Solving large sparse $Ax = b$.

Stopping criteria, & backward stability of MGS-GMRES.

Chris Paige (McGill University);

Miroslav Rozložník & Zdeněk Strakoš

(Academy of Sciences of the Czech Republic).

.pdf & .ps files of this talk are available from:

http://www.cs.mcgill.ca/∼chris/pub/list.html

Background

The talk starts on slide 9, after this background.

This talk discusses material that the three of us have been interested in for many years.

About 1/2 of this talk was given by Chris Paige at an excellent conference to celebrate Bob Russell —

The International Conference on Adaptivity and Beyond: Computational Methods for Solving Differential Equations. Vancouver, August 3–6, 2005.

The response motivated us to distribute it widely, & to encourage writers to present the ideas in texts that applications-oriented people might turn to.

Re: “Backward Errors”

The backward error (BE) material for this appears in the literature. The backward error theory and history is given elegantly by Higham, 2nd Edn., 2002: §1.10; pp. 29–30; Chapter 7, in particular §7.1, 7.2 and 7.7; and also by Stewart & Sun, 1990, Section III/2.3; Meurant, 1999, Section 2.7; among others — but this is not easily accessible to the non-expert.

The original BE references are:
Prager & Oettli, Num. Math. 1964,
for componentwise analysis, which led to:
Rigal & Gaches, J. Assoc. Comput. Mach. 1967,
for normwise analysis (used here).

Re: “Stopping Criteria” (part 1)

The relation of BEs to stopping criteria for $Ax = b$ was described by Rigal & Gaches, 1967, §5, and is explained and thoroughly discussed in Higham, 2nd Edn., 2002, §17.5; and in “Templates”, Barrett *et al.*, 1995, §4.2.

These ideas have been used for constructing stopping criteria for years. For example, in Paige & Saunders, ACM Trans. Math. Software 1982, the backward error idea is used to derive a family of stopping criteria which quantify the levels of confidence in A and b, and which are implemented in the generally available software realization of the LSQR method.

Re: “Stopping Criteria” (part 1)

For other general considerations, methodology and applications see
Arioli, Duff & Ruiz, SIAM J. Mat. An. Appl. 1992;
Arioli, Demmel & Duff,
SIAM J. Matrix Anal. Appl. 1989;
Chatelin & Frayssé, 1996;
Kasenally & Simoncini, SIAM J. Numer. An. 1997.

For more recent sources see
Arioli, Noulard & Russo, Calcolo, 2001;
Arioli, Loghin & Wathen, Numer. Math. 2005;
Paige & Strakoš, SIAM J. Sci. Comput. 2002;
Strakoš & Liesen, ZAMM, 2005.

Re: “Stopping Criteria” (part 1)

These ideas are not widely used by the applications community, apparently because very little attention has been paid to stopping criteria in some major numerical linear algebra or iterative methods text books (e.g. Watkins, Demmel, Bau & Trefethen, Saad), or reference books (e.g. Golub & Van Loan). They are not spelt out in some other leading books on iterative methods, (e.g. Axelsson, Greenbaum, Meurant), but references are given in van der Vorst. Deuflhard & Hohmann, §2.4.3, *do* introduce the topic.

It would be healthy for users and also for our community if stopping criteria were considered to be fundamental parts of iterative computations, rather than as miscellaneous issues (if at all).

Re: “Stopping Criteria” (part 1)

This talk presents the backward error ideas in a simple form for use in stopping criteria for iterative methods.

It emphasizes that the normwise relative backward error (NRBE) is the one to use when you know your algorithm is backward stable. It should convince the user that unless there is a good reason to prefer some other stopping criterion, NRBE should be used in science and engineering calculations.

For clarity we will mainly use the 2-norm here, but other subordinate matrix norms are possible. See *e.g.* Higham, 2nd Edn. 2002, §7.1.

Re: “Stopping Criteria” (part 1)

An example when some other stopping criteria are preferable:

Conjugate gradient methods
for solving discretized elliptic self-adjoint PDEs,

see:
Arioli, Numer. Math. 2004;
Dahlquist, Eisenstat & Golub, J.Math.Anal.Appl.’72;
Dahlquist, Golub & Nash, 1978;
Hestenes & Stiefel, J. Res. Nat. Bur. St. 1952;
Meurant, Numerical Algorithms 1999;
Meurant & Strakoš, Acta Numerica 2006;
Strakoš & Tichý, ETNA 2002;
Strakoš & Tichý, BIT 2005.

Notation

Vectors $a,\ b,\ x, \ldots;$ iterates $x_1,\ x_2, \ldots$

Vector norm $\|x\|_2 = \sqrt{x^T x}$

Notation

Vectors $a,\ b,\ x, \ldots;$ iterates $x_1,\ x_2, \ldots$

Vector norm $\|x\|_2 = \sqrt{x^T x}$

Matrices nonsingular $A \in \mathbf{R}^{n \times n};\ B, \ldots$

Singular values $\sigma_1(A) \geq \ldots \geq \sigma_n(A) > 0$

Notation

Vectors $a,\ b,\ x,\ldots;$ iterates $x_1,\ x_2,\ldots$

Vector norm $\|x\|_2 = \sqrt{x^T x}$

Matrices nonsingular $A \in \mathbf{R}^{n\times n};\ B,\ldots$

Singular values $\sigma_1(A) \geq \ldots \geq \sigma_n(A) > 0$

Condition number

$$\kappa_2(A) = \sigma_1(A)/\sigma_n(A).$$

Notation

Vectors $a,\ b,\ x, \ldots;$ iterates $x_1,\ x_2, \ldots$

Vector norm $\|x\|_2 = \sqrt{x^T x}$

Matrices nonsingular $A \in \mathbf{R}^{n\times n};\ B, \ldots$

Singular values $\sigma_1(A) \geq \ldots \geq \sigma_n(A) > 0$

Condition number

$$\kappa_2(A) = \sigma_1(A)/\sigma_n(A).$$

Computer precision $\epsilon \approx 10^{-16}$ (IEEE double).

Matrix norms

The results hold for general subordinate matrix norms.
For clarity, we just consider:

Spectral norm: $$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2 = \sigma_1(A).$$

Frobenius: $$\|A\|_F^2 = \mathrm{trace}(A^T A) = \sum_{i=1}^{n} \sigma_i^2(A).$$

Matrix norms for rank-one matrices: if $B = cd^T$:

$$\|B\|_2 = \|cd^T\|_2 = \|c\|_2 \, \|d\|_2 = \|cd^T\|_F = \|B\|_F$$

Iterative methods – large $Ax = b$

Produce approximations to the solution x :

$$x_1,\ x_2,\ \ldots,\ x_k,\ \ldots$$

with residuals

$$\ldots,\ r_k = b - Ax_k,\ \ldots$$

Each iteration is expensive, hope for $\ll n$ steps.

Iterative methods – large $Ax = b$

Produce approximations to the solution x :

$$x_1,\ x_2,\ \ldots,\ x_k,\ \ldots$$

with residuals

$$\ldots,\ r_k = b - Ax_k,\ \ldots$$

Each iteration is expensive, hope for $\ll n$ steps.

When do we STOP?

Data accurate to $O(\epsilon)$ (relatively).

We will first treat the case of finding an x_k about as good as we can hope for the given data A and b, using computer precision ϵ, and a numerically stable algorithm.

Later we will consider inaccurate data.

Basic stopping criteria

- Test the residual norm, e.g. $\|r_k\|_2 \leq O(\epsilon)$

Basic stopping criteria

- Test the residual norm, e.g. $\|r_k\|_2 \leq O(\epsilon)$

  What if $\|b\|_2$ is huge ?, or tiny ?

Basic stopping criteria

- Test the residual norm, e.g. $\|r_k\|_2 \leq O(\epsilon)$

  What if $\|b\|_2$ is huge ?, or tiny ?

- Test the *relative* residual, e.g.

$$\frac{\|r_k\|_2}{\|b\|_2} = \frac{\|b - Ax_k\|_2}{\|b\|_2} \leq O(\epsilon)$$

Basic stopping criteria

- Test the residual norm, e.g. $\|r_k\|_2 \leq O(\epsilon)$

  What if $\|b\|_2$ is huge ?, or tiny ?

- Test the *relative* residual, e.g.

$$\frac{\|r_k\|_2}{\|b\|_2} = \frac{\|b - Ax_k\|_2}{\|b\|_2} \leq O(\epsilon) \quad ???$$

Basic stopping criteria

- Test the residual norm, e.g. $\|r_k\|_2 \leq O(\epsilon)$

  What if $\|b\|_2$ is huge ?, or tiny ?

- Test the *relative* residual, e.g.

$$\frac{\|r_k\|_2}{\|b\|_2} = \frac{\|b - Ax_k\|_2}{\|b\|_2} \leq O(\epsilon) \quad ???$$

- Test the Normwise Relative Backward Error, e.g.

$$\frac{\|r_k\|_2}{\|b\|_2 + \|A\|_2\|x_k\|_2} \leq O(\epsilon)$$

Basic stopping criteria

- Test the residual norm, e.g. $\|r_k\|_2 \leq O(\epsilon)$

  What if $\|b\|_2$ is huge ?, or tiny ?

- Test the *relative* residual, e.g.

$$\frac{\|r_k\|_2}{\|b\|_2} = \frac{\|b - Ax_k\|_2}{\|b\|_2} \leq O(\epsilon) \quad ???$$

- Test the Normwise Relative Backward Error, e.g.

$$\frac{\|r_k\|_2}{\|b\|_2 + \|A\|_2\|x_k\|_2} \leq O(\epsilon)$$

- Why use NRBE? $(\|\cdot\|_1,\ \|\cdot\|_\infty,\ \|\cdot\|_F,\ etc.)$

A <u>Backward</u> Stable (BS) Alg.

will eventually give the exact answer to a nearby problem, *e.g.* for the 2-norm case: an iterate x_k satisfying

$$(A + \delta A_k)\, x_k = b + \delta b_k,$$
$$\|\delta A_k\|_2 \le O(\epsilon)\|A\|_2, \qquad \|\delta b_k\|_2 \le O(\epsilon)\|b\|_2.$$

A Backward Stable (BS) Alg.

will eventually give the exact answer to a nearby problem, *e.g.* for the 2-norm case: an iterate x_k satisfying

$$(A + \delta A_k)\, x_k = b + \delta b_k,$$
$$\|\delta A_k\|_2 \leq O(\epsilon)\|A\|_2, \qquad \|\delta b_k\|_2 \leq O(\epsilon)\|b\|_2.$$

(J.H. Wilkinson 1950's, for n step algorithms, *e.g.*

Cholesky: $(A + \delta A)\, x_c = b,\ \|\delta A\|_2 \leq 12n^2\epsilon\|A\|_2$).

Such an x_k is called a backward stable solution.

δA_k and δb_k can be called backward errors.

A Backward Stable (BS) Alg.

will eventually give the exact answer to a nearby problem, *e.g.* for the 2-norm case: an iterate x_k satisfying

$$(A + \delta A_k)\, x_k = b + \delta b_k,$$
$$\|\delta A_k\|_2 \le O(\epsilon)\|A\|_2, \qquad \|\delta b_k\|_2 \le O(\epsilon)\|b\|_2.$$

Then the true residual r_k will satisfy

$$r_k = b - A x_k = \delta A_k\, x_k - \delta b_k,$$
$$\|r_k\|_2 \le O(\epsilon)(\|A\|_2\, \|x_k\|_2 + \|b\|_2).$$

A Backward Stable (BS) Alg.

will eventually give the exact answer to a nearby problem, *e.g.* for the 2-norm case: an iterate x_k satisfying

$$(A + \delta A_k)\, x_k = b + \delta b_k,$$
$$\|\delta A_k\|_2 \le O(\epsilon)\|A\|_2, \qquad \|\delta b_k\|_2 \le O(\epsilon)\|b\|_2.$$

Then the true residual r_k will satisfy

$$r_k = b - Ax_k = \delta A_k\, x_k - \delta b_k,$$
$$\|r_k\|_2 \le O(\epsilon)(\|A\|_2\, \|x_k\|_2 + \|b\|_2).$$

$$\& \quad \text{NRBE} \;=\; \frac{\|r_k\|_2}{\|b\|_2 + \|A\|_2\|x_k\|_2} \le O(\epsilon),$$

satisfying the simple 2-norm NRBE test.

If and only if?

Here a backward stable solution x_k satisfies

$$\text{NRBE} = \frac{\|b - Ax_k\|_2}{\|b\|_2 + \|A\|_2 \|x_k\|_2} \leq O(\epsilon). \quad (*)$$

If and only if?

Here a backward stable solution x_k satisfies

$$\text{NRBE} = \frac{\|b - Ax_k\|_2}{\|b\|_2 + \|A\|_2 \|x_k\|_2} \leq O(\epsilon). \quad (*)$$

But if an x_k satisfies this,
is it necessarily a backward stable solution?

If and only if?

Here a backward stable solution x_k satisfies

$$\text{NRBE} = \frac{\|b - Ax_k\|_2}{\|b\|_2 + \|A\|_2 \|x_k\|_2} \leq O(\epsilon). \quad (*)$$

But if an x_k satisfies this,
is it necessarily a backward stable solution?

YES. Rigal & Gaches, JACM 1967:
If x_k satisfies $(*)$ then there exist
backward errors δA_k & δb_k such that

$$(A + \delta A_k)\, x_k = b + \delta b_k,$$
$$\|\delta A_k\|_2 \leq O(\epsilon)\|A\|_2, \qquad \|\delta b_k\|_2 \leq O(\epsilon)\|b\|_2.$$

Proof: **Suppose 2-norm NRBE $\leq O(\epsilon)$.**

Take
$$\delta A_k = \left\{ \frac{\|A\|_2 \|x_k\|_2}{\|b\|_2 + \|A\|_2 \|x_k\|_2} \right\} \frac{r_k x_k^T}{\|x_k\|_2^2},$$

and
$$\delta b_k = -\left\{ \frac{\|b\|_2}{\|b\|_2 + \|A\|_2 \|x_k\|_2} \right\} r_k .$$

Proof: **Suppose 2-norm NRBE $\le O(\epsilon)$.**

Take $\delta A_k = \left\{ \dfrac{\|A\|_2 \|x_k\|_2}{\|b\|_2 + \|A\|_2 \|x_k\|_2} \right\} \dfrac{r_k x_k^T}{\|x_k\|_2^2}$,

and $\delta b_k = -\left\{ \dfrac{\|b\|_2}{\|b\|_2 + \|A\|_2 \|x_k\|_2} \right\} r_k$.

Then $\delta A_k \, x_k - \delta b_k = r_k = b - A x_k$,

so $(A + \delta A_k)\, x_k = b + \delta b_k$, &

$\|\delta A_k\|_2 \le O(\epsilon) \|A\|_2$, $\|\delta b_k\|_2 \le O(\epsilon) \|b\|_2$.

Q.E.D.

Summary (2-norm case)

Stopping criterion: STOP IF

$$\text{NRBE} \;=\; \frac{\|b - Ax_k\|_2}{\|b\|_2 + \|A\|_2\|x_k\|_2} \leq O(\epsilon)\,.$$

Summary (2-norm case)

Stopping criterion: STOP IF

$$\text{NRBE} = \frac{\|b - Ax_k\|_2}{\|b\|_2 + \|A\|_2\|x_k\|_2} \le O(\epsilon)\,.$$

- A backward stable solution will trigger this stopping criterion.

Summary (2-norm case)

Stopping criterion: STOP IF

$$\text{NRBE} = \frac{\|b - Ax_k\|_2}{\|b\|_2 + \|A\|_2\|x_k\|_2} \leq O(\epsilon)\,.$$

- A backward stable solution will trigger this stopping criterion.
- If this stopping criterion is triggered, we have a backward stable solution. Optimal! (Minimum number of steps for the chosen $O(\epsilon)$).

Summary (2-norm case)

Stopping criterion: STOP IF

$$\text{NRBE} = \frac{\|b - Ax_k\|_2}{\|b\|_2 + \|A\|_2 \|x_k\|_2} \leq O(\epsilon)\,.$$

- A backward stable solution will trigger this stopping criterion.
- If this stopping criterion is triggered, we have a backward stable solution. Optimal!
- So use this stopping criterion for backward stable algorithms (with data accurate to $O(\epsilon)$).

A BS iterative computation

A is FS1836 from the Matrix market:
183 x 183, 1069 entries, real unsymmetric.
Condition number $\kappa_2(A) \approx 2 \times 10^{11}$.

(Chemical kinetics problem from atmospheric pollution studies. Alan Curtis, AERE Harwell, 1983).

A BS iterative computation

A is FS1836 from the Matrix market:
183 x 183, 1069 entries, real unsymmetric.
Condition number $\kappa_2(A) \approx 2 \times 10^{11}$.

Solve two artificial test problems:

$$1: \qquad x = e = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad b := Ae\,,$$

$$2: \qquad b := e\,,$$

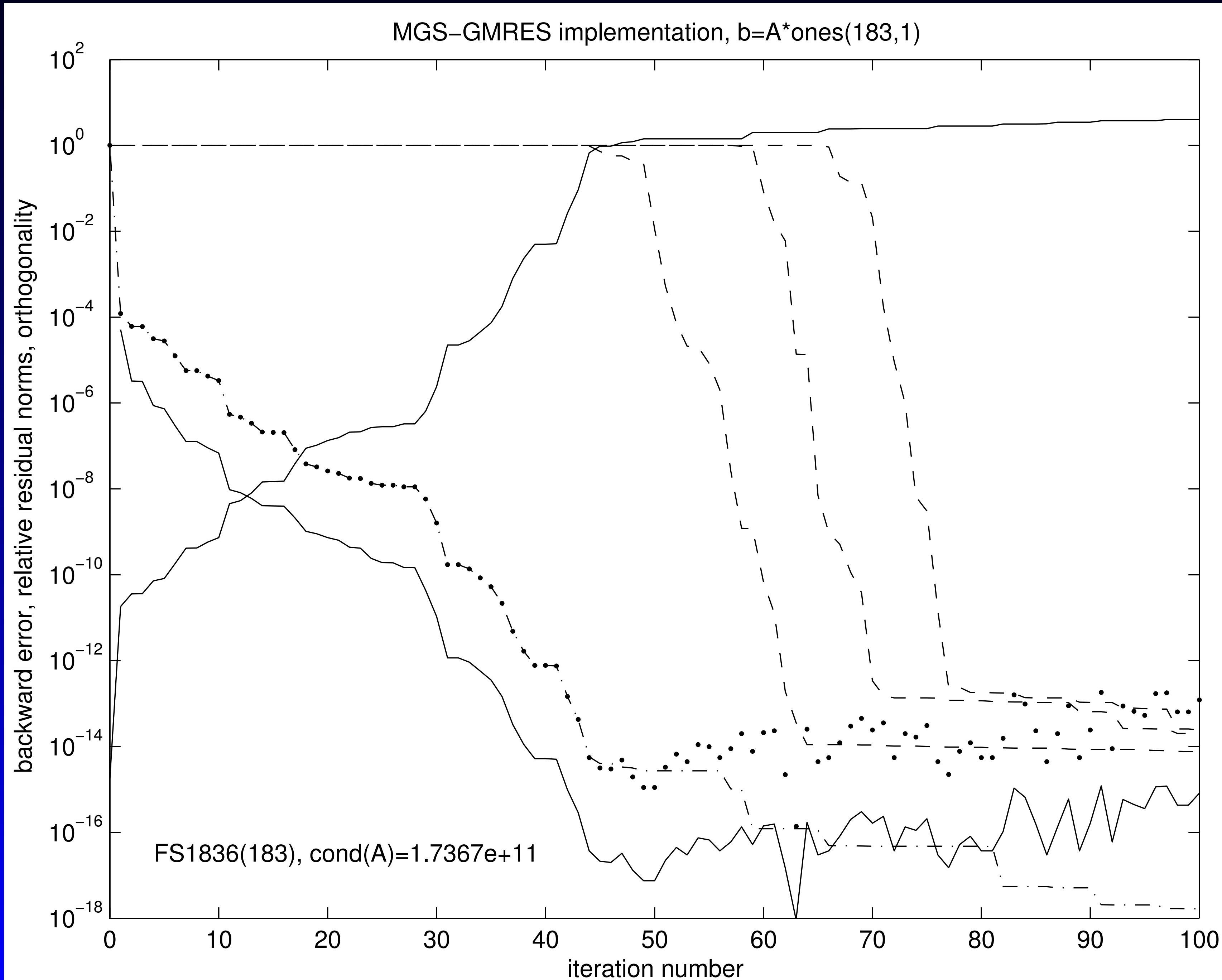
with the initial approximation $x_0 = 0$ (there must always be a good reason for using a nonzero x_0 !).

In the following two graphical slides,
concentrate on the two immediate $\searrow$ plots.

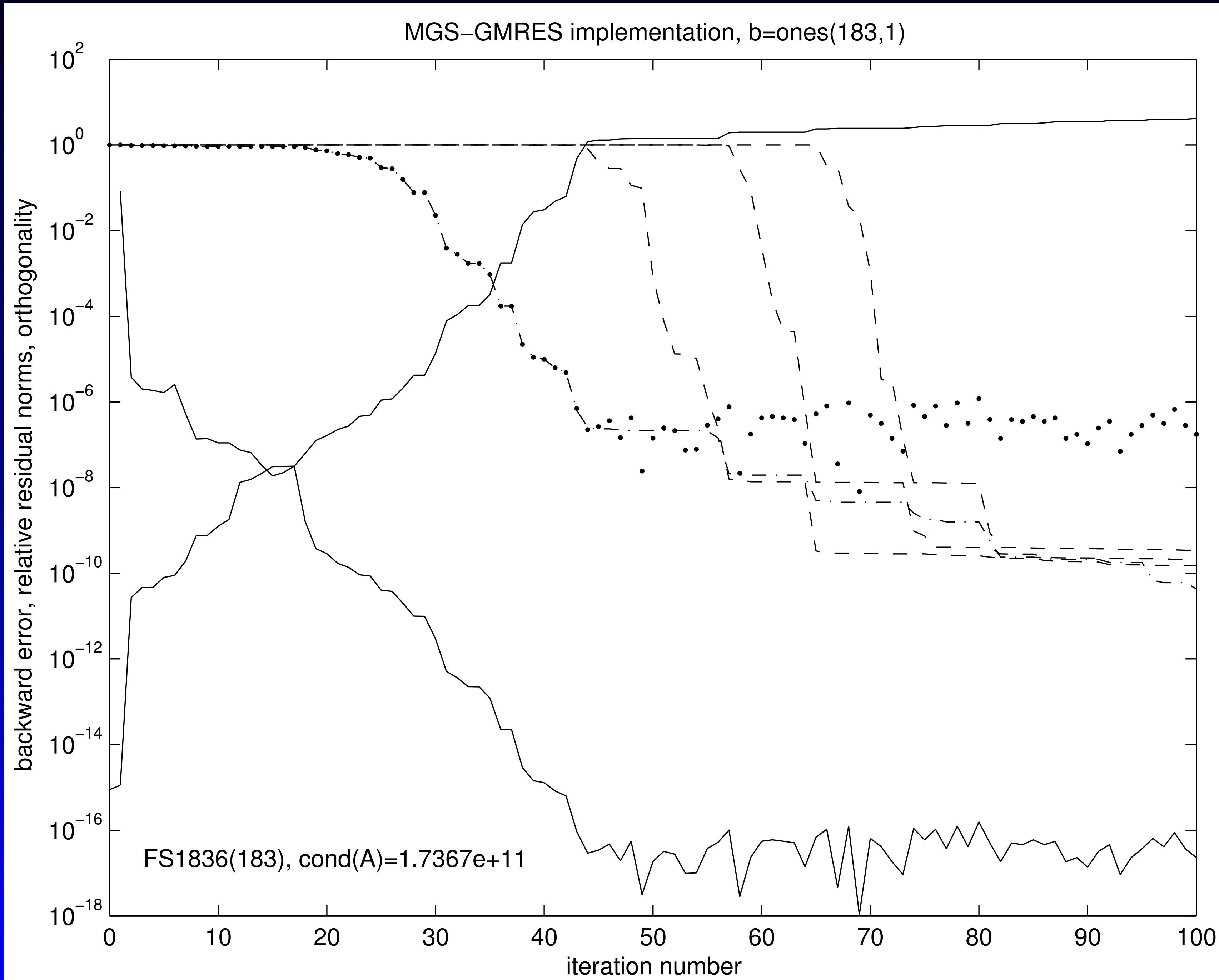
The $\nearrow$ plot denotes (loss of) orthogonality in MGS-GMRES.

The $- \ - \ - \ -$ plots denote singular values
of supposedly orthonormal matrices V_k,
& are of negligible interest to a general audience,
but crucial to the num. stability of MGS-GMRES.

1: $\|r_k\|_2/\|b\|_2$ ……. , $\|r_k\|_2/(\|b\|_2+\|A\|_2\|x_k\|_2)$ ——



2: $\|r_k\|_2/\|b\|_2$ ...., $\|r_k\|_2/(\|b\|_2+\|A\|_2\|x_k\|_2)$ ———



We see $\frac{\|r_k\|_2}{\|b\|_2}$ **can be VERY misleading.**

But the normwise relative backward error

$$\text{NRBE} = \frac{\|b - Ax_k\|_2}{\|b\|_2 + \|A\|_2 \|x_k\|_2}$$

is EXCELLENT, theoretically and computationally.

We see $\frac{\|r_k\|_2}{\|b\|_2}$ **can be VERY misleading.**

But the normwise relative backward error

$$\text{NRBE} = \frac{\|b - Ax_k\|_2}{\|b\|_2 + \|A\|_2\|x_k\|_2}$$

is EXCELLENT, theoretically and computationally.

A low at $k \approx 45$ for $n = 183$, then could *increase*!
A good stopping criterion is very important.

We see $\dfrac{\|r_k\|_2}{\|b\|_2}$ **can be VERY misleading.**

But the normwise relative backward error

$$\text{NRBE} \;=\; \frac{\|b - Ax_k\|_2}{\|b\|_2 + \|A\|_2\|x_k\|_2}$$

is EXCELLENT, theoretically and computationally.

A low at $k \approx 45$ for $n = 183$, then could *increase*!
A good stopping criterion is very important.

Similar ideas can apply to iterative methods
for <u>other</u> problems, *e.g.* NLE, SVD, EVP, . .

Inaccurate Data? Stop Early!

Usually $A \approx \tilde{A}, \; b \approx \tilde{b}$ where $\tilde{A}$ & $\tilde{b}$ are ideal unknowns.

Inaccurate Data? Stop Early!

Usually $A \approx \tilde{A}$, $b \approx \tilde{b}$ where $\tilde{A}$ & $\tilde{b}$ are ideal unknowns. Suppose we know α, β where

$$\begin{aligned} \tilde{A} &= A + \delta A, & \tilde{b} &= b + \delta b, \\ \|\delta A\|_2 &\le \alpha \|A\|_2, & \|\delta b\|_2 &\le \beta \|b\|_2. \end{aligned} \qquad (*)$$

Inaccurate Data? Stop Early!

Usually $A \approx \tilde{A}$, $b \approx \tilde{b}$ where $\tilde{A}$ & $\tilde{b}$ are ideal unknowns. Suppose we know α, β where

$$\tilde{A} = A + \delta A, \qquad \tilde{b} = b + \delta b,$$
$$\|\delta A\|_2 \le \alpha \|A\|_2, \quad \|\delta b\|_2 \le \beta \|b\|_2. \qquad (*)$$

Stopping criterion:

$$\frac{\|b - Ax_k\|_2}{\beta\|b\|_2 + \alpha\|A\|_2\|x_k\|_2} \le 1.$$

NOTE: Here “ ≤ 1 ”. Previously “ $\le O(\epsilon)$ ”.

Now the “accuracy measures” are α and β, and they appear in the denominator.

Inaccurate Data? Stop Early!

Usually $A \approx \tilde{A}$, $b \approx \tilde{b}$ where $\tilde{A}$ & $\tilde{b}$ are ideal unknowns. Suppose we know α, β where

$$\tilde{A} = A + \delta A, \qquad \tilde{b} = b + \delta b,$$
$$\|\delta A\|_2 \le \alpha \|A\|_2, \quad \|\delta b\|_2 \le \beta \|b\|_2. \qquad (*)$$

Justification for stopping criterion: If

$$\frac{\|b - Ax_k\|_2}{\beta \|b\|_2 + \alpha \|A\|_2 \|x_k\|_2} \le 1,$$

$\exists \quad \delta A_k, \ \delta b_k$ satisfying $(*)$, and

$$(A + \delta A_k)\, x_k = b + \delta b_k.$$

Inaccurate Data? Stop Early!

Usually $A \approx \tilde{A}$, $b \approx \tilde{b}$ where $\tilde{A}$ & $\tilde{b}$ are ideal unknowns. Suppose we know α, β where

$$\tilde{A} = A + \delta A, \qquad \tilde{b} = b + \delta b,$$
$$\|\delta A\|_2 \le \alpha \|A\|_2, \quad \|\delta b\|_2 \le \beta \|b\|_2. \qquad (*)$$

Justification for stopping criterion: If

$$\frac{\|b - Ax_k\|_2}{\beta\|b\|_2 + \alpha\|A\|_2\|x_k\|_2} \le 1,$$

$\exists$ δA_k, δb_k satisfying $(*)$, and

$$(A + \delta A_k)\, x_k = b + \delta b_k.$$

x_k the exact answer to a possible problem $\tilde{A} x_k = \tilde{b}$.

Proof: **If** **NRBE** $= \dfrac{\|b - Ax_k\|_2}{\beta\|b\|_2 + \alpha\|A\|_2\|x_k\|_2} \leq 1$,

$$\text{take} \quad \delta A_k \;=\; \left\{\frac{\alpha\|A\|_2\|x_k\|_2}{\beta\|b\|_2 + \alpha\|A\|_2\|x_k\|_2}\right\}\frac{r_k x_k^T}{\|x_k\|_2^2}\,,$$

$$\text{and} \quad \delta b_k \;=\; -\left\{\frac{\beta\|b\|_2}{\beta\|b\|_2 + \alpha\|A\|_2\|x_k\|_2}\right\} r_k\,.$$

Then $\quad \delta A_k\, x_k - \delta b_k = r_k = b - Ax_k\,,$

so $\quad (A + \delta A_k)\, x_k = b + \delta b_k\,, \qquad$ &

$$\|\delta A_k\|_2 \leq \alpha\|A\|_2\,, \qquad \|\delta b_k\|_2 \leq \beta\|b\|_2\,.$$

Q.E.D.

Computing $\|A\|_2$?

$\alpha = \beta = O(\epsilon)$ gives standard 2-norm BS criterion.

Write $\mu_k(\nu) \equiv \|b - Ax_k\|_2/(\beta\|b\|_2 + \alpha\nu\|x_k\|_2)$.

Eventually want $\nu = \|A\|_2, \quad \mu_k(\nu) \leq 1.$

Computing $\|A\|_2$?

$\alpha = \beta = O(\epsilon)$ gives standard 2-norm BS criterion.

Write $\mu_k(\nu) \equiv \|b - Ax_k\|_2/(\beta\|b\|_2 + \alpha\nu\|x_k\|_2)$. Many iterative methods produce a matrix B_k at step k such that to high accuracy $\|B_k\|_2 \nearrow \|A\|_2$ (almost always). In this case, although we do not always have $\|B_k\|_F \to \|A\|_F$, use the initial criterion $\mu_k(\|B_k\|_F) \leq 1$.

Computing $\|A\|_2$?

$\alpha = \beta = O(\epsilon)$ gives standard 2-norm BS criterion.

Write $\mu_k(\nu) \equiv \|b - Ax_k\|_2/(\beta\|b\|_2 + \alpha\nu\|x_k\|_2)$. Many iterative methods produce a matrix B_k at step k such that to high accuracy $\|B_k\|_2 \nearrow \|A\|_2$ (almost always). In this case, although we do not always have $\|B_k\|_F \to \|A\|_F$, use the initial criterion $\mu_k(\|B_k\|_F) \leq 1$. When that is met, estimate $\nu_k \approx \|B_k\|_2$ at this and further steps using some fast method, until $\mu_k(\nu_k) \leq 1$.

Computing $\|A\|_2$?

$\alpha = \beta = O(\epsilon)$ gives standard 2-norm BS criterion.

Write $\mu_k(\nu) \equiv \|b - Ax_k\|_2/(\beta\|b\|_2 + \alpha\nu\|x_k\|_2)$. Many iterative methods produce a matrix B_k at step k such that to high accuracy $\|B_k\|_2 \nearrow \|A\|_2$ (almost always). In this case, although we do not always have $\|B_k\|_F \to \|A\|_F$, use the initial criterion $\mu_k(\|B_k\|_F) \leq 1$. When that is met, estimate $\nu_k \approx \|B_k\|_2$ at this and further steps using some fast method, until $\mu_k(\nu_k) \leq 1$.

B_k is structured — for example:
GMRES: upper Hessenberg; LSQR: bidiagonal;

SYMMLQ & MINRES & CG: tridiagonal.

Direct use of $\|A\|_F$?

For the 2-norm case, the Rigal & Gaches minimal perturbations here are

$$\delta A_k = \left\{ \frac{\alpha \|A\|_2 \|x_k\|_2}{\beta \|b\|_2 + \alpha \|A\|_2 \|x_k\|_2} \right\} \frac{r_k x_k^T}{\|x_k\|_2^2},$$

$$\delta b_k = -\left\{ \frac{\beta \|b\|_2}{\beta \|b\|_2 + \alpha \|A\|_2 \|x_k\|_2} \right\} r_k\,.$$

δA_k is a rank one matrix, so $\|\delta A_k\|_2 = \|\delta A_k\|_F$.
And if $\|A\|_2$ is replaced by $\|A\|_F$,
they showed that the theory remains valid.
(They proved results for other norms too.)

Consequently:

A useful variant:

$$\eta_{\alpha,\beta}(x_k) \equiv \frac{\|b - Ax_k\|_2}{\beta\|b\|_2 + \alpha\|A\|_F\|x_k\|_2}$$

$$= \min_{\eta,\delta A,\delta b} \{\eta : (A + \delta A)\, x_k = b + \delta b,$$

$$\|\delta A\|_F \leq \eta\alpha\|A\|_F,\ \|\delta b\|_2 \leq \eta\beta\|b\|_2\}.$$

This gives the directly applicable NRBE’ criterion based on the Frobenius matrix norm.

MGS-GMRES for $Ax = b, \; A \in \mathbf{R}^{n \times n}$.

GMRES: "Generalized Minimum Residual" algorithm to solve $Ax = b, \; A \in \mathbf{R}^{n \times n}$, nonsing.

Y. SAAD & M. H. SCHULTZ, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.

Based on the algorithm by W. ARNOLDI, Quart. Appl. Math., 9 (1951), pp. 17–29.

MGS-GMRES for $Ax = b, \; A \in \mathbf{R}^{n \times n}$.

GMRES: "Generalized Minimum Residual" algorithm to solve $Ax = b, \; A \in \mathbf{R}^{n \times n}$, nonsing.

Y. SAAD & M. H. SCHULTZ, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.

Based on the algorithm by W. ARNOLDI, Quart. Appl. Math., 9 (1951), pp. 17–29.

The Modified Gram-Schmidt version (MGS-GMRES) is efficient, but looses orthogonality.

Some practitioners avoid it, or use reorthogonalization (*e.g.* Matlab). Is this necessary?

MGS-GMRES for $Ax = b, \; A \in \mathbb{R}^{n\times n}$.

Take $\varrho \equiv \|b\|_2, \;\; v_1 \equiv b/\varrho;$ generate columns of $V_{j+1} \equiv [v_1, \ldots, v_{j+1}]$ via the (MGS) Arnoldi alg.:

$$AV_j = V_{j+1}H_{j+1,j}, \qquad V_{j+1}^T V_{j+1} = I_{j+1}. \; *$$

Approximate solution $x_j \equiv V_j y_j$ has residual

$$\begin{aligned} r_j \;&\equiv\; b - Ax_j && = b - AV_j y_j \\ &=\; v_1\varrho - V_{j+1}H_{j+1,j}y_j && = V_{j+1}(e_1\varrho - H_{j+1,j}y_j). \end{aligned}$$

The *minimum* residual is found by taking

$$y_j \equiv \arg\min_y \{\|b - AV_j y\|_2 = \|e_1\varrho - H_{j+1,j}y\|_2\}. \; *$$

* DIFFICULTY: Computed $\bar{V}_{j+1}^T \bar{V}_{j+1} \neq I_{j+1}$.

Stability of MGS-GMRES

For some $k \leq n$, the MGS–GMRES method is backward stable for computing a solution $\bar{x}_k$ to

$$Ax = b, \quad A \in \mathbf{R}^{n \times n}, \quad \sigma_{min}(A) \gg n^2 \epsilon \|A\|_F;$$

Stability of MGS-GMRES

For some $k \leq n$, the MGS–GMRES method is backward stable for computing a solution $\bar{x}_k$ to

$$Ax = b, \quad A \in \mathbf{R}^{n \times n}, \quad \sigma_{min}(A) \gg n^2 \epsilon \|A\|_F;$$

as well as intermediate solutions $\bar{y}_j$ to the LLSPs:

$$\min_y \|b - A\bar{V}_j y\|_2, \qquad\qquad j = 1, \ldots, k,$$

where $\bar{x}_j \equiv fl(\bar{V}_j \bar{y}_j)$.

Stability of MGS-GMRES

For some $k \le n$, the MGS–GMRES method is backward stable for computing a solution $\bar{x}_k$ to

$$Ax = b, \quad A \in \mathbf{R}^{n \times n}, \quad \sigma_{min}(A) \gg n^2 \epsilon \|A\|_F;$$

as well as intermediate solutions $\bar{y}_j$ to the LLSPs:

$$\min_y \|b - A\bar{V}_j y\|_2, \qquad\qquad j = 1, \ldots, k,$$

where $\bar{x}_j \equiv fl(\bar{V}_j \bar{y}_j)$.

"Modified Gram-Schmidt (MGS), Least Squares, and backward stability of MGS-GMRES"
C. C. Paige, M. Rozložník, and Z. Strakoš,
Vol. 28, No. 1, 2006, pp. 264-284.

Stability of MGS-GMRES, ctd.

For some $k \leq n$, the MGS–GMRES method is backward stable for computing a solution $\bar{x}_k$ to

$$Ax = b, \quad A \in \mathbf{R}^{n\times n}, \quad \sigma_{min}(A) \gg n^2\epsilon\|A\|_F,$$

in that for some step $k \leq n$,
and some reasonable constant c,
the computed solution $\bar{x}_k$ satisfies

$$(A + \delta A_k)\,\bar{x}_k = b + \delta b_k,$$
$$\|\delta A_k\|_F \leq ckn\epsilon\|A\|_F, \quad \|\delta b_k\|_2 \leq ckn\epsilon\|b\|_2.$$

So we can use the F-norm NRBE' stopping criterion!

Conclusions. Solving $Ax = b$.

For a sufficiently nonsingular matrix, *e.g.*

$$\sigma_{min}(A) \gg n^2 \epsilon \|A\|_F,$$

(this is “rigorous”, but unnecessarily restrictive, a more practical requirement might be:

for large n, $\sigma_{min}(A) \geq 10\, n\, \epsilon \|A\|_F$)

Conclusions. Solving $Ax = b$.

For a sufficiently nonsingular matrix, *e.g.*

$$\sigma_{min}(A) \gg n^2 \epsilon \|A\|_F,$$

- we can happily use the efficient variant MGS-GMRES of the GMRES method,

Conclusions. Solving $Ax = b$.

For a sufficiently nonsingular matrix, *e.g.*

$$\sigma_{min}(A) \gg n^2 \epsilon \|A\|_F,$$

- we can happily use the efficient variant MGS-GMRES of the GMRES method,
- without reorthogonalization,

Conclusions. Solving $Ax = b$.

For a sufficiently nonsingular matrix, *e.g.*

$$\sigma_{min}(A) \gg n^2 \epsilon \|A\|_F,$$

- we can happily use the efficient variant MGS-GMRES of the GMRES method,
- without reorthogonalization,
- but with the optimal NRBE' stopping criterion,

Conclusions. Solving $Ax = b$.

For a sufficiently nonsingular matrix, *e.g.*

$$\sigma_{min}(A) \gg n^2 \epsilon \|A\|_F,$$

- we can happily use the efficient variant MGS-GMRES of the GMRES method,
- without reorthogonalization,
- but with the optimal NRBE' stopping criterion,
- since MGS-GMRES for $Ax = b$ is a backward stable iterative method.

Conclusions. Solving $Ax = b$.

For a sufficiently nonsingular matrix, *e.g.*

$$\sigma_{min}(A) \gg n^2 \epsilon \|A\|_F,$$

- we can happily use the efficient variant MGS-GMRES of the GMRES method,
- without reorthogonalization,
- but with the optimal NRBE' stopping criterion,
- since MGS-GMRES for $Ax = b$ is a backward stable iterative method.

Conclusions. Solving $Ax = b$.

For a sufficiently nonsingular matrix, *e.g.*

$$\sigma_{min}(A) \gg n^2 \epsilon \|A\|_F,$$

- we can happily use the efficient variant MGS-GMRES of the GMRES method,
- without reorthogonalization,
- but with the optimal NRBE' stopping criterion,
- since MGS-GMRES for $Ax = b$ is a backward stable iterative method.