The sensitivity of the normal equations for least squares problems with randomized preconditioners

Ilse C.F. Ipsen

Joint work with James Garrison

North Carolina State University
Raleigh, NC, USA

Context

If you perform numerical computations, you are beholden to the laws of

- Conditioning:
  Sensitivity of a mathematical problem to noise in the inputs (in exact arithmetic, independent of any algorithm)

- Numerical stability:
  Accuracy and reliability of an algorithm in floating point arithmetic (round off errors)

Conditioning (Sensitivity)

How does the output of a function change, if we change its input?

- Perturbation theory quantifies this

  $$\text{Change in output} \;\leq\; \text{amplifier} * \text{change in input}$$

  Function is illconditioned if the amplifier is large

- Why is this important?

  If a function is illconditioned, no algorithm can compute an accurate output

  Potential remedy:
  Change the function (preconditioning, regularization)

Overview

1. Conditioning of nonsingular linear systems
2. Conditioning of least squares problems is more difficult
3. Preconditioned least squares problems
4. Least squares problems as normal equations
5. Preconditioned normal equations

Conditioning of nonsingular linear systems

Nonsingular linear systems $\boldsymbol{A}\boldsymbol{x}_* = \boldsymbol{b}$

What is noise?

Measurement errors in the elements of $\boldsymbol{A}$ and $\boldsymbol{b}$
Roundoff errors when storing $\boldsymbol{A}$ and $\boldsymbol{b}$ in floating point format

Question:

How does the solution $\boldsymbol{x}_* = \boldsymbol{A}^{-1}\boldsymbol{b}$ change, if we change $\boldsymbol{A}$?

Answer:

Perturbed system $(\boldsymbol{A} + \boldsymbol{E})\hat{\boldsymbol{x}} = \boldsymbol{b}$

$$\underbrace{\frac{\|\boldsymbol{x}_* - \hat{\boldsymbol{x}}\|}{\|\hat{\boldsymbol{x}}\|}}_{\text{change in output}} \leq \underbrace{\|\boldsymbol{A}\|\|\boldsymbol{A}^{-1}\|}_{\text{amplifier}} \underbrace{\frac{\|\boldsymbol{E}\|}{\|\boldsymbol{A}\|}}_{\text{change in input}} \qquad \text{(two norm)}$$

Amplifier: Condition number $\kappa(\boldsymbol{A}) \equiv \|\boldsymbol{A}\|\|\boldsymbol{A}^{-1}\| \geq 1$
The larger $\kappa(\boldsymbol{A})$, the closer $\boldsymbol{A}$ is to being singular

Conditioning of least squares problems
is more difficult

Example 1: Matrix has linearly dependent columns

$$\min_{\boldsymbol{x}} \|\boldsymbol{A}(\epsilon)\boldsymbol{x} - \boldsymbol{b}\| \qquad \text{where} \quad \boldsymbol{A}(\epsilon) = \begin{bmatrix} 1 & 0 \\ 0 & \epsilon \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

- Exact problem: $\boldsymbol{A}(0)$ has linearly dependent columns

$$\boldsymbol{A}(0) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad \boldsymbol{x}_*(0) = \boldsymbol{A}(0)^\dagger \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{minimal norm solution}$$

- Perturbed problem: $\boldsymbol{A}(\epsilon)$ nonsingular

$$\text{Unique solution } \boldsymbol{x}_*(\epsilon) = \boldsymbol{A}(\epsilon)^{-1}\boldsymbol{b} = \begin{bmatrix} 1 \\ 1/\epsilon \end{bmatrix} \to \infty \quad \text{as} \quad \epsilon \to 0$$

Solution does not depend continuously on matrix elements
Least squares problem is ill-posed

Example 2: Matrix has linearly independent columns

- Exact problem $\min_x \|Ax - b\|$ (two-norm)

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 10^{-7} \\ 0 & 0 \end{bmatrix} \quad b = \begin{bmatrix} 1 \\ 0 \\ \beta \end{bmatrix} \qquad x_* = A^\dagger b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Condition number $\kappa(A) = \|A\| \|A^\dagger\| = 10^7$
Least squares residual $\|Ax_* - b\| = \beta > 0$

- Perturbed problem $\min_x \|(A + E)x - b\|$

$$E = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & \epsilon \end{bmatrix} \quad \hat{x} = (A + E)^\dagger b = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{10^{-7}}{10^{-14}+\epsilon^2} & \frac{\epsilon}{10^{-14}+\epsilon^2} \end{bmatrix} b = \begin{bmatrix} 1 \\ \frac{\epsilon\,\beta}{10^{-14}+\epsilon^2} \end{bmatrix}$$

- Error

$$\frac{\|\hat{x} - x_*\|}{\|x_*\|} \leq \frac{\beta}{10^{-14}}\,\epsilon = \underbrace{[\kappa(A)]^2}_{10^{14}} \underbrace{\frac{\|Ax_* - b\|}{\|A\|_2 \|x_*\|}}_{\beta} \underbrace{\frac{\|E\|}{\|A\|}}_{\epsilon}$$

Large least squares residual $\Longrightarrow$ noise amplified by squared condition number

Take-away: Two types of least squares problems

$\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|$ where $\boldsymbol{A}$ is $m \times n$

- Ill-posed: $\mathrm{rank}(\boldsymbol{A}) < n$ (linearly dependent columns)
  Solution does NOT depend continuously on data $\boldsymbol{A}$ and $\boldsymbol{b}$
  Infinitely many solutions
  Minimal norm solution $\boldsymbol{x} = \boldsymbol{A}^{\dagger}\boldsymbol{b}$

- Well-posed: $\mathrm{rank}(\boldsymbol{A}) = n$ (linearly independent columns)
  Solution depends continuously on data $\boldsymbol{A}$ and $\boldsymbol{b}$
  Unique solution $\boldsymbol{x} = \boldsymbol{A}^{\dagger}\boldsymbol{b}$ where $\boldsymbol{A}^{\dagger} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$

  But: Perturbation in $\boldsymbol{A}$ amplified by condition number $\kappa(\boldsymbol{A})$ or $[\kappa(\boldsymbol{A})]^2$

- Regularization: Make problem well posed or better conditioned
  Tikhonov, ridge regression: $\min_{\boldsymbol{x}} \{\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \lambda^2\|\boldsymbol{L}\boldsymbol{x}\|_2^2\}$
  Lasso: $\min_{\boldsymbol{x}} \{\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \lambda\|\boldsymbol{x}\|_1\}$
  Column subset selection
  Preconditioning

Conditioning of well-posed least squares problems

- $\boldsymbol{A}$ has full column rank
- Exact problem: $\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|$ has solution $\boldsymbol{x}_* = \boldsymbol{A}^\dagger \boldsymbol{b}$

- Perturbed matrix $\boldsymbol{A} + \boldsymbol{E}$ has full column rank
- Perturbed problem: $\min_{\boldsymbol{x}} \|(\boldsymbol{A} + \boldsymbol{E})\boldsymbol{x} - \boldsymbol{b}\|$ has solution $\hat{\boldsymbol{x}} = (\boldsymbol{A} + \boldsymbol{E})^\dagger \boldsymbol{b}$
- Relative error in solution

$$\underbrace{\frac{\|\hat{\boldsymbol{x}} - \boldsymbol{x}_*\|}{\|\hat{\boldsymbol{x}}\|}}_{\text{change in output}} \lesssim \underbrace{\kappa(\boldsymbol{A})\,(1 + \kappa(\boldsymbol{A})\,\rho)}_{\text{amplifier}} \underbrace{\frac{\|\boldsymbol{E}\|}{\|\boldsymbol{A}\|}}_{\text{change in input}} \qquad \rho \equiv \underbrace{\frac{\|(\boldsymbol{A} + \boldsymbol{E})\hat{\boldsymbol{x}} - \boldsymbol{b}\|}{\|\boldsymbol{A}\|\|\hat{\boldsymbol{x}}\|}}_{\text{normalized LS residual}}$$

Solutions can be illconditioned (sensitive) for two reasons:

Large condition number $\kappa(\boldsymbol{A})$

Large least squares residual ρ

If least squares residual large ($\rho \approx 1$), then error in solution amplified by squared condition number $[\kappa(A)]^2$

Preconditioned Least Squares Problems

Right preconditioning

Preconditioning $\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|$ with $\boldsymbol{R}_s$ gives

$$\min_{\boldsymbol{y}} \| \underbrace{\boldsymbol{A}\boldsymbol{R}_s^{-1}}_{\boldsymbol{A}_p} \underbrace{\boldsymbol{R}_s\boldsymbol{x}}_{\boldsymbol{y}} - \boldsymbol{b}\|$$

1. Precondition $\boldsymbol{A}_p \equiv \boldsymbol{A}\boldsymbol{R}_s^{-1}$
2. Solve preconditioned least squares problem

$$\min_{\boldsymbol{y}} \|\boldsymbol{A}_p\boldsymbol{y} - \boldsymbol{b}\| \quad \text{has solution} \quad \boldsymbol{y}_* = \boldsymbol{A}_p^{\dagger}\boldsymbol{b}$$

3. Recover original solution

$$\text{Solve} \quad \boldsymbol{R}_s\boldsymbol{x}_* = \boldsymbol{y}_*$$

Blendenpik [Avron, Maymounkov, Toledo 2011]:
Randomized preconditioner

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

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

Probabilistic condition number bounds for a good preconditioner

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

- Preconditioned matrix almost perfectly conditioned

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

- Preconditioner has almost same condition number as $\boldsymbol{A}$

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

Conditioning (deterministic) of preconditioned least squares problem

- $\boldsymbol{A}$ and $\boldsymbol{A}_p \equiv \boldsymbol{A}\boldsymbol{R}_s^{-1}$ have full column rank
- Exact problem: $\min_{\boldsymbol{y}} \|\boldsymbol{A}_p\boldsymbol{y} - \boldsymbol{b}\|$ has solution $\boldsymbol{y}_* = \boldsymbol{A}_p^\dagger \boldsymbol{b}$, solve $\boldsymbol{R}_s\boldsymbol{x}_* = \boldsymbol{y}_*$

- Perturbed preconditioned matrix: $\boldsymbol{A}_p + \boldsymbol{E}$
- Perturbed matrix has full rank: $\kappa(\boldsymbol{A}_p)\frac{\|\boldsymbol{E}\|}{\|\boldsymbol{A}_p\|} < 1$
- Perturbed preconditioned least squares problem:

$$\min_{\boldsymbol{y}} \|(\boldsymbol{A}_p + \boldsymbol{E})\boldsymbol{y} - \boldsymbol{b}\| \quad \text{has solution} \quad \hat{\boldsymbol{y}} = (\boldsymbol{A}_p + \boldsymbol{E})^\dagger \boldsymbol{b}$$

- Solve exactly $\boldsymbol{R}_s\hat{\boldsymbol{x}} = \hat{\boldsymbol{y}}$

- Relative error in solution

$$\underbrace{\frac{\|\boldsymbol{x}_* - \hat{\boldsymbol{x}}\|}{\|\hat{\boldsymbol{x}}\|}}_{\text{change in output}} \leq \underbrace{\kappa(\boldsymbol{R}_s)\,\kappa(\boldsymbol{A}_p)\,(1 + \kappa(\boldsymbol{A}_p)\,\rho)}_{\text{amplifier}} \underbrace{\frac{\|\boldsymbol{E}\|}{\|\boldsymbol{A}_p\|}}_{\text{change in input}} \qquad \underbrace{\rho \equiv \frac{\|\boldsymbol{b} - (\boldsymbol{A}_p + \boldsymbol{E})\hat{\boldsymbol{y}}\|}{\|\boldsymbol{A}_p\|\|\hat{\boldsymbol{y}}\|}}_{\text{normalized LS residual}}$$

Summary: Preconditioned least squares problems

1. Precondition: $\boldsymbol{A}_p \equiv \boldsymbol{A}\,\boldsymbol{R}_s^{-1}$
2. Solve preconditioned least squares problem: $\min_{\boldsymbol{y}} \|\boldsymbol{A}_p\boldsymbol{y} - \boldsymbol{b}\|$
3. Retrieve original solution: $\boldsymbol{R}_s\boldsymbol{x}_* = \boldsymbol{y}_*$

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

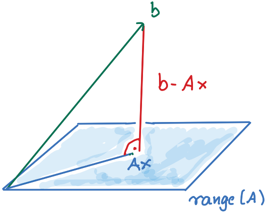
$$\frac{\|\boldsymbol{x}_* - \hat{\boldsymbol{x}}\|}{\|\hat{\boldsymbol{x}}\|} \lesssim \kappa(\boldsymbol{R}_s)\,\kappa(\boldsymbol{A}_p)\,\left(1 + \kappa(\boldsymbol{A}_p)\,\frac{\|\boldsymbol{b} - (\boldsymbol{A}_p + \boldsymbol{E})\hat{\boldsymbol{y}}\|}{\|\boldsymbol{A}_p\|\|\hat{\boldsymbol{y}}\|}\right)\frac{\|\boldsymbol{E}\|}{\|\boldsymbol{A}_p\|}$$

Preconditioned problem as well conditioned as original problem

Least Squares Problems as Normal Equations

Least squares problems as normal equations

$\min_{\boldsymbol{x}} \|\boldsymbol{A}\,\boldsymbol{x} - \boldsymbol{b}\|$ has solution $\boldsymbol{x}_* \;=\; \boldsymbol{A}^\dagger \boldsymbol{b}$
with least squares residual $\boldsymbol{A}\boldsymbol{x}_* - \boldsymbol{b}$



Least squares residual is orthogonal to $\mathrm{range}(\boldsymbol{A})$

$$\boldsymbol{A}^T \left(\boldsymbol{A}\boldsymbol{x}_* - \boldsymbol{b}\right) = \boldsymbol{0}$$

Least squares solution $\boldsymbol{x}_*$ solves the normal equations

$$\boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x}_* = \boldsymbol{A}^T \boldsymbol{b}$$

Conditioning of normal equations

- $\boldsymbol{A}$ has full column rank
- Exact system: $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}_* = \boldsymbol{A}^T\boldsymbol{b}$ has solution

$$\boldsymbol{x}_* = \underbrace{(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T}_{\boldsymbol{A}^\dagger}\boldsymbol{b}$$

- Perturbed system $(\boldsymbol{A}^T\boldsymbol{A} + \boldsymbol{E})\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$
- Relative error in solution

$$\underbrace{\frac{\|\boldsymbol{x}_* - \hat{\boldsymbol{x}}\|}{\|\hat{\boldsymbol{x}}\|}}_{\text{change in output}} \leq \underbrace{\kappa(\boldsymbol{A}^T\boldsymbol{A})}_{\text{amplifier}} \underbrace{\frac{\|\boldsymbol{E}\|}{\|\boldsymbol{A}^T\boldsymbol{A}\|}}_{\text{change in input}}$$

Condition number $\kappa(\boldsymbol{A}^T\boldsymbol{A}) = [\kappa(\boldsymbol{A})]^2$ is much worse

If $\kappa(\boldsymbol{A}) = 10^8$ then $\kappa(\boldsymbol{A}^T\boldsymbol{A}) = 10^{16}$, $\boldsymbol{A}^T\boldsymbol{A}$ numerically singular

Preconditioned Normal Equations

Two-sided preconditioning

Preconditioning $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}_* = \boldsymbol{A}^T\boldsymbol{b}$ with $\boldsymbol{R}_s$ gives

$$\underbrace{\boldsymbol{R}_s^{-T}\boldsymbol{A}^T}_{\boldsymbol{A}_p^T}\underbrace{\boldsymbol{A}\boldsymbol{R}_s^{-1}}_{\boldsymbol{A}_p}\underbrace{\boldsymbol{R}_s\boldsymbol{x}_*}_{\boldsymbol{y}_*} = \underbrace{\boldsymbol{R}_s^{-T}\boldsymbol{A}^T}_{\boldsymbol{A}_p^T}\boldsymbol{b}$$

1. Precondition: $\boldsymbol{A}_p \equiv \boldsymbol{A}\boldsymbol{R}_s^{-1}$
2. Solve preconditioned normal equations: $\boldsymbol{A}_p^T\boldsymbol{A}_p\boldsymbol{y}_* = \boldsymbol{A}_p^T\boldsymbol{b}$
3. Recover original solution: $\boldsymbol{R}_s\boldsymbol{x}_* = \boldsymbol{y}_*$

Existing Work

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

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[1]

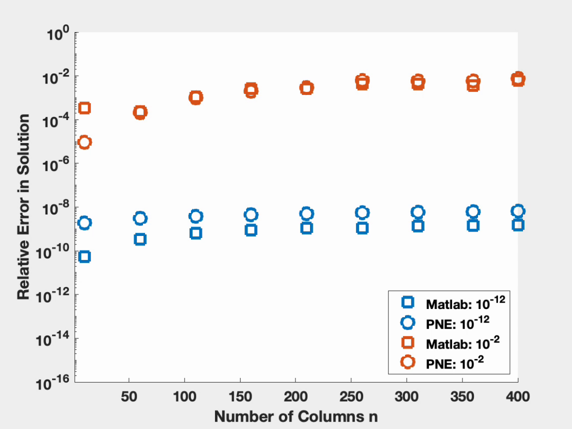
- Exact solution $\boldsymbol{x}_* = \texttt{randn}(n)$
- Least squares residual $\boldsymbol{e} \perp \mathrm{range}(\boldsymbol{A})$
- Righthand side $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}_* + \boldsymbol{e}$

Least squares solution method: QR decomposition

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

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

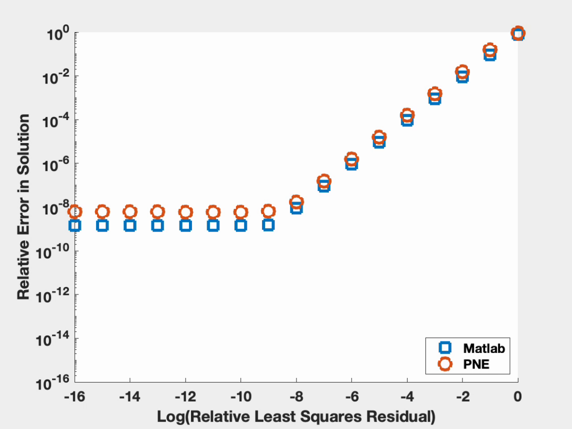
Preconditioned normal equations almost as accurate as Matlab backslash



Play/Pause

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

Accuracy depends on least squares residual



Play/Pause

Conditioning of preconditioned normal equations

- $\boldsymbol{A}$ and $\boldsymbol{A}_p \equiv \boldsymbol{A}\boldsymbol{R}_s^{-1}$ have full column rank
- Exact preconditioned normal equations: $\boldsymbol{A}_p^T \boldsymbol{A}_p \boldsymbol{y}_* = \boldsymbol{A}_p^T \boldsymbol{b}, \quad \boldsymbol{R}_s \boldsymbol{x}_* = \boldsymbol{y}_*$

- Perturbed preconditioned matrix:

$$\boldsymbol{A}_1 \equiv (\boldsymbol{A} + \boldsymbol{E})(\boldsymbol{R}_s + \boldsymbol{E}_s)^{-1}, \qquad \epsilon_s \equiv \frac{\|\boldsymbol{E}_s\|}{\|\boldsymbol{R}_s\|}$$

- Perturbed preconditioner is nonsingular: $\kappa(\boldsymbol{R}_s)\epsilon_s < 1$
- Perturbed preconditioned normal equations

$$\boldsymbol{A}_1^T \boldsymbol{A}_1 \hat{\boldsymbol{y}} = \boldsymbol{A}_1^T \boldsymbol{b}, \qquad (\boldsymbol{R}_s + \boldsymbol{E}_s)\hat{\boldsymbol{x}} = \hat{\boldsymbol{y}}$$

- Relative error in solution

$$\underbrace{\frac{\|\boldsymbol{x}_* - \hat{\boldsymbol{x}}\|}{\|\hat{\boldsymbol{x}}\|}}_{\text{change in output}} \lesssim \underbrace{\kappa(\boldsymbol{R}_s)\,\kappa(\boldsymbol{A}_p)\,(1 + \kappa(\boldsymbol{R}_s)\kappa(\boldsymbol{A}_p)\,\rho)}_{\text{amplifier}} \underbrace{\frac{\|\boldsymbol{E}\|}{\|\boldsymbol{A}\|}}_{\text{change in input}} \qquad \underbrace{\rho \equiv \frac{\|\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}\|}{\|\boldsymbol{A}\|\|\hat{\boldsymbol{x}}\|}}_{\text{normalized LS residual}}$$

Summary: Preconditioned normal equations

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

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

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

With a good preconditioner:

As well conditioned as least squares problem

Condition number amplifies least squares residual

Bound does not depend on perturbation in preconditioner

Preconditioner can be computed in lower precision

Half-preconditioned normal equations[2]

Dispense with triangular system solution

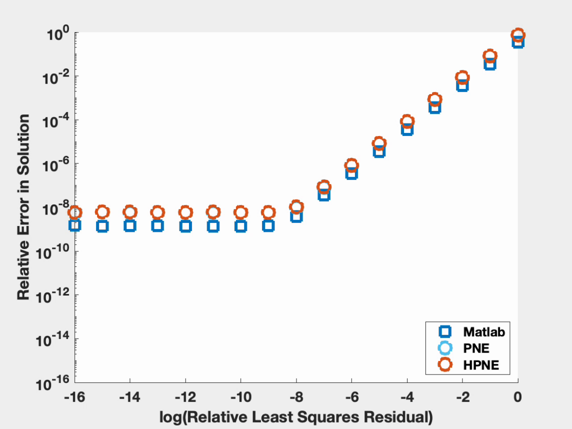
$$\boldsymbol{R}_s^{-T} \boldsymbol{A}^T \boldsymbol{A} \boldsymbol{x}_* = \boldsymbol{R}_s^{-T} \boldsymbol{A}^T \boldsymbol{b}$$

Not equivalent to a least squares problem

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

[2] Left-preconditioned CGNE

Half-preconditioned normal equations as accurate as preconditioned normal equations



Play/Pause

Why dependence on least squares residual, when there is no mathematically equivalent least squares problem?

Conditioning of half-preconditioned normal equations

- $\boldsymbol{A}$ and $\boldsymbol{A}_p \equiv \boldsymbol{A}\boldsymbol{R}_s^{-1}$ have full column rank
- Exact half-preconditioned normal equations: $\boldsymbol{A}_p^T \boldsymbol{A} \boldsymbol{x}_* = \boldsymbol{A}_p^T \boldsymbol{b}$

- Perturbed preconditioned matrix:

$$\boldsymbol{A}_1 \equiv \boldsymbol{A} + \boldsymbol{E}, \qquad \boldsymbol{A}_2 \equiv \boldsymbol{A}_1(\boldsymbol{R}_s + \boldsymbol{E}_s)^{-1}, \qquad \epsilon_s \equiv \frac{\|\boldsymbol{E}_s\|}{\|\boldsymbol{R}_s\|}$$

- Perturbed preconditioner is nonsingular: $\kappa(\boldsymbol{R}_s)\epsilon_s < 1$
- Perturbed preconditioned normal equations: $\boldsymbol{A}_2^T \boldsymbol{A}_1 \hat{\boldsymbol{x}} = \boldsymbol{A}_2^T \boldsymbol{b}$
- Relative error in solution

$$\underbrace{\frac{\|\boldsymbol{x}_* - \hat{\boldsymbol{x}}\|}{\|\hat{\boldsymbol{x}}\|}}_{\text{change in output}} \lesssim \underbrace{\kappa(\boldsymbol{A}_p^T \boldsymbol{A})\, \nu\, (1 + \kappa(\boldsymbol{R}_s)\, \rho)}_{\text{amplifier}} \underbrace{\frac{\|\boldsymbol{E}\|}{\|\boldsymbol{A}\|}}_{\text{change in input}}$$

where

$$\nu \equiv \frac{\|\boldsymbol{A}_p\|\|\boldsymbol{A}\|}{\|\boldsymbol{A}_p^T \boldsymbol{A}\|} \geq 1 \qquad \underbrace{\rho \equiv \frac{\|\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}\|}{\|\boldsymbol{A}\|\|\hat{\boldsymbol{x}}\|}}_{\text{normalized LS residual}}$$

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

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

With a good preconditioner:

As well conditioned as least squares problem

– although there is no corresponding least squares problem

Bound does not depend on perturbation in preconditioner

Preconditioner can be computed in lower precision

Overall Summary

- Solve $\min_x \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2$ by randomized preconditioning $\boldsymbol{A}_p \equiv \boldsymbol{A}\boldsymbol{R}_s^{-1}$
- Two types of preconditioned normal equations

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

  half-preconditioned: $\boldsymbol{A}_p^T \boldsymbol{A}\boldsymbol{x}_* = \boldsymbol{A}_p^T \boldsymbol{b}$
- With a good preconditioner:

  As well conditioned as original least squares problem

  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

  Bounds do not depend on perturbation of preconditioner

  Preconditioner can be computed in lower precision
- Not discussed:

  Computing the preconditioner in lower precision

  Speed up on GPUs

  Not-normal equations $\boldsymbol{B}^T \boldsymbol{A}\boldsymbol{x}_* = \boldsymbol{B}^T \boldsymbol{b}$ for $\mathrm{range}(\boldsymbol{B}) = \mathrm{range}(\boldsymbol{A})$