Backward Perturbation Analysis for Data Least Squares Problems

Xiao-Wen Chang

Computer Science, McGill University, Canada

Joint work with Gene Golub and Chris Paige

Workshop on TLS-EIV Modeling. July 24-27, 2006

Outline

- Introduction

Outline

- Introduction
- Backward perturbation analysis —
  a formula of a pseudo minimal backward error (PMBE)

Outline

- Introduction
- Backward perturbation analysis —
  a formula of a pseudo minimal backward error (PMBE)
- A lower bound on the PMBE

Outline

- Introduction
- Backward perturbation analysis —
  a formula of a pseudo minimal backward error (PMBE)
- A lower bound on the PMBE
- An asymptotic estimate for the PMBE

Outline

- Introduction
- Backward perturbation analysis —
  a formula of a pseudo minimal backward error (PMBE)
- A lower bound on the PMBE
- An asymptotic estimate for the PMBE
- Numerical examples

Outline

- Introduction
- Backward perturbation analysis —
  a formula of a pseudo minimal backward error (PMBE)
- A lower bound on the PMBE
- An asymptotic estimate for the PMBE
- Numerical examples
- Summary

Introduction

- Backward perturbation analysis (BPA):

  Problem: $f(d, x) = 0$, data d, solution x.

  Given an approximate solution y, BPA tries to solve

  $$\mu(y) \equiv \min \|\Delta d\|, \quad \text{s.t.} \quad f(d + \Delta d, y) = 0$$

  $\mu(y)$: minimal (normwise) backward error.

Introduction

- Backward perturbation analysis (BPA):

  Problem: $f(d, x) = 0$, data d, solution x.

  Given an approximate solution y, BPA tries to solve

  $$\mu(y) \equiv \min \|\Delta d\|, \quad \text{s.t.} \quad f(d + \Delta d, y) = 0$$

  $\mu(y)$: minimal (normwise) backward error.

- Uses of $\mu(y)$:

Introduction

- Backward perturbation analysis (BPA):

  Problem: $f(d, x) = 0$, data d, solution x.

  Given an approximate solution y, BPA tries to solve

  $$\mu(y) \equiv \min \|\Delta d\|, \quad \text{s.t.} \quad f(d + \Delta d, y) = 0$$

  $\mu(y)$: minimal (normwise) backward error.

- Uses of $\mu(y)$:
  - Check backward stability of an algorithm for a specific problem.

Introduction

- Backward perturbation analysis (BPA):

  Problem: $f(d, x) = 0$, data d, solution x.

  Given an approximate solution y, BPA tries to solve

  $$\mu(y) \equiv \min \|\Delta d\|, \quad \text{s.t.} \quad f(d + \Delta d, y) = 0$$

  $\mu(y)$: minimal (normwise) backward error.

- Uses of $\mu(y)$:
  - Check backward stability of an algorithm for a specific problem.
  - Design an effective stopping criterion for an iterative algorihm to solve the problem.

Introduction

Some work on BPA of linear systems

- Consistent systems (including structured ones) :
  Oettli & Prager (64), Rigal & Gaches (67),
  D. Higham & N. Higham (92), Varah (94),
  J.-G. Sun & Z. Sun (97), Sun (99) etc.

Some work on BPA of linear systems

- Consistent systems (including structured ones) :
  Oettli & Prager (64), Rigal & Gaches (67),
  D. Higham & N. Higham (92), Varah (94),
  J.-G. Sun & Z. Sun (97), Sun (99) etc.
- Ordinary overdetermined least squares:
  Stewart (77), Waldén, Karlson & Sun (95),
  Sun (96, 97), Gu (98), Grcar, Saunders & Su (04),
  Golub & Su (05)

Introduction

Some work on BPA of linear systems

- Consistent systems (including structured ones) :
  Oettli & Prager (64), Rigal & Gaches (67),
  D. Higham & N. Higham (92), Varah (94),
  J.-G. Sun & Z. Sun (97), Sun (99) etc.
- Ordinary overdetermined least squares:
  Stewart (77), Waldén, Karlson & Sun (95),
  Sun (96, 97), Gu (98), Grcar, Saunders & Su (04),
  Golub & Su (05)
- Contrained overdetermined least squares:
  Cox & Higham (99), Malyshev (01),
  Malyshev & Sadkane (02)

Introduction

Notation:

For any $a \in \mathbb{R}^n$: $\|a\| \equiv \|a\|_2 = (a^T a)^{1/2}$,

For any $A \in \mathbb{R}^{m \times n}$,

$A^\dagger$: the Moore-Penrose generalized inverse.

$\sigma_{\min}(A)$: the p-th largest singular value, $p = \min\{m, n\}$,

$\lambda_{\min}(A)$: the smallest eigenvalue when $m = n$.

Introduction

Date least squares (DLS):

Given $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

The DLS problem (Groat & Dowling (93)) :

$$\sigma_D = \min_{E,x} \|E\|_F \quad \text{subject to} \quad (A+E)x = b$$

If the solution $\hat{x} \neq 0$, the DLS problem is equivalent to

$$\sigma_D = \min_x \frac{\|b - Ax\|}{\|x\|}$$

Introduction — DLS

Paige & Strakoš (02):

- The DLS solution must exist and be unique if

$$\operatorname{rank}(A) = n, \quad b \not\perp \mathcal{U}_{\min}$$

where $\mathcal{U}_{\min}$ is the left singular vector subspace of A corresponding to $\sigma_{\min}(A)$.

Introduction — DLS

Paige & Strakoš (02):

- The DLS solution must exist and be unique if

$$\text{rank}(A) = n, \quad b \not\perp \mathcal{U}_{\min}$$

  where $\mathcal{U}_{\min}$ is the left singular vector subspace of A corresponding to $\sigma_{\min}(A)$.

- A vector $\hat{x}$ solves the DLS problem iff

$$f(A, \hat{x}) \equiv A^T(b - A\hat{x}) + \frac{\|b - A\hat{x}\|^2}{\|\hat{x}\|^2}\hat{x} = 0$$

$$\frac{\|b - A\hat{x}\|}{\|\hat{x}\|} < \sigma_{\min}(A)$$

Backward perturbation analysis for DLS

Suppose $y \neq 0$ is an approximate DLS solution.

The minimal backward error (MBE) problem :

$$\min_{\Delta A} \|\Delta A\|_F \quad \text{s.t.} \quad y = \arg\min_x \frac{\|b - (A + \Delta A)x\|}{\|x\|}$$

Aim: Derive a formula for the MBE $\min_{\Delta A} \|\Delta A\|_F$.

Set of backward perturbations

Define the set

$$\mathcal{C} \equiv \left\{ \Delta A : \; y = \arg\min_x \frac{\|b - (A + \Delta A)x\|}{\|x\|} \right\}$$

Therefore

$$\mathcal{C} = \left\{ \Delta A : f(A + \Delta A, y) = 0, \right.$$
$$\left. \frac{\|b - (A + \Delta A)y\|}{\|y\|} < \sigma_{\min}(A + \Delta A) \right\}$$

Set of backward perturbations

Difficult to give a general explicit expression for $\mathcal{C}$ due to the inequality.

Ignore the inequality and consider

$$\mathcal{C}_+ \equiv \{\Delta A : \ f(A + \Delta A, y) = 0\}$$

Note that $\mathcal{C} \subseteq \mathcal{C}_+$.

An expression for $\mathcal{C}_+$

Lemma: Any $\Delta A \in \mathcal{C}_+$ can be written as

$$\Delta A = vy^\dagger - vv^\dagger A + (I - vv^\dagger)(b - Ay)y^\dagger + (I - vv^\dagger)Z(I - yy^\dagger)$$

where $\quad b^T v = 0, \quad v \in \mathbb{R}^n, \quad Z \in \mathbb{R}^{m \times n}$.

Proof. Use a result of Chang, Paige, and Titley-Peloquin (the following talk by David Titley-Peloquin).

Pseudo minimal backward error (PMBE)

$$\mu_F(y) \equiv \min_{\Delta A \in \mathcal{C}_+} \|\Delta A\|_F = ?$$

Pseudo minimal backward error (PMBE)

$$\mu_F(y) \equiv \min_{\Delta A \in \mathcal{C}_+} \|\Delta A\|_F = ?$$

Theorem:

$$\boxed{\mu_F(y) = \left(\frac{\|r\|^2}{\|y\|^2} + \lambda_{\min}(M)\right)^{1/2} = \sigma_{\min}(N)}$$

where $r \equiv b - Ay$,

$$M \equiv (I - bb^\dagger)A(I - 2yy^\dagger)A^T(I - bb^\dagger)$$

$$N \equiv \left[(I - bb^\dagger)A(I - yy^\dagger), \frac{\|r\|}{\|y\|}(I - bb^\dagger)(I - rr^\dagger), \frac{\|r\|}{\|y\|}\frac{b}{\|b\|}\right]$$

Pseudo minimal backward error (PMBE)

$\mu_F(y)$ is reached by the optimal $\widehat{\Delta A}$:

$$\widehat{\Delta A} = \begin{cases} ry^\dagger & \text{if } \sigma_{\min}(N) = \|r\|/\|y\| \\ ry^\dagger - v_* v_*^T(A + 2ry^\dagger) & \text{if } \sigma_{\min}(N) < \|r\|/\|y\| \end{cases}$$

v_* is the unit eigenvector of M associated with $\lambda_{\min}(M)$, or the left singular vector of N associated with $\sigma_{\min}(N)$.

When is $\mu_F(y)$ the actual MBE?

Since $\mathcal{C} \subseteq \mathcal{C}_+$,

$$\mu_F(y) \equiv \min_{\Delta A \in \mathcal{C}_+} \|\Delta A\|_F \leq \min_{\Delta A \in \mathcal{C}} \|\Delta A\|_F$$

Since $\|b - A\hat{x}\|/\|\hat{x}\| < \sigma_{\min}(A)$,
when y is a “reasonable” approximation to $\hat{x}$,
it is likely that

$$\|b - (A + \widehat{\Delta A})y\|/\|y\| < \sigma_{\min}(A + \widehat{\Delta A})$$

i.e., it is likely that $\mu_F(y)$ is the MBE.

How to evaluate $\mu_F(y)$?

$$\mu_F(y) \equiv \min_{\Delta A \in \mathcal{C}_+} \|\Delta A\|_F = \sigma_{\min}(N)$$

$$N = \left[(I - bb^\dagger)A(I - yy^\dagger), \frac{\|r\|}{\|y\|}(I - bb^\dagger)(I - rr^\dagger), \frac{\|r\|}{\|y\|}\frac{b}{\|b\|} \right]$$

Computing the SVD of N to obtain $\mu_F(y)$ is expensive.

We want to estimate $\mu_F(y)$ in more efficient ways.

A lower bound on $\mu_2(y) \equiv \min_{\Delta A \in \mathcal{C}_+} \|\Delta A\|_2$

Theorem.

$$\boxed{\frac{1}{\sqrt{2}}\mu_F(y) \leq \mu_2(y) \leq \mu_F(y)}$$

A lower bound on $\mu_2(y) \equiv \min_{\Delta A \in \mathcal{C}_+} \|\Delta A\|_2$

Theorem.
$$\frac{1}{\sqrt{2}}\mu_F(y) \le \mu_2(y) \le \mu_F(y)$$

Theorem.
$$\mu_2^{\mathrm{lb}}(y) \equiv \frac{2\beta_0}{\beta_1 + \sqrt{4\beta_0 + \beta_1^2}} \le \mu_2(y)$$

where

$$\beta_0 \equiv \frac{\big\| \|y\|^2 A^T r + \|r\|^2 y \big\|}{2\|y\|^3}, \quad \beta_1 \equiv \frac{\|y\|^3 \|A\|_2 + 3\|y\|^2 \|r\|}{2\|y\|^3}$$

Note. $\mu_2^{\mathrm{lb}}(y)$ can be estimated in $O(mn)$ flops.

An asymptotic estimate for $\mu_F(y)$

Recall the optimal $\widehat{\Delta A}$ is a solution to

$$\mu_F(y) = \min \|\Delta A\|_F, \quad \text{s.t. } f(A + \Delta A, y) = 0 \qquad (*)$$

An asymptotic estimate for $\mu_F(y)$

Recall the optimal $\widehat{\Delta A}$ is a solution to

$$\mu_F(y) = \min \|\Delta A\|_F, \quad \text{s.t. } f(A + \Delta A, y) = 0 \qquad (*)$$

By Taylor's expansion,

$$f(A + \Delta A, y) \approx f(A, y) + \mathcal{J}_A f(A, y)\text{vec}(\Delta A),$$

$\mathcal{J}_A f(A, y) \in \mathbb{R}^{n \times mn}$ is the Jacobian of f w.r.t. vec(A).

An asymptotic estimate for $\mu_F(y)$

Recall the optimal $\widehat{\Delta A}$ is a solution to

$$\mu_F(y) = \min \|\Delta A\|_F, \quad \text{s.t. } f(A + \Delta A, y) = 0 \qquad (*)$$

By Taylor's expansion,

$$f(A + \Delta A, y) \approx f(A, y) + \mathcal{J}_A f(A, y)\text{vec}(\Delta A),$$

$\mathcal{J}_A f(A, y) \in \mathbb{R}^{n \times mn}$ is the Jacobian of f w.r.t. vec(A).

Instead of solving $(*)$, we solve

$$\tilde{\mu}_F(y) \equiv \min \|\Delta A\|_F,$$
$$\text{s.t. } f(A, y) + \mathcal{J}_A f(A, y)\text{vec}(\Delta A) = 0.$$

An asymptotic estimate for $\mu_F(y)$

$$\tilde{\mu}_F(y) \equiv \min \|\Delta A\|_F = \min \|\text{vec}(\Delta A)\|$$
$$\text{s.t.} \quad f(A, y) + \mathcal{J}_A f(A, y)\text{vec}(\Delta A) = 0$$

The solution $\widetilde{\Delta A}$ satisfies

$$\text{vec}(\widetilde{\Delta A}) = -[\mathcal{J}_A f(A, y)]^\dagger f(A, y)$$

Thus

$$\tilde{\mu}_F(y) \equiv \|\widetilde{\Delta A}\|_F = \|[\mathcal{J}_A f(A, y)]^\dagger f(A, y)\|$$

Theorem. $\tilde{\mu}_F(y)$ is an asymptotic estimate of $\mu_F(y)$, i.e.,

$$\boxed{\lim_{y \to \hat{x}} \frac{\tilde{\mu}_F(y)}{\mu_F(y)} = 1}$$

An asymptotic estimate for $\mu_F(y)$

Theorem. Let $B \equiv \begin{bmatrix} A + ry^\dagger \\ \frac{\|r\|}{\|y\|}(I - yy^\dagger) \end{bmatrix}, \quad c \equiv \begin{bmatrix} r \\ 0 \end{bmatrix} \in \mathbb{R}^{m+n}$.

Then

$$\boxed{\tilde{\mu}_F(y) = \frac{\|B(B^TB)^{-1}B^Tc\|}{\|y\|}}$$

Two methods for evaluating $\tilde{\mu}_F(y)$:

- QR factorization of B.
- Moment method.

Numerical tests

Goal: help to answer the following two questions.

- When is $\mu_F(y)$ the actual MBE
  or when will the following inequality hold ?

$$\|b - (A + \widehat{\Delta A})y\| / \|y\| < \sigma_{\min}(A + \widehat{\Delta A})$$

- Are $\mu_2^{\texttt{lb}}(y)$ and $\tilde{\mu}_F(y)$ good estimates of $\mu_F(y)$?

$$\mu_2^{\texttt{lb}}(y) \leq \mu_2(y) \leq \mu_F(y), \quad \lim_{y \to \hat{x}} \tilde{\mu}_F(y) / \mu_F(y) = 1$$

Numerical tests

Data:

- 100×40 "random" A, $\|A\|_F = 1$, $\kappa_2(A) = 10^5$.

Numerical tests

Data:

- 100×40 "random" A, $\|A\|_F = 1$, $\kappa_2(A) = 10^5$.
- $b = (A + E)x$, $x = [1, \ldots, 1]^T \in \mathbb{R}^{40}$,
  "random" E, $\|E\|_F \leq 10^{-6}$.

Numerical tests

Data:

- 100×40 "random" A, $\|A\|_F = 1$, $\kappa_2(A) = 10^5$.
- $b = (A + E)x$, $x = [1, \ldots, 1]^T \in \mathbb{R}^{40}$, "random" E, $\|E\|_F \le 10^{-6}$.
- "random" y, $\|y - \hat{x}\|/\|\hat{x}\| \le \delta_{\hat{x}}$, $\delta_{\hat{x}} = 0, 10^{-10}, 10^{-1}$.

Numerical tests

Data:

- 100×40 "random" A, $\|A\|_F = 1$, $\kappa_2(A) = 10^5$.
- $b = (A + E)x$, $x = [1, \ldots, 1]^T \in \mathbb{R}^{40}$,
  "random" E, $\|E\|_F \leq 10^{-6}$.
- "random" y, $\|y - \hat{x}\| / \|\hat{x}\| \leq \delta_{\hat{x}}$,
  $\delta_{\hat{x}} = 0, 10^{-10}, 10^{-1}$.
- For each $\delta_{\hat{x}}$, generated 1000 sample problems.

Numerical tests

The DLS solution :

$$\hat{x} = \frac{b^T b}{b^T A v_D} v_D$$

where v_D is the right singular vector corresponding to $\sigma_{\min}((I - bb^\dagger)A)$.

- We used the MATLAB built-in function `svd` to find v_D and then computed $\hat{x}$.
- We computed $\tilde{\mu}_F(y)$ by using the QR factn of B .

Numerical tests

Test results

Our tests showed the inequality

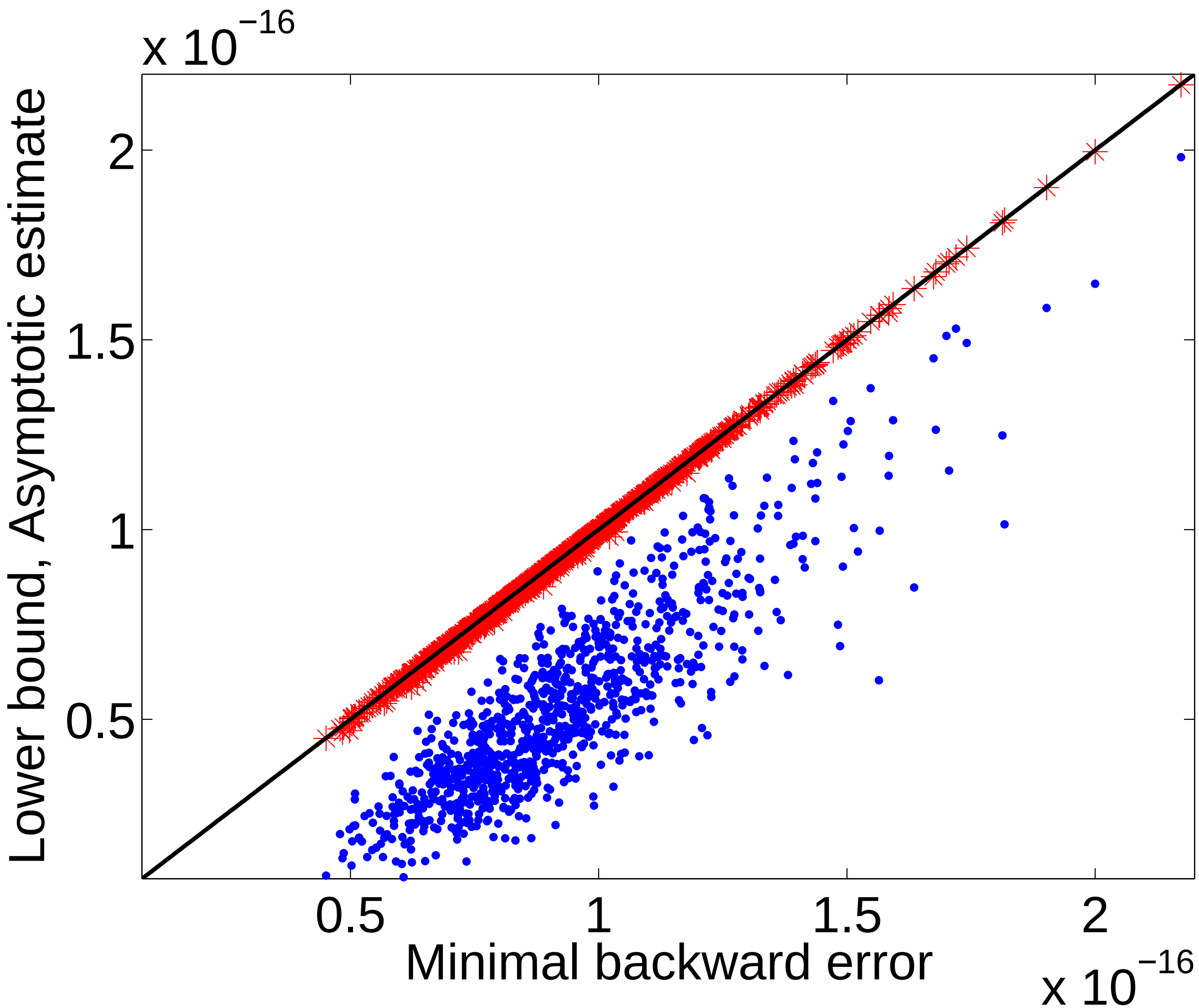
$$\|b - (A + \widehat{\Delta A})y\| / \|y\| < \sigma_{min}(A + \widehat{\Delta A})$$

holds for the test examples.

This confirms that $\mu_F(y)$ is likely to be the actual MBE if y is a reasonable approximation to $\hat{x}$.

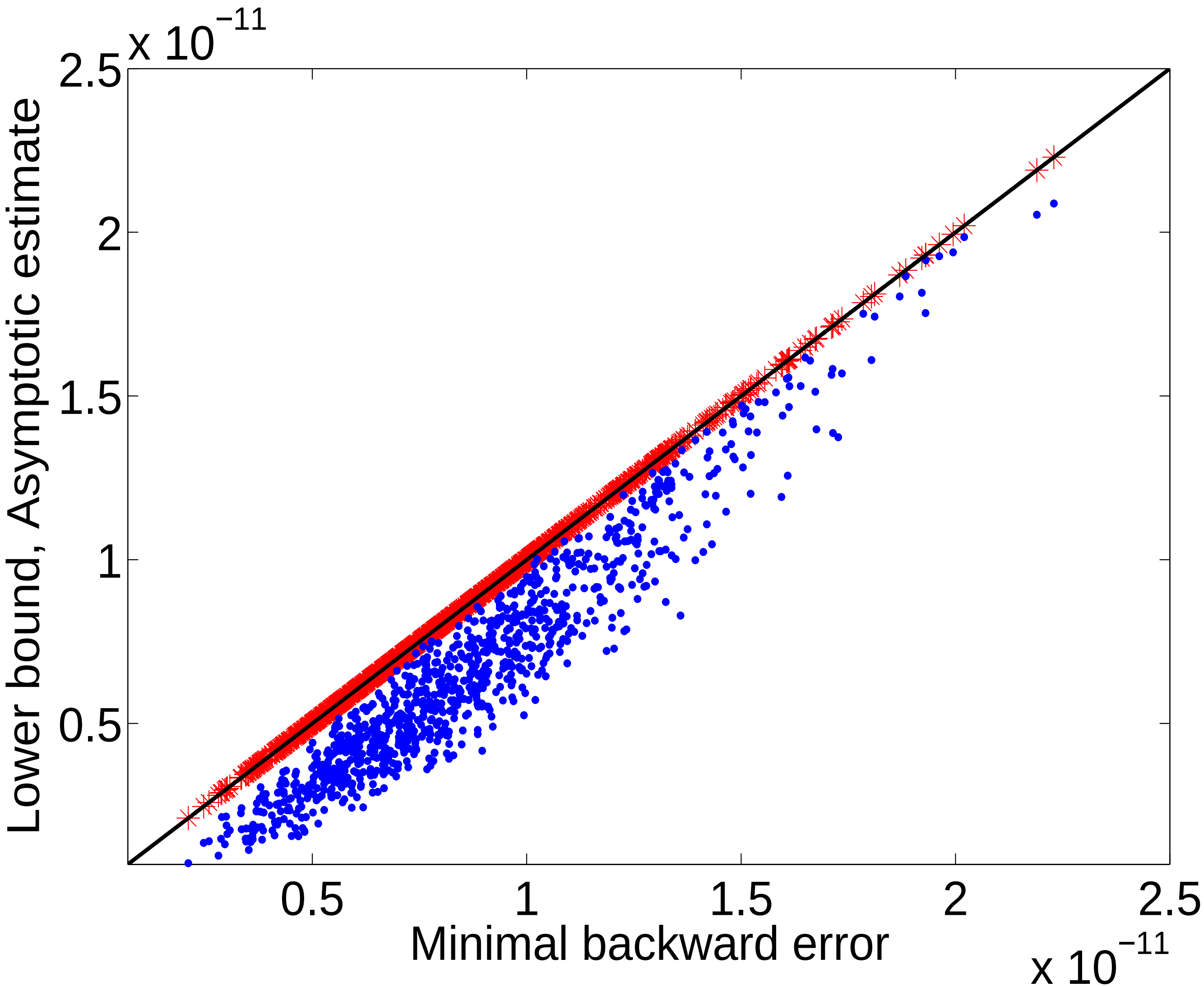
Numerical tests $(\delta_{\hat{x}} = 0)$

$\mu_F(y)$ vs $\mu_2^{\mathtt{lb}}(y)$: blue dots; $\mu_F(y)$ vs $\tilde{\mu}_F(y)$: red stars



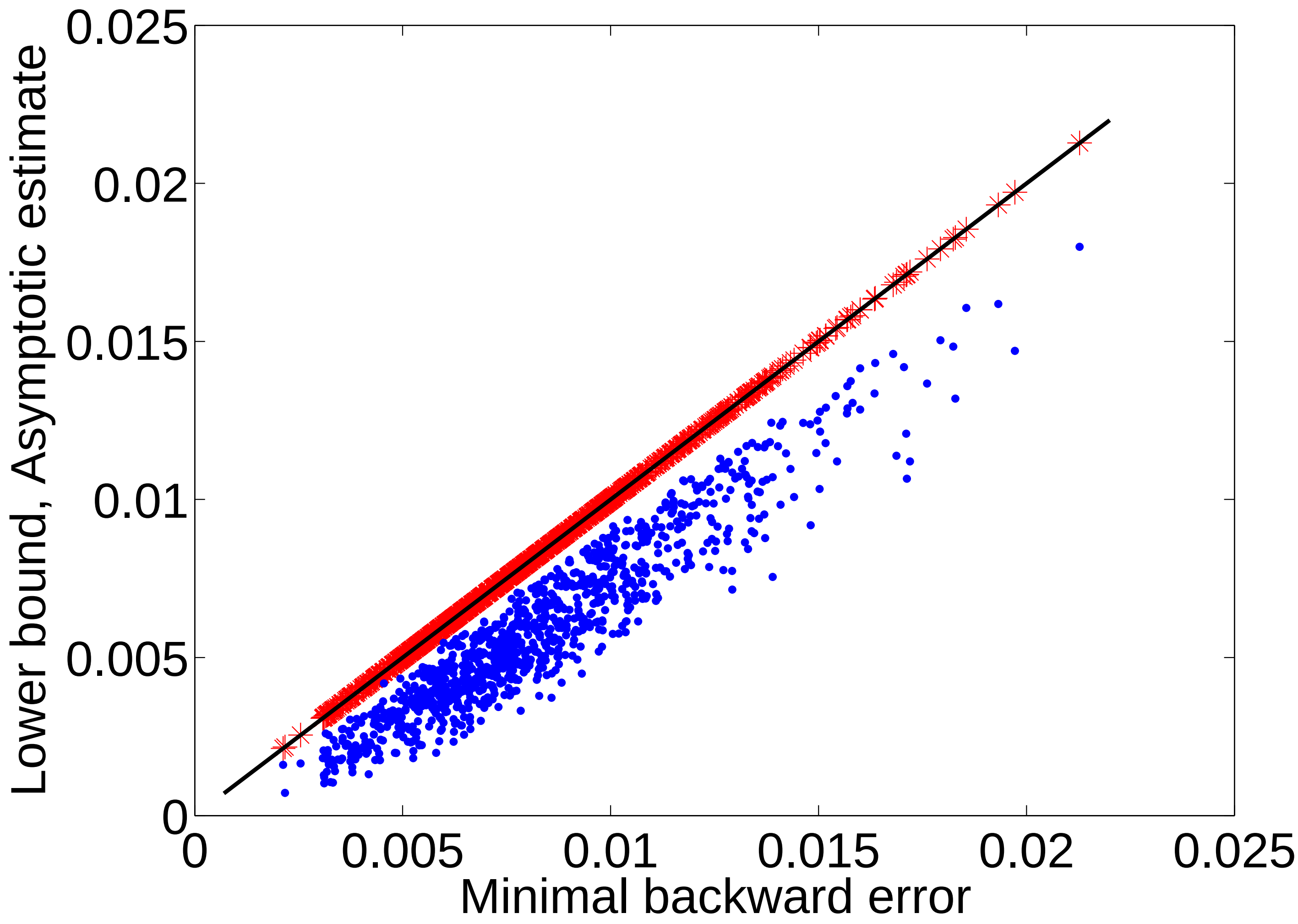
Numerical tests $(\delta_{\hat{x}} = 10^{-10})$

$\mu_F(y)$ vs $\mu_2^{\mathtt{lb}}(y)$: blue dots; $\mu_F(y)$ vs $\tilde{\mu}_F(y)$: red stars



Numerical tests $(\delta_{\hat{x}} = 10^{-1})$

$\mu_F(y)$ vs $\mu_2^{\mathtt{lb}}(y)$: blue dots; $\mu_F(y)$ vs $\tilde{\mu}_F(y)$: red stars



Summary

- Presented a formula for a pseudo minimal backward error $\mu_F(y)$ for DLS.
- If y is a reasonable approximation to the true solution, $\mu_F(y)$ is likely to be the true minimal backward error.
- Gave a lower bound on $\mu_F(y)$
  — a good approximation.
- Gave an asymptotic estimate for $\mu_F(y)$
  — an excellent approximation.