Randomized Linear Algebra
Lecture 9: High-dimensional Problems, Least-norm Solutions and Randomized Methods
Faster Least Squares Optimization: Random Projection

- **Left-sketching**

  Form $SA$ and $Sb$ where $S \in \mathbb{R}^{m \times n}$ is a random projection matrix

- Solve the smaller problem

  $$\min_{x \in \mathbb{R}^d} \left\| SAx - Sb \right\|_2^2$$

- using any classical method.

  Direct method complexity $md^2$
Let $S$ be $\frac{1}{m} \times$ i.i.d. Gaussian. $\mathbb{E}[S^T S] = I$

$\tilde{x} = \arg \min_{x \in \mathbb{R}^d} \|S A x - S b\|_2^2$

Unbiased $\mathbb{E}[\tilde{x}] = x_{LS}$

since $\tilde{x} = x_{LS} + (A^T S^T S A)^{-1} A^T S^T S b_{\perp}$

zero mean
Gaussian Sketch

- Let $S$ be $\frac{1}{m} \times$ i.i.d. Gaussian. $\mathbb{E}[S^T S] = I$

  $$\tilde{x} = \arg \min_{x \in \mathbb{R}^d} \| S Ax - Sb \|_2^2$$

- Unbiased $\mathbb{E}[\tilde{x}] = x_{LS}$
  
  Since $\tilde{x} = x_{LS} + \left( A^T S^T S A \right)^{-1} A^T S^T S b_{\perp}$

  zero mean

- Variance

  $$\mathbb{E} \| A(\tilde{x} - x_{LS}) \|_2^2 = f(x_{LS}) \frac{d}{m-d-1}$$

  valid for $m > d + 1$  where $f(x) = \| Ax - b \|_2^2$
Gaussian Sketch

- Let $S$ be $\frac{1}{m} \times$ i.i.d. Gaussian. $\mathbb{E}[S^T S] = I$
  
  $\tilde{x} = \arg\min_{x \in \mathbb{R}^d} \|SAx - Sb\|^2_2$

- Unbiased $\mathbb{E}[\tilde{x}] = x_{LS}$ since $\tilde{x} = x_{LS} + (A^T S^T S A)^{-1} A^T S^T S b_\perp$

- Variance $\mathbb{E}\|A(\tilde{x} - x_{LS})\|^2_2 = f(x_{LS}) \frac{d}{m-d-1}$ valid for $m > d + 1$ where $f(x) = \|Ax - b\|^2_2$

- Function value $f(\tilde{x}) = \|A\tilde{x} - b\|^2_2 = \|A(\tilde{x} - x_{LS})\|^2_2 + \|Ax_{LS} - b\|^2_2$

- $\mathbb{E}f(\tilde{x}) - f(x_{LS}) = f(x_{LS}) \frac{d}{m-d-1}$
Variance Reduction by Averaging

Let $S_1, \ldots, S_r$ be $\frac{1}{m} \times$ i.i.d. Gaussian. $\mathbb{E}[S^T S] = I$

$$\tilde{x}_i = \arg \min_{x \in \mathbb{R}^d} \| S_i Ax - S_i b \|^2$$

let $\tilde{x} = \frac{1}{r} \sum_{i=1}^{r} x_i$

Unbiased $\mathbb{E}[\tilde{x}] = x_{LS}$

Variance is reduced by $\frac{1}{r}$

$$\mathbb{E} \| A(\tilde{x} - x_{LS}) \|^2 = f(x_{LS}) \frac{1}{r} \frac{d}{m-d-1}$$
Variance Reduction by Averaging

- Let $S_1, \ldots, S_r$ be $\frac{1}{m} \times$ i.i.d. Gaussian. $\mathbb{E}[S^T S] = I$

\[ \tilde{x}_i = \arg \min_{x \in \mathbb{R}^d} \|S_i A x - S_i b\|_2^2 \]

- let $\tilde{x} = \frac{1}{r} \sum_{i=1}^r x_i$

- Unbiased $\mathbb{E}[\tilde{x}] = x_{LS}$

- Variance is reduced by $\frac{1}{r}$

- $\mathbb{E}\|A(\tilde{x} - x_{LS})\|_2^2 = f(x_{LS}) \frac{1}{r} \frac{d}{m-d-1}$

- $\mathbb{E}f(\tilde{x}) - f(x_{LS}) = f(x_{LS}) \frac{1}{r} \frac{d}{m-d-1}$
High-dimensional Least Squares Problems

- $A \in \mathbb{R}^{n \times d}$ where $d > n$
- no unique solution
High-dimensional Least Squares Problems

- $A \in \mathbb{R}^{n \times d}$ where $d > n$
- no unique solution
- minimum ($\ell_2$) norm solution is unique

$$x_{\text{min-norm}} = \arg\min_{Ax=b} \|x\|_2^2$$
Minimum norm solution and SVD

\[ x_{\text{min-norm}} = \arg \min_{Ax=b} \|x\|_2^2 \]
Random projection to reduce dimension: Right Sketch

\[ x_{\text{min-norm}} = \arg \min_{Ax=b} \|x\|_2^2 \]

- We can right multiply \( A \) and form \( AS \) where \( S \in \mathbb{R}^{d\times m} \) and solve

\[ \arg \min_{ASz=b} \|z\|_2^2 \]
Random projection to reduce dimension: Right Sketch

\[ x_{\text{min-norm}} = \arg \min_{Ax=b} \|x\|^2_2 \]

- We can right multiply \( A \) and form \( AS \) where \( S \in \mathbb{R}^{d \times m} \) and solve

\[ \arg \min_{ASz=b} \|z\|^2_2 \]

- How do we use \( z \in \mathbb{R}^m \)?
Right Sketch

\[ x_{\text{min-norm}} = \arg \min_{Ax=b} \|x\|_2^2 \]

approximation \[ \tilde{x} = S\tilde{z} \]

where \( \tilde{z} := \arg \min_{ASz=b} \|z\|_2^2 \)
Right Sketch

\[ x_{\text{min-norm}} = \arg \min_{Ax = b} \underbrace{\|x\|_2^2}_{f(x)} \]

approximation \( \tilde{x} = S\tilde{z} \)

where \( \tilde{z} := \arg \min_{ASz = b} \|z\|_2^2 \)

- Let \( S \) be i.i.d. Gaussian \( \mathcal{N}(0, \frac{1}{\sqrt{m}}) \)
- Is \( \tilde{x} \) unbiased, i.e., \( \mathbb{E}\tilde{x} = ? x_{\text{min-norm}} \)
Right Sketch

\[ x_{\text{min-norm}} = \arg \min_{Ax=b} f(x) \]

approximation \[ \tilde{x} = S\tilde{z} \]

where \[ \tilde{z} := \arg \min_{ASz=b} \|z\|_2^2 \]

- Let \( S \) be i.i.d. Gaussian \( \mathcal{N}(0, \frac{1}{\sqrt{m}}) \)
- Is \( \tilde{x} \) unbiased, i.e., \( \mathbb{E}\tilde{x} = ? x_{\text{min-norm}} \)
- Yes, conditioned on \( SA \)

\[ \tilde{x} \sim \mathcal{N}(x_{\text{min-norm}}, VV^T b^T (AS^T SA^T)^{-1} b) \]

- \( VV^T \) is the projection onto the null space of \( A \)
- error \( \tilde{x} - x_{\text{min-norm}} \in \text{Null}(A) \)
Right Sketch

\[ x_{\text{min-norm}} = \arg \min_{Ax=b} \|x\|_2^2 \]

approximation \( \tilde{x} = S\tilde{z} \)

where \( \tilde{z} := \arg \min_{ASz=b} \|z\|_2^2 \)

▶ Let \( S \) be i.i.d. Gaussian \( N(0, \frac{1}{\sqrt{m}}) \)
▶ Is \( \tilde{x} \) unbiased, i.e., \( \mathbb{E}\tilde{x} = ? x_{\text{min-norm}} \)
▶ Yes, conditioned on \( SA \)

\[ \tilde{x} \sim N(x_{\text{min-norm}}, VV^T b^T (AS^T SA^T)^{-1} b) \]

▶ \( VV^T \) is the projection onto the null space of \( A \)
▶ error \( \tilde{x} - x_{\text{min-norm}} \in \text{Null}(A) \)
▶ Using \( \mathbb{E}(AS^T SA^T)^{-1} = AA^T \frac{m}{m-n-1} \)

\[ \mathbb{E}\|\tilde{x} - x_{\text{min-norm}}\|_2^2 = \frac{d-n}{m-n-1} f(x_{\text{min-norm}}) = \frac{d-n}{m-n-1} \|x_{\text{min-norm}}\|_2^2 \]
Left Sketch vs Right Sketch Summary

- Both are unbiased using Gaussian projections
- $A$ is $n \times d$
- Left sketch $n \geq d$

\[\tilde{x} = \arg \min_{x \in \mathbb{R}^d} \|SAx - Sb\|_2^2\]

Variance: $\mathbb{E}\|A(\tilde{x} - x_{LS})\|_2^2 = f(x_{LS})\frac{d}{m-d-1}$

- Right sketch $d > n$

\[\tilde{x} = S\tilde{z} \quad \text{where} \quad \tilde{z} := \arg \min_{ASz=b} \|z\|_2^2\]

Variance: $\mathbb{E}\|\tilde{x} - x_{\text{min-norm}}\|_2^2 = f(x_{\text{min-norm}})\frac{d-n}{m-n-1}$
Back to Left Sketch: Which sketching matrices are good?

▶ We need to find conditions to guarantee approximate optimality
▶ Let $A = U\Sigma V^T$ SVD in compact form

some deterministic options
▶ $S = U^T$ is $d \times n$
▶ $S = A^T$

▶ For random $S$ matrices $A^T S^T S A$ needs to be invertible we want it to be close to $A^T A$
Approximate Matrix Multiplication

- Let the approximate product of $AB$ be $C = AS^T SB$

$$\Pr[\|AB - C\|_F > \epsilon \|A\|_F \|B\|_F] \leq \delta$$

- Follows from JL Moment property

- $S \in \mathbb{R}^{m \times n} \sim \frac{1}{\sqrt{m}} \times$ random i.i.d. sub-Gaussian, e.g., $\pm 1$, or $N(0, 1)$ with $m = \frac{c_1}{\epsilon^2} \log \frac{1}{\delta}$

- $S \in \mathbb{R}^{m \times n} \sim \frac{1}{\sqrt{m}} \times$ CountSketch matrix (one nonzero per column, which is $\pm 1$ at a uniformly random location) with $m = \frac{c_2}{\epsilon^2 \delta}$

- $S \in \mathbb{R}^{m \times n} \sim \frac{1}{\sqrt{m}} \times$ Fast JL Transform with $m = \frac{c_3}{\epsilon} \log \frac{1}{\delta}$
Approximate Matrix Multiplication

Let the approximate product of $AB$ be $C = AS^T SB$

$$\mathbb{P}[\|AB - C\|_F > \epsilon\|A\|_F\|B\|_F] \leq \delta$$

Follows from JL Moment property

$S \in \mathbb{R}^{m \times n} \sim \frac{1}{\sqrt{m}} \times \text{random i.i.d. sub-Gaussian, e.g., } \pm 1, \text{ or } N(0, 1)$ with $m = \frac{c_1}{\epsilon^2} \log \frac{1}{\delta}$

$S \in \mathbb{R}^{m \times n} \sim \frac{1}{\sqrt{m}} \times \text{CountSketch matrix (one nonzero per column, which is } \pm 1 \text{ at a uniformly random location)}$ with $m = \frac{c_2}{\epsilon^2 \delta}$

$S \in \mathbb{R}^{m \times n} \sim \frac{1}{\sqrt{m}} \times \text{Fast JL Transform with } m = \frac{c_3}{\epsilon} \log \frac{1}{\delta}$

Sparse JL and Fast JL are more efficient

advantages: doesn’t require any knowledge about matrices $A$ and $B$ (oblivious)

optimal sampling probabilities depend on the column/row norms of $A$ and $B$
Basic Inequality Method

- We minimize $\tilde{x} = \arg \min \|S(Ax - b)\|_2^2$
- $x_{LS}$ minimizes $\|Ax - b\|_2^2$
- How far is $\tilde{x}$ from $x_{LS}$?
- **Step 1.** Establish two optimality (in)equalities for these variables
  - $\|Ax_{LS} - b\|_2^2 \leq \|Ax' - b\|_2^2$ for any $x'$, i.e., $A^T(Ax_{LS} - b) = 0$
  - $\|S(A\tilde{x} - b)\|_2^2 \leq \|S(Ax_{LS} - b)\|_2^2$
Basic Inequality Method

- We minimize $\tilde{x} = \text{arg min} \|S(Ax - b)\|^2_2$
- $x_{LS}$ minimizes $\|Ax - b\|^2_2$
- How far is $\tilde{x}$ from $x_{LS}$?
- **Step 1.** Establish two optimality (in)equalities for these variables
  - $\|Ax_{LS} - b\|^2_2 \leq \|Ax' - b\|^2_2$ for any $x'$, i.e., $A^T(Ax_{LS} - b) = 0$
  - $\|S(A\tilde{x} - b)\|^2_2 \leq \|S(Ax_{LS} - b)\|^2_2$
- **Step 2.** Define error $\Delta = \tilde{x} - x_{LS}$ and re-write these inequalities in terms of $\delta$
  - $\|SA\Delta\|^2_2 \leq 2b^\perp^T(S^T S - I)A\Delta$
- **Step 3.** Argue $S^T S \approx I$
Leverage Scores
Questions?