Estimating the backward error in linear least squares problems

Pavel Jiránek\textsuperscript{1}

joint work with Serge Gratton\textsuperscript{1} and David Titley-Peloquin\textsuperscript{2}

\textsuperscript{1}CERFACS, Toulouse, France

\textsuperscript{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 $\text{rank}(A) = n$ and $b \in \mathbb{R}^m$.

**Linear LS problem**

Find $\hat{x} \in \mathbb{R}^n$ such that

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

(LS)

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

$$A^T A \hat{x} = A^T b \quad \Rightarrow \quad \hat{x} = (A^T A)^{-1} A^T b \equiv A^\dagger b.$$
Let $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = n$ and $b \in \mathbb{R}^m$.

**Linear LS problem**

Find $\hat{x} \in \mathbb{R}^n$ such that $\|b - A\hat{x}\|_2 = \min_{x \in \mathbb{R}^n} \|b - Ax\|_2$. \hspace{1cm} (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 $\alpha$ and $\beta$. 
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
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

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

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

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

Then

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

Recall that we want to have $E$ and $f$ such that

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

for some tolerances $\alpha$ and $\beta$.

This is achieved if

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

Chang, Paige, and Titley-Peloquin (2009)
Some properties of $\mu$:

- For inconsistent problems, the backward error is given entirely by $\sigma$:

  \[ b \notin \mathcal{R}(A) \implies \sigma < \omega \implies \mu = \sigma. \]

- The same holds for overdetermined problems:

  \[ \text{rank}(A) < m \implies \sigma \leq \omega \implies \mu = \sigma. \]

- For consistent problems, the backward error depends on the “relative error” associated with $x$:

  \[ b \in \mathcal{R}(A) \implies \sigma < \omega \iff \frac{\|A^\dagger r\|_2}{\sqrt{1 + \theta^2\|x\|_2^2}} > 1. \]
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

$$\mu = \min\{\omega, \sigma\}, \quad \sigma = \sigma_{\text{min}}(N) = \frac{\|Nr_*\|_2}{\|r_*\|_2}, \quad N \equiv \left[ A^T \omega(I - rr^\dagger) \right].$$

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
\[ \sigma = \sigma_{\min}(N) = \frac{\|Nr\|_2}{\|r\|_2}, \quad N \equiv \left[ \begin{array}{c} A^T \\ \omega(I - rr^\dagger) \end{array} \right]. \]

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 \( \sigma \) with \( r = b - Ax \) and \( \hat{r} = b - A\hat{x} \):
\[ \bar{\mu}_1 \equiv \frac{\|Nr\|_2}{\|r\|_2} = \frac{\|A^Tr\|_2}{\|r\|_2}, \quad \bar{\mu}_2 \equiv \frac{\|N\hat{r}\|_2}{\|\hat{r}\|_2} = \frac{\theta \|P_Ar\|_2}{\sqrt{1 + \theta^2}\|x\|_2^2} = \omega \frac{\|P_Ar\|_2}{\|r\|_2}, \]

where \( P_A \equiv AA^\dagger \) is the orthogonal projector onto \( \mathcal{R}(A) \).
Karlson-Waldén’s estimate

The LS backward error $\sigma$ is given implicitly by

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

The LS backward error $\sigma$ is given implicitly by

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

Karlson and Waldén (1997) proposed to estimate the backward error $\mu$ by the quantity

$$\nu \equiv \frac{\omega}{\|r\|_2} \left\| \begin{bmatrix} A \\ \omega I \end{bmatrix} \begin{bmatrix} A \\ \omega I \end{bmatrix}^\dagger \begin{bmatrix} r \\ 0 \end{bmatrix} \right\|_2.$$
The following properties of the estimate $\nu$ can be found in literature:

Karlson and Waldén (1997) showed that $\nu$ is (up to a constant) a lower bound on $\mu$:

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

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

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

Grcar (2003) showed that $\nu$ is asymptotically equivalent to $\mu$:

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

Altogether, these results indicate that $\nu$ is an accurate estimate of the backward error $\mu$ provided $x$ is a sufficiently accurate approximation of $\hat{x}$. 
Properties of the Karlson-Waldén’s estimate

The following properties of the estimate $\nu$ can be found in literature:

- Karlson and Waldén (1997) showed that $\nu$ is (up to a constant) a lower bound on $\mu$:
  \[
  0.5858 \nu \approx (2 - \sqrt{2}) \nu \leq \mu.
  \]
Properties of the Karlson-Waldén’s estimate

The following properties of the estimate $\nu$ can be found in literature:

- Karlson and Waldén (1997) showed that $\nu$ is (up to a constant) a lower bound on $\mu$:
  \[
  0.5858 \nu \approx (2 - \sqrt{2}) \nu \leq \mu.
  \]

- Gu (1998) improved the lower bound of Karlson and Waldén (1997) and provided an upper bound:
  \[
  0.6180 \nu \approx \frac{2}{1 + \sqrt{5}} \nu \leq \mu \leq \frac{\|r\|_2}{\|\hat{r}\|_2} \nu.
  \]
The following properties of the estimate $\nu$ can be found in literature:

- **Karlson and Waldén (1997)** showed that $\nu$ is (up to a constant) a lower bound on $\mu$:
  \[
  0.5858 \nu \approx (2 - \sqrt{2}) \nu \leq \mu.
  \]

- **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 $\nu$ is asymptotically equivalent to $\mu$:
  \[
  \lim_{x \to \hat{x}} \frac{\mu}{\nu} = 1.
  \]
Properties of the Karlson-Waldén’s estimate

The following properties of the estimate $\nu$ can be found in literature:

- **Karlson and Waldén (1997)** showed that $\nu$ is (up to a constant) a lower bound on $\mu$:

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

- **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 $\nu$ is asymptotically equivalent to $\mu$:

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

Altogether, these results indicate that $\nu$ is an accurate estimate of the backward error $\mu$ provided $x$ is a sufficiently accurate approximation of $\hat{x}$. 
However, a good estimate of the backward error should be accurate also for inaccurate approximations.
However, a good estimate of the backward error should be accurate also for inaccurate approximations.

New result

\[ \nu \leq \mu \leq \left( 2 - \frac{\|\hat{r}\|_2^2}{\|r\|_2^2} \right)^{1/2} \nu \leq \sqrt{2} \nu \]

Gratton, J, Titley-Peloquin (201?)

The estimate \( \nu \) is always a good approximation of the backward error \( \mu \).
Approximation properties of Stewart’s estimates

We can use the approximation properties of $\nu$ to analyze the accuracy of Stewart’s estimates

\[ \overline{\mu}_1 = \frac{\|A^T r\|_2}{\|r\|_2}, \quad \overline{\mu}_2 = \omega \frac{\|P_A r\|_2}{\|r\|_2}. \]

and get

\[ \frac{1}{\sqrt{1 + \sigma_{\text{max}}^2(A)/\omega^2}} \overline{\mu}_1 \leq \mu \leq \overline{\mu}_1, \quad \frac{1}{\sqrt{1 + \omega^2/\sigma_{\text{min}}^2(A)}} \overline{\mu}_2 \leq \mu \leq \overline{\mu}_2. \]
We can use the approximation properties of $\nu$ to analyze the accuracy of Stewart’s estimates

$$
\bar{\mu}_1 = \frac{\|A^T r\|_2}{\|r\|_2}, \quad \bar{\mu}_2 = \omega \frac{\|PA r\|_2}{\|r\|_2}.
$$

and get

$$
\frac{1}{\sqrt{1 + \sigma^2_{\text{max}}(A)/\omega^2}} \bar{\mu}_1 \leq \mu \leq \bar{\mu}_1,
\frac{1}{\sqrt{1 + \omega^2/\sigma^2_{\text{min}}(A)}} \bar{\mu}_2 \leq \mu \leq \bar{\mu}_2.
$$

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

$$
\frac{1}{\sqrt{2}} \bar{\mu} \leq \mu \leq \bar{\mu}.
$$

In the worst case, we get

$$
\frac{1}{\sqrt{1 + \kappa^2_2(A)}} \bar{\mu} \leq \mu \leq \bar{\mu}.
$$
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
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

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

\[ \Rightarrow \quad x_k = A_k^\dagger b, \quad A_k \equiv AV_k V_k^T. \]

\( \text{LSQR} \equiv \text{CG on } A^T Ax = A^T b. \)

Paige and Saunders (1982a,b), Hestenes and Stiefel (1952)
Golub-Kahan bidiagonalization:

\[ U_{k+1}(\beta_1 e_1) = b, \quad AV_k = U_{k+1}B_k, \quad A^T U_{k+1} = V_{k+1}B_k^T \]

Golub and Kahan (1965)

Solution of the reduced LS problem:

\[ \|b - Ax_k\|_2 = \|b - AV_ky_k\|_2 = \min_y \|\beta_1 e_1 - B_ky\|_2. \]

\[ Q_k[B_k, \beta_1 e_1] = \begin{bmatrix} R_k & f_k \\ 0 & \phi_{k+1} \end{bmatrix} \implies y_k = R_k^{-1} f_k \]
Stewart’s estimates

The estimate

\[ \bar{\mu}_1(x_k) = \frac{\|A^T r_k\|_2}{\|r_k\|_2} \]

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

The estimate

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

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

The estimate

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

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

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

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

The estimate

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

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 $O(d)$ operations ($+d$ iterations).

(J, Titley-Peloquin (2010))
Karlson-Waldén’s estimate

The value of

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

can be approximated by

\[ \nu_{k,d} = \frac{\omega_k}{\|r_k\|_2} \left\| \left( B_{k+d}^T B_{k+d} + \omega_k^2 I \right)^{-1/2} B_{k+d}^T t_k \right\|_2, \]

\[ \overline{\nu}_{k,d} = \frac{\omega_k}{\|r_k\|_2} \left\| \left( \overline{B}_{k+d}^T \overline{B}_{k+d} + \omega_k^2 I \right)^{-1/2} \overline{B}_{k+d}^T t_k \right\|_2 \]

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

\[ \nu_{k,d} \leq \nu_k \leq \overline{\nu}_{k,d}. \]

(J, Titley-Peloquin (2010))
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
Numerical experiments

Estimating the backward error in linear LS problems
Numerical experiments

Estimating the backward error in linear LS problems
Numerical experiments

Estimating the backward error in linear LS problems

\( \mu, \mu_1, \mu_2, \omega, \) relative error
Numerical experiments

\[ \mu(x_k, \theta) \] and its lower bounds

\begin{align*}
\text{iteration number} & = 50, 100, 150, 200, 250, 300 \\
10 & \quad -15 \\
10 & \quad -10 \\
10 & \quad 0 \\
\end{align*}

\[ d = 5, \ d = 10, \ d = 20 \]
Numerical experiments

$d = 5$, $d = 10$, $d = 20$
Thank you for your attention!


References II


