Stopping criteria in iterative methods – a miscellaneous issue?

Zdeněk Strakoš*

and

Chris C. Paige*†

*Institute of Computer Science AS CR,
†McGill University, Montreal, Canada.

Fair comparison of the direct and iterative principle

- Iterative principle
- Direct principle
- Accuracy needed
Combination of the direct and iterative principle

- In order to reduce the disadvantages and profit from the advantages.

- Principal advantage of the iterative part is in stopping the computation at the desired accuracy level.

- It requires a meaningful stopping criterion. The errors of the model, discretization error and the computational error should be of the same order.

- Due to difficulties with the previous point this (potential) principal advantage is often presented as a disadvantage (a need for a stopping criteria . . . ).
Stopping criteria and rounding error analysis

The graph illustrates the norm of the error which is to be estimated against the iteration number. It compares the results of finite precision computation (solid line) and exact arithmetic computation (dashed line). The y-axis represents the norm of the error on a logarithmic scale, while the x-axis represents the iteration number. The graph shows how the error decreases as the number of iterations increases, with the finite precision computation line generally following the exact arithmetic computation line but with some deviation at higher iterations due to rounding errors.
Backward error approach
How good is an approximate solution?

Consider a linear algebraic system $Ax = b$, and a computed approximation $x_n$. Then

$$Ax_n = b - r_n, \quad r_n = b - Ax_n.$$

Thus, $-r_n$ represents the (unique) perturbation $\Delta b$ of the right hand side $b$ such that $x_n$ is the exact solution of the perturbed system.

A simple one-sided example of the perturbation theory – backward error approach.
How good is an approximate solution?

[Goldstine, Von Neumann - 47], [Turing], [Wilkinson - 63, 65]

**Perturbation theory:** \((A + \Delta A) \hat{x} = b + \Delta b\).

**Normwise relative backward error:**

Given \(\hat{x}\), construct \(\Delta A\), \(\Delta b\) such that both
\[\frac{\|\Delta A\|_2}{\|A\|_2} \text{ and } \frac{\|\Delta b\|_2}{\|b\|_2}\]
are minimal;

\[
\hat{x} \rightarrow \frac{\|\Delta A\|_2}{\|A\|_2} = \frac{\|\Delta b\|_2}{\|b\|_2} = \frac{\|b - A\hat{x}\|_2}{\|b\|_2 + \|A\|_2\|\hat{x}\|_2}.
\]
We ask and answer the question

“How close is the problem \((A + \Delta A) x_n = b + \Delta b\), which is solved by \(x_n\) \textbf{accurately}, to the original problem \(Ax = b\)?”

Perhaps this is what we need – the matrix \(A\) and the right hand side \(b\) are \textbf{inaccurate anyway}.

\textbf{Is the computed convergence curve} close to the \textbf{exact} one?
Relative residual stopping criteria

Difference between the normwise relative backward error and the relative residual norm:

Backward error restricted to the right hand side only is given by

\[ \frac{\| r_n \|_2}{\| b \|_2} \cdot \]

Moreover, for an unwise choice of \( x_0 \) this may differ greatly from the frequently used relative residual norm

\[ \frac{\| r_n \|_2}{\| r_0 \|_2} \cdot \]
Literature
On the backward error

The 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 non-experts.

The original BE references are:

Rigal & Gaches, J. Assoc. Comput. Mach. 1967, for normwise analysis (used here);

Oettli & Prager, Num. Math. 1964, for componentwise analysis.
Relation to stopping criteria

Explained and thoroughly discussed in

Higham, 2nd Edn., 2002, §17.5; and in
“Templates”, Barrett et al., 1995, Section 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.
Stopping criteria

General considerations, methodology and applications:

Arioli, Demmel & Duff,
Chatelin & Frayssé, 1996;

Arioli, Noulard & Russo, Calcolo, 2001;
Strakoš & Liesen, ZAMM, 2005.
Stopping criteria

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, or reference books.

It would be healthy for users and also for our community if stopping criteria were considered to be fundamental parts of iterative computations, and not treated among the miscellaneous issues (or not treated at all).
Stopping criteria based on error estimates

An example when some more sophisticated stopping criteria may be preferable:

(Preconditioned) Conjugate gradient method for solving discretized elliptic self-adjoint PDEs,

see:

Arioli, Numer. Math. 2004;
Meurant, Numerical Algorithms 1999;
Strakoš & Tichý, ETNA 2002;
Strakoš & Tichý, BIT 2005.
Meurant & Strakoš, Acta Numerica 2006;
Inaccurate data:

Normwise Backward Error Summary
Inaccurate data – stop early!

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

$$
\begin{align*}
\tilde{A} &= A + \delta A, \\
\tilde{b} &= b + \delta b,
\end{align*}
$$

\[
\|\delta A\|_2 \leq \alpha \|A\|_2, \quad \|\delta b\|_2 \leq \beta \|b\|_2.
\]

Justification for stopping criterion: If

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

$$
\exists \ \delta A_k, \ \delta b_k \text{ satisfying } (*) \text{, and }
$$

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

$x_k$ the exact answer to a possible problem $\tilde{A}x_k = \tilde{b}$. 
Rigal & Gaches, J. Assoc. Comp. Mach. 1967, showed: given $E$, $f$,

$$
\eta_{E,f}(x_k) = \frac{\|b - Ax_k\|_2}{\|f\|_2 + \|E\|_2\|x_k\|_2}
$$

$$
= \min_{\eta,\delta A,\delta b} \{ \eta : (A + \delta A) x_k = b + \delta b, \\
\|\delta A\|_2 \leq \eta\|E\|_2, \|\delta b\|_2 \leq \eta\|f\|_2 \}.
$$

Take $E = \alpha A$, $f = \beta b$, and the result follows.
Important refinement

\[ \eta'_{E,f}(x_k) = \frac{\|b - Ax_k\|_2}{\|f\|_2 + \|E\|_F\|x_k\|_2} \]

\[ = \min_{\eta', \delta A, \delta b} \{ \eta' : (A + \delta A) x_k = b + \delta b, \]

\[ \|\delta A\|_F \leq \eta' \|E\|_F, \|\delta b\|_2 \leq \eta' \|f\|_2 \} \]

gives the directly applicable NRBE’ criterion based on the Frobenius matrix norm.
More details can be found in

“Modified Gram-Schmidt (MGS), Least Squares, and backward stability of MGS-GMRES”
C. C. Paige, M. Rozložník, and Z. Strakoš,
Thank you for your kind attention!