EE270
Large scale matrix computation, optimization and learning

Instructor : Mert Pilanci

Stanford University

Tuesday, Feb 9 2021

Randomized Linear Algebra and Optimization Lecture 10: Leverage Scores and Basic Inequality Method

Projected Least Squares Problems

- **Left-sketching**

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

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

- using any classical method.

  Direct method complexity md^2

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 $\tilde{x} = \arg\min \|S(Ax - b)\|_2^2$
- 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 δ
- $\|SA\Delta\|_2^2 \leq 2{b^{\perp}}^T(S^TS - I)A\Delta$
- **Step 3**. Argue $S^TS \approx I$

Basic Inequality Method

$A = U\Sigma V^T$

$A\Delta = U\Sigma V^T\Delta$

$z = \Sigma V^T \Delta$ or

$\Delta = V\Sigma^{-1} z$

- $\|SA\Delta\|_2^2 \leq 2b^{\perp T}(S^TS - I)A\Delta$

$\approx \|A\Delta\|_2^2$

$$\max_{\Delta} \left| \frac{\|SA\Delta\|_2^2}{\underbrace{\|A\Delta\|_2^2}_{\underbrace{\Delta V \Sigma \underbrace{U^T U}_{I} \Sigma V^T \Delta}_{z^T z}}} - 1 \right| = \max_z \left| \frac{\|SUz\|_2^2}{\|z\|_2^2} - 1 \right|$$

$$= \max_z \left| \frac{z^T}{\|z\|_2}(U^TS^TSU - I)\frac{z}{\|z\|_2} \right|$$

$$= \sigma^2_{\max}(U^TS^TSU - I)$$

$\max_z \frac{z^T}{\|z\|_2} Q \frac{z}{\|z\|_2} = \sigma_{max}(Q)^2$

$\mathbb{E}\ \Delta^T A^T S^T S A\Delta = \Delta^T A^T\, \mathbb{E} S^T S \cdot A\Delta = \|A\Delta\|_2^2$

FALSE!

this is incorrect

since Δ and S^TS are <u>not</u> independent

Basic Inequality Method

- $\|SA\Delta\|_2^2 \leq 2{b^{\perp}}^T(S^TS - I)A\Delta$

$$\max_{\Delta}\left|\frac{\|SA\Delta\|_2^2}{\|A\Delta\|_2^2} - 1\right| = \max_{z}\left|\frac{\|SUz\|_2^2}{\|z\|_2^2} - 1\right|$$

$$= \max_{z}\left|\frac{z^T}{\|z\|_2}(U^TS^TSU - I)\frac{z}{\|z\|_2}\right|$$

$Q = U\Sigma V^T \Rightarrow$

$$= \sigma^2_{\max}(U^TS^TSU - I) \leq \epsilon \cdot d < \epsilon'$$

$\sigma_{\max}(Q) \leq \|Q\|_F = \sqrt{\operatorname{tr} Q^TQ} = \sqrt{\operatorname{tr} V\Sigma U^TU\Sigma V^T} = \sqrt{\sum \sigma_i^2}$

- Approximate matrix multiplication (AMM):

$$\sigma_{\max}(U^TS^TSU - I) \leq \|U^TS^TSU - \underbrace{U^TU}_{I}\|_F \leq \epsilon\|\underbrace{U^TU}_{I_{d\times d}}\|_F^2 = \epsilon \cdot d$$

- This is called a Subspace Embedding

  we can rescale ϵ to get $\sigma_{\max}(U^TS^TSU - I) \leq \epsilon$ for appropriate m

AMM: $\|A^TS^TSA - A^TA\|_F \leq \epsilon \cdot \|A\|_F \cdot \|A\|_F$

Basic Inequality Method

if $m \geq \frac{1}{\epsilon} \cdot \delta^2$

- $\|SA\Delta\|_2^2 \leq 2{b^\perp}^T(S^TS - I)A\Delta$
- Now consider the right handside

$$= 2{b^\perp}^T(S^TS - I)\underbrace{UU^T}_{UU^TU\Sigma V^T}A\Delta \leq 2\|{b^\perp}^T(S^TS - I)UU^T\|_2\|A\Delta\|_2$$

- AMM again: $\|b^\perp S^TSUU^T - b^\perp UU^T\|_F \leq \frac{\epsilon}{\sqrt{m}}\|b^\perp\|_F\|\underbrace{UU^T}_{I_{d\times d}}\|_F \leq \frac{\epsilon}{\sqrt{m}}f(x_{LS})\sqrt{d}$ $= \sqrt{\frac{d}{m}}$. Optimal value

$\|b^\perp\|_2 = f(x_{LS}) = \|Ax_{LS} - b\|_2 = \|b^\perp\|_2$

Basic Inequality Method

- $\|SA\Delta\|_2^2 \leq 2{b^{\perp}}^T(S^TS - I)A\Delta$
- Now consider the right handside

$$= 2{b^{\perp}}^T(S^TS - I)UU^TA\Delta \leq 2\|{b^{\perp}}^T(S^TS - I)UU^T\|_2\|A\Delta\|$$

- AMM again: $\|b^{\perp}S^TSUU^T - b^{\perp}UU^T\|_F \leq \frac{\epsilon}{\sqrt{m}}\|b^{\perp}\|_F\|UU^T\|_F \leq \frac{\epsilon}{\sqrt{m}}f(x_{LS})\sqrt{d}$

Basic Inequality Method

- $\|SA\Delta\|_2^2 \leq 2{b^{\perp}}^T(S^TS - I)A\Delta$
- Summarizing two bounds:
- (1) $\max_{\Delta}\left|\frac{\|SA\Delta\|_2^2}{\|A\Delta\|_2^2} - 1\right| \leq \epsilon'$
- (2) $2{b^{\perp}}^T(S^TS - I)A\Delta \leq \frac{\epsilon}{\sqrt{m}} f(x_{LS})\sqrt{d}\|A\Delta\|_2$

Basic Inequality Method

- $\|SA\Delta\|_2^2 \leq 2{b^{\perp}}^T(S^TS - I)A\Delta$
- Summarizing two bounds:
- (1) $\max_{\Delta}\left|\frac{\|SA\Delta\|_2^2}{\|A\Delta\|_2^2} - 1\right| \leq \epsilon'$
- (2) $2{b^{\perp}}^T(S^TS - I)A\Delta \leq \frac{\epsilon}{\sqrt{m}} f(x_{LS})\sqrt{d}\|A\Delta\|_2$

(1) implies $-\epsilon'\|A\Delta\|_2^2 \leq \|SA\Delta\|_2^2 - \|A\Delta\|_2^2 \leq \epsilon'\|A\Delta\|_2^2$

hence $(1 - \epsilon')\|A\Delta\|_2^2 \leq \|SA\Delta\|_2^2$

Basic Inequality Method

- $\|SA\Delta\|_2^2 \leq 2{b^{\perp}}^T(S^TS - I)A\Delta$
- Summarizing two bounds:
- (1) $\max_{\Delta} \left| \frac{\|SA\Delta\|_2^2}{\|A\Delta\|_2^2} - 1 \right| \leq \epsilon'$
- (2) $2{b^{\perp}}^T(S^TS - I)A\Delta \leq \frac{\epsilon}{\sqrt{m}} f(x_{LS})\sqrt{d}\|A\Delta\|_2$

  (1) implies $-\epsilon'\|A\Delta\|_2^2 \leq \|SA\Delta\|_2^2 - \|A\Delta\|_2^2 \leq \epsilon'\|A\Delta\|_2^2$

  hence $(1 - \epsilon')\|A\Delta\|_2^2 \leq \|SA\Delta\|_2^2$
- Plugging in: $(1 - \epsilon')\|A\Delta\|_2^2 \leq \frac{\epsilon}{\sqrt{m}} f(x_{LS})\sqrt{d}\|A\Delta\|_2$

  $\|A\Delta\|_2 \leq \frac{\epsilon}{1-\epsilon'} f(x_{LS})\frac{\sqrt{d}}{\sqrt{m}}$

$\|A\tilde{x} - Ax^*\|_2 \leq \sqrt{\frac{d}{m}} \cdot \text{Opt Value} \cdot \epsilon$ true for any AMM method

Leverage Scores

- Intuition: Approximate Matrix Multiplication for $U^T U$ i.e,
  $\|U^T S^T S U - U^T U\|_F = \|U^T S^T S U - I\|_F \leq \epsilon$
  implies Least Squares cost approximation

Leverage Scores

$U^T U \approx U^T S^T S U$

optimal distribution

$p_i = \frac{\|u_i\|_2^2}{\underbrace{\sum_j \|u_j\|_2^2}_{=d}}$

- Intuition: Approximate Matrix Multiplication for $U^T U$ i.e,
  $\|U^T S^T S U - U^T U\|_F = \|U^T S^T S U - I\|_F \leq \epsilon$
  implies Least Squares cost approximation

$A = U \Sigma V^T$

$U = \begin{bmatrix} u_1^T \\ | \\ u_n^T \end{bmatrix}$

- We can pick a sampling matrix S
- Importance sampling: proportional to the rows norms of U
- Leverage scores: $\ell_i := \|u_i\|_2^2$ for $i = 1, \ldots n$
- $\sum_i \ell_i = \sum_i \|u_i\|_2^2 = \|U\|_F^2 = trU^T U = trI_d = d$ when A is full column rank
- Sampling probabilities: $p_i = \frac{1}{d}\|u_i\|_2^2$
  $\sum_i p_i = 1$
- Can be non-uniform or uniform $A = [I; 0]$

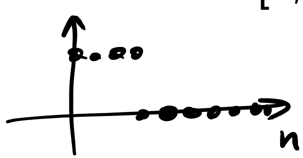
$A = \begin{bmatrix} \ddots \\ 0000 \\ 0000 \\ 0000 \\ \vdots \end{bmatrix} = \begin{bmatrix} I \\ 0 \end{bmatrix} \cdot I \cdot I = U \Sigma V^T$

$U^T U = I$

1) SVD : $O(nd^2)$ for l_i

2) Approx. leverage scores

3) Uniformize leverage scores.



Fast Johnson Lindenstrauss Transform

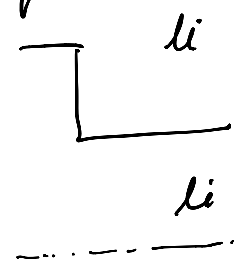
$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad 2\times 2$$

- Let H be the $n \times n$ Hadamard matrix
- Generate an $n \times n$ diagonal matrix of random ± 1 uniform signs : D
- Uniform $m \times n$ sub-sampling matrix P scaled with $\frac{\sqrt{n}}{\sqrt{m}}$ (Uniform sampling AMM matrix)
- Let $S = PHD$.
- Note that $\mathbb{E}S^T S = I$

$$H_4 = \begin{bmatrix} H_2 & H_2 \\ H_2 & -H_2 \end{bmatrix}$$

$$P = \begin{bmatrix} & & & 1 \\ & 1 & & \\ & & & 1 \\ 1 & & 1 & \end{bmatrix} \cdot \sqrt{\frac{n}{m}}$$

HD : leverage score uniformizer



$\mathbb{E}\, P^T P = I$

$$S \cdot A = \underbrace{PHD}_{S}A$$

$O(nd \log m)$ time to apply SA. based on FFT.

FJLT Preconditions Leverage Scores

- Fix a set X of n vectors in d-dimension. With high probability $\max_{x \in X} \|HDX\|_\infty \leq \sqrt{\frac{\log(n)}{d}}$

FJLT Preconditions Leverage Scores

- Fix a set X of n vectors in d-dimension. With high probability $\max_{x \in X} \|HDX\|_\infty \leq \sqrt{\frac{\log(n)}{d}}$
  Apply HD to data A
- $PHDA$ is uniformly sampled HDA
  Leverage scores of HDU are near uniform
  uniform sampling works!

Questions?