Randomized Solution of Least Squares Problems

Ilse C.F. Ipsen

Joint work with James Garrison

North Carolina State University
Raleigh, NC, USA

Research supported by DOE, NSF-CCF, NSF-DMS

Full Column-Rank Least Squares Problems

Given $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ with $\operatorname{rank}(\boldsymbol{A}) = n$

$$\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2$$

Normal equations: $\boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^T \boldsymbol{b}$

Condition number is $\kappa(\boldsymbol{A})^2$ where $\kappa(\boldsymbol{A}) \equiv \|\boldsymbol{A}\|_2 \, \|\boldsymbol{A}^\dagger\|_2$

Can be numerically unstable

Our remedy:

- Randomized preconditioning: $\boldsymbol{A}_p \equiv \boldsymbol{A}\boldsymbol{R}_s^{-1}$
- Preconditioned normal equations

$$\boldsymbol{A}_p^T \boldsymbol{A}_p \boldsymbol{y} = \boldsymbol{A}_p^T \boldsymbol{b}, \qquad \boldsymbol{R}_s \boldsymbol{x} = \boldsymbol{y}$$

- Half-preconditioned normal equations: $\boldsymbol{A}_p^T \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}_p^T \boldsymbol{b}$
  Special case of not-normal equations (Wathen 2025)

Existing Work on Preconditioned Normal Equations

- Preconditioners for accelerating iterative methods
- Improving solution accuracy with iterative refinement

  Wathen 2022, 2025
  Epperly, Greenbaum, Nakatsukasa 2025
  Lazzarino, Nakatsukasa, Zerbinati 2025
  Scott, Tuma 2025
  Carson, Daužickaitė 2025

This talk:

- Probabilistic condition number bounds
- Deterministic perturbation bounds

Probabilistic Condition Number Bounds for Randomized Preconditioners

Randomized Sampling

Sample c rows from smoothed matrix $\boldsymbol{A}_s \equiv \boldsymbol{S}\,\mathcal{F}\,\boldsymbol{A}$, uniformly and with replacement

- Random orthogonal matrix $\mathcal{F} = \boldsymbol{F}\boldsymbol{D}$
  $\boldsymbol{F}$ is discrete cosine transform (DCT-2)

$$\boldsymbol{F}_{ij} = \sqrt{\frac{2}{m}}\,\cos\left(\frac{\pi}{2m}(2j-1)(i-1)\right) \qquad 1 \leq i, j \leq m$$

  $\boldsymbol{D}$ is random diagonal where $\boldsymbol{D}_{jj} = \pm 1$ with probability $1/2$

- $\boldsymbol{S}$ samples c rows from identity, uniformly with replacement

$$\boldsymbol{I}_m = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}_1^T \\ \vdots \\ \boldsymbol{e}_m^T \end{bmatrix} \in \mathbb{R}^{m \times m} \qquad \boldsymbol{S} = \sqrt{\frac{m}{c}} \begin{bmatrix} \boldsymbol{e}_{k_1}^T \\ \vdots \\ \boldsymbol{e}_{k_c}^T \end{bmatrix} \in \mathbb{R}^{c \times m}$$

In expectation: $\mathbb{E}[\boldsymbol{S}^T\boldsymbol{S}] = \boldsymbol{I}_m$

Randomized Preconditioning

Given $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ with $\mathrm{rank}(\boldsymbol{A}) = n$

1. Sample c rows from smoothed matrix: $\boldsymbol{A}_s \equiv \boldsymbol{S}\mathcal{F}\boldsymbol{A}$
2. Compute preconditioner with QR: $\boldsymbol{A}_s = \boldsymbol{Q}_s \boldsymbol{R}_s$
3. Precondition matrix: $\boldsymbol{A}_p \equiv \boldsymbol{A}\, \boldsymbol{R}_s^{-1}$

Ingredients for probabilistic bounds

- Thin QR decomposition $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$ with $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}_n$
- Failure probability $0 < \delta < 1$
- Tolerance $0 < \epsilon < 1$
- Minimal sampling amount

$$c_0 \equiv 2m \max_{1 \leq i \leq m} \|\boldsymbol{e}_i^T \mathcal{F}\boldsymbol{Q}\|_2^2 \left(1 + \frac{\epsilon}{3}\right) \frac{\ln(n/\delta)}{\epsilon^2}$$

Probabilistic Condition Number Bounds

For any $0 < \epsilon < 1$ and $0 < \delta < 1$, sampling amount $c \geq c_0$ then with probability at least $1 - \delta$

- Preconditioned normal equations

$$\frac{1-\epsilon}{1+\epsilon} \leq \kappa(\boldsymbol{A}_p^T \boldsymbol{A}_p) \leq \frac{1+\epsilon}{1-\epsilon}$$

$$\sqrt{\frac{1-\epsilon}{1+\epsilon}}\,\kappa(\boldsymbol{A}) \leq \kappa(\boldsymbol{R}_s) \leq \sqrt{\frac{1+\epsilon}{1-\epsilon}}\,\kappa(\boldsymbol{A})$$

- Half-preconditioned normal equations

$$\sqrt{\frac{1-\epsilon}{1+\epsilon}}\,\kappa(\boldsymbol{A}) \leq \kappa(\boldsymbol{A}_p^T \boldsymbol{A}) \leq \sqrt{\frac{1+\epsilon}{1-\epsilon}}\,\kappa(\boldsymbol{A})$$

Summary: Randomized Preconditioner

With sampling amount $c = 3n$, condition number $\kappa(\boldsymbol{A}) \leq 10^8$

- Preconditioned matrix and Gram matrix very well conditioned

$$\kappa(\boldsymbol{A}_p) \leq 5, \qquad \kappa(\boldsymbol{A}_p^T \boldsymbol{A}_p) \leq 20$$

- Preconditioner and half-preconditioned matrix have same condition number as $\boldsymbol{A}$

$$\kappa(\boldsymbol{R}_s) \approx \kappa(\boldsymbol{A}), \qquad \kappa(\boldsymbol{A}_p^T \boldsymbol{A}) \approx \kappa(\boldsymbol{A})$$

The randomized preconditioner is effective

Deterministic Perturbation Bounds for Preconditioned Normal Equations (PNE)

Symmetric preconditioning of normal equations

$$R_s^{-T} A^T A R_s^{-1} \underbrace{R_s x}_{y} = R_s^{-T} A^T b$$

Equivalent to least squares problem[1] $\min_y \|A R_s^{-1} y - b\|_2$

Least squares residual remains the same

$$\min_y \|A R_s^{-1} y - b\|_2 = \min_x \|Ax - b\|_2$$

Preconditioned normal equations (PNE)

1. Precondition $A_p \equiv A R_s^{-1}$
2. Solve $A_p^T A_p y = A_p^T b$
3. Retrieve original solution $R_s x = y$

[1] Avron, Maymounkov, Toledo: Blendenpik (2010)

Numerical Experiments

Matrices $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ with $m = 6000$ rows

- Different number of columns $n = 10, \ldots, 400$
- Two-norm condition number $\kappa(\boldsymbol{A}) = 10^8$
- Sampling matrix $\boldsymbol{S}$ has $c = 3n$ rows

Least squares problem[2]

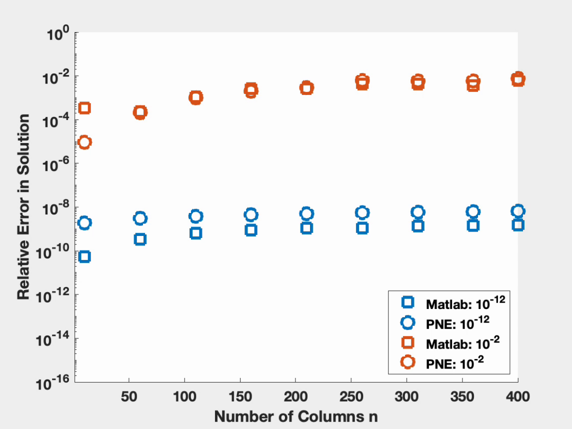
- Exact solution $\boldsymbol{x}_* = \texttt{randn}(n)$
- Least squares residual $\boldsymbol{e} \perp \mathrm{range}(\boldsymbol{A})$
- Righthand side $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}_* + \boldsymbol{e}$

Least squares solution method: QR decomposition

Graphs: Relative error in computed solution $\hat{\boldsymbol{x}}$

[2] Meier, Nakatsukasa, Townsend, Webb (2024)

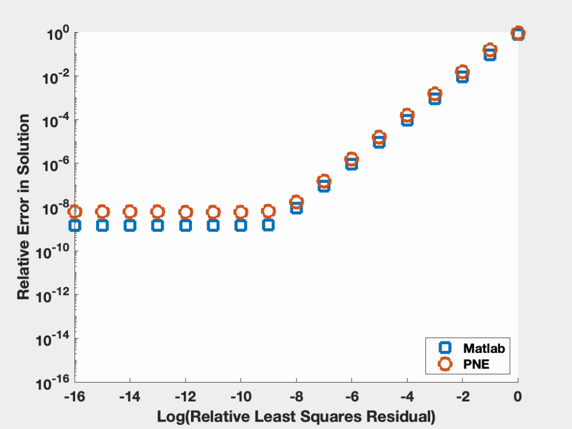
PNE Almost as Accurate as Matlab backslash



Play/Pause

Relative least squares residuals: 10^{-12} and 10^{-2}

PNE Accuracy Depends on Least Squares Residual



Play/Pause

Why does PNE accuracy depend on the least squares residual, when PNE 'does not know' about least squares problem???

Perturbation Bound: Overview

Goal: Explain why PNE depend on the least squares residual

Exact solutions:

$$\boldsymbol{A}_p^T \boldsymbol{A}_p \boldsymbol{y}_* = \boldsymbol{A}_p^T \boldsymbol{b} \qquad \boldsymbol{R}_s \boldsymbol{x}_* = \boldsymbol{y}_*$$

Computed solutions:

$$\boldsymbol{A}_p^T \boldsymbol{A}_p \hat{\boldsymbol{y}} \approx \boldsymbol{A}_p^T \boldsymbol{b} \qquad \boldsymbol{R}_s \hat{\boldsymbol{x}} = \hat{\boldsymbol{y}}$$

Two stages

1. Intermediate perturbation bound:
   Bound relative error in $\hat{\boldsymbol{y}}$ in terms of least squares residual
2. Final perturbation bound:
   Bound relative error in $\hat{\boldsymbol{x}}$ in terms of relative error in $\hat{\boldsymbol{y}}$

Intermediate Perturbation Bound

$\boldsymbol{R}_s$ is fixed nonsingular matrix

- Different perturbations for $\boldsymbol{A}$

$$\boldsymbol{A}_1 \equiv \boldsymbol{A}(\boldsymbol{R}_s + \boldsymbol{E}_s)^{-1}, \quad \boldsymbol{A}_2 \equiv \boldsymbol{A}_p + \boldsymbol{E}_p, \quad \epsilon \equiv \max\left\{\frac{\|\boldsymbol{E}_s\|}{\|\boldsymbol{R}_s\|}, \frac{\|\boldsymbol{E}_p\|}{\|\boldsymbol{A}_p\|}\right\}$$

- Perturbed system: $\quad \boldsymbol{A}_1^T \boldsymbol{A}_2 \hat{\boldsymbol{y}} = \boldsymbol{A}_1^T \boldsymbol{b}$

- Error bound

$$\frac{\|\boldsymbol{y}_* - \hat{\boldsymbol{y}}\|}{\|\hat{\boldsymbol{y}}\|} \leq \kappa(\boldsymbol{A}_p)\,\epsilon + \kappa(\boldsymbol{A}_p)^2\,\eta\left(\frac{\|\boldsymbol{b} - \boldsymbol{A}_p\hat{\boldsymbol{y}}\|}{\|\boldsymbol{A}_p\|\|\hat{\boldsymbol{y}}\|} + \epsilon\right)$$

where $\quad \eta \equiv \frac{\kappa(\boldsymbol{R}_s)\epsilon}{1-\kappa(\boldsymbol{R}_s)\,\epsilon}$

Intermediate PNE Perturbation Bound

To first order

$$\frac{\|\boldsymbol{y}_* - \hat{\boldsymbol{y}}\|}{\|\hat{\boldsymbol{y}}\|} \lesssim \max\left\{\kappa(\boldsymbol{A}_p)\,\epsilon,\ \kappa(\boldsymbol{A}_p)^2\,\kappa(\boldsymbol{R}_s)\,\epsilon\,\frac{\|\boldsymbol{b} - \boldsymbol{A}_p\hat{\boldsymbol{y}}\|}{\|\boldsymbol{A}_p\|\|\hat{\boldsymbol{y}}\|}\right\}$$

- If least squares residual is small then

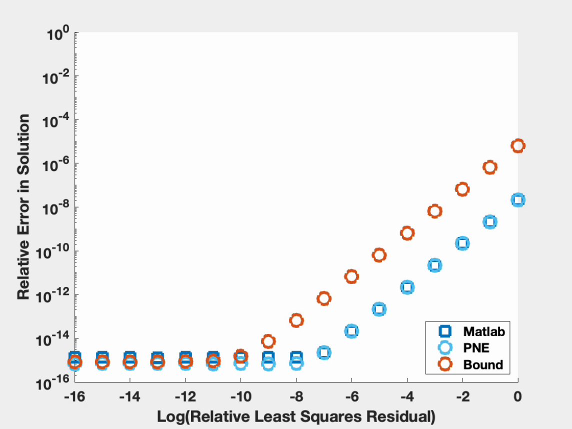
$$\frac{\|\boldsymbol{y}_* - \hat{\boldsymbol{y}}\|}{\|\hat{\boldsymbol{y}}\|} \lesssim \kappa(\boldsymbol{A}_p)\,\epsilon$$

- If least squares residual is large then

$$\frac{\|\boldsymbol{y}_* - \hat{\boldsymbol{y}}\|}{\|\hat{\boldsymbol{y}}\|} \lesssim \kappa(\boldsymbol{A}_p)^2\,\kappa(\boldsymbol{R}_s)\,\epsilon\,\frac{\|\boldsymbol{b} - \boldsymbol{A}_p\hat{\boldsymbol{y}}\|}{\|\boldsymbol{A}_p\|\|\hat{\boldsymbol{y}}\|}$$

Condition number of preconditioner amplifies LS residual

Intermediate Perturbation Bound is Informative



Play/Pause

Final Perturbation Bound

$\boldsymbol{R}_s$ is fixed nonsingular matrix

- Different perturbations for $\boldsymbol{A}$

$$\boldsymbol{A}_1 \equiv \boldsymbol{A}(\boldsymbol{R}_s + \boldsymbol{E}_s)^{-1}, \quad \boldsymbol{A}_2 \equiv \boldsymbol{A}_p + \boldsymbol{E}_p, \quad \epsilon \equiv \max\left\{\frac{\|\boldsymbol{E}_s\|}{\|\boldsymbol{R}_s\|}, \frac{\|\boldsymbol{E}_p\|}{\|\boldsymbol{A}_p\|}\right\}$$

- Perturbed PNE: $\boldsymbol{A}_1^T \boldsymbol{A}_2 \hat{\boldsymbol{y}} = \boldsymbol{A}_1^T \boldsymbol{b}, \quad \boldsymbol{R}_s \hat{\boldsymbol{x}} = \hat{\boldsymbol{y}}$

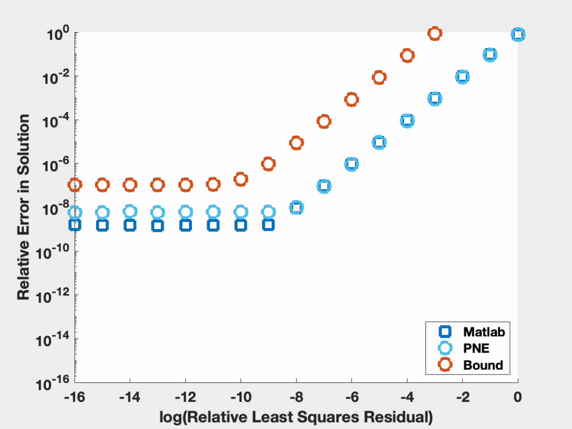
- Error bound

$$\frac{\|\boldsymbol{x}_* - \hat{\boldsymbol{x}}\|}{\|\hat{\boldsymbol{x}}\|} \leq \kappa(\boldsymbol{R}_s) \left(\kappa(\boldsymbol{A}_p)\epsilon + \kappa(\boldsymbol{A}_p)^2\, \eta \left(\frac{\|\boldsymbol{b} - \boldsymbol{A}_p\hat{\boldsymbol{y}}\|}{\|\boldsymbol{A}_p\|\|\hat{\boldsymbol{y}}\|} + \epsilon\right)\right)$$

where

$$\eta \equiv \frac{\kappa(\boldsymbol{R}_s)\epsilon}{1 - \kappa(\boldsymbol{R}_s)\,\epsilon}$$

Final Perturbation Bound is Informative



Play/Pause

In double precision: $\kappa(\boldsymbol{A})\epsilon \approx 10^{-8}$

Summary: Preconditioned Normal Equations

1. Precondition $\boldsymbol{A}_p \equiv \boldsymbol{A}\,\boldsymbol{R}_s^{-1}$
2. Solve $\boldsymbol{A}_p^T \boldsymbol{A}_p \boldsymbol{y} = \boldsymbol{A}_p^T \boldsymbol{b}$
3. Retrieve original solution $\boldsymbol{R}_s \boldsymbol{x} = \boldsymbol{y}$

Our randomization: $\kappa(\boldsymbol{A}_p) \leq 5$, $\kappa(\boldsymbol{R}_s) \approx \kappa(\boldsymbol{A})$

$$\frac{\|\boldsymbol{x}_* - \hat{\boldsymbol{x}}\|}{\|\hat{\boldsymbol{x}}\|} \lesssim \kappa(\boldsymbol{R}_s)\,\kappa(\boldsymbol{A}_p)\,\epsilon \left(1 + \kappa(\boldsymbol{A}_p)\,\kappa(\boldsymbol{R}_s)\,\frac{\|\boldsymbol{b} - \boldsymbol{A}_p\hat{\boldsymbol{y}}\|}{\|\boldsymbol{A}_p\|\|\hat{\boldsymbol{y}}\|}\right)$$

Condition number of preconditioner amplifies LS residual

LS residual dominates error if $\kappa(\boldsymbol{R}_s)\,\frac{\|\boldsymbol{b}-\boldsymbol{A}_p\hat{\boldsymbol{y}}\|}{\|\boldsymbol{A}_p\|\|\hat{\boldsymbol{y}}\|} > 1$

Deterministic Perturbation Bounds for Half-Preconditioned Normal Equations (HPNE)

Non-symmetric preconditioning of normal equations[3]

Dispense with triangular system solution

$$\boldsymbol{R}_s^{-T} \boldsymbol{A}^T \boldsymbol{A}\, \boldsymbol{x} = \boldsymbol{R}_s^{-T} \boldsymbol{A}^T \boldsymbol{b}$$

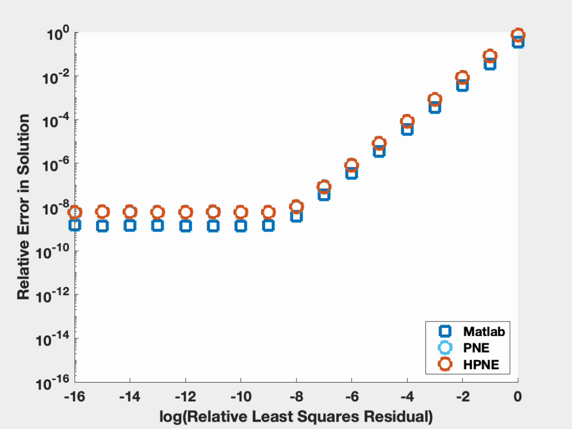
Not equivalent to a least squares problem

Half-preconditioned normal equations (HPNE)

1. Precondition $\boldsymbol{A}_p \equiv \boldsymbol{A}\, \boldsymbol{R}_s^{-1}$
2. Solve nonsymmetric system $\boldsymbol{A}_p^T \boldsymbol{A}\, \boldsymbol{x} = \boldsymbol{A}_p^T \boldsymbol{b}$

[3] Left-preconditioned CGNE

HPNE as Accurate as PNE



Play/Pause

Why does HPNE depend on the least squares residual, when there is no mathematically equivalent least squares problem?

HPNE Perturbation Bound

$\boldsymbol{R}_s$ is fixed nonsingular matrix

- Different perturbations for $\boldsymbol{A}$

$$\boldsymbol{A}_1 \equiv \boldsymbol{A}(\boldsymbol{R}_s + \boldsymbol{E}_s)^{-1}, \quad \boldsymbol{A}_2 \equiv \boldsymbol{A} + \boldsymbol{E}_A, \quad \epsilon \equiv \max\left\{\frac{\|\boldsymbol{E}_s\|}{\|\boldsymbol{R}_s\|}, \frac{\|\boldsymbol{E}_A\|}{\|\boldsymbol{A}\|}\right\}$$

- Perturbed HPNE: $\boldsymbol{A}_1^T \boldsymbol{A}_2 \hat{\boldsymbol{x}} = \boldsymbol{A}_1^T \boldsymbol{b}$

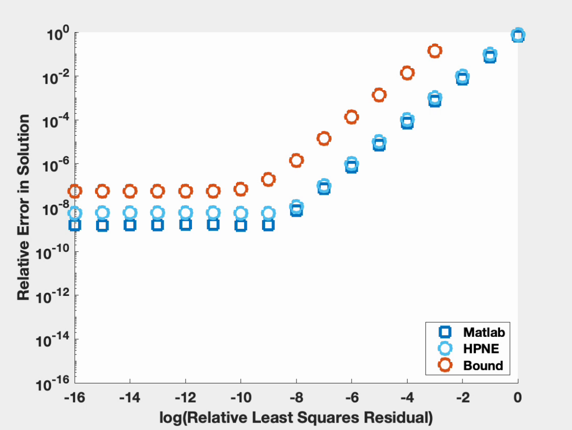
- Error bound

$$\frac{\|\boldsymbol{x}_* - \hat{\boldsymbol{x}}\|}{\|\hat{\boldsymbol{x}}\|} \leq \kappa(\boldsymbol{A}_p^T \boldsymbol{A})\,\mu\left(\eta\,\frac{\|\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}\|}{\|\boldsymbol{A}\|\|\hat{\boldsymbol{x}}\|} + (1+\eta)\epsilon\right)$$

where

$$\mu \equiv \frac{\|\boldsymbol{A}_p\|\|\boldsymbol{A}\|}{\|\boldsymbol{A}_p^T \boldsymbol{A}\|} \geq 1, \qquad \eta \equiv \frac{\kappa(\boldsymbol{R}_s)\epsilon}{1 - \kappa(\boldsymbol{R}_s)\epsilon}$$

HPNE Perturbation Bound is Informative



Play/Pause

In double precision: $\kappa(\boldsymbol{A})\epsilon \approx 10^{-8}$

Summary: Half-Preconditioned Normal Equations

1. Precondition $\boldsymbol{A}_p \equiv \boldsymbol{A}\,\boldsymbol{R}_s^{-1}$
2. Solve nonsymmetric system $\boldsymbol{A}_p^T \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}_p^T \boldsymbol{b}$

Our randomization: $\kappa(\boldsymbol{A}_p^T \boldsymbol{A}) \approx \kappa(\boldsymbol{A})$

$$\frac{\|\boldsymbol{x}_* - \hat{\boldsymbol{x}}\|}{\|\hat{\boldsymbol{x}}\|} \lesssim \kappa(\boldsymbol{A}_p^T \boldsymbol{A})\,\mu\,\epsilon \left(1 + \kappa(\boldsymbol{R}_s)\,\frac{\|\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}\|}{\|\boldsymbol{A}\|\,\|\hat{\boldsymbol{x}}\|}\right)$$

LS residual dominates error if $\kappa(\boldsymbol{R}_s)\,\frac{\|\boldsymbol{b}-\boldsymbol{A}\hat{\boldsymbol{x}}\|}{\|\boldsymbol{A}\|\|\hat{\boldsymbol{x}}\|} > 1$

although HPNE have no corresponding LS problem

Overall Summary

- Solve $\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2$ by
  Randomized preconditioning $\boldsymbol{A}_p = \boldsymbol{A}\boldsymbol{R}_s^{-1}$
- Two types of preconditioned normal equations
  PNE: $\boldsymbol{A}_p^T \boldsymbol{A}_p \boldsymbol{y} = \boldsymbol{A}_p^T \boldsymbol{b}, \quad \boldsymbol{R}_s \boldsymbol{x} = \boldsymbol{y}$
  HPNE: $\boldsymbol{A}_p^T \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}_p^T \boldsymbol{b}$
- PNE and HPNE can be as accurate as Matlab backslash
  Accuracy of solution depends on least squares residual
- Analysis
  Probabilistic condition number bounds
  Deterministic perturbation bounds: exact and informative
  Conditioning of preconditioner amplifies LS residual
- Not discussed:
  Accuracy of preconditioner in lower precision
  Speed up on GPUs