ROUNDING ERROR ANALYSIS OF THE CLASSICAL GRAM-SCHMIDT PROCESS USED FOR SOLVING THE LEAST SQUARES PROBLEMS

Miro Rozložník

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

joint results with
Luc Giraud, Julien Langou and Jasper van den Eshof

16.letní škola Software a Algoritmy Numerické Matematiky,
SANM'05, Hotel Srní, Srní, 12.-16. září 2005

THE LEAST SQUARES PROBLEM AND QR ORTHOGONALIZATION

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

$$m \geq rank(A) = n$$

$$\|b - Ax\| = \min_u \|b - Au\|$$

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 ($A^T A = R^T R$)

GRAM-SCHMIDT ALGORITHMS AND CHOLESKY QR DECOMPOSITION

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

  Schmidt, 1907,1908

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

  Laplace, 1816, Cauchy, 1837

- **Cholesky** QR decomposition

  Gauss, 1809, Cholesky, Benoit, 1927

all three schemes are mathematically equivalent, but they have **"different"** numerical properties, **classical** Gram-Schmidt can be **"quite unstable", ** forming the cross-product matrix can be potentially **"dangerous"**

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, Schwetlick, 1994

  Polish version of the book, 2nd edition

CHOLESKY QR DECOMPOSITION: COMPUTED ORTHOGONAL FACTOR

$$\|fl(A^TA) - A^TA\| \leq c_0 u \|A\|^2$$

$$\sigma_{min}(fl(A^TA)) \geq \sigma_{min}(A^TA) - c_0 u \|A\|^2 !$$

$$\bar{Q} = fl(A\bar{R}^{-1}) = A\bar{R}^{-1} + \Delta F_1$$

$$\|\Delta F_1\| \leq c_1 u \|A\| \|\bar{R}^{-1}\|$$

$$\bar{Q}\bar{R} = fl(A\bar{R}^{-1})\bar{R} = A + \bar{R}\Delta F_1 !$$

$$\bar{r}_{i,j} = \frac{fl\left[fl(a_i,a_j) - \sum_{k=1}^{i-1} \bar{r}_{k,j}\bar{r}_{k,i}\right]}{\bar{r}_{i,i}} + \Delta f_{i,j}^{(1)}$$

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

CHOLESKY QR DECOMPOSITION: COMPUTED TRIANGULAR FACTOR

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

$$A^T A + \Delta F_2 = \bar{R}^T \bar{R}!$$

$$\|\Delta F_2\| \leq c_1 u \|A\|^2$$

Cholesky factor of the cross-product matrix $A^T A$ is computed in a **backward stable** way!

CHOLESKY QR DECOMPOSITION: COMPUTED TRIANGULAR FACTOR

$$A^T A + \Delta F_2 = \bar{R}^T \bar{R} \ , \ \|\Delta F_2\| \leq c_1 u \|A\|^2$$

assuming $c_1 u \kappa^2(A) < 1$,

$$\|\bar{R}^{-1}\| \leq \frac{1}{\sigma_{min}(A)[1 - c_1 u \kappa^2(A)]^{1/2}}, \ \|\bar{R}\| \leq \|A\| \left[1 + c_1 u \kappa^2(A)\right]^{1/2}$$

$$A + \bar{R}\Delta F_1 = \bar{Q}\bar{R}, \ \|\bar{R}\Delta F_1\| \leq c_2 u \|A\| \kappa(A)!$$

CHOLESKY QR DECOMPOSITION: THE LOSS OF ORTHOGONALITY

$$A^T A + \Delta F_2 = \bar{R}^T \bar{R},\ \bar{Q} = A\bar{R}^{-1} + \Delta F_1$$

$$I - \bar{Q}^T \bar{Q} = -(\Delta F_1)^T A \bar{R}^{-1}$$
$$- \bar{R}^{-T} A^T \Delta F_1 - (\Delta F_1)^T \Delta F_1 + \bar{R}^{-T} \Delta F_2 \bar{R}^{-1}$$

assuming $c_1 u \kappa^2(A) < 1$

$$\|I - \bar{Q}^T \bar{Q}\| \leq \frac{c_3 u \kappa^2(A)}{1 - c_1 u \kappa^2(A)}$$

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 $A^T A = R^T R$

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

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

CLASSICAL GRAM-SCHMIDT PROCESS: COMPUTED TRIANGULAR FACTOR

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

$$A^T A + \Delta E_1 = \bar{R}^T \bar{R}!$$

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

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: COMPUTED TRIANGULAR FACTOR

$$A^T A + \Delta E_1 = \bar{R}^T \bar{R} \ , \ \|\Delta E_1\| \leq c_1 u \|A\|^2$$

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

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

$$\|\bar{R}^{-1}\| \leq \frac{1}{\sigma_{min}(A)[1-c_2 u\kappa(A)]^{1/2}}, \ \|\bar{R}\| \leq \|A\| \left[1 + c_1 u \kappa^2(A)\right]^{1/2}$$

A. Smoktunowicz, 2004

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

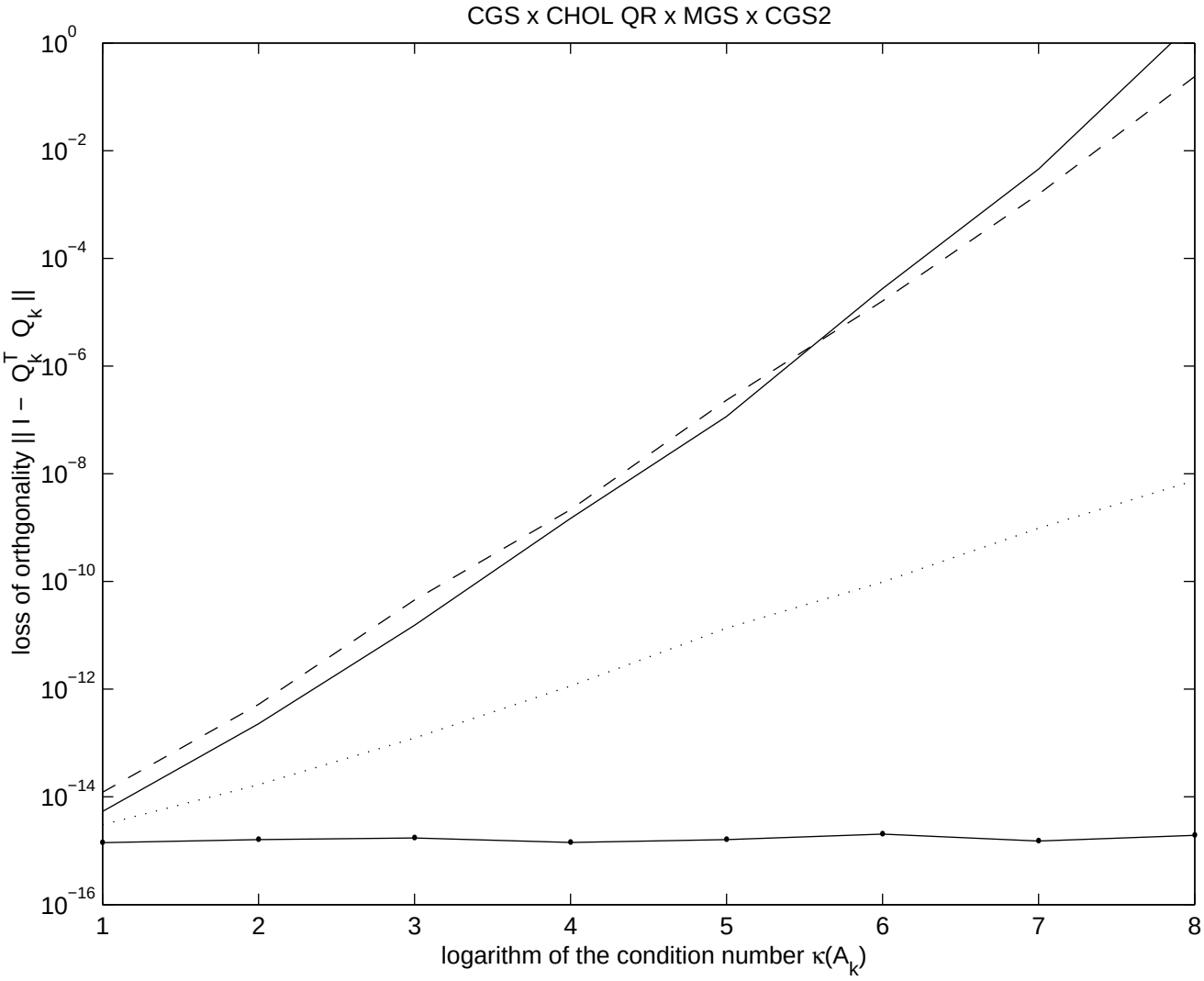
$$\|I - \bar{Q}^T \bar{Q}\| \leq \frac{\widehat{c}_2 u\kappa(A)}{1-\widehat{c}_1 u\kappa(A)}$$

Björck, 1967, Björck, Paige, 1992

- **classical Gram-Schmidt (CGS):** 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, 2004



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

LEAST SQUARES PROBLEM VIA CHOLESKY QR: SEMINORMAL EQUATIONS

$$\|b - Ax\| = \min_u \|b - Au\|$$

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

$$(\bar{R}^T \bar{R} + \Delta F_3)\bar{x} = A^T b + \Delta f_2$$

$$\|\Delta f_2\| \leq c_0 u \|A\| \|b\|, \ \|\Delta F_3\| \leq c_0 u \|\bar{R}\|^2$$

LEAST SQUARES PROBLEM VIA CHOLESKY QR: SEMINORMAL EQUATIONS

$$(\bar{R}^T\bar{R} + \Delta F_3)\bar{x} = A^T b + \Delta f_2$$

$$(A^T A + \Delta F_2 + \Delta F_3)\bar{x} = A^T b + \Delta f_2$$

$$(A^T A + \Delta F)\bar{x} = A^T b + \Delta f$$

$$\|\Delta F\| \leq c_4 u \|A\|^2, \ \|\Delta f\| \leq c_4 u \|A\|\|b\|$$

LEAST SQUARES PROBLEM: ERROR IN THE NORMAL EQUATIONS APPROACH

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

$$(A^T A + \Delta F)\bar{x} = A^T b + \Delta f$$

$$\frac{\|\bar{x}-x\|}{\|x\|} \leq \kappa^2(A) \frac{\frac{\|\Delta F\|}{\|A^T A\|} + \frac{\|\Delta f\|}{\|A^T b\|}}{1 - \kappa^2(A)\frac{\|\Delta F\|}{\|A^T A\|}} \sim \kappa^2(A) u$$

LEAST SQUARES PROBLEM WITH CLASSICAL GRAM-SCHMIDT

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

$$Rx = Q^T b,\ r = (I - QQ^T)b$$

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

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

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

LEAST SQUARES PROBLEM WITH CLASSICAL GRAM-SCHMIDT

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

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

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

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

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 almost the same forward error bound as the backward stable 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