What do we know about block Gram-Schmidt?

Erin C. Carson
Faculty of Mathematics and Physics, Charles University

February 24, 2021

E-NLA Seminar

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

slides: https://tinyurl.com/y2znwdfq

Collaborators

Kathryn Lund, formerly Charles University

Miro Rozložník, Czech Academy of Sciences

Stephen Thomas, National Renewable Energy Lab

The Gram-Schmidt process

Given a set of linear independent vectors $x_1, \dots, x_n$, we want to compute a set of orthogonal vectors $q_1, \dots, q_n$ such that $span\{x_1, \dots, x_n\} = span\{q_1, \dots, q_n\}$

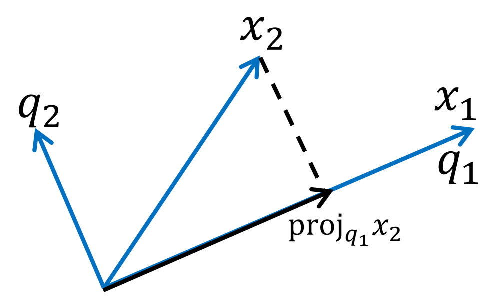
The Gram-Schmidt process

Given a set of linear independent vectors $x_1, \dots, x_n$, we want to compute a set of orthogonal vectors $q_1, \dots, q_n$ such that $span\{x_1, \dots, x_n\} = span\{q_1, \dots, q_n\}$

Gram-Schmidt process:

$$q_1 = x_1, \;\; q_k = x_k - \sum_{j=1}^{k-1} \frac{\langle q_j, x_k \rangle}{\|q_j\|^2} q_j, \quad k \geq 2$$

To get orthonormal vectors, $q_k = q_k / \|q_k\|$, for all k



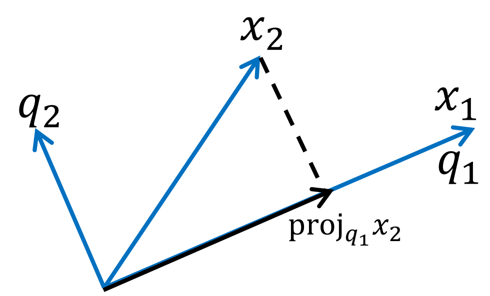
The Gram-Schmidt process

Given a set of linear independent vectors $x_1, \dots, x_n$, we want to compute a set of orthogonal vectors $q_1, \dots, q_n$ such that $span\{x_1, \dots, x_n\} = span\{q_1, \dots, q_n\}$

Gram-Schmidt process:

$$q_1 = x_1, \;\; q_k = x_k - \sum_{j=1}^{k-1} \frac{\langle q_j, x_k \rangle}{\|q_j\|^2} q_j \,, \quad k \geq 2$$

To get orthonormal vectors, $q_k = q_k / \|q_k\|$, for all k



Each vector x_k can be expressed as a linear combination of $q_1, \dots, q_k$.

So with $X = [x_1 \cdots x_n]$, $Q = [q_1 \cdots q_n]$, this means we can write

$$X = QR,$$

where columns of R give the coefficients of the aforementioned linear combinations, and thus R is upper triangular.

Typically require that the diagonal entries of R are positive; this gives a unique QR factorization.

Orthogonalization

- Many applications
  - Solving least squares problems $\min_x \|Ax - b\|_2^2 \;\rightarrow\; Rx = Q^T b$
  - Used within Krylov subspace methods
  - Etc.

Orthogonalization

- Many applications
  - Solving least squares problems $\min_x \|Ax - b\|_2^2 \quad \rightarrow \quad Rx = Q^T b$
  - Used within Krylov subspace methods
  - Etc.

- What happens in finite precision?
  - On a real computer, every time we perform a floating point operation, we may incur a small roundoff error
  - Over a whole computation, these tiny errors can accumulate or can be amplified!
  - The result:
    - $\bar{Q}$ no longer has exactly orthonormal columns!
    - $\bar{Q}\bar{R}$ is no longer exactly the same as X!
      - This can affect applications downstream

Measures of Error

Let $\bar{Q}$ and $\bar{R}$ denote computed QR factors of a matrix X.

How far is $\bar{Q}$ from having orthonormal columns?

"Loss of orthogonality": $\|I - \bar{Q}^T\bar{Q}\|$

How close is $\bar{Q}\bar{R}$ to X?

Relative residual norm: $\frac{\|X-\bar{Q}\bar{R}\|}{\|X\|}$

How close is $\bar{R}^T\bar{R}$ to X^TX?

Relative Cholesky residual norm: $\frac{\|X^TX-\bar{R}^T\bar{R}\|}{\|X\|^2}$

Gram-Schmidt algorithms

Classical Gram-Schmidt (CGS)

for $k = 1, \ldots, n$
  $w_k = x_k$
  for $j = 1, \ldots, k-1$
    $w_k = w_k - (q_j^T x_k) q_j$
  $q_k = w_k / \|w_k\|$

Modified Gram-Schmidt (MGS)

for $k = 1, \ldots, n$
  $w_k = x_k$
  for $j = 1, \ldots, k-1$
    $w_k = w_k - (q_j^T w_k) q_j$
  $q_k = w_k / \|w_k\|$

Gram-Schmidt algorithms

Classical Gram-Schmidt (CGS)

for $k = 1, \ldots, n$
  $w_k = x_k$
  for $j = 1, \ldots, k-1$
    $w_k = w_k - (q_j^T x_k) q_j$
  $q_k = w_k / \|w_k\|$

Modified Gram-Schmidt (MGS)

for $k = 1, \ldots, n$
  $w_k = x_k$
  for $j = 1, \ldots, k-1$
    $w_k = w_k - (q_j^T w_k) q_j$
  $q_k = w_k / \|w_k\|$

	CGS	MGS
Computation of entries of R	$r_{jk} = q_j^T x_k$	$r_{jk} = q_j^T \left(x_k - \sum_{i=1}^{k-1} r_{ik} q_k \right)$
Computation of next orthogonal vector	$(I - Q_{1:k-1} Q_{1:k-1}^T) x_k$	$(I - q_{k-1} q_{k-1}^T) \cdots (I - q_1 q_1^T) x_k$
Loss of orthogonality	$\|I - \bar{Q}^T \bar{Q}\| \leq O(\varepsilon) \kappa^{n-1}(X)$ if $O(\varepsilon)\kappa(X) < 1$	$\|I - \bar{Q}^T \bar{Q}\| \leq O(\varepsilon) \kappa(X)$ if $O(\varepsilon)\kappa(X) < 1$
Parallel messages/ synchronizations	$O(1)$	$O(k)$ in loop k

A bit of history…

[Leon, Björck, Gander, "Gram-Schmidt orthogonalization: 100 years and more", 2007]

- Method of orthogonalization popularized by a paper of Schmidt in 1907. The method here is what we know as "classical Gram-Schmidt"
- In a footnote, Schmidt credits an earlier paper by Gram, published in 1883, saying that this procedure is essentially equivalent.
  - The procedure in Gram's paper is what we know as "modified Gram-Schmidt"
- The linkage of the names "Gram" and "Schmidt" came along in 1935 in a paper by Wong

Jørgen Pedersen Gram

Erhard Schmidt

A bit of history…

[Leon, Björck, Gander, "Gram-Schmidt orthogonalization: 100 years and more", 2007]

- Method of orthogonalization popularized by a paper of Schmidt in 1907. The method here is what we know as "classical Gram-Schmidt"
- In a footnote, Schmidt credits an earlier paper by Gram, published in 1883, saying that this procedure is essentially equivalent.
    - The procedure in Gram's paper is what we know as "modified Gram-Schmidt"
- The linkage of the names "Gram" and "Schmidt" came along in 1935 in a paper by Wong

Jørgen Pedersen Gram

Erhard Schmidt

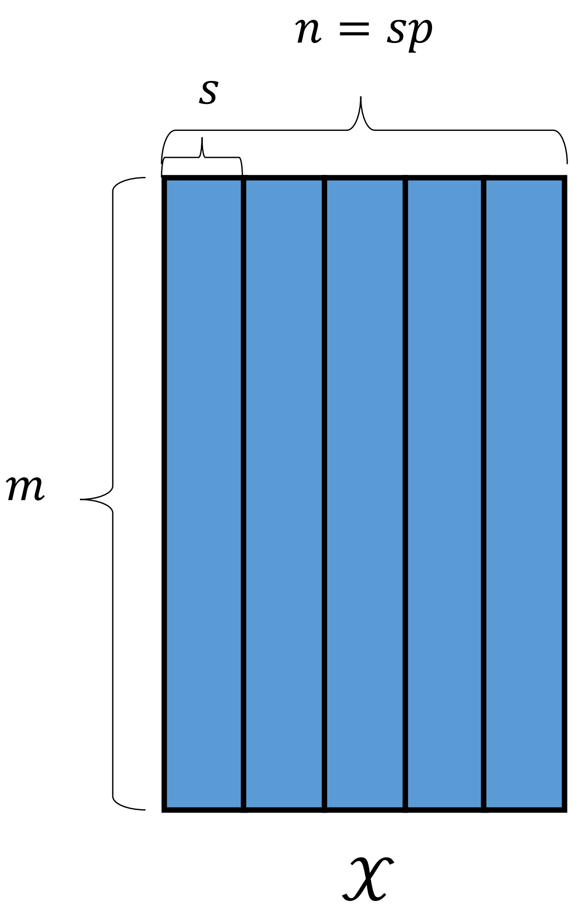
Pierre-Simon Laplace

- It turns out a procedure equivalent to modified Gram-Schmidt appears even in much earlier work of Laplace in 1820

Block Gram-Schmidt

- Sometimes we may want to use a block version of Gram-Schmidt

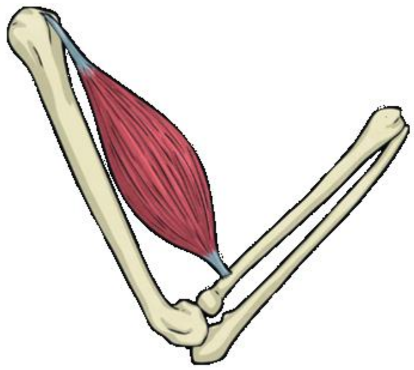
- Performance reasons (e.g., BLAS3)
- Block Krylov subspace methods
  - Better convergence
  - Simultaneously solve multiple RHSes
- s-step Krylov subspace methods



Muscle and Skeleton analogy [Hoemmen, 2010]

- How do we define a block Gram-Schmidt algorithm?

- We need 2 parts:
  - The "skeleton": A block Gram-Schmidt algorithm for interblock orthogonalization
  - The "muscle": A non-block orthogonalization algorithm for intrablock orthogonalization ("local QR", "panel factorization")
    - Need not be Gram-Schmidt-based
    - We will refer to this routine as "IntraOrtho()"



https://www.twinkl.com/illustration/contracted-muscle-arm-bone-skeleton-movement-anatomy-bicep-science-ks2

- For example: block MGS (BMGS) for orthogonalizing between blocks, Householder QR for orthogonalizing within blocks:

$$\text{BMGS} \circ \text{HouseQR}(\mathcal{X})$$

Notation I

- Use our own naming system of algorithms
    - Does suffix "2" mean reorthogonalized? BLAS-2 featuring? A second version of the algorithm?

- Suffixes:
    - +: run twice
    - I+: inner reorthogonalization
    - S+: selective reorthogonalization

Notation II

- Calligraphic letters for the whole block matrices $(\mathcal{X}, \mathcal{Q}, \mathcal{R})$
- Regular letters for the individual block quantities (X, Q, R)
- Bars denote computed (inexact) quantities

- m: number of rows in input matrix
- n: number of columns in input matrix $(n = ps)$
- p: number of blocks
- s: number of columns per block

$$m \geq n > p > s$$

$$\mathcal{X} = [X_1, X_2, \dots, X_p], \quad \mathcal{X} \in \mathbb{R}^{m \times n}, \ X_i \in \mathbb{R}^{m \times s}$$

Notation II

- Calligraphic letters for the whole block matrices $(\mathcal{X}, \mathcal{Q}, \mathcal{R})$
- Regular letters for the individual block quantities (X, Q, R)
- Bars denote computed (inexact) quantities

- m: number of rows in input matrix
- n: number of columns in input matrix $(n = ps)$
- p: number of blocks
- s: number of columns per block

$$m \geq n > p > s$$

$$\mathcal{X} = [X_1, X_2, \dots, X_p], \quad \mathcal{X} \in \mathbb{R}^{m \times n}, \quad X_i \in \mathbb{R}^{m \times s}$$

Economic QR factorization: $\mathcal{X} = \mathcal{Q}\mathcal{R}$, $\mathcal{Q} \in \mathbb{R}^{m \times n}, \mathcal{R} \in \mathbb{R}^{n \times n}$

$$\mathcal{Q} = [Q_1, Q_2, \dots, Q_p], \qquad \mathcal{R} = \begin{bmatrix} R_{1,1} & R_{1,2} & \cdots & R_{1,p} \\ & R_{2,2} & \cdots & R_{2,p} \\ & & \ddots & \vdots \\ & & & R_{p,p} \end{bmatrix}$$

$$\mathcal{Q}_{1:j} = [Q_1, \dots, Q_j], \qquad \mathcal{R}_{1:j,k} = \begin{bmatrix} R_{1,k} \\ \vdots \\ R_{j,k} \end{bmatrix}$$

Block Gram-Schmidt methods in practice

A few examples:

- [Boley and Golub, 1984]: Block Arnoldi with BMGSI+ ∘ MGSI+

  "However, since we obtain Z_k, by using a [block] Gram-Schmidt orthogonalization of W_k, against $Q_1, \dots, Q_k$ … there is little loss of stability by continuing to use Gram-Schmidt to orthogonalize Z_k. This was what we actually observed in our numerical experiments."

Block Gram-Schmidt methods in practice

A few examples:

- [Boley and Golub, 1984]: Block Arnoldi with BMGSI+ ∘ MGSI+

  "However, since we obtain Z_k, by using a [block] Gram-Schmidt orthogonalization of W_k, against $Q_1, \dots, Q_k$ ... there is little loss of stability by continuing to use Gram-Schmidt to orthogonalize Z_k. This was what we actually observed in our numerical experiments."

- [Simoncini and Gallapoulos, 1995]: Block GMRES with BMGS ∘ MGS
  - Also used in [Vital, 1990]

Block Gram-Schmidt methods in practice

A few examples:

- [Boley and Golub, 1984]: Block Arnoldi with BMGSI+ $\circ$ MGSI+
  "However, since we obtain Z_k, by using a [block] Gram-Schmidt orthogonalization of W_k, against $Q_1, \ldots, Q_k$ ... there is little loss of stability by continuing to use Gram-Schmidt to orthogonalize Z_k. This was what we actually observed in our numerical experiments."
- [Simoncini and Gallapoulos, 1995]: Block GMRES with BMGS $\circ$ MGS
  - Also used in [Vital, 1990]
- [Sadkane, 1993]: Block Arnoldi with BMGS $\circ$ "QR factorization"
- [Saad, 2003]: textbook provides block Arnoldi based on BCGS (Alg. 6.22) and BMGS (Alg. 6.23); IntraOrtho just specified as a "QR factorization"
- [Baker, Dennis, Jessup, 2006]: Block GMRES w/ BMGS $\circ$ "QR factorization"

Block Gram-Schmidt methods in practice

A few examples:

- [Boley and Golub, 1984]: Block Arnoldi with BMGSI+ ∘ MGSI+

  "However, since we obtain Z_k, by using a [block] Gram-Schmidt orthogonalization of W_k, against $Q_1, \ldots, Q_k$ ... there is little loss of stability by continuing to use Gram-Schmidt to orthogonalize Z_k. This was what we actually observed in our numerical experiments."
- [Simoncini and Gallapoulos, 1995]: Block GMRES with BMGS ∘ MGS
  - Also used in [Vital, 1990]
- [Sadkane, 1993]: Block Arnoldi with BMGS ∘ "QR factorization"
- [Saad, 2003]: textbook provides block Arnoldi based on BCGS (Alg. 6.22) and BMGS (Alg. 6.23); IntraOrtho just specified as a "QR factorization"
- [Baker, Dennis, Jessup, 2006]: Block GMRES w/ BMGS ∘ "QR factorization"

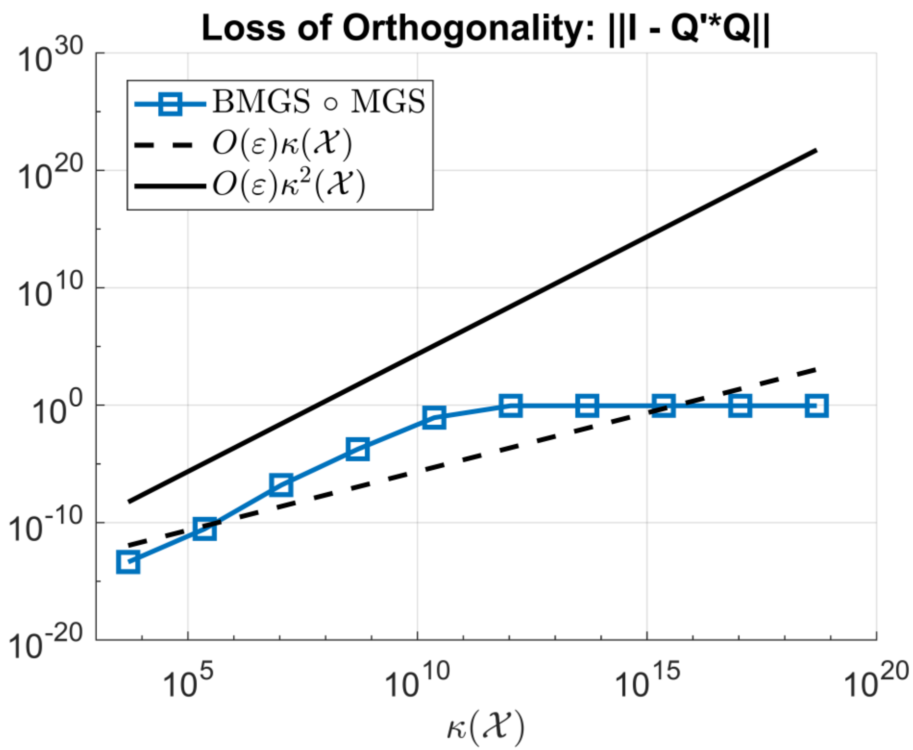
- Does it matter what we use for "QR factorization" (the IntraOrtho) within a block Gram-Schmidt method?

Does it matter?

- Recall: For MGS, $\|I - \bar{Q}^T \bar{Q}\| \leq O(\varepsilon)\kappa(X)$
- What is the bound on loss of orthogonality for BMGS $\circ$ MGS? (guess!)

Does it matter?

- Recall: For MGS, $\|I - \bar{Q}^T\bar{Q}\| \leq O(\varepsilon)\kappa(X)$
- What is the bound on loss of orthogonality for BMGS ∘ MGS? (guess!)



Läuchli matrix

$$\mathcal{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \eta & & & \\ & \eta & & \\ & & \ddots & \\ & & & \eta \end{bmatrix}$$

$\eta \in (\varepsilon, \sqrt{\varepsilon})$

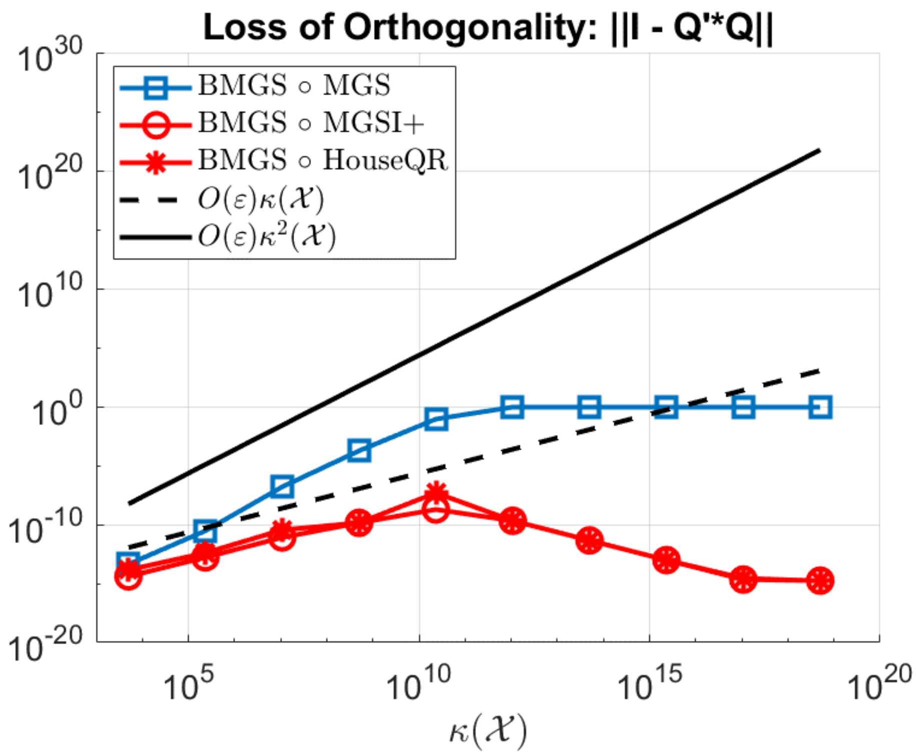
$m = 1000, p = 100,$
$s = 5$

[Jalby and Philippe, 1991]: For BMGS ∘ MGS, $\|I - \bar{Q}^T\bar{Q}\| \leq O(\varepsilon)\kappa^2(\mathcal{X})$


```
LaeuchliBlockKappaPlot([1000 100 5],  logspace(-1, -16, 10), {'BMGS'}, {'MGS'})
```

Does it matter?

- Recall: For MGS, $\|I - \bar{Q}^T \bar{Q}\| \leq O(\varepsilon)\kappa(X)$
- What is the bound on loss of orthogonality for BMGS ∘ MGS? (guess!)



Läuchli matrix

$$\mathcal{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \eta & & & \\ & \eta & & \\ & & \ddots & \\ & & & \eta \end{bmatrix}$$

$\eta \in (\varepsilon, \sqrt{\varepsilon})$

$m = 1000, p = 100,$
$s = 5$

[Jalby and Philippe, 1991]: For BMGS ∘ MGS, $\|I - \bar{Q}^T \bar{Q}\| \leq O(\varepsilon)\kappa^2(\mathcal{X})$

For BMGS ∘ MGS+, $\|I - \bar{Q}^T \bar{Q}\| \leq O(\varepsilon)\kappa(\mathcal{X})$


```
LaeuchliBlockKappaPlot([1000 100 5],  logspace(-1, -16, 10), {'BMGS'}, {'MGS', ‘MGS_IRO'})
```

Two Questions

If we use an intrablock orthogonalization routine (muscle) with $O(\varepsilon)$ loss of orthogonality and $O(\varepsilon)$ relative residual, what is the best a block Gram-Schmidt orthogonalization routine (skeleton) can do?

For a given block Gram-Schmidt variant (skeleton), what are the minimum requirements on the intra-block orthogonalization routine (muscle) such that the loss of orthogonality is good enough?

BlockStab MATLAB package

BlockStab (our code) has two simple drivers:

- BGS(XX, s, skel, musc, rpltol, verbose)
- IntraOrtho(X, musc, rpltol, verbose)
- Can also work directly with a skeleton or muscle, or implement your own

https://github.com/katlund/BlockStab

**For each plot, we list the function call needed to replicate the plot at the bottom of the slide*

Outline

1. Overview of muscles
2. BCGS skeletons
3. BMGS skeletons
4. Open questions

Overview of Muscles

CGS and CGS-P

- Pessimistic bound due to [Kiełbasiński, 1974]: If $O(\varepsilon)\kappa(X) < 1$,

$$\|I - \bar{Q}^T \bar{Q}\| \leq O(\varepsilon)\kappa^{s-1}(X)$$

  for $X \in \mathbb{R}^{m \times s}$

CGS and CGS-P

- Pessimistic bound due to [Kiełbasiński, 1974]: If $O(\varepsilon)\kappa(X) < 1$,

$$\|I - \bar{Q}^T \bar{Q}\| \leq O(\varepsilon)\kappa^{s-1}(X)$$

  for $X \in \mathbb{R}^{m \times s}$

- [Giraud, Langou, Rozložník, van den Eshof, 2005]: If $O(\varepsilon)\kappa^2(X) < 1$,

$$\|I - \bar{Q}^T \bar{Q}\| \leq O(\varepsilon)\kappa^2(X)$$

CGS and CGS-P

- Pessimistic bound due to [Kiełbasiński, 1974]: If $O(\varepsilon)\kappa(X) < 1$,

$$\|I - \bar{Q}^T \bar{Q}\| \leq O(\varepsilon)\kappa^{s-1}(X)$$

  for $X \in \mathbb{R}^{m \times s}$

- [Giraud, Langou, Rozložník, van den Eshof, 2005]: If $O(\varepsilon)\kappa^2(X) < 1$,

$$\|I - \bar{Q}^T \bar{Q}\| \leq O(\varepsilon)\kappa^2(X)$$

- [Smoktunowicz, Barlow, Langou, 2006]: The above only holds if we compute the diagonal entries of R using a Cholesky-like formula via the Pythagorean theorem (CGS-P)

CGS and CGS-P

- Pessimistic bound due to [Kiełbasiński, 1974]: If $O(\varepsilon)\kappa(X) < 1$,

$$\|I - \bar{Q}^T \bar{Q}\| \leq O(\varepsilon)\kappa^{s-1}(X)$$

  for $X \in \mathbb{R}^{m \times s}$

- [Giraud, Langou, Rozložník, van den Eshof, 2005]: If $O(\varepsilon)\kappa^2(X) < 1$,

$$\|I - \bar{Q}^T \bar{Q}\| \leq O(\varepsilon)\kappa^2(X)$$

- [Smoktunowicz, Barlow, Langou, 2006]: The above only holds if we compute the diagonal entries of R using a Cholesky-like formula via the Pythagorean theorem (CGS-P)

$$R_{1:k,k+1} = Q_{1:k}^T x_{k+1}$$
$$w = x_{k+1} - Q_{1:k} R_{1:k,k+1}$$

Let $\phi = \|x_{k+1}\|, \ \psi = \left\|R_{1:k,k+1}\right\|$

CGS:

$$R_{k+1,k+1} = \|w\| \left(= \sqrt{\phi^2 - \psi^2}\right)$$

CGS and CGS-P

- Pessimistic bound due to [Kiełbasiński, 1974]: If $O(\varepsilon)\kappa(X) < 1$,

$$\|I - \bar{Q}^T \bar{Q}\| \leq O(\varepsilon)\kappa^{s-1}(X)$$

  for $X \in \mathbb{R}^{m \times s}$

- [Giraud, Langou, Rozložník, van den Eshof, 2005]: If $O(\varepsilon)\kappa^2(X) < 1$,

$$\|I - \bar{Q}^T \bar{Q}\| \leq O(\varepsilon)\kappa^2(X)$$

- [Smoktunowicz, Barlow, Langou, 2006]: The above only holds if we compute the diagonal entries of R using a Cholesky-like formula via the Pythagorean theorem (CGS-P)

$$R_{1:k,k+1} = Q_{1:k}^T x_{k+1}$$
$$w = x_{k+1} - Q_{1:k} R_{1:k,k+1}$$

Let $\phi = \|x_{k+1}\|$, $\psi = \left\|R_{1:k,k+1}\right\|$

CGS:

$$R_{k+1,k+1} = \|w\| \left(= \sqrt{\phi^2 - \psi^2}\right)$$

CGS-P:

$$R_{k+1,k+1} = \sqrt{\phi - \psi} \cdot \sqrt{\phi + \psi}$$

Reorthogonalization

- "Twice is enough"

  [Parlett, 1987]; attributed to Kahan :

  An iterative Gram-Schmidt process on 2 vectors with one step of reorthogonalization produces 2 vectors orthonormal up to machine precision *if the matrix is not too ill-conditioned*.

Reorthogonalization

- "Twice is enough"

  [Parlett, 1987]; attributed to Kahan :

  An iterative Gram-Schmidt process on 2 vectors with one step of reorthogonalization produces 2 vectors orthonormal up to machine precision *if the matrix is not too ill-conditioned*.

- [Giraud, Langou, Rozložník, 2002], [Giraud, Langou, Rozložník, van den Eshof, 2005]: Twice is enough holds for $k > 2$ vectors
  - Previous work by [Abdelmalek, 1971]

$$\|I - \bar{Q}^T \bar{Q}\| \leq O(\varepsilon)$$

- Many variants: CGS+, CGSI+, CGSS+

Low-synchronization version

CGSI+LS [Świrydowicz, Langou, Ananthan, Yang, Thomas, 2020]

- One synchronization per column (CGSI+ up to 4)

$[\boldsymbol{Q}, R] = \texttt{CGSI+LS}(\boldsymbol{X})$

Allocate memory for $\boldsymbol{Q}$ and R
$\boldsymbol{u} = \boldsymbol{x}_1$
for $k = 2, \ldots s$ **do**
  if $k = 2$ **then**
    $[r_{k-1,k-1}^2 \;\; \rho] = \boldsymbol{u}^T [\boldsymbol{u} \;\; \boldsymbol{x}_k]$
  else if $k > 2$ **then**
    $\begin{bmatrix} \boldsymbol{w} & \boldsymbol{z} \\ \omega & \zeta \end{bmatrix} = [\boldsymbol{Q}_{1:k-2} \;\; \boldsymbol{u}]^T [\boldsymbol{u} \;\; \boldsymbol{x}_k]$
    $[r_{k-1,k-1}^2 \;\; \rho] = [\omega \;\; \zeta] - \boldsymbol{w}^T [\boldsymbol{w} \;\; \boldsymbol{z}]$
  end if
  $r_{k-1,k} = \rho / r_{k-1,k-1}$
  if $k = 2$ **then**
    $\boldsymbol{q}_{k-1} = \boldsymbol{u} / r_{k-1,k-1}$
  else if $k > 2$ **then**
    $R_{1:k-2,k-1} = R_{1:k-2,k-1} + \boldsymbol{w}$
    $R_{1:k-2,k} = \boldsymbol{z}$
    $\boldsymbol{q}_{k-1} = (\boldsymbol{u} - \boldsymbol{Q}_{1:k-2} \boldsymbol{w}) / r_{k-1,k-1}$
  end if
  $\boldsymbol{u} = \boldsymbol{x}_k - \boldsymbol{Q}_{1:k-1} R_{1:k-1,k}$
end for
$\begin{bmatrix} \boldsymbol{w} \\ \omega \end{bmatrix} = [\boldsymbol{Q}_{1:s-1} \;\; \boldsymbol{u}]^T \boldsymbol{u}$
$r_{s,s}^2 = \omega - \boldsymbol{w}^T \boldsymbol{w}$
$R_{1:s-1,s} = R_{1:s-1,s} + \boldsymbol{w}$
$\boldsymbol{q}_s = (\boldsymbol{u} - \boldsymbol{Q}_{1:s-1} \boldsymbol{w}) / r_{s,s}$
return $\boldsymbol{Q} = [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_s]$, $R = (r_{jk})$

Low-synchronization version

CGSI+LS [Świrydowicz, Langou, Ananthan, Yang, Thomas, 2020]

- One synchronization per column (CGSI+ up to 4)

Main ideas:

Orthogonal projector: $I - QTQ^T, \; T \approx (Q^T Q)^{-1}$

$$T = I - L - L^T$$

$[\boldsymbol{Q}, R] = \texttt{CGSI+LS}(\boldsymbol{X})$

Allocate memory for $\boldsymbol{Q}$ and R
$\boldsymbol{u} = \boldsymbol{x}_1$
for $k = 2, \ldots s$ **do**
  if $k = 2$ **then**
    $[r_{k-1,k-1}^2 \; \rho] = \boldsymbol{u}^T [\boldsymbol{u} \; \boldsymbol{x}_k]$
  else if $k > 2$ **then**
    $\begin{bmatrix} \boldsymbol{w} & \boldsymbol{z} \\ \omega & \zeta \end{bmatrix} = [\boldsymbol{Q}_{1:k-2} \; \boldsymbol{u}]^T [\boldsymbol{u} \; \boldsymbol{x}_k]$
    $[r_{k-1,k-1}^2 \; \rho] = [\omega \; \zeta] - \boldsymbol{w}^T [\boldsymbol{w} \; \boldsymbol{z}]$
  end if
  $r_{k-1,k} = \rho / r_{k-1,k-1}$
  if $k = 2$ **then**
    $\boldsymbol{q}_{k-1} = \boldsymbol{u} / r_{k-1,k-1}$
  else if $k > 2$ **then**
    $R_{1:k-2,k-1} = R_{1:k-2,k-1} + \boldsymbol{w}$
    $R_{1:k-2,k} = \boldsymbol{z}$
    $\boldsymbol{q}_{k-1} = (\boldsymbol{u} - \boldsymbol{Q}_{1:k-2} \boldsymbol{w}) / r_{k-1,k-1}$
  end if
  $\boldsymbol{u} = \boldsymbol{x}_k - \boldsymbol{Q}_{1:k-1} R_{1:k-1,k}$
end for
$\begin{bmatrix} \boldsymbol{w} \\ \omega \end{bmatrix} = [\boldsymbol{Q}_{1:s-1} \; \boldsymbol{u}]^T \boldsymbol{u}$
$r_{s,s}^2 = \omega - \boldsymbol{w}^T \boldsymbol{w}$
$R_{1:s-1,s} = R_{1:s-1,s} + \boldsymbol{w}$
$\boldsymbol{q}_s = (\boldsymbol{u} - \boldsymbol{Q}_{1:s-1} \boldsymbol{w}) / r_{s,s}$
return $\boldsymbol{Q} = [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_s]$, $R = (r_{jk})$

Low-synchronization version

CGSI+LS [Świrydowicz, Langou, Ananthan, Yang, Thomas, 2020]

- One synchronization per column (CGSI+ up to 4)

Main ideas:

Orthogonal projector: $I - QTQ^T, \;\; T \approx (Q^T Q)^{-1}$

$$T = I - L - L^T$$

1. Compute strictly lower triangular matrix L one row (or block of rows) at a time in single reduction to compute all inner products needed for current iteration
$$L_{k-1,1:k-2} = \left(Q_{1:k-2}^T q_{k-1}\right)^T$$

2. "Lag" reorthogonalization and normalization and merge it with this single reduction
   - Idea of lag also used in [Hernández, Román, Tomás, 2007]

$[\boldsymbol{Q}, R] = \texttt{CGSI+LS}(\boldsymbol{X})$

Allocate memory for $\boldsymbol{Q}$ and R
$\boldsymbol{u} = \boldsymbol{x}_1$
for $k = 2, \ldots s$ **do**
  if $k = 2$ **then**
    $[r_{k-1,k-1}^2 \; \rho] = \boldsymbol{u}^T [\boldsymbol{u} \; \boldsymbol{x}_k]$
  else if $k > 2$ **then**
    $\begin{bmatrix} \boldsymbol{w} & \boldsymbol{z} \\ \omega & \zeta \end{bmatrix} = [\boldsymbol{Q}_{1:k-2} \; \boldsymbol{u}]^T [\boldsymbol{u} \; \boldsymbol{x}_k]$
    $[r_{k-1,k-1}^2 \; \rho] = [\omega \; \zeta] - \boldsymbol{w}^T [\boldsymbol{w} \; \boldsymbol{z}]$
  end if
  $r_{k-1,k} = \rho / r_{k-1,k-1}$
  if $k = 2$ **then**
    $\boldsymbol{q}_{k-1} = \boldsymbol{u} / r_{k-1,k-1}$
  else if $k > 2$ **then**
    $R_{1:k-2,k-1} = R_{1:k-2,k-1} + \boldsymbol{w}$
    $R_{1:k-2,k} = \boldsymbol{z}$
    $\boldsymbol{q}_{k-1} = (\boldsymbol{u} - \boldsymbol{Q}_{1:k-2}\boldsymbol{w}) / r_{k-1,k-1}$
  end if
  $\boldsymbol{u} = \boldsymbol{x}_k - \boldsymbol{Q}_{1:k-1} R_{1:k-1,k}$
end for
$\begin{bmatrix} \boldsymbol{w} \\ \omega \end{bmatrix} = [\boldsymbol{Q}_{1:s-1} \; \boldsymbol{u}]^T \boldsymbol{u}$
$r_{s,s}^2 = \omega - \boldsymbol{w}^T \boldsymbol{w}$
$R_{1:s-1,s} = R_{1:s-1,s} + \boldsymbol{w}$
$\boldsymbol{q}_s = (\boldsymbol{u} - \boldsymbol{Q}_{1:s-1}\boldsymbol{w}) / r_{s,s}$
return $\boldsymbol{Q} = [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_s]$, $R = (r_{jk})$

Low-synchronization version

CGSI+LS [Świrydowicz, Langou, Ananthan, Yang, Thomas, 2020]

- One synchronization per column (CGSI+ up to 4)

Main ideas:

Orthogonal projector: $I - QTQ^T$, $T \approx (Q^TQ)^{-1}$

$$T = I - L - L^T$$

1. Compute strictly lower triangular matrix L one row (or block of rows) at a time in single reduction to compute all inner products needed for current iteration
$$L_{k-1,1:k-2} = \left(Q_{1:k-2}^T q_{k-1}\right)^T$$

2. "Lag" reorthogonalization and normalization and merge it with this single reduction
   - Idea of lag also used in [Hernández, Román, Tomás, 2007]

⇒Reorthogonalization happens "on the fly" instead of requiring complete second pass

$[\boldsymbol{Q}, R] = \texttt{CGSI+LS}(\boldsymbol{X})$

Allocate memory for $\boldsymbol{Q}$ and R
$\boldsymbol{u} = \boldsymbol{x}_1$
for $k = 2, \ldots s$ **do**
  if $k = 2$ **then**
    $[r_{k-1,k-1}^2 \; \rho] = \boldsymbol{u}^T[\boldsymbol{u} \; \boldsymbol{x}_k]$
  else if $k > 2$ **then**
    $\begin{bmatrix} \boldsymbol{w} & \boldsymbol{z} \\ \omega & \zeta \end{bmatrix} = [\boldsymbol{Q}_{1:k-2} \; \boldsymbol{u}]^T[\boldsymbol{u} \; \boldsymbol{x}_k]$
    $[r_{k-1,k-1}^2 \; \rho] = [\omega \; \zeta] - \boldsymbol{w}^T[\boldsymbol{w} \; \boldsymbol{z}]$
  end if
  $r_{k-1,k} = \rho / r_{k-1,k-1}$
  if $k = 2$ **then**
    $\boldsymbol{q}_{k-1} = \boldsymbol{u}/r_{k-1,k-1}$
  else if $k > 2$ **then**
    $R_{1:k-2,k-1} = R_{1:k-2,k-1} + \boldsymbol{w}$
    $R_{1:k-2,k} = \boldsymbol{z}$
    $\boldsymbol{q}_{k-1} = (\boldsymbol{u} - \boldsymbol{Q}_{1:k-2}\boldsymbol{w})/r_{k-1,k-1}$
  end if
  $\boldsymbol{u} = \boldsymbol{x}_k - \boldsymbol{Q}_{1:k-1}R_{1:k-1,k}$
end for
$\begin{bmatrix} \boldsymbol{w} \\ \omega \end{bmatrix} = [\boldsymbol{Q}_{1:s-1} \; \boldsymbol{u}]^T \boldsymbol{u}$
$r_{s,s}^2 = \omega - \boldsymbol{w}^T\boldsymbol{w}$
$R_{1:s-1,s} = R_{1:s-1,s} + \boldsymbol{w}$
$\boldsymbol{q}_s = (\boldsymbol{u} - \boldsymbol{Q}_{1:s-1}\boldsymbol{w})/r_{s,s}$
return $\boldsymbol{Q} = [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_s]$, $R = (r_{jk})$

Summary: CGS Variants

Algorithm	$\|I - \bar{Q}^T \bar{Q}\|$	Assumption on $\kappa(X)$	References
CGS	$O(\varepsilon)\kappa^{n-1}(X)$	$O(\varepsilon)\kappa(X) < 1$	[Kiełbasiński, 1974]
CGS-P	$O(\varepsilon)\kappa^2(X)$	$O(\varepsilon)\kappa^2(X) < 1$	[Smoktunowicz, Barlow, Langou, 2006]
CGS+	$O(\varepsilon)$	$O(\varepsilon)\kappa(X) < 1$	conjecture
CGSI+	$O(\varepsilon)$	$O(\varepsilon)\kappa(X) < 1$	[Abdelmalek, 1971] [Giraud, Langou, Rozložník, 2002] [Giraud, Langou, Rozložník, van den Eshof, 2005] [Barlow, Smoktunowicz, 2013]
CGSS+	$O(\varepsilon)$	$O(\varepsilon)\kappa(X) < 1$	[Daniel, Gragg, Kaufman, Stewart, 1976] [Hoffmann, 1989]
CGSI+LS	$O(\varepsilon)$	$O(\varepsilon)\kappa(X) < 1$	conjecture [Świrydowicz, Langou, Ananthan, Yang, Thomas, 2020]
CGSS+rpl	$O(\varepsilon)$	none	conjecture [Stewart, 2008]

Low-sync MGS

- Original low-sync muscles developed independently by [Barlow 2019] and [Świrydowicz et al., 2020]

- Perspectives:
  - Merge inner products and norms by batching and lagging
  - Cushion projectors with "error sponge" to achieve CGS-like communication with MGS-like stability

CGS: $(I - Q_k Q_k^T) x_{k+1}$

MGS: $(I - q_k q_k^T) \cdots (I - q_1 q_1^T) x_{k+1}$

Goal: $(I - Q_k C_k Q_k^T) x_{k+1}$

Low-sync variants

	Two synchronizations	One synchronization
$(I - Q_k T_k^T Q_k^T) x_{k+1}$ (matrix-matrix multiplication)	MGS-SVL [Barlow, 2019] (called "MGS2")	MGS-CWY [Swirydowicz et al., 2020]
$(I - Q_k T_k^{-T} Q_k^T) x_{k+1}$ (triangular solve)	MGS-LTS [Swirydowicz et al., 2020]	MGS-ICWY [Swirydowicz et al., 2020]

Low-sync variants

[Schreiber and Van Loan, 1989] + "Sheffield observation"

	Two synchronizations	One synchronization
$(I - Q_k T_k^T Q_k^T) x_{k+1}$ (matrix-matrix multiplication)	MGS-SVL [Barlow, 2019] (called "MGS2")	MGS-CWY [Swirydowicz et al., 2020]
$(I - Q_k T_k^{-T} Q_k^T) x_{k+1}$ (triangular solve)	MGS-LTS [Swirydowicz et al., 2020]	MGS-ICWY [Swirydowicz et al., 2020]

Low-sync variants

[Schreiber and Van Loan, 1989] + "Sheffield observation"

	Two synchronizations	One synchronization
$(I - Q_k T_k^T Q_k^T)x_{k+1}$ (matrix-matrix multiplication)	MGS-SVL [Barlow, 2019] (called "MGS2")	MGS-CWY [Swirydowicz et al., 2020]
$(I - Q_k T_k^{-T} Q_k^T)x_{k+1}$ (triangular solve)	MGS-LTS [Swirydowicz et al., 2020]	MGS-ICWY [Swirydowicz et al., 2020]

[Puglisi, 1992]
[Björck, 1994]

Low-sync variants

[Schreiber and Van Loan, 1989] + "Sheffield observation"

	Two synchronizations	One synchronization
$(I - Q_k T_k^T Q_k^T) x_{k+1}$ (matrix-matrix multiplication)	MGS-SVL [Barlow, 2019] (called "MGS2")	MGS-CWY [Swirydowicz et al., 2020]
$(I - Q_k T_k^{-T} Q_k^T) x_{k+1}$ (triangular solve)	MGS-LTS [Swirydowicz et al., 2020]	MGS-ICWY [Swirydowicz et al., 2020]

[Puglisi, 1992]
[Björck, 1994]

Note: For block variants, it will be a crucial point that these algorithms return T_k

MGS Variants

Algorithm	$\|I - \bar{Q}^T \bar{Q}\|$	Assumption on $\kappa(X)$	References
MGS	$O(\varepsilon)\kappa(X)$	$O(\varepsilon)\kappa(X) < 1$	[Björck, 1967]
MGS-SVL	$O(\varepsilon)\kappa(X)$	$O(\varepsilon)\kappa(X) < 1$	[Barlow, 2019]
MGS-LTS	$O(\varepsilon)\kappa(X)$	$O(\varepsilon)\kappa(X) < 1$	conjecture [Świrydowicz, Langou, Ananthan, Yang, Thomas, 2020]
MGS-CWY	$O(\varepsilon)\kappa(X)$	$O(\varepsilon)\kappa(X) < 1$	conjecture [Świrydowicz, Langou, Ananthan, Yang, Thomas, 2020]
MGS-ICWY	$O(\varepsilon)\kappa(X)$	$O(\varepsilon)\kappa(X) < 1$	conjecture [Świrydowicz, Langou, Ananthan, Yang, Thomas, 2020]
MGS+	$O(\varepsilon)$	$O(\varepsilon)\kappa(X) < 1$	[Jalby, Philippe, 1991] [Giraud, Langou, 2002]
MGSI+	$O(\varepsilon)$	$O(\varepsilon)\kappa(X) < 1$	[Hoffmann, 1989] [Gander, 1980] [Giraud, Langou, Rozložník, 2002]

Other Muscles

Algorithm	$\|I - \bar{Q}^T\bar{Q}\|$	Assumption on $\kappa(X)$	References
CholQR	$O(\varepsilon)\kappa^2(X)$	$O(\varepsilon)\kappa^2(X) < 1$	[Yamamoto, Nakatsukasa, Yanagisawa, Fukaya, 2015]
CholQR+	$O(\varepsilon)$	$O(\varepsilon)\kappa^2(X) < 1$	[Yamamoto, Nakatsukasa, Yanagisawa, Fukaya, 2015]
ShCholQR++	$O(\varepsilon)$	$O(\varepsilon)\kappa(X) < 1$	[Fukaya, Kannan, Nakatsukasa, Yamamoto, Yanagisawa, 2020]
HouseQR	$O(\varepsilon)$	none	[Wilkinson, 1965]
GivensQR	$O(\varepsilon)$	none	[Wilkinson, 1965]
TSQR	$O(\varepsilon)$	none	[Mori, Yamamoto, Zhang, 2012] [Demmel, Grigori, Hoemmen, Langou, 2012]

Block Classical Gram-Schmidt (BCGS)

Block CGS

$[\mathcal{Q}, \mathcal{R}] = \texttt{BCGS}(\boldsymbol{\mathcal{X}})$

1: $[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}\,(\boldsymbol{X}_1)$
2: **for** $k = 1, \ldots, p-1$ **do**
3: $\quad \mathcal{R}_{1:k,k+1} = \boldsymbol{\mathcal{Q}}_{1:k}^T \boldsymbol{X}_{k+1}$
4: $\quad \boldsymbol{W} = \boldsymbol{X}_{k+1} - \boldsymbol{\mathcal{Q}}_{1:k} \mathcal{R}_{1:k,k+1}$
5: $\quad [\boldsymbol{Q}_{k+1}, R_{k+1,k+1}] = \texttt{IntraOrtho}\,(\boldsymbol{W})$
6: **end for**
7: **return** $\boldsymbol{\mathcal{Q}} = [\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_p]$, $\mathcal{R} = (R_{jk})$

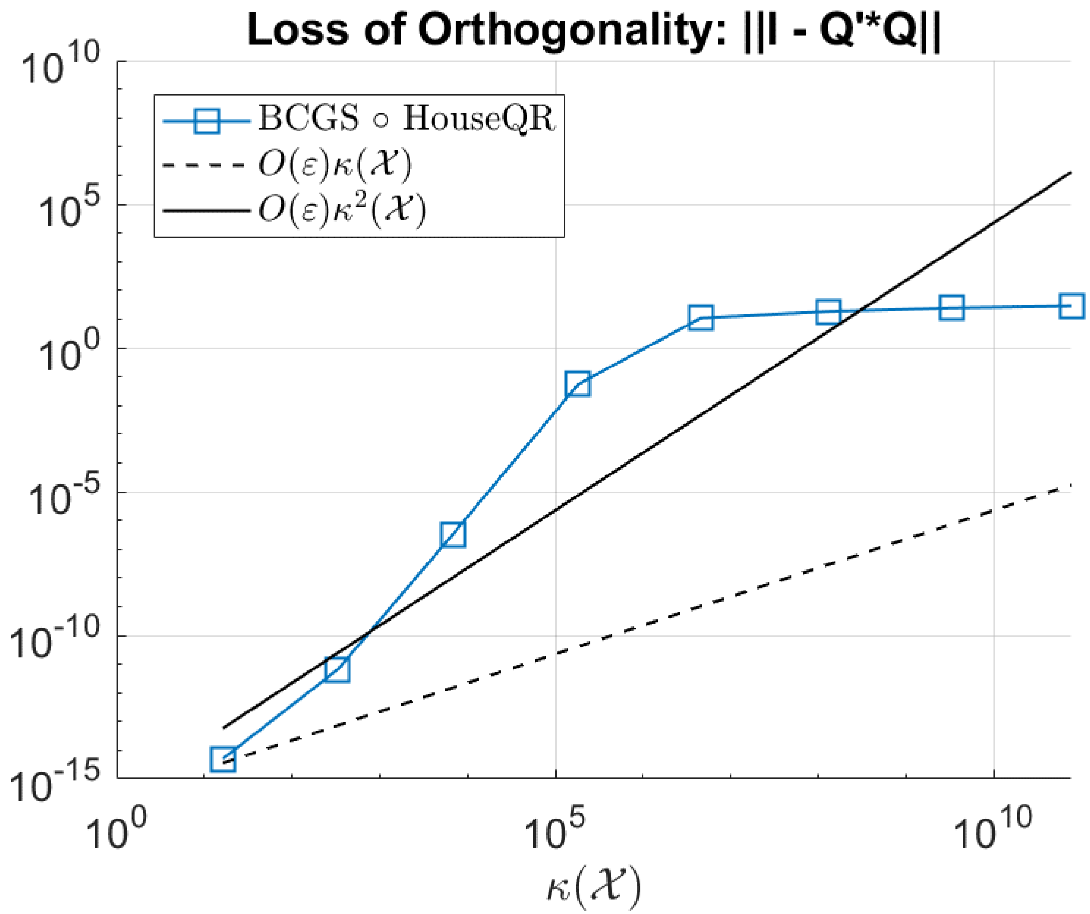
No existing proof of the loss of orthogonality in BCGS!

Conjecture: Even if our IntraOrtho has $O(\varepsilon)$ loss of orthogonality, BCGS is just as bad as CGS:

$$\|I - \bar{\mathcal{Q}}^T \bar{\mathcal{Q}}\| \leq O(\varepsilon)\kappa^{n-1}(\mathcal{X})$$

“Glued” matrices from [Smoktunowicz, Barlow, Langou, 2006]

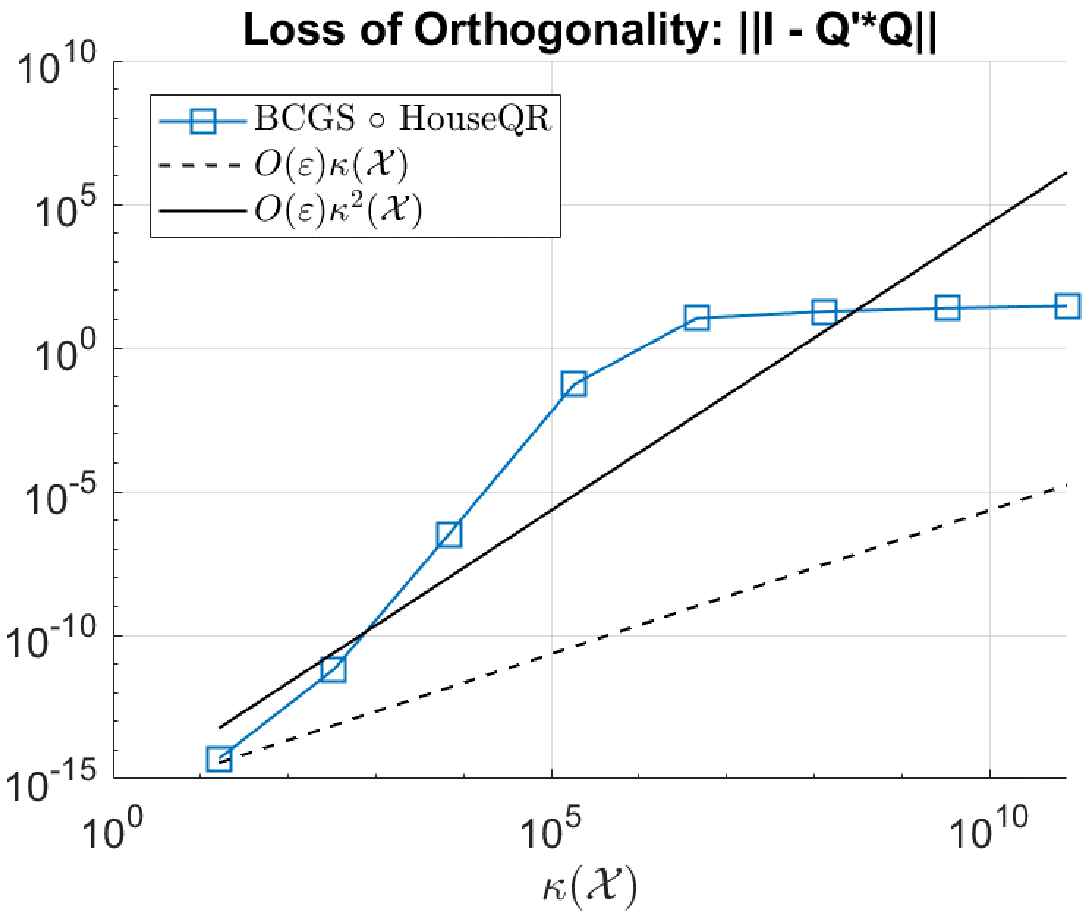
$m = 1000, p = 50, s = 4$



`GluedBlockKappaPlot([1000 50 4], 1:8, {'BCGS'}, {'HouseQR'})`

"Glued" matrices from [Smoktunowicz, Barlow, Langou, 2006]

$m = 1000, p = 50, s = 4$



BCGS loss of orthogonality is ***not*** $O(\varepsilon)\kappa^2(\mathcal{X})$!

```
GluedBlockKappaPlot([1000 50 4], 1:8, {'BCGS'}, {'HouseQR'})
```

Block Pythagorean CGS

$$W_{k+1} = X_{k+1} - \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1}$$

$$[Q_{k+1}, R_{k+1,k+1}] = \text{IntraOrtho}(W_{k+1})$$

$[\mathcal{Q}, \mathcal{R}] = \texttt{BCGS}(\boldsymbol{\mathcal{X}})$

1: $[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}\,(\boldsymbol{X}_1)$
2: **for** $k = 1, \ldots, p-1$ **do**
3: $\quad \mathcal{R}_{1:k,k+1} = \boldsymbol{\mathcal{Q}}_{1:k}^T \boldsymbol{X}_{k+1}$
4: $\quad \boldsymbol{W} = \boldsymbol{X}_{k+1} - \boldsymbol{\mathcal{Q}}_{1:k}\mathcal{R}_{1:k,k+1}$
5: $\quad [\boldsymbol{Q}_{k+1}, R_{k+1,k+1}] = \texttt{IntraOrtho}\,(\boldsymbol{W})$
6: **end for**
7: **return** $\boldsymbol{\mathcal{Q}} = [\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_p]$, $\mathcal{R} = (R_{jk})$

Block Pythagorean CGS

$$W_{k+1} = X_{k+1} - \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1}$$

$$[Q_{k+1}, R_{k+1,k+1}] = \text{IntraOrtho}(W_{k+1})$$

$[\boldsymbol{\mathcal{Q}}, \mathcal{R}] = \texttt{BCGS}(\boldsymbol{\mathcal{X}})$

$[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}(\boldsymbol{X}_1)$
for $k = 1, \ldots, p-1$ **do**
  $\mathcal{R}_{1:k,k+1} = \boldsymbol{\mathcal{Q}}_{1:k}^T \boldsymbol{X}_{k+1}$
  $\boldsymbol{W} = \boldsymbol{X}_{k+1} - \boldsymbol{\mathcal{Q}}_{1:k}\mathcal{R}_{1:k,k+1}$
  $[\boldsymbol{Q}_{k+1}, R_{k+1,k+1}] = \texttt{IntraOrtho}(\boldsymbol{W})$
end for
return $\boldsymbol{\mathcal{Q}} = [\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_p]$, $\mathcal{R} = (R_{jk})$

Block Pythagorean CGS

$$W_{k+1} = X_{k+1} - \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1}$$

$$[Q_{k+1}, R_{k+1,k+1}] = \text{IntraOrtho}(W_{k+1})$$

$$X_{k+1} = \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1} + W_{k+1}$$

$[\boldsymbol{\mathcal{Q}}, \mathcal{R}] = \texttt{BCGS}(\boldsymbol{\mathcal{X}})$

$[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}\,(\boldsymbol{X}_1)$

for $k = 1, \ldots, p-1$ **do**

  $\mathcal{R}_{1:k,k+1} = \boldsymbol{\mathcal{Q}}_{1:k}^T \boldsymbol{X}_{k+1}$

  $\boldsymbol{W} = \boldsymbol{X}_{k+1} - \boldsymbol{\mathcal{Q}}_{1:k}\mathcal{R}_{1:k,k+1}$

  $[\boldsymbol{Q}_{k+1}, R_{k+1,k+1}] = \texttt{IntraOrtho}\,(\boldsymbol{W})$

end for

return $\boldsymbol{\mathcal{Q}} = [\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_p]$, $\mathcal{R} = (R_{jk})$

Block Pythagorean CGS

$$W_{k+1} = X_{k+1} - \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1}$$

$$[Q_{k+1}, R_{k+1,k+1}] = \text{IntraOrtho}(W_{k+1})$$

$$X_{k+1} = \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1} + W_{k+1}$$

$[\mathcal{Q}, \mathcal{R}] = \texttt{BCGS}(\boldsymbol{\mathcal{X}})$

1: $[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}(\boldsymbol{X}_1)$
2: **for** $k = 1, \ldots, p-1$ **do**
3: $\quad \mathcal{R}_{1:k,k+1} = \boldsymbol{\mathcal{Q}}_{1:k}^T \boldsymbol{X}_{k+1}$
4: $\quad \boldsymbol{W} = \boldsymbol{X}_{k+1} - \boldsymbol{\mathcal{Q}}_{1:k}\mathcal{R}_{1:k,k+1}$
5: $\quad [\boldsymbol{Q}_{k+1}, R_{k+1,k+1}] = \texttt{IntraOrtho}(\boldsymbol{W})$
6: **end for**
7: **return** $\boldsymbol{\mathcal{Q}} = [\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_p]$, $\mathcal{R} = (R_{jk})$

Let $\mathcal{R}_{1:k,k+1} = Q_{\mathcal{R}} P_{k+1}$ be the QR factorization of $\mathcal{R}_{1:k,k+1}$

Let $X_{k+1} = Q_X T_{k+1}$ be the QR factorization of X_{k+1}

Block Pythagorean CGS

$$W_{k+1} = X_{k+1} - \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1}$$

$$[Q_{k+1}, R_{k+1,k+1}] = \text{IntraOrtho}(W_{k+1})$$

$$X_{k+1} = \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1} + W_{k+1}$$

$[\mathcal{Q}, \mathcal{R}] = \texttt{BCGS}(\boldsymbol{\mathcal{X}})$

1: $[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}\,(\boldsymbol{X}_1)$
2: **for** $k = 1, \ldots, p-1$ **do**
3: $\quad \mathcal{R}_{1:k,k+1} = \boldsymbol{\mathcal{Q}}_{1:k}^T \boldsymbol{X}_{k+1}$
4: $\quad \boldsymbol{W} = \boldsymbol{X}_{k+1} - \boldsymbol{\mathcal{Q}}_{1:k}\mathcal{R}_{1:k,k+1}$
5: $\quad [\boldsymbol{Q}_{k+1}, R_{k+1,k+1}] = \texttt{IntraOrtho}\,(\boldsymbol{W})$
6: **end for**
7: **return** $\boldsymbol{\mathcal{Q}} = [\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_p]$, $\mathcal{R} = (R_{jk})$

Let $\mathcal{R}_{1:k,k+1} = Q_{\mathcal{R}} P_{k+1}$ be the QR factorization of $\mathcal{R}_{1:k,k+1}$

Let $X_{k+1} = Q_X T_{k+1}$ be the QR factorization of X_{k+1}

$$T_{k+1}^T T_{k+1} = X_{k+1}^T X_{k+1}$$

Block Pythagorean CGS

$$W_{k+1} = X_{k+1} - \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1}$$

$$[Q_{k+1}, R_{k+1,k+1}] = \text{IntraOrtho}(W_{k+1})$$

$[\mathcal{Q}, \mathcal{R}] = \texttt{BCGS}(\mathcal{X})$

1: $[Q_1, R_{11}] = \texttt{IntraOrtho}(X_1)$
2: **for** $k = 1, \ldots, p-1$ **do**
3: $\quad \mathcal{R}_{1:k,k+1} = \mathcal{Q}_{1:k}^T X_{k+1}$
4: $\quad W = X_{k+1} - \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1}$
5: $\quad [Q_{k+1}, R_{k+1,k+1}] = \texttt{IntraOrtho}(W)$
6: **end for**
7: **return** $\mathcal{Q} = [Q_1, \ldots, Q_p]$, $\mathcal{R} = (R_{jk})$

$$X_{k+1} = \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1} + W_{k+1}$$

Let $\mathcal{R}_{1:k,k+1} = Q_{\mathcal{R}} P_{k+1}$ be the QR factorization of $\mathcal{R}_{1:k,k+1}$

Let $X_{k+1} = Q_X T_{k+1}$ be the QR factorization of X_{k+1}

$$\begin{aligned} T_{k+1}^T T_{k+1} &= X_{k+1}^T X_{k+1} \\ &= W_{k+1}^T W_{k+1} + \mathcal{R}_{1:k,k+1}^T \mathcal{R}_{1:k,k+1} \end{aligned}$$

Block Pythagorean CGS

$$W_{k+1} = X_{k+1} - \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1}$$

$$[Q_{k+1}, R_{k+1,k+1}] = \text{IntraOrtho}(W_{k+1})$$

$$X_{k+1} = \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1} + W_{k+1}$$

Let $\mathcal{R}_{1:k,k+1} = Q_{\mathcal{R}} P_{k+1}$ be the QR factorization of $\mathcal{R}_{1:k,k+1}$

Let $X_{k+1} = Q_X T_{k+1}$ be the QR factorization of X_{k+1}

$$\begin{aligned} T_{k+1}^T T_{k+1} &= X_{k+1}^T X_{k+1} \\ &= W_{k+1}^T W_{k+1} + \mathcal{R}_{1:k,k+1}^T \mathcal{R}_{1:k,k+1} \\ &= R_{k+1,k+1}^T R_{k+1,k+1} + P_{k+1}^T P_{k+1} \end{aligned}$$

$[\boldsymbol{\mathcal{Q}}, \mathcal{R}] = \texttt{BCGS}(\boldsymbol{\mathcal{X}})$

$[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}(\boldsymbol{X}_1)$
for $k = 1, \ldots, p-1$ **do**
  $\mathcal{R}_{1:k,k+1} = \boldsymbol{\mathcal{Q}}_{1:k}^T \boldsymbol{X}_{k+1}$
  $\boldsymbol{W} = \boldsymbol{X}_{k+1} - \boldsymbol{\mathcal{Q}}_{1:k}\mathcal{R}_{1:k,k+1}$
  $[\boldsymbol{Q}_{k+1}, R_{k+1,k+1}] = \texttt{IntraOrtho}(\boldsymbol{W})$
end for
return $\boldsymbol{\mathcal{Q}} = [\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_p]$, $\mathcal{R} = (R_{jk})$

Block Pythagorean CGS

$$W_{k+1} = X_{k+1} - \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1}$$
$$[Q_{k+1}, R_{k+1,k+1}] = \text{IntraOrtho}(W_{k+1})$$

$$X_{k+1} = \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1} + W_{k+1}$$

$[\boldsymbol{\mathcal{Q}}, \mathcal{R}] = \texttt{BCGS}(\boldsymbol{\mathcal{X}})$
$[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}(\boldsymbol{X}_1)$
for $k = 1, \ldots, p-1$ **do**
  $\mathcal{R}_{1:k,k+1} = \boldsymbol{\mathcal{Q}}_{1:k}^T \boldsymbol{X}_{k+1}$
  $\boldsymbol{W} = \boldsymbol{X}_{k+1} - \boldsymbol{\mathcal{Q}}_{1:k}\mathcal{R}_{1:k,k+1}$
  $[\boldsymbol{Q}_{k+1}, R_{k+1,k+1}] = \texttt{IntraOrtho}(\boldsymbol{W})$
end for
return $\boldsymbol{\mathcal{Q}} = [\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_p]$, $\mathcal{R} = (R_{jk})$

Let $\mathcal{R}_{1:k,k+1} = Q_{\mathcal{R}} P_{k+1}$ be the QR factorization of $\mathcal{R}_{1:k,k+1}$
Let $X_{k+1} = Q_X T_{k+1}$ be the QR factorization of X_{k+1}

$$\begin{aligned} T_{k+1}^T T_{k+1} &= X_{k+1}^T X_{k+1} \\ &= W_{k+1}^T W_{k+1} + \mathcal{R}_{1:k,k+1}^T \mathcal{R}_{1:k,k+1} \\ &= R_{k+1,k+1}^T R_{k+1,k+1} + P_{k+1}^T P_{k+1} \end{aligned}$$

$$R_{k+1,k+1} = \text{chol}\left(X_{k+1}^T X_{k+1} - \mathcal{R}_{1:k,k+1}^T \mathcal{R}_{1:k,k+1}\right) = \text{chol}\left(T_{k+1}^T T_{k+1} - P_{k+1}^T P_{k+1}\right)$$

Block Pythagorean CGS

$$W_{k+1} = X_{k+1} - \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1}$$

$$[Q_{k+1}, R_{k+1,k+1}] = \text{IntraOrtho}(W_{k+1})$$

$$X_{k+1} = \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1} + W_{k+1}$$

$[\mathcal{Q}, \mathcal{R}] = \texttt{BCGS}(\boldsymbol{\mathcal{X}})$

$[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}\,(\boldsymbol{X}_1)$
for $k = 1, \ldots, p-1$ **do**
  $\mathcal{R}_{1:k,k+1} = \boldsymbol{\mathcal{Q}}_{1:k}^T \boldsymbol{X}_{k+1}$
  $\boldsymbol{W} = \boldsymbol{X}_{k+1} - \boldsymbol{\mathcal{Q}}_{1:k}\mathcal{R}_{1:k,k+1}$
  $[\boldsymbol{Q}_{k+1}, R_{k+1,k+1}] = \texttt{IntraOrtho}\,(\boldsymbol{W})$
end for
return $\boldsymbol{\mathcal{Q}} = [\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_p]$, $\mathcal{R} = (R_{jk})$

Let $\mathcal{R}_{1:k,k+1} = Q_{\mathcal{R}} P_{k+1}$ be the QR factorization of $\mathcal{R}_{1:k,k+1}$

Let $X_{k+1} = Q_X T_{k+1}$ be the QR factorization of X_{k+1}

$$\begin{aligned} T_{k+1}^T T_{k+1} &= X_{k+1}^T X_{k+1} \\ &= W_{k+1}^T W_{k+1} + \mathcal{R}_{1:k,k+1}^T \mathcal{R}_{1:k,k+1} \\ &= R_{k+1,k+1}^T R_{k+1,k+1} + P_{k+1}^T P_{k+1} \end{aligned}$$

$$R_{k+1,k+1} = \underbrace{\text{chol}\left(X_{k+1}^T X_{k+1} - \mathcal{R}_{1:k,k+1}^T \mathcal{R}_{1:k,k+1}\right)}_{\text{BCGS-PIP}} = \underbrace{\text{chol}\left(T_{k+1}^T T_{k+1} - P_{k+1}^T P_{k+1}\right)}_{\text{BCGS-PIO}}$$

BCGS-PIP and BCGS-PIO

$[\mathcal{Q}, \mathcal{R}] = \texttt{BCGS}(\mathcal{X})$

$[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}\,(\boldsymbol{X}_1)$
for $k = 1, \ldots, p-1$ **do**
    $\mathcal{R}_{1:k,k+1} = \mathcal{Q}_{1:k}^T \boldsymbol{X}_{k+1}$
    $\boldsymbol{W} = \boldsymbol{X}_{k+1} - \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1}$
    $[\boldsymbol{Q}_{k+1}, R_{k+1,k+1}] = \texttt{IntraOrtho}\,(\boldsymbol{W})$
end for
return $\mathcal{Q} = [\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_p]$, $\mathcal{R} = (R_{jk})$

$[\mathcal{Q}, \mathcal{R}] = \texttt{BCGS-PIP}(\mathcal{X})$

$[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}\,(\boldsymbol{X}_1)$
for $k = 1, \ldots, p-1$ **do**
    $\begin{bmatrix} \mathcal{R}_{1:k,k+1} \\ \mathcal{Z}_{k+1} \end{bmatrix} = [\mathcal{Q}_{1:k}\ \ \boldsymbol{X}_{k+1}]^T \boldsymbol{X}_{k+1}$
    $R_{k+1,k+1} = \texttt{chol}(\mathcal{Z}_{k+1} - \mathcal{R}_{1:k,k+1}^T \mathcal{R}_{1:k,k+1})$
    $\boldsymbol{W}_{k+1} = \boldsymbol{X}_{k+1} - \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1}$
    $\boldsymbol{Q}_{k+1} = \boldsymbol{W}_{k+1} R_{k+1,k+1}^{-1}$
end for

$[\mathcal{Q}, \mathcal{R}] = \texttt{BCGS-PIO}(\mathcal{X})$

$[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}\,(\boldsymbol{X}_1)$
for $k = 1, \ldots, p-1$ **do**
    $\mathcal{R}_{1:k,k+1} = \mathcal{Q}_{1:k}^T \boldsymbol{X}_{k+1}$
    $\left[\sim, \begin{bmatrix} T_{k+1} & \\ & P_{k+1} \end{bmatrix}\right] = \texttt{IntraOrtho}\left(\begin{bmatrix} \boldsymbol{X}_{k+1} & \\ & \mathcal{R}_{1:k,k+1} \end{bmatrix}\right)$
    $R_{k+1,k+1} = \texttt{chol}(T_{k+1}^T T_{k+1} - P_{k+1}^T P_{k+1})$
    $\boldsymbol{W}_{k+1} = \boldsymbol{X}_{k+1} - \mathcal{Q}_{1:k}\mathcal{R}_{1:k,k+1}$
    $\boldsymbol{Q}_{k+1} = \boldsymbol{W}_{k+1} R_{k+1,k+1}^{-1}$
end for

- See [C., Lund, Rozložník, Thomas, 2020] and [C., Lund, Rozložník, 2021]
- BCGS-PIP also developed independently by [Yamazaki, Thomas, Hoemmen, Boman, Świrydowicz, Elliott, 2020]; called "CGS+CholQR"

New Stability Results for BCGS-PIP/PIO

Let $\mathcal{X} \in \mathbb{R}^{m \times n}$ be a matrix whose columns are organized into p blocks of size s, and assume that

$$O(\varepsilon)\kappa^2(\mathcal{X}) < 1.$$

Suppose we execute BCGS-PIP $\circ$ IntraOrtho($\mathcal{X}$) or BCGS-PIO $\circ$ IntraOrtho($\mathcal{X}$) on a machine with unit roundoff ε.

If for all X, IntraOrtho(X) computes factors $\bar{Q}$ and $\bar{R}$ that satisfy

$$\begin{aligned} \bar{R}^T\bar{R} &= X^TX + \Delta E, & \|\Delta E\| &\leq O(\varepsilon)\|X\|^2, \quad \text{and} \\ \bar{Q}\bar{R} &= X + \Delta D, & \|\Delta D\| &\leq O(\varepsilon)(\|X\| + \|\bar{Q}\|\|\bar{R}\|), \end{aligned}$$

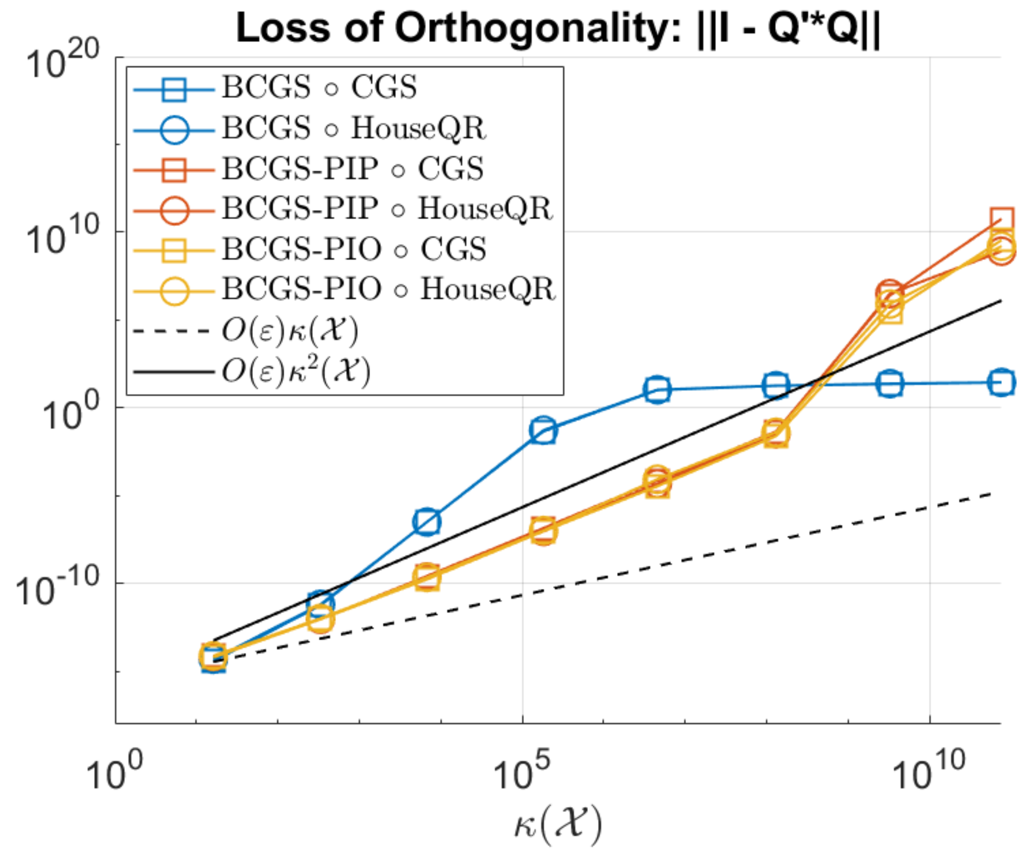
then the factors $\bar{\mathcal{Q}}$ and $\bar{\mathcal{R}}$ satisfy

$$\begin{aligned} &\|I - \bar{\mathcal{Q}}^T\bar{\mathcal{Q}}\| \leq O(\varepsilon)\kappa^2(\mathcal{X}), \quad \text{and} \\ &\bar{\mathcal{Q}}\bar{\mathcal{R}} = \mathcal{X} + \Delta\mathcal{D}, \quad \|\Delta\mathcal{D}\| \leq O(\varepsilon)\|\mathcal{X}\|. \end{aligned}$$

[C., Lund, Rozložník, 2021]

"Glued" matrices from [Smoktunowicz, Barlow, Langou, 2006]

$m = 1000, p = 50, s = 4$




```
GluedBlockKappaPlot([1000 50 4], 1:8, {'BCGS', 'BCGS_PIP_FREE', 'BCGS_PIO_FREE'}, {'CGS','HouseQR'})
```

Reorthogonalized Block Gram-Schmidt Variants

BCGSI+

[Barlow and Smoktunowicz, 2013]: If we have an IntraOrtho with $\|I - \bar{Q}^T\bar{Q}\| \leq O(\varepsilon)$ and if $O(\varepsilon)\kappa(\mathcal{X}) < 1$, then for BCGSI+,

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

$[\boldsymbol{\mathcal{Q}}, \mathcal{R}] = \texttt{BCGSI+}(\boldsymbol{\mathcal{X}})$

Allocate memory for $\boldsymbol{\mathcal{Q}}$ and $\mathcal{R}$
$[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}\,(\boldsymbol{X}_1)$
for $k = 1, \ldots, p-1$ **do**
  $\mathcal{R}^{(1)}_{1:k,k+1} = \boldsymbol{\mathcal{Q}}^T_{1:k}\boldsymbol{X}_{k+1}$ % first BCGS step
  $\boldsymbol{W} = \boldsymbol{X}_{k+1} - \boldsymbol{\mathcal{Q}}_{1:k}\mathcal{R}^{(1)}_{1:k,k+1}$
  $[\widehat{\boldsymbol{Q}}, R^{(1)}_{k+1,k+1}] = \texttt{IntraOrtho}\,(\boldsymbol{W})$
  $\mathcal{R}^{(2)}_{1:k,k+1} = \boldsymbol{\mathcal{Q}}^T_{1:k}\widehat{\boldsymbol{Q}}$ % second BCGS step
  $\boldsymbol{W} = \widehat{\boldsymbol{Q}} - \boldsymbol{\mathcal{Q}}_{1:k}\mathcal{R}^{(2)}_{1:k,k+1}$
  $[\boldsymbol{Q}_{k+1}, R^{(2)}_{k+1,k+1}] = \texttt{IntraOrtho}\,(\boldsymbol{W})$
  $\mathcal{R}_{1:k,k+1} = \mathcal{R}^{(1)}_{1:k,k+1} + \mathcal{R}^{(2)}_{1:k,k+1}R^{(1)}_{k+1,k+1}$
  $R_{k+1,k+1} = R^{(2)}_{k+1,k+1}R^{(1)}_{k+1,k+1}$
end for
return $\boldsymbol{\mathcal{Q}} = [\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_p]$, $\mathcal{R} = (R_{jk})$

BCGSI+

[Barlow and Smoktunowicz, 2013]: If we have an IntraOrtho with $\|I - \bar{Q}^T\bar{Q}\| \leq O(\varepsilon)$ and if $O(\varepsilon)\kappa(\mathcal{X}) < 1$, then for BCGSI+,

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

Key approach: Obtain bounds for the subproblem in every step of BCGSI+:

Given a near left-orthogonal matrix $\mathcal{U} \in \mathbb{R}^{m\times t}$ and a matrix $B \in \mathbb{R}^{m\times s}$, find $S \in \mathbb{R}^{t\times s}$, upper triangular $R_B \in \mathbb{R}^{s\times s}$, and left-orthogonal Q such that

$$B = \mathcal{U}S + QR_B \quad \text{and} \quad \mathcal{U}^TQ \approx 0.$$

*Requires the a priori assumption that $O(\varepsilon)\|B\|\left\|\bar{R}_B^{-1}\right\| < 1$

$[\mathcal{Q}, \mathcal{R}] = \texttt{BCGSI+}(\boldsymbol{\mathcal{X}})$

Allocate memory for $\boldsymbol{\mathcal{Q}}$ and $\mathcal{R}$
$[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}\,(\boldsymbol{X}_1)$
for $k = 1, \dots, p-1$ **do**
  $\mathcal{R}^{(1)}_{1:k,k+1} = \boldsymbol{\mathcal{Q}}^T_{1:k}\boldsymbol{X}_{k+1}$ % first BCGS step
  $\boldsymbol{W} = \boldsymbol{X}_{k+1} - \boldsymbol{\mathcal{Q}}_{1:k}\mathcal{R}^{(1)}_{1:k,k+1}$
  $[\widehat{\boldsymbol{Q}}, R^{(1)}_{k+1,k+1}] = \texttt{IntraOrtho}\,(\boldsymbol{W})$
  $\mathcal{R}^{(2)}_{1:k,k+1} = \boldsymbol{\mathcal{Q}}^T_{1:k}\widehat{\boldsymbol{Q}}$ % second BCGS step
  $\boldsymbol{W} = \widehat{\boldsymbol{Q}} - \boldsymbol{\mathcal{Q}}_{1:k}\mathcal{R}^{(2)}_{1:k,k+1}$
  $[\boldsymbol{Q}_{k+1}, R^{(2)}_{k+1,k+1}] = \texttt{IntraOrtho}\,(\boldsymbol{W})$
  $\mathcal{R}_{1:k,k+1} = \mathcal{R}^{(1)}_{1:k,k+1} + \mathcal{R}^{(2)}_{1:k,k+1}R^{(1)}_{k+1,k+1}$
  $R_{k+1,k+1} = R^{(2)}_{k+1,k+1}R^{(1)}_{k+1,k+1}$
end for
return $\boldsymbol{\mathcal{Q}} = [\boldsymbol{Q}_1, \dots, \boldsymbol{Q}_p]$, $\mathcal{R} = (R_{jk})$

BCGSI+

[Barlow and Smoktunowicz, 2013]: If we have an IntraOrtho with $\|I - \bar{Q}^T\bar{Q}\| \leq O(\varepsilon)$ and if $O(\varepsilon)\kappa(\mathcal{X}) < 1$, then for BCGSI+,

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

Key approach: Obtain bounds for the subproblem in every step of BCGSI+:

Given a near left-orthogonal matrix $\mathcal{U} \in \mathbb{R}^{m\times t}$ and a matrix $B \in \mathbb{R}^{m\times s}$, find $S \in \mathbb{R}^{t\times s}$, upper triangular $R_B \in \mathbb{R}^{s\times s}$, and left-orthogonal Q such that

$$B = \mathcal{U}S + QR_B \quad \text{and} \quad \mathcal{U}^T Q \approx 0.$$

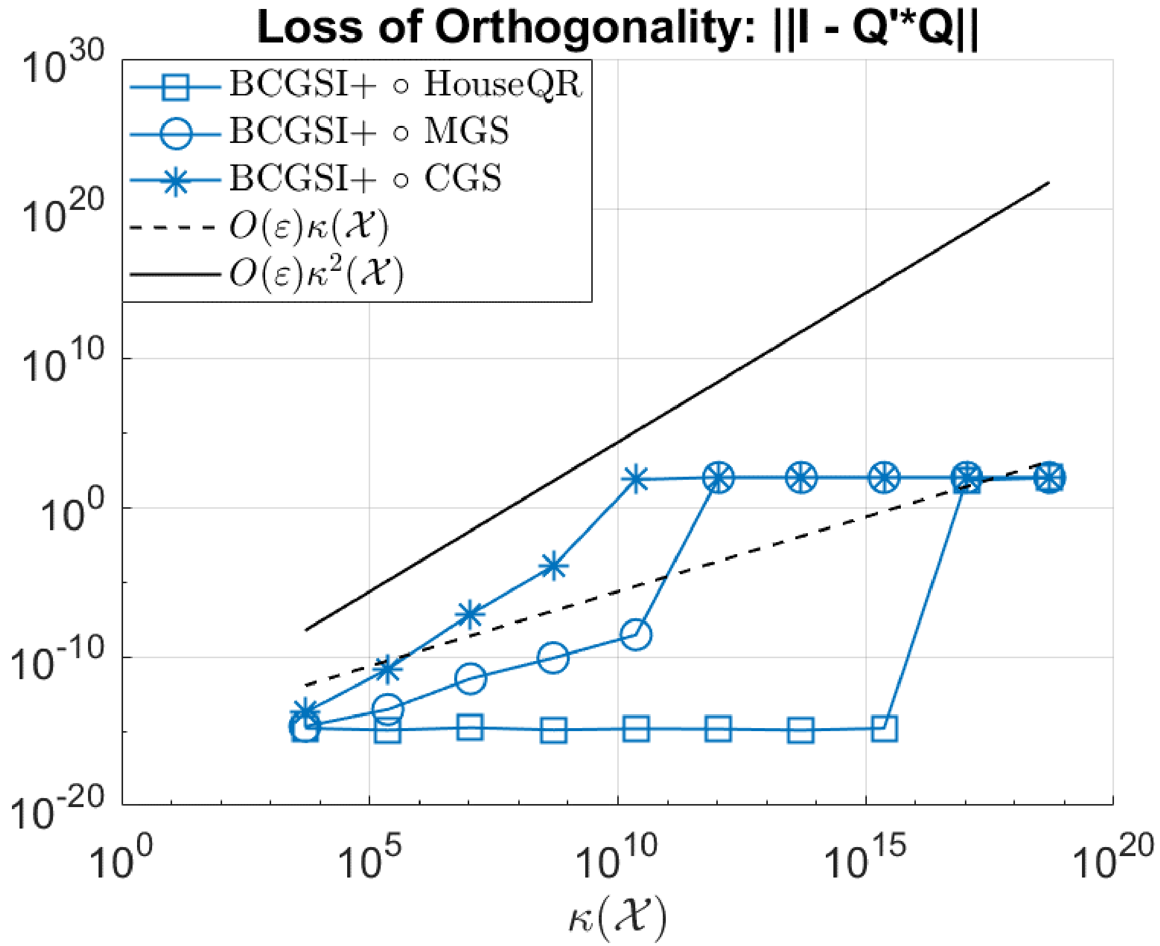
$[\boldsymbol{\mathcal{Q}}, \mathcal{R}] = \texttt{BCGSI+}(\boldsymbol{\mathcal{X}})$

Allocate memory for $\boldsymbol{\mathcal{Q}}$ and $\mathcal{R}$
$[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}\,(\boldsymbol{X}_1)$
for $k = 1, \ldots, p-1$ **do**
  $\mathcal{R}^{(1)}_{1:k,k+1} = \boldsymbol{\mathcal{Q}}^T_{1:k}\boldsymbol{X}_{k+1}$ % first BCGS step
  $\boldsymbol{W} = \boldsymbol{X}_{k+1} - \boldsymbol{\mathcal{Q}}_{1:k}\mathcal{R}^{(1)}_{1:k,k+1}$
  $[\widehat{\boldsymbol{Q}}, R^{(1)}_{k+1,k+1}] = \texttt{IntraOrtho}\,(\boldsymbol{W})$
  $\mathcal{R}^{(2)}_{1:k,k+1} = \boldsymbol{\mathcal{Q}}^T_{1:k}\widehat{\boldsymbol{Q}}$ % second BCGS step
  $\boldsymbol{W} = \widehat{\boldsymbol{Q}} - \boldsymbol{\mathcal{Q}}_{1:k}\mathcal{R}^{(2)}_{1:k,k+1}$
  $[\boldsymbol{Q}_{k+1}, R^{(2)}_{k+1,k+1}] = \texttt{IntraOrtho}\,(\boldsymbol{W})$
  $\mathcal{R}_{1:k,k+1} = \mathcal{R}^{(1)}_{1:k,k+1} + \mathcal{R}^{(2)}_{1:k,k+1}R^{(1)}_{k+1,k+1}$
  $R_{k+1,k+1} = R^{(2)}_{k+1,k+1}R^{(1)}_{k+1,k+1}$
end for
return $\boldsymbol{\mathcal{Q}} = [\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_p]$, $\mathcal{R} = (R_{jk})$

*Requires the a priori assumption that $O(\varepsilon)\|B\|\left\|\bar{R}_B^{-1}\right\| < 1$

*The requirement that *IntraOrtho has* $\|I - \bar{Q}^T\bar{Q}\| \leq O(\varepsilon)$ *is needed to guarantee that* Q_1 *is near left-orthogonal*; proof proceeds via induction.

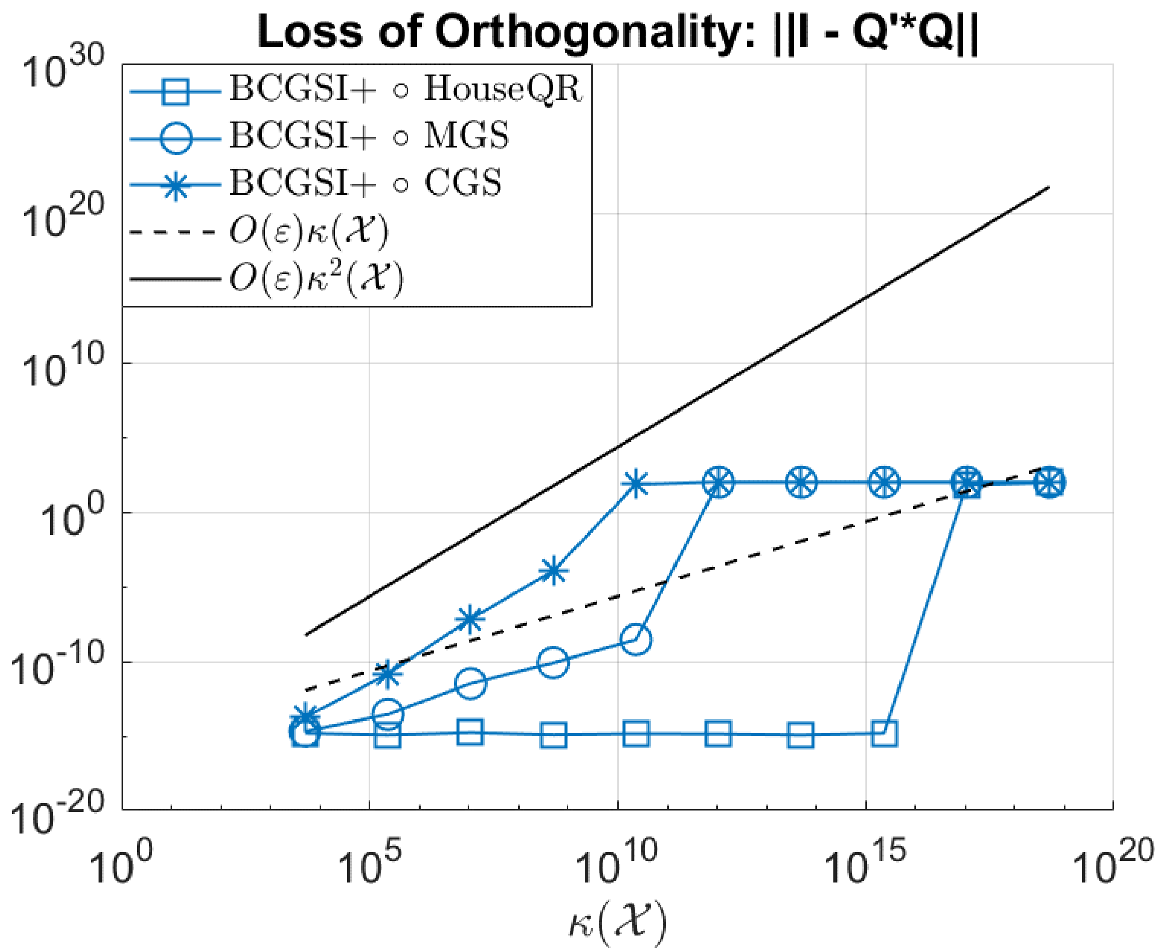
Läuchli matrix
$m = 1000, p = 100, s = 5$




```
LaeuchliBlockKappaPlot([1000 100 5],  logspace(-1, -16, 10), {'BCGS_IRO'}, {'HouseQR', 'MGS', 'CGS'})
```

Läuchli matrix

$m = 1000, p = 100, s = 5$

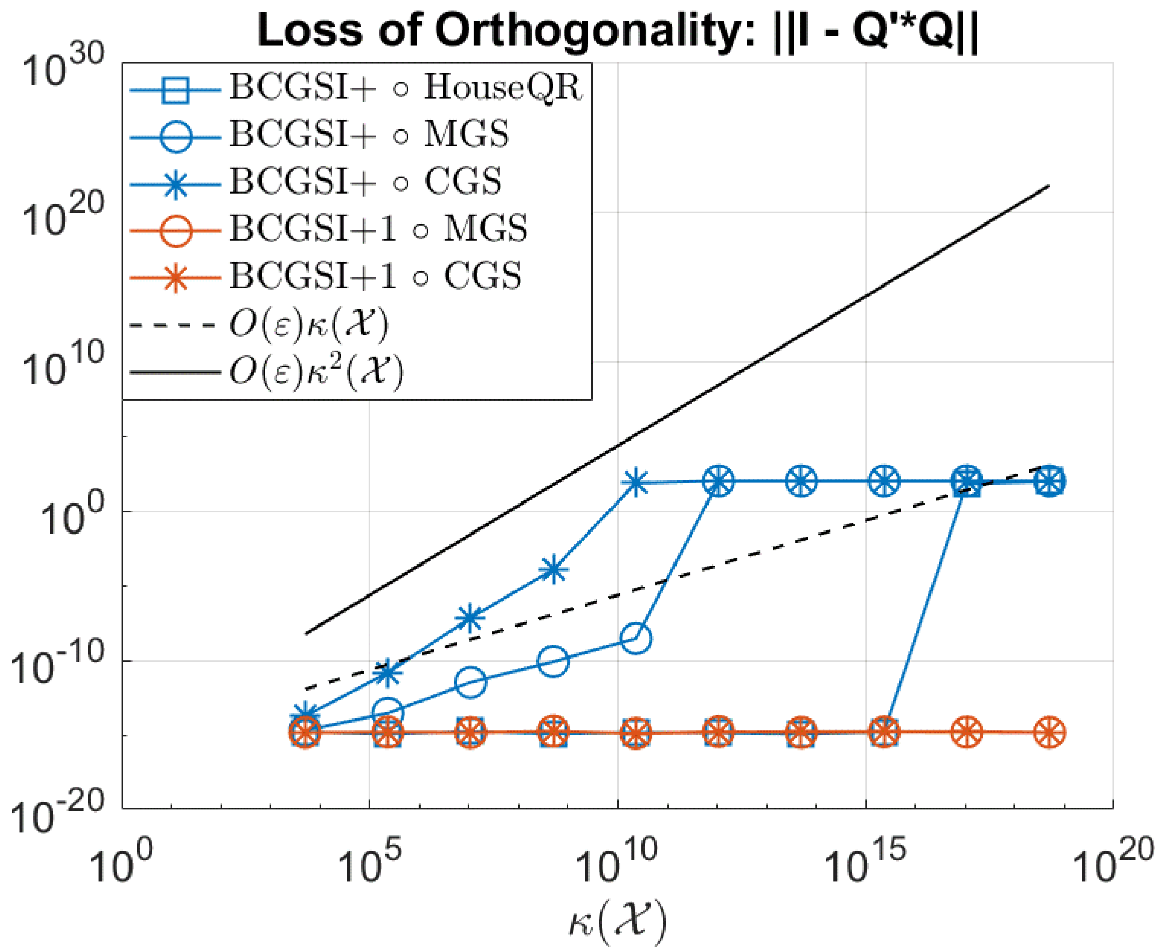


Recall: need $\left\| I - \bar{Q}_1{}^T \bar{Q}_1 \right\| \leq O(\varepsilon)$ to satisfy base case

⇒ Idea: What if we use a less stable IntraOrtho and just reorthogonalize the first block Q_1?

```
LaeuchliBlockKappaPlot([1000 100 5],  logspace(-1, -16, 10), {'BCGS_IRO'}, {'HouseQR', 'MGS', 'CGS'})
```

Läuchli matrix
$m = 1000, p = 100, s = 5$



BCGSI+1 [C., Lund, Rozložník, Thomas]:
Reorthogonalization on the first block to ensure $\left\|I - \bar{Q}_1^T \bar{Q}_1\right\| \leq O(\varepsilon)$

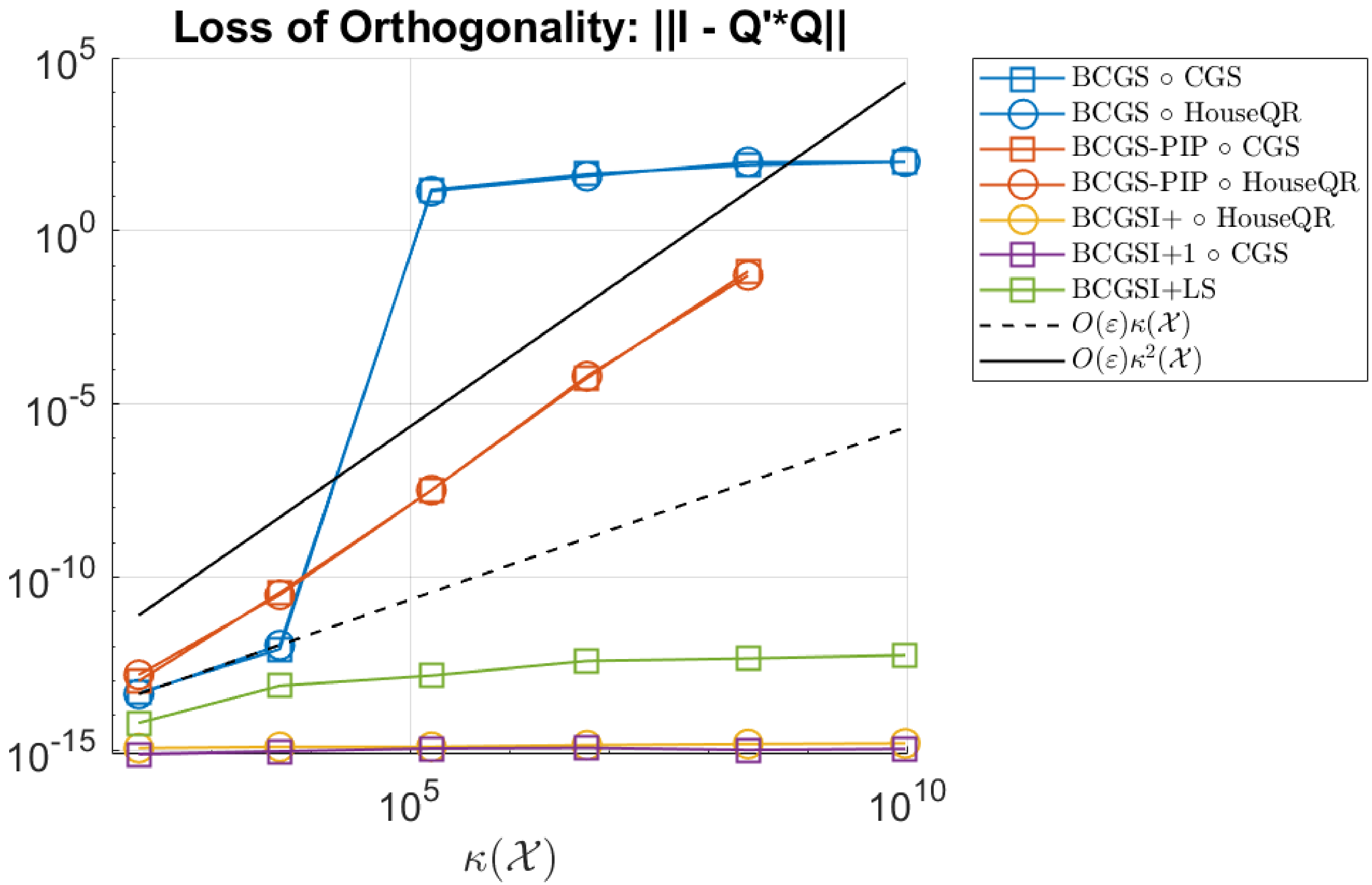
```
LaeuchliBlockKappaPlot([1000 100 5],  logspace(-1, -16, 10), {'BCGS_IRO','BCGS_IRO_1'}, {'HouseQR','MGS', 'CGS'})
```

BCGSI+LS

- Block generalization of CGSI+LS
- Equivalent algorithm given in [Yamazaki, Thomas, Hoemmen, Boman, Świrydowicz, Elliott, 2020] (see Figure 3)
- Notice: no IntraOrtho
- Conjectured that CGSI+LS has $O(\varepsilon)$ loss of orthogonality
  - What about BCGSI+LS?

$[\boldsymbol{\mathcal{Q}}, \mathcal{R}] = \texttt{BCGSI+LS}(\boldsymbol{\mathcal{X}})$

Allocate memory for $\boldsymbol{\mathcal{Q}}$ and $\mathcal{R}$
$\boldsymbol{U} = \boldsymbol{X}_1$
for $k = 2, \ldots p$ **do**
  if $k = 2$ **then**
    $[R_{k-1,k-1}^T R_{k-1,k-1} \; P] = \boldsymbol{U}^T [\boldsymbol{U} \; \boldsymbol{X}_k]$
  else if $k > 2$ **then**
    $\begin{bmatrix} \boldsymbol{W} & \boldsymbol{Z} \\ \Omega & Z \end{bmatrix} = [\boldsymbol{\mathcal{Q}}_{1:k-2} \; \boldsymbol{U}]^T [\boldsymbol{U} \; \boldsymbol{X}_k]$
    $[R_{k-1,k-1}^T R_{k-1,k-1} \; P] = [\Omega \; Z] - \boldsymbol{W}^T [\boldsymbol{W} \; \boldsymbol{Z}]$
  end if
  $R_{k-1,k} = R_{k-1,k-1}^{-T} P$
  if $k = 2$ **then**
    $\boldsymbol{Q}_{k-1} = \boldsymbol{U} R_{k-1,k-1}^{-1}$
  else if $k > 2$ **then**
    $\mathcal{R}_{1:k-2,k-1} = \mathcal{R}_{1:k-2,k-1} + \boldsymbol{W}$
    $\mathcal{R}_{1:k-2,k} = \boldsymbol{Z}$
    $\boldsymbol{Q}_{k-1} = (\boldsymbol{U} - \boldsymbol{\mathcal{Q}}_{1:k-2} \boldsymbol{W}) R_{k-1,k-1}^{-1}$
  end if
  $\boldsymbol{U} = \boldsymbol{X}_k - \boldsymbol{\mathcal{Q}}_{1:k-1} \mathcal{R}_{1:k-1,k}$
end for
$\begin{bmatrix} \boldsymbol{W} \\ \Omega \end{bmatrix} = [\boldsymbol{\mathcal{Q}}_{1:s-1} \; \boldsymbol{U}]^T \boldsymbol{U}$
$R_{s,s}^T R_{s,s} = \Omega - \boldsymbol{W}^T \boldsymbol{W}$
$\mathcal{R}_{1:s-1,s} = \mathcal{R}_{1:s-1,s} + \boldsymbol{W}$
$\boldsymbol{Q}_s = (\boldsymbol{U} - \boldsymbol{\mathcal{Q}}_{1:s-1} \boldsymbol{W}) R_{s,s}^{-1}$
return $\boldsymbol{\mathcal{Q}} = [\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_s]$, $\mathcal{R} = (R_{jk})$



"Monomial" matrices: Each block $X_k = [v_k, Av_k, \ldots, A^{s-1}v_k]$, where A is a diagonal $m \times m$ matrix with uniformly distributed eigenvalues in (.1,10) and v_k random

```
MonomialBlockKappaPlot([1000 120 2], 2:2:12, {'BCGS', 'BCGS_PIP','BCGS_IRO', 'BCGS_IRO_1','BCGS_IRO_LS'},
{'CGS', 'HouseQR'})
```

Summary: BCGS Variants

Algorithm	$\|I - \bar{Q}^T \bar{Q}\|$	Assumption on $\kappa(\mathcal{X})$	References
BCGS	$O(\varepsilon)\kappa^{n-1}(\mathcal{X})$	$O(\varepsilon)\kappa(\mathcal{X}) < 1$	conjecture
BCGS-P	$O(\varepsilon)\kappa^2(\mathcal{X})$	$O(\varepsilon)\kappa^2(\mathcal{X}) < 1$	[C., Lund, Rozložník, 2021]
BCGSI+	$O(\varepsilon)$	$O(\varepsilon)\kappa(\mathcal{X}) < 1$	[Barlow and Smoktunowicz, 2013]
BCGSI+1	$O(\varepsilon)$	$O(\varepsilon)\kappa(\mathcal{X}) < 1$	conjecture
BCGSS+rpl	$O(\varepsilon)$	none	conjecture, [Stewart, 2008]
BCGSI+LS	$O(\varepsilon)\kappa^2(\mathcal{X})$	?	conjecture

Block Modified Gram-Schmidt (BMGS)

Results of Jalby and Philippe (1991)

Intuition: From lines 4-8:

$$W = (I - Q_k Q_k^T) \cdots (I - Q_1 Q_1^T) X_{k+1}$$

Each projector $I - Q_j Q_j^T$ is equivalent to a step of CGS

- $\Rightarrow$ Underlying "CGS-like" nature of BMGS

$[\mathcal{Q}, \mathcal{R}] = \texttt{BMGS}(\mathcal{X})$

1: Allocate memory for $\mathcal{Q}$ and $\mathcal{R}$
2: $[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}(\boldsymbol{X}_1)$
3: **for** $k = 1, \ldots, p-1$ **do**
4: $\quad \boldsymbol{W} = \boldsymbol{X}_{k+1}$
5: $\quad$ **for** $j = 1, \ldots, k$ **do**
6: $\quad\quad R_{j,k+1} = \boldsymbol{Q}_j^T \boldsymbol{W}$
7: $\quad\quad \boldsymbol{W} = \boldsymbol{W} - \boldsymbol{Q}_j R_{j,k+1}$
8: $\quad$ **end for**
9: $\quad [\boldsymbol{Q}_{k+1}, R_{k+1,k+1}] = \texttt{IntraOrtho}(\boldsymbol{W})$
10: **end for**
11: **return** $\mathcal{Q} = [\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_p]$, $\mathcal{R} = (R_{jk})$

Results of Jalby and Philippe (1991)

Intuition: From lines 4-8:

$$W = \left(I - Q_k Q_k^T\right) \cdots \left(I - Q_1 Q_1^T\right) X_{k+1}$$

Each projector $I - Q_j Q_j^T$ is equivalent to a step of CGS

- $\Rightarrow$ Underlying "CGS-like" nature of BMGS

[Jalby and Philippe, 1991]:

- BMGS∘MGS behaves "like CGS"
- BMGS∘MGS+ is as stable as MGS

$[\mathcal{Q}, \mathcal{R}] = \texttt{BMGS}(\boldsymbol{\mathcal{X}})$

1: Allocate memory for $\mathcal{Q}$ and $\mathcal{R}$
2: $[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}(\boldsymbol{X}_1)$
3: **for** $k = 1, \ldots, p-1$ **do**
4: $\quad \boldsymbol{W} = \boldsymbol{X}_{k+1}$
5: $\quad$ **for** $j = 1, \ldots, k$ **do**
6: $\quad\quad R_{j,k+1} = \boldsymbol{Q}_j^T \boldsymbol{W}$
7: $\quad\quad \boldsymbol{W} = \boldsymbol{W} - \boldsymbol{Q}_j R_{j,k+1}$
8: $\quad$ **end for**
9: $\quad [\boldsymbol{Q}_{k+1}, R_{k+1,k+1}] = \texttt{IntraOrtho}(\boldsymbol{W})$
10: **end for**
11: **return** $\mathcal{Q} = [\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_p]$, $\mathcal{R} = (R_{jk})$

Results of Jalby and Philippe (1991)

Intuition: From lines 4-8:

$$W = (I - Q_k Q_k^T) \cdots (I - Q_1 Q_1^T) X_{k+1}$$

Each projector $I - Q_j Q_j^T$ is equivalent to a step of CGS

- $\Rightarrow$ Underlying "CGS-like" nature of BMGS

$[\mathcal{Q}, \mathcal{R}] = \texttt{BMGS}(\boldsymbol{\mathcal{X}})$

1: Allocate memory for $\mathcal{Q}$ and $\mathcal{R}$
2: $[\boldsymbol{Q}_1, R_{11}] = \texttt{IntraOrtho}(\boldsymbol{X}_1)$
3: **for** $k = 1, \ldots, p-1$ **do**
4: $\quad \boldsymbol{W} = \boldsymbol{X}_{k+1}$
5: $\quad$ **for** $j = 1, \ldots, k$ **do**
6: $\quad\quad R_{j,k+1} = \boldsymbol{Q}_j^T \boldsymbol{W}$
7: $\quad\quad \boldsymbol{W} = \boldsymbol{W} - \boldsymbol{Q}_j R_{j,k+1}$
8: $\quad$ **end for**
9: $\quad [\boldsymbol{Q}_{k+1}, R_{k+1,k+1}] = \texttt{IntraOrtho}(\boldsymbol{W})$
10: **end for**
11: **return** $\mathcal{Q} = [\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_p]$, $\mathcal{R} = (R_{jk})$

[Jalby and Philippe, 1991]:

- BMGS∘MGS behaves "like CGS"
- BMGS∘MGS+ is as stable as MGS

- Can manipulate proof of Theorem 4.1 from Jalby and Philippe's work to show that BMGS ∘ (any IntraOrtho with $\|I - \bar{Q}^T \bar{Q}\| \leq O(\varepsilon)$) is as stable as MGS

Block Low-Sync Variants

	Two synchronizations	One synchronization
$(I - Q_k T_k^T Q_k^T) x_{k+1}$ (matrix-matrix multiplication)	MGS-SVL [Barlow, 2019] (called "MGS2")	MGS-CWY [Świrydowicz et al., 2020]
$(I - Q_k T_k^{-T} Q_k^T) x_{k+1}$ (triangular solve)	MGS-LTS [Świrydowicz et al., 2020]	MGS-ICWY [Świrydowicz et al., 2020]

Block generalizations:

- BMGS-SVL∘MGS-SVL [Barlow, 2019] (called "MGS3")
- BMGS-SVL∘HouseQR [Barlow, 2019] (called "BMGS_H")
- BMGS-LTS: [C. Lund, Rozložník, Thomas, 2020]
- BMGS-CWY, BMGS-ICWY ([C. Lund, Rozložník, Thomas, 2020] and [Yamazaki et al., 2020])

Barlow's Analysis [2019]

Key quantities:

$$\Delta_{\mathcal{T}\mathcal{S}} = \bar{\mathcal{T}}\mathcal{S} - I$$
$$\Delta_{\bar{\mathcal{Q}}\bar{\mathcal{R}}} = \bar{\mathcal{Q}}\bar{\mathcal{R}} - \mathcal{X}$$
$$\Gamma_{\mathcal{T}\mathcal{R}} = (I - \bar{\mathcal{T}})\bar{\mathcal{R}}$$

$\mathcal{S} = \mathrm{triu}(\bar{\mathcal{Q}}^T\bar{\mathcal{Q}})$, and in exact arithmetic, $\mathcal{S} = \mathcal{T}^{-1}$.

Barlow's Analysis [2019]

Key quantities:

$$\Delta_{\mathcal{T}\mathcal{S}} = \bar{\mathcal{T}}\mathcal{S} - I$$
$$\Delta_{\bar{\mathcal{Q}}\bar{\mathcal{R}}} = \bar{\mathcal{Q}}\bar{\mathcal{R}} - \mathcal{X}$$
$$\Gamma_{\mathcal{T}\mathcal{R}} = (I - \bar{\mathcal{T}})\bar{\mathcal{R}}$$

$\mathcal{S} = \mathrm{triu}(\bar{\mathcal{Q}}^T\bar{\mathcal{Q}})$, and in exact arithmetic, $\mathcal{S} = \mathcal{T}^{-1}$.

[Barlow, 2019]: If $[\bar{Q}, \bar{R}, \bar{T}] = \mathrm{IntraOrtho}(X)$ satisfies

$$\Delta_{TS} = \bar{T}S - I_s, \quad \|\Delta_{TS}\|_F \leq O(\varepsilon)$$
$$\Delta_{\bar{Q}\bar{R}} = \bar{Q}\bar{R} - X, \quad \left\|\Delta_{\bar{Q}\bar{R}}\right\|_F \leq O(\varepsilon)\|X\|_F$$
$$\Gamma_{TR} = (I - \bar{T})\bar{R}, \quad \|\Gamma_{TR}\|_F \leq O(\varepsilon)\|X\|_F,$$

then BMGS-SVL $\circ$ IntraOrtho($\mathcal{X}$) satisfies

$$\|\Delta_{\mathcal{T}\mathcal{S}}\|_F \leq O(\varepsilon)$$
$$\|\Delta_{\bar{\mathcal{Q}}\bar{\mathcal{R}}}\|_F \leq O(\varepsilon)\|\mathcal{X}\|_F$$
$$\|\Gamma_{\mathcal{T}\mathcal{R}}\|_F \leq O(\varepsilon)\|\mathcal{X}\|_F.$$

Barlow's Analysis [2019]

Key quantities:

$$\Delta_{\mathcal{T}\mathcal{S}} = \bar{\mathcal{T}}\mathcal{S} - I$$
$$\Delta_{\bar{\mathcal{Q}}\bar{\mathcal{R}}} = \bar{\mathcal{Q}}\bar{\mathcal{R}} - \mathcal{X}$$
$$\Gamma_{\mathcal{T}\mathcal{R}} = (I - \bar{\mathcal{T}})\bar{\mathcal{R}}$$

$\mathcal{S} = \text{triu}(\bar{\mathcal{Q}}^T\bar{\mathcal{Q}})$, and in exact arithmetic, $\mathcal{S} = \mathcal{T}^{-1}$.

[Barlow, 2019]: If $[\bar{Q}, \bar{R}, \bar{T}] = \text{IntraOrtho}(X)$ satisfies

$$\Delta_{TS} = \bar{T}S - I_s, \quad \|\Delta_{TS}\|_F \leq O(\varepsilon)$$
$$\Delta_{\bar{Q}\bar{R}} = \bar{Q}\bar{R} - X\,, \quad \left\|\Delta_{\bar{Q}\bar{R}}\right\|_F \leq O(\varepsilon)\|X\|_F$$
$$\Gamma_{TR} = (I - \bar{T})\bar{R}, \quad \|\Gamma_{TR}\|_F \leq O(\varepsilon)\|X\|_F,$$

then BMGS-SVL $\circ$ IntraOrtho($\mathcal{X}$) satisfies

$$\left.\begin{aligned} &\|\Delta_{\mathcal{T}\mathcal{S}}\|_F \leq O(\varepsilon) \\ &\|\Delta_{\bar{\mathcal{Q}}\bar{\mathcal{R}}}\|_F \leq O(\varepsilon)\|\mathcal{X}\|_F \\ &\|\Gamma_{\mathcal{T}\mathcal{R}}\|_F \leq O(\varepsilon)\|\mathcal{X}\|_F. \end{aligned}\right\} \quad \begin{aligned} &\|I - \bar{\mathcal{Q}}^T\bar{\mathcal{Q}}\|_F \leq O(\varepsilon)\kappa(\mathcal{X}) \\ &\text{if } O(\varepsilon)\kappa(\mathcal{X}) < 1 \end{aligned}$$

Barlow proves directly that MGS-SVL satisfies the IntraOrtho constraints

Barlow's Analysis [2019]

- What about BMGS-SVL with IntraOrthos that don't produce T's?

Barlow's Analysis [2019]

- What about BMGS-SVL with IntraOrthos that don't produce T's?
- Implicitly, they produce $\bar{T} = I$
- In this case,

$$\|\Delta_{TS}\|_F = \|\bar{T}S - I\|_F = \|\mathrm{I} - \mathrm{triu}(\bar{Q}^T\bar{Q})\|_F \leq \|I - \bar{Q}^T\bar{Q}\|_F$$
$$\left\|\Delta_{\bar{Q}\bar{R}}\right\|_F = \|\bar{Q}\bar{R} - X\|_F$$
$$\|\Gamma_{TR}\|_F = \|(I - \bar{T})\bar{R}\|_F = 0$$

Barlow's Analysis [2019]

- What about BMGS-SVL with IntraOrthos that don't produce T's?
- Implicitly, they produce $\bar{T} = I$
- In this case,

$$\|\Delta_{TS}\|_F = \|\bar{T}S - I\|_F = \|\mathrm{I} - \mathrm{triu}(\bar{Q}^T\bar{Q})\|_F \leq \|I - \bar{Q}^T\bar{Q}\|_F$$
$$\left\|\Delta_{\bar{Q}\bar{R}}\right\|_F = \|\bar{Q}\bar{R} - X\|_F$$
$$\|\Gamma_{TR}\|_F = \|(I - \bar{T})\bar{R}\|_F = 0$$

For "non-T-based" IntraOrthos, must have $\|I - \bar{Q}^T\bar{Q}\|_F \leq O(\varepsilon)$

- HouseQR, CGSI+, TSQR, …

Barlow's Analysis [2019]

- What about BMGS-SVL with IntraOrthos that don't produce T's?
- Implicitly, they produce $\bar{T} = I$
- In this case,

$$\|\Delta_{TS}\|_F = \|\bar{T}S - I\|_F = \|\mathrm{I} - \mathrm{triu}(\bar{Q}^T\bar{Q})\|_F \le \|I - \bar{Q}^T\bar{Q}\|_F$$

$$\left\|\Delta_{\bar{Q}\bar{R}}\right\|_F = \|\bar{Q}\bar{R} - X\|_F$$

$$\|\Gamma_{TR}\|_F = \|(I - \bar{T})\bar{R}\|_F = 0$$

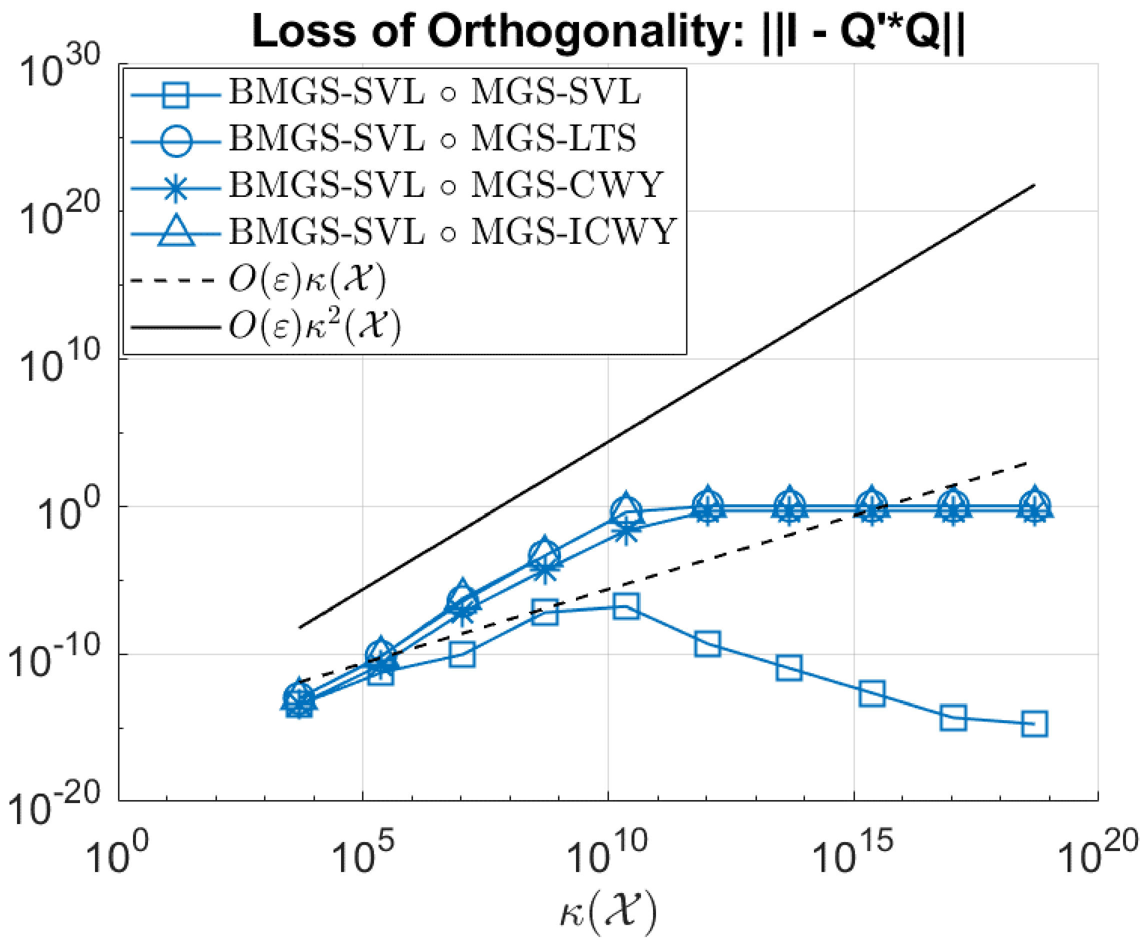
For "non-T-based" IntraOrthos, must have $\|I - \bar{Q}^T\bar{Q}\|_F \le O(\varepsilon)$

- HouseQR, CGSI+, TSQR, …

- What about BMGS with another T-variant IntraOrtho?

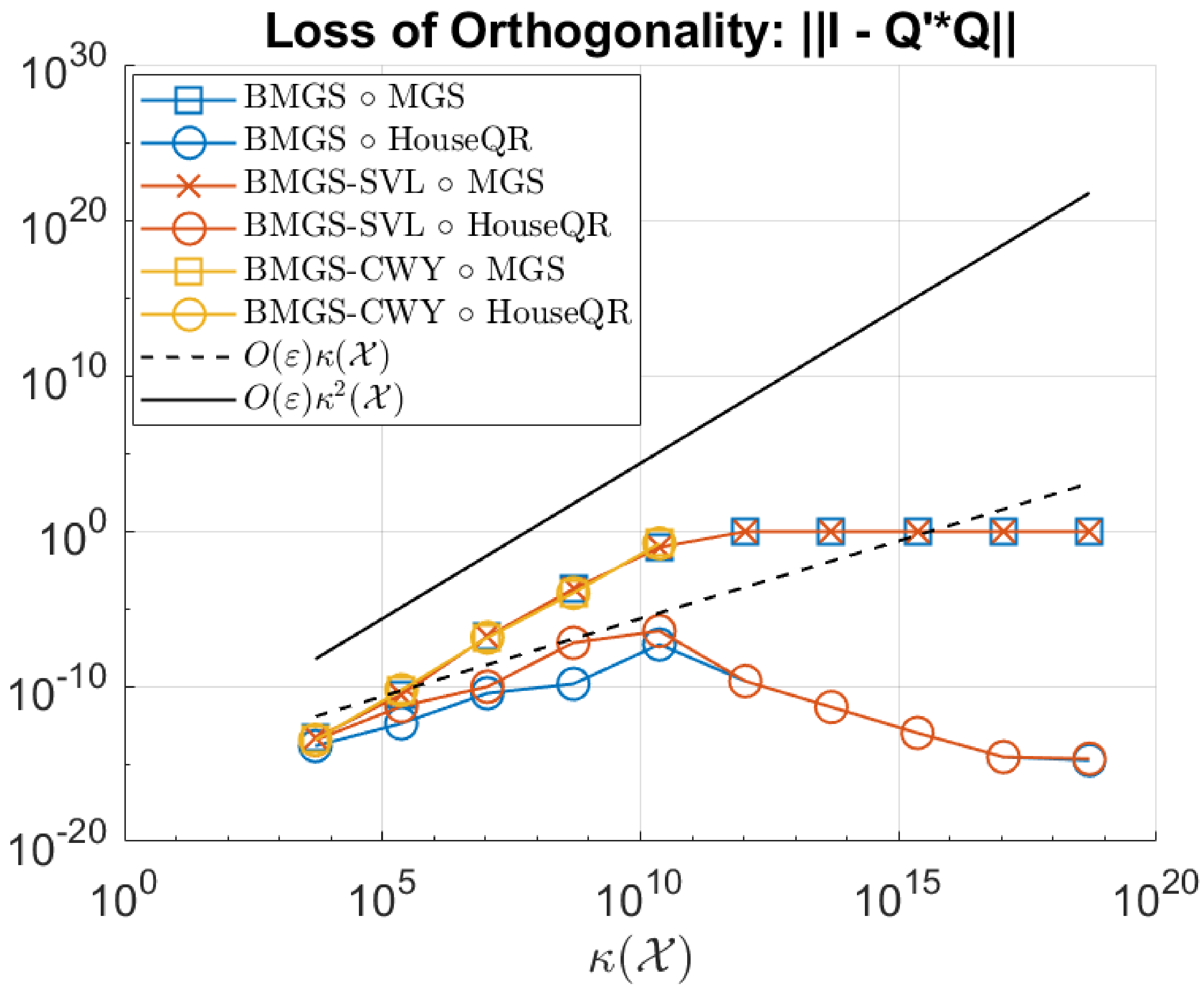
Läuchli matrix

$m = 1000, p = 100, s = 5$




```
LaeuchliBlockKappaPlot([1000 100 5],  logspace(-1, -16, 10), {'BMGS_SVL'}, {'MGS_SVL','MGS_LTS','MGS_CWY','MGS_ICWY'})
```

Läuchli matrix
$m = 1000, p = 100, s = 5$




```
LaeuchliBlockKappaPlot([1000 100 5],  logspace(-1, -16, 10), {'BMGS','BMGS_SVL','BMGS_CWY'}, {'MGS','HouseQR'})
```

BMGS Variants

Algorithm	IntraOrtho reqs.	$\|I - \bar{Q}^T\bar{Q}\|$	Assumption on $\kappa(\mathcal{X})$	References
BMGS	$O(\varepsilon)$	$O(\varepsilon)\kappa(\mathcal{X})$	$O(\varepsilon)\kappa(\mathcal{X}) < 1$	[Jalby and Philippe, 1991]
	$O(\varepsilon)\kappa(X)$	$O(\varepsilon)\kappa^2(\mathcal{X})$	$O(\varepsilon)\kappa(\mathcal{X}) < 1$	[Jalby and Philippe, 1991]
BMGS-SVL	$O(\varepsilon)$ or MGS-SVL	$O(\varepsilon)\kappa(\mathcal{X})$	$O(\varepsilon)\kappa(\mathcal{X}) < 1$	[Barlow, 2019]
BMGS-LTS	$O(\varepsilon)$ or MGS-LTS	$O(\varepsilon)\kappa(\mathcal{X})$	$O(\varepsilon)\kappa(\mathcal{X}) < 1$	conjecture
BMGS-CWY	any	$O(\varepsilon)\kappa^2(\mathcal{X})$	$O(\varepsilon)\kappa^2(\mathcal{X}) < 1$	conjecture
BMGS-ICWY	any	$O(\varepsilon)\kappa^2(\mathcal{X})$	$O(\varepsilon)\kappa^2(\mathcal{X}) < 1$	conjecture

Open Questions

Looking forward...

- Much work to do in proving stability, in particular for low-sync variants
  - What is the effect of normalization lag?
  - What skeletons work with what muscles?
- Randomized Gram-Schmidt variants [Balabanov, Grigori, 2020]
- Opportunities for mixed precision?
  - [Yamazaki, Tomov, Kurzak, Dongarra, Barlow, 2015]: mixed precision CholQR within BMGS and BCGS
  - [Yang, Fox, Sanders, 2019]: mixed precision HouseQR
- What are the necessary and sufficient conditions on degree of orthogonality in order to have backward stable block GMRES? Communication-avoiding GMRES?

Thank You!

carson@karlin.mff.cuni.cz
www.karlin.mff.cuni.cz/~carson/

Survey paper: https://arxiv.org/pdf/2010.12058.pdf
BCGS-P variants: http://www.math.cas.cz/fichier/preprints/IM_20210124200723_43.pdf
BlockStab MATLAB package: https://github.com/katlund/BlockStab