NUMERICS OF THE GRAM-SCHMIDT ORTHOGONALIZATION PROCESS

Miro Rozložník

Institute of Computer Science,
Czech Academy of Sciences,
Prague, Czech Republic
email: miro@cs.cas.cz

joint results with
Luc Giraud, Julien Langou and Jasper van den Eshof
but also **Alicja Smoktunowicz and Jesse Barlow!**

Seminarium Pierścienie, macierze i algorytmy numeryczne,
Politechnika Warszawska, Warszawa, November 13, 2009

OUTLINE

1. HISTORICAL REMARKS
2. CLASSICAL AND MODIFIED GRAM-SCHMIDT ORTHOGONALIZATION
3. GRAM-SCHMIDT IN FINITE PRECISION ARITHMETIC: LOSS OF ORTHOGONALITY IN THE CLASSICAL GRAM-SCHMIDT PROCESS
4. NUMERICAL NONSINGULARITY AND GRAM-SCHMIDT ALGORITHM WITH REORTHOGONALIZATION

GRAM-SCHMIDT PROCESS AS QR ORTHOGONALIZATION

$$A = (a_1, \ldots, a_n) \in \mathcal{R}^{m,n}$$
$$m \geq rank(A) = n$$

orthogonal basis Q of $span(A)$

$Q = (q_1, \ldots, q_n) \in \mathcal{R}^{m,n}$, $Q^T Q = I_n$

$A = QR$, R upper triangular ($A^T A = R^T R$)

HISTORICAL REMARKS I

origination of the QR factorization used for orthogonalization of functions:

J. P. Gram: Über die Entwicklung reeller Funktionen in Reihen mittelst der Methode der Kleinsten Quadrate. Journal f. r. a. Math., 94: 41-73, 1883.

algorithm of the QR decomposition but still in terms of functions:

E. Schmidt: Zur Theorie der linearen und nichlinearen Integralgleichungen. I Teil. Entwicklung willkürlichen Funktionen nach system vorgeschriebener. Mathematische Annalen, 63: 433-476, 1907.

name of the QR decomposition in the paper on nonsymmetric eigenvalue problem, rumor: the "Q" in QR was originally an "O" standing for orthogonal:

J.G.F. Francis: The QR transformation, parts I and II. Computer Journal 4:265-271, 332-345, 1961, 1962.

HISTORICAL REMARKS II

"modified" Gram-Schmidt (MGS) interpreted as an elimination method using weighted row sums not as an orthogonalization technique:

P.S. Laplace: Theorie Analytique des Probabilités. Courcier, Paris, third edition, 1820. Reprinted in P.S. Laplace. (Evres Compeétes. Gauthier-Vilars, Paris, 1878-1912).

"classical" Gram-Schmidt (CGS) algorithm to solve linear systems of infinitely many solutions:

E. Schmidt: Über die Auflösung linearen Gleichungen mit unendlich vielen Unbekanten, Rend. Circ. Mat. Palermo. Ser. 1, 25 (1908), pp. 53-77.

first application to finite-dimensional set of vectors:

G. Kowalewski: Einfuehrung in die Determinantentheorie. Verlag von Veit & Comp., Leipzig, 1909.

CLASSICAL AND MODIFIED GRAM-SCHMIDT ALGORITHMS

classical (CGS)

for $j = 1, \ldots, n$
  $u_j = a_j$
  for $k = 1, \ldots, j-1$

    $u_j = u_j - (a_j, q_k) q_k$

  end
  $q_j = u_j / \|u_j\|$
end

modified (MGS)

for $j = 1, \ldots, n$
  $u_j = a_j$
  for $k = 1, \ldots, j-1$

    $u_j = u_j - (u_j, q_k) q_k$

  end
  $q_j = u_j / \|u_j\|$
end

CLASSICAL AND MODIFIED GRAM-SCHMIDT ALGORITHMS

finite precision arithmetic:

$$\bar{Q} = (\bar{q}_1, \ldots, \bar{q}_n),\ \bar{Q}^T\bar{Q} \neq I_n,\ \|I - \bar{Q}^T\bar{Q}\| \leq ?$$

$$A \neq \bar{Q}\bar{R},\ \|A - \bar{Q}\bar{R}\| \leq ?$$

$$\bar{R}?\ ,\ \mathrm{cond}(\bar{R}) \leq ?$$

classical and **modified** Gram-Schmidt are mathematically equivalent, but they have **"different"** numerical properties

classical Gram-Schmidt can be **"quite unstable"**, can **"quickly" lose** all semblance of **orthogonality**

ILLUSTRATION, EXAMPLE

Läuchli, 1961, Björck, 1967: $A = \begin{pmatrix} 1 & 1 & 1 \\ \sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & \sigma \end{pmatrix}$

$$\kappa(A) = \sigma^{-1}(n + \sigma^2)^{1/2} \approx \sigma^{-1}\sqrt{n},\ \sigma \ll 1$$
$$\sigma_{min}(A) = \sigma,\ \|A\| = \sqrt{n + \sigma^2}$$

assume first that $\sigma^2 \leq u$, so $\mathrm{fl}(1 + \sigma^2) = 1$

if no other rounding errors are made, the matrices computed in CGS and MGS have the following form:

ILLUSTRATION, EXAMPLE

$$\begin{pmatrix} 1 & 0 & 0 \\ \sigma & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & 0 \\ \sigma & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{\sqrt{2}}{\sqrt{3}} \end{pmatrix}$$

CGS: $(\bar{q}_3, \bar{q}_1) = -\sigma/\sqrt{2}$, $(\bar{q}_3, \bar{q}_2) = 1/2$,
MGS: $(\bar{q}_3, \bar{q}_1) = -\sigma/\sqrt{6}$, $(\bar{q}_3, \bar{q}_2) = 0$

complete loss of orthogonality ($\Longleftrightarrow$ loss of lin. independence, loss of (numerical) rank): $\sigma^2 \leq u$ (CGS), $\sigma \leq u$ (MGS)

GRAM-SCHMIDT PROCESS VERSUS ROUNDING ERRORS

- **modified** Gram-Schmidt (MGS):

  assuming $\widehat{c}_1 u\kappa(A) < 1$

  $$\|I - \bar{Q}^T\bar{Q}\| \leq \frac{\widehat{c}_2 u\kappa(A)}{1-\widehat{c}_1 u\kappa(A)}$$

  Björck, 1967 , Björck, Paige, 1992

- **classical** Gram-Schmidt (CGS)?

  $$\|I - \bar{Q}^T\bar{Q}\| \leq \frac{\tilde{c}_2 u\kappa^{n-1}(A)}{1-\tilde{c}_1 u\kappa^{n-1}(A)}?$$

  Kielbasinski, Schwettlik, 1994

  Polish version of the book, 2nd edition

TRIANGULAR FACTOR FROM CLASSICAL GRAM-SCHMIDT VS. CHOLESKY FACTOR OF THE CROSS-PRODUCT MATRIX

exact arithmetic:

$$r_{i,j} = (a_j, q_i) \;=\; \left(a_j, \frac{a_i - \sum_{k=1}^{i-1} r_{k,i} q_k}{r_{i,i}}\right)$$

$$\;=\; \frac{(a_j, a_i) - \sum_{k=1}^{i-1} r_{k,i} r_{k,j}}{r_{i,i}}$$

The computation of R in the classical Gram-Schmidt is closely related to the left-looking Cholesky factorization of the cross-product matrix $A^T A = R^T R$

$$
\begin{aligned}
\bar{r}_{i,j} &= fl(a_j, \bar{q}_i) = (a_j, \bar{q}_i) + \Delta e^{(1)}_{i,j} \\
&= \left(a_j, \frac{fl(a_i - \sum_{k=1}^{i-1} \bar{q}_k \bar{r}_{k,i})}{\bar{r}_{i,i}} + \Delta e^{(2)}_i \right) + \Delta e^{(1)}_{i,j}
\end{aligned}
$$

$$
\begin{aligned}
\bar{r}_{i,i}\bar{r}_{i,j} &= \left(a_j, a_i - \sum_{k=1}^{i-1} \bar{q}_k \bar{r}_{k,i} + \Delta e^{(3)}_i \right) \\
&+ \bar{r}_{i,i} \left[(a_j, \Delta e^{(2)}_i) + \Delta e^{(1)}_{i,j} \right] \\
&= (a_i, a_j) - \sum_{k=1}^{i-1} \bar{r}_{k,i} [\bar{r}_{k,j} - \Delta e^{(1)}_{k,j}] \\
&+ (a_j, \Delta e^{(3)}_i) + \bar{r}_{i,i} \left[(a_j, \Delta e^{(2)}_i) + \Delta e^{(1)}_{i,j} \right]
\end{aligned}
$$

CLASSICAL GRAM-SCHMIDT PROCESS: COMPUTED TRIANGULAR FACTOR

$$\Sigma_{k=1}^{i} \bar{r}_{k,i}\bar{r}_{k,j} = (a_i, a_j) + \Delta e_{i,j},\ i < j$$

$$[A^T A + \Delta E_1]_{i,j} = [\bar{R}^T \bar{R}]_{i,j}!$$

$$\|\Delta E_1\| \leq c_1 u \|A\|^2$$

The CGS process is another way how to compute **a backward stable Cholesky factor of the cross-product matrix** $A^T A$!

CLASSICAL GRAM-SCHMIDT PROCESS: DIAGONAL ELEMENTS

$$u_j = (I - Q_{j-1}Q_{j-1}^T)a_j$$
$$\|u_j\| = \|a_j - Q_{j-1}(Q_{j-1}^T a_j)\| =$$
$$(\|a_j\|^2 - \|Q_{j-1}^T a_j\|^2)^{1/2}$$

computing $q_j = \frac{u_j}{(\|a_j\| - \|Q_{j-1}^T a_j\|)^{1/2}(\|a_j\| + \|Q_{j-1}^T a_j\|)^{1/2}}$,

$$\|a_j\|^2 = \Sigma_{k=1}^{j}\bar{r}_{k,j}^2 + \Delta e_{j,j},\ \|\Delta e_{j,j}\| \leq c_1 u \|a_j\|^2$$

J. Barlow, A. Smoktunowicz, Langou, 2006

CLASSICAL GRAM-SCHMIDT PROCESS: THE LOSS OF ORTHOGONALITY

$$A^T A + \Delta E_1 = \bar{R}^T \bar{R},\ A + \Delta E_2 = \bar{Q}\bar{R}$$

$$\bar{R}^T(I - \bar{Q}^T\bar{Q})\bar{R} = -(\Delta E_2)^T A - A^T \Delta E_2 - (\Delta E_2)^T \Delta E_2 + \Delta E_1$$

assuming $c_2 u\kappa(A) < 1$

$$\|I - \bar{Q}^T\bar{Q}\| \leq \frac{c_3 u\kappa^2(A)}{1 - c_2 u\kappa(A)}$$

GRAM-SCHMIDT PROCESS VERSUS ROUNDING ERRORS

- modified Gram-Schmidt (MGS): assuming $\hat{c}_1 u\kappa(A) < 1$

  $$\|I - \bar{Q}^T\bar{Q}\| \leq \frac{\hat{c}_2 u\kappa(A)}{1-\hat{c}_1 u\kappa(A)}$$

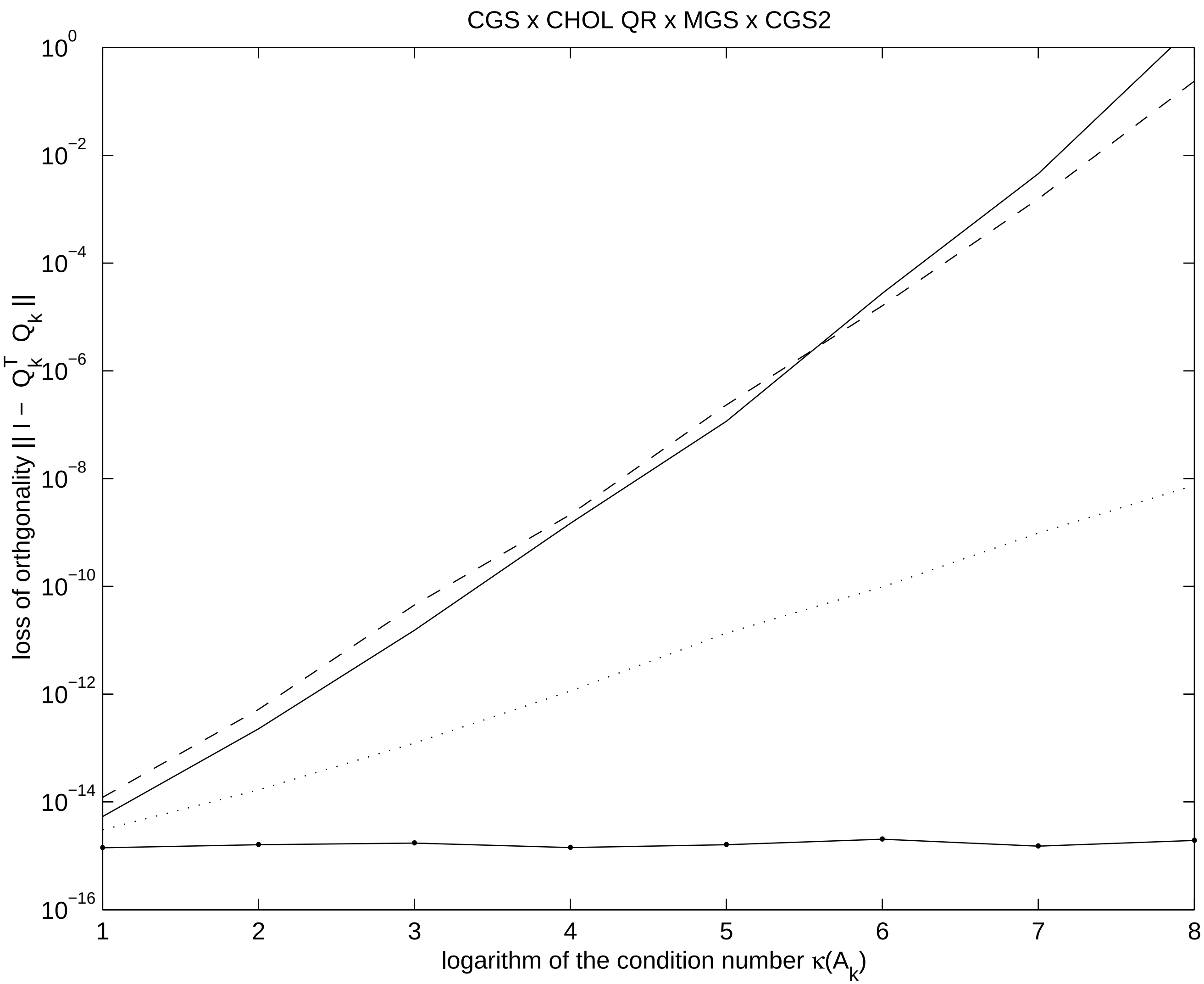
  Björck, 1967, Björck, Paige, 1992

- **classical Gram-Schmidt (CGS):** assuming $c_2 u\kappa(A) < 1$

  $$\|I - \bar{Q}^T\bar{Q}\| \leq \frac{c_3 u\kappa^2(A)}{1-c_2 u\kappa(A)}!$$

  Giraud, Van den Eshof, Langou, R, 2006

  J. Barlow, A. Smoktunowicz, Langou, 2006



Stewart, "Matrix algorithms" book, p. 284, 1998

GRAM-SCHMIDT ALGORITHMS WITH COMPLETE REORTHOGONALIZATION

classical (CGS2)

for $j = 1, \ldots, n$
  $u_j = a_j$
  for $i = 1, 2$
  for $k = 1, \ldots, j-1$

$$u_j = u_j - (a_j, q_k)q_k$$

  end
  end
  $q_j = u_j / \|u_j\|$
end

modified (MGS2)

for $j = 1, \ldots, n$
  $u_j = a_j$
  for $i = 1, 2$
  for $k = 1, \ldots, j-1$

$$u_j = u_j - (u_j, q_k)q_k$$

  end
  end
  $q_j = u_j / \|u_j\|$
end

GRAM-SCHMIDT PROCESS VERSUS ROUNDING ERRORS

- Gram-Schmidt (MGS, CGS): assuming $c_1 u\kappa(A) < 1$

  $$\|I - \bar{Q}^T\bar{Q}\| \leq \frac{c_2 u\kappa^{1,2}(A)}{1 - c_1 u\kappa(A)}$$

  Björck, 1967, Björck, Paige, 1992

  Giraud, van den Eshof, Langou, R, 2005

- Gram-Schmidt with **reorthogonalization** (CGS2, MGS2): assuming $c_3 u\kappa(A) < 1$

  $$\|I - \bar{Q}^T\bar{Q}\| \leq c_4 u$$

  Hoffmann, 1988

  Giraud, van den Eshof, Langou, R, 2005

ROUNDING ERROR ANALYSIS OF REORTHOGONALIZATION STEP

$$u_j = (I - Q_{j-1}Q_{j-1}^T)a_j,\ v_j = (I - Q_{j-1}Q_{j-1}^T)^2 a_j$$

$$\|u_j\| = |r_{j,j}| \geq \sigma_{min}(R_j) = \sigma_{min}(A_j) \geq \sigma_{min}(A)$$

$$\frac{\|a_j\|}{\|u_j\|} \leq \kappa(A),\ \frac{\|u_j\|}{\|v_j\|} = 1$$

$$A + \Delta E_2 = \bar{Q}\bar{R},\ \|\Delta E_2\| \leq c_2 u \|A\|$$

$$\frac{\|a_j\|}{\|\bar{u}_j\|} \leq \frac{\kappa(A)}{1 - \tilde{c}_1 u \kappa(A)},\ \frac{\|\bar{u}_j\|}{\|\bar{v}_j\|} \leq [1 - \tilde{c}_2 u \kappa(A)]^{-1}$$

FLOPS VERSUS PARALLELISM

- classical Gram-Schmidt (CGS): mn^2 saxpys
- classical Gram-Schmidt with reorthogonalization (CGS2): $2mn^2$ saxpys
- Householder orthogonalization: $2(mn^2 - n^3/3)$ saxpys

in parallel environment and using BLAS2, CGS2 may be faster than (plain) MGS!

Frank, Vuik, 1999, Lehoucq, Salinger, 2001

THANK YOU FOR YOUR ATTENTION!

THE ARNOLDI PROCESS AND THE GMRES METHOD WITH THE CLASSICAL GRAM-SCHMIDT PROCESS

$$V_n = [v_1, v_2, \ldots, v_n]$$

$$[r_0, AV_n] = V_{n+1}[\|r_0\|e_1, H_{n+1,n}]$$

$H_{n+1,n}$ is an upper Hessenberg matrix

Arnoldi process is a (recursive) column-oriented QR decomposition of the (special) matrix $[r_0, AV_n]$!

$$x_n = x_0 + V_n y_n, \; \|\|r_0\|e_1 - H_{n+1,n}y_n\| = \min_y \|\|r_0\|e_1 - H_{n+1,n}y\|$$

THE GRAM-SCHMIDT PROCESS IN THE ARNOLDI CONTEXT: LOSS OF ORTHOGONALITY

- modified Gram-Schmidt (MGS):

  $\|I - \bar{V}_{n+1}^T \bar{V}_{n+1}\| \leq \bar{c}_1 u \kappa([\bar{v}_1, A\bar{V}_n])$

Björck, Paige 1967, 1992

- classical Gram-Schmidt (CGS):
  $\|I - \bar{V}_{n+1}^T \bar{V}_{n+1}\| \leq \bar{c}_2 u \kappa^2([\bar{v}_1, A\bar{V}_n])$

Giraud, Langou, R, Van den Eshof 2004

CONDITION NUMBER IN ARNOLDI VERSUS RESIDUAL NORM IN GMRES

The loss of orthogonality in Arnoldi is controlled by the convergence of the residual norm in GMRES:

$$\|I - \bar{V}_{n+1}^T \bar{V}_{n+1}\| \leq \bar{c}_\alpha u \kappa^\alpha([\bar{v}_1, A\bar{V}_n]), \qquad \alpha = 1, 2$$

Björck 1967, Björck and Paige , 1992

Giraud, Langou, R, Van den Eshof 2003

$$\kappa([\bar{v}_1, A\bar{V}_n]) \leq \frac{\|[\bar{v}_1, A\bar{V}_n]\|}{\frac{\|\widehat{r}_n\|}{\|\bar{r}_0\|}[1 + \frac{\|\widehat{y}_n\|^2}{1-\delta_n^2}]^{1/2}}$$

$$\frac{\|\widehat{r}_n\|}{\|\bar{r}_0\|} = \|\bar{v}_1 - A\bar{V}_n\widehat{y}_n\| = min_y\|\bar{v}_1 - A\bar{V}_n y\| \ , \ \delta_n = \frac{\sigma_{n+1}([\bar{v}_1, A\bar{V}_n])}{\sigma_n(A\bar{V}_n)} < 1$$

Paige, Strakoš, 2000-2002

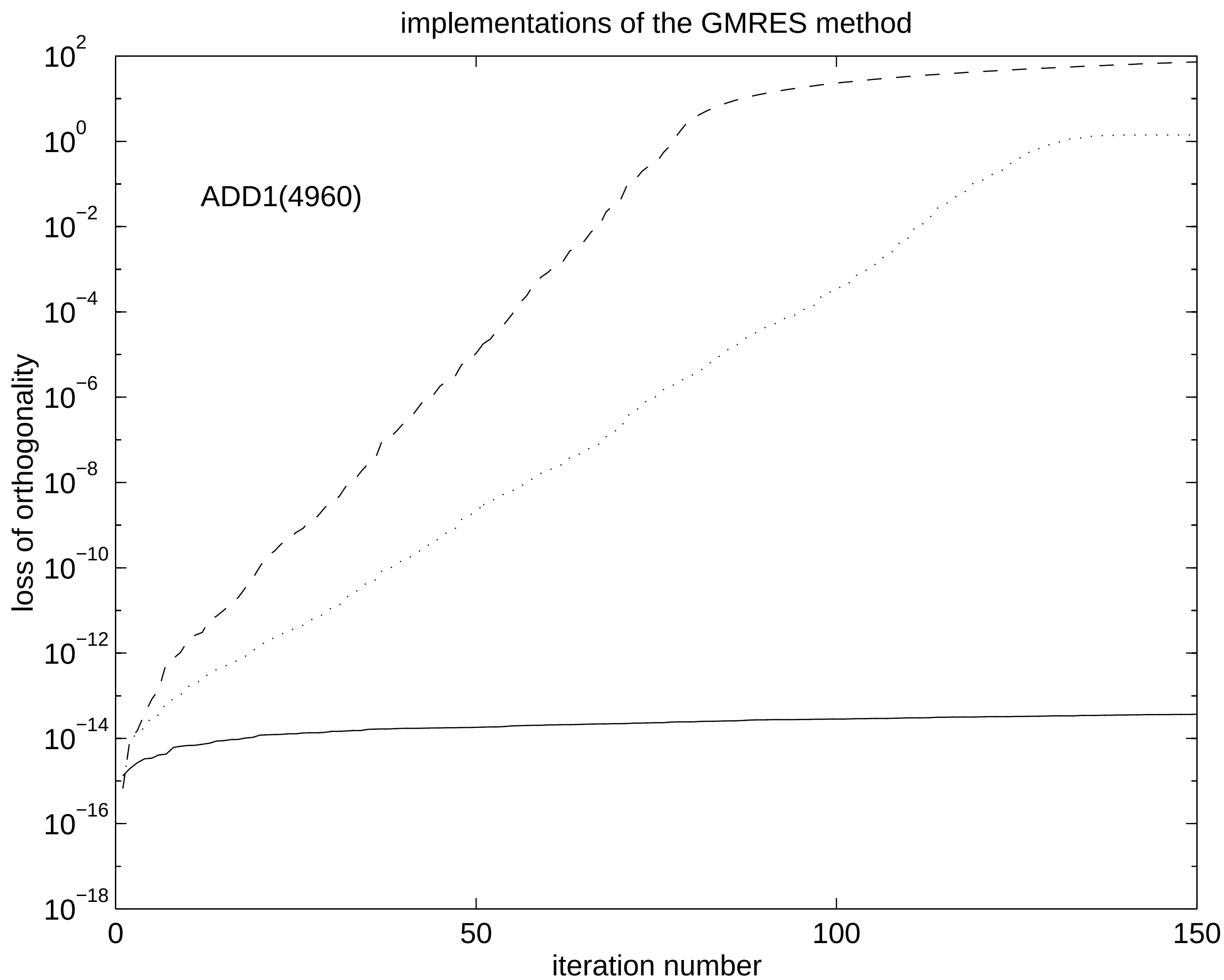
Greenbaum, R, Strakoš, 1997

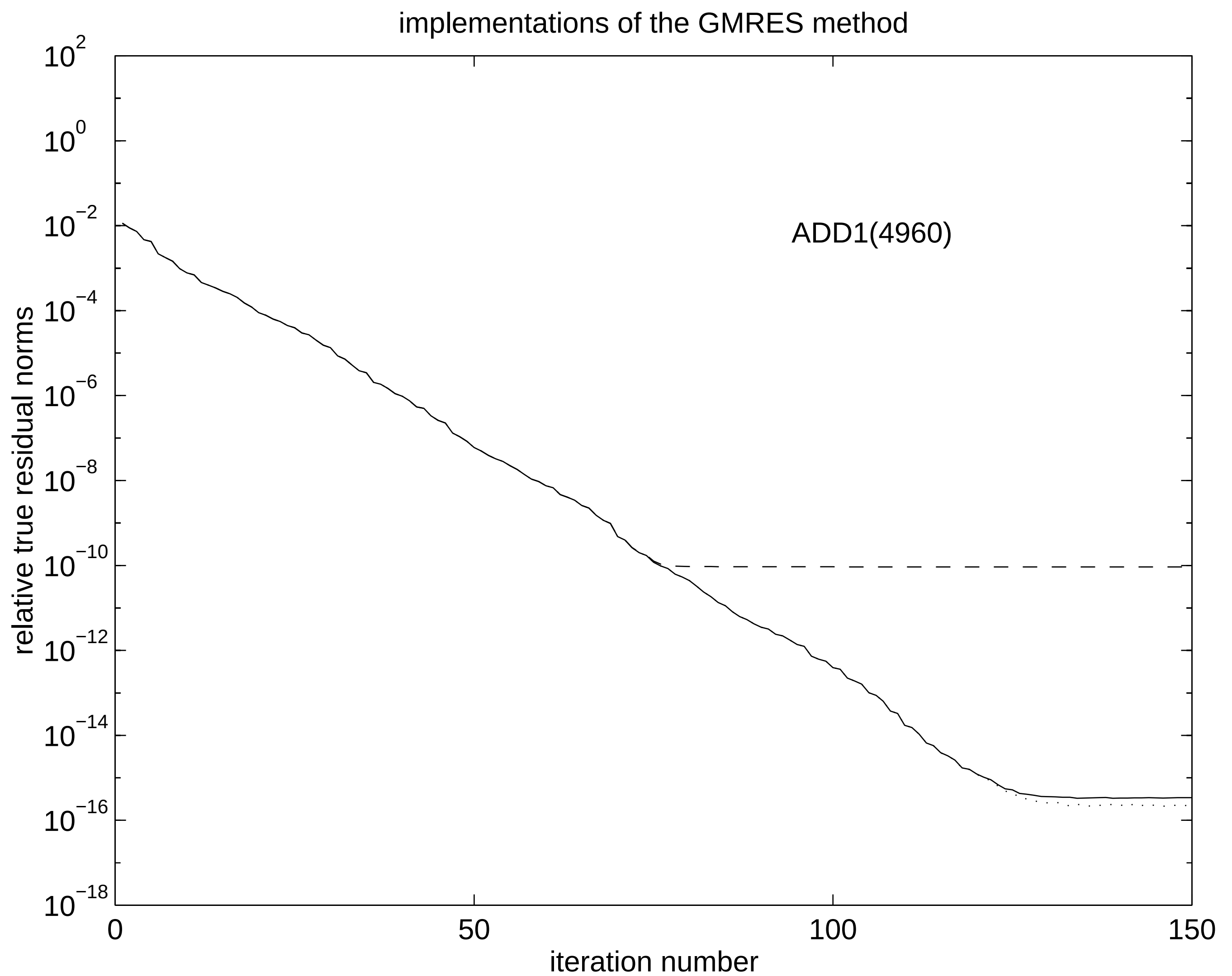
THE GMRES METHOD WITH THE GRAM-SCHMIDT PROCESS

The total loss of orthogonality (rank-deficiency) in the Arnoldi process with Gram-Schmidt can occur **only after** GMRES reaches its final accuracy level:

- modified Gram-Schmidt (MGS):
$$\frac{\|\hat{r}_n\|\|}{\|\bar{r}_0\|[1+\frac{\|\hat{y}_n\|^2}{1-\delta_n^2}]^{1/2}} \approx \bar{c}_1[\bar{v}_1, A\bar{V}_n]\|u$$

- classical Gram-Schmidt (CGS):
$$\frac{\|\hat{r}_n\|}{\|\bar{r}_0\|[1+\frac{\|\hat{y}_n\|^2}{1-\delta_n^2}]^{1/2}} \approx \left[\bar{c}_2\|[\bar{v}_1, A\bar{V}_n]\|u\right]^{1/2}$$


implementations of the GMRES method
ADD1(4960)
loss of orthogonality
iteration number
10^2
10^0
10^−2
10^−4
10^−6
10^−8
10^−10
10^−12
10^−14
10^−16
10^−18
0
50
100
150



implementations of the GMRES method
ADD1(4960)
relative true residual norms
iteration number
10^{2}
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
0
50
100
150

LEAST SQUARES PROBLEM WITH CLASSICAL GRAM-SCHMIDT

$$\|b - Ax\| = \min_u \|b - Au\|, \; r = b - Ax$$

$$A^T A x = A^T b$$

$$\bar{r} = (I - \bar{Q}\bar{Q}^T)b + \Delta e_1$$

$$(\bar{R} + \Delta E_3)\bar{x} = \bar{Q}^T b + \Delta e_2$$

$$\|\Delta e_1\|, \|\Delta e_2\| \leq c_0 u \|b\|, \; \|\Delta E_3\| \leq c_0 u \|\bar{R}\|$$

LEAST SQUARES PROBLEM WITH CLASSICAL GRAM-SCHMIDT

$$\bar{R}^T(\bar{R} + \Delta E_3)\bar{x} = (\bar{Q}\bar{R})^T b + \bar{R}^T \Delta e_2$$

$$(A^T A + \Delta E_1 + \bar{R}^T \Delta E_3)\bar{x} = (A + \Delta E_2)^T b + \bar{R}^T \Delta e_2$$

$$(A^T A + \Delta E)\bar{x} = A^T b + \Delta e$$

$$\|\Delta E\| \le c_4 u \|A\|^2, \; \|\Delta e\| \le c_4 u \|A\| \|b\|$$

LEAST SQUARES PROBLEM WITH CLASSICAL GRAM-SCHMIDT

$$\frac{\|\bar{r}-r\|}{\|b\|} \leq \kappa(A)(2\kappa(A)+1)\frac{c_5 u}{[1-c_1)u\kappa^2(A)]^{1/2}}$$

$$\frac{\|\bar{x}-x\|}{\|x\|} \leq \kappa^2(A)\left(2+\frac{\|r\|}{\|A\|\|x\|}\right)\frac{c_5 u}{1-c_1 u\kappa^2(A)}$$

The least square solution with classical Gram-Schmidt has the same forward error bound as the normal equation method:
$\bar{R}-\bar{Q}^T A = \bar{R}-\bar{R}^{-T}(A+\Delta E_2)^T A = -\bar{R}^{-T}[\Delta E_1 + (\Delta E_2)^T A]$

Björck, 1967

LEAST SQUARES PROBLEM WITH BACKWARD STABLE QR FACTORIZATION

$$\frac{\|\bar{r}-r\|}{\|b\|} \leq (2\kappa(A)+1)c_6 u$$

$$\frac{\|\bar{x}-x\|}{\|x\|} \leq \kappa(A)\left[2+(\kappa(A)+1)\frac{\|r\|}{\|A\|\|x\|}\right]\frac{c_6 u}{1-c_6 u\kappa(A)}$$

Householder QR factorization, modified Gram-Schmidt

Wilkinson, Golub, 1966, Björck, 1967