Estimating the backward error in linear least squares problems

Pavel Jiránek[1]

joint work with Serge Gratton[1] and David Titley-Peloquin[2]

[1]CERFACS, Toulouse, France

[2]Mathematical Institute, University of Oxford, UK

Seminar on Numerical Analysis SNA 2011

January 24–28, 2011

Outline

1. Introduction
2. Backward error in LS problems
3. Estimates of the LS backward error
4. LSQR algorithm and implementation of the estimates
5. Numerical experiments

Outline

1. Introduction
2. Backward error in LS problems
3. Estimates of the LS backward error
4. LSQR algorithm and implementation of the estimates
5. Numerical experiments

Let $A \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(A) = n$ and $b \in \mathbb{R}^m$.

Linear LS problem

Find $\hat{x} \in \mathbb{R}^n$ such that $\quad \|b - A\hat{x}\|_2 = \min_{x\in\mathbb{R}^n} \|b - Ax\|_2.$ (LS)

The unique solution $\hat{x}$ satisfies the system of normal equations

$$A^T A\hat{x} = A^T b \quad \Rightarrow \quad \hat{x} = (A^T A)^{-1} A^T b \equiv A^\dagger b.$$

Least squares problems & Acceptable solutions

Let $A \in \mathbb{R}^{m \times n}$ with $\operatorname{rank}(A) = n$ and $b \in \mathbb{R}^m$.

Linear LS problem

$$\text{Find } \hat{x} \in \mathbb{R}^n \text{ such that } \quad \|b - A\hat{x}\|_2 = \min_{x \in \mathbb{R}^n} \|b - Ax\|_2. \qquad \text{(LS)}$$

The unique solution $\hat{x}$ satisfies the system of normal equations

$$A^T A \hat{x} = A^T b \quad \Rightarrow \quad \hat{x} = (A^T A)^{-1} A^T b \equiv A^\dagger b.$$

We call an x to be an acceptable solution of (LS) if and only if x is the solution of a nearby LS problem

$$(A + E)^T [(b + f) - (A + E)x] = 0, \quad \|E\|_F \leq \alpha \|A\|_F, \quad \|f\|_2 \leq \beta \|b\|_2$$

for some given tolerances α and β.

Outline

Backward error in LS problems I

In backward error analysis, we interpret a given approximation x to the solution of a problem with A and b as the solution of a problem with perturbed data $A + E$ and $b + f$. In addition, we require E and f to be minimal in some sense.

Backward error for consistent systems

Let $x \neq 0$ and $\theta > 0$ be given and $r = b - Ax$. Then

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

Rigal and Gaches (1967), Higham (2002)

Backward error in LS problems II

For LS problems, we define the backward error associated with x by

$$\mu \equiv \min_{E,f}\{\|[E,\theta f]\|_F;\ (A+E)^T[(b+f)-(A+E)x]=0\}.$$

Backward error for LS problems

Let $x \neq 0$ and $\theta > 0$ be given and $r = b - Ax$. Let

$$\omega \equiv \frac{\theta\|r\|_2}{\sqrt{1+\theta^2\|x\|_2^2}}, \qquad N \equiv \begin{bmatrix} A^T \\ \omega(I - rr^\dagger) \end{bmatrix}.$$

Then

$$\mu = \min\{\omega,\ \sigma\}, \qquad \sigma \equiv \sigma_{\min}(N).$$

Waldén, Karlson, and Sun (1995), Higham (2002)

Backward error in LS problems III

Recall that we want to have E and f such that

$$\|E\|_F \leq \alpha\|A\|_F, \qquad \|f\|_2 \leq \beta\|b\|_2$$

for some tolerances α and β.

This is achieved if

$$\mu \leq \alpha\|A\|_F \qquad \text{with} \quad \theta = \frac{\alpha\|A\|_F}{\beta\|b\|_2}.$$

Chang, Paige, and Titley-Peloquin (2009)

Backward error in LS problems IV

Some properties of μ:

- For inconsistent problems, the backward error is given entirely by σ:

$$b \notin \mathcal{R}(A) \quad \Rightarrow \quad \sigma < \omega \quad \Rightarrow \quad \mu = \sigma.$$

- The same holds for overdetermined problems:

$$\operatorname{rank}(A) < m \quad \Rightarrow \quad \sigma \leq \omega \quad \Rightarrow \quad \mu = \sigma.$$

- For consistent problems, the backward error depends on the "relative error" associated with x:

$$b \in \mathcal{R}(A) \quad \Rightarrow \quad \sigma < \omega \quad \Leftrightarrow \quad \frac{\|A^\dagger r\|_2}{\sqrt{1 + \theta^2 \|x\|_2^2}} > 1.$$

Outline

1. Introduction
2. Backward error in LS problems
3. Estimates of the LS backward error
4. LSQR algorithm and implementation of the estimates
5. Numerical experiments

Stewart's estimates of the backward error

First upper bounds for the backward error in LS problems were given by Stewart (1975, 1977).

Recall that

$$\mu = \min\{\omega, \sigma\}, \qquad \sigma = \sigma_{\min}(N) = \frac{\|Nr_*\|_2}{\|r_*\|_2}, \qquad N \equiv \begin{bmatrix} A^T \\ \omega(I - rr^\dagger) \end{bmatrix}.$$

for some r_* which is equal to the residual in the optimally perturbed problem: $r_* \equiv (b + f_*) - (A + E_*)x$.

Stewart's estimates of the backward error

First upper bounds for the backward error in LS problems were given by Stewart (1975, 1977).

Recall that

$$\mu = \min\{\omega, \sigma\}, \qquad \sigma = \sigma_{\min}(N) = \frac{\|Nr_*\|_2}{\|r_*\|_2}, \qquad N \equiv \begin{bmatrix} A^T \\ \omega(I - rr^\dagger) \end{bmatrix}.$$

for some r_* which is equal to the residual in the optimally perturbed problem: $r_* \equiv (b + f_*) - (A + E_*)x$.

Stewart's bounds = Rayleigh quotient approximations of σ with $r = b - Ax$ and $\hat{r} = b - A\hat{x}$:

$$\overline{\mu}_1 \equiv \frac{\|Nr\|_2}{\|r\|_2} = \frac{\|A^T r\|_2}{\|r\|_2}, \qquad \overline{\mu}_2 \equiv \frac{\|N\hat{r}\|_2}{\|\hat{r}\|_2} = \frac{\theta\|P_A r\|_2}{\sqrt{1 + \theta^2\|x\|_2^2}} = \omega\frac{\|P_A r\|_2}{\|r\|_2},$$

where $P_A \equiv AA^\dagger$ is the orthogonal projector onto $\mathcal{R}(A)$.

Karlson-Waldén's estimate

The LS backward error σ is given implicitly by

$$\sigma = \frac{\omega}{\|r\|_2} \left\| \begin{bmatrix} A \\ \sqrt{\omega^2 - \sigma^2} I \end{bmatrix} \begin{bmatrix} A \\ \sqrt{\omega^2 - \sigma^2} I \end{bmatrix}^{\dagger} \begin{bmatrix} r \\ 0 \end{bmatrix} \right\|_2 .$$

Karlson-Waldén's estimate

The LS backward error σ is given implicitly by

$$\sigma = \frac{\omega}{\|r\|_2} \left\| \begin{bmatrix} A \\ \sqrt{\omega^2 - \sigma^2} I \end{bmatrix} \begin{bmatrix} A \\ \sqrt{\omega^2 - \sigma^2} I \end{bmatrix}^{\dagger} \begin{bmatrix} r \\ 0 \end{bmatrix} \right\|_2 .$$

Karlson and Waldén (1997) proposed to estimate the backward error μ by the quantity

$$\nu \equiv \frac{\omega}{\|r\|_2} \left\| \begin{bmatrix} A \\ \omega I \end{bmatrix} \begin{bmatrix} A \\ \omega I \end{bmatrix}^{\dagger} \begin{bmatrix} r \\ 0 \end{bmatrix} \right\|_2 .$$

Properties of the Karlson-Waldén's estimate

The following properties of the estimate ν can be found in literature:

Properties of the Karlson-Waldén's estimate

The following properties of the estimate ν can be found in literature:

- Karlson and Waldén (1997) showed that ν is (up to a constant) a lower bound on μ:

$$0.5858\,\nu \approx (2-\sqrt{2})\,\nu \leq \mu.$$

Properties of the Karlson-Waldén's estimate

The following properties of the estimate ν can be found in literature:

- Karlson and Waldén (1997) showed that ν is (up to a constant) a lower bound on μ:

$$0.5858\,\nu \approx (2-\sqrt{2})\,\nu \leq \mu.$$

- Gu (1998) improved the lower bound of Karlson and Waldén (1997) and provided an upper bound:

$$0.6180\,\nu \approx \frac{2}{1+\sqrt{5}}\,\nu \leq \mu \leq \frac{\|r\|_2}{\|\hat{r}\|_2}\,\nu.$$

Properties of the Karlson-Waldén's estimate

The following properties of the estimate ν can be found in literature:

- Karlson and Waldén (1997) showed that ν is (up to a constant) a lower bound on μ:

$$0.5858\,\nu \approx (2-\sqrt{2})\,\nu \leq \mu.$$

- Gu (1998) improved the lower bound of Karlson and Waldén (1997) and provided an upper bound:

$$0.6180\,\nu \approx \frac{2}{1+\sqrt{5}}\,\nu \leq \mu \leq \frac{\|r\|_2}{\|\hat{r}\|_2}\,\nu.$$

- Grcar (2003) showed that ν is asymptotically equivalent to μ:

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

Properties of the Karlson-Waldén's estimate

The following properties of the estimate ν can be found in literature:

- Karlson and Waldén (1997) showed that ν is (up to a constant) a lower bound on μ:

$$0.5858\,\nu \approx (2-\sqrt{2})\,\nu \leq \mu.$$

- Gu (1998) improved the lower bound of Karlson and Waldén (1997) and provided an upper bound:

$$0.6180\,\nu \approx \frac{2}{1+\sqrt{5}}\,\nu \leq \mu \leq \frac{\|r\|_2}{\|\hat{r}\|_2}\,\nu.$$

- Grcar (2003) showed that ν is asymptotically equivalent to μ:

$$\lim_{x\rightarrow\hat{x}} \frac{\mu}{\nu} = 1.$$

Altogether, these results indicate that ν is an accurate estimate of the backward error μ provided x is a sufficiently accurate approximation of $\hat{x}$.

Properties of the Karlson-Waldén's estimate

However, a good estimate of the backward error should be accurate also for inaccurate approximations.

Properties of the Karlson-Waldén's estimate

However, a good estimate of the backward error should be accurate also for inaccurate approximations.

New result

$$\nu \le \mu \le \left(2 - \frac{\|\hat{r}\|_2^2}{\|r\|_2^2}\right)^{1/2} \nu \le \sqrt{2}\,\nu$$

Gratton, J, Titley-Peloquin (201?)

The estimate ν is always a good approximation of the backward error μ.

Approximation properties of Stewart's estimates

We can use the approximation properties of ν to analyze the accuracy of Stewart's estimates

$$\overline{\mu}_1 = \frac{\|A^T r\|_2}{\|r\|_2}, \qquad \overline{\mu}_2 = \omega \frac{\|P_A r\|_2}{\|r\|_2}.$$

and get

$$\frac{1}{\sqrt{1+\sigma_{\max}^2(A)/\omega^2}}\,\overline{\mu}_1 \le \mu \le \overline{\mu}_1, \qquad \frac{1}{\sqrt{1+\omega^2/\sigma_{\min}^2(A)}}\,\overline{\mu}_2 \le \mu \le \overline{\mu}_2.$$

Approximation properties of Stewart's estimates

We can use the approximation properties of ν to analyze the accuracy of Stewart's estimates

$$\overline{\mu}_1 = \frac{\|A^T r\|_2}{\|r\|_2}, \qquad \overline{\mu}_2 = \omega \frac{\|P_A r\|_2}{\|r\|_2}.$$

and get

$$\frac{1}{\sqrt{1 + \sigma_{\max}^2(A)/\omega^2}}\,\overline{\mu}_1 \leq \mu \leq \overline{\mu}_1, \qquad \frac{1}{\sqrt{1 + \omega^2/\sigma_{\min}^2(A)}}\,\overline{\mu}_2 \leq \mu \leq \overline{\mu}_2.$$

Let $\overline{\mu} \equiv \min\{\overline{\mu}_1, \overline{\mu}_2\}$. If $\omega \geq \sigma_{\max}(A)$ or $\omega \leq \sigma_{\min}(A)$, then

$$\frac{1}{\sqrt{2}}\,\overline{\mu} \leq \mu \leq \overline{\mu}.$$

In the worst case, we get

$$\frac{1}{\sqrt{1 + \kappa_2(A)}}\,\overline{\mu} \leq \mu \leq \overline{\mu}.$$

Outline

LSQR algorithm I

We want to find $\hat{x}$ such that

$$\|b - A\hat{x}\|_2 = \min_{x \in \mathbb{R}^n} \|b - Ax\|_2.$$

Let V_k be an orthonormal basis of $\mathcal{K}_k \equiv \mathcal{K}_k(A^T A, A^T b)$. We look instead for $x_k = V_k y_k$ such that

$$\|b - Ax_k\|_2 = \min_{x \in \mathcal{K}_k} \|b - Ax\|_2 = \min_{y \in \mathbb{R}^k} \|b - AV_k y\|_2$$

$$\Rightarrow \quad x_k = A_k^{\dagger} b, \qquad A_k \equiv AV_k V_k^T.$$

LSQR $\equiv$ CG on $A^T Ax = A^T b$.

Paige and Saunders (1982a,b), Hestenes and Stiefel (1952)

LSQR algorithm II

Golub-Kahan bidiagonalization:

$$U_{k+1}(\beta_1 e_1) = b, \quad AV_k = U_{k+1}B_k, \quad A^T U_{k+1} = V_{k+1}\overline{B}_k^T$$

Golub and Kahan (1965)

Solution of the reduced LS problem:

$$\|b - Ax_k\|_2 = \|b - AV_k y_k\|_2 = \min_y \|\beta_1 e_1 - B_k y\|_2.$$

$$Q_k[B_k, \beta_1 e_1] = \begin{bmatrix} R_k & f_k \\ 0 & \bar{\phi}_{k+1} \end{bmatrix} \quad \Rightarrow \quad y_k = R_k^{-1} f_k$$

Stewart's estimates

The estimate

$$\overline{\mu}_1(x_k) = \frac{\|A^T r_k\|_2}{\|r_k\|_2}$$

can be easily evaluated in LSQR with the cost of $\mathcal{O}(1)$ operations.

Stewart's estimates

The estimate

$$\overline{\mu}_1(x_k) = \frac{\|A^T r_k\|_2}{\|r_k\|_2}$$

can be easily evaluated in LSQR with the cost of $\mathcal{O}(1)$ operations.

The estimate

$$\overline{\mu}_2(x_k) = \omega_k \frac{\|P_A r_k\|_2}{\|r_k\|_2}$$

needs to evaluate the norm of the projection of r_k onto the range of A, which is equal to the energy norm of the error in the underlying CG method,

$$\|P_A r_k\|_2 = \|A(\hat{x} - x_k)\|_2 = \|\hat{x} - x_k\|_{A^T A},$$

Stewart's estimates

The estimate

$$\overline{\mu}_1(x_k) = \frac{\|A^T r_k\|_2}{\|r_k\|_2}$$

can be easily evaluated in LSQR with the cost of $\mathcal{O}(1)$ operations.

The estimate

$$\overline{\mu}_2(x_k) = \omega_k \frac{\|P_A r_k\|_2}{\|r_k\|_2}$$

needs to evaluate the norm of the projection of r_k onto the range of A, which is equal to the energy norm of the error in the underlying CG method,

$$\|P_A r_k\|_2 = \|A(\hat{x} - x_k)\|_2 = \|\hat{x} - x_k\|_{A^T A},$$

and can be approximated, e.g., by

$$\|P_A r_k\|_2 \approx \|P_{A_{k+d}} r_k\|_2$$

with the cost of $\mathcal{O}(d)$ operations ($+d$ iterations).

(J, Titley-Peloquin (2010))

Karlson-Waldén's estimate

The value of

$$\nu_k = \frac{\omega_k}{\|r_k\|_2} \|(A^T A + \omega_k^2 I)^{-1/2} A^T r_k\|_2$$

can be approximated by

$$\underline{\nu}_{k,d} = \frac{\omega_k}{\|r_k\|_2} \|(B_{k+d}^T B_{k+d} + \omega_k^2 I)^{-1/2} B_{k+d}^T t_k\|_2,$$
$$\overline{\nu}_{k,d} = \frac{\omega_k}{\|r_k\|_2} \|(\overline{B}_{k+d}^T \overline{B}_{k+d} + \omega_k^2 I)^{-1/2} \overline{B}_{k+d}^T t_k\|_2$$

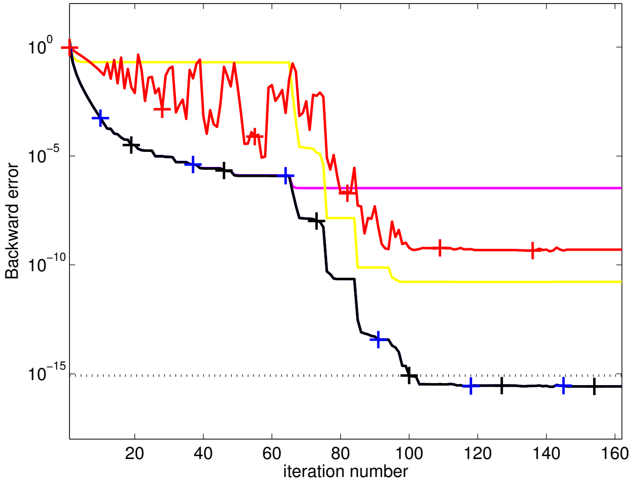
with the cost of $\mathcal{O}(k+d)$ operations ($+d$ iterations) and we have

$$\underline{\nu}_{k,d} \leq \nu_k \leq \overline{\nu}_{k,d}.$$

(J, Titley-Peloquin (2010))

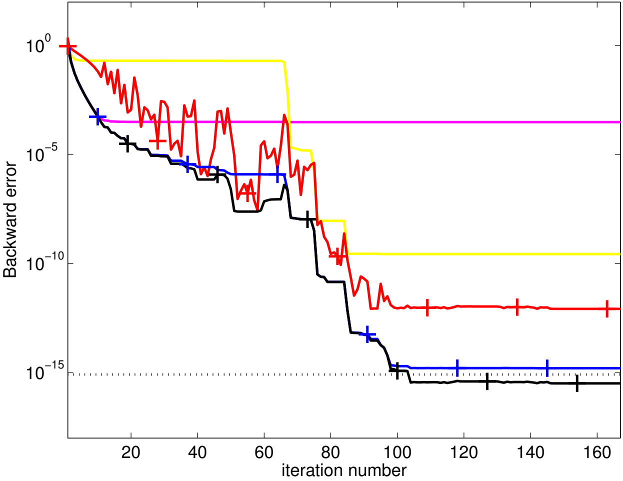
Outline

Numerical experiments



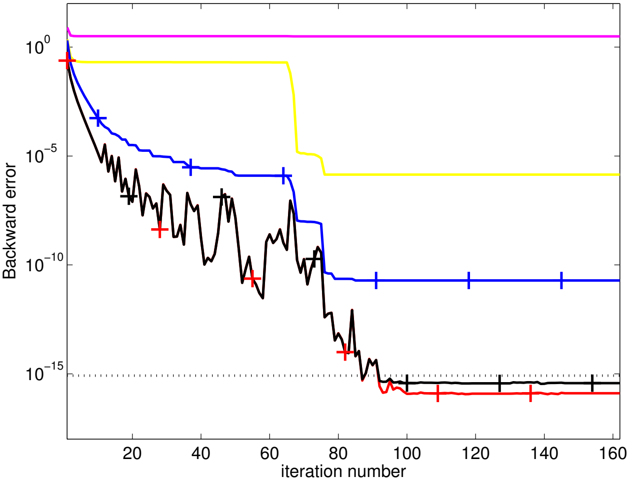
μ, $\overline{\mu}_1$, $\overline{\mu}_2$, ω, relative error

Numerical experiments



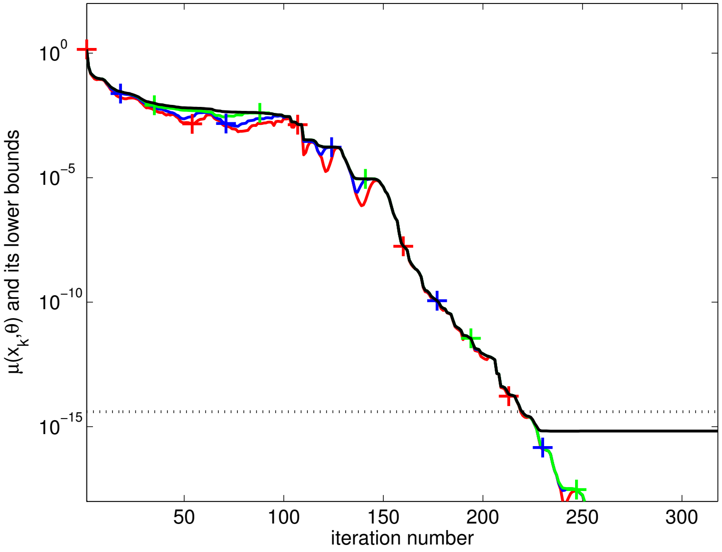
μ, $\overline{\mu}_1$, $\overline{\mu}_2$, ω, relative error

Numerical experiments



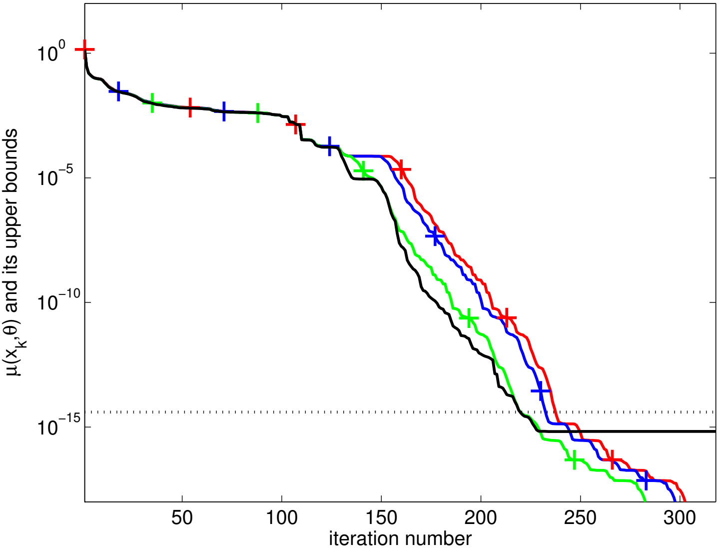
μ, $\overline{\mu}_1$, $\overline{\mu}_2$, ω, relative error

Numerical experiments



$d = 5$, $d = 10$, $d = 20$

Numerical experiments



$d = 5$, $d = 10$, $d = 20$

Thank you for your attention!

References I

X.-W. Chang, C. C. Paige, and D. Titley-Peloquin. Stopping criteria for the iterative solution of linear least squares problems. *SIAM J. Matrix Anal. Appl.*, 31(2):831–852, 2009.

G. H. Golub and W. M. Kahan. Calculating the singular values and pseudo-inverse of a matrix. *SIAM J. Numer. Anal. Ser. B*, 2(2):205–224, 1965.

S. Gratton, P. Jiránek, and D. Titley-Peloquin. On the accuracy of the Karlson-Waldén's estimate of the backward error in linear least squares problems. Submitted to *SIAM J. Matrix Anal. Appl.*

J. F. Grcar. Optimal sensitivity analysis of linear least squares. Technical Report LBNL-52434, Lawrence Berkeley National Laboratory, 2003.

M. Gu. Backward perturbation bounds for linear least squares problems. *SIAM J. Matrix Anal. Appl.*, 20(2):363–372, 1998.

M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Stand.*, 49(6):409–435, 1952.

N. J. Higham. *Accuracy and Stability of Numerical Algorithms*. SIAM, Philadelphia, PA, 2nd edition, 2002.

References II

P. Jiránek and D. Titley-Peloquin. Estimating the backward error in LSQR. *SIAM J. Matrix Anal. Appl.*, 31(4):2055–2074, 2010.

R. Karlson and B. Waldén. Estimation of optimal backward perturbation bounds for the linear least squares problem. *BIT*, 37(4):862–869, 1997.

C. C. Paige and M. A. Saunders. LSQR: an algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Software*, 8(1):43–71, 1982a.

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

J. L. Rigal and J. Gaches. On the compatibility of a given solution with the data of a linear system. *J. Assoc. Comput. Mach.*, 14(3):543–548, 1967.

G. W. Stewart. An inverse perturbation theorem for the linear least squares problems. *SIGNUM Newsletter*, 10:39–40, 1975.

G. W. Stewart. Research, development, and LINPACK. In *Mathematical Software III*, pages 1–14. Academic Press, New York, 1977.

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