigma^2)I\right]^{-1}$ and dividing by nonzero $\tilde{r}^T u_*$ leads to

$$1 = \omega^2 \tilde{r}^T \left[AA^T + (\omega^2 - \sigma^2)I\right]^{-1} \tilde{r}. \tag{3.12}$$

Using the Sherman-Morrison-Woodbury formula (see, e.g., [7, Section 2.1]) we get

$$\left[AA^T + (\omega^2 - \sigma^2)I\right]^{-1} = \frac{1}{\omega^2 - \sigma^2}\left\{I - A\left[A^T A + (\omega^2 - \sigma^2)I\right]^{-1} A^T\right\},$$

which by substituting to (3.12) gives

$$\sigma^2 = \omega^2 \tilde{r}^T A \left[A^T A + (\omega^2 - \sigma^2)I\right]^{-1} A^T \tilde{r}. \tag{3.13}$$

Since $\mu = \sigma < \omega$, we obtain

$$\mu^2 = \sigma^2 = \omega^2 \tilde{r}^T A \left[A^T A + (\omega^2 - \sigma^2)I\right]^{-1} A^T \tilde{r} \geq \omega^2 \tilde{r}^T A \left[A^T A + \omega^2 I\right]^{-1} A^T \tilde{r} = \nu^2 \tag{3.14}$$

and therefore $\nu \leq \mu$ in all cases, which proves the first part of (3.8).

Next we prove that $\mu \leq \sqrt{2}\nu$. Recall that when $x \neq \hat{x}$, $M$ has full column rank, so $M^T M$ is nonsingular. From the variational characterization of eigenvalues (see, e.g., [12, Theorem 4.2.2]) we have

$$\sigma^2 = \min_{\substack{y \in \mathbb{R}^m \\ \|y\|_2 = 1}} y^T M^T M y = \min_{\substack{z \in \mathbb{R}^m \\ \|z\|_2 = 1}} \frac{1}{z^T (M^T M)^{-1} z} \leq \frac{1}{\tilde{r}^T (M^T M)^{-1} \tilde{r}}. \tag{3.15}$$

Using the Sherman-Morrison-Woodbury formula, we find that

$$(AA^T + \omega^2 I)^{-1} = \frac{1}{\omega^2}\left[I - A(A^TA + \omega^2 I)^{-1}A^T\right] \tag{3.16}$$

and

$$(M^TM)^{-1} = \frac{1}{\omega^2}\left[\omega^2(AA^T + \omega^2 I)^{-1} + \frac{\omega^4(AA^T + \omega^2 I)^{-1}\tilde{r}\tilde{r}^T(AA^T + \omega^2 I)^{-1}}{1 - \omega^2\tilde{r}^T(AA^T + \omega^2 I)^{-1}\tilde{r}}\right]. \tag{3.17}$$

From (3.1) and (3.16) we find an alternative formula for $\nu$ in the form

$$\nu^2 = \omega^2\tilde{r}^TA(A^TA + \omega^2 I)^{-1}A^T\tilde{r} = \omega^2\left[1 - \omega^2\tilde{r}^T(AA^T + \omega^2 I)^{-1}\tilde{r}\right],$$

which gives

$$\omega^2\tilde{r}^T(AA^T + \omega^2 I)^{-1}\tilde{r} = 1 - \frac{\nu^2}{\omega^2}. \tag{3.18}$$

From (3.17) and (3.18) we get

$$\begin{aligned}\tilde{r}^T(M^TM)^{-1}\tilde{r} &= \frac{1}{\omega^2}\tilde{r}^T\left[\omega^2(AA^T + \omega^2 I)^{-1} + \frac{\omega^4(AA^T + \omega^2 I)^{-1}\tilde{r}\tilde{r}^T(AA^T + \omega^2 I)^{-1}}{1 - \omega^2\tilde{r}^T(AA^T + \omega^2 I)^{-1}\tilde{r}}\right]\tilde{r} \\ &= \frac{1}{\omega^2}\left[1 - \frac{\nu^2}{\omega^2} + \left(1 - \frac{\nu^2}{\omega^2}\right)^2\frac{\omega^2}{\nu^2}\right] = \frac{1}{\nu^2}\left(1 - \frac{\nu^2}{\omega^2}\right).\end{aligned} \tag{3.19}$$

Substituting (3.19) into (3.15) we obtain the inequality

$$\sigma^2 \le \nu^2\left(1 - \frac{\nu^2}{\omega^2}\right)^{-1}. \tag{3.20}$$

Because $\mu = \min\{\omega, \sigma\} \le \sigma$,

$$\mu^2\left(1 - \frac{\nu^2}{\omega^2}\right) \le \sigma^2\left(1 - \frac{\nu^2}{\omega^2}\right) \le \nu^2,$$

which, using the fact that $\mu = \min\{\omega, \sigma\} \le \omega$, gives

$$\mu^2 \le \nu^2\left(1 + \frac{\mu^2}{\omega^2}\right) \le 2\nu^2, \tag{3.21}$$

proving the second part of (3.8). ∎

We can also improve the result of Theorem 3.1 by obtaining a tighter upper bound of the form (3.6), with $\gamma$ depending on the ratio of the norms of the residual vectors $r$ and $\hat{r}$ corresponding to $x$ and $\hat{x}$, respectively.

COROLLARY 3.2. *Let the assumptions of Theorem 2.1 be satisfied, let $\mu$ and $\nu$ be defined by* (1.2) *and* (3.1)*, respectively, and let $r \neq 0$. Then*

$$\nu \le \mu \le \gamma_2\nu, \qquad \gamma_2 \equiv \sqrt{2 - \left(\frac{\|\hat{r}\|_2}{\|r\|_2}\right)^2} \le \sqrt{2}. \tag{3.22}$$

*Proof.* Suppose that $b \notin \mathcal{R}(A)$, i.e., $\hat{r} \neq 0$. In (3.21) we have

$$\frac{\mu^2}{\omega^2} \leq \frac{\sigma^2}{\omega^2} \leq \frac{1}{\omega^2} \frac{\|M\hat{r}\|_2^2}{\|\hat{r}\|_2^2} = \frac{\|(I - rr^\dagger)\hat{r}\|_2^2}{\|\hat{r}\|_2^2} = 1 - \frac{\|\hat{r}\|_2^2}{\|r\|_2^2}$$

using $A^T\hat{r} = 0$, $\hat{r} = (I - P_A)b = (I - P_A)r$, and $\|\hat{r}\|_2^2 = \|r\|_2^2 - \|P_A r\|_2^2$. The bound (3.22) clearly remains true if $b \in \mathcal{R}(A)$, i.e., if $\hat{r} = 0$, as it then reduces to (3.8). ∎

Analogously to (3.7), the bound (3.22) shows that $\nu \approx \mu$ if $\|r\|_2 \approx \|\hat{r}\|_2$. It also confirms (3.5), namely that $\mu/\nu \to 1$ as $x \to \hat{x}$, since $\|r\|_2 \to \|\hat{r}\|_2$ and $\gamma_2 \to 1$ as $x \to \hat{x}$. However, in contrast to previous results, our upper bound in (3.22) remains tight to within a factor $\sqrt{2}$ even when $x$ is not a good approximation to the LS solution $\hat{x}$ and $\|r\|_2 \gg \|\hat{r}\|_2$. It is therefore a significant improvement to (3.3), (3.4), and (3.5).

**4. Numerical experiments.** In this section we illustrate our theoretical results on several numerical experiments. We create three different types of LS problems, which we denote Type (a)–(c). In each case we create the matrix $A$ as follows: $A = U\Sigma V^T$, where $U \in \mathbb{R}^{m\times n}$ and $V \in \mathbb{R}^{n\times n}$ are the "thin" $Q$ factors in the QR decomposition of random matrices and $\Sigma \in \mathbb{R}^{n\times n}$ is diagonal and contains the singular values of $A$. We experiment with different distributions of singular values:

- Type (a): logarithmically equally-spaced between $10^{-2}$ and 1;
- Type (b): logarithmically equally-spaced between $10^{-5}$ and $10^3$;
- Type (c): linearly equally-spaced between $10^{-5}$ and $10^3$.

In these examples we use $m = 100$ and $n = 50$. For each problem type we set $b = Ae + \tau v$, where $e$ is a vector of ones and each element of $v$ is independently sampled from the standard normal distribution. We use $\tau = 0$, $\tau = 10^{-5}$, and $\tau = 10^{-1}$ to create compatible, nearly compatible, and fairly incompatible LS problems, respectively.

For each problem type and each vector $b$ we obtain several approximate LS solutions $x$ as follows. First we compute a backward stable LS solution $\hat{x}$ using MATLAB's "backslash" command. Then we set $x = \hat{x} + \delta_x\|\hat{x}\|_2 w$ for 25 logarithmically equally-spaced values of $\delta_x$ between $10^{-12}$ and $10^2$, where $w$ is a random vector created in the same way as $v$. We repeat this procedure 50 times, and in each case we compute the minimum and maximum of the ratio $\mu/\nu$. For simplicity we set $\theta = 1$ here. (Different choices of the parameter $\theta$ all gave very similar results.) Results are shown in Table 4. The ratio $\mu/\nu$ is always greater than 1 and less than $\sqrt{2} \approx 1.4142$, as expected.

| Type | $\tau$ | max $\mu/\nu$ | Type | $\tau$ | max $\mu/\nu$ | Type | $\tau$ | max $\mu/\nu$ |
|---|---|---|---|---|---|---|---|---|
| (a) | 0 | 1.1589 | (b) | 0 | 1.0353 | (c) | 0 | 1.2634 |
| | $10^{-5}$ | 1.2021 | | $10^{-5}$ | 1.0672 | | $10^{-5}$ | 1.3142 |
| | $10^{-1}$ | 1.1596 | | $10^{-1}$ | 1.0638 | | $10^{-1}$ | 1.3028 |

TABLE 1

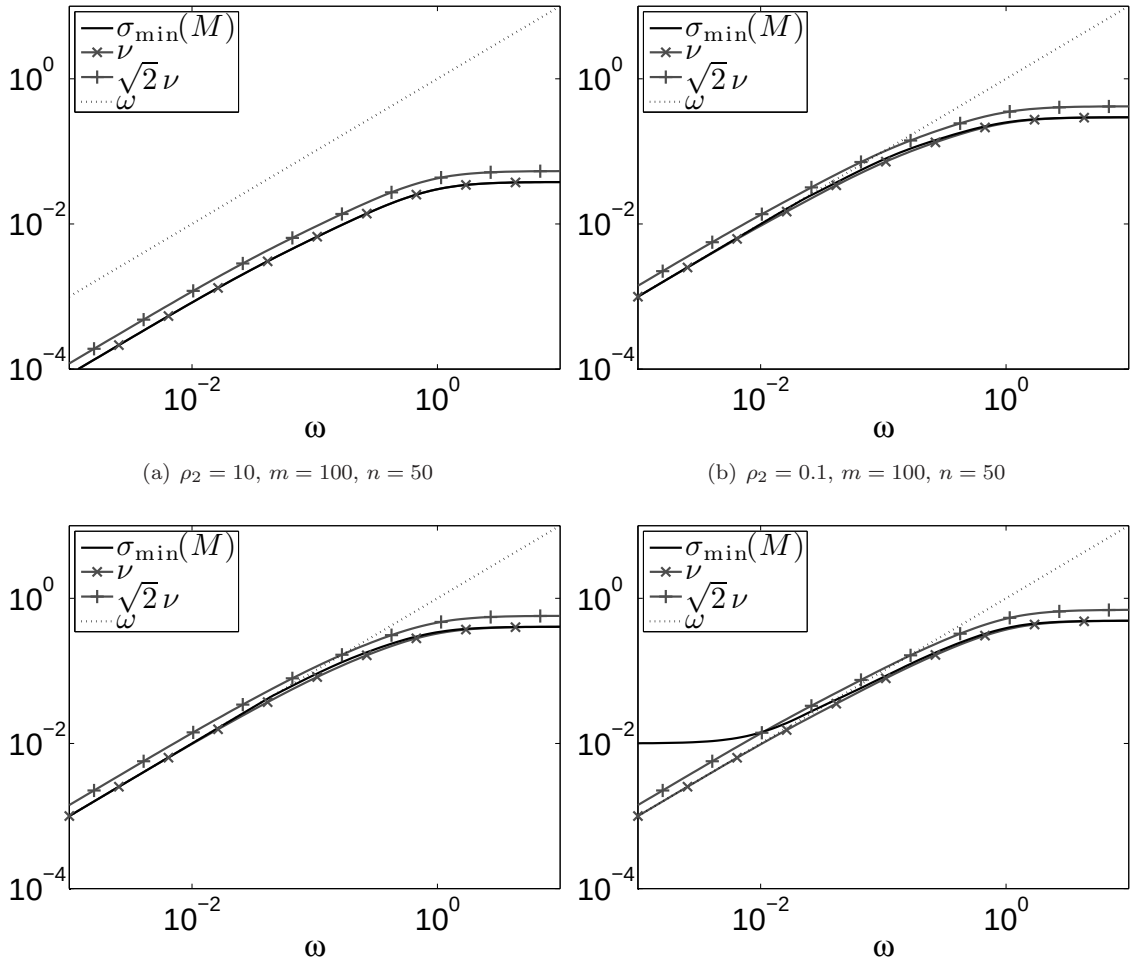
*Maxima of the ratio $\mu/\nu$ for each LS problem characterized by the matrix type and the size of the perturbation of the right-hand side $\tau$.*

The backward error $\mu$ in (1.2) and its estimate $\nu$ in (3.1) are, for a fixed matrix $A$, functions of the residual vector $r$ and the scalar $\omega$, which is related to $r$ through (1.3). However, one can consider $\mu$ and $\nu$ as functions of two independent arguments $r$ and $\omega$. The results of Section 3 do not depend on any particular relation between $r$ and $\omega$, and therefore

$$\nu \leq \mu = \min\{\omega, \sigma_{\min}(M)\} \leq \sqrt{2}\,\nu$$

holds for any $r \in \mathbb{R}^m$ and $\omega \geq 0$. In addition, if we consider $r$ and $\omega$ to be independent, $\mu$ and $\nu$ depend only on the direction of the vector $r$.

Let $A$ be the matrix of Type (a) specified above with logarithmically equally-spaced singular values between $10^{-2}$ and 1. In order to illustrate how $\nu$ approximates $\mu$, we choose a fixed vector $r = r_1 + \rho_2 r_2$, where $\rho_2$ is a scalar and $r_1$ and $r_2$ are random vectors lying in the range of $A$ and the null space of $A^T$, respectively (so that $r_1$ and $r_2$ are orthogonal). By taking suitable values of $\rho_2$ we can change the relative size of the component of the "residual" $r$ in the null-space of $A^T$, and therefore in a sense modify the level of incompatibility of the right-hand side of the LS problem. For each vector $r$, we look at how $\mu$ and $\nu$ vary as a function of $\omega$ taken from the interval $[10^{-3}, 10^{1}]$.




(a) $\rho_2 = 10$, $m = 100$, $n = 50$

(b) $\rho_2 = 0.1$, $m = 100$, $n = 50$

(c) $\rho_2 = 0$, $m = 100$, $n = 50$

(d) $\rho_2 = 1$, $m = 50$, $n = 50$

FIG. 1. *Plots of $\sigma_{\min}(M)$, $\omega$, and the bounds $\nu$ and $\sqrt{2}\,\nu$ of $\mu = \min\{\omega, \sigma_{\min}(M)\}$ for a fixed $r = r_1 + \rho_2 r_2$ with respect to $\omega$.*

We generate three vectors $r$ by setting $\rho_2 = 10$, $\rho_2 = 0.1$, and $\rho_2 = 0$, and choose $m = 100$ and $n = 50$. In Figures 1(a)-1(c), we plot $\sigma_{\min}(M)$, $\omega$, and the bounds $\nu$ and $\sqrt{2}\,\nu$ on $\mu$ as functions of the scalar variable $\omega$. Since $\operatorname{rank}(A) < m$, we have always $\sigma_{\min}(M) \leq \omega$ and hence $\mu = \sigma_{\min}(M)$ (see the comments after Theorem 2.2). In each case the lower bound $\nu$ and the upper bound $\sqrt{2}\,\nu$ are good approximations of $\mu$. This is true also in the test problem reported in Figure 1(d), where we set $m = n = 50$ so that $\operatorname{rank}(A) = m$. In this case, even though for small $\omega$ the minimal singular value of $M$ is greater than $\omega$, the estimate $\nu$ is still an excellent approximation to $\mu = \min\{\omega, \sigma_{\min}(M)\}$.

**5. Conclusion.** The projection $\nu$ given in (3.1), or estimates of $\nu$, are often used to approximate the LS backward error $\mu$ in (1.2); see, e.g., [10, 13, 6]. The known bounds (3.3) and (3.4) on the ratio $\mu/\nu$ and the asymptotic equivalence (3.5) of $\mu$ and $\nu$ suggest that $\nu$ is indeed a good estimate of $\mu$ whenever the approximate solution $x$ is sufficiently close to a LS solution $\hat{x}$. These results, however, provide no guarantees when $x$ is not close to $\hat{x}$. (In practice, of course, we do not know how close $x$ is to $\hat{x}$.) In particular, from (3.3)–(3.5) alone, it might be possible that $\nu \ll \mu$ when $\|r\|_2 \gg \|\hat{r}\|_2$. The main contribution of this work was to give constant bounds, and in particular a constant upper bound, on the ratio $\mu/\nu$ (see Theorem 3.1 and Corollary 3.2). These results prove that $\nu$ can always safely be used instead of $\mu$ in practical applications.

REFERENCES


[1] M. Arioli and S. Gratton, *Least-squares problems, normal equations, and stopping criteria for the conjugate gradient method*, Tech. Rep. RAL-TR-2008-008, Rutherford Appleton Laboratory, 2008. 1

[2] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996. 1

[3] X.-W. Chang, C. C. Paige, and D. Titley-Peloquin, *Characterizing matrices that are consistent with given solutions*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1406–1420. 2

[4] ———, *Stopping criteria for the iterative solution of linear least squares problems*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 831–852. 1, 2

[5] X.-W. Chang and D. Titley-Peloquin, *Backward perturbation analysis for scaled total least squares problems*, Numer. Linear Algebra Appl., 16 (2009), pp. 627–648. 3

[6] D. C.-L. Fong and M. A. Saunders, *LSMR: an iterative algorithm for sparse least-squares problems*, Tech. Rep. SOL 2010-2, Department of Management Science and Engineering, Stanford University, 2010. 1, 10

[7] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, third ed., 1996. 1, 2, 6

[8] J. F. Grcar, *Optimal sensitivity analysis of linear least squares*, Tech. Rep. LBNL-52434, Lawrence Berkeley National Laboratory, 2003. 2, 5

[9] J. F. Grcar, M. A. Saunders, and Z. Su, *Estimates of optimal backward perturbations for linear least squares problems*, Tech. Rep. SOL 2007-1, Department of Management Science and Engineering, Stanford University, 2007. 2, 5

[10] M. Gu, *Backward perturbation bounds for linear least squares problems*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 363–372. 1, 2, 5, 10

[11] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, PA, 2nd ed., 2002. 2, 3

[12] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, New York, NY, 1985. 4, 6

[13] P. Jiránek and D. Titley-Peloquin, *Estimating the backward error in LSQR*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2055–2074. 1, 2, 5, 10

[14] R. Karlson and B. Waldén, *Estimation of optimal backward perturbation bounds for the linear least squares problem*, BIT, 37 (1997), pp. 862–869. 2, 5

[15] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ, 1974. 1

[16] C. C. Paige and M. A. Saunders, *ALGORITHM 583; LSQR: Sparse linear equations and least-squares problems*, ACM Trans. Math. Software, 8 (1982), pp. 195–209. 2

[17] ———, *LSQR: an algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Software, 8 (1982), pp. 43–71. 1, 2

[18] Z. Su, *Computational Methods for Least Squares Problems and Clinical Trials*, PhD thesis, Stanford University, 2005. 2, 5

[19] B. Waldén, R. Karlson, and J.-G. Sun, *Optimal backward perturbation bounds for the linear least squares problem*, Numer. Linear Algebra Appl., 2 (1995), pp. 271–286. 2, 3