Stable and efficient QR factorization and least-squares solver based on CholeskyQR

Yuji Nakatsukasa
University of Oxford

Based on joint work with

T. Fukaya (Hokkaido), R. Kannan (Arup/Manchester),
Y. Yamamoto (Elec.-Comm.), Y. Yanagisawa (Waseda)

Advances in NLA: Celebrating J. H. Wilkinson
May 2019

Preface: power method for Jordan blocks

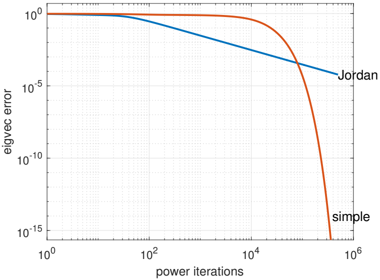
Power method: $x = Ax$, $x = \frac{x}{||x||}$, $\lambda = x^T Ax$, repeat

- Basis for QR algorithm, Krylov methods
- Converges geometrically $(\lambda, x) \to (\lambda_1, x_1)$ if $|\lambda_1| > |\lambda_2|$
- What if $|\lambda_1| = |\lambda_2|$?
  - If $\lambda_1 \neq \lambda_2$, nonconvergence
  - If $\lambda_1 = \lambda_2$ semisimple, $x \to$ eigvec $c_1 x_1 + c_2 x_2$, geometrically
  - **What if $\lambda_1 = \lambda_2$ defective?** (Jordan blocks)

Preface: power method for Jordan blocks

Power method: $x = Ax$, $x = \frac{x}{\|x\|}$, $\lambda = x^T Ax$, repeat

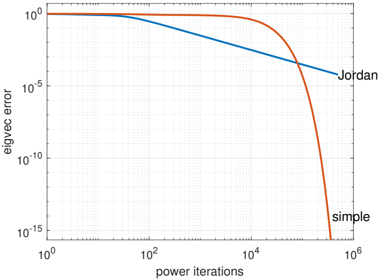
- Basis for QR algorithm, Krylov methods
- Converges geometrically $(\lambda, x) \to (\lambda_1, x_1)$ if $|\lambda_1| > |\lambda_2|$
- What if $|\lambda_1| = |\lambda_2|$?
  - If $\lambda_1 \neq \lambda_2$, nonconvergence
  - If $\lambda_1 = \lambda_2$ semisimple, $x \to$ eigvec $c_1 x_1 + c_2 x_2$, geometrically
  - **What if $\lambda_1 = \lambda_2$ defective?** (Jordan blocks)
    - Converges, but only algebraically



Preface: power method for Jordan blocks

Power method: $x = Ax$, $x = \frac{x}{\|x\|}$, $\lambda = x^T Ax$, repeat

- Basis for QR algorithm, Krylov methods
- Converges geometrically $(\lambda, x) \to (\lambda_1, x_1)$ if $|\lambda_1| > |\lambda_2|$
- What if $|\lambda_1| = |\lambda_2|$?
  - If $\lambda_1 \neq \lambda_2$, nonconvergence
  - If $\lambda_1 = \lambda_2$ semisimple, $x \to$ eigvec $c_1x_1 + c_2x_2$, geometrically
  - **What if $\lambda_1 = \lambda_2$ defective?** (Jordan blocks)
    - Converges, but only algebraically [Wilkinson AEP (65)]



Lesson: Always look at AEP for NLA results!

Cholesky QR for QR factorization

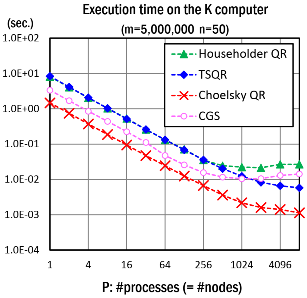
To compute $A = QR \in \mathbb{R}^{m\times n} (m \geq n)$, $Q^T Q = I_n$, R upper triangular

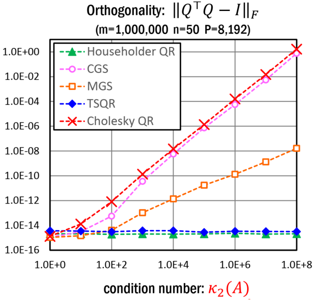
$$G = A^T A$$

$$R = \mathrm{chol}(G)$$

$$Q = AR^{-1}$$

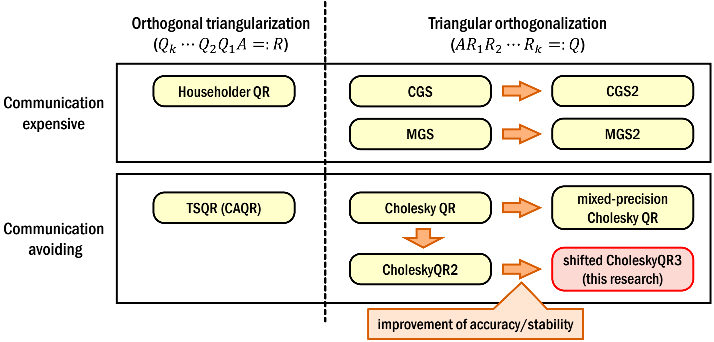
- Trivial extension to oblique inner product, quasimatrices etc
- Fast but unstable!
  - $\kappa_2(G) = \kappa_2(A)^2$, $\mathrm{chol}(G)$ often breaks down if $\kappa_2(A) \geq \mathbf{u}^{-1/2}$
  - Even if $\mathrm{chol}(G)$ computed, $\kappa_2(Q) - 1 \gg \mathbf{u}$ unless $\kappa_2(A) = O(1)$





This talk: improve stability and revive CholeskyQR

QR factorization algorithms



- Repeat to improve: "Twice is enough" [Kahan-Parlett]
- Orthogonal triangularization vs. triangular orthogonalization [Trefethen-Bau NLA (96)]
- Mixed-precision CholeskyQR: selective use of quadruple precision arithmetic [Yamazaki-Tomov-Dongarra (2015)]
- This work: use only double precision

CholeskyQR2

Observe that

- If $\kappa_2(A) < 10^8$, CholeskyQR avoids breakdown, but $\hat{Q}$ not orthogonal
- But $\kappa_2(\hat{Q}) \le 1.1$! and $\|A - \hat{Q}\hat{R}\|/\|A\| = O(u)$
- So repeat! Compute $\hat{Q}$'s QR fact. via CholeskyQR: $R' = \text{chol}(\hat{Q}^T\hat{Q})$, $\hat{Q} := \hat{Q}(R')^{-1}$, $\hat{R} := R'\hat{R}$ then $\kappa_2(\hat{Q}) \approx 1$

Pseudocode CholeskyQR2

```
[Q,R] = cholqr(A);
[Q2,R2] = cholqr(Q);
Q = Q2; R = R2*R;  % then A=QR, Q^TQ=I
```

Excellent stability: Computed $\hat{Q}, \hat{R}$ satisfy [Yamamoto et. al. (ETNA 15)]

$$\|\hat{Q}^T\hat{Q} - I_n\|_2 \le 7mn\mathbf{u}, \quad \frac{\|A - \hat{Q}\hat{R}\|_F}{\|A\|_F} \le 5n^3\mathbf{u}$$

CholeskyQR2

Observe that

- If $\kappa_2(A) < 10^8$, CholeskyQR avoids breakdown, but $\hat{Q}$ not orthogonal
- But $\kappa_2(\hat{Q}) \leq 1.1$**!** and $\|A - \hat{Q}\hat{R}\|/\|A\| = O(u)$
- So repeat! Compute $\hat{Q}$'s QR fact. via CholeskyQR: $R' = \text{chol}(\hat{Q}^T\hat{Q})$, $\hat{Q} := \hat{Q}(R')^{-1}$, $\hat{R} := R'\hat{R}$ then $\kappa_2(\hat{Q}) \approx 1$

Pseudocode CholeskyQR2

```
[Q,R] = cholqr(A);
[Q2,R2] = cholqr(Q);
Q = Q2; R = R2*R;  % then A=QR, Q^TQ=I
```

Excellent stability: Computed $\hat{Q}, \hat{R}$ satisfy [Yamamoto et. al. (ETNA 15)]

$$\|\hat{Q}^T\hat{Q} - I_n\|_2 \leq 7mn\mathbf{u}, \quad \frac{\|A - \hat{Q}\hat{R}\|_F}{\|A\|_F} \leq 5n^3\mathbf{u}$$

Can we remove $\kappa_2(A) < 10^8$ requirement?

shiftedCholeskyQR3

[Fukaya-Kannan-N.-Yamamoto-Yanagisawa ArXiv]

$$G = A^T A,$$

choose $s > 0$

$$R = \text{chol}(G + sI),$$

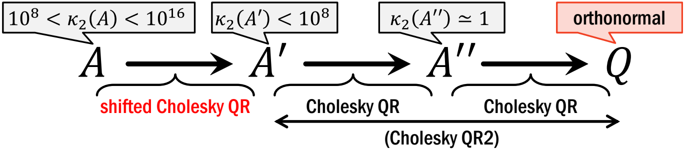
$$A' = AR^{-1}$$

under mild conditions $\kappa_2(A') < 10^8$, so use CholeskyQR2:

$$\hat{R} = \text{chol}(A'^T A'), \qquad A'' := A'\hat{R}^{-1}, \quad R := \hat{R}R$$

$$\hat{R} = \text{chol}(A''^T A''), \qquad Q := A''\hat{R}^{-1}, \quad R := \hat{R}R$$

yields Q orthonormal, R triangular



Essentially three CholeskyQR: **"thrice is enough"**

Choice of shift s

$$G = A^T A,$$
$$\text{choose} \quad s > 0$$
$$R = \text{chol}(G+sI),$$
$$A' = AR^{-1}$$

- small s: reduces $\kappa_2(A')$
- large s: $\text{chol}(G + sI)$ avoids breakdown This choice guarantees this:

$$s = 11(mn + n(n+1))\mathbf{u}\|A\|_2^2.$$

- Based on worst-case analysis; often smaller s suffices

Stability analysis sketch

$$G = A^T A,$$
$$\text{choose} \quad s > 0$$
$$R = \text{chol}(G + sI),$$
$$A' = AR^{-1}$$

Main goal: prove $\kappa_2(A) \leq 10^{16} \Rightarrow \kappa_2(\hat{A}') \lesssim 10^8$ with computed $\hat{A}'$

- $\|\hat{A}'\|_2 \leq 2$ straightforward
- for $\sigma_{\min}(A'^{-1})$,

1. write $\hat{R}^{-\top}(A^\top A + sI)\hat{R}^{-1} = I - \hat{R}^{-\top}(E_1 + E_2)\hat{R}^{-1}$
2. show $\|\hat{R}^{-\top}(E_1 + E_2)\hat{R}^{-1}\|_2 \leq \frac{1}{9}$
3. $\sigma_n(A\hat{R}^{-1}) \geq \frac{\sigma_n(A)}{\sqrt{(\sigma_n(A))^2+s}} \sqrt{1 - \|\hat{R}^{-\top}(E_1 + E_2)\hat{R}^{-1}\|} \geq 0.9 \frac{\sigma_n(A)}{\sqrt{(\sigma_n(A))^2+s}}$
4. show $\sigma_{\min}(\hat{A}') \geq (1 - 0.4)\sigma_n(A\hat{R}^{-1})$
5. take $s = O(\mathbf{u})$ to get $\sigma_{\min}(\hat{A}') \geq O(\mathbf{u}^{-\frac{1}{2}})$

When is thrice enough?

1. One shiftedCholeskyQR gives $\kappa_2(\hat{A}') \leq 2\sqrt{3(1 + \frac{s}{\sigma_{\min}(A)^2})}$
2. CholeskyQR2($\hat{A}'$) succeeds if $\kappa_2(\hat{A}') \leq \frac{1}{8\sqrt{(mn+n(n+1))\mathbf{u}}}$

Hence shiftedCholQR3 works if

$$\kappa(A) \leq \frac{1}{96(mn + n(n+1))}\mathbf{u}^{-1}.$$

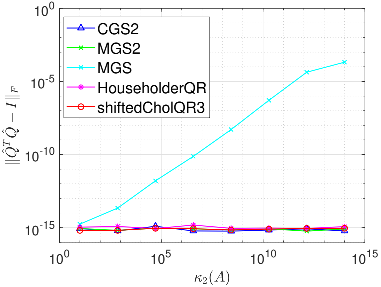
Then computed $\hat{Q}\hat{R} \approx A$ satisfy

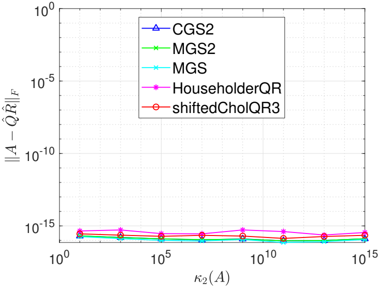
$$\|\hat{Q}^T\hat{Q} - I\|_F \leq 6(mn + n(n+1))\mathbf{u}, \qquad \frac{\|A - \hat{Q}\hat{R}\|_F}{\|A\|_2} \leq 15n^2\mathbf{u}.$$

- Comparable to CholeskyQR2
- “Better” bounds than (Modified)Gram-Schmidt2

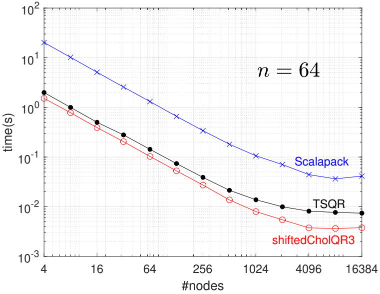
Experiments

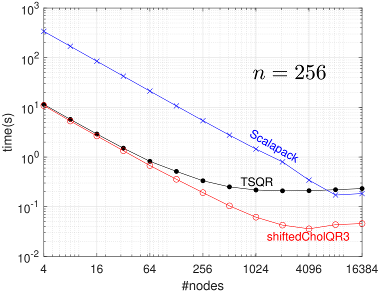
Stability ($m = 300, n = 50$)





Speed ($m \approx 1.6 \times 10^7$)





B-orthogonalization (oblique inner product)

Goal: given $A \in \mathbb{R}^{m\times n}$ and $B \succ 0 \in \mathbb{R}^{n\times n}$,
compute factorization $A = QR$, where $Q^T B Q = I_n$

- used e.g. in generalized eigenproblems

CholeskyQR straightforward to extend!

$$G = A^T B A$$

$$R = \text{chol}(G)$$
$$Q = AR^{-1}$$

B-orthogonalization (oblique inner product)

Goal: given $A \in \mathbb{R}^{m \times n}$ and $B \succ 0 \in \mathbb{R}^{n \times n}$,
compute factorization $A = QR$, where $Q^T B Q = I_n$

- used e.g. in generalized eigenproblems

shiftedCholeskyQR3 straightforward to extend!

$$G = A^T B A$$

choose $s > 0$

$$R = \text{chol}(G + sI)$$

$$A' = AR^{-1}$$

B-orthogonalization (oblique inner product)

Goal: given $A \in \mathbb{R}^{m\times n}$ and $B \succ 0 \in \mathbb{R}^{n\times n}$,
compute factorization $A = QR$, where $Q^T BQ = I_n$

- used e.g. in generalized eigenproblems

shiftedCholeskyQR3 straightforward to extend!

$$G = A^T BA$$
$$\text{choose} \quad s > 0$$
$$R = \text{chol}(G+sI)$$
$$A' = AR^{-1}$$

then (B-)CholeskyQR2

$$\hat{R} = \text{chol}(A'^T BA'), \qquad A'' := A'\hat{R}^{-1}, \quad R := \hat{R}R$$
$$\hat{R} = \text{chol}(A''^T BA''), \qquad Q := A''\hat{R}^{-1}, \quad R := \hat{R}R$$

Numerical stability for $B \neq I$

Provided that

$$\frac{\|A\|_2 \sqrt{\|B\|_2}}{\sqrt{\sigma_n(A^\top BA)}} \cdot \sqrt{\kappa_2(B)} \leq \frac{\mathbf{u}^{-1}}{96(2m\sqrt{mn} + n(n+1))},$$

1. B-orthogonality

$$\|\hat{Q}^T B\hat{Q} - I\|_F \leq 8[m\sqrt{mn}\mathbf{u} + n(n+1)\mathbf{u}]\kappa_2(B),$$

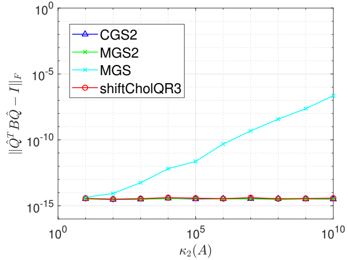
2. backward error

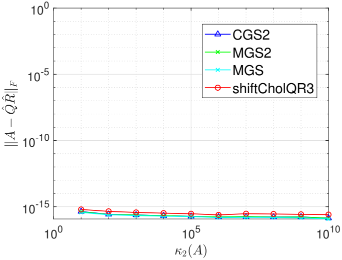
$$\frac{\|A - \hat{Q}\hat{R}\|_F}{\|\hat{A}\|_2} \leq 16n^2\mathbf{u}(\kappa_2(B))^{3/2}.$$

- condition reduces to $\kappa_2(A) \leq C\mathbf{u}^{-1}$ when $B = I$
- $\kappa_2(B) \gg 1$ apparent issue (not observed in practice)

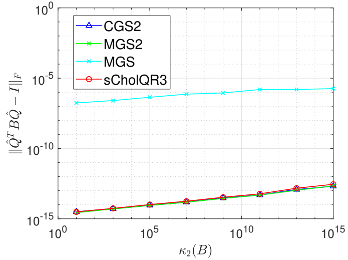
Experiments $B \neq I$

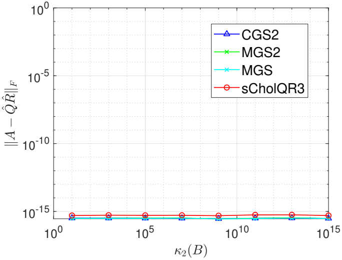
$\kappa_2(B) = 10^{10}$, vary $\kappa_2(A)$





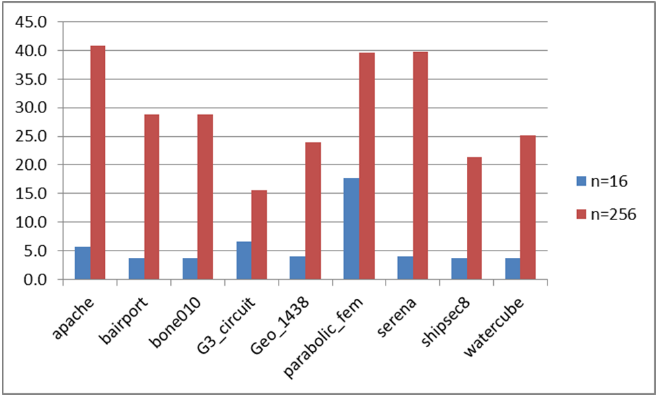
$\kappa_2(A) = 10^{10}$, vary $\kappa_2(B)$





Experiments $B \neq I$

Speedup factor over CGS2



n: number of CPUs

Least-squares problem

$$\min_x ||Ax - b||_2$$

(Direct) solvers

- Normal equation $(A^T A)x = A^T b$: unstable
- QR-based: $A = QR$, $Q^T Ax = Q^T b$
  - Houserholder QR: stable
  - classical Gram-Schmidt: unstable
  - modified Gram-Schmidt: better but unstable
  - modified Gram-Schmidt on $[A, b]$: stable
  - CholeskyQR: unstable $\Rightarrow$ fix!
    - But not with CholeskyQR2—prefer to not form $Q = AR^{-1}$

Least-squares problem via CholeskyQR

[w/ T. Fukaya]

$$\min_x \|Ax - b\|_2$$

If $\kappa_2(A) \leq 10^8$, with CholeskyQR $A = \widehat{Q}\widehat{R} + \epsilon$, $\kappa_2(\widehat{Q}) \leq 1.1$

- $x = \widehat{R}^{-1}\widehat{Q}^T b$ unstable
- but since $\kappa_2(\widehat{Q}) \leq 1.1$, $\widehat{Q}^T(Ax - b) = 0 \Leftrightarrow (\widehat{Q}^T A)x = \widehat{Q}^T b$ stable!

  (simple to prove via [Higham ASNA (02)])
- can we avoid forming $\widehat{Q}$?

Least-squares problem via CholeskyQR

[w/ T. Fukaya]

$$\min_x \|Ax - b\|_2$$

If $\kappa_2(A) \leq 10^8$, with CholeskyQR $A = \widehat{Q}\widehat{R} + \epsilon$, $\kappa_2(\widehat{Q}) \leq 1.1$

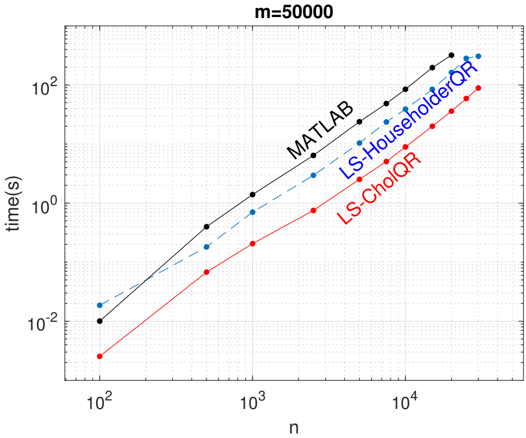
- $x = \widehat{R}^{-1}\widehat{Q}^T b$ unstable
- but since $\kappa_2(\widehat{Q}) \leq 1.1$, $\widehat{Q}^T(Ax - b) = 0 \Leftrightarrow (\widehat{Q}^T A)x = \widehat{Q}^T b$ stable!

  (simple to prove via [Higham ASNA (02)])
- can we avoid forming $\widehat{Q}$?
- LS-CholQR: $\widehat{Q}^T A = (A\widehat{R}^{-1})^T(A\widehat{R}^{-1})\widehat{R} =: G^T G\widehat{R}$, so $(\widehat{Q}^T A)x = \widehat{Q}^T b \Leftrightarrow G^T G y = \widehat{Q}^T b$ by CG, and $\widehat{R}x = y$
  - $\kappa_2(G^T G) \leq 1.2$! plain CG converges in a few iterations
  - each G-multiplication requires R-solve
  - stable with **iterative refinement** [Golub-Wilkinson (66)]:
    $\hat{x} = \text{argmin}_x \|(A + \Delta A)x - (b + \Delta b)\|_2, \quad \frac{\|\Delta A\|}{\|A\|} \leq \epsilon, \frac{\|\Delta b\|}{\|b\|} \leq \epsilon$

Combines normal eqn speed (almost) + Householder stability

Least-squares experiments

$\min_x \|Ax - b\|_2$, $\kappa_2(A) = 10^6$, $A \in \mathbb{R}^{50000\times n}$



- Cholesky-based LS solver 3-5 times faster than Householder
- About 10x faster than MATLAB's \ (pivots)
- Stability measures $\|Ax - b\|_2$, $\|x - x_*\|_2$ always comparable

Summary

CholeskyQR

- can be made stable by shifting+repeating for $\kappa_2(A) = O(u^{-1})$
- can be efficient+stably for least-squares if $\kappa_2(A) = O(u^{-1/2})$

postface:

- Wilkinson: The Algebraic Eigenvalue Problem (1965)
- Parlett: The Symmetric Eigenvalue Problem (1980)
- N.: The low-rank eigenvalue problem (arXiv yesterday, 5 pp.)

$$\mathrm{eig}_{\lambda\neq0}\left(A \; B \right) = \mathrm{eig}_{\lambda\neq0}\left(B \; A \right) = \mathrm{eig}_{\lambda\neq0}\left(BA \right)$$

Least-squares conditioning: a revisit

Suppose $\|A\| \approx 1$.

1. Error in computed solution $\hat{x}$ for

$$Ax = b \tag{1}$$

   is (of course) $\frac{\|x-\hat{x}\|}{\|x\|} = O(\mathbf{u}\kappa_2(A))$.

2. Error in computed solution $\hat{x}$ for [Golub-Wilkinson 66]

$$\min_x \|Ax - b\|_2 \tag{2}$$

   is $\frac{\|x-\hat{x}\|}{\|x\|} = O(\mathbf{u}\kappa_2(A)^2)$ if $b \notin \text{range}(A)$.

Least-squares conditioning: a revisit

Suppose $\|A\| \approx 1$.

1. Error in computed solution $\hat{x}$ for

$$Ax = b \qquad (1)$$

   is (of course) $\frac{\|x-\hat{x}\|}{\|x\|} = O(\mathbf{u}\kappa_2(A))$. But $\|x - \hat{x}\| = O(\mathbf{u}\kappa_2(A)^2)$!

2. Error in computed solution $\hat{x}$ for [Golub-Wilkinson 66]

$$\min_x \|Ax - b\|_2 \qquad (2)$$

   is $\frac{\|x-\hat{x}\|}{\|x\|} = O(\mathbf{u}\kappa_2(A)^2)$ if $b \notin \text{range}(A)$. And $\|x - \hat{x}\| = O(\mathbf{u}\kappa_2(A)^2)$

Least-squares conditioning: a revisit

Suppose $\|A\| \approx 1$.

1. Error in computed solution $\hat{x}$ for

$$Ax = b \tag{1}$$

   is (of course) $\frac{\|x-\hat{x}\|}{\|x\|} = O(\mathbf{u}\kappa_2(A))$. But $\|x - \hat{x}\| = O(\mathbf{u}\kappa_2(A)^2)$!

2. Error in computed solution $\hat{x}$ for [Golub-Wilkinson 66]

$$\min_x \|Ax - b\|_2 \tag{2}$$

   is $\frac{\|x-\hat{x}\|}{\|x\|} = O(\mathbf{u}\kappa_2(A)^2)$ if $b \notin \text{range}(A)$. And $\|x - \hat{x}\| = O(\mathbf{u}\kappa_2(A)^2)$

Absolute error is $O(\mathbf{u}\kappa_2(A)^2)$ generically (e.g. random b), for both (1) and (2)

- For (1), error large precisely when $\|x\| \gg 1$
- For (1), error large even when $\|x\| = O(1)$