NUMERICS OF THE GRAM-SCHMIDT PROCESS: FROM THE STANDARD INNER PRODUCT TO THE SR DECOMPOSITION

Miro Rozložník

joint work with Alicja Smoktunowicz, Felicja Okulicka-Dłużewska and Heike Fassbender

Czech Academy of Sciences, Prague, Czech Republic

MAT TRIAD 2017 - International Conference on MATRIX Analysis and its Applications Będlewo, Poland, September 25-29, 2017

Orthogonalization with respect to the standard inner product

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

orthogonal basis Q of $span(A)$:

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

$A = QR$, $R \in \mathcal{R}^{n,n}$ upper triangular,
factorization uniqueness: positive diagonal entries

$$\kappa(Q) = 1,\ \|R\| = \|A\|,\ \|R^{-1}\| = 1/\sigma_n(A),\ (\kappa(R) = \kappa(A))$$

$$C = A^T A = R^T R$$

CLASSICAL AND MODIFIED GRAM-SCHMIDT ALGORITHMS

- **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**

classical Gram-Schmidt process:

for $j = 1, \ldots, n$

$\quad u_j = a_j$

$\quad$ for $i = 1, \ldots, j-1$

$\quad\quad r_{i,j} = \langle a_j, q_i \rangle$

$\quad\quad u_j = u_j - r_{i,j} q_i$

$\quad r_{j,j} = \|u_j\|$

$\quad q_j = u_j / r_{j,j}$

modified Gram-Schmidt process:

for $j = 1, \ldots, n$

$\quad u_j = a_j$

$\quad$ for $i = 1, \ldots, j-1$

$\quad\quad r_{i,j} = \langle u_j, q_i \rangle$

$\quad\quad u_j = u_j - r_{i,j} q_i$

$\quad r_{j,j} = \|u_j\|$

$\quad q_j = u_j / r_{j,j}$

GRAM-SCHMIDT PROCESS VERSUS ROUNDING ERRORS

- **modified** Gram-Schmidt (MGS):
  assuming $O(u)\kappa(A) < 1$
  $$\|I - \bar{Q}^T\bar{Q}\| \leq \frac{O(u)\kappa(A)}{1-O(u)\kappa(A)}$$

Björck, 1967 , Björck, Paige, 1992

- **classical** Gram-Schmidt (CGS)?
  $$\|I - \bar{Q}^T\bar{Q}\| \leq \frac{O(u)\kappa^{n-1}(A)}{1-O(u)\kappa^{n-1}(A)}?$$

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:

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

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

Cholesky QR algorithm: the triangular factor computed as the Cholesky factor of the **cross-product** matrix C and the orthogonal vectors recovered from the inverse of the triangular factor as $Q = AR^{-1}$

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 $O(u)\kappa(A) < 1$

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

Giraud, van den Eshof, Langou, R, 2005
Barlow, Smoktunowicz, Langou, 2006

ITERATED GRAM SCHMIDT OR GRAM-SCHMIDT PROCESS WITH REORTHOGONALIZATION

- **Iterated** Gram-Schmidt algorithm: Gram-Schmidt process can be applied iteratively to improve the orthogonality between the computed vectors

- Gram-Schmidt with **reorthogonalization**: **"two-steps are enough"** to preserve the orthogonality to working accuracy

classical Gram-Schmidt:

for $j = 1, \ldots, n$

$u_j = a_j$

for $i = 1, \ldots, j-1$

$r_{i,j} = \langle a_j, q_i \rangle$

$u_j = u_j - r_{i,j} q_i$

$r_{j,j} = \sqrt{\|a_j\|^2 - \sum_{i=1}^{j-1} r_{i,j}^2}$

$q_j = u_j / r_{j,j}$

classical Gram-Schmidt with **reorthogonalization**:

for $j = 1, \ldots, n$

$u_j = a_j$

for $k = 1, 2$

$a_j^{(k)} = u_j$

for $i = 1, \ldots, j-1$

$r_{i,j}^{(k)} = \langle a_j^{(k)}, q_i \rangle$

$u_j = u_j - r_{i,j}^{(k)} q_i$

$r_{j,j} = \|u_j\|$

$q_j = u_j / r_{j,j}$

GRAM-SCHMIDT WITH THE REORTHOGONALIZATION

$$u_j = (I - Q_{j-1}Q_{j-1}^T)a_j,\; v_j = (I - Q_{j-1}Q_{j-1}^T)^2 a_j$$
$$\|u_j\| = |r_{j,j}| \geq \sigma_{min}(R_j) = \sigma_{min}(A_j) \geq \sigma_{min}(A)$$
$$\frac{\|a_j\|}{\|u_j\|} \leq \kappa(A),\; \frac{\|u_j\|}{\|v_j\|} = 1,\; Q_{j-1}^T\left(\frac{v_j}{\|v_j\|}\right) = 0$$
$$A + \Delta E_2 = \bar{Q}\bar{R},\; \|\Delta E_2\| \leq O(u)\|A\|$$

$$\frac{\|a_j\|}{\|\bar{u}_j\|} \leq \frac{\kappa(A)}{1 - O(u)\kappa(A)},\; \frac{\|\bar{u}_j\|}{\|\bar{v}_j\|} \leq \frac{1}{1 - O(u)\kappa(A)},\; \frac{\|\bar{Q}_{j-1}^T \bar{v}_j\|}{\|\bar{v}_j\|} \leq ?$$

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

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

Giraud, van den Eshof, Langou, R, 2005

STANDARD INNER PRODUCT: ROUNDING ERRORS

- **modified** Gram-Schmidt:
  assuming $\mathcal{O}(u)\kappa(A) < 1$
  $\|I - \bar{Q}^T\bar{Q}\| \leq \frac{\mathcal{O}(u)\kappa(A)}{1-\mathcal{O}(u)\kappa(A)}$

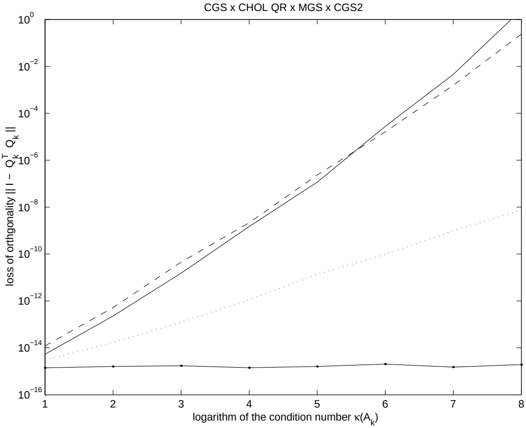
  Björck, 1967, Björck, Paige, 1992

- **classical** Gram-Schmidt:
  assuming $\mathcal{O}(u)\kappa(A) < 1$
  $\|I - \bar{Q}^T\bar{Q}\| \leq \frac{\mathcal{O}(u)\kappa^2(A)}{1-\mathcal{O}(u)\kappa(A)}$

  Giraud, van den Eshof, Langou, R, 2005
  Barlow, Smoktunowicz, Langou, 2006

- classical or modified Gram-Schmidt with **reorthogonalization**:
  assuming $\mathcal{O}(u)\kappa(A) < 1$
  $\|I - \bar{Q}^T\bar{Q}\| \leq \mathcal{O}(u)$

  Giraud, van den Eshof, Langou, R, 2005
  Barlow, Smoktunowicz, 2011



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

Orthogonalization with respect to a non-standard inner product

$$B \in \mathcal{R}^{m,m} \text{ symmetric positive definite, inner product } \langle \cdot, \cdot \rangle_B$$

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

B-orthonormal basis of $span(A)$:

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

$$A = QR,\ R \in \mathcal{R}^{n,n} \text{ upper triangular with positive diagonal entries}$$

$$B^{1/2}A = (B^{1/2}Q)R,\ \|B^{1/2}Q\| = \sigma_n(B^{1/2}Q) = 1\ (\kappa(Q) \leq \kappa^{1/2}(B))$$

$$\|R\| = \|B^{1/2}A\|,\ \|R^{-1}\| = 1/\sigma_n(B^{1/2}A)\ (\kappa(R) = \kappa(B^{1/2}A)$$

$$C = A^T B A = R^T R$$

classical Gram-Schmidt:

for $j = 1, \ldots, n$

$u_j = a_j$

for $i = 1, \ldots, j-1$

$r_{i,j} = \langle a_j, q_i \rangle_B$

$u_j = u_j - r_{i,j} q_i$

$r_{j,j} = \sqrt{\|a_j\|_B^2 - \sum_{i=1}^{j-1} r_{i,j}^2}$

$q_j = u_j / r_{j,j}$

classical Gram-Schmidt with **reorthogonalization**:

for $j = 1, \ldots, n$

$u_j = a_j$

for $k = 1, 2$

$a_j^{(k)} = u_j$

for $i = 1, \ldots, j-1$

$r_{i,j}^{(k)} = \langle a_j^{(k)}, q_i \rangle_B$

$u_j = u_j - r_{i,j}^{(k)} q_i$

$r_{j,j} = \|u_j\|_B$

$q_j = u_j / r_{j,j}$

LOSS OF B-ORTHOGONALITY IN GRAM-SCHMIDT

modified Gram-Schmidt:

$$\mathcal{O}(u)\kappa(B)\kappa(B^{1/2}A) < 1$$

$$\|I - \bar{Q}^T B\bar{Q}\| \leq \frac{\mathcal{O}(u)\|B\|\|\bar{Q}\|^2\kappa(B^{1/2}A)}{1 - \mathcal{O}(u)\|B\|\|\bar{Q}\|^2\kappa(B^{1/2}A)}$$

classical Gram-Schmidt and AINV algorithm:

$$\mathcal{O}(u)\kappa(B)\kappa(B^{1/2}A)\kappa(A) < 1$$

$$\|I - \bar{Q}^T B\bar{Q}\| \leq \frac{\mathcal{O}(u)\|B\|^{1/2}\|\bar{Q}\|\kappa(B^{1/2}A)\kappa^{1/2}(B)\kappa(A)}{1 - \mathcal{O}(u)\|B\|^{1/2}\|\bar{Q}\|\kappa(B^{1/2}A)\kappa^{1/2}(B)\kappa(A)}$$

classical Gram-Schmidt with reorthogonalization:

$$\mathcal{O}(u)\kappa^{1/2}(B)\kappa(B^{1/2}A) < 1$$

$$\|I - \bar{Q}^T B\bar{Q}\| \leq \mathcal{O}(u)\|B\|\|\bar{Q}\|\|\bar{Q}^{(1)}\|$$

THE LOCAL ERRORS IN A NON-STANDARD INNER PRODUCTS

general positive definite B :

$$|\mathrm{fl}[\langle \bar{u}_i, \bar{q}_j \rangle_B] - \langle \bar{u}_i, \bar{q}_j \rangle_B| \leq \mathcal{O}(u) \|B\| \|\bar{u}_i\| \|\bar{q}_j\|$$
$$|1 - \|\bar{q}_j\|_B^2| \leq \mathcal{O}(u) \|B\| \|\bar{q}_j\|^2$$

diagonal positive (weight matrix) B :

$$|\mathrm{fl}[\langle \bar{u}_i, \bar{q}_j \rangle_B] - \langle \bar{u}_i, \bar{q}_j \rangle_B| \leq \mathcal{O}(u) \|\bar{u}_i\|_B \|\bar{q}_j\|_B$$
$$|1 - \|\bar{q}_j\|_B^2| \leq \mathcal{O}(u)$$

DIAGONAL CASE IS SIMILAR TO STANDARD CASE

modified Gram-Schmidt:

$$\mathcal{O}(u)\kappa(B^{1/2}A) < 1$$

$$\|I - \bar{Q}^T B\bar{Q}\| \leq \frac{\mathcal{O}(u)\kappa(B^{1/2}A)}{1 - \mathcal{O}(u)\kappa(B^{1/2}A)}$$

classical Gram-Schmidt and AINV algorithm

$$\mathcal{O}(u)\kappa^2(B^{1/2}A) < 1$$

$$\|I - \bar{Q}^T B\bar{Q}\| \leq \frac{\mathcal{O}(u)\kappa^2(B^{1/2}A)}{1 - \mathcal{O}(u)\kappa^2(B^{1/2}A)}$$

classical Gram-Schmidt with reorthogonalization:

$$\mathcal{O}(u)\kappa(B^{1/2}A) < 1$$

$$\|I - \bar{Q}^T B\bar{Q}\| \leq \mathcal{O}(u)$$

Gulliksson, Wedin 1992, Gulliksson 1995

Orthogonalization with respect to a symmetric bilinear form

$B \in \mathcal{R}^{m,m}$ symmetric indefinite and nonsingular

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

B-orthonormal basis of $span(A)$:

$$Q = (q_1, \ldots, q_n) \in \mathcal{R}^{m,n},\ Q^T B Q = \Omega \in \mathrm{diag}(\pm 1)$$

$A = QR$, $R \in \mathcal{R}^{n,n}$ upper triangular with positive diagonal

if no principal minor of C vanishes (if C is strongly nonsingular)

$$C = A^T B A = R^T \Omega R$$

Bunch 1971, Bunch-Parlett 1971
Della Dora 1975, Elsner 1979, Bunse-Gerstner 1981
Slapnicar 1999, Singer and Singer 2000, Singer 2006

classical Gram-Schmidt:

for $j = 1, \dots, n$

$u_j = a_j$

for $i = 1, \dots, j-1$

$r_{i,j} = \omega_i \langle B a_j, q_i \rangle$

$u_j = u_j - r_{i,j} q_i$

$\omega_j = \text{sign}\left[\langle B a_j, a_j \rangle - \sum_{i=1}^{j-1} \omega_i r_{i,j}^2\right], \; r_{j,j} = \sqrt{\left|\langle B a_j, a_j \rangle - \sum_{i=1}^{j-1} \omega_i r_{i,j}^2\right|}$

$q_j = u_j / r_{j,j}$

classical Gram-Schmidt with **reorthogonalization**:

for $j = 1, \dots, n$

$u_j = a_j$

for $k = 1, 2$

$a_j^{(k)} = u_j$

for $i = 1, \dots, j-1$

$r_{i,j}^{(k)} = \omega_i \langle B a_j^{(k)}, q_i \rangle$

$u_j = u_j - r_{i,j}^{(k)} q_i$

$\omega_j = \text{sign}[\langle B u_j, u_j \rangle], \; r_{j,j} = \sqrt{|\langle B u_j, u_j \rangle|}$

$q_j = u_j / r_{j,j}$

Conditioning of the factors R and Q

$$\|R_j^{-1}\|^2 \leq \|C_j^{-1}\| + 2 \sum_{i=1,\ldots,j-1;\ \omega_{i+1}\neq\omega_i} \|C_i^{-1}\|$$

$$C = R^T \Omega R \Rightarrow \|R\| \leq \|C\|\|R^{-1}\|$$

$$\kappa(R) \leq \|C\| \left(\|C^{-1}\| + 2 \sum_{j;\ \omega_{j+1}\neq\omega_j} \|C_j^{-1}\|\right)$$

$$\|Q\| \leq \|A\|\|R^{-1}\|, \quad \sigma_{min}(Q) \geq \frac{\sigma_{min}(A)}{\|R\|}$$

$$\kappa(Q) \leq \kappa(A)\kappa(R)$$

R, Okulicka-Dluzewska, Smoktunowicz, 2015

N. Higham, J-orthogonal matrices, SIAM Review 2003

Example with well-conditioned principal submatrix

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \; B = \begin{pmatrix} 1 & \sqrt{\varepsilon} \\ \sqrt{\varepsilon} & -\varepsilon \end{pmatrix}$$

$$Q = R^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & \frac{1}{\sqrt{\varepsilon}} \end{pmatrix}, \quad R = Q^{-1} = \begin{pmatrix} 1 & \sqrt{\varepsilon} \\ 0 & \sqrt{\varepsilon} \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$\|B\| \approx 1 + \varepsilon$ and $\sigma_{min}(B) = 2\varepsilon$

$$\|R\| \approx \sqrt{1+\varepsilon}, \;\; \sigma_{min}(R) \approx \sqrt{\varepsilon}, \;\; \kappa(R) = \kappa(Q) \approx \frac{1}{\sqrt{\varepsilon}}$$

Example with ill-conditioned principal submatrix

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ B = \begin{pmatrix} \varepsilon & 1 \\ 1 & -\varepsilon \end{pmatrix}$$

$$Q = R^{-1} = \begin{pmatrix} \frac{1}{\sqrt{\varepsilon}} & -\frac{1}{\sqrt{\varepsilon(1+\varepsilon^2)}} \\ 0 & \frac{\sqrt{\varepsilon}}{\sqrt{1+\varepsilon^2}} \end{pmatrix}, \quad R = Q^{-1} = \begin{pmatrix} \sqrt{\varepsilon} & \frac{1}{\sqrt{\varepsilon}} \\ 0 & \frac{\sqrt{1+\varepsilon^2}}{\sqrt{\varepsilon}} \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\|B\| = \sigma_{min}(B) = \sqrt{1+\varepsilon^2}$$

$$\|R\| \approx \frac{\sqrt{2}}{\sqrt{\varepsilon}},\ \sigma_{min}(R) \approx \frac{\sqrt{\varepsilon}}{\sqrt{2}}, \quad \kappa(R) = \kappa(Q) \approx \frac{2}{\varepsilon}$$

Classical Gram-Schmidt computes a Cholesky-like factor of C

Cholesky-like factorization:

assuming $\mathcal{O}(u)\|A\|^2\|B\|(\|C^{-1}\| + \max_{j,\ \bar{\omega}_{j+1}\neq\bar{\omega}_j} \|C_j^{-1}\|) < 1$

$$C + \Delta C = \bar{R}^T \bar{\Omega} \bar{R},$$
$$\|\Delta C\| \leq \mathcal{O}(u)[\|\bar{R}\|^2 + \|B\|\|A\|^2]$$

Bunch 1971, Bunch-Parlett 1971
Slapnicar, 1999

Classical Gram-Schmidt (B-CGS) process :

$$C + \Delta C = \bar{R}^T \bar{\Omega} \bar{R},$$
$$\|\Delta C\| \leq \mathcal{O}(u)[\|\bar{R}\|^2 + \|B\|\|A\|\|\bar{Q}\|\|\bar{R}\| + \|B\|\|A\|^2]$$

The loss of B-orthogonality between computed vectors

Cholesky-like B-QR factorization: $\bar{Q} = \mathrm{fl}(A\bar{R}^{-1})$

$$\|\bar{Q}^T B\bar{Q} - \bar{\Omega}\| \le \mathcal{O}(u)[\kappa^2(\bar{R}) + \|\bar{R}^{-1}\|^2\|A\|^2\|B\| + 2\|B\bar{Q}\|\|\bar{Q}\|\kappa(\bar{R})]$$

Classical Gram-Schmidt (B-CGS) process :

$$\|\bar{Q}^T B\bar{Q} - \bar{\Omega}\| \le$$
$$\mathcal{O}(u)\left[\kappa^2(\bar{R}) + \|\bar{R}^{-1}\|^2\|A\|^2\|B\| + 3\|BA\|\|\bar{R}^{-1}\|\|\bar{Q}\|\kappa(\bar{R})\right]$$

The loss of B-orthogonality between computed vectors

CGS with reorthogonalization (B-CGS2):

$\mathcal{O}(u)\|A\|^2\|B\|\|C\|(\|C^{-1}\| + \max_{j,\ \bar{\omega}_{j+1}\neq\bar{\omega}_j} \|C_j^{-1}\|)^2 < 1$

$$\|\bar{Q}^T B\bar{Q} - \bar{\Omega}\| \leq \mathcal{O}(u)\|B\|\|\bar{Q}\|^2$$

Numerical experiments - model examples

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} = \begin{pmatrix} R_{11}^T & 0 \\ R_{12}^T & R_{22}^T \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix},$$

1. $\kappa(C_{11}) = 100 \ll \kappa(C) \approx 10^{2i}$, $\kappa(C_{12}) = 10^i$ for $i = 0, \ldots, 8$; $C_{22} = 0$ ($\|C_{11}\| = \|C_{12}\| = 1$)

2. $\kappa(C_{11}) = 10^i \gg \kappa(C) = 1$ for $i = 0, \ldots, 16$; $C_{11}^2 + C_{12}^2 = I$ $C_{22} = -C_{11}$ ($\|C_{11}\| = 1/2$)

The spectral properties of computed factors with respect to the conditioning of the submatrix C_{12} for Problem 1.

$\|C_{12}^{-1}\|$	$\|C^{-1}\|$	$\|S_{22}\|$	$\|\bar{R}\| = \|\bar{Q}^{-1}\|$	$\|\bar{R}^{-1}\| = \|\bar{Q}\|$
10^0	1.6180e+00	1.0000e+02	1.4142e+01	1.4142e+01
10^1	1.0099e+02	1.0000e+02	1.4142e+01	1.4142e+01
10^2	1.0001e+04	1.0000e+02	1.4142e+01	1.0001e+02
10^3	1.0000e+06	1.0000e+02	1.4142e+01	1.0000e+03
10^4	1.0000e+08	1.0000e+02	1.4142e+01	1.0000e+04
10^5	1.0000e+10	1.0000e+02	1.4142e+01	1.0000e+05
10^6	1.0000e+12	1.0000e+02	1.4142e+01	1.0000e+06
10^7	9.9808e+13	1.0000e+02	1.4142e+01	1.0000e+07
10^8	1.8925e+16	1.0000e+02	1.4142e+01	1.0000e+08

The loss of B-orthogonality $\|\bar{\Omega} - \bar{Q}^T B \bar{Q}\|$ with respect to the conditioning of the submatrix C_{12} for Problem 1.

$\|C_{12}^{-1}\|$	Cholesky B-QR	Cholesky B-QR2	B-CGS	B-CGS2
10^0	6.9767e-15	3.1373e-15	4.5838e-15	3.1956e-15
10^1	8.5940e-14	6.6516e-15	5.1740e-14	7.1550e-15
10^2	1.8989e-12	5.6400e-14	4.4021e-12	5.1951e-14
10^3	4.8268e-10	3.2421e-13	1.5760e-10	4.4188e-13
10^4	2.9594e-08	4.9631e-12	1.1656e-08	2.6936e-12
10^5	1.5621e-06	3.7820e-11	1.8274e-06	2.9007e-11
10^6	2.4082e-05	2.0335e-10	2.3673e-04	2.8010e-10
10^7	3.7036e-02	2.5207e-09	9.6352e-03	2.9913e-09
10^8	6.5241e-01	2.0603e-08	4.1306e-01	2.4907e-08

The spectral properties of computed factors with respect to the conditioning of the submatrix C_{11} for Problem 2.

$\|C_{11}^{-1}\|$	$\|C^{-1}\|$	$\|S_{22}\|$	$\|\bar{R}\| = \|\bar{Q}^{-1}\|$	$\|\bar{R}^{-1}\| = \|\bar{Q}\|$
10^0	1.0000e+00	2.0000e+00	1.9319e+00	1.9319e+00
10^1	1.0000e+00	2.0000e+01	6.3226e+00	6.3226e+00
10^2	1.0000e+00	2.0000e+02	2.0000e+01	2.0000e+01
10^3	1.0000e+00	2.0000e+03	6.3246e+01	6.3246e+01
10^4	1.0000e+00	2.0000e+04	2.0000e+02	2.0000e+02
10^5	1.0000e+00	2.0000e+05	6.3246e+02	6.3246e+02
10^6	1.0000e+00	2.0000e+06	2.0000e+03	2.0000e+03
10^7	1.0000e+00	2.0000e+07	6.3246e+03	6.3246e+03
10^8	1.0000e+00	2.0000e+08	2.0000e+04	2.0000e+04
10^9	1.0000e+00	2.0000e+09	6.3246e+04	6.3246e+04
10^{10}	1.0000e+00	2.0000e+10	2.0000e+05	2.0000e+05
10^{11}	1.0000e+00	2.0000e+11	6.3246e+05	6.3246e+05
10^{12}	1.0000e+00	2.0000e+12	2.0000e+06	2.0000e+06
10^{13}	1.0000e+00	1.9999e+13	6.3245e+06	6.3245e+06
10^{14}	1.0000e+00	2.0004e+14	2.0188e+07	2.0520e+07
10^{15}	1.0000e+00	2.0011e+15	6.6349e+07	5.2040e+07

The loss of B-orthogonality $\|\bar{\Omega} - \bar{Q}^T B\bar{Q}\|$ with respect to the conditioning of the principal submatrix C_{11} for Problem 2.

$\|C_{11}^{-1}\|$	Cholesky B-QR	Cholesky B-QR2	B-CGS	B-CGS2
10^0	5.0322e-16	3.2067e-16	5.3413e-16	3.9373e-16
10^1	1.2883e-15	8.7715e-16	1.5521e-15	1.2610e-15
10^2	4.5583e-15	3.5957e-15	4.6097e-15	3.2657e-15
10^3	1.9874e-14	1.6704e-14	2.6765e-14	2.2026e-14
10^4	1.5159e-13	1.2480e-13	1.4222e-13	1.3054e-13
10^5	1.0447e-12	8.1751e-13	1.1241e-12	1.2374e-12
10^6	1.0511e-11	7.1311e-12	1.6597e-11	6.4763e-12
10^7	5.8440e-11	5.0812e-11	2.1037e-10	5.1101e-11
10^8	3.5174e-10	2.3857e-10	6.4724e-10	5.8383e-10
10^9	5.6336e-09	4.7359e-09	8.5080e-09	3.2390e-09
10^{10}	6.4206e-08	4.7271e-08	1.8162e-07	4.7073e-08
10^{11}	3.3127e-07	2.8293e-07	1.0061e-06	4.2164e-07
10^{12}	3.4508e-06	2.6920e-06	7.6409e-06	6.0936e-06
10^{13}	2.2361e-05	5.5208e-05	1.3357e-04	4.7861e-03
10^{14}	5.4077e-04	3.6470e-04	6.8111e-04	2.1676e+00
10^{15}	5.4339e-03	2.9211e-03	1.0174e-02	4.1463e+00

Orthogonalization with respect to a skew-symmetric bilinear form

$$A = (a_1, \ldots, a_{2n}) \in \mathcal{R}^{2m,2n},\ m \geq n = rank(A)/2$$

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \in \mathcal{R}^{2m,2m} \text{ skew-symmetric and orthogonal}$$

J-orthonormal basis of $span(A)$: $Q = (q_1, \ldots, q_{2n}) \in \mathcal{R}^{2m,2n}$

$$Q^T J Q = \mathrm{diag}(\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}) \in \mathcal{R}^{2n,2n}$$

$A = QR$, $R \in \mathcal{R}^{n,n}$ upper triangular with positive diagonal

if no minor of C with even dimension vanishes

$$C = A^T J A = R^T \mathrm{diag}(\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}) R$$

Della Dora 1975, Elsner 1979, Bunse-Gerstner 1981
Mehrmann 1979, Bunse-Gerstner and Mehrmann 1986
Benner, Byers, Fassbender, Mehrmann, Watkins 2000

Orthogonalization with respect to a skew-symmetric bilinear form

classical Gram-Schmidt (CGS)

for $j = 1, \ldots, n$

$[u_{2j-1}, u_{2j}] = [a_{2j-1}, a_{2j}]$

for $i = 1, \ldots, j-1$

$$[u_{2j-1}, u_{2j}] = [u_{2j-1}, u_{2j}] - \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{-1} [q_{2i-1}, q_{2i}]^T J [a_{2j-1}, a_{2j}]$$

$$\begin{pmatrix} r_{11} & 0 \\ r_{12} & r_{22} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix} = [u_{2j-1}, u_{2j}]^T J [u_{2j-1}, u_{2j}]$$

$$[q_{2j-1}, q_{2j}] = [u_{2j-1}, u_{2j}] \begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix}^{-1}$$

Uniqueness of the Cholesky-like factorization?

$$C = \begin{pmatrix} 0 & \pm\|C\| \\ \mp\|C\| & 0 \end{pmatrix} = R^T J R$$

$$= \begin{pmatrix} r_{11} & 0 \\ r_{12} & r_{22} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix} = \begin{pmatrix} 0 & r_{11}r_{22} \\ -r_{11}r_{22} & 0 \end{pmatrix}$$

How to compute the (normalization) factor $R = \begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix}$?

Mehrmann 1979, Bunse-Gerstner and Mehrmann 1986
Fassbender 2000, Benner 2003
Salam 2005
Ferng, Lin, Wang 1997
Bhatia 1994, Chang, 1998

Local minimization of the condition number of R

$$\kappa^2(R) = \frac{\|R\|_F^2 + \sqrt{\|R\|_F^4 - 4r_{11}^2 r_{22}^2}}{\|R\|_F^2 - \sqrt{\|R\|_F^4 - 4r_{11}^2 r_{22}^2}}$$

As $r_{11}r_{22} = d$ is fixed and $\kappa(R)$ is an increasing function of $\|R\|_F$, it is minimized if $r_{12} = 0$ and $|r_{11}| = |r_{22}|$. Then

$$R^T R = |d| \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \kappa(R) = 1$$

.

Mehrmann 1979, Bunse-Gerstner and Mehrmann 1986

Local minimization of the condition number of Q

$$Q = AR^{-1},\ \kappa^2(Q) = \frac{\|Q\|_F^2 + \sqrt{\|Q\|_F^4 - 4\frac{(\|A\|\sigma_{min}(A))^2}{d^2}}}{\|Q\|_F^2 - \sqrt{\|Q\|_F^4 - 4\frac{(\|A\|\sigma_{min}(A))^2}{d^2}}}$$

As $\|A\|\sigma_{min}(A)/d$ is fixed and $\kappa(Q)$ is an increasing function of $\|Q\|_F$, it is minimized if r_{12} is chosen so that $q_1 \perp q_2$ with $\|q_1\| = \|q_2\|$. Then

$$Q^T Q = \frac{\|A\|\sigma_{min}(A)}{|d|} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \kappa(Q) = 1$$

.

Fassbender, R 2016

Conditioning of the factors R and Q

$$\|R^{-1}\|^2 \leq \|C^{-1}\| + \sqrt{2} \sum_{k=1}^{n-1} (\|C_{2(k-1)}^{-1} C_{2(k-1),k}\| + 1)^2 \|R_{k,k}^{-1}\|$$

$$\|R\| \leq \|C\| \|R^{-1}\|$$

$$\|Q\| \leq \|A\| \|R^{-1}\|, \quad \sigma_{min}(Q) \geq \frac{\sigma_{min}(A)}{\|R\|}$$

$$\kappa(Q) \leq \kappa(A) \kappa(R)$$

Xu 2003

Example with $\kappa(R) \gg \kappa(A) \approx \kappa(C) \approx 1$

$$A = \begin{pmatrix} \sqrt{\varepsilon} & 1 & 0 & 0 \\ 1 & 0 & 0 & -\varepsilon \\ 0 & \sqrt{\varepsilon} & 0 & 1 \\ 0 & 0 & 1 & -\sqrt{\varepsilon} \end{pmatrix}, \; C = \begin{pmatrix} 0 & \varepsilon & 1 & 0 \\ -\varepsilon & 0 & 0 & 1 \\ -1 & 0 & 0 & \varepsilon \\ 0 & -1 & -\varepsilon & 0 \end{pmatrix}$$

$$Q = \begin{pmatrix} 1 & 0 & \frac{1}{\sqrt{\varepsilon}\sqrt{1-\varepsilon^2}} & -\frac{1}{\sqrt{1-\varepsilon^2}} \\ \frac{1}{\sqrt{\varepsilon}} & 0 & 0 & -\frac{\sqrt{1-\varepsilon^2}}{\sqrt{\varepsilon}} \\ 0 & 0 & -\frac{1}{\sqrt{1-\varepsilon^2}} & -\frac{\sqrt{\varepsilon}}{\sqrt{1-\varepsilon^2}} \\ 0 & \frac{1}{\sqrt{\varepsilon}} & \frac{\sqrt{\varepsilon}}{\sqrt{1-\varepsilon^2}} & \frac{\varepsilon}{\sqrt{1-\varepsilon^2}} \end{pmatrix},$$

$$R = \begin{pmatrix} \sqrt{\varepsilon} & 0 & 0 & -\frac{1}{\sqrt{\varepsilon}} \\ 0 & \sqrt{\varepsilon} & \frac{1}{\sqrt{\varepsilon}} & 0 \\ 0 & 0 & \frac{\sqrt{1-\varepsilon^2}}{\sqrt{\varepsilon}} & 0 \\ 0 & 0 & 0 & -\frac{\sqrt{1-\varepsilon^2}}{\sqrt{\varepsilon}} \end{pmatrix}$$

$$\sigma(A) \approx 1, \; \kappa(A) \approx 1,$$

$$\sigma(R) \approx \frac{\sqrt{2}}{\sqrt{\varepsilon}}, \frac{\sqrt{\varepsilon}}{\sqrt{2}}, \; \kappa(R) \approx \frac{2}{\varepsilon}, \; \sigma(Q) \approx \frac{\sqrt{2}}{\sqrt{\varepsilon}}, \frac{\sqrt{\varepsilon}}{\sqrt{2}}$$

Conditioning of factors in the SR decomposition

C skew-symmetric, Bunch decomposition of C

$L \in \mathcal{R}^{2n,2n}$ block unit lower triangular, $D \in \mathcal{R}^{2n,2n}$ block diagonal

$$C = LDL^T, \quad D = \begin{pmatrix} d_1 \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} & & \\ & \ddots & \\ & & d_n \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \end{pmatrix}$$

$$U \in \mathcal{R}^{2n,2n},\ U = AL^{-T},\ U^T JU = D$$

$$R = \operatorname{diag}(R_{1,1}, \ldots, R_{n,n})L^T, \quad Q = U\operatorname{diag}(R_{1,1}^{-1}, \ldots, R_{n,n}^{-1})$$

Bunch 1982, Benner, Byers, Fassbender, Mehrmann, Watkins 2000, Singer, Singer 2003

Towards the global minimization of the condition number of R

$$L^T = \begin{pmatrix} I & \dots & L_{n,1}^T \\ & \ddots & \vdots \\ & & I \end{pmatrix} = \begin{pmatrix} L_1 \\ \vdots \\ L_n \end{pmatrix}, \quad L_n = \begin{pmatrix} \ell_{n,1} \\ \ell_{n,2} \end{pmatrix}.$$

1. For each n minimize $\|R_{n,n}L_n\|_F^2$ subject to $r_{11}r_{22} = d_n$:
$\|R_{n,n}L_n\|_F^2 = 2|d_n|\sqrt{\|\ell_{n,1}\|^2\|\ell_{n,2}\|^2 - (\ell_{n,1}, \ell_{n,2})^2} = 2\beta_n^2$
and $\| \begin{pmatrix} 1 & 0 \end{pmatrix} R_{n,n}L_n\| = \| \begin{pmatrix} 0 & 1 \end{pmatrix} R_{n,n}L_n\| = \beta_n$.
2. Set $\beta = \max_n \beta_n$. For each n compute $R_{n,n}$ so that
$\|R_{n,n}L_n\|_F^2 = 2\beta^2$ and $r_{11}r_{22} = d_n$.

We can find a block scaling such that all rows of the matrix $R = \mathrm{diag}(R_{1,1}, \dots, R_{n,n})L^T$ have the same norm equal to β.
Optimality of block scaling? $2n\beta^2 = \|R_{n,n}^R L_n\|_F^2 \leq \|R_{n,n}L_n\|_F^2$

Van der Sluis 1969, Shapiro 1982, 1985

Towards the global minimization of the condition number of Q

$$U = (U_1, \dots, U_n)\,, \quad U_n = (u_{n,1}, u_{n,2})\,.$$

1. For each n minimize $\|U_n R_{n,n}^{-1}\|_F^2$ subject to $r_{11}r_{22} = d_n$:
$\|U_n R_{n,n}^{-1}\|_F^2 = 2\frac{\sqrt{\|u_{n,1}\|^2\|u_{n,2}\|^2 - (u_{n,1},u_{n,2})^2}}{|d_n|} = 2\beta_n^2$
and $\|U_n R_{n,n}^{-1}\begin{pmatrix}1\\0\end{pmatrix}\| = \|U_n R_{n,n}^{-1}\begin{pmatrix}0\\1\end{pmatrix}\| = \beta_n$.
2. Set $\beta = \max_n \beta_n$. For each n compute $R_{n,n}$ so that
$\|U_n R_{n,n}^{-1}\|_F^2 = 2\beta^2$ and $r_{11}r_{22} = d_n$

We can find a block scaling such that all collumns of the matrix $Q = U\mathrm{diag}(R_{1,1}^{-1}, \dots, R_{n,n}^{-1})$ have the same norm equal to β.
Optimality? $2n\beta^2 = \|AL_n^{-T}(R_{n,n}^C)^{-1}\|_F^2 \le \|AL_n^{-T}(R_{n,n})^{-1}\|_F^2$

Van der Sluis 1969, Shapiro 1982, 1985

Dopico, Johnson 2009

Example with $\kappa(A) \approx \kappa(C)$

$$C = \begin{pmatrix} 0 & \varepsilon & 0 & 1 \\ -\varepsilon & 0 & -1 & 0 \\ 0 & 1 & 0 & 1+\frac{1}{\varepsilon} \\ -1 & 0 & -(1+\frac{1}{\varepsilon}) & 0 \end{pmatrix} = LDL^T$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{1}{\varepsilon} & 0 & 1 & 0 \\ 0 & \frac{1}{\varepsilon} & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & \varepsilon & 0 & 0 \\ -\varepsilon & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & \frac{1}{\varepsilon} & 0 \\ 0 & 1 & 0 & \frac{1}{\varepsilon} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$A = \begin{pmatrix} 0 & 1 & 0 & 1+\frac{1}{\varepsilon} \\ 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & \varepsilon & 0 & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & -\frac{1}{\varepsilon} & 0 \\ 0 & 0 & -1 & 0 \\ 0 & \varepsilon & 0 & 0 \end{pmatrix}$$

$$\|C\| \approx \frac{1}{\varepsilon},\ \|C^{-1}\| \approx \frac{1}{\varepsilon^2},\ \kappa(C) \approx \frac{1}{\varepsilon^3}.$$
$$\|A\| \approx \frac{1}{\varepsilon},\ \|A^{-1}\| \approx \frac{1}{\varepsilon^2},\ \kappa(A) \approx \frac{1}{\varepsilon^3}.$$
$$\|U\| \approx \frac{1}{\varepsilon},\ \|U^{-1}\| \approx \frac{1}{\varepsilon},\ \kappa(U) \approx \frac{1}{\varepsilon^2}$$

Example: triangular factor local minimization vs. equilibration

$$R_1 = \begin{pmatrix} \sqrt{\varepsilon} & 0 & 0 & 0 \\ 0 & \sqrt{\varepsilon} & & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \frac{1}{\varepsilon} & 0 \\ 0 & 1 & 0 & \frac{1}{\varepsilon} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \sqrt{\varepsilon} & 0 & \frac{1}{\sqrt{\varepsilon}} & 0 \\ 0 & \sqrt{\varepsilon} & 0 & \frac{1}{\sqrt{\varepsilon}} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\|R_1\| \approx \frac{1}{\sqrt{\varepsilon}},\ \|R_1^{-1}\| \approx \frac{1}{\varepsilon},\ \kappa(R_1) \approx \frac{1}{\varepsilon\sqrt{\varepsilon}}$$

$$\begin{aligned} R_2 &= \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & \frac{\sqrt{\varepsilon}}{\sqrt{1+\varepsilon^2}} & \frac{\sqrt{(1+\varepsilon^2)^2-\varepsilon^2}}{\sqrt{\varepsilon}\sqrt{1+\varepsilon^2}} \\ 0 & 0 & 0 & \frac{\sqrt{1+\varepsilon^2}}{\sqrt{\varepsilon}} \end{pmatrix} \begin{pmatrix} 1 & 0 & \frac{1}{\varepsilon} & 0 \\ 0 & 1 & 0 & \frac{1}{\varepsilon} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &= \frac{\sqrt{1+\varepsilon^2}}{\sqrt{\varepsilon}} \begin{pmatrix} \frac{\varepsilon}{\sqrt{1+\varepsilon^2}} & 0 & \sqrt{1+\varepsilon^2} & 0 \\ 0 & \frac{\varepsilon}{\sqrt{1+\varepsilon^2}} & 0 & \sqrt{1+\varepsilon^2} \\ 0 & 0 & \frac{\varepsilon}{1+\varepsilon^2} & \frac{\sqrt{(1+\varepsilon^2)^2-\varepsilon^2}}{1+\varepsilon^2} \\ 0 & 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

$$\|R_2\| \approx \frac{\sqrt{1+\varepsilon^2}}{\sqrt{\varepsilon}},\ \|R_2^{-1}\| \approx \frac{\sqrt{\varepsilon}}{\sqrt{1+\varepsilon^2}}\frac{1}{\varepsilon},\ \kappa(R_2) \approx \frac{1}{\varepsilon}.$$

Example: semi-symplectic factor local minimization vs. equilibration

$$Q_1 = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & -\frac{1}{\varepsilon} & 0 \\ 0 & 0 & -1 & 0 \\ 0 & \varepsilon & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{\varepsilon}} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{\varepsilon}} & & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{\sqrt{\varepsilon}} & 0 & 1 \\ \frac{1}{\sqrt{\varepsilon}} & 0 & -\frac{1}{\varepsilon} & 0 \\ 0 & 0 & -1 & 0 \\ 0 & \sqrt{\varepsilon} & 0 & 0 \end{pmatrix}$$

$$\|Q_1\| \approx \tfrac{1}{\varepsilon},\ \|Q_1^{-1}\| \approx \tfrac{1}{\varepsilon},\ \kappa(Q_1) \approx \tfrac{1}{\varepsilon^2}$$

$$Q_2 = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & -\frac{1}{\varepsilon} & 0 \\ 0 & 0 & -1 & 0 \\ 0 & \varepsilon & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{\sqrt[4]{1+\varepsilon^2}}{\sqrt{\varepsilon}} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt[4]{1+\varepsilon^2}\sqrt{\varepsilon}} & & 0 \\ 0 & 0 & \frac{\sqrt{\varepsilon}}{\sqrt[4]{1+\varepsilon^2}} & 0 \\ 0 & 0 & 0 & \frac{\sqrt[4]{1+\varepsilon^2}}{\sqrt{\varepsilon}} \end{pmatrix}$$

$$= \frac{\sqrt[4]{1+\varepsilon^2}}{\sqrt{\varepsilon}} \begin{pmatrix} 0 & \frac{1}{\sqrt{1+\varepsilon^2}} & 0 & 1 \\ 1 & 0 & -\frac{1}{\sqrt{1+\varepsilon^2}} & 0 \\ 0 & 0 & -\frac{\varepsilon}{\sqrt{1+\varepsilon^2}} & 0 \\ 0 & \frac{\varepsilon}{\sqrt{1+\varepsilon^2}} & 0 & 0 \end{pmatrix}.$$

$$\|Q_2\| \approx \frac{\sqrt[4]{1+\varepsilon^2}}{\sqrt{\varepsilon}}\sqrt{2},\ \|Q_2^{-1}\| \approx \frac{\sqrt{\varepsilon}}{\sqrt[4]{1+\varepsilon^2}} \frac{\sqrt{2}\sqrt{1+\varepsilon^2}}{\varepsilon},\ \kappa(Q_2) \approx \frac{2\sqrt{1+\varepsilon^2}}{\varepsilon}.$$

Thank you for your attention!!!

References:
L. Giraud, J. Langou, R, and J. van den Eshof: Rounding error analysis of the classical Gram-Schmidt orthogonalization process, Numerische Mathematik (2005) 101: 87-100.
R, J. Kopal, M. Tuma, A. Smoktunowicz: Numerical stability of orthogonalization methods with a non-standard inner product, BIT Numerical Mathematics (2012) 52:1035–1058.
R, F. Okulicka-Dluzewska, A. Smoktunowicz: Cholesky-like factorization of symmetric indefinite matrices and orthogonalization with respect to bilinear forms, SIAM J. Matrix Anal. and Appl. (2015), Vol. 36, No. 2, pp. 727—751.
H. Faßbender, R: On the conditioning of factors in the SR decomposition, Linear Algebra Appl. (2016), Vol. 505, pp. 224-244.

Orthogonalization with respect to a skew-symmetric bilinear form

$B \in \mathcal{R}^{2m,2m}$ skew-symmetric and nonsingular

$A = (a_1, \ldots, a_{2n}) \in \mathcal{R}^{2m,2n}$, $m \geq n = rank(A)/2$
$A = QR$, $R \in \mathcal{R}^{2n,2n}$ upper triangular with positive diagonal

B-orthonormal basis of $span(A)$: $Q = (q_1, \ldots, q_{2n}) \in \mathcal{R}^{2m,2n}$

$$Q^T B Q = \mathrm{diag}(\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}) \in \mathcal{R}^{2n,2n}$$

$$C = A^T B A = R^T \mathrm{diag}(\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}) R$$

Orthogonalization with respect to a skew-symmetric bilinear form

Schur-like factorization of skew-symmetric and nonsingular B

$$B = V \begin{pmatrix} 0 & \Sigma^2 \\ -\Sigma^2 & 0 \end{pmatrix} V^T$$

$V \in \mathcal{R}^{2m,2m}$ orthogonal with $V^T V = V V^T = I$
$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_m) \in \mathcal{R}^{m,m}$ with positive entries

$$B = V \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix} \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix} V^T$$

$\begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix} V^T A$ is a J-orthogonal matrix with

$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \in \mathcal{R}^{2m,2m}$ skew-symmetric and orthogonal

Example with $\kappa(R) \approx \kappa(A) \gg \kappa(C) \approx 1$

$$A = \begin{pmatrix} \sqrt{\varepsilon} & 0 & 0 & -\frac{1}{\sqrt{\varepsilon}} \\ 0 & 0 & 0 & \frac{\sqrt{1-\varepsilon^2}}{\sqrt{\varepsilon}} \\ 0 & \sqrt{\varepsilon} & \frac{1}{\sqrt{\varepsilon}} & 0 \\ 0 & 0 & \frac{\sqrt{1-\varepsilon^2}}{\sqrt{\varepsilon}} & 0 \end{pmatrix}, C = \begin{pmatrix} 0 & \varepsilon & 1 & 0 \\ -\varepsilon & 0 & 0 & 1 \\ -1 & 0 & 0 & \varepsilon \\ 0 & -1 & -\varepsilon & 0 \end{pmatrix}$$

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad R = \begin{pmatrix} \sqrt{\varepsilon} & 0 & 0 & -\frac{1}{\sqrt{\varepsilon}} \\ 0 & \sqrt{\varepsilon} & \frac{1}{\sqrt{\varepsilon}} & 0 \\ 0 & 0 & \frac{\sqrt{1-\varepsilon^2}}{\sqrt{\varepsilon}} & 0 \\ 0 & 0 & 0 & -\frac{\sqrt{1-\varepsilon^2}}{\sqrt{\varepsilon}} \end{pmatrix}$$

$$\sigma(A) \approx \frac{\sqrt{2}}{\sqrt{\varepsilon}}, \frac{\sqrt{\varepsilon}}{\sqrt{2}}, \ \kappa(A) \approx \frac{2}{\varepsilon},$$
$$\sigma(R) \approx \frac{\sqrt{2}}{\sqrt{\varepsilon}}, \frac{\sqrt{\varepsilon}}{\sqrt{2}}, \ \kappa(R) \approx \frac{2}{\varepsilon}, \ \kappa(Q) = 1$$

Local minimization of the condition number of R

$$\kappa^2(R) = \frac{\|R\|_F^2 + \sqrt{\|R\|_F^4 - 4r_{11}^2 r_{22}^2}}{\|R\|_F^2 - \sqrt{\|R\|_F^4 - 4r_{11}^2 r_{22}^2}}$$

As $r_{11}r_{22} = d$ is fixed and $\kappa(R)$ is an increasing function of $\|R\|_F$, it is minimized if $r_{12} = 0$ and $|r_{11}| = |r_{22}|$. Then

$$R^T R = |d| \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \kappa(R) = 1$$

.

$$A = \begin{pmatrix} \sqrt{\varepsilon} & 1 \\ 1 & 0 \\ 0 & \sqrt{\varepsilon} \\ 0 & 0 \end{pmatrix}, A^T J A = \begin{pmatrix} 0 & \varepsilon \\ -\varepsilon & 0 \end{pmatrix}, R = \begin{pmatrix} \sqrt{\varepsilon} & 0 \\ 0 & \sqrt{\varepsilon} \end{pmatrix}$$

$$Q = AR^{-1} = \begin{pmatrix} 1 & 1/\sqrt{\varepsilon} \\ 1/\sqrt{\varepsilon} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

Mehrmann 1979, Bunse-Gerstner and Mehrmann 1986

Orthogonal factor Q?

$$r_{11} = \|a_1\| = \sqrt{1+\varepsilon},\ q_1 = \frac{1}{\sqrt{1+\varepsilon}} \begin{pmatrix} \sqrt{\varepsilon} \\ 1 \\ 0 \\ 0 \end{pmatrix},\ r_{12} = q_1^T a_2 = \frac{\sqrt{\varepsilon}}{\sqrt{1+\varepsilon}},$$

$$r_{22} = \frac{a_1^T J a_2}{r_{11}} = \frac{\varepsilon}{\sqrt{1+\varepsilon}},\ q_2 = \frac{1}{r_{22}}(a_2 - r_{12}q_1) = \frac{1}{\varepsilon} \begin{pmatrix} \frac{1}{\sqrt{1+\varepsilon}} \\ -\frac{\sqrt{\varepsilon}}{\sqrt{1+\varepsilon}} \\ \sqrt{\varepsilon}\sqrt{1+\varepsilon} \\ 0 \end{pmatrix}$$

$$Q^T Q = \begin{pmatrix} 1 & 0 \\ 0 & \approx \frac{1}{\varepsilon^2(1+\varepsilon)} \end{pmatrix},\ \kappa(Q) \approx \frac{1}{\varepsilon}$$

$$R = \begin{pmatrix} \sqrt{1+\varepsilon} & \frac{\sqrt{\varepsilon}}{\sqrt{1+\varepsilon}} \\ 0 & \frac{\varepsilon}{\sqrt{1+\varepsilon}} \end{pmatrix},\ \lambda(R^T R) \approx 1 + 2\varepsilon, \varepsilon^2/16,\ \kappa(R) \approx \frac{4}{\varepsilon}$$

Local minimization of the condition number of Q

$$Q = AR^{-1},\ \kappa^2(Q) = \frac{\|Q\|_F^2 + \sqrt{\|Q\|_F^4 - 4\frac{(\|A\|\sigma_{min}(A))^2}{d^2}}}{\|Q\|_F^2 - \sqrt{\|Q\|_F^4 - 4\frac{(\|A\|\sigma_{min}(A))^2}{d^2}}}$$

As $\|A\|\sigma_{min}(A)/d$ is fixed and $\kappa(Q)$ is an increasing function of $\|Q\|_F$, it is minimized if r_{12} is chosen so that $q_1 \perp q_2$ with $\|q_1\| = \|q_2\|$. Then

$$Q^TQ = \frac{\|A\|\sigma_{min}(A)}{|d|}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \kappa(Q) = 1$$

.

$$R = \begin{pmatrix} \frac{\sqrt{\varepsilon}\sqrt{1+\varepsilon}}{\sqrt[4]{1+\varepsilon+\varepsilon^2}} & \frac{\varepsilon}{\sqrt{1+\varepsilon}\sqrt[4]{1+\varepsilon+\varepsilon^2}} \\ 0 & \frac{\sqrt{\varepsilon}\sqrt[4]{1+\varepsilon+\varepsilon^2}}{\sqrt{1+\varepsilon}} \end{pmatrix}$$

$$Q = AR^{-1} = \begin{pmatrix} \frac{\sqrt[4]{1+\varepsilon+\varepsilon^2}}{\sqrt{1+\varepsilon}} & \frac{1}{\sqrt{\varepsilon}\sqrt{1+\varepsilon}\sqrt[4]{1+\varepsilon+\varepsilon^2}} \\ \frac{\sqrt[4]{1+\varepsilon+\varepsilon^2}}{\sqrt{\varepsilon}\sqrt{1+\varepsilon}} & -\frac{1}{\sqrt{1+\varepsilon}\sqrt[4]{1+\varepsilon+\varepsilon^2}} \\ 0 & \frac{\sqrt{1+\varepsilon}}{\sqrt[4]{1+\varepsilon+\varepsilon^2}} \\ 0 & 0 \end{pmatrix}$$