

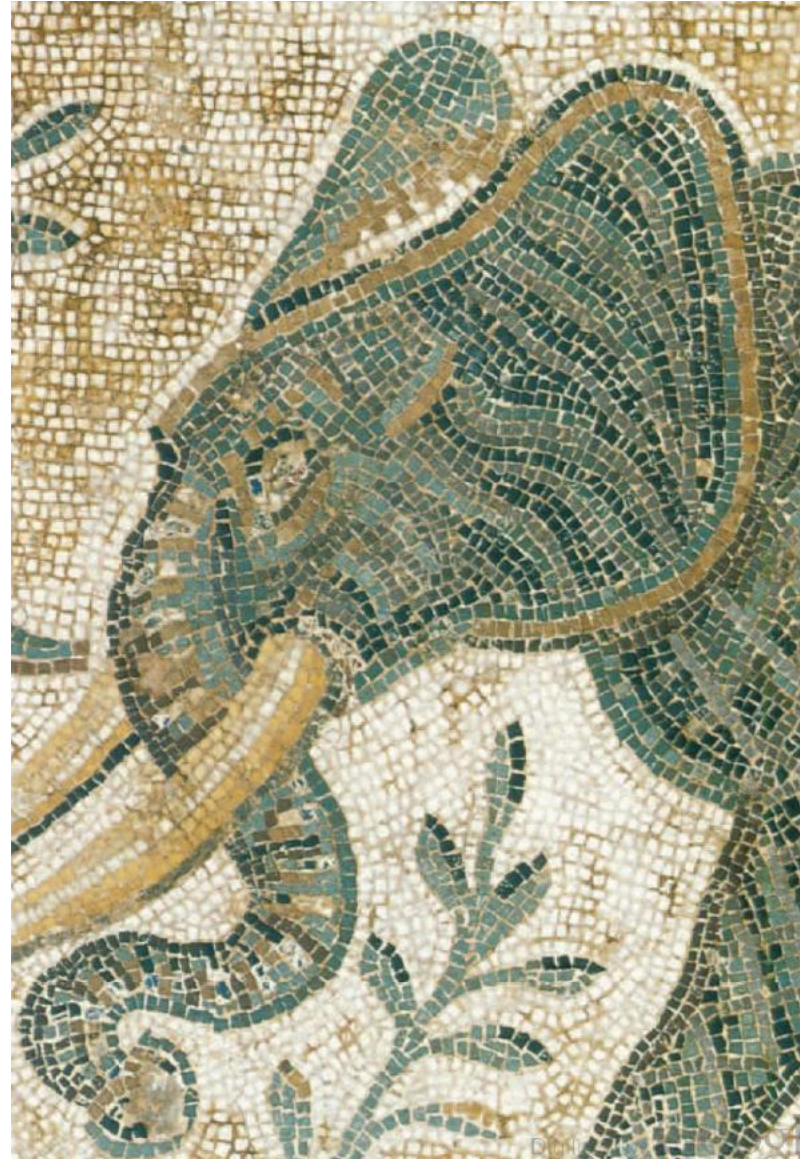
Piecing together the mosaic: Remarks on the mathematics of Zdeněk Strakoš

Jörg Liesen
Institute of Mathematics, TU Berlin
August 1, 2017

On mosaics

mosaic. A picture or pattern produced by arranging together small pieces of stone, tile, glass, etc.

(Oxford English Dictionary)



African elephant,
limestone, marble and glass,
Roman, 4th century,
Bardo Museum, Tunisia

From (A. B. Abed, ed.,
Stories in Stone, 2006)

June 1997 and June 2017



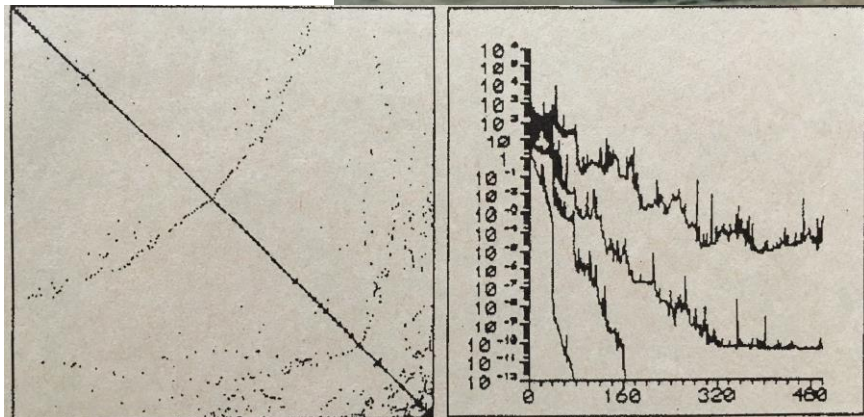
Put things into their proper context!



The Milovy Meeting in June 1997



9 skal	
HOTEL (0616) 66 42 82	
DEVĚT SKAL 66 42 84	
MILOVY 592 03 Sněžné na Moravě	
Jméno: Jörg Liesen	
HOTELOVÝ PRŮKAZ	
Pobyt	Číslo pokoje: AMS
od: do:	
Číslo stolu:	Směna:
Přejeme příjemný pobyt na Vysoké ***	



From: Daniel Szyld <impc97@uivt.cas.cz>
Date: Wed, 9 Jul 1997 16:20:04 -0400 (EDT)
Subject: Report on Czech-U.S. Workshop on Iterative Methods

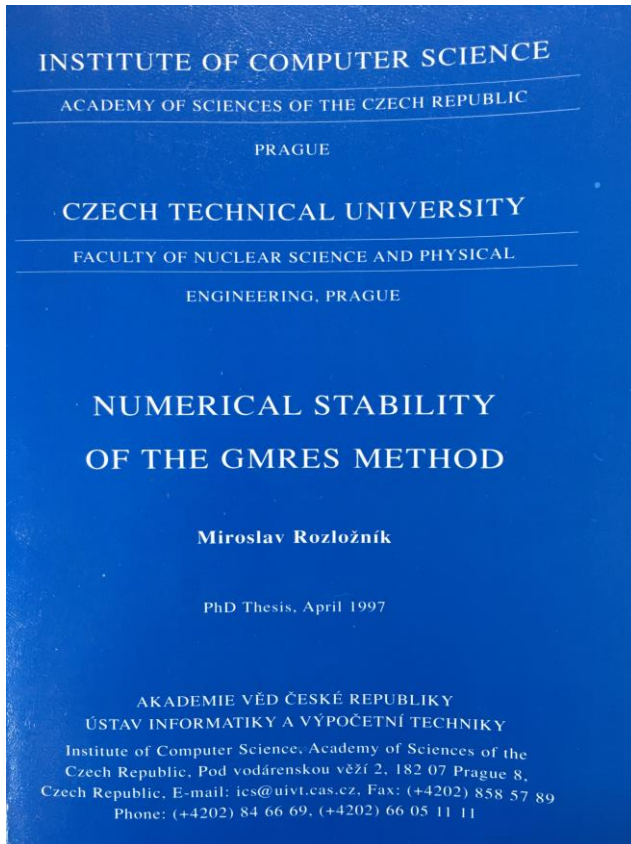
Report on the
Czech-U.S. Workshop on Iterative Methods and Parallel Computing (IMPC'97),
June 16-21, 1997, Milovy, Czech Republic.

Submitted by Daniel Szyld and Zdenek Strakos

Over 120 scientists from 20 countries met at the hotel Devet Skal in Milovy (literally "Nine Rocks") near the Bohemian-Moravian Highlands (Central part of the Czech Republic). About half the participants were from the Czech Republic and the United States. The rest came from almost every (Eastern and Western) European country as well as Turkey, and Israel. All enjoyed the moderate weather, the wonderful atmosphere of camaraderie, and the lake view (and a few even dared a swim in it).

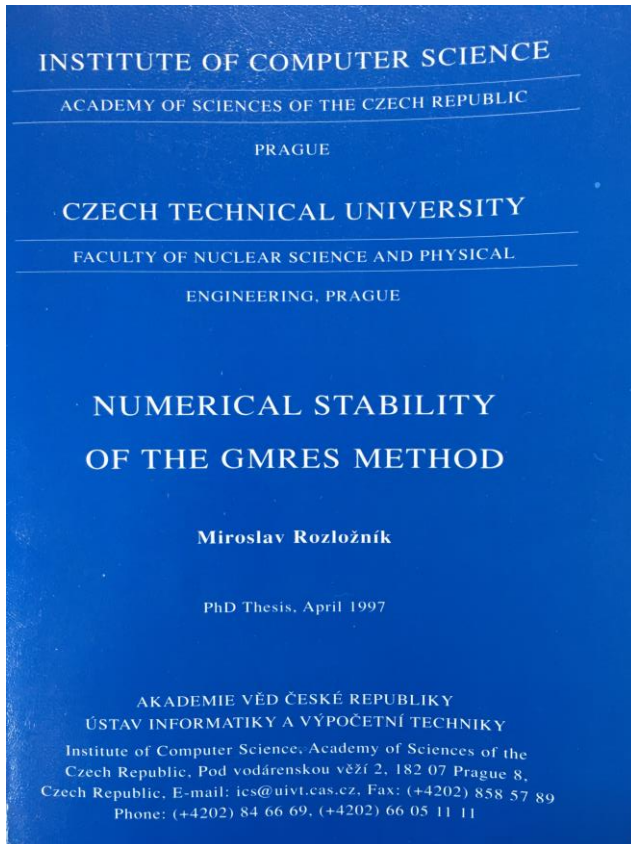
Miro's PhD thesis

- Miro's PhD thesis (completed in April 1997) served as a role model for my own thesis writing (completed in November 1998).
- Quoting Miro: “I also want to give special thanks to Zdeněk for setting high standards and then expecting them from me.”



Miro's PhD thesis

- Miro's PhD thesis (completed in April 1997) served as a role model for my own thesis writing (completed in November 1998).
- Quoting Miro: “I also want to give special thanks to Zdeněk for setting high standards and then expecting them from me.”



High standards and attention to detail

Theorem 3.3.4 gives the formula for the continued fraction (3.3.15) in terms of the orthogonal polynomials $P_n(x)$ and their derivatives.

The partial fraction expansion given in Theorem 3.3.5 enables the following development.

3.3. Orthogonal polynomials and continued fractions

with the analogous identity for $a_n R_{n-1}$. Setting $P_n = a_n P_{n-1}$ and $R_n = a_n R_{n-1}$, then shows

$$F_n = \frac{P_n}{P_{n-1}} = \frac{a_n R_n}{a_n P_{n-1}} = \frac{R_n}{P_{n-1}}$$

which proves (3.3.16) and (3.3.18). Now assume that (3.3.17) holds for $n-2$ and $n-1$. Substituting this assumption into (3.3.18) gives

$$R_n(x) = \int_{-1}^1 (\lambda - x) \omega_n(x) (1 - \psi_n(x)) d\omega(x) - \frac{d\omega(x)}{dx} (\psi_n(x) - \psi_{n-1}(x)) d\omega(x)$$

$$= \int_{-1}^1 \frac{\psi_n(x) - \psi_{n-1}(x)}{\lambda - x} d\omega(x) - \int_{-1}^1 \psi_{n-1}(x) d\omega(x)$$

where the last term is zero due to the orthogonality of $\psi_{n-1}(x)$ to the constant polynomial $\psi_0(x) = 1$, which finishes the proof.

very important paper
of the partial fraction expansion of $F_n(x)$. It is important to note that the expansion is in terms of the zeros of the orthogonal polynomials.

Theorem 3.3.5 Using the previous notation, the n th convergent $F_n(x)$ of the continued fraction corresponding to $\omega(x)$ can be decomposed into the partial fraction

$$F_n(x) = \frac{R_n(x)}{\psi_n(x)} = \sum_{j=1}^n \frac{a_j}{\lambda - \lambda_j^{(n)}} \quad (3.3.19)$$

where $\lambda_1^{(n)}, \dots, \lambda_n^{(n)}$ and a_1, \dots, a_n are the nodes and weights of the n -node Gauss-Chebyshev quadrature associated with $\omega(x)$.

Proof. The polynomials $R_n(x)$ and $\psi_n(x)$ are of degree $n-1$ and n , respectively. Thus, $R_n(x)$ can be uniquely expressed as a linear combination of the n linearly independent polynomials of degree $n-1$ given by

$$\frac{\psi_n(x)}{(\lambda - \lambda_j^{(n)})} \prod_{i \neq j} (\lambda - \lambda_i^{(n)})$$

This means that the partial fraction decomposition

$$F_n(x) = \frac{R_n(x)}{\psi_n(x)} = \sum_{j=1}^n \frac{\eta_j}{\lambda - \lambda_j^{(n)}} \quad \text{where } \eta_j = \frac{R_n(\lambda_j^{(n)})}{\psi_n'(\lambda_j^{(n)})} \quad j=1, \dots, n$$

exists, and by construction it is unique. Substituting for $R_n(x)$ from (3.3.17), and using the fact that $\psi_n(\lambda_j^{(n)}) = 0$, $j=1, \dots, n$,

$$\eta_j = \int_{-1}^1 \frac{\psi_n(\lambda_j^{(n)}) - \psi_n(x)}{\lambda_j^{(n)} - x} d\omega(x) = \int_{-1}^1 \frac{\psi_n(x)}{\psi_n'(\lambda_j^{(n)}) (\lambda_j^{(n)} - x)} d\omega(x)$$

Due year before Chebyshev
It was published in a remarkable paper by Christoffel (1859).

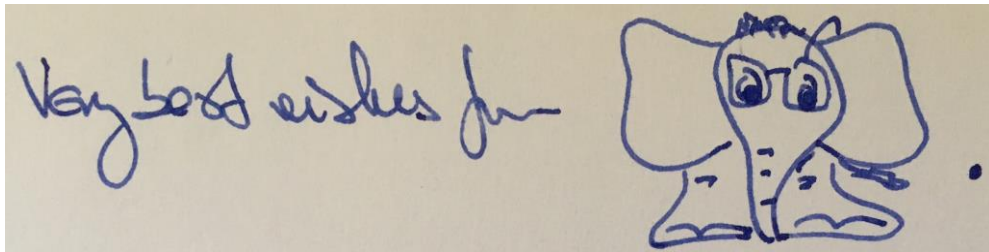
Chapter II, sections 7 and 8, and Section 5

THIS IS UNPLEASANT. IF YOU CHANGE THE WAY OF REFERENCING TO THE OLD PAPERS THEN PLEASE EXPLAIN WHY DO TO THINK THE PROPOSED CHANGE IS BETTER.

IT IS NEARLY IMPOSSIBLE THAT YOUR CHANGE IS NOT HARMFUL TO EVERYBODY. IT CREATES A MULTITUDE

SW

How to link the sentence with Shil'ya?



4.5. Necessary conditions

polynomial of A with respect to \mathcal{V} , and let n be a nonnegative integer, $s+2 \leq n$, $d_{\min}(A)$ be given. If

$$A^n v \in K_{s+1}(A, v) \quad \text{for all vectors } v \text{ of grade } d_{\min}(A), \quad (4.5.1)$$

then A is normal(t) for some $t \leq s$.

at
of the proof

PROOF. First suppose that A is invertible. We will extend the result to the general case at the end. We denote $\delta = d_{\min}(A)$ for convenience. Let $v \in \mathcal{V}$ be any vector of grade δ with respect to A . Since A is invertible, Av is of grade δ with respect to A . Consider any (fixed) scalar γ that is not an eigenvalue of A . Then the operator $A - \gamma I$ is invertible and the vector $w \equiv (A - \gamma I)v$ is of grade δ with respect to A . When (4.5.1) holds, there exist polynomials p, q , and r of degree at most s , which satisfy

$$A^n w = p_s(A)w, \quad A^s(Av) = q(A)(Av), \quad A^s v = r(A)v. \quad (4.5.2) \quad \text{(eq:pol1)}$$

Note that the polynomial p_s depends on γ , but the polynomials q and r do not. Using (4.5.2) and $w = (A - \gamma I)v$ we obtain

$$A^n w = p_s(A)w = p_s(A)(A - \gamma I)v = A p_s(A)v - \gamma p_s(A)v,$$

and

$$A^s(Av) = A^s(A - \gamma I)v = A^s(Av) - \gamma A^s v = A q(A)v - \gamma r(A)v.$$

Combining the last two identities gives

$$t_s(A)v = 0, \quad \text{where } t_s(\lambda) \equiv \lambda(p_s(\lambda) - q(\lambda)) - \gamma(p_s(\lambda) - r(\lambda)).$$

The polynomial t_s is of degree at most $s+1$. Since v is of grade $\delta \geq s+2$ and $t_s(A)v = 0$, we see that t_s must be the zero polynomial. A straightforward algebraic manipulation then gives

$$\gamma(q(\lambda) - r(\lambda)) = (\lambda - \gamma)(p_s(\lambda) - q(\lambda)), \quad (4.5.3) \quad \text{(eq:roots)}$$

giving
 $\lambda(p_s(\lambda) - q(\lambda)) = \lambda(p_s(\lambda) - q(\lambda))$
Since the polynomial on the left hand side is of degree at most s .

the operators AA^T and AA^T must be equal

In this way we proved that

$j=1, \dots, N$

and therefore

finite elements. The same reasoning can, however, be applied to a general case. The particular choice of the model problem and helps to make the point transparent. With the choice of the model problem where the first set of N nodes separates the "bubbles" over the individual elements. Clearly, the nodes can not describe the shape of the error.

234

5. Code of computations using Krylov subspace methods

No footnote too long!

Figure 5.6: Algebraic error $u_h - u_h^{(n)}$ (left) and total error $u - u_h^{(n)}$ (right) when CG is stopped with $\gamma = 0.02$. The vertical axes are scaled by 10^{-6} (left) and 10^{-4} (right); see also Remark 5.1.1. (fig:Ag0.02)

the solution ∇u is not piecewise constant
condition: we can even write \mathcal{E}_3 .

$$\|\nabla(u - u_h)\|^2 = \|\nabla u\|^2 - \|\nabla u_h\|^2$$

see (2.5.30) in Section 2.5.2. This suggests that the local distributions of the algebraic error and the discretization error can be very different. Indeed demonstrated by our experiment. Despite the comparable size of the values

$$\|\nabla(u - u_h)\|^2 \quad \text{and} \quad \|\nabla(u_h - u_h^{(n)})\|^2 = \|x - x_h\|_A^2$$

the shape of the total error for $\alpha = 3$ and $\gamma = 50$ is fully determined by its algebraic part. With decreasing γ the algebraic error gets smaller and it eventually becomes insignificant. Still, the convergence of the total error happens only after $\|x - x_h\|_A^2$ drops seven orders of magnitude below the squared energy norm of the discretisation error $\|\nabla(u - u_h)\|^2$. We can also observe that rather

however
the number of iterations of error reduction is not the same as the number of iterations of error reduction. The components of the error in the invariant subspaces corresponding to the large eigenvalues of A play a minor role. We will return to the oscillation pattern of the CG error in the proposed example in Section 5.9.4.

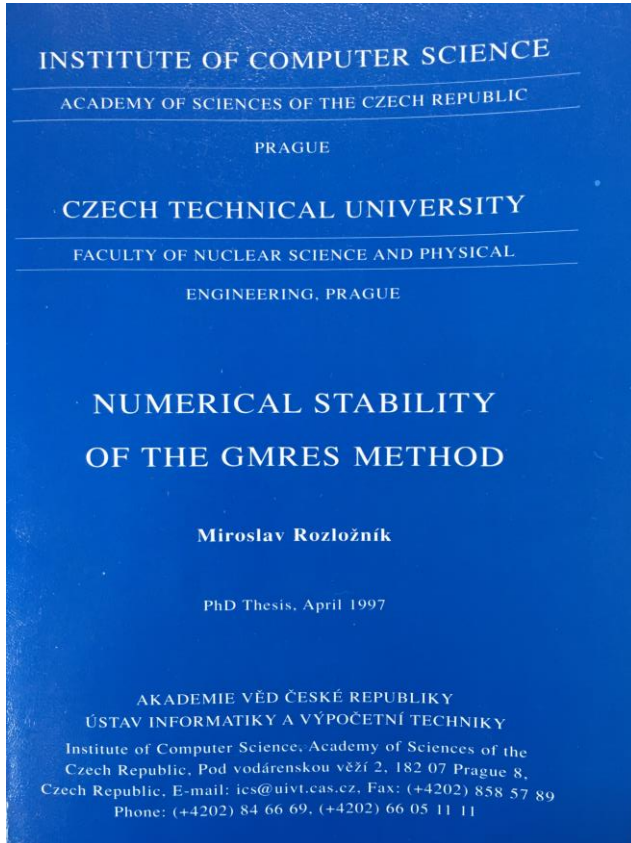
In conclusion, this simple example demonstrates:

(1) For the given discretisation there is some iteration count n for which the contribution of the discretisation and the algebraic errors to the total error are in balance. It is, however, crucial to evaluate such balance in

integrates over the individual elements

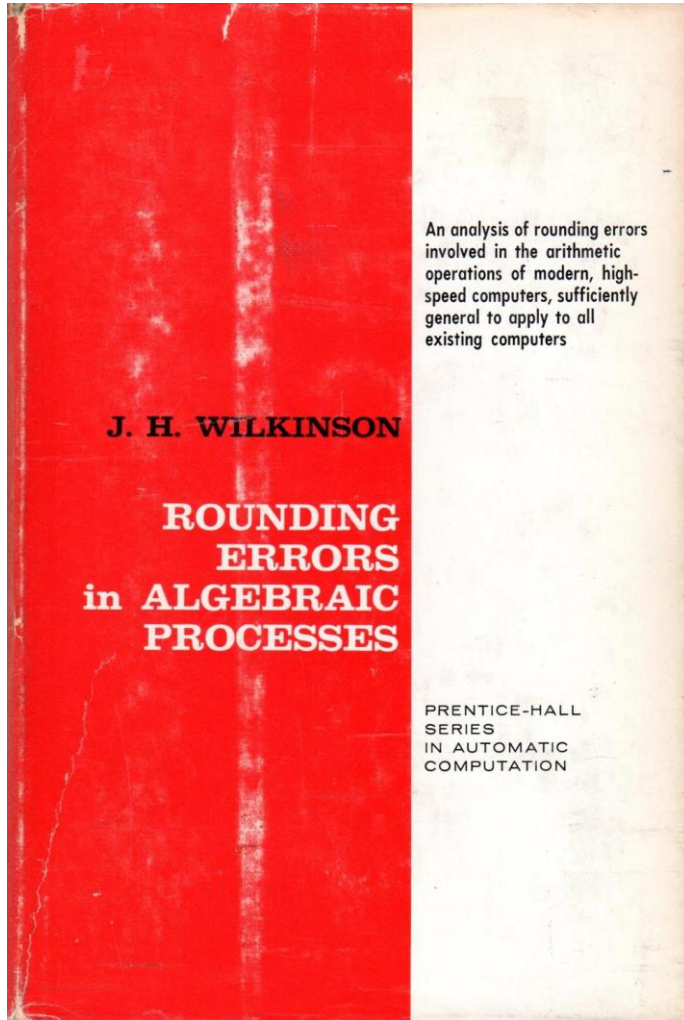
The norm $\|\nabla(u - u_h)\|$ again compares the error to the approximation of the gradient ∇u by the constant ∇u_h . Since the approximation of ∇u by ∇u_h is of quadrilateral (bilinear) character, the difference of two constants ∇u_h and ∇u_h on each element, there is no reason why the local distribution of the discretization and the algebraic error should in general be similar. This is

Miro's PhD thesis



- One of the main results:
Householder GMRES is backward stable.
- Important open question:
Is modified Gram-Schmidt GMRES backward stable as well?

Origins of backward error analysis



An analysis of rounding errors involved in the arithmetic operations of modern, high-speed computers, sufficiently general to apply to all existing computers

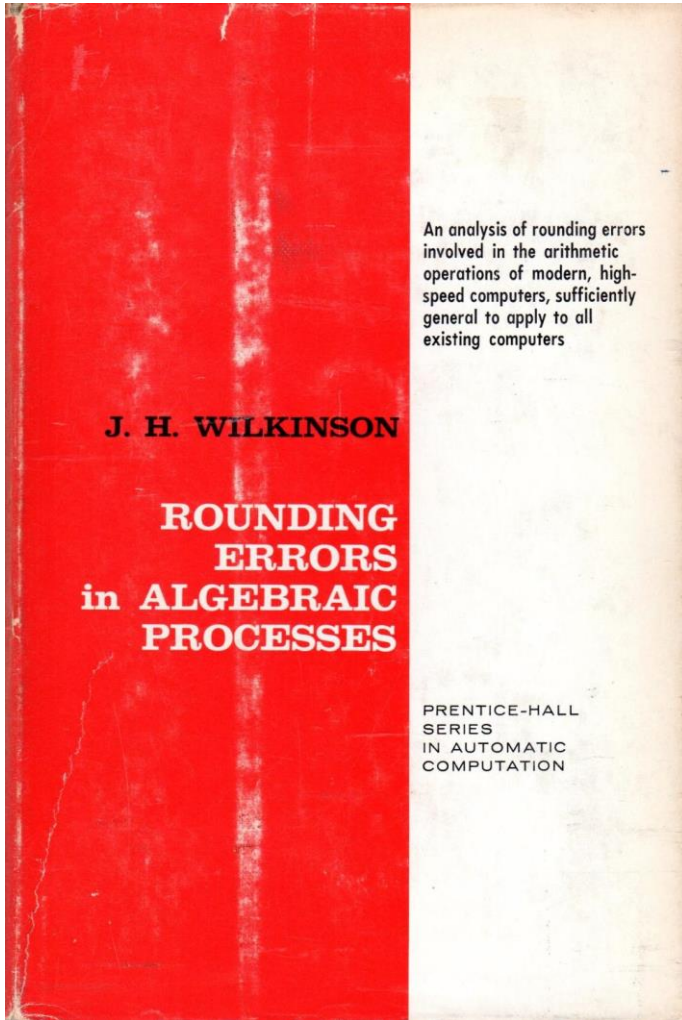
PRENTICE-HALL
SERIES
IN AUTOMATIC
COMPUTATION

(1963)

The idea of a *backward* error analysis, was to some extent implicit in the papers of von Neumann and Goldstine [18] and Turing [23]. It was described explicitly in Givens' paper [8] in the section on the calculation of the eigenvalues of a tri-diagonal matrix by the Sturm-sequence process. The error analysis in that paper seems not to have attracted as much attention as it deserved, possibly because it was not published in a readily available journal. Backward analysis has been used extensively by the author for the treatment of algebraic processes and has the advantage of suggesting automatically a convenient basis for comparison with the computed values.



Origins of backward error analysis



(1963)

The idea of a *backward* error analysis, was to some extent implicit in the papers of von Neumann and Goldstine [18] and Turing [23]. It was described explicitly in Givens' paper [8] in the section on the calculation of the eigenvalues of a tri-diagonal matrix by the Sturm-sequence process. The error analysis in that paper seems not to have attracted as much attention as it deserved, possibly because it was not published in a readily available journal. Backward analysis has been used extensively by the author for the treatment of algebraic processes and has the advantage of suggesting automatically a convenient basis for comparison with the computed values.



(1964)

NUMERICAL COMPUTATION OF THE CHARACTERISTIC VALUES OF A REAL SYMMETRIC MATRIX

Wallace Givens

3.2 INVERSION OF THE ERROR PROBLEM

Considered in broad terms, a method of computation should be regarded as an operator which is applied to a selected one of a class of permissible M -component data vectors and which yields an ordered set of N numbers, that is, a solution vector.⁽¹⁾ This single-valued mapping of a region of the "data space" onto some region of the "solution space" is what we mean by a numerical method.

(1954)

A different section of Givens paper from 1954

TABLE 6

Order of the Given Matrix	Time to Reduce to Jacobi Form $\approx \frac{1}{3}(n-2)(n-1)(4n+83)10^{-3}$ seconds
10	3 seconds
20	19 seconds
30	55 seconds
40*	2 minutes
100*	26 minutes
1000*	15.7 days (of 24 hr)

*For matrices of order greater than about 36, the internal memory will not suffice, and the estimates will require sharp upward revision. The reliability of present machines would probably not permit a matrix of order 1000 to be reduced in any reasonable time.

- In 1954 it took **1.356.480 seconds** to tridiagonalize a real symmetric matrix of order 1000.

A different section of Givens paper from 1954

TABLE 6

Order of the Given Matrix	Time to Reduce to Jacobi Form $\approx \frac{1}{3}(n-2)(n-1)(4n+83)10^{-3}$ seconds
10	3 seconds
20	19 seconds
30	55 seconds
40*	2 minutes
100*	26 minutes
1000*	15.7 days (of 24 hr)

*For matrices of order greater than about 36, the internal memory will not suffice, and the estimates will require sharp upward revision. The reliability of present machines would probably not permit a matrix of order 1000 to be reduced in any reasonable time.

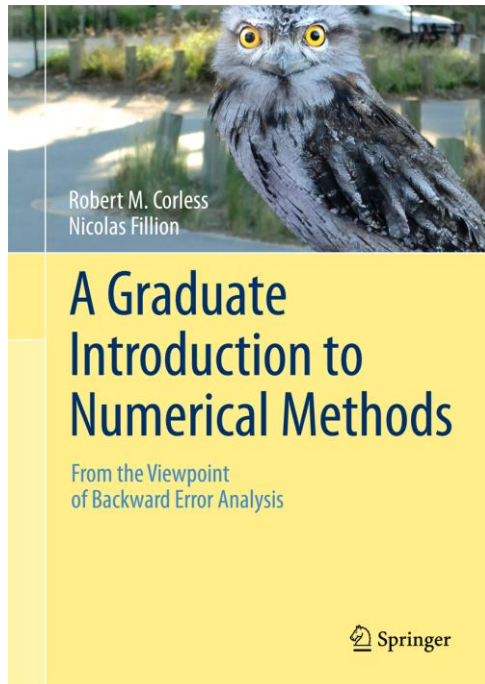
```
>> A=randn(1000);A=A+A';  
>> tic, [Anew,Q]=tridiagonalize(A); toc,  
Elapsed time is 28.552937 seconds.  
>> norm(eye(1000)-Q*Q')  
ans =  
    1.2264e-14  
>> norm(Anew-Q*A*Q')  
ans =  
    3.6638e-13
```

- In 1954 it took **1.356.480 seconds** to tridiagonalize a real symmetric matrix of order 1000.
- Today it takes **28 seconds** in MATLAB on my notebook with an ad hoc implementation based on Householder transformations.
- Computational devices (quickly) become outdated. **Mathematics is timeless.**

An analysis of rounding errors involved in the arithmetic operations of modern, high-speed computers, sufficiently general to apply to all existing computers

The ubiquitous nature of backward error analysis

- Backward error analysis was originally considered by Givens in the context of tridiagonalizing a real symmetric matrix.
- Through [synthetization](#), [abstraction](#) and [generalization](#) it has become a mature theory which is applied throughout numerical mathematics.
- This is the strength of mathematics (cf. Mehrmann, Schilders & Strakoš, ICIAM Newsletter, 01/2017).



(2014)

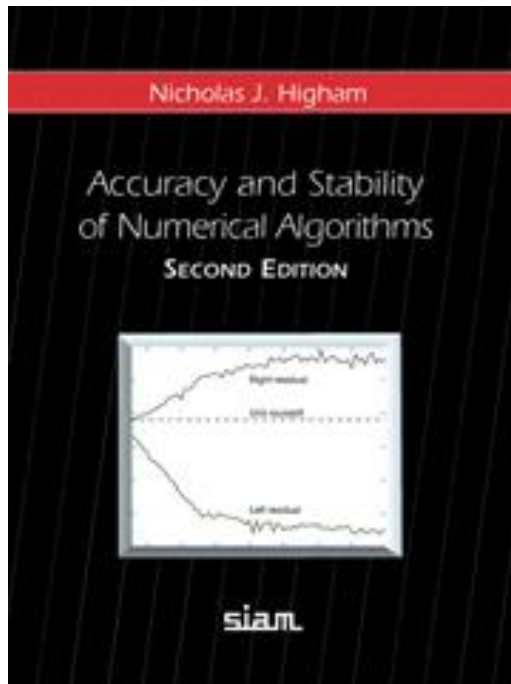
“A good numerical method gives you nearly the right solution to nearly the right problem.”

Backward error analysis for $Ax = b$

- Consider a linear algebraic system $Ax = b$ with $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$.
- If \hat{x} is an approximate solution, then $\|x - \hat{x}\|$ is the (absolute) **forward error**.
- Backward error idea: Which linear algebraic system is solved exactly by the approximation \hat{x} ?
- If $(A + \Delta A)\hat{x} = b + \Delta b$, then $\|\Delta A\|$ and $\|\Delta b\|$ are called the (absolute) **backward errors** in A and b .

Backward error analysis for $Ax = b$

- Consider a linear algebraic system $Ax = b$ with $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$.
- If \hat{x} is an approximate solution, then $\|x - \hat{x}\|$ is the (absolute) **forward error**.
- Backward error idea: Which linear algebraic system is solved exactly by the approximation \hat{x} ?
- If $(A + \Delta A)\hat{x} = b + \Delta b$, then $\|\Delta A\|$ and $\|\Delta b\|$ are called the (absolute) **backward errors** in A and b .



(2002)

“The data frequently contains uncertainties due to measurements, previous computations, or errors committed in storing machine numbers on the computer.

If the backward error is no larger than these uncertainties, then the computed solution may hardly be criticised – it may be the solution we are seeking – for all we know.”

Backward error analysis for $Ax = b$

- Consider a linear algebraic system $Ax = b$ with $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$.
- If \hat{x} is an approximate solution, then $\|x - \hat{x}\|$ is the (absolute) **forward error**.
- Backward error idea: Which linear algebraic system is solved exactly by the approximation \hat{x} ?
- If $(A + \Delta A)\hat{x} = b + \Delta b$, then $\|\Delta A\|$ and $\|\Delta b\|$ are called the (absolute) **backward errors** in A and b .

Theorem (Rigal & Gaches, 1967)

Let $r = b - A\hat{x}$ be the residual, then the **normwise relative backward error** of \hat{x} is given by

$$\beta(\hat{x}) = \min \{ \varepsilon : (A + \Delta A)\hat{x} = b + \Delta b \text{ with } \|\Delta A\|/\|A\| \leq \varepsilon \text{ and } \|\Delta b\|/\|b\| \leq \varepsilon \} = \frac{\|r\|}{\|b\| + \|A\|\|\hat{x}\|},$$

where the second equality is attained by explicitly known the perturbations.

- A numerical method is **normwise backward stable** when its computed approximation \hat{x} satisfies

$$\beta(\hat{x}) \approx \mathbf{u} \quad (= \text{machine precision}).$$

The GMRES method

- Consider $Ax = b$ with $A \in \mathbb{R}^{n \times n}$ nonsingular and $b \in \mathbb{R}^n$. Let $x_0 \in \mathbb{R}^n$ and $r_0 = b - Ax_0$.
- The **GMRES method** (Saad & Schultz, 1986) for $Ax = b$ generates a sequence x_k , $k = 1, 2, \dots$, with

$$x_k \in x_0 + \mathcal{K}_k(A, r_0) \quad \text{and} \quad \|r_k\| = \|b - Ax_k\| = \min_{z \in x_0 + \mathcal{K}_k(A, r_0)} \|b - Az\|,$$

where $\mathcal{K}_k(A, r_0) = \text{span}\{r_0, Ar_0, \dots, A^{k-1}r_0\}$ is the k th Krylov subspace generated by A and r_0 .

- GMRES is **one of the most important iterative methods** for general linear algebraic systems.

The GMRES method

- Consider $Ax = b$ with $A \in \mathbb{R}^{n \times n}$ nonsingular and $b \in \mathbb{R}^n$. Let $x_0 \in \mathbb{R}^n$ and $r_0 = b - Ax_0$.
- The **GMRES method** (Saad & Schultz, 1986) for $Ax = b$ generates a sequence x_k , $k = 1, 2, \dots$, with

$$x_k \in x_0 + \mathcal{K}_k(A, r_0) \quad \text{and} \quad \|r_k\| = \|b - Ax_k\| = \min_{z \in x_0 + \mathcal{K}_k(A, r_0)} \|b - Az\|,$$

where $\mathcal{K}_k(A, r_0) = \text{span}\{r_0, Ar_0, \dots, A^{k-1}r_0\}$ is the k th Krylov subspace generated by A and r_0 .

- GMRES is **one of the most important iterative methods** for general linear algebraic systems.
- The paper of Saad & Schultz has **10.500 citations** on Google Scholar (as of July 2017).
- Note: The 1992 paper of van der Vorst on Bi-CGStab, which in 2000 was named “the most-cited mathematics paper of the last decade” by the Institute of Scientific Information (ISI), currently has 5.300 citations on Google Scholar.

Backward stability of Householder GMRES

- GMRES uses Arnoldi's method for computing orthogonal bases of $\mathcal{K}_k(A, r_0)$.
- The [Arnoldi decomposition](#) in step k is given by $AV_k = V_{k+1}H_{k+1,k}$.
- Ideally, $V_k^T V_k = I_k$, but in finite precision computations we lose orthogonality.
- In the [Householder implementation](#) of Arnoldi we have $\|I_k - V_k^T V_k\| \approx \mathbf{u}$.
- One can show that the (final) computed x_n satisfies

$$\beta(x_n) = \frac{\|b - Ax_n\|}{\|b\| + \|A\|\|x_n\|} \approx \mathbf{u},$$

i.e., Householder GMRES is normwise backward stable (see Miro's PhD thesis).

- In practice the cheaper [modified Gram-Schmidt \(MGS\) implementation](#) is used.

GMRES and linear least squares

- The GMRES minimization problem is a **linear least squares problem**:

$$\|r_k\| = \min_{z \in x_0 + \mathcal{K}_k(A, r_0)} \|b - Az\| = \min_{y \in \mathbb{R}^k} \|r_0 - By\| \iff By \approx r_0,$$

where the columns of $B \in \mathbb{R}^{n \times k}$ form any basis of $A\mathcal{K}_k(A, r_0)$.

Theorem

If $\gamma > 0$ is a scaling parameter, then

$$\|r_k\| = \frac{\sigma_{\min}([r_0\gamma, B])}{\gamma} \prod_{j=1}^k \frac{\sigma_j([r_0\gamma, B])}{\sigma_j(B)}.$$

GMRES and linear least squares

- The GMRES minimization problem is a **linear least squares problem**:

$$\|r_k\| = \min_{z \in x_0 + \mathcal{K}_k(A, r_0)} \|b - Az\| = \min_{y \in \mathbb{R}^k} \|r_0 - By\| \iff By \approx r_0,$$

where the columns of $B \in \mathbb{R}^{n \times k}$ form any basis of $A\mathcal{K}_k(A, r_0)$.

Theorem

If $\gamma > 0$ is a scaling parameter, then

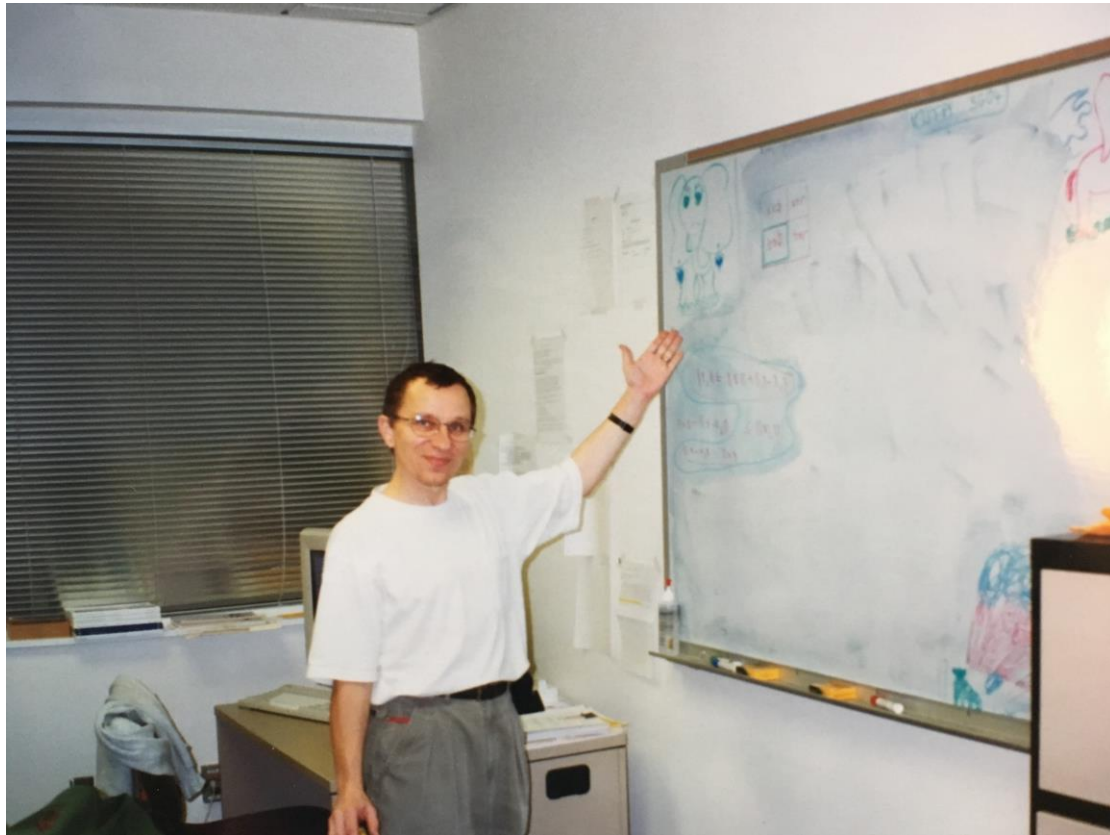
$$\|r_k\| = \frac{\sigma_{\min}([r_0\gamma, B])}{\gamma} \prod_{j=1}^k \frac{\sigma_j([r_0\gamma, B])}{\sigma_j(B)}.$$

- With $\gamma = \|r_0\|^{-1}$ and $B = QR$ we obtain

$$\sigma_{\min}([r_0/\|r_0\|, Q]) \leq \frac{\|r_k\|}{\|r_0\|} \leq \sqrt{2}\sigma_{\min}([r_0/\|r_0\|, Q]),$$

which is useful for analyzing certain implementations, e.g., Simpler GMRES (Walker & Zhou, 1994).

Written in Bielefeld, Urbana, Atlanta, Zürich & Prague



(Emory University, 1999)



(University of Illinois, 2001)

Least squares and total least squares

- The least squares (LS) distance for the problem $By \approx r_0$ is given by

$$\min_{r,y} \|r\| \quad \text{subject to} \quad By = r_0 - r.$$

- This is a backward error like interpretation:

The LS residual r is a minimal correction to the right hand side r_0 in order to make the corrected system compatible.

Least squares and total least squares

- The **least squares (LS) distance** for the problem $By \approx r_0$ is given by

$$\min_{r,y} \|r\| \quad \text{subject to} \quad By = r_0 - r.$$

- This is a **backward error like interpretation**:

The LS residual r is a minimal correction to the right hand side r_0 in order to make the corrected system compatible.

- If we allow **corrections to both B and r_0** , we get the total least squares problem.
- For each parameter $\gamma > 0$, the **scaled total least squares (STLS) distance** is given by

$$\min_{s,E,z} \|[s, E]\|_F \quad \text{subject to} \quad (B + E)z\gamma = r_0\gamma - s.$$

- The STLS distance is equal to $\sigma_{\min}([r_0\gamma, B])$, which is an important quantity in the GMRES context.

Laying the foundations

- In order to understand this situation, Zdeněk and Chris Paige completely reworked the foundations of LS and STLS problems:

Theorem

For $\gamma \rightarrow 0$ the STLS solution becomes the LS solution, and

$$\lim_{\gamma \rightarrow 0} \frac{\sigma_{\min}([r_0 \gamma, B])}{\gamma} = \|r\| \quad (\text{LS distance}).$$

Numer. Math. (2002) 91: 117–146
Digital Object Identifier (DOI) 10.1007/s002110100314

Numerische
Mathematik

Scaled total least squares fundamentals

Christopher C. Paige^{1,*}, Zdeněk Strakoš^{2,**}

Laying the foundations

- In order to understand this situation, Zdeněk and Chris Paige completely reworked the foundations of LS and STLS problems:

Theorem

For $\gamma \rightarrow 0$ the STLS solution becomes the LS solution, and

$$\lim_{\gamma \rightarrow 0} \frac{\sigma_{\min}([r_0\gamma, B])}{\gamma} = \|r\| \quad (\text{LS distance}).$$

Numer. Math. (2002) 91: 117–146
Digital Object Identifier (DOI) 10.1007/s002110100314

Numerische
Mathematik

Scaled total least squares fundamentals

Christopher C. Paige^{1,*}, Zdeněk Strakoš^{2,**}

Theorem

With $\theta(\gamma) = \sigma_{\min}([r_0\gamma, B])/\sigma_{\max}(B)$ and $\delta(\gamma) = \sigma_{\min}([r_0\gamma, B])/\sigma_{\min}(B)$,

$$\left(1 + \frac{\gamma^2 \|y\|^2}{1 - \theta(\gamma)^2}\right)^{1/2} \leq \frac{\|r\|\gamma}{\sigma_{\min}([r_0\gamma, B])} \leq \left(1 + \frac{\gamma^2 \|y\|^2}{1 - \delta(\gamma)^2}\right)^{1/2}.$$

Numer. Math. (2002) 91: 93–115
Digital Object Identifier (DOI) 10.1007/s002110100317

Numerische
Mathematik

**Bounds for the least squares distance
using scaled total least squares**

Christopher C. Paige^{1,*}, Zdeněk Strakoš^{2,**}

Backward error and loss of orthogonality in MGS-GMRES

- A classical result of Björk (1967) implies that in the finite precision MGS-Arnoldi algorithm we have

$$\|I - V_k^T V_k\|_F \approx \kappa([r_0 \gamma, AV_{k-1}]) \mathbf{u}.$$

Theorem

With certain (optimal) scalings $\gamma > 0$ and diagonal $D_{k-1} > 0$, and when $\delta_k \ll 1$,

$$\frac{1}{\sqrt{2}} \leq \kappa([r_0 \gamma, AV_{k-1} D_{k-1}]) \frac{\|b - Ax_k\|}{\|b\| + \|A\| \|x_k\|} \leq 2.$$

SIAM J. SCI. COMPUT.
Vol. 23, No. 6, pp. 1898–1923

© 2002 Society for Industrial and Applied Mathematics

RESIDUAL AND BACKWARD ERROR BOUNDS IN MINIMUM
RESIDUAL KRYLOV SUBSPACE METHODS*

CHRISTOPHER C. PAIGE† AND ZDENEK STRAKOŠ‡

- Thus, the product of the loss of orthogonality and the normwise relative backward error satisfies

$$\|I - V_k^T V_k\|_F \frac{\|b - Ax_k\|}{\|b\| + \|A\| \|x_k\|} \approx \mathcal{O}(1) \mathbf{u}.$$

All this holds in finite precision MGS-GMRES

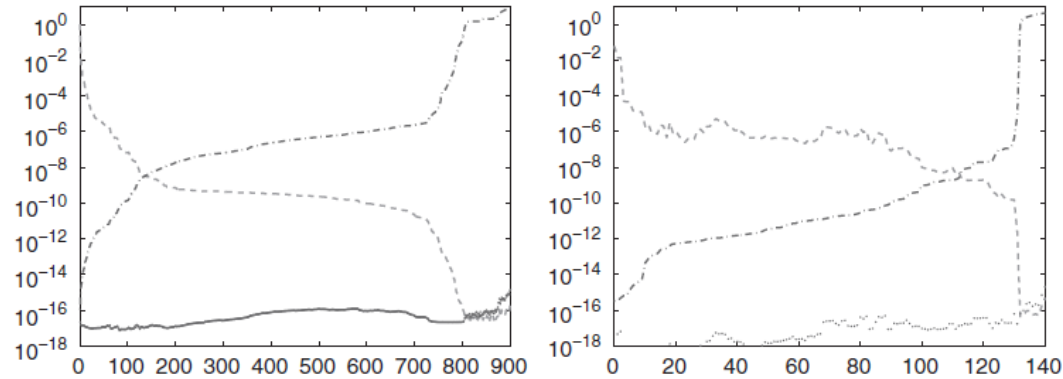


Figure 5.30 Results of MGS GMRES computations with the matrices Sherman2 (left) and West132 (right) from Matrix Market. Throughout the computation the product (dots) of the normwise relative backward error (dashed line) and the loss of orthogonality among the MGS Arnoldi vectors (dashed-dotted line) are close to (or below) machine precision.

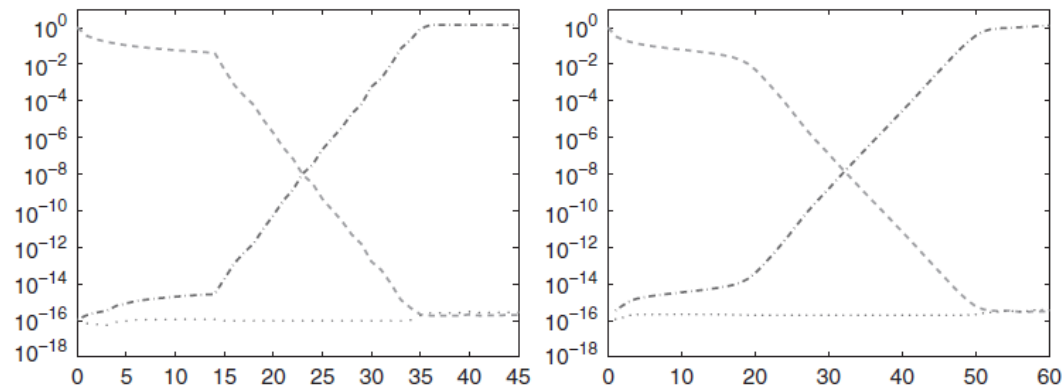
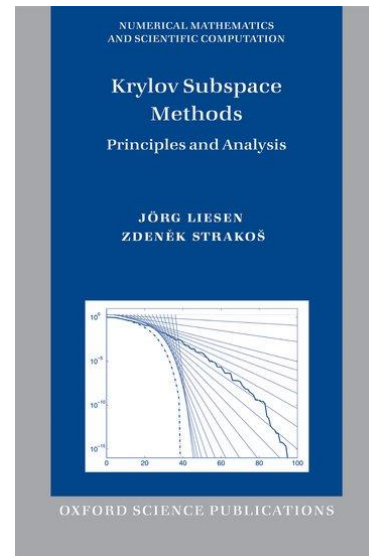


Figure 5.31 The same as in Figure 5.30 for matrices from the convection–diffusion model problem with dominating convection (see Section 5.7.5), the discontinuous inflow boundary conditions, the vertical wind (left) and the curl wind (right).

Pictures from (L. & Strakoš, Krylov Subspace Methods, Oxford University Press, 2013)



Completing the MGS-GMRES mosaic

SIAM J. MATRIX ANAL. APPL.
Vol. 28, No. 1, pp. 264–284

© 2006 Society for Industrial and Applied Mathematics

MODIFIED GRAM–SCHMIDT (MGS), LEAST SQUARES, AND BACKWARD STABILITY OF MGS-GMRES*

CHRISTOPHER C. PAIGE[†], MIROSLAV ROZLOŽNÍK[‡], AND ZDENĚK STRAKOŠ[‡]

8.2. Backward stability of MGS-GMRES for $Ax = b$ in (1.1). Even though MGS-GMRES always computes a backward stable solution \bar{y}_k for the least squares problem (7.3), see section 8.1, we still have to prove that $\bar{V}_k \bar{y}_k$ will be a backward stable solution for the original system (1.1) for some k (we take this k to be $\hat{m} - 1$ in (6.1)), and this is exceptionally difficult. Usually we want to show we have a backward stable solution when we *know* we have a small residual. The analysis here is different in that we will first prove that $\bar{B}_{\hat{m}}$ is numerically rank deficient, see (8.4), but to prove backward stability, we will then have to *prove* that our residual will be small, amongst other things, and this is far from obvious. Fortunately two little known researchers have studied this arcane area, and we will take ideas from [17]; see Theorem 2.4. To simplify the development and expressions we will absorb all small constants into the $\tilde{\gamma}_{kn}$ terms below.

Completing the MGS-GMRES mosaic

Using the usual approach of combining (8.15) with the definitions

$$\Delta b'_k \equiv -\frac{\|b\|_2}{\|b\|_2 + \|A\|_F \|\bar{x}_k\|_2} \tilde{r}_k(\bar{y}_k), \quad \Delta A'_k \equiv \frac{\|A\|_F \|\bar{x}_k\|_2}{\|b\|_2 + \|A\|_F \|\bar{x}_k\|_2} \frac{\tilde{r}_k(\bar{y}_k) \bar{x}_k^T}{\|\bar{x}_k\|_2^2},$$

shows $(A + \Delta A_k + \Delta A'_k) \bar{x}_k = b + \Delta b_k(\bar{y}_k) + \Delta b'_k$,

$$\|\Delta A_k + \Delta A'_k\|_F \leq \tilde{\gamma}_{kn} \|A\|_F, \quad \|\Delta b_k(\bar{y}_k) + \Delta b'_k\|_2 \leq \tilde{\gamma}_{kn} \|b\|_2,$$

proving that the MGS-GMRES solution \bar{x}_k is backward stable for (1.1).

Completing the MGS-GMRES mosaic

Using the usual approach of combining (8.15) with the definitions

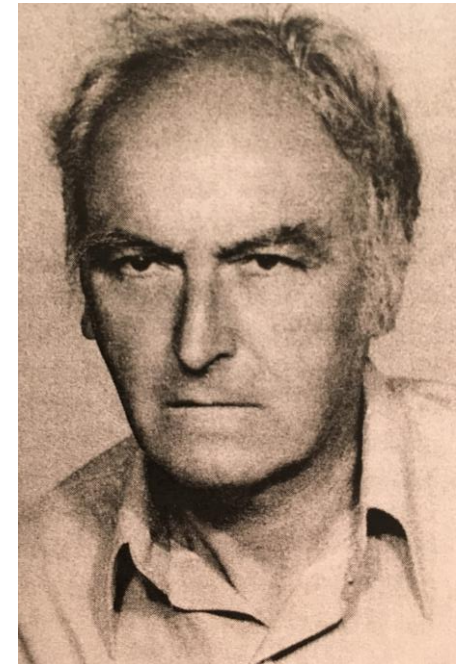
$$\Delta b'_k \equiv -\frac{\|b\|_2}{\|b\|_2 + \|A\|_F \|\bar{x}_k\|_2} \tilde{r}_k(\bar{y}_k), \quad \Delta A'_k \equiv \frac{\|A\|_F \|\bar{x}_k\|_2}{\|b\|_2 + \|A\|_F \|\bar{x}_k\|_2} \frac{\tilde{r}_k(\bar{y}_k) \bar{x}_k^T}{\|\bar{x}_k\|_2^2},$$

shows $(A + \Delta A_k + \Delta A'_k) \bar{x}_k = b + \Delta b_k(\bar{y}_k) + \Delta b'_k$,

$$\|\Delta A_k + \Delta A'_k\|_F \leq \tilde{\gamma}_{kn} \|A\|_F, \quad \|\Delta b_k(\bar{y}_k) + \Delta b'_k\|_2 \leq \tilde{\gamma}_{kn} \|b\|_2,$$

proving that the MGS-GMRES solution \bar{x}_k is backward stable for (1.1).

A higher level of understanding –
when Truth and Beauty become one.



Vlastimil Pták
(1925–1999)

Completing the MGS-GMRES mosaic

Using the usual approach of combining (8.15) with the definitions

$$\Delta b'_k \equiv -\frac{\|b\|_2}{\|b\|_2 + \|A\|_F \|\bar{x}_k\|_2} \tilde{r}_k(\bar{y}_k), \quad \Delta A'_k \equiv \frac{\|A\|_F \|\bar{x}_k\|_2}{\|b\|_2 + \|A\|_F \|\bar{x}_k\|_2} \frac{\tilde{r}_k(\bar{y}_k) \bar{x}_k^T}{\|\bar{x}_k\|_2^2},$$

shows $(A + \Delta A_k + \Delta A'_k) \bar{x}_k = b + \Delta b_k(\bar{y}_k) + \Delta b'_k$,

$$\|\Delta A_k + \Delta A'_k\|_F \leq \tilde{\gamma}_{kn} \|A\|_F, \quad \|\Delta b_k(\bar{y}_k) + \Delta b'_k\|_2 \leq \tilde{\gamma}_{kn} \|b\|_2,$$

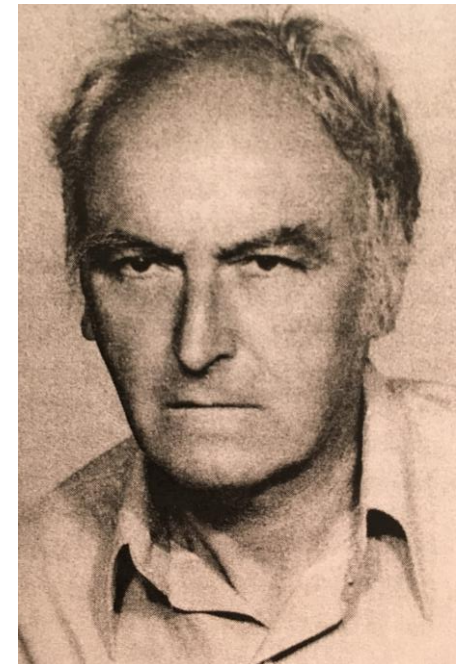
proving that the MGS-GMRES solution \bar{x}_k is backward stable for (1.1).

A higher level of understanding –
when Truth and Beauty become one.

V. Pták, *Circumstances of the submission of my paper in 1956*,
LAA 310 (2000):

“It is a comforting thought that the validity of mathematical
theorems cannot be affected by ideological disputes ...

Even on the shelves of mathematical libraries the 1956 volume of
Acta Szeged is conspicuous by the poorer quality of the paper.
The presence of a foreign army on the territory of Hungary
made it difficult to keep the usual standard ...”



Vlastimil Pták
(1925–1999)

The convergence behavior of GMRES

BIT
1998, Vol. 38, No. 4, pp. 636–643

0006-3835/98/3804-0636 \$12.00
© Swets & Zeitlinger

KRYLOV SEQUENCES OF MAXIMAL LENGTH AND CONVERGENCE OF GMRES *

M. ARIOLI¹, V. PTÁK² and Z. STRAKOŠ² †

The convergence behavior of GMRES

BIT
1998, Vol. 38, No. 4, pp. 636–643

0006-3835/98/3804-0636 \$12.00
© Swets & Zeitlinger

KRYLOV SEQUENCES OF MAXIMAL LENGTH AND CONVERGENCE OF GMRES *

M. ARIOLI¹, V. PTÁK² and Z. STRAKOŠ² †

- Complete characterization of all matrices A with prescribed eigenvalues and right hand sides b , such that GMRES attains a prescribed convergence curve.
- In particular, any nonincreasing convergence curve is possible for GMRES for a matrix having any eigenvalues.

Theorem 5.7.8

Consider N given positive numbers $f_0 \geq f_1 \geq \dots \geq f_{N-1} > 0$ and N nonzero complex numbers $\lambda_1, \dots, \lambda_N$, not necessarily distinct. Let $A \in \mathbb{C}^{N \times N}$ and $b \in \mathbb{C}^N$. Then the following three assertions are equivalent:

- (1) The eigenvalues of A are $\lambda_1, \dots, \lambda_N$ and GMRES applied to $Ax = b$ with $x_0 = 0$ yields the residual norms $\|r_n\| = f_n$ for $n = 0, 1, \dots, N - 1$.
- (2) $A = W_N R_N C R_N^{-1} W_N^*$ and $b = W_N h$, where
 - W_N is a unitary matrix,
 - C is the companion matrix of the polynomial

$$q(\lambda) \equiv (\lambda - \lambda_1) \cdots (\lambda - \lambda_N) \equiv \lambda^N - \sum_{j=0}^{N-1} \alpha_j \lambda^j$$

(this matrix is stated in (5.7.24) above),

- $h = [g_1, \dots, g_N]^T$, where $g_n \equiv (f_{n-1}^2 - f_n^2)^{1/2}$, $n = 1, \dots, N$, and we set $f_N \equiv 0$,
- R_N is nonsingular and upper triangular such that $R_N s = h$, where

$$s = [\xi_1, \dots, \xi_N]^T \quad \text{and} \quad p(\lambda) \equiv \left(1 - \frac{\lambda}{\lambda_1}\right) \cdots \left(1 - \frac{\lambda}{\lambda_N}\right) \equiv 1 - \sum_{j=1}^N \xi_j \lambda^j,$$

$$\text{i.e. } \xi_n = -\alpha_n / \alpha_0, \quad n = 1, \dots, N - 1, \quad \text{and} \quad \xi_N = 1 / \alpha_0.$$

- (3) $A = W_N Y C Y^{-1} W_N^*$ and $b = W_N h$, where W_N , C and h are defined as in (2),

$$Y \equiv R_N C^{-1} = \left[\begin{array}{c|c} g_1 & R \\ \vdots & \\ g_{N-1} & \\ \hline g_N & 0 \end{array} \right],$$

where g_1, \dots, g_N are defined as in (2), and R is an $(N - 1) \times (N - 1)$ nonsingular upper triangular matrix.

The convergence behavior of GMRES

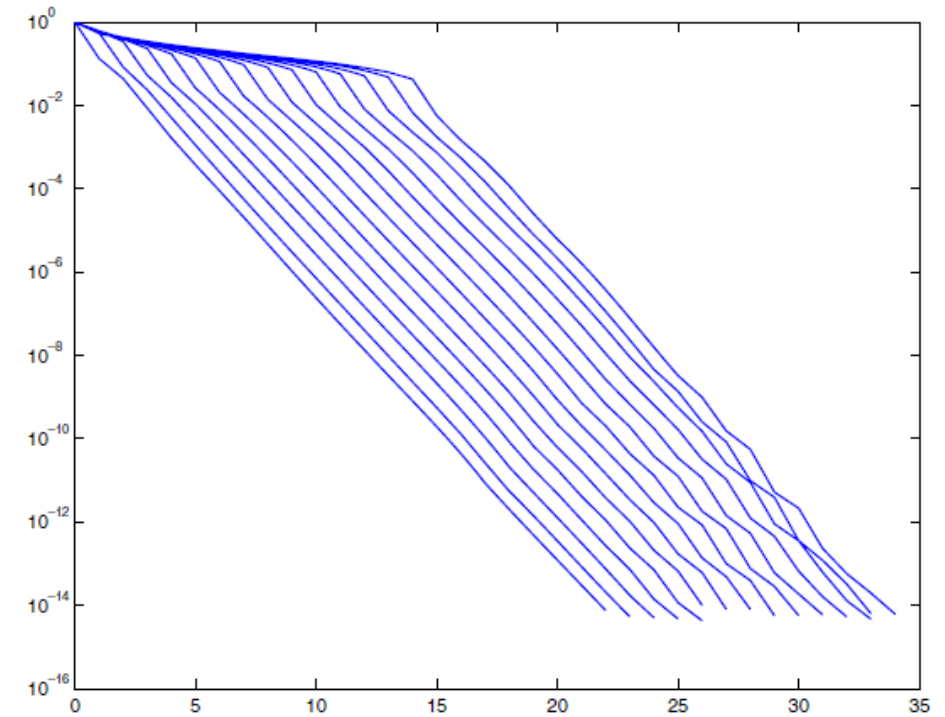
SIAM J. SCI. COMPUT.
Vol. 26, No. 6, pp. 1989–2009

© 2005 Society for Industrial and Applied Mathematics

GMRES CONVERGENCE ANALYSIS FOR A CONVECTION-DIFFUSION MODEL PROBLEM*

J. LIESEN[†] AND Z. STRAKOŠ[‡]

- The length initial phase of stagnation depends on the boundary conditions in the model problem.
- The convergence of GMRES in particular for nonnormal matrices may depend strongly on the right hand side.



The convergence behavior of GMRES

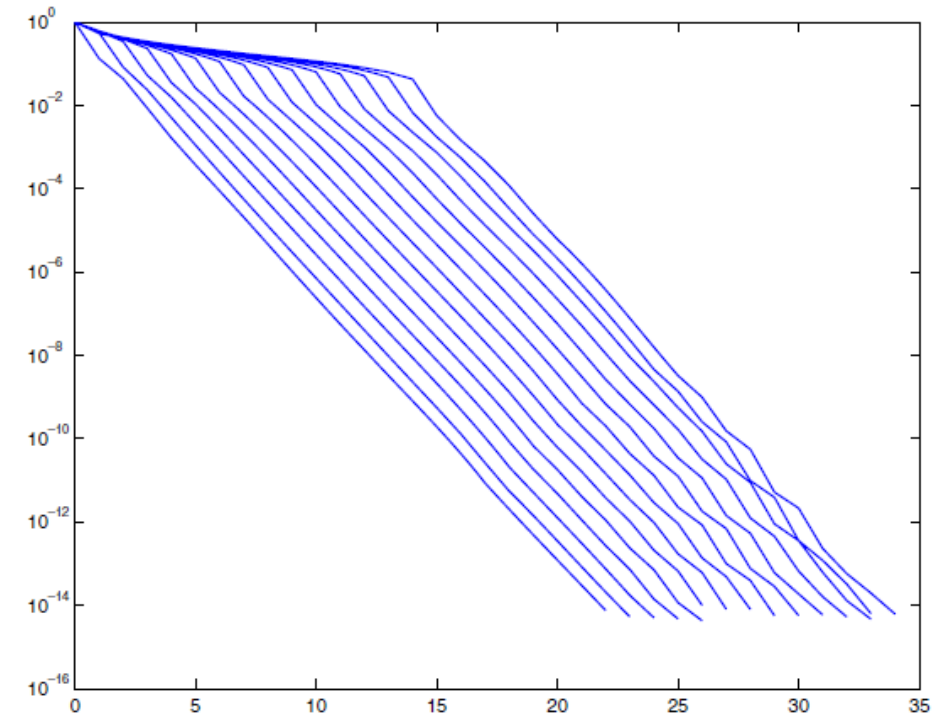
SIAM J. SCI. COMPUT.
Vol. 26, No. 6, pp. 1989–2009

© 2005 Society for Industrial and Applied Mathematics

GMRES CONVERGENCE ANALYSIS FOR A CONVECTION-DIFFUSION MODEL PROBLEM*

J. LIESEN[†] AND Z. STRAKOŠ[‡]

- The length initial phase of stagnation depends on the boundary conditions in the model problem.
- The convergence of GMRES in particular for nonnormal matrices may depend strongly on the right hand side.



Some pieces have been placed in the GMRES convergence mosaic, but the overall picture is still unclear.

Piecing together the mosaic

“Goal: To get some *understanding* when and why things work, and when and why they do not.”

GMRES convergence for
convection-diffusion problems

Backward stability
of GMRES

All nonincreasing convergence
curves are possible for GMRES

Piecing together the mosaic

“Goal: To get some *understanding* when and why things work, and when and why they do not.”

The Strakoš matrix &
finite precision CG

Existence of short
Arnoldi recurrences

GMRES convergence for
convection-diffusion problems

Error estimation
in the CG method

Backward stability
of GMRES

Gauss quadrature &
the CG method

All nonincreasing convergence
curves are possible for GMRES

Piecing together the mosaic

“Goal: To get some *understanding* when and why things work, and when and why they do not.”

The Strakoš matrix &
finite precision CG

Existence of short
Arnoldi recurrences

GMRES convergence for
convection-diffusion problems

Error estimation
in the CG method

Backward stability
of GMRES

Discretization and algebraic
errors in elliptic PDE problems

Gauss quadrature &
the CG method

All nonincreasing convergence
curves are possible for GMRES

A posteriori error
estimators & stopping criteria

Piecing together the mosaic

“Goal: To get some *understanding* when and why things work, and when and why they do not.”

The Strakoš matrix & finite precision CG

Error estimation in the CG method

Gauss quadrature & the CG method

Existence of short Arnoldi recurrences

Backward stability of GMRES

All nonincreasing convergence curves are possible for GMRES

GMRES convergence for convection-diffusion problems

Discretization and algebraic errors in elliptic PDE problems

A posteriori error estimators & stopping criteria

Acta Numerica (2006), pp. 471–542
doi: 10.1017/S096249290626011X © Cambridge University Press, 2006
Printed in the United Kingdom

The Lanczos and conjugate gradient algorithms in finite precision arithmetic

Gérard Meurant
CEA/DRF,
BP 12,

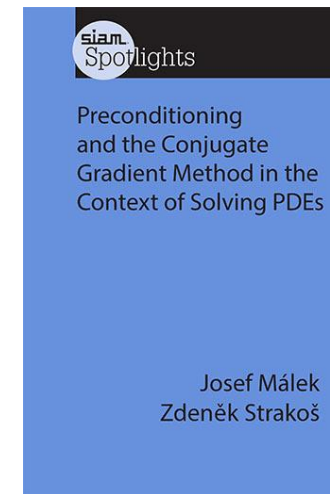
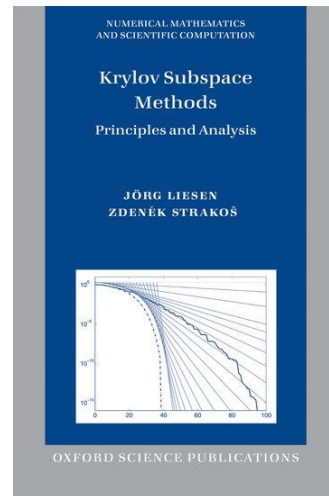
91680, Bruyères le Châtel, France
E-mail: gerard.meurant@cea.fr

Zdeněk Strakoš
Institute of Computer Science,
Academy of Sciences of the Czech Republic,
Pod Václavskými 2,
182 07 Praha 8, Czech Republic
E-mail: strako@cs.cas.cz

Dedicated to Chris Paige for his fundamental contributions to the rounding error analysis of the Lanczos algorithm

The Lanczos and conjugate gradient algorithms were introduced more than five decades ago as tools for numerical computation of dominant eigenvalues of symmetric matrices and for solving linear algebraic systems with symmetric positive definite matrices, respectively. Because of their fundamental relationship with the theory of orthogonal polynomials and Gauss quadrature of the Riemann–Stieltjes integral, the Lanczos and conjugate gradient algorithms represent very interesting general mathematical objects, with highly nonlinear properties which can be conveniently translated from algebraic language into the language of mathematical analysis, and vice versa. The algorithms are also very interesting numerically, since their numerical behaviour can be explained by an elegant mathematical theory, and the interplay between analysis and algebra is useful there too.

Motivated by this view, the present contribution wishes to pay a tribute to those who have made an understanding of the Lanczos and conjugate gradient algorithms possible through their pioneering work, and to review recent solutions of several open problems that have also contributed to knowledge of the subject.



Piecing together the mosaic

“Goal: To get some *understanding* when and why things work, and when and why they do not.”



Srdečné blahopraní k narodeninám!

