Linear Algebra and its Applications 605 (2020) 227–238

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Estimating the backward error for the least-squares problem with multiple right-hand sides

Eric Hallman [1]

2311 Stinson Dr, Raleigh, NC, United States of America

ARTICLE INFO

Article history:
Received 17 November 2019
Accepted 17 July 2020
Available online 24 July 2020
Submitted by V. Mehrmann

MSC:
15A06
65F99

Keywords:
Linear least-squares problem
Backward error

ABSTRACT

Let A and B be $m \times n$ and $m \times d$ matrices, and let $\widetilde{X}$ be an approximate solution to the problem $\min_X \|AX - B\|_F$. In 1996, Sun found an explicit expression for the *optimal backward error*—the size of the smallest perturbation to A (and possibly B) such that $\widetilde{X}$ is an exact solution to the perturbed problem. The expression requires finding the difference of two potentially close numbers, and so its numerical evaluation can be unstable. We offer an estimate of the backward error that can be evaluated stably and when $d = 1$ is identical to the Karlson-Waldén estimate of 1997. We prove that this estimate always approximates the optimal backward error to within a factor of $\sqrt{2}$.

Published by Elsevier Inc.

1. Introduction

Let $\widetilde{X} \in \mathbb{C}^{n \times d}$ be an approximate solution to the problem

$$\min_X \|AX - B\|_F, \tag{LS}$$

E-mail address: erhallma@ncsu.edu.

[1] This research was supported in part by the National Science Foundation through grant DMS-1745654.

https://doi.org/10.1016/j.laa.2020.07.020
0024-3795/Published by Elsevier Inc.

where $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{m\times d}$. One method for evaluating the quality of $\widetilde{X}$ as a solution is to find the smallest perturbation to A (and possibly B) such that $\widetilde{X}$ solves the perturbed problem exactly. The *optimal backward error* $\mu(\widetilde{X}, \tau)$ is thus defined in [1] as

$$\mu(\widetilde{X}, \tau) := \min\{\|[E, \tau G]\|_F \,:\, \widetilde{X} \text{ minimizes } \|(A+E)X - (B+G)\|_F\}, \tag{1}$$

where $\tau \in (0, \infty]$. If $\tau = \infty$, only perturbations to A are permitted.

In 1995, Waldén, Karlson, and Sun [2] found an exact formula for $\mu(\tilde{x}, \tau)$ in the case where $\widetilde{X} = \tilde{x}$ and $B = b$ have only a single column (i.e., $d = 1$), and Higham [3, (20.21)] proposed a stable method for computing it. There are several known estimates in the case $d = 1$, including a few cheaply computable upper bounds [4,5]. The most accurate of these estimates is the Karlson-Waldén estimate $\nu(\tilde{x}, \tau)$ [6, Eqn. 2.6], which in 2012 Gratton et al. [7] proved will always approximate $\mu(\tilde{x}, \tau)$ to within a factor of $\sqrt{2}$.

In 1996, Sun [1] gave a general formula for $\mu(\widetilde{X}, \tau)$. He observed that the formula required computing the difference of two potentially close numbers, and that its numerical evaluation could therefore be unstable [1, §5.1]. To the best of our knowledge, no stable method for computing $\mu(\widetilde{X}, \tau)$ has been discovered for $d > 1$.

In this paper we extend the Karlson-Waldén estimate to the general case $d \geq 1$ (21), and offer a stable method for computing it (17) that generalizes the expression in [8, (2.1)]. As was done for $d = 1$, we prove (Theorem 4.8) that our estimate always approximates the backward error to within a factor of $\sqrt{2}$. Although the problem of stably computing $\mu(\widetilde{X}, \tau)$ remains open, the backward error can thus at least be stably approximated.

1.1. Notation

If a matrix A has the compact SVD $U\Sigma V^*$, then the projection onto the column space of A is denoted by $\Pi_A = UU^*$ and the pseudoinverse of A by $A^\dagger = V\Sigma^{-1}U^*$. The nuclear norm of A is denoted by $\|A\|_*$. For a symmetric matrix S with eigensystem $S = \sum_i \lambda_i q_i q_i^*$, let $S_- = \sum_{\lambda_i<0} \lambda_i q_i q_i^*$.

2. Formulas and estimates for $d = 1$

In this section we provide an overview of existing formulas and estimates for the least-squares backward error in the case $d = 1$, which has received the most attention. In 1995, Waldén, Karlson, and Sun [2] showed that

$$\mu(\tilde{x}, \tau) = \left(\omega^2 + \min\left\{0, \lambda_{\min}\left(AA^* - \omega^2 rr^\dagger\right)\right\}\right)^{1/2}, \tag{2}$$

where $r = b - A\tilde{x}$ and ω is defined by Rigal and Gaches [9] as

$$\omega(\tilde{x},\tau) = \min_{E,g}\{\|[E,\tau g]\|_F \,:\, (A+E)\tilde{x} = b+g\} \tag{3}$$

$$= \frac{\|r\|_2}{\sqrt{\tau^{-2}+\|\tilde{x}\|_2^2}}. \tag{4}$$

Noting that formula (2) was "mathematically elegant, ... [but] not suitable for computation", Waldén et al. [2] offered the alternative formulation

$$\mu(\tilde{x},\tau) = \min\left\{\omega, \sigma_{\min}\left(\left[A, \omega(I-rr^\dagger)\right]\right)\right\}.$$

This expression nominally involves the computation of the smallest singular value of an $m\times(m+n)$ matrix, but Karlson and Waldén showed in [6, Lemma 3.1] that with a QR factorization of A it can be reduced to the problem of finding the smallest singular value of an $(n+1)\times 2n$ matrix.

In 1975 and 1977, Stewart [4,5] gave the respective backward perturbations

$$E_0 = \frac{r\tilde{x}^*}{\|\tilde{x}\|_2^2},\ \|E_0\|_F = \frac{\|r\|_2}{\|\tilde{x}\|_2},\quad E_1 = \frac{(\Pi_A r)\tilde{x}^*}{\|\tilde{x}\|_2^2},\ \|E_1\|_F = \frac{\|\Pi_A r\|_2}{\|\tilde{x}\|_2}$$

and

$$E_2 = -\frac{rr^*A}{\|r\|_2^2},\quad \|E_2\|_F = \frac{\|A^*r\|_2}{\|r\|_2}.$$

By modifying E_0 and E_1 to handle cases where $\tau<\infty$, we may define

$$\mu_0 := \omega,\quad \mu_1 := \omega\frac{\|\Pi_A r\|_2}{\|r\|_2},\quad \mu_2 := \frac{\|A^*r\|_2}{\|r\|_2},$$

where $\omega=\omega(\tilde{x},\tau)$ in (4). All of these quantities are upper bounds on $\mu(\tilde{x},\tau)$, and μ_0 and μ_2 are used in practice as stopping rules for iterative least-squares solvers such as LSQR and LSMR [10,11]. In 2013, Gratton et al. [12] showed that while $\min\{\mu_1,\mu_2\}$ is often close to μ, it can also overestimate the error by a factor as large as the square root of the condition number of A.

In 1997, Karlson and Waldén [6] proposed the estimate

$$\nu(\tilde{x},\tau) := \omega\|(A^*A+\omega^2 I)^{-1/2}A^*r\|_2/\|r\|_2 = \frac{\omega}{\|r\|_2}\left\|\begin{bmatrix}A\\ \omega I\end{bmatrix}\begin{bmatrix}A\\ \omega I\end{bmatrix}^\dagger\begin{bmatrix}r\\ 0\end{bmatrix}\right\|_2, \tag{5}$$

where $\omega=\omega(\tilde{x},\tau)$ as before. In the subsequent years various authors [13,14,8,15] worked to prove or experimentally verify bounds on the accuracy of this estimate. The tightest known bounds were given in 2012 by Gratton et al. [7], who proved that the bounds

$$1 \le \frac{\mu(\tilde{x},\tau)}{\nu(\tilde{x},\tau)} \le \sqrt{1+\|\Pi_A r\|_2^2/\|r\|_2^2} \le \sqrt{2} \tag{6}$$

always hold. If $r = 0$, then $\mu(\tilde{x}, \tau) = \nu(\tilde{x}, \tau) = 0$. If $r \neq 0$ but b is in the column space of A, then $\|\Pi_A r\|_2 / \|r\|_2 = 1$. If b is not in the column space of A, then

$$\lim_{A^T r \to 0} \frac{\|\Pi_A r\|_2}{\|r\|_2} = 0.$$

Thus $\nu(\tilde{x}, \tau)$ is always a good estimate of $\mu(\tilde{x}, \tau)$, and the estimate becomes increasingly accurate as $\tilde{x}$ converges to the true solution, provided the system is inconsistent.

3. Sun's results

Here we summarize Sun's main theorems from [1], with somewhat modified notation. The first theorem covers the case where only perturbations to A are permitted, but $\widetilde{X}$ has full column rank. The second theorem is a generalization of the first, allowing $\widetilde{X}$ to have any rank. The third theorem applies whenever perturbations to B are permitted.

Theorem 3.1. *Let $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{m\times d}$, and $\widetilde{X} \in \mathbb{C}^{n\times d}$ with $\operatorname{rank}(\widetilde{X}) = d$. Let $R = B - A\widetilde{X}$, and define $N = R\widetilde{X}^\dagger$. Then*

$$\mu(\widetilde{X}, \infty) = \left[\|N\|_F^2 + \operatorname{Tr}(AA^* - NN^*)_-\right]^{1/2}. \tag{7}$$

If $d = 1$ then $N = r\tilde{x}^\dagger$, in which case (7) is equivalent to (2). Since $\operatorname{Tr}(AA^* - NN^*)_-$ is equal to the sum of the negative eigenvalues of $(AA^* - NN^*)$, evaluating the right-hand side may be unstable if $\|N\|_F$ is much larger than $\mu(\widetilde{X}, \infty)$.

If $\widetilde{X}$ does not have full column rank, the formula becomes slightly more complicated.

Theorem 3.2. *Let $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{m\times d}$, and $\widetilde{X} \in \mathbb{C}^{n\times d}$. Define $R = B - A\widetilde{X}$, $N = R\widetilde{X}^\dagger$, and $M = B(I - \widetilde{X}^\dagger\widetilde{X})$. Then*

$$\mu(\widetilde{X}, \infty) = \left[\|\Pi_M A\|_F^2 + \|\bar{N}\|_F^2 + \operatorname{Tr}(\bar{A}\bar{A}^* - \bar{N}\bar{N}^*)_-\right]^{1/2}, \tag{8}$$

where $\bar{A} = (I - \Pi_M)A$ and $\bar{N} = (I - \Pi_M)N$.

Some commentary on the importance of the rank of $\widetilde{X}$: Sun notes that if $\widetilde{X}$ does not have full column rank, then we may without loss of generality write $\widetilde{X} = [\widetilde{X}_1, 0]$ where $\widetilde{X}_1$ has full column rank. We may correspondingly split $B = [B_1, B_2]$. It follows that a backward perturbation E is valid iff

$$(A + E)^*[B_1 - (A + E)\widetilde{X}_1] = 0 \quad \text{and} \quad (A + E)^* B_2 = 0. \tag{9}$$

Defining $R_1 = B_1 - A\widetilde{X}_1$, we find that $N = R_1\widetilde{X}_1^\dagger$ and $M = [0, B_2]$. The appearance of the term $\|\Pi_M A\|_F$ in (8) is therefore due to the second constraint in (9).

We present an example to illustrate the significance of the terms $\bar{A}$ and $\bar{N}$ in (8).

Example 3.3. Let

$$A = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \widetilde{X} = [1], \quad \text{and} \quad B = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

It follows from (7) that $\mu(\widetilde{X}, \infty) = \frac{\sqrt{5}-1}{2} \approx 0.618$. The optimal backward perturbation is $E = \begin{bmatrix} 0, \frac{1-\sqrt{5}}{2} \end{bmatrix}^*$.

By contrast, let

$$A = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad X = [1 \quad 0], \quad \text{and} \quad B = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}.$$

It follows from (8) that $\mu(\widetilde{X}, \infty) = 1$. The optimal backward perturbation is $E = \pm [1, 0]^*$. Even though $A^* B_2 = 0$, the backward error is different because the optimal perturbation must satisfy the constraint $(A + E)^* B_2 = 0$.

The third theorem applies when perturbations to B are permitted, in which case $\widetilde{X}$ may have arbitrary rank.

Theorem 3.4. *Let $A \in \mathbb{C}^{m \times n}$, $B \in \mathbb{C}^{m \times d}$, $\widetilde{X} \in \mathbb{C}^{n \times d}$, and $\tau \in (0, \infty)$. Let $R = B - A\widetilde{X}$ and define $\widetilde{X}_\tau = [\widetilde{X}^*, \frac{1}{\tau} I]^*$ and $N_\tau = R\widetilde{X}_\tau^\dagger$. Then*

$$\mu(\widetilde{X}, \tau) = \left[\|N_\tau\|_F^2 + \operatorname{Tr}(AA^* - N_\tau N_\tau^*)_- \right]^{1/2}. \tag{10}$$

Sun notes that when $\widetilde{X}$ has full column rank, $\mu(\widetilde{X}, \infty) = \lim_{\tau \to \infty} \mu(\widetilde{X}, \tau)$.

4. Extending the Karlson-Waldén estimate

In order to derive our extension of the Karlson-Waldén estimate, we focus on reliably estimating the quantity $\hat{\mu}(A, N)$ defined by

$$\hat{\mu}(A, N) := \left[\|N\|_F^2 + \operatorname{Tr}(AA^* - NN^*)_- \right]^{1/2} \tag{11}$$

for arbitrary matrices $A \in \mathbb{C}^{m \times n}$ and $N \in \mathbb{C}^{m \times p}$. Here we use N to emphasize the connection to $\mu(\widetilde{X}, \tau)$ via the relation $N = R\widetilde{X}^\dagger$. We begin with the following lemma, which holds for matrices of arbitrary rank and dimension.

Lemma 4.1. *For matrices $A \in \mathbb{C}^{m \times n}$ and $N \in \mathbb{C}^{m \times p}$, $A^* N = 0$ iff there is a matrix Y such that $A^* Y = 0$ and $(I - YY^\dagger)N = 0$.*

Proof. If $A^* N = 0$ then choose Y to have the same column space as N. Conversely, if Y exists, then $A^* N = (A^* Y Y^\dagger) N + A^*((I - YY^\dagger)N) = 0$. □

This lemma allows us to reformulate $\hat{\mu}(A, N)$ in a useful way.

Theorem 4.2. *Let $A \in \mathbb{C}^{m\times n}$ and $N \in \mathbb{C}^{m\times p}$. Then*

$$\hat{\mu}(A, N) = \min_{E,F}\{\|[E, F]\|_F \,:\, (A + E)^*(N + F) = 0\}. \tag{12}$$

Proof. By the preceding lemma, we can rewrite the square of the right-hand side of (12) as

$$\min_{Y,E,F}\{\|[E, F]\|_F^2 \,:\, (A + E)^*Y = 0 \text{ and } (I - YY^\dagger)(N + F) = 0\}.$$

For fixed Y the optimal perturbations are $E = -YY^\dagger A$ and $F = -(I - YY^\dagger)N$, and this minimization problem may therefore be further reduced to the problem

$$\min_Y \|YY^\dagger A\|_F^2 + \|(I - YY^\dagger)N\|_F^2. \tag{13}$$

From the properties of the trace function and Frobenius norms, it follows that

$$\begin{aligned}\min_Y \|YY^\dagger A\|_F^2 + \|(I - YY^\dagger)N\|_F^2 &= \min_Y \operatorname{Tr} Y^\dagger AA^*Y + \operatorname{Tr} N^*(I - YY^\dagger)N \\ &= \|N\|_F^2 + \min_Y \operatorname{Tr} Y^\dagger(AA^* - NN^*)Y \\ &= \|N\|_F^2 + \operatorname{Tr}(AA^* - NN^*)_- \\ &= \hat{\mu}^2(A, N).\end{aligned}$$

Taking square roots then gives the desired result. □

From the proof above, it can be seen that the optimal perturbations E and F naturally satisfy $E^*F = 0$. By rearranging the right-hand side of (12), we obtain

$$\hat{\mu}(A, N) = \min_{E,F}\{\|[E, F]\|_F \,:\, A^*F + E^*N = -A^*N \text{ and } E^*F = 0\}.$$

By removing the constraint $E^*F = 0$, we can obtain a lower bound on $\hat{\mu}(A, N)$. We define $\hat{\nu}(A, N)$ to be the solution to this relaxed problem:

$$\hat{\nu}(A, N) := \min_{E,F}\{\|[E, F]\| \,:\, A^*F + E^*N = -A^*N\}. \tag{14}$$

Thus $\hat{\nu}(A, N) \le \hat{\mu}(A, N)$ by construction.

If the singular value decompositions of A and N are $U\Sigma V^*$ and $W\Lambda Z^*$, the optimal E and F may be written as $W\hat{E}^*V^*$ and $U\hat{F}Z^*$, and so

$$\hat{\nu}(A, N) = \min_{\hat{E},\hat{F}}\{\|[\hat{E}, \hat{F}]\|_F \,:\, \Sigma\hat{F} + \hat{E}\Lambda = -\Sigma(U^*W)\Lambda\}. \tag{15}$$

If A and N have ranks r_A and r_N, both $\hat{E}$ and $\hat{F}$ will be $r_A \times r_N$ matrices. Computing their entries one coordinate at a time gives

$$\hat{E}_{ij} = \frac{-\sigma_i^2 \lambda_j (u_i^* w_j)}{\sigma_i^2 + \lambda_j^2} \quad \text{and} \quad \hat{F}_{ij} = \frac{-\sigma_i \lambda_j^2 (u_i^* w_j)}{\sigma_i^2 + \lambda_j^2},$$

and therefore

$$\hat{\nu}(A, N) = \left[\sum_{i=1}^{r_A} \sum_{j=1}^{r_N} \frac{\sigma_i^2 \lambda_j^2}{\sigma_i^2 + \lambda_j^2} (u_i^* w_j)^2 \right]^{1/2} \tag{16}$$

$$= \left[\sum_{j=1}^{r_N} \lambda_j^2 \left\| (\Sigma^2 + \lambda_j^2 I)^{-1/2} \Sigma U^* w_j \right\|_2^2 \right]^{1/2} \tag{17}$$

$$= \left[\sum_{j=1}^{r_N} \lambda_j^2 \left\| (A^* A + \lambda_j^2 I)^{-1/2} A^* w_j \right\|_2^2 \right]^{1/2} \tag{18}$$

Remark 4.3. The expression in (17) generalizes an expression found in [8, (2.1)] and elsewhere. It is a sum of nonnegative quantities and may therefore be computed stably, at least to the extent that the products $U^* w_j$ can be computed accurately.

Remark 4.4. When $d = 1$, the matrix $N = r\tilde{x}_\tau^\dagger$ has rank one. Then $\lambda_1 = \|r\|_2 / \|\tilde{x}_\tau\|_2 = \omega(\tilde{x}, \tau)$ and $w_1 = r/\|r\|$, and it follows from (18) that

$$\hat{\nu}(A, N) = \omega \| (A^* A + \omega^2 I)^{-1/2} A^* r \|_2 / \|r\|_2, \tag{19}$$

which coincides with the definition of $\nu(\tilde{x}, \tau)$ in (5). This justifies our calling $\nu(\widetilde{X}, \tau)$ (Definition 4.6 below) an extension of the Karlson-Waldén estimate.

4.1. Definition of the Karlson-Waldén estimate

We may use the definition of $\hat{\nu}(A, N)$ (14) to obtain an estimate $\nu(\widetilde{X}, \tau)$ for the backward error $\mu(\widetilde{X}, \tau)$. First, we condense Sun's results to a single theorem.

Theorem 4.5. *Let $A \in \mathbb{C}^{m \times n}$, $B \in \mathbb{C}^{m \times d}$, $\widetilde{X} \in \mathbb{C}^{n \times d}$, and $\tau \in (0, \infty]$. Let $R = B - A\widetilde{X}$ and define $\widetilde{X}_\tau = [\widetilde{X}^*, \frac{1}{\tau} I]^*$, $N_\tau = R\widetilde{X}_\tau^\dagger$, and $M = R(I - \widetilde{X}_\tau^\dagger \widetilde{X}_\tau)$. Then*

$$\mu(\widetilde{X}, \tau) = \left[\|\Pi_M A\|_F^2 + \hat{\mu}^2(\bar{A}, \bar{N}_\tau) \right]^{1/2}, \tag{20}$$

where $\bar{A} = (I - \Pi_M)A$ and $\bar{N}_\tau = (I - \Pi_M)N_\tau$.

We then obtain the definition of $\nu(\widetilde{X},\tau)$ by replacing the term $\hat{\mu}(\bar{A},\bar{N}_\tau)$ in (20) with its estimate $\hat{\nu}(\bar{A},\bar{N}_\tau)$.

Definition 4.6. With the notation used in Theorem 4.5, and $\hat{\nu}$ defined as in (14),

$$\nu(\widetilde{X},\tau) := \left[\|\Pi_M A\|_F^2 + \hat{\nu}^2(\bar{A},\bar{N}_\tau)\right]^{1/2}. \tag{21}$$

Note that if $\widetilde{X}$ has full column rank or if $\tau < \infty$, then $M = 0$, $\bar{A} = A$, and $\bar{N}_\tau = N_\tau$.

4.2. Accuracy of the Karlson-Waldén estimate

Here we prove that the Karlson-Waldén estimate is always a good estimate of $\mu(\widetilde{X},\tau)$. We do so by proving that $\hat{\nu}(A,N)$ (14) is always a good estimate of $\hat{\mu}(A,N)$ (12), and the corresponding result for $\nu(\widetilde{X},\tau)$ follows almost immediately.

Theorem 4.7. *For any matrices A and N,*

$$1 \le \frac{\hat{\mu}(A,N)}{\hat{\nu}(A,N)} \le \sqrt{1+\|\Pi_A\Pi_N\|_2} \le \sqrt{2}.$$

Proof. The first inequality is true by the way $\hat{\nu}(A,N)$ was defined. To establish the second inequality, we note that it follows from (11) that

$$\hat{\mu}(A,N) = \hat{\mu}\left(\begin{bmatrix}A\\0\end{bmatrix},\begin{bmatrix}N\\0\end{bmatrix}\right). \tag{22}$$

Let (E,F) be such that $A^*F + E^*B = -A^*B$ and $\|[E,F]\|_F = \hat{\nu}(A,N)$. Then if G_E and G_F are chosen to satisfy $G_E^*G_F = -E^*F$, the pair $\left(\begin{bmatrix}E\\G_E\end{bmatrix},\begin{bmatrix}F\\G_F\end{bmatrix}\right)$ will be a valid backward perturbation to the augmented problem (22), implying that

$$\hat{\mu}^2(A,N) \le \|[E,F]\|_F^2 + \|[G_E,G_F]\|_F^2.$$

The pair (G_E,G_F) with smallest norm satisfies $\|[G_E,G_F]\|_F^2 = 2\|E^*F\|_*$ [16, Lemma 5.1]. The optimal E and F satisfy $E = \Pi_N E$ and $F = \Pi_A F$, and so by a generalized version of Holder's inequality for Schatten norms [17, §3] we find that

$$\|E^*F\|_* = \|E^*\Pi_N\Pi_A F\|_* \le \|E^*\Pi_N\Pi_A\|_F\|F\|_F \le \|\Pi_A\Pi_N\|_2\|E\|_F\|F\|_F.$$

By the RMS-GM inequality, $2\|E\|_F\|F\|_F \le \|[E,F]\|_F^2$. Therefore,

$$\hat{\mu}^2(A,N) \le (1+\|\Pi_A\Pi_N\|_2)\|[E,F]\|_F^2 = (1+\|\Pi_A\Pi_N\|_2)\hat{\nu}^2(A,N),$$

and the desired inequality follows. The final inequality of the theorem holds because $\|\Pi_A\Pi_N\|_2 \le 1$. □

Using the formula for $\mu(\widetilde{X},\tau)$ from (20) and the definition of $\nu(\widetilde{X},\tau)$ from (21), we get our main result.

Theorem 4.8. *Let $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{m\times d}$, $\widetilde{X} \in \mathbb{C}^{n\times d}$, and $\tau \in (0,\infty]$. Then*

$$1 \leq \frac{\mu(\widetilde{X},\tau)}{\nu(\widetilde{X},\tau)} \leq \sqrt{1+\|\Pi_{\bar{A}}\Pi_{\bar{N}_\tau}\|_2} \leq \sqrt{2},$$

where $\bar{A}$ and $\bar{N}_\tau$ are defined as in Theorem 4.5.

This bound is slightly weaker than the one from (6) given by Gratton et al. [7], but still strong enough to show that the estimate $\nu(\widetilde{X},\tau)$ is increasingly accurate as $\|\Pi_{\bar{A}}\Pi_{\bar{N}_\tau}\|_2 \to 0$. In particular, we get the following corollary.

Corollary 4.9. *For $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{m\times d}$, let X_{opt} be any solution to (LS) and let $R_{opt} = B - AX_{opt}$. If $\widetilde{X}$ is constrained so that* $\operatorname{rank}(R) = \operatorname{rank}(R_{opt})$*, then*

$$\lim_{\widetilde{X}\to X_{opt}} \frac{\mu(\widetilde{X},\tau)}{\nu(\widetilde{X},\tau)} = 1$$

for any $\tau \in (0,\infty)$.

Proof. Since $\tau < \infty$, $\widetilde{X}_\tau$ has full column rank. Therefore, $\bar{N}_\tau = N_\tau = R\widetilde{X}_\tau^\dagger$, and so $\Pi_{N_\tau} = \Pi_R$. Since R has the same column rank as R_{opt} by assumption, it follows that Π_R converges to $\Pi_{R_{\text{opt}}}$ as $\widetilde{X}$ converges to X_{opt}. Since $A^*R_{\text{opt}} = 0$, we conclude that

$$\lim_{\widetilde{X}\to X_{\text{opt}}} \Pi_A\Pi_{N_\tau} = 0,$$

and the corollary then follows from Theorem 4.8. □

5. Simple backward error bounds

In the minimization problem (13) each choice of a matrix Y yields a particular backward perturbation, and by extension an upper bound on the backward error. Assuming for simplicity that $\widetilde{X}$ has full column rank, the choices $Y = 0$, $\Pi_Y = (I - \Pi_A)$, and $\Pi_Y = \Pi_R$ correspond to the respective perturbations

$$[E_0, \tau G_0] = N_\tau = R\widetilde{X}_\tau^\dagger, \quad [E_1, \tau G_1] = \Pi_A N_\tau = \Pi_A R\widetilde{X}_\tau^\dagger,$$

and

$$[E_2, \tau G_2] = [-\Pi_R A, 0] = [-RR^\dagger A, 0].$$

These are the natural extensions of Stewart's perturbations from Section 2. In particular, the pair (E_0, G_0) is the optimal backward perturbation for the consistent problem $AX = B$ and arises in the context of the total least squares problem [18].

6. Practical computation of $\nu(\widetilde{X}, \tau)$

Although we describe our estimate $\nu(\widetilde{X}, \tau)$ in terms of $\hat{\nu}(A, N_\tau)$, it is not necessary to compute N_τ explicitly. This fact is reflected in Sun's original formulas for $\mu(\widetilde{X}, \tau)$, which used the $m \times d$ matrix $\tau R(I + \tau^2 \widetilde{X}^* \widetilde{X})^{-1/2}$ in place of the $m \times n$ matrix $N_\tau = R\widetilde{X}_\tau^\dagger$.

Nor is it necessary to compute the SVD of A, despite the form of (17). Instead, we can compute the singular values Λ and left singular vectors W of N_τ, then use the close relation between formulas (18) and (5). From there, Chapter 2 of Zheng Su's thesis [8] discusses in detail methods for computing the Karlson-Waldén estimate when $d = 1$. If A is sparse then it is possible to use sparse QR methods to compute (18). If A is too large to permit direct methods, it is possible to use LSQR [10] or LSMR [11] to do the same.

We emphasize that formula (18) is *not* equivalent to computing the Karlson-Waldén estimate for each column of (LS), as the following example illustrates.

Example 6.1. Let

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \widetilde{X} = \begin{bmatrix} 1 & 1 \\ 1 & 1+\varepsilon \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

where ε is a small nonzero number. If we consider the backward error for each column individually, then the error for the first column is zero and the error for the second column is $\mathcal{O}(\varepsilon)$. If we consider the backward error for the entire system, however, we find that

$$N = RX^\dagger = \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 \\ \sqrt{2} \end{bmatrix}}_{\lambda_1 w_1} [\sqrt{2}/2, -\sqrt{2}/2],$$

and it follows from (7) and (18) that $\mu(\widetilde{X}, \infty) = 1$ and $\nu(\widetilde{X}, \infty) = \sqrt{2/3}$.

Remark 6.2. In the case $d = 1$, taking $b = 0$ and $\tilde{x} \neq 0$ gives results similar to those in the above example. One general conclusion we may draw is that if $\operatorname{rank}(\widetilde{X}) > \operatorname{rank}(B)$ then the approximate solution $\widetilde{X}$ is fundamentally flawed.

Remark 6.3. If we consider $\widetilde{X}$ as a function of ε in the example above, then

$$\lim_{\varepsilon \to 0} \frac{\mu(\widetilde{X}, \infty)}{\nu(\widetilde{X}, \infty)} = \sqrt{3/2} \neq 1.$$

The example therefore also demonstrates the importance of the condition that $\operatorname{rank}(R) = \operatorname{rank}(R_{\text{opt}})$ in Corollary 4.9.

Finally, in the event that $\operatorname{rank}(\widetilde{X}) < d$ and $\tau = \infty$, it is not necessary to compute $\bar{A} = (I - \Pi_M)A$ explicitly in order to evaluate formula (21). Using Su's work [8, §2.6], we may rewrite (18) as

$$\hat{\nu}(\bar{A}, \bar{N}_\tau) = \left[\sum_{j=1}^{r_N} \left\| \begin{bmatrix} \bar{A} \\ \lambda_j I \end{bmatrix} y_j \right\|_2^2 \right]^{1/2},$$

where each y_j solves the least-squares problem

$$\min_y \left\| \begin{bmatrix} \bar{A} \\ \lambda_j I \end{bmatrix} y - \begin{bmatrix} \lambda_j w_j \\ 0 \end{bmatrix} \right\|_2. \tag{23}$$

In this case, Λ and W are the singular values and left singular vectors of $\bar{N}_\tau = (I - \Pi_M)N_\tau$.

If we use an iterative method such as LSQR or LSMR to solve (23), we do not need to form $\bar{A} = (I - \Pi_M)A$ explicitly, but only need to compute products of the form $\bar{A}v = (I - \Pi_M)Av$ and $\bar{A}^T u = A^T(I - \Pi_M)u$. The method outlined above still requires us to compute the singular values and left singular vectors of $(I - \Pi_M)N_\tau$, but if $d \ll \min\{m, n\}$ then doing so will be inexpensive compared to the cost of forming $\bar{A}$.

Declaration of competing interest

The author declares that he has no competing interest.

Acknowledgements

The author would like to thank Tim Kelley and Ilse Ipsen for their helpful suggestions on writing this manuscript, and the referee for comments that greatly improved its presentation.

References

[1] J.-G. Sun, Optimal backward perturbation bounds for the linear least-squares problem with multiple right-hand sides, IMA J. Numer. Anal. 16 (1) (1996) 1–11, https://doi.org/10.1093/imanum/16.1.1.
[2] B. Waldén, R. Karlson, J.-G. Sun, Optimal backward perturbation bounds for the linear least squares problem, Numer. Linear Algebra Appl. 2 (3) (1995) 271–286, https://doi.org/10.1002/nla.1680020308.
[3] N.J. Higham, Accuracy and Stability of Numerical Algorithms, 2nd edition, SIAM, Philadelphia, 2002.
[4] G.W. Stewart, An inverse perturbation theorem for the linear least squares problem, SIGNUM Newsl. 10 (2–3) (1975) 39–40.

[5] G.W. Stewart, Research, development, and LINPACK, in: Mathematical Software, III, 1977, pp. 1–14.
[6] R. Karlson, B. Waldén, Estimation of backward perturbation bounds for the linear least squares problem, BIT Numer. Math. 37 (1997) 862–869, https://doi.org/10.1007/BF02510356.
[7] S. Gratton, P. Jiránek, D. Titley-Peloquin, On the accuracy of the Karlson-Walden estimate of the backward error in linear least squares problems, SIAM J. Matrix Anal. Appl. 33 (2012) 822–836, https://doi.org/10.1137/110825467.
[8] Z. Su, Computational Methods for Least Squares Problems and Clinical Trials, Ph.D. thesis, Stanford University, Stanford, CA, USA, 2005.
[9] J.L. Rigal, J. Gaches, On the compatibility of a given solution with the data of a linear system, J. ACM 14 (1967) 543–548, https://doi.org/10.1145/321406.321416.
[10] C.C. Paige, M.A. Saunders, LSQR: an algorithm for sparse linear equations and sparse least squares, ACM Trans. Math. Softw. 8 (1) (1982) 43–71, https://doi.org/10.1145/355984.355989.
[11] D. Fong, M.A. Saunders, LSMR: an iterative algorithm for sparse least squares problems, SIAM J. Sci. Comput. 33 (5) (2011) 2950–2971, https://doi.org/10.1137/10079687X.
[12] S. Gratton, P. Jiránek, D. Titley-Peloquin, Simple backward error bounds for linear least-squares problems, Linear Algebra Appl. 439 (2013) 78–89, https://doi.org/10.1016/j.laa.2013.03.007.
[13] M. Gu, Backward perturbation bounds for linear least squares problems, SIAM J. Matrix Anal. Appl. 20 (2) (1998) 363–372, https://doi.org/10.1137/S0895479895296446.
[14] J.F. Grcar, Optimal sensitivity analysis of linear least squares, Tech. Rep. LBNL-52434, Lawrence Berkeley National Laboratory, Berkeley, CA, 2003.
[15] J.F. Grcar, M.A. Saunders, 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, Stanford, CA, 2007.
[16] B. Recht, M. Fazel, P.A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, SIAM Rev. 52 (3) (2010) 471–501, https://doi.org/10.1137/070697835.
[17] K. Ball, E.A. Carlen, E.H. Lieb, Sharp uniform convexity and smoothness inequalities for trace norms, Invent. Math. 115 (1) (1994) 463–482, https://doi.org/10.1007/BF01231769.
[18] S. Van Huffel, J. Vandewalle, The Total Least Squares Problem: Computational Aspects and Analysis, SIAM, Philadelphia, 1991.